SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0001 $01.00/0

SEQUENCES OF MEROMORPHIC FUNCTIONS CORRESPONDING TO A FORMAL LAURENT SERIES*

WILLIAM B. JONES† AND W. J. THRON†

Abstract. A general theory is developed for sequences of functions $\{R_n(z)\}$ meromorphic at the origin which correspond to a formal Laurent series (fLs) L in the sense that the Laurent expansion of $R_n(z)$ agrees with L up to the ν_n power of z, where ν_n tends to infinity with n. Included are necessary and sufficient conditions for the existence of an fLs to which a given sequence corresponds. Also methods are described for obtaining sequences of meromorphic functions which correspond to a given fLs. As consequences of the property of correspondence it is shown that (under suitable restrictions) uniform convergence of a sequence is equivalent to uniform boundedness and that, when a sequence converges uniformly, its limit is a function whose Laurent expansion is L. Applications are considered for Padé approximants, continued fractions of various types and certain special functions.

1. Introduction. Following Henrici [5] we call

$$L = c_m z^m + c_{m+1} z^{m+1} + c_{m+2} z^{m+2} + \cdots, \qquad c_m \neq 0 \tag{1.1}$$

where the c_k, $k \geqq m$, are complex numbers, a *formal Laurent series* (fLs). $L = 0$ is also considered an fLs. The set $\mathscr{L}$ of all fLs forms a field with respect to addition and multiplication defined in the manner suggested by (1.1) (see, for example, [5, § 1.8]). If $f(z)$ is a function meromorphic at the origin (i.e., in an open disk containing the origin), then its Laurent expansion (convergent in a deleted neighborhood of the origin) will be denoted by $L(f)$. A sequence $\{R_n(z)\}$ of functions meromorphic at the origin will be said to *correspond to an fLs L* (*at* $z = 0$) if

$$\lim_{n\to\infty} \lambda(L - L(R_n)) = \infty, \tag{1.2}$$

where λ is the function defined as follows: $\lambda : \mathscr{L} \to \mathbb{R} \cup [\infty]$; if $L = 0$ then $\lambda(L) = \infty$; if $L \neq 0$ then $\lambda(L) = m$ where m is defined by (1.1).

Correspondence of sequences of meromorphic functions plays a key role in the theory of Padé approximants as well as in the problem of expanding functions, or fLs, in various types of continued fractions. In the past, various consequences of correspondence, as well as sufficient conditions for a sequence to correspond to an fLs, have been obtained case by case. We present here a general theory which, not only contains many known results as special cases, but from which new applications are derived. The formulation presented here clarifies the relationships involved in the method as well as the techniques for its application. In Theorem 1 (§ 2) we prove that a necessary and sufficient condition for a sequence $\{R_n(z)\}$ to correspond to some fLs is that

$$\lim_{n\to\infty} \lambda(L(R_{n+1}) - L(R_n)) = \infty.$$

In Theorems 2 and 3 (§ 2) it is shown that the approximants of a continued fraction formed from a system of three term recursion relations will correspond to a formal solution of the system under suitable restrictions. Moreover, these restrictions are shown to be invariant under equivalence transformations of the continued fraction. For sequences of functions $\{R_n(z)\}$ meromorphic at the origin and holomorphic in a deleted neighborhood of the origin, it is established in Theorem 4 that if $\{R_n(z)\}$

* Received by the editors November 11, 1976.
† Department of Mathematics, University of Colorado, Boulder, Colorado 80309. This research was supported by the National Science Foundation under Grant MCS 74-22111.

corresponds to an fLs L, then: 1) uniform convergence of $\{R_n(z)\}$ is equivalent to uniform boundedness, and 2) if $\{R_n(z)\}$ converges uniformly to a function $f(z)$ then $L = L(f)$. In many instances the second part of this result may be even more important than the first, for frequently the convergence can be deduced from other known criteria, while it may be otherwise difficult to determine what the limit is. Theorem 5 (§ 3) is an application of Theorem 4 explicitly for sequences of Padé approximants. It is an improvement of a result [8] previously given by the present authors. Applications of each result are discussed following the proofs of the theorems. Before proceeding to § 2, we summarize briefly a few facts and definitions that are used.

Every function $f(z)$ meromorphic at the origin has a unique fLs expansion $L(f)$. The one-to-one mapping L thus provides an embedding of the field $\mathcal{M}$ of all functions meromorphic at the origin in the field $\mathcal{L}$. If $\{R_n(z)\}$ corresponds to an fLs L, then the *order of correspondence* of $R_n(z)$ is defined to be

$$\nu_n = \lambda(L - L(R_n)).$$

It can be seen that if $\{R_n(z)\}$ corresponds to L, then L and $L(R_n)$ agree term-by-term up to and including the term involving $z^{\nu_n - 1}$.

We further extend the definition of correspondence as follows: A sequence of functions $\{R_n(z)\}$ meromorphic at $z = \infty$ (i.e., in an open neighborhood of $z = \infty$) will be said to *correspond to an fLs*

$$L = c_m w^m + c_{m+1} w^{m+1} + c_{m+2} w^{m+2} + \cdots, \qquad c_m \neq 0, \quad w = 1/z, \tag{1.3}$$

at $z = \infty$, if

$$\lim_{n \to \infty} \lambda\left(L - L\left(R_n\left(\frac{1}{w}\right)\right)\right) = \infty.$$

Similarly correspondence at $z = a$ $(a \in \mathbb{C})$ can be defined by considering $z = w + a$.

The following properties are easily deduced: For L_1 and L_2 in $\mathcal{L}$,

$$\lambda(L_1 L_2) = \lambda(L_1) + \lambda(L_2), \tag{1.4}$$

$$\lambda(L_1/L_2) = \lambda(L_1) - \lambda(L_2) \quad \text{if } L_2 \neq 0, \tag{1.5}$$

$$\lambda(L_1 \pm L_2) \geqq \min[\lambda(L_1), \lambda(L_2)], \tag{1.6}$$

$$\lambda(L_1 \pm L_2) = \min[\lambda(L_1), \lambda(L_2)] \quad \text{if } \lambda(L_1) \neq \lambda(L_2). \tag{1.7}$$

We note two observations due to Baker [1] in the case of Padé approximants: If $\{R_n(z)\}$ corresponds to L and if A, B, C, D are functions meromorphic at the origin such that $L(C) + L(D)L \neq 0$, $AD - BC \neq 0$ and $\lambda(L(C) + L(D)L(R_n)) \leq k$ for all n and some fixed k, then

$$\frac{A + BR_n}{C + DR_n} \text{ corresponds to } \frac{L(A) + L(B)L}{L(C) + L(D)L}. \tag{1.8}$$

Similarly

$$R_n\left(\frac{az}{cz + d}\right) \text{ corresponds to } L\left(\frac{az}{cz + d}\right) \quad \text{if } ad \neq 0. \tag{1.9}$$

By an (infinite) *continued fraction* is meant an ordered pair $\langle\langle\{a_n\}, \{b_n\}\rangle, \{f_n\}\rangle$, where $a_1, a_2, \cdots$ and $b_0, b_1, b_2, \cdots$ are complex numbers $(a_n \neq 0)$ and where $\{f_n\}$ is defined as follows:

$$f_n = S_n(0), \qquad n = 0, 1, 2, \cdots, \tag{1.10a}$$

where

$$S_0(w) = s_0(w) = b_0 + w, \tag{1.10b}$$

$$S_n(w) = S_{n-1}(s_n(w)), \qquad n = 1, 2, 3, \cdots, \tag{1.10c}$$

and

$$s_n(w) = \frac{a_n}{b_n + w}, \qquad n = 1, 2, 3, \cdots. \tag{1.10d}$$

For convenience we use the equivalent symbols

$$b_0 + \mathop{K}_{n=1}^{\infty} (a_n/B_n), \qquad b_0 + K(a_n/b_n) \tag{1.11}$$

and

$$b_0 + \frac{a_1}{b_1 +} \frac{a_2}{b_2 +} \frac{a_3}{b_3 +} \cdots \tag{1.12}$$

to denote the continued fraction $\langle\langle\{a_n\}, \{b_n\}\rangle, \{f_n\}\rangle$. The numbers a_n and b_n are called the *elements* of the continued fraction and f_n is called the nth *approximant.* A continued fraction $b_0 + K(a_n/b_n)$ is said to *converge* if its sequence of approximants $\{f_n\}$ converges (to a finite limit). When convergent, the *value* of the continued fraction is defined to be $\lim f_n$ and is sometimes denoted by (1.11) or (1.12). The nth *numerator* A_n and *denominator* B_n are defined by the second order linear *difference equations*

$$A_{-1} = 1, \quad A_0 = b_0, \quad B_{-1} = 0, \quad B_0 = 1, \tag{1.13a}$$

$$A_n = b_n A_{n-1} + a_n A_{n-2}, \qquad n = 1, 2, 3, \cdots, \tag{1.13b}$$

$$B_n = b_n B_{n-1} + a_n B_{n-2}, \qquad n = 1, 2, 3, \cdots. \tag{1.13c}$$

The following are well known [12], [18]:

$$S_n(w) = \frac{A_n + A_{n-1}w}{B_n + B_{n-1}w} = b_0 + \frac{a_1}{b_1 +} \cdots \frac{a_{n-1}}{+ b_{n-1} +} \frac{a_n}{b_n + w}, \qquad n = 0, 1, 2, \cdots, \tag{1.14}$$

$$f_n = S_n(0) = \frac{A_n}{B_n}, \qquad n = 0, 1, 2, \cdots, \tag{1.15}$$

$$A_n B_{n-1} - B_n A_{n-1} = (-1)^{n-1} \prod_{k=1}^{n} a_k \neq 0, \qquad n = 1, 2, 3, \cdots. \tag{1.16}$$

Equation (1.16) is sometimes called the *determinant formula.*

A continued fraction is said to *correspond to an fLs L* if its sequence of approximants corresponds to L.

An fLs (1.1) is called a *formal power series* (fps) if $m \geqq 0$. An open connected subset of the complex plane $\mathbb{C}$ is called a *domain.* We denote the closure of a bounded subset K of $\mathbb{C}$ by $\bar{K}$.

2. Correspondence. To motivate further the definition of correspondence we observe that, for $L \in \mathscr{L}$, the function β defined by

$$\beta(L) = \begin{cases} 0, & L = 0 \\ 2^{-\lambda(L)}, & L \neq 0 \end{cases} \tag{2.1}$$

is a valuation on $\mathscr{L}$ (for definition and properties of valuations on fields see, for example, [19]). We summarize a few properties of the valuation β that are employed in the proof of Theorem 1. Valuations are not used in the remainder of this paper. In terms of the valuation β one can define a metric ρ on $\mathscr{L}$ by

$$\rho(L_1, L_2) = \beta(L_1 - L_2), \quad \text{for } L_1, L_2 \in \mathscr{L}. \tag{2.2}$$

In terms of the metric ρ (which has been used previously in a similar context by Franzen [3]), the statement that a sequence $\{R_n(z)\}$ corresponds to L is equivalent to saying that $\{L(R_n)\}$ converges to L (with respect to the metric ρ). It is also the case that $\mathscr{L}$ is the completion of $L(\mathscr{M})$ with respect to ρ and hence $\mathscr{L}$ is ρ-complete. For $R_n(z) \in \mathscr{M}$, $\{L(R_n)\}$ is a Cauchy sequence (with respect to ρ) if for a given $\epsilon > 0$ there exists an n_ϵ such that

$$\rho(L(R_{n+k}), L(R_n)) < \epsilon, \quad \text{for } n \geqq n_\epsilon \text{ and } k \geqq 0. \tag{2.3}$$

Certain sequences $\{L(R_n)\}$ are Cauchy sequences (with respect to ρ) and, since $\mathscr{L}$ is ρ-complete, every Cauchy sequence converges (with respect to ρ) to some element $L \in \mathscr{L}$.

THEOREM 1. (A) *Given a sequence* $\{R_n(z)\}$ *of functions meromorphic at the origin, there exists an fLs L such that* $\{R_n(z)\}$ *corresponds to L if and only if*

$$\lim_{n\to\infty} \lambda(L(R_{n+1}) - L(R_n)) = \infty. \tag{2.4}$$

(B) *If* (2.4) *holds, then the L to which* $\{R_n(z)\}$ *corresponds is determined uniquely.* (C) *If the sequence* $\{\lambda(L(R_{n+1}) - L(R_n))\}$ *tends monotonically to* ∞, *then the order of correspondence of* $R_n(z)$ *is given by*

$$\nu_n = \lambda(L(R_{n+1}) - L(R_n)).$$

Proof. In view of the preceding discussion, to prove (A) it suffices to show that $\{L(R_n)\}$ is a Cauchy sequence (with respect to the metric ρ) if and only if (2.4) holds. Further we note that condition (2.3) is equivalent to

$$\lambda(L(R_{n+k}) - L(R_n)) > \operatorname{Log}_2\left(\frac{1}{\epsilon}\right), \quad \text{for } n \geqq n_\epsilon \text{ and } k \geqq 0. \tag{2.5}$$

By (1.6) we see that

$$\lambda\left(\sum_{j=1}^{k} L_j\right) \geqq \min_{1 \leq j \leq k} \lambda(L_j). \tag{2.6}$$

Thus

$$\begin{aligned}\lambda(L(R_{n+k}) - L(R_n)) &= \lambda\left(\sum_{j=1}^{k} (L(R_{n+j}) - L(R_{n+j-1}))\right) \\ &\geqq \min_{1 \leqq j \leqq k} \lambda(L(R_{n+j}) - L(R_{n+j-1})).\end{aligned} \tag{2.7}$$

It follows from (2.5) and (2.7) that $\{L(R_n)\}$ is a Cauchy sequence (with respect to ρ) if and only if given $N > 0$, there exists an n_N such that

$$\lambda(L(R_{n+1}) - L(R_n)) > N, \quad \text{for } n \geqq n_N.$$

Hence (2.4) holds if and only if $\{L(R_n)\}$ is a Cauchy sequence with respect to ρ. This proves (A). Part (B) follows immediately from properties of a complete metric space and (C) is a direct consequence of the definitions. This completes the proof.

Example 1 (T-fractions [15]). If $A_n(z)$ and $B_n(z)$ denote the nth numerator and denominator of a T-*fraction*

$$1+d_0z+\underset{n=1}{\overset{\infty}{K}}\left(\frac{z}{1+d_nz}\right),\qquad d_n\in\mathbb{C}, \tag{2.8}$$

then if follows from the difference equations (1.13) that $B_n(0)=1$, and from the determinant formula (1.16), that

$$\frac{A_{n+1}(z)}{B_{n+1}(z)}-\frac{A_n(z)}{B_n(z)}=\frac{(-1)^nz^{n+1}}{B_n(z)B_{n+1}(z)},\qquad n\geqq 0. \tag{2.9}$$

Therefore

$$\lambda\left(L\left(\frac{A_{n+1}}{B_{n+1}}\right)-L\left(\frac{A_n}{B_n}\right)\right)=n+1,\qquad n\geqq 0$$

and hence by Theorem 1 there exists an fLs L to which (2.8) corresponds. Since the order of correspondence of A_n/B_n is $\nu_n=n+1$, it follows that $L(A_n/B_n)$ agrees with L through the term involving z^n. Thus L is an fps of the form

$$L=1+c_1z+c_2z^2+c_3z^3+\cdots. \tag{2.10}$$

If it is further assumed that

$$d_n\neq 0,\quad \text{for } n\geqq 1, \tag{2.11}$$

then it follows from the difference equations (1.13) that $B_n(z)$ is a polynomial of exactly degree n with leading coefficient $d_1d_2\cdots d_n$. Letting $z=1/w$ and $R_n^*(w)=A_n(1/w)/B_n(1/w)$, we obtain from (2.9) that

$$\lambda(L(R_{n+1}^*)-L(R_n^*))=n,\qquad n\geqq 0.$$

Thus by Theorem 1 there exists another fLs L^* (in w) to which $\{R_n^*(w)\}$ corresponds, the order to correspondence of $R_n^*(w)$ being n [7], [12], [17]. Since $L(R_n^*)$ agrees with L^* up to and including the term involving w^{n-1}, it follows that L^* has the form

$$\begin{aligned}L^*&=\frac{d_0}{w}+c_0^*+c_1^*w+c_2^*w^2+\cdots\\&=d_0z+c_0^*+\frac{c_1^*}{z}+\frac{c_2^*}{z^2}+\cdots.\end{aligned} \tag{2.12}$$

It can be seen that the degrees of $A_n(z)$ and $B_n(z)$ do not exceed $n+1$ and n, respectively. Thus it follows from the results shown above that, when (2.11) holds, the nth approximant $A_n(z)/B_n(z)$ is the $(n+1,n)$ entry in the two-point Padé table of the L and L^* (see, for example, [14] for a brief discussion of two-point Padé tables). The relation between these tables and T-fractions is discussed more fully in [9].

Example 2 (P-fractions [11]). A continued fraction of the form

$$b_0(z)+\underset{n=1}{\overset{\infty}{K}}\left(\frac{1}{b_n(z)}\right), \tag{2.13a}$$

where each $b_n(z)$ is a polynomial in $1/z$,

(2.13b) $b_n(z)=\sum_{k=-N_n}^{0}a_{-k}^{(n)}z^k,\qquad N_0\geqq 0;\qquad N_n\geqq 1\ \text{ and }\ a_{-N_n}^{(n)}\neq 0,\quad n\geqq 1,$

is called a *P-fraction*. As in Example 1 it can be shown that, if $R_n(z)$ denotes the nth approximant of (2.13) then

$$\lambda(L(R_{n+1})-L(R_n))=2\sum_{k=1}^{n} N_k+N_{n+1}, \qquad n\geqq 0. \tag{2.14}$$

Hence by Theorem 1 there exists an fLs L to which (2.13) corresponds and L has the form

$$L=\sum_{k=-N_0}^{\infty} a_{-k}^{(0)}z^k. \tag{2.15}$$

We shall return to T-fractions and P-fractions in later applications. Before proceeding, however, we mention that similar statements can be made for continued fractions of the form

$$1+\mathop{K}_{n=1}^{\infty}\left(\frac{a_n z^{\alpha_n}}{1}\right), \qquad a_n\neq 0, \quad a_n\in\mathbb{C}, \quad \alpha_n\in[1,2,3,\cdots], \tag{2.16}$$

called *C-fractions* [10].

THEOREM 2. *Let* $\{a_n(z)\}$ *and* $\{b_n(z)\}$ *be sequences of functions meromorphic at the origin, with*

$$a_n(z)\not\equiv 0, \qquad n\geqq 1, \tag{2.17}$$

and let L_0 *be an fLs. Let* $\{L_n\}$ *be a sequence of fLs defined recursively as follows:*

$$L_{n+1}=\frac{L(a_{n+1})}{L_n-L(b_n)}, \qquad n\geqq 0, \tag{2.18a}$$

provided

$$L_n\neq L(b_n), \qquad n\geqq 0, \tag{2.18b}$$

(otherwise see (B)). *Then*:

(A) *If* $R_n(z)$ *denotes the n-th approximant of the continued fraction*

$$b_0(z)+\mathop{K}_{n=1}^{\infty}\left(\frac{a_n(z)}{b_n(z)}\right), \tag{2.19}$$

then $\{R_n(z)\}$ *corresponds to* L_0 *provided that*

$$\lambda(L(b_n))+\lambda(L(b_{n-1}))<\lambda(L(a_n)), \qquad n\geqq 1, \tag{2.20a}$$

and

$$\lambda(L_n)+\lambda(L(b_{n-1}))<\lambda(L(a_n)), \qquad n\geqq 1. \tag{2.20b}$$

If (2.20) *holds, then the order of corresondence of* $R_n(z)$ *is*

$$\nu_0=\lambda(L(a_1))-\lambda(L_1), \tag{2.21a}$$

$$\nu_n=\sum_{k=1}^{n}\lambda(L(a_k))-2\sum_{k=1}^{n-1}\lambda(L(b_k))-\lambda(L_n), \qquad n\geqq 1. \tag{2.21b}$$

(B) *If in defining* $\{L_n\}$ *by* (2.18a) *we obtain*

$$L_k\neq L(b_k) \quad \textit{for} \quad 0\leqq k\leqq m-1, \quad \textit{and} \quad L_m=L(b_m), \tag{2.22}$$

then

$$L_0=L\left(b_0(z)+\frac{a_1(z)}{b_1(z)+}\cdots\frac{a_m(z)}{+b_m(z)}\right). \tag{2.33}$$

Remark. Before proving Theorem 2, we point out that *conditions* (2.20) *are invariant under equivalence transformations of the continued fraction* (2.19) in the following sense: Let $\{r_n(z)\}$ be an arbitrary sequence of nonvanishing functions meromorphic at the origin and define

(2.24a) $\quad a_n^*(z)=r_n(z)r_{n-1}(z)a_n(z), \qquad n\geqq 1 \quad (r_0(z)\equiv 1)$

(2.24b) $\quad b_n^*(z)=r_n(z)b_n(z), \qquad n\geqq 0$

(2.24c) $\quad L_n^*=L(r_n)L_n, \qquad n\geqq 0.$

Then in view of (2.17) we have $a_n^*(z)\not\equiv 0$. Moreover,

(2.25) $$L_n^*\neq L(b_n^*), \qquad n\geqq 0,$$

if and only if (2.18b) holds;

(2.26) $$L_{n+1}^*=\frac{L(a_{n+1}^*)}{L_n^*-L(b_n^*)}, \qquad n\geqq 0,$$

provided (2.25) holds; and

(2.27a) $$\lambda(L(b_n^*))+\lambda(L(b_{n-1}^*))<\lambda(L(a_n^*)), \qquad n\geqq 1,$$

and

(2.27b) $$\lambda(L_n^*)+\lambda(L(b_{n-1}^*))<\lambda(L(a_n^*)), \qquad n\geqq 1,$$

hold if and only if (2.20a) and (2.20b) hold, respectively. In consequence, the continued fraction

(2.28) $$b_0^*(z)+\mathop{K}_{n=1}^{\infty}\left(\frac{a_n^*(z)}{b_n^*(z)}\right)$$

is equivalent to (2.19), and hence (2.28) corresponds to L_0^* provided the conditions (2.27) and (2.25) hold. Thus it is unnecessary to search for an equivalence transformation of a continued fraction for the purpose of making Theorem 2 applicable.

Proof of Theorem 2. (A) Suppose that (2.18) holds and let $A_n(z)$ and $B_n(z)$ denote the nth numerator and denominator of (2.19). From (2.18a) we have

(2.29) $$L_n=L(b_n)+\frac{L(a_{n+1})}{L_{n+1}}, \qquad n\geqq 1.$$

Here division by L_{n+1} is possible since $L_{n+1}\neq 0$ follows from (2.17) and (2.18a). Thus

$$L_0=L(b_0)+\frac{L(a_1)}{L(b_1)+}\cdots\frac{L(a_{n-1})}{+L(b_{n-1})+}\frac{L(a_n)}{L_n}, \qquad n\geqq 1,$$

and hence by (1.14)

(2.30) $$L_0=\frac{L(a_n)L(A_{n-2})+L_nL(A_{n-1})}{L(a_n)L(B_{n-2})+L_nL(B_{n-1})}, \qquad n\geqq 2.$$

Then applying (2.30), the determinant formula (1.16) and $L(R_{n-1})=L(A_{n-1})/L(B_{n-1})$, we obtain

$$L_0-L(R_{n-1})=\frac{(-1)^{n-1}\prod_{k=1}^{n}L(a_k)}{L(B_{n-1})(L(a_n)L(B_{n-2})+L_nL(B_{n-1}))}, \qquad n\geqq 2$$

and hence from (1.4) and (1.5)

$$\lambda(L_0 - L(R_{n-1})) = \sum_{k=1}^{n} \lambda(L(a_k)) - \lambda(L(B_{n-1})) \tag{2.31}$$
$$- \lambda(L(a_n)L(B_{n-2}) + L_n L(B_{n-1})), \qquad n \geqq 2.$$

A simple induction argument based on (1.7), (1.13), and (2.20) can be used to establish the following formulas:

$$\lambda(L(B_0)) = 0, \tag{2.32a}$$

$$\lambda(L(B_n)) = \sum_{k=1}^{n} \lambda(L(b_k)), \qquad n \geqq 1. \tag{2.32b}$$

Then by use of (1.7), (2.20), and (2.32) we can prove that

$$\lambda(L(a_n)L(B_{n-2}) + L_n L(B_{n-1})) = \lambda(L_n) + \sum_{k=1}^{n-1} \lambda(L(b_k)), \qquad n \geqq 2. \tag{2.33}$$

Now substituting (2.32) and (2.33) into (2.31) gives

$$\lambda(L_0 - L(R_{n-1})) = \sum_{k=1}^{n} \lambda(L(a_k)) - 2 \sum_{k=1}^{n-1} \lambda(L(b_k)) - \lambda(L_n), \qquad n \geqq 2. \tag{2.34}$$

Rearranging the terms in (2.34), one obtains

$$\lambda(L_0 - L(R_{n-1})) = (\lambda(L(a_1) - \lambda(L(b_1)))$$
$$+ \sum_{k=2}^{n-1} (\lambda(L(a_k)) - \lambda(L(b_k)) - \lambda(L(b_{k-1}))) \tag{2.35}$$
$$+ (\lambda(L(a_n)) - \lambda(L(b_{n-1})) - \lambda(L_n)), \qquad n \geqq 2.$$

It follows from (2.20) that each term in the sum in (2.35) and the last term are positive integers and hence

$$\lim_{n\to\infty} \lambda(L_0 - L(R_n)) = \infty,$$

so that $\{R_n(z)\}$ corresponds to L_0. Equation (2.21b) follows from (2.34); (2.21a) is an immediate consequence of the definitions. This proves (A).

(B) It follows from (2.18a) and (22) that

$$L_0 = L(b_0) + \frac{L(a_1)}{L(b_1)+} \cdots \frac{L(a_{m-1})}{+L(b_{m-1})+} \frac{L(a_m)}{L_m}$$

from which (2.23) follows since, by (2.22), $L_m = L(b_m)$. This completes the proof.

The following examples illustrate a method by which Theorem 2 can be applied to show that, for a given fLs L, there exists a continued fraction which corresponds to L.

Example 3 (T-fractions, continued). Let

$$L_0 = 1 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots$$

be a given fps and let

$$a_n(z) = z, \quad \text{for } n \geqq 1,$$
$$b_n(z) = 1 + d_n z, \quad \text{for } n \geqq 0.$$

We now show (by means of Theorem 2) that the d_n can be chosen so that the resulting T-fraction (2.8) corresponds to L_0. In accordance with (2.18) we set

$$L_1 = \frac{z}{L_0 - (1 + d_0 z)} = \frac{1}{(c_1 - d_0) + c_2 z + c_3 z^2 + \cdots}.$$

By choosing $d_0 = c_1 - 1$, we obtain L_1 in the form

$$L_1 = 1 + c_1^{(1)} z + c_2^{(1)} z^2 + \cdots.$$

The remaining d_n can be defined successively in a similar manner, and the resulting L_n are defined by (2.18a). Therefore, since

$$\lambda(z) = 1, \quad \lambda(1 + d_n z) = 0, \quad \lambda(L_n) = 0, \quad \text{for } n \geqq 1,$$

it follows from Theorem 2 that (2.8) corresponds to L_0.

Example 4. In a manner completely analogous to that used in Example 3, it can be shown that, for a given fLs

$$L^* = z + c_0^* + \frac{c_1^*}{z} + \frac{c_2^*}{z^2} + \cdots,$$

there exists a continued fraction of the form

$$e_0 + z + \mathop{K}_{n=1}^{\infty} \left(\frac{z}{e_n + z} \right), \qquad e_n \in \mathbb{C},$$

corresponding to L^* at $z = \infty$.

Example 5 (P-fractions, continued). Let

$$L_0 = \sum_{k=-N_0}^{\infty} a_{-k}^{(0)} z^k, \qquad N_0 \geqq 0$$

be a given fLs. We define $b_0(z)$ by

$$b_0(z) = \sum_{k=-N_0}^{0} a_{-k}^{(0)} z^k,$$

and, assuming that $L_0 \neq b_0(z)$, we define

$$L_1 = \frac{1}{L_0 - b_0(z)}.$$

Let $a_{-N_1}^{(0)} z^{N_1}$ denote the first nonzero term in $L_0 - b_0(z)$, so that $N_1 \geq 1$ and $a_{-N_1}^{(0)} \neq 0$. Thus L_1 can be expressed in the form

$$L_1 = \sum_{k=-N_1}^{\infty} a_{-k}^{(1)} z^k, \qquad a_{-N_1}^{(1)} = 1/a_{-N_1}^{(0)} \neq 0.$$

Next we define

$$b_1(z) = \sum_{k=-N_1}^{0} a_{-k}^{(1)} z^k.$$

Continuing in this manner, we obtain (via (2.18a)) a sequence of fLs $\{L_n\}$, provided $L_n \neq b_n(z)$ for all $n \geqq 0$. In that case, since

$$\lambda(1) = 0, \quad \lambda(b_n(z)) = -N_n, \quad \lambda(L_n) = -N_n, \quad n \geqq 1$$

where $N_n \geqq 1$ for $n \geqq 1$, it follows from Theorem 2 that the resulting P-fraction (2.13) corresponds to L_0. In the case that there exists an m satisfying (2.22), then (2.23) holds. Similar statements can be made for C-fractions (2.16).

The following result (a simple consequence of Theorem 2) is particularly useful for sequences of nonzero fLs satisfying certain systems of three-term recurrence relations.

THEOREM 3. *Let $\{a_n(z)\}$ and $\{b_n(z)\}$ be sequences of functions meromorphic at the origin with*

$$a_n(z) \not\equiv 0, \qquad \text{for } n \geqq 1. \tag{2.36}$$

Let $\{P_n\}$ be a sequence of nonzero fLs satisfying the three-term recurrence relations

$$P_n = L(b_n)P_{n+1} + L(a_{n+1})P_{n+2}, \qquad n \geqq 0. \tag{2.37}$$

Then the continued fraction

$$b_0(z) + \underset{n=1}{\overset{\infty}{K}} \left(\frac{a_n(z)}{b_n(z)}\right) \tag{2.38}$$

corresponds to the fLs $L = P_0/P_1$ provided the following conditions are satisfied:

$$\lambda(L(b_n)) + \lambda(L(b_{n-1})) < \lambda(L(a_n)), \qquad n \geqq 1, \tag{2.39a}$$

$$\lambda(P_n/P_{n+1}) + \lambda(L(b_{n-1})) < \lambda(L(a_n)), \qquad n \geqq 1. \tag{2.39b}$$

Proof. Letting $L_n = P_n/P_{n+1}$, for $n \geqq 0$, we obtain from (2.37) that

$$L_n - L(b_n) = \frac{L(a_{n+1})}{L_{n+1}}, \qquad n \geqq 0. \tag{2.40}$$

But (2.36) implies that $L(a_{n+1}) \neq 0$ for $n \geqq 0$ and hence (from (2.40)) $L_n \neq L(b_n)$, $n \geqq 0$. Thus $\{L_n\}$ satisfies all of the conditions of Theorem 2(A) and our assertion follows.

Example 6. (Hypergeometric functions and the continued fraction of Gauss). The *hypergeometric function* $F(a, b, c; z)$ is defined by the power series

$$F(a, b, c; z) = 1 + \frac{ab}{c}\frac{z}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots, \tag{2.41}$$

where a, b, c are complex constants, $c \notin [0, -1, -2, -3, \cdots]$. If a or b is in the set $[0, -1, -2, -3, \cdots]$, then $F(a, b, c; z)$ is a polynomial. Otherwise the power series in (2.41) has radius of convergence equal to one. We define

$$\begin{aligned}
P_{2n} &= F(a+n, b+n, c+2n; z), && n \geqq 0,\\
P_{2n+1} &= F(a+n, b+n+1, c+2n+1; z), && n \geqq 0,\\
a_{2n}(z) &= \frac{(a+n)(c-b+n)}{(c+2n-1)(c+2n)}z, && n \geqq 1,\\
a_{2n+1}(z) &= \frac{(a+n)(c-b+n)}{(c+2n)(c+2n+1)}z, && n \geqq 0.
\end{aligned}$$

Then it can be shown [2] that

$$P_n = P_{n+1} + a_{n+1}(z)P_{n+2}, \quad \text{for } n \geqq 0.$$

If a, b, c are chosen so that $a_n(z) \not\equiv 0$ for $n \geqq 1$, then

$$\lambda(a_n(z)) = 1, \quad \lambda(1) = 0, \quad \lambda(P_n/P_{n+1}) = 0, \quad n \geqq 1,$$

and hence by Theorem 3, the continued fraction

$$1+\underset{n=1}{\overset{\infty}{K}}\left(\frac{a_n(z)}{1}\right)$$

corresponds to $P_0/P_1 = F(a, b, c; z)/F(a, b+1, c+1; z)$. By means of (1.8) (with $A = D = 1$, $C = B = 0$) it follows that the *continued fraction of Gauss*

$$\frac{1}{1+K_{n=1}^{\infty}\ (a_n(z)/1)}$$

corresponds to P_1/P_0.

Numerous continued fraction expansions of functions have been obtained from Example 6. In a similar manner many more can be obtained for the confluent hypergeometric functions [18]. It is not our purpose here to elaborate on the applications of Theorem 3 but merely to illustrate the method. However, we will treat one further application, the Legendre functions of the second kind. Although Gautschi [4] has proved convergence of the continued fraction by an application of Pincherle's theorem, the following proof of correspondence is new.

Example 7. Legendre functions of the second kind

(2.42a) $$Q^m_{a+n}(z) = K_n z^{-m-a-n-1} F(\tfrac{1}{2}n+\tfrac{1}{2}a+\tfrac{1}{2}m+1, \tfrac{1}{2}n+\tfrac{1}{2}a+\tfrac{1}{2}m+\tfrac{1}{2}; n+a+\tfrac{3}{2}; z^{-2}),$$

where

(2.42b) $$K_n = e^{mi\pi}2^{-n-a-1}\pi^{1/2}\frac{\Gamma(a+n+m+1)}{\Gamma(a+n+3/2)}$$

are used, for example, to solve Laplace's equation by spherical harmonics [2]. Here m and n are nonnegative integers and a an arbitrary complex number. F denotes the hypergeometric function (2.41). We shall consider m and a as fixed, let $z = 1/w$ and then define $P_n(w)$ by

$$P_n(w) = Q^m_{a+n}(z).$$

Then the $P_n(w)$ satisfy the system of three-term recurrence relations [2]

(2.43) $$P_n(w) = \frac{2n+2a+3}{n+a+m+1}\cdot\frac{1}{w}P_{n+1}(w) - \frac{n+a-m+2}{n+a+m+1}P_{n+2}(w), \qquad n \geqq 0.$$

Letting $b_n(w)$ denote the coefficient of P_{n+1} and $a_{n+1}(w)$ the coefficient of P_{n+2} in (2.43), we obtain

$$\lambda(a_n) = 0, \quad \lambda(b_n) = -1, \quad \lambda(P_n/P_{n+1}) = -1, \quad n \geqq 0.$$

Therefore it follows from Theorem 3 that

$$b_0(w)+\underset{k=1}{\overset{\infty}{K}}\left(\frac{a_k(w)}{b_k(w)}\right),$$

corresponds to P_0/P_1 at $w = 0$; or, more generally,

(2.44) $$b_n(w)+\underset{k=1}{\overset{\infty}{K}}\left(\frac{a_{k+n}(w)}{b_{k+n}(w)}\right)$$

corresponds (at $w = 0$) to P_n/P_{n+1}, $n = 0, 1, 2, \cdots$. Thus the continued fraction (2.52) with w replaced by $1/z$ corresponds (at $z = \infty$) to $Q^m_{a+n}(z)/Q^m_{a+n+1}(z)$. The convergence of (2.44) will be dealt with in the following section.

3. Uniform convergence. A sequence $\{R_n(z)\}$ of functions meromorphic in a domain D is said to *converge uniformly on a compact subset* K *of* D if and only if:

(i) there exists $N(K)$ such that $R_n(z)$ is holomorphic in some domain containing K for all $n \geqq N(K)$, and

(ii) given $\epsilon > 0$ there exists $N_\epsilon > N(K)$ such that

$$\sup_{z\in K} |R_{n+k}(z) - R_n(z)| < \epsilon \quad \text{for } n \geqq N_\epsilon, \quad k \geqq 0. \tag{3.1}$$

The sequence $\{R_n(z)$ is said to be *uniformly bounded on a compact subset* K *of* D if and only if there exist $M(K)$ and $B(K)$ such that

$$\sup_{z\in K} |R_n(z)| \leqq B(K) \quad \text{for } n \geqq M(\mathrm{K}) \tag{3.2}$$

In Theorem 4 we shall show that (subject to certain restrictions) a sequence $\{R_n(z)\}$, which corresponds to an fLs L, will be uniformly convergent if and only if it is uniformly bounded. First, however, it should be pointed out that a sequence $\{R_n(z)\}$ may converge uniformly to a holomorphic function $f(z)$ in a neighborhood of the origin without necessarily corresponding to $L(f)$, the Taylor series expansion of $f(z)$ at $z = 0$. For example, consider

$$L(e^z) = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

and

$$R_n(z) = 1 + \sum_{k=1}^{\infty} \left(\frac{1}{k!} + \frac{1}{n^2}\right) z^k, \qquad n = 1, 2, 3, \cdots.$$

It is easily seen that $\{R_n(z)\}$ converges uniformly to e^z on $|z| \leqq \rho < 1$, but $\{R_n(z)\}$ does not correspond to $L(e^z)$.

THEOREM 4. *Let* $\{R_n(z)\}$ *be a sequence of functions meromorphic at the origin and corresponding to an fLs*

$$L = c_m z^m + c_{m+1} z^{m+1} + \cdots, \qquad c_m \neq 0. \tag{3.3}$$

Further suppose that there exists a deleted neighborhood of the origin $D^* = [z : 0 < |z| < \delta]$ *such that each* $R_n(z)$ *is holomorphic in* D^*. *Let* D *be a domain containing* D^*; *if* $m < 0$ *we require that* $0 \notin D$. *Then*:

(A) $\{R_n(z)\}$ *converges uniformly on every compact subset of* D *if and only if* $\{R_n(z)\}$ *is uniformly bounded on every compact subset of* D.

(B) *If* $\{R_n(z)\}$ *converges uniformly on every compact subset of* D, *then* $f(z) = \lim_{n\to\infty} R_n(z)$ *is holomorphic in* D *and* $L = L(f)$.

Proof. (A) That uniform convergence implies uniform boundedness follows from a standard argument which need not be repeated here. Now suppose that $\{R_n(z)\}$ is uniformly bounded on all compact subsets of D. Let K be an arbitrary compact subset of D. Let K_0 be an open, connected, bounded subset of D such that $K \subset K_0 \subset \bar{K}_0 \subset D$ and such that K_0 contains an annulus $\eta < |z| < 5\eta < \delta$. Since K_0 is assumed to be bounded, $\bar{K}_0$ is a compact subset of D. There then exists an n_0 and an M, both depending on $\bar{K}_0$, such that

$$\sup_{z\in \bar{K}_0} |R_n(z)| < M \quad \text{for } n \geqq n_0. \tag{3.4}$$

Since $R_n(z)$ is meromorphic at the origin and holomorphic in D^*, it can be represented by its convergent Laurent series

$$L(R_n) = \sum_{k=m_n}^{\infty} \gamma_k^{(n)} z^k, \qquad \eta < |z| < 5\eta < \delta, \quad \gamma_{m_n}^{(n)} \neq 0, \tag{3.5}$$

where

$$\gamma_k^{(n)} = \frac{1}{2\pi i}\int_c \frac{R_n(\zeta)}{\zeta^{k+1}}\,d\zeta, \qquad k = m_n, m_n+1, \cdots, \tag{3.6}$$

and where c is the circle $|\zeta| = 4\eta$ traversed once in the counterclockwise direction. It follows from (3.4) and (3.6) that

$$|\gamma_k^{(n)}| \leqq \frac{M}{(4\eta)^k}, \qquad k \geqq m_n, \quad n \geqq n_0. \tag{3.7}$$

We note in passing that the assumption that each $R_n(z)$ is holomorphic in D^* was made to insure that $L(R_n)$ would be the Laurent expansion of $R_n(z)$ in the annulus $\eta < |z| < 5\eta$. From (3.5) we see that

$$\lambda(L(R_n)) = m_n, \qquad n \geqq 0. \tag{3.8}$$

Further,

$$\begin{aligned}\lambda(L(R_{m+n}) - L(R_n)) &= \lambda(L(R_{m+n}) - L + L - L(R_n)) \\ &\geqq \min[\lambda(L(R_{m+n}) - L), \lambda(L - L(R_n))], \qquad n \geqq 0, \quad m \geqq 0.\end{aligned} \tag{3.9}$$

Therefore, since $\lambda(L - L(R_n)) \to \infty$ as $n \to \infty$, given any N, we can find an $n_1 > n_0$ such that

$$\lambda(L(R_{m+n}) - L(R_n)) \geqq N, \quad \text{for } n \geqq n_1, \quad m \geqq 0. \tag{3.10}$$

It follows from (3.5), (3.7), and (3.10) that, for $n \geqq n_1$ and $m \geqq 0$,

$$\begin{aligned}\sup_{\eta<|z|<2\eta} |R_{m+n}(z) - R_n(z)| &\leqq \sup_{\eta<|z|<2\eta} \sum_{k=N}^{\infty} |(\gamma_k^{(m+n)} - \gamma_k^{(n)})z^k| \\ &\leqq \sum_{k=N}^{\infty} \frac{2M}{(4\eta)^k}(2\eta)^k = \frac{4M}{2^N}.\end{aligned} \tag{3.11}$$

Thus we see that $\{R_n(z)\}$ is a uniform Cauchy sequence on $\eta < |z| < 2\eta$. An application of the Stieltjes–Vitali theorem (see, for example, [6, p. 251] or [16, p. 142]) completes the proof that $\{R_n(z)\}$ converges uniformly on $\bar{K}_0$ and hence on K.

(B) Suppose now that $\{R_n(z)\}$ is uniformly convergent on all compact subsets of D. Define

$$L_n = c_m z^m + c_{m+1} z^{m+1} + \cdots + c_{m+n} z^{m+n}, \qquad n \geqq 0. \tag{3.12}$$

Then

$$\begin{aligned}\lambda(L_n - L(R_n)) &= \lambda((L_n - L) + (L - L(R_n))) \\ &\geqq \min[m+n+1, \lambda(L - L(R_n))].\end{aligned}$$

Hence

$$\lim_{n\to\infty} \lambda(L_n - L(R_n)) = \infty \tag{3.13}$$

and, for every $k \geqq m$, there exists l_k such that

$$c_k = \gamma_k^{(l)}, \quad \text{for } l \geqq l_k. \tag{3.14}$$

Now let K^* be an arbitrary compact subset of D^*. Then there exist ϵ and η such that

$$0 < \epsilon < |z| < \eta < \delta, \quad \text{for } z \in K^*.$$

The set $K_1 = [z : \epsilon \leqq |z| \leqq \eta]$ is a compact subset of D^*. Let M_1 belong to K_1; that is,

$$\sup_{z \in K_1} |R_n(z)| \leqq M_1 \quad \text{for } n \geqq M(K_1). \tag{3.15}$$

(We note that the existence of bound M_1 follows from part (A) already proved.) Now the coefficients $\gamma_k^{(n)}$ in (3.5) can be written as

$$\gamma_k^{(n)} = \frac{1}{2\pi i} \int_{c_1} \frac{R_n(\zeta)}{\zeta^{k+1}} \, d\zeta, \qquad k = m_n, m_n+1, \cdots, \tag{3.16}$$

where c_1 is the circle $|z| = \rho$, $\epsilon < \rho < \eta$, traversed one time in the counterclockwise direction, and such that

$$\mu = \max_{z \in K^*} \left|\frac{z}{\rho}\right| < 1. \tag{3.17}$$

It follows from (3.14), (3.15) and (3.16) that

$$\begin{aligned} |c_k| &= |\gamma_k^{(l)}|, && \text{for } l \geqq l_k, \\ &= \frac{1}{2\pi} \left| \int_{c_1} \frac{R_l(\zeta)}{\zeta^{k+1}} \, d\zeta \right| \leqq \frac{M_1}{\rho^k}, && l \geqq l_k. \end{aligned} \tag{3.18}$$

Thus we have, for all $z \in K^*$,

$$\begin{aligned} |L_n(z)| &\leqq \sum_{k=m}^{m+n} |c_k z^k| \\ &\leqq \sum_{k=m}^{m+n} |\gamma_k^{(l)} z^k|, \qquad l \geqq l_k, \\ &\leqq M_1 \sum_{k=m}^{m+n} \left|\frac{z}{\rho}\right|^k. \end{aligned} \tag{3.19}$$

If $m < 0$, then by (3.17) and (3.19),

$$|L_n(z)| \leqq M_1 \left[\sum_{k=m}^{-1} \mu^k + \frac{1}{1-\mu} \right], \quad \text{for } z \in K^*, \tag{3.20a}$$

and if $m = 0$, then the

$$|L_n(z)| \leqq \frac{M_1}{1-\mu}, \quad \text{for } z \in K^*. \tag{3.20b}$$

Thus we have shown that the sequence $\{L_n(z)\}$ is uniformly bounded on every compact subset of D^*. Since the $L_n(z)$ are rational functions, holomorphic in D^*, and since $\lambda(L - L_n) = m + n + 1 \to \infty$ as $n \to \infty$, by part (A) of the theorem, it follows that $\{L_n(z)\}$ converges uniformly on all compact subsets of D^* to a function $f(z)$ which is holomorphic in D^*. Clearly $L = L(f)$. Now

$$\begin{aligned} |L_n(z) - R_n(z)| &\leqq \sum_{k=\tau_n}^{\infty} |(c_k - \gamma_k^{(n)}) z^k|, \qquad \tau_n = \lambda(L_n - L(R_n)), \\ &\leqq \sum_{k=\tau_n}^{\infty} 2M_1 \left|\frac{z}{\rho}\right|^k, \qquad \text{by (3.18)}, \\ &\leqq \frac{2M_1 \mu^{\tau_n}}{1-\mu}, \quad \text{for } z \in K^*, \qquad \text{by (3.17)}. \end{aligned} \tag{3.21}$$

Since $0<\mu<1$ and $\tau_n\to\infty$ as $n\to\infty$ (by (3.13)), it follows from (3.21) that $\{L_n(z)-R_n(z)\}$ converges uniformly to 0 on K^*. But

$$|f(z)-R_n(z)|\leqq|f(z)-L_n(z)|+|L_n(z)-R_n(z)|,$$

and hence we can conclude that $\{R_n(z)\}$ converges uniformly to $f(z)$ on all compact subsets of D^*. The extension to D can be obtained by analytic continuation. This completes the proof.

The following is an immediate corollary of Theorem 4.

COROLLARY 4.1. *Let $\{R_n(z)\}$ and $\{Q_n(z)\}$ be two sequences of functions meromorphic at the origin which correspond to the same fLs L. If both sequences converge uniformly on every compact subset of a domain D containing a deleted neighborhood of the origin, say D^*, and if $\lambda(L)<0$ implies $0\notin D$ and if all $R_n(z)$ and $Q_n(z)$ are holomorphic in D^*, then both sequences converge to the same function $f(z)$ which is holomorphic in D and for which $L(f)=L$.*

For $\lambda(L)=m\geqq 0$, the statement of Theorem 4 becomes sufficiently simpler that it is worth stating separately.

THEOREM 4′. *Let $\{R_n(z)\}$ be a sequence of functions meromorphic at the origin which corresponds to a formal power series*

$$P=c_0+c_1z+c_2z^2+\cdots.$$

Let D be a domain containing a neighborhood of the origin. Then:

(A) *$\{R_n(z)\}$ converges uniformly on every compact subset of D if and only if $\{R_n(z)\}$ is uniformly bounded on every compact subset of D.*

(B) *If $\{R_n(z)\}$ converges uniformly on every compact subset of D, then $f(z)=\lim_{n\to\infty}R_n(z)$ is holomorphic in D and P is the Taylor series expansion of $f(z)$ about $z=0$.*

Part (A) of Theorem 4′ was proved for regular and associated continued fractions by Pringsheim [13]. For C-fractions (2.16), (A), (B) and Corollary 4.1 were established by Leighton and Scott [10]; for T-fractions (2.8) the same three results can be found in [15]. The corresponding results for P-fractions (2.13) have not been previously known. As a further application of Theorem 4, we consider again the Legendre functions of the second kind (2.42).

Example 8. In Example 7 it was shown that the continued fraction (2.44) corresponds to P_n/P_{n+1} at $w=0$. If we set

$$a_n^*=r_nr_{n-1}a_n,\qquad n\geqq 1,$$
$$b_n^*=r_nb_n=1,\qquad n\geqq 0$$

where the a_n and b_n are defined as in Example 7, and where

$$r_n=\frac{n+a+m+1}{2n+2a+3}w,\qquad n\geqq 1\quad (r_0=1)$$

then the continued fraction

$$b_n^*(w)+\mathop{K}_{k=1}^{\infty}\left(\frac{a_{k+n}^*(w)}{1}\right)\tag{3.22}$$

is equivalent to (2.44). But

$$\lim_{n\to\infty}a_n^*(w)=-\frac{w^2}{4}$$

and hence it follows from Worpitzky's theorem [18, p. 42] that for sufficiently large n, (3.22) converges uniformly on compact subsets of $|w|<1$. Thus it follows from Theorem 4′ that (3.22) converges to P_n/P_{n+1}, at least for $|w|<1$ and sufficiently large n. By replacing w by $1/z$, we see that (3.22) converges to $Q_{a+n}^{(m)}(z)/Q_{a+n+1}^{m}(z)$ in $|z|>1$, n sufficiently large.

For Padé approximants the present authors [8] gave a result which, using Theorem 4′ and a tighter estimate of $\lambda(P-L(R_\nu))$, we can now improve somewhat. Let P be a formal power series and let $R_{m,n}(z)$ denote the (m, n) Padé approximant to P, so that $R_{m,n}(z)=A_{m,n}(z)/B_{m,n}(z)$, where $A_{m,n}$ and $B_{m,n}$ are polynomials of degrees at most m and n, respectively, and

$$\lambda(PB_{m,n}-A_{m,n})\geqq m+n+1.$$

It then follows that

$$\lambda(P-L(A_{m,n}/B_{m,n}))\geqq m+n+1-\min[m, n]=1+\max[m, n].$$

In the transition from $PB_{m,n}-A_{m,n}$ to $P-L(A_{m,n}/B_{m,n})$ the value of λ may decrease by a value r which is such that z^r is the highest power of z contained as a factor in $B_{m,n}$. Hence $r\leqq n$. If z^r is a factor of $B_{m,n}$ then it must also be a factor of $A_{m,n}$, so that $r\leqq m$.

Our theorem now becomes

THEOREM 5. *Let $\{m_\nu\}$ and $\{n_\nu\}$ be sequences of nonnegative integers such that*

$$\lim_{\nu\to\infty}\max[m_\nu, n_\nu]=\infty.$$

Let $R_\nu(z)$ denote the (m_ν, n_ν) Padé approximant of the formal power series

$$P=c_0+c_1z+c_2z^2+\cdots.$$

Let D be a domain containing a neighborhood of the origin. Then:

(A) *$\{R_\nu(z)\}$ converges uniformly on every compact subset of D if and only if $\{R_\nu(z)\}$ is uniformly bounded on every compact subset of D.*

(B) *If $\{R_\nu(z)\}$ converges uniformly on every compact subset of D then $f(z)=\lim_{\nu\to\infty}R_\nu(z)$ is holomorphic in D and P is the Taylor series expansion of $f(z)$ about $z=0$.*

The following example throws some light on what can happen if some of our conditions (in Theorems 4, 4′, and 5) are not met.

Example 9. Let

$$R_n(z)=\frac{1}{1-(nz)^n},\qquad n\geqq 1.$$

Then

$$L(R_n)=1+(nz)^n+(nz)^{2n}+\cdots,$$

so that $\{R_n(z)\}$ corresponds to $L=1$. Each $R_n(z)$ has n poles on the circle $|z|=1/n$, but $\{R_n(z)\}$ is uniformly bounded on every compact subset of $0<|z|$. Finally, $R_n(z)\}$ converges to 0 for $0<|z|$.

REFERENCES

[1] GEORGE A. BAKER, JR., *Essentials of Padé Approximants*, Academic Press, New York, 1975.
[2] A. ERDÉLYI, ET. AL., *Higher Transcendental Functions*, vol. I, McGraw-Hill, New York, 1953.

[3] N. R. FRANZEN, *Some convergence results for Padé approximants*, J. Approximation Theory, 6 (1972), pp. 254–263.

[4] W. GAUTSCHI, *Computational aspects of three-term recurrence relations*, SIAM Rev., 9 (1967), pp. 24–82.

[5] P. HENRICI, *Applied and Computational Complex Analysis*, vol. 1, John Wiley, New York, 1974.

[6] E. HILLE, *Analytic Function Theory*, vol. II, Ginn, New York, 1962.

[7] THOMAS H. JEFFERSON, JR., *Some additional properties of T-fractions*, Ph.D. thesis, Univ. of Colorado, Boulder, CO, 1969.

[8] WILLIAM B. JONES AND W. J. THRON, *On convergence of Padé approximants*, This Journal, 6 (1975), pp. 9–16.

[9] ———, *Two-point Padé tables and T-fractions*, Bull. Amer. Math Soc., 83 (1977), pp. 388–390.

[10] W. LEIGHTON AND W. T. SCOTT, *A general continued fraction expansion*, Ibid., 45 (1939), pp. 596–605.

[11] A. MAGNUS, *Certain continued fractions associated with the Padé table*, Math. Z., 78 (1962), pp. 361–374.

[12] O. PERRON, *Die Lehre von den Kettenbrüchen*, Band II, B. G. Teubner, Stuttgart, 1957.

[13] A. PRINGSHEIM, *Über Konvergenz und funktionentheoretischen Charakter gewisser limitärperiodischer Kettenbrüche*, Sitzungsberichte, der (kgl.) Bayerischen Akademie der Wissenschaffen zu München, 1910.

[14] P. SHENG AND J. D. DOW, *Intermediate coupling theory: Padé approximants for Polarona*, Phys. Rev. B, 4 (1971), pp. 1343–1359.

[15] W. J. THRON, *Some properties of continued fraction* $1+d_0z+K(z/(1+d_nz))$, Bull. Amer. Math. Soc., 54, (1948), pp. 206–218.

[16] ———, *The Theory of Functions of a Complex Variable*, John Wiley, New York, 1953.

[17] HAAKON WAADELAND, *On T-fractions of certain functions with a first order pole at the point of infinity*, Det Kongelige Norske Videnskabers, Selskabs Forhandlinger, Bind 40 (1967), Nr. 1, pp. 1–6.

[18] H. S. WALL, *Analytic Theory of Continued Fractions*, Van Nostrand, New York, 1948.

[19] B. L. VAN DER WAERDEN, *Algebra*, vol. 2., translation of 5th ed., Frederick Ungar, New York, 1970.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0002 $01.00/0

NECESSARY AND SUFFICIENT CONDITION FOR MAINTAINING OSCILLATIONS AND NONOSCILLATIONS IN GENERAL FUNCTIONAL EQUATIONS AND THEIR ASYMPTOTIC PROPERTIES*

BHAGAT SINGH†

Abstract. A necessary and sufficient condition is found for the nonoscillation of

$$(r(t)y'(t))^{(n-1)}+F(h(y(g(t))), t)=0, \qquad n \geqq 2.$$

Case study for the asymptotic oscillatory behavior of solutions of the equations

$$(r(t)y'(t))'+a(t)h(y(g(t)))=f(t)$$

and

$$(r(t)y'(t))'+p(t)y(t)+a(t)h(y(g(t)))=f(t)$$

is made for the two cases when $\int^{\infty} 1/r(t)\,dt=\infty$ and

$$\int^{\infty} 1/r(t)\,dt<\infty, \qquad r(t)>0.$$

1. Introduction. In [12], this author found conditions on $a(t)$, $r(t)$ and $f(t)$ to ensure that all nonoscillatory solutions of the equation

$$(1) \qquad (r(t)y'(t))'+a(t)y(t-\tau(t))=f(t)$$

approach finite limits asymptotically. A similar set of conditions was found in [13] by this author to force all oscillatory solutions of a slightly more general equation

$$(2) \qquad (r(t)y'(t))'+a(t)y^{\alpha}(t-\tau(t))=f(t)$$

to approach zero.

So far, for the most part, the main thrust of results in oscillation theory for equations of type (1) and (2) is in the direction of $\int^{\infty} 1/r(t)\,dt=\infty$. Indeed many interesting applications of these equations such as variable mass problems result when $\int^{\infty} 1/r(t)\,dt=\infty$. For this case the results are numerous and the interested reader is referred to the works of T. Burton and R. Grimmer [1], Hammett [5], Kusano and Onose [6], this author [12], [14]–[17], Staikos and Sficas [18] and Tuefel [20]. The list is by no means complete. The literature is very scanty on results when $\int^{\infty} 1/r(t)\,dt<\infty$. In our works on [12] and [13] we obtained some asymptotic results for the latter and observed that more need to be said about the case $\int^{\infty} 1/r(t)\,dt=\infty$. The latter half of this paper is devoted to the study of the equation

$$(3) \qquad (r(t)y'(t))'+p(t)y'(t)+a(t)h(y(g(t)))=f(t)$$

for which we obtain conditions so that all solutions of (3) are nonoscillatory; and another set of conditions that allows all oscillatory solutions of (3) to approach zero as $t\to\infty$, one set requiring $\int^{\infty} 1/r(t)\,dt=\infty$.

With regard to the solutions of these equations we would like to study only those solutions which can be continuously extended on some positive half real line, say for $t>t_0$ where $t_0>0$. In § 2, our first theorem, therefore is to show that any nontrivial

---

* Received by the editors February 18, 1977.

† Department of Mathematics, University of Wisconsin Center, Manitowoc, Wisconsin 54220.

solution of

$$(r(t)y'(t))^{(n-1)}+F(h(y(g(t))),t)=0, \qquad n \geqq 2 \tag{4}$$

can be continued indefinitely to the right of t_0. The generality of equation (4) and (in it) of function F allows easy extensions to cover equations (1), (2) and (3). In the process this theorem generalizes a similar theorem of Graef and Spikes [3], [4]. In § 3 we prove a necessary and sufficiency type theorem which essentially gives a strong criterion for the nonoscillation of equation (4) subject to one of the conditions as $\int^\infty 1/r(t)\,dt<\infty$. Throughout this whole work ample examples demonstrate the applicability of results. Incidently, we call any equation (under study here) oscillatory when all of its infinitely continuable solutions are oscillatory. Otherwise it is called nonoscillatory. Again a function $\theta(t)\in C[t_0,\infty)$ is said to be oscillatory if $\theta(t)$ has arbitrarily large zeros; otherwise $\theta(t)$ is called nonoscillatory.

As a result of our Theorem 1 in § 2, the term "solution" will be used in this work only to refer to continuously extendable solutions of equations under study.

In § 4, we find conditions, one being $\int^\infty 1/r(t)\,dt<\infty$, that force all solutions of equation

$$(r(t)y'(t))'+a(t)h(y(g(t)))=f(t) \tag{5}$$

to be nonoscillatory.

We would like to remark in passing that the methods discovered to deal with ordinary differential equations usually do not carry over to similar equations containing a delay term. See Travis [19] this author [12].

For readers interested in practical applications of similar equations we suggest Norkin [10] who gives specific equations that arise naturally in perturbed combustion phenomena inside rocket engines.

The entire study in this work is subject to the following assumptions:

(i) $a(t)$, $r(t)$, $p(t)$, $f(t)$, $h(t)$ and $g(t)$ are continuous on R, the real line;

(ii) $r(t)>0$, $g(t)>0$, $g(t)\leqq t$, $0<g'(t)\leqq S$ for some S, $g(t)\to\infty$ as $t\to\infty$;

(iii) $\lim_{t\to\infty}\sup\,(t-g(t))=\infty$;

(iv) h is odd, sign $h(t)=$ sign t, $0<h(t)/t\leqq m_0$ for some m_0 on R;

(v) $F: R\times R\to R$ continuous, odd and increasing in the first argument; sign $F(z,t)=$ sign z for all $t>0$.

2. Continuability.

THEOREM 1. *In addition to previous conditions suppose further that*

$$0<F(z,t)/z<m_0' \qquad \text{on } R. \tag{6}$$

Let $y(t)$ be a solution of equation (4) *such that $y(g(t))\in C^{(n)}(-\infty,T_0)$, $T_0>0$. Then $y(t)$ can be continuously extended to all of R.*

Proof. Suppose to the contrary that $y(g(t))$ cannot be continued past T_0, i.e.,

$$\lim_{t\to T_0-}\sup|y(g(t))|=\infty. \tag{7}$$

Now let $P_0<g(a_0)$ where $a_0=g(T_0)-b$, $b>0$. Integrating equation (4) over $[a_0,t]$, $t<T_0$ we have

$$y'(t)=\frac{1}{r(t)}\left[K_1+K_2t+\cdots+K_{n-1}t^{n-2}-\int_{a_0}^{t}\frac{(t-s)^{n-2}}{(n-2)!}F(h(y(g(s))),s)\,ds\right] \tag{8}$$

where K_i, $i=1,\cdots,n-1$ are appropriately chosen constants.

Integrating (8) between P_0 and $g(t)$ we get

$$y(g(t)) = K_0 + K_1 \int_{P_0}^{g(t)} \frac{1}{r(x)}\,dx + K_2 \int_{P_0}^{g(t)} \frac{x}{r(x)}\,dx + \cdots + K_{n-1} \int_{P_0}^{g(t)} \frac{x^{n-2}}{r(x)}\,dx \tag{9}$$
$$- \int_{P_0}^{g(t)} \frac{1}{r(s)} \int_{a_0}^{s} \frac{(s-x)^{n-2}}{(n-2)!} F(h(y(g(x))), x)\,dx\,ds$$

where $K_0 = y(g(P_0))$.

(9) yields in view of (6)

$$|y(g(t))| \leqq K_0 + c_1 t + c_2 t^2 + \cdots + c_{n-1} t^{n-1} \tag{10}$$
$$+ \int_{P_0}^{t} \frac{(t-s)^{n-1}}{(n-1)!} \frac{F(h, x)}{h} \cdot \frac{h(y(g(s)))}{y(g(s))} \cdot |y(g(s))|\,ds$$

where by continuity

$$K_i \int_{P_0}^{g(t)} \frac{x^{i-1}}{r(x)}\,dx \leqq c_i t^i, \qquad i = 1, 2, \cdots, n-1 \quad \text{for } t \in [P_0, T_0].$$

Dividing (10) by t^{n-1} and using (6) and the condition on h we have

$$\frac{|y(g(t))|}{t^{n-1}} \leqq M + \int_{P_0}^{t} \frac{(t-s)^{n-1}}{t^{n-1}} \cdot \frac{1}{(n-1)!} m_0 m_0' \cdot s^{n-1} \frac{|y(g(s))|}{s^{n-1}}\,ds \tag{11}$$
$$\leqq M + L \int_{P_0}^{t} s^{n-1} \cdot \frac{|y(g(s))|}{s^{n-1}}\,ds$$

where $L = m_0 m_0'/(n-1)!$ and $M > (K_0/(t^{n-1}) + c_1/(t^{n-2}) + \cdots + c_{n-1})$ for $t \in [P_0, T_0]$. It should be noted that $y(g(p_0))$ is defined. By Gronwall's inequality, there exists a positive L_0 such that for $t \in [P_0, T_0]$

$$|y(g(t))| \leqq L_0$$

contradicting (7). The proof is now complete.

Remark. From here on we shall use the term "solution" only for continuously extendable solutions of equations on some positive half real line.

3. Necessary and sufficient condition.

THEOREM 2. *Suppose* $\int_t^\infty 1/r(t)\,dt = \infty$. *Let*

$$Q(t, T) = \int_T^t \frac{1}{r(s)} \int_T^s \frac{(s-x)^{n-3}}{(n-3)!}\,dx\,ds. \tag{A}$$

Then a necessary and sufficient condition for equation (4) *to have a nonoscillatory solution asymptotic to* $dQ(t, T)$, $d \neq 0$ *is*

$$\int_T^\infty F(cQ(g(t), T), t)\,dt < \infty \quad \textit{for some } c > 0. \tag{B}$$

Proof. (*Necessity*). Let $y(t)$ be a nonoscillatory solution of (4) with the property that

$$\lim_{t\to\infty} (y(t)/Q(t, T)) = d \neq 0. \tag{12}$$

Without any loss of generality we can assume that there exists $t_0 > 0$ such that $y(g(t))$, $y(t)$ and d are positive for $t \geqq t_0$. Let $T > g(t_0)$. From equation (4) on repeated integration we have

$$
\begin{aligned}
y(t)=y(T)+K_1\int_T^t\frac{1}{r(s)}\,ds+K_2\int_T^t\frac{(s-T)}{r(s)}\,ds+K_3\int_T^t\frac{1}{r(s)}\int_T^s(s-x)\,dx\,ds\\
+\cdots+K_{n-1}\int_T^t\frac{1}{r(s)}\int_T^s\frac{(s-x)^{n-3}}{(n-3)!}\,dx\,ds\\
-\int_T^t\frac{1}{r(s)}\int_T^s\frac{(s-x)^{n-2}}{(n-2)!}F(h(y(g(x))),x)\,dx\,ds,
\end{aligned}
\tag{13}
$$

where $K_i=(r(T)y'(T))^{(i-1)}$, $i=1,2,\cdots,n-1$.

Define

$$\psi_T(y(g(t)))\equiv\int_T^t\frac{1}{r(s)}\int_T^s\frac{(s-x)^{n-2}}{(n-2)!}F(h(y(g(x))),x)\,dx\,ds,$$

$$\psi_T(CR(g(t)))\equiv\int_T^t\frac{1}{r(s)}\int_T^s\frac{(s-x)^{n-2}}{(n-2)!}F(CR(g(x)),x)\,dx\,ds.$$

Dividing (13) by $Q(t, T)$ and taking limit we have

$$d=K_{n-1}-\lim_{t\to\infty}\frac{\psi_T(y(g(t)))}{Q(t,T)}. \tag{14}$$

Now let $Q(g(t), T)\equiv Q(g(p))$ then

$$\lim_{t\to\infty}\frac{\psi_T(y(g(t)))}{Q(t,T)}\leqq\lim_{t\to\infty}\frac{\psi_T(CQ(g(t)))}{Q(t,T)}$$

where $C=dm_0$ due to the condition on h and increasing (in first argument) nature of $F(z, t)$. Now by l'Hôpital's rule

$$\lim_{t\to\infty}\frac{\psi_T(CQ(g(t))}{Q(t,T)}=\int_T^\infty F(CQ(g(t)),t)\,dt. \tag{15}$$

The conclusion about necessity now follows from (15).

(*Sufficiency*). Without any loss of generality let T above be large enough so that for $t>T'>T$ we have

$$\psi_{T'}(CQ(g(t))/Q<\int_T^\infty F(CQ(g(t)),t)\,dt<d/2. \tag{16}$$

Since

$$\lim_{t\to\infty}\frac{\psi_T(y(g(t)))}{Q(t,T)}=\int_T^\infty F(h(y(g(t))),t)\,dt$$

and is finite, let $T_2>T$ be so large that for $t\geqq T_2$

$$\frac{\psi_{T_2}(y(g(t)))}{Q(t,T)}<\int_T^\infty F(h(y(g(t))),t)\,dt. \tag{17}$$

In fact, for $T_2>T$

$$\lim_{t\to\infty}\frac{\psi_{T_2}(y(g(t)))}{Q(t,T)}=\int_{T_2}^\infty F(h(y(g(t))),t)\,dt<\int_T^\infty F(h(y(g(t))),t)\,dt,$$

which justifies (17).

Consider the class S of all continuous functions $f(t)$ in $[T_2, \infty)$ such that $|f(t)|/Q(t, T)$ is bounded. We define the norm $\|\cdot\|$ as

$$\|f(t)\| = \sup\{(Q(t, T))^{-1}|f(t)|, t \geqq T_2\} < \infty.$$

It is easily verified that S with the norm $\|\cdot\|$ is a Banach space. Let $W \subset S$ with the property that

$$d/2 \leqq x/Q(t) \leqq d \tag{18}$$

for $x \in S$. We observe that W is a closed, bounded and convex subset of S. Now we define an operator ϕ on S as

$$\phi(y(t)) = dQ(t, T) - \psi_{T_2}(y(g(t))). \tag{19}$$

We shall then seek a fixed point of the operator ϕ via Schauder fixed point theorem. We, first, notice that $\phi(W) \subseteq W$. In fact, if $y(t) \in W$ then

$$\phi(y(t)) = dQ(t, T) - dQ(t, T)(\psi_{T_2}(y(g(t)))/dQ(t, T)). \tag{20}$$

Now $y(g(t)) \leqq dQ(g(t), T)$. Therefore from (20) we have

$$\phi(y(t)) \geqq dQ(t, T) - dQ(t, T)\frac{\psi_{T_2}(CQ(g(t))}{dQ(t, T)}. \tag{21}$$

This gives $\phi(y(t))/Q(t, T) \geqq d/2$ by (16). Also $\phi(y)/Q(t, T) \leqq d$. Hence $\phi(W) \subseteq W$. Next we shall show that ϕ is continuous. Let $y_n(t) \to y(t)$ in norm as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \|y_n(t) - y(t)\| = 0. \tag{22}$$

Now

$$\begin{aligned}\left|\frac{\phi(y_n(t))}{Q(t, T)} - \frac{\phi(y(t))}{Q(t, T)}\right| &\leqq \left[\int_{T_2}^t \frac{1}{r(s)} \int_{T_2}^s \frac{(s-x)^{n-2}}{(n-2)!}|F(h(y_n(g(x))), x) \right.\\ &\qquad \left. - F(h(y(g(x))), x)|\, dx\, ds\right] \Big/ Q(t, T) \\ &\leqq \int_T^\infty |F(h(y_n(g(x))), x) - F(h(y(g(x))), x)|\, dx\end{aligned}$$

by (17). Thus

$$\|\phi(y_n) - \phi(y)\| \leqq \int_T^\infty G_n(t)\, dt \tag{23}$$

where

$$G_n(t) = |F(h(y_n(g(t))), t) - F(h(y(g(t))), t)| \leqq 2F(CQ(g(t), t)).$$

By the Lebesgue dominated convergence theorem, the right hand side of (23) approaches zero and the continuity of ϕ is established.

Next we shall show that $\overline{\phi(W)}$ is compact. To this end it suffices to show that the family $D = \{\phi(y)/Q(t, T): y \in W\}$ is uniformly bounded and equicontinuous. Uniform boundedness is obvious. To show equicontinuity we follow Levitan [8] according to whom it will be achieved if we could subdivide $[T_2, \infty)$ into finite number of subintervals on each of which all functions of family D have oscillations approaching zero.

Let then $t_2 > t_1 > T_2$. Now

$$\left\|\left[\frac{\phi(y(t_2))}{Q(t_2, T)} - \frac{\phi(y(t_1))}{Q(t_1, T)}\right]\right\| = \left|\frac{\psi_{T_2}(y(g(t_2)))}{Q(t_2, T)} - \frac{\psi_{T_2}(y(g(t_1)))}{Q(t_1, T)}\right|$$

$$\to 0 \quad \text{as } t_1 \text{ (and hence } t_2) \to \infty \text{ uniformly.}$$

Hence for a given $\varepsilon > 0$, there exists a number $T_3 > T_2$ such that for all pairs $t_2 > t_1 > T_3$

$$\left|\frac{\phi(y(t_2))}{Q(t_2, T)} - \frac{\phi(y(t_1))}{Q(t_1, T)}\right| \leqq \varepsilon.$$

Now consider the closed interval $[T_2, T_3]$. It is easy to see that the family D has uniformly bounded derivative and hence is equicontinuous in $[T_2, T_3]$. Thus we have succeeded in subdividing $[T_2, \infty)$ into finitely many subintervals on each of which oscillations of members of D die out. Hence $\overline{\phi(W)}$ is compact.

Applying Schauder's fixed point theorem, ϕ has a fixed point in W. Suppose $\phi(y_0(t)) = y_0(t)$, then it follows from (19) that

(24) $$y_0(t) = dQ(t, T) - \psi_{T_2}(y(g(t)));$$

i.e.,

(25) $$\begin{aligned} y(t) = d\left(\int_T^t \frac{1}{r(s)} \int_T^s \frac{(s-x)^{n-3}}{(n-3)!} \, dx \, ds\right) \\ - \int_{T_2}^t \frac{1}{r(s)} \int_{T_2}^s \frac{(s-x)^{n-2}}{(n-2)!} F(h(y(g(x))), x) \, dx \, ds. \end{aligned}$$

Simple differentiation shows that $y_0(t)$ is a solution of equation (4) which is nonoscillatory and asymptotic to $Q(t, T)$. In fact, since $y_0(t) > 0$, it follows from equation (4) that $y_0(t)$ is monotonic.

The proof of Theorem 2 is now complete.

THEOREM 3. *A necessary condition for equation* (4) *to be oscillatory is*

(26) $$\int^\infty F(CQ_0(g(t), t)) \, dt = \infty \qquad \text{for some } C > 0,$$

where we define

$$Q_0(t) = \int^t \frac{1}{r(s)} \int^s \frac{(s-x)^{n-3}}{(n-3)!} \, dx \, ds.$$

Proof. This follows from Theorem 2.

4. Further remarks on nonoscillation. In this section we take up equation (5), namely

(5) $$(r(t)y'(t))' + a(t)h(y(g(t))) = f(t).$$

Theorem 2, in the case of equation (4), guarantees the existence of a nonoscillatory solution. Here we shall prove a theorem by which all solutions of equation (5) become nonoscillatory. We shall need the following lemma which is Theorem 2 of this author [13, p. 40].

LEMMA 1. *Suppose*

$$\int^{\infty} |f(t)|\,dt<\infty, \tag{27}$$

$$\int^{\infty} |a(t)|\,dt<\infty, \tag{28}$$

and

$$\int^{\infty} \frac{1}{r(t)}\,dt<\infty. \tag{29}$$

Then all oscillatory solutions of equation (5) *approach zero asymptotically.*

Proof. The presence of h and g in equation (5) here requires trivial modifications in the proof in [13].

THEOREM 4. *In addition to conditions of Lemma* 1 *suppose*

$$\liminf_{t\to\infty} \int^{t} [f(t)-|a(t)|]\,dt>0; \tag{30}$$

then all solutions of equation (5) *are nonoscillatory.*

Proof. Suppose to the contrary that $y(t)$ is an oscillatory solution of equation (5). By Lemma 1, $y(t)\to 0$ as $t\to\infty$. Let P be large enough so that for $t\geqq P$, $|h(y(g(t)))|<1$. Since $y(t)$ is oscillatory (ry') is also oscillatory. Let $P_0>P$ be a zero of $(r(t)y'(t))$. From equation (5), on integration, we have

$$(r(t)y'(t))+\int_{P_0}^{t} ah(y(g(x)))\,dx=\int_{P_0}^{t} f(x)\,dx,$$

or

$$(r(t)y'(t))+\int_{P_0}^{t} |a(x)|\,|h(y(g(x)))|\,dx\geqq\int_{P_0}^{t} f(x)\,dx. \tag{31}$$

Since $|h(y(g(t)))|<1$, (31) yields

$$(r(t)y'(t))\geqq\int_{P_0}^{t} (f(x)-|a(x)|)\,dx>0. \tag{32}$$

But (32) implies that $y(t)$ is nonoscillatory. The proof is now complete by contradiction.

Example 1. Consider the equation

$$(e^{t}y'(t))'+e^{-2t}y(t)=2e^{-t}+e^{-4t}. \tag{33}$$

It is easily verified that all conditions of Theorem 4 are satisfied. Hence all solutions of (33) are nonoscillatory. $y(t)=e^{-2t}$ is one nonoscillatory solution of (33).

Example 2. Consider the equation

$$(e^{t}y'(t))'+e^{-t}\sin t\,y(\sqrt{t})=2e^{-t}, \qquad t>0. \tag{34}$$

Since here again the coefficients meet the conditions of Theorem 4, all solutions of (34) are nonoscillatory.

Our next theorem gives sufficient conditions for all solutions of equation (5) to approach nonzero limits on the extended real line.

THEOREM 5. *Suppose for* $a(t)>0$, *all conditions of Theorem* 4 *hold. Further suppose that there exists a continuously differentiable function* $\lambda(t)$ *such that*

$$\lambda(t)\to 0 \quad \textit{as}\ t\to\infty, \tag{35}$$

$$(r(t)\lambda'(t))' = f(t). \tag{36}$$

Let $y(t)$ *be a solution of* (5). *Then* $\lim |y(t)| = \beta \leqq \infty$.

Proof. By Theorem 4 $y(t)$ is nonoscillatory. All that we need to show is that $\lim_{t\to\infty} |y(t)|$ exists on the extended real line.

From equation (5)

$$(r(t)(y(t)-\lambda(t))')' + a(t)h(y(g(t))) = 0. \tag{37}$$

Without any loss of generality suppose P_3 is large enough so that for $t \geqq P_3$ both $y(t)$ and $y(g(t))$ are positive and condition (36) holds. From (37), $(r(t)(y(t)-\lambda(t))')' < 0$ and hence $(y(t)-\lambda(t))$ is monotonic. Thus

$$\lim_{t\to\infty} (y(t)-\lambda(t)) \quad \text{exists on extended } R.$$

Now

$$\liminf_{t\to\infty} y(t) = \liminf_{t\to\infty} [(y(t)-\lambda(t))+\lambda(t)]$$
$$= \lim_{t\to\infty} (y(t)-\lambda(t)).$$

Similarly

$$\limsup_{t\to\infty} y(t) = \lim_{t\to\infty} (y(t)-\lambda(t)),$$

and the proof is complete.

Example 3. Consider the equation

$$(e^t y'(t))' + e^{-t} y(t) = 2e^{-t} + e^{-3t}. \tag{38}$$

It has $y(t) = e^{-2t}$ as a solution. For function $\lambda(t)$ we take

$$\lambda(t) = e^{-2t} + \tfrac{1}{12} e^{-4t}.$$

Thus all conditions of Theorem 5 are satisfied. All solutions of (38) approach limits on extended real line.

5. Asymptotic nonoscillation.

LEMMA 2. *Suppose* $r(t) > 0$, $f(t) > 0$, $a(t) > 0$;

$$\int^{\infty} \frac{1}{f(t)}\, dt < \infty; \tag{39}$$

$$\int^{\infty} \frac{a(t)f(t)}{r(t)}\, dt < \infty; \tag{40}$$

$$\left(\frac{f(t)r'(t)}{r(t)} - f'(t) + \frac{f(t)p(t)}{r}\right)' \geqq 0 \tag{41}$$

for $t \geqq A$, *for some* $A > t_0 > 0$. *Let* $y(t)$ *be an oscillatory solution of equation* (3); *then* $y(t)$ *is bounded above on* R^+.

Proof. Let $y^+(t) = \max(y(t), 0)$. Let $T > A > t_0$ to be large enough so that

$$m \int_T^{\infty} \frac{a(t)f(t)}{r(t)}\, dt < 1 \tag{42}$$

and

$$\int_T^\infty \frac{1}{f(t)}\,dt<1. \tag{43}$$

Let $t_2>t_1>T$ be two consecutive zeros of $y(t)$ such that $y(t)>0$ in (t_1, t_2). Let $M_0=\max y(t)$, $t_0\in[t_1, t_2]$ be such that $M_0=y(t_0)$. Now $M_0=\int_{t_1}^{t_0} y'(t)\,dt$ also $M_0=-\int_{t_0}^{t_2} y'(t)\,dt$.

Together these give

$$2M_0\leqq\int_{t_1}^{t_2}|y'(t)|\,dt. \tag{44}$$

Now equation (3) can be written as

$$(f(t)y'(t))'+\left(\frac{fr'}{r}-f'+\frac{fp}{r}\right)y'+\frac{afh(y(g(t)))}{r}=\frac{f^2}{r}; \tag{45}$$

we shall now closely follow the proof of Theorem 1 of this author [13, p. 38–40]. The proof requires some changes due to the presence of the term H_2y' where

$$H_2=(fr'/r-f'+fp/r). \tag{46}$$

From (44) we obtain

$$2M_0\leqq\int_t^{t_2}[f(t)]^{-1/2}[f(t)|y'(t)|]^{1/2}|y'(t)|^{1/2}\,dt \tag{47}$$

which yields as in [13, p. 39, conclusion 15]

$$4M_0^2\leqq\left[\int_{t_1}^{t_2}\frac{1}{f(t)}\,dt\right]\left[-\int_{t_1}^{t_2}y(t)(f(t)y'(t))'\,dt\right]. \tag{48}$$

From (45) and (48) we get

$$\begin{aligned}4M_0^2\left[\int_{t_1}^{t_2}\frac{1}{f(t)}\,dt\right]^{-1}\leqq\int_{t_1}^{t_2}y(t)H_2(t)y'(t)\,dt+\int_{t_1}^{t_2}\frac{a(t)f(t)y(t)h(y(g(t)))}{r(t)}\,dt\\ -\int_{t_1}^{t_2}\frac{y(t)f^2(t)}{r(t)}\,dt.\end{aligned} \tag{49}$$

The rightmost term in (49) is nonnegative; further integration of the first term on the right of (49) yields

$$4M_0^2\left[\int_{t_1}^{t_2}\frac{1}{f(t)}\,dt\right]^{-1}\leqq-\frac{1}{2}\int_{t_1}^{t_2}H_2'(t)y^2(t)\,dt+\int_{t_1}^{t_2}\frac{a(t)f(t)y(t)h(y(g(t)))}{r(t)}\,dt.$$

Hence

$$4M_0^2\left[\int_{t_1}^{t_2}\frac{1}{f(t)}\,dt\right]^{-1}\leqq\int_{t_1}^{t_2}\frac{a(t)f(t)y(t)h(y(g(t)))}{r(t)}\,dt, \tag{50}$$

Since $H_2'\geqq0$ by condition (41). Setting

$$H_3(t)=(a(t)f(t))/r(t), \tag{51}$$

we obtain from (50)

$$4M_0^2\left[\int_{t_1}^{t_2}\frac{1}{f(t)}\,dt\right]^{-1}\leqq\int_{t_1}^{t_2}H_3(t)h(y(g(t)))y(t)\,dt. \tag{52}$$

Let $q > p$ be large enough consecutive zeros of $y(t)$ such that $p - g(q) > T$. Let $t \in (p, q)$. Suppose $M_q = \max|y(t)|$ for $t \in (p, q)$. If $y(t)$ is not bounded above then $\limsup_{t\to\infty} y(t) = \infty$. Let $r_0 > q$ be the smallest number such that

$$|y(r_0)| = M_q + 1. \tag{53}$$

Let T_1 be the greatest zero of $y(t)$ less than r_0 and T_2 be the smallest zero of $y(t)$ greater than r_0. Then

$$q \leqq T_1 < r_0 < T_2 \tag{54}$$

and T_1, T_2 are consecutive zeros of $y(t)$. Let $M_2 = \max|y(t)|$, $t \in [T, T_2]$ and $M_2 = y(t_{M_2})$, $t_{M_2} \in [T, T_2]$. In a manner of this author [13, p. 40] it now follows that

$$M_2 = \max|y(t)|, \qquad t \in [T_1, T_2]. \tag{55}$$

In inequality (52) we replace t_1 and t_2 by T_1 and T_2 respectively. It is clear that for $t \in [T_1, T_2]$, $g(t) \in [T, q]$. Hence

$$|y(g(t))| \leqq M_2, \qquad t \in [T_1, T_2]. \tag{56}$$

Let now $y(t) > 0$ in (T_1, T_2). From (52) and (56) and the fact that $0 < h(x)/x \leqq m$, we have

$$4M_2^2\left[\int_{T_1}^{T_2} \frac{1}{f(t)}\,dt\right]^{-1} \leqq \int_{T_1}^{T_2} \frac{y(t)|y(g(t))|H_3(t)h(y(g(t)))}{y(g(t))}\,dt,$$

which gives

$$4 \leqq \left[\int_{T_1}^{T_2} \frac{1}{f(t)}\,dt\right]\left[m\int_{T_1}^{T_2} H_3(t)\,dt\right]. \tag{57}$$

Now (57) gives a contradiction in view of (42) and (43). This shows that $y(t) < 0$ in (T_1, T_2). Thus $y^+(t)$ doesn't exceed a finite bound. The proof is now complete.

THEOREM 6. *Suppose conditions of Lemma 2 hold. Further suppose that* $p(t) > 0$, $p'(t) \leqq 0$, $\int^\infty a(t)\,dt < \infty$, *and* $\int^\infty f(t)\,dt = \infty$. *Then all solutions of* (3) *are nonoscillatory.*

Proof. Suppose to the contrary that $y(t)$ is an oscillatory solution of equation (3). By Lemma 2, $y^+(t)$ is bounded. Now $y'(t)$ must be oscillatory. Let $T > A > t_0$ be large enough as before. Let $y_0 > T$ be a zero of $y'(t)$. From equation (3)

$$r(t)y'(t) + \int_{y_0}^t p(x)y'(x)\,dx + \int_{y_0}^t a(x)\frac{h(y(g(x)))}{y(g(x))}y(g(x))\,dx = \int_{y_0}^t f(x)\,dx.$$

This gives

$$\begin{aligned} r(t)y'(t) + p(t)y(t) - p(y_0)y(y_0) - \int_{y_0}^t p'(x)y(x)\,dx \\ + \int_{y_0}^t a(x)\frac{h(y(g(x)))}{y(g(x))}y(g(x))\,dx = \int_{y_0}^t f(x)\,dx. \end{aligned} \tag{58}$$

From (58)

$$\begin{aligned} r(t)y'(t) + p(t)y^+(t) - p(y_0)y(y_0) - \int_{y_0}^t p'(x)y(x)\,dx + m\int_{y_0}^t a(x)y^+(g(x))\,dx \\ \geqq \int_{y_0}^t f(x)\,dx \end{aligned} \tag{59}$$

since $p'(t) \leqq 0$. Now $\lim_{t\to\infty} \int_{y_0}^t f(x)\,dx = \infty$. Due to the bounded nature of $y^+(t)$ and other conditions of this theorem we get from (59)

$$r(t)y'(t)\to\infty \quad \text{as } t\to\infty \tag{60}$$

since $r(t)>0$, $y'(t)>0$ eventually, a contradiction. The proof of Theorem 6 is now complete.

Example 4. Consider the equation

$$y''(t)+(2+e^{-t})y'(t)+e^{-35t}y(t)=e^{2t}. \tag{61}$$

All conditions of Theorem 6 are satisfied. Hence all solutions of (61) are nonoscillatory.

Example 5. Consider the retarded equation

$$y''(t)+(2+e^{-t})y'(t)+e^{-3t+\pi}(y(t-\pi))=3e^{t}+e^{-2t}+1. \tag{62}$$

Here $r(t)\equiv 1$, $f(t)=e^{-2t}+3e^{t}+1$, $p(t)=2+e^{-t}$, $a(t)=e^{-3t+\pi}$. It is easily verified that conditions of Theorem 6 are satisfied. This equation has all solutions nonoscillatory. In fact $y(t)=e^{t}$ is one such solution.

Remark 1. The boundedness or unboundedness of $r(t)$ has not played any role in this theorem. We take up this matter in the next section.

THEOREM 7. *Suppose conditions of Theorem* 6 *hold. Further suppose that* $r(t)$ *is bounded. Then all solutions of* (3) *are unbounded and nonoscillatory* (*positive*).

Proof. We follow the proof of Theorem 6 to prove the nonoscillatory nature of solutions. Integrating equation 3 for $t\geqq T$ where T is large we have

$$\begin{aligned} &r(t)y'(t)-r(T)y'(T)+p(t)y(t)-p(T)y(T)\\ &\quad -\int_T^t p'(x)y(x)\,dx+\int_T^t a(x)\frac{h(y(g(x)))}{y(g(x))}y(g(x))\,dx=\int_T^t f(x)\,dx. \end{aligned} \tag{63}$$

Suppose now $y(t)$ is bounded; then (63) yields

$$r(t)y'(t)\to\infty \quad \text{as } t\to\infty. \tag{64}$$

This in view of boundedness of $r(t)$ gives the desired contradiction. The proof of Theorem 7 is now complete.

Remark 2. Coming back to Examples 4 and 5 we see that all solutions of equations (61) and (62) are nonoscillatory and unbounded.

Remark 3. It also follows from (63) that the specific nonoscillatory nature of these solutions is positive.

Remark 4. Boundedness of $r(t)$ in Theorem 7 cannot be weakened as indicated by our next example.

Example 6. Consider the equation

$$(e^{3t}y'(t))'+(1+e^{-t})y'(t)+e^{-t-\pi}y(t-\pi)=2e^{2t}+e^{-t}. \tag{65}$$

This equation has $y(t)=-e^{-t}$ as a negative nonoscillatory solution. For condition (41) of Lemma 2 we have

$$\begin{aligned} \left(\frac{fr'}{r}-f'+\frac{fp}{r}\right)' &= (6e^{2t}+3e^{-t}-4e^{2t}+e^{-t}+(2e^{2t}+2e^{t}+e^{-t}+e^{-2t})/e^{3t})'\\ &=(2e^{2t}+6e^{-t}+2e^{-2t}+e^{-4t}+e^{-5t})'\\ &>0 \quad \text{eventually.} \end{aligned}$$

All other conditions of Theorem 7 can be easily verified. Here unbounded $r(t)$ is causing the problem.

THEOREM 8. *Suppose conditions of Theorem* 6 *hold. Further suppose that* $r(t)$ *satisfies*

$$\liminf_{t\to\infty} \frac{\int^t f(s)\,ds}{r(t)} > 0. \tag{66}$$

Then all solutions of equation (3) *are unbounded and positive.*

Proof. Let $y(t)$ be a solution of equation (3). By Theorem 6 $y(t)$ is nonoscillatory. Suppose $y(t)$ is bounded. Let T be large enough so that $|y(t)| \leqq N$ for $t \geqq T$. Integrating equation (3) for $t \geqq T$ we get equation (63). Dividing (63) by $r(t)$ we have

$$y'(t) - \frac{r(T)y'(T)}{r(t)} + \frac{p(t)N + p(T)N}{r(t)} - \frac{N\int_T^t p'(x)\,dx}{r(t)} + \frac{mN}{r(t)}\int_T^t a(x)\,dx \geqq \frac{\int_T^t f(x)\,dx}{r(t)} \tag{67}$$

and a contradiction easily follows from (67) in view of condition (66). Hence $y(t)$ is unbounded. Suppose now that $y(t) < 0$ for $t \geqq T$. (63) reveals in view of $p(t) > 0$ and $p'(t) < 0$ that

$$r(t)y'(t) - r(T)y'(T) - p(T)y(T) \geqq \int_T^t f(x)\,dx. \tag{68}$$

Dividing (68) by $r(t)$ and taking the limit as $t \to \infty$ we find that $\liminf_{t\to\infty} y'(t) > 0$, a contradiction to negative $y(t)$. The proof of Theorem 8 is now complete.

Remark 5. Condition (66) cannot be weakened if all other conditions of Theorem 8 are satisfied. Example 6 testifies to this. In fact in Example 6

$$\liminf_{t\to\infty} \frac{\int^t f(x)\,dx}{r(t)} = 0,$$

violating (66).

6. Asymptotic oscillation. In this section we prove a theorem that gives conditions so that all oscillatory solutions of equation (3) approach zero. In fact Theorem 2 of this author [13, p. 40] states that all oscillatory solutions of equation (1) eventually vanish if $\int^\infty 1/r(t)\,dt < \infty$, $\int^\infty |a(t)|\,dt < \infty$ and $\int^\infty |f(t)|\,dt < \infty$. However consider the following example.

Example 7. The equation

$$\begin{aligned}(\exp(2t)y'(t))' + \exp((t/2) - 3\pi))y(t-\pi) \\ = 2\exp(-t)\sin t - 4\exp(-t)\cos t - \exp((-5/2)t)\sin t\end{aligned} \tag{69}$$

has $y = e^{-3t}\sin t$ as an oscillatory solution approaching zero. In fact it will be shown via Theorem 9 that all oscillatory solutions of (69) approach zero as $t \to \infty$. But this equation is not covered by our Theorem 2 in [13]. This motivation leads us to the following theorem.

THEOREM 9. *Suppose* $a(t) > 0$, $r(t) > 0$ *and*

$$\int^\infty \frac{1}{a(t)}\,dt < \infty, \tag{70}$$

$$H_2' > 0 \quad \textit{where} \quad H_2 = (ar'/r - a' + ap/r), \tag{71}$$

$$\int^\infty \frac{a^2(t)}{r(t)}\,dt < \infty, \tag{72}$$

(73) $$\int^{\infty} \frac{a(t)|f(t)|}{r(t)}\,dt < \infty.$$

Then all oscillatory solutions of equation (3) *approach zero asymptotically.*

Proof. Let $T \geqq t_0'$ be sufficiently large so that for $t \geqq T$, $H_2' > 0$ and we can write

(74) $$\int_T^{\infty} \frac{1}{a(t)}\,dt < \frac{1}{2},$$

(75) $$\int_T^{\infty} \frac{(a(t)|f(t)|)}{r(t)}\,dt < 1,$$

(76) $$\int_T^{\infty} \frac{a^2(t)}{r(t)}\,dt < 1.$$

Let $t_2 > t_1 > T$ be two consecutive zeros of $y(t)$ and without any loss of generality, suppose $y(t) > 0$ in (t_1, t_2); we rewrite equation (3) as

(77) $$(a(t)y'(t))' + H_2(t)y'(t) + \frac{a^2(t)}{r(t)}h(y(g(t))) = (a(t)f(t))/r.$$

Let $K_0 = \max y(t)$, $t \in [t_1, t_2]$. Following proof of Theorem 1 in [13, p. 39] from conclusions (9) through (14) we arrive at

$$4K_0^2 \leqq \left[\int_{t_1}^{t_2} \frac{1}{a(t)}\,dt\right]\left[-\int_{t_1}^{t_2} (a(t)y'(t))'y(t)\,dt\right].$$

Using (77) we get

$$4K_0^2 \leqq \left[\int_{t_1}^{t_2} \frac{1}{a(t)}\,dt\right]\left[\int_{t_1}^{t_2} H_2(t)y(t)y'(t)\,dt + \int_{t_1}^{t_2} \frac{a^2(t)}{r(t)}h(y(g(t)))y(t)\,dt - \int_{t_1}^{t_2} \frac{(a(t)f(t)y(t))}{r(t)}\,dt\right].$$

Since $H_2'(t) > 0$ and $y(t_1) = y(t_2) = 0$, $\int_{t_1}^{t_2} H_2 yy'\,dt = -\int_{t_1}^{t_2} H_2' y^2\,dt$; adding this to the right hand side and dividing by K_0 we get

(78) $$4K_0\left[\int_{t_1}^{t_2} \frac{1}{a(t)}\,dt\right]^{-1} \leqq \int_{t_1}^{t_2} \frac{a^2(t)}{r(t)}h(y(g(t)))\,dt + \int_{t_1}^{t_2} \frac{a(t)|f(t)|}{r(t)}\,dt.$$

From here on the proof of Theorem 2 (and Theorem 1) of [13, p. 39–41] i.e. from conclusion (16) through conclusion (29) and down on pages 39–41 applies verbatim. The proof is now complete.

Remark 6. Coming back to Example 7 we notice

$$r(t) = e^{2t}, \quad a(t) = e^{t/2-3\pi}, \quad f(t) = 2e^{-t}\sin t - 4e^{-t}\cos t - e^{(-5/2)t}\sin t,$$

$$p = 0, \quad H_2 = (ar'/r - a' + ap/r) = 2e^{t/2-3\pi} - \tfrac{1}{2}e^{t/2-3\pi} = \tfrac{3}{2}\exp(t/2 - 3\pi).$$

$$H_2' > 0, \quad \int^{\infty} \frac{1}{a(t)}\,dt < \infty, \quad \int^{\infty} \frac{a^2(t)}{r(t)}\,dt < \infty \quad \text{and} \quad \int^{\infty} \frac{a|f|}{r(t)}\,dt < \infty.$$

Thus all conditions of Theorem 9 are satisfied. Hence all oscillatory solutions of equation (69) vanish asymptotically.

REFERENCES

[1] T. BURTON AND R. GRIMMER, *On the asymptotic behavior of solutions of* $x''(t)+a(t)f(x)=e(t)$, Pacific J. Math., 41 (1972), pp. 77–88.

[2] K. FOSTER, *Criteria for oscillation and growth of nonoscillatory solutions of forced differential equations of even order*, J. Differential Equations, 20 (1976), pp. 115–132.

[3] J. R. GRAEF AND P. W. SPIKES, *Continuity, boundedness and asymptotic behavior of solutions of* $x''+q(t)f(x)=r(t)$, Ann. Mat. Pura Appl., CI (1974), pp. 307–320.

[4] ———, *Asymptotic behavior of solutions of a second order nonlinear differential equation*, J. Differential Equations, 17 (1975), pp. 461–476.

[5] M. E. HAMMETT, *Nonoscillation properties of a nonlinear differential equation*, Proc. Amer. Math. Soc., 30 (1971), pp. 92–96.

[6] T. KUSANO AND H. ONOSE, *Oscillations of functional differential equations with retarded arguments*, J. Differential Equations, 15 (1974), pp. 269–277.

[7] T. KUSANO AND M. NAITO, *Nonlinear oscillations of second order differential equations with retarded argument*, Ann. Mat. Pura Appl. CVI (1975), pp. 171–185.

[8] B. M. LEVITAN, *Some problems of the theory of almost periodic functions*, I, Uspehi Mat. Nauk., 2 (1947), pp. 133–192.

[9] S. LONDEN, *Some nonoscillation theorems for a second order nonlinear differential equation*, this Journal, 4 (1973), pp. 460–465.

[10] S. B. NORKIN, *Differential Equations of Second Order with Retarded Arguments*, American Mathematical Society, Providence, RI, 1972, pp. 1–8.

[11] S. RANKIN, *Oscillation of a forced second order nonlinear differential equation*, Proc. Amer. Math. Soc., to appear.

[12] BHAGAT SINGH, *Asymptotic nature of nonoscillatory solutions of nth order retarded differential equations*, this Journal, 6 (1975), pp. 784–795.

[13] ———, *Asymptotically vanishing oscillatory trajectories in second order retarded equations*, this Journal, 6 (1975), pp. 784–795.

[14] ———, *Impact of delays on oscillation in general functional equations*, Hiroshima Math. J., 5 (1975), pp. 351–361.

[15] ———, *Nonoscillation of forced fourth order retarded equations*, SIAM J. Appl. Math., 28 (1975), pp. 265–269.

[16] ———, *A necessary and sufficient condition for the oscillation of an even order nonlinear delay differential equation*, Canad. J. Math., 25 (1973), pp. 1078–1089.

[17] ———, *Forced oscillations in general ordinary differential equations with deviating arguments*, Hiroshima Math. J., 6 (1976), pp. 7–14.

[18] V. A. STAIKOS AND Y. G. SFICAS, *Forced oscillations for differential equations of arbitrary order*, J. Differential Equations, 17 (1975), pp. 1–11.

[19] C. C. TRAVIS, *Oscillation theorems for second order differential equations with functional arguments*, Proc. Amer. Math. Soc., 30 (1972), pp. 199–201.

[20] H. TUEFEL, *Forced second order nonlinear oscillations*, J. Math. Anal. Appl., 40 (1972), pp. 148–152.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001/-0003 $01.00/0

A NONLINEAR SINGULAR PERTURBATION PROBLEM FOR SECOND ORDER SYSTEMS*

WALTER G. KELLEY†

Abstract. The existence and asymptotic behavior as $\varepsilon \to 0^+$ of solutions of nonlinear boundary value problems for second order systems are studied using differential inequality techniques. Conditions are given under which two point problems for $\varepsilon x'' = f(t, x, x', \varepsilon)$ have unique solutions which converge uniformly as $\varepsilon \to 0^+$ outside boundary layers at each endpoint of width $\sqrt{\varepsilon}$ to a solution of the reduced equation $0 = f(t, x, x', 0)$.

1. Introduction. Recently, Chang [2] has studied the quasilinear boundary value problem

$$\varepsilon x'' + C(t, x, \varepsilon)x' = h(t, x, \varepsilon), \tag{1}$$

$$x(0, \varepsilon) = A(\varepsilon), \qquad x(1, \varepsilon) = B(\varepsilon), \tag{2}$$

where x, h, A and B are vector-valued and C is a matrix function. Under the assumptions that the reduced problem

$$C(t, x, 0)x' = h(t, x, 0),$$
$$x(1) = B(0)$$

has a C^2 solution $u(t)$ and every eigenvalue of $C(t, u(t), 0)$ has real part greater than or equal to $8\mu > 0$ for $0 \leqq t \leqq 1$, plus additional assumptions, he proves that for ε sufficiently small, the problem (1), (2) has a solution $x(t, \varepsilon)$ on $[0, 1]$ such that

$$x(t, \varepsilon) = u(t) + \mathcal{O}(\varepsilon) + \mathcal{O}(e^{-\mu t/\varepsilon}),$$
$$x'(t, \varepsilon) = u'(t) + \mathcal{O}(\varepsilon) + \mathcal{O}(e^{-\mu t/\varepsilon}),$$

where the Landau order symbol holds uniformly in t as $\varepsilon \to 0$.

A natural question is: What can be concluded if $C = 0$ or if C is "small" in a neighborhood of $u(t)$? In the case of a scalar equation, Howes [4] has shown that if $C = 0$, if the reduced equation

$$0 = h(t, x, 0) \tag{3}$$

has a C^2 solution u on $[0, 1]$, if $\partial h/\partial x \geqq m > 0$ in a neighborhood of u and if certain other conditions are satisfied, then (1), (2) has a unique solution $x(t, \varepsilon)$ for ε sufficiently small, and

$$|x(t, \varepsilon) - u(t)| \leqq |A(\varepsilon) - u(0)|\, e^{-\sqrt{m/\varepsilon}\, t} + |B(\varepsilon) - u(1)|\, e^{-\sqrt{m/\varepsilon}(1-t)} + c\varepsilon$$

for $0 \leqq t \leqq 1$, where c is a positive constant independent of ε.

In this note we show that Howes' result can be extended to vector equations under suitable hypotheses. More generally, we also demonstrate that similar conclusions can be reached for the vector boundary value problem consisting of (2) and

$$\varepsilon x'' = f(t, x, x', \varepsilon), \tag{4}$$

provided that $\partial f/\partial x'$ is "small" in a sense to be made precise below.

* Received by the editors November 24, 1976, and in final revised form May 5, 1977.
† Department of Mathematics, University of Oklahoma, Norman, Oklahoma 73019.

2. Some preliminary results. In this section we collect for the convenience of the reader the result to be used in the proofs of our main theorems. Let us consider the two point boundary value problem

$$x''=g(t,x,x'), \tag{5}$$

$$x(0)=\alpha, \qquad x(1)=\beta, \tag{6}$$

where $g\colon [0,1]\times R^d\times R^d\to R^d$ is continuous and $\alpha,\beta\in R^d$.

For $i=1,\cdots,N$, let $r_i(t,x)$ be of class C^2 on $[0,1]\times R^d$, $w_i(t,x)$ the gradient vector of r_i, $v_i(t,x)$ the gradient vector of $\partial r_i/\partial t$, where these gradients are taken with respect to x, and $P_i(t,x)$ the Hessian of r_i with respect to x. Let the first and second derivatives of r_i with respect to (5) be denoted by

$$r_i'=\frac{\partial r_i}{\partial t}+w_i\cdot x', \tag{7}$$

$$r_i''=\frac{\partial^2 r_i}{\partial t^2}+2v_i\cdot x'+x'P_i\cdot x'+w_i\cdot g, \tag{8}$$

for $i=1,\cdots,N$, where the dot indicates the usual scalar product in R^d. Define $D=\{(t,x,y)\colon 0\leqq t\leqq 1,\ r_i(t,x)<0 \text{ for } i=1,\cdots,N,\ y\in R^d\}$.

We give two types of Nagumo conditions for g.

N_1: There exists a sequence $\{\phi_i\}_{i=1}^d$ of positive, nondecreasing continuous functions on $(0,\infty)$ such that

$$\int^\infty \frac{s\,ds}{\phi_i(s)}=\infty$$

and

$$|g^i(t,x,y)|\leqq\phi_i(|y^i|) \quad \text{for } (t,x,y)\in D, \quad i=1,\cdots,d.$$

N_2: There is a positive, nondecreasing, continuous function ϕ on $(0,\infty)$ such that

$$\frac{s^2}{\phi(s)}\to\infty \quad \text{as } s\to\infty$$

and

$$\|f(t,x,y)\|\leqq\phi(\|y\|) \quad \text{for } (t,x,y)\in D.$$

The following theorem is a special case of Theorem 4 in [5].

THEOREM 1. *Assume* $\{(t,x)\colon 0\leqq t\leqq 1,\ r(t,x)\leqq 0\}$ *is a bounded set and*

(a) *the functions* r_i *described above satisfy for* $i=1,\cdots,d$

$$r_i''>0 \quad \textit{when } r_i=0 \textit{ and } r_i'=0; \tag{9}$$

(b) *there is a function of class* C^2 *on* $[0,1]$ *which satisfies* (6) *and whose trajectory is contained in* D;

(c) *initial value problems for* (5) *have unique solutions*;

(d) g *satisfies either* N_1 *or* N_2 *on* D.

Then the boundary value problem (5), (6) *has a solution* $x(t)$ *with* $r_i(t,x(t))<0$ *for* $0\leqq t\leqq 1$ *and* $i=1,\cdots,N$.

We note that the proof of this theorem requires only that each r_i be of class C^2 in a neighborhood of the set $\{(t,x)\colon 0\leqq t\leqq 1,\ r(t,x)=0\}$. Also, assumption (a) can be relaxed and assumption (c) can be omitted if additional conditions are placed on the functions r_i. See Theorem 5 of [5].

If $\ell : R^d \to R^d$ is a linear operator, then its adjoint will be denoted ℓ^*, and $\ell \geqq 0$ means that $x \cdot \ell(x) \geqq 0$ for all $x \in R^d$. The next theorem is proved in [3].

THEOREM 2. *Suppose that g has continuous partial derivatives with respect to x and x', and*

$$4\frac{\partial g}{\partial x} - \frac{\partial g}{\partial x'}\left(\frac{\partial g}{\partial x'}\right)^* \geqq 0, \tag{10}$$

where the partials are computed at an arbitrary point (t, x, x') in $[0, 1] \times R^d \times R^d$. Then (5), (6) *has at most one solution.*

The proof of this theorem shows that if a priori bounds, say $\|x\| \leqq R$, $\|x'\| \leqq \mu$, are known for solutions of (5), (6), then the hypotheses of the theorem need hold only for these restricted values of x and x'.

3. Singular perturbation problems. We begin by considering the system (1) with $C = 0$, namely,

$$\varepsilon x'' = h(t, x, \varepsilon), \tag{11}$$

on the interval [0, 1] with boundary conditions (2) and the reduced equation (3). The following theorem is a generalization of Theorem 3.1 in [4].

THEOREM 3. *Assume*:

(a) *equation* (3) *has a $C^{(2)}[0, 1]$ solution u*;

(b) *h is continuous in (t, x, ε) and is of class $C^{(1)}$ with respect to x in*

$$E = \{(t, x, \varepsilon): 0 \leqq t \leqq 1, \|x - u(t)\| \leqq \|A(\varepsilon) - u(0)\| e^{-\sqrt{m/\varepsilon}t} + \|B(\varepsilon) - u(1)\| e^{-\sqrt{m/\varepsilon}(1-t)} + c\varepsilon, 0 < \varepsilon \leqq \varepsilon_1\},$$

for some positive constants m, c and ε_1;

(c) *for all $(t, x, \varepsilon) \in E$, $(\partial h/\partial x)(t, x, \varepsilon) - mI \geqq 0$, where I is the identity*;

(d) *there is a $\gamma > 0$ so that $\|h(t, u(t), \varepsilon)\| < \gamma\varepsilon$ for $0 \leqq t \leqq 1$, $0 < \varepsilon \leqq \varepsilon_1$.*

Then for each ε, $0 < \varepsilon \leqq \varepsilon_1$, there is a unique solution $x(t, \varepsilon)$ of (11), (2) *which satisfies*

$$\|x(t, \varepsilon) - u(t)\| < \|A(\varepsilon) - u(0)\| e^{-\sqrt{m/\varepsilon}t} + \|B(\varepsilon) - u(1)\| e^{-\sqrt{m/\varepsilon}(1-t)} + c\varepsilon,$$

for $0 \leqq t \leqq 1$.

Proof. For the first part of the proof, we assume $u(t) = 0$ for $0 \leqq t \leqq 1$. Fix ε so that $0 < \varepsilon \leqq \varepsilon_1$. Define

$$r(t, x) = \|x\| - \|A(\varepsilon)\| e^{-\sqrt{m/\varepsilon}t} - \|B(\varepsilon)\| e^{-\sqrt{m/\varepsilon}(1-t)} - \frac{\varepsilon\gamma}{m},$$

for $0 \leqq t \leqq 1$ and $x \in R^d$. Note that r is of class C^2 except when $x = 0$ (see the remark following Theorem 1).

We will apply Theorem 1. All the hypotheses are easily seen to be satisfied, except (a). In formula (8), we have $w(x) = x/\|x\|$ $(x \neq 0)$ and $v(t, x) = 0$ for all t and all x. Furthermore, $P(x) \geqq 0$ for all $x \neq 0$ since the function $\|x\|$ is convex. Thus (9) will be satisfied if we show

$$\frac{\partial^2 r}{\partial t^2} + \frac{x}{\|x\|} \cdot \frac{1}{\varepsilon} h(t, x, \varepsilon) > 0$$

whenever $r = 0$.

Assume $r(t, x)=0$ and compute

$$\frac{\partial^2 r}{\partial t^2}+\frac{x}{\|x\|}\cdot\frac{1}{\varepsilon}h(t, x, \varepsilon)=-\|A(\varepsilon)\|\frac{m}{\varepsilon}e^{-\sqrt{m/\varepsilon}t}-\|B(\varepsilon)\|\frac{m}{\varepsilon}e^{-\sqrt{m/\varepsilon}(1-t)}$$

$$+\frac{x}{\|x\|}\cdot\frac{h}{\varepsilon}(t, 0, \varepsilon)+\frac{x}{\|x\|}\cdot\frac{1}{\varepsilon}[h(t, x, \varepsilon)-h(t, 0, \varepsilon)].$$

By applying a version of the mean value theorem to the last expression (see [1]), we have

$$\frac{x}{\|x\|}\cdot\frac{1}{\varepsilon}[h(t, x, \varepsilon)-h(t, 0, \varepsilon)]=\frac{x}{\|x\|}\cdot\frac{1}{\varepsilon}\frac{\partial h}{\partial x}(t, z, \varepsilon)(x)\geqq\frac{m}{\varepsilon}\|x\|.$$

by hypothesis (c), where z is on the line segment from 0 to x. Thus

$$\frac{\partial^2 r}{\partial t^2}+\frac{x}{\|x\|}\cdot\frac{1}{\varepsilon}h(t, x, \varepsilon)\geqq-\|A(\varepsilon)\|\frac{m}{\varepsilon}e^{-\sqrt{m/\varepsilon}t}-\|B(\varepsilon)\|\frac{m}{\varepsilon}e^{-\sqrt{m/\varepsilon}(1-t)}+\frac{x}{\|x\|}\cdot\frac{h}{\varepsilon}(t, 0, \varepsilon)$$

$$+\frac{m}{\varepsilon}\left[\|A(\varepsilon)\|e^{-\sqrt{m/\varepsilon}t}+\|B(\varepsilon)\|e^{-\sqrt{m/\varepsilon}(1-t)}+\frac{\varepsilon\gamma}{m}\right]$$

$$=\frac{x}{\|x\|}\cdot\frac{h}{\varepsilon}(t, 0, \varepsilon)+\gamma>0$$

by hypothesis (d). From Theorem 1, we conclude that the boundary value problem (11), (2) has a solution $x(t, \varepsilon)$ with

$$\|x(t, \varepsilon)\|<\|A(\varepsilon)\|e^{-\sqrt{m/\varepsilon}t}+\|B(\varepsilon)\|e^{-\sqrt{m/\varepsilon}(1-t)}+\frac{\varepsilon\gamma}{m}$$

for $0\leqq t\leqq 1$.

If the reduced solution $u(t)$ is not zero, then we can make the change of variable $y=x-u(t)$, and the transformed boundary value problem satisfies the hypotheses of the theorem with reduced solution zero, so that the first part of the proof is applicable. The uniqueness follows immediately from Theorem 2. Q.E.D.

In [4], Howes also considers the fourth order problem

$$\varepsilon x^{(4)}=F(t, x, x'', \varepsilon),$$

$$x(0, \varepsilon)=A_1(\varepsilon), \qquad x(1, \varepsilon)=B_1(\varepsilon),$$

$$x''(0, \varepsilon)=A_2(\varepsilon), \qquad x''(1, \varepsilon)=B_2(\varepsilon).$$

His results for this scalar problem can be extended to systems by following the basic outline of his argument but applying Theorem 1 instead of the corresponding theorem for scalar equations. Since our proof of Theorem 3 illustrates the type of modifications which must be made in analyzing the vector problem, we omit the details.

We now consider equation (4), where

$$f:[0, 1]\times R^d\times R^d\times[0, \infty)\to R^d.$$

THEOREM 4. *Assume*:

(a) *the reduced problem*, $0=f(t, x, x', 0)$, *has a* $C^{(2)}[0, 1]$ *solution* u;

(b) f *is continuous in* (t, x, x', ε) *and is of class* $C^{(1)}$ *with respect to* x *and* x' *in*

$$F=\{(t, x, x', \varepsilon): 0\leqq t\leqq 1, \|x-u(t)\|^2\leqq\|A(\varepsilon)-u(0)\|^2 e^{-\sqrt{2m/\varepsilon}t}$$
$$+\|B(\varepsilon)-u(1)\|^2 e^{-\sqrt{2m/\varepsilon}(1-t)}+c\varepsilon^2, 0<\varepsilon\leqq\varepsilon_1\},$$

for some positive constants m, c and ε_1;

(c) *there is a* $\delta > 0$ *so that*

$$\left(\frac{\partial f}{\partial x} - \frac{1}{4\varepsilon}\left(\frac{\partial f}{\partial x'}\right)\left(\frac{\partial f}{\partial x'}\right)^*\right)(t, x, x', \varepsilon) - (m+\delta)I \geqq 0$$

for all $(t, x, x', \varepsilon) \in F$;

(d) F *satisfies* N_1 *or* N_2 *in* F;

(e) *there is a* $\gamma > 0$ *so that* $\|f(t, u(t), u'(t), \varepsilon)\| < \gamma\varepsilon$ *for* $0 \leqq t \leqq 1$, $0 \leqq \varepsilon \leqq \varepsilon_1$.

Then for each ε, $0 < \varepsilon \leqq \varepsilon_1$, *there exists a unique solution* $x(t, \varepsilon)$ *of* (4), (2) *which satisfies*

$$\|x(t, \varepsilon) - u(t)\|^2 < \|A(\varepsilon) - u(0)\|^2 e^{-\sqrt{2m/\varepsilon}t} + \|B(\varepsilon) - u(1)\|^2 e^{-\sqrt{2m/\varepsilon}(1-t)} + c\varepsilon^2$$

for $0 \leqq t \leqq 1$.

Proof. As in the proof of Theorem 3, we assume that $u(t) = 0$ for $0 \leqq t \leqq 1$, and this assumption gives no loss of generality. Fix ε so that $0 < \varepsilon \leqq \varepsilon_1$ and define

$$r(t, x) = \|x\|^2 - \|A(\varepsilon)\|^2 e^{-\sqrt{2m/\varepsilon}t} - \|B(\varepsilon)\|^2 e^{-\sqrt{2m/\varepsilon}(1-t)} - \left(\frac{\gamma}{\delta}\varepsilon\right)^2$$

for $0 \leqq t \leqq 1$ and $x \in R^d$. Among the hypotheses of Theorem 1, only (a) is not immediate.

For this choice of r, we have $w(x) = 2x$, $v(x) = 0$ and $P(x) = 2I$ for all $x \in R^d$. Thus in F we obtain

$$r'' = \frac{\partial^2 r}{\partial t^2} + 2x' \cdot x' + 2x \cdot \frac{f}{\varepsilon}(t, x, x', \varepsilon).$$

Let f' denote the differential of f with respect to (x, x') for fixed values of (t, ε). By applying the mean value theorem used in the proof of Theorem 3, we have

$$\begin{aligned} x \cdot f(t, x, x', \varepsilon) &= x \cdot f(t, 0, 0, \varepsilon) + x \cdot [f(t, x, x', \varepsilon) - f(t, 0, 0, \varepsilon)] \\ &= x \cdot f(t, 0, 0, \varepsilon) + x \cdot f'(t, z, z', \varepsilon)(x, x') \\ &= x \cdot f(t, 0, 0, \varepsilon) + x \cdot \frac{\partial f}{\partial x}(t, z, z', \varepsilon)(x) \\ &\qquad + x \cdot \frac{\partial f}{\partial x'}(t, z, z', \varepsilon)(x'), \end{aligned}$$

for some point (z, z') on the line segment between $(0, 0)$ and (x, x'). Thus the expression

$$x' \cdot x' + x \cdot \frac{f}{\varepsilon}(t, x, x', \varepsilon)$$

can be written in the form

$$x \cdot \frac{f}{\varepsilon}(t, 0, 0, \varepsilon) + \left|x' + \frac{1}{2\varepsilon}\left(\frac{\partial f}{\partial x'}\right)^*(x)\right|^2 + \frac{x}{\varepsilon} \cdot \left(\frac{\partial f}{\partial x} - \frac{1}{4\varepsilon}\left(\frac{\partial f}{\partial x'}\right)\left(\frac{\partial f}{\partial x'}\right)^*\right)(x),$$

where all partials are evaluated at (t, z, z', ε). From hypotheses (c) and (e), it follows that whenever $r(t, x)=0$, we have

$$\begin{aligned} r'' &\geqq \frac{\partial^2 r}{\partial t^2} - \frac{2}{\varepsilon}\|x\|\,\|f(t, 0, 0, \varepsilon)\| + 2\frac{m+\delta}{\varepsilon}\|x\|^2 \\ &= -\frac{2m}{\varepsilon}\|A(\varepsilon)\|^2 e^{-\sqrt{2m/\varepsilon}t} - \frac{2m}{\varepsilon}\|B(\varepsilon)\|^2 e^{-\sqrt{2m/\varepsilon}(1-t)} - \frac{2}{\varepsilon}\|x\|\,\|f(t, 0, 0, \varepsilon)\| \\ &\quad + \frac{2\delta}{\varepsilon}\|x\|^2 + \frac{2m}{\varepsilon}\left[\|A(\varepsilon)\|^2 e^{-\sqrt{2m/\varepsilon}t} + \|B(\varepsilon)\|^2 e^{-\sqrt{2m/\varepsilon}(1-t)} + \frac{\gamma^2}{\delta^2}\varepsilon^2\right] \\ &= \frac{2}{\varepsilon}\left[\|x\|(\delta\|x\| - \|f(t, 0, 0, \varepsilon)\|) + \frac{m\gamma^2}{\delta^2}\varepsilon^2\right] > 0, \end{aligned}$$

since $\delta\|x\| > \delta(\gamma\varepsilon/\delta) = \gamma\varepsilon \geqq \|f(t, 0, 0, \varepsilon)\|$ if $r(t, x)=0$.

Theorem 1 applies, and the problem (4), (2) has a solution $x(t, \varepsilon)$ which satisfies $r(t, x(t, \varepsilon))<0$ for $0\leqq t\leqq 1$. Since $(\partial f/\partial x')(\partial f/\partial x')^* \geqq 0$, hypothesis (c) implies that (10) is satisfied for $g=f$, and the solution $x(t, \varepsilon)$ is unique. Q.E.D.

In the case of the quasilinear equation (1), where $f(t, x, x', \varepsilon) = -C(t, x, \varepsilon)x' + h(t, x, \varepsilon)$, hypothesis (c) of Theorem 4 essentially requires $\partial f/\partial x$ to be positive definite and $C(t, x, \varepsilon) = \mathcal{O}(\sqrt{\varepsilon})$ for x in a neighborhood of u. However, this condition may hold even if (4) is not quasilinear. For example, in the scalar problem

$$\varepsilon y'' = y + \varepsilon y(y')^2,$$
$$y(0)=\tfrac{1}{2}, \quad y(1)=\tfrac{1}{2},$$

we have

$$\frac{\partial f}{\partial y} - \frac{1}{4\varepsilon}\left(\frac{\partial f}{\partial y'}\right)^2 = 1 + \varepsilon(y')^2(1-y) > 1$$

for y sufficiently close to the reduced solution $u(t)=0$, $0\leqq t\leqq 1$.

Finally, we should add that differential inequality methods of the type used above do not appear to be effective for singularly perturbed systems in which the derivative x' plays a substantial role in the differential equation.

REFERENCES

[1] T. M. APOSTOL, *Mathematical Analysis*, Addison-Wesley, Reading, MA, 1974.

[2] K. W. CHANG, *Singular perturbations of a boundary value problem for a vector second order differential equation*, Siam J. Appl. Math., 30 (1976), pp. 42–54.

[3] P. HARTMAN, *On the boundary value problems for systems of ordinary, nonlinear, second order differential equations*, Trans. Amer. Math. Soc., 96 (1960), pp. 493–509.

[4] F. A. HOWES, *Differential inequalities and applications to nonlinear singular perturbation problems*, J. Differential Equations, 20 (1976), pp. 133–148.

[5] W. G. KELLEY, *A geometric method of studying two point boundary value problems for second order systems*, Rocky Mountain J. Math., 7 (1977), pp. 251–263.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0004 $01.00/0

EXTENSIONS OF SHEFFER POLYNOMIAL SETS*

WM. R. ALLAWAY†

Abstract. A Sheffer A type zero polynomial set $\{P_n(x)\}_{n=0}^{\infty}$ is one in which its generating function is of the form $A(t)\, e^{xH(t)} = \sum_{n=0}^{\infty} P_n(x)t^n/n!$. For any Sheffer A type zero polynomial set $\{P_n(x)\}_{n=0}^{\infty}$, a method is given for constructing a formal Newton series expansion $\psi(z, s)$ such that $a^{-s}\psi(x+a, s) = \sum_{k=0}^{\infty} P_k(x)[s^{(k)}/(a^k k!)]$, where $s^{(k)} = s(s-1)\cdots(s-k+1)$ and $\psi(z, s)$ is an extension of $P_n(x)$ in the sense that $\psi(x, n) = P_n(x)$ for $n = 0, 1, 2, \cdots$. The extensions of the Bernoulli and Euler polynomial sets are given in terms of the Hurwitz zeta functions. These extensions are shown to be formal solutions of some finite difference equations with nonconstant coefficients. These finite difference equations are then used to linearize the product $P_n(x)P_m(x+a)$ for the Hermite and Euler polynomial set. For the Hermite case the inverse formulas of these linearizations are obtained.

1. Introduction. L. Poli [13] in 1954 showed that for the Hermite polynomial set $\{H_n(x)\}_{n=0}^{\infty}$

$$yH_n(x) = \sum_{k=0}^{n} \binom{n}{k} t^{n-k} D_t^k y, \tag{1.1}$$

where y is the generating function of the Hermite polynomial set defined by

$$y := \exp(xt - t^2/2) = \sum_{k=0}^{\infty} H_k(x)\frac{t^k}{k!}.$$

Haradze [7], 1964, obtained a similar result for the ultraspherical polynomial set. For this case the differential operator is not linear. Allaway [1] has done the Laguerre, Meixner, and Poisson–Charlier polynomial case. Ismail [8] has generalized Poli's result to the class of Appell polynomial sets. These types of results are interesting because they are useful in finding the solutions in close form of some differential equations with nonconstant coefficients. (See Allaway [1].)

In this paper, we are interested in finding the finite difference analogue of Poli's formula for some well-known polynomial sets.

It is obvious that Poli's formula for $\{x^n\}_{n=0}^{\infty}$ is

$$e^{xt}x^n = D_t^n\, e^{xt}, \tag{1.2}$$

where the generating function of $\{x^n\}_{n=0}^{\infty}$ is

$$e^{xt} = \sum_{k=0}^{\infty} \frac{x^k t^k}{k!}.$$

The finite difference analogue for the polynomial set $\{x^n\}_{n=0}^{\infty}$ is

$$(1+x)^s x^n = \Delta_s^n (1+x)^s, \tag{1.3}$$

where the Newton series generating function of $\{x^n\}_{n=0}^{\infty}$ is

$$(1+x)^s = \sum_{n=0}^{\infty} \frac{x^n s^{(n)}}{n!}.$$

* Received by the editors September 15, 1977 and in revised form November 3, 1977. Presented to the American Mathematical Society January 23, 1976 at the Winter Annual Meeting A.M.S., San Antonio, Texas.

† Department of Mathematical Sciences, Lakehead University, Thunder Bay, Ontario, Canada P7B 5E1. This work was supported in part by the National Research Council of Canada under Grant A8721.

Throughout this paper, we will always use the notation

$$s^{(n)} = s(s-1)(s-2)\cdots(s-n+1).$$

In order to generalize what is done in going from (1.2) to (1.3), that is to find finite difference analogues to (1.1), we study

1) polynomial sets $\{P_n(x)\}_{n=0}^{\infty}$ such that

$$\psi(x+1, s) = \sum_{k=0}^{\infty} P_k(x)\frac{s^{(k)}}{k!},$$

where $\psi(x, n) = P_n(x)$ $n = 0, 1, 2, \cdots$ and

2) a transformation T such that $T(e^{xt}) = (1+x)^s$.

2. Extensions and Appell polynomial sets. Let $\{P_n(x)\}_{n=0}^{\infty}$ be a polynomial set. That is, $P_n(x)$ is a real polynomial of degree exactly equal to n. An *extension* of $\{P_n(x)\}_{n=0}^{\infty}$ is a function $\psi(z, s)$ of the two complex variables z and s such that $\psi(z, n) = p_n(z)$ for $n = 0, 1, 2, \cdots$, and z belonging to the complex numbers. Mathematical physics abounds with examples of polynomial sets and their extensions. For example, z^s is an extension of $\{x^n\}_{n=0}^{\infty}$, the Hermite functions are an extension of the Hermite polynomial set (see [10, p. 285]), and $s\zeta(1-s, z)$ is an extension of the Bernoulli polynomial set $\{B_n(x)\}_{n=0}^{\infty}$, (see [6, vol. 1, p. 27]), where $\zeta(s, z)$ is the Hurwitz zeta function defined for Re $(s) > 1$, by

$$\zeta(s, z) = \sum_{n=0}^{\infty} (z+n)^{-s}, \qquad z \neq 0, -1, -2, \cdots. \tag{2.1}$$

Let $\Phi(z, s, v)$ be defined by

$$\Phi(z, s, v) = \sum_{n=0}^{\infty} (v+n)^{-s} z^n.$$

See [6, vol. 1, p. 27] for some of its properties. It is easy to show that $2\Phi(-1, -s, z)$ is an extension of the Euler polynomial set. Indeed, by using the complex contour integral representation of $\zeta(s, z)$ and $\Phi(-1, s, z)$ (see [6, vol. 1, pp. 25, 28]) it follows that for $s \neq 1, 2, 3, \cdots$ and Re $(z) > 0$

$$\Phi(-1, s, z) = 2^{-s}[\zeta(s, z/2) - \zeta(s, (z+1)/2)].$$

By using this fact and the fact that $s\zeta(1-s, z)$ is an extension of the Bernoulli polynomial set, we obtain

$$\begin{aligned} 2\Phi(-1, -m, z) &= 2^{m+1}[\zeta(-m, z/2) - \zeta(-m, (z+1)/2)] \\ &= 2^{m+1}\frac{[B_{m+1}(z/2) - B_{m+1}((z+1)/2)]}{-(m+1)} \\ &= E_m(z). \end{aligned}$$

From our definition of extension it is obvious that a given polynomial set has an infinite number of extensions.

Sheffer [16] in 1939 studied polynomial sets $\{P_n(x)\}_{n=0}^{\infty}$ that have a generating function of the form

$$A(t)\exp(xH(t)) = \sum_{n=0}^{\infty} P_n(x)\frac{t^n}{n!}, \tag{2.2}$$

with

$$A(t)=\sum_{n=0}^{\infty} a_nt^n, \quad H(t)=\sum_{n=1}^{\infty} h_nt^n, \quad a_0h_1\neq 0.$$

Such polynomial sets are now known as Sheffer A type zero polynomial sets. Many of the classical polynomial sets are Sheffer A type zero polynomial sets. For example, $\{x^n\}_{n=1}^{\infty}$, the Bernoulli polynomial set, the Euler polynomial set, the Hermite polynomial set, the Laguerre Polynomial set, the Poisson–Charlier polynomial set and the Meixner polynomial set are all Sheffer A type zero polynomial sets.

Let $\{Q_n(x)\}_{n=0}^{\infty}$ have a generating function of the form

$$\exp(xH(t))=\sum_{n=0}^{\infty} Q_n(x)\frac{t^n}{n!}, \tag{2.3}$$

where $H(t)=\sum_{n=1}^{\infty} h_nt^n$, $h_1\neq 0$. Define

$$\phi(x, y, s)=\sum_{k=0}^{\infty} P_k(x)Q_{s-k}(y)\frac{s^{(k)}}{k!}, \tag{2.4}$$

where $Q_s(x)$ is any extension of $\{Q_n(x)\}_{n=0}^{\infty}$.

We first wish to show that $\{P_n(x)\}_{n=0}^{\infty}$ is a Sheffer A type zero polynomial set if and only if there exists a polynomial set $\{Q_n(x)\}_{n=0}^{\infty}$ having a generating function of the form (2.3), such that for all complex numbers c, $\phi(x-c, c, s)$ as given by (2.4) is an extension of $\{P_n(x)\}_{n=0}^{\infty}$.

First, let us show that every Sheffer A type zero polynomial set $\{P_n(x)\}_{n=0}^{\infty}$ has an extension of the form as given by (2.4). Indeed, if $\{P_n(x)\}_{n=0}^{\infty}$ is a Sheffer A type zero polynomial set, then it follows from the generating function (2.2) that for all complex numbers c and z,

$$\begin{aligned} P_n(z)&=\sum_{k=0}^{n}\binom{n}{k}Q_{n-k}(c)P_k(z-c)\\ &=\phi(z-c, c, n), \end{aligned}$$

for $n=0, 1, 2, \cdots$. Thus $\phi(x-c, c, s)$ is an extension of $\{P_n(x)\}_{n=0}^{\infty}$.

Conversely, let $\phi(x-c, c, s)$ be an extension of $\{P_n(x)\}_{n=0}^{\infty}$. Thus

$$\begin{aligned} P_n(x)&=\phi(x-c, c, n)\\ &=\sum_{k=0}^{n}\binom{n}{k}P_k(x-c)Q_{n-k}(c). \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{n=0}^{\infty} P_n(x)\frac{t^n}{n!}&=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n}{k}P_k(x-c)Q_{n-k}(c)\frac{t^n}{n!}\\ &=\sum_{n=0}^{\infty} P_n(x-c)\frac{t^n}{n!}\sum_{k=0}^{\infty} Q_k(c)\frac{t^k}{k!}. \end{aligned} \tag{2.5}$$

By hypothesis,

$$\exp(cH(t))=\sum_{k=0}^{\infty} Q_k(c)\frac{t^k}{k!}.$$

Therefore, if we let $F(x, t)$ be the generating function of $\{P_n(x)\}_{n=0}^{\infty}$ we obtain from (2.5)

$$F(x, t) = F(x-c, t) \exp(cH(t)). \tag{2.6}$$

Let $\log_e F(x, t) = u(x, t)$. Thus (2.6) becomes

$$u(x, t) - u(x-c, t) = cH(t),$$

which has a solution

$$u(x, t) = xH(t) + g(x, t),$$

where $g(x, t) = g(x-c, t)$ for all complex numbers c. Therefore,

$$F(x, t) = G(x, t) \exp(xH(t)), \tag{2.7}$$

where $G(x, t) = G(x-c, t)$ for all complex numbers c. From (2.7) and the fact that $F(x, t)$ is the generating function of $\{P_n(x)\}_{n=0}^{\infty}$ we have that

$$G(x, t) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} P_{n-k}(x) Q_k(-x) \frac{t^n}{n!}.$$

Thus, for all complex numbers c,

$$\sum_{k=0}^{n} \binom{n}{k} P_{n-k}(x) Q_k(-x)$$

is periodic with period c and therefore

$$\sum_{k=0}^{n} \binom{n}{k} P_{n-k}(x) Q_k(-x) = a_n.$$

Thus,

$$G(x, t) = \sum_{n=0}^{\infty} a_n \frac{t^n}{n!}$$
$$= A(t).$$

Therefore, $\{P_n(x)\}_{n=0}^{\infty}$ is a Sheffer A type zero polynomial set.

In what follows we will restrict our attention to the special case

$$H(t) = t.$$

Sheffer A type zero polynomial sets for which $H(t) = t$ are known as Appell polynomial sets (see [2]). From the above we know that an Appell polynomial set $\{P_n(x)\}_{n=0}^{\infty}$ is characterized by the fact that it has an extension $\psi(z, s)$ of the form

$$\psi(x+a, s) = \sum_{k=0}^{\infty} P_k(x) a^{s-k} \frac{s^{(k)}}{k!}. \tag{2.8}$$

3. Some transformations. In the preceeding section, we have shown how to find (see (2.8)) an extension of any Appell polynomial set. The other thing we need in order to find finite difference analogues of Poli's formula is to study the transformation T such that

$$T(e^{xt}) = (1+x)^s.$$

By writing the power series expansion of e^{xt} in terms of t^n and the Newton series expansion of $(1+x)^s$ in terms of $s^{(n)}$ we see that

$$T(t^n)=s^{(n)}. \tag{3.1}$$

Instead of using the transformation as defined by (3.1), we use a mild generalization

$$T_a(t^n)=\frac{s^{(n)}}{a^n}, \tag{3.2}$$

where a is a complex number not equal to zero. It turns out, that when we use this definition for our transform, we can generalize some previously known results. For example, one of the finite difference analogues of Poli's formulas contain as a special case

$$H_m(x)H_n(x)=\sum_{k=0}^{\min(m,n)} k!\binom{m}{k}\binom{n}{k}H_{m+n-2k}(x),$$

which was first proved in 1918 by Nielson [11].

We take the domain space of T_a to be the set of all formal power series $\sum_{k=0}^{\infty} a_k t^k$, where a_k's are complex numbers. We will denote this space by F_1. The range space of T_a which we will denote by F_2, is the set of all formal Newton series $\sum_{k=0}^{\infty} b_k s^{(k)}$ where b_k is a complex number. Let addition and multiplication on F_1 be defined by

$$\alpha+\beta=\sum_{k=0}^{\infty}(a_k+b_k)t^k,$$

$$\alpha\beta=\sum_{n=0}^{\infty}\sum_{k=0}^{n} a_{n-k}b_k t^n,$$

where $\alpha=\sum_{k=0}^{\infty} a_k t^k$ and $\beta=\sum_{k=0}^{\infty} b_k t^k$. See Niven [12] for a discussion of the algebra of this integral domain. As usual, we will define equality on F_2 by

$$\sum_{k=0}^{\infty} a_k s^{(k)}=\sum_{k=0}^{\infty} b_k s^{(k)} \quad \text{iff} \quad a_k=b_k \quad \text{for } 0,1,2,\cdots.$$

It is easy to show that in F_2

$$\sum_{k=0}^{\infty} a_k s^{(k)}=\sum_{k=0}^{\infty} b_k s^{(k)} \quad \text{iff} \quad \sum_{k=0}^{\infty} a_k n^{(k)}=\sum_{k=0}^{\infty} b_k n^{(k)}$$

for $n=0,1,2,3,\cdots$.

From now on, we will denote equality in F_2 by $\doteqdot$. Let us define addition in F_2 in a manner similar to what we did in F_1. That is,

$$\sum_{k=0}^{\infty} a_k s^{(k)}+\sum_{k=0}^{\infty} b_k s^{(k)} \doteqdot \sum_{k=0}^{\infty}(a_k+b_k)s^{(k)}.$$

When we try to define a multiplication on F_2, such that T_a as defined by (3.2) is an isomorphism between F_1 and F_2, we encounter the annoying fact that $s^{(k)}s^{(n)}\neq s^{(n+k)}$. For this reason, we introduce a set of operators

$$\mathcal{O}=\left\{\sum_{k=0}^{\infty} a_k\left(\frac{sE^{-1}}{a}\right)^k \middle| a \text{ and } a_k \text{ are complex numbers and } a\neq 0\right\},$$

where E^{-1} is the backward shift operator acting on s and defined on F_2. We note that $(sE^{-1}/a)^k=s^{(k)}E^{-1}/a^k$ and $\mathcal{O}$ is a set of operators that map F_2 into F_2 by means of the following formula,

$$\begin{aligned}\sum_{n=0}^{\infty} a_n\left(\frac{sE^{-1}}{a}\right)^n \sum_{k=0}^{\infty} b_k s^{(k)} &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} a_n b_k \frac{(sE^{-1})^n s^{(k)}}{a^n}\\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} a_n b_k \frac{s^{(n+k)}}{a^n}\\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{a_{n-k} b_k s^{(n)}}{a^{n-k}}.\end{aligned}$$

Again, by using Niven's [12] approach, $\mathcal{O}$ can be made into an integral domain by defining addition by

$$\sum_{k=0}^{\infty} a_k\left(\frac{sE^{-1}}{a}\right)^k + \sum_{k=0}^{\infty} b_k\left(\frac{sE^{-1}}{a}\right)^k = \sum_{n=0}^{\infty} (a_n+b_n)\left(\frac{sE^{-1}}{a}\right)^n.$$

and multiplication by composition, that is

$$\sum_{n=0}^{\infty} a_n\left(\frac{sE^{-1}}{a}\right)^n \cdot \sum_{k=0}^{\infty} b_k\left(\frac{sE^{-1}}{a}\right)^k = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_{n-k} b_k\left(\frac{sE^{-1}}{a}\right)^n.$$

It is easy to see that F_1 and $\mathcal{O}$ are isomorphic under the isomorphism $\mathcal{T}$ defined by

$$\mathcal{T}\left(\sum_{k=0}^{\infty} a_k t^k\right) = \sum_{k=0}^{\infty} a_k\left(\frac{sE^{-1}}{a}\right)^k.$$

Of course, we are interested in the elements of F_2 for it is these elements that will be used in finding finite difference analogues to Poli's formula.

The natural mapping between $\mathcal{O}$ and F_2 denoted by $\tilde{\mathcal{T}}$ is defined by

$$\begin{aligned}\tilde{\mathcal{T}}\left(\sum_{k=0}^{\infty} a_k\left(\frac{sE^{-1}}{a}\right)^k\right) &= \sum_{k=0}^{\infty} a_k\left(\frac{sE^{-1}}{a}\right)^k \circ 1\\ &= \sum_{k=0}^{\infty} \frac{a_k s^{(k)}}{a^k}.\end{aligned}$$

T_a can be written as

$$T_a = \tilde{\mathcal{T}} \circ \mathcal{T}. \tag{3.3}$$

It is easy to show that T_a as defined by (3.3) is an isomorphism between F_1 and F_2.

The integral representation of T_a is

$$\begin{aligned}T_a(f(t)) &= \frac{1}{\Gamma(-s)}\int_0^{\infty} e^{-t} t^{-s-1} f\left(\frac{-t}{a}\right) dt\\ &= {}_{-\infty}D_0^s\left[e^x f\left(\frac{x}{a}\right)\right],\end{aligned} \tag{3.4}$$

where ${}_{-\infty}D_0^s$ is the Liouville fractional derivative at 0 (see [15]), Re $(s)<0$ and $f(t)$ is a polynomial. Transformations related to the inverse of T_1 have been studied by many authors (see [4], [5] and [9]).

4. Finite difference analogues to Poli's formula. We have developed everything that is required to find the finite difference analogue to Poli's formula

$$yH_n(x) = \sum_{k=0}^{n} \binom{n}{k} t^{n-k} \frac{d^k y}{dt^k}, \tag{4.1}$$

where $\{H_n(x)\}_{n=0}^{\infty}$ is the Hermite polynomial set defined by

(4.2)
$$y := e^{xt-t^2/2} = \sum_{n=0}^{\infty} H_n(x)t^n/n!.$$

Let T_a be the transform as defined by (3.3) and $y(t)=\sum_{n=0}^{\infty} b_n t^n$ belongs to F_1 (see § 3). If $d^k y/dt^k$ is defined on F_1 by

$$\frac{d^k y}{dt^k} = \sum_{n=0}^{\infty} \frac{b_{n+k}(n+k)!}{n!} t^n$$

then

(4.3)
$$T_a\left(\frac{d^k y}{dt^k}\right) \doteqdot a^k \Delta^k T_a(y),$$

where Δ is the forward difference operator defined on F_2 by

$$\Delta\left(\sum_{n=0}^{\infty} b_n s^{(k)}\right) \doteqdot \sum_{n=0}^{\infty} b_{n+1}(n+1)s^{(n)}.$$

We also note that if $f(x, t)$ is the generating function for an Appell polynomial set $\{P_n(x)\}_{n=0}^{\infty}$, then

$$T_a(f(x, t)) \doteqdot \sum_{k=0}^{\infty} \frac{P_k(x)s^{(k)}}{a^k k!}$$
$$\doteqdot a^{-s} \sum_{k=0}^{\infty} \frac{P_k(x)a^{s-k}s^{(k)}}{k!}.$$

If $\{P_n(x)\}$ is an Appell polynomial set then by (2.8) this becomes

(4.4)
$$T_a(f(x, t)) \doteqdot a^{-s}\psi(a+x, s),$$

where $\psi(x, s)$ is any extension of $\{P_n(x)\}_{n=0}^{\infty}$.

By using (4.3) and (4.4), we take the T_a transform of both sides of (4.1), and we obtain

(4.5)
$$a^{-s}\mathcal{H}(a+x, s)H_n(x) \doteqdot \left\{\sum_{k=0}^{n} \binom{n}{k} s^{(k)} E_s^{-k} a^{n-2k} \Delta_s^{n-k}\right\} a^{-s}\mathcal{H}(a+x, s),$$

where $\mathcal{H}(x, s)$ is any extension of the Hermite polynomial set. By using the well-known fact that

$$\Delta_s^k = \sum_{i=0}^{k} \binom{k}{i} (-1)^i E_s^{k-i},$$

it is easy to show that for any function $f(s)$

(4.6)
$$\Delta_s^k a^{-s} f(s) = a^{-s-k}(E_s - a)^k f(s).$$

By using (4.6) and the fact that E_s^{-1} and Δ_s commute we have, from (4.5), that

(4.7)
$$\mathcal{H}(a+x, s)H_n(x) \doteqdot \sum_{k=0}^{n} \binom{n}{k} s^{(k)}(E-a)^{n-k}\mathcal{H}(a+x, s-k).$$

Ismail [8] generalized Poli's results (equation (4.1)) to Appell polynomial sets. He showed that if the Appell polynomial set $\{P_n(x)\}_{n=0}^{\infty}$ has the generating function

$$Y := A(t)\, e^{xt} = \sum_{n=0}^{\infty} P_n(x) \frac{t^n}{n!}, \tag{4.8}$$

where $A(t)$ is analytic at the origin, $A(0) \neq 0$ and (4.8) is true for t in a neighborhood of the origin and for all x, then

$$yP_n(x) = n! \sum_{k=0}^{n} \frac{b_{n-k}(t)}{k!} D_t^k y, \tag{4.9}$$

where $\{b_n(t)\}_{n=0}^{\infty}$ is given by

$$\sum_{n=0}^{\infty} b_n(t) u^n = \frac{A(t)A(u)}{A(t+u)}. \tag{4.10}$$

By taking the T_a transform of (4.9) we obtain

$$a^{-s}\psi(a+x, s)P_n(x) \doteqdot n! \sum_{k=0}^{n} \frac{b_{n-k}(a^{-1}sE_s^{-1})}{k!} a^k \Delta^k a^{-s}\psi(a+x, s), \tag{4.11}$$

where $\psi(x, s)$ is any extension of $\{P_n(x)\}_{n=0}^{\infty}$.

In the case of the Euler polynomial set $\{E_n(x)\}_{n=0}^{\infty}$, equation (4.9) has the form

$$yE_n(x) = \sum_{k=0}^{n} \binom{n}{k} E_k(0) D^{n-k} y - \frac{2e^t}{e^t+1} \sum_{k=1}^{n} \binom{n}{k} E_k(0) D^{n-k} y, \tag{4.12}$$

where

$$y := \frac{2e^{xt}}{e^t+1} = \sum_{k=0}^{\infty} \frac{E_k(x)t^k}{k!}.$$

By taking the T_a transform of both sides of (4.8) we obtain

$$\begin{aligned} a^{-s}\mathscr{E}(x+a, s)E_n(x) \doteqdot{}& \sum_{k=0}^{n} \binom{n}{k} E_k(0) a^{n-k} \Delta^{n-k} a^{-s} \mathscr{E}(x+a, s) \\ & - \sum_{k=0}^{\infty} \frac{E_k(1)s^{(k)}E^{-k}}{a^k k!} \sum_{i=1}^{n} \binom{n}{i} E_i(0) a^{n-i} \Delta^{n-i} a^{-s} \mathscr{E}(x+a, s), \end{aligned} \tag{4.13}$$

where $\mathscr{E}(x, s)$ is any extension of the Euler polynomial set, E^{-1} is the unit backward shift operator acting on s and $\Delta = E - 1$. By using (4.6) this becomes

$$\begin{aligned} \mathscr{E}(x+a, s)E_n(x) \doteqdot{}& \sum_{k=0}^{n} \binom{n}{k} E_k(0)(E-a)^{n-k} \mathscr{E}(x+a, s) \\ & - \sum_{k=0}^{\infty} \frac{E_k(1)s^{(k)}}{k!} \sum_{i=1}^{n} \binom{n}{i} E_i(0)(E-a)^{n-i} \mathscr{E}(x+a, s-k). \end{aligned} \tag{4.14}$$

By using the same method a formula similar to (4.9) can be obtained for the Bernoulli polynomial sets, but it is much more complicated and is thus omitted.

5. Analytic considerations. We recall from § 3 that the symbol "$\doteqdot$" was used for equality in F_2. In this section, we wish to consider when "$\doteqdot$" can be replaced by "$=$".

From the definition of "$\doteqdot$" we know that when s is restricted to a nonnegative integer, then $F(s) \doteqdot G(s)$ if and only if $F(s) = G(s)$. Equation (4.7) becomes

$$H_m(x+a)H_n(x) = \sum_{k=0}^{\min(m,n)} \binom{n}{k}\binom{m}{k} k!(E-a)^{n-k} H_{m-k}(x+a), \tag{5.1}$$

where E is the forward unit shift operator acting on the m of $H_{m-k}(x+a)$, and $\{H_n(x)\}_{n=0}^{\infty}$ is the Hermite polynomial set as defined by (4.2). It is interesting to note that by letting $a=0$ we get the well-known result due to Nielson [11], namely

$$H_m(x)H_n(x) = \sum_{k=0}^{\min(m,n)} \binom{n}{k}\binom{m}{k} k! H_{m+n-2k}(x).$$

In a similar manner (4.11) becomes

$$P_m(a+x)P_n(x) = n!a^m \sum_{k=0}^{n} \frac{b_{n-k}(a^{-1}sE_s^{-1})}{k!} a^k \Delta_m^k a^{-m} P_m(a+x) \tag{5.2}$$

and (4.14) becomes

$$\begin{aligned} E_m(x+a)E_n(x) &= \sum_{k=0}^{n} \binom{n}{k} E_k(0)(E-a)^{n-k} E_m(x+a) \\ &\quad - \sum_{k=0}^{m} \binom{m}{k} E_k(1) \sum_{i=1}^{n} \binom{n}{i} E_i(0)(E-a)^{n-i} E_{m-k}(x+a). \end{aligned} \tag{5.3}$$

Equations (5.1), (5.2) and (5.3) are analogous to linearization of the product of two polynomials that have been studied by many authors (see [3, Lecture 5]).

We know that the Hermite function $H_v(x)$, defined by

$$H_v(x) = e^{x^2/4} D_v(x),$$

where $D_v(x)$ is the parabolic cylinder function defined in [6, vol. 2, p. 116], is also an extension of the Hermite polynomial $\{H_n(x)\}_{n=0}^{\infty}$. By using this fact and (4.7), we obtain

$$H_v(a+x)H_n(x) \doteqdot \sum_{k=0}^{n} \binom{n}{k} v^{(k)} (E-a)^{n-k} H_{v-k}(x+a). \tag{5.4}$$

It is well known that $H_v(x)$ is an entire function in both the variable x and the parameter v [10, p. 285]. We will show by mathematical induction on n that "$\doteqdot$" in (5.4) can be replaced by "$=$" to obtain

$$H_v(x+a)H_n(x) = \sum_{k=0}^{n} \binom{n}{k} v^{(k)} E^{-k} (E-a)^{n-k} H_v(x+a), \tag{5.5}$$

for v and x arbitrary complex numbers and $n = 0, 1, 2, \cdots$. By direct substitution, it is easy to see that (5.5) is true for $n = 0, 1$. Now make the induction hypothesis that (5.5) is true for $n = 0, 1, 2, \cdots, m$, and by using the three term recursion relation for $H_v(x+a)$ and $H_{m+1}(x)$ we obtain

$$\begin{aligned} H_v(x+a)H_{m+1}(x) &= H_{v+1}(x+a)H_m(x) + vH_{v-1}(x+a)H_m(x) \\ &\quad - aH_v(x+a)H_m(x) - mH_v(x+a)H_{m-1}(x). \end{aligned}$$

By using the induction hypothesis, we obtain

$$\begin{aligned}
H_v(x+a)H_{m+1}(x) &= (E-a)^{m+1}H_v(x+a)+(vE^{-1})^{m+1}H_v(x+a) \\
&\quad + \sum_{k=1}^{m}\left\{\binom{m}{k}(v+1)^{(k)}E^{-k+1}(E-a)^{m-k}\right. \\
&\qquad \left. -\frac{m!}{(m-k)!(k-1)!}v^{(k-1)}E^{-k+1}(E-a)^{m-k}\right\}H_v(x+a) \\
&\quad + v\sum_{k=1}^{m}\binom{m}{k-1}(v-1)^{(k-1)}E^{-k+1}(E-a)^{m-k+1}H_{v-1}(x+a) \\
&\quad - a\sum_{k=1}^{m}\binom{m}{k}(vE^{-1})^{k}(E-a)^{m-k}H_v(x+a) \\
&= (E-a)^{m+1}H_v(x+a)+(vE^{-1})^{m+1}H_v(x+a) \\
&\quad + \sum_{k=1}^{m}\left\{\binom{m}{k}v^{(k)}E^{-k+1}(E-a)^{m-k}\right. \\
&\qquad \left. -a\binom{m}{k}(vE^{-1})^{k}(E-a)^{m-k}\right\}H_v(x+a) \\
&\quad + \sum_{k=1}^{m}\binom{m}{k-1}v^{(k)}E^{-k}(E-a)^{m+1-k}H_v(x+a) \\
&= (E-a)^{m+1}H_v(x+a)+(vE^{-1})^{m+1}H_v(x+a) \\
&\quad + \sum_{k=1}^{m}\left(\binom{m}{k}+\binom{m}{k-1}\right)v^{(k)}E^{-k}(E-a)^{m+1-k}H_v(x+a) \\
&= \sum_{k=0}^{m+1}\binom{m+1}{k}(vE^{-1})^{k}(E-a)^{m+1-k}H_v(x+a).
\end{aligned}$$

It follows directly from (5.5) that for each zero x_i, $i=1,2,\cdots,n$, of $H_n(x)$, $H_v(x_i+a)$ is a solution of the finite difference equation

$$\sum_{k=0}^{n}\binom{n}{k}v^{(k)}E^{-k}(E-a)^{n-k}y(v)=0.$$

6. Inverse formulas. In this section we shall apply the T_a-transform technique as previously developed to obtain inverse type formulas for (4.7), (5.1) and (5.5). That is, we will show that

$$(6.1)\qquad (E-a)^n\mathscr{H}(x+a,s) \doteqdot \sum_{k=0}^{n}(-1)^k\binom{n}{k}s^{(k)}\mathscr{H}(x+a,s-k)H_{n-k}(x)$$

$$(6.2)\qquad (E-a)^nH_m(x+a)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{m}{k}k!H_{m-k}(x+a)H_{n-k}(x)$$

and

$$(6.3)\qquad (E-a)^nH_v(x+a)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}v^{(k)}H_{v-k}(x+a)H_{n-k}(x),$$

where $\{H_n(x)\}_{n=0}^{\infty}$ is the Hermite polynomial set, $H_v(x)$ is the Hermite function and

$\mathscr{H}(x, s)$ is any extension of the Hermite polynomial set. To do this we note that if

$$y := e^{xt-t^2/2} = \sum_{k=0}^{\infty} H_k(x)\frac{t^k}{k!},$$

then

$$\begin{aligned}\frac{d^n y}{dt^n} &= H_n(x-y)y \\ &= \sum_{k=0}^{n} (-1)^k\binom{n}{k} H_{n-k}(x)t^k y.\end{aligned}$$

Now take the T_a transform of both sides of this equation to obtain

$$a^n \Delta^n a^{-s}\mathscr{H}(a+x, s) \doteqdot \sum_{k=0}^{n} (-1)^k \binom{n}{k} H_{n-k}(x)\left(\frac{sE^{-1}}{a}\right)^k a^{-s}\mathscr{H}(a+x, s).$$

By using (4.6) on this equation we obtain (6.1). Equations (6.2) and (6.3) follow from (6.1) by using the same technique as was previously used to obtain (5.1) and (5.5) from (4.7).

If we let $a = 0$ in (6.2) we obtain a result that was proved by G. N. Watson [17]. L. Carlitz, in a private communication, has obtained (6.2) and (6.3). He proved the former by a generating function technique, similar to what Watson used, and the latter by an induction argument similar to what we used to obtain (5.5).

Acknowledgment. I wish to express my thanks to Richard Askey, L. Carlitz, and Mourad Ismail for their helpful comments.

REFERENCES

[1] WM. R. ALLAWAY, *The representation of orthogonal polynomials in terms of a differential operator*, J. Math. Anal. Appl., 56 (1976), pp. 288–293.
[2] P. APPELL, *Sur une class de polynomes*, Ann. Sci. École Norm. Sup. (Paris), 9 (1880), pp. 119–144.
[3] R. ASKEY, *Orthogonal Polynomials and Special Functions*, Society for Industrial and Applied Mathematics, Philadelphia, 1975.
[4] L. BERG, *Zur operatorenrechnung fur Funktionen einer diskreten Veränderlichen*, Studia Math., 20 (1961), pp. 227–243.
[5] V. A. DITKIN AND A. P. PRUDNIKOV, *Operational calculus*, Progress in Mathematics, 1 (1968), pp. 1–74.
[6] A. ERDÉLYI ET AL, *Higher Transcental Functions*, vols. 1, 2, McGraw-Hill, New York, 1953.
[7] A. K. HARADZE, *On the representation of ultraspherical polynomials by a differential operator containing the generating function of the polynomials*, Soviet Math. Dokl., 5 (1964), pp. 1191–1193.
[8] M. E. H. ISMAIL, *On solving certain differential equations with variable coefficients*, Aequationes Math., to appear.
[9] ———, *A family of operational calculi*, Ph.D. dissertation, University of Alberta, Edmonton, 1974.
[10] N. N. LEBEDEV, *Special functions and their application*, Prentice-Hall, Englewood Cliffs, NJ, 1965.
[11] N. NIELSON, *Recherches sur les polynomes d'Hermite*, Det. Kgl. Danske Viden, Selskab. Math. fys. Medd. I, 6 (1918), pp. 1–78.
[12] I. NIVEN, *Formal Power Series*, Amer. Math. Monthly, 76 (1969), pp. 871–889.
[13] L. POLI, *Polynomes d'Hermite de equation differentielles*, Mathesis, 63 (1954), pp. 319–325.
[14] E. D. RAINVILLE, *Special Functions*, Macmillan, New York, 1960.
[15] B. ROSS, *Fractional Calculus and its Applications*, Springer-Verlag, New York, 1975.
[16] I. M. SHEFFER, *Some properties of polynomial sets of type zero*, Duke Math. J., 5 (1939), pp. 590–622.
[17] G. N. WATSON, *A note on the polynomials of Hermite and Laguerre*, J. London Math. Soc., 13 (1938), pp. 29–32.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0005 $01.00/0

SOME PROPERTIES OF SOLUTIONS OF $(r(t)\psi(x)x')' + a(t)f(x) = 0$*

W. E. MAHFOUD† AND S. M. RANKIN†

Abstract. The equation

$$(r(t)\psi(x)x')' + a(t)f(x) = b(t) \tag{1}$$

is considered, where a, f, and b are continuous, r and ψ are continuously differentiable, $r(t)>0$, and $\psi(x)$ and $xf(x)$ are positive for $x \neq 0$. It is shown by a transformation of variables that (1) can be reduced to

$$x'' + a(t)f(x) = b(t) \tag{2}$$

and hence results concerning (2) can be extended without difficulty to (1). Illustrative results on uniqueness, continuation, and oscillation of solutions of $(r(t)\psi(x)x')' + a(t)f(x) = 0$ are obtained and the case $\psi(0) = 0$ is discussed.

Introduction. Consider the equation

$$(r(t)\psi(x)x')' + a(t)f(x) = b(t) \tag{1}$$

where $r: [0, \infty) \to (0, \infty)$; a, $b: [0, \infty) \to (-\infty, \infty)$, and ψ, $f: (-\infty, \infty) \to (-\infty, \infty)$. We assume a, f, and b are continuous, r and ψ are continuously differentiable, and $xf(x) > 0$ for $x \neq 0$.

By a solution of (1) at $t_0 \geqq 0$ is meant a function $x: [t_0, t_1) \to (-\infty, \infty)$, $t_0 < t_1$, which satisfies (1) for all $t \in [t_0, t_1)$. We assume the existence of solutions of (1) at t_0 for every $t_0 \geqq 0$. A solution $x(t)$ of (1) at t_0 is said to be continuable if $x(t)$ exists for all $t \geqq t_0$. A continuable solution $x(t)$ of (1) is said to be oscillatory if $x(t)$ has zeros for arbitrarily large t and nonoscillatory if there exists $t^* \geqq 0$ such that $x(t) \neq 0$ for all $t \geqq t^*$. Equation (1) is said to be oscillatory if every continuable solution of (1) is oscillatory.

Equation (1) has been discussed in [4], [5], and [7] and some results have been extended from the equation

$$x'' + a(t)f(x) = b(t) \tag{2}$$

to equation (1). In this paper, we show, that under appropriate conditions on r and ψ, equation (1) can be reduced to (2) and hence results such as that of [4], [5], and [7] can easily be extended from (2) to (1).

Main results. Let $R = (-\infty, \infty)$ and define $h: R \to R$ by

$$h(x) = \int_0^x \psi(u)\, du. \tag{3}$$

If ψ is assumed to satisfy $\psi(x) > 0$ for $x \neq 0$, then, clearly, h is increasing, continuously differentiable, and $xh(x) > 0$ for $x \neq 0$. Furthermore, the function $g: h(R) \to R$ defined by

$$g = f \circ h^{-1} \tag{4}$$

is continuous and satisfies $xg(x) > 0$ for $x \neq 0$.

* Received by the editors November 24, 1976, and in final revised form May 31, 1977.
† Department of Mathematics, Murray State University, Murray, Kentucky 42071.

THEOREM 1. *Suppose* $\psi(x)>0$ *for* $x\neq 0$. *If* $x=\phi(t)$ *is a solution of* (1) *on some interval* I, *then* $z=h\circ\phi(t)$ *is a solution of the equation*

$$(r(t)z')'+a(t)g(z)=b(t) \tag{5}$$

where h *and* g *are as defined in* (3) *and* (4).

Conversely, if $z=\lambda(t)$ *is a nontrivial solution of* (5) *on some interval* I, *then* $x=h^{-1}\circ\lambda(t)$ *is a nontrivial solution of* (1) *on some interval* $J\subset I$. *If, in addition* $\psi(0)\neq 0$ *or* $z(t)\neq 0$ *for all* $t\in I$, *then* $J=I$.

Proof. It is easy to verify, by use of (3) and (4), that z and x defined above are respectively solutions of (5) and (1) and that $z'=\psi(x)x'$ and $J\subset I$. In fact, this equality together with (3) shows that if $\psi(0)=0$ and z vanishes at some $t_1\in I$, then x' may not exist at t_1; in this case, J is a proper subset of I. The last statement of the theorem follows also at once.

COROLLARY 1. *Every oscillatory solution of* (1) *generates an oscillatory solution of* (5).

COROLLARY 2. *There is a one-to-one correspondence between the nonoscillatory solutions of* (1) *and* (5).

COROLLARY 3. *Suppose* $b(t)\equiv 0$ *and* $\psi(0)=0$. *If the solution* $z(t)$ *of* (5), *with* $z(t_0)=z'(t_0)=0$, *for all* $t_0\geqq 0$, *is unique, then* (1) *has no nontrivial oscillatory solutions. If, in addition,* (5) *is oscillatory, then* (1) *has no nontrivial continuable solutions.*

Proof. Suppose $x(t)$ is a nontrivial solution of (1); then, by Theorem 1 and (3), $z(t)=h(x(t))$ is a nontrivial solution of (5) such that

$$z'(t)=\psi(x(t))x'(t). \tag{6}$$

By (3), $x(t)$ vanishes if and only if $z(t)$ vanishes. Since $\psi(0)=0$, it follows from (6) that if $z(t_1)=0$ for some $t_1\geqq 0$ then $z'(t_1)=0$ and hence by the uniqueness assumption $z(t)\equiv 0$ and so is $x(t)$, a contradiction.

Example 1. Consider the equation

$$(x^2x')'+\tfrac{1}{3}x^3=0 \tag{7}$$

and let $x(t)$ be a solution of (7); then $z(t)=[x(t)]^3/3$ is a solution of the linear equation

$$z''+z=0.$$

Thus $z(t)=A\sin(t+B)$ for some constants A and B and hence $x(t)=C\sin^{1/3}(t+B)$. As $x'(t)$ does not exist for $t=k\pi-B$, $k=1,2,\cdots$, then (7) has no nontrivial continuable solution.

Example 2. Consider the equation

$$(x^{2n}x')'+[k/t^2]x^{2n+1}=0 \tag{8}$$

where n is a positive integer and k is a constant. Here, $h(x)=x^{2n+1}/(2n+1)$ and the associated equation is

$$z''+[k(2n+1)/t^2]z=0. \tag{9}$$

As (9) is oscillatory for $k>1/(8n+4)$ and nonoscillatory for $k\leqq 1/(8n+4)$, then, by Corollary 3, equation (8) can have continuable solutions only when $k\leqq 1/(8n+4)$ and hence by Corollary 2, no nontrivial solution of (8) is oscillatory.

We now consider the unforced equation

$$(r(t)\psi(x)x')+a(t)f(x)=0 \tag{10}$$

subject to the additional condition $\psi(x)>0$ for $x\neq 0$. The transformation h in (3) reduces (10) to

$$(r(t)z')' + a(t)g(z) = 0 \tag{11}$$

where g is defined in (4). If we now let

$$s = \nu(t) = \int_0^t [1/r(u)]\,du \tag{12}$$

equation (11) is reduced to

$$\ddot{w} + R(s)A(s)g(w) = 0 \tag{13}$$

where $R(s) = r(t(s))$, $A(s) = a(t(s))$, $w(s) = z(t(s))$, and $\dot{} = d/ds$.

THEOREM 2. *Suppose* $\psi(x) > 0$ *for* $x \neq 0$, $a(t) < 0$ *on* $[t_1, t_2]$, $t_1 \geqq 0$, *and* $x(t)$ *is a solution of* (10) *on* $[t_1, t_2]$ *such that* $x(t_1) = x'(t_1) = 0$. *Then* $x(t) = x'(t) = 0$ *for* $t \in [t_1, t_2]$ *is and only if*

$$\int_{0^+}^{1} \psi(x)[F(x)]^{-1/2}\,dx = \infty \tag{i}$$

and

$$\int_{0^-}^{-1} \psi(x)[F(x)]^{-1/2}\,dx = -\infty \tag{ii}$$

where $F(x) = \int_0^x \psi(u)f(u)\,du$.

Proof. We may assume without loss of generality that $r(t) \equiv 1$ since the transformation ν in (12) is one-to-one. Let $z(t) = h(x(t))$; then, by Theorem 1, $z(t)$ is a solution of (11) such that $z(t_1) = z'(t_1) = 0$. Let $G(v) = \int_0^v g(u)\,du$; then

$$G(v) = \int_0^v f(h^{-1}(u))\,du = \int_0^{h^{-1}(v)} f(u)\psi(u)\,du = F(h^{-1}(v))$$

and hence, for every $\varepsilon > 0$, we have $\int_{\pm\varepsilon}^{\pm 1} [G(v)]^{-1/2}\,dv = \int_{\pm\varepsilon}^{\pm 1} [F(h^{-1}(v))]^{-1/2}\,dv$. Let $w = h^{-1}(v)$; then

$$\int_{\pm\varepsilon}^{\pm 1} [G(v)]^{-1/2}\,dv = \int_{h^{-1}(\pm\varepsilon)}^{h^{-1}(\pm 1)} [F(w)]^{-1/2}\psi(w)\,dw.$$

By (3), $\varepsilon \to 0$ if and only if $h^{-1}(\pm\varepsilon) \to 0$. Hence (i) and (ii) hold if and only if $\int_{0\pm}^{\pm 1} [G(v)]^{-1/2}\,dv = \pm\infty$. Thus, by [3, Thm. 4], $z(t) \equiv z'(t) \equiv 0$ on $[t_1, t_2]$ and hence, by (3), $x(t) \equiv x'(t) \equiv 0$ on $[t_1, t_2]$ if and only if (i) and (ii) hold. The proof is now complete.

The next result is another illustration of Theorem 1 by which we extend a noncontinuation result of Burton and Grimmer [2] to (10) under the additional assumption

$$\int_0^{\pm\infty} \psi(x)\,dx = \pm\infty. \tag{14}$$

THEOREM 3. *Suppose* $\psi(x) > 0$ *for* $x \neq 0$, (14) *holds and* $a(t) < 0$ *on* $[t_1, t_2]$, $t_1 \geqq 0$. *Then* (10) *has a solution* $x(t)$ *at* t_1 *such that* $\lim_{t \to T} |x(t)| = \infty$ *for some* $T \in (t_1, t_2]$ *if and only if*

$$\int_0^{\infty} \psi(x)[1 + F(x)]^{-1/2}\,dx < \infty \tag{i}$$

or

$$\text{(ii)} \qquad \int_0^{-\infty} \psi(x)[1+F(x)]^{-1/2}\,dx > -\infty$$

where $F(x)=\int_0^x \psi(u)f(u)\,du$.

Proof. Let $r=1$ and define G as in the proof of Theorem 2. Then it follows that

$$\int_0^v [1+G(u)]^{-1/2}\,du = \int_0^{h^{-1}(v)} [1+F(w)]^{-1/2}\psi(w)\,dw.$$

By (3) and (14) $v\to\pm\infty$ if and only if $h^{-1}(v)\to\pm\infty$. Thus, by [2, Thm. 2], equation (11) has a solution $z(t)$ at t_1 such that $\lim_{t\to T^-}|z(t)|=\infty$, for some $T\in(t_1, t_2]$, if and only if (i) or (ii) holds. It is not hard to see from the proof of [2, Thm. 2] that $z(t)$ can be chosen so that $z(t_1)\neq 0$. In fact that has been shown in [6]. In this case, by Theorem 1, $z(t)$ generates a solution $x(t)$ of (10) at t_1 and, by (3), $\lim_{t\to T^-}|x(t)|=\infty$ if and only if $\lim_{t\to T^-}|z(t)|=\infty$. The proof is now complete.

Remark. Part (i) of Theorem 3 has been proved in [4] without condition (14) by using Burton and Grimmer's argument. Since the main interest in studying (10) is indeed equations of the form $(x^\alpha x')'+a(t)f(x)=0$, α a nonnegative integer, we feel that condition (14) is not a significant restriction. We also point out that the author in [4] assumed that $x\psi(x)>0$ for $x\neq 0$ and claimed that part (ii) of the theorem holds by an argument similar to that of part (i). Apparently, he overlooked the fact that the assumption $x\psi(x)>0$ for $x\neq 0$ implies that $F(x)<0$ when $x<0$ and hence the integral condition in (ii) may not make sense. Even though $[1+|F(x)|]^{-1/2}$ is defined, the integral $\int_0^{-\infty}\psi(x)[1+|F(x)|]^{-1/2}\,dx$ is positive and hence (ii) is satisfied automatically. Thus, the difficulty may not be overcome unless we assume $\psi(x)>0$ for $x\neq 0$. This observation applies also to [4, Thm. 3].

The following theorems extend some results in [8] and [9] to equation (10) and consequently improve the result in [7]. We also point out that the requirement that f be differentiable implies that the function g defined by (4) is also differentiable whenever $\psi(x)\neq 0$.

THEOREM 4. *Suppose* $\psi(x)>0$ *for* $x\neq 0$ *and the following conditions are satisfied*

$$\text{(i)} \qquad \int^\infty [1/r(u)]\,du=\infty \quad \text{and} \quad f'(x)\geqq 0,$$

$$\text{(ii)} \qquad \left|\int_0^{\pm\infty}\psi(x)\,dx\right|<\infty,$$

and

$$\text{(iii)} \qquad \int^\infty a(t)\int_0^t [1/r(u)]\,du\,dt=\infty.$$

Then every continuable solution of (1) *is oscillatory.*

Proof. Let $x(t)$ be a solution of (10) on $[t_0,\infty)$, $t_0\geqq 0$, and let $X(s)=x(t(s))$, where t and s are related by (12). Then, by (i) and Theorem 1, $z(s)=h(X(s))$ is a solution of (13) on $[s_0,\infty)$, where $s_0=\nu(t_0)$ and h is defined by (3). By (ii), $z(s)$ is bounded on $[s_0,\infty)$ and by [8, Thm. 2.2] bounded solutions of (13) oscillate if $\int^\infty sR(s)A(s)\,ds=\infty$. Since, by the transformations $s=\nu(t)$, this integral condition is precisely the integral condition in (iii), then, by [8, Thm. 2.2] $z(s)$ oscillates and so does $x(t)$. The proof is now complete.

Remark. It is possible by Corollary 3 that the only continuable solution of (10) is the trivial solution.

THEOREM 5. *Suppose* $\psi(x)>0$ *for* $x\neq 0$ *and the following conditions are satisfied*

$$\text{(i)} \qquad \int^{\infty} [1/r(u)]\,du=\infty \quad \textit{and} \quad f'(x)\geqq 0,$$

$$\text{(ii)} \qquad \int_{\pm 1}^{\pm\infty} [\psi(x)/f(x)]\,dx<\infty,$$

and

$$\text{(iii)} \qquad \int^{\infty} a(t)\int_0^t [1/r(u)]\,du\,dt=\infty.$$

Then every continuable solution of (10) *is oscillatory.*

Proof. Let $x(t)$ be a solution of (10) on $[t_0,\infty)$, $t_0\geqq 0$. Then, as in the proof of Theorem 4, $z(s)$ is a solution of (13) on $[s_0,\infty)$. We assume $\int_0^{\pm\infty}\psi(x)\,dx=\pm\infty$; otherwise, the result follows from Theorem 4. Let $v=\int_0^u \psi(w)\,dw$; then for $\tau>0$

$$\int_1^{\tau} [\psi(u)/f(u)]\,du=\int_{h(1)}^{h(\tau)} [1/f\circ h^{-1}(v)]\,dv=\int_{h(1)}^{h(\tau)} [1/g(v)]\,dv.$$

As $\tau\to\infty$ if and only if $h(\tau)\to\infty$, then (ii) is satisfied if and only if $\int_{\pm 1}^{\pm\infty}[1/g(v)]\,dv<\infty$. Hence by [8, Thm. 2.1], $z(s)$ oscillates and hence $x(t)$. The proof is now complete.

Remark. Condition (iii) in Theorem 5 can be replaced by the slightly more general condition

$$\text{(iv)} \qquad \int^{\infty} a(t)\left(\int^t [1/r(u)]\,du\right)^{\alpha} dt=\infty, \qquad 0\leqq\alpha\leqq 1,$$

if we use [9, Corollary 1] and Theorem 1. See also [9, pp. 305].

Although results of (2) seem to extend to (1), the effect of the singularity $\psi(0)=0$ on the behavior of solutions of (1) is quite clear from Examples 1 and 2. The next Theorem describes the oscillatory solutions of (10) when $\psi(0)=0$.

THEOREM 6. *Suppose* $x(t)$ *is a solution of* (10) *on* $[t_1,t_2]$ *such that* $x(t_1)=x(t_2)=0$. *If* $\psi(0)=0$ *and* $a(t)$ *does not change sign on* $[t_1,t_2]$, *then* $x(t)=0$ *on* $[t_1,t_2]$.

Proof. Suppose there exists $t^*\in(t_1,t_2)$ such that $x(t^*)\neq 0$; then there exists $T_1, T_2\in[t_1,t_2]$ such that $x(T_1)=x(T_2)=0$ and $x(t)\neq 0$ on (T_1,T_2). Integrate (10) from T_1 to T_2 to obtain

$$\int_{T_1}^{T_2} [r(t)\psi(x(t))x'(t)]'\,dt+\int_{T_1}^{T_2} a(t)f(x(t))\,dt=0.$$

As $x(T_1)=x(T_2)=0$ and $\psi(0)=0$, then the first integral is zero and hence $\int_{T_1}^{T_2} a(t)f(x(t))\,dt=0$. As the integrand is of one sign and a continuous function of t, then $a(t)f(x(t))=0$ for all $t\in[T_1,T_2]$ and hence $x(t)\equiv 0$ on $[T_1,T_2]$, a contradiction.

COROLLARY. *Suppose* $\psi(0)=0$ *and* $a(t)$ *does not change sign. Then the only oscillatory solutions of* (10) *are the solutions which are eventually identically zero.*

Acknowledgment. The authors would like to thank the referees for their valuable suggestions.

REFERENCES

[1] L. E. BOBISUD, *Oscillation of nonlinear second-order equations*, Proc. Amer. Math. Soc., 23 (1969), pp. 501–505.

[2] T. BURTON AND R. GRIMMER, *On continuability of solutions of second order differential equations*, Ibid., 29 (1971), pp. 277–283.

[3] ———, *On the asymptotic behavior of solutions of* $x''+a(t)f(x)=e(t)$, Pacific J. Math., 41 (1972), pp. 43–55.

[4] VADIM KOMKOV, *Continuability and estimates of solutions of* $(a(t)\psi(x)x')'+c(t)f(x)=0$, Ann. Polon. Math., 30 (1974), no. 2, pp. 125–137.

[5] ———, *On the behavior of solutions of* $(a(t)\psi(x)x')'+c(t)f(x)=0$, to appear.

[6] W. E. MAHFOUD, *Noncontinuation and uniqueness of solutions of the delay equation* $x^{(n)}(t)+a(t)\phi(x(t))f(x(q(t)))=0$, J. Math. Anal. Appl., 59 (1977), pp. 617–627.

[7] V. A. STAIKOS AND Y. G. SFICAS, *Oscillations for forced second order nonlinear differential equations*, Atti. Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez. I^a, 55, no. 8 (1973), pp. 25–30.

[8] C. C. TRAVIS, *A note on second order nonlinear oscillations*, Math. Japon., 18 (1973), pp. 261–264.

[9] JAMES S. W. WONG, *Oscillation theorems for second order nonlinear differential equations*, Bull. Inst. Math. Acad. Sinica, 3 (1975), pp. 283–309.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0006 $01.00/0

GROWTH AND OSCILLATION PROPERTIES OF SECOND ORDER LINEAR DIFFERENCE EQUATIONS*

WILLIAM T. PATULA†

Abstract. This paper establishes the existence of recessive and dominant solutions for nonoscillatory and certain types of oscillatory second order homogeneous linear difference equations. Growth properties concerning these solutions are established. This information is then used to obtain two limit point results. Several sufficient conditions for oscillation are also presented.

1. Introduction and preliminary remarks. We will be considering linear homogeneous second order difference equations of the form

$$c_n x_{n+1} + c_{n-1} x_{n-1} = b_n x_n, \qquad c_n > 0. \tag{1}$$

Usually, but not always, we will also assume $b_n \geqq 0$. This is not nearly as restrictive as it may first appear to be. By means of the substitution $x_n = (-1)^n y_n$, the equation $c_n y_{n+1} + c_{n-1} y_{n-1} = d_n y_n$, $d_n \leqq 0$, is equivalent to (1) with $b_n = -d_n \geqq 0$.

Motivated by recent results in Hinton and Lewis [6], we will investigate growth properties of certain solutions of (1). The particular form of the equation (1) appears in [6] and also Atkinson [1, p. 15]. Note that by means of the substitution $p_n = b_n - c_n - c_{n-1}$, equation (1) is equivalent to the self-adjoint form

$$-\Delta(c_{n-1}\Delta x_{n-1}) + p_n x_n = 0, \tag{2}$$

where the forward difference operator Δ is defined by $\Delta x_n = x_{n+1} - x_n$.

The analogue of (2) in the continuous case is the differential equation

$$(-r(t)x'(t))' + q(t)x(t) = 0, \qquad r(t) > 0, \quad t \geqq a. \tag{3}$$

Many of the properties usually associated with (3) also hold for the difference equation (1). For example, specifying two consecutive values x_k, x_{k+1} of a solution $x = \{x_n\}$, $n \geqq 0$, uniquely determines all other values x_n. Equation (1) has two linearly independent solutions, say $u = \{u_n\}$ and $v = \{v_n\}$, such that $c_n(u_n v_{n+1} - u_{n+1} v_n) = 1$, for all n. A solution x of (1) will be called bounded if $|x_n| \leqq M$, for all n. A nontrivial solution will be called oscillatory if for any N, there exists a $k \geqq N$ such that $x_k x_{k+1} \leqq 0$. Oscillation can also be defined in terms of nodes. See [2, pp. 131 and 224]. The recurrence relation (1) will be called oscillatory if all solutions are oscillatory. However, if one solution of (1) oscillates, all solutions oscillate [2, p. 221]. For other interesting properties we refer the reader to the books [1] and [2].

We would first like to state a result of Hartman and Wintner [4] in a slightly different setting. For convenience, we include the proof.

LEMMA 1 (Hartman and Wintner [4]). *If* (1) *is such that any nontrivial solution x can have at most one value $x_k = 0$, then any two values x_n, x_m, $n \neq m$, uniquely determine the solution x.*

Proof. Let u and v be the solutions defined by $u_0 = 0$, $u_1 = 1$ and $v_0 = 1$, $v_1 = 0$. Then u and v are linearly independent. Consider the system of equations

$$x_m = \alpha_1 u_m + \alpha_2 v_m,$$

$$x_n = \alpha_1 u_n + \alpha_2 v_n.$$

* Received by the editors April 12, 1977, and in final revised form November 16, 1977.
† Department of Mathematics, Southern Illinois University, Carbondale, Illinois 62901.

If this system does not have a unique solution in terms of α_1 and α_2, then there exist values of α_1 and α_2, not both zero, such that $0=\alpha_1 u_m+\alpha_2 v_m=\alpha_1 u_n+\alpha_2 v_n$. However, this implies that the solution $(\alpha_1 u+\alpha_2 v)$ assumes the value 0 twice, a contradiction.

The next result is basically a lemma of Olver and Sookne [8], which we state in a slightly expanded form.

LEMMA 2 [8]. *Suppose* $|b_n|\geqq c_{n-1}+c_n$. *If* v *is a solution such that* $|v_{N+1}|\geqq|v_N|$, *for some integer* N, *then* $|v_{n+1}|\geqq|v_n|$, *for all* $n\geqq N$. *If there exists a sequence* $\{\varepsilon_n\}$ *of nonnegative numbers such that*

$$|b_n|\geqq(1+\varepsilon_n)c_n+c_{n-1} \quad \text{and} \quad \sum\varepsilon_n=\infty \tag{4}$$

then $|v_n|\to\infty$, *as* $n\to\infty$.

Proof. Use induction and assume it is true for some $n\geqq N+1$, that is, $|v_n|\geqq|v_{n-1}|$. Then

$$\begin{aligned}|v_{n+1}|&=|b_n v_n-c_{n-1}v_{n-1}|/c_n\\ &\geqq[(|b_n|-c_{n-1})/c_n]|v_n|\geqq|v_n|, \quad \text{for all } n\geqq N.\end{aligned} \tag{5}$$

If in addition we have (4), then (5) becomes

$$|v_{n+1}|\geqq(1+\varepsilon_n)|v_n|.$$

However, assuming $n\geqq N$, the above inequality implies that $v_{n+1}\geqq v_N\prod_j(1+\varepsilon_j)$, $N\leqq j\leqq n$, which means $|v_n|\to\infty$, as $n\to\infty$. This completes the proof.

Remark 1. We always assume $c_n>0$. If we also have $b_n\geqq c_n+c_{n-1}$, then the solution v defined by $v_0=1$, $v_1=1$ must have $v_{n+1}\geqq v_n\geqq 1$. That is, the absolute value signs in the preceding proof can be dropped. Clearly v is nonoscillatory, so that the condition $b_n-c_n-c_{n-1}\geqq 0$ is sufficient for nonoscillation. This is not surprising in view of equation (2) and the analogous result for the continuous case (3). This result can also be found in [6] and [2, p. 224]. Also, Lemma 2 implies the hypothesis of Lemma 1 is satisfied if $|b_n|\geqq c_n+c_{n-1}$.

Remark 2. Lemma 2 is sharp for the case of constant coefficients. Assume $b_n=b$ and $c_n=1$, for all n. Assuming $c_n=1$ is allowed because of linearity. Using the techniques as in [2, p. 125] for solving difference equations with constant coefficients, we see that all solutions of (1) are bounded if and only if $|b|<2$.

Our next result is elementary but useful.

LEMMA 3. *If there exists a subsequence* $b_{n_k}\leqq 0$, *where* $n_k\to\infty$ *as* $k\to\infty$, *then* (1) *is oscillatory.*

Proof. Suppose not. Then we may assume the existence of a solution x such that $x_n>0$, for all n sufficiently large. However, the left side of (1) will always be positive while the right side will be $\leqq 0$ for all values of n_k, a contradiction.

2. Properties of nonoscillatory solutions. In Olver and Sookne [8], the following definition is made.

DEFINITION. If there exist two linearly independent solutions u and v of (1) such that $u_n/v_n\to 0$, as $n\to\infty$, then u is called *recessive* or *sub-dominant* and v is called *dominant.*

If (1) is nonoscillatory, then the terms recessive and dominant are the analogues of the terms principal and nonprincipal for the corresponding differential equation (3). See Hartman [3, p. 355]. Note that in [3] the leading coefficient is positive while here it is negative. We remark that principal or recessive solutions are unique up to a constant factor. Following the arguments in [3, p. 355] for the continuous case, we have the following theorem.

THEOREM 1. *If* (1) *is nonoscillatory, there exists a recessive solution u and a dominant solution v such that*

$$\sum^{\infty} \frac{1}{(c_n u_n u_{n+1})} = \infty \quad \textit{and} \quad \sum^{\infty} \frac{1}{(c_n v_n v_{n+1})} < \infty.$$

Proof. Let u and v be two linearly independent solutions of (1). Then $c_n(v_n u_{n+1} - v_{n+1}u_n) = d$, for some constant d and for all n. Choose k large enough so that $u_n \neq 0$, $v_n \neq 0$, for all $n \geqq k$. Then

$$\Delta(u_n/v_n) = [(v_n u_{n+1} - u_n v_{n+1})/v_n v_{n+1}] \cdot [c_n/c_n] \tag{6}$$
$$= d/(c_n v_n v_{n+1}).$$

Since $d/(c_n v_n v_{n+1})$ is of one sign for $n \geqq k$, we conclude that u_n/v_n is monotone. Let $L = \lim (u_n/v_n)$, as $n \to \infty$, where L could be infinite. If $L = \pm\infty$, then v is recessive and u is dominant. If $L = 0$, then u is recessive and v is dominant. If L is a real nonzero constant, define the solution $x = u - Lv$. Note that x and v are linearly independent. Then $(x_n/v_n) \to 0$, as $n \to \infty$. Thus, renaming if necessary, we can always find a recessive solution u and a dominant solution v.

From (6), we have

$$\frac{u_n}{v_n} = \frac{u_k}{v_k} + \sum_{j=k}^{n-1} \frac{d}{c_j v_j v_{j+1}}.$$

Since $u_n/v_n \to 0$, as $n \to \infty$, we conclude that $\sum^{\infty} 1/(c_j v_j v_{j+1}) < \infty$. Starting with $\Delta(v_n/u_n)$, a similar argument proves $\sum^{\infty} 1/(c_j u_j u_{j+1}) = \infty$. This completes the proof.

Based on what happens in the continuous case (3), it is not surprising that Theorem 1 has several corollaries.

COROLLARY 1. *Suppose* (1) *is nonoscillatory. If v is a solution of* (1) *such that $\sum^{\infty} 1/(c_n v_n v_{n+1}) < \infty$, then v is dominant and u defined by $u_n = v_n \sum_{j=n}^{\infty} 1/(c_j v_j v_{j+1})$ is recessive. Similarly, if u is a solution such that $\sum^{\infty} 1/(c_n u_n u_{n+1}) = \infty$, then u is recessive, and v defined by $v_n = u_n \sum_{j=k}^{n-1} 1/(c_j u_j u_{j+1})$ is a dominant solution where k is large enough so that $u_j \neq 0$, $j \geqq k$.*

Proof. Suppose we have a solution v such that $\sum^{\infty} 1/(c_n v_n v_{n+1}) < \infty$. Then define u as stated in the hypothesis. Note that u is a solution. Then $u_n/v_n = \sum_n^{\infty} 1/(c_j v_j v_{j+1}) \to 0$, as $n \to \infty$, so that u is recessive and v is dominant. A similar argument proves the other case.

COROLLARY 2. *If $\sum^{\infty} 1/c_n = \infty$ and if all solutions of* (1) *are bounded, then* (1) *must oscillate.*

Proof. Suppose (1) is nonoscillatory. Then Theorem 1 implies the existence of a dominant solution v such that $\sum^{\infty} 1/(c_n v_n v_{n+1}) < \infty$. By hypothesis, $v_n v_{n+1} \leqq M$, for all n. Thus we have

$$\infty > \sum^{\infty} \frac{1}{c_n v_n v_{n+1}} \geqq \sum^{\infty} \frac{1}{c_n M} = \frac{1}{M} \sum^{\infty} \frac{1}{c_n}, \quad \text{a contradiction.}$$

If $1/c_n$ is summable, then the conclusion of Corollary 2 may no longer be true. See the example preceding Theorem 3.

If $p_n \geqq 0$ in (2), that is, if $b_n - c_n - c_{n-1} \geqq 0$ in (1), then we can be more precise about the behavior of the recessive and dominant solutions.

THEOREM 2. *If $b_n - c_n - c_{n-1} \geqq 0$, then there exist a recessive solution u and a dominant solution v such that $u_n > 0$, $u_{n+1} \leqq u_n$ and $v_n > 0$, $v_{n+1} \geqq v_n$. Suppose there* *exists a nonnegative sequence* $\{\varepsilon_n\}$ *such that*

$$b_n-(1+\varepsilon_n)c_n-c_{n-1}\geqq 0 \quad \text{and} \quad \sum \varepsilon_n=\infty. \tag{7}$$

Then $v_n\to\infty$. *If there exists a nonnegative sequence* $\{\gamma_n\}$ *such that*

$$b_n-c_n-(1+\gamma_n)c_{n-1}\geqq 0 \quad \text{and} \quad \sum \gamma_n=\infty, \tag{8}$$

then $u_n\to 0$.

Proof. Let v be the solution of (1) defined by $v_0=1$, $v_1=2$. Remark 1 then implies $v_{n+1}\geqq v_n$, for all n. Next, by an actually stronger result of Hartman and Wintner [4], we have the existence of a solution u such that $u_n>0$ and $u_{n+1}\leqq u_n$. (Note that there is a misprint on the first page of [4]. It should be $\Delta y_k<0$, not >0.) Therefore, we can say u_n/v_n is positive and monotone decreasing to some limit L. If $L=0$, then u is recessive and v is dominant. Suppose $L>0$. Then $u_n-Lv_n\geqq 0$, for all n. Also, $(u_{n+1}-Lv_{n+1})-(u_n-Lv_n)=(u_{n+1}-u_n)-L(v_{n+1}-v_n)\leqq 0$. If for some integer k, $u_k-Lv_k=0$, then $u_n-Lv_n=0$, for all $n\geqq k$, a contradiction to u and v being linearly independent. Thus $(u_n-Lv_n)>0$ and $\Delta(u_n-Lv_n)\leqq 0$, for all n. Clearly $(u_n-Lv_n)/v_n\to 0$, as $n\to\infty$. Renaming if necessary, we have the existence of a dominant solution v and a recessive solution u.

If condition (7) is satisfied, Lemma 2 implies $v_n\to\infty$, as $n\to\infty$.

Suppose condition (8) is satisfied. As previously mentioned, the arguments in [4] establish the existence of a solution u such that $u_{n+1}\leqq u_n$. Thus we may write $u_{n-1}=(b_nu_n-c_nu_{n+1})/c_{n-1}\geqq[(b_n-c_n)/c_{n-1}]u_n\geqq(1+\gamma_n)u_n$, and hence $u_n/u_0\leqq 1/\prod(1+\gamma_j)$, $1\leqq j\leqq n$. Assumption (8) then implies $u_n\to 0$, as $n\to\infty$. Clearly u must be recessive, because u_n/v_n tends to zero where v is the dominant solution defined earlier. This completes the proof.

Some examples illustrating Theorem 2 follow. Let $b_n=2$ and $c_n=1$, for all n. Then $u_n=1$ and $v_n=n$ are the recessive and dominant solutions. Clearly $u_n\not\to 0$.

A second example, from [6], has $b_n=n(2n^2-1)/(n+1)$, $c_n=n^2$, for $n\geqq 1$. Then $b_n-c_n-c_{n-1}=-1/(n+1)<0$.

It is easy to verify that $u_n=1/n$ is a solution, and Corollary 1 implies it is recessive. Also, from Corollary 1, v defined by

$$v_n=u_n\sum_{j=1}^{n-1}\frac{1}{c_ju_ju_{j+1}}=\frac{1}{n}\sum_{j=1}^{n-1}\left(1+\frac{1}{j}\right) \tag{9}$$

is dominant. However, $v_n\leqq 2$, for all n, so that $v_n\not\to\infty$, as $n\to\infty$.

The previous example is actually indicative of a more general result.

THEOREM 3. *Assume* $b_n-c_n-c_{n-1}\leqq 0$, *for all* n. *If* (1) *is nonoscillatory and if* $\sum^\infty 1/c_n<\infty$, *then all solutions of* (1) *are bounded.*

Proof. Let x be any solution of (1). We may assume $x_n>0$, for all $n\geqq k$, for some integer k. Note that $(b_n-c_{n-1})/c_n\leqq 1$. Rewriting (1) yields

$$\begin{aligned} x_{n+1}&=[(b_n-c_{n-1})/c_n]x_n+[c_{n-1}/c_n][x_n-x_{n-1}]\\ &\leqq x_n+(c_{n-1}/c_n)(x_n-x_{n-1}), \end{aligned}$$

or

$$x_{n+1}-x_n\leqq(c_{n-1}/c_n)(x_n-x_{n-1}), \quad \text{for any } n\geqq k. \tag{10}$$

Repeated application of (10) yields

$$x_{n+1}-x_n\leqq(c_{n-1}/c_n)(c_{n-2}/c_{n-1})\cdots(c_k/c_{k+1})(x_{k+1}-x_k),$$

or

$$x_{n+1} \leqq x_n + M/c_n, \tag{11}$$

where M is a constant independent of n. Repeated application of (11) yields

$$\begin{aligned} x_{n+1} &\leqq M/c_n + M/c_{n-1} + \cdots + M/c_k + x_k \\ &\leqq x_k + M \sum_{j=0}^{n-k} \frac{1}{c_{k+j}}. \end{aligned}$$

Our hypotheses yield that x_{n+1} is bounded independently of n or that the solution x is bounded. This proves the theorem.

Note that the first of the two examples preceding Theorem 3 shows that the conclusion of Theorem 3 may no longer be true if $\sum^{\infty} 1/c_n = \infty$.

Based on Theorem 1, we can conclude that the nonoscillation of (1) implies the existence of two linearly independent solutions u and v of (1) such that

$$\sum^{\infty} \frac{1}{c_n(u_n u_{n+1} + v_n v_{n+1})} < \infty. \tag{12}$$

It should be pointed out that the converse of this is not true. That is, one can easily construct examples where the sum in (12) is finite but (1) is oscillatory. For instance, if $c_n = 1$ and $b_n = -2$, then $(-1)^n$ and $n(-1)^n$ are solutions which satisfy (12). This differs from the continuous case (3), where the convergence of the integral analogue of the sum in (12) is equivalent to nonoscillation. See [3, p. 354].

For completeness, we conclude this section by reformulating a comparison theorem found in [7] for equations of the form (1).

Consider (1) and the following equation:

$$r_n w_{n+1} + r_{n-1} w_{n-1} = d_n w_n, \qquad r_n > 0. \tag{13}$$

THEOREM 4. *Suppose*

$$(r_{n-1}/r_n) \geqq (c_{n-1}/c_n)$$

and

$$(d_n/r_n) - (r_{n-1}/r_n) \geqq (b_n/c_n) - (c_{n-1}/c_n) \geqq 1.$$

Suppose also that $w_1 \geqq x_1 \geqq 0$, $w_0 \geqq x_0 \geqq 0$ *and* $w_1 - w_0 \geqq x_1 - x_0 \geqq 0$. *Then* $w_{n+1} - w_n \geqq x_{n+1} - x_n$ *and* $w_{n+1} \geqq x_{n+1}$, *for* $n \geqq 1$.

Proof. Divide (1) by c_n and (13) by r_n. The result then follows from [7, Thm. 1].

3. Limit point results. Consider the following equation:

$$c_n x_{n+1} + c_{n-1} x_{n-1} = b_n x_n + \lambda x_n, \qquad \lambda \text{ real or complex.} \tag{14}$$

Equation (14) is called limit circle, L.C., if all solutions are square summable. Otherwise it is called limit point, L.P. See [1, p. 127] for an explanation and development of the related theory.

It is proven in [1] that if (14) is L.C. for one value of λ, then it is L.C. for any value of λ, including $\lambda = 0$.

THEOREM 5. *If* $\sum^{\infty} 1/\sqrt{c_n} = \infty$ *and if* (1) *is nonoscillatory, then* (14) *is L.P.*

Proof. Suppose not, so that (14) and hence (1) are L.C. Since (1) is nonoscillatory, let u and v be two linearly independent solutions as found in Theorem 1. Then

$$\begin{aligned}\sum^k \frac{1}{(c_n)^{1/2}} &= \sum^k \frac{(u_n u_{n+1} + v_n v_{n+1})^{1/2}}{(c_n(u_n u_{n+1} + v_n v_{n+1}))^{1/2}} \\ &\leqq \left[\sum^k (u_n u_{n+1} + v_n v_{n+1})\right]^{1/2} \left[\sum^k \frac{1}{c_n(u_n u_{n+1} + v_n v_{n+1})}\right]^{1/2}.\end{aligned} \tag{15}$$

The first term on the right in (15) is bounded because u and v are square summable. The second term is bounded by Theorem 1. Thus $1/\sqrt{c_n}$ is summable, a contradiction.

THEOREM 6. *If* $|b_n| \geqq c_n + c_{n-1}$, *then* (14) *is L.P.*

Proof. Lemma 2 implies the solution of (1) defined by $v_0 = 0$, $v_1 = 1$ is not square summable. Hence (1) and thus (14) are not L.C.

It is interesting to compare Theorem 6 with Theorem 11 of [6].

4. Some comments on oscillation. If $b_n \leqq 0$, Lemma 3 implies that (1) is oscillatory. In addition, if $b_n \leqq -c_n - c_{n-1} < 0$, the substitution $y_n = (-1)^n x_n$ transforms (1) to

$$c_n y_{n+1} + c_{n-1} y_{n-1} = (-b_n) y_n, \tag{16}$$

where $-b_n \geqq c_n + c_{n-1} > 0$. We can now apply many of the results of § 2 to (16) and obtain information about solutions of (1). In particular, Theorems 1 and 2 yield existence and growth properties of recessive and dominant solutions, where now the only difference is that eventually solutions must oscillate and actually alternate in sign. A comparison principle based on Theorem 4 could also be formulated between certain types of oscillatory equations. This all follows from the hypothesis $b_n \leqq -c_n - c_{n-1} < 0$ and the substitution $y_n = (-1)^n x_n$. We leave such formulations to the interested reader.

We conclude with several sufficient conditions for oscillation, where there is no restriction on the sign of b_n.

THEOREM 7. *If* $b_n \leqq \min(c_n, c_{n-1})$, *for all n sufficiently large, then* (1) *is oscillatory.*

Proof. Let x be any solution of (1). Suppose (1) is nonoscillatory. Then for n large enough, we may assume $x_n > 0$. Also, we may assume $b_n > 0$, by Lemma 3. Since $c_{n-1}x_{n-1} > 0$, equation (1) yields that $c_n x_{n+1} < b_n x_n$, or $c_n/b_n < x_n/x_{n+1}$. However, $c_n/b_n \geqq 1$, so that we have

$$x_n > x_{n+1}, \quad \text{for all } n \text{ sufficiently large.} \tag{17}$$

In a similar fashion, (1) implies $c_{n-1}x_{n-1} < b_n x_n$, or $c_{n-1}/b_n < x_n/x_{n-1}$. Since $c_{n-1}/b_n \geqq 1$, we have

$$x_n > x_{n-1}, \quad \text{for all } n \text{ sufficiently large.} \tag{18}$$

However, (17) and (18) are contradictory, and so (1) must oscillate.

COROLLARY 3. *If* $b_n \leqq c_n$ *and if* c_n *is eventually nonincreasing, then* (1) *is oscillatory.*

COROLLARY 4. *If* $b_n \leqq c_{n-1}$ *and if* c_n *is eventually nondecreasing, then* (1) *is oscillatory.*

Note that Corollary 4 also follows from Theorem 3 of [6].

THEOREM 8. *If* $b_n \leqq c_{n-1}$ *and if* $\sum^\infty 1/c_n < \infty$, *then* (1) *is oscillatory.*

Proof. Assume not. Then (1) is nonoscillatory. Repeating the appropriate argument in Theorem 7, we again arrive at (18), for any solution. By Theorem 1, if (1) is nonoscillatory, there exists a recessive solution u such that $\sum^\infty 1/(c_n u_n u_{n+1}) = \infty$.

However, (18) implies that $u_{n+1} \geqq u_n$, for all n sufficiently large, so that $u_{n+1} \geqq u_k$, for $n \geqq k$, for some integer k. Thus

$$\sum_k^\infty \frac{1}{c_n u_n u_{n+1}} \leqq \sum_k \frac{1}{c_n u_k^2} = \frac{1}{u_k^2} \sum_k^\infty \frac{1}{c_n} < \infty,$$

a contradiction. This completes the proof.

Many of the ideas in this paper extend to the nonhomogeneous case. These will appear in a sequel.

Acknowledgment. We are grateful to the referee for many valuable comments, especially those which improved Theorem 2. We are also grateful to Professor Don Hinton for several very helpful suggestions.

REFERENCES

[1] F. V. ATKINSON, *Discrete and Continuous Boundary Value Problems*, Academic Press, New York, 1964.

[2] T. FORT, *Finite Differences and Difference Equations in the Real Domain*, Oxford University Press, London, 1948.

[3] P. HARTMAN, *Ordinary Differential Equations*, John Wiley, New York, 1964.

[4] P. HARTMAN AND A. WINTNER, *On linear difference equations of the second order*, Amer. J. Math., 72 (1950), pp. 124–128.

[5] ———, *Linear differential and difference equations with monotone solutions*, Ibid., 75 (1953), pp. 731–743.

[6] D. HINTON AND R. LEWIS, *Spectral analysis of second order difference equations*, J. Math. Anal. Appl., 63 (1978), pp. 421–438.

[7] F. W. J. OLVER, *Bounds for the solutions of second order linear difference equations*, J. Res. Nat. Bur. Standards Sect. B, 71 (1967), pp. 161–166.

[8] F. W. J. OLVER AND D. J. SOOKNE, *Note on backward recurrence algorithms*, Math. Comput., 26, no. 120 (1972), pp. 941–947.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0007 $01.00/0

REPRODUCING-KERNEL HILBERT SPACES OF DISTRIBUTIONS AND GENERALIZED STOCHASTIC PROCESSES*

REUVEN MEIDAN†

Abstract. In this work we consider reproducing-kernel Hilbert spaces (RKHSs) of distributions and pursue their connection with generalized stochastic processes (GSPs). It is shown that every Hilbert space of distributions is an RKHS and that there exists a one-to-one correspondence between the class of positive definite kernel operators (PDKOs) and RKHSs, where the PDKO induces the reproducing kernel of the RKHS. The RKHS is then shown to represent the GSP in the sense of an isometrical isomorphism between two Hilbert spaces. As applications of the theory we consider generalized integral equations, the problem of linear least-squares estimation and series expansions of the GSP.

1. Introduction. Let $t \to x_t$ be a second-order ordinary stochastic process (OSP), i.e., a mapping from the subset T of the reals into $L^2(\Omega)$, the space of second-order random variables on the probability space Ω. With this process we can associate the correlation function defined by

$$R(s, t) = E[x_t \bar{x}_s], \tag{1}$$

where E denotes the statistical average with respect to Ω. $R(s, t)$ enjoys the property of positive definiteness which qualifies it as a reproducing kernel of a *reproducing-kernel Hilbert space* (RKHS). This association of a unique RKHS of functions on T with a given OSP has been treated quite extensively in the literature. Some of the notable workers are Loéve (cf. Lévy [5, Appendix I]), Parzen [6] and Kailath [4].

Consider now the converse problem, i.e., let H be a given Hilbert space of functions on T. We are looking for a process whose RKHS is H. Naturally a necessary, but also sufficient, condition for the existence of such a process is that H should be an RKHS. However, it is well known that not every Hilbert space of functions enjoys the reproducing property, a commonly-quoted counter-example being $L^2(T)$, the space of square-integrable functions on T. We therefore propose the following approach.

1) We shall deal with Hilbert spaces of distributions rather than ordinary functions. Should H be a given Hilbert space of functions on T it can usually be interpreted in terms of distributions. The advantage of this approach is that every Hilbert space of distributions is actually an RKHS.

2) We consider *generalized* rather than *ordinary* stochastic processes (GSPs instead of OSPs).

A GSP is a linear and continuous operator from a space of test functions into a space of random variables, and by considering GSPs two advantages are gained. First, the family of processes amenable to the analysis is increased. Second, since the GSP is a linear operator, linear operator theory is applicable. It will be shown that the Hilbert space of distributions H associated with a GSP is isometrically isomorphic to the closure in $L^2(\Omega)$ of the range space of the GSP. Hence H serves as an interfacing space between the space of test functions (the domain of the GSP) and the space of random variables (the range of the GSP).

By adopting the above approach the following correspondence between stochastic processes and Hilbert spaces is obtained. Given a GSP, a unique Hilbert space of distributions can be constructed such that the former's correlation operator serves as

* Received by the editors October 4, 1976, and in final revised form August 30, 1977.

† National Research Institute for Mathematical Sciences of the CSIR, Pretoria, South Africa, and Department of Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa.

the latter's reproducing operator. Conversely, a given Hilbert space of distributions uniquely defines its reproducing operator, which in turn can serve as the correlation operator of a GSP which is thus determined modulo a unitary operator.

We pursue two applications of the theory:

1) Coordinate-free series expansion of the GSP (§ 4). These are given in terms of *complete orthonormal systems* (CONSs) in H. They are coordinate-free in the sense that there is a freedom of choice in selecting the CONSs.

2) Generalized integral equations are investigated in § 5.

It should be noted that Hilbert spaces of (vector-valued) distributions have also been discussed by Schwartz [7] who applied them to the theory of elementary particles in quantum mechanics.

2. Preliminaries. The following is a review of the concepts relevant to this work. Let T denote an open set in R^n and $D(T)$ the space of infinitely differentiable test functions whose supports are compact and contained in T. $D(T)$ is equipped with Schwartz's testing function topology. Let $D'(T)$ denote the space of distributions in T, i.e., $D'(T)$ is the dual of $D(T)$. It is equipped with the weak topology generated by $D(T)$. We say that H is a *subspace of distributions* if it is a subspace of $D'(T)$ and if, in addition, its intrinsic topology is stronger than the relative topology induced by the (weak) topology of $D'(T)$. Let R be a linear and continuous operator from $D(T)$ into $D'(T)$. We refer to R as a *kernel operator* in view of Schwartz's kernel theorem, according to which these operators enjoy representations in terms of kernels which are distributions on $T\times T$.

A sesquilinear form on $D(T)\times D(T)$

$$S(\varphi, \psi)=\langle R\varphi, \bar{\psi}\rangle_T, \qquad \varphi, \psi\in D(T), \tag{2}$$

can be associated with R. $\langle\cdot\,,\,\cdot\rangle_T$ denotes the scalar product of $D(T)$ and its dual $D'(T)$. By sesquilinearity we mean that S is linear with respect to the first variable and anti-linear with respect to the other. We say that R is *positive definite* if

$$S(\varphi, \varphi)\geqq 0 \tag{3}$$

for every $\varphi\in D(T)$. We propose the term 'positive', rather than 'nonnegative' in order to simplify the terminology and because this use of the term has been proposed in notable treatises (e.g. Gel'fand and Vilenkin [2, p. 26]). In this work we deal with *positive definite kernel operators* (PDKOs), i.e., linear and continuous operators from $D(T)$ into $D'(T)$ which, in addition, are positive definite.

Let R^t denote the transpose operator of R. It too is linear and continuous from $D(T)$ into $D'(T)$. It can easily be verified that the positive definiteness of R implies that it is self-transposed in the sense that

$$R=R^t.$$

Let $(\Omega, \mathscr{A}, P)$ denote a fixed probability space, where Ω is a set, $\mathscr{A}$ a σ-algebra of subsets of Ω and P a probability measure on $\mathscr{A}$. By $L^2(\Omega)$ we denote the space of second-order random variables on Ω, i.e., $L^2(\Omega)$ consists of scalar-valued functions on Ω which are square-integrable. $L^2(\Omega)$ is a Hilbert space when equipped with the scalar product

$$(f, g)_\Omega=\int_\Omega f\bar{g}\,dP(\omega)=E[f\bar{g}]. \tag{4}$$

A *generalized stochastic process* (GSP) is a linear and continuous operator from a space of testing functions on T into a space of random variables. We shall adopt Itô's approach [3], which is somewhat more restrictive than the definition proposed by Gel'fand [1], [2]. According to Itô's definition, a (second-order) GSP is a linear and continuous operator from $D(T)$ into $L^2(\Omega)$. In this work it is necessary for the range of the operator to be a Hilbert space, in order that it should be possible to construct the correlation operator associated with the GSP.

Let $t \to x_t$ be a (second-order) ordinary stochastic process (OSP), i.e. $t \to x_t$ is a mapping from T into $L^2(\Omega)$. If the OSP is locally Bochner-integrable, it induces a GSP by

$$u\varphi = \int_T x_t\varphi(t)\,dt, \qquad \varphi \in D(T). \tag{5}$$

This verifies that the GSP is indeed a generalization of the OSP concept. Let u be a GSP in the sense of the above definition and let u^t denote its transpose. u^t is thus a linear and continuous operator from the dual of $L^2(\Omega)$ into $D'(T)$. In view of Riesz's representation theorem, the dual of $L^2(\Omega)$ can be identified with $L^2(\Omega)$ itself. Hence u^t can be viewed as operating from $L^2(\Omega)$, and the composite operator u^tu can be constructed. We use the following notation

$$R = u^tu \tag{6}$$

and refer to R as the *correlation operator* associated with the GSP u. If u is a GSP induced by an OSP, the operator R is an integral operator,

$$R\varphi = \int_T R(s, t)\varphi(s)\,ds, \qquad s, t \in T. \tag{7}$$

The kernel $R(s, t)$ is now an ordinary function,

$$R(s, t) = (x_t, x_s)_\Omega \tag{8}$$

and within the framework of the OSP is called the correlation function of the process. Hence the correlation operator is the generalization of the concept of the correlation function.

Clearly R is a linear and continuous operator from $D(T)$ into $D'(T)$. Moreover, it is positive definite. This follows from the fact that for every $\varphi \in D(T)$,

$$\langle R\varphi, \bar{\varphi}\rangle_T = \langle (u^tu)\varphi, \bar{\varphi}\rangle_T = (u\varphi, u\varphi)_\Omega \geqq 0.$$

Hence R is a PDKO. In regards to the above chain equality one should note that

$$\langle u^tf, \bar{\varphi}\rangle_T = (f, u\varphi)_\Omega, \qquad f \in L^2(\Omega), \quad \varphi \in D(T)$$

which follows from the anti-isomorphism between $L^2(\Omega)$ and its dual.

Now to the converse argument. Given a PDKO R from $D(T)$ into $D'(T)$, a (not necessarily unique) GSP u can be found such that R is its correlation operator. In other words R can always be factored as in (6). This follows from a theorem by Gel'fand and Vilenkin [2]. However it will follow also from Theorem 3 of this paper.

Next we review the subject of reproducing-kernel Hilbert spaces. Let X be an abstract set and let H be a Hilbert space of functions on X. H is called a reproducing-kernel Hilbert space (RKHS) if it enjoys the following reproducing property. There exists a complex-valued function $K(x, y)$ on $X \times X$, called the reproducing kernel, such that

(i) for any fixed $y \in X$, $K(\cdot, y)$ is in H;

(ii) $K(x, y)$ induces the reproducing property by

$$(f(\cdot), K(\cdot, y))_H = f(y)$$

for each $f(\cdot) \in H$, and $y \in X$. $(\cdot, \cdot)_H$ denotes the scalar product in H. As is well known, there exists a one-to-one correspondence between the family of RKHSs on X and the family of positive definite complex-valued functions on $X \times X$, the correspondence being the (unique) connection between an RKHS and its reproducing kernel.

We shall need the following fact about RKHSs. Every $x \in X$ induces the reproducing functional $F_x: f \to f(x)$, $f \in H$. F_x is a linear functional on H. H is an RKHS if and only if all the reproducing functionals $\{F_x - x \in X\}$ are continuous on H.

We shall also need the family of Sobolev spaces of type 2. Let K be a compact set of R^n and suppose we define the following family of scalar products on $D(K)$,

$$(\varphi, \psi)_m = \sum_{|i| \leqq m} \int_K (D^i \varphi)\overline{(D^i \psi)}\, dt$$

where i is an n-vector $(i_1, \cdots, i_n)$ of nonnegative integers where

$$|i| = \sum_{k=1}^{n} i_k$$

and m a nonnegative integer. If we complete $D(K)$ with respect to the norm generated by the mth scaler product we obtain the Sobolev space $W_m(K)$ of order m.

3. RKHSs of distributions. When dealing with RKHSs of distributions the underlying space X is identified with $D(T)$, in contrast to the case of RKHSs of ordinary functions where X is identified with T. We start this section by constructing an RKHS of distributions around a given PDKO R. R is shown to be the reproducing operator in the sense that the sesquilinear form generated by it (eq. (2)) serves as the reproducing kernel of the space. Next we show that every Hilbert subspace of distributions is in fact an RKHS and determine its reproducing operator.

THEOREM 1. *Let R be a PDKO from $D(T)$ into $D'(T)$. A unique Hilbert subspace of distributions H can be constructed such that R is its reproducing operator.*

Proof. We first establish the construction of H. Consider the range space Ra (R) of R. Clearly it is a subspace of $D'(T)$. Let f and g be in Ra (R). Then there exist elements φ and ψ (not necessarily unique) such that $f = R\varphi$ and $g = R\psi$. We associate with the pair (f, g) a scalar product

$$(f, g)_H = \langle f, \overline{R^{-1}g} \rangle_T = \langle R\varphi, \bar{\psi} \rangle_T \triangleq S(\varphi, \psi). \tag{9}$$

Naturally R^{-1} denotes the inverse of R. However, since R is not necessarily one-to-one, R^{-1} does not exist as a (unique-valued) operator but should rather be interpreted as a binary relation. Nevertheless we claim that the form (9) is unique, since Ra (R), as a subspace of $D'(T)$, is orthogonal to N (R), the null space in $D(T)$ of the operator R.

Let $f \in$ Ra (R); then there exists a $\varphi \in D(T)$ such that $f = R\varphi$. Hence for every $\psi \in D(T)$,

$$\langle f, \psi \rangle_T = \langle R\varphi, \psi \rangle_T. \tag{10}$$

In view of the self-transposeness of R, we have that

$$\langle R\varphi, \psi \rangle_T = \langle R\psi, \varphi \rangle_T. \tag{11}$$

But the right-hand side of (11) is equal to zero for $\psi \in \mathrm{N}(R)$. Combination of the last two equations verifies that Ra (R) is orthogonal to N (R). It follows that R^{-1} is unique on Ra (R) modulo null elements of R, which establishes the fact that the form (9) is indeed unique. We can consider the quotient space $D(T)$ over N (R). As an operator from Ra (R) into this quotient space, R^{-1} is unique.

We still have to verify that the form $(\cdot, \cdot)_H$ qualifies as a scalar product. Indeed, its sesquilinearity has been mentioned above. Its Hermitian symmetry follows from the self-transposeness of R (eq. (11)) and is positive definiteness is a result of the positive definiteness of R. It also follows from the above argument that $(f, f)_H$ is equal to zero if and only if $f=0$. Indeed, $(f, f)_H = \langle R\varphi, \bar{\varphi}\rangle_T$ where φ is unique in the quotient space $D(T)/\mathrm{N}(R)$ such that $f = R\varphi$. But, as was established above, Ra (R) is ortrogonal to N (R). Hence $\langle R\varphi, \bar{\varphi}\rangle_T$ is equal to zero iff $\varphi = 0$ from which follows $f=0$.

Consequently the form (9) satisfies the requirements of a scalar product rendering Ra (R) a pre-Hilbert space. We complete it in the norm induced by the scalar product and obtain the Hilbert space H. We claim that H is a space of distributions i.e., that it is a subspace of $D'(T)$, and that, in addition, its topology is stronger than the relative topology induced by $D'(T)$.

Let $\psi \in D(T)$ and $f \in \mathrm{Ra}\,(R)$. There exists a $g \in D'(T)$ such that $g = R\psi$ and

$$\langle f, \bar{\psi}\rangle_T = (f, g)_H. \tag{12}$$

Consider a strongly-bounded set B in Ra (R). This set is also weakly bounded. Hence the right-hand side of the last equation (12) is bounded over $f \in B$. Consequently the left-hand side of (12) is bounded over B for every $\psi \in D(T)$. It follows that B is weakly bounded in $D'(T)$. Let I denote the natural injection of Ra (R) into $D'(T)$. We have just established that I is bounded. But boundedness in this case implies continuity. It follows that I can be uniquely and continuously extended from Ra (R) onto its completion H, as a continuous operator into $D'(T)$. Consequently H qualifies as a Hilbert space of distributions.

We next verify that H is the RKHS associated with the operator R. Indeed, consider the sequilinear form $S(\varphi, \psi)$ generated by R (eq. (3)). It is a complex-valued function on $D(T) \times D(T)$. S satisfies the properties of the reproducing kernel of H, i.e.,

$$S(\cdot, \varphi) = R\bar{\varphi} \tag{13}$$

is an element of H for every $\varphi \in D(T)$ and

$$(g(\cdot), S(\cdot, \varphi))_H = g(\varphi) \tag{14}$$

for every $g \in H$.

We now approach the problem from the opposite direction.

THEOREM 2. *Let H be a Hilbert subspace of distributions on T. Then H enjoys the reproducing property and its reproducing operator is II^t, where I is the natural injection of H into $D'(T)$ and I^t its transpose.*

Proof. As mentioned in § 2, a necessary and sufficient condition for a Hilbert space to enjoy the reproducing property is that every reproducing functional should be continuous. In this case the underlying space is $D(T)$; hence the reproducing functional associated with $\varphi \in D(T)$ is $F_\varphi : f \to f(\varphi)$, where f traverses H. Now for every $\varphi \in D(T)$, F_φ is a continuous functional on $D'(T)$. Since H is a subspace of $D'(T)$ with a stronger topology, F_φ is also continuous on H. Hence H is an RKHS.

Next, since I is, by hypothesis, continuous from H into $D'(T)$, its transpose I^t is continuous from $D(T)$ into H. The positive definiteness of II^t is obvious; hence it is a PDKO. Now consider $I^t\varphi$. It is in H and for every $g \in H$,

$$(g, I^t\bar{\varphi})_H = \langle Ig, \varphi\rangle_T = g(\varphi)$$

which express the reproducing property. This completes the proof.

We conclude this section by noting that H is separable. This follows from the fact that $D(T)$ is separable. The separability will be helpful for the series expansions of § 5.

4. The association of RKHSs of distributions and GSPs. Let u be a GSP and L denote the closure in $L^2(\Omega)$ of its range space. L is often called the linear space of the process in the sense that it represents the random variables attainable by linear operations, including limits, on the measurements of the process. We associate with the GSP u the RKHS of distributions H generated by the correlation operator $R = u^t u$.

THEOREM 3. *Let u be a GSP, R its correlation operator and H the RKHS of distributions generated by R. Then H is isometrically isomorphic to L, the closure in $L^2(\Omega)$ of the range of u. Conversely, let H be an RKHS of distributions, then a GSP u can be found, such that*

$$R = u^t u = II^t. \tag{15}$$

In other words the diagram of Fig. 1 *is commutative.*

The theorem establishes the following polar decompositions of u and u^t

$$u = JUI^t, \tag{16}$$

$$u^t = IU^tJ^t \tag{17}$$

where U is a unitary operator, i.e. $UU^t = U^tU =$ identity operator in H, J is the injection of L into $L^2(\Omega)$ and J^t the projection obtained by transposing J.

$$\begin{tikzcd}
D(T) \arrow[r, "I^t"] \arrow[d, "R"'] \arrow[rrr, bend left, "u"] & H \arrow[r, "U"] & L \arrow[r, "J"] & L^2(\Omega) \arrow[lll, bend left=60, "u^t"] \\
D'(T) & H \arrow[l, "I"'] & L \arrow[l, "U^t"'] & L^2(\Omega) \arrow[l, "J^t"']
\end{tikzcd}$$

FIG. 1. *The spaces and operators involved in the discussion of Theorem* 3.

Proof of the theorem. We show first that U^t is a unitary operator from L onto H. This will be established by verifying that U^t maps isometrically a dense subspace in L onto a dense subspace in H. Indeed, consider Ra (u), the range of u. It is dense in L. Let f and g be two elements of Ra (u) then there exist φ, ψ in $D(T)$ such that $f = u\varphi$, $g = u\psi$. Now by the following chain equality

$$(U^tf, U^tg)_H = \langle R\varphi, \bar{\psi}\rangle_T = \langle U^tu\varphi, \bar{\psi}\rangle_T = (u\varphi, u\psi)_\Omega = (f, g)_\Omega,$$

we conclude that U^t is an isometric operator and that it maps Ta (u) in L onto Ra (R) in H. Since Ra (R) is dense in H, the direct part of the proof is complete.

Conversely, given H, I^t is found by transposing the injection I of H into $D'(T)$, following the discussion of Theorem 2. Take a complete orthonormal system in H and a sequence of zero mean, unit variance independent random variables $\{g_n\}$. Define the operator U on $\{f_n\}$ as

$$g_n = Uf_n.$$

Then U determines the requested unitary operator and the requested GSP is the composition $u = UI^t$.

We now consider the null space of u in $D(T)$. In view of the commutativity of the diagram in Fig. 1 it is easily verified that $\mathrm{N}(u) = \mathrm{N}(I^t) = \mathrm{N}(R)$. We established earlier that $\mathrm{Ra}(R)$ in $D'(T)$ is contained in the orthogonal of $\mathrm{N}(R)$. Clearly this holds also for H. In fact we can state the following.

THEOREM 4. *Let H^{wc} denote the closure of H in the weak topology of $D'(T)$. Then H^{wc} is the orthogonal of* $\mathrm{N}(R)$, *i.e.*,

$$\mathrm{Ra}(R) \subset H \subset H^{\mathrm{wc}} = \mathrm{N}(R)^{\perp}. \tag{18}$$

Proof. This is a consequence of Proposition 35.4 of Trèves [8].

5. Application to generalized integral equations. Let us consider the following generalized integral equation: suppose u is a given GSP from $D(T)$ into $L^2(\Omega)$ and f a given distribution in $D'(T)$. We wish to find an $x \in L^2(\Omega)$ such that for every $\varphi \in D(T)$

$$(x, u\varphi)_\Omega = f(\varphi). \tag{19}$$

Clearly if a solution exists, it is not necessarily unique. In order to render it unique we consider the solution of minimum norm which is the solution in L_u. The proof of the following theorem is trivial.

THEOREM 5. *The generalized integral equation admits a unique solution $x \in L_u$ if and only if $f \in H$.*

The generalized integral equation (19) is encountered in the theory of linear mean-squares estimation. We are given a GSP u from $D(T)$ into $L^2(\Omega)$. We take a subset A of the testing function in $D(T)$ which spans $D(T)$. Let B be the image of A in $L^2(\Omega)$. B consists of the data obtained by observation of the process. L_u is spanned by B and constitutes the random variables in $L^2(\Omega)$ which are accessible by linear operations on the data.

Next assume that a random variable x is given in $L^2(\Omega)$. We wish to provide an estimate $\hat{x}$ for x such that

(i) $\hat{x}$ is a linear estimate, i.e., it is in L_u, and
(ii) the mean square error $\|x - \hat{x}\|_\Omega$ is minimal.

We call $\hat{x}$ the linear least-squares estimator of x. The problem is, of course, solved by the projection theorem, i.e. the estimator $\hat{x}$ is the projection of x into L_u.

For our problem it is sufficient to determine x by its scalar product over the range of u. Hence we assume that $f(\varphi) = (x, u\varphi)_\Omega$ is given. It follows that the minimum norm solution of equation (19) as discussed in Theorem 4 is our estimator $\hat{x}$.

The following observation is instructive. Suppose $\hat{x}$ is an element of L_u. The set of all $x \in L^2(\Omega)$ whose estimator is $\hat{x}$, is mapped by u^t into a single element in H.

6. Series expansions of the GSP. In the previous section we were concerned with the bottom row of Fig. 1. In this section we deal with the upper row expressed in the decomposition (16) of u. We utilize it in order to obtain coordinate-free series expansions of u. These are expressed in terms of *complete orthonormal systems* (CONSs) in H, which are countable in view of the separability of H.

THEOREM 6. *Let $\{f_n\}$ be any choice of a CONS in H. U admits the representation*

$$\sum_{n=1}^{\infty} f_n g_n \tag{20}$$

where $\{g_n = Uf_n\}$ is a CONS in L, the tensor product notation $f_n g_n$ stands for the operator

$$(f_n g_n)(f) = (f, f_n)_H g_n, \qquad f \in H, \tag{21}$$

and the limit of the partial sums of (20) *exists in the weak operator topology, i.e. for every $f \in H$,*

$$U_N f = \left(\sum_{n=1}^{N} f_n g_n\right) f \tag{22}$$

converges to Uf in the norm topology of $L^2(\Omega)$.

Proof. Let f be in H. By Parseval's theorem it can be decomposed by

$$f = \sum_{n=1}^{\infty} (f, f_n)_H f_n, \tag{23}$$

the convergence of the series (23) holding in the norm topology of H. We now operate with U on both sides of (23):

$$Uf = U\left[\sum_{n=1}^{\infty} (f, f_n)_H f_n\right] = \sum_{n=1}^{\infty} (f, f_n)_H UF_n. \tag{24}$$

The interchange of the order between the summation and operation with U is justifiable in view of the continuity of U. Since U is unitary, the system $\{Uf_n\}$ is orthonormal and spans L; hence it is complete in L.

We now comment on the operator I^t of the decomposition (16). Since I is one-to-one, the image of $D(T)$ under I^t is dense in H. Hence I^t is a one-to-one operator from the quotient space $D(T)$ over $\mathrm{N}(R)$ into H. Another property of I^t is its nuclearity, which follows from the nuclearity of $D(T)$.

Next we consider the composite operator UI^t. By the series expansion we have

$$(UI^t)(\varphi) = \sum_{n=1}^{\infty} (I^t\varphi, f_n)_H Uf_n = \sum_{n=1}^{\infty} (I^t\phi, f_n)_H g_n \tag{25}$$

for any choice of a CONS $\{f_n\}$ in H. By the definition of the transpose,

$$(I^t\varphi, f_n)_H = \langle \overline{If_n}, \varphi\rangle_T. \tag{26}$$

Hence we have the following theorem.

THEOREM 7. *Let u be a* GSP *from $D(T)$ into $L^2(\Omega)$. Then it can be expanded as*

$$u\varphi = \sum_{n=1}^{\infty} \langle \bar{f}_n, \varphi\rangle_T g_n, \tag{27}$$

where $\{f_n\}$ is any choice of a CONS in $H(R)$ considered as distributions in $D'(T)$ and $\{g_n\}$ is an orthonormal system (ONS) in $L^2(\Omega)$ given by $g_n = Uf_n$.

In view of the above, the following expansion of R is clear.

THEOREM 8. *Let R be a positive definite kernel operator from $D(T)$ into $D'(T)$. Then R can be expanded as*

$$R = \sum_{n=1}^{\infty} f_n \bar{f}_n \tag{28}$$

where $\{f_n\}$ *is as in the previous theorem and the notation* $f_n\bar{f}_n$ *stands for the operator*

$$(f_n\bar{f}_n)(\varphi)=\langle\bar{f}_n,\varphi\rangle_T f_n,\qquad \varphi\in D(T). \tag{29}$$

As mentioned earlier, the merit of the expansions lies in the freedom of choice in regard to the basis $\{f_n\}$ of the expansion. On the other hand, the convergence of the expansion holds only in the mean-square sense of $L^2(\Omega)$. Pointwise convergence can be established based on a different approach. However, this is accomplished at the cost of restricting the basis system eligible for the expansion. Based on the discussion in Gel'fand and Vilenkin, we consider the family W_m of the Sobolev space of order m and type 2. For every compact set K we can find an order m such that I^t is continuously extendible as nuclear operator from $W_m(K)$ into H. Hence I^t can be given a eigen-expansion (Gel'fand and Vilenkin [2, p. 75]). Which converges pointwise for almost every $\omega\in\Omega$. We state this result in the following theorem.

THEOREM 9. *Let K be a compact subset of T and* $W_m(K)$ *the Sobolev space of order m and type* 2 *over K. There exist an m, a CONS* $\{h_n\}$ *in* $W_m(K)$, *a CONS* $\{g_n\}$ *in L, and a sequence of nonnegative numbers* $\{\lambda_n\}$ *with* $\sum_{n=1}^{\infty}\lambda_n<\infty$, *such that for every* $\varphi\in D(K)$,

$$u\varphi=\sum_{n=1}^{\infty}\lambda_n(h_n,\varphi)_m g_n(\omega). \tag{30}$$

The convergence in $L^2(\Omega)$ *holds both pointwise almost surely and in the mean-square sense.* $D(K)$ *denotes the subspace of* $D(T)$ *consisting of functions whose supports are contained in K, and* $(\cdot,\cdot)_m$ *denotes the scalar product in* $W_m(K)$.

REFERENCES

[1] I. M. GEL'FAND, *Generalized random processes*, Dokl. Akad. Nauk SSSR, 100 (1956), pp. 855–856. (In Russian.)

[2] I. M. GEL'FAND AND N. YA. N. VILENKIN, *Generalized Functions, vol.* 4, Academic Press, New York, 1964.

[3] K. ITÔ, *Stationary random distributions*, Mem. Coll. Sci. Kyoto Univ. Ser A, 28, no. 3 (1954), pp. 209–223.

[4] T. KAILATH, *RKHS-approach to detection and estimation problems—Part I: Deterministic signals in Gaussian noise*, IEEE Trans. Information Theory, IT-17 (1971), pp. 530–579.

[5] P. LÉVY, *Processes Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris, 1964.

[6] E. PARZEN, *Time Series Analysis Papers*, Holden-Day, San Francisco, 1967.

[7] L. SCHWARTZ, *Application of Distributions to the Theory of Elementary Particles in Quantum Mechanics*, Gordon and Breach, New York, 1967.

[8] F. TRÈVES, *Topological Vector Spaces, Distributions and Kernels*, Academic Press, New York, 1967.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0008 $01.00/0

BOUNDED SOLUTIONS OF FINITE DIMENSIONAL APPROXIMATIONS TO THE BOUSSINESQ EQUATIONS*

JAMES H. CURRY†

Abstract. A class of finite systems of nonlinear ordinary differential equations are derived which yield finite dimensional approximation to the solutions of the Boussinesq equations. Such finite solutions, to these evolution equations, can be associated with trajectories in a phase space defined by the amplitudes of the components occurring in the differential equations. Once we make a passage to a phase space description of a system, a natural question which arises concerns the long time behavior of trajectories. In this paper we give sufficient conditions for a large class of approximate solutions to the Boussinesq equations to remain bounded for all time.

Introduction. In a fundamental paper [3] E. N. Lorenz establishes that a certain quadratic dynamical system has only bounded solutions. As a consequence of this result and the observation that the divergence of the vector field is nonpositive, he was able to prove the existence of a limiting surface for the system which he studied. Then using numerical techniques, Lorenz gave strong evidence for nontrivial dynamics for a physically interesting system arising from atmospheric convection. Recently Guckenheimer in [4] and Williams [12] have been able to prove that there are uncountably many topologically distinct attractors of the type discovered by Lorenz.

In this paper we establish that Lorenz's results on the existence of a limit surface is not an artifact of the small system of equation which he studies. Specifically we prove that a general class of quadratic dynamical systems arising from the equations of atmospheric convection do have only bounded solutions. It is an easy consequence of our results that the system studied by Lorenz has only bounded solutions. It is also the case that the divergence of all the vector fields which have bounded solution is nonpositive and therefore all trajectories must tend to an attracting set. Whether it is possible to extend the results of Guckenheimer and Williams to these higher dimensional limit surfaces is unknown.

Recent numerical experiments by the author, which shall be reported on elsewhere, indicate that the limiting surface first discovered by Lorenz most likely has higher dimensional analogues; however, one major difference is how the system transitions to turbulence.

In § 1 we introduce the equations from which Lorenz derived his three component system. Also in this section we provide a brief discussion of the evolution equations, geometry, and boundary conditions which we shall assume throughout the remainder of this paper. In § 2 we prove our main result, which establishes sufficient conditions for a large class of truncations (called complete) to the equations of motion to have only bounded solutions. Then in Appendix A we establish the definitions of the relevant physical parameters and indicate how the governing equations may be transformed into a suitable dimensionless form.

1. Preliminaries. The equations which govern convective motion in a fluid layer heated from below are the Navier–Stokes and heat conduction equations. We consider these equations in the Boussinesq approximation. In Appendix A it is shown

* Received by the editors April 19, 1977, and in revised form August 2, 1977.

† National Center for Atmospheric Research, Boulder, Colorado 80307. On leave from Department of Mathematics, Howard University, Washington, DC. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

how to rewrite the two dimensional Boussinesq equations into the following dimensionless form:

$$
(1.1)\qquad
\begin{aligned}
(\Delta\psi)_t &= -\frac{\partial(\psi, \Delta\psi)}{\partial(x, z)} + \sigma\Delta^2\psi + \sigma\frac{\partial\theta}{\partial x},\\
(\theta)_t &= -\frac{\partial(\psi, \theta)}{\partial(x, z)} + R\frac{\partial\psi}{\partial x} + \Delta\theta
\end{aligned}
$$

where $\psi(x, z, t)$ and $\theta(x, z, t)$ denote the stream function and departure of temperature from a linear profile while σ and R denote the Prandtl and Rayleigh numbers respectively (for the definition of the Prandtl and Rayleigh numbers, we refer the reader to Appendix A). In the remainder of this paper we shall consider the convection equations in the form (1.1).

We shall assume that all fluid motion is confined to the region $\mathscr{D} = \{(x, z): 0 \leqq x \leqq 2\pi/a, 0 \leqq z \leqq \pi\}$, and impose periodic boundary conditions (period $2\pi/a$) in the horizontal direction and free boundary conditions on the surfaces $z = 0$, $z = \pi$. Specifically these latter conditions are

$$\theta(x, 0) = \theta(x, \pi) = 0, \quad \psi(x, 0) = \psi(x, \pi) = 0, \quad \Delta\psi(x, 0) = \Delta\psi(x, \pi) = 0.$$

Although the free boundary conditions are perhaps not the most interesting from the physical point of view, we have chosen them in preference to rigid boundary conditions because they greatly simplify computations.

Also in the interest of simplicity, we have restricted our attention to a particular class of solutions—those for which ψ and $\partial\theta/\partial x$ vanish identically for $x = 0$ and $x = 2\pi/a$, i.e., for which the horizontal components of the velocity and temperature gradient vanish. It can be shown that if these conditions hold initially, they hold at later times as well.

A general representation for ψ and θ which is consistent with the requirements of the previous paragraphs is

$$
(1.2)\qquad
\begin{aligned}
\psi(x, z, t) &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \psi_{mn} \sin(amx)\sin(nz),\\
\theta(x, z, t) &= \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} \theta_{mn} \cos(amx)\sin(nz)
\end{aligned}
$$

where ψ_{mn} and θ_{mn} are functions of time alone.

If we now substitute (1.2) into (1.1) we get a system of infinitely many coupled nonlinear ordinary differential equations for the components of $\psi(x, z, t)$ and $\theta(x, z, t)$.

Let $\tilde{N} = N \cup \{0\}$ be the set of nonnegative integers. Suppose that $\mathscr{T}$ is a nonempty subset of $\tilde{N} \times \tilde{N}$. By a truncation of the convection equations, we shall mean a system of ordinary differential equations obtained by substituting ψ and θ from (1) into (1.1) and setting all ψ_{mn}, θ_{mn} terms equal to zero unless $(m, n) \in \mathscr{T}$. Specifically we shall consider those solutions $\psi(x, z, t)$, $\theta(x, z, t)$ to (1.1) such that

$$\psi(x, z, t) = \sum_{\substack{(m,n)\in\mathscr{T}\\ m\neq 0}} \psi_{mn} \sin(amx)\sin(nz),$$

$$\theta(x, z, t) = \sum_{(m,n)\in\mathscr{T}} \theta_{mn} \cos(amx)\sin(nz)$$

where $\mathscr{T}$ is a finite subset of $\tilde{N} \times \tilde{N}$.

We may write our system of ordinary differential equations symbolically as

$$\dot{X} = F(X) \tag{1.3}$$

where $\dot{X}$ is a vector whose entries are the time derivatives of the components of ψ and θ, $F(X) = \begin{pmatrix} \dot{\psi} \\ \dot{\theta} \end{pmatrix}$ where $\dot{\psi} = (\dot{\psi}_{mn})$ for $m \neq 0$ and $(m, n) \in \mathscr{T}$, and where $\dot{\theta}$ is defined analogously.

The set of all those points $X = \begin{pmatrix} \psi \\ \theta \end{pmatrix}$ will be the phase space for our system, and we shall denote by $T^t(X)$ the solution curve to (1.3) which passes through X at time zero.

2. Main result. Prior to establishing our main result we state three lemmas whose proofs are elementary and a proposition which will be an essential step in the proof of our main result.

LEMMA 1. *Let f, g be twice continuously differentiable functions on $\{(x, z): 0 \leqq x \leqq 2\pi/a, 0 \leqq z \leqq \pi\}$ which are periodic in x with period $2\pi/a$ and which vanish when $z = 0$ and $z = \pi$. Then*

$$\int_0^{2\pi/a} dx \int_0^{\pi} dz \frac{\partial(f, g)}{\partial(x, z)} = 0.$$

COROLLARY. *If f, g are defined as in Lemma 1, then*

$$\int_0^{2\pi/a} dx \int_0^{\pi} dz\, f \frac{\partial(f, g)}{\partial(x, z)} = 0.$$

LEMMA 2. *If*

$$f(x, z) = \sum_{(m,n) \in \mathscr{T}} a_{mn} \cos(amx) \sin(nz),$$

$$g(x, z) = \sum_{(m,n) \in \mathscr{T}} b_{mn} \sin(amx) \sin(nz),$$

then

1) $$\int_0^{2\pi/a} dx \int_0^{\pi} dz\, f^2 = \frac{\pi^2}{2a} \sum_{(m,n) \in \mathscr{T}} a_{mn}^2$$

2) $$\int_0^{2\pi/a} dx \int_0^{\pi} dz\, g\, \Delta g = -\frac{\pi^2}{2a} \sum_{\substack{(m,n) \in \mathscr{T} \\ m \neq 0}} \rho_{mn} b_{mn}^2 \quad \textit{where } \rho_{mn} = a^2 m^2 + n^2$$

3) $$\int_0^{2\pi/a} dx \int_0^{\pi} dz\, f \frac{\partial g}{\partial x} = -\int_0^{2\pi/a} dx \int_0^{\pi} dz\, g \frac{\partial f}{\partial x} = \frac{\pi^2}{2} \sum_{(m,n) \in \mathscr{T}} a_{mn} b_{mn} m.$$

By $\partial(\psi, \theta)^T / \partial(x, z)$ we mean the function obtained by expanding the Jacobian of ψ and θ in a Fourier series and dropping those terms with $(m, n) \notin \mathscr{T}$.

LEMMA 3. *Let $\mathscr{T}$, ψ, and θ be defined as above; then*

$$\int_0^{2\pi/a} dx \int_0^{\pi} dz\, \theta \frac{\partial(\psi, \theta)^T}{(x, z)} = 0, \qquad \int_0^{2\pi/z} dx \int_0^{\pi} dz\, \psi \frac{\partial(\psi, \Delta\psi)^T}{\partial(x, z)} = 0.$$

PROPOSITION. *Let Q be a continuously differentiable function defined on the state space of our system and suppose that there exists a constant Q_1 such that*

$$\frac{d}{dt}Q(T^tx)\Big|_{t=0}<0$$

whenever $Q(x)\geqq Q_1$. Then for every x in our phase space, $\sup_{t\geqq 0} Q(T^tx)<\infty$.

Proof. If x_0 is any point in our phase space, then either $Q(x_0)\geqq Q_1$ or $Q(x_0)<Q_1$. Suppose that $Q(x_0)\geqq Q_1$; then by hypothesis

$$\frac{d}{dt}Q(T^tx_0)\Big|_{t=0}<0;$$

therefore in a neighborhood about $t=0$ $Q(T^tx_0)$ is a decreasing function for all t in this neighborhood. Further, $Q(T^tx_0)$ must continue to decrease until $Q(T^tx_0)<Q_1$.

If on the other hand, $Q(x_0)<Q_1$, then $Q(T^tx_0)\leqq Q_1$ for all t. Since if for some t_0, $Q(T^{t_0}x_0)=Q_1$, then by hypothesis

$$\frac{d}{dt}Q(T^sy_0)\Big|_{s=0}<0 \quad \text{where } y_0=T^{t_0}x_0.$$

If we now argue as above, we conclude that if $Q(x_0)<Q_1$ then $Q(T^tx_0)\leqq Q_1$ for all t, so the proof of the Proposition is now complete.

Now let $\mathcal{T}$ be as previously defined and $N=\sup\{n\colon (m,n)\in\mathcal{T} \text{ for some } m\neq 0\}$. We will say that *$\mathcal{T}$ defines a complete truncation* if $(0,2n)\in\mathcal{T}$ for $n=1,2,3,\cdots,N$. We may now state our main result.

THEOREM. *Every solution of the equations of motion for a complete truncation is bounded for $t\geqq 0$.*

Proof. In order to establish the theorem it is sufficient to show that an equivalent system of equations has all solutions bounded for $t\geqq 0$.

Let us examine the complete truncation to the convection equations associated with $\mathcal{T}$. For $(m,n)\in\mathcal{T}$ we define S_{mn}, T_{mn} by the following rules: for m different from zero,

$$S_{mn}=\psi_{mn}, \qquad T_{mn}=\theta_{mn};$$

for m equal to zero and n odd,

$$T_{0n}=\theta_{0n},$$

while for the remaining case

$$T_{02n}+C_{02n}=\theta_{02n}.$$

Given the above definitions of S_{mn}, T_{mn} we define

$$T(x,z,t)=\sum_{(m,n)\in\mathcal{T}} T_{mn}\cos(amx)\sin(nz)$$

$$S(x,z,t)=\sum_{\substack{(m,n)\in\mathcal{T}\\ m\neq 0}} S_{mn}\sin(amx)\sin(nz)$$

From the above definitions it is apparent that

$$\theta(x,z,t)=T(x,z,t)-c(z), \qquad \psi(x,z,t)=S(x,z,t)$$

where

$$c(z)=\sum_{n=1}^{N} C_{02n}\sin(2nz),$$

N is the number in the definition of a complete truncation and $C_{02n}=-2R/n$. If we now rewrite the convection equations in terms of S and T we have

$$(\Delta S)_t=-\frac{\partial(S,\Delta S)^T}{\partial(x,z)}+\sigma\Delta^2 S+\sigma\frac{\partial T}{\partial x},$$

$$(T)_t=-\frac{\partial(S,T)^T}{\partial(x,z)}+\frac{\partial(c(z),S)^T}{\partial(x,z)}+R\frac{\partial S}{\partial x}+\Delta(T-c(z)).$$

Let

$$Q=\frac{1}{2}\int_0^{2\pi/a}dx\int_0^{\pi}dz\left\{T^2-\frac{R}{\sigma}S\,\Delta S\right\}.$$

From Lemmas 1 and 3 it follows that

$$\frac{dQ}{dt}=\int_0^{2\pi/a}dx\int_0^{\pi}dz\left\{T\left[-\frac{\partial c}{\partial z}\frac{\partial S}{\partial x}+R\frac{\partial S}{\partial x}+\Delta(T-c(z))\right]-RS\left[\frac{\partial T}{\partial x}+\Delta^2 S\right]\right\}.$$

Let us make a closer examination of the first term in the expression for dQ/dt, i.e., $-T(\partial S/\partial x)\,\partial c/\partial z$. If we recall the definitions of T, S and c, while at the same time making use of elementary trigonometric identities to change products of sines into sums or differences of cosine terms, we find that

$$-T\frac{\partial S}{\partial x}\frac{\partial c}{\partial z}=-a\sum_{(\alpha,\beta)\in\tau}\sum_{(\gamma,\delta)\in\mathscr{T}}\sum_{n=1}^{N}\cdot(C_{0zn}\cdot n\gamma T_{\alpha\beta}\delta_{\gamma\delta})\cdot(\cos((\beta-\delta)z)$$
$$-\cos((\beta+\delta)z))\cos(2n)\cdot\cos(a\alpha x)\cos(a\gamma x).$$

Hence

$$\int_0^{2\pi/a}dx\;T\frac{\partial S}{\partial x}\frac{\partial c}{\partial z}=2\pi R\sum_{(m,\beta)\in\mathscr{T}}\sum_{(m,\delta)\in\mathscr{T}}\sum_{n=1}^{N}(mT_{m\beta}S_{m\delta})$$
$$\cdot(\cos((\beta-\delta)z)-\cos((\beta+\delta)z))\cos(2n).$$

Prior to performing the integrations in the z-variable we make several observations: for fixed β, δ,

$$\int_0^{\pi}dz\sum_{n=1}\cos(\beta-\delta)z-\cos(\beta+\delta)z\cos(2n)=0$$

unless, β, δ have the same parity. Further if (m,β) and (m,δ) are elements of $\mathscr{T}$ for $m\neq 0$, and β, δ are distinct but have the same parity, then $\exists n_1, n_2\in\{1,2,\cdots,N\}$ such that $2n_1=\beta+\delta$ and $2n_2=|\beta-\delta|$. If we now exploit the above observations, it is apparent that

$$-\int_0^{2\pi/a}dx\int_0^{\pi}dz\;T\frac{\partial S}{\partial x}\frac{\partial c}{\partial z}=\frac{\pi^2 R}{2}\sum_{(m,n)\in\mathscr{T}}mT_{mn}S_{mn}.$$

With the aid of Lemma 2 and the above remarks we have that

$$\frac{dQ}{dt}=\int_0^{2\pi/a} dx \int_0^{\pi} dz\, \{T\,\Delta T - RS\,\Delta^2 S - T\,\Delta c(z)\}$$

$$= -\frac{\pi^2}{2a} \sum_{\substack{(m,n)\in\mathcal{T}\\ m\neq 0}} (R\rho_{mn}^2 S_{mn}^2 + \rho_{mn} T_{mn}^2) - \frac{\pi^2}{2a} \sum_{n\ \mathrm{odd}} \rho_{0n} T_{0n}^2 - \frac{\pi^2}{a} \sum_{n=1}^{N} \rho_{02n} T_{02n}^2$$

$$+\frac{\pi^2}{a} \sum_{N=1}^{N} \rho_{02n} T_{02n} C_{02n}.$$

If we now call the last term above B_2 and the remaining terms $-B_1$, then

$$\frac{dQ}{dt} = -B_1 + B_2,$$

where B_1 is a strictly positive definite quadratic form in the S_{mn}, T_{mn} and B_2 is linear in T_{mn}. Therefore there exist finite values P_1, P_2 such that $P_1 B_1 \geqq Q$ and $|B_2| \leqq P_2\sqrt{Q}$ so

$$\frac{dQ}{dt} = -B_1 + B_2 \leqq -\frac{1}{P_1} Q + P_2\sqrt{Q}$$

and the right-hand side of the last inequality will be less than zero provided $Q > P_1^2 P_2^2$. The boundedness of solutions now follows from the previous proposition.

Appendix A. In this Appendix we transform the Boussinesq equations into a dimensionless form; the relevant equations are:

$$\text{(A.1)}\qquad \begin{aligned} (\tilde{\Delta}\tilde{\psi})_{\tilde{t}} &= -\frac{\partial(\tilde{\psi}, \tilde{\Delta}\tilde{\psi})}{(\tilde{x}, \tilde{z})} + \tilde{\nu}\tilde{\Delta}^2\tilde{\psi} + g\alpha\frac{\partial\tilde{\theta}}{\partial\tilde{x}}, \\ (\tilde{\theta})_{\tilde{t}} &= -\frac{\partial(\tilde{\psi}, \tilde{\theta})}{(\tilde{x}, \tilde{z})} + \frac{\tau}{H}\frac{\partial\psi}{\partial x} + \kappa\tilde{\Delta}\tilde{\psi} \end{aligned}$$

where $\tilde{\psi}$ and $\tilde{\theta}$ are functions of $\tilde{x}, \tilde{z}$ and $\tilde{t}$ which denote the stream function and departure of the temperature from linearity respectively. The constants: g, α, ν, κ, τ, and H denote respectively the acceleration due to gravity, the coefficient of thermal expansion, kinematic viscosity, thermal conductivity, the temperature difference between the plates and depth of the fluid layer. We shall use Δ to denote the Laplacian in the variables x and z.

We now introduce the dimensionless variables x, z, and t, where

$$\text{(A.2)}\qquad x = \tilde{x}/c_1, \qquad z = \frac{\tilde{z}}{c_1}, \qquad t = \frac{\tilde{t}}{c_2};$$

c_1 and c_2 are constants having the dimensions of length and time, and shall be determined later.

If we now substitute (A.2) into (A.1) we have

$$\text{(A.3)}\qquad \begin{aligned} (\tilde{\Delta}\tilde{\psi})_{\tilde{t}} &= -\frac{c_2}{c_1}\frac{\partial(\psi, \Delta\psi)}{(x, z)} + \nu\frac{c_2}{c_1}\tilde{\Delta}^2\tilde{\psi} + g\alpha c_1 c_2\frac{\partial\tilde{\theta}}{\partial x}, \\ (\tilde{\theta})_{\tilde{t}} &= -\frac{c_2}{c_1}\frac{\partial(\tilde{\psi}, \tilde{\theta})}{(x, z)} + \frac{\tau}{H}\frac{c_2}{c_1}\frac{\partial\tilde{\psi}}{x} + \kappa\frac{c_2}{c_1}\Delta\tilde{\theta}. \end{aligned}$$

It will be convenient for us to rescale $\tilde{\psi}$ and $\tilde{\theta}$ and call the rescaled functions ψ and θ where

$$\psi = \tilde{\psi}/k_1 \quad \text{and} \quad \theta = \tilde{\theta}/k_2.$$

Here k_1 and k_2 are constants which shall be determined shortly.

If we now rewrite (A.3) using the definition of ψ and θ given above, we have

$$\text{(A.4)} \qquad \begin{aligned} (\Delta\psi)_t &= -\frac{c_2}{c_1} k_1 \frac{\partial(\psi, \Delta\psi)}{\partial(x, z)} + \frac{c_2}{c_1} \Delta^2 \psi + g\alpha \frac{c_1 c_2}{\kappa} \frac{k_2}{k_1} \frac{\partial \theta}{\partial x} \\ (\theta)_t &= -\frac{c_2}{c_1^2} k_1 \frac{\partial(\psi, \theta)}{\partial(x, z)} + \frac{\tau}{H} \frac{c_2}{c_1} \frac{k_1}{k_2} \frac{\partial \psi}{\partial x} + \frac{c_2}{c_1^2} \kappa \, \Delta\psi \end{aligned}$$

where we have used Δ to denote the Laplacian in the dimensionless variables.

There are many possible relations which can be imposed on the constants in (A.4); if we assume that $c_2 = c_1^2 k_1$ and $k_1 = \kappa$ and suppose that $k_2 = \nu/(gc_1c_2)$ where we let $c_1 = H/\pi$, then the dimensionless convection equations beome

$$\text{(A.5)} \qquad \begin{aligned} (\Delta\psi)_t &= -\frac{\partial(\psi, \Delta\psi)}{\partial(x, z)} + \frac{\nu}{\kappa} \Delta^2 \psi + \frac{\nu}{\kappa} \frac{\partial \theta}{\partial x}, \\ (\theta)_t &= -\frac{\partial(\psi, \theta)}{(x, z)} + \frac{\tau g \alpha}{\kappa \nu \pi} \frac{H^3}{\pi} \frac{\partial \psi}{\partial x} + \Delta\theta. \end{aligned}$$

The ratio of viscosity to conductivity (ν/κ) is called the *Prandtl number*, which shall be denoted by σ. The coefficient for the $\partial\psi/\partial x$ term which appears in the equality for $(\theta)_t$ is called the *Rayleigh number* and shall be denoted by R.

Acknowledgment. The author would like to take this opportunity to express his thanks to R. C. J. Somerville for several helpful discussions, and to Professor A. Chorin for helpful discussions and pointing out the importance and beauty of the work of D. D. Joseph. Finally, the author would like to express his highest thanks and regards to his dissertation advisor, Professor O. E. Lanford III, for many stimulating discussions and for making the period in which this work was done très amusante.

REFERENCES

[1] P. HARTMAN, *Ordinary Differential Equations*, John Wiley, New York, 1964.
[2] O. E. LANFORD III, unpublished notes, University of California, Berkeley.
[3] E. N. LORENZ, *Deterministic Nonperiodic Flows*, J. Atmospheric Sci., 20 (1963), pp. 130–141.
[4] J. MARSDEN AND M. MCCRACKEN, *The Hopf Bifurcation and its Applications*, Lecture Notes in the Applied Mathematical Sciences, vol. 19, Springer-Verlag, New York, 1976.
[5] J. B. MCLAUGHLIN AND P. C. MARTIN, Phys. Rev. A, 13 (1975), pp. 186–203.
[6] D. RUELLE AND F. TAKENS, *On the nature of turbule*, Comm. Math. Phys., 20 (1971), pp. 167–192.
[7] GEORGE VERONIS, *Large-amplitude Benard convection*, J. Fluid Mech., 20 (1966), Part 1, pp. 49–68.
[8] NEMYTSKII AND STEPANOV, *Qualitative Theory of Differential Equations*, Princeton University Press, Princeton, NJ, 1960.
[9] D. D. JOSEPH, *On the stability of the Boussinesq equations*, Arch. Rational Mech. Anal., 20 (1965), pp. 59–71.
[10] ———, *Stability of Fluid Motions I*, Tracts in Natural Philosophy, vol. 27, Springer-Verlag, New York, 1976.
[11] S. CHANDRASEKHAR, *Hydrodynamic and Hydromagnetic Stability*, Oxford University Press, Oxford, England, 1961.
[12] R. F. WILLIAMS, *Structure of Lorenz attractors*, preprint.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0009 $01.00/0

A NOTE ON A PAPER BY L. CARLITZ*

J. CIGLER†

Abstract. We give a simple derivation of the generating functions $\sum [s_n(nx+y)/n!]z^n$ for Sheffer sets $\{s_n(x)\}$.

In a recent paper [1] L. Carlitz has obtained some interesting generating functions. The purpose of this note is to show that some of his formulas can be best understood within Rota's theory of Sheffer sets [2]. In order to avoid repetitions of well-known facts we follow the notation and terminology of Rota's paper [2].

Let

$$Q = g(D) = \frac{a_1}{1!}D + \frac{a_2}{2!}D^2 + \cdots$$

be a delta operator and let $G(D)$ be the inverse formal power series which satisfies $G(g(D)) = g(G(D)) = D$. Then the set of polynomials of binomial type corresponding to Q, which we denote by $\{q_n(x)\}$ has the generating function

$$\sum_{n=0}^{\infty} \frac{q_n(x)}{n!} D^n = e^{xG(D)}. \qquad \int_0^{2\pi/a} dx\, T$$

It is obvious that $q_n(y) = (e^{yG(D)}x^n)_{x=0}$.

Let now

$$s(D) = s_0 + \frac{s_1}{1!}D + \frac{s_2}{2!}D^2 + \cdots, \qquad s_0 \neq 0,$$

be an invertible operator and consider the polynomials $s_n(x)$ defined by the generating function

$$\frac{1}{s(D)} e^{xG(D)} = \sum_{n=0}^{\infty} \frac{s_n(x)}{n!} D^n.$$

This set of polynomials $\{s_n(x)\}$ is called the Sheffer set relative to Q and the invertible operator $s(D)$.

THEOREM. *Let $\{s_n(x)\}$ be the Sheffer set relative to $Q = g(D)$ and the invertible operator $s(D)$. Then*

$$\sum_{n=0}^{\infty} \frac{s_n(nx+y)}{n!} (De^{-xG(D)})^n = \frac{1}{s(D)} \frac{e^{yG(D)}}{1 - xDG'(D)}.$$

Proof. For each complex number a the operator $R = De^{-aG(D)}$ is a delta operator. Let $\{p_n(x)\}$ be the corresponding set of polynomials of binomial type. By [2, Thm. 4], we have

$$p_n(x) = R' e^{(n+1)aG(D)} x^n = (1 - aDG'(D)) e^{naG(D)} x^n.$$

Let now

$$\sum_{n=0}^{\infty} \frac{c_n}{n!} R^n = \frac{1}{s(D)} \frac{e^{bG(D)}}{1 - aDG'(D)}$$

* Received by the editors August 5, 1977.

† Mathematisches Institut, Universität Wien, Vienna, Austria.

be the expansion of the right-hand side with respect to the delta operator R. Then we have

$$c_n = \left[\frac{1}{s(D)} \frac{e^{bG(D)}}{1-aDG'(D)} p_n(x)\right]_{x=0}$$

by the first expansion theorem [2, Thm. 2]. This gives

$$c_n = \left[\frac{1}{s(D)} \frac{e^{bG(D)}}{1-aDG'(D)} (1-aDG'(D))\, e^{naG(D)} x^n\right]_{x=0}$$

$$= \left[\frac{1}{s(D)} e^{(na+b)G(D)} x^n\right]_{x=0} = s_n(na+b),$$

which proves the theorem.

By suitable choices of $G(D)$ and $s(D)$ most of the concrete examples of Carlitz's paper appear as special cases of our theorem.

REFERENCES

[1] L. CARLITZ, *A class of generating functions*, this Journal, 8 (1977), pp. 518–532.
[2] G. C. ROTA, D. KAHANER. AND A. ODLYZKO, *Finite operator calculus*, J. Math. Anal. Appl., 42 (1973), pp. 684–760.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0010 $01.00/0

IDEAL INVERSION FORMULAE FOR THE FOURIER TRANSFORM*

F. J. WILSON†

Abstract. In this paper a method is given for generating new ideal inversion formulae for the Fourier transform starting from known Laplace inversion formulae. Two new ideal inversion formulae for the Fourier transform are developed.

1. Introduction. In [1] Cooper considers integral transforms of the form

$$f_\lambda(u) = \int_{-\infty}^{\infty} k(u, v, \lambda) F(v)\, dv \tag{1.1}$$

for a function F in $L^{p'}(-\infty, \infty)$, where $1/p + 1/p' = 1$ and $1 \leqq p \leqq 2$, and shows that for certain kernels, $k(u, v, \lambda)$, the boundedness of the set of functions $\{f_\lambda\}$ in $L^p(-\infty, \infty)$ is necessary and sufficient for F to be the Fourier transform of a function of $L^p(-\infty, \infty)$. Cooper calls the integral transform (1.1) an ideal inversion formula for the Fourier transform if it satisfies the two conditions:

(i) the boundedness of the set of functions $f_\lambda(u)$ in $L^p(-\infty, \infty)$ is necessary and sufficient for F to be the Fourier transform of an f in $L^p(-\infty, \infty)$,

(ii) if F is the Fourier transform of an f in $L^p(-\infty, \infty)$ then $f_\lambda \to f$ strongly in $L^p(-\infty, \infty)$ for $1 \leqq p \leqq 2$ and $f_\lambda(t) \to f(t)$ as $\lambda \to \infty$ at every point in the Lebesgue set of f.

We will call a kernel, $k(u, v, \lambda)$, which gives rise to an ideal inversion formula an ideal inversion kernel for the Fourier transform.

One particular ideal inversion formula considered by Cooper in [1] is

$$f_\lambda(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\lambda^{\lambda+1}}{(\lambda - iuv)^{\lambda+1}} F(v)\, dv. \tag{1.2}$$

He also shows that if we consider $F(v)$ as an analytic function and λ as a positive integer, then (1.2) reduces to

$$\frac{\sqrt{2\pi}}{\lambda!} \frac{(-1)^\lambda}{i} \left(\frac{\lambda}{iu}\right)^{\lambda+1} F^{(\lambda)}\left(\frac{\lambda}{iu}\right) \to f(u) \tag{1.3}$$

over the positive reals as $\lambda \to \infty$. Cooper notes that (1.3) is clearly related to the Post–Widder inversion formula for the Laplace transform [4, p. 288]

$$\frac{(-1)^\lambda}{\lambda!} \left(\frac{\lambda}{u}\right)^{\lambda+1} \bar{f}^{(\lambda)}\left(\frac{\lambda}{u}\right) \to f(u) \tag{1.4}$$

over the positive reals as $\lambda \to \infty$, in which we are trying to find the inverse Laplace transform of the function $\bar{f}$.

In this paper we show that starting with known Laplace inversion formulae we can generate, in certain cases, new ideal inversion formulae for the Fourier transform. To illustrate the method if, for example, we started with the Laplace inversion formula (1.4), working nonrigorously, we could by substitution obtain a possible inversion formula for the Fourier transform, (1.3), and then using Cauchy's integral formula obtain (1.2), which is an integral transform of the required type. We would then have

* Received by the editors March 29, 1977.

† Department of Mathematics, University of Wales Institute of Science and Technology, Cardiff CF1 3NU, Wales, U.K.

to check that the kernel so obtained was an ideal inversion kernel for the Fourier transform.

In this paper we consider two inversion formulae for the Laplace transform. The first inversion formula, (4.1), gives rise directly to an ideal inversion kernel. The second inversion formula, (5.1), does not directly give rise to an ideal inversion kernel but by adjusting the kernel it is possible to form an ideal inversion kernel.

2. Notation. Throughout the paper p will be restricted to $1 \leqq p \leqq 2$ and p' will be given by $1/p + 1/p' = 1$.

The Fourier transform of f will be denoted by $\hat{f}$ where

$$\hat{f}(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)\, e^{-ixy}\, dx$$

and the Laplace transform of f will be denoted by $\mathring{f}$ where

$$\mathring{f}(w) = \int_0^{\infty} f(t)\, e^{-wt}\, dt.$$

For the Fourier transforms of the kernel $k(u, v, \lambda)$ we will write

$$K_1(t, v, \lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k(u, v, \lambda)\, e^{-iut}\, du$$

and

$$K_2(u, t, \lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k(u, v, \lambda)\, e^{-ivt}\, dv.$$

If $K_1(t, v, \lambda)$ satisfies

$$\lim_{\lambda \to \infty} \int_{-\infty}^{\infty} f(t)\, dt \int_{-\infty}^{\infty} K_1(t, v, \lambda) g(v)\, dv = \int_{-\infty}^{\infty} f(t) g(t)\, dt$$

for all f in Ch, where Ch is the set of characteristic functions of finite intervals, and g in $L^{p'}(-\infty, \infty)$ then we write that $K_1(t, v, \lambda)$ is an approximation kernel on $\{\mathrm{Ch}, L^{p'}\}$.

If

$$\int_{-\infty}^{\infty} K_2(u, t, \lambda) g(t)\, dt \to g(u)$$

as $\lambda \to \infty$ strongly in L^p for all g in L^p, we write that $K_2(u, t, \lambda)$ is a strong approximation kernel on $\{\cdot, L^p\}$.

3. Conditions for an ideal inversion kernel. In [1] Cooper considers several different conditions for $k(u, v, \lambda)$ to be an ideal inversion kernel for the Fourier transform depending on the form of the kernel and whether $K_1(t, v, \lambda)$ exists only as a generalized function. We will be using two of these conditions.

Firstly, $k(u, v, \lambda)$ satisfies condition A if

A1. $k(u, v, \lambda)$ belongs to $L^p(-\infty, \infty)$ as a function of v for almost all u and all $\lambda > \lambda_0$ and $k(u, v, \lambda)$ belongs to $L^r(-\infty, \infty)$ as a function of u for almost all v and all $\lambda > \lambda_0$, where $1 < r \leqq 2$.

A2. $\|k(u, v, \lambda)\|_p$ and $\|K_1(u, v, \lambda)\|_p$ as functions of v are $O(e^{mu^2})$ as $|u| \to \infty$ for some $m < \frac{1}{2}$ and all $\lambda > \lambda_0$.

A3. $K_1(t, v, \lambda)$ is an approximation kernel on $\{\mathrm{Ch}, L^{p'}\}$ where Ch is the set of characteristic functions of finite intervals.

A4. $K_2(u, t, \lambda)$ is a strong approximation kernel on $\{\cdot, L^p\}$.

Secondly, $k(u, v, \lambda)$ satisfies condition B if $k(u, v, \lambda)=e^{iuv}l(v, \lambda)$ and

B1. $k(u, v, \lambda)$ belongs to $L^p(-\infty, \infty)$ as a function of v for almost all u and all $\lambda > \lambda_0$.

B2. $\sqrt{2\pi}\, l(v, \lambda)\to 1$ as $\lambda\to\infty$ boundedly on every finite interval.

B3. $K_2(u, t, \lambda)$ is a strong approximation kernel on $\{\cdot, L^p\}$.

4. Extended Post–Widder ideal inversion kernel. Starting from the extended Post–Widder inversion formula for the Laplace transform [4, p. 295]

$$f(u)=\lim_{\lambda\to\infty}\frac{(-1)^\lambda(\lambda+\theta_\lambda)^{\lambda+1}}{\lambda!u^{\lambda+1}}\bar{f}^{(\lambda)}\left(\frac{\lambda+\theta_\lambda}{u}\right) \tag{4.1}$$

where $\theta_\lambda=o(\lambda)$ as $\lambda\to\infty$, we obtain, nonrigorously, an inversion formula for the Fourier transform by substituting

$$\hat{f}(w)=\frac{1}{\sqrt{2\pi}}\bar{f}(iw)$$

giving

$$f(u)=\lim_{\lambda\to\infty}\frac{\sqrt{2\pi}(-1)^\lambda(\lambda+\theta_\lambda)^{\lambda+1}}{i^\lambda\lambda!u^{\lambda+1}}\hat{f}^{(\lambda)}\left(\frac{\lambda+\theta_\lambda}{iu}\right). \tag{4.2}$$

Assuming that $\hat{f}$ is analytic and satisfies certain boundedness conditions we use Cauchy's integral formula for $\hat{f}^{(\lambda)}(z)$ to turn the possible inversion formula for the Fourier transform, (4.2), into the required form

$$f(u)=\lim_{\lambda\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{(\lambda+\theta_\lambda-iuv)^{\lambda+1}}\hat{f}(v)\,dv.$$

We call the kernel

$$k(u, v, \lambda)=\frac{1}{\sqrt{2\pi}}\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{(\lambda+\theta_\lambda-iuv)^{\lambda+1}}$$

where $\lambda_\lambda=o(\lambda)$ as $\lambda\to\infty$, the *extended Post–Widder kernel* and, as one would expect, it is closely related to the kernel used by Cooper [1, p. 292].

We now show that the extended Post–Widder kernel is an ideal inversion kernel for the Fourier transform by proving that it satisfies condition A.

First we need the Fourier transforms of $k(u, v, \lambda)$, one of which is given by

$$K_1(t, v, \lambda)=\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iut}}{(\lambda+\theta_\lambda-iuv)^{\lambda+1}}\,du.$$

Letting $z=t(\lambda+\theta_\lambda-iuv)/v$ we obtain

$$\begin{aligned}K_1(t, v, \lambda)&=\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{2\pi i|v|}e^{-t(\lambda+\theta_\lambda)/v}\left(\frac{t}{v}\right)^\lambda\int_{(t/v)(\lambda+\theta_\lambda-i\infty)}^{(t/v)(\lambda+\theta_\lambda+i\infty)}e^z z^{-\lambda-1}\,dz\\&=\begin{cases}\dfrac{(\lambda+\theta_\lambda)^{\lambda+1}e^{-t(\lambda+\theta_\lambda)/v}|t|^\lambda}{\Gamma(\lambda+1)|v|^{\lambda+1}}, & \dfrac{t}{v}\geqq 0,\\[2ex] 0, & \dfrac{t}{v}<0,\end{cases}\end{aligned}$$

and by symmetry $K_1(t, v, \lambda)=K_2(v, t, \lambda)$.

Clearly for large enough λ condition A1 is satisfied. Next

$$\|k(u,\cdot,\lambda)\|_p{}^p=\int_{-\infty}^{\infty}\left|\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{\sqrt{2\pi}\,(\lambda+\theta_\lambda-iuv)^{\lambda+1}}\right|^p dv \leqq \frac{2(2\pi)^{-p/2}(\lambda+\theta_\lambda)}{u}\frac{\pi}{2} \quad \text{for } \lambda>1+\frac{4}{p},$$

which is $O(e^{mu^2})$ as $|u|\to\infty$ for some $m<\frac{1}{2}$ for large enough λ, and

$$\|K_1(t,\cdot,\lambda)\|_p{}^p=\int_{-\infty}^{\infty}|K_1(t,v,\lambda)|^p\,dv=\frac{p(\lambda+\theta_\lambda)t^{1-\lambda p}}{\{\Gamma(\lambda+1)\}^p p^{p(\lambda+1)}}\Gamma(p\lambda+p-1),$$

which is $O(e^{mt^2})$ as $|t|\to\infty$ for some $m<\frac{1}{2}$ for large enough λ, and therefore condition A2 is satisfied. We now show that $K_1(t,v,\lambda)$ and $K_2(u,t,\lambda)$ are both strong approximation kernels on $\{\cdot, L^p\}$ as they satisfy the conditions of Lemma 2 in [1, p. 292]. Firstly,

$$\int_{-\infty}^{\infty}|K_1(t,v,\lambda)|\,dt=\int_{-\infty}^{\infty}|K_2(u,t,\lambda)|\,dt=1$$

and

$$\int_{-\infty}^{\infty}|K_1(t,v,\lambda)|\,dv=\int_{-\infty}^{\infty}|K_2(u,t,\lambda)|\,du=1+\frac{\theta_\lambda}{\lambda},$$

which are bounded in u, v, t and λ, for large enough λ. Secondly, for $0\leqq a<u<b$

$$\int_a^b K_2(u,t,\lambda)\,dt=\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{\Gamma(\lambda+1)}(I_1+I_2)$$

where

$$I_1=\int_{a/u}^{\lambda/(\lambda+\theta_\lambda)} e^{-x(\lambda+\theta_\lambda)}x^\lambda\,dx \quad\text{and}\quad I_2=\int_{\lambda/(\lambda+\theta_\lambda)}^{b/u} e^{-x(\lambda+\theta_\lambda)}x^\lambda\,dx.$$

For large λ we simplify I_1 and I_2 by using Laplace's asymptotic method for integrals [4, Thm. 2a, p. 277], giving

$$I_1\sim\frac{e^{-\lambda}\lambda^\lambda\lambda^{1/2}}{(\lambda+\theta_\lambda)^{\lambda+1}}\sqrt{\pi/2}$$

and similarly for I_2. Therefore for $0\leqq a<u<b$

$$\int_a^b K_2(u,t,\lambda)\,dt\to 1$$

as $\lambda\to\infty$. Similar results can be proved for $a<u<b\leqq 0$ and hence for $a<u<b$. If $u<a$ then we have

$$\int_a^b K_2(u,t,\lambda)\,dt\leqq\int_a^{\infty}\frac{(\lambda+\theta_\lambda)^{\lambda+1}e^{-t(\lambda+\theta_\lambda)/u}t^\lambda}{\Gamma(\lambda+1)u^{\lambda+1}}\,dt$$

$$\sim\frac{(\lambda+\theta_\lambda)^{\lambda+1}}{\Gamma(\lambda+1)}\lambda^{-1}\left(\frac{u}{a}-1-\frac{\theta_\lambda}{\lambda}\right)^{-\lambda}e^{-a(\lambda+\theta_\lambda)/u}\left(\frac{a}{u}\right)^\lambda,$$

using Laplace's asymptotic method for integrals [2, p. 65], and hence the integral is dominated by a multiple of $u^{-\lambda}$ for large λ and tends to zero as $\lambda\to\infty$. A similar result holds for $\int_a^b K_2(u,t,\lambda)\,dt$ for $b<u$. Therefore $\int_a^b K_2(u,t,\lambda)\,dt$ is dominated by a function of $L^p(-\infty,\infty)$ and tends to $\chi_{(a,b)}(u)$ almost everywhere as $\lambda\to\infty$. Hence by [1, Lemma 2, p. 292], $K_2(u,t,\lambda)$ is a strong approximation kernel on $\{\cdot, L^p\}$. Since

$K_1(t, v, \lambda) = K_2(v, t, \lambda)$ the same results hold for $K_1(t, v, \lambda)$. Therefore $k(u, v, \lambda)$ satisfies conditions A3 and A4 and hence we have proved that the extended Post–Widder kernel

$$k(u, v, \lambda) = \frac{1}{\sqrt{2\pi}} \frac{(\lambda + \theta_\lambda)^{\lambda+1}}{(\lambda + \theta_\lambda - iuv)^{\lambda+1}}$$

where $\theta_\lambda = o(\lambda)$ as $\lambda \to \infty$ is an ideal inversion kernel for the Fourier transform.

5. Extended Phragmen ideal inversion kernel. Starting from the Phragmen inversion formula for the Laplace transform [3, p. 133]

$$\int_0^u f(t)\, dt = \lim_{\lambda \to \infty} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!} e^{n\lambda u} \hat{f}(n\lambda) \tag{5.1}$$

for $u > 0$, we obtain, nonrigorously, an inversion formula for the Fourier transform by substituting

$$\hat{f}(w) = \frac{1}{\sqrt{2\pi}} \hat{f}(iw)$$

and, as before, we turn the possible inversion formula for the Fourier transform into the required form by using Cauchy's integral formula, assuming that $\hat{f}$ is analytic and satisfies certain boundedness conditions, giving

$$\int_0^u f(t)\, dt = \lim_{\lambda \to \infty} \frac{1}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!} e^{n\lambda u} \int_{-\infty}^{\infty} \frac{\hat{f}(v)}{(iv - n\lambda)} dv.$$

Differentiating with respect to u and interchanging the integral and summation signs we get

$$f(u) = \lim_{\lambda \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(v) \left\{ \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \lambda\, e^{n\lambda u}}{(n-1)!(iv - n\lambda)} \right\} dv$$

and we obtain the kernel

$$k'(u, v, \lambda) = \frac{1}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \lambda\, e^{n\lambda u}}{(n-1)!(iv - n\lambda)}.$$

We have labeled this kernel $k'(u, v, \lambda)$ because even though it is not an ideal inversion kernel itself, it can be used to generate one. We can show that $k'(u, v, \lambda)$ is not an ideal inversion kernel for the Fourier transform by examining its Fourier transform with respect to v, $K_2'(u, t, \lambda)$:

$$K_2'(u, t, \lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ivt} k'(u, v, \lambda)\, dv = \begin{cases} \lambda\, e^{\lambda(u-t)} e^{-e^{\lambda(u-t)}}, & t \geqq 0, \\ 0, & t < 0, \end{cases}$$

which cannot be a strong approximation kernel.

To obtain an ideal inversion kernel for the Fourier transform we define

$$K_2(u, t, \lambda) = \lambda\, e^{\lambda(u-t)} e^{-e^{\lambda(u-t)}} \quad \text{for all } t$$

and we find the corresponding kernel, $k(u, v, \lambda)$, by taking the inverse Fourier transform of $K_2(u, t, \lambda)$:

$$k(u, v, \lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lambda\, e^{\lambda(u-t)} e^{-e^{\lambda(u-t)}} e^{ivt}\, dt = \frac{e^{iuv}}{\sqrt{2\pi}} \Gamma\left(1 - \frac{iv}{\lambda}\right).$$

We now show that this kernel, which we will call the *extended Phragmen kernel*, is an ideal inversion kernel for the Fourier transform by proving that it satisfies condition B.

First, for any finite interval (A, B)

$$\int_A^B |k(u, v, \lambda)|^p \, dv = (2\pi)^{-p/2} \int_A^B \left| e^{iuv} \Gamma\left(1 - \frac{iv}{\lambda}\right) \right|^p dv \leqq (2\pi)^{-p/2}(B - A).$$

For large v/λ

$$\left| \Gamma\left(1 - \frac{iv}{\lambda}\right) \right| \sim \sqrt{2\pi} \left| \frac{v}{\lambda} \right|^{1/2} e^{-(\pi/2)|v/\lambda|}.$$

Therefore for large positive B, $\int_B^\infty |\Gamma(1 - iv/\lambda)|^p \, dv$ is bounded, and similarly for $\int_{-\infty}^A |\Gamma(1 - iv/\lambda)|^p \, dv$ for large negative A. Therefore $k(u, v, \lambda)$ satisfies condition B1. Next,

$$\left| \Gamma\left(1 - \frac{iv}{\lambda}\right) \right| \leqq 1 \quad \text{and} \quad \lim_{\lambda \to \infty} \Gamma\left(1 - \frac{iv}{\lambda}\right) = 1$$

where taking the limit inside the integral sign is justified by dominated convergence. Therefore $k(u, v, \lambda)$ satisfies condition B2. Last, we show that $K_2(u, t, \lambda)$ is a strong approximation kernel on $\{\cdot, L^p\}$. Firstly,

$$\int_{-\infty}^{\infty} |K_2(u, t, \lambda)| \, dt = \int_{-\infty}^{\infty} |K_2(u, t, \lambda)| \, du = 1.$$

Secondly, for any finite interval (a, b)

$$\int_a^b K_2(u, t, \lambda) \, dt = e^{-e^{\lambda(u-b)}} - e^{-e^{\lambda(u-a)}} \to \chi_{(a,b)}(u) \quad \text{as } \lambda \to \infty$$

and $e^{-e^{\lambda(u-b)}} - e^{-e^{\lambda(u-a)}}$ is dominated by a function of $L^p(-\infty, \infty)$. Hence by [1, Lemma 2, p. 292] $K_2(u, t, \lambda)$ is a strong approximation kernel on $\{\cdot, L^p\}$. Therefore $k(u, v, \lambda)$ satisfies condition B3 and we have proved that the extended Phragmen kernel

$$k(u, v, \lambda) = \frac{e^{iuv}}{\sqrt{2\pi}} \Gamma\left(1 - \frac{iv}{\lambda}\right)$$

is an ideal inversion kernel for the Fourier transform.

REFERENCES

[1] J. L. B. COOPER, *Fourier transforms and inversion formulae for L^p functions*, Proc. London Math. Soc., 14 (1964), pp. 271–298.

[2] N. G. DE BRUIJN, *Asymptotic Methods in Analysis*, North-Holland, Amsterdam, 1970.

[3] G. DOETSCH, *Laplace-Transformation*, Dover, New York, 1943.

[4] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1946.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0011 $01.00/0

OSCILLATION AND ASYMPTOTIC BEHAVIOR OF FORCED NONLINEAR EQUATIONS*

A. G. KARTSATOS† AND J. TORO‡

Abstract. The oscillation and the asymptotic behavior of the solutions of the equation

$$x^{(n)}+H(t, x(q(t)))=Q(t)$$

are studied under assumptions of smallness or periodicity for $Q(t)$. Recent results of Mahfoud concerning the case $Q(t)\equiv 0$ are extended via a transformation introduced recently by the first author.

Introduction. In this paper we study equations of the form

$$x^{(n)}+H(t, x(q(t)))=Q(t), \qquad n \text{ even}, \tag{I}$$

where $H(t, u)$, $Q(t)$ are defined and continuous on $[0, +\infty)\times(-\infty, +\infty)$, $[0, +\infty)$ respectively, and $uH(t, u)>0$ for any (t, u) with $u\neq 0$. The function q is also defined and continuous on $[0, +\infty)$ and satisfies $\lim_{t\to+\infty} q(t)=+\infty$. Our main purpose here is to show that the results of Mahfoud in [7] concerning the homogeneous case ($Q(t)\equiv 0$) can be extended to the case (I) by use of a method introduced by the first author in [3], [4]. According to this method, Equation (I) is reduced to a "homogeneous-like" equation which can be treated much more easily than (I). The forcings Q will be assumed to be "small" or "periodic-like" and oscillatory. The reader is referred to the survey article of the first author in [6], where an account is given of several criteria for oscillation of forced and perturbed equations, as well as an almost complete bibliography on the subject.

In § 1 we establish some background information, § 2 is devoted to the main results of this paper, and in § 3 we discuss some possible extensions of the present results.

1. Preliminaries. In what follows, $R=(-\infty, \infty)$, $R_+=[0, \infty)$, $R_-=(-\infty, 0]$ and, for every $\alpha>0$, $R_\alpha^+=[\alpha, \infty)$, $R_\alpha^-=(-\infty, \alpha]$. By $C[A, B]$ we shall denote the space of all continuous functions from the set A into the set B. The following group of hypotheses will be referred to as Condition (S):

(i) $H\in C[R_+\times R, R]$, $uH(t, u)>0$ for every $(t, u)\in R_+\times R$ with $u\neq 0$;
(ii) $q\in C[R_+, R]$ and $\lim_{t\to\infty} q(t)=+\infty$;
(iii) $Q\in C[R_+, R]$;

By a solution of (I) (under (S)) we mean any function $x(t)$, $t\in[t_x, +\infty)$, $t_x\geqq 0$ which is n times continuously differentiable on $[t_x, +\infty)$ and satisfies (I) on the same interval. The number t_x depends on the particular solution under consideration. A function $f\in C[R_\alpha^+, R]$, for some $\alpha>0$, is "oscillatory" if it has an unbounded set of zeros in R_α^+. We denote by R_α the set $R_\alpha^+\cup R_\alpha^-$, for any $\alpha>0$, and we consider the spaces:

$C(R)=\{f\in C[R, R];\ uf(u)>0 \text{ for any } u\neq 0\}$;
$C^1(R_\alpha)=\{f\in C(R);\ f \text{ is continuously differentiable on } R_\alpha\}$;
$C_p(R_\alpha)=\{f\in C(R);\ f \text{ is of bounded variation on every } [a, b]\subset R_\alpha\}$.

Lemmas 1, 2 below can be found in Mahfoud's paper [7].

---

* Received by the editors June 12, 1976, and in final revised form September 22, 1977.

† Department of Mathematics, University of South Florida, Tampa, Florida 33620.

‡ Department of Mathematics, University of South Florida, Tampa, Florida 33620. The work of this author was supported by the Instituto Politecnico Nacional, Mexico City, Mexico D.F.

LEMMA 1. *Let* $\alpha > 0$ *and* $f \in C(R)$. *Then* $f \in C_p(R_\alpha)$ *if and only if* $f(x) = g(x)h(x)$ *for all* $x \in R_\alpha$, *where* $g: R_\alpha \to (0, +\infty)$ *is increasing on* $(-\infty, -\alpha]$ *and decreasing on* $[\alpha, \infty)$, *and* $h: R_\alpha \to R$ *is increasing in* R_α.

DEFINITION 1. The function h in Lemma 1 will be called an *increasing component* of f while g will be called a *positive component* of f.

We also consider the following space:

$C_I(R_\alpha) = \{f \in C_p(R_\alpha);\ f$ has a positive component bounded away from zero$\}$.

The importance of the above spaces in the present considerations is made clear by the following.

LEMMA 2. *Let* $\alpha > 0$. *Then if* $f \in C_I(R_\alpha)$, *there exists* $\beta > 0$ *such that* $f(x_1) \geqq \beta f(x_2)$ *whenever* $x_1 \geqq x_2 \geqq \alpha$ *and* $f(x_1) \leqq \beta f(x_2)$ *whenever* $x_1 \leqq x_2 \leqq -\alpha$.

For the forcing Q in (I) we will always assume one of the following conditions:

(iv) there exists $P \in C[R_+, R]$ such that $P^{(n)}(t) = Q(t)$, $t \in R_+$, P is oscillatory and $\lim_{t\to\infty} P(t) = 0$.

(v) there exist $P_j \in C[R_+, R]$, $j = 1, 2$, such that P_j is oscillatory, $\liminf_{t\to\infty} P_1(t) = 0$, $\limsup_{t\to\infty} P_2(t) = 0$, and $P_j^{(n)}(t) \equiv Q(t)$, $t \in R_+$.

2. Main results. The following theorem extends to the case (I) Theorem 1 in Mahfoud's paper [7] and provides, even in Mahfoud's case, a much simpler proof than the one in [7].

THEOREM 1. *Suppose that Condition* (S) *holds. Furthermore, suppose that for each* $\alpha > 0$ *there exists a function* $P_{1,\alpha} \in C[R_+, R_+]$ *and a function* $P_{2,\alpha} \in C[R_-, R_-]$ *such that*

$$H(t, u) \geqq P_{1,\alpha}(t) \quad \text{for every } u \in R_\alpha^+,$$

$$H(t, u) \leqq P_{2,\alpha}(t) \quad \text{for every } u \in R_\alpha^-$$

and, for some integer i *with* $0 \leqq i \leqq n-1$,

$$\int_0^\infty t^i P_{1,\alpha}(t)\,dt = +\infty, \qquad \int_0^\infty t^i P_{2,\alpha}(t)\,dt = -\infty.$$

Then if the forcing Q *satisfies* (v) *with* $P_j^{(n-i-1)}(t)$ *bounded for* $j = 1, 2$, *every solution of* (I) *with bounded* $(n-i-1)$*st derivative is oscillatory.*

Proof. Let $x(t)$, $t \in [\lambda, +\infty)$, $\lambda > 0$ be a solution of (I) with bounded $(n-i-1)$st derivative, and assume that $x(t)$ is nonoscillatory. Then $x(t)$ is eventually positive or negative. We assume that $x(t)$ is positive for all large t, say for $t \geqq \lambda$, and we reach a contradiction. The reader should have in mind that a very similar proof covers the case of a negative $x(t)$. Now let $u(t) \equiv x(t) - P_1(t)$, $t \geqq \lambda$, where $P_1(t)$ is the function in (v). Then $u(t)$ has its $(n-i-1)$st derivative bounded and satisfies

$$u^{(n)} + H(t, u(q(t))) + P_1(q(t)) = 0, \qquad t \geqq \lambda. \tag{1}$$

We shall show that $u(t)$ has to be negative for all large t, which, in view of the oscillatory character of $P_1(t)$, will imply a contradiction to the positiveness of $x(t)$. In fact, since $x(t) = u(t) + P_1(t) > 0$ for $t \geqq \lambda$, and $q(t) \to +\infty$ as $t \to +\infty$, there exists $t_1 \geqq \lambda$ such that $q(t) \geqq \lambda$ for every $t \geqq t_1$. Consequently, $u(q(t)) + P_1(q(t)) > 0$ for every $t \geqq t_1$. This in turn implies along with (1) that $u^{(n)}(t) < 0$ for every $t \geqq t_1$. Consequently, all the derivatives $u^{(k)}(t)$ are monotonic (hence of constant sign) for all large t and $0 \leqq k \leqq n-1$. Now suppose $u(t) > 0$ for all large t. Then since n is even, we may take t_1 above to be such that $u'(t) > 0$ for $t \geqq t_1$. Now given positive $\varepsilon < u(t_1)$ there exists $t_2 \geqq t_1$ such that $P_1(t) \geqq -\varepsilon$ for all $t \geqq t_2$. Let $t_3 \geqq t_2$ be such that $q(t) \geqq t_2$ for every $t \geqq t_3$. Then

$u(q(t))+P_1(q(t))\geqq u(t_2)-\varepsilon\geqq u(t_1)-\varepsilon>0$ for all $t\geqq t_3$. Now consider the function $F(t)\equiv t^i u^{(n-1)}(t)$. Then

$$F'(t)=-t^iH(t,u(q(t))+P_1(q(t)))+it^{i-1}u^{(n-1)}(t) \tag{2}$$

for every $t\geqq t_3$. Let us remark now that for $n-i\leqq k\leqq n$,

$$(-1)^k u^{(k)}(t)<0, \qquad t\in[t_3,\infty).$$

In fact, if this is not true, then two consecutive derivatives of order between $n-i$ and n would be eventually positive or eventually negative for all large t, but this would contradict the boundedness of $u^{(n-i-1)}(t)$. Now we integrate (2) from t_3 to $t\geqq t_3$ to obtain

$$\begin{aligned} t^iu^{(n-1)}(t)-i\int_{t_3}^{t} s^{i-1}u^{(n-1)}(s)\,ds \\ =t_3^iu^{(n-1)}(t_3)-\int_{t_3}^{t} s^iH(s,u(q(s))+P_1(q(s)))\,ds \\ \leqq t_3^iu^{(n-1)}(t_3)-\int_{t_3}^{t} s^iP_{1,\mu}(s)\,ds, \qquad \mu=u(t_1)-\varepsilon. \end{aligned} \tag{3}$$

Since the last member of (3) tends to $-\infty$ as $t\to+\infty$, and since $u^{(n-1)}(t)>0$, $t\geqq t_3$ we must have

$$\int_{t_3}^{\infty} s^{i-1}u^{(n-1)}(s)\,ds=+\infty.$$

Now the proof continues exactly as in Kartsatos [2, Thm. 1] and leads to

$$(-1)^{m+1}\int_{t_3}^{+\infty} s^{i-m}u^{(n-m)}(s)\,ds=+\infty, \qquad 1\leqq m\leqq i. \tag{4}$$

Actually, (4) is stronger than the corresponding estimate of Kartsatos in [2]. Letting $m=i$ in (4) we find

$$\lim_{t\to\infty}(-i)^{i+1}[u^{(n-i-1)}(t)-u^{(n-i-1)}(t_3)]=+\infty,$$

a contradiction to the boundedness of the function $u^{(n-i-1)}(t)$. This completes the proof.

COROLLARY 1. *Let* (S) *hold and assume that* $H(t,u)\equiv P_0(t)f(u)$ *with* $P_0\in C[R_+,R_+\backslash\{0\}]$, $f\in C(R)$ *and* $\liminf_{|u|\to\infty}|f(u)|>0$,

$$\int_0^{\infty} t^iP_0(t)\,dt=+\infty \quad \textit{for some } i \textit{ with } 0\leqq i\leqq n-1.$$

Then if Q is as in Theorem 1, *the conclusion of Theorem* 1 *holds.*

The proof follows easily from the proof of Theorem 1. This corollary, for $Q\equiv 0$, was given by Mahfoud in [7]. Of course we may assume instead of $P_0(t)>0$ on $[0,+\infty)$ that $P_0(t)\geqq 0$ and not identically equal to zero in any infinite subinterval of R_+.

THEOREM 2. *Let Condition* (S) *hold and assume that for every $\alpha>0$ there exist functions* $P_{1,\alpha}\in C[R_+\times R_\alpha^+,R_+]$, $P_{2,\alpha}\in C[R_+\times R_\alpha^-,R_-]$ *such that*

$$\begin{aligned} H(t,u_1)\geqq P_{1,\alpha}(t,u_2) \quad \textit{for } t\in R_+, \quad u_1\geqq u_2\geqq\alpha, \\ H(t,u_1)\leqq P_{2,\alpha}(t,u_2) \quad \textit{for } t\in R_+, \quad u_1\leqq u_2\leqq-\alpha, \end{aligned} \tag{4a}$$

and, for every $\mu > 0$ *and some i with* $0 \leqq i \leqq n-1$,

$$\int^{\infty} t^i P_{1,\alpha}(t, \mu q^{n-i-1}(t))\,dt = +\infty$$

$$\int^{\infty} t^i P_{2,\alpha}(t, -\mu q^{n-i-1}(t))\,dt = -\infty.^{1}$$

Let $P_j \in C[R_+, R]$, $j = 1, 2$, *satisfy* $\liminf_{t\to\infty} P_1^{(n-i-1)}(t) = 0$, $\limsup_{t\to\infty} P_2^{(n-i-1)}(t) = 0$, $P_j^{(n)} \equiv Q$ *and be bounded and oscillatory. Then every solution* $x(t)$ *of* (I) *with* $x^{(n-i-1)}(t)$ *bounded is either oscillatory, or such that* $\lim_{t\to\infty} [x^{(n-i-1)}(t) - P_j^{(n-i-1)}(t)] = 0$ *for* $j = 1$ *or* 2.

Proof. Let $x(t)$ be a solution of (I) assumed to be positive on $[\lambda, +\infty)$, $\lambda > 0$, and such that the function $x^{(n-i-1)}(t)$ is bounded on $[\lambda, +\infty)$. Then, as in the proof of Theorem 1, if $u(t)$, $t \geqq \lambda$, denotes the function $x(t) - P_1(t)$, there exists $t_1 \geqq \lambda$ such that all the derivatives $u^{(k)}(t)$ are of constant sign for all $t \geqq t_1$, $0 \leqq k \leqq n$, and $u(q(t)) + P_1(q(t)) > 0$ for all $t \geqq t_1$. Now since $u^{(n-i-1)}(t)$ is monotonic, $\lim_{t\to\infty} u^{(n-i-1)}(t) = L$ exists with $0 \leqq |L| < +\infty$.

The proof of this theorem follows immediately from Theorem 1 if $i = n-1$ because we can easily obtain the functions $P_{1,\alpha}(t)$, $P_{2,\alpha}(t)$ therein from the present functions $P_{1,\alpha}(t, u)$, $P_{2,\alpha}(t, u)$. In fact, if $L = 0$, the theorem is proved. If $L > 0$, then $H(t, u(q(t)) + P_1(q(t))) \geqq P_{1,\alpha}(t, \mu)$ for some suitable constants $\alpha > 0$, $\mu > 0$, by our assumptions, where

$$\int^{\infty} t^{n-1} P_{1,\alpha}(t, \mu) = +\infty.$$

Since $P_{1,\alpha}(t, \mu)$ does not depend on u, the proof of Theorem 1 applies. Similarly one argues in the case $L < 0$ by using the function $P_{2,\alpha}(t, \mu)$. Thus, we shall assume that $i < n-1$. Now let $L < 0$. Then we easily obtain by successive integration that $\lim_{t\to\infty} u(t) = -\infty$, a contradiction to the positiveness of $x(t)$. Thus, $L \geqq 0$. Let $L > 0$ and $0 < \varepsilon < L$. Then there exists $t_2 \geqq t_1$ such that $u^{(n-i-1)}(t) + P_1^{(n-i-1)}(t) \geqq L - \varepsilon$ for all $t \geqq t_2$. By successive integration we can find a constant $m > 0$ such that $u(t) + P_1(t) \geqq mt^{n-i-1}$ for all $t \geqq$ (say) $t_3 \geqq t_2$. Let $t_4 \geqq t_3$ be such that $q(t) \geqq t_3$ for every $t \geqq t_4$. Then $u(q(t)) + P_1(q(t)) \geqq mq^{n-i-1}(t)$, $t \geqq t_4$. Consequently, if $F(t)$ is the function of the proof of Theorem 1, we obtain similarly, for $\alpha = mt_3^{n-i-1}$,

$$\begin{aligned} F(t) - F(t_4) &= -\int_{t_4}^{t} s^i H(s, u(q(s)) + P_1(q(s)))\,ds + i\int_{t_4}^{t} s^{i-1} u^{(n-1)}(s)\,ds \\ &\leqq -\int_{t_4}^{t} s^i P_{1,\alpha}(s, mq^{n-i-1}(s))\,ds + i\int_{t_4}^{t} s^{i-1} u^{(n-1)}(s)\,ds. \end{aligned} \tag{5}$$

Now the proof follows as in Theorem 1 and leads to a contradiction. Thus, $\lim_{t\to\infty} u^{(n-i-1)}(t) = 0$ and a similar argument covers the case $x(t) < 0$ for all large t.

COROLLARY 2. *Let Condition* (S) *hold and let* $H(t, u) \equiv P_0(t) f(u)$ *with* $P_0 \in C[R_+, R_+ \backslash \{0\}]$, $f \in C_I(R_\alpha)$ *for some* $\alpha > 0$. *Let Q be as in Theorem* 2. *Then the conclusion of Theorem* 2 *holds if*

$$\int_0^{\infty} t^i P_0(t) f(\pm\mu q^{n-i-1}(t))\,dt = \pm\infty \quad \textit{for some } i, \quad 0 \leqq i \leqq n-1$$

and every $\mu > 0$.

[1] Whenever the lower limit of an integral is omitted, it is meant that the condition is valid for the smallest $\bar{t} \geqq 0$ such that the integrand is well defined for all $t \geqq \bar{t}$.

To prove the above corollary we simply take into consideration the bounds obtained for $f(u(q(t))+P_1(q(t)))$ for Lemma 2. In fact, in (5) we would now have

$$-\int_{t_4}^{t} s^i P_0(s) f(u(q(s))+P_1(q(s)))\,ds$$
$$\leqq -\beta \int_{t_4}^{t} s^i P_0(s) f(mq^{n-i-1}(s))\,ds \to -\infty \quad \text{as } t\to+\infty.$$

Naturally, $mq^{n-i-1}(t) \geqq \alpha$ for all large t. Thus, t_4 can be properly chosen sufficiently large so that the above inequality makes sense.

Mahfoud gave in [7] the above corollary for $Q(t)\equiv 0$. Actually a stronger conclusion holds above but we shall not concern ourselves with it here.

THEOREM 3. *Suppose that H, Q satisfy Condition* (S). *Suppose further that*

$$\text{(6)} \qquad \int^{\infty} t^i P_{1,\alpha}(t, \mu q^{n-i-2}(t))\,dt = +\infty, \qquad \int^{\infty} P_{1,\alpha}(t, \mu q^{n-1}(t))\,dt = +\infty,$$

$$\text{(7)} \qquad \int^{\infty} t^i P_{2,\alpha}(t, -\mu q^{n-i-2}(t))\,dt = -\infty, \qquad \int^{\infty} P_{2,\alpha}(t, -\mu q^{n-1}(t))\,dt = -\infty$$

for every $i \in \{0, 1, \cdots, n-2\}$ and every $\mu > 0$, where the functions $P_{j,\alpha}$ satisfy the inequalities (4a) *of Theorem* 2. *Then if* (iv) *holds and $P^{(k)}(t)\to 0$ as $t\to\infty$ for every $k = 0, 1, 2, \cdots, n-1$, every solution of* (I) *is oscillatory.*

Proof. Let $x(t)$ be a solution of (I) with $x(t)>0$, $t \geqq \lambda > 0$. Then the last integral conditions in (6) and (7) imply that $\lim_{t\to\infty} x^{(n-1)}(t) = 0$. In fact, this follows from Theorem 2. To show this, let $u(t) \equiv x(t) - P(t)$, $t \geqq \lambda$. Then $u(t)$ satisfies the inequality

$$\text{(8)} \qquad u^{(n)}(t) = -H(t, u(q(t)) + P(q(t))) < 0$$

for all t (say) $\geqq t_1 \geqq \lambda$. It follows that $u^{(n-1)}(t) > 0$ for all large t, otherwise $\lim_{t\to\infty} u(t) = -\infty$, which implies $x(t) - P(t) < 0$ for all large t, a contradiction to $x(t) > 0$. Consequently, $u^{(n-1)}(t)$ is decreasing and positive for all large t, thus bounded. This fact, along with $\lim_{t\to\infty} P^{(n-1)}(t) = 0$, implies that $x^{(n-1)}(t)$ is bounded. Thus, Theorem 2 applies now for $i = 0$ to obtain $\lim_{t\to\infty} x^{(n-1)}(t) = 0$. Now let i be any integer so that $0 \leqq i \leqq n-2$ and $\lim_{t\to\infty} x^{(n-i-1)}(t) = 0$. We show that $\lim_{t\to\infty} x^{(n-i-2)}(t) = 0$. Suppose this is not true. Let $u(t)$ be as above so that $u(q(t)) + P(q(t)) > 0$ for all $t \geqq$ (some) $t_1 \geqq \lambda$ and (8) holds for all $t \geqq t_1$. Since $u^{(n)}(t) < 0$, we may assume that all the intermediate derivatives $u^{(k)}(t)$ are of constant sign for every $t \geqq t_1$. Here $0 \leqq k \leqq n-1$. In particular, $u^{(n-i-1)}(t)$ is monotonic for $t \in [t_1, +\infty)$ and $\lim_{t\to\infty} u^{(n-i-1)}(t) = \lim_{t\to\infty} [x^{(n-i-1)}(t) - P^{(n-i-1)}(t)] = 0$. Now let $\lim_{t\to\infty} u^{(n-i-2)}(t) = L$. If $L < 0$ then $\lim_{t\to\infty} x(t) < 0$, for $i = n-2$, a contradiction, or $\lim_{t\to\infty} u(t) = \lim_{t\to\infty} x(t) = -\infty$, for $i < n-2$, a contradiction again. Let $L > 0$. Since $\lim_{t\to\infty} u^{(n-i-1)}(t) = 0$, it follows from Lemma 2 in Mahfoud [7] that $\sum_{k=0}^{i} (-1)^{i+k} t^k u^{(n-i+k)}(t)/k! > 0$ for all $t \geqq t_1$. Consequently, from equation (4) of [7] we obtain

$$\text{(9)} \qquad \frac{1}{i!} \int_{t_1}^{t} s^i H(s, u(q(s)) + P(q(s)))\,ds \leqq C_i$$

for all $t \geqq t_1$, where C_i are constants. Now let L be finite, and let $i = n-2$. Then $u^{(n-i-2)}(t) \equiv u(t)$ is bounded and this implies the boundedness of $x(t)$. Since

$$\int^{\infty} t^{n-2} P_{1,\alpha}(t, \mu)\,dt = +\infty, \qquad \int^{\infty} t^{n-2} P_{2,\alpha}(t, -\mu)\,dt = -\infty,$$

it follows immediately as in the Corollary of Kartsatos [3] that $x(t)$ is oscillatory, a contradiction. Now let $L=+\infty$ and $i=n-2$. Then $H(t, u(q(t))+P(q(t)))\geqq P_{1,M}(t, M)$ for every $t\geqq t_2\geqq t_1$ where $M>0$ and t_2 are chosen so that $u(q(t))+P(q(t))\geqq M$ for $t\geqq t_2$. Consequently, from (9) we obtain

$$\int_{t_2}^{\infty} t^{n-2}P_{1,M}(t, M)\,dt<+\infty,$$

a contradiction to our assumptions. Now if $i<n-2$, there exists $t_2\geqq t_1$ such that $u(q(t))+P(q(t))\geqq kq^{n-i-2}(t)$ for every $t\geqq t_2$, where k is a positive constant. Then (9) implies again

$$\int_{t_2}^{\infty} t^iP_{1,\alpha}(t, kq^{n-i-2}(t))\,dt<+\infty, \tag{10}$$

where α is an appropriate constant. From (10) we have again a contradiction, and this proves our claim. Thus, by induction on i, we actually obtain $\lim_{t\to\infty} u(t)=0$, an impossibility, for if $u(t)>0$ then $u'(t)>0$ and if $u(t)<0$, $x(t)<P(t)$ which contradicts the positiveness of $x(t)$.

COROLLARY 3. *Under Condition* (S), *suppose further that* $f\in C_I(R_\alpha)$ *for some* $\alpha>0$, $P_0\in C[R_+, R_+\backslash\{0\}]$, *and* P *is as in Theorem* 3. *Then if* $H(t, u)\equiv P_0(t)f(u)$ *for* $(t, u)\in R_+\times R$, *and*

$$\int_0^{\infty} t^iP_0(t)f(\pm cq^{n-i-2}(t))\,dt=\pm\infty$$

for every $c>0$ *and every* $i\in\{0, 1, \cdots, n-2\}$, *every solution of* (I) *is oscillatory.*

This corollary was shown for $Q\equiv 0$ by Mahfoud in [7]. It should be noted, in view of Lemma 2, that the above integral conditions actually imply

$$\int_0^{\infty} P_0(t)f(\pm\mu q^{n-1}(t))\,dt=\pm\infty$$

for every $\mu>0$.

In the following theorem we assume that for every $t_0\geqq 0$, $\phi\in C[[0, t_0], R]$, and every $n-1$-tuple $(c_1, c_2, \cdots, c_{n-1})$ of real numbers, Equation (I) has at least one solution $x(t)$ valid for all $t\geqq t_0$ and such that $x(t)=\phi(t)$, $t\leqq t_0$, and $x^{(i)}(t_0)=c_i$, $i=1, 2, \cdots, n-1$.

THEOREM 4. *Let Condition* (S) *be satisfied with* $q(t)\leqq t$, *let* Q, P *be as in Theorem* 3 *and suppose that for every* $\alpha>0$ *there exist functions* $P_{\alpha,1}\in C[R_+\times R_\alpha^+, R_+]$, $P_{\alpha,2}\in C[R_+\times R_\alpha^-, R_-]$ *such that*

$$H(t, x_1)\leqq P_{\alpha,1}(t, x_2)\quad \text{for } t\in R_+,\quad \alpha\leqq x_1\leqq x_2,$$
$$H(t, x_1)\geqq P_{\alpha,2}(t, x_2)\quad \text{for } t\in R_+,\quad x_2\leqq x_1\leqq -\alpha.$$

Then if every solution $x(t)$ *of* (I) *is oscillatory we have*

$$\int^{\infty} P_{\alpha,1}(t, \mu q^{n-1}(t)+P(q(t)))\,dt=+\infty,$$

$$\int^{\infty} P_{\alpha,2}(t, -\mu q^{n-1}(t)+P(q(t)))\,dt=-\infty$$

for every $\mu>0$.

Proof. Let $\mu q^{n-1}(t)+P(q(t))\geqq\alpha$ for $t\geqq\beta\geqq\alpha$ and

$$\int_\beta^\infty P_{\alpha,1}(t, \mu q^{n-1}(t)+P(q(t))\,dt<+\infty$$

for some $\mu>0$ and some $\beta>0$. Now let $t_0\geqq 1+\beta$ and ε with $0<\varepsilon<\mu/4$ be such that $t_0^{n-1}\geqq(\beta+\varepsilon)/(\mu-\varepsilon)$ and $|P^{(k)}(t)|<\varepsilon/(n-1)!$ for every $t\geqq t_0$ and every $k=0, 1, 2, \cdots, n-1$. Furthermore, choose $t_1\geqq t_0$ such that $q(t)\geqq t_0$ for every $t\geqq t_1$, and

$$\int_{t_1}^\infty P_{\alpha,1}(t, \mu q^{n-1}(t)+P(q(t)))\,dt<(\mu-2\varepsilon)/2.$$

The above are consequences of the fact that $\lim_{t\to\infty}P^{(k)}(t)=0$, $\lim_{t\to\infty}q(t)=+\infty$ and the integral conditions in the assumptions. Now let $x(t)$ be a solution of (I) such that

$$x(t)=(\mu-\varepsilon)t^{n-1}, \qquad t\leqq t_1.$$

Then $x^{(n-k)}(t)=\alpha_1 t^{k-1}/(k-1)!$, $k=1, 2, \cdots, n$, $t\leqq t_1$, where $\alpha_1=(n-1)!(\mu-\varepsilon)$. Since $x^{(n-1)}(t_1)=\alpha_1>0$, there exists $t_2>t_1$ such that $x^{(n-1)}(t)>0$ for $t\in[t_1, t_2)$. This implies that $x^{(n-2)}(t)$ is increasing on $[t_1, t_2)$ and since $x^{(n-2)}(t_1)=\alpha_1 t_1>0$, we must have $x^{(n-2)}(t)>0$ on $[t_1, t_2)$. Similarly, one shows by induction that $x^{(n-k)}(t)$ is increasing and positive on $[t_1, t_2)$ for every $k=2, 3, \cdots, n$. Now consider $x(q(t))$, $t\in[t_1, t_2)$. If $q(t)\in[t_1, t_2)$, then $x(q(t))>0$ by the previous argument. If $q(t)\notin[t_1, t_2)$, then $t_0\leqq q(t)<t_1$. Thus, $x(q(t))=(\mu-\varepsilon)q^{n-1}(t)>0$ for $t\in[t_1, t_2)$. It follows that $x(q(t))>0$ for every $t\in[t_1, t_2)$. Now consider the function $u(t)\equiv x(t)-P(t)$, $t\geqq t_0$. Then $u^{(n-1)}(t_1)=\alpha_1-P^{(n-1)}(t_1)\geqq\mu-2\varepsilon>0$ and $u^{(n-k)}(t_1)=\alpha_1 t_1^{k-1}/(k-1)!-P^{(n-k)}(t_1)\geqq\mu-2\varepsilon>0$ for every $k=2, 3, \cdots, n$. Consequently, the argument above about $x(q(t))$ can be repeated now for $u(q(t))$ to obtain the existence of some t_3 with $t_1<t_3\leqq t_2$ with $u(q(t))>0$ in $[t_1, t_3)$. Assume that t_3 must be $<t_2$. Then $u(q(t))$ has a zero t_4 in the interval $[t_3, t_2)$. Then $x(q(t_4))=P(q(t_4))$. If $q(t_4)\in[t_1, t_4]$ (we should have in mind here that $q(t)\leqq t$), then $0=x(q(t_4))-P(q(t_4))\geqq x(t_1)-P(q(t_4))=\alpha_1 t_1^{n-1}/(n-1)!-P(q(t_4))\geqq\mu-\varepsilon-P(q(t_4))>0$, a contradiction. If $q(t_4)\notin[t_1, t_4]$, then $t_0\leqq q(t_4)<t_1$ and $0=x(q(t_4))-P(q(t_4))=(\mu-\varepsilon)q^{n-1}(t_4)-P(q(t_4))\geqq\mu-\varepsilon-P(q(t_4))>0$, a contradiction again. Consequently, $u(q(t))>0$, $t\in[t_1, t_2)$. Now since $u^{(n)}(t)=-H(t, u(q(t))+P(q(t)))<0$, we have $u^{(n-1)}(t)\leqq u^{(n-1)}(t_1)=\alpha_1-P^{(n-1)}(t_1)\leqq\alpha_1+\varepsilon$. Now by integration (observing that $u^{(n-k)}(t_1)\leqq(\alpha_1+\varepsilon)t_1^{k-1}/((k-1)!)$) we obtain, for $1\leqq k\leqq n$, $u^{(n-k)}(t)\leqq(\alpha_1+\varepsilon)t^{k-1}/((k-1)!))$, which for $k=n$ yields $u(t)\leqq(\alpha_1+\varepsilon)t^{n-1}/(n-1)!\leqq\mu t^{n-1}$ for all $t\in[t_1, t_2)$. Since $u(q(t))=(\mu-\varepsilon)q^{n-1}(t)-P(q(t))=\mu q^{n-1}(t)-(\varepsilon q^{n-1}(t)+P(q(t)))\leqq\mu q^{n-1}(t)$ for $q(t)\notin[t_1, t_2)$, it follows that $u(q(t))\leqq\mu q^{n-1}(t)$ for $t\in[t_1, t_2)$. Now we also have $u(q(t))+P(q(t))=x(q(t))\geqq x(t_1)\geqq(\mu-\varepsilon)t_0^{n-1}\geqq\beta+\varepsilon\geqq\beta$ whenever $q(t)\in[t_1, t_2)$ and $u(q(t))+P(q(t))=(\mu-\varepsilon)q^{n-1}(t)+P(q(t))\geqq(\mu-\varepsilon)t_0^{n-1}+P(q(t))\geqq\beta-\varepsilon+\varepsilon=\beta$ for $q(t)\notin[t_1, t_2)$. It follows that $u(q(t))+P(q(t))\geqq\beta$ for every $t\in[t_1, t_2)$. Consequently, by integrating the equation in u once from t_1 to $t\geqq t_1$ we obtain

$$\begin{aligned} u^{(n-1)}(t)-u^{(n-1)}(t_1)&=-\int_{t_1}^t H(s, u(q(s))+P(q(s)))\,ds\\ &\geqq-\int_{t_1}^t P_{\alpha,1}(s, \mu q^{n-1}(s)+P(q(s)))\,ds,\end{aligned}$$

which implies

$$u^{(n-1)}(t) \geqq \alpha_1 - P^{(n-1)}(t_1) - \int_{t_1}^{t} P_{\alpha,1}(s, \mu q^{n-1}(s) + P(q(s)))\, ds$$

$$\geqq \mu - 2\varepsilon - \int_{t_1}^{\infty} P_{\alpha,1}(s, \mu q^{n-1}(s) + P(q(s)))\, ds$$

$$> \frac{(\mu - 2\varepsilon)}{2}$$

for $t \in [t_1, t_2)$. Consequently, as long as $x^{(n-1)}(t)$ remains positive for $t \geqq t_1$, $x^{(n-1)}(t) = u^{(n-1)}(t) + P^{(n-1)}(t) \geqq [(\mu - 2\varepsilon)/2] - \varepsilon = (\mu - 4\varepsilon)/2 > 0$. This actually implies that $x^{(n-1)}(t) \geqq (\mu - 4\varepsilon)/2$ for all $t \geqq t_1$. Thus $x(t)$ is positive for all large t, and this completes the proof.

COROLLARY 4. *Let Condition* (S) *be satisfied with* $H(t, u) \equiv P_0(t) f(u)$, *where* $P_0 \in C[R_+, R_+ \backslash \{0\}]$, $f \in C_p(R_\alpha)$ *for some* $\alpha > 0$, *and* P, q *in Theorem* 4. *Then if every solution of* (I) *is oscillatory we must have*

$$\int_0^{\infty} P_0(t) h[\pm \mu q^{n-1}(t) + P(q(t))]\, dt = \pm\infty$$

for every increasing component h *of* f *and every* $\mu > 0$.

To show the above assertion it suffices to take in Theorem 4 $P_{\alpha,1}(t, u) = P_0(t) g(\alpha) h(u)$ and $P_{\alpha,2}(t, u) = P_0(t) g(-\alpha) h(u)$ where g is the corresponding positive component of f. The above corollary is a special case of Mahfoud's Theorem 4 in [7] if $Q(t) \equiv 0$.

COROLLARY 5. *Let Condition* (S) *hold and* $H(t, u) \equiv P_0(t) f(u)$ *with* $P_0 \in C[R_+, R_+ \backslash \{0\}]$ *and* $f \in C_I(R_\alpha)$ *for some* $\alpha > 0$. *Let* $P(t)$, $q(t)$ *be as in Theorem* 4. *Then if every solution of* (I) *is oscillatory we must have*

$$\int_0^{\infty} P_0(t) f(\pm \mu q^{n-1}(t) + P(q(t)))\, dt = \pm\infty$$

for any $\mu > 0$.

In fact, this follows easily from Corollary 4 as in the proof of Corollary 4 of Mahfoud [7]. In the following result $H(t, u)$ is supposed to be linear in u. This allows us to adapt the proof of Mahfoud [7, Thm. 5] to the present forced case.

THEOREM 5. *Let Condition* (S) *be satisfied with* $H(t, u) \equiv P_0(t) u$, *where* $P_0 \in C[R_+, R_+ \backslash \{0\}]$. *Let* $P(t)$ *be as in Theorem* 4. *Then if every solution of* (I) *oscillates for any choice of* $q(t)$ *satisfying* (ii) *and* $q(t) \leqq t$, $t \in R_+$, *we must have*

$$\int_0^{\infty} P_0(t)\, dt = +\infty.$$

Proof. Assume that every solution of (I) oscillates. Then

$$\int_0^{\infty} P_0(t)[\mu q^{n-1}(t) + P(q(t))]\, dt = +\infty$$

by Theorem 4 for every $\mu > 0$ and for every $q(t)$ as in the assumptions of the theorem. If

$$\int_0^{\infty} P_0(t)\, dt < +\infty,$$

then it follows from Theorem 11 of Burton and Grimmer [1] that there exists a function $P_1 \in C[R_+, [1, \infty)]$ which is increasing, onto, and such that

$$\int_1^\infty P_0(t)P_1(t)\,dt < +\infty.$$

The proof now follows as in Theorem 5 of [7] and is based upon constructing a suitable $q(t)$ from the function $P_1(t)$ for which

$$\int_{t_1}^\infty P_0(t)q^{n-1}(t)\,dt \leqq \int_{t_1}^\infty P_0(t)P_1(t)\,dt < +\infty,$$

for some $t_1 \geqq 1$, a contradiction, because $\lim_{t\to\infty} P(q(t)) = 0$ implies also

$$\int_{t_2}^\infty P_0(t)[q^{n-1}(t)+\mu P(q(t))]\,dt \leqq 2\int_{t_2}^\infty P_0(t)q^{n-1}(t)\,dt < +\infty$$

for large enough $t_2 \geqq t_1$.

This theorem can be easily extended to the case $H(t, u) \equiv P_0(t)f(u)$ under the rest of the assumptions of Corollary 4, if we further assume that for some increasing component $h(u)$ of f we have

$$\lim_{\substack{u_1, u_2 \to \infty \\ u_1 - u_2 \to 0}} [h(u_1)/h(u_2)] = 1.$$

This assumption is needed, if we adopt the above method, in order to ensure that

$$\int_{t_1}^\infty P_0(t)h(q^{n-1}(t))\,dt < +\infty$$

implies

$$\int_{t_1}^\infty P_0(t)h(\mu q^{n-1}(t)+P(t))\,dt < +\infty$$

where $q(t)$ is as in the proof of Theorem 5.

3. Discussion. All the results of this paper can be easily extended to odd values of n as it is usual in nonlinear oscillation theory. It would also be interesting to see versions of the results here covering perturbed cases of the form

$$x^{(n)} + H(t, x(q(t))) = Q(t, x(q_1(t))). \tag{11}$$

For the first results in this direction and for $q(t) \equiv q_1(t) \equiv t$, the reader is referred to [5] where perturbations $Q(t, u)$ are considered with $|Q(t, u)| \leqq Q_0(t)|u|^r$ with Q_0 sufficiently small and $r \geqq 1$. Naturally, extensions to equations with "middle terms" of the form $S(t)x^{(n-k)}(q_1(t))$ are desirable. For an account of some results in this direction and for $k = 1, 2$, the reader is referred to [6].

Acknowledgment. The authors wish to express their thanks to the referee for his helpful suggestions.

REFERENCES

[1] T. BURTON AND R. GRIMMER, *Oscillation, continuation, and uniqueness of solutions of retarded differential equations*, Trans. Amer. Math. Soc., 179 (1973), pp. 193–209.

[2] A. G. KARTSATOS, *On oscillation of solutions of even order nonlinear differential equations*, J. Differential Equations, 6 (1969), pp. 232–237.

[3] ———, *On the maintenance of oscillations of nth order equations under the effect of a small forcing term*, Ibid., 10 (1971), pp. 355–363.

[4] ———, *Maintenance of oscillations under the effect of a periodic forcing term*, Proc. Amer. Math. Soc., 33 (1972), pp. 377–383.

[5] ———, *Oscillation of nth order equations with perturbations*, J. Math. Anal. Appl., 57 (1977), pp. 161–169.

[6] ———, *Recent results on oscillation of solutions of forced and perturbed nonlinear differential equations of even order*, Stability of Dynamical Systems, NSF–CBMS Conference, Mississippi State Univ., J. Graef, ed., Marcel Dekker, 1977.

[7] W. E. MAHFOUD, *Oscillation and asymptotic behaviour of solutions of nth order nonlinear delay differential equations*, J. Differential Equations, 24 (1977), pp. 75–98.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0012 $01.00/0

ON SOME MAXIMUM PRINCIPLES INVOLVING HARMONIC FUNCTIONS AND THEIR DERIVATIVES*

L. E. PAYNE† AND G. A. PHILIPPIN‡

Abstract. In this paper we establish sufficient conditions on $f(h)$ to guarantee that given two harmonic functions H and h, the quotient $|\text{grad } H|^2/f(h)$ satisfies a maximum principle. The principle is then used to derive isoperimetric bounds for derivatives of the Green's function and for the force field in electrostatics.

1. Introduction. Maximum principles for harmonic functions and for solutions of certain classes of second order elliptic equations have been known for more than a century. The two fundamental maximum principles for such functions are usually referred to in the literature as Hopf's first and second principles [2], [3]. In this paper we make use of these two Hopf principles to derive new maximum principles for certain combinations of harmonic functions and their derivatives. For an account of previous results on maximum principles see the book of Protter and Weinberger [6].

Throughout we shall be concerned with functions defined on a bounded region $D \subset R^N$ (or its complement). The boundary ∂D will be assumed to be a $C^{2+\varepsilon}$ surface so that the governing equation will be satisfied on the boundary. This boundary smoothness can frequently be relaxed, but we do not attempt in this paper to determine the minimum smoothness requirements on ∂D. The symbol Δ will be used to denote the Laplace operator and a comma will be used to indicate differentiation, i.e.,

$$u_{,ij} \stackrel{\text{def}}{\equiv} \frac{\partial^2 u}{\partial x_i\, \partial x_j}. \tag{1.1}$$

We also make use of the summation convention, in which a repeated index indicates summation over that index from 1 to N, i.e.,

$$u_{,ij}u_{,ij} \stackrel{\text{def}}{\equiv} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\partial^2 u}{\partial x_i\, \partial x_j}\right)^2. \tag{1.2}$$

2. A maximum principle. Let H and h be two harmonic functions in D. We would like to be able to characterize those functions $\varphi(x)$ of the form

$$\varphi(x) = g(|\text{grad } H|^2, h),$$

for which the maximum principle holds. We shall, however, be somewhat less ambitious and seek specific combinations g of the form $|\text{grad H}|^2/f(h)$ with f to be chosen in such a way that g satisfies a maximum principle. More precisely we establish the following:

THEOREM. *Let H and h be two harmonic functions in $D \subset R^N$ with $H \in C^1(\bar{D})$ and $h \in C^0(\bar{D})$, and let $f(h)$ be a positive C^2 function over the range of admissible values of h.*

* Received by the editors February 28, 1977, and in final revised form June 22, 1977.

† Department of Mathematics, Cornell University, Ithaca, New York, 14853. The research of this author was supported by the National Science Foundation under Grant NSF MPS 72-04511.

‡ Department of Mathematics, Cornell University, Ithaca, New York 14853. The research of this author was supported by the Swiss Nationalfunds.

Assume further that on this range f satisfies

$$[f^{N-2/2(N-1)}]'' \leqq 0, \qquad N \geqq 3, \tag{2.1}$$

$$[\log f]'' \leqq 0, \qquad N = 2. \tag{2.2}$$

Then the function

$$\psi = \frac{H_{,i}H_{,i}}{f(h)}, \tag{2.3}$$

assumes its maximum value on ∂D.

We note that (2.2) is satisfied in particular for exponential functions and positive powers of h.

The proof of the theorem is based on the derivation of an elliptic differential inequality for ψ. For $N \geqq 3$, this derivation makes use of the following:

LEMMA. *Let* $v \in C^2(D)$ *and* $w \in C^1(D)$. *Then at points in D where* $|\text{grad } v|$ *and* $|\text{grad } w|$ *are positive the following inequality holds*:

$$\begin{aligned} v_{,ij}v_{,ij} \geqq &\frac{1}{N-1}\left(\frac{v_{,ik}v_{,i}v_{,k}}{v_{,l}v_{,l}}\right)^2 + \frac{v_{,ik}v_{,k}v_{,ij}v_{,j}}{v_{,l}v_{,l}} \\ &+ \frac{1}{2}\frac{(v_{,ik}v_{,i}w_{,k})^2}{v_{,l}v_{,l}w_{,j}w_{,j}} + \frac{1}{2}\frac{(v_{,ik}v_{,i}w_{,k})^2(v_{,s}w_{,s})^2}{(v_{,l}v_{,l})^2(w_{,j}w_{,j})^2} \\ &- \frac{v_{,ik}v_{,i}w_{,k}v_{,jr}v_{,j}v_{,r}v_{,m}w_{,m}}{(v_{,l}v_{,l})^2 w_{,s}w_{,s}} - \frac{2}{N-1}\frac{\Delta v v_{,ik}v_{,i}v_{,k}}{v_{,l}v_{,l}}. \end{aligned} \tag{2.4}$$

The proof of this lemma follows from the fact that $\chi_{ij}\chi_{ij} \geqq 0$, where χ_{ij} is defined as

$$\begin{aligned} \chi_{ij} = v_{,ij} &- \frac{v_{,ik}v_{,k}v_{,j}}{v_{,l}v_{,l}} + \frac{1}{N-1}\frac{v_{,ks}v_{,k}v_{,s}}{v_{,l}v_{,l}}\left\{\delta_{ij} - \frac{v_{,i}v_{,j}}{v_{,l}v_{,l}}\right\} \\ &+ \frac{1}{2}\frac{v_{,kl}v_{,k}w_{,l}}{v_{,s}v_{,s}w_{,m}w_{,m}}\{w_{,i}v_{,j} - w_{,j}v_{,i}\}. \end{aligned} \tag{2.5}$$

Here δ_{ij} is the Kronecker symbol. If $v \equiv w$, then (2.4) reduces to

$$\begin{aligned} v_{,ij}v_{,ij} \geqq &\frac{1}{N-1}\frac{(v_{,ik}v_{,i}v_{,k})^2}{(v_{,l}v_{,l})^2} + \frac{v_{,ik}v_{,k}v_{,ij}v_{,j}}{v_{,l}v_{,l}} \\ &- \frac{2}{N-1}\frac{\Delta v v_{,ik}v_{,i}v_{,k}}{v_{,l}v_{,l}}. \end{aligned} \tag{2.4'}$$

Proof of the theorem. By differentiating (2.3) we obtain

$$\psi_{,k} = 2f^{-1}H_{,ik}H_{,i} - f^{-2}f'H_{,i}H_{,i}h_{,k}, \tag{2.6}$$

and

$$\Delta\psi = 2f^{-1}H_{,ik}H_{,ik} - 4f^{-2}f'H_{,i}H_{,ik}h_{,k} - (f^{-2}f')'H_{,i}H_{,i}h_{,k}h_{,k}. \tag{2.7}$$

Insertion of (2.4) (with $v = H$ and $w = h$) into (2.7) gives

$$
\begin{aligned}
\Delta\psi \geqq f^{-1}\Big\{ & \frac{2}{N-1}\Big(\frac{H_{,ik}H_{,i}H_{,k}}{H_{,l}H_{,l}}\Big)^2 + 2\frac{H_{,ik}H_{,k}H_{,ij}H_{,j}}{H_{,l}H_{,l}} \\
& + \frac{(H_{,ik}H_{,i}h_{,k})^2}{H_{,l}H_{,l}h_{,j}h_{,j}} + \frac{(H_{,ik}H_{,k}h_{,i})^2(H_{,s}h_{,s})^2}{(H_{,l}H_{,l})^2(h_{,j}h_{,j})^2} \\
& - 2\frac{H_{,ik}H_{,i}h_{,k}H_{,jr}H_{,j}H_{,r}H_{,m}h_{,m}}{(H_{,l}H_{,l})^2 h_{,s}h_{,s}}\Big\} \\
& - 4f'f^{-2}H_{,i}H_{,ik}h_{,k} - (f^{-2}f')'H_{,i}H_{,i}h_{,k}h_{,k}.
\end{aligned} \tag{2.8}
$$

We now insert the following identities:

$$H_{,ik}H_{,i}H_{,k} = \frac{f^{-1}f'}{2}H_{,i}H_{,i}H_{,k}h_{,k} + \frac{f}{2}\psi_{,k}H_{,k}, \tag{2.9}$$

$$H_{,ik}H_{,i}h_{,k} = \frac{f^{-1}f'}{2}H_{,i}H_{,i}h_{,k}h_{,k} + \frac{f}{2}\psi_{,k}h_{,k}, \tag{2.10}$$

$$
\begin{aligned}
H_{,ik}H_{,k}H_{,ij}H_{,j} = {} & \frac{f^2}{4}\psi_{,k}\psi_{,k} + \frac{f'}{2}H_{,i}H_{,i}h_{,k}\psi_{,k} \\
& + \frac{f^{-2}f'^2}{4}(H_{,i}H_{,i})^2 h_{,k}h_{,k},
\end{aligned} \tag{2.11}
$$

into the right hand side of (2.8). This leads to

$$\Delta\psi + W_k\psi_{,k} \geqq \frac{3-N}{4(N-1)}f'^2f^{-3}(H_{,k}h_{,k})^2 + (\tfrac{3}{4}f'^2 - ff'')f^{-3}H_{,i}H_{,i}h_{,k}h_{,k}, \tag{2.12}$$

where the term $W_k\psi_{,k}$ includes all terms involving first derivatives of ψ. We note of course that W_k may become unbounded at points at which $|\text{grad}\, H| = 0$. However, W_k remains bounded in the limit at those points where $|\text{grad}\, h| = 0$ (assuming $|\text{grad}\, H| \neq 0$ there). Since $N \geqq 3$ and

$$H_{,i}H_{,i}h_{,k}h_{,k} \geqq (H_{,i}h_{,i})^2, \tag{2.13}$$

it follows that the right hand side will remain positive provided

$$ff'' - \frac{N}{2(N-1)}f'^2 \leqq 0. \tag{2.14}$$

But this is equivalent to (2.1). Hence if $N \geqq 3$ and if (2.1) is satisfied, it follows that

$$\Delta\psi + W_k\psi_{,k} \geqq 0 \quad \text{in } D, \tag{2.15}$$

and we conclude from Hopf's first principle [2] that ψ takes its maximum value either at a critical point of H, or on ∂D. But the maximum can occur at an interior critical point of H iff $|\text{grad}\, H| \equiv 0$ ($H = \text{const}$). In this latter case, the theorem is trivially true. This establishes the first part of the theorem.

In the case $N = 2$ we replace the inequality (2.4) by the following identity:

$$v_{,ij}v_{,ij} \equiv (\Delta v)^2 + \frac{2}{v_{,l}v_{,l}}v_{,i}v_{,ij}v_{,k}v_{,kj} - \frac{2\Delta v}{v_{,l}v_{,l}}v_{,i}v_{,j}v_{,ij}. \tag{2.16}$$

The same kind of computation as before gives now the following *equality* for ψ:

$$\Delta\psi - \frac{f\psi_{,k}\psi_{,k}}{H_{,j}H_{,j}} = -\frac{1}{f}(\log f)''H_{,i}H_{,i}h_{,k}h_{,k}, \tag{2.17}$$

which establishes the second part of the theorem.

3. Applications.

a) Bounds for derivatives of the Green's function. The first application involves the Green's function $G(Q, P)$ for the Laplace equation in D. For fixed Q in D, $G(Q, P)$ vanishes for $P \in \partial D$, and in D has the representation

$$G(Q, P) = \frac{r^{2-N}}{(N-2)\omega_N} - g(Q, P), \qquad N \geqq 3. \tag{3.1}$$

Here r is the distance between P and Q, and $\omega_N = 2\pi^{N/2}/\Gamma(N/2)$ is the surface of the unit sphere in N-dimensions. The regular part of the Green's function, $g(Q, P)$, is for fixed Q a solution of the following boundary value problem:

$$\begin{cases} \Delta g = 0 & \text{in } D, \\ g = \dfrac{r^{2-N}}{(N-2)\omega_N} & \text{on } \partial D. \end{cases} \tag{3.2}$$

In $\mathbb{R}^2$ we have instead

$$G(Q, P) = \frac{1}{2\pi}\log\frac{1}{r} - g(Q, P), \tag{3.1'}$$

where $g(Q, P)$ satisfies for fixed Q

$$\begin{cases} \Delta g = 0 & \text{in } D, \\ g = \dfrac{1}{2\pi}\log\dfrac{1}{r} & \text{on } \partial D. \end{cases} \tag{3.2'}$$

For fixed Q in D we now apply the theorem with the following choices for H, h, and $f(h)$:

$$H(P) = G(Q, P); \qquad h = r^{-(N-2)}, \qquad f(h) = \begin{cases} h^{2(N-1)/(N-2)}, & N \geqq 3, \\ r^{-2}, & N = 2. \end{cases} \tag{3.3}$$

According to the theorem established in § 2 the quantity

$$\phi \equiv r^{2N-2}G_{,i}G_{,i} \tag{3.4}$$

takes its maximum value either on ∂D, or at the point Q. We now show that unless D is an N-ball and Q is the center point, the maximum of ϕ cannot occur at Q.

Using (3.1) or (3.1′) we observe that ϕ may be rewritten as

$$\phi = \omega_N^{-2} + 2\omega_N^{-1}r^{N-2}x_ig_{,i} + r^{2(N-1)}g_{,i}g_{,i}, \tag{3.5}$$

so that our theorem implies

$$\phi \leqq \max\left\{\max_{\partial D} r^{2(N-1)}\left(\frac{\partial G}{\partial n}\right)^2, \omega_N^{-2}\right\}, \tag{3.6}$$

where $\partial G/\partial n$ denotes the outward normal derivative of G on ∂D. We now show that

$$\omega_N^{-2} \leqq \max_{\partial D}\left[r^{2(N-1)}\left(\frac{\partial G}{\partial n}\right)^2\right], \tag{3.7}$$

with equality iff $g \equiv$constant, i.e. iff D is an N-ball and Q is its center. To establish (3.7) it suffices to show that in a neighborhood of Q there are points at which $\phi > \omega_N^{-2}$. Let K_ε be a ball of radius ε with center at Q such that $K_\varepsilon \subset D$. Clearly, the last term in (3.5) is everywhere nonnegative. We need only show that for $g \not\equiv$constant the second term on the right of (3.5) changes sign in K_ε. For any value of ρ in $[0, \varepsilon]$ we have

$$\oint_{\partial K_\rho} r^{N-2} x_i g_{,i}\, ds = \rho^{N-1} \oint_{\partial K_\rho} \frac{\partial g}{\partial n}\, ds = \rho^{N-1} \int_{K_\rho} \Delta g\, dx = 0. \tag{3.8}$$

Thus unless $g \equiv$constant, the quantity $r^{N-2} x_i g_{,i}$ must change sign in K_ε. This establishes (3.7), from which it follows that

$$\phi \leqq \max_{\partial D} r^{2(N-1)} \left(\frac{\partial G}{\partial n} \right)^2. \tag{3.6'}$$

The equality sign can hold in (3.7) only if $g \equiv$constant in K_ε (and hence by analyticity, in D), in which case D must be an N-ball and Q its center.

We now seek an upper bound for $\max_{\partial D} r^{2(N-1)} (\partial G / \partial n)^2$. As indicated in [6], it is known that if we choose a domain $\tilde{D}$ such that $D \subset \tilde{D}$ and let ∂D and $\partial \tilde{D}$ share a common point P, then

$$\left| \frac{\partial G_D}{\partial n} \right| \leqq \left| \frac{\partial G_{\tilde{D}}}{\partial n} \right| \quad \text{at } P. \tag{3.9}$$

If D is convex, then for any point P of ∂D we may choose $\tilde{D}$ as the half space which contains D and shares the common boundary point P. This yields for any boundary point P the estimate

$$\left| \frac{\partial G_D}{\partial n} \right| \leqq \frac{2 n_r}{\omega_N r^{N-1}}, \qquad N \geqq 2, \tag{3.10}$$

where n_r is the projection of the unit outward normal vector at P onto the straight line joining P and Q. Since ϕ takes its maximum value on ∂D, we conclude by means of (3.10) that

$$|\operatorname{grad} G| \leqq \frac{2}{\omega_N} r^{1-N}, \qquad N \geqq 2, \tag{3.11}$$

an inequality which holds for convex D at any point $P \neq Q$ in D. The equality in (3.9) is never realized for any bounded domain D; so we cannot expect (3.11) to be sharp. Nevertheless it appears to be sharper than similar bounds obtained by Bramble and Payne in [1]. This result can easily be extended to nonconvex domains by choosing for $\tilde{D}$ the exterior of a sphere.

From (3.11) we can compute a bound for $G(Q, R)$ at a point R in D. Consider the ray from Q through R which intersects ∂D at P_1. Let d denote the distance from Q to P_1, and r the distance from Q to R. Then by integrating (3.11) along the ray we obtain for D convex

$$G(Q, R) \leqq \begin{cases} \dfrac{2}{(N-2)\omega_N} [r^{-(N-2)} - d^{-(N-2)}], & N \geqq 3, \\[2ex] \dfrac{1}{\pi} \log \dfrac{d}{r}, & N = 2. \end{cases} \tag{3.12}$$

b) Bounds for electrostatic capacity and charge density. A classical problem of electrostatics asks for the solution of the following exterior Dirichlet problem:

$$\Delta u = 0 \quad \text{in} \quad D^* \equiv R^3 - \bar{D}, \qquad u = 1 \quad \text{on} \quad \partial D,$$

(3.13)

$$u = O\left(\frac{1}{r}\right) \quad \text{as} \quad r \to \infty.$$

Here u is the electrostatic potential of the conductor and r measures the distance from some convenient origin inside D. The charge density on ∂D is given by $|\text{grad } u|$. We introduce the function

(3.14)
$$\psi(x) = \frac{u_{,i}u_{,i}}{u^4}, \qquad x \in D^*,$$

which according to the theorem established in § 2 takes its maximum value either on ∂D, or at infinity. An easy computation shows that

(3.15)
$$\lim_{r \to \infty} \psi(x) = C^{-2},$$

where C is the capacity of the conductor. This follows from the well-known expansion of u for large r. Thus

(3.16)
$$\psi(x) \leqq \max\,[\max_{\partial D} \psi(x), C^{-2}].$$

We show now that

(3.17)
$$C^{-2} \leqq \max_{\partial D} \psi(x),$$

with equality iff D is a sphere. To effect the proof of (3.17), consider the level surface $u = \tilde{u}$ where $\tilde{u} \in (0, 1)$. From the expansion of u in the neighborhood of infinity, we know that for sufficiently small $\tilde{u}$, the surface $u = \tilde{u}$ will be starshaped with respect to the origin. Consider such a starshaped surface $\partial D(\tilde{u})$ and denote by $D^*(\tilde{u})$ the region exterior to $\partial D(\tilde{u})$. From the identity

(3.18)
$$0 = \int_{D^*(\tilde{u})} x_i u_{,i} \Delta u \, dx = \oint_{\partial D(\tilde{u})} x_i u_{,i} \frac{\partial u}{\partial n} ds + \frac{1}{2} \int_{D^*(\tilde{u})} u_{,i} u_{,i} \, dx$$
$$- \frac{1}{2} \oint_{\partial D(\tilde{u})} x_i n_i u_{,j} u_{,j} \, ds,$$

used previously by Payne and Weinberger in [4], we obtain

(3.19)
$$-\oint_{\partial D(\tilde{u})} x_i n_i \left(\frac{\partial u}{\partial n}\right)^2 ds = \tilde{u} \oint_{\partial D(\tilde{u})} \frac{\partial u}{\partial n} ds = 4\pi C \tilde{u}.$$

(Note that in the boundary integrals, the normal vector points outward from $D^*(\tilde{u})$, and thus $x_i n_i$ is negative.) It follows then from (3.19) that

(3.20)
$$\frac{4\pi C}{\tilde{u}^3} = -\oint_{\partial D(\tilde{u})} x_i n_i \psi \, ds \leqq 3 \tilde{V} \max_{\partial D(\tilde{u})} \psi,$$

where $\tilde{V}$ is the volume of $D(\tilde{u})$, the region interior to $\partial D(\tilde{u})$. Now the classical Poincaré inequality (see Pólya and Szegö [5]) states that

(3.21)
$$\tilde{V} \leqq \frac{4\pi}{3} \tilde{C}^3,$$

where $\tilde{C}$ is the capacity of the condenser with boundary $\partial D(\tilde{u})$. We must now relate $\tilde{C}$ to C. Since $\partial D(\tilde{u})$ is a level surface of u it is easily seen that

$$C = \tilde{C}\tilde{u}. \tag{3.22}$$

From (3.20), (3.21), (3.22) we obtain

$$C^{-2} \leqq \max_{\partial D(\tilde{u})} \psi. \tag{3.23}$$

It is again easily seen that the equality sign can hold for all such starshaped level surfaces iff D is a sphere. But (3.23) shows that there are points in D^* where $\psi \geqq C^{-2}$, from which (3.17) follows (making use of 3.16). We have thus established that

$$\psi(x) \leqq \max_{\partial D} \psi(x), \tag{3.24}$$

with equality iff D is a sphere.

At a point P_0 on ∂D where ψ assumes its maximum value it follows from Hopf's second principle that either D is a sphere or

$$\frac{\partial \psi}{\partial n} = 2\frac{\partial u}{\partial n}\frac{\partial^2 u}{\partial n^2} - 4\left(\frac{\partial u}{\partial n}\right)^3 > 0. \tag{3.25}$$

From the differential equation (3.13) evaluated on ∂D:

$$\frac{\partial^2 u}{\partial n^2} = 2K\frac{\partial u}{\partial n}, \tag{3.26}$$

and (3.24), we obtain

$$0 < \frac{\partial u(P_0)}{\partial n} < K(P_0).\dagger \tag{3.27}$$

Since $\psi(x)$ and $\partial u/\partial n$ take their maximum values at the same point on ∂D it follows that either D is a sphere or

$$\max_{\partial D} \frac{\partial u}{\partial n} < K(P_0) < \max_{\partial D} K \equiv K_0. \tag{3.28}$$

Inequalities (3.17) and (3.28) now give the following bound for the capacity of D:

$$C \geqq K_0^{-1}, \tag{3.29}$$

where the equality sign holds iff D is a sphere.

Other isoperimetric inequalities result from integration of the inequality

$$\psi(x) \leqq K_0^2 \tag{3.30}$$

in various ways. For instance if we take the square root of both sides and integrate over ∂D, we obtain

$$4\pi C \leqq K_0 S, \tag{3.31}$$

where S is the surface area of ∂D. On the other hand, if D is starshaped, multiplication of (3.30) by $-x_i n_i$ and integration over ∂D yields the inequality

$$4\pi C \leqq 3VK_0^2, \tag{3.32}$$

with equality holding in (3.31) and (3.32) iff D is a sphere. Similarly, an integration of

† The symbol $K(P)$ denotes the average curvature of ∂D at P.

$|\text{grad } u| \leqq K_0 u^2$ along a straight line joining any point $A \in D^*$ to the nearest point on ∂D leads to

$$u(A) \geqq (1 + K_0 d)^{-1}, \tag{3.33}$$

where d is the distance from A to ∂D.

Another inequality which follows directly from (3.30) is

$$4\pi C \leqq K_0^2 \int_{D^*} u^4 \, dx. \tag{3.34}$$

We may also use (3.28) to obtain some idea of the location of the point P_0 of maximum charge density on ∂D. In fact from (3.28) we find that at P_0 the following inequalities must be satisfied (assuming D is not a sphere):

$$K(P_0) > \max_{\partial D} \frac{\partial u}{\partial n} \geqq \frac{1}{S} \oint_{\partial D} \frac{\partial u}{\partial n} \, ds = \frac{4\pi C}{S}, \tag{3.35}$$

and in case D is starshaped with respect to the origin

$$K(P_0) > \left\{ \frac{-\oint_{\partial D} x_i n_i (\partial u / \partial n)^2 \, ds}{-\oint_{\partial D} x_i n_i \, ds} \right\}^{1/2} = \left(\frac{4\pi C}{3V} \right)^{1/2}. \tag{3.36}$$

Inequality (3.35) holds for nonstarshaped regions.

Similar inequalities can be derived for the expression

$$\psi_1 = r^4 u_{,i} u_{,i}, \tag{3.37}$$

where u is the solution of (3.13), but the results in this case are of less interest. Likewise we could apply our theorem to

$$\phi_1 = \frac{G_{,i} G_{,i}}{[G + \gamma]^{2(N-1)/(N-2)}}, \qquad N \geqq 3, \tag{3.38}$$

where G is the Green's function introduced earlier, and γ is an appropriate constant. This would give a lower bound for $\max_{\partial D} (\partial G / \partial n)$. However, such a bound seems to be of limited interest.

4. Extensions and conclusions. It is clear from the proof of the theorem established in § 2 that for $N = 2$, if $(\log f)'' \geqq 0$ in D and if grad H does not vanish in D, then the function ψ defined in (2.3) will take its *minimum value* on ∂D. In particular, if f is an exponential function of h and if grad $H \neq 0$ in D, then ψ will assume both its maximum and minimum value on ∂D. This would yield additional information on the Green's function and on logarithmic potential problems in two dimensions. One would expect that for $N \geqq 3$, a suitable hypothesis on $f(h)$ might insure that in cases in which $H_{,i} H_{,i}$ does not vanish in D, ψ will satisfy a minimum principle, but results in this direction appear to be less useful.

We have considered in this paper a particular combination of harmonic functions and their gradients. Many other combinations can easily be shown to satisfy a maximum principle. For instance, if H_i, $i = 0, 1, \cdots, m$ are harmonic functions and $H_0 > 0$

in $\bar{D}$, then it follows rather easily from the generalized maximum principle in [6] that the quantity $H_0^{-2} \sum_{i=1}^{m} H_i H_i$ takes its maximum value on ∂D. A similar result holds of course for solutions of more general second order elliptic equations.

Acknowledgment. The authors wish to thank the referee for his helpful comments and suggestions.

REFERENCES

[1] J. H. Bramble and L. E. Payne, *Bounds for the first derivatives of Green's function*, Acad. Naz. dei Lincei Ser. 8, 42 (1967) pp. 604–610.
[2] E. Hopf, *Elementare Bemerkung über die Lösung partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus*, Sitzungsber. Preuss. Akad. Wiss., 19 (1927), pp. 147–152.
[3] ———, *A remark on linear elliptic differential equations of the second order*, Proc. Amer. Math. Soc., 3 (1952), pp. 791–793.
[4] L. E. Payne and H. F. Weinberger, *New bounds in harmonic and biharmonic problems*, J. Math. Phys. 33 (1955), pp. 291–307.
[5] G. Pólya and G. Szegö, *Isoperimetric inequalities in mathematical physics*, Princeton University Press, Princeton, NJ, 1951.
[6] M. H. Protter and H. F. Weinberger, *Maximum principles in differential equations*, Prentice-Hall, Englewood Cliffs, NJ, 1967.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/10001-0013 $01.00/0

ON SOLUTIONS OF A TRANSCENDENTAL EQUATION BASIC TO THE THEORY OF VIBRATING PLATES*

C. E. SIEWERT† AND J. S. PHELPS, III†

Abstract. The theory of complex variables is used to develop exact closed-form solutions of the transcendental equation $a \tan \xi + \tanh \xi = 0$.

1. Introduction. As discussed by Leissa [1] and Marguerre [2], the study of the vibration of elastic plates invariably leads to eigenfunction expansions. In many such cases the required eigenvalues are established as the solutions of transcendental equations. One such problem is that of the dually clamped oscillating plate. Here we seek a solution to

$$(\nabla^4 - k^4)W(x, y) = 0, \tag{1}$$

with $W(\alpha, y) = W(0, y) = 0$, $W(x, 0) = f(x)$, and $W(x, \beta) = g(x)$. The solution for $W(x, y)$ can be established by separation of variables, with the x component expressed as

$$X(x) = \sinh\left(\frac{\gamma}{2}\right) \cos\left[\gamma\left(\frac{x}{\alpha} - \frac{1}{2}\right)\right] + \sin\left(\frac{\gamma}{2}\right) \cosh\left[\gamma\left(\frac{x}{\alpha} - \frac{1}{2}\right)\right], \tag{2a}$$

with

$$\tanh(\gamma/2) + \tan(\gamma/2) = 0, \tag{2b}$$

or

$$X(x) = \sinh\left(\frac{\gamma}{2}\right) \sin\left[\gamma\left(\frac{x}{\alpha} - \frac{1}{2}\right)\right] - \sin\left(\frac{\gamma}{2}\right) \sinh\left[\gamma\left(\frac{x}{\alpha} - \frac{1}{2}\right)\right], \tag{3a}$$

with

$$\tanh(\gamma/2) - \tan(\gamma/2) = 0. \tag{3b}$$

We wish here to investigate the transcendental equation

$$a \tan \xi + \tanh \xi = 0, \tag{4}$$

which clearly contains the foregoing as special cases.

2. General analysis: $|\xi| \leqq \pi/2$. In order to find the real and imaginary solutions of

$$a \tan \xi + \tanh \xi = 0, \qquad a \in (-\infty, \infty), \tag{5}$$

we first wish to introduce and study the sectionally analytic function

$$F(z) = \mathrm{Log}(z+1) - \mathrm{Log}(z-1) - i\frac{|a|}{a}[\log(z + i|a|) - \log(z - i|a|)]. \tag{6}$$

Here we use the standard notation Log (ζ) to represent the principal branch of the log function, i.e.,

$$\mathrm{Log}(\zeta) = \ln|\zeta| + i \arg(\zeta), \qquad \arg(\zeta) \in (-\pi, \pi). \tag{7}$$

---

* Received by the editors January 28, 1977, and in revised form July 18, 1977.

† Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina 27607.

For the functions $\log(z \pm i|a|)$ appearing in (6) we use branches of the log function such that

$$\log(\zeta) = \ln|\zeta| + i \arg(\zeta), \qquad \arg(\zeta) \in \left(\frac{-\pi}{2}, \frac{3\pi}{2}\right). \tag{8}$$

With these choices of the log functions it is clear that $F(z)$ is analytic in the complex z plane cut from -1 to 1 along the real axis and from $-i|a|$ to $i|a|$ along the imaginary axis. It is a simple matter to show that

$$F(z) = \frac{2}{z}(1+a) + \frac{2}{3z^3}(1-a^3) + O\left(\frac{1}{z^5}\right), \qquad \text{as } |z| \to \infty. \tag{9}$$

We now wish to use the argument principle [3] to establish the number of zeros of $F(z)$ inside the contours C_1 and C_2, shown in Fig. 1, as $R \to \infty$ and $\varepsilon \to 0$. Since in general $F(z)$ vanishes as $1/z$ as $|z| \to \infty$, we find that the argument of $F(z)$ decreases by 2π as the contour C_1 is traversed (in the positive sense). For the special case $a = -1$, $F(z)$ vanishes as $1/z^3$ as $|z| \to \infty$, and thus for this case the argument of $F(z)$ decreases by 6π as C_1 is traversed.

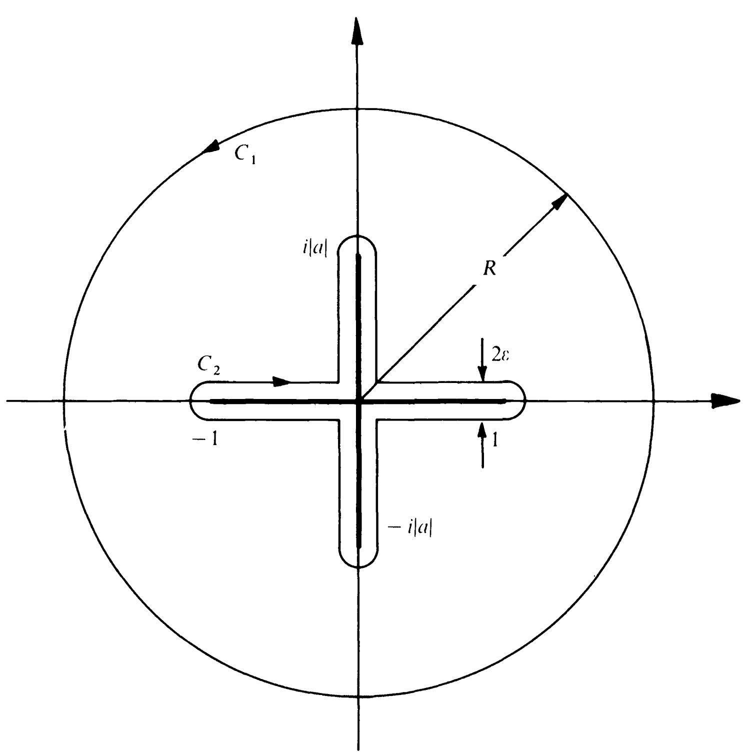

FIG. 1. *The contours C_1 and C_2.*

To compute the change in the argument of $F(z)$ as the contour C_2 is traversed, we first require the limiting values $F^{\pm}(x)$ of $F(z)$ as z approaches the cut $[-1, 1]$ from above (+) and below (−) and the limiting values $F^{\pm}(iy)$ as z approaches the cut $[-i|a|, i|a|]$ from the left (+) and the right (−). It is a relatively straightforward matter to show that

$$F^{\pm}(x) = R(x) \mp i\pi, \qquad x \in [-1, 1], \tag{10}$$

where

(11) $$R(x)=2\tanh^{-1}(x)+\frac{|a|}{a}\left[\operatorname{sgn}(x)\pi-2\operatorname{Tan}^{-1}\left(\frac{x}{|a|}\right)\right].$$

We use here the convention that $\operatorname{Tan}^{-1}(x)$ denotes the principal branch of the arctan function. In a similar manner, we can compute the limiting values of $F(z)$ as z approaches the cut along the imaginary axis. We find

(12) $$F^{\pm}(iy)=\mp\frac{|a|}{a}\pi+iI_0(y),\qquad y\in[-|a|,|a|],$$

where

(13) $$I_0(y)=-\frac{2|a|}{a}\tanh^{-1}\left(\frac{y}{|a|}\right)+2\operatorname{Tan}^{-1}(y)-\operatorname{sgn}(y)\pi.$$

If we use Δ_2 to denote the change in the argument of $F(z)$ as the contour C_2 is traversed, in limit as $\varepsilon\to 0$, we can use (10) and (12) to deduce that $\Delta_2=2\pi$, for $a>0$, and that $\Delta_2=6\pi$ for $a<0$. (The special case $a=0$ clearly is not interesting.) We thus conclude that for $a>0$, $F(z)$ has only a zero at infinity; on the other hand, for $a<0$ we note that, in general, $F(z)$ has two zeros in the finite plane plus one zero at infinity. For the special case of $a=-1$, $F(z)$ clearly has only a triple zero at infinity.

In order to relate the zeros of $F(z)$ to the desired solutions of (5), let us first deduce the special forms of $F(z)$ for $z=x\in(-\infty,-1)\cup(1,\infty)$ and for $z=iy$, $y\in(-\infty,-|a|)\cup(|a|,\infty)$. Evaluating $F(z)$ on that part of the real axis that excludes the cut, we find

(14) $$F(x)=2\tanh^{-1}\left(\frac{1}{x}\right)+\frac{|a|}{a}\left[\operatorname{sgn}(x)\pi-2\operatorname{Tan}^{-1}\left(\frac{x}{|a|}\right)\right],\qquad x\in(-\infty,-1)\cup(1,\infty).$$

On that part of the imaginary axis that excludes the cut we find

(15)
$$F(iy)=i\left[\frac{-2|a|}{a}\tanh^{-1}\left(\frac{|a|}{y}\right)+2\operatorname{Tan}^{-1}(y)-\operatorname{sgn}(y)\pi\right],\qquad y\in(-\infty,-|a|)\cup(|a|,\infty).$$

If we now consider $a\in(-\infty,-1)$, we can deduce from (14) that $F(z)$ has (in addition to a zero at infinity) two real zeros $\pm x_0$, $x_0\in(1,\infty)$. It follows therefore that $\pm\hat{\xi}_0$, where

(16) $$\hat{\xi}_0=i\operatorname{Tan}^{-1}\left(\frac{|a|}{x_0}\right),\qquad a\in(-\infty,-1),$$

are two of the desired solutions of (5) for this case. Considering now $a\in(-1,0)$, we conclude that $F(z)$ has (in addition to a zero at infinity) two imaginary zeros $\pm iy_0$, $y_0\in(|a|,\infty)$. Thus we observe from (15) that $\pm\xi_0$, where

(17) $$\xi_0=\tanh^{-1}\left(\frac{|a|}{y_0}\right),\qquad a\in(-1,0),$$

are two of the desired solutions for the considered values of the parameter a. To summarize our conclusions thus far we note that for $a>0$, $F(z)$ has only a zero at infinity which corresponds to the trivial solution ($\xi_0=0$) of (5). For $a\in(-\infty,-1)$, $F(z)$ has two additional real zeros $\pm x_0$ which correspond to the imaginary solutions $\pm\hat{\xi}_0$, where $\hat{\xi}_0$ is given by (16). For $a\in(-1,0)$, $F(z)$ has, in addition to a zero at infinity, two imaginary zeros $\pm iy_0$ which correspond to the real solutions $\pm\xi_0$, where ξ_0 is given by (17).

We note from (5) that

$$\hat{\xi}_0(a) = i\xi_0\left(\frac{1}{a}\right), \qquad a \in (-\infty, -1), \tag{18}$$

and thus we need here only $\xi_0(a)$, $a\in(-1, 0)$, in order to establish the real and imaginary solutions of (5) such that $|\xi_0| \leqq \pi/2$.

If we now consider only $a<0$ and let $\pm z_0$ denote the finite zeros of $F(z)$, then we note that the function

$$T(z) \triangleq \frac{F(z)}{z^2 - z_0^2} \tag{19}$$

is analytic in the complex plane cut along $L = [-1, 1] \cup [-i|a|, i|a|]$. In addition, $T(z)$ is nonvanishing in the finite plane, and the limiting values of $T(z)$ satisfy the Riemann–Hilbert problem [4]

$$T^+(\tau) = \left[\frac{F^+(\tau)}{F^-(\tau)}\right] T^-(\tau), \qquad \tau \in L. \tag{20}$$

It thus follows [4] that $T(z)$ can differ from any canonical solution of the Riemann–Hilbert problem by no more than a constant multiple. Thus we can write

$$\frac{F(z)}{z^2 - z_0^2} = KX(z), \tag{21}$$

where $X(z)$ is a canonical solution to the considered Riemann–Hilbert problem and K is a constant to be established. The desired canonical solution $X(z)$ can be constructed from the work of Muskhelishvili [4]; some care is required, however, to be sure that the "endpoint behavior" is correct. We find

$$X(z) = \frac{1}{z^3} \exp\left[\frac{2}{\pi}\int_0^1 x\theta_0(x)\frac{dx}{x^2 - z^2} + \frac{2}{\pi}\int_0^{|a|} y\phi_0(y)\frac{dy}{y^2 + z^2}\right], \tag{22}$$

where

$$\theta_0(x) = \tan^{-1}\left[\frac{-\pi}{R(x)}\right] \tag{23a}$$

and

$$\phi_0(y) = \tan^{-1}\left[\frac{-\pi}{I_0(y)}\right]. \tag{23b}$$

Here $\theta_0(x)$ and $\phi_0(y)$ are continuous, with $\theta_0(0) = \phi_0(0) = -3\pi/4$ and $\theta_0(1) = \phi_0(|a|) = 0$.

We can now substitute (22) into (21) and let $|z| \to \infty$ to find to find $K = 2(1+a)$, and thus we can solve (21) to obtain the general result

$$z_0^2 = z^2 - \frac{F(z)}{2(1+a)X(z)}, \qquad a<0. \tag{24}$$

Equation (24) represents a general solution for the zeros of $F(z)$ and is valid for any value of z. We can let $|z| \to \infty$ in (24) to find the specific result

$$z_0^2 = \frac{2}{\pi}\int_0^{|a|} \phi_0(y) y\, dy - \frac{2}{\pi}\int_0^1 \theta_0(x) x\, dx - \frac{(1-a^3)}{3(1+a)}, \qquad a<0. \tag{25}$$

It is clear that (25) can be used in

(26a) $$\xi_0 = \tanh^{-1}\left(\frac{|a|}{|z_0|}\right), \qquad a \in (-1, 0),$$

or

(26b) $$\xi_0 = i\,\mathrm{Tan}^{-1}\left(\frac{|a|}{|z_0|}\right), \qquad a \in (-\infty, -1),$$

to give exact analytical results for the desired solutions ($\pm\xi_0$, $|\xi_0| \leqq (\pi/2)$) of (5). In the next section we develop similar expressions for the solutions such that $|\xi_0| \geqq (\pi/2)$.

3. General analysis: $|\xi| \geqq (\pi/2)$. Here we wish to generalize the analysis of the previous section in order to find additional real and imaginary solutions of (5). If we let

(27) $$F_k(z) = F(z) + 2k\pi i, \qquad k = 1, 2, 3 \cdots,$$

then we conclude that $F_k(z)$ is analytic in the plane cut along L and has limiting values

(28) $$F_k^{\pm}(x) = R(x) + i(2k \mp 1)\pi, \qquad x \in [-1, 1],$$

and

(29) $$F_k^{\pm}(iy) = \mp\frac{|a|}{a}\pi + iI_k(y), \qquad y \in [-|a|, |a|].$$

Here $R(x)$ is given by (11) and

(30) $$I_k(y) = I_0(y) + 2k\pi.$$

If we use the argument principle again, we find that $F_k(z)$ has exactly one zero in the finite plane. Note that if we were to allow k to be negative we could write $F_{-k}(z) = -F_k(-z)$; thus the zeros corresponding to negative values of k are just the negative of the zeros corresponding to positive values of k. If now we evaluate $F_k(z)$ on the imaginary axis, but not on the cut, we find that

(31) $$F_k(iy) = F(iy) + 2k\pi i, \qquad y \in (-\infty, -|a|) \cup (|a|, \infty),$$

always has one simple zero, say y_k. It follows from (31) that $\pm\xi_k$, where

(32) $$\xi_k = \left(k - \frac{1}{2}\frac{|a|}{a}\right)\pi + \frac{|a|}{a}\mathrm{Tan}^{-1}(|y_k|), \qquad k = 1, 2, 3 \cdots,$$

are the additional real solutions of (5) that we seek. As in the previous section, we can generate the imaginary solutions $\pm\hat{\xi}_k$ of (5) by

(33) $$\hat{\xi}_k(a) = i\xi_k\left(\frac{1}{a}\right), \qquad a \in (-\infty, \infty).$$

We now observe that

(34) $$\frac{F_k(z)}{z - iy_k} = K_k X_k(z),$$

where K_k is a constant to be established and $X_k(z)$ is a canonical solution of the

Riemann–Hilbert problem defined by

$$X_k^+(\tau)=\left[\frac{F_k^+(\tau)}{F_k^-(\tau)}\right]X_k^-(\tau), \qquad \tau\in L. \tag{35}$$

We find that $X_k(z)$ can be written as

$$X_k(z)=\frac{1}{z-ia}\exp\left[\frac{1}{2\pi i}\int_0^1 [z\ln M_k(x)+2ix\theta_k(x)]\frac{dx}{x^2-z^2} + \frac{1}{\pi}\int_{-|a|}^{|a|}\phi_k(y)\frac{dy}{y+iz}\right], \tag{36}$$

where

$$M_k(x)=\frac{R^2(x)+(2k-1)^2\pi^2}{R^2(x)+(2k+1)^2\pi^2}, \tag{37}$$

$$\theta_k(x)=\tan^{-1}\left[\frac{-2\pi R(x)}{R^2(x)+\pi^2(4k^2-1)}\right], \tag{38}$$

and

$$\phi_k(y)=\tan^{-1}\left[\frac{(|a|/a)\pi}{I_k(y)}\right]. \tag{39}$$

The angle defined by (38) is continuous for $x\in(0,1)$, with $\theta_k(0)=\tan^{-1}(-|a|/a)/(2k^2)$. As y varies from $-|a|$ to $|a|$, the angle $\phi_k(y)$ varies from $-\pi\to 0$, for $a<0$, and from $0\to\pi$, for $a>0$; we note that $\phi_k(y)$ has a discontinuity at $y=0$.

If we now substitute (36) into (34) and let $|z|\to\infty$, we find that $K_k=2k\pi i$. Thus we can solve (34) to obtain the explicit result

$$y_k=-iz+\frac{F_k(z)}{2k\pi X_k(z)}, \qquad a\in(-\infty,\infty). \tag{40}$$

Equation (40) is valid for any z, and thus can be substituted into (32) to give the remaining real solutions of (5). To obtain a specific form of (40), we can let $|z|\to\infty$ to find

$$y_k=a+\frac{1+a}{k\pi}+\frac{1}{2\pi}\int_0^1 \ln M_k(x)\,dx-\frac{|a|}{\pi}\int_{-1}^1 \phi_k(|a|x)\,dx. \tag{41}$$

4. Conclusions. We have successfully found all of the real and imaginary solutions of (5). The real solution corresponding to $k=0$ is given by (25) and (26a) for $a\in(-1,0)$, and the imaginary solutions are given by (25) and (26b) for $a\in(-\infty,-1)$. For $a>0$ there are no real or imaginary solutions corresponding to $k=0$. For $a\in(-\infty,\infty)$ and $k=1,2,3\cdots$, the real solutions of (5) are given by (32) and (41); the imaginary solutions are given by (33). Of course, if ξ is a solution, so is $-\xi$.

To be sure that our final results are free of errors, we have evaluated (25), (26), (32) and (41) numerically for various values of a and k; without difficulty solutions correct to six significant figures were obtained.

Acknowledgment. The authors are grateful to Dr. M. N. Özişik for suggesting this problem.

REFERENCES

[1] A. W. LEISSA, *Vibration of Plates*, NASA SP-160 National Aeronautic and Space Administration, Washington, DC, 1969.
[2] K. MARGUERRE, *Elastic Plates*, Blaisdell, Waltham, MA, 1969.
[3] L. V. AHLFORS, *Complex Analysis*, McGraw-Hill, New York, 1953.
[4] N. I. MUSKHELISHVILI, *Singular Integral Equations*, Noordhoff, Groningen, The Netherlands, 1953.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0014 $01.00/0

A VARIATIONAL FORMULA FOR THE GROWTH RATE OF A POSITIVE OPERATOR SEMIGROUP*

THOMAS G. KURTZ†

Abstract. Let B be a Banach space, M a convex subset of B^* such that $\|\nu\| \leqq 1$ for all $\nu \in M$,

$$K = \{f \in B: \nu f \geqq 0, \text{ for all } \nu \in M\}$$

and

$$K_0 = \{f \in B: \nu f > 0 \text{ for all } \nu \in M\}.$$

Let $T(t)$ be a strongly continuous semigroup on B with infinitesimal generator A. If $T(t)$: $K \to K$ then

$$\Psi_2 \equiv \inf_{u \in D(A) \cap K_0} \sup_{\nu \in M} \frac{\nu A u}{\nu u} = \inf\{\lambda : \exists u \in D(A) \cap K_0 \ni \lambda u - Au \in K\} \equiv \lambda_1.$$

If, in addition, $\|f\| = \sup_{\nu \in M} \nu f$ for all $f \in K_0$ then

$$\Psi_2 = \lambda_1 = \lambda_2 \equiv \inf_{f \in K_0} \overline{\lim_{t \to \infty}} \frac{1}{t} \log \|T(t)f\|.$$

Other similarly defined quantities are also considered.

Introduction. In [1] Donsker and Varadhan gave a variational formula for the principal eigenvalue of an operator $A = L + V$ where L generates a positive semigroup on $C(X)$, the space of continuous functions on a compact metric space, satisfying $T(t)1 = 1$, and V is multiplication by a function $V \in C(X)$. After a minor transformation their variational formula can be seen to be an analog of the variational (minimax) formulas for the spectral radius of a positive operator and the analogous quantity for more general order preserving mappings. See for example [2], [3], [5], [6].

In this paper we consider this variational formula applied to more general operators A, requiring only that A generate a strongly continuous semigroup of linear operators $T(t)$ on a Banach space B which satisfies the following positivity condition:

Let M be a convex subset of B^* such that $\|\nu\| \leqq 1$ for all $\nu \in M$ and define

$$K = \{f \in B : \nu f \geqq 0 \text{ for all } \nu \in M\}$$

and

$$K_0 = \{f \in B : \nu f > 0 \text{ for all } \nu \in M\}.$$

Positivity condition.

$$T(t): K \to K. \tag{1}$$

While it is not necessarily true that $T(t)$: $K_0 \to K_0$, it does follow that

$$(\lambda - A)^{-1} = \int_0^\infty e^{-\lambda t} T(t)\, dt: K_0 \to K_0. \tag{2}$$

In [1] Donsker and Varadhan defined $\lim_{t \to \infty} (1/t) \log \|T(t)\|$ to be the "principal eigenvalue" of A. This limit always exists due to the subadditivity

$$\log \|T(t+s)\| \leqq \log \|T(t)\| + \log \|T(s)\|, \tag{3}$$

but, as Donsker and Varadhan point out, it need not be an eigenvalue of A in the usual sense. It does provide a measure of the rate of growth of $T(t)$; however, in our

* Received by the editors June 27, 1977.

† Mathematics Department, University of Wisconsin—Madison, Madison, Wisconsin 53706.

more general setting it may not be the most appropriate measure. With this in mind we define the following:

$$\lambda_0 \equiv \sup\left\{\lambda : \sup_{\nu\in M}\int_0^\infty e^{-\lambda t}\nu T(t)f\,dt = \infty \text{ for all } f\in K_0\right\}; \tag{4}$$

$$\lambda_1 = \inf\{\lambda : \exists u \in D(A)\cap K_0 \ni \lambda u - Au \in K\}; \tag{5}$$

$$\lambda_2 \equiv \inf_{f\in K_0}\overline{\lim_{t\to\infty}}\frac{1}{t}\log\|T(t)f\|; \tag{6}$$

$$\lambda_3 = \lim_{t\to\infty}\frac{1}{t}\log\|T(t)\|. \tag{7}$$

Our primary concern is the relationship of these quantities to the variational formulas

$$\Psi_1 \equiv \sup_{\nu\in M}\inf_{u\in D(A)\cap K_0}\frac{\nu Au}{\nu u}, \tag{8}$$

and

$$\Psi_2 \equiv \inf_{u\in D(A)\cap K_0}\sup_{\nu\in M}\frac{\nu Au}{\nu u}.$$

These relationships are given by the following theorem and its corollaries.

THEOREM I. *Let $T(t)$ be a strongly continuous semigroup satisfying the positivity condition* [(1) *above*]. *Then*

$$\Psi_1 \leqq \lambda_0 \leqq \Psi_2 = \lambda_1 \leqq \lambda_2 \leqq \lambda_3. \tag{9}$$

COROLLARY II. *If $\|f\| = \sup_{\nu\in M}|\nu f|$ for every $f\in K_0$ then $\lambda_0 = \Psi_2 = \lambda_1 = \lambda_2$. If $\|f\| = \sup_{\nu\in M}(|\nu f|/\|\nu\|)$ for every $f\in K_0$ then $\Psi_2 = \lambda_1 = \lambda_2$.*

COROLLARY III. *If M is weak* compact then $\Psi_1 = \lambda_0 = \Psi_2$.*

COROLLARY IV. *If the conditions of Corollaries* II *and* III *hold then $\Psi_1 = \lambda_0 = \Psi_2 = \lambda_1 = \lambda_2 = \lambda_3$.*

Remark. Corollary IV includes the situation considered by Donsker and Varadhan.

The condition of Corollary III can be weakened somewhat to give:

COROLLARY V. *If there is a weak* compact subset $N\subset M$ such that*

$$\sup_{\nu\in M}\frac{\nu Au}{\nu u} = \sup_{\nu\in N}\frac{\nu Au}{\nu u}$$

for all $u\in D(A)\cap K_0$ then $\Psi_1 = \lambda_0 = \Psi_2$.

Proof. The inequality $\lambda_2 \leqq \lambda_3$ is immediate. If $\lambda > \lambda_2$ then there is an $f\in K_0$ such that $u = \int_0^\infty e^{-\lambda t}T(t)f\,dt \in D(A)\cap K_0$ and $\lambda u - Au = f\in K$. Hence $\lambda \geqq \lambda_1$ and $\lambda_2 \geqq \lambda_1$. To see that $\Psi_2 \leqq \lambda_1$ observe that if $u\in D(A)\cap K_0$ and $(\lambda - A)u\in K$ then

$$\frac{\nu Au}{\nu u} = \lambda - \frac{\nu(\lambda - A)u}{\nu u} \leqq \lambda \tag{10}$$

for all $\nu\in M$. Consequently $\Psi_2 \leqq \lambda$ for any such λ and hence $\Psi_2 \leqq \lambda_1$ (the infimum of all such λ). On the other hand if $\lambda > \Psi_2$ there is a $u\in K_0$ such that

$$\sup_{\nu}\frac{\nu Au}{\nu u} < \lambda.$$

Consequently

$$\nu(\lambda u - Au) = \lambda \nu u - \frac{\nu A u}{\nu u} \nu u \geqq 0.$$

Therefore $\lambda u - Au \in K$ and $\lambda > \lambda_1$. It follows that $\Psi_2 \geqq \lambda_1$.

To complete the proof we need the following lemma.

LEMMA 1. *If* $u \in K \cap D(A)$ *then*

$$\nu u \geqq \int_0^\tau e^{-\lambda t} \nu T(t)(\lambda - A) u \, dt.$$

Proof. Since

$$\int_0^\tau e^{-\lambda t} T(t)(\lambda - A) u \, dt = \lambda \int_0^\tau e^{-\lambda t} T(t) u \, dt - A \int_0^\tau e^{-\lambda t} T(t) u \, dt$$
$$= u - e^{-\lambda \tau} T(\tau) u,$$

the inequality follows from the positivity conditions.

Let $u \in D(A) \cap K_0$ and $\lambda < \lambda_0$. Then

$$\sup_{\nu \in M} \frac{\nu A u}{\nu u} = \lambda - \inf_{\nu} \frac{\nu(\lambda - A) u}{\nu u}.$$

Either there exists ν such that $\nu(\lambda - A) u \leqq 0$ or $(\lambda - A) u \in K_0$. If $(\lambda - A) u \in K_0$ then Lemma 1 implies

$$\sup_{\nu \in M} \int_0^\infty e^{-\lambda t} \nu T(t)(\lambda - A) u \, dt \leqq \sup_{\nu \in M} \nu u \leqq \|u\|.$$

But this contradicts the fact that $\lambda < \lambda_0$. Consequently $\Psi_2 \geqq \lambda_0$.

Finally suppose $\lambda > \lambda_0$. Then there is an $f \in K_0$ such that

$$\int_0^\infty e^{-\lambda t} \nu T(t) f \, dt < \infty \tag{11}$$

for every $\nu \in M$.

Observe that $u_\tau = \int_0^\tau e^{-\lambda t} T(t) f \in D(A) \cap K_0$ and $(\lambda - A) u_\tau = f - e^{-\lambda \tau} T(\tau) f$. By (11) $\underline{\lim}_{\tau \to \infty} e^{-\lambda \tau} \nu T(\tau) f = 0$ for every $\nu \in M$. Consequently

$$\inf_\tau \frac{\nu A u_\tau}{\nu u_\tau} = \inf_\tau \left(\lambda - \frac{\nu(\lambda - A) u_\tau}{\nu u_\tau} \right)$$
$$= \inf_\tau \left(\lambda + \frac{\nu e^{-\lambda \tau} T(\tau) f}{\nu u_\tau} - \frac{\nu f}{\nu u_\tau} \right) \leqq \lambda.$$

Consequently $\Psi_1 \leqq \lambda$ for every $\lambda > \lambda_0$ and hence $\Psi_1 \leqq \lambda_0$.

Proof of Corollary II. If $\mu > \lambda_0$ there is an $f \in K_0$ such that

$$\sup_{\nu \in M} \int_0^\infty e^{-\mu t} \nu T(t) f \, dt < \infty. \tag{12}$$

Let $\lambda > \mu$ and define

$$u_\tau = \int_0^\tau e^{-\lambda t} T(t) f \, dt.$$

Then for $\tau_1 < \tau_2$

$$\|u_{\tau_1} - u_{\tau_2}\| = \sup_{\nu} \nu \int_{\tau_1}^{\tau_2} e^{-\lambda t} T(t) f \, dt \leqq e^{-(\lambda-\mu)\tau_1} \sup_{\nu} \int_0^{\infty} e^{-\mu t} \nu T(t) f \, dt.$$

Consequently $\lim_{\tau\to\infty} u_\tau = u$ exists and it follows that $(\lambda - A)u = f$. Finally,

$$e^{-\lambda\tau} T(\tau) u = u - \int_0^{\tau} e^{-\lambda t} T(t) f \, dt.$$

Since $f \in K_0$ we have

$$\sup_{\nu} \nu\, e^{-\lambda\tau} T(\tau) u = \|e^{-\lambda\tau} T(\tau) u\| \leqq \sup_{\nu} \nu u = \|u\|.$$

Consequently $\lambda_2 \leqq \lambda$ and hence $\lambda_2 = \lambda_0$. The second part of the corollary follows from the fact that Ψ_2, λ_1 and λ_2 do not change if we replace M by

$$\hat{M} = \left\{ \frac{\nu}{\|\nu\|} : \nu \in M \right\}.$$

Proof of Corollary III. Since $\nu A u/(\nu u)$ is a quasi-concave-convex function in the terminology of Sion [7], Sion's theorem implies $\Psi_1 = \Psi_2$.

Proof of Corollary IV. By the weak* compactness of M, if $f \in K_0$ then $\inf_{\nu\in M} \nu f > 0$. Consequently, if $f \in K_0$ then for every $g \in B$ there is a C such that $Cf - \frac{1}{2}g \in K_0$ and $Cf + \frac{1}{2}g \in K_0$. It follows that

$$\begin{aligned} \|e^{-\lambda t} T(t) g\| &\leqq \|e^{-\lambda t} T(t)(Cf - \tfrac{1}{2}g)\| + \|e^{-\lambda t} T(t)(Cf + \tfrac{1}{2}g)\| \\ &= \sup_{\nu\in M} \nu\, e^{-\lambda t} T(t)(Cf - \tfrac{1}{2}g) + \sup_{\nu\in M} \nu\, e^{-\lambda t} T(t)(Cf + \tfrac{1}{2}g) \\ &\leqq 4 \sup_{\nu\in M} \nu\, e^{-\lambda t} T(t) Cf \\ &= 4C\|e^{-\lambda t} T(t) f\|. \end{aligned} \tag{13}$$

If $\lambda > \lambda_2$ there is an $f \in K_0$ such that $\sup_t \|e^{-\lambda t} T(t) f\| < \infty$. Consequently $\sup_t \|e^{-\lambda t} T(t) g\| < \infty$ for all $g \in B$. The uniform boundedness theorem implies

$$\sup_t \|e^{-\lambda t} T(t)\| < \infty$$

and hence $\lambda \geqq \lambda_3$. It follows that $\lambda_2 = \lambda_3$.

Proof of Corollary V. Sion's theorem implies

$$\begin{aligned} \Psi_2 &= \inf_{u\in D(A)\cap K_0} \sup_{\nu\in M} \frac{\nu A u}{\nu u} = \inf_{u\in D(A)\cap K_0} \sup_{\nu\in N} \frac{\nu A u}{\nu u} \\ &= \sup_{\nu\in N} \inf_{u\in D(A)\cap K_0} \frac{\nu A u}{\nu u} \leqq \Psi_1. \end{aligned}$$

Examples. We now give examples showing that strict inequality is possible for each of the inequalities in Theorem I.

Example A $(\lambda_3 > \lambda_2)$. Let $B = \hat{C}(\mathbb{R})$, the space of continuous functions vanishing at infinity with the sup norm.

Let M be the set of measures ν with $\nu(\mathbb{R})=1$ and $\nu(U)>0$ for every open set U. Then

$$K_0=\{f\in \hat{C}(\mathbb{R}): f\geqq 0, f\neq 0\}.$$

Let $T(t)f(x)=e^{t\sin x}f(x)$. Then $\lambda_3=1$ and $\lambda_2=-1$.

Example B ($\lambda_2>\lambda_1$). Let $B=\hat{C}(\mathbb{R})$ and M be the space of positive measures with $\nu(-\infty,0)=1$ and $\nu[0,\infty)=0$. Then $K_0=\{f\in B: f(x)>0 \text{ for } x<0\}$.

Let $T(t)=f(x-t)$. Then $Af(x)=-f'(x)$. If $\lambda<0$ and $f(x)\in D(A)$ with $f(x)=e^{-\lambda x}$ for $x<0$, $\lambda f(x)-Af(x)=0$ for $x<0$ and hence $\lambda_1\leqq\lambda$. Therefore $\lambda_1=-\infty$. However, $\lambda_2=0$.

Example C ($\lambda_1=\Psi_2>\lambda_0$). Let $B=L^p(0,\infty)$, $p'<p$, $1/q+1/p=1$ and $1/q'+1/p'=1$. Let $M=\{g: g\geqq 0 \text{ a.e. } \|g\|_q\leqq 1, \|g\|_{q'}\leqq 1\}$. Then $K=\{f\in L^p: f\geqq 0 \text{ a.e.}\}$. Define $T(t)f(x)=f(e^{\mu t}x)$ for some $\mu>0$. Then

$$\|T(t)f\|_p=\left(\int_0^\infty |f(e^{\mu t}x)|^p\,dx\right)^{1/p}=e^{-(\mu/p)t}\|f\|_p.$$

If $-\mu/p>\lambda>-\mu/p'$ and $f\in L^p\cap L^{p'}\cap K_0$ then

$$\sup_{g\in M}\int_0^\infty e^{-\lambda t}\int_0^\infty g(x)T(t)f(x)\,dx\,dt<\infty.$$

This follows from the inequality

$$\int_0^\infty g(x)T(t)f(x)\,dx\leqq \|g\|_{q'},e^{-(\mu/p')t}\|f\|_{p'}.$$

Therefore $\lambda\geqq\lambda_0$ and $\lambda_0<-\mu/p$. On the other hand, if $u\in D(A)\cap K_0$ and $\lambda u-Au\in K$ we have by Lemma 1 that

$$u\geqq e^{-\lambda t}T(t)u \quad \text{a.e.}$$

Consequently

$$\|u\|_p\geqq\|e^{-\lambda t}T(t)u\|_p=e^{-\lambda t}e^{-(\mu/p)t}\|u\|_p.$$

Therefore we must have $\lambda\geqq-\mu/p$ and hence $\lambda_1\geqq-\mu/p$.

Example D ($\lambda_0>\Psi_1$). Let $B=\hat{C}(\mathbb{R})$, M be the collection positive measure with $\nu(\mathbb{R})=1$ and $\nu([-a,a]^c)=0$ for some a, and let $T(t)f(x)=f(x-t)$. Then $\lambda_0=0$ and $\Psi_1=-\infty$. To see that $\Psi_1=-1$, note that if $u(x)=e^{\lambda x}$ on the support of ν, then

$$\frac{\nu Au}{\nu u}=-\lambda.$$

Relationship to other work. In [1], with $A=L+V$, Donsker and Varadhan define

$$\Psi_1=\sup_{\mu\in M}\left[\int_X V(x)\mu(dx)+\inf_{u\in D(A)\cap K_0}\int\left(\frac{Lu}{u}\right)(x)\mu(dx)\right] \tag{14}$$

where M is the space of positive measures with $\mu(X)=1$. This is the same as

$$\begin{aligned}\Psi_1&=\sup_{\mu\in M}\inf_{u\in D(A)\cap K_0}\int_X[V(x)u(x)+Lu(x)]\frac{1}{u(x)}\mu(dx).\\ &=\sup_{\mu\in M}\inf_{u\in D(A)\cap K_0}\int_X\frac{Au(x)}{u(x)}\mu(dx).\end{aligned} \tag{15}$$

Similarly their definition of Ψ_2 is

$$\Psi_2 = \inf_{u \in D(A) \cap K_0} \sup_{\mu \in M} \int_X \frac{Au(x)}{u(x)} \mu(dx). \tag{16}$$

Taking

$$\nu(dx) = \frac{1}{u(x)} \mu(dx) \Big/ \int_X \frac{1}{u} \, d\mu \tag{17}$$

we have

$$\int_X \frac{Au(x)}{u(x)} \mu(dx) = \int_X Au(x) \nu(dx) \Big/ \int_X u(x) \nu(dx). \tag{18}$$

For fixed u, (17) defines a one-to-one mapping of M onto M and the sup over μ on the left of (18) is the same as the sup over ν on the right. It follows that the definition of Ψ_2 in (16) is the same as the previous definition.

Unfortunately there doesn't appear to be a simple way of showing that the two definitions of Ψ_1 are the same.

Finally, without a positivity assumption, Lumer and Phillips [4] have shown

$$\lim_{t \to \infty} \frac{1}{t} \log \|T(t)\| = \sup \{\nu A u \colon u \in D(A), \nu \in B^*, \|u\| = \|\nu\| = \nu u = 1\}.$$

REFERENCES

[1] MONROE D. DONSKER AND S. R. S. VARADHAN, *On a variational formula for the principle eigenvalue for operators with maximum principle*, Proc. Nat. Acad. Sci. U.S.A., 72 (1975), pp. 780–783.

[2] G. P. BARKER AND HANS SCHNEIDER, *Algebraic Perron–Frobenius theory*, Linear Algebra Appl., 11 (1975), pp. 219–233.

[3] SAMUEL KARLIN, *Positive operators*, J. Math. Mech., 8 (1959), pp. 907–937.

[4] G. LUMER AND R. S. PHILLIPS, *Dissapative operators in a Banach space*, Pacific J. Math., 11 (1961), pp. 679–698.

[5] IVO MAREK, *Frobenious theory of positive operators: Comparison theorems and applications*, SIAM J. Appl. Math., 19 (1970), pp. 607–628.

[6] HANS SCHNEIDER AND R. E. L. TURNER, *Positive eigenvectors of order-preserving maps*, J. Math. Anal., 37 (1972), pp. 506–515.

[7] MAURICE SION, *On general minimax theorems*, Pacific J. Math., 8 (1958), pp. 171–176.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0015 $01.00/0

FREQUENCY DOMAIN STABILITY FOR A CLASS OF EQUATIONS ARISING IN REACTOR DYNAMICS*

D. WEXLER†

Abstract. We establish Liapunov type stability properties for an evolution equation in a Hilbert space by using the Popov frequency domain method. Some systems arising in reactor dynamics may be viewed as specializations of the equation discussed in this paper.

1. Introduction. In recent years, an increasing interest has been taken in the system of integro-differential equations

$$\begin{aligned}&\frac{\partial}{\partial t}T(t,\xi)=\frac{\partial}{\partial \xi}\left[p_1(\xi)\frac{\partial}{\partial \xi}T(t,\xi)\right]+p_2(\xi)T(t,\xi)+\varphi(\sigma(t))a(\xi),\\ &\frac{d}{dt}\sigma(t)=\int_{\gamma_1}^{\gamma_2}b(\xi)T(t,\xi)\,d\xi,\quad \text{for all } t\in\,]0,+\infty[\quad\text{and almost all }\xi\text{ in }]\gamma_1,\gamma_2[,\end{aligned}\tag{1.1}$$

subject to boundary conditions

$$\begin{aligned}&\delta_1T(t,\gamma_1)+\delta_2\frac{\partial T}{\partial\xi}(t,\gamma_1)=0,\\ &\delta_3T(t,\gamma_2)+\delta_4\frac{\partial T}{\partial\xi}(t,\gamma_2)=0\quad\text{for all } t\in\,]0,+\infty[,\end{aligned}\tag{1.2}$$

and to initial conditions

$$\sigma(0)=\sigma_0,\qquad T(t,\cdot)\to T_0\quad\text{in } L^2(\gamma_1,\gamma_2)\quad\text{as } t\to 0,\tag{1.3}$$

where the real constants δ_j satisfy

$$|\delta_1|+|\delta_2|>0,\qquad |\delta_3|+|\delta_4|>0,\tag{1.4}$$

the nonlinear function $\varphi: \mathbb{R}\to\mathbb{R}$ is continuous with $r\varphi(r)>0$ for all $r\in\mathbb{R}$, $r\neq 0$ and a, b are elements of the space $L^2(\gamma_1,\gamma_2)$ of real-valued square-integrable (classes of) functions on $]\gamma_1,\gamma_2[$; when $]\gamma_1,\gamma_2[\,=\mathbb{R}$, the boundary conditions are replaced by appropriate L^2 conditions. Also, conditions on δ_j and on the real functions p_1, p_2 were required in order to insure the associated to (1.1), (1.2) Sturm–Liouville operator A,

$$Ax(\xi)=\frac{d}{d\xi}\left[p_1(\xi)\frac{d}{d\xi}x(\xi)\right]+p_2(\xi)x(\xi),\tag{1.5}$$

to be self-adjoint and negative in $L^2(\gamma_1,\gamma_2)$.

Systems of this type arise as dynamic models of one-dimensional continuous medium nuclear reactors and one is interested in the asymptotic behavior of the solutions as $t\to+\infty$, mainly in Liapunov type stability; the reader is referred to [7], [9] for the physical significance of the various parameters of the system.

This problem has been studied most notably by Levin and Nohel [9], [10], [11], [12], Nohel [16], Miller [13], [14]. By eliminating the unknown function T they obtained a scalar nonlinear Volterra integro-differential equation which was discussed by means of energy functions and/or transform methods.

* Received by the editors January 13, 1977, and in revised form July 13, 1977.
† Department of Mathematics, Facultés Universitaires N.D. de la Paix, Namur, Belgium.

Bronikowski, Hall and Nohel [2] used a Galerkin procedure to approximate the problem by a set of ordinary differential equations which is then discussed by Liapunov function method. The stability conditions in [2], [11] are expressed in terms of the Fourier coefficients of a and b with respect to the system of eigenfunctions of A.

The theory of C_0-semigroups is applied by Suhadolc [20] and Infante and Walker [7]. In [20], a linear variant of the above problem is considered with $]\gamma_1, \gamma_2[= \mathbb{R}$ and $A = d^2/d\xi^2$; the theory of analytical semigroups is used to obtain existence and regularity of the solutions and transform methods are applied to discuss stability. In [7], the abstract evolution problem

$$\frac{du}{dt} = Au + \varphi(\sigma)a, \qquad \frac{d\sigma}{dt} = \langle b, u\rangle \tag{1.6}$$

is considered in $L^2(\gamma_1, \gamma_2) \times \mathbb{R}$, where A is a negative self-adjoint operator in $L^2(\gamma_1, \gamma_2)$, $\langle \cdot, \cdot \rangle$ is the inner product in $L^2(\gamma_1, \gamma_2)$ and $a, b \in L^2(\gamma_1, \gamma_2)$. When A is the Sturm–Liouville operator (1.5), system (1.6) is an L^2-version of problem (1.1), (1.2). For system (1.6), Infante and Walker [7] have established stability conditions by means of the theory of nonlinear C_0-semigroups combined with some estimates obtained on the basis of a Liapunov function which is much similar to that used previously in the theory of absolute stability of differential equations in finite-dimensional spaces. Their approach applies also when the effect of delayed neutrons is included and the heat conduction is nonlinear [21].

Here we choose to consider the abstract evolution problem (1.6) under more general assumptions on A. It will be assumed that A is a linear operator in a real Hilbert space H and that A generates a differentiable exponentially stable C_0-semigroup on H (A is not necessarily self-adjoint, nor negative); the case in which 0 is a simple isolated eigenvalue of the complexification A^c of A and the case in which φ is linear will be discussed elsewhere. We establish stability conditions by applying Popov type frequency domain methods to the associated Volterra integral equation; our approach is an extension to the Hilbert setting of the approach used previously by Corduneanu for differential equations in finite-dimensional spaces [cf. 3, chap. 3]. Note that Levin and Nohel made mention previously [11] of the possibility to study in this way the Volterra integral equation associated with (1.1), (1.2), but only for some rather special cases.

In § 2, we give the precise description of our setting and in § 3, we establish the stability criteria; for shortness, we limit our discussion to asymptotic stability. We also establish some exponential estimates for the solutions and discuss the sensitivity of our stability conditions with respect to small perturbations in the parameters of the system.

As usual, when applying frequency domain methods to differential equations, the stability conditions we obtain are expressed in terms of positivity of a function involving a, b and the resolvent of A^c and they do not depend on the nonlinear function φ belonging to a specified class. In general, to check the frequency domain condition (condition (ii) of Theorem 1) requires the knowledge of the resolvent of A^c, which is far from being an easy matter. When A is defined by (1.5), it amounts to solving a Sturm–Liouville problem for an ordinary second order differential operator; in some special cases (say if p_1 and p_2 are constants), this problem may be effectively solved. On the other hand, for significant systems, some easy-to-check stability conditions which do not make explicit use of the resolvent are available [7]. In § 4, we show that the asymptotic stability result established previously in [7] is closely related to a specialization of Corollary 1.

For some significant integro-differential systems, the solutions of the associated abstract version are in fact classical ones and one may derive also stability results under the norm of the uniform convergence. This is illustrated in § 4 for the case of problem (1.1), (1.2). Furthermore, for problem (1.1), (1.2), we related the frequency domain stability conditions to the asymptotic stability conditions established previously in [2].

2. Setting of the abstract problem. In the sequel H is a real Hilbert space with inner product $\langle \cdot . \cdot \rangle$ and the norm $|\cdot|$, I is the identity operator on H and A is a linear operator with domain $D(A) \subset H$ and range $R(A) \subset H$. We consider the differential system

$$\frac{du}{dt} = Au + \varphi(\sigma)a, \qquad \frac{d\sigma}{dt} = \langle b, u \rangle, \tag{2.1}$$

where a and b are given elements in H and $\varphi: \mathbb{R} \to \mathbb{R}$ is a given (nonlinear) locally Lipschitz function. The above system will be viewed in the Hilbert space $\mathcal{H} = H \times \mathbb{R}$ with inner product

$$\langle (x_1, r_1), (x_2, r_2) \rangle_{\mathcal{H}} = \langle x_1, x_2 \rangle + r_1 r_2 \qquad (x_j, r_j) \in \mathcal{H}, \quad j = 1, 2.$$

In §§ 2 and 3, it is assumed that A generates a C_0-semigroup S on H, which satisfies the following conditions:

$$S \text{ is differentiable (i.e. } S(t)H \subset D(A) \text{ for all } t > 0) \tag{2.2}$$

and there exist $M \geqq 1$, $\alpha > 0$ such that

$$|S(t)|_{\mathcal{L}(H)} \leqq M e^{-\alpha t} \quad \text{for all } t \geqq 0, \tag{2.3}$$

where $\mathcal{L}(H)$ denotes the Banach space of bounded linear operators from H to H; for the theory of semigroups of linear operators, we refer the reader to [8, chap. IX] and [22, chap. IX]. Condition (2.2) implies that for each $x \in H$ the function $S(\cdot)x: \mathbb{R}^+ \to H$ is of class C^∞ on $]0, +\infty[$, hence by using the Banach–Steinhaus theorem and the Taylor formula, we may see that the operator-valued function $S: \mathbb{R}^+ \to \mathcal{L}(H)$ is also of class C^∞ on $]0, +\infty[$. Then, by using (2.3) and

$$\frac{dS}{dt}(t) = S(t-\theta)\frac{dS}{dt}(\theta) \quad \text{for all } t, \theta \in \mathbb{R}^+, \quad t \geqq \theta > 0,$$

it follows that there exists a continuous, decreasing function $f:]0, +\infty[\to \mathbb{R}^+$ with $f(t) \to 0$ as $t \to +\infty$, such that

$$\left|\frac{dS}{dt}(t)\right|_{\mathcal{L}(H)} = |AS(t)|_{\mathcal{L}(H)} \leqq f(t), \quad \text{for all } t > 0. \tag{2.4}$$

It is useful to consider also the complexification H^c of H; the elements of H^c will be written as $x + iy$, $x \in H$, $y \in H$ and the inner product of H^c will be denoted by $\langle \cdot, \cdot \rangle_{H^c}$. For any linear operator U in H, we denote by U^c the linear operator in H^c defined by

$$U^c(x + iy) = Ux + iUy \quad \text{with domain} \quad D(U^c) = D(U) + iD(U).$$

In particular, I^c is the identity operator on H^c. By the Hille–Yosida theorem, saying that A generates a C_0-semigroup which satisfies (2.3) is equivalent to the following: A

is densely defined, closed, the resolvent set $P(A^c)$ of A^c contains the half plane $\operatorname{Re}\lambda > -\alpha$ and for each $n = 1, 2, 3, \cdots$,

$$(2.5) \qquad |(\lambda I^c - A^c)^{-n}|_{\mathscr{L}(H^c)} \leqq M(\operatorname{Re}\lambda + \alpha)^{-n} \quad \text{for all } \lambda \in \mathbb{C}, \quad \operatorname{Re}\lambda > -\alpha.$$

We see in particular that A^{-1} is a bounded linear operator.

An important case in which our assumptions on A hold is the case in which for some $\alpha > 0$, the operator $A^c + \alpha I^c$ generates a bounded holomorphic semigroup. This condition is in turn satisfied when for some $\alpha > 0$ the operator $-(A^c + \alpha I^c)$ is m-sectorial with vertex 0 [8, pp. 490–491] (note that significant differential operators are m-sectorial [8, p. 280]). The latter condition holds in particular when A is self-adjoint and there exists $\alpha > 0$ such that $A + \alpha I$ is negative.

The function (u, σ) from the interval $[0, \theta]$ to $\mathscr{H}$ is said to be a solution of (2.1) on $[0, \theta]$ with initial data $(u_0, \sigma_0) \in \mathscr{H}$, if it satisfies the following conditions: (i) u is continuous on $[0, \theta]$, of class C^1 on $]0, \theta]$, $u(0) = u_0$ and

$$u(t) \in D(A), \qquad \frac{du}{dt}(t) = Au(t) + \varphi(\sigma(t))a, \quad \text{for all } t \in [0, \theta];$$

(ii) σ is of class C^1 on $[0, \theta]$, $\sigma(0) = \sigma_0$ and

$$\frac{d\sigma}{dt}(t) = \langle b, u(t)\rangle \quad \text{for all } t \in [0, \theta].$$

The function (u, σ) from $\mathbb{R}^+$ to $\mathscr{H}$ is said to be a solution of (2.1) on $\mathbb{R}^+$ with initial data $(u_0, \sigma_0) \in \mathscr{H}$, if it is for all $\theta > 0$ a solution of (2.1) on $[0, \theta]$ with initial data (u_0, σ_0).

Proposition 1 below reduces the initial data problem for (2.1) to the scalar nonlinear Volterra integral equation (2.6).

PROPOSITION 1. *For any $(u_0, \sigma_0) \in \mathscr{H}$ and any $\theta > 0$ there exists at most one solution of* (2.1) *on* $[0, \theta]$ *with initial data* (u_0, σ_0). *This solution is* (u, σ) *if and only if σ is a solution of the integral equation*

$$(2.6) \quad \sigma(t) = \sigma_0 + \langle b, S(t)A^{-1}u_0 - A^{-1}u_0\rangle + \int_0^t \varphi(\sigma(s))\langle b, S(t-s)A^{-1}a - A^{-1}a\rangle\, ds$$

continuous on $[0, \theta]$ *and for all* $t \in [0, \theta]$,

$$(2.7) \qquad u(t) = S(t)u_0 + \int_0^t \varphi(\sigma(s))S(t-s)a\, ds.$$

Proof. Assume (u, σ) is a solution of (2.1) on $[0, \theta]$ with initial data (u_0, σ_0). Then, for each $\tau \in]0, \theta[$, u is a solution on $[\tau, \theta]$ of the inhomogeneous problem

$$\frac{dw}{dt} = Aw + \varphi(\sigma(t))a, \qquad w(\tau) = u(\tau),$$

so that, according to [8, p. 486], we have

$$u(t) = S(t-\tau)u(\tau) + \int_\tau^t \varphi(\sigma(s))a\, ds \quad \text{for all } \tau, t \in]0, \theta], \quad \tau \leqq t.$$

By letting $\tau \to 0$, it follows that (2.7) holds on $[0, \theta]$. Then by integrating on $[0, t]$ the second equation of (2.1), we see that

$$\sigma(t) = \sigma_0 + \left\langle b, \int_0^t S(s)u_0\, ds\right\rangle + \left\langle b, \int_0^t d\tau \int_0^\tau \varphi(\sigma(s))S(\tau-s)a\, ds\right\rangle.$$

Since $A^{-1}u_0 \in D(A)$, the function $S(\cdot)A^{-1}u_0$ is a primitive of $S(\cdot)u_0$, so that

$$\int_0^t S(s)u_0\,ds = S(t)A^{-1}u_0 - A^{-1}u_0. \tag{2.8}$$

By applying Fubini's theorem and (2.8), we see that

$$\int_0^t d\tau \int_0^\tau \varphi(\sigma(s))S(\tau - s)a\,ds = \int_0^t \varphi(\sigma(s))[S(t-s)A^{-1}a - A^{-1}a]\,ds,$$

whence σ satisfies (2.6) on $[0, \theta]$.

Conversely, assume that u and σ satisfy (2.6) and (2.7) respectively on $[0, \theta]$. Clearly, $u(0) = u_0$ and $\sigma(0) = \sigma_0$. Since $A^{-1}u_0$ and $A^{-1}a$ belong to $D(A)$, the functions $S(\cdot)A^{-1}u_0$ and $S(\cdot)A^{-1}a$ are of class C^1. Then we see easily that σ is also of class C^1 and satisfies on $[0, \theta]$ the second equation of (2.1). Since S is of class C^∞ on $]0, +\infty[$ and $S(t)u_0 \in D(A)$ for all $t > 0$, to prove that u is of class C^1 on $]0, \theta]$ and satisfies the first equation in (2.1) on $]0, \theta]$, it suffices to verify that this holds true for the function v defined on $[0, \theta]$ by

$$v(t) = \int_0^t \varphi(\sigma(s))S(t-s)a\,ds.$$

Indeed, for any $x \in H$ define the function v_x on $[0, \theta]$ by

$$v_x(t) = \int_0^t \varphi(\sigma(s))S(t-s)x\,ds = \int_0^t \varphi(\sigma(t-s))S(s)x\,ds;$$

clearly, $v_a = v$. By using the fact that $\varphi \circ \sigma$ is Lipschitz, we see easily that $v_x \in C^1([0, \theta])$ and

$$\frac{dv_x}{dt}(t) = \varphi(\sigma_0)S(t)x + \int_0^t \frac{d(\varphi \circ \sigma)}{dt}(s)S(t-s)x\,ds \tag{2.9}$$

for all $t \in [0, \theta]$, $x \in H$. If $x \in D(A)$, we may use [6, p. 153] to see that

$$v_x(t) \in D(A) \quad \text{and} \quad \int_0^t \varphi(\sigma(s))AS(t-s)x\,ds = Av_x(t),$$

and then an integration by parts yields

$$\int_0^t \frac{d(\varphi \circ \sigma)}{dt}(s)S(t-s)x\,ds = \varphi(\sigma(t))x - \varphi(\sigma_0)S(t)x + Av_x(t). \tag{2.10}$$

Since A is closed and $D(A)$ is dense in H, it follows that for all $t \in [0, \theta]$, $v(t) \in D(A)$ and (2.10) with $x = a$ holds true. Then, (2.9) and (2.10) with $x = a$ imply

$$\frac{dv}{dt}(t) = Av(t) + \varphi(\sigma(t))a \quad \text{for all } t \in [0, \theta]. \tag{2.11}$$

To prove uniqueness, note that if (u, σ) and $(\hat{u}, \hat{\sigma})$ are two solutions of (2.1) on $[0, \theta]$ with the same initial data (u_0, σ_0), (2.6) implies

$$\sigma(t) - \hat{\sigma}(t) = \int_0^t [\varphi(\sigma(s)) - \varphi(\hat{\sigma}(s))]\langle b, S(t-s)A^{-1}a - A^{-1}a\rangle\,ds.$$

Since φ is locally Lipschitz and σ, $\hat{\sigma}$ are bounded, we may then apply the Gronwall–Bellman lemma to see that $\sigma = \hat{\sigma}$, hence by (2.7), $u = \hat{u}$. The proof of Proposition 1 is complete.

Remark 1. By using (2.10) with $x=a$ and (2.11), we see that if (u, σ) is a solution of (2.1) on $[0, \theta]$ with initial data (u_0, σ_0) then the function $\varphi\circ\sigma$ is Lipschitz on $[0, \theta]$ and

$$\frac{du}{dt}(t)=AS(t)u_0+\varphi(\sigma_0)S(t)a+\int_0^t \frac{d(\varphi\circ\sigma)}{dt}(s)S(t-s)a\,ds$$

for all $t\in\,]0, \theta]$. In addition to this, when $u_0\in D(A)$, the function u is of class C^1 on $[0, \theta]$ and

$$\frac{du}{dt}(0)=Au_0+\varphi(\sigma_0)a.$$

We shall make use of the following "local semigroup property":

PROPOSITION 2. *If (u_1, σ_1) is a solution of* (2.1) *on $[0, \theta_1]$ with initial data (u_0, σ_0) and if $(\hat{u}, \hat{\sigma})$ is a solution of* (2.1) *on $[0, \theta_2]$ with initial data $(u_1(\theta_1), \sigma_1(\theta_1))$ then the function (u, σ) defined on $[0, \theta_1+\theta_2]$ by*

$$u(t)=u_1(t)\quad \text{if } t\in[0, \theta_1],\qquad u(t)=\hat{u}(t-\theta_1)\quad \text{if } t\in\,]\theta_1, \theta_1+\theta_2],$$

$$\sigma(t)=\sigma_1(t)\quad \text{if } t\in[0, \theta_1],\qquad \sigma(t)=\hat{\sigma}(t-\theta_1)\quad \text{if } t\in\,]\theta_1, \theta_1+\theta_2]$$

is a solution of (2.1) *on $[0, \theta_1+\theta_2]$ with initial data (u_0, σ_0).*

Proof. It suffices to observe that, since $u_1(\theta_1)\in D(A)$ the function $\hat{u}$ is differentiable at 0 and

$$\frac{du_1}{dt}(\theta_1)=\frac{d\hat{u}}{dt}(0)=Au(\theta_1)+\varphi(\sigma(\theta_1))a.$$

3. Stability. Assume $\varphi(0)=0$, so that system (2.1) admits the zero solution. The zero solution of (2.1) is said to be stable in the large if: (i) for each $(u_0, \sigma_0)\in\mathcal{H}$ there exists a solution of (2.1) on $\mathbb{R}^+$ with initial data (u_0, σ_0) (uniqueness is insured by Proposition 1); and (ii) there exists a continuous strictly increasing function Π: $\mathbb{R}^+\to\mathbb{R}^+$ with $\Pi(0)=0$ such that, for any solution (u, σ) with initial data (u_0, σ_0) and any $r>0$,

$$|(u_0, \sigma_0)|_{\mathcal{H}}\leqq r \quad\text{implies}\quad |(u(t), \sigma(t))|_{\mathcal{H}}\leqq \Pi(r)\quad \text{for all } t\geqq 0.$$

The zero solution is said to be uniformly asymptotically stable in the large if it is stable in the large and if, for any bounded set $\mathcal{B}$ in $\mathcal{H}$, the solution (u, σ) of (2.1) with initial data (u_0, σ_0) tends to 0 as $t\to+\infty$, uniformly with respect to $(u_0, \sigma_0)\in\mathcal{B}$.

We now state our main result.

THEOREM 1. *Assume the following conditions hold*:

(i) $r\varphi(r)>0$ *for all* $r\in\mathbb{R}$, $r\neq 0$ (*so that* $\varphi(0)=0$);

(ii) $\langle b, A^{-1}a\rangle>0$ *and there exists* $q\geqq 0$ *such that*

$$\operatorname{Re}(1-isq)\langle b, (isI^c-A^c)^{-1}A^{-1}a\rangle_{H^c}-q\langle b, A^{-1}a\rangle\leqq 0\quad \textit{for all } s\geqq 0.$$

Then, the zero solution of (2.1) *is uniformly asymptotically stable in the large.*

Proof. Our proof is in two steps, but let us first introduce some functions we use in the sequel. Since φ is locally Lipschitz, there exists a continuous strictly increasing function ℓ: $\mathbb{R}^+\to\mathbb{R}^+$ such that for each $r\geqq 0$,

$$|\varphi(r_1)-\varphi(r_2)|\leqq \ell(r)|r_1-r_2|\quad \text{for all } r_1, r_2\in[-r, r].$$

Define ν_1: $\mathbb{R}^+\to\mathbb{R}^+$ by

$$\nu_1(r)=2r(1+(M+1)|b||A^{-1}|_{\mathcal{L}(H)}).$$

Next, for any $r>0$, denote by θ the unique solution of

$$\theta+\int_0^\theta |\langle b, S(s)A^{-1}a-A^{-1}a\rangle|\,ds=(2\ell(\nu_1(r)))^{-1}$$

and put $\tau(r)=\inf(\theta, 1)$, $\tau(0)=\lim_{r\to 0}\tau(r)$, so that the function $\tau\colon \mathbb{R}^+\to]0, 1]$ is continuous and decreasing.

Step I. We prove that the zero solution of (2.1) is stable in the large and all the solutions of (2.1) tend to zero as $t\to+\infty$.

Fix an arbitrary $r>0$ and choose an arbitrary $(u_0, \sigma_0)\in\mathscr{H}$ with

$$|(u_0, \sigma_0)|_{\mathscr{H}}\leqq r. \tag{3.1}$$

Denote by $\mathscr{C}$ the subspace of $C([0, \tau(r)])$ consisting of functions σ which satisfy $\sigma(0)=\sigma_0$ and $|\sigma|\leqq\nu_1(r)$ on $[0, \tau(r)]$. Clearly, $\mathscr{C}$ is a complete metric space. By using (2.3) and the choice of ℓ, ν_1 and τ, we see easily that the operator $\mathscr{U}$ defined by

$$\mathscr{U}\sigma(t)=\sigma_0+\langle b, S(t)A^{-1}u_0-A^{-1}u_0\rangle+\int_0^t\varphi(\sigma(s))\langle b, S(t-s)A^{-1}a-A^{-1}a\rangle\,ds$$

is a strict contraction in $\mathscr{C}$. It follows that the integral equation (2.6) possesses a unique solution σ_1 on $[0, \tau(r)]$ and $|\sigma_1|\leqq\nu_1(r)$ on $[0, \tau(r)]$. Let u_1 be the function defined on $[0, \tau(r)]$ by (2.7), so that $|u_1|\leqq\mu_1(r)$ on $[0, \tau(r)]$, where $\mu_1\colon\mathbb{R}^+\to\mathbb{R}^+$ is defined by

$$\mu_1(\theta)=M(\theta+\alpha^{-1}|a|\nu_1(\theta)\ell(\nu_1(\theta))).$$

According to Proposition 1, (u_1, σ_1) is a solution of (2.1) on $[0, \tau(r)]$ with initial data (u_0, σ_0). It is of importance to the remainder of the proof to note that the interval of existence of this solution and the estimates

$$|u_1|\leqq\mu_1(r),\qquad |\sigma_1|\leqq\nu_1(r)\quad\text{on}\quad [0, \tau(r)] \tag{3.2}$$

do not depend on the initial data satisfying (3.1). Note also that μ_1 and ν_1 are continuous, strictly increasing and $\mu_1(0)=\nu_1(0)=0$.

Put $u_1(\tau(r))=\hat{u}_0$ and $\sigma_1(\tau(r))=\hat{\sigma}_0$. By applying the local semi-group property, we see that, if $(\hat{u}, \hat{\sigma})$ is a solution of (2.1) on $\mathbb{R}^+$ with initial data $(\hat{u}_0, \hat{\sigma}_0)$, the function (u, σ) defined on $\mathbb{R}^+$ by

$$\begin{aligned} &u(t)=u_1(t)\quad \textit{if } t\in[0, \tau(r)], &&u(t)=\hat{u}(t-\tau(r))\quad \textit{if } t>\tau(r),\\ &\sigma(t)=\sigma_1(t)\quad \textit{if } t\in[0, \tau(r)], &&\sigma(t)=\hat{\sigma}(t-\tau(r))\quad \textit{if } t>\tau(r),\end{aligned} \tag{3.3}$$

is a solution of (2.1) on $\mathbb{R}^+$ with initial data (u_0, σ_0). According to Proposition 1, it follows then that, in order to prove existence of a solution (u, σ) of (2.1) on $\mathbb{R}^+$ with initial data (u_0, σ_0), we have to prove existence of a continuous on $\mathbb{R}^+$ solution $\hat{\sigma}$ of the integral equation

$$\hat{\sigma}(t)=\hat{\sigma}_0+\langle b, S(t)A^{-1}\hat{u}_0-A^{-1}\hat{u}_0\rangle+\int_0^t\varphi(\hat{\sigma}(s))\langle b, S(t-s)A^{-1}a-A^{-1}a\rangle\,ds. \tag{3.4}$$

The main advantage in considering (3.4) instead of (2.6) consists in the fact that, since $\hat{u}_0\in D(A)$, the function h defined on $\mathbb{R}^+$ by

$$h(t)=\hat{\sigma}_0+\langle b, S(t)A^{-1}\hat{u}_0-A^{-1}\hat{u}_0\rangle$$

belongs to $C^2(\mathbb{R}^+)$ and by (2.3), the derivatives dh/dt, d^2h/dt^2 belong to $L^1(\mathbb{R}^+)$, while, if $u_0\notin D(A)$, the second derivative of the similar to h term in (2.6) may have a nonintegrable singularity at 0; these properties of h will allow to apply to (3.4) the results of [3, pp. 90–95].

Define the function k_0 on $\mathbb{R}^+$ by $k_0(t)=\langle b, S(t)A^{-1}a\rangle$ and put $\rho=\langle b, A^{-1}a\rangle$, $k=k_0-\rho$, so that (3.4) is written as

$$\hat{\sigma}(t)=h(t)+\int_0^t \varphi(\hat{\sigma}(s))k(t-s)\,ds. \tag{3.5}$$

Clearly, h, ρ, k and φ verify conditions 1, 2 and 3 of [3, Thm. 3.1, p. 91]. Since the resolvent of A^c is the Laplace transform of S^c [8, p. 482], we may write the Fourier transform $\tilde{k}_0$ of k_0 as

$$\tilde{k}_0(s)=\int_0^{+\infty} e^{ist}\langle b, S(t)A^{-1}a\rangle_{H^c}=\langle b, (isI^c-A^c)^{-1}A^{-1}a\rangle_{H^c},$$

so that, by (ii) all of the conditions in [3, Th. 3.1, p. 91] hold. It follows that (3.5) admits a continuous on $\mathbb{R}^+$ solution $\hat{\sigma}$ and $\hat{\sigma}(t)\to 0$ as $t\to+\infty$. Since $\varphi(\hat{\sigma}(\cdot))$ tends also to 0 and S satisfies (2.3), the function $\hat{u}$ defined on $\mathbb{R}^+$ by (2.7) with σ replaced by $\hat{\sigma}$ and u_0 by $\hat{u}_0$ tends to 0 as $t\to+\infty$. We see so that for any $(u_0, \sigma_0)\in\mathcal{H}$, there exists a (unique) solution (u, σ) of (2.1) on $\mathbb{R}^+$ with initial data (u_0, σ_0) and $(u(t), \sigma(t))\to 0$ as $t\to+\infty$.

To prove stability in the large, we establish estimates on u and σ which are expressed in terms of r and the parameters of the system (2.1), but do not depend on the initial data satisfying (3.1). An integration by parts in (3.5) yields

$$\hat{\sigma}(t)=h(t)+\int_0^t \Phi(s)\frac{dk_0}{dt}(t-s)\,ds,$$

where

$$\Phi(t)=\int_0^t \varphi(\hat{\sigma}(s))\,ds,$$

so that, by using (2.3), we obtain

$$|\hat{\sigma}(t)|\leqq|h(t)|+\alpha^{-1}M|a||b|\sup_{s\geqq 0}|\Phi(s)|, \quad \text{for all } t\geqq 0. \tag{3.6}$$

By using (2.3) and (3.2), we see that

$$|h(t)|\leqq \nu_2(r) \quad \text{for all } t\geqq 0, \tag{3.7}$$

where the function $\nu_2\colon \mathbb{R}^+\to\mathbb{R}^+$, $\nu_2(\theta)=\nu_1(\theta)+(M+1)|b|\mu_1(\theta)|A^{-1}|_{\mathcal{L}(H)}$, is continuous, strictly increasing and $\nu_2(0)=0$. According to [3, (3.22), p. 93]

$$|\Phi(t)|\leqq\psi(F(h(0))) \quad \text{for all } t\geqq 0,$$

where F is defined on $\mathbb{R}$ by

$$F(\theta)=\int_0^\theta \varphi(s)\,ds \tag{3.8}$$

and ψ is defined on $\mathbb{R}^+$ by

$$\psi(\theta)=K\rho^{-1}+(K^2\rho^{-2}+2q\rho^{-1}\theta)^{1/2}, \tag{3.9}$$

with

$$K=\|h\|_{L^\infty(\mathbb{R}^+)}+q\left\|\frac{dh}{dt}\right\|_{L^\infty(\mathbb{R}^+)}+\left\|\frac{dh}{dt}\right\|_{L^1(\mathbb{R}^+)}+q\left\|\frac{d^2h}{dt^2}\right\|_{L^1(\mathbb{R}^+)};$$

clearly, $F(\theta) \geqq 0$. Since φ is continuous and satisfies (i), the function

$$\varphi_1 \colon \mathbb{R}^+ \to \mathbb{R}^+, \qquad \varphi_1(\theta) = \nu_1(\theta) \sup \{|\varphi(s)| : |s| \leqq \nu_1(\theta)\},$$

is continuous, strictly increasing and $\varphi_1(0) = 0$. Clearly, $F(h(0)) \leqq \varphi_1(r)$ and then, since ψ is increasing, we have

$$|\Phi(t)| \leqq \psi(\varphi_1(r)) \quad \text{for all } t \geqq 0. \tag{3.10}$$

Let us estimate K. By using (2.3), we see easily that

$$\left\| \frac{d^2 h}{dt^2} \right\|_{L^1(\mathbb{R}^+)} \leqq \alpha^{-1} M |b| |A\hat{u}_0|.$$

By using the first equation of (2.1), Remark 1 and $\hat{u}_0 = u_1(\tau(r))$, we see that

$$A\hat{u}_0 = AS(\tau(r))u_0 + \varphi(\sigma_0)S(\tau(r))a - \varphi(\sigma_1(\tau(r)))a + \int_0^{\tau(r)} \frac{d(\varphi \circ \sigma_1)}{dt}(s) S(\tau(r) - s) a \, ds.$$

Use (3.2), $d\sigma_1/dt = \langle b, u_1 \rangle$ and the Lipschitz property of $\varphi \circ \sigma_1$ to see that

$$\left| \frac{d(\varphi \circ \sigma_1)}{dt}(s) \right| \leqq |b| \ell(\nu_1(r)) \mu_1(r) \quad \text{a.e. in } [0, \tau(r)].$$

Then, by (2.3), (2.4), (3.2), $0 < \tau(r) \leqq 1$ and the Lipschitz property of φ it follows that

$$\left\| \frac{d^2 h}{dt^2} \right\|_{L^1(\mathbb{R}^+)} \leqq \nu_3(r), \tag{3.11}$$

where $\nu_3 \colon \mathbb{R}^+ \to \mathbb{R}^+$ is defined by

$$\nu_3(\theta) = \alpha^{-1} M |b| \{[f(\tau(\theta)) + M|a|\ell(\theta)]\theta + |a|\ell(\nu_1(\theta))[\nu_1(\theta) + M|b|\mu_1(\theta)]\};$$

note that, by the properties of the functions involved in the expression of ν_3 and since, by (ii), we have $a \neq 0$, $b \neq 0$, it follows that ν_3 is also continuous, strictly increasing and $\nu_3(0) = 0$. By (2.3) and (3.2)

$$\left\| \frac{dh}{dt} \right\|_{L^\infty(\mathbb{R}^+)} \leqq M|b|\mu_1(r) \quad \text{and} \quad \left\| \frac{dh}{dt} \right\|_{L^1(\mathbb{R}^+)} \leqq \alpha^{-1} M |b| \mu_1(r)$$

so that, by taking into account (3.7) and (3.11), we have

$$K \leqq \nu_4(r), \tag{3.12}$$

where $\nu_4 \colon \mathbb{R}^+ \to \mathbb{R}^+$ is defined by

$$\nu_4(\theta) = \nu_2(\theta) + (q + \alpha^{-1}) M |b| \mu_1(\theta) + q\nu_3(\theta).$$

By collecting (3.9), (3.10) and (3.12), we see that $|\Phi(t)| \leqq \nu_5(r)$ for all $t \geqq 0$, where $\nu_5 \colon \mathbb{R}^+ \to \mathbb{R}^+$ is defined by

$$\nu_5(\theta) = \nu_4(\theta)\rho^{-1} + [\nu_4^2(\theta)\rho^{-2} + 2q\rho^{-1}\varphi_1(\theta)]^{1/2}.$$

Then, by (3.6) and (3.7), it follows that $|\hat{\sigma}(t)| \leqq \nu_6(r)$ for all $t \geqq 0$ with $\nu_6 \colon \mathbb{R}^+ \to \mathbb{R}^+$ defined by

$$\nu_6(\theta) = \nu_2(\theta) + \alpha^{-1} M |a| |b| \nu_5(\theta)$$

and so, by (3.2) and (3.3), we deduce

$$|\sigma(t)| \leqq \nu_6(r) \quad \text{for all } t \geqq 0.$$

Note that ν_6 is also strictly increasing, continuous and $\nu_6(0)=0$. By using (2.3), (2.7) and the Lipschitz property of φ, we see then that

$$|u(t)| \leqq \mu_2(r) \quad \text{for all } t \geqq 0,$$

with μ_2: $\mathbb{R}^+ \to \mathbb{R}^+$ defined by

$$\mu_2(\theta) = M(\theta + \alpha^{-1}\nu_6(\theta)\ell(\nu_6(\theta))).$$

So, for any $(u_0, \sigma_0) \in \mathscr{H}$ satisfying (3.1), the solution (u, σ) of (2.1) on $\mathbb{R}^+$ with initial data (u_0, σ_0) satisfies

$$|(u(t), \sigma(t))|_{\mathscr{H}} \leqq \Pi(r),$$

where $\Pi = (\mu_2^2 + \nu_6^2)^{1/2}$. Clearly, Π is continuous, strictly increasing and $\Pi(0)=0$. Since $r>0$ is arbitrary, we deduce that the zero solution of (2.1) is stable in the large.

Step II. To complete the proof of Theorem 1, we establish that for each bounded set $\mathscr{B}$ in $\mathscr{H}$ and each $\varepsilon>0$, there exists $T \geqq 0$ such that any solution (u, σ) of (2.1) on $\mathbb{R}^+$ with initial data $(u_0, \sigma_0) \in \mathscr{B}$ satisfies $|(u(t), \sigma(t))|_{\mathscr{H}} \leqq \varepsilon$ for all $t \geqq T$.

Suppose the above claim does not hold. There exist then a bounded set $\mathscr{B}$ in $\mathscr{H}$, $\varepsilon>0$, an increasing sequence (θ_n) of positive numbers with $\theta_n \to +\infty$ and a sequence $((u^n, \sigma^n))$ of solutions on $\mathbb{R}^+$ with initial data $(u_0^n, \sigma_0^n) \in \mathscr{B}$ such that

$$|(u^n(\theta_n), \sigma^n(\theta_n))|_{\mathscr{H}} > \varepsilon \quad \text{for all } n \in \mathbb{N}. \tag{3.13}$$

Since (u_0^n, σ_0^n) belongs to the bounded set $\mathscr{B}$, the global stability insures that the sequences of functions (u^n) and (σ^n) are uniformly bounded on $\mathbb{R}^+$ and then, by $d\sigma^n/dt = \langle b, u^n \rangle$ it follows that the sequence $(d\sigma^n/dt)$ is also uniformly bounded on $\mathbb{R}^+$. Apply then the Ascoli–Arzela theorem to the sequence of real functions (σ^n) to see that it admits a subsequence which converges uniformly on the compact sets in $\mathbb{R}^+$ to a continuous function $\hat{\sigma}$. Clearly, we may assume that the initial sequence (σ^n) possesses this property. Now the sequence $((u_0^n, \sigma_0^n))$ is bounded in $\mathscr{H}$, so that it admits a subsequence which converges weakly to an element $(\hat{u}_0, \hat{\sigma}_0) \in \mathscr{H}$. Clearly, we may assume $(u_0^n, \sigma_0^n) \to (\hat{u}_0, \hat{\sigma}_0)$ weakly in $\mathscr{H}$, so that $u_0^n \to \hat{u}_0$ weakly in H and $\sigma_0^n \to \hat{\sigma}_0$ in $\mathbb{R}$.

According to Proposition 1, we have, for all $t \geqq 0$ and all $n \in \mathbb{N}$

$$\sigma^n(t) = \sigma_0^n + \langle b, S(t)A^{-1}u_0^n - A^{-1}u_0^n \rangle + \int_0^t \varphi(\sigma^n(s))\langle b, S(t-s)A^{-1}a - A^{-1}a \rangle \, ds.$$

We let $n \to \infty$. Since we have

$$\langle b, S(t)A^{-1}u_0^n - A^{-1}u_0^n \rangle = \langle (A^{-1})^* S^*(t)b - (A^{-1})^* b, u_0^n \rangle$$

(the symbol * stands here for the adjoint operator), $u_0^n \to \hat{u}_0$ weakly in H, $\sigma_0^n \to \hat{\sigma}_0$ in $\mathbb{R}$, $\sigma^n \to \hat{\sigma}$ and $\varphi(\sigma^n(\cdot)) \to \varphi(\hat{\sigma}(\cdot))$ uniformly on $[0, t]$, it follows that $\hat{\sigma}$ is a solution of the integral equation (2.6), continuous on $\mathbb{R}^+$, with σ_0 replaced by $\hat{\sigma}_0$ and u_0 by $\hat{u}_0$. Define $\hat{u}$ on $\mathbb{R}^+$ by (2.7) with u_0 replaced by $\hat{u}_0$ and σ by $\hat{\sigma}$, so that, by Proposition 1, $(\hat{u}, \hat{\sigma})$ is the solution of (2.1) on $\mathbb{R}^+$ with initial data $(\hat{u}_0, \hat{\sigma})$. Now we have seen in Step I that all of the solutions tend to 0 as $t \to +\infty$, hence there exists $T' \geqq 0$ such that

$$|(\hat{u}(t), \hat{\sigma}(t))|_{\mathscr{H}} \leqq 2^{-1}\Pi^{-1}(\varepsilon) \quad \text{for all } t \geqq T', \tag{3.14}$$

where Π^{-1} is the reciprocal of the continuous strictly increasing function Π constructed in Step I.

Since $\sigma^n \to \hat{\sigma}$ uniformly on compact sets in $\mathbb{R}^+$, it follows by (3.14) that, for any $t \geqq T'$, there exists $N'_t \in \mathbb{N}$ such that

$$(3.15) \qquad |\sigma^n(t)| \leqq 2^{-1/2}\Pi^{-1}(\varepsilon) \quad \text{for all } n \geqq N'_t.$$

Since the sequence $((u_0^n, \sigma_0^n))$ is bounded, there exists $r > 0$ such that $|(\hat{u}_0, \hat{\sigma}_0)| \leqq r$ and $|(u_0^n, \sigma_0^n)| \leqq r$, so that the global stability implies

$$(3.16) \quad |(\hat{u}(t), \hat{\sigma}(t))|_{\mathcal{H}} \leqq \Pi(r) \quad \text{and} \quad |(u^n(t), \sigma^n(t))|_{\mathcal{H}} \leqq \Pi(r) \quad \text{for all } t \geqq 0 \text{ and all } n \in \mathbb{N}.$$

Since

$$\hat{u}(t) = S(t)\hat{u}_0 + \int_0^t \varphi(\hat{\sigma}(s))S(t-s)a\,ds$$

and

$$u^n(t) = S(t)u_0^n + \int_0^t \varphi(\sigma^n(s))S(t-s)a\,ds,$$

we may use the Lipschitz property of φ, (2.3) and (3.16) to see that

$$|u^n(t) - \hat{u}(t)| \leqq 2Mre^{-\alpha t} + \alpha^{-1}|a|M\ell(\Pi(r)) \sup_{s \in [0,t]} |\sigma^n(s) - \hat{\sigma}(s)|,$$

for all $t \geqq 0$ and all $n \in \mathbb{N}$. Combine the above estimate, (3.14) and the fact that $\sigma^n \to \hat{\sigma}$ uniformly on compact sets in $\mathbb{R}^+$ to deduce that there exist $T \geqq T'$ and $N \in \mathbb{N}$, $N \geqq N'_{T'}$, such that

$$|u^n(T)| \leqq 2^{-1/2}\Pi^{-1}(\varepsilon) \quad \text{for all } n \geqq N.$$

Then, by taking into account (3.15), we see that

$$|(u^n(T), \sigma^n(T))|_{\mathcal{H}} \leqq \Pi^{-1}(\varepsilon) \quad \text{for all } n \geqq N,$$

whence global stability combined with the semigroup property implies

$$|(u^n(t), \sigma^n(t))|_{\mathcal{H}} \leqq \varepsilon \quad \text{for all } n \geqq N \quad \text{and all } t \geqq T.$$

But the above estimate contradicts (3.13) for large n. The proof of Theorem 1 is complete.

Remark 2. Theorem 1 allows us to see that the derivative of a solution (u, σ) with initial data (u_0, σ_0) tends to zero as $t \to +\infty$, uniformly with respect to (u_0, σ_0) in a fixed bounded set $\mathcal{B} \subset \mathcal{H}$. Indeed, since $d\sigma/dt = \langle b, u\rangle$, one has $d\sigma(t)/dt \to 0$ as $t \to +\infty$, uniformly with respect to $(u_0, \sigma_0) \in \mathcal{B}$. Next, the Lipschitz property of φ and stability in the large imply

$$\left|\frac{d(\varphi \circ \sigma)}{dt}(t)\right| \leqq \ell(\Pi(r)) \left|\frac{d\sigma}{dt}(t)\right| \quad \text{a.e. in } \mathbb{R}^+,$$

where $r = \sup\{|x| : x \in \mathcal{B}\}$. Use then the above estimate, Remark 1, (2.3), (2.4) and $d\sigma(t)/dt \to 0$ as $t \to +\infty$, uniformly with respect to $(u_0, \sigma_0) \in \mathcal{B}$, to deduce that $du(t)/dt \to 0$ as $t \to +\infty$ uniformly with respect to $(u_0, \sigma_0) \in \mathcal{B}$. Since $d^2\sigma/dt^2 = \langle b, du/dt\rangle$, it follows also that $d^2\sigma(t)/dt^2 \to 0$ as $t \to +\infty$, uniformly with respect to $(u_0, \sigma_0) \in \mathcal{B}$.

Remark 3. One may express the frequency domain condition in terms of A and A^2 instead of A^c and, in this form, it is sometimes easier to check. Note first that, since the resolvent set of A contains the imaginary axis, one has for all $s \geqq 0$ and $x \in H$

$$(sI + A^2)^{-1} \in \mathcal{L}(H) \quad \text{and} \quad (isI^c - A^c)^{-1}x = -A(s^2I + A^2)^{-1}x - is(s^2I + A^2)^{-1}x.$$

Then we see by standard calculations that condition (ii) of Theorem 1 may be written:

$$\text{(3.17)}\qquad \begin{aligned}&\langle b, A^{-1}a\rangle > 0 \quad \text{and there exists } q \geqq 0 \text{ such that}\\ &\langle b, (I+qA)(sI+A^2)^{-1}a\rangle \geqq 0 \quad \text{for all } s \geqq 0.\end{aligned}$$

Note that conditions (3.17) hold in the case in which for some $\alpha > 0$ and $k < 0$ the operator $-(A^c + \alpha I^c)$ is m-sectorial with vertex 0 and $b = ka \neq 0$; this case, with A a differential operator, is significant in reactor dynamics. Indeed, in this case

$$\langle b, A^{-1}a\rangle = k\langle AA^{-1}a, A^{-1}a\rangle > 0$$

(use the fact that $\langle Ax, x\rangle \leqq -\alpha|x|^2$ for all $x \in D(A)$). Next, put $(sI+A^2)^{-1}a = a_s$, so that for all $s \geqq 0$ and all $q > 0$ we have

$$\langle b, (I+qA)(sI+A^2)^{-1}a\rangle = qk\{(s+q^{-2})\langle (A+q^{-1}I)a_s, a_s\rangle \\ + \langle (A-q^{-1}I)(A+q^{-1}I)a_s, (A+q^{-1}I)a_s\rangle\}.$$

It follows that (3.17) holds with any $q \geqq \alpha^{-1}$ (use the fact that $A - q^{-1}I$ is negative for all $q > 0$ and $A + q^{-1}I$ is negative for all $q \geqq \alpha^{-1}$).

One may obtain stability results under weaker regularity assumptions on φ, provided that A satisfies some stronger conditions. Let us mention the following one.

Remark 4. Assume that φ is continuous, A is selfadjoint and for some $\alpha > 0$, $A + \alpha I$ is negative. Note that then $-A$ is the subdifferential of a function which is proper, convex and lower semicontinuous on H [1, p. II–29]. Modify the solution concept as follows. The function (u, σ) from $[0, \theta]$ to $\mathscr{H}$ is said to be a solution of (2.1) on $[0, \theta]$ with initial data (u_0, σ_0) if it satisfies the conditions: (i) u is continuous on $[0, \theta]$, absolutely continuous on any compact interval in $]0, \theta[$, $u(0) = u_0$ and

$$u(t) \in D(A), \qquad \frac{du}{dt}(t) = Au(t) + \varphi(\sigma(t))a, \quad \text{a.e. in }]0, \theta[;$$

(ii) σ is of class C^1 on $[0, \theta]$, $\sigma(0) = \sigma_0$ and

$$\frac{d\sigma}{dt}(t) = \langle b, u(t)\rangle \quad \text{on } [0, \theta].$$

It is clear that Proposition 2 still holds true and, by using [1, p. III–20], we may see that Proposition 1 holds also true but without uniqueness. Assume further that conditions (i) and (ii) of Theorem 1 are satisfied. We may deduce local existence of the solutions by using, for instance, [15, p. 36]. Then proceed as in Step I of the proof of Theorem 1 to obtain the integral equation (3.4) with $\hat{u}_0 \in D(A)$. By applying again [3, p. 91], we see that for each $(u_0, \sigma_0) \in \mathscr{H}$ there exists at least one solution of (2.1) on $\mathbb{R}^+$ with initial data (u_0, σ_0) and, in addition to this, any solution of (2.1) on $\mathbb{R}^+$ tends to zero as $t \to +\infty$.

Remark 5. One may obtain results similar to those of Theorem 1 by invoking instead of [3, Thm. 3.1] the frequency domain criteria for Volterra integral or integro-differential equations established recently in [17], [19]. In this way it is possible to drop the regularity condition (2.2) provided that one uses a weaker solution concept and one requires some additional growth conditions on φ.

Exponential estimates may be obtained by linearisation when φ satisfies some additional conditions.

COROLLARY 1. *Assume the conditions of Theorem* 1 *hold and denote by* $\mathscr{T}(\cdot)X_0$ *the solution of* (2.1) *on* $\mathbb{R}^+$ *with initial data* $X_0 = (u_0, \sigma_0)$; *assume moreover that* φ *is* *differentiable at* 0 *with* $(d\varphi/dr)(0) > 0$. *Then there exists* $\omega > 0$ *such that for each bounded set* $\mathcal{B}$ *in* $\mathcal{H}$ *we may find* $C \geqq 1$ *with*

$$(3.18) \qquad |\mathcal{T}(t)X_0|_{\mathcal{H}} \leqq C e^{-\omega t}|X_0|_{\mathcal{H}} \quad \textit{for all } t \geqq 0 \quad \textit{and} \quad X_0 \in \mathcal{B}.$$

If, in addition to the above assumptions, the function

$$(3.19) \qquad (r_1, r_2) \mapsto \frac{\varphi(r_1) - \varphi(r_2)}{r_1 - r_2} \quad \textit{admits a limit as } (r_1, r_2) \to (0, 0),$$

then for each bounded set $\mathcal{B}$ *in* $\mathcal{H}$ *we may find* $C' \geqq 1$ *with*

$$(3.20) \qquad |\mathcal{T}(t)X_0 - \mathcal{T}(t)\hat{X}_0|_{\mathcal{H}} \leqq C' e^{-\omega t}|X_0 - \hat{X}_0|_{\mathcal{H}} \quad \textit{for all } t \geqq 0 \quad \textit{and} \quad X_0, \hat{X}_0 \in \mathcal{B}.$$

Proof. Consider first the special case in which φ is linear, $\varphi(r) = mr$, with $m > 0$ and denote then by $\mathcal{S}(\cdot)X_0$ the solution of (2.1) on $\mathbb{R}^+$ with initial data X_0. By using Proposition 2 and the linearity of (2.1), we see that $\mathcal{S}$ is a linear C_0-semigroup on $\mathcal{H}$ and then Theorem 1 combined with [5, Lemma 1] implies the existence of $C_1 \geqq 1$ and $\omega_1 > 0$ such that

$$(3.21) \qquad |\mathcal{S}(t)|_{\mathcal{L}(\mathcal{H})} \leqq C_1 e^{-\omega_1 t}, \quad \text{for all } t \geqq 0.$$

Consider now the case of a nonlinear φ and put $(d\varphi/dr)(0) = m$. Clearly, $\varphi(r) = mr + r\varphi^*(r)$ for all $r \in \mathbb{R}$, with φ^* continuous on $\mathbb{R}$ and $\varphi^*(0) = 0$. Define the operator $\mathcal{A}$ in $\mathcal{H}$ by $\mathcal{A}(x, r) = (Ax + mra, \langle b, x\rangle)$, with $D(\mathcal{A}) = D(A) \times \mathbb{R}$ and define $\Phi^*: \mathcal{H} \to \mathcal{H}$ by $\Phi^*(x, r) = (r\varphi^*(r)a, 0)$. Then $\mathcal{T}(\cdot)X_0$ is the solution of the inhomogeneous problem

$$\frac{dX}{dt} = \mathcal{A}X + \Phi^*(\mathcal{T}(t)X_0), \qquad X(0) = X_0.$$

As seen above, $\mathcal{A}$ is the generator of a linear C_0-semigroup $\mathcal{S}$ which satisfies (3.21) for some $C_1 \geqq 1$ and $\omega_1 > 0$. It follows then by [8, p. 486] that

$$(3.22) \qquad \mathcal{T}(t)X_0 = \mathcal{S}(t)X_0 + \int_0^t \mathcal{S}(t-s)\Phi^*(\mathcal{T}(s)X_0)\, ds, \quad \text{for all } t \geqq 0 \quad \text{and} \quad X_0 \in \mathcal{H}.$$

Fix $\varepsilon \in]0, \omega_1[$ and let $\mathcal{B}$ be a bounded set in $\mathcal{H}$. By using the properties of φ^*, it follows that there exists $\delta > 0$ with

$$(3.23) \qquad C_1|\Phi^*(X)|_{\mathcal{H}} \leqq \varepsilon|X|_{\mathcal{H}}, \quad \text{for all } X \in \mathcal{H}, \quad |X|_{\mathcal{H}} \leqq \delta.$$

By Theorem 1, there exists $\tau \geqq 0$ such that

$$(3.24) \qquad |\mathcal{T}(t+\tau)X_0|_{\mathcal{H}} \leqq \delta, \quad \text{for all } t \geqq 0 \text{ and } X_0 \in \mathcal{B}.$$

By using (3.21), (3.22), (3.23), (3.24) and the semigroup property of $\mathcal{T}$, we see that for all $t \geqq 0$ and $X_0 \in \mathcal{B}$,

$$|e^{\omega_1 t}\mathcal{T}(t+\tau)X_0|_{\mathcal{H}} \leqq C_1|\mathcal{T}(\tau)X_0|_{\mathcal{H}} + \varepsilon \int_0^t e^{\omega_1 s}|\mathcal{T}(s+\tau)X_0|_{\mathcal{H}}\, ds,$$

hence, by the Gronwall–Bellman lemma,

$$(3.25) \qquad |\mathcal{T}(t+\tau)X_0|_{\mathcal{H}} \leqq C_2 e^{-\omega(t+\tau)}|\mathcal{T}(\tau)X_0|_{\mathcal{H}}, \quad \text{for all } t \geqq 0 \text{ and } X_0 \in \mathcal{B},$$

where $\omega = \omega_1 - \varepsilon > 0$ and $C_2 = C_1 e^{\omega\tau}$. On the other hand, by using (3.21), (3.22), the fact that Φ^* is locally Lipschitz with $\Phi^*(0, 0) = (0, 0)$ and again the Gronwall–Bellman lemma, we see easily that

$$(3.26) \qquad |\mathcal{T}(t)X_0|_{\mathscr{H}} \leqq C_3|X_0|_{\mathscr{H}}, \quad \text{for all } t\in[0,\tau] \quad \text{and} \quad X_0\in\mathscr{B}.$$

Clearly, (3.25) and (3.26) imply (3.18) provided that $C \geqq 1$ is sufficiently large.

Assume now that (3.19) holds, so that there exists a function $\psi^*: \mathbb{R}^2 \to \mathbb{R}$ with $\psi^*(r_1, r_2) \to 0$ as $(r_1, r_2) \to 0$, satisfying for all $(r_1, r_2) \in \mathbb{R}^2$,

$$(3.27) \qquad r_1\varphi^*(r_1) - r_2\varphi^*(r_2) = \varphi(r_1) - \varphi(r_2) - m(r_1 - r_2) = (r_1 - r_2)\psi^*(r_1, r_2).$$

By (3.22), it follows that for all $t \geqq 0$ and $X_0, \hat{X}_0 \in \mathscr{H}$, we have

$$(3.28) \quad \mathcal{T}(t)X_0 - \mathcal{T}(t)\hat{X}_0 = \mathcal{S}(t)(X_0 - \hat{X}_0) + \int_0^t \mathcal{S}(t-s)[\Phi^*(\mathcal{T}(s)X_0) - \Phi^*(\mathcal{T}(s)\hat{X}_0)]\, ds.$$

Choose $\delta' > 0$ such that

$$C_1|a||\psi^*(r_1, r_2)| \leqq \varepsilon, \quad \text{for all } (r_1, r_2)\in\mathbb{R}^2 \quad \text{with} \quad \sup(|r_1|, |r_2|) \leqq \delta'$$

and choose $\tau' \geqq 0$ such that

$$|\mathcal{T}(t+\tau')X|_{\mathscr{H}} \leqq \delta' \quad \text{for all } t \geqq 0 \quad \text{and} \quad X\in\mathscr{B};$$

hence, by using (3.21), (3.27), (3.28) and the semigroup property of $\mathcal{T}$, we see that

$$\begin{aligned} &e^{\omega_1 t}|\mathcal{T}(t+\tau')X_0 - \mathcal{T}(t+\tau')\hat{X}_0|_{\mathscr{H}} \\ &\quad \leqq C_1|\mathcal{T}(\tau')X_0 - \mathcal{T}(\tau')\hat{X}_0|_{\mathscr{H}} + \varepsilon \int_0^t e^{\omega_1 s}|\mathcal{T}(s+\tau')X_0 \\ &\qquad\qquad - \mathcal{T}(s+\tau')\hat{X}_0|_{\mathscr{H}}\, ds, \quad \text{for all } t\geqq 0 \quad \text{and} \quad X_0, \hat{X}_0 \in \mathscr{B}. \end{aligned}$$

Further on the proof of (3.20) follows as in the final part of the proof of (3.18).

Note that condition (3.19) holds when, for instance, φ is of class C^1 on some neighborhood of 0; in particular, it holds in the physically interesting case $\varphi(r) = e^r - 1$.

A question which seems of importance when (2.1) serves as a mathematical model of a physical system is the sensitivity of its stability property with respect to small perturbations in parameters, in other words, the structural stability of (2.1). Theorem 2 below furnishes a partial answer to this question by using the following version of the frequency domain criterion for integral equations:

LEMMA 1 (cf. [3, p. 134]). *Consider the integral equation*

$$(3.29) \qquad \sigma(t) = h(t) + \int_0^t \varphi(\sigma(s))k(t-s)\, ds$$

and assume that the following conditions hold:

(i) $dh/dt \in L^1(\mathbb{R}^+)$ *and* $d^2h/dt^2 \in L^1(\mathbb{R}^+)$;

(ii) $k = k_0 - \rho$, *where ρ is a strictly positive constant,* $k_0 \in L^1(\mathbb{R}^+)$ *and* $dk_0/dt \in L^1(\mathbb{R}^+)$.

Assume in addition to this that there exist $q \geqq 0$ and $\gamma > 0$ such that

(iii) *the continuous function φ: $\mathbb{R} \to \mathbb{R}$ satisfies $0 < r\varphi(r) < r^2\gamma$ for all $r \in \mathbb{R}$, $r \neq 0$; and*

(iv) $\operatorname{Re}(1-isq)[\tilde{k}_0(s)+(is)^{-1}\rho] \leqq \gamma^{-1}$ *for all* $s>0$.

Then there exists a solution of (3.29) *continuous on* $\mathbb{R}^+$. *Moreover, any solution* σ *of* (3.29) *continuous on* $\mathbb{R}^+$ *tends to* 0 *as* $t \to +\infty$ *and satisfies*

$$\left|\int_0^t \varphi(\sigma(s))\,ds\right| \leqq \psi(F(h(0))) \quad \textit{for all } t \geqq 0,$$

where F *and* ψ *are defined by* (3.8) *and* (3.9) *respectively.*

For each $\gamma>0$ consider the set G_γ of couples $(a,b)\in H\times H$ satisfying

$$(3.30)\qquad \begin{aligned}&\langle b, A^{-1}a\rangle>0 \quad \text{and there exists } q\geqq 0 \text{ such that}\\ &\operatorname{Re}(1-isq)\langle b,(isI^c-A^c)^{-1}A^{-1}a\rangle_{H^c}-q\langle b,A^{-1}a\rangle<\gamma^{-1} \quad \text{for all } s\geqq 0\end{aligned}$$

and the class $\mathcal{F}_\gamma$ of locally Lipschitz real functions φ satisfying

$$0<r\varphi(r)<\gamma r^2 \quad \text{for all } r\in\mathbb{R},\ r\neq 0.$$

THEOREM 2. *Let* γ *be a strictly positive number. Then* G_γ *is open and for each* $(a,b)\in G_\gamma$ *and each* $\varphi\in\mathcal{F}_\gamma$ *the zero solution of* (2.1) *is uniformly asymptotically stable in the large.*

Proof. For any $q\in\mathbb{R}$ define the function $\Psi_q\colon \mathbb{R}\times H\times H\to\mathbb{R}$ by

$$(3.31)\qquad \Psi_q(s,a,b)=\operatorname{Re}(1-isq)\langle b,(isI^c-A^c)^{-1}A^{-1}a\rangle_{H^c}-q\langle b,A^{-1}a\rangle$$

and put

$$G_\gamma^q=\{(a,b)\in H\times H:\langle b,A^{-1}a\rangle>0 \text{ and } \Psi_q(s,a,b)<\gamma^{-1} \text{ for all } s\geqq 0\}.$$

Since

$$(isI^c-A^c)^{-1}A^{-1}a=(is)^{-1}[A^{-1}a+(isI^c-A^c)^{-1}a] \quad \text{for all } s>0,$$

we have

$$(3.32)\qquad \begin{aligned}\Psi_q(s,a,b)&=\operatorname{Re}(q+is^{-1})\langle b,(isI^c-A^c)^{-1}a\rangle_{H^c}\\ &=\operatorname{Re}(q+is^{-1})\int_0^{+\infty} e^{ist}\langle b,S(t)a\rangle\,dt\end{aligned}$$

$$\text{for all } s>0 \quad \text{and} \quad (a,b)\in H\times H,$$

hence, by applying the Riemann–Lebesgue lemma, it follows that for all $(a,b)\in H\times H$, $\Psi_q(s,a,b)\to 0$ as $s\to+\infty$. Put $\Psi_q(+\infty,a,b)=0$ and then, since $\Psi_q(\cdot,a,b)$ is continuous on $[0,+\infty]$, it attains its lower upper bound $m_q(a,b)$ on $[0,+\infty]$. Clearly

$$G_\gamma^q=\{(a,b)\in H\times H:\langle b,A^{-1}a\rangle>0 \text{ and } m_q(a,b)<\gamma^{-1}\}.$$

Apply (2.5) with $n=1$ to see that $\Psi_q(s,\cdot,\cdot)$ is continuous on $H\times H$, uniformly with respect to $s\in[0,+\infty]$ (use (3.31) for $s\in[0,1]$ and (3.32) for $s\in[1,+\infty]$). It follows that m_q is continuous on $H\times H$, hence G_γ^q is open and then $G_\gamma=\bigcup_{q\geqq 0}G_\gamma^q$ is also open.

To see that the zero solution of (2.1) is uniformly asymptotically stable in the large, we may follow the same way as in the proof of Theorem 1; we have just to invoke Lemma 1 instead of [3, Thm 3.1, p. 91].

Remark 6. Since $\Psi_0(\cdot,a,b)$ is bounded from above on $\mathbb{R}^+$, for each $(a,b)\in H\times H$ with $\langle b,A^{-1}a\rangle>0$, conditions (3.30) hold with $q=0$ provided that $\gamma>0$ is small enough. It follows then, by Theorem 2, that for each $(a,b)\in H\times H$ with $\langle b,A^{-1}a\rangle>0$ the zero solution of (2.1) is uniformly asymptotically stable in the large provided that $\varphi\in\mathcal{F}_\gamma$ with sufficiently small $\gamma>0$.

4. Further considerations. Proposition 3 below shows that the conditions on a, b and A required in [7] to insure for all functions φ belonging to a specified class uniform asymptotic stability and exponential estimates may be viewed as a specialization of the conditions on a, b and A in Theorem 1. Note also that the class of functions φ considered in [7] is more restrictive than the one of Theorem 1 and closely related to that of Corollary 1. Let us first state the following:

LEMMA 2. *Let C be a linear operator in a complex Hilbert space X. If C generates a bounded holomorphic semigroup, then for all $s>0$ one has $(I+s^{-1}C^2)^{-1}\in\mathcal{L}(X)$ and*

$$(I+s^{-1}C^2)^{-1}x\to x \quad \text{as} \quad s\to+\infty \quad \text{for all } x\in X. \tag{4.1}$$

Proof. By [8, pp. 488–489], saying that C generates a bounded holomorphic semigroup is equivalent to the following: C is densely defined, there exists $\omega>0$ such that the resolvent set $P(C)$ of C contains the sector $|\arg\lambda|<2^{-1}\pi+\omega$ and for each $\varepsilon\in]0,\omega[$ there exists $K_\varepsilon\geqq 0$ such that

$$|(\lambda I-C)^{-1}|_{\mathcal{L}(X)}\leqq K_\varepsilon|\lambda|^{-1} \quad \text{for all } \lambda\in\mathbb{C}, \quad |\arg\lambda|\leqq 2^{-1}\pi+\omega-\varepsilon. \tag{4.2}$$

Then, since

$$(I+s^{-1}C^2)x=s^{-1}(-is^{1/2}I-C)(is^{1/2}I-C)x \quad \text{for all } s>0 \quad \text{and} \quad x\in D(C^2),$$

it follows that $(I+s^{-1}C^2)^{-1}\in\mathcal{L}(X)$ for all $s>0$.

Let us show that for all $\varepsilon\in]0,\omega[$ and $x\in X$,

$$(I-\lambda^{-1}C)^{-1}x\to x \quad \text{as} \quad |\lambda|\to\infty, \quad \lambda\in\mathbb{C}, \quad |\arg\lambda|\leqq 2^{-1}\pi+\omega-\varepsilon. \tag{4.3}$$

If $x\in D(C)$, we have

$$|(I-\lambda^{-1}C)^{-1}x-x|_X=|(\lambda I-C)^{-1}Cx|_X\leqq K_\varepsilon|\lambda|^{-1}|Cx|_X,$$

hence (4.3) holds for all $x\in D(C)$; then use the fact that $D(C)$ is dense in X and the fact that, by (4.2), the operators $\lambda(\lambda I-C)^{-1}$, $\lambda\in\mathbb{C}$, $|\arg\lambda|\leqq 2^{-1}\pi+\omega-\varepsilon$, are uniformly bounded, to see that (4.3) holds for all $x\in X$.

For all $x\in X$ and $s>0$ we have

$$\begin{aligned}(I+s^{-1}C^2)^{-1}x-x=-is^{1/2}(-is^{1/2}I-C)^{-1}[(I+is^{-1/2}C)^{-1}x-x]\\ +(I-is^{-1/2}C)^{-1}x-x,\end{aligned}$$

hence (4.1) follows by applying (4.3) and the fact that, by (4.2), the operators $-is^{1/2}(-is^{1/2}I-C)^{-1}$, $s>0$, are uniformly bounded.

PROPOSITION 3. *Let A be a linear operator in H and let a, b be elements in H. Assume the following conditions hold: $a\neq 0$; A^c generates a bounded holomorphic semigroup; there exists an operator $B\in\mathcal{L}(H)$ which is selfadjoint and such that $Ba=-b$, $\langle x,Bx\rangle\geqq\gamma|x|^2$ for some $\gamma>0$ and all $x\in H$ and $\langle x,BAx\rangle\leqq-\mu|x|^2$ for some $\mu>0$ and all $x\in D(A)$. Then there exists $\alpha>0$ such that $A^c+\alpha I^c$ generates a bounded holomorphic semigroup and conditions* (ii) *of Theorem* 1 *hold.*

Proof. For each $(x,y)\in H\times H$ put $\langle Bx,y\rangle=\langle x,y\rangle_1$. By using the properties of B, we see that $\langle\cdot,\cdot\rangle_1$ is an inner product on H and the norm $|\cdot|_1$ associated to this inner product is equivalent to the initial one $|\cdot|$. Denote by H_1 the space H endowed with inner product $\langle\cdot,\cdot\rangle_1$, so that H_1 is also a real Hilbert space and denote by H_1^c the complexification of H_1.

By using the properties of B, we see easily that the operator $A^c+\beta I^c$ with $\beta=\mu|B|^{-1}_{\mathcal{L}(H)}$ is dissipative in H_1^c. Since A^c generates a bounded holomorphic semigroup, it follows that A^c is densely defined, there exists $\omega>0$ such that the resolvent set $P(A^c)$ of A^c contains the sector $|\arg\lambda|<2^{-1}\pi+\omega$ and for each $\varepsilon\in]0,\omega[$ there exists $K_\varepsilon>0$ such that

$$|(\lambda I^c-A^c)^{-1}|_{\mathscr{L}(H^c)}\leqq K_\varepsilon|\lambda|^{-1}\quad\text{for all }\lambda\in\mathbb{C},\quad |\arg\lambda|\leqq 2^{-1}\pi+\omega-\varepsilon. \tag{4.4}$$

We see that for all $\lambda>\beta$, the range of $\lambda I^c-(A^c+\beta I^c)$ is all of H_1^c. It follows then that $A^c+\beta I^c$ is maximal dissipative in H_1^c, so that $A^c+\beta I^c$ generates a contraction semigroup on H_1^c, so that $A^c+\beta I^c$ generates a contraction semigroup on H_1^c [18]. Since the norms $|\cdot|_{H^c}$ and $|\cdot|_{H_1^c}$ are equivalent, we see then that $A^c+\beta I^c$ generates a bounded C_0-semigroup on H^c. Choose now $\alpha\in]0,\beta[$. By applying the Hille–Yosida theorem to $A^c+\beta I^c$, we see then that $P(A^c+\alpha I^c)$ contains the half-plane $\operatorname{Re}\lambda>-(\beta-\alpha)$ and there exists $K>0$ such that

$$|[\lambda I^c-(A^c+\alpha I^c)]^{-1}|_{\mathscr{L}(H^c)}\leqq K(\operatorname{Re}\lambda+\beta-\alpha)^{-1}\qquad\text{for all }\lambda\in\mathbb{C},\quad \operatorname{Re}\lambda>-(\beta-\alpha). \tag{4.5}$$

By the above properties of A^c, it follows that $P(A^c+\alpha I^c)$ contains the sector $|\arg(\lambda-\alpha)|<2^{-1}\pi+\omega$ and, by (4.4), one has for each $\varepsilon\in]0,\omega[$

$$|[\lambda I^c-(A^c+\alpha I^c)]^{-1}|_{\mathscr{L}(H^c)}\leqq K_\varepsilon|\lambda-\alpha|^{-1}\qquad\text{for all }\lambda\in\mathbb{C},\quad |\arg(\lambda-\alpha)|\leqq 2^{-1}\pi+\omega-\varepsilon. \tag{4.6}$$

Since $P(A^c+\alpha I^c)$ contains the sector $|\arg(\lambda-\alpha)|<2^{-1}\pi+\omega$ and the half-plane $\operatorname{Re}\lambda>-(\beta-\alpha)$, where $\beta-\alpha>0$, we may find $\omega'\in]0,\omega[$ such that $P(A^c+\alpha I^c)$ contains the sector $|\arg\lambda|<2^{-1}\pi+\omega'$. We claim moreover that for each $\varepsilon\in]0,\omega'[$ there exists $K'_\varepsilon>0$ such that

$$|[\lambda I^c-(A^c+\alpha I^c)]^{-1}|_{\mathscr{L}(H^c)}\leqq K'_\varepsilon|\lambda|^{-1}\quad\text{for all }\lambda\in\mathbb{C},\quad |\arg\lambda|\leqq 2^{-1}\pi+\omega'-\varepsilon, \tag{4.7}$$

so that $A^c+\alpha I^c$ generates a bounded holomorphic semigroup. To establish the existence of $K'_\varepsilon>0$ such that (4.7) holds, denote by λ_0 the intersection of the lines $\arg\lambda=2^{-1}\pi+\omega'-\varepsilon$ and $\operatorname{Re}\lambda=-(\beta-\alpha)$ and denote by Q the compact set in $\mathbb{C}$ deliminated by the lines $|\arg\lambda|=2^{-1}\pi+\omega'-\varepsilon$, $|\operatorname{Im}\lambda|=\operatorname{Im}\lambda_0$ and $\operatorname{Re}\lambda=2\alpha$; then use (4.5) for $\lambda\in Q$ and (4.6) for $\lambda\in\mathbb{C}$ satisfying

$$|\arg\lambda|\leqq 2^{-1}\pi+\omega'-\varepsilon,\qquad |\arg(\lambda-\alpha)|\leqq\arg\lambda_0,\quad \lambda\notin Q.$$

To see that conditions (ii) of Theorem 1 hold, we apply Remark 3. Clearly,

$$\langle b,A^{-1}a\rangle=-\langle Ba,A^{-1}a\rangle=-\langle A^{-1}a,BAA^{-1}a\rangle\geqq\mu|A^{-1}a|^2$$

and then, since $a\neq 0$, it follows $\langle b,A^{-1}a\rangle>0$. Next, by using the properties of B, we see that

$$\begin{aligned}\langle b,&(I+qA)(sI+A^2)^{-1}a\rangle\\ &=-qs\langle(sI+A^2)^{-1}a,BA(sI+A^2)^{-1}a\rangle\\ &\qquad-q\langle A(sI+A^2)^{-1}a,BA^2(sI+A^2)^{-1}a\rangle-\langle Ba,(sI+A^2)^{-1}a\rangle,\end{aligned}$$

hence

$$\begin{aligned}\langle b,(I+qA)(sI+A^2)^{-1}a\rangle&\geqq\mu qs|(sI+A^2)^{-1}a|^2+\mu q|A(sI+A^2)^{-1}a|\\ &\qquad-|Ba||(sI+A^2)^{-1}a|\quad\text{for all }s\geqq 0\quad\text{and}\quad q\geqq 0.\end{aligned} \tag{4.8}$$

Then, since the functions $s\mapsto(sI+A^2)^{-1}a$ and $s\mapsto A(sI+A^2)^{-1}a$ are continuous, there exists $q_1>0$ such that

$$\langle b,(I+qA)(sI+A^2)^{-1}a\rangle\geqq 0\quad\text{for all }s\in[0,1]\quad\text{and}\quad q\geqq q_1.$$

On the other hand, for all $s>0$ and $q \geqq 0$ we have, by (4.8)

$$\langle b, (I+qA)(sI+A^2)^{-1}a\rangle \geqq s^{-1}\{q\mu|(I+s^{-1}A^2)^{-1}a|^2 - |Ba||(I+s^{-1}A^2)^{-1}a|\}$$

and then, by Lemma 2, we may find $q_2 \geqq q_1$ such that

$$\langle b, (I+q_2A)(sI+A^2)^{-1}\rangle \geqq 0 \quad \text{for all } s>1,$$

hence conditions (3.17) hold with $q=q_2$.

To illustrate the application of the results in § 3 to special systems, we consider problem (1.1), (1.2), although integro-differential systems with more general elliptic operator A may be discussed in a similar way. In the remainder of this section, we assume that the parameters of (1.1), (1.2) satisfy the following conditions: $p_1 \in C^1([\gamma_1, \gamma_2])$ and $p_1>0$ on $[\gamma_1, \gamma_2]$; $p_2 \in C([\gamma_1, \gamma_2])$ and $p_2 \leqq 0$ on $[\gamma_1, \gamma_2]$; the real constants δ_j, $j=1, \cdots, r$, satisfy (1.4) and $\delta_1\delta_2 \leqq 0$, $\delta_3\delta_4 \geqq 0$; either $p_2 \not\equiv 0$, or $|\delta_1|+|\delta_3|>0$; a and b belong to $L^2(\gamma_1, \gamma_2)$; the function $\varphi: \mathbb{R} \to \mathbb{R}$ is locally Lipschitz and $r\varphi(r)>0$, for all $r \in \mathbb{R}$, $r \neq 0$.

Put $H=L^2(\gamma_1, \gamma_2)$ and define A in H by (1.5) with domain $D(A)$ consisting of functions $x \in C^1([\gamma_1, \gamma_2])$ such that $d^2x/d\xi^2 \in L^2(\gamma_1, \gamma_2)$ and

$$\delta_1 x(\gamma_1)+\delta_2 \frac{dx}{d\xi}(\gamma_1)=0, \qquad \delta_3 x(\gamma_2)+\delta_4 \frac{dx}{d\xi}(\gamma_2)=0.$$

According to [4, chap. V, § 14 and chap. VI, § 2], the above assumptions imply the following: A is selfadjoint; the spectrum of A^c consists of a strictly decreasing sequence $(\lambda_n)_{n\geqq 0}$ of simple, strictly negative eigenvalues; the sequence $(\lambda_n n^{-2})_{n\geqq 0}$ possesses a finite nonzero limit; the system $(e_n)_{n\geqq 0}$ of associated eigenvectors with $|e_n|=1$ is a Hilbert basis of H; the functions e_n, $n \in \mathbb{N}$, are real, uniformly bounded and belong to $C^2[(\gamma_1, \gamma_2])$; $A-\lambda_0 I$ is negative.

Clearly, system (2.1) may then be viewed as an L^2-version of problem (1.1), (1.2) and since A generates a C_0-semigroup S which satisfies (2.2) and (2.3), we may apply to (2.1) the results in § 3. Denote the Fourier coefficients of a and b with respect to e_n by a_n and b_n respectively. Then, by using Remark 3, we see that conditions (ii) of Theorem 1 may be written:

$$\sum_{n=0}^{\infty} \lambda_n^{-1} a_n b_n > 0 \text{ and there exists } q \geqq 0 \text{ such that}$$

$$\sum_{n=0}^{\infty} (1+q\lambda_n)(s+\lambda_n^2)^{-1} a_n b_n \geqq 0 \qquad \text{for all } s \geqq 0. \tag{4.9}$$

In [2] it was established that if

$$\text{for all } n \in \mathbb{N},\ a_n b_n \leqq 0, \text{ and there exists } m \in \mathbb{N} \text{ with } a_m b_m < 0, \tag{4.10}$$

and if a_n, b_n and φ verify some additional conditions, then the zero solution of (1.1), (1.2) is asymptotically stable in a specified sense. We note that the assumption (4.10) already insures uniform asymptotic stability in our setting, for (4.10) implies (4.9) with $q=|\lambda_0|^{-1}$.

Theorems 1, 2 and Corollary 1 furnish stability results for the L^2-version of problem (1.1), (1.2) under the norm of $L^2(\gamma_1, \gamma_2)$. Most of the previously established stability results for the problem (1.1), (1.2) are in terms of classical solutions and stronger norms (cf. [2], [9], [11], [12]). That is why it seems interesting to show that in fact, the solutions of the L^2-version satisfy (1.1), (1.2) in some classical sense and stability results under a stronger norm hold. Given (T_0, σ_0) in $\mathcal{H}=L^2(\gamma_1, \gamma_2)\times\mathbb{R}$, we say that the couple of functions (T, σ),

$$T:]0, +\infty[\times[\gamma_1, \gamma_2]\to\mathbb{R}, \qquad \sigma: [0, +\infty[\to\mathbb{R},$$

is a classical solution of (1.1), (1.2) with initial data (T_0, σ_0) if the following conditions hold: $T\in C^1(]0, +\infty[\times[\gamma_1, \gamma_2])$ and $\sigma\in C^1(\mathbb{R}^+)$; for each $t\in]0, +\infty[$ one has $\partial^2T(t, \cdot)/\partial\xi\in L^2(\gamma_1, \gamma_2)$; (T, σ) satisfies (1.1), (1.2) and (1.3).

We limit ourselves to derive a "classical" version for Theorem 1. Note however that such versions may be obtained also for Theorem 2 and Corollary 1 and clearly they have physical significance.

THEOREM 3. *Assume conditions* (4.9) *hold. Let* (T_0, σ_0) *be an arbitrary element in* $\mathcal{H}$ *and denote by* (u, σ) *the solution of* (2.1) *on* $\mathbb{R}^+$ *with initial data* (T_0, σ_0). *Then* (T, σ), *where* $T(t, \cdot)=u(t)$, $t\in]0, +\infty[$, *is the unique classical solution of problem* (1.1), (1.2) *with initial data*, (T_0, σ_0). *Moreover, for any bounded set* $\mathcal{B}\subset\mathcal{H}$, *the solution* (T, σ) *of* (1.1), (1.2) *with initial data* (T_0, σ_0) *satisfies*

$$(4.11)\qquad \begin{aligned} &\sigma(t)\to 0, \qquad d\sigma(t)/dt\to 0, \qquad d^2\sigma(t)/dt^2\to 0, \qquad \sup_{\xi\in[\gamma_1, \gamma_2]} |T(t, \xi)|\to 0,\\ &\sup_{\xi\in[\gamma_1, \gamma_2]}\left|\frac{\partial T}{\partial t}(t, \xi)\right|\to 0, \qquad \sup_{\xi\in[\gamma_1, \gamma_2]}\left|\frac{\partial T}{\partial \xi}(t, \xi)\right|\to 0 \quad \text{and} \quad \int_{\gamma_1}^{\gamma_2}\left|\frac{\partial^2 T}{\partial \xi^2}(t, \xi)\right|^2 d\xi\to 0\end{aligned}$$

as $t\to+\infty$, *uniformly with respect to* $(T_0, \sigma_0)\in\mathcal{B}$.

Proof. Clearly, (T, σ) satisfies (1.3). Since for all $t\in]0, +\infty[$ we have $u(t)\in D(A)$, it follows that $T(t, \cdot)\in C^1([\gamma_1, \gamma_2])$, $\partial^2T(t, \cdot)/\partial\xi^2\in L^2(\gamma_1, \gamma_2)$ and (1.2) holds. For each $(t, \xi, \eta)\in]0, +\infty[\times[\gamma_1, \gamma_2]^2$ put

$$K(t, \xi, \eta)=\sum_{n=0}^{\infty} e_n(\xi)e_n(\eta)\exp(\lambda_n t).$$

By using the classical theory of the heat equation, we see that for all $x\in L^2(\gamma_1, \gamma_2)$ and all $(t, \xi)\in]0, +\infty[\times[\gamma_1, \gamma_2]$,

$$(4.12)\qquad (S(t)x)(\xi)=\int_{\gamma_1}^{\gamma_2} K(t, \xi, \eta)x(\eta)\,d\eta=\sum_{n=0}^{\infty} x_n e_n(\xi)\exp(\lambda_n t),$$

where x_n is the Fourier coefficient of x with respect to e_n. Hence by using (2.7), we may see that for all $(t, \xi)\in]0, +\infty[\times[\gamma_1, \gamma_2]$,

$$(4.13)\quad T(t, \xi)=\int_{\gamma_1}^{\gamma_2} K(t, \xi, \eta)T_0(\eta)\,d\eta+\int_0^t \varphi(\sigma(t-s))\,ds\int_{\gamma_1}^{\gamma_2} K(s, \xi, \eta)a(\eta)\,d\eta.$$

Then it is easy to see that T is continuous. Since $\varphi\circ\sigma$ is Lipschitz, it follows also that $\partial T/\partial t$ exists on $]0, +\infty[\times[\gamma_1, \gamma_2]$ and

$$(4.14)\qquad \begin{aligned}\frac{\partial T}{\partial t}(t, \xi)=&\int_{\gamma_1}^{\gamma_2}\frac{\partial K}{\partial t}(t, \xi, \eta)T_0(\eta)\,d\eta+\varphi(\sigma(0))\int_{\gamma_1}^{\gamma_2} K(t, \xi, \eta)a(\eta)\,d\eta\\ &+\int_0^t\frac{d(\varphi\circ\sigma)}{dr}(s)\,ds\int_{\gamma_1}^{\gamma_2} K(t-s, \xi, \eta)a(\eta)\,d\eta;\end{aligned}$$

hence we may see that $\partial T/\partial t$ is continuous on $]0, +\infty[\times[\gamma_1, \gamma_2]$. The existence of the derivatives $\partial T/\partial t$ (in the classical sense) and du/dt (in the L^2-sense) implies for each

$t \in]0, +\infty[$

$$du(t)/dt = \partial T(t, \cdot)/\partial t \quad \text{a.e. in }]\gamma_1, \gamma_2[.$$

Then, since (u, σ) is a solution of (2.1), we deduce that (T, σ) satisfies (1.1).

By integrating the first equality in (1.1) on $[0, \xi]$ with respect to the second argument, we see that $\partial T/\partial \xi$ is continuous on $]0, +\infty[\times [\gamma_1, \gamma_2]$ whenever $\partial T(\cdot, 0)/\partial \xi$ is continuous on $]0, +\infty[$. Then divide the equality obtained in this way by p_1 and integrate once more to see that $\partial T(\cdot, 0)/\partial \xi$ is continuous.

Uniqueness of the solution with initial data (T_0, σ_0) follows by the Lipschitz property of φ and the uniqueness for the classical solutions of the associated homogeneous heat equation.

The claims concerning the first five limits in (4.11) follow easily by applying Theorem 1, Remark 2, (4.12), (4.13), (4.14) and the obvious estimate

$$\left|\frac{d(\varphi \circ \sigma)}{dr}(t)\right| \leqq \ell(\sup_{s \geqq 0} |\sigma(s)|)\left|\frac{d\sigma}{dt}(t)\right|, \qquad \text{a.e. in } \mathbb{R}^+.$$

Then, by using the first equation in (1.1), we see that

$$\frac{\partial}{\partial \xi}\left[p_1(\cdot)\frac{\partial T}{\partial \xi}(t, \cdot)\right] \to 0 \quad \text{in } L^2(\gamma_1, \gamma_2) \quad \text{as } t \to +\infty, \tag{4.15}$$
$$\text{uniformly with respect to } (T_0, \sigma_0) \in \mathcal{B}.$$

An integration by parts yields for all $t \in]0, +\infty[$,

$$\int_{\gamma_1}^{\gamma_2} p_1(\xi)\left|\frac{\partial T}{\partial \xi}(t, \xi)\right|^2 d\xi$$
$$= p_1(\gamma_2)T(t, \gamma_2)\frac{\partial T}{\partial \xi}(t, \gamma_2) - p_1(\gamma_1)T(t, \gamma_1)\frac{\partial T}{\partial \xi}(t, \gamma_1)$$
$$- \int_{\gamma_1}^{\gamma_2} T(t, \xi)\frac{\partial}{\partial \xi}\left[p_1(\xi)\frac{\partial T}{\partial \xi}(t, \xi)\right] d\xi.$$

By (1.2), (1.4) and $\delta_1\delta_2 \leqq 0$, $\delta_3\delta_4 \geqq 0$, it follows that the integrated term in the above equality is negative. Then use the Cauchy–Schwartz inequality, (4.15), the fourth limit in (4.11) and the properties of p_1 to see that

$$p_1(\cdot)\frac{\partial T}{d\xi}(t, \cdot) \to 0 \quad \text{in } L^2(\gamma_1, \gamma_2) \quad \text{as } t \to t + \infty \tag{4.16}$$
$$\text{uniformly with respect to } (T_0, \sigma_0) \in \mathcal{B}.$$

Now, (4.15) implies that for some $\tau > 0$, the functions $p_1(\cdot)[\partial T(t, \cdot)/\partial \xi]$, $t \geqq \tau$, $(T_0, \sigma_0) \in \mathcal{B}$, are uniformly equicontinuous and this, combined with (4.16) and the properties of p_1, yields the claim concerning the sixth limit in (4.11). The claim concerning the seventh limit in (4.11) follows then by (4.15).

REFERENCES

[1] H. BREZIS, *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les espaces de Hilbert*, Lecture Notes No. 5, North-Holland, Amsterdam/New York, 1973.

[2] T. A. BRONIKOWSKI, J. E. HALL AND J. A. NOHEL, *Quantitative estimates for a nonlinear system of integrodifferential equations arising in reactor dynamics*, this Journal, 3 (1972), pp. 567–588.

[3] C. CORDUNEANU, *Integral Equations and Stability of Feedback Systems*, Academic Press, New York/London, 1973.

[4] R. COURANT AND D. HILBERT, *Methods of Mathematical Physics*, vol. 1, Interscience, New York-London, 1953.
[5] R. DATKO, *Uniform asymptotic stability of evolutionary processes in a Banach space*, this Journal, 3 (1972), pp. 428–445.
[6] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators*, part I, Interscience, New York, 1966.
[7] E. F. INFANTE AND J. A. WALKER, *On the stability properties of an equation arising in reactor dynamics*, J. Math. Anal. Appl., 55 (1976), pp. 112–124.
[8] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.
[9] J. J. LEVIN AND J. A. NOHEL, *On a system of integro-differential equations occurring in reactor dynamics*, J. Math. Mech., 9 (1960), pp. 347–368.
[10] ———, *A system of nonlinear integro-differential equations*, Michigan Math. J., 13 (1966), pp. 257–270.
[11] ———, *A nonlinear system of integro-differential equations*, Mathematical Theory of Control, Academic Press, New York, 1966, pp. 398–405.
[12] ———, *The integrodifferential equations of a class of nuclear reactors with delayed neutrons*, Arch. Rational Mech. Anal., 31 (1968), pp. 151–171.
[13] R. K. MILLER, *On the linearization of Volterra integral equations*, J. Math. Anal. Appl., 23, (1968), pp. 198–208.
[14] ———, *An unstable nonlinear integrodifferential system*, Proceedings, U.S.-Japan Seminar on Differential and Functional Equations, Benjamin, New York, 1967, pp. 479–489.
[15] ———, *Nonlinear Volterra Integral Equations*, Benjamin, New York, 1971.
[16] J. A. NOHEL, *Some problems in nonlinear Volterra integral equations*, Bull. Amer. Math. Soc., 68 (1962), pp. 323–329.
[17] J. A. NOHEL AND D. F. SHEA, *Frequency domain methods for Volterra equations*, Advances in Math., 22 (1976), pp. 278–304.
[18] R. S. PHILLIPS, *Dissipative operators and hyperbolic systems of partial differential equations*, Trans. Amer. Math. Math. Soc., 90 (1959), pp. 193–254.
[19] O. J. STAFFANS, *Positive definite measures with applications to a Volterra equation*, Ibid., 218 (1976), 219–237.
[20] A. SUHADOLC, *On a system of integro-differential equations*, SIAM J. Appl. Math., 21 (1971), pp. 195–206.
[21] J. A. WALKER AND E. F. INFANTE, *On the stability of an operator equation modeling nuclear reactors with delayed neutrons*, Quart. Appl. Math., 34 (1977) 4, pp. 421–427.
[22] K. YOSIDA, *Functional Analysis*, Springer-Verlag, Berlin-New York, 1974.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0016 $01.00/0

CARDINAL-TYPE APPROXIMATIONS OF A FUNCTION AND ITS DERIVATIVES*

L. LUNDIN† AND F. STENGER‡

Abstract. Whittaker's cardinal function is used to approximate certain analytic functions in Sobolev norm. L^∞ is of primary interest, although attention is also given to $L^2(-\infty,\infty)$. Results are given for functions defined on a general contour in the complex plane, and special treatment is given to the important real domains $(-\infty,\infty)$ $(-1,1)$ and $(0,\infty)$. In all cases, it is shown that the approximations converge to the function at the rate $C\exp(-cn^{1/2})$, where n is the number of points of interpolation and C and c are positive constants.

1. Introduction. Recently a family of approximation methods has been derived in [10] which is based on Whittaker's cardinal function. These methods are extremely accurate when applied to functions that are analytic on an interval (a, b), but which may have singularities at a and b. Formulas were derived in [10] for interpolating f over (a, b), for the approximate integration of f over (a, b), for approximating the Hilbert transform of f over (a, b), and for approximating the Fourier transform of f over $(-\infty,\infty)$. The important cases of $(a, b)=(-1, 1)$, $(0,\infty)$ and $(-\infty,\infty)$ were given special consideration in [10].

The formulas derived in [10] as well as those derived in [9], [12] are shown in [11] to be optimal in a certain sense. Let $\mathscr{D}$ be a region in the complex plane having boundary points a and b, and let $\mathscr{F}$ denote the class of functions which are analytic in $\mathscr{D}$ and which have a finite norm, the norm depending on the particular method of approximation. For example, $\|f\|=[\int_{\partial\mathscr{D}}|f(z)|^p|dz|]^{1/p}$, $p\geqq 1$, for the case of quadrature, while $\|f\|=[\int_{\partial\mathscr{D}}|\phi'(z)|\,|f(z)|^p|dz|]^{1/p}$ for the case of interpolation, where ϕ is a conformal map of $\mathscr{D}$ onto $\{w:|\mathrm{Im}\, w|<\pi/2\}$ with $\phi(a)=-\infty$ and $\phi(b)=\infty$. It was shown in [11] that the rate of convergence of an n-point optimal method of approximation is $C\|f\|\exp(-cn^{1/2})$ where C and c are constants depending only on $\mathscr{D}$ and the particular norm. The error of the formulas derived in [9], [10], [12] is also of the form $C\|f\|\exp(-cn^{1/2})$.

Optimal rules for solving elliptic boundary value problems by integral equation methods are also being sought by Soviet mathematicians, notably Sobolev (see e.g. [8]). However, these mathematicians are seeking optimal rules relative to Sobolev norms, which demand only the existence of the integral of some power of the first m derivatives of the function. Thus, their rules do not take advantage of a.e. analyticity of solutions of elliptic boundary value problems that one encounters in applications. The resulting rate of convergence which they achieve using an n-point approximation is $Cn^{-c}\|f\|$.

In the present paper we shall derive various n-point methods of interpolating a function over an interval (a, b) which are analogous to those in [10]. The resulting approximations may be used to approximate f, as well as a finite number of derivatives of f over $[a, b]$, to within an error which is uniformly bounded by $C\|f\|\exp(-cn^{1/2})$. This was not possible by means of the approximation given in [10] for the case of a finite or semi-infinite interval.

In § 2 we establish our notation. In § 3 we obtain some explicit expressions for approximating derivatives over $(-\infty,\infty)$ by means of Whittaker's cardinal function.

* Received by the editors October 26, 1976, and in revised form September 19, 1977. This work was supported by the United States Army under Research Grant DAHC-04-G-0175.

† Department of Mathematics, University of Delaware, Newark, Delaware 19711.

‡ Department of Mathematics, University of Utah, Salt Lake City, Utah 84112.

In §§ 4 and 5 we obtain some bounds on the error of L^2 and L^∞ approximation of derivatives over $(-\infty, \infty)$. In § 6 we consider the general case of approximation of derivatives over a contour in the complex plane, and we derive a basis for approximation, which is very nearly orthogonal and for which the coefficients of the approximate orthogonal expansion may be very simply expressed. Finally, in §§ 7 and 8 we consider the important special cases of approximation over $(-1, 1)$ and $(0, \infty)$.

Due to the simplicity and accuracy the approximations that we shall derive are particularly useful for obtaining approximate solutions to differential and integral equations via the Galerkin method. For example, given an elliptic differential equation $Lu = g$ over a region Ω, the solution will be an analytic function of each independent variable at each point in the interior of Ω whenever g and the coefficients of Lu are analytic. If $\Omega \subset R^2$ and if $\partial\Omega$ can be subdivided into a finite collection of analytic arcs $\{S_i\}$ such that the boundary data, g and the coefficients in Lu are analytic on the interior of each S_i, then the solution u will also be analytic there. Approximate n-point methods which converge more rapidly than $C \exp(-cn^{1/2})$ could be derived if we knew the behavior of u at its singularities; however, this behavior is usually unknown [2, Chap. 4]. It is in these instances that the approximations derived in this paper are particularly powerful; since we know where the singularities of u occur, we can subdivide Ω (or $\partial\Omega$, depending on the method of approximate solution) into a finite number of subsets Ω_i such that singularities of u occur only on the boundary of each Ω_i, and we can then approximate u on each Ω_i by the methods of this paper. The resulting approximation is accurate to within an error bounded by $C \exp(-cn^{1/(2d)})$, where d denotes the dimensions of the region.

Let us briefly list some references in which approximations of this type have been carried out. The solution of the problem $\Delta u - k^2 u = 0$ via an integral equation method was considered in [1], for a solution u in the exterior of a bounded domain Ω in the plane subject to the condition $u = f$ on $\partial\Omega$, where $\partial\Omega$ is the union of a finite number of analytic arcs and f is analytic on the interior of each arc. An n-point Galerkin approximate solution was obtained using approximations derived in [10], for which the resulting error at any point exterior to Ω is bounded by $C \exp(-cn^{1/2})$.

In [7], a Galerkin method was derived for solving the integral equation problem

$$f(x) = \lambda \int_{\partial\Omega} K(x, y) f(y)\, dy + g(x), \qquad x \in \partial\Omega, \tag{1.1}$$

in which $\partial\Omega$ is the surface of a body and consists of a finite number of "analytic" patches $\{S_i\}$, $K(x, y)$ is any one of several potential theory-type kernels (e.g., $(\partial/\partial n) \cdot (1/r)$), where λ may or may not be an eigenvalue of (1.1) and where g is an analytic function of two independent variables on each S_i. An n-point Galerkin approximate solution was obtained using basis functions of the form $\phi_i(x)\psi_i(y)$ where ϕ_i and ψ_i were derived in [10]. The resulting error is bounded by $C \exp(-cn^{1/4})$ at all points on $\partial\Omega$.

An n-point approximate solution was obtained in [3] for the Hilbert problem

$$F_+ = GF_- \quad \text{on } \partial\Omega,$$

in which $\partial\Omega$ consists of a finite number of analytic arcs $\{S_i\}$ and G is analytic on each S_i. The approximate solution is accurate to within $C \exp(-cn^{1/2})$ at all points of $\Omega \cup \partial\Omega$.

In [13] a method was derived for obtaining an approximate solution of Wiener–Hopf integral equations. The methods of [10] were used there also. Under suitable assumptions of analyticity of the Fourier transform of the kernel of the equation and of the nonhomogeneous term, the error of an n-term approximate solution may be shown bounded by $C \exp(-cn^{1/2})$.

In the references cited thus far, it was not necessary to approximate derivatives of solutions. In [4] an n-point approximate solution was obtained for the problem

$$y'' = y - y^3/x^2, \qquad y(0) = y(\infty) = 0,$$

using approximations derived in the present paper. The resulting error is bounded by $C \exp(-cn^{1/2})$.

In [6] the scattered field due to an axially symmetric body in 3-space was approximated via the Galerkin method of solution of a vector integral equation. The assumptions under which the problem was solved are similar to those used in [7]; however, the integral equation formulation demanded an accurate approximation of both the solution and its first partial derivatives in order to get a final n-point Galerkin approximation of the scattered field for which the error at all points outside the body was bounded by $C \exp(-cn^{1/2})$. An approximation derived in § 7 of the present paper was used to carry out the procedure.

2. Background and notation. It is convenient to introduce several notations and definitions which are used throughout the remainder of the paper. As usual, $\mathbb{C}$ denotes the complex plane and R the real line.

For h positive, N a positive integer and f defined on R, we will be considering each of the following functions:

$$S(k, h)(z) = \left[\frac{\pi}{h}(z - kh)\right]^{-1} \sin\left[\frac{\pi}{h}(z - kh)\right], \tag{2.1}$$

for $z \in \mathbb{C}$ and $k = 0, \pm 1, \pm 2, \cdots$;

$$C(f, h) = \sum_{k=-\infty}^{\infty} f(kh) S(k, h); \tag{2.2}$$

$$C_N(f, h) = \sum_{k=-N}^{N} f(kh) S(k, h); \tag{2.3}$$

$$\varepsilon(f, h) = f - C(f, h); \tag{2.4}$$

$$\varepsilon_N(f, h) = f - C_N(f, h). \tag{2.5}$$

The functions $S(k, h)$ satisfy the orthogonality relation

$$\int_R S(k, h)(x) S(j, h)(x)\, dx = \begin{cases} h & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases} \tag{2.6}$$

Also, it is easily seen that

$$S(k, h)(jh) = \delta_{k,j}, \tag{2.7}$$

and for this reason, any function g of the form

$$g = \sum_{k=-\infty}^{\infty} a_k S(k, h) \tag{2.8}$$

is usually called a (Whittaker) cardinal function.

We consider n-point rules of approximation where $n = 2N + 1$, $N \geqq 1$, and the initial approximations are as given in (2.3). Our error bounds will be essentially of the form $C \exp(-c_1 N^{1/2})$. One obtains a bound of the form $C \exp(-cn^{1/2})$ (as promised in the previous section) via the following:

$$c_1 N^{1/2} = c_1 \left(\frac{N}{2N+1}\right)^{1/2} n^{1/2} \geqq \left(\frac{1}{3}\right)^{1/2} c_1 n^{1/2}.$$

DEFINITION 2.1. Let $B(h)$ denote the family of all functions f defined on $\mathbb{C}$ such that f is entire, $f \in L^2(R)$ and

$$|f(z)| \leqq C \exp[\pi |y|/h],$$

for all $z = x + iy \in \mathbb{C}$ and for some positive constant C.

The importance of the function $C(f, h)$ with regards to the class $B(h)$ is indicated in the following theorem [5].

THEOREM 2.1. *If $f \in B(h)$, then $f \equiv C(f, h)$ and*

$$\int_R f^2(x)\, dx = h \sum_{k=-\infty}^{\infty} f^2(kh). \tag{2.9}$$

The sequence $\{h^{-1/2} S(k, h)\}_{k=-\infty}^{\infty}$ is therefore a complete orthonormal sequence in $B(h)$.

DEFINITION 2.2. For d positive and $p \geqq 1$, let B_d^p denote the collection of all functions f such that f is analytic in

$$\mathcal{D}_d = \{x + iy : |y| < d\}, \tag{2.10}$$

$$\int_{-d}^{d} |f(x+iy)|\, dy \to 0 \quad \text{as } x \to \pm\infty, \tag{2.11}$$

and $N(f, p, \mathcal{D}_d) < \infty$, where

$$N(f, p, \mathcal{D}_d) = \lim_{y \to d^-} \left\{ \left[\int |f(x+iy)|^p\, dx \right]^{1/p} + \left[\int |f(x-iy)|^p\, dx \right] \right\}. \tag{2.12}$$

Our next theorem is found in [10], and gives a very useful representation of the function defined in (2.4).

THEOREM 2.2. *Let d and h be positive and let $p = 1$ or 2. Assume $f \in B_d^p$. Then for each $x \in R$,*

$$\begin{aligned} \varepsilon(f, h)(x) = \frac{\sin(\pi x/h)}{2\pi i} \int_R \Bigg\{ & \frac{f(t - id^-)}{(t - x - id) \sin[\pi(t - id)/h]} \\ & - \frac{f(t + id^-)}{(t - x + id) \sin[\pi(t + id)/h]} \Bigg\}\, dt. \end{aligned} \tag{2.13}$$

We shall obtain some explicit bounds on $\varepsilon(f, h)$ in § 4. It will be seen that not only $\varepsilon(f, h)$ but all derivatives of $\varepsilon(f, h)$ decrease to zero very rapidly as $h \to 0$. Indeed, if h is sufficiently small, then "for all practical purposes" the sequence $\{h^{-1/2} S(k, h)\}_{k=-\infty}^{\infty}$ is an orthonormal sequence in B_d^p. This interpretation is a consequence of the following result, the proof of which is similar to that of Theorem 6.4, and is omitted.

THEOREM 2.3. *If $f \in B_d^1$, then*

$$\left| \int_R f(x) S(k, h)(x)\, dx - hf(k\ h) \right| \leqq \frac{N(f, 1, d)}{2d} e^{-\pi d/h}. \tag{2.14}$$

In order to facilitate the use of formula (2.11), we shall write simply

$$\varepsilon(f, h) = S_h \cdot I(f, h), \tag{2.15}$$

where

$$S_h(x)=\frac{1}{2\pi i}\sin\left(\frac{\pi x}{h}\right), \tag{2.16}$$

and thus $I(f, h)(x)$ is the integral term in (2.13).

3. Function and derivative evaluation. In applying the results of this paper, one would need to evaluate certain derivatives of a given cardinal function at the points $x=kh$, k an integer. We demonstrate a method for accomplishing such evaluations by extending an idea put forth in [5].

Let us consider a function

$$g=\sum_{k=-\infty}^{\infty} a_k S(k, h), \tag{3.1}$$

where we assume the series in (3.1) converges. Then for each $a \in R$, Theorem 2.1 implies that

$$g(x)=\sum_{k=-\infty}^{\infty} g(a+kh)S(k, h)(x-a), \qquad x\in R. \tag{3.2}$$

From this, it is easily verified that

$$g'(x)=-\sum_{\substack{k=-\infty\\ k\neq 0}}^{\infty}\frac{(-1)^k}{kh}g(x+kh), \tag{3.3}$$

and so we are led to the relation

$$g^{(n)}(x)=-\frac{1}{h}\sum_{\substack{k=-\infty\\ k\neq 0}}^{\infty}\frac{(-1)^k}{k}g^{(n-1)}(x+kh). \tag{3.4}$$

In particular, we have

$$g^{(n)}(kh)=-\frac{(-1)^k}{h}\sum_{\substack{j=-\infty\\ j\neq k}}^{\infty}\frac{(-1)^j}{j-k}g^{(n-1)}(jh). \tag{3.5}$$

The result (3.5) may then be used to obtain $g^{(n)}(kh)$ in terms of the constants $\{a_j\}_{j=-\infty}^{\infty}$. For example, we immediately obtain

$$g'(kh)=-\frac{(-1)^k}{h}\sum_{\substack{j=-\infty\\ j\neq 0}}^{\infty}\frac{(-1)^j}{j-k}a_j. \tag{3.6}$$

For the case $n=2$, one may substitute (3.6) in (3.5) and employ partial fraction techniques to obtain

$$g''(kh)=-\frac{1}{h^2}\left[\frac{\pi^2}{3}a_k+2(-1)^k\sum_{\substack{j=-\infty\\ j\neq k}}^{\infty}\frac{(-1)^j}{(j-k)^2}a_j\right]. \tag{3.7}$$

As motivation for the formulas given here, as well as for results given in § 4, we present a theorem regarding the nature of certain functions as in (3.1).

THEOREM 3.1. *Let h be positive and let N be a positive integer. Let g denote the function $\sum_{k=-N}^{N} a_k S(k, h)$. Then for each nonnegative integer n, $g^{(n)} \in B(h)$ and hence*

$$g^{(n)} \equiv \sum_{k=-\infty}^{\infty} g^{(n)}(kh) S(k, h), \tag{3.8}$$

and

$$\int_R [g^{(n)}(x)]^2 \, dx = h \sum_{k=-\infty}^{\infty} [g^{(n)}(kh)]^2. \tag{3.9}$$

Proof. First, since g is entire, so is $g^{(n)}$. Calculating $S(0, h)^{(n)}$ directly, we find

$$S(0, h)^{(n)}(z)$$
$$= \frac{h}{2\pi i} \sum_{j=0}^{n} \binom{n}{j} \left(\frac{i\pi}{h}\right)^j [e^{i\pi z/h} + (-1)^{j+1} e^{-i\pi z/h}](-1)^{n-j}(n-j)! z^{-(n-j+1)}.$$

So, for z real, we see that $|S(0, h)^{(n)}| = O(|z|^{-1})$ as $|z| \to \infty$. Hence, the continuity of $S(0, h)^{(n)}$ implies $S(0, h)^{(n)} \in L^2(R)$. Since $g^{(n)}$ is a finite sum of multiples of translations of $S(0, h)^{(n)}$, we conclude that $g^{(n)} \in L^2(R)$.

Next, we see that for $z = x + iy$,

$$|S(0, h)^{(n)}(z)| \leqq \frac{h}{2\pi} \sum_{j=0}^{n} \binom{n}{j} \left(\frac{\pi}{h}\right)^j [e^{-\pi y/h} + e^{\pi y/h}](n-j)! |z|^{-(n-j+1)}$$
$$\leqq \frac{hn!}{\pi |z|} e^{\pi |y|/h} \sum_{j=0}^{n} \binom{n}{j} \left(\frac{\pi}{h}\right)^j |z|^{-(n-j)}$$
$$= \frac{hn!}{\pi |z|} e^{\pi |y|/h} \left(\frac{\pi}{h} + \frac{1}{|z|}\right)^n.$$

Thus, the continuity of $S(0, h)^{(n)}$ implies the existence of a constant L such that

$$|S(0, h)^{(n)}(z)| \leqq L \, e^{\pi |y|/h}, \quad \text{for all } z = x + iy \in \mathbb{C}.$$

Then for $C = L \sum_{k=-N}^{N} |a_k|$, we have

$$|g^{(n)}(z)| \leqq C e^{\pi |y| h},$$

which concludes the proof that $g^{(n)} \in B(h)$. Equations (3.8) and (3.9) follow from Theorem 2.1. Q.E.D.

We close this section with a corollary to Theorem 2.1. The corollary treats only the case $n = 1$, but similar results for larger n are of course possible.

COROLLARY. *Let h, N and g be as in Theorem 2.1. Then*

$$\int_R [g'(x)]^2 \, dx = h^{-1} \sum_{k=-N}^{N} \sum_{j=-N}^{N} c_{k,j} a_k a_j, \tag{3.10}$$

where

$$c_{k,j} = \begin{cases} \dfrac{\pi^2}{3} & \textit{if } i = j, \\ 2(-1)^{i+j}(j-i)^{-2} & \textit{if } i \neq i. \end{cases} \tag{3.11}$$

Proof. Substituting (3.6) in (3.8), we find

$$g' = -\frac{1}{h} \sum_{k=-\infty}^{\infty} (-1)^k \left[\sum_{j=-N}^{N} \frac{(-1)^j}{j-k} a_j\right] S(k, h). \tag{3.12}$$

Hence, we conclude from (3.9) that

$$\int_R [g'(x)]^2\,dx = h^{-1} \sum_{k=-\infty}^{\infty} \left[\sum_{\substack{j=-N \\ j\neq k}}^{N} \frac{(-1)^j}{j-k} a_j \right]^2, \tag{3.13}$$

which may be written

$$\int_R [g'(x)]^2\,dx = h^{-1} \sum_{i=-N}^{N} \sum_{j=-N}^{N} \left[(-1)^{i+j} a_i a_j \sum_{\substack{k=-\infty \\ k\neq i,j}}^{\infty} \frac{1}{(i-k)(j-k)} \right]. \tag{3.14}$$

An application of partial fractions to (3.14) yields (3.10). Q.E.D.

4. Approximation in $H_d^m(R)$.

DEFINITION 4.1. For m a nonnegative integer and d positive, let H_d^m denote the family of all functions f such that

(a) $f \in B_d^2$,

(b) $f^{(j)} \in L^2(R)$, for $j = 0, 1, \cdots, m$.

For $f \in H_d^m$,

$$\|f\|_{m,2} = \left[\sum_{j=0}^{m} \|f^{(j)}\|_2^2 \right]^{1/2}. \tag{4.1}$$

We consider next approximations in $H_d^m(R)$.

THEOREM 4.1. *Let h and d be positive, let m be a nonnegative integer and let $f \in H_d^m(R)$. Then if $\pi d/h \geqq 1$,*

$$\|\varepsilon(f, h)^{(n)}\|_2 \leqq \frac{n!eN(f, 2, \mathcal{D}_d)}{2\pi \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n, \tag{4.2}$$

for $n = 0, 1, \cdots, m$.

Proof. For each $j = 0, 1, \cdots, n$, we define the functions ϕ, Φ and Φ_+ by

$$\Phi(t) = \frac{f(t - id^-)}{(t - x - id)^j \sin[\pi(t - id)/h]},$$

$$\phi(t) = \frac{1}{2\pi} \int_R e^{-ixt}\,\Phi(x)\,dx,$$

and

$$\Phi_+(x + iy) = \int_0^{\infty} e^{ixt} e^{-yt} \phi(t)\,dt, \qquad y > 0.$$

Then

$$\Phi_+(x + iy) = \frac{1}{i} \int_R \frac{\Phi(t)}{t - x - iy}\,dt,$$

and we may use Parseval's theorem as in [10] to obtain

$$\begin{aligned} \int_R |\Phi_+(x + iy)|^2\,dx &= 2\pi \int_0^{\infty} e^{-2yt} |\phi(t)|^2\,dt \\ &\leqq 2\pi \int_R |\phi(t)|^2\,dt \\ &= \int_R |\Phi(x)|^2\,dx. \end{aligned}$$

Thus, taking $I_{+,j}(x)=\Phi_+(x+id)$ and noting that $|\sin[\pi(t-id)/h]|\geqq\sinh(\pi d/h)$, we have

$$\|I_{+,j}\|_2^2\leqq\int_R|\Phi(t)|^2\,dt$$

$$\leqq d^{-2j}[\sinh(\pi d/h)]^{-2}\int_R|f(t-id^-)|^2\,dt.$$

We may repeat the argument above for

$$\Phi(t)=\frac{f(t+id^-)}{(t-x+id)^j\sin[\pi(x+id/h)]}$$

to obtain a similar bound on the corresponding $I_{-,j}$. Recalling the notation of (2.15), we note that $(1/i)I^{(j)}=j!(I_{+,j}+I_{-,j})$, where $I=I(f,h)$, and so

$$\|I^{(j)}\|_2^2\leqq(j!)^2d^{-2j}[\sinh(\pi d/h)]^{-2}N(f,2,\mathscr{D}_d)^2.$$

So, using Leibnitz's formula in (2.15), we find, recalling (2.16), that

$$\|\varepsilon(f,h)^{(n)}\|_2\leqq\sum_{j=0}^{n}\binom{n}{j}\|S_h^{(j)}I(f,h)^{(n-j)}\|_2$$

$$\leqq\frac{1}{2\pi}\sum_{j=0}^{n}\binom{n}{j}\left(\frac{\pi}{h}\right)^j\|I^{(n-j)}\|_2$$

$$\leqq\frac{n!N(f,2,\mathscr{D}_d)}{2\pi\sinh(\pi d/h)d^n}\sum_{j=0}^{n}\left(\frac{\pi d}{h}\right)^j\frac{1}{j!}$$

$$\leqq\frac{n!N(f,2,\mathscr{D}_d)}{2\pi\sinh(\pi d/h)}\left(\frac{\pi}{h}\right)^n\sum_{j=0}^{n}\frac{1}{j!},$$

which yields (4.2). Q.E.D.

The corollary which follows is now immediate.

COROLLARY. *Let h, d, m and f be as in Theorem 4.1. Then there exists a positive constant C, depending only on f, d and m, such that*

$$\|\varepsilon(f,h)\|_{m,2}\leqq Ch^{-m}e^{-\pi d/h}. \tag{4.3}$$

Next, we use the results above to obtain a bound on $\varepsilon_N(f,h)$. This bound is of course more useful in applications since, in general, the practical evaluation of a function at an infinite collection of points is impossible. First a lemma.

LEMMA 4.2. *Let $S(k,h)$ be as defined in (2.1). Then if n is a nonnegative integer,*

$$\|S(k,h)^{(n)}\|_2=\left(\frac{\pi}{2n+1}\right)^{1/2}\left(\frac{\pi}{h}\right)^{n-1/2} \tag{4.4}$$

and

$$\|S(k,h)^{(n)}\|_\infty\leqq\frac{1}{n+1}\left(\frac{\pi}{h}\right)^n. \tag{4.5}$$

Proof. We have

$$S(k,h)(x)=\frac{h}{2\pi}\int_{-\pi/h}^{\pi/h}e^{-ikht}e^{ixt}\,dt,$$

so that

$$S(k,h)^{(n)}(x)=\frac{h}{2\pi}\int_{-\pi/h}^{\pi/h}(it)^n e^{-ikht}e^{ixt}\,dt. \tag{4.6}$$

Parseval's theorem then yields

$$\|S(k,h)^{(n)}\|_2=(2\pi)^{1/2}\left\{\int_{-\pi/h}^{\pi/h}\left(\frac{h}{2\pi}\right)^2 t^{2n}\,dt\right\}^{1/2},$$

from which (4.4) easily follows. Also, (4.6) gives

$$|S(k,h)^{(n)}(x)|\leqq\frac{h}{2\pi}\int_{-\pi/h}^{\pi/h}|t|^n\,dt=\frac{1}{n+1}\left(\frac{\pi}{h}\right)^n,$$

and so (4.5) holds. Q.E.D.

THEOREM 4.3. *Let h, d, m and f be as in Theorem 4.1. Furthermore, assume that there exist positive constants C and α such that*

$$|f(x)|\leqq Ce^{-\alpha|x|} \tag{4.7}$$

for all $x\in(-\infty,\infty)$. Then if N is a positive integer,

$$\begin{aligned}\|\varepsilon_N(f,h)^{(n)}\|_2\leqq\left(\frac{\pi}{h}\right)^n\Bigg[&\frac{n!\,eN(f,2,\mathcal{D}_d)}{2\pi\sinh(\pi d/h)}\\&+2C\alpha^{-1}[h(2n+1)]^{-1/2}e^{-\alpha Nh}\Bigg]\end{aligned} \tag{4.8}$$

for $n=0,1,\cdots,m$.

Proof. Applying the triangle inequality to the nth derivative of (2.5) and using (2.4), we find that

$$\begin{aligned}&\|\varepsilon_N(f,h)^{(n)}\|_2\\&\leqq\|\varepsilon(f,h)^{(n)}\|_2+\left\|\sum_{|k|>N}f(kh)S(k,h)^{(n)}\right\|_2\\&\leqq\|\varepsilon(f,h)^{(n)}\|_2+\left(\frac{\pi}{2n+1}\right)^{1/2}\left(\frac{\pi}{h}\right)^{n-1/2}\sum_{k=N+1}^{\infty}[|f(-kh)|+|f(kh)|]\\&\leqq\|\varepsilon(f,h)^{(n)}\|_2+2\left(\frac{\pi}{2n+1}\right)^{1/2}\left(\frac{\pi}{h}\right)^{n-1/2}Ce^{-\alpha(N+1)h}[1-e^{-\alpha h}]^{-1}.\end{aligned}$$

The inequality $\alpha h\leqq e^{\alpha h}-1$ and (4.2) now yield (4.8). Q.E.D.

Again, we immediately obtain a corollary regarding approximation in $H_d^m(R)$.

COROLLARY. *Let f, d and m be as in the Corollary to Theorem 4.1, and assume that f also satisfies (4.5). Let N be a positive integer, and take $h=(\pi d/\alpha N)^{1/2}$. Then there exists a constant K depending on f, d and m such that*

$$\|\varepsilon_N(f,h)\|_{m,2}\leqq KN^{1/2m+1/4}\exp[-(\pi d\alpha N)^{1/2}]. \tag{4.9}$$

We pause to note that the bound in (4.9) approaches zero (as $N\to\infty$) faster than any power of N^{-1}. In particular, we note that the error converges to zero at the rate $C\exp(-cN^{1/2})$.

We remark that if we also knew $\|\varepsilon(f,h)^{(n)}\|_\infty$, then the inequality

$$\|\varepsilon(f,h)^{(n)}\|_p\leqq\|\varepsilon(f,h)^{(n)}\|_2^{2/p}\|\varepsilon(f,h)^{(n)}\|_\infty^{1-2/p} \tag{4.10}$$

would yield a bound on $\|\varepsilon(f, h)^{(n)}\|_p$ for any $p \in [2, \infty]$. We shall obtain bounds on $\|\varepsilon(f, h)^{(n)}\|_\infty$ in the following section.

5. Supremum norm approximation.

THEOREM 5.1. *Let h and d be positive, and assume $f \in B_d^p$, for $p = 1$ or 2. For n a nonnegative integer,*

$$\varepsilon(f, h)^{(n)} = \sum_{j=0}^{n} \binom{n}{j} S_h^{(j)} I(f, h)^{(n-j)}. \tag{5.1}$$

Furthermore, assuming $\pi d/h \geqq 1$, if $p = 1$,

$$\|\varepsilon(f, h)^{(n)}\|_\infty \leqq \frac{n!\, eN(f, 1, \mathscr{D}_d)}{2\pi d \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n, \tag{5.2}$$

and if $p = 2$ and $h < \pi$, letting $\bar{d} = \max(1, d^{-n})$, we have

$$\|\varepsilon(f, h)^{(n)}\|_\infty \leqq \frac{\bar{d} n!\, e^{1/2} N(f, 2, \mathscr{D}_d)}{2\pi^{1/2} d^{1/2} \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n (1 - h^2/\pi^2)^{-1/2}. \tag{5.3}$$

Proof. We have obtained (5.1) previously, and it is easily seen from (2.16) that $\|S_h^{(j)}\|_\infty \leqq (2\pi)^{-1}(\pi/h)^j$. Also, (2.13) yields

$$I(f, h)^{(j)}(x) = j! \int_R \left\{ \frac{f(t - id^-)}{(t - x - id)^{j+1} \sin[\pi(t - id)/h]} - \frac{f(t + id^-)}{(t - x + id)^{j+1} \sin[\pi(t + id)/h]} \right\} dt.$$

So if $f \in B_d^1$, we use the inequalities $|t - x + id| \geqq d$ and $|\sin(u + iv)| \geqq \sinh|v|$ for u and v real to obtain

$$\|I(f, h)^{(j)}\|_\infty \leqq \frac{j!\, N(f, 1, \mathscr{D}_d)}{d^{j+1} \sinh(\pi d/h)}.$$

Hence, if $f \in B_d^1$ and $\pi d/h \geqq 1$,

$$\begin{aligned} \|\varepsilon(f, h)^{(n)}\|_\infty &\leqq \frac{n!\, N(f, 1, \mathscr{D}_d)}{2\pi d^{n+1} \sinh(\pi d/h)} \sum_{j=0}^{n} \left(\frac{\pi d}{h}\right)^j \frac{1}{j!} \\ &\leqq \frac{n!\, eN(f, 1, \mathscr{D}_d)}{2\pi d \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n, \end{aligned}$$

which is (5.2).

Now, if $f \in B_d^2$, we have

$$|I(f, h)^{(j)}(x)| \leqq \frac{j!}{\sinh(\pi d/h)} \int_R \left\{ \left| \frac{f(t - id^-)}{(t - x - id)^{j+1}} \right| + \left| \frac{f(t + id^-)}{(t - x + id)^{j+1}} \right| \right\} dt.$$

Schwarz's inequality then implies

$$\begin{aligned} \|I(f, h)^{(j)}\|_\infty &\leqq \frac{j!\, N(f, 2, \mathscr{D}_d)}{\sinh(\pi d/h)} \left[\int_R \frac{dt}{(t^2 + d^2)^{j+1}} \right]^{1/2} \\ &= \frac{[\pi j!\, (\frac{1}{2})_j]^{1/2}}{d^{j+1/2}} \frac{N(f, 2, \mathscr{D}_d)}{\sinh(\pi d/h)}. \end{aligned}$$

So, it follows that if $\pi d/h \geqq 1$ and $\pi/h > 1$,

$$\|\varepsilon(f, h)^{(n)}\|_\infty \leqq \frac{N(f, 2, \mathcal{D}_d)}{2\pi^{1/2} d^{1/2} \sinh(\pi d/h)} \sum_{j=0}^{n} \binom{n}{j} \left(\frac{\pi}{h}\right)^{n-j} d^{-j} \left[j! \left(\frac{1}{2}\right)_j\right]^{1/2}$$

$$\leqq \frac{\bar{d} n! N(f, 2, \mathcal{D}_d)}{2\pi^{1/2} d^{1/2} \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n \left[\sum_{j=0}^{n} \frac{(\frac{1}{2})_j}{j!} \left(\frac{h}{\pi}\right)^{2j}\right]^{1/2}$$

$$\cdot \left[\sum_{j=0}^{n} \left(\frac{1}{j!}\right)^2\right]^{1/2}$$

$$\leqq \frac{\bar{d} n! N(f, 2, \mathcal{D}_d)}{2\pi^{1/2} d^{1/2} \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n (1 - h^2/\pi^2)^{-1/2} e^{1/2},$$

which is (5.3). Q.E.D.

THEOREM 5.2. *Let h and d be positive, and let N be a positive integer. Assume f is such that*

$$|f(x)| \leqq C e^{-\alpha|x|}, \qquad x \in R, \tag{5.4}$$

where C and α are positive constants, and let $\pi d/h \geqq 1$. Then, if $f \in B_d^1$,

$$\|\varepsilon_N(f, h)^{(n)}\|_\infty \leqq \frac{n!\, e N(f, 1, \mathcal{D}_d)}{2\pi d \sinh(\pi d/h)} \left(\frac{\pi}{h}\right)^n + \frac{2c}{(n+1)\alpha\pi} \left(\frac{\pi}{h}\right)^{n+1} e^{-\alpha N h}, \tag{5.5}$$

while if $f \in B_d^2$ and $h < \pi$,

$$\|\varepsilon_N(f, h)^{(n)}\|_\infty \leqq \frac{\bar{d} n!\, e^{1/2} N(f, 2, \mathcal{D}_d)(\pi/h)^n}{2\pi^{1/2} d^{1/2} (1 - h^2/\pi^2)^{1/2} \sinh(\pi d/h)} + \frac{2C}{(n+1)\alpha\pi} \left(\frac{\pi}{h}\right)^{n+1} e^{-\alpha N h}, \tag{5.6}$$

where $\bar{d} = \max(1, d^{-n})$.

Proof. Applying the triangle inequality and using (4.5), we find

$$\|\varepsilon_N(f, h)^{(n)}\|_\infty \leqq \|\varepsilon(f, h)^{(n)}\|_\infty + \left\|\sum_{|k|>N} f(kh) S(k, h)^{(n)}\right\|_\infty$$

$$\leqq \|\varepsilon(f, h)^{(n)}\|_\infty + \frac{1}{n+1} \left(\frac{\pi}{h}\right)^n \sum_{k=N+1}^{\infty} [|f(kh)| + |f(-kh)|].$$

We may use the method of proof of Theorem 4.3 to obtain

$$\|\varepsilon_N(f, h)^{(n)}\|_\infty \leqq \|\varepsilon(f, h)^{(n)}\|_\infty + \frac{2C}{(n+1)\alpha\pi} \left(\frac{\pi}{h}\right)^{n+1} e^{-\alpha N h}.$$

Hence, (5.5) and (5.6) follow from Theorem 5.1. Q.E.D.

COROLLARY. *Let the hypotheses of Theorem 5.2 hold for $p = 1$ or 2 with $h = (\pi d/\alpha N)^{1/2}$. Then there exists a positive constant K depending only on n, d, f and α, such that*

$$\|\varepsilon_N(f, h)^{(n)}\|_\infty \leqq K N^{(n+1)/2} \exp[-(\pi d \alpha N)^{1/2}]. \tag{5.7}$$

6. Applications of conformal mappings. The results of the previous sections may be extended to yield interpolation and approximation formulas over an arbitrary interval or even a contour. We consider the general case here.

Let $\mathscr{D}$ be a simply connected domain in the complex plane $\mathbb{C}$, and let $\mathscr{D}_d$ be defined as in (2.10). Let ϕ be a conformal map of $\mathscr{D}$ onto $\mathscr{D}_d$ and set $\psi = \phi^{-1}$. Let $a = \psi(-\infty)$ and $b = \psi(\infty) \neq a$ be boundary points of $\mathscr{D}$, and let us take

$$\Gamma = \{w \in \mathscr{D}\colon\ w = \psi(x),\ -\infty \leqq x \leqq \infty\}. \tag{6.1}$$

Let $B(\mathscr{D})$ denote the family of all functions f that are analytic in $\mathscr{D}$, such that

$$N(f, \mathscr{D}) = \int_{\partial\mathscr{D}} |f(z)\,dz| < \infty, \tag{6.2}$$

and such that (for u real)

$$\int_{\psi(L+u)} |f(z)\,dz| \to 0 \quad \text{as } u \to \pm\infty, \tag{6.3}$$

where

$$L = \{iy\colon\ -d \leqq y \leqq d\}. \tag{6.4}$$

We also set

$$x_k = \psi(kh), \qquad k = 0, \pm 1, \pm 2, \cdots, \tag{6.5}$$

and let g be a function which is analytic in $\mathscr{D}$, whose properties we shall determine more explicitly in what follows.

THEOREM 6.1. *If $f\phi'/g \in B(\mathscr{D})$, then for each $x \in \Gamma$,*

$$\begin{aligned}\frac{g(x)\sin[\pi\phi(x)/h]}{2\pi i} &\int_{\partial\mathscr{D}} \frac{[f(z)\phi'(z)/g(z)]\,dz}{[\phi(z)-\phi(x)]\sin[\pi\phi(z)/h]} \\ &= f(x) - \sum_{k=-\infty}^{\infty} \frac{f(x_k)}{g(x_k)}\, g(x) S(k, h) \circ \phi(x).\end{aligned} \tag{6.6}$$

We omit the proof of Theorem 6.1, since it is a simple application of (2.13) and change of variables.

In applications it is desirable to be able to choose g so that the infinite sum in (6.6) is an accurate approximation of f on Γ; in addition, we shall also want to approximate the nth derivative of f on Γ, for $n = 1, 2, \cdots, m$. We shall thus assume not only the existence and boundedness of $f^{(n)}$ on Γ, but also the possibility of choosing g so that the nth derivative of the infinite sum in (6.6) exists and is bounded on Γ. From (6.6) it is clear that the introduction of the following notation will be helpful.

$$S(g, \phi, k, h) = g \cdot [S(k, h) \circ \phi], \tag{6.7}$$

$$C(f, g, \phi, h) = \sum_{k=-\infty}^{\infty} \frac{f(x_k)}{g(x_k)} S(g, \phi, k, h), \tag{6.8}$$

and

$$C_N(f, g, \phi, h) = \sum_{k=-N}^{N} \frac{f(x_k)}{g(x_k)} S(g, \phi, k, h). \tag{6.9}$$

THEOREM 6.2. *Let m be a nonnegative integer and let $f\phi'/g \in B(\mathcal{D})$. If there exists a constant C' depending on m and g such that for all $x \in \Gamma$ and $z \in \partial\mathcal{D}$,*

$$\left|\left(\frac{d}{dx}\right)^n\left\{\frac{g \cdot S_h \circ \phi}{\phi(z)-\phi}\right\}(x)\right| \leqq C'h^{-n}, \tag{6.10}$$

for $n = 0, 1, \cdots, m$, then there exists a constant C depending only on m, g and f such that for all $x \in \Gamma$,

$$|f^{(n)}(x) - C(f, g, \phi, h)^{(n)}(x)| \leqq Ch^{-n} e^{-\pi d/h}, \tag{6.11}$$

for $n = 0, 1, \cdots, m$.

Proof. Differentiating both sides of (6.6) n times with respect to x, using (6.10) and noting that for $z \in \partial\mathcal{D}$, $|\text{Im}\, \phi(z)| = d$ so that $|\sin[\pi\phi(z)/h]| \geqq \sinh(\pi d/h)$, we find that the left-hand side of (6.11) is bounded by

$$\frac{C'h^{-n}}{\sinh(\pi d/h)} \int_{\partial\mathcal{D}} |[f(z)\phi'(z)/g(z)]\, dz|.$$

By assumption $f\phi'/g \in B(\mathcal{D})$, which yields (6.11). Q.E.D.

THEOREM 6.3. *Let the conditions of Theorem 6.2 be satisfied, and let*

$$|S(g, \phi, 0, h)^{(n)}(x)| \leqq C_1 h^{-n}, \tag{6.12}$$

for $n = 0, 1, \cdots, m$, for all $x \in \Gamma$, where C_1 is a positive constant. Assume furthermore that there exist constants C_2 and α such that for all $x \in \Gamma$,

$$|f(x)/g(x)| \leqq C_2\, e^{-\alpha|\phi(x)|}. \tag{6.13}$$

Let N be a positive integer and take $h = (\pi d/\alpha N)^{1/2}$. Then there exists a positive constant K depending only on m, d, f and α such that for all $x \in \Gamma$,

$$|f^{(n)}(x) - C_N(f, g, \phi, h)^{(n)}(x)| \leqq KN^{(n+1)/2} \exp[-(\pi d\alpha N)^{1/2}], \tag{6.14}$$

for $n = 0, 1, \cdots, m$.

Proof. Using the triangle inequality, we have

$$\begin{aligned} |f^{(n)}(x) - C_N(f, g, \phi, h)^{(n)}(x)| &\leqq |f^{(n)}(x) - C(f, g, \phi, h)^{(n)}(x)| \\ &\quad + \left|\sum_{|k|>N} \frac{f(x_k)}{g(x_k)} S(g, \phi, k, h)^{(n)}(x)\right|. \end{aligned} \tag{6.15}$$

Using (6.12) and (6.13) and recalling that $\phi(x_k) = kh$, we find that

$$\begin{aligned} \left|\sum_{|k|>N} \frac{f(x_k)}{g(x_k)} S(g, \phi, k, h)^{(n)}(x)\right| &\leqq \frac{2C_1C_2}{h^n} \sum_{k=N+1}^{\infty} e^{-\alpha kh} \\ &\leqq \frac{2C_1C_2}{\alpha h^{n+1}} e^{-\alpha Nh}. \end{aligned}$$

Hence, taking $h = (\pi d/\alpha N)^{1/2}$ and bounding $|f^{(n)}(x) - C(f, g, \phi, h)^{(n)}(x)|$ by means of (6.11), using the inequality (6.15) we obtain (6.14). Q.E.D.

We show next that if h is sufficiently small, then for "purposes of applications" the sequence $\{S(g, \phi, k, h)\}_{k=-\infty}^{\infty}$ may be considered to be an orthogonal sequence when used for approximating functions f such that $fg\phi' \in B(\mathcal{D})$.

If we replace g in (6.6) by 1 and then replace f by $f \cdot S(g, \phi, j, h)$, we arrive at the identity

$$\begin{aligned}\frac{\sin[\pi\phi(x)/h]}{2\pi i}\int_{\partial\mathscr{D}}\frac{f(z)S(g,\phi,j,h)(z)\phi'(z)\,dz}{[\phi(z)-\phi(x)]\sin[\pi\phi(z)/h]}\\ =f(x)S(g,\phi,j,h)(x)-\frac{f(x_j)g(x_j)}{g(x)}S(g,\phi,j,h)(x),\end{aligned}\tag{6.16}$$

which is valid, provided $fg' \in B(\mathscr{D})$. (This may be shown by replacing $f\phi'/g$ in Theorem 6.1 by $fgS(g, \phi, j, h)\phi'/g$.)

If $z \in \partial\mathscr{D}$, then $[\operatorname{Im}\phi(z)]\operatorname{sgn}[\operatorname{Im}\phi(z)] = d$, so that

$$\begin{aligned}\frac{1}{2\pi i}\int_\Gamma\frac{\sin[\pi\phi(x)/h]}{\phi(z)-\phi(x)}\phi'(x)\,dx\\ =\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\sin[\pi x/h]}{\phi(z)-x}\,dx\\ =\frac{i}{2}\exp\left[\frac{i\pi\phi(z)}{h}\operatorname{sgn}\operatorname{Im}\phi(z)\right],\end{aligned}\tag{6.17}$$

where $\operatorname{sgn}\operatorname{Im}\phi(z)=1$ (respectively -1) if $\operatorname{Im}\phi(z)>0$ (respectively <0). Furthermore, we have

$$\int_\Gamma S(j,h)\circ\phi(x)\phi'(x)\,dx=\int_{-\infty}^{\infty}S(j,h)(x)\,dx=h.\tag{6.18}$$

Hence, multiplying (6.16) by ϕ', integrating over Γ, interchanging the order of integration and using (6.17) and (6.18), we arrive at

$$\begin{aligned}\frac{i}{2}\int_{\partial\mathscr{D}}\frac{f(z)S(g,\phi,j,h)(z)\exp[i\pi\phi(z)\operatorname{sgn}\operatorname{Im}\phi(z)/h]\phi'(z)\,dz}{\sin[\pi\phi(z)/h]}\\ =\int_\Gamma f(x)S(g,\phi,j,h)(x)\phi'(x)\,dx-hf(x_j)g(x_j).\end{aligned}\tag{6.19}$$

We bound the integral on the left-hand side of (6.19) by noting that if $z \in \partial\mathscr{D}$, then

$$\frac{\pi}{h}|S(j,h)\circ\phi(z)/\sin[\pi\phi(z)/h]|=\left|\frac{(-1)^j}{\phi(z)-jh}\right|\leqq\frac{1}{d}\tag{6.20a}$$

and

$$|\exp[i\pi\phi(z)\operatorname{sgn}\operatorname{Im}\phi(z)/h]|=e^{-\pi d/h}.\tag{6.20b}$$

Recalling that $S(g, \phi, j, h) = g \cdot [S(k, h)\circ\phi]$ and substituting (6.20) into (6.19), we obtain the following result.

THEOREM 6.4. *If* $fg\phi' \in B(\mathscr{D})$, *then*

$$\begin{aligned}\left|\int_\Gamma f(x)S(g,\phi,j,h)(x)\phi'(x)\,dx-hf(x_j)g(x_j)\right|\\ \leqq hN(fg\phi',\mathscr{D})\,e^{-\pi d/h}/(2\pi d).\end{aligned}\tag{6.21}$$

In the remaining sections, we shall restrict our attention to two important special cases. In particular, we consider the case when $\mathscr{D}$ is an eye-shaped region, and the case when $\mathscr{D}$ is an infinite wedge.

7. Approximation over $(-1, 1)$. In this case, our maps ϕ and ψ have the form

$$\phi(z)=\log\frac{1+z}{1-z}, \qquad \psi(w)=\tanh\frac{w}{2}, \tag{7.1}$$

and so

$$\phi'(z)=\frac{2}{1-z^2}. \tag{7.2}$$

$\mathscr{D}$ has the special form (see Fig. 1)

$$\begin{aligned}\mathscr{D}&=[\{z:|z+i\cot d|<\csc d\}\cap\{z:\operatorname{Im} z\geqq 0\}]\\&\quad\cup[\{z:|z-i\cot d|<\csc d\}\cap\{z:\operatorname{Im} z\leqq 0\}],\end{aligned} \tag{7.3}$$

where we assume $0<d\leqq\pi/2$. Furthermore, we see that

$$\Gamma=\{z\in\mathscr{D}:\ \phi(z)\in[-\infty,\infty]\}=\{x:-1\leqq x\leqq 1\}, \tag{7.4}$$

and $a=-1$ and $b=1$.

For g, we consider the function

$$g(x)=(1-x^2)^\beta, \tag{7.5}$$

where $\beta\geqq 0$.

We now state the primary result of this section. The proof is based on the sequence of lemmas which follow.

THEOREM 7.1. *Let ϕ be as in* (7.1), *let m be a nonnegative integer, and let g be as in* (7.5) *where $\beta\geqq m$. Assume $f\phi'/g\in B(\mathscr{D})$ where $\mathscr{D}$ is the eye-shaped region of Fig.* 1, *and that there exist positive constants C and α such that for all $x\in[-1,1]$,*

$$|f(x)/g(x)|\leqq C(1-x^2)^\alpha. \tag{7.6}$$

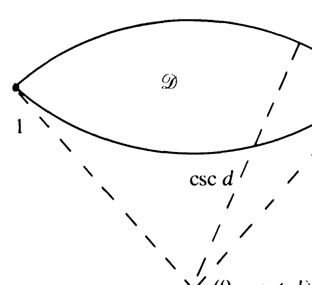

FIG. 1. *The region $\mathscr{D}$.*

Let N be a positive integer, $h=(\pi d/\alpha N)^{1/2}$ and

$$x_k=\tanh(kh/2),\quad k=0,\pm 1,\pm 2,\cdots. \tag{7.7}$$

If $f^{(m)}$ exists on $[-1, 1]$, *then* (6.14) *holds for all $x\in[-1,1]$ and for $n=0,1,\cdots,m$, where K is a constant depending only on m, d, f and α.*

LEMMA 7.2. *Let g be defined as in* (7.5), *where β is a nonnegative constant. If j is any nonnegative integer, there exists a constant C depending on β and j such that for all*

$x \in (-1, 1)$,

(7.8) $$|g^{(j)}(x)| \leqq C(1-x^2)^{\beta-j}.$$

Proof. Leibnitz's formula gives us

$$\left|\left(\frac{d}{dx}\right)^j (1-x^2)^\beta\right| = \sum_{k=0}^{j} \binom{j}{k} [(1+x)^\beta]^{(k)} [(1-x)^\beta]^{(j-k)}.$$

We note that for $x \in [-1, 1]$, $|1+x| \leqq 2$ and $|1-x| \leqq 2$, and so for $0 \leqq k \leqq j$,

$$(1+x)^{\beta-k}(1-x)^{\beta-j+k} = (1-x^2)^{\beta-j}(1+x)^{j-k}(1-x)^k \leqq 2^j(1-x^2)^{\beta-j}.$$

Hence, it follows that

$$\left|\left(\frac{d}{dx}\right)^j (1-x^2)^\beta\right| \leqq \left\{\sum_{k=0}^{j} 2^j \binom{j}{k} \left|\beta(\beta-1)\cdots(\beta-k+1)\right|\left|\beta(\beta-1)\cdots(\beta-j+k+1)\right|\right\} \cdot (1-x^2)^{\beta-j},$$

which yields (7.8). Q.E.D.

LEMMA 7.3. *Let j and k be positive integers with $j \geqq k$, and let $\phi(x) = \log[(1+x)/(1-x)]$. Then for all $x \in (-1, 1)$,*

(7.9) $$|\phi^{(j)}(x)[\phi'(x)]^{k-j}| \leqq C(1-x^2)^{-k}$$

where C is a constant depending only on j and k.

Proof. Differentiating ϕ, we obtain

$$\phi'(x) = \frac{2}{1-x^2} = \frac{1}{1+x} + \frac{1}{1-x},$$

so that if $j \geqq 1$, then

$$\phi^{(j)}(x) = (j-1)!\left[\frac{(-1)^{j-1}}{(1+x)^j} + \frac{1}{(1-x)^j}\right].$$

Hence, we immediately obtain

$$\frac{\phi^{(j)}(x)}{[\phi'(x)]^j} = \frac{(j-1)!}{2^j}[(-1)^{j-1}(1-x)^j + (1+x)^j],$$

and so since each of $|1+x|$ and $|1-x|$ is bounded by 2 on $[-1, 1]$, we conclude

$$|\phi^{(j)}(x)| \leqq 2(j-1)![\phi'(x)]^j.$$

By combining this inequality with (7.8) and (7.2), we arrive at (7.9). Q.E.D.

LEMMA 7.4. *Let j be any nonnegative integer. Then there exists a constant C depending only on j such that for all $x \in (-1, 1)$ and for all uniformly bounded positive h,*

(7.10) $$\left|\left(\frac{d}{dx}\right)^j \sin\left[\frac{\pi\phi(x)}{h}\right]\right| \leqq Ch^{-j}(1-x^2)^{-j}.$$

Proof. First, from the definition of ϕ,

$$\left|\left(\frac{d}{dx}\right)^j \sin\left[\frac{\pi\phi(x)}{h}\right]\right| \leqq \left|\left(\frac{d}{dx}\right)^j e^{i\pi\phi(x)/h}\right|$$

$$= \left|\left(\frac{d}{dx}\right)^j [(1+x)^{i\pi/h}(1-x)^{-i\pi/h}]\right|.$$

We now use Leibnitz's rule, as we did in the proof of Lemma 7.2, to obtain

$$\left|\left(\frac{d}{dx}\right)^j \sin\left[\frac{\pi\phi(x)}{h}\right]\right| \leqq h^{-j}(1-x^2)^{-j}\left|\sum_{k=0}^{j}\binom{j}{k}\left(\frac{i\pi}{h}-k+1\right)_k h^k\right.$$

$$\left.\cdot(-1)^{j-k}\left(-\frac{i\pi}{h}-j+k+1\right)_{j-k} h^{j-k}(1-x)^{k-i\pi/h}(1+x)^{j-k+i\pi/h}\right|.$$

For all $x \in [-1, 1]$, $|(1\pm x)^{k\pm i\pi/h}| \leqq 2^k$ for $0 \leqq k \leqq j$. Moreover, for $0 \leqq k \leqq j$,

$$\left|h^k\left(\pm\frac{i\pi}{h}-k+1\right)_k\right| \leqq \gamma_j,$$

where γ_j depends only on j, since by assumption h is uniformly bounded. Hence, we arrive at (7.10). Q.E.D.

LEMMA 7.5. *Let j be a nonnegative integer. There exists a positive constant C depending only on j and d, such that if $x \in (-1, 1)$ and $z \in \partial\mathcal{D}$, then*

$$\left|\left(\frac{d}{dx}\right)^j\left[\frac{1}{\phi(z)-\phi(x)}\right]\right| \leqq C(1-x^2)^{-j}. \tag{7.11}$$

Proof. The inequality (7.11) is clearly valid for $j=0$, since ϕ is real on $(-1, 1)$ and $\operatorname{Im}\phi(z) = \pm d$. If $j>0$, it is easily shown by induction that

$$\left(\frac{d}{dx}\right)^j \frac{1}{\phi(z)-\phi(x)} = \sum_{k=1}^{j} \frac{p_{jk}(\phi)(x)}{[\phi(z)-\phi(x)]^{k+1}}, \tag{7.12}$$

where

$$p_{jk}(\phi) = \sum_{s=0}^{k} c_{jks} \prod_{q=0}^{k} [\phi^{(q)}]^{l_{jksq}}$$

and where c_{jks} and l_{jksq} are integers; furthermore, $l_{jksq} \geqq 0$ for all j, k, s and q, and $\sum_{q=0}^{k} q l_{jksq} = j$. Hence, by Lemma 7.3,

$$\left|\prod_{q=0}^{k} [\phi^{(q)}(x)]^{l_{jksq}}\right| \leqq C'(1-x^2)^{-j}, \tag{7.13}$$

where C' depends only on j, k and s. Finally, for $z \in \partial\mathcal{D}$ and $x \in (-1, 1)$, $|\phi(z)-\phi(x)| \geqq d$, and so (7.10) follows from (7.12) once (7.13) is applied. Q.E.D.

LEMMA 7.6. *Let j be a nonnegative integer, and let $h>0$ be uniformly bounded. Then there exists a constant C depending only on j such that*

$$|[S(k, h)\circ\phi]^{(j)}(x)| \leqq Ch^{-j}(1-x^2)^{-j}, \tag{7.14}$$

for all $x \in (-1, 1)$, and for every integer k.

Proof. The proof is in many ways similar to the proof of Lemma 7.4. We apply Leibnitz's rule to the representation

$$S(k, h)\circ\phi(x) = \frac{1}{2}\int_{-1}^{1} e^{-ik\pi t} e^{i\pi\phi(x)t/h}\,dt$$

$$= \frac{1}{2}\int_{-1}^{1} e^{-i\pi kt}\,(1+x)^{i\pi t/h}(1-x)^{-i\pi t/h}\,dt$$

and find

$$[S(k,h)\circ\phi]^{(j)}(x)$$
$$=\tfrac{1}{2}h^{-j}(1-x^2)^{-j}\int_{-1}^{1} e^{-i\pi kt}$$
$$\cdot\left[\sum_{s=0}^{j}(-1)^{j-s}\binom{j}{s}(1+x)^{j-s+i\pi t/h}(1-x)^{s-i\pi t/h}\left(\frac{i\pi t}{h}-s+1\right)_s h^s\right.$$
$$\left.\cdot\left(-\frac{i\pi t}{h}-j+s+1\right)_{j-s} h^{j-s}\right]dt.$$

The estimates

$$|(1+x)^{j-s+i\pi t/h}|\leqq 2^{j-s},\qquad -1\leqq x\leqq 1,$$
$$|(1-x)^{s-i\pi t/h}|\leqq 2^{s},\qquad -1\leqq x\leqq 1,$$

and

$$\left|\left(\pm\frac{i\pi t}{h}-s+1\right)_s h^2\right|\leqq \gamma_{js},\qquad 0\leqq s\leqq j,$$

where γ_{js} is a constant depending on j and s, may now be used to obtain (7.14). Q.E.D.

We now complete the proof of Theorem 7.1. To the identity

$$\left[\frac{gS_h\circ\phi}{\phi(z)-\phi}\right]^{(n)}(x)=\sum_{k=0}^{n}\sum_{j=0}^{k}\binom{n}{k}\binom{k}{j}\left[\frac{1}{\phi(z)-\phi(x)}\right]^{(n-k)}[S_h\circ\phi]^{(j)}(x)$$
$$\cdot g^{(k-j)}(x),$$

which is a consequence of Leibnitz's rule, we apply the results of Lemmas 7.2, 7.4 and 7.5, to obtain

$$\left|\left[\frac{(g\cdot S_h\circ\phi)}{\phi(z)-\phi}\right]^{(n)}(x)\right|$$
$$\leqq\sum_{k=0}^{n}\sum_{j=0}^{k}C_1(1-x^2)^{k-n}C_2(1-x^2)^{-j}h^{-j}C_3(1-x^2)^{\beta+j-k}$$
$$\leqq C(1-x^2)^{\beta-n}h^{-n},$$

where C is a constant depending on m and β.

In Theorem 7.1 we assume $\beta\geqq m$. Hence, the conditions of Theorem 6.2 are satisfied, and so (6.11) holds.

If we apply Leibnitz's rule to obtain the nth derivative of $S(g,\phi,0,h)$ and then use Lemmas 7.2 and 7.6 to bound each term in the resulting expression, we find that (6.12) is satisfied, since $\beta\geqq n$.

Finally, if $|f(x)/g(x)|\leqq C(1-x^2)^{\alpha}$ on $(-1,1)$, where $\alpha>0$, then

$$|f(x)/g(x)|\leqq C(1+|x|)^{\alpha}(1-|x|)^{\alpha}\leqq 2^{2\alpha}C\left(\frac{1-|x|}{1+|x|}\right)^{\alpha}.$$

But it is easily shown that for $x \in (-1, 1)$, we have

$$\log \frac{1-|x|}{1+|x|} = -\left|\log \frac{1+x}{1-x}\right|.$$

Hence, if (7.6) is satisfied, then letting $C' = 2^{2\alpha}C$,

$$\begin{aligned} |f(x)/g(x)| &\leqq C' \exp\left\{-\alpha\left|\log \frac{1+x}{1-x}\right|\right\} \\ &= C' \exp\{-\alpha|\phi(x)|\}, \end{aligned}$$

and so (6.10) holds. Thus, Theorem 7.1 follows from Theorem 6.3.

We observe that an application of Theorem 7.1 requires that f approach zero quite rapidly as $x \to \pm 1$. Of course, f must also be m times differentiable on $[-1, 1]$. If f is sufficiently differentiable, but f does not approach zero as necessary, we consider in place of f a new function F. The function F is obtained from f by subtracting a certain polynomial. In particular, we may choose $F = f - p_M$, $M > \beta$, where p_M is of the form

$$(7.15) \qquad p_M(x) = \sum_{k=0}^{M} \left[a_k\left(\frac{1-x}{2}\right)^{k+1}\left(\frac{1+x}{2}\right)^{k} + b_k\left(\frac{1-x}{2}\right)^{k}\left(\frac{1+x}{2}\right)^{k+1}\right].$$

The polynomial p_M is obtained inductively by defining the sequence of polynomials $\{p_k\}_{k=0}^{M}$ as follows:

$$(7.16) \qquad p_0(x) = a_0\left(\frac{1-x}{2}\right) + b_0\left(\frac{1+x}{2}\right),$$

where

$$(7.17) \qquad a_0 = f(-1), \qquad b_0 = f(1),$$

and

$$(7.18) \quad p_{k+1}(x) = p_k(x) + a_{k+1}\left(\frac{1-x}{2}\right)^{k+2}\left(\frac{1+x}{2}\right)^{k+1} + b_{k+1}\left(\frac{1-x}{2}\right)^{k+1}\left(\frac{1+x}{2}\right)^{k+2},$$

where

$$(7.19) \qquad \begin{aligned} a_{k+1} &= \frac{2^{k+1}}{(k+1)!}[f^{(k+1)}(-1) - p_k^{(k+1)}(-1)], \\ b_{k+1} &= \frac{(-2)^{k+1}}{(k+1)!}[f^{(k+1)}(1) - p_k^{(k+1)}(1)]. \end{aligned}$$

8. Approximation over $(0, \infty)$. Here, $\mathscr{D}$ will denote the sector

$$(8.1) \qquad \mathscr{D} = \{z : |\arg z| < d\}$$

where $0 < d \leqq \pi$. The functions ϕ and ψ are given by

$$(8.2) \qquad \phi(z) = \log z, \qquad \psi(w) = e^w,$$

and

$$(8.3) \qquad \phi'(z) = \frac{1}{z}.$$

Further,

$$\Gamma=\psi([-\infty,\infty])=\{x: 0\leqq x\leqq\infty\}. \tag{8.4}$$

We let $B(\mathscr{D})$ denote the family of all functions f analytic in $\mathscr{D}$ such that

$$N(f,d)=\lim_{\theta\to d^-}\int_0^\infty \{|f(r\,e^{i\theta})|+|f(r\,e^{-i\theta})|\}\,dr<\infty, \tag{8.5}$$

and

$$\int_{-d}^{d}|f(r\,e^{i\theta})|r\,d\theta\to 0 \qquad \text{as } r\to\infty. \tag{8.6}$$

We shall consider the points

$$x_k=e^{kh}, \qquad k=0,\pm1,\pm2,\cdots, \tag{8.7}$$

and the function

$$g(z)=z^\beta, \tag{8.8}$$

where $\beta\geqq 0$.

It is much simpler to prove the validity of Theorems 6.2 and 6.3 for approximation over $(0,\infty)$ than it was over $(-1,1)$. We shall therefore give the equivalent of Lemmas 7.2 to 7.6 in a single lemma. In this lemma, which follows, C_j, $C_{j,\beta}$ and $C_{j,d}$ are positive constants depending only on j, j and β, and j and d, respectively.

LEMMA 8.1. *Let j be a nonnegative integer, let $\phi(z)=\log z$, and let g be defined as in* (8.8). *Then for all $x\in(0,\infty)$,*

$$|g^{(j)}(x)|\leqq C_{j,\beta}x^{\beta-j} \qquad (j\geqq 0); \tag{8.9}$$

$$|\phi^{(j)}(x)|\leqq C_jx^{-j} \qquad (j>0); \tag{8.10}$$

$$\left|\left(\frac{d}{dx}\right)^j[\phi(z)-\phi(x)]^{-1}\right|\leqq C_{j,d}x^{-j} \qquad (z\in\partial\mathscr{D},\quad j\geqq 0); \tag{8.11}$$

$$\left|\left(\frac{d}{dx}\right)^j\sin[\pi\phi(x)/h]\right|\leqq C_jh^{-j}x^{-j} \qquad (0<h\leqq h_0,\quad j\geqq 0); \tag{8.12}$$

$$|[S(k,h)\circ\phi]^{(j)}(x)|\leqq C_jh^{-j}x^{-j} \qquad (0<h\leqq h_0,\quad j\geqq 0,\ k \text{ an integer}). \tag{8.13}$$

Proof. Inequality (8.9) is a consequence of the identity

$$g^{(j)}(x)=\left(\frac{d}{dx}\right)^jx^\beta=(-1)^j(\beta)_jx^{\beta-j},$$

while (8.10) follows from

$$\left(\frac{d}{dx}\right)^j\log x=(-1)^{j-1}(j-1)!x^{-j}.$$

If $j=0$, (8.11) is clearly satisfied, since $z\in\partial\mathscr{D}$ and $x\in(0,\infty)$ imply $|\phi(z)-\phi(x)|\geqq d$. If $j>0$, it is readily verified by induction that there exist constants $C_{1j}, C_{2j},\cdots, C_{jj}$, depending only on j, such that

$$\left(\frac{d}{dx}\right)^j[\phi(z)-\phi(x)]^{-1}=x^{-j}\sum_{k=1}^{j}C_{kj}[\phi(z)-\phi(x)]^{-k-1}$$

which yields (8.11).

Next, we have for $x \in (0, \infty)$ and $0 < h \leqq h_0$,

$$\begin{aligned}\left|\left(\frac{d}{dx}\right)^j \sin[\pi\phi(x)/h]\right| &= \left|\operatorname{Im}\left(\frac{d}{dx}\right)^j \exp[i\pi(\log x)/h]\right| \\ &\leqq \left|\left(\frac{d}{dx}\right)^j \exp[i\pi(\log x)/h]\right| \\ &= \left|\left(\frac{d}{dx}\right)^j x^{i\pi/h}\right| \\ &= \left|\frac{i\pi}{h}\left(\frac{i\pi}{h}-1\right)\cdots\left(\frac{i\pi}{h}-j+1\right)x^{-j+i\pi/h}\right| \\ &= x^{-j}\prod_{k=0}^{j-1}\left(k^2+\frac{\pi^2}{h^2}\right)^{1/2} \\ &= x^{-j}\frac{\pi}{h}(j-1)!\prod_{k=1}^{j-1}\left(1+\frac{\pi^2}{k^2h^2}\right)^{1/2} \\ &\leqq Ch^{-j}x^{-j},\end{aligned}$$

where C depends only on j. This completes the proof of (8.12).

The proof of (8.13) is similar to the proof of Lemma 4.2, starting with the identity

$$S(k, h)\circ\phi(x) = \frac{1}{2}\int_{-1}^{1} e^{-i\pi kt}x^{i\pi t/h}\,dt.$$

We omit the details. Q.E.D.

THEOREM 8.2. *Let $\phi(z) = \log z$ and let g be as in* (8.8) *with $\beta \geqq m$, where m is a nonnegative integer. Let $f\phi'/g \in B(\mathscr{D})$, where $\mathscr{D}$ is the sector* (8.1), *and $B(\mathscr{D})$ is defined by* (8.5)–(8.6). *Let*

$$|f(x)/g(x)| \leqq C_1x^{\alpha}, \qquad 0 \leqq x \leqq 1,$$
$$|f(x)/g(x)| \leqq C_1x^{-\alpha-1}, \qquad 1 \leqq x \leqq \infty,$$

where C_1 and α are positive constants. Let $x_k = e^{kh}$, $k = 0, \pm 1, \pm 2, \cdots$, and take $h = (\pi d/\alpha N)^{1/2}$, where N is a positive integer. If $f^{(m)}$ exists on $[0, \infty]$, then (6.11) *holds for all $x \in [0, \infty]$ and for $n = 0, 1, \cdots, m$, where K is a constant depending only on m, d, f and ∞.*

The proof of Theorem 8.2 is similar to the proof of Theorem 7.1, and we omit it. In order to apply Theorem 8.2, it is required that f approach zero quite rapidly as $x \to 0$. If f does not satisfy this criterion, but is sufficiently differentiable on $[0, \infty]$, then one may instead approximate F, where

$$F(x) = f(x) - e^{-x}\sum_{k=0}^{[\beta+1]} a_kx^k, \tag{8.14}$$

where

$$\begin{aligned}&a_0 = f(0),\\ &a_k = \frac{(-1)^k}{k!}\left[f^{(k)}(0) - \sum_{j=0}^{k-1} a_jj!\right], \qquad k = 1, 2, \cdots, \beta+1.\end{aligned} \tag{8.15}$$

REFERENCES

[1] Y. H. CHIU, *An integral equation method of solution of $\Delta u + k^2 u = 0$ in the exterior of a bounded domain*, Ph.D. thesis, University of Utah, Salt Lake City, UT, 1976.

[2] R. P. GILBERT, *Function Theoretic Methods in Partial Differential Equations*, Academic Press, New York, 1969.

[3] Y. IKEBE, T. Y. LI AND F. STENGER, *The numerical solution of the Hilbert problem*, Theory of Approximation with Applications, A. G. Law and B. N. Sahney, eds., Academic Press, New York, 1976.

[4] L. LUNDIN, *A cardinal function method of solution of $\Delta u = u - u^3$*, submitted for publication.

[5] J. MCNAMEE, F. STENGER AND E. L. WHITNEY, *Whittaker's cardinal function in retrospect*, Math. Comput., 25 (1971), pp. 141–154.

[6] W. PETRICK, J. SCHWING AND F. STENGER, *Algorithm for the electromagnetic scattering from rotationally symmetric bodies* (presented at the SIAM fall meeting in San Francisco (1975)), J. Math. Anal. Appl., to appear.

[7] J. SCHWING, *Eigensolutions of potential theory problems in R^3*, Ph.D. thesis, University of Utah, Salt Lake City, UT, 1976.

[8] S. L. SOBOLEV, *Convergence of approximate integration formulas for functions from $L_2^{(m)}$*, Soviet Math. Dokl., 6 (1965), pp. 865–867.

[9] F. STENGER, *Integration formulas based on the trapezoidal formula*, J. Inst. Math., 12 (1973), pp. 103–114.

[10] ———, *Approximation via Whittaker's cardinal function*, J. Approximation Theory, 17 (1976), pp. 222–240.

[11] ———, *Optimal convergence of minimum norm approximations*, Numer. Math., to appear.

[12] ———, *An analytic function which is an approximate characteristic function*, SIAM J. Numer. Anal., 12 (1975), pp. 239–254.

[13] ———, *Approximate solution of convolution-type integral equations*, this Journal, 4 (1973), pp. 536–555.

SIAM. J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0017 $01.00/0

BOUNDARY VALUE PROBLEMS OF MIXED TYPE ARISING IN THE KINETIC THEORY OF GASES*

HANS G. KAPER†

Abstract. The linear integro-differential equation

$$\frac{\partial xu}{\partial t}(t, x)+u(t, x)-\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-y^2)u(t, y)\, dy = f(t, x),$$

$t \in (0, \tau)$, $x \in \mathbb{R}$, is interpreted in functional form as a relation between the mappings $f: [0, \tau] \to L_2(\mathbb{R}, \mu)$ and $u: [0, \tau] \to L_2(\mathbb{R}, \mu)$ in a weighted Hilbert space $L_2(\mathbb{R}, \mu)$. It is shown that this equation, together with the appropriate boundary conditions, defines a well-posed problem in $L_2(\mathbb{R}, \mu)$.

1. Introduction. In this article we are concerned with boundary value problems described by the linear integro-differential equation

$$\frac{\partial xu}{\partial t}(t, x)+u(t, x)-\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-y^2)u(t, y)\, dy = f(t, x), \tag{1.1}$$

$t \in (0, \tau)$, $x \in \mathbb{R}$. This equation arises in the kinetic theory of gases. It is a linearized form of the stationary one-dimensional Boltzmann equation, in which the collision operator is represented by the BGK-model; cf. [1], [2]. The unknown function u represents the perturbation of the molecular distribution function in phase space relative to an absolute Maxwellian. In the present case, the phase space is two-dimensional; u depends on one position coordinate (t) and the t-component of the velocity vector (x). The function f is supposed to be given.

Equation (1.1) is of mixed type in the sense that although it is parabolic for $x \neq 0$, the preferred t-direction is towards increasing values of t for $x > 0$, towards decreasing values of t for $x < 0$. This property of the equation is reflected in the boundary conditions that are normally specified, viz. $u(0, x)$ given for $x > 0$, $u(\tau, x)$ given for $x < 0$.

Equations similar to (1.1) arise in the theories of neutron transport and radiative transfer, where the unknown function u is defined on the domain $[0, \tau] \times [-1, 1]$, rather than $[0, \tau] \times \mathbb{R}$. In his thesis [3], Hangelbroek presented a functional-analytic study of boundary value problems associated with the neutron transport equation. His results were subsequently extended by Lekkerkerker [4] to include the case of neutron transport in a conservative medium, which is most relevant in the context of radiative transfer theory. Recently, Beals [5] presented a more abstract approach to this class of boundary value problems.

Utilizing the methods of Hangelbroek, Lekkerkerker and Beals, we present, in this article, a functional-analytic approach to boundary value problems described by (1.1). Compared to [3] and [4], two new features arise. One is the infinite range of the variable x, which implies that the multiplication operator T ("multiplication with the function x") is unbounded. Another is the nature of the collision operator A which is a projection operator. A similar situation occurred in [4], but there the resulting point spectrum of the relevant operator (AT^{-1}) was separated from the continuous spectrum. In the present situation, the point spectrum is imbedded in the continuous spectrum and a different procedure for singling out the eigenspace is called for. The

* Received by the editors October 15, 1976, and in final revised form September 30, 1977.

† Applied Mathematics Division, Argonne National Laboratory, Argonne, Illinois 60439. This work was supported by the U.S. Energy Research and Development Administration.

reader will also observe that we have adopted the operator AT^{-1} (in our notation) as the fundamental operator, rather than $T^{-1}A$, as was done in [4]. This choice is natural if one writes the streaming term in Boltzmann's equation in divergence form, as we prefer to do and as we have done in (1.1). Taking the variable x from under the differentiation symbol $\partial/\partial t$ would imply that we solve a stronger form of the transport equation then is strictly justifiable.

In § 2 we give an abstract statement of a boundary value problem for (1.1). In § 3 we analyze the operator AT^{-1} in some detail. In § 4 we prove the existence of a diagonalizing operator for AT^{-1} (some of these results were announced previously in [6].) In § 5 we establish the existence and uniqueness of a solution for the abstract boundary value problem. In the final § 6 we discuss some aspects involved in the actual construction of the solution of a given boundary value problem. These aspects are covered in detail in two forthcoming articles [7], [8], in which we also relate our results to those obtained earlier by Cercignani in his analysis of the slip flow problem via the method of "singular eigenfunction expansions" [9].

Notation. The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. Moreover, $\mathbb{R}_+ = \{x \in \mathbb{R}: x \geqq 0\}$ and $\mathbb{R}_- = \{x \in \mathbb{R}: x \leqq 0\}$. The identity operator in any linear space is denoted by I. The domain, range and kernel of an operator A in a Banach space X are denoted by $D(A)$, $R(A)$ and $N(A)$, respectively. The resolvent set $\rho(A)$ and the spectrum $\sigma(A)$ of A are defined as in [10, § III.6.1]. For $z \in \rho(A)$, the symbol $R(z; A)$ is used for the resolvent operator $(zI - A)^{-1}$. A point $z \in \sigma(A)$ is said to belong to the point spectrum $\sigma_p(A)$ of A if the equation $(zI - A)f = 0$ has a nontrivial solution $f \in D(A)$; it is said to belong to the continuous spectrum $\sigma_c(A)$ of A if $R(zI - A)$ is dense in X, $zI - A$ is invertible, but $(zI - A)^{-1}$ is unbounded. The restriction of the operator A to a subspace M of X is denoted by $A|M$.

2. Statement of the problem. Let $L_2(\mathbb{R}, \mu)$ denote the Hilbert space of (equivalence classes of) all complex-valued functions defined on $\mathbb{R}$, whose moduli are square integrable with respect to the (finite) measure μ,

$$\mu(dx) = \pi^{-1/2} \exp(-x^2)\, dx. \tag{2.1}$$

We denote the inner product and norm in $L_2(\mathbb{R}, \mu)$ by $(\cdot, \cdot)_\mu$ and $\|\cdot\|_\mu$, respectively,

$$(f, g)_\mu = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\mu(dx) \qquad f, g \in L_2(\mathbb{R}, \mu), \tag{2.2}$$

$$\|f\|_\mu = (f, f)_\mu^{1/2} \qquad f \in L_2(\mathbb{R}, \mu). \tag{2.3}$$

The symbol e denotes the element of $L_2(\mathbb{R}, \mu)$ defined by

$$e(x) = 1 \qquad x \in \mathbb{R}. \tag{2.4}$$

We define the multiplication operator T in $L_2(\mathbb{R}, \mu)$ by the expression

$$Tf(x) = xf(x) \qquad x \in \mathbb{R}, \quad f \in D(T) \tag{2.5}$$

on the domain $D(T) = \{f \in L_2(\mathbb{R}, \mu): \int_{\mathbb{R}} |xf(x)|^2 \mu(dx) < \infty\}$. Its inverse T^{-1} is defined on $D(T^{-1}) = R(T)$, such that $T^{-1}g = f$ if $g = Tf$. Both T and T^{-1} are unbounded and selfadjoint. Furthermore, we define the operator A in $L_2(\mathbb{R}, \mu)$,

$$Af = f - (f, e)_\mu e \qquad f \in L_2(\mathbb{R}, \mu). \tag{2.6}$$

A is a bounded projection, $N(A) = \mathrm{sp}\,(e)$.

We interpret (1.1) as an equation in $L_2(\mathbb{R}, \mu)$,

$$(Tu)'(t)+Au(t)=f(t) \qquad t\in(0,\tau), \tag{2.7a}$$

i.e., as a relation between $f:[0,\tau]\to L_2(\mathbb{R},\mu)$ and $u:[0,\tau]\to D(T)\subset L_2(\mathbb{R},\mu)$, such that $Tu\in C^1(0,\tau;L_2(\mathbb{R},\mu))$. Here, $'$ denotes differentiation with respect to t, i.e., $(Tu)'(t)=\lim_{h\to 0}(Tu(t+h)-Tu(t))/h$, the limit being taken in $L_2(\mathbb{R},\mu)$.

Let $L_2(\mathbb{R}_-,\mu)$ and $L_2(\mathbb{R}_+,\mu)$ be the linear vector spaces of (equivalence classes of) all complex-valued functions defined on $\mathbb{R}_-$ and $\mathbb{R}_+$, respectively, whose moduli are square integrable with respect to the measure μ. We identify these vector spaces with the closed subspaces of $L_2(\mathbb{R},\mu)$ which consist of those elements that vanish on $\mathbb{R}_+$ and $\mathbb{R}_-$, respectively. Thus, they are Banach spaces with respect to the norm $\|\cdot\|_\mu$. We decompose $L_2(\mathbb{R},\mu)$ into a topological direct sum,

$$L_2(\mathbb{R},\mu)=L_2(\mathbb{R}_-,\mu)\oplus L_2(\mathbb{R}_+,\mu)$$

and denote the projection operator which maps $L_2(\mathbb{R},\mu)$ onto $L_2(\mathbb{R}_\pm,\mu)$ along $L_2(\mathbb{R}_\mp,\mu)$ by $Q_\pm$. In the following analysis we show that a boundary value problem described by (2.7a) together with the boundary conditions

$$Q_+Tu(0)=Q_+g, \qquad Q_-Tu(\tau)=Q_-g, \qquad g\in D(T^{-1}) \tag{2.7b}$$

admits one and only one solution.

In terms of the function $v(t)=Tu(t)$, equation (2.7a) becomes

$$v'(t)+AT^{-1}v(t)=f(t) \qquad t\in(0,\tau),$$

which is an ordinary differential equation for v in $L_2(\mathbb{R},\mu)$. As its evolution is determined by the operator AT^{-1}, it is apparent that this operator is fundamental in the study of boundary value problems described by (1.1). In the following section we present its relevant properties.

3. The operator AT^{-1}. Consider the operator AT^{-1}. Its domain is $D(AT^{-1})=D(T^{-1})=R(T)$. Since $R(AT^{-1})\subset R(A)$, $(f,e)_\mu=0$ whenever $f\in R(AT^{-1})$.

AT^{-1} differs from T^{-1} by a degenerate operator of rank one. Hence, its spectrum is most easily studied by comparing it with the spectrum of T^{-1}. The latter coincides with the real axis and is a purely continuous spectrum. An important role is played by the function ω defined by the integral

$$\omega(z)=\int_{-\infty}^{\infty} x(x-z)^{-1}\mu(dx) \qquad z\in\mathbb{C}\backslash\mathbb{R}. \tag{3.1}$$

It has the following properties:

(3.2)
- (i) ω is analytic in the complex plane cut along the real axis;
- (ii) $\omega(z)\sim-(2z^2)^{-1}[1+(1\cdot 3)(2z^2)^{-1}+(1\cdot 3\cdot 5)(2z^2)^{-2}+\cdots]$ as $z\to\infty$ uniformly in $\arg z$, $0<|\arg z|<\pi$;
- (iii) $\omega(z)$ does not vanish in the cut complex plane ($\operatorname{Im}\omega(z)=0$ implies $\operatorname{Re}\omega(z)>0$);
- (iv) the limiting values $\omega^\pm(\lambda)=\lim_{\varepsilon\to 0+}\omega(\lambda\pm i\varepsilon)$, $\lambda\in\mathbb{R}$, exist and are given by

$$\omega^\pm(\lambda)=\omega(\lambda)\pm i\lambda\sqrt{\pi}\exp(-\lambda^2), \tag{3.3}$$

where we have defined

$$\omega(\lambda)=⨍_{-\infty}^{\infty} x(x-\lambda)^{-1}\mu(dx) \qquad \lambda\in\mathbb{R}. \tag{3.4}$$

Here, $⨍$ denotes the Cauchy principal value integral. The integral (3.4) can be evaluated in terms of Dawson's integral

$$D(\lambda)=\exp(-\lambda^2)\int_0^x \exp(x^2)\,dx \qquad \lambda\in\mathbb{R},$$

$$\omega(\lambda)=1-2\lambda D(\lambda); \tag{3.5}$$

see [11, § 7.1].

LEMMA 3.1. *$\sigma(AT^{-1})=\mathbb{R}$; for $z\in\rho(AT^{-1})$ we have*

$$R(z;AT^{-1})f=R(z;T^{-1})\left[f-\frac{1}{\omega(1/z)}(T^{-1}R(z;T^{-1})f,e)_\mu e\right], \tag{3.6}$$

where ω is defined by (3.1).

Proof. Suppose $z\in\rho(T^{-1})$ Consider the equation

$$(zI-AT^{-1})g=f \tag{3.7}$$

for some $g\in D(AT^{-1})$. It is equivalent to

$$(zI-T^{-1})g+(T^{-1}g,e)_\mu e=f.$$

Since $z\in\rho(T^{-1})$, $R(z;T^{-1})e\in D(T^{-1})$ and $g\in D(T^{-1})$, we may operate on both sides with $T^{-1}R(z;T^{-1})$. Then, taking inner products with e we obtain the identity

$$[1+(T^{-1}R(z;T^{-1})e,e)_\mu](T^{-1}g,e)_\mu=(T^{-1}R(z;T^{-1})f,e)_\mu.$$

The expression in brackets is readily evaluated. Its value is $\omega(1/z)$, which is nonzero for $z\in\rho(T^{-1})$, so we can solve for $(T^{-1}g,e)_\mu$. Thus,

$$g=R(z;T^{-1})f-[\omega(1/z)]^{-1}(T^{-1}R(z;T^{-1})f,e)_\mu R(z;T^{-1})e. \tag{3.8}$$

Hence, $z\in\rho(T^{-1})$ implies $z\in\rho(AT^{-1})$, so $\sigma(AT^{-1})\subset\sigma(T^{-1})$.

Next, suppose $\lambda\in\rho(AT^{-1})$ for some $\lambda\in\mathbb{R}=\sigma(T^{-1})$. For any $h\in D(T^{-1})$, the equation $(\lambda I-AT^{-1})f=(\lambda I-T^{-1})h$ has a solution $f_{\lambda,h}\in D(T^{-1})$. The element $g_{\lambda,h}=h-f_{\lambda,h}\in D(T^{-1})$ is such that

$$(\lambda I-T^{-1})g_{\lambda,h}=(T^{-1}(h-g_{\lambda,h}),e)_\mu e. \tag{3.9}$$

If $\lambda=0$, it follows from the identity (3.9) that $T^{-1}g_{\lambda,h}=(T^{-1}g_{\lambda,h},e)_\mu e-(T^{-1}h,e)_\mu e$ or, since $(e,e)_\mu=1$, that $(T^{-1}h,e)_\mu=0$. Suppose $\lambda\neq 0$ ($\lambda\in\mathbb{R}$). The right-hand side of the identity (3.9) represents a continuous function. A continuous function which is in the range of $\lambda I-T^{-1}$ must have a zero at $x=1/\lambda$. Hence, the constant-valued function $(T^{-1}(h-g_{\lambda,h}),e)_\mu e$ is identically equal to zero, i.e., $(T^{-1}(h-g_{\lambda,h}),e)_\mu=0$. Then it follows from (3.9) that $g_{\lambda,h}=0$ and, hence, $(T^{-1}h,e)_\mu=0$. However, here we have arrived at a contradiction: the element $h=Te$ clearly belongs to $D(T^{-1})$, yet $(T^{-1}h,e)_\mu=1\neq 0$. Hence, $\lambda\in\sigma(T^{-1})$ implies $\lambda\in\sigma(AT^{-1})$, so $\sigma(AT^{-1})\supset\sigma(T^{-1})$. This completes the proof of the first part of the lemma. The identity (3.6) follows immediately from (3.7) and (3.8). □

We observe that both A and T^{-1} are selfadjoint in $L_2(\mathbb{R},\mu)$. However, since A and T^{-1} do not commute, the product AT^{-1} is not selfadjoint. But, as we will show next, the nonselfadjointness can be localized in a finite-dimensional subspace of $L_2(\mathbb{R},\mu)$.

For $\lambda\in\sigma_p(AT^{-1})$, let $X_\lambda^{(i)}=N((\lambda I-AT^{-1})^i)$. If l is such that $X_\lambda^{(l-1)}\neq X_\lambda^{(l)}$ and $X_\lambda^{(i)}=X_\lambda^{(l)}$ for $i=l+1,l+2,\cdots$, we say that λ is an eigenvalue of AT^{-1} of rank l;

we call $X_\lambda^{(l)}$ the generalized eigensubspace of the operator AT^{-1} for the eigenvalue λ, and $\dim(X_\lambda^{(l)})$ the multiplicity of the eigenvalue λ.

LEMMA 3.2. *$\sigma_p(AT^{-1})=\{0\}$; 0 is a multiple eigenvalue of rank two and multiplicity two; the generalized eigensubspace is spanned by the vectors Te and T^2e.*

Proof. Suppose $(\lambda I - AT^{-1})f=0$ for some $\lambda\in\sigma(AT^{-1})$, $\lambda\neq 0$. Then $(\lambda I - T^{-1})f = -(T^{-1}f, e)_\mu e$. A continuous function which is in the range of $\lambda I - T^{-1}$ must have a zero at $x=1/\lambda$. Hence, the constant-valued function $(T^{-1}f, e)_\mu e$ is identically equal to zero, $(T^{-1}f, e)_\mu = 0$ and, consequently, $(\lambda I - T^{-1})f=0$. However, T^{-1} has a pure continuous spectrum, so $f=0$ and no $\lambda\in\sigma(AT^{-1})$ with $\lambda\neq 0$ can be an eigenvalue of AT^{-1}. For $\lambda=0$, the eigenvalue equation $(\lambda I - AT^{-1})f=0$ has a nontrivial solution, $f=Te$, so $\sigma_p(AT^{-1})$ consists of the single point $\lambda=0$.

A is a projection operator with $N(A)=\mathrm{sp}\,(e)$. Hence, any solution of the equation $AT^{-1}f=0$ is necessarily a constant multiple of Te, so $X_0^{(1)}=\mathrm{sp}\,(Te)$. It follows that $(AT^{-1})^2f = AT^{-1}(AT^{-1}f)=0$ if and only if $AT^{-1}f=\alpha Te$ for some $\alpha\in\mathbb{C}$, which in turn, implies $f=\alpha T^2e+\beta Te$ for some $\beta\in\mathbb{C}$, so $X_0^{(2)}=\mathrm{sp}\,(Te, T^2e)$. Suppose now that f satisfies the equation $(AT^{-1})^3f=0$. Then $AT^{-1}f\in X_0^{(2)}$, so $AT^{-1}f=\alpha Te+\beta T^2e$ for some pair $\alpha,\beta\in\mathbb{C}$. A necessary condition for $\alpha Te+\beta T^2e$ to be in the range of A is that its inner product with e be zero. Hence, $\beta=0$ and we obtain, as before, that f must be a linear combination of Te and T^2e. Thus, $X_0^{(3)}=X_0^{(2)}$ and also $X_0^{(i)}=X_0^{(2)}$ for $i=3,4,\cdots$. This proves that the eigenvalue 0 has rank two. Since $X_0^{(2)}$ is spanned by the two linearly independent vectors Te and T^2e, its multiplicity is also equal to two. $\square$

On the basis of Lemmas 3.1 and 3.2 we introduce a direct sum decomposition of the Hilbert space $L_2(\mathbb{R},\mu)$, viz., $L_2(\mathbb{R},\mu)=N((AT^{-1})^2)\oplus\overline{R((AT^{-1})^2)}$. Since $D(AT^{-1})$ is dense in $L_2(\mathbb{R},\mu)$ and both A and T^{-1} are selfadjoint in $L_2(\mathbb{R},\mu)$, we have $(AT^{-1})^*=T^{-1}A$ and, therefore, $\overline{R((AT^{-1})^2)}=N((AT^{-1})^{*2})^\perp=N((T^{-1}A)^2)^\perp$; see [10, § III.5.5]. It is easily verified that $N((T^{-1}A)^2)=\mathrm{sp}\,(e, Te)$. We put

$$G=\mathrm{sp}\,(Te, T^2e), \tag{3.10}$$

$$H=\{f\in L_2(\mathbb{R},\mu):(f,e)_\mu=(f,Te)_\mu=0\}. \tag{3.11}$$

G and H are closed linear manifolds in $L_2(\mathbb{R},\mu)$. The (nonorthogonal) projection P which maps $L_2(\mathbb{R},\mu)$ onto H along G is given by

$$Pf=f-2(f,Te)_\mu Te-2(f,e)_\mu T^2e \qquad f\in L_2(\mathbb{R},\mu). \tag{3.12}$$

THEOREM 3.1. (i) *$L_2(\mathbb{R},\mu)=G\oplus H$ and the pair $\{G,H\}$ reduces AT^{-1};*

(ii) *$AT^{-1}|G$ is defined on G and has the representation $\begin{pmatrix}0&1\\0&0\end{pmatrix}$ relative to the basis (Te, T^2e) of G;*

(iii) *$AT^{-1}|H$ is invertible and $(AT^{-1}|H)^{-1}=PT|H$.*

Proof. (i) The first part of the theorem follows immediately from the way the manifolds G and H have been introduced. We recall that AT^{-1} is said to be reduced by the pair $\{G,H\}$ if each of the subspaces G, H is invariant under AT^{-1} and if, moreover, $PD(AT^{-1})\subset D(AT^{-1})$; see [10, § III.5.6]. The latter property is readily verified from (3.12).

(ii) Any $f\in G$ can be represented in the form $f=\alpha Te+\beta T^2e$ for some pair $\alpha,\beta\in\mathbb{C}$. Obviously, $f\in D(AT^{-1})$ and $AT^{-1}f=\alpha Ae+\beta ATe=\beta Te$.

(iii) Let $f\in D(T^{-1})\cap H$. Then, $PTAT^{-1}f=Pf-(T^{-1}f,e)_\mu PTe$. Since $f\in H$, we have $Pf=f$. Furthermore, $Te\in G$, so $PTe=0$. Hence, $PTAT^{-1}f=f$ for all $f\in D(T^{-1})\cap H$. Next, let $f\in D(T)\cap H$. Then $AT^{-1}PTf=Af-2(Tf,Te)_\mu Ae$. Since $f\in H$, we have $(f, e)_\mu = 0$, so $Af = f$. Furthermore, $Ae = 0$. Hence, $AT^{-1}PTf = f$ for all $f \in D(T) \cap H$. Combining these two results we see that $AT^{-1}|H$ and $PT|H$ are each other's inverse. □

In conclusion, we observe that the operator $PT|H$ coincides with the operator PTP on H. The domain of the latter operator is $D(PTP) = D(T)$, since $G \subset D(T)$. For an operator of the form PTP where T is closed and P is a projection whose null-space is finite-dimensional and contained in $D(T)$, one can define the Weinstein–Aronszajn determinant of the second kind, $\omega_P(z; T)$; see [10, § IV.6.1]. It is readily verified that, in our case, $\omega_P(z; T) = -2\omega(z)$, where ω is defined by (3.1). Hence, apart from a numerical factor, the function ω is the W.A.-determinant of the second kind associated with the operators T and P. We will further elaborate upon this observation in § 6 and in our subsequent article [8].

4. The isomorphism $\hat{F}: L_2(\mathbb{R}, \mu) \to \mathbb{C}^2 \oplus L_2(\mathbb{R}, \nu)$. In the previous section we obtained a reduction of the operator AT^{-1} by introducing the pair of manifolds $\{G, H\}$. The action of AT^{-1} on G is completely described by the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. In the present section we devote ourselves to a study of the restriction $AT^{-1}|H$, or rather to a study of its inverse which we denote by B,

$$B = (AT^{-1}|H)^{-1}. \tag{4.1}$$

From Theorem 3.1 we know that $B = PT|H$ and $D(B) = D(T) \cap H$. The operator B differs from $T|H$ by a degenerate operator of rank one,

$$Bf = Tf - 2(Tf, Te)_\mu Te \qquad f \in D(B) = D(T) \cap H. \tag{4.2}$$

Let $P_H(\mathbb{R})$ denote the set of all polynomial functions on $\mathbb{R}$ which belong to H.

LEMMA 4.1. *$P_H(\mathbb{R})$ is dense in H.*

Proof. Take any $f \in H$. Then also $f \in L_2(\mathbb{R}, \mu)$. The Hermite polynomials span a dense subset in $L_2(\mathbb{R}, \mu)$; see [12, § 5.7]. Hence, there exists a sequence of polynomials $\{p_n : n = 1, 2, \cdots\}$ on $\mathbb{R}$ such that $\|p_n - f\|_\mu \to 0$ as $n \to \infty$. But $Pf = f$, so $\|Pp_n - f\|_\mu = \|P(p_n - f)\|_\mu \leqq \|P\| \, \|p_n - f\|_\mu$ and, consequently, $\|Pp_n - f\|_\mu \to 0$ as $n \to \infty$. Since $Pp_n \in P_H(\mathbb{R})$ for $n = 1, 2, \cdots$, it follows that $P_H(\mathbb{R})$ is dense in H. □

From Lemma 4.1 we conclude that $D(B)$, which contains the set $P_H(\mathbb{R})$, is dense in H.

LEMMA 4.2. *B is selfadjoint in H, with $\sigma(B) = \mathbb{R}$.*

Proof. B maps H into itself. Since $D(B)$ is dense in H, B^* is uniquely defined as an operator in H. For any $f, g \in D(B)$ we have $(Bf, g)_\mu = (Tf - 2(Tf, Te)_\mu Te, g)_\mu = (Tf, g)_\mu = (f, Tg)_\mu = (f, Tg - 2(Tg, Te)_\mu Te) = (f, Bg)_\mu$, so B is symmetric and $D(B) \subset D(B^*)$. To show that B is selfadjoint it suffices to show that $D(B^*) \subset D(B)$. Suppose that $g \in D(B^*)$. The mapping $f \mapsto (Bf, g)$ defines a continuous linear functional on $D(B)$. Since $D(T) = G \oplus D(B)$ and G is finite dimensional, we can use the identity $(Bf, g)_\mu = (Tf, g)_\mu$, which holds for any pair $f \in D(B)$, $g \in H$, to extend the functional $f \mapsto (Bf, g)_\mu$ to a continuous linear functional on $D(T)$. Hence, $g \in D(T^*) \cap H = D(T) \cap H = D(B)$ and $(Tf, g)_\mu = (f, Tg)_\mu = (Pf, Tg)_\mu = (f, PTg)_\mu$, i.e., $D(B^*) \subset D(B)$ and $B^* = PT|H$, which proves the first part of the lemma.

Consider the operator $B^{-1} = AT^{-1}|H$. Since AT^{-1} is reduced by $\{G, H\}$, $\sigma(B^{-1}) = \sigma(AT^{-1}) \backslash \sigma(AT^{-1}|G) = \mathbb{R}\backslash\{0\} = \mathbb{R}$. B^{-1} is a closed invertible operator in H, so its extended spectrum $\tilde{\sigma}(B^{-1})$ and the extended spectrum $\tilde{\sigma}(B)$ of its inverse are mapped onto each other by the mapping $z \mapsto z^{-1}$ of the extended complex plane; see [10, § III.6.3]. Consequently, $\tilde{\sigma}(B) = \mathbb{R} \cup \{\infty\}$ and therefore $\sigma(B) = \mathbb{R}$. □

Since B is selfadjoint in H it follows from Von Neumann's general spectral theorem—cf. [13, Thm. XII.2.3]—that there exists a uniquely determined resolution of the identity E for B with support $\sigma(B)=\mathbb{R}$ in terms of which we can develop an operational calculus. In doing so we will adhere to the convention that a function f defined on $\mathbb{R}$ is denoted by $\tilde{f}$ whenever $\mathbb{R}$ is considered as $\sigma(B)$.

Let $P(\mathbb{R})$ be the vector space of all polynomial functions defined on $\mathbb{R}$. With each element $\tilde{p}\in P(\mathbb{R})$ there corresponds a unique operator $\tilde{p}(B)$ such that $\tilde{p}(B)=\alpha_0 I+\alpha_1 B+\cdots+\alpha_n B^n$ if $\tilde{p}(\lambda)=\alpha_0+\alpha_1\lambda+\cdots+\alpha_n\lambda^n$; see [13, Corollary XII.2.8]. The domain of $\tilde{p}(B)$ is $D(\tilde{p}(B))=D(B^n)$. Since $P_H(\mathbb{R})\subset D(B^n)$ and $P_H(\mathbb{R})$ is dense in H (see Lemma 4.1), $D(\tilde{p}(B))$ is dense in H for any polynomial operator $\tilde{p}(B)$. We let $\mathcal{P}=\{\tilde{p}(B): \tilde{p}\in P(\mathbb{R})\}$.

Next, let e_H be the projection in H of the element $e\in L_2(\mathbb{R},\mu)$,

$$e_H=Pe=e-2T^2e. \tag{4.3}$$

Clearly, $e_H\in D(\tilde{p}(B))$ for all $\tilde{p}(B)\in\mathcal{P}$. Let $\mathcal{P}e_H$ denote the vector space of functions defined on $\mathbb{R}$ which can be written as $\tilde{p}(B)e_H$ for some $\tilde{p}(B)\in\mathcal{P}$. The following lemma gives a characterization of this vector space.

LEMMA 4.3. *If $\tilde{p}\in P(\mathbb{R})$ is a polynomial of degree n $(n\geqq 0)$ on $\mathbb{R}$, then $\tilde{p}(B)e_H$ is a polynomial of degree $n+2$ on $\mathbb{R}$ and $\tilde{p}(B)e_H\in H$; furthermore, $\mathcal{P}e_H=P_H(\mathbb{R})$.*

Proof. First, we show that B^ne_H $(n\geqq 0)$ is a polynomial of degree $n+2$ on $\mathbb{R}$. We use induction on n. For $n=0$, the statement is true. Suppose the statement is true for $n=k-1$. Then $B^ke_H=Bp_{k+1}$, where p_{k+1} is a polynomial of degree $k+1$ on $\mathbb{R}$. Using (4.2) we find that $B^ke_H=Tp_{k+1}-2(Tp_{k+1},Te)_\mu Te$, which is a polynomial of degree $k+2$ on $\mathbb{R}$. Hence, the statement is true for $n=k$. By induction, it is then true for all $n\geqq 0$. Thus, generally, if $\tilde{p}\in P(\mathbb{R})$ is a polynomial of degree n, then $\tilde{p}(B)e_H$ is a polynomial of degree $n+2$. Moreover, since $e_H\in H$ and B is a mapping of H into itself, $\tilde{p}(B)e_H\in H$. Denote by $P_{H,n}(\mathbb{R})$ the subspace of polynomials in $P_H(\mathbb{R})$ of degree $\leqq n$. Then $\dim P_{H,2}(\mathbb{R})=1$. Since $P_{H,n}(\mathbb{R})$ contains a polynomial of degree n for all $n\geqq 2$, $\dim P_{H,n+1}(\mathbb{R})=\dim P_{H,n}(\mathbb{R})+1$ $(n\geqq 2)$. Hence, $\dim P_{H,n}(\mathbb{R})=n-1$. It follows that the mapping $\tilde{p}\mapsto\tilde{p}(B)e_H$, which is linear and injective by the result above, is a mapping onto $P_H(\mathbb{R})$. $\square$

The mapping $\tilde{p}(B)\leftrightarrow\tilde{p}(B)e_H$ defines a one-to-one relationship between the elements of $\mathcal{P}$ and the elements of $P_H(\mathbb{R})$. Combining this result with the remarks made in the discussion preceding Lemma 4.3 we obtain the result expressed in the following theorem.

THEOREM 4.1. *There exists a linear isomorphism $F: P_H(\mathbb{R})\to P(\mathbb{R})$ defined by*

$$Fp=\tilde{p}\quad \textit{for}\ p=\tilde{p}(B)e_H. \tag{4.4}$$

Under the isomorphism a polynomial of degree $n+2$ in $P_H(\mathbb{R})$ is mapped onto a polynomial of degree n in $P(\mathbb{R})$. In particular, $Fe_H=\tilde{e}$, where $\tilde{e}\in P(\mathbb{R})$ denotes the polynomial function that is identically equal to one on $\mathbb{R}$.

Next, we introduce the finite positive Borel measure ν on $\mathbb{R}$,

$$\nu(d\lambda)=\|E(d\lambda)e_H\|_\mu^2. \tag{4.5}$$

It is actually possible to give an explicit representation of this measure. To this end we first establish the following lemma.

LEMMA 4.4. *For $z\in\mathbb{C}\backslash\mathbb{R}$ we have*

$$R(z;B)f=R(z;T)\{f+[\omega(z)]^{-1}(R(z;T)f,e)_\mu Te\}.$$

Proof. Take $g \in D(B)$ and let $f=(zI-B)g$. Since $Bg=Tg-2(Tg, Te)_\mu Te$ we have $f=(zI-T)g+2(Tg, Te)_\mu Te$. For $z \in \mathbb{C}\backslash\mathbb{R}$, the resolvent $R(z;T)=(zI-T)^{-1}$ exists and is bounded in $L_2(\mathbb{R}, \mu)$, so g must be of the form $g=R(z;T)f+\alpha_z R(z;T)Te$ for some $\alpha_z \in \mathbb{C}$. Since $g \in H$, we have $(g, e)_\mu=0$, i.e., $(R(z;T)f+\alpha_z R(z;T)Te, e)_\mu=(R(z;T)f, e)_\mu-\alpha_z\omega(z)=0$, so $\alpha_z=[\omega(z)]^{-1}(R(z;T)f, e)_\mu$. □

We now prove the following result concerning the measure ν.

LEMMA 4.5. *The measure ν defined by* (4.5) *is such that*

$$\int_a^b \nu(d\lambda)=\int_a^b \frac{\mu(d\lambda)}{\omega^+\omega^-(\lambda)}$$

for any interval $(a, b)\subset\mathbb{R}$, where the functions ω^+, ω^- are defined by (3.3).

Proof. First we establish the identity for an arbitrary finite interval (a, b). In the strong operator topology we have the result

$$E((a,b))=\lim_{\delta\to 0+}\lim_{\varepsilon\to 0+}\frac{1}{2\pi i}\int_{a+\delta}^{b-\delta}[R(\lambda-i\varepsilon;B)-R(\lambda+i\varepsilon;B)]\,d\lambda;$$

see [13, Thm. XII.2.10]. Hence, in the topology of $L_2(\mathbb{R}, \mu)$,

$$E((a,b))e_H=\lim_{\delta\to 0+}\lim_{\varepsilon\to 0+}\frac{1}{2\pi i}\int_{a+\delta}^{b-\delta}[R(\lambda-i\varepsilon;B)e_H-R(\lambda+i\varepsilon;B)e_H]\,d\lambda.$$

Consequently,

$$\begin{aligned}\int_a^b \nu(d\lambda)&=(E((a,b))e_H, e_H)_\mu\\ &=\lim_{\delta\to 0+}\lim_{\varepsilon\to 0+}\frac{1}{2\pi i}\int_{a+\delta}^{b-\delta}[(R(\lambda-i\varepsilon;B)e_H, e_H)_\mu-(R(\lambda+i\varepsilon;B)e_H, e_H)_\mu]\,d\lambda,\end{aligned}$$

since taking the inner product $(\cdot, e_H)_\mu$ is a continuous operation in H. Using lemma 4.4 one finds

$$(R(z;B)e_H, e_H)_\mu=\frac{1-\omega(z)}{z\omega(z)}+2z \qquad z\in\mathbb{C}\backslash\mathbb{R}.$$

The limiting values $\omega^\pm(\lambda)=\lim_{\varepsilon\to 0+}\omega(\lambda\pm i\varepsilon)$, $\lambda\in\mathbb{R}$, of $\omega(z)$ exist and are nonzero. Moreover, from the theory of Cauchy integrals we know that the convergence is uniform in λ on every compact subset of $\mathbb{R}$; see [14, §§ 15, 16]. Hence,

$$\begin{aligned}&\lim_{\varepsilon\to 0+}[(R(\lambda-i\varepsilon;B)e_H, e_H)_\mu-(R(\lambda+i\varepsilon;B)e_H, e_H)_\mu]\\ &\qquad=\frac{\omega^+(\lambda)-\omega^-(\lambda)}{\lambda(\omega^+\omega^-)(\lambda)}=\frac{2\pi i}{\omega^+\omega^-(\lambda)} \qquad \lambda\in\mathbb{R},\end{aligned}$$

uniformly on compact subsets of $\mathbb{R}$. Consequently,

$$\int_a^b \nu(d\lambda)=\lim_{\delta\to 0+}\int_{a+\delta}^{b-\delta}[(\omega^+\omega^-)(\lambda)]^{-1}\mu(d\lambda)=\int_a^b[(\omega^+\omega^-)(\lambda)]^{-1}\mu(d\lambda).$$

This proves the statement of the lemma for the case of a finite interval (a, b). Since the measure ν is finite, the extension to the case of an infinite interval ($a=-\infty$ and/or $b=\infty$) is straightforward.

We use the measure ν to define an inner product $(\cdot,\cdot)_\nu$ and a norm $\|\cdot\|_\nu$ in $P(\mathbb{R})$; thus,

$$(4.6)\qquad (\tilde{p}_1,\tilde{p}_2)=\int_{-\infty}^{\infty}\tilde{p}_1(\lambda)\overline{\tilde{p}_2(\lambda)}\nu(d\lambda)\qquad \tilde{p}_1,\tilde{p}_2\in P(\mathbb{R}),$$

$$(4.7)\qquad \|\tilde{p}\|_\nu=(\tilde{p},\tilde{p})_\nu^{1/2}\qquad \tilde{p}\in P(\mathbb{R}).$$

The map $F: P_H(\mathbb{R})\to P(\mathbb{R})$ is an isometry if $P_H(\mathbb{R})$ and $P(\mathbb{R})$ are endowed with the norms $\|\cdot\|_\mu$ and $\|\cdot\|_\nu$, respectively; cf. [13, Thm. XII.2.6].

Let $L_2(\mathbb{R},\nu)$ be the completion of $P(\mathbb{R})$ with respect to the norm $\|\cdot\|_\nu$. We define the multiplication operator $\tilde{T}$ in $L_2(\mathbb{R},\nu)$ by the expression

$$(4.8)\qquad \tilde{T}\tilde{f}(\lambda)=\lambda\tilde{f}(\lambda)\qquad \lambda\in\mathbb{R},\quad \tilde{f}\in D(\tilde{T})$$

on the domain $D(\tilde{T})$, where $D(\tilde{T})=\{\tilde{f}\in L_2(\mathbb{R},\nu): \int_{-\infty}^{\infty}|\lambda\tilde{f}(\lambda)|^2\nu(d\lambda)<\infty\}$. Its inverse $\tilde{T}^{-1}$ is defined on $D(\tilde{T})$, such that $\tilde{T}^{-1}\tilde{g}=\tilde{f}$ if $\tilde{g}=\tilde{T}\tilde{f}$. Both $\tilde{T}$ and $\tilde{T}^{-1}$ are unbounded, densely defined and selfadjoint in $L_2(\mathbb{R},\nu)$.

THEOREM 4.2. *The isomorphism F introduced in Theorem* 4.1 *can be extended uniquely to a unitary transformation from H onto $L_2(\mathbb{R},\nu)$. This extension, which we denote by the same symbol F, has the following properties*:

(i) *F maps $D(B)$ onto $D(\tilde{T})$, $FBf=\tilde{T}Ff$ for all $f\in D(B)$,*

(ii) *F maps $D(B^{-1})$ onto $D(\tilde{T}^{-1})$, $FB^{-1}f=\tilde{T}^{-1}Ff$ for all $f\in D(B^{-1})$.*

Proof. Since $P_H(\mathbb{R})$ is dense in H (relative to the norm $\|\cdot\|_\mu$) and $P(\mathbb{R})$ is dense in $L_2(\mathbb{R},\nu)$ (relative to the norm $\|\cdot\|_\nu$), the extension can be accomplished by continuity. As the proofs of properties (i) and (ii) are entirely analogous, we will give only the former. For any $f\in D(B)$ there exists a sequence of polynomials $\{\tilde{p}_n: n=1,2,\cdots\}$ in $P(\mathbb{R})$ such that $\|f-\tilde{p}_n(B)e_H\|_\mu\to 0$ and $\|Ff-\tilde{p}_n\|_\nu\to 0$ as $n\to\infty$. Given such a sequence we have, for any $\tilde{q}\in P(\mathbb{R})$, $(\tilde{q},FBf)_\nu=(\tilde{q}(B)e_H,Bf)_\mu=(B\tilde{q}(B)e_H,f)_\mu=\lim_{n\to\infty}(\tilde{T}\tilde{q},\tilde{p}_n)_\nu=\lim_{n\to\infty}(\tilde{q},\tilde{T}\tilde{p}_n)_\nu$. Since $P(\mathbb{R})$ is dense in $L_2(\mathbb{R},\nu)$, it follows that $\{\tilde{T}\tilde{p}_n: n=1,2,\cdots\}$ is a Cauchy sequence in $L_2(\mathbb{R},\nu)$. But $\tilde{T}$ is closed and $\tilde{p}_n\to Ff$ in $L_2(\mathbb{R},\nu)$, so $Ff\in D(\tilde{T})$ and $\tilde{T}\tilde{p}_n\to\tilde{T}Ff$ in $L_2(\mathbb{R},\nu)$. Consequently, $(\tilde{q},FBf)_\nu=(\tilde{q},\tilde{T}Ff)_\nu$ for any $\tilde{q}\in P(\mathbb{R})$. Again using the fact that $P(\mathbb{R})$ is dense in $L_2(\mathbb{R},\nu)$ we conclude that $FBf=\tilde{T}Ff$. $\square$

Because of the properties (i) and (ii) we say that the transformation F diagonalizes the operator B. Theorem 4.2 is actually a special instance of the spectral theorem for unbounded selfadjoint operators in a Hilbert space; cf. [13, Thm. XII.3.5].

By putting

$$(4.9)\qquad F_0f=FPf\qquad f\in L_2(\mathbb{R},\mu)$$

we obtain an extension F_0 of F which is defined on all of $L_2(\mathbb{R},\mu)$. F_0 has the property that it maps the element $e\in L_2(\mathbb{R},\mu)$ onto the element $\tilde{e}\in L_2(\mathbb{R},\nu)$. Of course, the mapping F_0 thus defined is not injective. An extension which is injective is defined in the following theorem.

THEOREM 4.3. *The mapping $\hat{F}f: L_2(\mathbb{R},\mu)\to\mathbb{C}^2\oplus L_2(\mathbb{R},\nu)$ defined by*

$$(4.10)\qquad \hat{F}f=(2(f,Te)_\mu, 2(f,e)_\mu, FPf)\qquad f\in L_2(\mathbb{R},\mu)$$

is a linear isomorphism. For any $f\in D(T^{-1})$ we have

$$(4.11)\qquad \hat{F}AT^{-1}f=(2(f,e)_\mu, 0, \tilde{T}^{-1}FPf).$$

$\hat{F}$ is unitary if $L_2(\mathbb{R}, \mu)$ and $\mathbb{C}^2 \oplus L_2(\mathbb{R}, \nu)$ are endowed with the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, where

$$\|f\|_X^2 = \|(I-P)f\|_\mu^2 + \|Pf\|_\mu^2 \qquad f \in L_2(\mathbb{R}, \mu), \tag{4.12}$$

$$\|(a, b, \tilde{f})\|_Y^2 = \tfrac{1}{2}|a|^2 + \tfrac{3}{4}|b|^2 + \|\tilde{f}\|_\nu^2 \qquad (a, b, \tilde{f}) \in \mathbb{C}^2 \oplus L_2(\mathbb{R}, \nu). \tag{4.13}$$

Proof. The first part of the theorem follows from the direct sum decomposition $L_2(\mathbb{R}, \mu) = G \oplus H$ and the fact that F is an isomorphism from H onto $L_2(\mathbb{R}, \nu)$. To prove (4.11) we observe that $\hat{F}AT^{-1}f = \hat{F}AT^{-1}Pf + \hat{F}AT^{-1}(I-P)f = \hat{F}B^{-1}Pf + \hat{F}AT^{-1}(I-P)f = \hat{F}F^{-1}\tilde{T}^{-1}FPf + \hat{F}AT^{-1}(I-P)f = (2(f, e)_\mu,\ 0,\ \tilde{T}^{-1}FPf)$. Finally, for any $f \in L_2(\mathbb{R}, \mu)$ we have $\|\hat{F}f\|_Y^2 = \|(2(f, Te)_\mu,\ 2(f, e)_\mu,\ FPf)\|_Y^2 = 2|(f, Te)_\mu|^2 + 3|(f, e)_\mu|^2 + \|FPf\|_\nu^2 = \|I-P)f\|_\mu^2 + \|Pf\|_\mu^2 = \|f\|_X^2$. □

At the end of § 3 we already observed that the operator $B = PT|H$ coincides with the operator PTP on H. The maximal domain of PTP in $L_2(\mathbb{R}, \mu)$ coincides with $D(T)$. Hence, by putting

$$B_0 f = BPf \qquad f \in D(T) \subset L_2(\mathbb{R}, \mu) \tag{4.14}$$

we obtain an extension B_0 of B which is defined on all of $D(T)$. It is such that $FB_0 f = FBPf = \tilde{T}FPf = \tilde{T}F_0 f$ for any $f \in D(T)$.

5. Existence and uniqueness theory. We now turn to a discussion of the boundary value problem (2.7). First, we treat the homogeneous equation

$$(Tu)'(t) + Au(t) = 0 \qquad t \in (0, \tau) \tag{5.1}$$

for $u: [0, \tau] \to D(T) \subset L_2(\mathbb{R}, \mu)$, $Tu \in C^1(0, \tau;\ L_2(\mathbb{R}, \mu))$, subject to the boundary conditions

$$Q_+Tu(0) = Q_+g, \qquad Q_-Tu(\tau) = Q_-g \qquad g \in D(T^{-1}). \tag{5.2}$$

In terms of the function $v(t) = Tu(t)$, equations (5.1) and (5.2) become

$$v'(t) + AT^{-1}v(t) = 0 \qquad t \in (0, \tau), \tag{5.3}$$

$$Q_+v(0) = Q_+g, \qquad Q_-v(\tau) = Q_-g \qquad g \in D(T^{-1}). \tag{5.4}$$

Application of the transformation $\hat{F}$, which is linear and continuous in $L_2(\mathbb{R}, \mu)$ and, hence, commutes with the differentiation operator $(')$ leads to the following set of differential equations in $\mathbb{C}^2 \oplus L_2(\mathbb{R}, \nu)$,

$$m_1'(t) + m_0(t) = 0, \qquad m_0'(t) = 0, \tag{5.5}$$

$$\tilde{v}'(t) + \tilde{T}^{-1}\tilde{v}(t) = 0, \tag{5.6}$$

for the two scalar-valued functions $m_0, m_1 \in C^1([0, \tau])$ and the vector-valued function $\tilde{v}: [0, \tau] \to D(\tilde{T}^{-1}) \subset L_2(\mathbb{R}, \nu)$, $\tilde{v} \in C^1(0, \tau; L_2(\mathbb{R}, \nu))$, where $m_i(t) = (v(t), T^i e)_\mu$, $i = 0, 1$, and $\tilde{v}(t) = FPv(t)$. The equations (5.5) are satisfied by

$$m_0(t) = \beta, \qquad m_1(t) = \alpha + \beta(\tfrac{1}{2}\tau - t) \qquad t \in [0, \tau], \tag{5.7}$$

where $\alpha, \beta \in \mathbb{C}$ are arbitrary constants. (The constant $\frac{1}{2}\tau$ has been inserted for later convenience.) We use semigroup theory to solve the differential equation (5.6).

To this end we introduce the spaces $L_2(\mathbb{R}_-, \nu)$ and $L_2(\mathbb{R}_+, \nu)$ in the same way as we introduced the spaces $L_2(\mathbb{R}_-, \mu)$ and $L_2(\mathbb{R}_+, \mu)$ earlier. We decompose $L_2(\mathbb{R}, \nu)$,

$$L_2(\mathbb{R}, \nu) = L_2(\mathbb{R}_-, \nu) \oplus L_2(\mathbb{R}_+, \nu) \tag{5.8}$$

and let $\tilde{Q}_\pm$ be the projection operators which map $L_2(\mathbb{R}, \nu)$ onto $L_2(\mathbb{R}_\pm, \nu)$ along

$L_2(\mathbb{R}_\mp, \nu)$. Since $\tilde{T}^{-1}$ is reduced by $Q_\pm$, equation (5.6) is solved in $L_2(\mathbb{R}, \nu)$ by a function $\tilde{v} = \tilde{v}_- + \tilde{v}_+$, $\tilde{v}_\pm \in L_2(\mathbb{R}_\pm, \nu)$, if and only if it is solved in $L_2(\mathbb{R}_-, \nu)$ by $\tilde{v}_-$ and in $L_2(\mathbb{R}_+, \nu)$ by $\tilde{v}_+$.

LEMMA 5.1. *The operator* $-\tilde{T}^{-1}$ *is the infinitesimal generator of a holomorphic semigroup* $\{\exp(\mp t\tilde{T}^{-1})\colon \operatorname{Re} t > 0\}$ *in* $L_2(\mathbb{R}_\pm, \nu)$,

$$(\exp(\mp t\tilde{T}^{-1})\tilde{f}_\pm)(\lambda) = \exp(\mp t\lambda^{-1})\tilde{f}_\pm(\lambda) \qquad \nu\text{-a.e.}, \quad \lambda \in \mathbb{R}_\pm, \tag{5.9}$$

for every $\tilde{f}_\pm \in L_2(\mathbb{R}_\pm, \nu)$. *The operator* $\exp(\mp t\tilde{T}^{-1})$ *is uniformly bounded for* $|\arg t| \leqq \pi/2 - \varepsilon$ $(\varepsilon > 0)$ *and strongly continuous* (*within this smaller sector*) *at* $t = 0$ *with* $\exp(\mp 0 \cdot \tilde{T}^{-1}) = I$.

Proof. We prove the lemma only for the upper sign; the proof for the lower sign is entirely analogous. $\tilde{T}^{-1}$ is nonnegative and selfadjoint in $L_2(\mathbb{R}_+, \nu)$; hence $\exp(-t\tilde{T}^{-1})$ is holomorphic for $\operatorname{Re} t > 0$ and $\|\exp(-t\tilde{T}^{-1})\| \leqq 1$, cf. [10, § IX.1.6]. That $\exp(\mp t\tilde{T}^{-1})$ may be identified with the multiplication operator (5.9) follows from the fact that $\exp(-t\lambda^{-1})$, continuously extended to $\lambda = 0$, defines a continuous function of λ on $\mathbb{R}_+$. The remaining properties of $\exp(-t\tilde{T}^{-1})$ follow from the general theory; see [10, § IX.1.6]. □

It follows from Lemma 5.1 that any function $\tilde{v}(t)$ of the form

$$\tilde{v}(t) = [\exp(-t\tilde{T}^{-1})\tilde{Q}_+ + \exp((\tau - t)\tilde{T}^{-1})\tilde{Q}_-]\tilde{h} \qquad t \in [0, \tau] \tag{5.10}$$

where $\tilde{h} \in D(\tilde{T}^{-1})$ is arbitrary, satisfies the differential equation (5.6). Conversely, each solution of (5.6) has the form (5.10).

Let Q_p and Q_m be the orthogonal projections of H onto the positive and negative subspaces for the selfadjoint operator $AT^{-1}|H$, viz.,

$$Q_p f = F^{-1}\tilde{Q}_+ F f, \qquad Q_m f = F^{-1}\tilde{Q}_- F f \qquad f \in H. \tag{5.11}$$

In view of the results (5.7) and (5.10) we introduce the family of operators $\{U(t)\colon 0 < \operatorname{Re} t < \tau\}$ in $L_2(\mathbb{R}, \mu)$,

$$U(t) = \exp((\tfrac{1}{2}\tau - t)AT^{-1})(I - P) + [\exp(-tAT^{-1})Q_p + \exp((\tau - t)AT^{-1})Q_m]P. \tag{5.12}$$

THEOREM 5.1. $U(t)$ *is holomorphic in the open strip* $S = \{t\colon 0 < \operatorname{Re} t < \tau\}$, *uniformly bounded for* $t \in S_\varepsilon = \{t\colon |\arg t| \leqq \pi/2 - \varepsilon, |\arg(\tau - t) - \pi| \leqq \pi/2 - \varepsilon\}$, $\varepsilon > 0$, *and strongly continuous* (*within* S_ε) *at* $t = 0$ *and* $t = \tau$. *The function* $v(t) = U(t)h$ *satisfies* (5.3) *for* $0 < t < \tau$ *for any* $h \in D(T^{-1})$. *Conversely, each solution of* (5.3) *has the form* $v(t) = U(t)h$ *for some* $h \in D(T^{-1})$.

Proof. On G, $U(t)$ is represented by the matrix

$$\begin{pmatrix} 1 & \frac{1}{2}\tau - 1 \\ 0 & 1 \end{pmatrix}$$

relative to the basis (Te, T^2e), which is clearly holomorphic for all $t \in \mathbb{C}$. On H, $U(t)$ corresponds via F to the transformation $\exp(-t\tilde{T}^{-1})\tilde{Q}_+ + \exp((\tau - t)\tilde{T}^{-1})\tilde{Q}_-$ in $L_2(\mathbb{R}, \nu)$, which is holomorphic in S, uniformly bounded in S_ε and strongly continuous at $t = 0$ and $t = \tau$, according to Lemma 5.1. The range of $U(t)$ is contained in $D(T^{-1})$. If $v(t) = U(t)h$, $h \in D(T^{-1})$, then $m_1(t)$, $m_0(t)$ have the form (5.7) and $\tilde{v}(t)$ has the form (5.10), so $\hat{F}v(t)$ satisfies (5.5), (5.6). □

In view of the boundary conditions (5.4) we define two operators V_τ and W_τ in $L_2(\mathbb{R}, \mu)$,

$$V_\tau = Q_+ U(0) + Q_- U(\tau), \tag{5.13}$$

$$W_\tau = Q_- U(0) + Q_+ U(\tau). \tag{5.14}$$

As can be seen from (5.4), the question of the existence of a solution of the boundary value problem (5.3), (5.4) depends upon the invertibility of the operator V_τ on $D(T^{-1})$. Using the expression (5.12) for $U(t)$ we see that V_τ is given by

$$V_\tau=[Q_+\exp(\tfrac{1}{2}\tau AT^{-1})+Q_-\exp(-\tfrac{1}{2}\tau AT^{-1})](I-P)$$
$$+[Q_+Q_p+Q_-Q_m+Q_+\exp(\tau AT^{-1})Q_m+Q_-\exp(-\tau AT^{-1})Q_p]P.$$

We put

$$|B^{-1}|=AT^{-1}Q_p-AT^{-1}Q_m. \tag{5.15}$$

$|B^{-1}|$ is a positive operator in H. Its exponential $\exp(-\tau|B^{-1}|)$ is well defined, $\|\exp(-\tau|B^{-1}|)\|\leqq 1$ for $\tau\geqq 0$. We rewrite the expression for V_τ in the following form,

$$\begin{aligned}V_\tau&=[Q_+\exp(\tfrac{1}{2}\tau AT^{-1})+Q_-\exp(-\tfrac{1}{2}\tau AT^{-1})](I-P)\\&\quad+[(Q_+Q_p+Q_-Q_m)+(Q_+Q_m+Q_-Q_p)\exp(-\tau|B^{-1}|)]P.\end{aligned}\tag{5.16}$$

Let $|T^{-1}|$ be the positive operator defined in $L_2(\mathbb{R},\mu)$ by the relation

$$|T^{-1}|=T^{-1}Q_+-T^{-1}Q_-. \tag{5.17}$$

LEMMA 5.2. *The operator V_τ is an injective map of $D(T^{-1})$ into itself.*

Proof. Take any $h\in D(T^{-1})$. Then $V_\tau h,\ W_\tau h\in D(T^{-1})$ and

$$\begin{aligned}(|T^{-1}|V_\tau h, V_\tau h)_\mu-(|T^{-1}|W_\tau h, W_\tau h)_\mu&=(T^{-1}U(0)h, U(0)h)_\mu-(T^{-1}U(\tau)h, U(\tau)h)_\mu\\&=-\int_0^\tau\frac{d}{dt}(T^{-1}U(t)h, U(t)h)_\mu\,dt\\&=2\int_0^\tau\|AT^{-1}U(t)h\|_\mu^2\,dt.\end{aligned}$$

Now,

$$\begin{aligned}\|AT^{-1}U(t)h\|_\mu^2&=2|(h,e)_\mu|^2+\int_0^\infty\lambda^{-2}[\exp(-2t/\lambda)]|FPh(\lambda)|^2\nu(d\lambda)\\&\quad+\int_{-\infty}^0\lambda^{-2}[\exp(2(\tau-t)/\lambda)]|FPh(\lambda)|^2\nu(d\lambda).\end{aligned}$$

Using Fubini's theorem we find

$$\begin{aligned}2\int_0^\tau\|AT^{-1}U(t)h\|_\mu^2\,dt&=4|(h,e)_\mu|^2+\int_0^\infty\lambda^{-1}(1-\exp(-2\tau/\lambda))|FPh(\lambda)|^2\nu(d\lambda)\\&\quad-\int_{-\infty}^0\lambda^{-1}(1-\exp(2\tau/\lambda))|FPh(\lambda)|^2\nu(d\lambda)\\&=4|(h,e)_\mu|^2+(|B^{-1}|(I-\exp(-2\tau|B^{-1}|))Ph, Ph)_\mu.\end{aligned}$$

Hence, for any $h\in D(T^{-1})$ we have the inequality

$$(|T^{-1}|V_\tau h, V_\tau h)_\mu\geqq 4|(h,e)_\mu|^2+(C_\tau Ph, Ph)_\mu, \tag{5.18}$$

where C_τ is a positive operator, $C_\tau=|B^{-1}|(I-\exp(-2\tau|B^{-1}|))$. Now, suppose $V_\tau h=0$ for some $h\in D(T^{-1})$. Then it follows from (5.18) that $(h,e)_\mu=0$ and $Ph=0$, i.e., $h=\alpha Te$ for some $\alpha\in\mathbb{C}$. However, $V_\tau Te=Te$. Hence, $V_\tau h=0$ implies $h=0$. □

Next, we want to show that V_τ is a surjective map of $D(T^{-1})$ onto itself. We will do so in a somewhat indirect way suggested to us by R. Beals. It is based upon the idea that one first studies, instead of V_τ, another operator V_τ^ε ($\varepsilon>0$) which coincides with V_τ on $D(T^{-1})\cap H$ for every $\varepsilon>0$, but which may differ from V_τ on G. (We recall that $D(T^{-1})=G\oplus D(T^{-1})\cap H$.)

Let A_ε be the following perturbation of A in $L_2(\mathbb{R},\mu)$,

$$A_\varepsilon f = Af + 2\varepsilon(f, T^2e)_\mu T^2e \qquad f\in L_2(\mathbb{R},\mu), \tag{5.19}$$

where $\varepsilon>0$. This operator is bounded, positive, selfadjoint in $L_2(\mathbb{R},\mu)$. Its inverse A_ε^{-1} exists and is given by

$$A_\varepsilon^{-1}f = f + ((2\varepsilon^{-1}+3)(f,e)_\mu - 2(f,T^2e)_\mu)e - 2(f,e)_\mu T^2e \qquad f\in L_2(\mathbb{R},\mu).$$

We use this operator A_ε^{-1} to define a new inner product $(\cdot,\cdot)_{A_\varepsilon}$ and norm $\|\cdot\|_{A_\varepsilon}$ in $L_2(\mathbb{R},\mu)$,

$$(f,g)_{A_\varepsilon} = (A_\varepsilon^{-1}f, g)_\mu \qquad f,g\in L_2(\mathbb{R},\mu), \tag{5.20}$$

$$\|f\|_{A_\varepsilon} = (f,f)_{A_\varepsilon}^{1/2} \qquad f\in L_2(\mathbb{R},\mu). \tag{5.21}$$

Since A_ε^{-1} is bounded and positive in $L_2(\mathbb{R},\mu)$, the norm $\|\cdot\|_{A_\varepsilon}$ is equivalent to the norm $\|\cdot\|_\mu$ in $L_2(\mathbb{R},\mu)$. On H, A_ε^{-1} is given by the expression

$$A_\varepsilon^{-1}f = f - 2(f,T^2e)_\mu e \qquad f\in H,$$

which is independent of ε, so for all $\varepsilon>0$ we have the identity

$$(f,g)_{A_\varepsilon} = (f,g) \qquad f,g\in H. \tag{5.22}$$

The properties of the operator $A_\varepsilon T^{-1}$ in $L_2(\mathbb{R},\mu)$ are very similar to those of AT^{-1}, as may be seen from the following lemma.

LEMMA 5.3. (i) *The operator $A_\varepsilon T^{-1}$ is reduced by the pair $\{G,H\}$ for all $\varepsilon>0$.*

(ii) *$A_\varepsilon T^{-1}|G$ is defined on G and has the representation $\begin{pmatrix}0&1\\ \varepsilon&0\end{pmatrix}$ relative to the basis (Te, T^2e) of G.*

(iii) *$A_\varepsilon T^{-1}|H$ coincides with $AT^{-1}|H$ for all $\varepsilon>0$.*

(iv) *$A_\varepsilon T^{-1}$ is invertible in $L_2(\mathbb{R},\mu)$, its inverse $B_\varepsilon=(A_\varepsilon T^{-1})^{-1}=TA_\varepsilon^{-1}$ is given by*

$$B_\varepsilon f = 2\varepsilon^{-1}(f,e)_\mu Te + 2(f,Te)_\mu T^2e + BPf. \tag{5.23}$$

Proof. For any $f\in D(A_\varepsilon T^{-1})=D(T^{-1})$ we have the identity

$$A_\varepsilon T^{-1}f = AT^{-1}f + 2\varepsilon(f,Te)_\mu T^2e.$$

The lemma is readily verified if one uses the properties of AT^{-1} given in Theorem 3.1. The expression for B_ε follows from the representation $(A_\varepsilon T^{-1}|G)^{-1}=\begin{pmatrix}1&1/\varepsilon\\ 1&0\end{pmatrix}$ relative to the basis (Te, T^2e) of G and from the fact that $(A_\varepsilon T^{-1}|H)^{-1}=(AT^{-1}|H)^{-1}=PT|H=B$. □

Observe that B_ε coincides with the operator B in H for all $\varepsilon>0$. In Lemma 4.2 we established the selfadjointness of B in H relative to the inner product $(\cdot,\cdot)_\mu$. In the following lemma we establish a similar property for the operator B_ε.

LEMMA 5.4. *B_ε is selfadjoint in $L_2(\mathbb{R},\mu)$ relative to the inner product $(\cdot,\cdot)_{A_\varepsilon}$; $\sigma(B_\varepsilon)=\mathbb{R}$ and $\sigma_p(B_\varepsilon)=\{\pm1/\sqrt{\varepsilon}\}$.*

Proof. Since $D(B_\varepsilon)=D(T)$, $D(B_\varepsilon)$ is dense in $L_2(\mathbb{R},\mu)$ in the topology induced by the norm $\|\cdot\|_\mu$ and, hence, also in the topology induced by the norm $\|\cdot\|_\varepsilon$. For any

$f, g \in D(B_\varepsilon)$ we have

$$(B_\varepsilon f, g)_{A_\varepsilon} = (A_\varepsilon^{-1} T A_\varepsilon^{-1} f, g)_\mu = (f, A_\varepsilon^{-1} T A_\varepsilon^{-1} g)_\mu = (f, B_\varepsilon g)_{A_\varepsilon},$$

so B_ε is symmetric relative to the inner product $(\cdot,\cdot)_{A_\varepsilon}$ and $D(B_\varepsilon) \subset D(B_\varepsilon^*)$. Suppose $g \in D(B_\varepsilon^*)$. The mapping $f \mapsto (B_\varepsilon f, g)_{A_\varepsilon}$ defines a continuous linear functional on $D(B_\varepsilon)$. But $(B_\varepsilon f, g)_{A_\varepsilon} = (A_\varepsilon^{-1} T A_\varepsilon^{-1} f, g)_\mu$ for any pair $f \in D(B_\varepsilon) = D(T)$, $g \in L_2(\mathbb{R}, \mu)$, so the mapping $f \mapsto (B_\varepsilon f, g)_{A_\varepsilon}$ defines, at the same time, a continuous linear functional on $D(T)$. Consequently, $g \in D(T^*)$. Since $D(T^*) = D(T) = D(B_\varepsilon)$, it follows that $g \in D(B_\varepsilon)$. Hence, $D(B_\varepsilon^*) \subset D(B_\varepsilon)$ and B_ε is selfadjoint in $L_2(\mathbb{R}, \mu)$ relative to the inner product $(\cdot,\cdot)_{A_\varepsilon}$. The last two statements of the lemma follow from the fact that B_ε coincides with B on H, so $\sigma(B_\varepsilon | H) = \sigma(B) = \mathbb{R}$, and from the fact that B_ε is represented on G by the matrix $\begin{pmatrix} 0 & 1/\varepsilon \\ 1 & 0 \end{pmatrix}$ whose eigenvalues are $\pm 1/\sqrt{\varepsilon}$, so $\sigma(B_\varepsilon | G) = \sigma_p(B_\varepsilon) = \{\pm 1/\sqrt{\varepsilon}\}$. In fact, it is readily verified that the eigenfunction $f_\pm$ associated with the eigenvalue $\pm 1/\sqrt{\varepsilon}$ is given by $f_\pm = Te \pm \sqrt{\varepsilon} T^2 e$. □

The spectral measure E_ε defined by the selfadjoint operator B_ε can be used to develop an operational calculus. Thus, one establishes the existence of a linear isomorphism F_ε which diagonalizes B_ε on its domain. This isomorphism coincides with the isomorphism F on H. Let Q_p^ε and Q_m^ε be the orthogonal projections which map $L_2(\mathbb{R}, \mu)$ onto the subspaces positive and negative for B_ε, respectively. We define the operator V_τ^ε in $L_2(\mathbb{R}, \mu)$ by the expression

$$V_\tau^\varepsilon = (Q_+ Q_p^\varepsilon + Q_- Q_m^\varepsilon) + (Q_+ Q_m^\varepsilon + Q_- Q_p^\varepsilon) \exp(-\tau |A_\varepsilon T^{-1}|), \tag{5.24}$$

where

$$|A_\varepsilon T^{-1}| = A_\varepsilon T^{-1} (Q_p^\varepsilon - Q_m^\varepsilon); \tag{5.25}$$

cf. (5.15) and (5.16). Then $V_\tau^\varepsilon | H = V_\tau | H$ for all $\varepsilon > 0$.

We wish to study V_τ^ε on $D(T^{-1})$. To this end, however, it is convenient to first study V_τ^ε as an operator in a complete vector space. Therefore, we use the positive operators $|A_\varepsilon T^{-1}|$ and $|T^{-1}|$ to define the inner products $(\cdot,\cdot)_{B_\varepsilon}$ and $(\cdot,\cdot)_T$ on $D(T^{-1})$,

$$(f, g)_{B_\varepsilon} = (|A_\varepsilon T^{-1}| f, g)_{A_\varepsilon} = ((Q_p^\varepsilon - Q_m^\varepsilon) f, T^{-1} g)_\mu \qquad f, g \in D(T^{-1}),$$

$$(f, g)_T = (|T^{-1}| f, g)_\mu = ((Q_+ - Q_-) f, T^{-1} g)_\mu \qquad f, g \in D(T^{-1}),$$

with the corresponding norms $\|\cdot\|_{B_\varepsilon}$ and $\|\cdot\|_T$,

$$\|f\|_{B_\varepsilon} = (f, f)_{B_\varepsilon}^{1/2} \qquad f \in D(T^{-1}),$$

$$\|f\|_T = (f, f)_T^{1/2} \qquad f \in D(T^{-1}),$$

and consider the completion of $D(T^{-1})$ with respect to each of these norms. We denote these completions by H_{B_ε} and H_T, respectively. We extend Q_p^ε and Q_m^ε by continuity to bounded selfadjoint projections in H_{B_ε} and, similarly, Q_+ and Q_- to bounded selfadjoint projections in H_T. The following lemma holds.

LEMMA 5.5. *The norms $\|\cdot\|_{B_\varepsilon}$ and $\|\cdot\|_T$ are equivalent on $D(T^{-1})$.*

Proof. Suppose $f \in D(T^{-1})$. Let $u(t) = \exp(-t|T|) f$. Then $u(t) \in D(T^{-1})$ for all $t \geqq 0$. Moreover, $u(t)$ and $T^{-1} u(t) \to 0$ in $L_2(\mathbb{R}, \mu)$ as $t \to \infty$, so $(|A_\varepsilon T^{-1}| u(t), u(t))_{A_\varepsilon} =$

$((Q_p^\varepsilon - Q_m)u(t), T^{-1}u(t))_\mu \to 0$ as $t \to \infty$. Therefore,

$$\|f\|_{B_\varepsilon}^2 = -\int_0^\infty \frac{d}{dt}(|A_\varepsilon T^{-1}|u(t), u(t))_{A_\varepsilon}\, dt$$

$$= 2\,\mathrm{Re} \int_0^\infty (|A_\varepsilon T^{-1}|u(t), |T|u(t))_{A_\varepsilon}\, dt$$

$$= 2\,\mathrm{Re} \int_0^\infty ((Q_p^\varepsilon - Q_m^\varepsilon)u(t), (Q_+ - Q_-)u(t))_\mu\, dt$$

$$\leqq 2C^2 \int_0^\infty \|u(t)\|_\mu^2\, dt = C^2\|f\|_T^2.$$

Then, also, $\|f\|_T^2 = (T^{-1}(Q_+ - Q_-)f, f)_\mu = (A_\varepsilon T^{-1}(Q_+ - Q_-)f, f)_{A_\varepsilon} = ((Q_+ - Q_-)f, (Q_p^\varepsilon - Q_m^\varepsilon)f)_{B_\varepsilon} \leqq \|(Q_+ - Q_-)f\|_{B_\varepsilon}\|(Q_p^\varepsilon - Q_m^\varepsilon)f\|_{B_\varepsilon} \leqq C\|(Q_+ - Q_-)f\|_T\|f\|_{B_\varepsilon} \leqq C\|f\|_T\|f\|_{B_\varepsilon}$, so $C^{-1}\|f\|_{B_\varepsilon} \leqq \|f\|_T \leqq C\|f\|_{B_\varepsilon}$ for any $f \in D(T^{-1})$. □

It follows that H_{B_ε} and H_T coincide as sets, that Q_p^ε and Q_m^ε are bounded as mappings from H_{B_ε} into H_T, that Q_+ and Q_- are bounded as mappings from H_T into H_{B_ε}, etc. We consider the map V_τ^ε defined by (5.24) as an operator in H_T and prove the following result.

LEMMA 5.6. *V_τ^ε is a topological automorphism of H_T.*

Proof. Let

$$V^\varepsilon = Q_+Q_p^\varepsilon + Q_-Q_m^\varepsilon, \qquad W^\varepsilon = Q_+Q_m^\varepsilon + Q_-Q_p^\varepsilon.$$

Then

$$V_\tau^\varepsilon = V^\varepsilon + W^\varepsilon \exp(-\tau|A_\varepsilon T^{-1}|).$$

Since H_T and H_{B_ε} coincide as sets and have equivalent topologies, we may consider V^ε and W^ε as mappings from H_{B_ε} into H_T. Their adjoints $V^{\varepsilon *}$ and $W^{\varepsilon *}$ are defined by the relations

$$(V^\varepsilon f, g)_T = (f, V^{\varepsilon *}g)_{B_\varepsilon}, \qquad (W^\varepsilon f, g)_T = (f, W^{\varepsilon *}g)_{B_\varepsilon}$$

for all $f \in H_{B_\varepsilon}$, $g \in H_T$. Given the selfadjointness of Q_+ and Q_- relative to the inner product $(\cdot,\cdot)_T$ and the selfadjointness of Q_p^ε and Q_m^ε relative to the inner product $(\cdot,\cdot)_{B_\varepsilon}$, one verifies that

$$V^{\varepsilon *} = Q_p^\varepsilon Q_+ + Q_m^\varepsilon Q_-, \qquad W^{\varepsilon *} = -Q_m^\varepsilon Q_+ - Q_p^\varepsilon Q_-.$$

Now, for any $h \in H_{B_\varepsilon}$,

$$\|V^\varepsilon h\|_T^2 - \|W^\varepsilon h\|_T^2 = \|h\|_{B_\varepsilon}^2, \tag{5.26}$$

so $\|V^\varepsilon h\|_T \geqq \|h\|_{B_\varepsilon}$, i.e., V^ε is injective with closed range. Similarly, for any $g \in H_T$,

$$\|V^{\varepsilon *}g\|_{B_\varepsilon}^2 - \|W^{\varepsilon *}g\|_{B_\varepsilon}^2 = \|g\|_T^2, \tag{5.27}$$

so $\|V^{\varepsilon *}g\|_{B_\varepsilon} \geqq \|g\|_T$, i.e. $V^{\varepsilon *}$ is injective. It follows that V^ε is a topological isomorphism from H_{B_ε} onto H_T. Since H_{B_ε} and H_T coincide as sets, we conclude that V^ε is a topological automorphism of H_T. The same is true of its adjoint, $V^{\varepsilon *}$.

Using the identity (5.27) with $g = (V^{\varepsilon *})^{-1}h$, $h \in H_T$, we find

$$\|W^{\varepsilon *}(V^{\varepsilon *})^{-1}h\|_{B_\varepsilon}^2 = \|h\|_{B_\varepsilon}^2 - \|(V^{\varepsilon *})^{-1}h\|_T^2 \leqq \gamma\|h\|_{B_\varepsilon}^2,$$

where $0 \leqq \gamma < 1$, so the operator $W^{\varepsilon *}(V^{\varepsilon *})^{-1}$: $H_{B_\varepsilon} \to H_{B_\varepsilon}$ has norm less than one. The same is then true of its adjoint, $(V^\varepsilon)^{-1}W^\varepsilon: H_{B_\varepsilon} \to H_{B_\varepsilon}$. Since the operator $\exp(-\tau|A_\varepsilon T^{-1}|): H_{B_\varepsilon} \to H_{B_\varepsilon}$ has norm at most equal to one, it follows that the series

$$\sum_{n=0}^{\infty} (-1)^n[(V^\varepsilon)^{-1}W^\varepsilon \exp(-\tau|A_\varepsilon T^{-1}|)]^n \tag{5.28}$$

converges in norm. The operator given by this series is defined on all of H_T. As an operator on H_T, it is the inverse of

$$(V^\varepsilon)^{-1}V_\tau^\varepsilon = I + (V^\varepsilon)^{-1}W^\varepsilon \exp(-\tau|A_\varepsilon T^{-1}|).$$

It follows that the mapping $(V^\varepsilon)^{-1}V_\tau^\varepsilon: H_T \to H_T$, and therefore also the mapping V_τ^ε is surjective. □

In particular, considering the operator V_τ^ε on $D(T^{-1})$ we have the following result.

LEMMA 5.7. *V_τ^ε is a surjective map of $D(T^{-1})$ onto itself.*

Proof. The operators $\exp(-\tau|A_\varepsilon T^{-1}|)$, W^ε, V^ε and $(V^\varepsilon)^{-1}$ map $D(T^{-1})$ into itself. It follows that the operator $((V^\varepsilon)^{-1}V_\tau^\varepsilon)^{-1}$, which is defined by the infinite series (5.28), maps $D(T^{-1})$ into itself. The same is then true for $(V_\tau^\varepsilon)^{-1}$, so V_τ^ε as an operator in $D(T^{-1})$ is surjective. □

The proof of the corresponding property of V_τ is now almost trivial: V_τ maps $D(T^{-1})$ into itself, V_τ is injective by Lemma 5.2, and V_τ coincides with V_τ^ε on a subspace of finite codimension; hence, V_τ is a surjective map of $D(T^{-1})$ onto itself. As this result is crucial to the proof of the main results of this paper, we state it in the form of a theorem.

THEOREM 5.2. *The operator V_τ defined by* (5.16) *is a bijective map of $D(T^{-1})$ onto itself.*

Combining Theorems 5.1 and 5.2 and taking into account the relation $u(t) = T^{-1}v(t)$ between the solution $u(t)$ of the boundary value problem (5.1), (5.2) and the solution $v(t)$ of the boundary value problem (5.3), (5.4), we arrive at the main result of our investigation.

THEOREM 5.3. *The boundary value problem* (5.1), (5.2) *has a unique solution $u(t)$ for every $g \in D(T^{-1})$. It is given by $u(t) = T^{-1}U(t)V_\tau^{-1}g$ for $t \in [0, \tau]$, where $U(t)$ is defined by* (5.12) *and V_τ^{-1} is the inverse of the operator V_τ defined by* (5.16).

To conclude this section we briefly discuss the solution of the inhomogeneous boundary value problem (2.7).

Let $\mathrm{Lip}_\alpha(0, \tau; L_2(\mathbb{R}, \mu))$ denote the class of mappings $f: [0, \tau] \to L_2(\mathbb{R}, \mu)$ which are Lipschitz continuous with exponent α, i.e., for which there exist two constants L and α with $0 < \alpha \leqq 1$ such that $\|f(s) - f(t)\|_\mu \leqq L(t-s)^\alpha$ for all $0 \leqq s \leqq t \leqq \tau$.

THEOREM 5.4. *The boundary value problem* (2.7) *has a unique solution $u(t)$ for every $f \in \mathrm{Lip}_\alpha(0, \tau; L_2(\mathbb{R}, \mu))$, $g \in D(T^{-1})$. It is given by $u(t) = T^{-1}v_0(t) + u_1(t)$ for $t \in [0, \tau]$, where*

$$\begin{aligned} v_0(t) = &\int_0^t \exp(-(t-s)AT^{-1})(I-P)f(s)\, ds \\ &+ \int_0^t \exp(-(t-s)AT^{-1})Q_pPf(s)\, ds \\ &- \int_t^\tau \exp(-(t-s)AT^{-1})Q_mPf(s)\, ds \qquad t \in [0, \tau], \end{aligned}$$

and $u_1(t)$ is the (*unique*) *solution of the homogeneous equation* (5.1) *which satisfies the* *boundary conditions* $Q_+Tu_1(0)=Q_+g_1$, $Q_-Tu_1(\tau)=Q_-g_1$, *where* $g_1=g-Q_+T^{-1}v_0(0)-Q_-T^{-1}v_0(\tau)$.

Proof. Under the stated conditions on f the function $u_0(t)$ represents a particular solution of (2.7a); cf. [10, § IX.1.7]. The theorem is then an immediate consequence of the linearity of the problem and Theorem 5.3. □

6. Discussion. The theorems established at the end of the previous section provide a complete answer to the question of the solubility of boundary value problems described by (1.1). It must be observed, however, that the proofs of the theorems are not constructive. For the actual solution of boundary value problems one needs to know, for example, the representations of the transformations F and F^{-1} on (subspaces of) H and $L_2(\mathbb{R}, \nu)$, respectively, cf. Theorem 4.2, and also the representation of the transformation V_τ or, rather, of its inverse V_τ^{-1} on $D(T^{-1})$, cf. Theorem 5.2. We discuss these and other matters in two forthcoming articles [7], [8]. At this point we mention that the representations of F and F^{-1} are established rather easily, at least on certain subspaces of Lipschitz functions. These representations provide the connection between our Theorem 4.2 and the so-called full range completeness theorem enunciated by Cercignani in a study of the slip flow problem; cf. [9, Thm. IV]. A representation for V_τ^{-1} is much more difficult to establish. It is tied to the construction of a unique factorization of the function ω defined in (3.1) into a product of two functions analytic in the left and right half of the complex plane, respectively. It is indeed possible to construct such a factorization and to derive a representation for V_τ^{-1} from it. This representation, in turn, provides a connection between our Theorem 5.2 and Cercignani's half range completeness theorem; cf. [9, Thm. VI].

Acknowledgments. The author would like to express his appreciation to Dr. R. J. Hangelbroek (Univ. of Delaware) and Dr. R. Beals (Univ. of Chicago) for many valuable remarks and suggestions. The author is indebted to the referee who provided detailed comments on the original version of the manuscript. The final form of the manuscript was prepared during the Summer semester 1976/77, while the author was on foreign assignment to the Mathematics Department of the University of Vienna (Austria).

REFERENCES

[1] C. CERCIGNANI, *Theory and Applications of the Boltzmann Equation*, Elsevier, New York, 1975.

[2] J. H. FERZIGER AND H. G. KAPER, *Mathematical Theory of Transport Processes in Gases*, North-Holland, Amsterdam, 1972.

[3] R. J. HANGELBROEK, *A functional-analytic approach to the linear transport equation*, Ph.D. thesis, Groningen, 1973; TTSP, 5 (1976), pp. 1–85.

[4] C. G. LEKKERKERKER, *The linear transport equation. The degenerate case* $c=1$, Proc. Royal Soc. Edinburgh Sect. A, 75A (1975/76), pp. 259–282.

[5] R. BEALS, personal communication.

[6] H. G. KAPER, *Full-range boundary value problems in the kinetic theory of gases*, Proc. Fourth Conf. on Transport Theory, Blacksburg, VA (March 1975), to appear.

[7] ———, *Spectral representation of an unbounded linear transformation arising in the kinetic theory of gases*, this Journal, 10, pp. 179–191.

[8] ———, *A constructive approach to the solution of a class of boundary value problems of mixed type*, J. Math. Anal. Appl., to appear.

[9] C. CERCIGNANI, *Elementary solutions of the linearized gas-dynamics Boltzmann equation and their application to the slip-flow problem*, Ann. Physics, 20 (1962), pp. 219–233.

[10] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.

[11] M. ABRAMOWITZ AND I. A. STEGUN, eds., *Handbook of Mathematical Functions*, NBS Appl. Math. Ser., 55, National Bureau of Standards, Washington, D.C., 1964.
[12] G. SZEGÖ, *Orthogonal Polynomials*, American Mathematical Society, Providence, RI, 1939.
[13] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators*, Interscience, New York, 1963.
[14] N. I. MUSCHELISCHVILI, *Singulaere Intergralgleichungen*, Akademie-Verlag, Berlin, 1965.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0018 $01.00/0

SPECTRAL REPRESENTATION OF AN UNBOUNDED LINEAR TRANSFORMATION ARISING IN THE KINETIC THEORY OF GASES*

HANS G. KAPER†

Abstract. The linear integro-differential equation

$$\frac{\partial xu}{\partial t}(t, x)+u(t, x)-\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-y^2)u(t, y)\,dy=f(t, x),$$

$t\in(0, \tau)$, $x\in\mathbb{R}$, is interpreted in functional form as an ordinary differential equation for the mapping $u: [0, \tau]\to L_2(\mathbb{R}, \mu)$ in a weighted Hilbert space $L_2(\mathbb{R}, \mu)$. In the article *Boundary value problems of mixed type arising in the kinetic theory of gases* [SIAM J. Math. Anal., this issue, pp. 161–178] the author proved the existence of a diagonalizing transformation for the (unbounded) evolution operator associated with the differential equation. In the present article explicit representations for this transformation are established.

1. Introduction. In this article we are concerned with some constructive aspects of the solution of boundary value problems described by the linear integro-differential equation

$$\frac{\partial xu}{\partial t}(t, x)+u(t, x)-\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-y^2)u(t, y)\,dy=f(t, x), \tag{1.1}$$

$t\in(0, \tau)$, $x\in\mathbb{R}$. In [1], we presented a Hilbert space approach based on an interpretation of (1.1) as an ordinary differential equation for the mapping $u:[0, \tau]\to L_2(\mathbb{R}, \mu)$,

$$(Tu)'(t)+Au(t)=0, \qquad t\in(0, \tau). \tag{1.2}$$

Here, $L_2(\mathbb{R}, \mu)$ is a suitably weighted L_2-space, $\mu(dx)=\pi^{-1/2}\exp(-x^2)\,dx$. (See [1] for the notations and definitions.) The results presented in [1] are, however, not constructive. For the actual solution of boundary value problems one needs to know the representations of certain isomorphisms which play a role in the abstract theory. In this and a subsequent article we address ourselves to these more constructive aspects. The knowledgeable reader will observe that the approach of the following sections parallels the approach developed by Hangelbroek [2] and Lekkerkerker [3] in their study of the neutron transport equation. This, of course, is due to the similarity of the underlying equations. However, as we pointed out in the Introduction to our article [1], there are significant differences between the two equations. The most significant one concerns the range of the independent variable x, which is finite in the neutron transport case and infinite in the present case. This implies that unbounded operators are the rule, rather than the exception. Needless to say, it significantly complicates the discussion.

We refer to [1, §§ 2–4], for the formal statement of the problem and a spectral analysis of the (unbounded) operator AT^{-1} which determines the evolution of the function $u(t)$ in $L_2(\mathbb{R}, \mu)$. An important result expressed in [1, Thm. 4.2], concerned the existence of an isomorphism $F: H\to L_2(\mathbb{R}, \nu)$, where H is a closed subspace of $L_2(\mathbb{R}, \mu)$ with finite codimension and $L_2(\mathbb{R}, \nu)$ is another weighted L_2-space; F diagonalizes the restriction $AT^{-1}|H$, or rather its inverse $B=(AT^{-1}|H)^{-1}$, on its domain. The existence of this isomorphism implies the existence of an isomorphism

* Received by the editors September 2, 1976, and in final revised form September 30, 1977.

† Applied Mathematics Division, Argonne National Laboratory, Argonne, Illinois 60439. This work was supported by the U.S. Energy Research and Development Administration.

$\hat{F}: L_2(\mathbb{R}, \mu) \to \mathbb{C}^2 \oplus L_2(\mathbb{R}, \nu)$ which plays an essential role in the solution of boundary value problems described by (1.1). In this article we address ourselves to the problem of establishing representations for the isomorphism F and its inverse.

In § 2 we show that F and F^{-1} are represented by certain Cauchy principal value integrals on polynomial subspaces. Using these representations we define two families of linear functionals $\{\phi_\lambda : \lambda \in \mathbb{R}\}$ and $\{\chi_x : x \in \mathbb{R}\}$ and show in § 3 that these functionals can be extended continuously to certain locally convex vector spaces of locally Lipschitz continuous functions which possibly have a jump discontinuity at the origin. In § 4 we show that on these spaces the isomorphisms F and F^{-1} are again represented by the functionals $\{\phi_\lambda : \lambda \in \mathbb{R}\}$ and $\{\chi_x : x \in \mathbb{R}\}$, respectively. These results provide a rigorous framework for the so-called full-range completeness theorem enunciated by Cercignani, [4, Thm. IV], in his study of the slip flow problem. We formulate such a theorem in the final § 5.

Notation. The notation in this article is the same as in [1]. References to formulas in [1] are preceeded by the Roman numeral I.

2. The transformations F and F^{-1} on polynomial subspaces. In this section we address ourselves to the problem of finding representations of F and F^{-1} on polynomial subspaces. Let $P(\mathbb{R})$ be the vector space of all polynomial functions on $\mathbb{R}$. We recall that P is the (nonorthogonal) projection operator which maps $L_2(\mathbb{R}, \mu)$ onto H along G; cf. I-(3.12).

THEOREM 2.1. *For any $p \in P(\mathbb{R})$ we have*

$$FPp(\lambda) = -\lambda \int_{-\infty}^{\infty} \frac{p(x)-p(\lambda)}{x-\lambda}\mu(dx) + p(\lambda), \qquad \lambda \in \mathbb{R}, \tag{2.1}$$

or, equivalently,

$$FPp(\lambda) = \lambda \fint_{-\infty}^{\infty} \frac{p(x)}{\lambda - x}\mu(dx) + \omega(\lambda)p(\lambda), \qquad \lambda \in \mathbb{R}, \tag{2.2}$$

where $\fint$ denotes the Cauchy principal value integral and $\omega(\lambda)$ is given by I-(3.4).

Proof. First we verify (2.1) for $p \in P(\mathbb{R}) \cap H$. The subspace H does not contain any polynomial of degree less than two. The only polynomials of degree two in H are constant multiples of $Pe = e - 2T^2e$, for which (2.1) is readily verified. (We recall that $FPe = \tilde{e}$.) For any polynomial $p \in P(\mathbb{R}) \cap H$ of degree three or more there exists a polynomial $\tilde{p}_k \in P(\mathbb{R})$ of degree $k \geqq 1$ such that $Fp = \tilde{p}_k$. Such a polynomial $\tilde{p}_k$ can be written in the form

$$\tilde{p}_k = (\tilde{T} - \lambda I)\tilde{p}_{k-1} + \tilde{p}_k(\lambda)\tilde{e}, \tag{2.3}$$

where $\tilde{p}_{k-1} \in P(\mathbb{R})$ is a polynomial of degree $k-1$ ($\lambda \in \mathbb{R}$, λ fixed). Applying the operator F^{-1} to both sides we obtain

$$\begin{aligned} F^{-1}\tilde{p}_k &= (B - \lambda I)F^{-1}\tilde{p}_{k-1} + \tilde{p}_k(\lambda)Pe \\ &= (T - \lambda I)F^{-1}\tilde{p}_{k-1} - 2(TF^{-1}\tilde{p}_{k-1}, Te)_\mu Te + \tilde{p}_k(\lambda)Pe. \end{aligned} \tag{2.4}$$

When we evaluate $F^{-1}\tilde{p}_k$ at some (arbitrary) point $x \in \mathbb{R}$ and at the (fixed) point $\lambda \in \mathbb{R}$, subtract the two expressions, and divide both sides of the resulting expression by $x - \lambda$, we obtain the identity

$$[F^{-1}\tilde{p}_k(x) - F^{-1}\tilde{p}_k(\lambda)](x-\lambda)^{-1} = F^{-1}\tilde{p}_{k-1}(x) - 2(TF^{-1}\tilde{p}_{k-1}, Te)_\mu - 2\tilde{p}_k(\lambda)(x+\lambda).$$

Since $F^{-1}\tilde{p}_{k-1} \in H$ we have $(F^{-1}\tilde{p}_{k-1}, Te)_\mu = 0$, so upon taking inner products with Te

we obtain

$$\tilde{p}_k(\lambda) = -\int_{-\infty}^{\infty} [F^{-1}\tilde{p}_k(x) - F^{-1}\tilde{p}_k(\lambda)]x(x-\lambda)^{-1}\mu(dx)$$

$$= -\lambda \int_{-\infty}^{\infty} [F^{-1}\tilde{p}_k(x) - F^{-1}\tilde{p}_k(\lambda)](x-\lambda)^{-1}\mu(dx) + F^{-1}\tilde{p}_k(\lambda),$$

which proves the validity of the representation (2.1) for $p \in P(\mathbb{R}) \cap H$. Since $P(\mathbb{R}) = G \oplus P(\mathbb{R}) \cap H$ and the right-hand side of (2.1) is zero for $p = Te$ and $p = T^2 e$, the validity of the representation (2.1) for all $p \in P(\mathbb{R})$ follows. The representation (2.2) is an immediate consequence of (2.1) and the definition I-(3.4) of $\omega(\lambda)$ for $\lambda \in \mathbb{R}$. $\square$

THEOREM 2.2. *For any $\tilde{p} \in P(\mathbb{R})$ we have*

$$F^{-1}\tilde{p}(x) = x \int_{-\infty}^{\infty} \frac{\tilde{p}(\lambda) - \tilde{p}(x)}{\lambda - x} \nu(d\lambda) + (Pe)(x)\tilde{p}(x), \qquad x \in \mathbb{R}, \tag{2.5}$$

or, equivalently,

$$F^{-1}\tilde{p}(x) = x \fint_{-\infty}^{\infty} \frac{\tilde{p}(\lambda)}{\lambda - x} \frac{\mu(d\lambda)}{(\omega^+\omega^-)(\lambda)} + \frac{\omega(x)}{(\omega^+\omega^-)(x)}\tilde{p}(x), \qquad x \in \mathbb{R}, \tag{2.6}$$

Proof. Equation (2.5) holds trivially for all polynomials of degree 0 on $P(\mathbb{R})$, which are constant multiples of $\tilde{e}$. (We recall that $F^{-1}\tilde{e} = Pe$.) For any polynomial $\tilde{p}_k \in P(\mathbb{R})$ of degree $k \geqq 1$ we start with the same relation (2.3), this time written in the form $\tilde{p}_k = (\tilde{T} - xI)\tilde{p}_{k-1} + \tilde{p}_k(x)\tilde{e}$, with $x \in \mathbb{R}$, x fixed. Evaluating $\tilde{p}_k$ at some (arbitrary) point $\lambda \in \mathbb{R}$ we obtain $\tilde{p}_{k-1}(\lambda) = (\lambda - x)^{-1}(\tilde{p}_k(\lambda) - \tilde{p}_k(x))$, so

$$(\tilde{p}_{k-1}, \tilde{e})_\nu = \int_{-\infty}^{\infty} [\tilde{p}_k(\lambda) - \tilde{p}_k(x)](\lambda - x)^{-1}\nu(d\lambda). \tag{2.7}$$

Now, $(\tilde{p}_{k-1}, \tilde{e})_\nu = (F^{-1}\tilde{p}_{k-1}, Pe)_\mu = -2(F^{-1}\tilde{p}_{k-1}, T^2 e)_\mu$. On the other hand, $F^{-1}\tilde{p}_k = (T - xI)F^{-1}\tilde{p}_{k-1} - 2(TF^{-1}\tilde{p}_{k-1}, Te)_\mu Te + \tilde{p}_k(x)Pe$, (cf. (2.4)) so $F^{-1}\tilde{p}_k = (T - xI)F^{-1}\tilde{p}_{k-1} + (\tilde{p}_{k-1}, \tilde{e})_\nu Te + \tilde{p}_k(x)Pe$. Evaluation at x and substitution of the expression (2.7) leads to the representation (2.5).

To establish the representation (2.6) we need an expression for Pe which we will now derive. Consider (2.2) for $p = Pe$. Since $FPe = \tilde{e}$, we have the identity

$$\lambda \fint_{-\infty}^{\infty} (Pe)(x)(\lambda - x)^{-1}\mu(dx) + \omega(\lambda)(Pe)(\lambda) = 1, \qquad \lambda \in \mathbb{R}. \tag{2.8}$$

We view this identity as a singular integral equation for Pe. Let $\Phi(z)$ be defined for $z \in \mathbb{C} \backslash \mathbb{R}$ as the Cauchy integral

$$\Phi(z) = z \int_{-\infty}^{\infty} Pe(x)(z - x)^{-1}\mu(dx), \qquad z \in \mathbb{C} \backslash \mathbb{R}.$$

Φ is analytic in the complex plane cut along the real axis. Near infinity its behavior is given by $\Phi(z) \sim -z^{-2} - (51/12)z^{-4} + \cdots$ as $z \to \infty$. The limiting values on the cut, $\Phi^\pm(\lambda) = \lim_{\varepsilon \to 0+} \Phi(\lambda \pm i\varepsilon)$, $\lambda \in \mathbb{R}$, exist and are given by

$$\Phi^\pm(\lambda) = \lambda \fint_{-\infty}^{\infty} (Pe)(x)(\lambda - x)^{-1}\mu(dx) \mp \tfrac{1}{2}(\omega^+ - \omega^-)(\lambda)(Pe)(\lambda), \qquad \lambda \in \mathbb{R}.$$

From this result we obtain the identities

$$\lambda \oint_{-\infty}^{\infty} (Pe)(x)(\lambda - x)^{-1}\mu(dx) = \tfrac{1}{2}(\Phi^+ + \Phi^-)(\lambda), \tag{2.9a}$$

$$(Pe)(x) = -(\Phi^+ - \Phi^-)(\lambda)/((\omega^+ - \omega^-)(\lambda)), \tag{2.9b}$$

which we use to rewrite (2.8) in terms of $\Phi^{\pm}(\lambda)$ and $\omega^{\pm}(\lambda)$,

$$(\Phi/\omega)^+(\lambda) - (\Phi/\omega)^-(\lambda) = -(\omega^+ - \omega^-)(\lambda)/((\omega^+\omega^-)(\lambda)), \qquad \lambda \in \mathbb{R}. \tag{2.10}$$

Thus, we have transformed the original singular integral equation (2.8) into the following boundary value problem: to determine a complex function $N = \Phi/\omega$, which is analytic in the complex plane cut along the real axis and whose limiting values at the cut satisfy the jump condition (2.10). The general solution of this boundary value problem having finite degree at infinity is

$$N(z) = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\omega^+ - \omega^-}{\omega^+\omega^-}(\lambda)\frac{d\lambda}{\lambda - z} + p(z),$$

where p is an arbitrary polynomial; see [5, § 38]. We determine p by noting that the ratio Φ/ω is bounded at infinity. In fact, $\lim_{z\to\infty}(\Phi/\omega)(z) = 2$, so

$$\Phi(z) = \left[2 - \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\omega^+ - \omega^-}{\omega^+\omega^-}(\lambda)\frac{d\lambda}{\lambda - z}\right]\omega(z).$$

Since $(2\pi i)^{-1}(\omega^+ - \omega^-)(\lambda)\,d\lambda = \lambda\mu(d\lambda)$ and $\int_{-\infty}^{\infty}[\omega^+\omega^-(\lambda)]^{-1}\mu(d\lambda) = \int_{-\infty}^{\infty}\nu(d\lambda) = \|Pe\|_\mu^2 = 2$, we can rewrite this result,

$$\Phi(z) = -z\omega(z)\int_{-\infty}^{\infty}[\omega^+\omega^-(\lambda)]^{-1}(\lambda - z)^{-1}\mu(d\lambda).$$

Using (2.9b) we readily obtain the following expression for $(Pe)(x)$,

$$(Pe)(x) = x\oint_{-\infty}^{\infty}\frac{1}{\lambda - x}\frac{\mu(d\lambda)}{(\omega^+\omega^-)(\lambda)} + \frac{\omega(x)}{(\omega^+\omega^-)(x)}, \qquad x \in \mathbb{R}. \tag{2.11}$$

The representation (2.6) for F^{-1} now follows immediately from the representation (2.5) if one takes into account the relation between the measures μ and ν, cf. [1, Lemma 4.5], and substitutes $(Pe)(x)$ according to (2.11). □

3. Lipschitz spaces and the functionals ϕ_λ ($\lambda \in \mathbb{R}$) and χ_x ($x \in \mathbb{R}$). The expressions (2.2) and (2.6) define two linear functionals ϕ_λ and χ_x on $P(\mathbb{R})$,

$$\langle \phi_\lambda, p\rangle = FPp(\lambda), \qquad \lambda \in \mathbb{R}, \quad p \in P(\mathbb{R}), \tag{3.1}$$

$$\langle \chi_x, \tilde{p}\rangle = F^{-1}\tilde{p}(x), \qquad x \in \mathbb{R}, \quad \tilde{p} \in P(\mathbb{R}). \tag{3.2}$$

In this section we extend these functionals to certain Lipschitz spaces.

We recall that a complex-valued function f defined on a subset J of $\mathbb{R}$ is said to satisfy a (uniform) Lipschitz condition with exponent α $(0 < \alpha \leqq 1)$ if the asymptotic relation $f(x) - f(y) = O(|x-y|^\alpha)$ as $|x-y| \to 0$ is satisfied for all $x, y \in J (x \neq y)$. A function is called Lipschitz continuous on J if it is bounded and satisfies a (uniform) Lipschitz condition there. For $0 < \alpha \leqq 1$, Lip (J, α) will denote the vector space of all Lipschitz continuous functions on J with exponent α. Lip (J, α) is a Banach space if the norm $\|\cdot\|_{J,\alpha}$ is defined by

$$\|f\|_{J,\alpha} = \sup_{x\in J}|f(x)| + p_{J,\alpha}(f), \qquad f \in \text{Lip}\,(J, \alpha),$$

where

$$p_{J,\alpha}(f)=\sup_{x,y\in J}\{|f(x)-f(y)|/|x-y|^{\alpha}\}.$$

Furthermore, we will say that f satisfies a (uniform) lipschitz (lower case l) condition with exponent $\alpha(0<\alpha\leqq 1)$ if the (stronger) asymptotic relation $f(x)-f(y)=o(|x-y|^{\alpha})$ as $|x-y|\to 0$ is satisfied for all $x, y\in J(x\neq y)$, and call a function lipschitz (lower case l) continuous on J if it is bounded and satisfies a (uniform) lipschitz condition there. If $\mathrm{lip}\,(J,\alpha)$ denotes the vector space of all lipschitz continuous functions on J with exponent $\alpha(0<\alpha\leqq 1)$, then $\mathrm{lip}\,(J,\alpha)$ is a closed subspace of $\mathrm{Lip}\,(J,\alpha)$. One has the following inclusion relations,

$$\mathrm{Lip}\,(J,\alpha)\subset \mathrm{lip}\,(J,\beta)\subset \mathrm{Lip}\,(J,\beta)\quad \text{for } 0<\beta<\alpha\leqq 1.$$

For $0<\alpha\leqq 1$, we define $\Lambda(\mathbb{R},\alpha,\mu)$ as the linear vector space of all functions f defined and continuous on $\mathbb{R}$, which belong to $L_2(\mathbb{R},\mu)$ and which are locally lipschitz continuous with exponent α on $\mathbb{R}$ in the sense that f is lipschitz continuous with exponent α on every compact subset K of $\mathbb{R}$. For any compact subset K of $\mathbb{R}$ we define the seminorm $p_{K,\alpha,\mu}(\cdot)$,

$$p_{K,\alpha,\mu}(f)=\|f\|_{\mu}+\|f\|_{K,\alpha},\qquad f\in\Lambda(\mathbb{R},\alpha,\mu).$$

$\Lambda(\mathbb{R},\alpha,\mu)$ is a locally convex vector space by the family of these seminorms. We refer to the topology thus defined as the lipschitz topology of $\Lambda(\mathbb{R},\alpha,\mu)$. It is stronger than the topology induced on $\Lambda(\mathbb{R},\alpha,\mu)$ by the norm $\|\cdot\|_{\mu}$. If $\Lambda(\mathbb{R},\mu)$ denotes the linear vector space of all functions f defined and continuous on $\mathbb{R}$ which belong to $L_2(\mathbb{R},\mu)$ and which are locally lipschitz continuous, then $\Lambda(\mathbb{R},\mu)=\bigcup_{0<\alpha\leqq 1}\Lambda(\mathbb{R},\alpha,\mu)$. $\Lambda(\mathbb{R},\mu)$ endowed with the inductive limit topology is again a locally convex vector space; see [6, § I.8]. The spaces $\Lambda(\mathbb{R},\alpha,\nu)$ $(0<\alpha\leqq 1)$ and $\Lambda(\mathbb{R},\nu)$ are defined in an entirely similar manner.

For future applications we will also consider spaces of piecewise continuous functions, in particular, continuous functions which have a jump discontinuity at the origin. To this end we introduce the vector spaces $\Lambda(\mathbb{R}_{\pm},\alpha,\mu)$, $0<\alpha\leqq 1$, and $\Lambda(\mathbb{R}_{\pm},\mu)$, whose definitions are similar to those of $\Lambda(\mathbb{R},\alpha,\mu)$ and $\Lambda(\mathbb{R},\mu)$, respectively, and form the direct sums $\Lambda(\mathbb{R}_{-},\alpha,\mu)\oplus\Lambda(\mathbb{R}_{+},\alpha,\mu)$ and $\Lambda(\mathbb{R}_{-},\mu)\oplus\Lambda(\mathbb{R}_{+},\mu)$. These are again locally convex vector spaces by the family of seminorms $\{p_{K,\alpha,\mu}(\cdot)$: $K\subset\mathbb{R}$, K compact$\}$, and the direct sums are topological. Obviously,

$$\Lambda(\mathbb{R},\mu)\subset\Lambda(\mathbb{R}_{-},\mu)\oplus\Lambda(\mathbb{R}_{+},\mu)\subset L_2(\mathbb{R},\mu). \tag{3.3}$$

Since the set $\{0\}\subset\mathbb{R}_{-}\cap\mathbb{R}_{+}$ has μ-measure zero, $\Lambda(\mathbb{R}_{-},\mu)\oplus\Lambda(\mathbb{R}_{+},\mu)$ can be mapped identically into $L_2(\mathbb{R},\mu)$. We consider in particular the restriction e_{+} of e to $\mathbb{R}_{+}$. As an element of $\Lambda(\mathbb{R}_{-},\mu)\oplus\Lambda(\mathbb{R}_{+},\mu)$ it satisfies

$$e_{+}(x)=\begin{cases}0 & x\in\mathbb{R}_{-},\\ 1 & x\in\mathbb{R}_{+}.\end{cases} \tag{3.4}$$

Each $f\in\Lambda(\mathbb{R}_{-},\mu)\oplus\Lambda(\mathbb{R}_{+},\mu)$ can be written in a unique way as $f=h+ae_{+}$, where $h\in\Lambda(\mathbb{R},\mu)$ and $a=f(0+)-f(0-)\in\mathbb{C}$, so

$$\Lambda(\mathbb{R}_{-},\mu)\oplus\Lambda(\mathbb{R}_{+},\mu)=\Lambda(\mathbb{R},\mu)\oplus\mathbb{C}e_{+}, \tag{3.5}$$

where $\mathbb{C}e_{+}$ is the subspace spanned by e_{+}.

The spaces $\Lambda(\mathbb{R}_{-},\alpha,\nu)\oplus\Lambda(\mathbb{R}_{+},\alpha,\nu)$, $0<\alpha\leqq 1$, and $\Lambda(\mathbb{R}_{-},\nu)\oplus\Lambda(\mathbb{R}_{+},\nu)$ are defined in an entirely similar manner. They have similar properties as the μ-spaces. In particular $\Lambda(\mathbb{R}, \nu) \subset \Lambda(\mathbb{R}_-, \nu) \oplus \Lambda(\mathbb{R}_+, \nu) \subset L_2(\mathbb{R}, \nu)$, where $\Lambda(\mathbb{R}_-, \nu) \oplus \Lambda(\mathbb{R}_+, \nu)$ can be mapped identically into $L_2(\mathbb{R}, \nu)$ since the set $\{0\} \subset \mathbb{R}_- \cap \mathbb{R}_+$ has ν-measure zero, and $\Lambda(\mathbb{R}_-, \nu) \oplus \Lambda(\mathbb{R}_+, \nu) = \Lambda(\mathbb{R}, \nu) \oplus \mathbb{C}\tilde{e}_+$, where $\tilde{e}_+$ is the restriction of $\tilde{e}$ to $\mathbb{R}_+$.

Having established these definitions we now set out towards our first goal, namely to prove that the polynomials on $\mathbb{R}$ are dense in $\Lambda(\mathbb{R}, \alpha, \mu)$ and $\Lambda(\mathbb{R}, \alpha, \nu)$ relative to their respective lipschitz topologies. This goal will be achieved in a number of steps.

LEMMA 3.1. *The set* $\mathrm{lip}_0(\mathbb{R}, \alpha)$ *of all functions in* $\mathrm{lip}(\mathbb{R}, \alpha)$ *which have compact support is dense in* $\Lambda(\mathbb{R}, \alpha, \mu)$ *and* $\Lambda(\mathbb{R}, \alpha, \nu)$ *relative to the respective lipschitz topologies* $(0 < \alpha \leqq 1)$.

Proof. Let $f \in \Lambda(\mathbb{R}, \alpha, \mu)$. We take a real-valued function ψ which is defined and infinitely differentiable on $\mathbb{R}$, which has compact support, and which is such that $\psi(x) = 1$ for $|x| \leqq 1$. The function f_ε defined by $f_\varepsilon(x) = f(x)\psi(\varepsilon x)$ $(x \in \mathbb{R})$ belongs to $\mathrm{lip}_0(\mathbb{R}, \alpha)$ for any $\varepsilon > 0$. Furthermore,

$$\begin{aligned}\int_{-\infty}^{\infty} |(f - f_\varepsilon)(x)|^2 \mu(dx) &= \int_{-\infty}^{\infty} |1 - \psi(\varepsilon x)|^2 |f(x)|^2 \mu(dx) \\ &= \int_{|x| > 1/\varepsilon} |1 - \psi(\varepsilon x)|^2 |f(x)|^2 \mu(dx) \\ &\leqq \{\sup_{x \in \mathbb{R}} |1 - \psi(x)|^2\} \int_{|x| > 1/\varepsilon} |f(x)|^2 \mu(dx).\end{aligned}$$

Since $f \in L_2(\mathbb{R}, \mu)$, the last integral converges to zero as $\varepsilon \to 0$, so $f_\varepsilon \to f$ in $L_2(\mathbb{R}, \mu)$.

Next, for any compact subset K of $\mathbb{R}$ we have $f|K \in \mathrm{lip}(K, \alpha)$, so

$$|(f - f_\varepsilon)(x)| \leqq \{\sup_{x \in K} |f(x)|\} |1 - \psi(\varepsilon x)| \leqq \|f\|_{K,\alpha} |1 - \psi(\varepsilon x)|.$$

By choosing ε sufficiently small we can achieve the identity $\psi(\varepsilon x) = 1$ for all $x \in K$, so $\lim_{\varepsilon \to 0} \sup_{x \in K} |(f - f_\varepsilon)(x)| = 0$. Furthermore, for any pair of points $x, y \in K$ we have

$$\begin{aligned}|(f - f_\varepsilon)(x) - (f - f_\varepsilon)(y)| &\leqq |1 - \psi(\varepsilon x)|\, |f(x) - f(y)| + |f(y)|\, |\psi(\varepsilon x) - \psi(\varepsilon y)| \\ &\leqq [|1 - \psi(\varepsilon x)| + \varepsilon\{\sup_{x,y \in K} |x - y|^{1-\alpha}\}] \|f\|_{K,\alpha} |x - y|^\alpha.\end{aligned}$$

The expression in the square brackets can be made arbitrarily small by letting ε go to zero, so $\lim_{\varepsilon \to 0} p_{K,\alpha}(f - f_\varepsilon) = 0$. Thus, $f_\varepsilon \to f$ in the topology of $L_2(\mathbb{R}, \mu)$, as well as in the topology of $\mathrm{lip}(K, \alpha)$ for any compact subset K of $\mathbb{R}$. Hence, $f_\varepsilon \to f$ in the lipschitz topology of $\Lambda(\mathbb{R}, \alpha, \mu)$.

Similarly one shows that, for any $\tilde{f} \in \Lambda(\mathbb{R}, \alpha, \nu)$, one can construct functions $\tilde{f}_\varepsilon \in \mathrm{lip}_0(\mathbb{R}, \alpha)$ $(\varepsilon > 0)$ such that $f_\varepsilon \to f$ in the lipschitz topology of $\Lambda(\mathbb{R}, \alpha, \nu)$. □

LEMMA 3.2. *The set* $C^1_{\mathrm{B}}(\mathbb{R})$ *of bounded functions on* $\mathbb{R}$ *which have a bounded continuous derivative is dense in* $\Lambda(\mathbb{R}, \alpha, \mu)$ *and in* $\Lambda(\mathbb{R}, \alpha, \nu)$ *relative to the respective lipschitz topologies* $(0 < \alpha < 1)$.

Proof. Let $\phi \in C^1_{\mathrm{B}}(\mathbb{R})$. Then $\phi \in L_2(\mathbb{R}, \mu)$ and $\phi \in L_2(\mathbb{R}, \nu)$. Furthermore, $|\phi(x) - \phi(y)| \leqq \{\sup |\phi'(x)|\} |x - y|$ for all pairs $x, y \in \mathbb{R}$, so ϕ satisfies the asymptotic relation $\phi(x) - \phi(y) = o(|x - y|^\alpha)$ as $|x - y| \to 0$ if $0 < \alpha < 1$. Since ϕ is bounded on $\mathbb{R}$, $\phi \in \mathrm{lip}(\mathbb{R}, \alpha)$ for $0 < \alpha < 1$ and, a fortiori, $\phi|K \in \mathrm{lip}(K, \alpha)$ for $0 < \alpha < 1$ for any compact subset K of $\mathbb{R}$. Hence, $C^1_{\mathrm{B}}(\mathbb{R}) \subset \Lambda(\mathbb{R}, \alpha, \mu)$ and $C^1_{\mathrm{B}}(\mathbb{R}) \subset \Lambda(\mathbb{R}, \alpha, \nu)$ for $0 < \alpha < 1$.

Because of Lemma 3.1 it is sufficient to prove that any $f \in \mathrm{lip}_0(\mathbb{R}, \alpha)$ can be approximated arbitrarily closely in the lipschitz topology of either $\Lambda(\mathbb{R}, \alpha, \mu)$ or $\Lambda(\mathbb{R}, \alpha, \nu)$ by elements of $C^1_{\mathrm{B}}(\mathbb{R})$. In [2, Lemma 6.2], it is shown that $C^1_{\mathrm{B}}(\mathbb{R})$ is a dense subspace of lip $(\mathbb{R}, \alpha)$ $(0<\alpha<1)$, so for given $\varepsilon>0$ there exists a $\phi_\varepsilon \in C^1_{\mathrm{B}}(\mathbb{R})$ such that $\|f-\phi_\varepsilon\|_{\mathbb{R},\alpha}<\varepsilon$. Then, $\|f-\phi_\varepsilon\|_\mu \leqq \sup_{x\in\mathbb{R}} |(f-\phi_\varepsilon)(x)| \leqq \|f-\phi_\varepsilon\|_{\mathbb{R},\alpha}<\varepsilon$ and, obviously, $\|f-\phi_\varepsilon\|_{K,\alpha} \leqq \|f-\phi_\varepsilon\|_{\mathbb{R},\alpha}<\varepsilon$ for any compact subset K of $\mathbb{R}$. Hence, $\phi_\varepsilon \to f$ in the lipschitz topology of $\Lambda(\mathbb{R}, \alpha, \mu)$. The proof for $\Lambda(\mathbb{R}, \alpha, \nu)$ is entirely similar. □

Now we come to the following important result.

LEMMA 3.3. *The set $P(\mathbb{R})$ of polynomial functions on $\mathbb{R}$ is dense in $\Lambda(\mathbb{R}, \alpha, \mu)$ and in $\Lambda(\mathbb{R}, \alpha, \nu)$ relative to the respective lipschitz topologies* $(0<\alpha<1)$.

Proof. Because of Lemma 3.2 it suffices to prove that any $f \in C^1_{\mathrm{B}}(\mathbb{R})$ can be approximated arbitrarily closely in either of the lipschitz topologies by elements of $P(\mathbb{R})$. Obviously, $P(\mathbb{R}) \subset \Lambda(\mathbb{R}, \alpha, \mu)$ and $P(\mathbb{R}) \subset \Lambda(\mathbb{R}, \alpha, \nu)$ for $0<\alpha<1$. Let $f \in C^1_{\mathrm{B}}(\mathbb{R})$ be given. According to the theory of weighted uniform approximation of continuous functions on $\mathbb{R}$, cf. [7, § 1.8], there exists for any $\varepsilon>0$ a polynomial $q_\varepsilon \in P(\mathbb{R})$ such that $\sup\{|f'(x)-q_\varepsilon(x)| \exp(-\frac{1}{4}x^2): x \in \mathbb{R}\}<\frac{1}{2}\varepsilon$. Then the polynomial $p_\varepsilon \in P(\mathbb{R})$ defined by $p_\varepsilon(x)=f(0)+\int_0^x q_\varepsilon(t)\,dt$ $(x \in \mathbb{R})$ is such that

$$\begin{aligned} |f(x)-p_\varepsilon(x)| &\leqq \left|\int_0^x |f'(t)-q_\varepsilon(t)|\,dt\right| \\ &\leqq \tfrac{1}{2}\varepsilon\left|\int_0^x \exp(-\tfrac{1}{4}t^2)\,dt\right| \\ &\leqq \varepsilon|D(\tfrac{1}{2}x)| \exp(-\tfrac{1}{4}x^2). \end{aligned}$$

for all $x \in \mathbb{R}$, where $D(x)$ is Dawson's integral, $D(x)=\exp(-x^2)\int_0^x \exp(t^2)\,dt$ for $x \in \mathbb{R}$; cf. [8, § 7.1]. Dawson's integral is bounded, $|D(x)|<1$ for $x \in \mathbb{R}$, so the polynomial $p_\varepsilon \in P(\mathbb{R})$ provides a weighted uniform ε-approximation of f and f' in the sense that, simultaneously,

$$\sup_{x\in\mathbb{R}} \{|(f-p_\varepsilon)(x)| \exp(-\tfrac{1}{4}x^2)\}<\varepsilon$$

and

$$\sup_{x\in\mathbb{R}} \{|(f'-p'_\varepsilon)(x)| \exp(-\tfrac{1}{4}x^2)\}<\varepsilon.$$

Then,

$$\|f-p_\varepsilon\|_\mu \leqq \sup_{x\in\mathbb{R}} \{|(f-p_\varepsilon)(x)| \exp(-\tfrac{1}{4}x^2)\}\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)^{1/2}<2^{1/4}\varepsilon. \tag{3.6}$$

Furthermore, on any compact subset K of $\mathbb{R}$,

$$|(f-p_\varepsilon)(x)| \leqq \{\sup_{x\in K} e^{x^2/4}\}|(f-p_\varepsilon)(x)| \exp(-\tfrac{1}{4}x^2)<\{\sup_{x\in K} e^{x^2/4}\}\varepsilon$$

and

$$\begin{aligned} |(f-p_\varepsilon)(x)-(f-p_\varepsilon)(y)| &\leqq \left|\int_y^x (f'-p'_\varepsilon)(t)\,dt\right| \leqq \varepsilon\left|\int_y^x e^{t^2/4}\,dt\right| \\ &< \{\sup_{x\in K} e^{x^2/4}\}\{\sup_{x,y\in K} |x-y|^{1-\alpha}\}\varepsilon|x-y|^\alpha, \end{aligned}$$

so there exists a positive constant $C=C(K, f, \varepsilon)$ such that $\|f-p_\varepsilon\|_{K,\varepsilon}<C_\varepsilon$. These results imply that $p_\varepsilon \to f$ in the lipschitz topology of $\Lambda(\mathbb{R}, \alpha, \mu)$. The proof for $\Lambda(\mathbb{R}, \alpha, \nu)$ is entirely similar. □

THEOREM 3.1. *The linear functions $\phi_\lambda (\lambda \in \mathbb{R})$ defined on $P(\mathbb{R})$ by equation* (3.1) *can be extended uniquely by continuity to* $\Lambda(\mathbb{R}, \alpha, \mu)$, $0<\alpha<1$. *The extended functionals, denoted by ϕ_λ again, are given by*

$$(3.7)\qquad \langle \phi_\lambda, f\rangle = -\lambda \int_{-\infty}^{\infty} \frac{f(x)-f(x)}{x-\lambda}\mu(dx)+f(\lambda), \qquad \lambda \in \mathbb{R}, \quad f\in \Lambda(\mathbb{R}, \alpha, \mu),$$

or, equivalently,

$$(3.8)\qquad \langle \phi_\lambda, f\rangle = \lambda \oint_{-\infty}^{\infty} \frac{f(\lambda)}{\lambda - x}\mu(dx)+\omega(\lambda)f(\lambda), \qquad \lambda \in \mathbb{R}, \quad f\in \Lambda(\mathbb{R}, \alpha, \mu).$$

Similarly, the functionals $\chi_x (x \in \mathbb{R})$ defined on $P(\mathbb{R})$ by equation (3.2) *can be extended uniquely by continuity to* $\Lambda(\mathbb{R}, \alpha, \nu)$, $0<\alpha<1$. *The extended functionals, denoted by χ_x again, are given by*

$$(3.9)\qquad \langle \chi_x, \tilde f\rangle = x \int_{-\infty}^{\infty} \frac{\tilde f(\lambda)-\tilde f(x)}{\lambda - x}\nu(d\lambda)+(Pe)(x)\tilde f(x), \qquad x \in \mathbb{R}, \quad \tilde f\in \Lambda(\mathbb{R}, \alpha, \nu),$$

or, equivalently,

$$(3.10)\quad \langle \chi_x, \tilde f\rangle = x \int_{-\infty}^{\infty} \frac{\tilde f(\lambda)}{\lambda - x}\,\frac{\mu(d\lambda)}{\omega^+\omega^-(\lambda)}+\frac{\omega(x)}{\omega^+\omega^-(x)}\tilde f(x), \qquad x \in \mathbb{R}, \quad \tilde f\in \Lambda(\mathbb{R}, \alpha, \nu).$$

Proof. The expression in the right member of (3.7) is well defined for any $f \in \Lambda(\mathbb{R}, \alpha, \mu)$. It therefore defines a linear functional on $\Lambda(\mathbb{R}, \alpha, \mu)$ for each $\lambda \in \mathbb{R}$, which is an extension of the functional ϕ_λ on $P(\mathbb{R})$. We denote this extension by ϕ_λ again and prove that ϕ_λ is, in fact, a continuous linear functional on $\Lambda(\mathbb{R}, \alpha, \mu)$. Let $K_n = [-n, n]$ for $n = 1, 2, \cdots$. For a fixed $\lambda \in \mathbb{R}$ there certainly exists a value of n, $n = n_\lambda$ say, such that $|\lambda| \leqq n_\lambda - 1$. Let $K_\lambda = K_{n_\lambda}$. Then

$$\begin{aligned}
|\langle \phi_\lambda, f\rangle| &\leqq |\lambda| \int_{K_\lambda} \frac{|f(x)-f(\lambda)|}{|x-\lambda|}\mu(dx) + |\lambda| \int_{\mathbb{R}\setminus K_\lambda} |f(x)|\mu(dx) + |\lambda|\,|f(\lambda)| \int_{\mathbb{R}\setminus K_\lambda} \mu(dx) + |f(\lambda)| \\
&\leqq \left\{ \sup_{x,y\in K_\lambda} \frac{|f(x)-f(y)|}{|x-y|^\alpha}\right\} |\lambda| \int_{K_\lambda} |x-\lambda|^{-1+\alpha}\mu(dx) + \|f\|_\mu |\lambda| \left(\int_{-\infty}^{\infty} \mu(dx)\right)^{1/2} \\
&\quad + \{\sup_{x\in K_\lambda} |f(x)|\}\left\{|\lambda| \int_{-\infty}^{\infty} \mu(dx) + 1\right\} \\
&\leqq C(\lambda)(\|f\|_\mu + \|f\|_{K_\lambda, \alpha})
\end{aligned}$$

for some positive constant $C(\lambda)$. Hence, if $f \to 0$ in $\Lambda(\mathbb{R}, \alpha, \mu)$, then $\langle \phi_\lambda, f\rangle \to 0$ in $\mathbb{C}$, i.e., ϕ_λ is a continuous linear functional in $\Lambda(\mathbb{R}, \alpha, \mu)$. That ϕ_λ is the unique extension to $\Lambda(\mathbb{R}, \alpha, \mu)$ of the original functional defined on $P(\mathbb{R})$ follows from the fact that $P(\mathbb{R})$ is dense in $\Lambda(\mathbb{R}, \alpha, \mu)$ $(0<\alpha<1)$.

Similarly, the expression in the right member of (3.9), where the measure ν is given in [1, Lemma 4.5], is well defined for any $\tilde f \in \Lambda(\mathbb{R}, \alpha, \nu)$. It therefore defines a linear functional on $\Lambda(\mathbb{R}, \alpha, \nu)$ for each $x \in \mathbb{R}$, which is an extension of the functional χ_x on $P(\mathbb{R})$. We denote this extension by χ_x again. The proof that χ_x is, in fact, a continuous linear functional on $\Lambda(\mathbb{R}, \alpha, \nu)$ is entirely similar to the proof above that ϕ_λ is a continuous linear functional on $\Lambda(\mathbb{R}, \alpha, \mu)$. □

Finally, in order to obtain appropriate extensions of ϕ_λ to $\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$ and of χ_x to $\Lambda(\mathbb{R}_-,\alpha,\nu)\oplus\Lambda(\mathbb{R}_+,\alpha,\nu)$ we define

$$\langle\phi_\lambda, e_+\rangle=\begin{cases}\lambda\displaystyle\int_0^\infty\frac{\mu(dx)}{\lambda-x} & \lambda\in\mathbb{R}_-,\\[2ex] -\lambda\displaystyle\int_{-\infty}^0\frac{\mu(dx)}{\lambda-x}+1 & \lambda\in\mathbb{R}_+,\end{cases}\tag{3.11}$$

and

$$\langle\chi_x,\tilde{e}_+\rangle=\begin{cases}x\displaystyle\int_0^\infty\frac{\nu(d\lambda)}{\lambda-x} & x\in\mathbb{R}_-,\\[2ex] -x\displaystyle\int_{-\infty}^0\frac{\nu(d\lambda)}{\lambda-x}+(Pe)(x) & x\in\mathbb{R}_+.\end{cases}\tag{3.12}$$

We then have the following result.

THEOREM 3.2. *The linear functionals ϕ_λ ($\lambda\in\mathbb{R}$) defined on $\Lambda(\mathbb{R},\alpha,\mu)$ in the previous theorem can be extended to $\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$. The extended functionals, denoted by ϕ_λ again, are given by the same expressions* (3.7) *and* (3.8). *Similarly, the linear functionals χ_x ($x\in\mathbb{R}$) defined on $\Lambda(\mathbb{R},\alpha,\nu)$ in the previous theorem can be extended to $\Lambda(\mathbb{R}_-,\alpha,\nu)\oplus\Lambda(\mathbb{R}_+,\alpha,\nu)$. The extended functionals, denoted by χ_x again, are given by the same expressions* (3.9) *and* (3.10).

Proof. Let $f\in\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$. Then there exist a (unique) $h\in\Lambda(\mathbb{R},\alpha,\mu)$ and a (unique) $a=(f(0+)-f(0-))\in\mathbb{C}$, such that $f=h+ae_+$. We define

$$\langle\phi_\lambda,f\rangle=\langle\phi_\lambda,h\rangle+a\langle\phi_\lambda,e_+\rangle,\qquad\lambda\in\mathbb{R}.$$

We observe that the expressions in the right members of (3.7) and (3.8) are well defined when $f=e_+$ and equal to $\langle\phi_\lambda,e_+\rangle$ as defined by (3.11). Hence, (3.7) and (3.8) can be used to calculate $\langle\phi_\lambda,f\rangle$ for $f\in\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$. This proves the first part of the theorem. The second part can be proven in a similar manner. □

Notice that, at this point, there are two functionals ϕ_0 and two functionals χ_0.

In the next section we investigate the relation between the functionals ϕ_λ ($\lambda\in\mathbb{R}$) and χ_x ($x\in\mathbb{R}$) and the transformations F and F^{-1}.

4. The transformations F amd F^{-1} on Lipschitz spaces. We now turn to a detailed investigation of the transformation F and its inverse on the Lipschitz spaces introduced in the previous section. To this end we first investigate the relation between the transformation F and its inverse and the functionals ϕ_λ and χ_x, respectively.

THEOREM 4.1. *On $\Lambda(\mathbb{R},\mu)$ the composite transformation FP is represented by the functionals $\{\phi_\lambda:\lambda\in\mathbb{R}\}$. On $\Lambda(\mathbb{R},\nu)$ the transformation F^{-1} is represented by the functionals $\{\chi_x: x\in\mathbb{R}\}$.*

Proof. Let $f\in\Lambda(\mathbb{R},\mu)$. Then $f\in\Lambda(\mathbb{R},\alpha,\mu)$ for some α, $0<\alpha\leqq 1$. For $\alpha=1$, $\Lambda(\mathbb{R},\alpha,\mu)$ contains only those elements of $L_2(\mathbb{R},\mu)$ which are constant on $\mathbb{R}$. For such elements we certainly have $FPf(\lambda)=\langle\phi_\lambda,f\rangle$, ν-a.e. on $\mathbb{R}$. If $f\in\Lambda(\mathbb{R},\alpha,\mu)$ for some α, $0<\alpha<1$, then there exists a sequence $\{p_n: n=1,2,\cdots\}$ with $p_n\in P(\mathbb{R})$ such that $p_n\to f$ as $n\to\infty$ in the lipschitz topology. Since the lipschitz topology is stronger than the topology induced by the norm $\|\cdot\|_\mu$ in $\Lambda(\mathbb{R},\alpha,\mu)$, $\|p_n-f\|_\mu\to 0$ as $n\to\infty$. Consequently, $\|FP(p_n-f)\|_\nu\to 0$ as $n\to\infty$, so there exists a subsequence $\{p_{n_k}: k=1,2,\cdots\}$ such that $FPp_{n_k}(\lambda)\to FPf(\lambda)$ ν-almost everywhere on $\mathbb{R}$ as $k\to\infty$. Now,

$FPp_{n_k}(\lambda)=\langle\phi_\lambda, p_{n_k}\rangle$ for all $\lambda\in\mathbb{R}$, so

$$FPf(\lambda)=\lim_{k\to\infty}\langle\phi_\lambda, p_{n_k}\rangle=\lim_{n\to\infty}\langle\phi_\lambda, p_n\rangle=\langle\phi_\lambda, f\rangle$$

ν-a.e. on $\mathbb{R}$. In the same way we prove that if $\tilde{f}\in\Lambda(\mathbb{R},\nu)$ then

$$F^{-1}\tilde{f}(x)=\langle\chi_x,\tilde{f}\rangle$$

μ-a.e. on $\mathbb{R}$. $\square$

If $f\in\Lambda(\mathbb{R},\mu)$ then FPf, as an element of $L_2(\mathbb{R},\nu)$, may be identified with the function with values $\langle\phi_\lambda, f\rangle$, $\lambda\in\mathbb{R}$. In the following theorem we will show that this function is a locally lipschitz continuous function on $\mathbb{R}$. Hence, FPf may be taken to be continuous and we have the identity

$$FPf(\lambda)=\langle\phi_\lambda, f\rangle,\qquad \lambda\in\mathbb{R},\quad f\in\Lambda(\mathbb{R},\mu). \tag{4.1}$$

Similarly

$$F^{-1}\tilde{f}(x)=\langle\chi_x,\tilde{f}\rangle,\qquad x\in\mathbb{R},\quad \tilde{f}\in\Lambda(\mathbb{R},\nu). \tag{4.2}$$

THEOREM 4.2. *If $0<\alpha<1$ and $f\in\Lambda(\mathbb{R},\alpha,\mu)$, then $\langle\phi_\lambda, f\rangle$ $(\lambda\in\mathbb{R})$ defines a function in $\Lambda(\mathbb{R},\beta,\nu)$ for each β satisfying $0<\beta<\alpha$. The composite map FP: $\Lambda(\mathbb{R},\alpha,\mu)\to\Lambda(\mathbb{R},\beta,\nu)$ is continuous $(0<\beta<\alpha)$. Similarly, if $0<\alpha<1$ and $\tilde{f}\in\Lambda(\mathbb{R},\alpha,\nu)$, then $\langle\chi_x,\tilde{f}\rangle$ $(x\in\mathbb{R})$ defines a function in $\Lambda(\mathbb{R},\beta,\mu)$ for each β satisfying $0<\beta<\alpha$. The map F^{-1}: $\Lambda(\mathbb{R},\alpha,\nu)\to\Lambda(\mathbb{R},\beta,\mu)$ is continuous $(0<\beta<\alpha)$.*

Proof. We prove the theorem only for FP: $\Lambda(\mathbb{R},\alpha,\mu)\to\Lambda(\mathbb{R},\beta,\nu)$, since the proof for F^{-1}: $\Lambda(\mathbb{R},\alpha,\nu)\to\Lambda(\mathbb{R},\beta,\mu)$ is entirely analogous.

Let $f\in\Lambda(\mathbb{R},\alpha,\mu)$. Then, by (4.1), $\langle\phi_\cdot, f\rangle=FPf$, so $\langle\phi_\cdot, f\rangle\in L_2(\mathbb{R},\nu)$ and

$$\|\langle\phi_\cdot, f\rangle\|_\nu=\|Pf\|_\mu\leqq\|f\|_\mu. \tag{4.3}$$

Given any compact subset K' of $\mathbb{R}$, we can certainly find another compact subset K of $\mathbb{R}$ such that $K'\subset K$ and $d=\{\sup|x-y|: x\in K', y\in\mathbb{R}\backslash K\}>0$. For $\lambda\in K'$ we have the inequalities

$$\begin{aligned}|\langle\phi_\lambda, f\rangle|&\leqq|\lambda|\int_K\frac{|f(x)-f(\lambda)|}{|x-\lambda|}\mu(dx)+\frac{1}{d}\left[|\lambda|\int_{\mathbb{R}\backslash K}|f(x)|\mu(dx)+|\lambda||f(\lambda)|\int_{\mathbb{R}\backslash K}\mu(dx)\right]+|f(\lambda)|\\
&\leqq\left\{\sup_{\lambda\in K'}\sup_{x\in K}\frac{|f(x)-f(\lambda)|}{|x-\lambda|^\alpha}\right\}\left\{\sup_{\lambda\in K'}|\lambda|\int_K|x-\lambda|^{-1+\alpha}\mu(dx)\right\}\\
&\quad+\frac{1}{d}\{\sup_{\lambda\in K'}|\lambda|\}\left[\int_{\mathbb{R}\backslash K}|f(x)|\mu(dx)+\{\sup_{\lambda\in K'}|f(\lambda)|\}\int_{\mathbb{R}\backslash K}\mu(dx)\right]+\sup_{\lambda\in K'}|f(\lambda)|\\
&\leqq\left\{\sup_{x,\lambda\in K}\frac{|f(x)-f(\lambda)|}{|x-\lambda|^\alpha}\right\}\left\{\sup_{\lambda\in K'}|\lambda|\int_K|x-\lambda|^{-1+\alpha}\mu(dx)\right\}+\frac{1}{d}\{\sup_{\lambda\in K'}|\lambda|\}\\
&\quad\times[\|f\|_\mu+\sup_{\lambda\in K'}|f(\lambda)|]+\sup_{\lambda\in K'}|f(\lambda)|,\end{aligned}$$

so

$$\sup_{\lambda\in K'}|\langle\phi_\lambda, f\rangle|\leqq C_1(\|f\|_\mu+\|f|K\|_{K,\alpha}) \tag{4.4}$$

for some positive constant C_1. Furthermore, for any pair of points $\lambda, \xi \in K'$,

$$
\begin{aligned}
|\langle\phi_\lambda, f\rangle - \langle\phi_\xi, f\rangle| \leqq & \left|\lambda \int_K \frac{f(x)-f(\lambda)}{x-\lambda}\mu(dx) - \xi\int_K \frac{f(x)-f(\xi)}{x-\xi}\mu(dx)\right| \\
& + \frac{|\lambda-\xi|}{d^2}\{\sup_{\lambda\in K'}|\lambda|\}\left[\int_{\mathbb{R}\setminus K}|f(x)|\mu(dx) + \{\sup_{\lambda\in K'}|f(\lambda)|\}\int_{\mathbb{R}\setminus K}\mu(dx)\right] \\
& + f(\lambda) - f(\xi).
\end{aligned} \tag{4.5}
$$

The expression between the square brackets is estimated as before, so

$$
\begin{aligned}
& \frac{|\lambda-\xi|}{d^2}\{\sup_{\lambda\in K'}|\lambda|\}\left[\int_{\mathbb{R}\setminus K}|f(x)|\mu(dx) + \{\sup_{\lambda\in K'}|f(\lambda)|\}\int_{\mathbb{R}\setminus K}\mu(dx)\right] \\
& \leqq C_2(\|f\|_\mu + \|f|K\|_{K,\alpha})|\lambda-\xi|
\end{aligned} \tag{4.6}
$$

for some positive constant C_2. The difference of the two integrals over K can be estimated by means of a slight modification of the proof of the Plemelj–Privalov theorem given in [5, § 18], along the same lines as the proof of the inequalities (6.14) of [2]. Since the modifications are quite obvious we omit the details and give only the result, which reads

$$
\begin{aligned}
& \left|\lambda \int_K \frac{f(x)-f(\lambda)}{x-\lambda}\mu(dx) - \xi\int_K \frac{f(x)-f(\xi)}{x-\xi}\mu(dx)\right| \\
& \leqq \begin{cases} C_3\|f|K\|_{K,\alpha}|\lambda-\xi|^\alpha & \text{if } |\lambda-\xi| \geqq \frac{1}{2}\operatorname{diam}(K) \\ C_4\|f|K\|_{K,\alpha}|\lambda-\xi|^\alpha(1-\ln|\lambda-\xi|) & \text{if } |\lambda-\xi| < \frac{1}{2}\operatorname{diam}(K), \end{cases}
\end{aligned} \tag{4.7}
$$

where C_3 and C_4 are some positive constants and $\operatorname{diam}(K) = \sup_{x,y\in K}|x-y|$. Combining (4.5), (4.6) and (4.7) we obtain the asymptotic relation

$$
\langle\phi_\lambda, f\rangle - \langle\phi_\xi, f\rangle = o(|\lambda-\xi|^\beta) \quad \text{as } |\lambda-\xi| \to 0, \tag{4.8}
$$

for each β satisfying $0<\beta<\alpha$. From (4.4) and (4.8) we see that $\langle\phi_\cdot, f\rangle|K'$ is lipschitz continuous with exponent β. Since the compact subset K' was arbitrarily chosen on $\mathbb{R}$ we conclude that $\langle\phi_\cdot, f\rangle \in \Lambda(\mathbb{R}, \beta, \nu)$ for each β satisfying $0<\beta<\alpha$.

From (4.3) through (4.7) and the definition of the seminorms in $\Lambda(\mathbb{R}, \alpha, \mu)$ and $\Lambda(\mathbb{R}, \alpha, \nu)$ it follows that for any compact subset K' of $\mathbb{R}$,

$$
p_{K',\beta,\nu}(FPf) \leqq Cp_{K,\alpha,\mu}(f),
$$

where K is any compact subset of $\mathbb{R}$ which contains K' in its interior. Hence, $FPf\to 0$ in $\Lambda(\mathbb{R}, \beta, \nu)$ if $f\to 0$ on $\Lambda(\mathbb{R}, \alpha, \mu)$, i.e., the map $FP\colon \Lambda(\mathbb{R}, \alpha, \mu) \to \Lambda(\mathbb{R}, \beta, \nu)$ is continuous ($0<\beta<\alpha$). □

The following statement is an immediate consequence of Theorem 4.2.

COROLLARY. *F is a topological isomorphism from $\Lambda(\mathbb{R}, \mu)\cap H$ onto $\Lambda(\mathbb{R}, \nu)$.*

Next we turn our attention to functions which have a jump discontinuity at the origin.

THEOREM 4.3. *On $\Lambda(\mathbb{R}_-, \mu)\oplus\Lambda(\mathbb{R}_+, \mu)$ the transformation FP is represented by the functionals $\{\phi_\lambda\colon \lambda\in\mathbb{R}\}$. On $\Lambda(\mathbb{R}_-, \nu)\oplus\Lambda(\mathbb{R}_+, \nu)$, F^{-1} is represented by the functionals $\{\chi_x\colon x\in\mathbb{R}\}$.*

Proof. First, consider any $f\in\Lambda(\mathbb{R}_-, \mu)\oplus\Lambda(\mathbb{R}_+, \mu)$ which is in $D(T)$. Then $TPf\in\Lambda(\mathbb{R}, \mu)$ and by (4.1) we have the identity $\langle\phi_\lambda, TPf\rangle = FPTPf(\lambda)$ for all $\lambda\in\mathbb{R}$. Since $FPTPf = \tilde{T}FPf$ for $f\in D(T)$, it follows that $\langle\phi_\lambda, TPf\rangle = \lambda FPf(\lambda)$, ν-a.e. on $\mathbb{R}$. On the other hand, from (3.7) we readily obtain the identity $\langle\phi_\lambda, TPf\rangle = \lambda\langle\phi_\lambda, f\rangle$, which holds for all $\lambda \in \mathbb{R}$. Thus, $FPf(\lambda)=\langle\phi_\lambda, f\rangle$, ν-a.e. on $\mathbb{R}\backslash\{0\}$, at least for any $f\in D(T)$. However, $D(T)$ contains the set $P(\mathbb{R})$ which is dense in $\Lambda(\mathbb{R},\mu)$, and the element e_+, so $D(T)$ is dense in $\Lambda(\mathbb{R},\mu)\oplus\mathbb{C}e_+=\Lambda(\mathbb{R}_-,\mu)\oplus\Lambda(\mathbb{R}_+,\mu)$ and the identity $FPf(\lambda)=\langle\phi_\lambda, f\rangle$, ν-a.e. on $\mathbb{R}\backslash\{0\}$ can be extended by continuity to all $f\in\Lambda(\mathbb{R}_-,\mu)\oplus\Lambda(\mathbb{R}_+,\mu)$.

Next, let $\tilde f\in\Lambda(\mathbb{R}_-,\nu)\oplus\Lambda(\mathbb{R}_+,\nu)$ be in $D(\tilde T)$. Then $\tilde T\tilde f\in\Lambda(\mathbb{R},\nu)$, so by (4.2) we have the identity $\langle\chi_x, \tilde T\tilde f\rangle=F^{-1}\tilde T\tilde f(x)$ for all $x\in\mathbb{R}$. Since $F^{-1}\tilde T=BF^{-1}$ on $D(\tilde T)$ it follows that $\langle\chi_x,\tilde T\tilde f\rangle=BF^{-1}\tilde f(x)=xF^{-1}\tilde f(x)-2(TF^{-1}\tilde f, Te)_\mu x$, μ-a.e. on $\mathbb{R}$. On the other hand, from (3.9) we obtain the identity $\langle\chi_x,\tilde T\tilde f\rangle=x\langle\chi_x,\tilde f\rangle+(\tilde f,\tilde e)_\nu x$, which holds for all $x\in\mathbb{R}$. But $(\tilde f,\tilde e)_\nu=(F^{-1}\tilde f, Pe)_\mu=-2(F^{-1}\tilde f, T^2e)_\mu=-2(TF^{-1}\tilde f, Te)_\mu$, so we find that $F^{-1}\tilde f(x)=\langle\chi_x,\tilde f\rangle$, μ-a.e. in $\mathbb{R}\backslash\{0\}$, at least for any $\tilde f\in D(\tilde T)$. But $D(\tilde T)$ contains the set $P(\mathbb{R})$, which is dense in $\Lambda(\mathbb{R},\nu)$, and the element $\tilde e_+$, so $D(\tilde T)$ is dense in $\Lambda(\mathbb{R},\nu)\oplus\mathbb{C}\tilde e_+=\Lambda(\mathbb{R}_-,\nu)\oplus\Lambda(\mathbb{R}_+,\nu)$ and the identity $F^{-1}\tilde f(x)=\langle\chi_x,\tilde f\rangle$, μ-a.e. on $\mathbb{R}\backslash\{0\}$ can be extended to all $\tilde f\in\Lambda(\mathbb{R}_-,\nu)\oplus\Lambda(\mathbb{R}_+,\nu)$. □

If $f\in\Lambda(\mathbb{R}_-,\mu)\oplus\Lambda(\mathbb{R}_+,\mu)$ then FPf, as an element of $L_2(\mathbb{R},\nu)$ may be identified with the function with values $\langle\phi_\lambda, f\rangle$, $\lambda\in\mathbb{R}$. In the following theorem we will show that this function is a locally lipschitz continuous function on $\mathbb{R}_-$ and $\mathbb{R}_+$, so FPf may be taken to be continuous on $\mathbb{R}$ except at the origin and we have

$$FPf(\lambda)=\langle\phi_\lambda, f\rangle,\qquad \lambda\in\mathbb{R}\backslash\{0\},\quad f\in\Lambda(\mathbb{R}_-,\mu)\oplus\Lambda(\mathbb{R}_+,\mu). \tag{4.9}$$

Similarly,

$$F^{-1}\tilde f(x)=\langle\chi_x,\tilde f\rangle,\qquad x\in\mathbb{R}\backslash\{0\},\quad \tilde f\in\Lambda(\mathbb{R}_-,\nu)\oplus\Lambda(\mathbb{R}_+,\nu). \tag{4.10}$$

THEOREM 4.4. *If $0<\alpha<1$ and $f\in\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$ then $\langle\phi_\lambda, f\rangle$ $(\lambda\in\mathbb{R})$ defines a function in $\Lambda(\mathbb{R}_-,\beta,\nu)\oplus\Lambda(\mathbb{R}_+,\beta,\nu)$ for each β satisfying $0<\beta<\alpha$; FP is a continuous map from the first into the second locally convex vector space. Similarly, if $0<\alpha<1$ and $f\in\Lambda(\mathbb{R}_-,\alpha,\nu)\oplus\Lambda(\mathbb{R}_+,\alpha,\nu)$ then $\langle\chi_x, f\rangle$ $(x\in\mathbb{R})$ defines a function in $\Lambda(\mathbb{R}_-,\beta,\mu)\oplus\Lambda(\mathbb{R}_+,\beta,\mu)$ for each β satisfying $0<\beta<\alpha$; F^{-1} is a continuous map from the first into the second locally convex vector space.*

Proof. We prove only the first half of the theorem. Take any $f\in\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$. We can decompose f, $f=h+ae_+$, with $h\in\Lambda(\mathbb{R},\alpha,\mu)$ and $a=(f(0+)-f(0-))\in\mathbb{C}$. In view of Theorem 4.2, $\langle\phi_\lambda, h\rangle$ $(\lambda\in\mathbb{R})$ defines a function in $\Lambda(\mathbb{R},\beta,\nu)$ for each β satisfying $0<\beta<\alpha$. The (regular) integrals occurring in (3.11) define functions which are lipschitz continuous with exponent β on $\mathbb{R}_-$ and $\mathbb{R}_+$ for any β satisfying $0<\beta<1$, so $\langle\phi_., e_+\rangle\in\Lambda(\mathbb{R}_-,\beta,\nu)\oplus\Lambda(\mathbb{R}_+,\beta,\nu)$ for $0<\beta<1$. Combining these two results we see that $\langle\phi_., f\rangle\in\Lambda(\mathbb{R}_-,\beta,\nu)\oplus\Lambda(\mathbb{R}_+,\beta,\nu)$ for any β satisfying $0<\beta<\alpha$.

Next, choose K and K' as in the proof of Theorem 4.2. Again using the decomposition $f=h+ae_+$, we have the inequality

$$p_{K',\beta,\nu}(FPf)\leqq Cp_{K,\alpha,\mu}(h)+|a|p_{K',\beta,\nu}(FPe_+)$$

for some positive constant C. Both $p_{K,\alpha,\mu}(h)$ and $|a|$ can be estimated by $3p_{K,\alpha,\mu}(f)$, so there exists a positive constant $C'=3(C+p_{K',\beta,\nu}(FPe_+))$ such that

$$p_{K',\beta,\nu}(FPf)\leqq C'p_{K,\alpha,\mu}(f),$$

which proves that FP is a continuous map from $\Lambda(\mathbb{R}_-,\alpha,\mu)\oplus\Lambda(\mathbb{R}_+,\alpha,\mu)$ into $\Lambda(\mathbb{R}_-,\beta,\nu)\oplus\Lambda(\mathbb{R}_+,\beta,\nu)$ $(0<\beta<\alpha<1)$. □

As an immediate consequence of this theorem we have the following result.

COROLLARY. *F is a topological isomorphism from $(\Lambda(\mathbb{R}_-,\mu)\oplus\Lambda(\mathbb{R}_+,\mu))\cap H$ onto $\Lambda(\mathbb{R}_-,\nu)\oplus\Lambda(\mathbb{R}_+,\nu)$.*

5. A full-range completeness theorem. In his study of the slip flow problem Cercignani [4] formulated a full-range completeness theorem for which we are now able to give a precise interpretation.

THEOREM 5.1. *An arbitrary function* $f \in \Lambda(\mathbb{R}, \mu)$ *admits the full-range expansion*

$$f(x) = 2(f, Te)_\mu x + 2(f, e)_\mu x^2 + \langle \chi_x, FPf \rangle, \qquad x \in \mathbb{R},$$

where $FPf(\lambda) = \langle \phi_\lambda, f \rangle$ *for* $\lambda \in \mathbb{R}$. *Here*, $\{\phi_\lambda : \lambda \in \mathbb{R}\}$ *and* $\{\chi_x : x \in \mathbb{R}\}$ *are families of continuous linear functionals on* $\Lambda(\mathbb{R}, \mu)$ *and* $\Lambda(\mathbb{R}, \nu)$, *respectively, which are given by the expressions* (3.7) *and* (3.9).

Proof. The theorem is an immediate consequence of the definition of the isomorphism $\hat{F}: L_2(\mathbb{R}, \mu) \to \mathbb{C}^2 \oplus L_2(\mathbb{R}, \nu)$, cf. [1, Thm. 4.3], and of Theorems 4.1 and 4.2 and the Corollary to Theorem 4.2. $\square$

Acknowledgment. The final form of the manuscript was prepared during the Summer Semester 1976/77, while the author was on foreign assignment to the Mathematics Department of the University of Vienna (Austria).

REFERENCES

[1] H. G. KAPER, *Boundary value problems of mixed type arising in the kinetic theory of gases*, this Journal, 10, pp. 161–178.

[2] R. J. HANGELBROEK, *A functional-analytic approach to the linear transport equation*, Ph.D. thesis, Groningen, the Netherlands, 1973; TTSP, 5 (1976), pp. 1–85.

[3] C. G. LEKKERKERKER, *The linear transport equation. The degenerate case c = 1*, Proc. Roy. Soc. Edinburgh, Sect. A, 75A (1975/76), pp. 259–282.

[4] C. CERCIGANI, *Elementary solutions of the linearized gas-dynamics Boltzmann equation and their application to the slip-flow problem*, Ann. Physics, 20 (1962), pp. 219–233.

[5] N. J. MUSCHELISCHVILI, *Singulaere Integralgleichungen*, Akademie-Verlag, Berlin, 1965.

[6] K. MAURIN, *General Eigenfunctions Expansions and Unitary Representations of Topological Groups*, Warszawa, 1968.

[7] A. F. TIMAN, *Theory of Approximation of Functions of a Real Variable*, Pergamon Press, Macmillan, New York, 1963.

[8] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, eds., NBS Appl. Math. Ser. vol. 55, National Bureau of Standards, Washington, D.C., 1964.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/10019 $01.00/0

THE SUMMATION OF SERIES OF HYPERBOLIC FUNCTIONS*

I. J. ZUCKER†

Abstract. A method of summing sixteen series of hyperbolic functions is presented. The method is based on expressing the series in terms of the nome, q, of the Jacobian elliptic functions. The q-series thus obtained are then expressed in closed form in terms of complete elliptic integrals of the first and second kind and the corresponding modulus. It is shown that when a certain parameter in these series is the square root of a rational number the series are summable in terms of Γ functions.

1. Introduction. Recently Ling [1] proposed the summation of four series of hyperbolic functions, namely

$$\mathrm{I}_s(c)=\sum_1^\infty \operatorname{cosech}^s(n\pi c),\qquad \mathrm{II}_s(c)=\sum_1^\infty \operatorname{sech}^s(n\pi c),$$
(1)
$$\mathrm{III}_s(c)=\sum_1^\infty \operatorname{cosech}^s[(2n-1)\pi c/2],\qquad \mathrm{IV}_s(c)=\sum_1^\infty \operatorname{sech}^s[(2n-1)\pi c/2].$$

Ling [1] used Weierstrassian elliptic functions to obtain closed forms for these series for even s, and summed the series in terms of Γ functions for $c=1$, $\sqrt{3}$ and $1/\sqrt{3}$. The object of this communication is to extend these results so that

(a) II_s and IV_s may be expressed in closed forms for all integral s,

(b) many other series of hyperbolic functions may be expressed in closed form e.g.

$$\mathrm{V}_s=\sum_1^\infty n^s \operatorname{cosech}(n\pi c),\qquad \mathrm{VI}_s=\sum_1^\infty n^s \operatorname{sech}(n\pi c),$$
(2)
$$\mathrm{VII}_s=\sum_1^\beta n^s \operatorname{cosech}[(2n-1)\pi c/2],\qquad \mathrm{VIII}_s=\sum_1^\infty n^s \operatorname{sech}[(2n-1)n\pi c/2],$$
$$\mathrm{IX}_s=\sum_1^\infty (-1)^{n+1}\operatorname{cosech}^s(n\pi c),\qquad \text{etc.}$$

(c) all series which can be put in closed form may be summed in terms of Γ functions whenever c is the square root of a rational number. The results for closed forms of various series are exhibited in Tables 1(i–xvi) in terms of K, E, k and k' where K and E are the complete elliptic integrals of the first and second kind respectively, k is the modulus of the elliptic integrals and $k^2+k'^2=1$.

2. Deduction of results in Table 1. Let

$$K=\int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}},\qquad k^2+k'^2=1,\quad K'=K(k'),$$
(3)
$$E=\int_0^1\left(\frac{1-k^2x^2}{1-x^2}\right)^{1/2}dx,\qquad \frac{K'}{K}=c,\qquad q=e^{-\pi c}.$$

All the hyperbolic series may be expressed in terms of q-series, and all the q-series may be generated from Fourier expansions of the Jacobian elliptic functions or their

* Received by the editors May 9, 1977.

† Department of Physics, University of Surrey, Guildford, Surrey, GU2 5XH, England.

squares. As an example consider

$$\mathrm{I}_{2s}=\sum_1^\infty \operatorname{cosech}^{2s}(n\pi c)=\sum_1^\infty\left(\frac{2}{q^{-n}-q^n}\right)^{2s}$$

(4)
$$=2^{2s}\sum_1^\infty\frac{q^{2ns}}{(1-q^{2n})^{2s}}=2^{2s}\sum_{m=0}^\infty\frac{(m+2s-1)!}{m!(2s-1)!}\sum_1^\infty q^{2n(m+s)}$$
$$=\frac{2^{2s}}{(2s-1)!}\sum_{m=0}^\infty\frac{(m+2s-1)!}{m!}\frac{q^{2(m+s)}}{1-q^{2(m+s)}}.$$

Now let $(m+2s-1)!/m!\equiv(m+s)^{2s-1}+\alpha_1(s)(m+s)^{2s-3}+\alpha_2(s)(m+s)^{2s-5}\cdots$.

For $m=1-s,\quad 0=1+\alpha_1(s)+\alpha_2(s)\cdots,$

(5) $m=2-s,\quad 0=2^{2s-1}+\alpha_1(s)2^{2s-3}+\alpha_2(s)2^{2s-5}\cdots,$

$\vdots$

$m=-1,\quad 0=(s-1)^{2s-1}+\alpha_1(s)(s-1)^{2s-3}+\alpha_2(s)(s-1)^{2s-5}\cdots.$

These equations may be solved for $\alpha_m(s)$ to give the solutions

(6)
$$\alpha_m(s)=(-1)^m\sum_{j_1\neq j_2\neq\cdots\neq j_m=1}^{s-1}\prod_{m=1}^{m}j_m^2.$$

Thus

$$\alpha_1(s)=-\sum_{j=1}^{s-1}j^2=-\frac{1}{6}(s-1)s(2s-1),$$
$$\alpha_2(s)=\sum_{j_1\neq j_2=1}^{s-1}j_1^2j_2^2=\frac{1}{2}\left[\left(\sum_{j=1}^{s-1}j^2\right)^2-\sum_{j=1}^{s-1}j^4\right]$$
$$=\frac{(s-2)(s-1)s(2s-1)(2s-3)(5s+1)}{360}$$

and so on. Some values of $\alpha_m(s)$ are tabulated in Table 2. Hence (4) reduces to

(7)
$$\mathrm{I}_{2s}=\frac{2^{2s}}{(2s-1)!}[A_{2s-1}+\alpha_1(s)A_{2s-3}+\alpha_2(s)A_{2s-5}\cdots]$$

where

$$A_s=\sum_1^\infty\frac{n^sq^{2n}}{1-q^{2n}}.$$

Thus

$$\mathrm{I}_2=4A_1,\qquad \mathrm{I}_4=\tfrac{8}{3}(A_3-A_1),$$
$$\mathrm{I}_6=\tfrac{8}{15}(A_5-5A_3+4A_1),\qquad \mathrm{I}_8=\tfrac{16}{315}(A_7-14A_5+49A_3-36A_1)\quad\text{etc.}$$

The A_s are obtained from the Fourier expansion of the square of the Jacobian elliptic function *ns*. Whittaker and Watson [2, p. 535] quote Jacobi [3] and give

(8)
$$\left(\frac{2K}{\pi}\right)^2 ns^2\left(\frac{2Kx}{\pi}\right)=\frac{4K}{\pi^2}(K-E)+\operatorname{cosec}^2 x-8\sum_1^\infty\frac{nq^{2n}}{1-q^{2n}}\cos(2nx).$$

Expanding $\cos(2nx)$ as a power series then yields

$$\left(\frac{2K}{\pi}\right)ns^2\left(\frac{2Kx}{\pi}\right)=\frac{4K}{\pi^2}(K-E)+\operatorname{cosec}^2 x-8\sum_{s=0}^{\infty}(-1)^s A_{2s-1}\frac{(2x)^{2s}}{(2s)!}. \tag{9}$$

Now the power series exansions of the functions $sn(x)$, $cn(x)$ and $dn(x)$ are easily found. They are given to quite high orders by Hancock [4, p. 486]. The power series expansion of any Jacobian elliptic function may thus be deduced. For example.

$$sn(x)=x-(1+k^2)\frac{x^3}{3!}+(1+14k^2+k^4)\frac{x^5}{5!}-(1+135k^2+135k^4+k^6)\frac{x^7}{7!}\cdots.$$

Hence

$$\begin{aligned} ns(x)=\frac{1}{sn(x)}=\frac{1}{x}+(1+k^2)\frac{x}{3!}+(7-22k^2+7k^4)\frac{x^5}{3\cdot 5!}\\ +(31-15k^2-15k^4+31k^6)\frac{x^7}{3\cdot 7!}\cdots,\end{aligned}$$

$$ns^2(x)=\frac{1}{x^2}+\frac{1+k^2}{3}+(1-k^2+k^4)\frac{x^2}{15}+(1+k^2)(1-2k^2)(2-k^2)\frac{x^4}{189}\cdots. \tag{10}$$

Then by comparing coefficients of x in equation (9) using (10) in the left hand side one has

$$1-24A_1=\left(\frac{2K}{\pi}\right)^2\left(\frac{3E}{K}+k^2-2\right),$$

$$1+240A_3=\left(\frac{2K}{\pi}\right)^4(1-k^2+k^4),\ \text{etc.}$$

Ramanujan [5] has evaluated A_{2s-1} to A_{31}! From these results and by use of (7) the hyperbolic series I_s may be summed in closed form, that is in terms of K, E and k, for all even s. The series II_s, III_s, IV_s, IX_s, X_s, XI_s and XII_s may all be obtained in a similar manner, sometimes only for even s and sometimes for only odd s and occasionally for both. For odd s the coefficients $\beta_m(s)$ which appear in equations analogous to (7) are defined by

$$\beta_m(s)=(-1)^m\sum_{j_1\neq j_2\neq\cdots\neq j_m=1}^{s}\prod_{m=1}^{m}(2j_m-1)^2$$

and some values for these are given in Table 2.

The other series such as V_s may be put into closed form more easily. For example

$$\mathrm{V}_s=\sum_1^{\infty}n^s\operatorname{cosech}(n\pi c)=\sum_1^{\infty}\frac{2n^s}{q^{-n}-q^n}=2\sum_1^{\infty}\frac{n^s q^n}{1-q^{2n}}=2C_s \tag{11}$$

where the generating function for C_s is

$$\left(\frac{2K}{\pi}\right)dn^2\left(\frac{2Kx}{\pi}\right)=\frac{4KE}{\pi^2}+8\sum_{s=0}^{\infty}(-1)^s C_{2s+1}\frac{(2x)^{2s}}{(2s)!}. \tag{12}$$

Tables 1(i–xvi) thus show in order a q-series, its generating function (a Fourier expansion of a Jacobian elliptic function), the value of the q-series in terms of K, E, k and k' for a few values of s, and the hyperbolic series which can be summed in closed form with these series. The Fourier expansions for the squares of the elliptic functions used in Tables 1(i–iv) are all found in Jacobi [3]. The other Fourier expansions may be obtained from standard text books e.g. Abramowitz and Stegun [6, p. 575]. The values of the q-series for particular values of c are also given.

3. Summation of series in Γ functions when c^2 is rational. Whenever a series may be written in closed form in terms of K and k then the sum may be written in terms of Γ-functions and algebraic numbers whenever c^2 is rational. For when c^2 is rational a theorem of Abel states (Whittaker and Watson [2, p. 525]) that k is an algebraic number. Selberg and Chowla [7] further showed that when c^2 is rational K is expressible in terms of Γ functions. The results for K in terms of Γ functions when $c^2 = 1$, 3 and 4 are well known [2, pp. 524–526]. The result for $c^2 = 2$ is implicit in some work of Ramanujan [8] and has been given explicitly by Glasser and Wood [9]. Selberg and Chowla [7] gave the results for $c^2 = 5$ and 7. Zucker [10] has given K in Γ functions for all c^2 from 1 to 16 excluding $c^2 = 14$. These results are given in Table 3. Corresponding results for k^2 are given in Table 4.

Now whenever c^2 is rational E can be expressed in terms of K (Ramanujan [8]). This is not stated explicitly by him but his results imply that

$$E = \frac{\pi}{4Kc} + \frac{K}{3}[2 - k^2 + f_c(k)] \tag{13}$$

where $f_c(k)$ is an *algebraic* function of k depending on c. Ramanujan unfortunately does not give details of how to compute $f_c(k)$, but the author has deduced expressions of E in terms of K for several c and these are given in Table 5. It is clear that with sufficient labor E may be expressed in terms of K and hence in Γ functions. Hence whenever a hyperbolic series is expressible in closed form in terms of K, E, k and k' then it may be summed in terms of Γ functions and algebraic numbers.

4. Discussion. It is immediately evident from Tables 1(i–xvi) that all the series A_s to U_s inclusive can only be found in closed form for either s even or s odd, but never for both. Thus the hyperbolic series can usually only be found in closed form for either even or odd s. However, some curiosities appear. For example, II_{2s} is expressible in terms of B_{2s+1} and II_{2s+1} is expressible in terms of G_{2s} and expressions for B_{2s+1} and G_{2s} may be found in closed form. A similar situation occurs for IV_s. So both II_s and IV_s are expressible in closed form for all integrals s. But for I_s and III_s although expressions may be found for them for both even and odd s in terms of q-series—Tables 1(i), (iii), (v), (vii)—only for s even are the appropriate q-series expressible in closed form.

Whenever the factor $(1-2k^2)$ appears, the term including it will vanish when $2k^2 = 1$. For this value of k, $k = k'$ and $K = K'$, hence $c = 1$. Thus certain sums will vanish. For example from Table 1(xvi) it is observed that

$$2U_3(1) = \sum_1^{\infty} (-1)^{n+1}(2n-1)^3 \operatorname{sech}\left[(2n-1)\frac{\pi}{2}\right] = 0. \tag{14}$$

This result is part of a general theorem stated by Ramanujan [11] that

$$\sum_1^{\infty} (-1)^{n+1}(2n-1)^{4m-1} \operatorname{sech}\left[(2n-1)\frac{\pi}{2}\right] = 0 \tag{15}$$

for all positive integral m. This would suggest that a possible way of proving this general result would be to show that U_{4m-1} always contained a factor $(1-2k^2)$. We have confirmed this for U_7, in fact

$$4U_7=\left(\frac{2K}{\pi}\right)^8 kk'(1-2k^2)(1-136k^2-136k^4)$$

but we have not been able to prove the general result. The general result has been proved recently by Berndt [12] using a different approach.

Another result in Ramanujan's famous letter to Hardy [13] has its solution in the q-series considered here. This is

$$\frac{1^{13}}{e^{2\pi}-1}+\frac{2^{13}}{e^{4\pi}-1}+\frac{3^{13}}{e^{6\pi}-1}+\cdots=\frac{1}{24}. \tag{16}$$

This series is just $A_{13}(1)$ and (16) follows from the fact that A_{4n+1} always contains the factor $(1-2k^2)$; hence

$$A_{4n+1}(1)=0.$$

Indeed it follows that

$$\sum_{k=1}^{\infty}\frac{k^{4n+1}}{e^{2\pi k}-1}=\frac{B_{4n+2}}{4(2n+1)} \tag{17}$$

a result proved by Berndt [12] who quotes Glaisher [14] as first having given this result.

It is also evident in Ramanujan's letter that he had attempted to find closed form expressions for A_s when s is even, but was not successful. Thus in his letter he gives

$$\frac{1^2}{e^x-1}+\frac{2^2}{e^{2x}-1}+\frac{3^2}{e^{3x}-1}+\cdots=\frac{2}{x^3}\zeta(3)-\frac{1}{12x}-\frac{x}{1440}+\frac{x^3}{181440}\cdots. \tag{18}$$

The left hand side of (18) is essentially A_2.

The formula for U_5 yields a result similar to (18). U_5 contains the factor $(1-16k^2+16k^4)$ and this vanishes when $k^2=(2-\sqrt{3})/4$. For this value of k^2, $c=\sqrt{3}$; hence

$$\sum_1^{\infty}(-1)^{n+1}(2n-1)^5\operatorname{sech}[(2n-1)\sqrt{3}\pi/2]=0. \tag{19}$$

Other interesting results deducible from Table 1 are

$$\sum_1^{\infty}\operatorname{cosech}^2(n\pi)=\frac{1}{6}-\frac{1}{2\pi} \tag{20}$$

$$\sum_1^{\infty}(-1)^{n+1}n\operatorname{cosech}(n\pi)=\frac{1}{4\pi}=\frac{1}{2}\sum_1^{\infty}\operatorname{sech}^2\left[(2n-1)\frac{\pi}{2}\right] \tag{21}$$

$$\sum_1^{\infty}(-1)^{n+1}n^5\operatorname{cosech}(n\pi)=0. \tag{22}$$

Certain factors in Table 1 will obviously become zero for other values of k. For example, in Table 1(viii) it is apparent that J_4 vanishes when $k^2=1/5$ but we do not know what c is for this value of k. The determination of c when k is given seems an interesting problem.

Acknowledgments. The continuous stimulation and encouragement of Professor M. L. Glasser (Waterloo) is most gratefully acknowledged. Professor B. C. Berndt's (Illinois) comments and interest are much appreciated.

Note. Berndt (Private communication) comments that (19) is a special case of

$$\sum_1^\infty (-1)^{n+1}(2n-1)^{6m-1} \operatorname{sech}[(2n-1)\sqrt{3}\pi/2 = 0], \qquad m>0,$$

and (22) a special case of

$$\sum_1^\infty (-1)^{n+1} n^{4m+1} \operatorname{cosech}(n\pi) = 0, \qquad m>0,$$

both of which he has proved. Equation (21) goes back to Cauchy.

TABLE 1(i)

$$A_s(c) = A_s = \sum_{n=1}^\infty \frac{n^s q^{2n}}{1-q^{2n}}$$

$$\left(\frac{2K}{\pi}\right)^2 ns^2\left(\frac{2Kx}{\pi}\right) = \frac{4K}{\pi^2}(K-E) + \operatorname{cosec}^2 x - 8\sum_0^\infty (-1)^s A_{2s+1}\frac{(2x)^{2s}}{(2s)!}$$

$$1-24A_1 = \left(\frac{2K}{\pi}\right)^2\left(\frac{3E}{K}+k^2-2\right) \qquad A_1(1) = \frac{1}{24}-\frac{1}{8\pi}$$

$$1+240A_3 = \left(\frac{2K}{\pi}\right)^4(1-k^2+k^4)$$

$$1-504A_5 = \left(\frac{2K}{\pi}\right)^6\frac{1}{2}(1+k^2)(1-2k^2)(2-k^2) \qquad A_5(1) = 1/504$$

$$1+480A_7 = (1+240A_3)^2$$

$$\mathrm{I}_{2s} = \sum_1^\infty \operatorname{cosech}^{2s}(n\pi c) = \frac{2^{2s}}{(2s-1)!}\sum_{m=0}^{s-1} \alpha_m A_{2s-2m-1}$$

TABLE 1(ii)

$$B_s(c) = B_s = \sum_{n=1}^\infty \frac{(-1)^{n+1} n^s q^{2n}}{1-q^{2n}}$$

$$\left(\frac{2K}{\pi}\right)^2(1-k^2)\left[nc^2\left(\frac{2Kx}{\pi}\right)-1\right] = -\frac{4KE}{\pi^2} + \sec^2 x + 8\sum_{s=0}^\infty (-1)^s B_{2s+1}\frac{(2x)^{2s}}{(2s)!}$$

$$1+8B_1 = \frac{4EK}{\pi^2}$$

$$1-16B_3 = \left(\frac{2K}{\pi}\right)^4(1-k^2)$$

$$1+8B_5 = \left(\frac{2K}{\pi}\right)^6\frac{1}{2}(1-k^2)(2-k^2)$$

$$\mathrm{II}_{2s} = \sum_1^\infty \operatorname{sech}^{2s}(n\pi c) = (-1)^{s-1}\frac{2^{2s}}{(2s-1)!}\sum_{m=0}^{s-1} \alpha_m B_{2s-2m-1}$$

TABLE 1(iii)

$$C_s(c) = C_s = \sum_{n=1}^{\infty} \frac{n^s q^n}{1-q^{2n}}$$

$$\left(\frac{2K}{\pi}\right)^2 dn^2\left(\frac{2Kx}{\pi}\right) = \frac{4EK}{\pi^2} + 8 \sum_{s=0}^{\infty} (-1)^s C_{2s+1} \frac{(2x)^{2s}}{(2s)!}$$

$$8C_1 = \left(\frac{2K}{\pi}\right)^2 \left(1 - \frac{E}{K}\right)$$

$$16C_3 = \left(\frac{2K}{\pi}\right)^4 k^2$$

$$16C_5 = \left(\frac{2K}{\pi}\right)^6 k^2(1+k^2)$$

$$\mathrm{III}_{2s} = \sum_1^{\infty} \operatorname{cosech}^{2s}\left[(2n-1)\frac{\pi c}{2}\right] = \frac{2^{2s}}{(2s-1)!} \sum_{m=0}^{s-1} C_{2s-2m-1}$$

$$\mathrm{V}_s = \sum_1^{\infty} n^s \operatorname{cosech}(n\pi c) = 2C_s$$

TABLE 1(iv)

$$D_s(c) = D_s = \sum_1^{\infty} \frac{(-1)^{n+1} n^s q^n}{1-q^{2n}}$$

$$\left(\frac{2K}{\pi}\right)^2 (1-k^2) nd^2\left(\frac{2Kx}{\pi}\right) = \frac{4EK}{\pi^2} - 8 \sum_{s=0}^{\infty} (-1)^s D_{2s+1} \frac{(2x)^{2s}}{(2s)!}$$

$$8D_1 = \left(\frac{2K}{\pi}\right)^2 \left(\frac{E}{K} + k^2 - 1\right) \qquad D_1(1) = \frac{1}{8\pi}$$

$$16D_3 = \left(\frac{2K}{\pi}\right)^4 k^2(1-k^2)$$

$$16D_5 = \left(\frac{2K}{\pi}\right)^6 k^2(1-k^2)(1-2k^2) \qquad D_5(1) = 0$$

$$\mathrm{IV}_{2s} = \sum_1^{\infty} \operatorname{sech}^{2s}\left[(2n-1)\frac{\pi c}{2}\right] = (-1)^{s-1} \frac{2^{2s}}{(2s-1)!} \sum_{m=0}^{s-1} \alpha_m D_{2s-2m-1}$$

$$\mathrm{XIII}_s = \sum_1^{\infty} (-1)^{n+1} n^s \operatorname{cosech}(n\pi c) = 2D_s$$

TABLE 1(v)

$$F_s(c) = F(s) = \sum_1^{\infty} \frac{(2n-1)^s q^{2n-1}}{1-q^{2n-1}}$$

$$\left(\frac{2K}{\pi}\right) ns\left(\frac{2Kx}{\pi}\right) = \operatorname{cosec} x - 4 \sum_{s=0}^{\infty} (-1)^s F_{2s+1} \frac{x^{2s+1}}{(2s+1)!}$$

$$1 + 24F_1 = \left(\frac{2K}{\pi}\right)^2 (1+k^2)$$

$$7 - 240F_3 = \left(\frac{2K}{\pi}\right)^4 (7 - 22k^2 + 7k^4)$$

$$31 + 504F_5 = \left(\frac{2K}{\pi}\right)^6 (1+k^2)(31 - 46k^2 + 31)$$

$$\mathrm{I}_{2s+1} = \frac{2}{(2s)!} \sum_{m=0}^{\infty} \beta_m F_{2s-2m}$$

TABLE 1(vi)

$$G_s(c) = G_s = \sum_1^\infty \frac{(-1)^{n+1}(2n-1)^s q^{2n-1}}{1-q^{2n-1}}$$

$$\left(\frac{2K}{\pi}\right) dc\left(\frac{2Kx}{\pi}\right) = \sec x + 4 \sum_{s=0}^{\infty} (-1)^s G_{2s} \frac{x^{2s}}{(2s)!}$$

$$1 + 4G_0 = \frac{2K}{\pi}$$

$$1 - 4G_2 = \left(\frac{2K}{\pi}\right)^3 (1-k^2)$$

$$5 + 4G_4 = \left(\frac{2K}{\pi}\right)^5 (1-k^2)(5-k^2)$$

$$61 - 4G_6 = \left(\frac{2K}{\pi}\right)^7 (1-k^2)(61-46k^2+k^4)$$

$$\mathrm{II}_{2s+1} = (-1)^s \frac{2}{(2s)!} \sum_{m=0}^{s} \beta_m G_{2s-2m}$$

TABLE 1(vii)

$$H_s(c) = H_s = \sum_1^\infty \frac{(2n-1)^s q^{n-1/2}}{1-q^{2n-1}}$$

$$\left(\frac{2Kk}{\pi}\right) sn\left(\frac{2Kx}{\pi}\right) = 4 \sum_{s=0}^{\infty} (-1)^s H_{2s+1} \frac{x^{2s+1}}{(2s+1)!}$$

$$4H_1 = \left(\frac{2K}{\pi}\right)^2 k$$

$$4H_3 = \left(\frac{2K}{\pi}\right)^4 k(1+k^2)$$

$$4H_5 = \left(\frac{2K}{\pi}\right)^6 k(1+14k^2+k^4)$$

$$\mathrm{III}_{2s+1} = \frac{2}{(2s)!} \sum_{m=0}^{s} \beta_m H_{2s-2m}$$

$$\mathrm{VII}_s = \sum_1^\infty (2n-1)^s \operatorname{cosech}\left[(2n-1)\frac{\pi c}{2}\right] = 2H_s$$

TABLE 1(viii)

$$J_s(c) = J_s = \sum_1^\infty \frac{(-1)^{n+1}(2n-1)^s q^{n-1/2}}{1-q^{2n-1}}$$

$$\left(\frac{2Kk}{\pi}\right) cd\left(\frac{2Kx}{\pi}\right) = 4 \sum_{s=0}^\infty (-1)^s J_{2s} \frac{x^{2s}}{(2s)!}$$

$$4J_0 = \left(\frac{2K}{\pi}\right) k$$

$$4J_2 = \left(\frac{2K}{\pi}\right)^3 k(1-k^2)$$

$$4J_4 = \left(\frac{2K}{\pi}\right)^5 k(1-k^2)(1-5k^2)$$

$$4J_6 = \left(\frac{2K}{\pi}\right)^7 k(1-k^2)(1-46k^2+61k^4)$$

$$\mathrm{IV}_{s+1} = (-1)^s \frac{2}{(2s)!} \sum_{m=0}^s \beta_m J_{2s-2m}$$

$$\mathrm{XV}_s = \sum_1^\infty (-1)^{n+1}(2n-1)^s \operatorname{cosech}\left[(2n-1)\frac{\pi c}{2}\right] = 2J_s$$

TABLE 1(ix)

$$L_s(c) = L_s = \sum_1^\infty \frac{n^s q^{2n}}{1+q^{2n}}$$

$$\left(\frac{2K}{\pi}\right) cs\left(\frac{2Kx}{\pi}\right) = \cot x - 4 \sum_{s=0}^\infty (-1)^s L_{2s+1} \frac{(2x)^{2s+1}}{(2s+1)!}$$

$$1+24L_1 = \left(\frac{2K}{\pi}\right)^2 \frac{1}{2}(2-k^2)$$

$$1-240L_3 = \left(\frac{2K}{\pi}\right)^4 \frac{1}{8}(8-8k^2-7k^4)$$

$$1+504L_5 = \left(\frac{2K}{\pi}\right)^6 \frac{1}{32}(2-k^2)(16-16k^2+31k^4)$$

$$\mathrm{IX}_{2s} = \sum_1^\infty (-1)^{n+1} \operatorname{cosech}^{2s}(n\pi c) = \frac{2^{2s}}{(2s-1)!} \sum_{m=0}^{s-1} \alpha_m L_{2s-2m-1}$$

TABLE 1(x)

$$M_s(c)=M_s=\sum_1^\infty \frac{(-1)^{n+1}n^s q^{2n}}{1+q^{2n}}$$

$$\left(\frac{2Kk'}{\pi}\right)sc\left(\frac{2Kx}{\pi}\right)=\tan x+4\sum_{s=0}^{\infty}(-1)^s M_{2s+1}\frac{(2x)^{2s+1}}{(2s+1)!}$$

$$1-8M_1=\left(\frac{2K}{\pi}\right)^2 k'$$

$$1+16M_3=\left(\frac{2K}{\pi}\right)^4\frac{1}{2}k'(2-k^2)$$

$$1-8M_5=\left(\frac{2K}{\pi}\right)^6\frac{k'}{16}(16-16k^2+k^4)$$

$$\mathrm{X}_{2s}=\sum_1^\infty(-1)^n\operatorname{sech}^{2s}(n\pi c)=(-1)^{s-1}\frac{2^{2s}}{(2s-1)!}\sum_{m=0}^{s-1}\alpha_m M_{2s-2m-1}$$

TABLE 1(xi)

$$N_s(c)=N_s=\sum_1^\infty\frac{n^s q^n}{1+q^{2n}}$$

$$\left(\frac{2K}{\pi}\right)dn\left(\frac{2Kx}{\pi}\right)=1+4\sum_{s=0}^{\infty}(-1)^s N_{2s}\frac{(2x)^{2s}}{(2s)!}$$

$$1+4N_0=\frac{2K}{\pi}$$

$$16N_2=\left(\frac{2K}{\pi}\right)^3 k^2$$

$$64N_4=\left(\frac{2K}{\pi}\right)^5 k^2(4+k^2)$$

$$\mathrm{XI}_{2s}=\sum_1^\infty(-1)^{n+1}\operatorname{cosech}^{2s}\left[(2n-1)\frac{\pi c}{2}\right]=\frac{2^{2s}}{(2s-1)!}\sum_{m=0}^{s-1}\alpha_m N_{2s-2m-1}$$

$$\mathrm{VI}_s=\sum_1^\infty n^s\operatorname{sech}(n\pi c)=2N_s$$

TABLE 1(xii)

$$P_s(c)=P_s=\sum_1^\infty \frac{(-1)^{n+1}n^s q^n}{1+q^{2n}}$$

$$\left(\frac{2Kk'}{\pi}\right)nd\left(\frac{2Kx}{\pi}\right)=1+4\sum_{s=0}^{\infty}(-1)^s P_{2s}\frac{(2x)^{2s}}{(2s)!}$$

$$1-4P_0=\frac{2Kk'}{\pi}$$

$$16P_2=\left(\frac{2K}{\pi}\right)^3 k'k^2$$

$$64P_4=\left(\frac{2K}{\pi}\right)^5 k'k^2(4-5k^2)$$

$$\mathrm{XII}_{2s}=\sum_1^\infty (-1)^{n+1}\operatorname{sech}^{2s}\left[(2n-1)\frac{\pi c}{2}\right]=(-1)^{s-1}\frac{2^{2s}}{(2s-1)!}\sum_{m=0}^{s-1}\alpha_m P_{2s-2m-1}$$

$$\mathrm{XIV}_s=\sum_1^\infty (-1)^{n+1}n^s\operatorname{sech}(n\pi c)=2P_s$$

TABLE 1(xiii)

$$Q_s(c)=Q_s=\sum_1^\infty \frac{(2n-1)^s q^{2n-1}}{1+q^{2n-1}}$$

$$\left(\frac{2K}{\pi}\right)1s\left(\frac{2Kx}{\pi}\right)=\operatorname{cosec}x-4\sum_{s=0}^{\infty}(-1)^s Q_{2s+1}\frac{x^{2s+1}}{(2s+1)!}$$

$$1-24Q_1=\left(\frac{2K}{\pi}\right)^2(1-2k^2)\qquad Q_1(1)=\frac{1}{24}$$

$$7+240Q_3=\left(\frac{2K}{\pi}\right)^4(7+8k^2-8k^4)$$

$$31-504Q_5=\left(\frac{2K}{\pi}\right)^6(1-2k^2)(31-16k^2+16k^4)\qquad Q_5(1)=\frac{31}{504}$$

$$\mathrm{IX}_{2s+1}=\sum_1^\infty (-1)^{n+1}\operatorname{cosech}^{2s+1}(n\pi c)=\frac{2}{(2s)!}\sum_{m=0}^{s}\beta_m Q_{2s-2m}$$

TABLE 1(xiv)

$$R_s(c)=R_s=\sum_1^\infty \frac{(-1)^{n+1}(2n-1)^s q^{2n-1}}{1+q^{2n-1}}$$

$$\left(\frac{2Kk'}{\pi}\right)nc\left(\frac{2Kx}{\pi}\right)=\sec x-4\sum_0^\infty (-1)^s R_{2s}\frac{x^{2s}}{(2s)!}$$

$$1-4R_0=\frac{2Kk'}{\pi}$$

$$1+4R_2=\left(\frac{2K}{\pi}\right)^3 k'$$

$$5-4R_4=\left(\frac{2K}{\pi}\right)^5 k'(5-4k^2)$$

$$61+4R_6=\left(\frac{2K}{\pi}\right)^7 k'(61-76k^2+16k^4)$$

$$\mathrm{X}_{2s+1}=\sum_1^\infty (-1)^{n+1}\operatorname{sech}^{2s+1}(n\pi c)=(-1)^s\frac{2}{(2s)!}\sum_{m=0}^{s}\beta_m R_{2s-2m}$$

TABLE 1(xv)

$$T_s(c)=T_s=\sum_1^\infty \frac{(2n-1)^s q^{n-1/2}}{1+q^{2n-1}}$$

$$\left(\frac{2Kk}{\pi}\right)cn\left(\frac{2Kx}{\pi}\right)=4\sum_0^\infty (-1)^s T_{2s}\frac{x^{2s}}{(2s)!}$$

$$4T_0=\frac{2Kk}{\pi}$$

$$4T_2=\left(\frac{2K}{\pi}\right)^3 k$$

$$4T_4=\left(\frac{2K}{\pi}\right)^5 k(1+4k^2)$$

$$\mathrm{XI}_{2s+1}=\sum_1^\infty (-1)^{n+1}\operatorname{cosech}^{2s+1}\left[(2n-1)\frac{\pi c}{2}\right]=\frac{2}{(2s)!}\sum_{m=0}^{s}\beta_m T_{2s-2m}$$

$$\mathrm{VIII}_s=\sum_1^\infty (2n-1)^s\operatorname{sech}^s\left[(2n-1)\frac{\pi c}{2}\right]=2T_s$$

TABLE 1(xvi)

$$U_s(c) = U_s = \sum_1^\infty \frac{(-1)^{n+1}(2n-1)^s q^{n-1/2}}{1+q^{2n-1}}$$

$$\left(\frac{2Kkk'}{\pi}\right) sd\left(\frac{2Kx}{\pi}\right) = 4\sum_0^\infty (-1)^s U_{2s+1} \frac{x^{2s+1}}{(2s+1)!}$$

$$4U_1 = \left(\frac{2K}{\pi}\right)^2 kk'$$

$$4U_3 = \left(\frac{2K}{\pi}\right)^4 kk'(1-2k^2) \qquad U_3(1) = 0$$

$$4U_5 = \left(\frac{2K}{\pi}\right)^6 kk'(1-16k^2+16k^4) \qquad U_5(\sqrt{3}) = 0$$

$$\mathrm{XII}_{2s+1} = \sum_1^\infty (-1)^{n+1} \operatorname{sech}\left[(2n-1)\frac{\pi c}{2}\right] = (-1)^s \frac{2}{(2s)!} \sum_{m=0}^s \beta_m U_{2s-2m}$$

$$\mathrm{XVI}_s = \sum_1^\infty (-1)^{n+1}(2n-1)^s \operatorname{sech}\left[(2n-1)\frac{\pi c}{2}\right] = 2U_s$$

TABLE 2

s	1	2	3	4	5
α_0	1	1	1	1	1
α_1	0	−1	−5	−14	−30
α_2	0	0	4	49	273
α_3	0	0	0	−36	−820
α_4	0	0	0	0	576
β_0	1	1	1	1	1
β_1	0	−1	−10	−35	−84
β_2	0	0	9	259	1974
β_3	0	0	0	−225	−12916
β_4	0	0	0	0	11025

TABLE 3

c^2	K
1	$\dfrac{[\Gamma(\frac{1}{4})]^2}{4\pi^{1/2}}$
2	$\dfrac{(\sqrt{2}+1)^{1/2}\Gamma(\frac{1}{8})\Gamma(\frac{3}{8})}{2^{13/4}\pi^{1/2}}$
3	$\dfrac{3^{1/4}[\Gamma(\frac{1}{3})]^3}{2^{7/3}\pi}$
4	$\dfrac{(\sqrt{2}+1)[\Gamma(\frac{1}{4})]^2}{2^{7/2}\pi^{1/2}}$
5	$(\sqrt{5}+2)^{1/4}\left[\dfrac{\Gamma(\frac{1}{20})\Gamma(\frac{3}{20})\Gamma(\frac{7}{20})\Gamma(\frac{9}{20})}{180\pi}\right]^{1/2}$
6	$\left[\dfrac{(\sqrt{2}-1)(\sqrt{3}+\sqrt{2})(2+\sqrt{3})\Gamma(\frac{1}{24})\Gamma(\frac{5}{24})\Gamma(\frac{7}{24})\Gamma(\frac{11}{24})}{384\pi}\right]^{1/2}$

TABLE 3 (*continued*)

7	$\dfrac{\Gamma(\frac{1}{7})\Gamma(\frac{2}{7})\Gamma(\frac{4}{7})}{7^{1/4}4\pi}$
8	$\left[\dfrac{2\sqrt{2}+(1+5\sqrt{2})^{1/2}}{4\sqrt{2}}\right]^{1/2}\dfrac{(\sqrt{2}+1)^{1/4}\Gamma(\frac{1}{8})\Gamma(\frac{3}{8})}{8\pi^{1/2}}$
9	$\dfrac{3^{1/4}(2+\sqrt{3})^{1/2}}{12\pi^{1/2}}[\Gamma(\frac{1}{4})]^2$
10	$\left[\dfrac{(2+3\sqrt{2}+\sqrt{5})\Gamma(\frac{1}{40})\Gamma(\frac{7}{40})\Gamma(\frac{9}{40})\Gamma(\frac{11}{40})\Gamma(\frac{13}{40})\Gamma(\frac{19}{40})\Gamma(\frac{23}{40})\Gamma(\frac{37}{40})}{2560\pi^3}\right]^{1/2}$
11	$[2+(17+3\sqrt{33})^{1/3}+(17-3\sqrt{33})^{1/3}]^2\dfrac{\Gamma(\frac{1}{11})\Gamma(\frac{3}{11})\Gamma(\frac{4}{11})\Gamma(\frac{5}{11})\Gamma(\frac{9}{11})}{11^{1/4}144\pi^2}$
12	$\dfrac{(\sqrt{2}+1)(\sqrt{3}+\sqrt{2})(2-\sqrt{3})^{1/2}3^{1/4}[\Gamma(\frac{1}{3})]^3}{2^{13/3}\pi}$
13	$(18+5\sqrt{13})^{1/4}\left[\dfrac{\Gamma(\frac{1}{52})\Gamma(\frac{7}{52})\Gamma(\frac{9}{52})\Gamma(\frac{11}{52})\Gamma(\frac{15}{52})\Gamma(\frac{17}{52})\Gamma(\frac{19}{52})\Gamma(\frac{25}{52})\Gamma(\frac{29}{52})\Gamma(\frac{31}{52})\Gamma(\frac{47}{52})\Gamma(\frac{49}{52})}{6656\pi^5}\right]^{1/2}$
15	$\left[\dfrac{(\sqrt{5}+1)\Gamma(\frac{1}{15})\Gamma(\frac{2}{15})\Gamma(\frac{4}{15})\Gamma(\frac{8}{15})}{240\pi}\right]^{1/2}$
16	$\dfrac{(\sqrt{2}+1)^{1/2}(9-3\sqrt{2}+4\cdot 2^{1/4})^{1/2}[\Gamma(\frac{1}{4})]^2}{2^{9/2}\pi^{1/2}}$

TABLE 4

c^2	k^2
1	$1/2$
2	$(\sqrt{2}-1)^2$
3	$(2-\sqrt{3})/4$
4	$(\sqrt{2}-1)^4$
5	$[1-2(\sqrt{5}-2)^{1/2}]/2$
6	$(\sqrt{3}-\sqrt{2})^2(2-\sqrt{3})^2$
7	$(8-3\sqrt{7})/16$
8	$113+80\sqrt{2}-4(5+4\sqrt{2})(14+10\sqrt{2})^{1/2}$
9	$(2-\sqrt{3})(\sqrt{2}-3^{1/4})^2/2$
10	$(\sqrt{10}-3)^2(3-2\sqrt{2})^2$
11	$[1-\{(x^{24}-64)/x^{24}\}^{1/2}]/2 \qquad x=[2+(3\sqrt{33}+17)^{1/3}-(3\sqrt{33}-17)^{1/3}]/3$
12	$(\sqrt{3}-\sqrt{2})^4(\sqrt{2}-1)^4$
13	$[1-6(5\sqrt{13}-18)^{1/2}]/2$
15	$(3-\sqrt{5})(4-\sqrt{15})(2-\sqrt{3})^2/64$
16	$(9-3\sqrt{2}-4\cdot 2^{1/4})^2(\sqrt{2}+1)^6$

TABLE 5

c^2	E
1	$\frac{\pi}{4K}+\frac{K}{2}$
2	$\frac{\pi}{4\sqrt{2}K}+\frac{K}{\sqrt{2}}$
3	$\frac{\pi}{4\sqrt{3}K}+\frac{\sqrt{3}+1}{2\sqrt{3}}K$
4	$\frac{\pi}{8K}+2(\sqrt{2}-1)K$
5	$\frac{\pi}{4\sqrt{5}K}+\left[1+\frac{2}{3}(\sqrt{5}-2)^{1/2}+\frac{(2+2\sqrt{5})^{1/2}}{3\sqrt{5}}\right]\frac{K}{2}$
7	$\frac{\pi}{4\sqrt{7}K}+\frac{7+2\sqrt{7}}{14}K$

REFERENCES

[1] C. B. LING, *On summation of series of hyperbolic functions*, this Journal, 5 (1974), pp. 551–562.

[2] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, 4th ed., Cambridge University Press, Cambridge, 1927.

[3] C. G. J. JACOBI, *Fundamenta Nova Theoriae Functionum Ellipticarum*, Koengsburg, 1829.

[4] H. HANCOCK, *Theory of Elliptic Functions*, Dover, New York, 1958.

[5] S. RAMANUJAN, *On certain arithmetical functions*, Trans. Cambridge Philos. Soc., 22(1916), pp. 159–184.

[6] M. ABRAMOWITZ AND IRENE A. STEGUN, *Handbook of Mathematical Functions*, Dover, New York, 1965.

[7] A. SELBERG AND S. CHOWLA, *On Epstein's zeta function*, J. Reine Angew. Math., 227 (1967), pp. 86–110.

[8] S. RAMANUJAN, *Modular equations and approximations to π*, Quart. J. Math. Oxford, 45 (1914), pp. 350–372.

[9] M. L. GLASSER AND V. E. WOOD, *A closed form evaluation of the elliptic integral*, Math. Comput. 25 (1971), pp. 535–536.

[10] I. J. ZUCKER, *The evaluation in terms of Γ functions of the periods of elliptic curves admitting complex multiplication*. Math. Proc. Cambridge Philos. Soc., 82 (1977), pp. 111–118.

[11] S. RAMANUJAN, *Question 358*, J. Indian Math. Soc., 4 (1912), p. 78.

[12] BRUCE C. BERNDT, *Modular transformations and generalisations of several formulae of Ramanujan*, Rocky Mountain J. Math., 7 (1977), pp. 147–189.

[13] G. H. HARDY, *Obituary notice*, Proc. Roy. Soc. Ser. A, 99 (1921), pp. xiii–xxix.

[14] J. W. L. GLAISHER, *On the series which represent the twelve elliptic and the four zeta functions*, Mess. Math., 18 (1889), pp. 1–84.

SIAM J. MATH. ANAL.
Vol. 10, No. 1, January 1979

0036-1410/79/1001-0020 $01.00/0

ON THE DERIVATIVE OF THE DRAZIN INVERSE OF A COMPLEX MATRIX*

ROBERT E. HARTWIG† AND JIM SHOAF†

Abstract. It is shown that the derivative of the Drazin inverse of a differentiable matrix $A(t)$ exists for all values of t in the domain of definition except for the kernels of the nontrivial eigenvalues. Expressions are found for this derivative in terms of the characteristic polynomial, the spectral components and the matrices A, A^0 and A^d respectively. A short proof is given for Stewart's theorem on the continuity of the Moore–Penrose inverse, and a formula corresponding to Wedin's formula is given for the Drazin inverse.

1. Introduction. In the last few years the concept of the Drazin generalized inverse of a matrix [1, p. 169–172], which is defined as the unique solution to the equations

$$A^{m_1+1}XA = A^{m_1}, \quad XAX = X, \quad AX = XA, \quad m_1 = \text{index}(A), \tag{1.1}$$

has received increasing attention. Applications have been found in such unrelated fields as Markov chains [16], differential equations [17] and Neumann iterations [13]. Most of the continuous applications of the Drazin inverse $(\cdot)^d$, in such fields as optimal control [4] and singular perturbations [8], [5], are due to its use in solving explicitly the linear system $A\mathring{\mathbf{x}}(t)+B\mathbf{x}(t)=\mathbf{f}(t)$, where A and B are constant matrices and A is singular. For example, the system $A\mathring{\mathbf{x}}=\mathbf{x}$, with $t \geqq 0$, has a solution if and only if $\mathbf{x}(0)$ lies in the range $R(A^d)$ of A^d, in which case the general solution is given by $\mathbf{x}(t)=e^{A^dt}AA^d\mathbf{q}$, with $\mathbf{q}$ arbitrary [7]. Similarly the singular autonomous system

$$\begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}\begin{bmatrix} \mathring{\mathbf{x}} \\ \mathring{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} A_1(\varepsilon) & A_2(\varepsilon) \\ B_1(\varepsilon) & B_2(\varepsilon) \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \qquad \mathbf{x}(\varepsilon,0)=\mathbf{x}_0, \quad \mathbf{y}(\varepsilon,0)=\mathbf{y}_0$$

has the property that every solution has a pointwise limit at $t>0$, as $\varepsilon \to 0^+$, precisely when the matrix

$$\begin{bmatrix} 0 & 0 \\ B_1(0) & B_2(0) \end{bmatrix}$$

has a group inverse [14], and nonzero eigenvalues have negative real parts [8]. While the Drazin inverse enters the first type of problem through the core-nilpotent decomposition, it enters the second type of problem through its contour integral representation. Since all of these applications dealt with constant matrices, it is to be expected that a study of Drazin inverses of variable matrices will be useful in any future generalizations of these applications to the time dependent case.

It is the purpose of this paper to initiate such a study by showing that the Drazin inverse A^d for a square differentiable matrix $A(t)$, defined by (1.1) exists for all t in the domain of definition except for the kernels of the nontrivial eigenvalues $\lambda_k(t) \not\equiv 0$, and may also be expressed in terms of A, A^0 and A^d. Furthermore, denoting time differentiation by $(\cdot)^0$, we shall show that $(A^d)^0$ satisfies for these values of t,

$$A(A^d)^0A = AA^dA^0AA^d + N_A(A^{m_1-1})^0(A^d)^{m_1-1} + (A^d)^{m_1-1}(A^{m_1-1})^0N_A, \tag{1.2}$$

where $N_A = A(I-AA^d)$ is the nilpotent part of A and m_1 is the index of A. This reduces

* Received by the editors May 10, 1976, and in final revised form September 26, 1977.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650.

to the well-known result

$$A(A^{-1})^0A = -A^0 \tag{1.3}$$

when A^{-1} exists.

The motivation for this work was given by the recent papers of Decell [9] and Golub and Pereyra [11], who derived an expression for the derivative of the Moore–Penrose inverse $A^\dagger$ for a differentiable matrix $A(t)$, in terms of A, $A^0 = (d/dt)\,A(t)$ and $A^\dagger$, for the cases where the derivative exists.

We shall throughout this paper denote the real or complex (open) domain of definition by $\mathscr{D}$, and shall use the notation $R(\cdot)$ and $N(\cdot)$ to denote the range and nullspace of a matrix $(\cdot)$. A solution to the equation $AXA = A$ will be called an inner or 1-inverse of A and is denoted by A^-. As usual, a matrix will be called continuous, differentiable, etc. if all its entries are continuous, differentiable, etc. We shall assume A to be permanently degenerate and permenantly singular on $\mathscr{D}$, [15, p. 64]. For later convenience let us write

$$\Delta_A(\lambda) = |\lambda I - A(t)| = \lambda^{n_1(t)} \prod_{i=2}^{s(t)} [\lambda - \lambda_i(t)]^{n_i(t)},$$

and let the minimal polynomial of $A(t)$ be denoted by $\psi_A(\lambda) = \lambda^{m_1(t)} \prod_{i=2}^{s(t)} [\lambda - \lambda_i(t)]^{m_i(t)}$, where the nontrivial eigenvalues $\lambda_i(t)$ are distinct for at least one value of $t \in \mathscr{D}$. The eigenvalues will always be continuous functions at $t \in \mathscr{D}$, [17, p. 221], but in general little can be said about their differentiability. When $A(t)$ is analytic for $t \in \mathscr{D} \subseteq \mathbb{C}$, however, it is known that the eigenvalues are continuous in $\mathscr{D}$ and will also be analytic except for the exceptional points of which there can only be a finite number in each compact subset of $\mathscr{D}$, [15, p. 64]. The term "near" shall as always stand for "in a sufficiently small neighborhood of", and the interior, boundary and complement of a set $(\cdot)$ will be denoted by $\mathrm{int}(\cdot)$, $\partial(\cdot)$ and $(\cdot)^c$ respectively.

2. The existence of $[A(t)^d]^0$. It is well-known that for a square real or complex matrix A, its Drazin inverse A^d always exists and is a polynomial in A, [1, p. 172]. We shall examine its dependence on t as t ranges over the domain $\mathscr{D}$. Suppose that

$$\Delta_{A(t)}(\lambda) = \lambda^n - \sigma_1(t)\lambda^{n-1} + \cdots (-1)^{n-k}\sigma_{n-k}(t)\lambda^k, \tag{2.1}$$

where the coefficients $\sigma_i(t)$ are *fixed* functions of t, and $\sigma_{n-k}(t)$ is the largest coefficient not *identically* zero for *all* t in $\mathscr{D}$. That is, $\sigma_{n-k}(t') \neq 0$ for some $t' \in \mathscr{D}$. For other values of t it is possible that $\sigma_{n-k}(t) = 0$ and so we may define for all $t \in \mathscr{D}$:

$$\begin{aligned}\Delta_{A(t)}(\lambda) &= \lambda^n - \sigma_1(t)\lambda^{n-1} + \cdots (-1)^{n-n_1(t)}\sigma_{n-n_1(t)}(t)\lambda^{n_1(t)} \\ &= \lambda^{n_1(t)} \prod_{i=2}^{s(t)} [\lambda - \lambda_i(t)]^{n_i(t)},\end{aligned} \tag{2.2}$$

where $\sigma_{n-n_1(t)}(t) = \prod_{i=2}^{s(t)} \lambda_i(t)^{n_i(t)}$ is the highest coefficient not zero at *this* particular value of t. It is a "variable" function which by definition *never* vanishes on $\mathscr{D}$. The integer $n - n_1(t)$ is commonly called the *core rank* of $A(t)$, that is $n - n_1(t) = \mathrm{rank}\,(A^2A^d(t))$. The fact that $\sigma_{n-k}(t') \neq 0$ combined with the definition of $\sigma_{n-n_1(t)}(t)$ implies that

$$k = n_1(t') = \min_{t\in\mathscr{D}} n_1(t) \geqq m_1(t') \geqq \min_{t\in\mathscr{D}} m_1(t).$$

Incidentally, it is unknown whether $m_1(t') = \min_t m_1(t)$. Since $\sigma_{n-n_1(t)}(t) \neq 0$, we may write for all $t \in \mathcal{D}$:

$$\Delta_{A(t)}(\lambda) = (-1)^{n-k}\sigma_{n-n_1(t)}(t)\lambda^{n_1(t)}\left[1 - \frac{\lambda q(\lambda, t)}{\sigma_{n-n_1(t)}}\right]$$

and

$$A(t)^d = A^l(t)\left[\frac{q(A)}{\sigma_{n-n_1(t)}}\right]^{l+1}, \qquad l \geqq m_1(t). \tag{2.3}$$

Let us now assume that $A(t)$ is continuous on $\mathcal{D}$. Since the coefficients $\sigma_i(t)$ are sums of leading principal minors, they are also continuous on $\mathcal{D}$. Consequently the sets $N_i = \ker \sigma_{n-k-i}(t), i = 0, 1, \cdots, n-k$ are all relative *closed* subsets of $\mathcal{D}$. Our main observation is the following result upon which all our later conclusions in this section are based.

LEMMA 1. *Let $A(t)$ be continuous on $\mathcal{D}$ and let $N_i = \ker \sigma_{n-k-i}(t)$, $i = 0, \cdots, n-k$. Then*

(α) *the core rank of $A(t)$ is an integer valued function with discontinuities exactly at $S = \bigcup_{i=0}^{n-k} \partial S$, where $S_0 = N_0^c$ and $S_i = (N_0 \cap N_1 \cdots \cap N_{i-1})\backslash N_i$, $i = 1, \cdots, n-k$. The set S is closed with no interior.*

(β) *Near a point of discontinuity the core-rank of $A(t)$ cannot decrease.*

Proof. (α). Let $A(t)$ be continuous on $\mathcal{D}$ and suppose that $\Delta_{A(t)}(\lambda)$ is given locally by (2.2) and globally by (2.1). As observed earlier core-rank $A(t) \leqq n-k$ with core-rank $A(t) < n-k$ for $t \in \ker \sigma_{n-k}(t)$. By how much the core-rank will drop depends on which of the other $\sigma_i(t)$ also do vanish at t and in which order. Now consider the sets N_i and S_j as defined above. Their most important property is that the sets $\{S_i, S_{i+1}, \cdots, S_{n-k}\}$ form a *partition* of the set $N_0 \cap N_1 \cdots \cap N_{i-1}$, $i = 1, \cdots, n-k$. This means not only that the sets $\{S_0, S_1, S, \cdots, S_{n-k}\}$ partition $\mathcal{D}$, but also that the set

$$S_0 \cup S_1 \cup \cdots \cup S_i = (S_{i+1} \cup S_{i+2} \cdots \cup S_{n-k})^c = (N_0 \cap N_1 \cdots \cap N_i)^c$$

will be *open* for each $i = 0, 1, \cdots, n-k$. From the definition of S_i, we may conclude that core-rank $A(t) = n-k-i \Leftrightarrow t \in S_i$ $i = 0, 1, \cdots, n-k$. This implies that the jump discontinuities of core-rank $A(t)$ occur *precisely* at the boundaries of the S_i, $i = 0, 1, \cdots, n-k$. Indeed, $t \in \partial S_i$ means that every neighborhood of t contains points of S_i and S_i^c; that is, points with core-rank $(A) = i$ and core-rank $A \neq i$ respectively. In other words, core-rank (A) has a discontinuity at t. Conversely, if t_0 is a discontinuity of core-rank $A(t)$, then $t_0 \in S_i$ for some i, since the S_i partition $\mathcal{D}$. But $t \notin \operatorname{int} S_i$ because core-rank $A(t)$ is constant on $\operatorname{int} S_i$. Hence $t_0 \in \partial S_i \cap S_i$. The set $S = \bigcup_{i=0}^{n-k} \partial S_i \cap S_i = \bigcup_{i=0}^{n-1} \partial S_i$ is clearly closed and has (Baire's theorem) no interior.

(β). Suppose t_0 is a point of discontinuity of core-rank $A(t)$ and that $t_0 \in \partial S_i \cap S_i$ for some i. Now $\partial S_i \cap S_i \subseteq S_0 \cup S_1 \cdots \cup S_i$, which is an *open* set, and thus all sufficiently small neighborhoods of t_0 only contain points of $S_0, S_1, \cdots, S_i$. Hence core-rank (A) cannot decrease at these points, thus completing the proof.

We shall now show that most of the main results of [6], [19] and [3] follow as easy corollaries of this Lemma, once we recall [2, p. 38] that an idempotent matrix $E(t)$ is continuous near $t_0 \in \mathcal{D}$ if and only if its rank is continuous near t_0.

COROLLARY 1. *If $A_{m\times n}(t)$ is continuous on $\mathcal{D}$, then*

(i) rank $A(t)$ *is an integer valued function with discontinuities at a closed set with no interior.*

(ii) *near a point of discontinuity* rank $A(t)$ *cannot decrease.*

Proof. Observe that rank $A(t) =$ core-rank $(A^*A(t))$.

We remark in passing that rank $(A)=$ core-rank (A) precisely when A equals its core A^2A^d.

COROLLARY 2. *Let $A(t)$ be continuous (differentiable, C^m, analytic) on $\mathscr{D}$. Then $A^d(t)$ is continuous (differentiable, C^m, analytic) near $t_0 \in \mathscr{D}$ if and only if* core-rank *$A(t)$ is constant near t_0, that is $t_0 \in \mathscr{D}\backslash S$.*

Proof. If core-rank $A(t)$ is constant, say $n-p$, near t_0, then $\sigma_{n-p}(t)\neq 0$ near t_0 and consequently $\sigma_{n-p}(t)^{-1}$ is continuous (differentiable, C^m, analytic) near t_0. For these values of t,

$$A^d(t)=A^l(t)\left[\frac{q(A)}{\sigma_{n-p}(t)}\right]^{l+1}, \qquad l \geqq p.$$

Now because the coefficients in $q(A)$ are among the $\sigma_i(t)$, it follows that they and hence $A^d(t)$ are also continuous (differentiable, C^m, analytic) near t_0. Conversely, if $A^d(t)$ is continuous near t_0, then so is the idempotent matrix $AA^d(t)$. Hence rank $(AA^d)=$ core-rank $A(t)$ is constant near t_0.

COROLLARY 3. *Let $A_{m\times n}(t)$ be continuous (differentiable, C^m, analytic) on $\mathscr{D}$. Then $A^\dagger(t)$ is continuous (differentiable, C^m, analytic) near t_0 if and only if rank $A(t)$ is constant near t_0.*

Proof. Since A^*A is an E.P. matrix [1, p. 163], with rank $A=$ rank $A^*A=$ core-rank A^*A and $(A^*A)^d=(A^*A)^{\#}=(A^*A)^\dagger$, it follows with aid of Corollary 2, and the properties of $A^\dagger$ (such as $A^\dagger=(A^*A)^\dagger A^*$, $(A^*A)^\dagger=A^\dagger A^{*\dagger}$), that the following are equivalent:

(i) rank $A(t)$ is constant near t_0,
(ii) core-rank (A^*A) is constant near t_0,
(iii) $(A^*A)^\dagger$ is continuous (differentiable, etc.) near t_0,
(iv) $A^\dagger(t)$ is continuous (differentiable, etc.) near t_0.

Remarks. (i) Both the real and complex cases are taken care of in the above analysis.

(ii) If $A^\dagger(t)$ has a discontinuity at t_0, then so do all $A^-(t)$ (otherwise AA^- would be continuous near t_0 implying that rank $A(t)$ is constant near t_0). On the other hand, if $A^\dagger(t)$ is continuous near t_0 there may exist $A^-(t)$ which are discontinuous at t_0, as seen from the example where $t_0=1$ and

$$A(t)=\begin{bmatrix}0 & t\\ 0 & 0\end{bmatrix}, \quad A^\dagger(t)=\begin{bmatrix}0 & 0\\ t^{-1} & 0\end{bmatrix}, \quad A^-(t)=\begin{bmatrix}(1-t)^{-1} & 0\\ t^{-1} & 0\end{bmatrix}.$$

(iii) The question of whether or not A^d is continuous near t_0 is *independent* of the index $m_1(t_0)$. Indeed, if

$$A=\begin{bmatrix}1 & 1 & 1\\ 0 & t & 0\\ 0 & 0 & 0\end{bmatrix},$$

then $\Delta=\psi=\lambda(\lambda-1)(\lambda-t)$ for all $t\neq 0$, while for $t=0$, $\Delta_0=\lambda^2(\lambda-1)$, $\psi_0=\lambda(\lambda-1)$. Thus the index remains the same through $t=0$, yet

$$A^d_{(t)}=\begin{cases}A(t)[(1+t)I-A(t)]^2/t^2, & t\neq 0,\\ A(0)=A^2(0), & t=0,\end{cases}$$

which fails to be continuous at $t=0$. On the other hand, if

$$A=\begin{bmatrix}0 & t & 1\\ 0 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},$$

then $\Delta=\psi=\lambda^2(\lambda-1)$ for all $t\neq 0$, while for $t=0$, $\Delta_0=\lambda^2(\lambda-1)$, $\psi_0=\lambda(\lambda-1)$. Hence the *index* changes through $t=0$, yet A^d is differentiable for all t! Indeed, $A^d=A^2(t)$, for all t, which may be seen from (2.3) on taking $l=2$ for all t.

(iv) The results of Lemma and Corollaries 1–3 are easily adapted to the case where $A=A(\mathbf{t})$ is a continuous function of, say, m variables $(t_1, t_2, \cdots, t_m)\in\mathcal{D}$. This includes as a special case the additive perturbation $A+E=A(e_{ij})$, c.f. [19, p. 50].

(v) If $A^d(t)$ is differentiable at $t_0\in\mathcal{D}\backslash S$, and $\sigma_{n-p}(t)\neq 0$ near t_0, then

$$[A^d(t)]^0=(A^l)^0\left[\frac{q(A)}{\sigma_{n-p}(t)}\right]^{l+1}+A^l\left\{\left[\frac{q(A)}{\sigma_{n-p}(t)}\right]^{l+1}\right\}^0,\qquad l\geqq p.$$

Because this is not too convenient a form we shall give two alternative expressions for $(A^d)^0$, one of which uses the spectral theorem, and the second of which expresses the derivative in terms of A, A^0 and A^d.

3. A^d and the spectral theorem. Suppose first that we fix $t\in\mathcal{D}$. The spectral theorem for matrices states that

$$f(A)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1} f^{(j)}(\lambda_k)Z_k^j\qquad(\lambda_1=0)\tag{3.1}$$

for any function $f(\lambda)$ for which the scalars $f^{(j)}(\lambda_k)$ are well defined, and where the Z_k^j are the spectral components of A.

Using an obvious generalization of the method of Englefield [10], for calculating the group inverse for a matrix, we have the following. If

$$P^{-1}AP=C\oplus N,$$

where C is invertible, $N\in\mathscr{C}_{n_1\times n_1}$ is nilpotent of index m_1 and $\oplus$ denotes a direct sum, then it is well-known that

$$A^d=P[C^{-1}\oplus 0_{n_1\times n_1}]P^{-1}.\tag{3.2}$$

Consider now the resolvent

$$(\lambda I-A)^{-1}=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\frac{j!Z_k^j}{(\lambda-\lambda_k)^{j+1}},\qquad \lambda_1=0.\tag{3.3}$$

If we multiply this by $1/\lambda$ and integrate along a contour Γ in the complex plane, enclosing *all* the nonzero eigenvalues of A, but *excluding* $\lambda_1=0$, then we obtain the residue theorem

$$\begin{aligned}\frac{1}{2\pi i}\int_\Gamma(\lambda I-A)^{-1}\frac{d\lambda}{\lambda}&=\sum_{k,j}\frac{j!Z_k^j}{2\pi i}\int_\Gamma\frac{d\lambda}{\lambda(\lambda-\lambda_k)^{j+1}}\\ &=\sum_{k=2}^{s}\sum_{j=0}^{m_k-1}(-1)^j\frac{j!Z_k^j}{\lambda_k^{j+1}}\end{aligned}\tag{3.4}$$

On the other hand,

$$\frac{1}{2\pi i}\int_\Gamma (\lambda I-A)^{-1}\frac{d\lambda}{\lambda}=P\left[\frac{1}{2\pi i}\int_\Gamma (\lambda I-C)^{-1}\frac{d\lambda}{\lambda}\oplus\frac{1}{2\pi i}\int_\Gamma (\lambda I-N)^{-1}\frac{d\lambda}{\lambda}\right]P^{-1}$$
$$(3.5)\qquad =P[C^{-1}\oplus 0]P^{-1}=A^d,$$

since N has only zero eigenvalues and C^{-1} exists. Combining these we get

$$(3.6)\qquad A^d=\sum_{k=2}^{s}\sum_{j=0}^{m_k-1}\frac{(-1)^j j!}{\lambda_k^{j+1}}Z_k^j=g(A),$$

where $g(\lambda)$ is any differentiable function of λ such that

$$g^{(j)}(\lambda_k)=\begin{cases}\left(\dfrac{1}{\lambda}\right)^{(j)}\Bigg|_{\lambda-\lambda_k}, & \lambda_k\neq 0,\quad k=1,\cdots,s,\\ 0, & \lambda_k=0,\quad j=0,\cdots,m_{k-1}.\end{cases}$$

If we set $f(\lambda)=1-\lambda g(\lambda)$, then $f(0)=1$ and $f^{(j)}(\lambda_k)=0$ otherwise. Hence

$$(3.7)\qquad I-AA^d=Z_1^0\quad\text{and}\quad N_A=A(I-AA^d)=Z_1^1,$$

that is, the nilpotent part of A is *exactly* the principal nilpotent matrix of A associated with $\lambda_1=0$! Moreover $Z_1^j=(A^j/j!)(I-AA^d)$, $j=0,\cdots,m_1-1$, while the core of A is given by

$$(3.8)\qquad C_A=A^2A^a=h(A)=\sum_{k=2}^{s}(\lambda_k Z_k^0+Z_k^1),$$

where $h(\lambda)=\lambda^2 g(\lambda)$. In a similar fashion using any function $f(\lambda)$ such that $f(\lambda_l)=1$, $f^{(j)}(\lambda_l)=0$, $j=1,\cdots,m_l-1$ *and* $f^{(j)}(\lambda_k)=(1/\lambda)^{(j)}|_{\lambda=\lambda_k}$, if $k\neq l$, we obtain

$$(3.9)\qquad Z_l^0=I-(A-\lambda_l I)(A-\lambda_l I)^d=\text{projection on } N(A-\lambda_l I)^{m_l}.$$

We may subsequently rewrite the spectral theorem (3.1) in terms of Drazin inverse as:

$$f(A)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\frac{f^{(j)}(\lambda_k)}{j!}(A-\lambda_k I)^j[I-(A-\lambda_k I)(A-\lambda_k I)^d]$$

in which for any eigenvalue λ_l,

$$(A-\lambda_l I)^d=\sum_{k\neq l}\sum_{j=0}^{m_k-1}\frac{(-1)^j j!}{(\lambda_k-\lambda_l)^{j+1}}Z_k^j.$$

Let us now return to the case where $A=A(t)$ and let us assume that A is *analytic* in $\mathscr{D}\subseteq\mathbb{C}$. The spectral representation (3.6) for $A^d(t)$ is valid for all $t\in\mathscr{D}$ except those for which $\lambda_k(t)=0$ or for which the spectral components $Z_k^j(t)$ do have an algebraic singularity. The latter set of points is contained in the set of exceptional points and depends on the order of the branch point of $\lambda_k(t)$ at these exceptional points [15, p. 70]. Hence for any nonexceptional value of t in $\mathscr{D}\backslash S$, we may differentiate (3.6) to give

$$(3.10)\qquad [A^d(t)]^0=\sum_{k=2}^{s}\sum_{j=0}^{m_k-1}\frac{(-1)^j j!}{\lambda_k(t)^{j+2}}[\lambda_k(t)\mathring{Z}_k^j-(j+1)\mathring{\lambda}_k Z_k^j(t)].$$

Again this formula is not very easy to handle in practice. Before we turn to our last expression for $(A^d)^0$, we remark that for matrices over a field $\mathscr{F}$, the Drazin inverse A^d may be characterized as follows.

LEMMA 2. *If* $m_1 = \text{index}(A)$, $A^{l+1}X = A^l$ *and* $X \in R(A^p)$ *for some* $0 \leqq p \leqq m_1$, *then* $A^d = A^{l-p}X^{l-p+1} = A^{m_1-p}X^{m_1-p+1}$.

Proof. Since $A^{l+1}X = A^l$ implies $l \geqq m_1$, and $A^{l+q}X^q = A^l$, we have $A^{l+1}(A^{l-p}X^{l-p+1}) = A^{l+(l-p+1)}X^{l-p+1} = A^l$. Also $A^{l-p}XX^{l-p} = A^lYX^{l-p} \in R(A^l)$ and thus $(A^{l-p}X^{l-p+1} - A^d) \in R(A^l) \cap N(A^{l+1}) = \{0\}$. The second result follows similarly. In particular, if $p = 0$, then $A^{l+1}X = A^l$ implies that $A^d = A^lX^{l+1}$, which shows that the commutivity assumption in Lemma 5, p. 171 of [1], may be dropped! If we set $p = m_1$, we see that $A^{l+1}X = A^l$ and $X \in R(A^{m_1})$ imply that $A^d = X$. The above result is useful for example, when $\Delta_A = (-1)^{n_1}\sigma_{n_1}\lambda^{n_1}[1 - \lambda^{p+1}r(\lambda)]$ with $\sigma_{n_1} \neq 0$, since then $A^d = A^{m_1-p}r(A)^{m_1-p+1} = A^{n_1-p}r(A)^{n_1-p+1}$.

4. The derivative of A^d, in terms of A, $\mathring{A}$ and A^d. Let us first consider the case where $m_1 = \text{index}(A) = 1$, and $t_0 \in \mathscr{D}\backslash S$. Then $(A^d)^0 = (A^{\#})^0$ exists near t_0 and we may differentiate the defining equations (1.1) to give with $X = A^{\#}$,

$$(4.1)\qquad \begin{aligned} &\text{(a)}\ \ \mathring{A}XA + A\mathring{X}A + AX\mathring{A} = \mathring{A},\\ &\text{(b)}\ \ \mathring{X}AX + X\mathring{A}X + XA\mathring{X} = \mathring{X},\\ &\text{(c)}\ \ \mathring{A}X + A\mathring{X} = \mathring{X}A + X\mathring{A}. \end{aligned}$$

Multiplying (b) on the left and right by A we obtain

$$(4.2)\qquad A\mathring{X}A = A\mathring{X}A + AX\mathring{A}XA + A\mathring{X}A$$

and hence

$$(4.3)\qquad A\mathring{X}A = -AX\mathring{A}XA,$$

which reduces to (1.3) when A^{-1} exists. Next, multiplying (4.1c) by A on the left yields

$$A^2\mathring{X} = A\mathring{X}A + AX\mathring{A} - A\mathring{A}X$$

which on multiplication by X^2 on the left yields

$$(4.4)\qquad XA\mathring{X} = X^2(A\mathring{X}A) + X^2A - X\mathring{A}X.$$

Similarly multiplying (4.1c) on the right by A gives

$$\mathring{X}A^2 = \mathring{A}XA + A\mathring{X}A - X\mathring{A}A,$$

so that a multiplication on the right by X^2 produces

$$(4.5)\qquad \mathring{X}AX = \mathring{A}X^2 + (A\mathring{X}A)X^2 - X\mathring{A}X.$$

Hence substitution of (4.4) and (4.5) into (4.1b) gives

$$(4.6)\qquad \begin{aligned} \mathring{X} = \mathring{A}X^2 + (A\mathring{X}A)X^2 - X\mathring{A}X + X\mathring{A}X\\ + X^2(A\mathring{X}A) + X^2\mathring{A} - X\mathring{A}X, \end{aligned}$$

which reduces, with aid of (4.3), to

$$(4.7)\qquad \mathring{X} = -X\mathring{A}X + (I - AX)\mathring{A}X^2 + X^2\mathring{A}(I - AX).$$

In the general case when $m_1 \geqq 2$, we use the fact that

$$A^d = A^{m_1-1}(A^{m_1})^{\#} \quad \text{and} \quad (A^{m_1})^{\#} = (A^d)^{m_1},$$

so that

$$(4.8)\qquad (A^d)^0 = (A^{m_1-1})^0(A^{m_1})^{\#} + A^{m_1-1}[(A^{m_1})^{\#}]^0.$$

Using (4.7) we get

$$(A^d)^0=(A^{m_1-1})^0(A^d)^{m_1}+A^{m_1-1}(I-AA^d)(A^{m_1})^0(A^d)^{2m_1} \\ +(A^d)^{m_1+1}(A^{m_1})^0(I-AA^d)-A^d(A^{m_1})^0(A^d)^{m_1}. \tag{4.9}$$

Since $(A^k)^0=\mathring{A}A^{k-1}+A(A^{k-1})^0$, $k=1,2,\cdots$, this reduces to

$$(A^d)^0=-A^d\mathring{A}A^d+(I-AA^d)(A^{m_1-1})^0(A^d)^{m_1}+A^{m_1-1}(I-AA^d)\mathring{A}(A^d)^{m_1+1} \\ +(A^d)^{m_1+1}\mathring{A}(I-AA^d)A^{m_1-1}+(A^d)^{m_1}(A^{m_1-1})^0(I-AA^d), \tag{4.10}$$

which only requires the calculation of $(A^{m_1-1})^0$ from A^{m_1-1} which has to be computed anyway. Finally, multiplying this result by A on both sides immediately yields (1.2).

When A is E.P., that is $R(A^*)=R(A)$, then $A^{\#}$ exists and equals $A^\dagger$, [1, p. 163], so that $(A^\dagger)^0$ then becomes by (4.7)

$$(A^\dagger)^0=-A^\dagger AA^\dagger+(I-AA^\dagger)\mathring{A}A^{\dagger 2}+A^{\dagger 2}\mathring{A}(I-AA^\dagger). \tag{4.11}$$

If we now use the identity

$$(A^\dagger A)^0=[(A^\dagger A)^*]^0 \tag{4.12}$$

we obtain

$$[A^\dagger A^0-(\mathring{A}A^\dagger)^*](I-AA^\dagger)=\{[A(A^\dagger)^0]^*-(A^\dagger)^0A\}(I-AA^\dagger)=0,$$

so that (4.11) reduces to

$$(A^\dagger)^0=-A^\dagger AA^\dagger+(I-AA^\dagger)(A^\dagger\mathring{A})^*A+A^\dagger(\mathring{A}A^\dagger)^*(I-A^\dagger A), \tag{4.13}$$

which is the result given in [9], [11] for the derivative of the Moore–Penrose inverse. It was remarked in [11] that this formula could be obtained from Wedin's formula [20], [21]

$$B^\dagger-A^\dagger=-B^\dagger(B-A)A^\dagger \\ +(I-B^\dagger B)(B-A)^*A^{*\dagger}A^\dagger+B^\dagger B^{*\dagger}(B-A)^*(I-AA^\dagger), \tag{4.14}$$

on letting B tend to A. A similar formula is

$$B^d-A^d=-B^d(B-A)A^d \\ +(I-B^dB)(B^d-A^d)AA^d+B^dB(B^d-A^d)(I-AA^d) \tag{4.15}$$

which cannot, however, be used to calculate the Fréchet derivative of A at t, since it involves B^d-A^d on both sides. The exact analog needed is the following formula:

$$B^d-A^d=-B^d(B-A)A^d+(I-B^dB)\sum_{i=0}^{l-1}B^i(B-A)(A^d)^{i+2} \\ +\sum_{i=0}^{l-1}(B^d)^{i+2}(B-A)A^i(I-AA^d), \tag{4.16}$$

which is true in any finite dimensional algebra provided $l\geqq\min\{\text{index}(A),\text{index}(B)\}$.

On setting $B=A+dA$ and letting $dA\to 0$ we obtain the Fréchet derivative

$$d(A(t)^d)=-A^d(dA)A^d+(I-AA^d)\sum_{i=0}^{l-1}A^i(dA)(A^d)^{i+2} \\ +\sum_{i=0}^{l-1}(A^d)^{i+2}(dA)A^i(I-AA^d). \tag{4.17}$$

Indeed, if $\|\cdot\|$ is any norm on $\mathbb{C}_{n\times n}$, then

$$\lim_{\|dA\|\to 0} \frac{B^d - A^d - d(A^d)}{\|dA\|}$$

$$\lim_{\|dA\|\to 0} \frac{1}{\|dA\|}\Big\{ -(B^d - A^d)(dA)A^d + \sum_{i=0}^{l-1} [(I - BB^d)B^i - (I - AA^d)A^i]$$

$$\cdot (dA)(A^d)^{i+2} + \sum_{i=0}^{l-1} [(B^d)^{i+2} - (A^d)^{i+2}](dA)A^i(I - AA^d)\Big\}.$$

Since $(I - BB^d)B^i - (I - AA^d)A^i = (B^i - A^i) - (B^d - A^d)B^{i+1} + A^d(B^{i+1} - A^{i+1})$, this limit is zero provided again that $B^d \to A^d$, that is, that $t \in \mathcal{D}\backslash S$. If we replace $d(\cdot)$ by $(d/dt)(\cdot)$ in (4.17) we obtain the presentation of $(A^d)^0$ as given by Campbell [3].

We thus have that while the right hand sides of (4.10), (4.13) and (4.17) are linear functions that exist for *all* $t \in \mathcal{D}$ they become the Fréchet differentials of A^d and $A^\dagger$ respectively, only where the latter are differentiable. Let us conclude this section with an example of the use of (4.10).

Example. Let $A = \begin{bmatrix} t & 1 \\ 0 & 0 \end{bmatrix}$. Then $\Delta = -t\lambda[1 - (1/t)\lambda]$, $t \neq 0$, and so $A^d = A/t^2$. Since $A^2 = tA$, $AA^d = A/t$, $\mathring{A}A = A$ and $A\mathring{A} = \begin{bmatrix} t & 0 \\ 0 & 0 \end{bmatrix}$, we have by (4.10)

$$\begin{aligned}(A^d)^0 &= -A^d\mathring{A}A^d + (I - AA^d)\mathring{A}(A^d)^2 + (A^d)^2\mathring{A}(I - AA^d) \\ &= -A^2/t^4 + (I - A/t)\mathring{A}A^2/t^4 + (A^2/t^4)\mathring{A}(I - A/t) \\ &= -2A/t^3 + A\mathring{A}/t^3 = \begin{bmatrix} -1/t^2 & -2/t^3 \\ 0 & 0 \end{bmatrix},\end{aligned}$$

which is exactly $(d/dt)[A/t^2]$.

REFERENCES

[1] A. BEN ISRAEL AND T. N. GREVILLE, *Generalized Inverses, Theory and Applications*, Wiley, New York, 1974.
[2] T. L. BOULLION AND P. L. ODELL, *Generalized Inverse Matrices*, Wiley Interscience, New York, 1971.
[3] S. L. CAMPBELL, *Differentiation of the Drazin inverse*, this Journal, 30 (1976), pp. 703–707.
[4] ———, *Optimal control of autonomous linear process with singular matrices in the quadratic cost functional*, SIAM J. Control 14 (1976), pp. 1092–1106.
[5] ———, *Singular perturbations of autonomous linear systems, II*, submitted for publication.
[6] S. L. CAMPBELL AND C. D. MEYER, *Continuity properties of the Drazin pseudoinverse*, Linear Algebra and Appl., 10 (1975), pp. 77–84.
[7] S. L. CAMPBELL, C. D. MEYER AND N. J. ROSE, *Applications of the Drazin inverse to linear systems of differential equations*, SIAM J. Appl. Math., 31 (1976), pp. 411–425.
[8] S. L. CAMPBELL AND N. J. ROSE, *Singular perturbations of autonomous linear systems*, this Journal, to appear.
[9] H. P. DECELL, *On the derivative of the generalized inverse of a matrix*, J. Linear and Multilinear Algebra, 1 (1974), pp. 357–359.
[10] M. J. ENGLEFIELD, *The commuting inverses of a square matrix*, Proc. Cambridge Philos. Soc., 62 (1966), pp. 667–671.
[11] G. H. GOLUB AND V. PEREYRA, *The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate*, SIAM J. Numer. Anal., 10 (1973), pp. 413–432.
[12] R. E. HARTWIG, *Some properties of Hypercompanion matrices*, Industrial Math., 24 (1974), pp. 77–84.
[13] R. E. HARTWIG AND F. J. HALL, *Applications of the Drazin inverse to Cesaro–Neumann iterations*, Recent Advances in Generalized Inverses, M. Z. Nashed ed., Pitman, London, San-Francisco, to appear.

[14] R. E. HARTWIG AND J. M. SHOAF, *Group inverses and Drazin inverses of bidiagonal and triangular Toeplitz matrices*, J. Austral. Math. Soc., 24 (Ser. A) (1977), pp. 10–34.

[15] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.

[16] C. D. MEYER, *The role of the group generalized inverse in the theory of finite Markov chains*, SIAM Rev., 17 (1975), pp. 443–464.

[17] A. M. OSTROWSKI, *Solutions of Equations and Systems of Equations*, 2nd ed., Academic Press, New York, 1966.

[18] C. R. RAO AND S. K. MITRA, *Generalized Inverses of Matrices and its Applications*, John Wiley, New York, 1971.

[19] G. W. STEWART, *On the continuity of the generalized inverse*, SIAM J. Appl. Math., 17 (1969), pp. 33–45.

[20] PER-ÅKE WEDIN, *Perturbation bounds in connection with singular value decomposition*, BIT, 12 (1972), pp. 99–111.

[21] ———, *On pseudoinverses of perturbed matrices*, Tech. Rep. Department of Computer Science, Lund Univ. Sweden, 1969, 56 pp.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0001 $01.00/0

STABILITY OF INTERCONNECTED SYSTEMS DESCRIBED BY STOCHASTIC NONLINEAR VOLTERRA INTEGRAL EQUATIONS*

R. L. GUTMANN† AND A. N. MICHEL†

Abstract. New results are established for the stochastic stability of interconnected systems (also called composite systems or large scale systems) described by nonlinear Volterra integral equations with uncertain weighting functions, uncertain nonlinearities and stochastic inputs. We utilize the viewpoint of analyzing complex systems in terms of lower order (and simpler) subsystems and in terms of the system interconnecting structure. Whenever appropriate, frequency domain techniques (including circle criteria and Popov-like conditions) are emphasized. The results are applied to systems described by stochastic nonlinear differential equations. In addition, a specific example is also presented.

1. Introduction. In this paper we establish new stability results for a large class of dynamical systems described by *nonlinear Volterra integral equations with random driving functions and random coefficients.* Such systems, which arise naturally in applications, have for some time been the subject of several investigations (see, e.g., the survey paper by Kozin [4]) and are of current interest (see, e.g., Morozan [15], Tsokos [22], and Tsokos and Padgett [23]). As in the case of deterministic dynamical systems, difficulties can often arise in applying these results to complex systems of high dimension and intricate structure. In the present paper we circumvent most of these difficulties by viewing such systems as an interconnection of lower order (and usually simpler) subsystems; and by accomplishing the stability analysis at different hierarchical levels (i.e., in terms of the subsystems and systems interconnecting structure). In the current literature, such systems are often referred to as *large scale systems, composite systems, interconnected systems,* and in certain applications, *decentralized systems.*

Although the present results are phrased in terms of general operator gains, we emphasize frequency domain techniques whenever appropriate (including *circle conditions* and *Popov-type criteria*). In arriving at the present results, we use several existing results as motivation: single-loop results for deterministic dynamical systems (see, e.g., Sandberg [19], [20]); single-loop results for stochastic dynamical systems (see Tsokos [22] and Tsokos and Padgett [23]); results for deterministic composite systems (see Lasley and Michel [5]). For additional related work dealing with qualitative analysis of interconnected dynamical systems, refer to Michel [8]–[10], Michel and Rasmussen [11], Rasmussen and Michel [18], Lasley and Michel [6], Porter and Michel [17], and Miller and Michel [13], [14]. As a final note we emphasize that several of the present results, when reduced to corresponding deterministic results, have *not* been reported elsewhere.

Notation used throughout this paper is established in the next section. The main results are presented in the third section. In the fourth section these results are applied to dynamical systems described by stochastic differential equations endowed with

* Received by the editors February 1, 1977, and in revised form August 29, 1977. This work was supported by the National Science Foundation under Grant ENG 75-14093 and by the Engineering Research Institute, Iowa State University.

† Department of Electrical Engineering and Engineering Research Institute, Iowa State University, Ames, Iowa 50011.

random coefficients. A specific control system is considered in the fifth section. All results are proved in the Appendix.

2. Preliminaries. Let $A=[a_{ij}]$ denote an $n\times m$ matrix and let A^T denote its transpose. Let A^* denote the complex conjugate transpose of A. The inverse of a nonsingular $n\times n$ matrix, A, is denoted by A^{-1}. If C and D are real $n\times m$ matrices, then $C\geqq D$ means $c_{ij}\geqq d_{ij}$ for all i and j and $C\geqq 0$ means $c_{ij}\geqq 0$ for all i and j. Let $I_{(N)}$ denote the $N\times N$ identity matrix. Let $\Lambda[M]$ denote the positive square root of the largest eigenvalue of M^*M. If the elements of a real matrix, B, depend on a real parameter, t, we say that B is bounded if there exists a real number, K, such that $|b_{ij}(t)|\leqq K<\infty$ for all allowable t. We define $R=(-\infty,\infty)$, $R^N=R\times R\times\cdots,\times R$, and $R^+=[0,\infty)$. If $x=[x^{(1)},x^{(2)},\cdots,x^{(N)}]^T\in R^N$, then $|x|=[(x^{(1)})^2+(x^{(2)})^2+\cdots+(x^{(N)})^2]^{1/2}$. The set of all real, Lebesgue-measurable N-vector-valued functions of the real variable $t\in R^+$ is denoted by $H_{(N)}(R^+)$ and $L_{p(N)}(R^+)=\{f\in H_{(N)}(R^+)$: $\int_0^\infty |f(t)|^p\,dt<\infty\}$, $1\leqq p<\infty$. The inner product of two elements, f and g, of $L_{2(N)}(R^+)$ is denoted by

$$\langle f,g\rangle=\int_0^\infty f^Tg\,dt.$$

The norm of $f\in L_{2(N)}(R^+)$ is defined by $\|f\|=\langle f,f\rangle^{1/2}$. The norm of a linear operator, T, on $L_{2(N)}(R^+)$ is denoted by $\|T\|$. If $A(t)=[a_{ij}(t)]$ is an arbitrary real $N_1\times N_2$ matrix-valued Lebesgue-measurable function of $t\in R^+$, we say $A\in K_{p(N_1\times N_2)}(R^+)$, $1\leqq p<\infty$, if $\int_0^\infty |a_{ij}(t)|^p\,dt<\infty$ for all i and j.

Given a probability space $(\Omega,\mathcal{F},P)$, denote by $X_{(N)}$ the space of all N-dimensional real-valued random vectors over Ω which are square integrable with respect to P, that is, if $x(\omega)=[x^{(1)}(\omega),\cdots,x^{(N)}(\omega)]^T\in X_{(N)}$, then $x^{(i)}(\omega)$ is $\mathcal{F}$-measurable, $i=1,2,\cdots,N$, and

$$\int_\Omega x^T(\omega)x(\omega)\,dP(\omega)<\infty.$$

Let $H_{(N)}(R^+,\Omega)$ denote the space of all real-valued N-dimensional random vector processes over $R^+\times\Omega$ such that if $x\in H_{(N)}(R^+,\Omega)$, then $x(\cdot,\omega)\in H_{(N)}(R^+)$ for fixed $\omega\in\Omega$, and $x(t,\cdot)\in X_{(N)}$ for fixed $t\in R^+$. Define $L_{p(N)}(R^+,L_\infty(\Omega))$ by

$$L_{p(N)}(R^+,L_\infty(\Omega))=\left\{x\in H_{(N)}(R^+,\Omega)\text{: }\operatorname*{ess\,sup}_{\omega\in\Omega}\int_0^\infty |x(t,\omega)|^p\,dt<\infty\right\},$$

$1\leqq p<\infty$. Let $A(t,\omega)=[a_{ij}(t,\omega)]$ denote an arbitrary real $N_1\times N_2$-matrix-valued stochastic process with $a_{ij}\in H_{(1)}(R^+,\Omega)$. Then $K_{p(N_1\times N_2)}(R^+,L_\infty(\Omega))$, $1\leqq p<\infty$, denotes matrix processes $A(t,\omega)$ such that

$$\operatorname*{ess\,sup}_{\omega\in\Omega}\int_0^\infty |a_{ij}(t,\omega)|^p\,dt<\infty\quad\text{for all } i \text{ and } j.$$

Let $T\in R^+$ and define, for $x\in H_{(N)}(R^+,\Omega)$,

$$x_T(t,\omega)=\begin{cases}x(t,\omega), & 0\leqq t<T,\\ 0, & t>T.\end{cases}$$

Define $\mathscr{E}_{p(N)}=\{x\in H_{(N)}(R^+,\Omega)\text{: } x_T\in L_{p(N)}(R^+,L_\infty(\Omega))\}$, $1\leqq p<\infty$. Define the truncation operator, Π_T, on $H_{(N)}(R^+,\Omega)$ by $\Pi_Tx=x_T$. Let $\mathscr{E}_{s(N)}$ denote those processes in $\mathscr{E}_{2(N)}$ with time-derivatives in $\mathscr{E}_{2(N)}$.

Let $\eta_{(N)}$ denote the collection of memoryless nonlinearities of the type

$$\psi(x(t,\omega),t,\omega)=[\psi^{(1)}(x^{(1)}(t,\omega),t,\omega),\cdots,\psi^{(N)}(x^{(N)}(t,\omega),t,\omega)]^T,$$

with $x=[x^{(1)},\cdots,x^{(N)}]^T\in H_{(N)}(R^+,\Omega)$, $t\in R^+$, $\omega\in\Omega$, where $\psi^{(i)}(g,t,\omega)$, $i=1,2,\cdots,N$, are real-valued functions of the real variables $g\in R$ and $t\in R^+$, and the variable $\omega\in\Omega$ such that

(i) $P\{\omega: \psi^{(i)}(0,t,\omega)=0,\ t\in R^+,\ i=1,\cdots,N\}=1$;

(ii) there exist real numbers a and b with the property

$$P\left\{\omega: a\leqq\frac{\psi^{(i)}(g,t,\omega)}{g}\leqq b,\ t\in R^+,\ g\neq 0,\ i=1,\cdots,N\right\}=1;$$

(iii) $\psi^{(i)}(x(t,\omega),t,\omega)$ is a Lebesgue-measurable function of t and an $\mathcal{F}$-measurable function of ω whenever x is a Lebesgue-measurable function of t and an $\mathcal{F}$-measurable function of ω; and

(iv) $\psi^{(i)}(x(t,\omega),t,\omega)\in H_{(N)}(R^+,\Omega)$ whenever $x\in H_{(N)}(R^+,\Omega)$. Define $C_1(\psi)=\{\omega\in\Omega$: (i) and (ii) above are true for $\psi\in\eta_{(N)}\}$.

DEFINITION 1. The stochastic $N\times N$ matrix, $A(\omega)$, whose elements are measurable functions of ω, is said to be *stochastically stable* if for some positive real γ,

$$P\{\omega: \mathrm{Re}\,(\lambda_k(\omega))<-\gamma,\ k=1,2,\cdots,N\}=1$$

where $\lambda_k(\omega)$, $k=1,2,\cdots,N$, are the eigenvalues of $A(\omega)$.

We consider *composite systems* described by the following stochastic integral operator equations:

$$\text{(S)}\qquad \begin{aligned} e_i(t,\omega)&=u_i(t,\omega)-y_i(t,\omega),\\ y_i(t,\omega)&=\int_0^t k_i(t-\tau,\omega)\psi_i(e_i(\tau,\omega),t,\omega)\,d\tau,\\ u_i(t,\omega)&=r_i(t,\omega)+\sum_{j=1}^m B_{ij}e_j(t,\omega)+\sum_{j=1}^m D_{ij}y_j(t,\omega),\end{aligned}$$

for $i,j\in M=\{1,2,\cdots,m\}$. For each $i\in M$ we assume that r_i, e_i, y_i, and u_i belong to $\mathscr{E}_{2(N_i)}$; $\psi_i\in\eta_{(N_i)}$; $k_i\in K_{1(N_i\times N_i)}$. For each $i,j\in M$, B_{ij} and D_{ij} are operators on $\mathscr{E}_{2(N_j)}$ with values in $\mathscr{E}_{2(N_i)}$. These operators are assumed to be one of two types: either a *Type A operator* with

$$B_{ij}e_j(t,\omega)=b_{Aij}(t,\omega)\cdot e_j(t,\omega),\quad\text{or}$$

$$D_{ij}y_j(t,\omega)=d_{Aij}(t,\omega)\cdot y_j(t,\omega),$$

where $b_{Aij}(t,\omega)$ and $d_{Aij}(t,\omega)$ are $N_i\times N_j$-dimensional matrix-valued random processes with elements in $L_{2(1)}(R^+,L_\infty(\Omega))$; or a *Type B operator* with

$$B_{ij}e_j(t,\omega)=\int_0^t b_{Bij}(t-\tau,\omega)\phi_{ij}(e_j(\tau,\omega),\tau,\omega)\,d\tau,\quad\text{or}$$

$$D_{ij}y_j(t,\omega)=\int_0^t d_{Bij}(t-\tau,\omega)\phi_{ij}(y_j(\tau,\omega),\tau,\omega)\,d\tau,$$

where b_{Bij} and d_{Bij} belong to $K_{1(N_i\times N_j)}(R^+,L_\infty(\Omega))\cap K_{2(N_i\times N_j)}(R^+,L_\infty(\Omega))$ and $\phi_{ij}\in\eta_{(N_j)}$. We define the following $\sum_{j=1}^m N_j\times\sum_{j=1}^m N_j$-dimensional matrices of operators:

$$B_A=[B_{Aij}]$$

where

$$B_{Aij}=\begin{cases} B_{ij} & \text{if } B_{ij} \text{ is of Type } A,\\ 0 & \text{if } B_{ij} \text{ is of Type } B,\ i, j\in M,\end{cases}$$

and

$$B_B=[B_{Bij}]$$

where

$$B_{Bij}=\begin{cases} B_{ij} & \text{if } B_{ij} \text{ is of Type } B,\\ 0 & \text{if } B_{ij} \text{ is of Type } A,\ i, j\in M.\end{cases}$$

We also define the operators K_i and Q_i on $\mathscr{E}_{2(N_i)}$, $i\in M$, by

$$K_i x(t, \omega)=\int_0^t k_i(t-\tau, \omega)x(\tau, \omega)\,d\tau, \qquad t\in R^+, \quad x\in \mathscr{E}_{2(N_i)}$$

and

$$Q_i x(t, \omega)=\psi_i(x(t, \omega), t, \omega), \qquad t\in R^+, \quad x\in \mathscr{E}_{2(N_i)}.$$

Furthermore we define the Laplace transform $\tilde{K}_i(s, \omega)$ by

$$\tilde{K}_i(s, \omega)=\int_0^\infty k_i(t, \omega)\,e^{-st}\,dt.$$

Recall that an operator, H, on $\mathscr{E}_{2(N)}$ is causal if, for an arbitrary $T\in R^+$,

$$\Pi_T Hx(t)=\Pi_T H \Pi_T x(t), \qquad t\in R^+, \quad x\in \mathscr{E}_{2(N)}$$

where Π_T is the truncation operator on $\mathscr{E}_{2(N)}$. It is assumed throughout this paper that K_i, Q_i, B_{ij}, and D_{ij} are causal operators for all i and j.

System (S) may be viewed as the interconnection of m *free* or *isolated subsystems*, each of dimension N_i and each described by an equation of the form

$$(S_i) \qquad e_i(t, \omega)=r_i(t, \omega)-\int_0^t k_i(t-\tau, \omega)\psi_i(e_i(\tau, \omega), \tau, \omega)\,d\tau, \qquad i\in M.$$

We now define the type of stochastic stability we will be concerned with.

DEFINITION 2. System (S) is said to be *stochastically absolutely stable* if (see Tsokos [22] and Tsokos and Padgett [23, Chap. 9])

$$P\{\omega: \lim_{t\to\infty} e_i(t, \omega)=0\}=1, \qquad i\in M,$$

and

$$P\{\omega: \lim_{t\to\infty} y_i(t, \omega)=0\}=1, \qquad i\in M.$$

We now present two lemmas which are used throughout this paper. The first deals with Minkowski matrices or M-matrices (see Ostrowski [16] and Fiedler and Ptak [1]).

DEFINITION 3. A square matrix $A=[a_{ij}]$ is said to be an *M-matrix* if the off-diagonal elements of A are all nonpositive ($a_{ij}\leqq 0$, $i\neq j$) and if all principal minors of A are positive.

LEMMA 1. *Let $A=[a_{ij}]$ be an $n\times n$ matrix with nonpositive off-diagonal elements. Then the following conditions are mutually equivalent.*

(i) *The principal minors of A are all positive (i.e. A is an M-matrix).*

(ii) *The successive principal minors of A are all positive.*

(iii) *The matrix A is nonsingular and $A^{-1}\geqq 0$.*

Lemma 1 is proved in Ostrowski [16] and Fiedler and Ptak [1].

The second lemma concerns the asymptotic behavior of the function $f(t)$ defined by

$$f(t)=\int_0^t k(t-\tau)h(\tau)\,d\tau, \tag{1}$$

where $k\in L_{1(1)}(R^+)$ and $h\in L_{2(1)}(R^+)$. The following lemma and associated proof are presented in Sandberg [19] (see also Hewitt and Stromberg [2, p. 398, Thm. 22.31]).

LEMMA 2. *If in* (1) $k\in L_{1(1)}(R^+)\cap L_{2(1)}(R^+)$, *then* $f(t)\to 0$ *as* $t\to\infty$.

3. Main results. The following theorems are proved in the Appendix.

THEOREM 1. *System* (S) *is* stochastically absolutely stable *if the following conditions hold*:

(i) $r_i\in L_{2(N_i)}(R^+, L_\infty(\Omega))$ *and* $|r_i(t,\omega)|\to 0$ *as* $t\to\infty$ *a.e.* $[P]$, $i\in M$;

(ii) $\det[I_{(N_i)}+\frac{1}{2}(a_i+b_i)\tilde{K}_i(s,\omega)]\neq 0$ *for* $\operatorname{Re}(s)\geqq 0$, *a.e.* $[P]$, $i\in M$;

(iii) *there exist constants* α_i, γ_{ik}, ξ_{ik} *and* μ_i *with*

$$\alpha_i\geqq\tfrac{1}{2}(b_i-a_i)\sup_{\lambda\in R^+}\Lambda[(I_{(N_i)}+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega))^{-1}\tilde{K}_i(j\lambda,\omega)]$$

a.e. $[P]$, $i\in M$,

$$\gamma_{ik}\geqq\|(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}B_{ik}\|\quad \text{a.e. } [P],\quad i,k\in M,$$

$$\xi_{ik}\geqq\|(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}D_{ik}\|\quad \text{a.e. } [P],\quad i,k\in M,$$

$$\mu_i=\max(|a_i|,|b_i|)\ \text{where } I \text{ denotes the identity operator,}$$

such that the matrix $A=[a_{ij}]$ has positive successive principal minors, where

$$a_{ik}=\begin{cases}1-\alpha_i-\gamma_{ii}-\xi_{ii}\mu_i, & i=k,\\ -\gamma_{ik}-\xi_{ik}\mu_k, & i\neq k,\quad i,k\in M\end{cases}$$

(the matrix A is referred to as a *test matrix*); *and*

(iv) *the operator $(I-B_A)$ has a bounded inverse on $L_{2(\sum N_i)}(R^+, L_\infty(\Omega))$ for $t>T^*$, a.e. $[P]$ for some $T^*\in R^+$.*

Remark 1. If B_{ik} or D_{ik} are Type B operators with $\phi_{ik}e_k=e_k$ or $\phi_{ik}y_k=y_k$, e_k, $y_k\in\mathscr{E}_{2(N_k)}$, then the A-matrix elements γ_{ik} or ξ_{ik} may be found from

$$\gamma_{ik}\geqq\sup_{\lambda\in R^+}\Lambda\{(I_{(N_i)}+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega))^{-1}\tilde{B}_{ik}(j\lambda,\omega)\}\quad\text{a.e. } [P],\quad\text{or}$$

$$\xi_{ik}\geqq\sup_{\lambda\in R^+}\Lambda\{(I_{(N_i)}+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega))^{-1}\tilde{D}_{ik}(j\lambda,\omega)\}\quad\text{a.e. } [P],$$

where $\tilde{K}_i(s,\omega)$, $\tilde{B}_{ik}(s,\omega)$, and $\tilde{D}_{ik}(s,\omega)$ represent the Laplace transforms of $k_i(t,\omega)$, $b_{Bik}(t,\omega)$, and $d_{Bik}(t,\omega)$, respectively. For the case where $N_i=N_k=1$, the above A-matrix elements may be determined graphically. It can be seen that under these conditions γ_{ik} is the smallest number, b, such that the locus of $(1/b)\tilde{B}_{ik}(j\lambda,\omega)$, $\lambda\in R$, is interior to the locus of $(1+\frac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega))$, $\lambda\in R^+$, for almost every ω. For further frequency-domain interpretations, see Remark 4.

Remark 2. The deterministic version of this theorem (which has not previously appeared) may be obtained by taking $\Omega=\{1\}$ and $P\{1\}=1$.

Remark 3. For $m=N_i=1$, $B_{11}=D_{11}=0$, and for the deterministic case (see Remark 2), Theorem 1 reduces to a version of the familiar circle theorem introduced by Sandberg [19] and Zames [24], [25].

Remark 4. For $N_i=1$, the A-matrix terms α_i may be determined from the Nyquist locus of $\tilde{K}_i(j\lambda,\omega)$. Note that we desire to find an α_i such that for almost every ω

$$\tfrac{1}{2}(b_i-a_i)|\tilde{K}_i(j\lambda,\omega)|\leqq \alpha_i|1+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega)|$$

or

$$\left(\frac{b_i-a_i}{2\alpha_i}\right)^2|\tilde{K}_i(j\lambda,\omega)|^2\leqq|1+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega)|^2,$$

that is,

$$\left(\frac{b_i-a_i}{2\alpha_i}\right)^2\tilde{K}_i(j\lambda,\omega)\tilde{K}_i^*(j\lambda,\omega)\leqq(1+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega))(1+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i^*(j\lambda,\omega)).$$

We may write, for almost every $\omega\in\Omega$,

$$0\leqq 1+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i^*\left[\left(\frac{a_i+b_i}{2}\right)^2-\left(\frac{b_i-a_i}{2\alpha_i}\right)^2\right]\tilde{K}_i\tilde{K}_i^*,$$

where the arguments of $\tilde{K}_i$, $\tilde{K}_i^*$ are understood to be $(j\lambda,\omega)$. Defining

$$\Delta_i=\frac{b_i-a_i}{2\alpha_i},\qquad \sigma_i=\frac{a_i+b_i}{2},\qquad \rho_i=\sigma_i^2-\Delta_i^2, \tag{2}$$

we have

$$0\leqq 1+\sigma_i\tilde{K}_i+\sigma_i\tilde{K}_i^*+\rho_i\tilde{K}_i\tilde{K}_i^*. \tag{3}$$

We now consider three cases.

(a) If $\rho_i>0$, (3) may be written, for almost every $\omega\in\Omega$, as

$$0\leqq\frac{1}{\rho_i}+\left(\frac{\sigma_i}{\rho_i}\right)\tilde{K}_i+\left(\frac{\sigma_i}{\rho_i}\right)\tilde{K}_i^*+\tilde{K}_i\tilde{K}_i^*$$

or equivalently as

$$\left|\tilde{K}_i+\frac{\sigma_i}{\rho_i}\right|\geqq\left[\left(\frac{\sigma_i}{\rho_i}\right)^2-\left(\frac{1}{\rho_i}\right)\right]^{1/2}=\left|\frac{\Delta_i}{\rho_i}\right|. \tag{4}$$

The implication of (4) is that the locus of $\tilde{K}_i(j\lambda,\omega)$ avoids the circle with center $-\sigma_i/\rho_i$ and radius $|\Delta_i/\rho_i|$ for almost every $\omega\in\Omega$. A minimum $\alpha_i\in R^+$ is sought so that this condition is met.

(b) If $\rho_i<0$, (3) may be written, for almost every $\omega\in\Omega$, as

$$0\geqq\frac{1}{\rho_i}+\left(\frac{\sigma_i}{\rho_i}\right)\tilde{K}_i+\left(\frac{\sigma_i}{\rho_i}\right)\tilde{K}_i^*+\tilde{K}_i\tilde{K}_i^*,$$

and we proceed to (4) with the inequality reversed. This implies that we are seeking a minimum $\alpha_i\in R^+$ such that the locus of $\tilde{K}_i(j\lambda,\omega)$, $\lambda\in R$, is contained in a circle with center $-\sigma_i/\rho_i$ and radius $|\Delta_i/\rho_i|$ for almost every $\omega\in\Omega$.

(c) If $\rho_i = 0$ (if ρ_i changes sign for $0 < \sigma_i < 1$ we need to consider the possibility of this case) we may write (3), for almost every $\omega \in \Omega$, as

$$0 \leqq \sigma_i \tilde{K}_i + \alpha_i K_i^* + 1,$$

or equivalently as

$$\sigma_i \operatorname{Re}(\tilde{K}_i) + \tfrac{1}{2} \geqq 0.$$

That is, we require

$$\operatorname{Re} \tilde{K}_i(j\lambda, \omega) \geqq -\frac{1}{2\sigma_i}, \qquad \lambda \in R, \quad \text{a.e. } [P]$$

(for ρ_i to change signs for $0 < \alpha_i < 1$, it is necessary that $\sigma_i > 0$).

Remark 5. Condition (ii) of Theorem 1 may be checked graphically if $N_i = 1$, by applying the principle of the argument (see, for instance, Holtzman [3]) for complex functions. To satisfy the condition

$$1 + \tfrac{1}{2}(a_i + b_i)\tilde{K}_i(s, \omega) \neq 0 \quad \text{for } \operatorname{Re}(s) \geqq 0, \quad \text{a.e. } [P],$$

we require that the locus of $\tilde{K}_i(j\lambda, \omega)$, $\lambda \in R$, does not encircle the point $(2/(a_i + b_i), 0)$ with probability one.

An example using these graphical techniques is worked in § 5.

The following theorem is a composite stochastic system version of the *Popov stability criterion.*

THEOREM 2. *System* (S) *is stochastically absolutely stable if the following conditions hold*:

(i) $N_i = 1$, $i \in M$;
(ii) $D_{ij} = 0$, $i, j \in M$, *B_{ij} is a Type B operator with* $\phi_{ij} e_j(t, \omega) = e_j(t, \omega)$, $i, j \in M$;
(iii) $\psi_i \in \eta_{(1)}$ *with* $a_i = 0$, $b_i > 0$, $i \in M$;
(iv) $k_i, \dot{k}_i \in L_{1(1)}(R^+, L_\infty(\Omega))$, $k_i \in L_2(R^+, L_\infty(\Omega))$, $i \in M$;
(v) $r_i, \dot{r}_i \in L_{2(1)}(R^+, L_\infty(\Omega))$, $|r_i(t, \omega)| \to 0$ *as* $t \to \infty$, *a.e.* $[P]$, $i \in M$;
(vi) *there exists a* $q_i > 0$ *such that*

$$\operatorname{Re}[(1 + j\lambda q_i)\tilde{K}_i(j\lambda, \omega)] + b_i^{-1} \geqq \delta_i > 0, \quad \textit{a.e. } [P],$$

for some real δ_i; *and*

(vii) *there exist constants* α_i, γ_{ij}, *and* β_{ij} *with*

$$\alpha_i \geqq \sup_{\lambda \in R^+} |\tilde{K}_i(j\lambda, \omega)| \quad \textit{a.e. } [P],$$

$$\gamma_{ij} \geqq \sup_{\lambda \in R^+} |(1 + j\lambda q_i)\tilde{B}_{ij}(j\lambda, \omega)| \quad \textit{a.e. } [P], \quad \textit{and}$$

$$\beta_{ij} \geqq \sup_{\lambda \in R^+} |\tilde{B}_{ij}(j\lambda, \omega)| \quad \textit{a.e. } [P],$$

such that the matrix $A = [a_{ij}]$ *has positive successive principal minors, where*

$$a_{ij} = \begin{cases} 1 - (\beta_{ii} + \alpha_i \delta_i^{-1} \gamma_{ii}), & i = j \\ -\alpha_i \delta_i^{-1} \gamma_{ij} - \beta_{ij}, & i \neq j, \quad i, j \in M, \end{cases}$$

(the matrix A is referred to as a *test matrix*).

Remark 6. The comments of Remark 2 hold for Theorem 2 as well. For the deterministic case with $m = N_1 = 1$, Theorem 2 reduces to a version of the Popov-like theorems of Sandberg [21] and Zames [24], [25]. For $m = 1$, Theorem 2 is somewhat similar to Theorem 9.2.1 of Tsokos and Padgett [23]. We do not, however, require boundedness or continuity of the nonlinearity, ψ_i, as in [23].

Remark 7. Condition (vi) of Theorem 2 is the familiar Popov condition. The value of δ_i may be determined graphically. It is the minimum distance, parallel to the real axis, between the graph of the modified Nyquist plot of the linear operator K_i and the Popov line with intercept $-b_i^{-1}$ and slope q_i^{-1}.

Remark 8. In setting the D_{ij}, $i, j \in M$, terms of (S) to zero, we are allowing the subsystems to be interconnected only through the error terms, $e_i(t, \omega)$. This, however, is quite natural when applying the theorem to interconnected systems described by differential equations (see § 4) and is rather flexible in control system work (see McClamroch [7]).

4. Applications to stochastic nonlinear differential equations. In this section we present conditions for stability of interconnected stochastic systems governed by one of the following two types of differential equations,

$$\frac{dx_i(t, \omega)}{dt} = A_i(\omega)x_i(t, \omega) + \psi_i(x_i(t, \omega), t, \omega) + f_i(t, \omega) + \sum_{j=1}^{m} d_{ij}(\omega)x_j(t, \omega), \tag{5}$$

or

$$\frac{dx_i(t, \omega)}{dt} = A_i(\omega)x_i(t, \omega) + v_i(\omega)\psi_i(x_i(t, \omega), t, \omega) + f_i(t, \omega) + \sum_{j=1}^{m} d_{ij}(\omega)\sigma_i(t, \omega), \tag{6}$$

with

$$\sigma_i(t, \omega) = c_i^T(\omega)x_i(t, \omega),$$

where for both equations $i, j \in M = \{1, 2, \cdots, m\}$. It is assumed that for (5) and (6) with $i, j \in M$, $A_i(\omega)$ is an $N_i \times N_i$ matrix whose elements are $\mathscr{F}$-measurable functions of ω; $x_i(t, \omega)$, $c_i(\omega)$, $v_i(\omega)$, and $f_i(t, \omega)$ are $N_i \times 1$ vectors whose elements are random variables for each $t \in R^+$, and where the elements of $c_i(\omega)$ and $v_i(\omega)$ are essentially bounded; $d_{ij}(\omega)$ is an $N_i \times N_j$ random matrix; $\sigma_i(t, \omega)$ is a scalar random variable for each $t \in R^+$. For $M = \{1\}$, (6) is similar to one studied by Tsokos [22] and Tsokos and Padgett [23].

We apply Theorem 1 to determine conditions for stochastic absolute stability of systems governed by (5) and Theorem 2 to determine conditions for the stability of systems governed by (6).

THEOREM 3. *Differential system* (5) *is* stochastically absolutely stable *if the following conditions hold*:

(i) $A_i(\omega)$ *is a stochastically stable matrix*, $i \in M$;

(ii) $f_i \in L_{2(N_i)}(R^+, L_\infty(\Omega))$, $i \in M$;

(iii) $\psi_i \in \eta_{(N_i)}$, $i \in M$;

(iv) $\det[(s + \frac{1}{2}(a_i + b_i))I_{(N_i)} - A_i] \neq 0$ *for* $\operatorname{Re}(s) \geqq 0$ *a.e.* $[P]$, $i \in M$; *and*

(v) *there exist constants* γ_i, δ_{ik} *with*

$$\gamma_i \geqq \tfrac{1}{2}(b_i - a_i) \sup_{\lambda \in R^+} \Lambda\{[(j\lambda + \tfrac{1}{2}(a_i + b_i))I_{(N_i)} + A_i(\omega)]^{-1}\} \quad a.e.\ [P], \quad and$$

$$\delta_{ik} \geqq \sup_{\lambda \in R^+} \Lambda\{[(j\lambda + \tfrac{1}{2}(a_i + b_i))I_{(N_i)} + A_i(\omega)]^{-1}[j\lambda I - A_i(\omega)]\, d_{ij}(\omega)\} \quad a.e.\ [P],$$

such that the matrix $C = [c_{ik}]$ *has positive successive principal minors, where*

$$c_{ik} = \begin{cases} 1 - \gamma_i, & i = k \\ -\delta_{ik}, & i \neq k, \quad i, k \in M. \end{cases}$$

THEOREM 4. *Differential system* (6) *is* stochastically absolutely stable *if the following conditions hold*:

(i) $\psi_i \in \eta_{(1)}$, *with* $a_i = 0$, $b_i > 0$, $i \in M$;

(ii) $A_i(\omega)$ *is a stochastically stable matrix*, $i \in M$;

(iii) $f_i \in L_{2(1)}(R^+, L_\infty(\Omega))$, $i \in M$;

(iv) *there exists a* $q_i > 0$ *such that*

$$\operatorname{Re}\,[(1+j\lambda q_i)c_i^T(\omega)(j\lambda I - A_i(\omega))^{-1}v_i(\omega)] + b_i \geqq \delta_i > 0,$$

a.e. $[P]$, *for some real* δ_i; *and*

(v) *there exist constants* α_i, γ_{ik}, B_{ik} *with*

$$\alpha_i \geqq \sup_{\lambda \in R^+} |c_i^T(\omega)(j\lambda I_{(N_i)} - A_i(\omega))^{-1}v_i(\omega)| \quad a.e.\ [P],$$

$$\gamma_{ik} \geqq \sup_{\lambda \in R^+} |(1+j\lambda q_i)c_i^T(\omega)(j\lambda - A_i(\omega))^{-1}v_i(\omega)\, d_{ik}(\omega)| \quad a.e.\ [P],$$

and

$$B_{ik} \geqq \sup_{\lambda \in R^+} |c_i^T(\omega)(j\lambda I_{(N_i)} - A_i(\omega))^{-1}v_i(\omega)\, d_{ik}(\omega)| \quad a.e.\ [P],$$

such that the matrix $C = [c_{ik}]$ *has positive successive principal minors, where*

$$c_{ik} = \begin{cases} 1 - \alpha_i\delta_i^{-1}\gamma_{ii} + B_{ii}, & i = k \\ -\alpha_i\delta_i^{-1}\gamma_{ik} - B_{ik}, & i \neq k, \quad i, k \in M. \end{cases}$$

5. Example. We apply Theorem 1 to the control system shown in Fig. 1, which is described functionally by

$$\begin{aligned}
e_1(t, \omega) &= r_1(t, \omega) - F_1 y_1(t, \omega) + y_2(t, \omega), \\
y_1(t, \omega) &= \psi_1 N_1 e_1(t, \omega), \\
e_2(t, \omega) &= -F_2 y_2(t, \omega) + y_1(t, \omega) + S_3 y_3(t, \omega), \\
y_2(t, \omega) &= \psi_2 N_2 e_2(t, \omega), \\
e_3(t, \omega) &= r_3(t, \omega) - y_3(t, \omega) + y_1(t, \omega), \\
y_3(t, \omega) &= N_3 \psi_3 e_3(t, \omega),
\end{aligned}$$

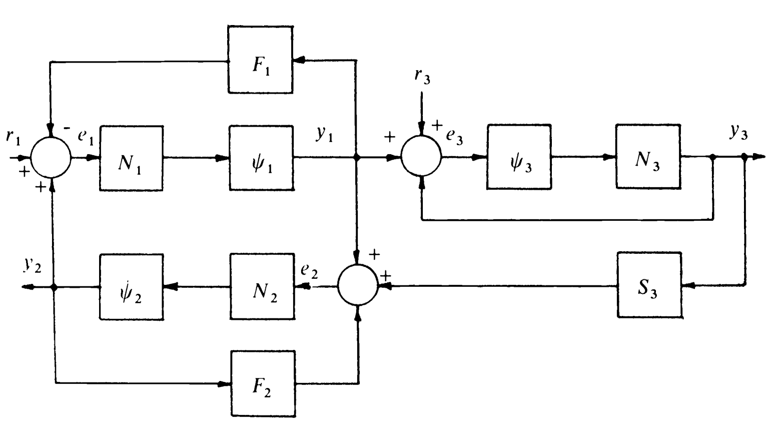

FIG. 1. *Block diagram of the system for the Example.*

where N_1, N_2, N_3, F_2, and S_3 are random convolution operators on $\mathscr{E}_{2(1)}(R^+, L_\infty(\Omega))$ characterized by their transforms:

$$\tilde{N}_1(s,\omega)=\frac{G_1(s+2)}{(s+1)(s+3)},\qquad \tilde{N}_2(s,\omega)=\frac{10}{(s+2)(s+5)},$$

$$\tilde{N}_3(s,\omega)=\frac{(s+5)}{(s+6)(s+2)},\qquad \tilde{F}_2(s,\omega)=\frac{1}{s+1},$$

$$\tilde{S}_3(s,\omega)=\frac{1}{s+d(\omega)},\qquad P[\omega\colon 4\leqq d(\omega)\leqq 6]=1;$$

F_1 is a Type A operator with

$$F_1y_1(t,\omega)=\Theta(\omega)y_1(t,\omega),\qquad P[\omega\colon \Pi\leqq\theta(\omega)\leqq 2\Pi]=1;$$

$\psi_1, \psi_2, \psi_3\in\eta_{(1)}$ with

$$\psi_1x_1(t,\omega)=\sin(x_1(t,\omega)),\qquad \psi_2x_2(t,\omega)=a_1(\omega)x_2(t,\omega),$$

$$\psi_3x_3(t,\omega)=0.5\sin(a_2(\omega)x_3(t,\omega)),$$

where $a_1(\omega)$ is a uniform random variable on $[0, 1]$, and $a_2(\omega)$ is a standard normal random variable; $e_1, e_2, e_3, y_1, y_2, y_3\in\mathscr{E}_{2(1)}(R^+, L_\infty(\Omega))$ and $r_1, r_3\in L_{2(1)}(R^+, L_\infty(\Omega))$ (we will define $r_2(t,\omega)\equiv 0$). The gain $G_1\in R^+$ is to be determined in such a fashion as to insure a stochastically absolutely stable system. Note that in the present form, the system is not of the form of (S). In order to restructure the problem, define

$$e_i'(t,\omega)=N_ie_i(t,\omega),\qquad i=1,2,\quad\text{and}$$

$$r_i'(t,\omega)=N_ir_i(t,\omega),\qquad i=1,2.$$

We now find

$$\begin{aligned} &e_1'(t,\omega)=r_1'(t,\omega)-N_1F_1\psi_1e_1'(t,\omega)+N_1\psi_2e_2'(t,\omega),\\ (7)\qquad &e_2'(t,\omega)=r_2'(t,\omega)-N_2F_2\psi_2e_2'(t,\omega)+N_2\psi_1e_1'(t,\omega)+N_2S_3y_3(t,\omega),\\ &e_3(t,\omega)=r_3(t,\omega)-N_3\psi_3e_3(t,\omega)+\psi_1e_1'(t,\omega). \end{aligned}$$

This system is depicted in Fig. 2.

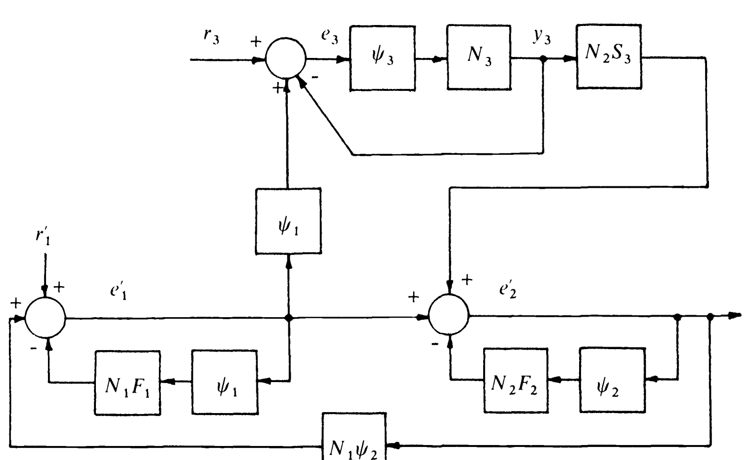

FIG. 2. *Modified block diagram of the system for the Example.*

Note that if modified system (7) is stochastically absolutely stable, then the original system is also. By comparison with (S) we may make the following identifications: $K_1 = N_1F_1$, $K_2 = N_2F_2$, $K_3 = N_3$, $B_{12} = N_1\psi_2$, $B_{21} = N_2\psi_1$, $B_{31} = \psi_1$, $D_{23} = N_2S_3$, and $B_{11} = B_{13} = B_{22} = B_{23} = B_{32} = B_{33} = D_{11} = D_{12} = D_{13} = D_{21} = D_{22} = D_{31} = D_{32} = D_{33} = 0$. We check conditions (ii) and (iii) of Theorem 1 by the graphical method of Remark 4. The Nyquist plots of K_2 and K_3 are shown in Figs. 4 and 5. Note that since $K_1 = N_1F_1$ depends explicitly on ω through F_1, we show in Fig. 3 a region where the locus of $\tilde{K}_1(j\lambda, \omega)$ will fall with probability one. From the nonlinear

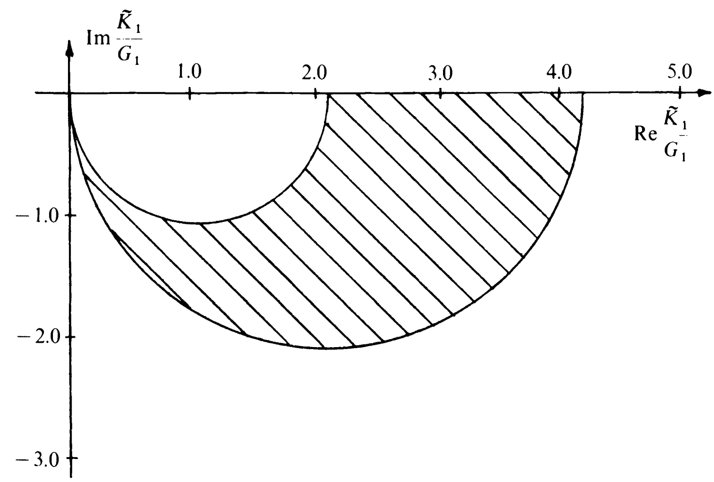

FIG. 3. *Nyquist plot of $\tilde{K}_1(j\lambda, \omega)/G_1$ for the Example.*

elements, we determine the following parameters: $a_1 = -0.2122$, $b_1 = 1.0$, $a_2 = 0$, $b_2 = 1.0$, $a_3 = -0.5$, and $b_3 = 0.5$. We apply Remark 4 to determine the A-matrix terms α_i, $i = 1, 2, 3$. For the first subsystem (S_1) we have from (2) that

$$\Delta_1 = \frac{b_1 - a_1}{2\alpha_1} = \frac{0.6061}{\alpha_1}, \qquad \sigma_1 = \frac{a_1 + b_1}{2} = 0.3939, \quad \text{and}$$

$$\rho_1 = \sigma_1^2 - \Delta_1^2 = 0.1552 - \frac{0.3674}{\alpha_1^2}.$$

Since $\rho_1 < 0$ for $0 < \alpha < 1$, we apply case (b) of Remark 4 and therefore we are looking for a circle that contains the locus of $\tilde{K}_1(j\lambda, \omega)$, $\lambda \in R$, with probability one. The center and radius of allowable circles are given by c_1 and r_1 respectively, and may be computed from Remark 4:

$$c_1 = -\frac{\sigma_1}{\rho_1} = \frac{0.3939\alpha_1^2}{0.3674 - 0.1552\alpha_1^2}, \quad \text{and} \quad r_1 = \left|\frac{\Delta_1}{\rho_1}\right| = \frac{0.6061\alpha_1}{0.3674 - 0.1552\alpha_1^2}.$$

Note that c_1 and r_1 are monotone increasing functions of α_1 for $0 < \alpha_1 < 1$. Using this fact and Fig. 3, we seek the minimum α_1 such that

$$c_1 + r_1 = \max_{\omega \in \Omega} \tilde{K}_1(0, \omega) = 4.1888G_1.$$

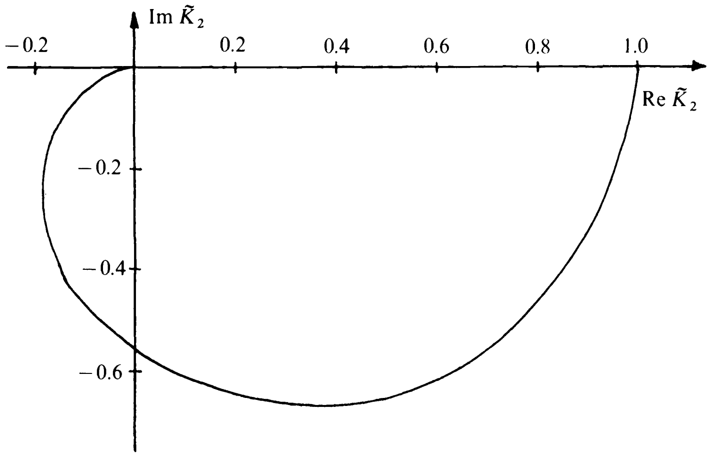

FIG. 4. *Nyquist plot of $\tilde{K}_2(j\lambda)$ for the Example.*

This is equivalent to the relation for α_1,

$$\alpha_1 = \frac{\sqrt{0.3674 + 2.4248G_1 + 4.0G_1^2} - 0.6061}{0.7878 + 13.002G_1}. \tag{8}$$

In a similar fashion we solve for α_2 and α_3,

$$\alpha_2 = 0.3333, \quad \text{and} \quad \alpha_3 = 0.2084.$$

In order to determine the A-matrix coefficient γ_{12}, observe that

$$\left\| (1 + \tfrac{1}{2}(a_1 + b_1)K_1)^{-1} \frac{K_1}{\theta} \right\| \leqq \frac{\alpha_1}{\theta} \leqq \frac{\alpha_1}{\Pi}.$$

So we choose $\gamma_{12} = \alpha_1/\Pi$. Similarly we choose $\gamma_{21} = 0.825$, $\gamma_{31} = 1.0$, $\xi_{23} = 0.2063$, and $\mu_2 = 1$. The remaining A-matrix parameters are zero. We compute the test matrix, A, as

$$A = \begin{bmatrix} 1-\alpha_1 & -\gamma_{12} & 0 \\ -\gamma_{21} & 1-\alpha_2 & -\xi_{23} \\ -\gamma_{31} & 0 & 1-\alpha_3 \end{bmatrix} = \begin{bmatrix} 1-\alpha_1 & -\alpha_1/\Pi & 0 \\ -0.825 & 0.6667 & -0.2063 \\ -1 & 0 & 0.7916 \end{bmatrix}.$$

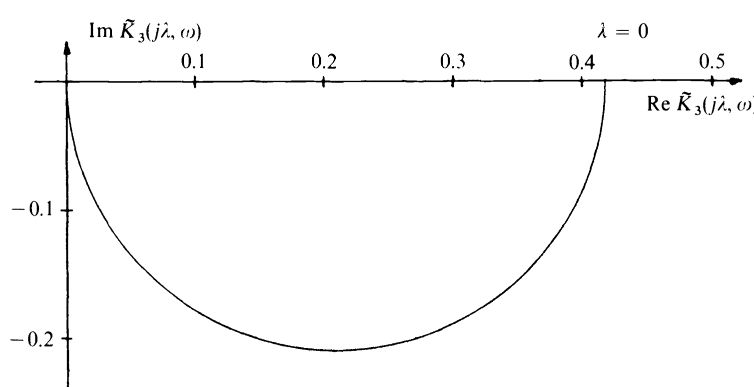

FIG. 5. *Nyquist plot of $\tilde{K}_3(j\lambda, \omega)$ for the Example.*

In order to satisfy condition (iii) of Theorem 1, we need positive successive principal minors of A, that is

$$1-\alpha_1>0, \qquad (1-\alpha_1)(0.6667)-\frac{0.825\alpha_1}{\Pi}>0, \quad \text{and}$$

$$0.7916\left((1-\alpha_1)(0.6667)-\frac{0.825\alpha_1}{\Pi}\right)-\frac{0.2063\alpha_1}{\Pi}>0.$$

It can be seen that all three inequalities are satisfied if the third one is satisfied. This happens for $\alpha_1<0.6585$. In order to compute G_1, we must satisfy (8), that is

$$\frac{\sqrt{0.3674+2.4248G_1+4.0G_1^2}-0.6061}{0.7878+1.3002G_1}<0.6585$$

or

$$0<G_1<0.4535.$$

6. Concluding remarks. New stochastic absolute stability results for a large class of multi-input-multi-output systems modeled by stochastic Volterra-type equations were established. Whenever appropriate, frequency domain interpretations were emphasized. The deterministic results obtained when the probability space $(\Omega, \mathcal{F}, P)$ is trivial, are also new.

As in related existing results for deterministic interconnected systems, [5], [8], [13], etc., the objective here was always the same: to analyze composite systems in terms of lower order subsystems and in terms of the interconnecting structure.

To demonstrate the method of analysis advanced, we applied the results to two types of systems described by differential equations and to a control system.

Appendix. Proofs of theorems.

Proof of Theorem 1. Define the following subsets of Ω:

$C_2=\{\omega\in\Omega$: (i), (ii), (iii), (iv) of the theorem are satisfied for all $i\in M\}$;

$C_{3,i}=\{\omega\in\Omega$: $e_i\in\mathscr{E}_{2(N_i)}$ satisfies (S)$\}$; and

$C_{4,i}=\{\omega\in\Omega$: $k_i\in L_{1(N_i)}(R^+)\cap L_{2(N_i)}(R^+)\}$.

Recall the definition of $C_1(\psi)$ from § 2, and define

$$D=\{\bigcap_{i\in M}C_1(\psi_i)\}\cap\{C_2\}\cap\{\bigcap_{i\in M}C_{3,i}\}\cap\{\bigcap_{i\in M}C_{4,i}\}.$$

Note that $P[D]=1$. Using the definition of the operators K_i, Q_i, and Π_T given in § 2, we may rewrite (S) as

$$e_i(t,\omega)=u_i(t,\omega)-K_iQ_ie_i(t,\omega), \qquad t\in R^+, \quad \omega\in D.$$

Truncating at $T\in R^+$, we have

$$e_{iT}(t,\omega)=u_{iT}(t,\omega)-\Pi_TK_iQ_ie_i(t,\omega), \qquad t, T\in R^+, \quad \omega\in D.$$

Since K_i and Q_i are causal operators, $i\in M$, we may write

$$e_{iT}(t,\omega)=u_{iT}(t,\omega)-\Pi_TK_i\Pi_TQ_ie_{iT}(t,\omega), \qquad t, T\in R^+, \quad \omega\in D$$

or

$$\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)e_{iT}(t,\omega)=u_{iT}(t,\omega)-\Pi_T K_i \Pi_T(Q_i-\tfrac{1}{2}(a_i+b_i)I)e_{iT}(t,\omega)$$

where I denotes the identity operator on $L_{2(N_i)}(R^+, L_\infty(\Omega))$. For $\omega \in D$, condition (ii) of the theorem assures that $(I+\frac{1}{2}(a_i+b_i)K_i)^{-1}$ exists on $L_{2(N_i)}(R^+)$ and is causal (see Sandberg [19] or Miller [12]), so that

$$e_{iT}(t,\omega)=\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}u_{iT}(t,\omega)-\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1} \cdot \Pi_T K_i \Pi_T(Q_i-\tfrac{1}{2}(a_i+b_i)I)e_{iT}(t,\omega), \qquad t, T\in R^+, \quad \omega \in D, \quad i \in M,$$

and hence

$$\|e_{iT}\| \leqq \|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}u_{iT}\|+\|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}\Pi_T K_i\| \cdot \|\Pi_T(Q_i-\tfrac{1}{2}(a_i+b_i)I)\| \cdot \|e_{iT}\|, \qquad \omega\in D, \quad T \in R^+, \quad i \in M.$$

Since Π_T is a projection on $L_{2(N_i)}(R^+)$ (for fixed $\omega \in D$) we have

$$\|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}\Pi_T K_i\| \leqq \|(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}K_i\| \leqq \sup_{\lambda \in R^+} \Lambda[(I+\tfrac{1}{2}(a_i+b_i)\tilde{K}_i(j\lambda,\omega))^{-1}\tilde{K}_i(j\lambda,\omega)]$$

(see Sandberg [19]). Also note that

$$\|Q_i-\tfrac{1}{2}(a_i+b_i)\| \leqq \tfrac{1}{2}(b_i-a_i), \qquad \omega \in D, \quad i \in M.$$

Thus

$$\|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}\Pi_T K_i\| \cdot \|\Pi_T(Q_i-\tfrac{1}{2}(a_i+b_i)I)\| \leqq \alpha_i$$

(by the definition of α_i given in (iii) of the theorem). Defining δ_i by

$$\|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}\| \leqq \delta_i, \qquad \omega \in D, \quad i \in M,$$

we have

$$\begin{aligned}\|e_{iT}\| &\leqq \|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}u_{iT}\|+\alpha_i\|e_{iT}\| \\ &\leqq \delta_i\|r_{iT}\|+\sum_{j=1}^m \|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}\Pi_T B_{ij}e_{jT}\| \\ &\quad +\sum_{j=1}^m \|\Pi_T(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}\Pi_T D_{ij}y_{jT}\|+\alpha_i\|e_{iT}\|\end{aligned}$$

for $T \in R^+$, $i \in M$, and $\omega \in D$. This implies that

$$\|e_{iT}\| \leqq \delta_i\|r_{iT}\|+\sum_{j=1}^m \|(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}B_{ij}\| \cdot \|e_{jT}\| + \sum_{j=1}^m \|(I+\tfrac{1}{2}(a_i+b_i)K_i)^{-1}D_{ij}K_j\| \cdot \|Q_j\| \cdot \|e_{jT}\|+\alpha_i\|e_{iT}\|.$$

Using the fact that $\|Q_i\| \leqq \max(|b_i|, |a_i|)$ and the definitions of μ_i, γ_{ik}, and ξ_{ik} as given in (iii) of the theorem, we have

$$\|e_{iT}\| \leqq \delta_i\|r_{iT}\|+\sum_{j=1}^m \gamma_{ij}\|e_{jT}\|+\sum_{j=1}^m \mu_j\xi_{jj}\|e_{jT}\|+\alpha_i\|e_{iT}\|, \qquad i, j \in M.$$

Using vector notation with $\|e_T\| = [\|e_{1T}\|, \cdots, \|e_{mT}\|]^T$, and $\|r_T\|$ defined similarly, we have

$$A\|e_T\| \leqq [\operatorname{diag}(\delta_i)]\|r_T\|$$

by the definition of A given in (iii) of the theorem. Since by condition (iii) A is an M-matrix, it possesses an inverse consisting of all nonnegative elements, and hence

$$\|e_T\| \leqq A^{-1}[\operatorname{diag}(\delta_i)]\|r_T\|.$$

For $\omega \in D$ we have $r_i \in L_{2(N_i)}(R^+)$ and by letting $T \to \infty$ we have $e_i \in L_{2(N_i)}(R^+)$.

Using the matrix notation $[e] = [e_1(t, \omega), \cdots, e_m(t, \omega)]^T$, with $[r]$ and $[y]$ being defined similarly, and defining the operators on $L_{2(\sum N_i)}(R^+, L_\infty(\Omega))$

$$K = [\operatorname{diag}(K_i)], \qquad Q = [\operatorname{diag}(Q_i)], \qquad D = [D_{ij}],$$

and recalling the definitions of B_A and B_B from § 2, we write (S) as

$$[e] = [r] - KQ[e] + B_A[e] + B_B[e] + DKQ[e].$$

By condition (iv), $(I - B_A)$ has a bounded inverse for $t \geqq T^*$, $\omega \in D$, for some $T^* \in R^+$. We therefore have

$$[e] = (I - B_A)^{-1}\{[r] - KQ[e] + B_B[e] + D[y]\}.$$

Observe that since $e_i \in L_{2(N_i)}(R^+)$ for $\omega \in D$, then $Q_i e_i \in L_{2(N_i)}(R^+)$ for $\omega \in D$. Note that $(KQ[e] + B_B[e] + DKQ[e])$ may be written as a linear combination of integrals of the form $\int_0^t g_i(t-\tau, \omega)h_i(\tau, \omega)\,d\tau$, where for $\omega \in D$, $g_i \in K_{1(1\times 1)}(R^+) \cap K_{2(1\times 1)}(R^+)$ and $h_i \in L_2(R^+)$. Therefore using Lemma 2 it follows that these integrals approach zero as $t \to \infty$. Since $|r_i| \to 0$ as $t \to \infty$ by hypothesis, the theorem follows. □

Proof of Theorem 2. Define the following subsets of Ω:

$$C_2 = \{\omega \in \Omega\text{: (iv), (v), (vi), and (vii) of the theorem are satisfied}\};$$

$$C_{3i} = \{\omega \in \Omega\text{: } e_i(t, \omega) \text{ satisfies (S)}\};$$

$$C_{4i} = \{\omega \in \Omega\text{: } k_i, \dot{k}_i \in K_{1(1\times 1)}(R^+), k_i \in K_{2(1\times 1)}(R^+)\}; \text{ and}$$

$$C_{5i} = \{\omega \in \Omega\text{: } r_i, \dot{r}_i \in L_{2(1)}(R^+), |r_i(t, \omega)| \to 0 \text{ as } t \to \infty\}.$$

Recall the definition of $C_1(\psi)$ from § 2, and define

$$D = \{\bigcap_{i \in M} C_1(\psi_i)\} \cap \{C_2\} \cap \{\bigcap_{i \in M} C_{3i}\} \cap \{\bigcap_{i \in M} C_{4i}\} \cap \{\bigcap_{i \in M} C_{5i}\}.$$

Note that $P(D) = 1$. Condition (iv) guarantees that for $\omega \in D$ we may write the operator K_i as

$$K_i = K_{2i}K_{1i},$$

where K_{1i} is a linear mapping of $\mathscr{E}_{2(1)}$ into itself and K_{2i} maps $\mathscr{E}_{2(1)}$ into $\mathscr{E}_{s(1)}$ characterized by its transform:

$$K_{2i}(j\lambda, \omega) = (1 + j\lambda q_i)^{-1}.$$

Also note that there exists a time-invariant linear mapping K_{i3} of $\mathscr{E}_{s(1)}$ into $\mathscr{E}_{2(1)}$ such that

$$K_{3i}K_{2i} = I \text{ (the identity operator on } \mathscr{E}_{2(1)}\text{); and}$$

$$K_{2i}K_{3i} = I \text{ (the identity operator on } \mathscr{E}_{s(1)}\text{).}$$

Note that K_{i3} is characterized by the transform

$$K_{i3}(j\lambda, \omega) = (1 + j\lambda q_i).$$

We define the new variables $v_i(t, \omega) = K_{3i}e_i(t, \omega)$ (or $e_i(t, \omega) = K_{2i}v_i(t, \omega)$) and $z_i(t, \omega) = Q_i K_{2i} v_i(t, \omega)$. From (S) we have

$$K_{3i}e_i(t, \omega) = K_{3i}u_i(t, \omega) - K_{3i}K_{2i}K_{1i}Q_i e_i(t, \omega)$$

or

$$v_i(t, \omega) = K_{3i}u_i(t, \omega) - K_{1i}Q_i K_{2i} v_i(t, \omega)$$

or finally

$$v_i(t, \omega) = K_{3i}u_i(t, \omega) - K_{1i}z_i(t, \omega),$$

from which we may write

$$\text{(A.1)} \qquad \begin{aligned} \langle (K_{3i}u_i)_T, z_{iT}\rangle &= \langle (K_{1i}z_i)_T, z_{iT}\rangle + \langle v_{iT}, z_{iT}\rangle \\ &= \langle (K_{1i}z_i)_T, z_{iT}\rangle + \langle v_{iT}, (Q_i K_{2i} v_i)_T\rangle. \end{aligned}$$

Since $Q_i x(t, \omega) \leqq b_i x(t, \omega)$ for $x \in L_2(R^+, L_\infty(\Omega))$, $\omega \in D$ and $K_{2i}(j\lambda, \omega) = (1 + j\lambda q_i)^{-1}$, by an application of Lemma 2 of Zames [25], we have

$$\langle v_{iT}, (Q_i K_{2i} v_i)_T\rangle \geqq b_i^{-1}\|z_{iT}\|^2, \qquad T \in R^+, \quad v_i \in \mathscr{E}_{2(1)}, \quad \omega \in D, \quad i \in M.$$

Also note that

$$\begin{aligned} \langle (K_{1i}z_i)_T, z_{iT}\rangle &= \langle (K_{1i}z_{iT}), z_{iT}\rangle \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{K}_{1i}(j\lambda, \omega)\tilde{z}_{iT}(j\lambda, \omega)\tilde{z}_{iT}^*(j\lambda, \omega)\, d\lambda \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \operatorname{Re}\{(1 + j\lambda q_i)\tilde{K}_{1i}(j\lambda, \omega)\}|\tilde{z}_{iT}(j\lambda, \omega)|^2\, d\lambda \\ &\geqq \frac{1}{2\pi}\int_{-\infty}^{\infty} (\delta_i - b_i^{-1})|\tilde{z}_i(j\lambda, \omega)|^2\, d\lambda \\ &= (\delta_i - b_i^{-1})\|z_{iT}\|^2. \end{aligned}$$

The above result uses condition (vi) of the theorem. Equation A.1 becomes

$$\begin{aligned} b_i^{-1}\|z_{iT}\|^2 + (\delta_i - b_i^{-1})\|z_{iT}\|^2 &\leqq \langle (K_{3i}u_i)_T, z_{iT}\rangle \\ &\leqq \|K_{i3}u_{iT}\| \cdot \|z_{iT}\| \end{aligned}$$

or

$$\|z_{iT}\| \leqq \delta_i^{-1}\|K_{3i}u_{iT}\|.$$

We may now write

$$\begin{aligned} \|e_{iT}\| \leqq \|(K_i z_i)_T\| + \|u_{iT}\| &\leqq \|K_i\|\delta_i^{-1}\|K_{3i}u_{iT}\| + \|u_{iT}\| \\ &\leqq \|K_i\|\delta_i^{-1}\left\{\|K_{3i}r_{iT}\| + \sum_{j=1}^{m}\|K_{3i}B_{ij}e_{jT}\|\right\} + \|r_{iT}\| + \sum_{j=1}^{m}\|B_{ij}e_{jT}\|. \end{aligned}$$

Using the notation $\alpha_i = \|K_i\|$, $\gamma_{ij} = \|K_{3i}B_{ij}\|$, $\beta_{ij} = \|B_{ij}\|$, we have

$$\|e_{iT}\| \leqq \alpha_i\delta_i^{-1}\|K_{3i}r_{iT}\| + \alpha_i\delta_i^{-1}\sum_{j=1}^{m}(\gamma_{ij} + \beta_{ij})\|e_{jT}\| + \|r_{iT}\|$$

or, using the vector notation used in the proof of Theorem 1, we have

$$A\|e_T\| \leqq [\operatorname{diag}(\alpha_i \delta_i^{-1})]\|K_3 r_T\| + \|r_T\|,$$

where the matrix A is given in (vii) of the theorem. Since by hypothesis $r_i \in L_2(R^+, L_\infty(\Omega))$, for $\omega \in D$ we have $r_i \in L_2(R^+)$ and hence $K_{3i} r_i \in L_2(R^+)$ for $\omega \in D$ (see Holtzman [3, Chap. VIII]). Since A is an M-matrix by condition (vii), we have as a result that $e_i \in L_{2(1)}(R^+)$ for $\omega \in D$, $i \in M$. By the argument in the proof of Theorem 1, since $e_i \in L_2(R^+)$ for $\omega \in D$, then $Q_i e_i \in L_2(R^+)$ for $\omega \in D$ and $y_i \in L_2(R^+)$ for $\omega \in D$; and since $k_i \in K_{1(1\times 1)}(R^+) \cap K_{2(1\times 1)}(R^+)$ for $\omega \in D$, then by Lemma 2 and the fact that $|r_i(t, \omega)| \to 0$ as $t \to \infty$ we have $|e_i(t, \omega)| \to 0$ as $t \to \infty$ a.e. $[P]$.

Proof of Theorem 3. From (5) we have

$$\frac{dx_i(t, \omega)}{dt} - A_i(\omega) x_i(t, \omega) = -\psi_i(x_i(t, \omega), t, \omega) + f_i(t, \omega) + \sum_{\substack{j=1 \\ j \neq i}}^{m} d_{ij}(\omega) x_j(t, \omega),$$

or, using the usual differential equation techniques,

$$\frac{d}{dt}[e^{-A_i(\omega)t} x_i(t, \omega)] = -e^{-A_i(\omega)t} \psi_i(x_i(t, \omega), t, \omega) + e^{-A_i(\omega)t} f_i(t, \omega)$$
$$+ \sum_{\substack{j=1 \\ j \neq i}}^{M} e^{-A_i(\omega)t} d_{ij}(\omega) x_j(t, \omega)$$

or

$$x_i(t, \omega) = e^{A_i(\omega)t} x_i(0) - \int_0^t e^{A_i(\omega)(t-\tau)} \psi_i(x_i(\tau, \omega), \tau, \omega)\, d\tau$$
$$+ \int_0^t e^{A_i(\omega)(t-\tau)} f_i(\tau, \omega)\, d\tau$$
$$+ \sum_{\substack{j=1 \\ j \neq i}}^{m} \int_0^t e^{A_i(\omega)(t-\tau)} d_{ij}(\omega) x_j(\tau, \omega)\, d\tau$$

which is of the form of (S) with the following assignments:

$$e_i(t, \omega) = x_i(t, \omega);$$
$$r_i(t, \omega) = e^{A_i(\omega)t} x_i(0) + \int_0^t e^{A_i(\omega)(t-\tau)} f_i(\tau, \omega)\, d\tau;$$
$$k_i(t-\tau, \omega) = e^{A_i(\omega)(t-\tau)};$$
$$B_{ij} e_j(t, \omega) = \int_0^t e^{A_i(\omega)(t-\tau)} d_{ij}(\omega) e_j(\tau, \omega)\, d\tau, \qquad i \neq j.$$

We will show that, under the conditions of this theorem, the hypotheses of Theorem 1 are satisfied.

Note that the elements of $k_i(t, \omega)$ are linear combinations of $t^k \exp(\rho_i(\omega)t)$, $t^k \exp(\rho_j(\omega)t) \sin \sigma_j(\omega)t$, and $t^k \exp(\rho_j(\omega)t) \cos \sigma_j(\omega)t$, where $k \in \{0, 1, \cdots, N_i\}$, $\rho_j(\omega)$ denotes the real part of the jth eigenvalue of $A_i(\omega)$ and $\sigma_j(\omega)$ is related to the jth eigenvalue of $A_i(\omega)$. Recall from the definition of a stochastically stable matrix that $\operatorname{Re}(\lambda_k(\omega)) \leqq -\gamma < 0$ a.e. $[P]$. Hence we have $k_i \in K_{1(N_i \times N_i)}(R^+, L_\infty(\Omega))$, $i \in M$.

Clearly, $r_i \in L_{2(N_i)}(R^+, L_\infty(\Omega))$ by the same argument. Also

$$\int_0^t k_i(t-\tau, \omega) f_i(\tau, \omega) \to 0 \quad \text{as } t \to \infty \quad \text{a.e. } [P]$$

by Lemma 2, since $k_i \in K_{1(N_i \times N_i)}(R^+) \cap K_{2(N_i \times N_i)}(R^+)$ for almost every ω and $f_i \in L_{2(N_i)}(R^+)$ for almost every ω. Obviously $|e^{A_i(\omega)} x_i(0)| \to 0$ as $t \to \infty$ a.e. $[P]$. Hence condition (ii) of Theorem 1 is satisfied. In this case

$$\begin{aligned} \det[I + \tfrac{1}{2}(a_i + b_i)\tilde{K}_i(s, \omega)] &= \det[I + \tfrac{1}{2}(a_i + b_i)(sI - A_i(\omega))^{-1}] \\ &= \det[sI - A_i(\omega) + \tfrac{1}{2}(a_i + b_i)I] \cdot \det[(sI - A_i(\omega))^{-1}] \\ &\neq 0, \qquad \mathrm{Re}\,(s) \geqq 0 \quad \text{a.e. } [P]. \end{aligned}$$

The above relation is due to conditions (i) and (iv) of Theorem 3. Conditions (iii) of Theorem 1 and condition (v) of Theorem 3 are equivalent as may be verified from

$$\|D_{ij}\| = 0, \qquad i, j \in M,$$

and

$$\begin{aligned} &\sup_{\lambda \in R^+} \Lambda[(I + \tfrac{1}{2}(a_i + b_i)\tilde{K}_i(j\lambda, \omega))^{-1} K_i(j\lambda, \omega)] \\ &\quad = \sup_{\lambda \in R^+} \Lambda[(I + \tfrac{1}{2}(a_i + b_i)(j\lambda I + A_i(\omega))^{-1})^{-1} (j\lambda I + A_i(\omega))^{-1}] \\ &\quad = \sup_{\lambda \in R^+} \Lambda[(j\lambda I + A_i(\omega) + \tfrac{1}{2}(a_i + b_i)I)^{-1}], \end{aligned}$$

and similarly

$$\begin{aligned} &\sup_{\lambda \in R^+} \Lambda[(I + \tfrac{1}{2}(a_i + b_i)\tilde{K}_i(j\lambda, \omega))^{-1} \tilde{B}_{ik}(j\lambda, \omega)] \\ &\quad = \sup_{\lambda \in R^+} \Lambda[(I + \tfrac{1}{2}(a_i + b_i)(j\lambda I + A_i(\omega))^{-1})^{-1} (j\lambda I + A_i(\omega))^{-1}\, d_{ik}(\omega)] \\ &\quad = \sup_{\lambda \in R^+} \Lambda[(j\lambda I + A_i(\omega) + \tfrac{1}{2}(a_i + b_i)I)^{-1}\, d_{ik}(\omega)]. \end{aligned}$$

Since B_{ij}, $i, j \in M$ is a Type B operator, condition (iv) of Theorem 1 is satisfied. The proof is now complete. □

Proof of Theorem 4. From (6) we have

$$\frac{dx_i(t, \omega)}{dt} - A_i(\omega) x_i(t, \omega) = v_i(\omega)\psi_i(\sigma_i(t, \omega), t, \omega) + f_i(t, \omega) + \sum_{\substack{j=1 \\ j \neq i}}^{m} d_{ij}(\omega)\sigma_j(t, \omega)$$

with

$$\sigma_i(t, \omega) = c_i^T(\omega) x_i(t, \omega).$$

As in the proof of Theorem 3 we arrive at

$$\begin{aligned} x_i(t, \omega) = e^{A_i(\omega)t} x_i(0, \omega) &+ \int_0^t e^{A_i(\omega)(t-\tau)} v_i(\omega)\psi_i(\sigma_i(\tau, \omega)\tau, \omega)\, d\tau \\ &+ \int_0^t e^{A_i(\omega)(t-\tau)} f_i(\tau, \omega)\, d\tau \\ &+ \sum_{\substack{j=1 \\ j \neq i}}^{m} \int_0^t e^{A_i(\omega)(t-\tau)}\, d_{ij}(\omega)\sigma_j(\tau, \omega)\, d\tau. \end{aligned}$$

Noting the definition of $\sigma_i(t, \omega)$, it is obvious that

$$\begin{aligned}\sigma_i(t, \omega) = c_i^T(\omega)\, e^{A_i(\omega)t} x_i(0, \omega) \\ &+ c_i^T(\omega) \int_0^t e^{A_i(\omega)(t-\tau)} v_i(\omega) \psi_i(\sigma_i(\tau, \omega), \tau, \omega)\, d\tau \\ &+ c_i^T(\omega) \int_0^t e^{A_i(\omega)(t-\tau)} f_i(\tau, \omega)\, d\tau \\ &+ \sum_{\substack{j=1 \\ j \neq i}}^{m} c_i^T(\omega) \int_0^t e^{A_i(\omega)(t-\tau)}\, d_{ij}(\omega) \sigma_j(\tau, \omega)\, d\tau.\end{aligned}$$

This equation is of the form of (S) with the following identifications:

$$e_i(t, \omega) = \sigma_i(t, \omega),$$

$$r_i(t, \omega) = c_i^T(\omega)\, e^{A_i(\omega)t} x_i(0, \omega) + c_i^T(\omega) \int_0^t e^{A_i(\omega)(t-\tau)} f_i(\tau, \omega)\, d\tau$$

$$k_i(t, \omega) = c_i^T(\omega)\, e^{A_i(t)} v_i(\omega)$$

$$B_{ij} e_j(t, \omega) = \int_0^t c_i^T(\omega)\, e^{A_i(\omega)(t-\tau)}\, d_{ij} e_j(\tau, \omega)\, d\tau.$$

We will show that, under the conditions of Theorem 4, the hypotheses of Theorem 2 are satisfied. As in the proof of Theorem 3 we know that $e^{A_i(\omega)t} \in K_{1(N_i \times N_i)}(R^+, L_\infty(\Omega)) \cap K_{2(N_i \times N_i)}(R^+, L_\infty(\Omega))$. Since the elements of $C_i(\omega)$ and $v_i(\omega)$ are essentially bounded, we have that $k_i \in K_{1(1\times 1)}(R^+, L_\infty(\Omega)) \cap K_{2(1\times 1)}(R^+, L_\infty(\Omega))$, and $\dot{k}_i \in K_{2(1\times 1)}(R^+, L_\infty(\Omega))$. In a similar fashion r_i, $\dot{r}_i \in L_2(R^+, L_\infty(\Omega))$. The fact that

$$c_i^T(\omega)\, e^{A_i(\omega)t} x_i(0, \omega) \to 0 \quad \text{as } t \to \infty \quad \text{a.e. } [P], \quad i \in M$$

may be seen by recalling the constituents of $e^{A_i(\omega)t}$ as given in the proof of Theorem 3. Also, the term

$$c_i^T(\omega) \int_0^t e^{A_i(\omega)(t-\tau)} f_i(\tau, \omega)\, d\tau \to 0 \quad \text{as } t \to \infty \quad \text{a.e. } [P], \quad i \in M$$

by the same argument as in the proof of Theorem 3. Since

$$\tilde{K}_i(j\lambda, \omega) = c_i^T(\omega)(j\lambda I - A_i(\omega))^{-1} v_i(\omega),$$

it may be seen that condition (vi) of Theorem 2 is satisfied by condition (iv) of Theorem 4. Condition (vii) of Theorem 2 follows as a direct consequence of the form of $\tilde{K}_i(j\lambda, \omega)$ as given above and by condition (v) of Theorem 4.

REFERENCES

[1] M. FIEDLER AND V. PTAK, *On matrices with non-positive off-diagonal elements and positive principal minors*, Czechoslovak. Math. J., 12 (1962), pp. 382–400.

[2] E. HEWITT AND K. STROMBERG, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965.

[3] J. M. HOLTZMAN, *Nonlinear System Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1970.

[4] F. KOZIN, *A survey of stability of stochastic systems*, Automatika, 5 (1969), pp. 95–112.

[5] E. L. LASLEY AND A. N. MICHEL, *Input–output stability of interconnected systems*, IEEE Trans. Automatic Control, 21 (1976), pp. 84–89.

[6] ———, *L_∞- and l_∞-stability of interconnected systems*, IEEE Trans. Circuits and Systems, 23 (1976), pp. 261–270.

[7] N. H. McClamroch, *A representation for multivariable feedback systems and its use in stability analysis. Part II: Nonlinear systems*, Report no. CUED/B—Control TR/98, Department of Engineering, Cambridge University, Cambridge, England, 1975.
[8] A. N. Michel, *Stability analysis of interconnected systems*, SIAM J. Control, 12 (1974), pp. 554–579.
[9] ———, *Stability analysis and trajectory behavior of composite systems*, IEEE Trans. Circuits and Systems, 22 (1975), pp. 305–312.
[10] ———, *Stability analysis of stochastic composite systems*, IEEE Trans. Automatic Control, 20 (1975), pp. 246–250.
[11] A. N. Michel and R. D. Rasmussen, *Stability of stochastic composite systems*, Ibid., 21 (1976), pp. 89–94.
[12] R. K. Miller, *Nonlinear Volterra Integral Equations*, Benjamin, New York, 1971.
[13] R. K. Miller and A. N. Michel, *L_2-Stability and instability of large scale systems described by integrodifferential equations*, this Journal, 8 (1977), pp. 547–557.
[14] ———, *Stability of multivariable feedback systems containing elements which are open-loop unstable*, Proceedings of the 1975 Allerton Conference on Circuit and System Theory, Urbana, Illinois (October, 1975), pp. 580–589.
[15] T. Morozan, *The method of V. M. Popov for control systems with random parameters*, J. Math. Anal. Appl., 16 (1966), pp. 201–205.
[16] A. Ostrowski, *Determinanten mit überweigender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen*, Comment. Math. Helv., 30 (1956), pp. 175–210.
[17] D. W. Porter and A. N. Michel, *Input-output stability of time-varying nonlinear multiloop feedback systems*, IEEE Trans. Automatic Control, 19 (1974), pp. 422–427.
[18] R. D. Rasmussen and A. N. Michel, *On vector Lyapunov functions for stochastic dynamical systems*, Ibid., 21 (1976), pp. 250–254.
[19] I. W. Sandberg, *On the L_2-boundedness of solutions of nonlinear functional equations*, Bell System Tech. J., 43 (1964), pp. 1581–1599.
[20] ———, *Some results on the theory of physical systems governed by nonlinear functional equations*, Ibid., 44 (1965), pp. 871–898.
[21] ———, *Some stability results related to those of V. M. Popov*, Ibid., 44 (1965), pp. 2133–2148.
[22] C. P. Tsokos, *The method of V. M. Popov for differential systems with random parameters*, J. Appl. Probability, 8 (1971), pp. 289–310.
[23] C. P. Tsokos and W. J. Padgett, *Random Integral Equations with Applications to Life Sciences and Engineering*, Academic Press, New York, 1974.
[24] G. Zames, *On the input-output stability of time-varying nonlinear feedback systems—Part I: Conditions derived using concepts of loop, gain, conicity, and positivity*, IEEE Trans. Automatic Control, 11 (1966), pp. 228–239.
[25] ———, *On the input-output stability of time-varying nonlinear feedback systems—Part II: Conditions involving circles in the frequency plane and sector nonlinearities*, Ibid., 11 (1966), pp. 465–476.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036–1410/79/1002–0002 $01.00/0

STABILIZATION OF FLOWS THROUGH POROUS MEDIA*

B. H. GILDING†

Abstract. Let $u(x, t)$ be the solution of the Cauchy–Dirichlet problem for the porous media equation in one space dimension in the quarter-plane $(0, \infty)\times(0, \infty)$. An estimate is obtained for the asymptotic behaviour of $u(x, t)$ as $t\to\infty$, in terms of the prescribed lateral boundary value of u. The paper is an enlargement on a result of Peletier [SIAM J. Appl. Math, 1971].

1. Introduction. In this paper we shall discuss the asymptotic behaviour, as time tends to infinity, of solutions of a class of mixed initial boundary value problems for the porous media equation in one space dimension.

Let S_T denote the half-strip $(0, \infty)\times(0, T]$ for some fixed number $T>0$. We consider the following mixed initial boundary value problem:

$$u_t=(u^m)_{xx} \quad \text{in } S_T, \tag{1}$$

$$u(x, 0)=u_0(x) \quad \text{for } 0\leqq x<\infty, \tag{2}$$

$$u(0, t)=\psi(t) \quad \text{for } 0\leqq t\leqq T. \tag{3}$$

Here, the subscripts x and t denote partial differentiation with respect to the space and time variable, respectively, and m is a constant greater than 1. The functions u_0 and ψ are assumed to be bounded, nonnegative and continuous on their respective intervals of definition, and it is assumed that they satisfy the first compatibility condition: $u_0(0)=\psi(0)$.

Equation (1) arises in the theory of flows through porous media. Let u denote the density of a polytropic gas flowing through a homogeneous porous media. Then if the flow is in one dimension, u satisfies (1) [4].

It is well known that problem (1)–(3) need not have a classical solution [5]. However a class of weak solutions has been defined by Oleinik, Kalashnikov and Yui-Lin [5]. In their definition a function $u(x, t)$ defined on $\bar{S}_T$ is said to be a weak solution of problem (1)–(3) if: (a) u is bounded, nonnegative and continuous in $\bar{S}_T$; (b) $u(0, t)=\psi(t)$ for all $t\in[0, T]$; (c) (u^m) has a generalized derivative with respect to x in S_T, which is bounded in every set of the form $(\delta, \infty)\times(0, T]$, $\delta>0$, and square-integrable in every bounded subset of S_T; and (d) u satisfies the identity

$$\int\int_{S_T}\{\phi_x(u^m)_x-\phi_t u\}\,dx\,dt=\int_0^\infty \phi(x, 0)u_0(x)\,dx \tag{4}$$

for all $\phi\in C^1(\bar{S}_T)$ which vanish for $x=0$, large x and $t=T$. Under the assumption that u_0^m and ψ^m are both uniformly Lipschitz continuous, Oleinik, Kalashnikov and Yui-Lin have proved that problem (1)–(3) has a unique weak solution.

If a function u is a weak solution of problem (1)–(3) in any half-strip S_T, i.e. T may be an arbitrary positive number, then we say that u is a weak solution of problem (1)–(3) in the domain $S=(0, \infty)\times(0, \infty)$.

It will be useful later if we state here some properties of the weak solution $u(x, t)$. The weak solution can be constructed as the pointwise limit of a decreasing sequence of positive classical solutions of (1) $\{u_n(x, t)\}_{n=1}^\infty$, in an expanding sequence of cylinders

* Received by the editors September 18, 1975, and in revised form March 11, 1977.

† University of Sussex, United Kingdom, and Technische Hogeschool Delft, Netherlands; presently at Hydrological Research Institute, Department of Water Affairs, Pretoria, Republic of South Africa.

$Q_T^n=(0,n)\times(0,T]$, $n=1, 2, 3, \cdots$, [5]. By applying the theory of uniformly parabolic equations to this sequence it is then possible to deduce some properties of the weak solution. For instance, it can be shown that u is a classical solution of (1) in a neighbourhood of any point where it is positive. Also, one can show that if u_1 and u_2 are two weak solutions of problem (1)–(3) with respective data $u_{01}(x)$, $\psi_1(t)$ and $u_{02}(x)$, $\psi_2(t)$ satisfying

$$u_{01}(x)\geqq u_{02}(x) \quad \text{for all } x\in[0,\infty),$$

and

$$\psi_1(t)\geqq\psi_2(t) \quad \text{for all } t\in[0,T],$$

then $u_1\geqq u_2$ everywhere in S_T. Consequently, by comparison with a family of explicit weak solutions it is possible to prove that if $u_0(x)$ has compact support in $[0,\infty)$ then $u(x,t)$ has compact support in $\bar{S}_T$. Following on from the work of Oleinik, Kalashnikov and Yui-Lin it has been shown that the generalized derivative $(u^m)_x$ is actually continuous in S_T, and that the function $u^{m-1}(x,t)$ is Lipschitz continuous with respect to x for any $t>0$ [1], [3].

Throughout the paper, we shall make the following assumptions about u_0 and ψ.

A1. *u_0 is a nonnegative function defined on $[0,\infty)$ with compact support, such that u_0^{m-1} is Lipschitz continuous on $[0,\infty)$.*

A2. *ψ is a nonnegative function defined on $[0,\infty)$, such that ψ^m is Lipschitz continuous on $[0,\infty)$ and $\psi(0)=u_0(0)$.*

Assumption A1 is stronger than we need for existence of weak solutions of problem (1)–(3). However, in view of the regularity properties of u, and the fact that we are only interested in the behaviour of u as $t\to\infty$, it involves no loss of generality. Assumption A2 is precisely the one needed for existence of weak solutions in S.

We consider now a class of similarity solutions of equation (1) in S. Suppose that there exist solutions of equation (1) in S of the following forms:

$$\text{(5a)}\quad \text{I.} \qquad u_1(x,t)=(t+1)^\alpha f_1(\eta), \qquad \eta=x(t+1)^{-\frac{1}{2}\{1+(m-1)\alpha\}};$$

$$\text{(5b)}\quad \text{II.} \qquad u_2(x,t)=\exp(\alpha t)f_2(\eta), \qquad \eta=x\exp\{-\tfrac{1}{2}\alpha(m-1)t\}.$$

Then, formally, f_1 and f_2 should satisfy the equations:

$$\text{(6a)}\quad \text{I.} \qquad (f_1^m)''+\tfrac{1}{2}\{1+(m-1)\alpha\}\eta f_1'=\alpha f_1, \qquad 0<\eta<\infty;$$

$$\text{(6b)}\quad \text{II.} \qquad (f_2^m)''+\tfrac{1}{2}\alpha(m-1)\eta f_2'=\alpha f_2, \qquad 0<\eta<\infty;$$

respectively.

Similarity solutions of these types have been investigated in a recent paper [2]. Instead of considering equations (6a, b) the authors studied the more general equation

$$\text{(7)} \qquad (f^m)''+p\eta f'=qf, \qquad 0<\eta<\infty,$$

where p and q are arbitrary real constants. They defined a weak solution of equation (7) by analogy to the definition of a weak solution of problem (1)–(3) and searched for weak solutions satisfying

$$\text{(8)} \qquad f(0)=U>0, \qquad f(\infty)=0.$$

It was shown that problem (7), (8) has a weak solution with compact support if and only if $p\geqq0$ and $2p+q>0$. Moreover, problem (7), (8) can have at most one weak solution with compact support.

For the moment we restrict our attention to equation (7) when $p \geqq 0$ and $2p+q > 0$. Denote by $f(\eta; U)$ the weak solution of problem (7), (8) with compact support. Then there exists a point $a(U) \in (0, \infty)$ such that $f(\eta; U)$ is a positive classical solution of (7) in $[0, a(U))$ and $f(\eta; U) \equiv 0$ in $[a(U), \infty)$. Moreover, if $U_1 > U_2$ then $f(\eta; U_1) \geqq f(\eta; U_2)$ for all $\eta \in [0, \infty)$. We note furthermore; that as $U \to \infty$, $a(U) \to \infty$ and $f(\eta; U) \to \infty$ uniformly on compact subsets of $[0, \infty)$; whereas as $U \to 0$, $a(U) \to 0$ and $f(\eta; U) \to 0$ uniformly on $[0, \infty)$ [2].

If we return to the discussion of equations (6a, b) we see that problem (6a), (8) has a weak solution with compact support if and only if $\alpha > -1/m$, whereas problem (6b), (8) has a weak solution with compact support if and only if $\alpha > 0$. In both cases there is at most one such weak solution.

It is easy to verify that under the transformations (5a, b) the weak solutions of problems (6a, b), (8) with compact support become weak solutions of equations (1) in S. Thus, for any $\alpha > -1/m$ and $U > 0$ there is a unique (weak) similarity solution of (1) in S of type (5a) which vanishes for large x and satisfies

$$u_1(0, t) = U(t+1)^\alpha \quad \text{for } t \geqq 0.$$

Whereas, for any $\alpha > 0$ and $U > 0$ there is a unique (weak) similarity solution of (1) in S of type (5b) which vanishes for large x and satisfies

$$u_2(0, t) = U \exp(\alpha t) \quad \text{for } t \geqq 0.$$

The object of this paper is to discuss the asymptotic behaviour as $t \to \infty$, of the solution u of problem (1)–(3) in S, if either

$$\psi(t) \sim U(t+1)^\alpha, \qquad U > 0, \quad \alpha > -1/m, \quad \text{as } t \to \infty,$$

or

$$\psi(t) \sim U \exp(\alpha t), \qquad U > 0, \quad \alpha > 0, \quad \text{as } t \to \infty.$$

Specifically, we shall show that u converges towards the appropriate similarity solution.

The main result of this paper is contained in the following two theorems.

THEOREM 1. *Let u_1 and u_2 be two weak solutions of problem* (1)–(3) *in S with corresponding sets of boundary data u_{01}, ψ_1 and u_{02}, ψ_2 which satisfy the assumptions* A1 *and* A2. *Suppose that for some $\alpha > -1/m$ there exist positive constants A and B such that*

$$A(t+1)^\alpha \leqq \psi_1(t), \ \psi_2(t) \leqq B(t+1)^\alpha \quad \textit{for } t \geqq 0.$$

Then, for all $(x, t) \in S$

$$|u_1(x,t) - u_2(x,t)| \leqq C(t+1)^\alpha \left\{ (t+1)^{-(m\alpha+1)} \left(1 + \int_0^t |\psi_1^m(s) - \psi_2^m(s)|\, ds \right) \right\}^\lambda,$$

where

$$\lambda = \min\{\tfrac{1}{3}, 1/(2m-1)\}$$

and C is a constant which only depends on m, α, A, B, u_{01} and u_{02}.

THEOREM 2. *Let u_1 and u_2 be two weak solutions of problem* (1)–(3) *in S with corresponding sets of boundary data u_{01}, ψ_1 and u_{02}, ψ_2 which satisfy the assumptions* A1 *and* A2. *Suppose that for some $\alpha > 0$ there exist positive constants A and B such that*

$$A \exp(\alpha t) \leqq \psi_1(t), \ \psi_2(t) \leqq B \exp(\alpha t) \quad \textit{for } t \geqq 0.$$

Then, for all $(x, t) \in S$

$$|u_1(x,t)-u_2(x,t)| \leqq C \exp(\alpha t)\left\{\exp(-m\alpha t)\left(1+\int_0^t |\psi_1^m(s)-\psi_2^m(s)|\,ds\right)\right\}^{\lambda},$$

where

$$\lambda = \min\{\tfrac{1}{3}, 1/(2m-1)\}$$

and C is a constant which only depends on m, α, A, B, u_{01} *and* u_{02}.

Thus, suppose that $u(x, t)$ is a solution of problem (1)–(3) with data satisfying assumptions A1 and A2, such that for some $\alpha > -1/m$ and some $U > 0$:

$$\psi(t) > 0 \quad \text{for } t \geqq 0, \qquad \psi(t) \sim U(t+1)^{\alpha} \quad \text{as } t \to \infty.$$

Then, since

$$\int_0^t |\psi^m(s) - U^m(s+1)^{m\alpha}|\,ds = o(t^{m\alpha+1}) \quad \text{as } t \to \infty,$$

subject to the normalizing factor $(t+1)^{\alpha}$, $u(x, t)$ converges to the similarity solution of equation (1) with lateral boundary data $U(t+1)^{\alpha}$ as $t \to \infty$ uniformly with respect to x.

Similarly, suppose that $u(x, t)$ is a weak solution of problem (1)–(3) with data satifying assumptions A1 and A2, such that for some $\alpha > 0$ and some $U > 0$:

$$\psi(t) > 0 \quad \text{for } t \geqq 0, \qquad \psi(t) \sim U \exp(\alpha t) \quad \text{as } t \to \infty.$$

Then, since

$$\int_0^t |\psi^m(s) - U^m \exp(m\alpha s)|\,ds = o(\exp(m\alpha s)) \quad \text{as } t \to \infty,$$

subject to the normalizing factor $\exp(\alpha t)$, $u(x, t)$ converges to the similarity solution of equation (1) with lateral boundary data $U \exp(\alpha t)$ as $t \to \infty$ uniformly with respect to x.

The theorems extend a result of Peletier [7]. He considered a weak solution of problem (1)–(3) whose initial data satisfied A1 and $\psi(t) \equiv U > 0$. He showed that this solution converged to the similarity solution with lateral boundary data identically U. His work may be considered a special case of Theorem 1 with $\alpha = 0$ and $\psi_1 \equiv \psi_2$, and produces the same rate of convergence. The method which we use to prove our theorems is based largely on Peletier's ideas.

We shall prove the theorems in three stages. In § 2 we shall prove an integral identity for weak solutions of problem (1)–(3) and then introduce two lemmas which allow us to convert this integral identity into a pointwise estimate. However, we shall first require some results on the Hölder continuity of weak solutions of problem (1)–(3). In § 3 we shall therefore derive an estimate of the Hölder continuity of a positive classical solution of (1). As a weak solution of problem (1)–(3) can be constructed as the limit of a sequence of positive classical solutions of (1), this allows us to derive our required estimate for weak solutions of problem (1)–(3) in § 4, and thus to prove the theorems.

2. An integral identity. The proof of Theorems 1 and 2 rests on the following basic identity.

LEMMA 1. *Let u be a weak solution of problem* (1)–(3) *in S_T for some $T>0$. Then for any $t_0 \in (0, T]$*:

$$(9) \qquad \int_0^\infty xu(x, t_0)\,dx = \int_0^\infty xu_0(x)\,dx + \int_0^{t_0} \psi^m(t)\,dt.$$

Proof. Because u_0 has compact support, there exists a number $\rho>0$ such that for all $t\in[0, T]$, $u(x, t)=0$ if $x \geqq \rho$. Since $(u^m)_x \in C(S_T)$, this implies that $(u^m)_x=0$ on $[\rho, \infty)\times(0, T]$. It follows that (4) continues to hold for all $\phi \in C^1(\bar{S}_T)$ which only vanish for $x=0$ and $t=T$.

We first assume that $t_0<T$. Let

$$k(s)=\begin{cases}\exp\{-1/(1-s^2)\} & \text{for } |s|<1,\\ 0 & \text{for } |s|\geqq 1.\end{cases}$$

Then we can define, for any $n\geqq 1$, a function J_n by

$$J_n(t)=\left(\int_{n(t-t_0)}^{n(T-t_0)} k(s)\,ds\right)\left(\int_{-\infty}^{\infty} k(s)\,ds\right)^{-1}.$$

In view of the preceding remarks it is easy to verify that for u, $\phi(x, t)=xJ_n(t)$ is an admissible test function in (4). Substitution of ϕ yields for any $n\geqq 1$, remembering the continuity of $(u^m)_x$:

$$-\int\int_{S_T} xJ'_n(t)u(x, t)\,dx\,dt = J_n(0)\int_0^\infty xu_0(x)\,dx + \int_0^T J_n(t)\psi^m(t)\,dt.$$

If we now pass to the limit, and use the dominated convergence theorem, we obtain (9) for $t_0<T$. For $t_0=T$ (9) now follows by continuity.

We use this identity to prove the following inequality.

LEMMA 2. *Let u_1 and u_2 be weak solutions of problem* (1)–(3) *in S, with corresponding data u_{01}, ψ_1 and u_{02}, ψ_2. Then for any $t_0\in(0, \infty)$*:

$$(10) \quad \int_0^\infty x|u_1(x, t_0)-u_2(x, t_0)|\,dx \leqq \int_0^\infty x|u_{01}(x)-u_{02}(x)|\,dx + \int_0^{t_0} |\psi_1^m(t)-\psi_2^m(t)|\,dt.$$

Proof. Set $u_0^+(x)=\max\{u_{01}(x), u_{02}(x)\}$ and $\psi^+(t)=\max\{\psi_1(t), \psi_2(t)\}$, similarly set $u_0^-(x)=\min\{u_{01}(x), u_{02}(x)\}$ and $\psi^-(t)=\min\{\psi_1(t), \psi_2(t)\}$. Then as the functions u_{01}, ψ_1 and u_{02}, ψ_2 satisfy assumptions A1 and A2 so also do u_0^+, ψ^+ and u_0^-, ψ^-. Thus equation (1) has a weak solution $u^+(x, t)$ in S satisfying

$$u^+(x, 0)=u_0^+(x) \quad \text{for } 0\leqq x<\infty,$$
$$u^+(0, t)=\psi^+(t) \quad \text{for } 0\leqq t<\infty,$$

and a weak solution $u^-(x, t)$ in S satisfying

$$u^-(x, 0)=u_0^-(x) \quad \text{for } 0\leqq x<\infty,$$
$$u^-(0, t)=\psi^-(t) \quad \text{for } 0\leqq t<\infty.$$

Moreover, by Lemma 1, (9) holds for both u^+ and u^- for any $t_0 \in (0, \infty)$. Hence

$$\begin{aligned} \int_0^\infty x\{u^+(x, t_0) - u^-(x, t_0)\}\, dx &= \int_0^\infty x\{u_0^+(x) - u_0^-(x)\}\, dx \\ &\quad + \int_0^{t_0} \{(\psi^+)^m(t) - (\psi^-)^m(t)\}\, dt \\ &= \int_0^\infty x|u_{01}(x) - u_{02}(x)|\, dx + \int_0^{t_0} |\psi_1^m(t) - \psi_2^m(t)|\, dt \end{aligned} \tag{11}$$

for any $t_0 \in (0, \infty)$. Now, there is a maximum principle for weak solutions of problem (1)–(3) which implies that

$$u^-(x, t) \leqq u_1(x, t),\ u_2(x, t) \leqq u^+(x, t)$$

for all $(x, t) \in S$. Hence

$$|u_1(x, t) - u_2(x, t)| \leqq u^+(x, t) - u^-(x, t) \tag{12}$$

for all $(x, t) \in S$. Combining (11) and (12) we have proved (10).

Remark. If $u_{01}(x) \geqq u_{02}(x)$ *for all* $x \in [0, \infty)$ *and* $\psi_1(t) \geqq \psi_2(t)$ *for all* $t \in [0, t_0]$ *inequality* (10) *is in fact an equality. Hence* (10) *is sharp.*

To prove Theorems 1 and 2 we need to convert the integral estimate obtained in Lemma 2 to the pointwise estimate. To do this we use a modification of a result due to Peletier [7]. The alteration is only slight and presents no extra complexity of proof. We shall therefore omit the proof. The modified result is the following.

LEMMA 3. *Let* $\theta(x)$ *be a nonnegative function defined on* $[0, \infty)$ *satisfying the conditions;*

(i) θ *is uniformly Hölder continuous on* $[0, \infty)$ *with exponent* $\gamma \in (0, 1]$ *and coefficient* K,
(ii) $\int_0^\infty x\theta(x)\, dx \leqq L < \infty$,
(iii) $\theta(x_0) = 0$ *for some* $x_0 \in [0, \infty)$.

Then for all $x \in [0, \infty)$

$$\theta(x) \leqq C_0 K^{2/(\gamma+2)} L^{\gamma/(\gamma+2)},$$

where

$$C_0 = \{2(\gamma+2)/\gamma\}^{\gamma/(\gamma+2)}.$$

For two weak solutions u_1 and u_2 of problem (1)–(3) in S, we wish to apply Lemma 3 to $|u_1(x, t) - u_2(x, t)|$ for fixed t. Lemma 2 implies that (ii) holds, whereas because u_1 and u_2 have compact support condition (iii) is satisfied. We therefore only have to show that $|u_1(x, t) - u_2(x, t)|$ is uniformly Hölder continuous with respect to x on $[0, \infty)$.

3. A Hölder continuity estimate. In this section we shall deal only with positive classical solutions of equation (1). In the next section we shall apply the results we obtain to the sequence of positive classical solutions of equation (1) from which a weak solution of problem (1)–(3) may be constructed.

An estimate of Hölder continuity for positive classical solutions of (1) was obtained by Aronson [1]. However, this estimate is derived in terms of the supremum of a solution in its domain of definition. Here, we wish ultimately to apply our result to weak solutions of equation (1) in S which have a definite asymptotic behaviour on the lateral boundary of S. We look for a method by which we may capitalize on the asymptotic behaviour on the lateral boundary. It turns out that by considering the similarity transformations of equation (1) we can derive the estimate that we require.

Denote by R the rectangle $(0, H)\times(0, T]$ for some $H>0$, $T>0$, and by R^* the rectangle $(0, \frac{1}{2}H)\times(0, T]$.

LEMMA 4. *Assume $p\geqq 0$ and $2p+q>0$. Let $v\in C^{2,1}(\bar{R})$ be a positive solution of the equation*

$$v_\tau=(v^m)_{\eta\eta}+p\eta v_\eta-qv \tag{13}$$

in $\bar{R}$, such that $v_\eta\in C^{2,1}(R)$. Suppose also that there are positive constants A, M and C_1 such that

$$v(\eta,\tau)\leqq M \quad \text{for all } (\eta,\tau)\in\bar{R}, \qquad v(0,\tau)\geqq A \quad \text{for all } \tau\in[0,T], \tag{14}$$

and

$$|(v^{m-1})_\eta(\eta,0)|\leqq C_1 \quad \text{for all } \eta\in[0,H]. \tag{15}$$

Then

$$|v(\eta_1,\tau)-v(\eta_2,\tau)|\leqq K|\eta_1-\eta_2|^\nu, \tag{16}$$

where

$$\nu=\min\{1, 1/(\mathrm{m}-1)\},$$

for all (η_1,τ), $(\eta_2,\tau)\in\overline{R^*}$. Here K is a positive constant which only depends on m, p, q, A, M, C_1 and H.

Proof. Conditions (14) and (15) imply that $v(\eta,0)\geqq\{A^{m-1}-C_1\eta\}^{1/(m-1)}$ for all $\eta\in[0,H]$ where $\eta\leqq A^{m-1}/C_1$. We choose a $U\in(0,A]$ such that problem (7), (8) has a weak solution with compact support, $f(\eta;U)$, with the property that $f(\eta;U)\leqq\{A^{m-1}-C^1\eta\}^{1/(m-1)}$ on $[0, A^{m-1}/C_1]$ and $a(U)\leqq H$. Now, $f(\eta;U)$ is easily verified to be a steady-state solution of equation (13) in $(0,a(U))\times(0,T]$. Moreover

$$v(0,\tau)\geqq A\geqq U=f(0;U) \quad \text{for } \tau\in[0,T],$$

$$v(a(U),\tau)>0=f(a(U);U) \quad \text{for } \tau\in[0,T],$$

and, by the above

$$v(\eta,0)>\{A^{m-1}-C_1\eta\}^{1/(m-1)}\geqq f(\eta;U) \quad \text{for all } \eta\in[0,a(U)].$$

By applying the maximum principle to equation (13) in $(0,a(U))\times(0,T]$ we therefore deduce that $v(\eta,\tau)\geqq f(\eta;U)$ everywhere in this domain. It follows that equation (13) is uniformly parabolic in $[0,\delta]\times[0,T]$, where $\delta=\frac{1}{2}a(U)$. By a standard barrier function argument [6] we can subsequently estimate $|v_\eta|$ in $[0,\delta]\times[0,T]$ in terms of m, p, q, A, M and C_1. Plainly this implies that (16) holds in $[0,\delta]\times[0,T]$.

To prove the lemma we now only have to show that (16) holds in $[\delta,\frac{1}{2}H]\times[0,T]$, that is assuming that $\delta<\frac{1}{2}H$. Observe that if $p=q=0$ then equation (13) reduces to the porous media equation (1), and a result of the kind we are seeking has already been obtained for this equation by Aronson [1]. In fact, his work extends to equation (13) with little extra difficulty. As the argument is almost identical we shall not present it here.

Now, suppose that $v(\eta,\tau)$ is a positive solution of (13) in R satisfying the conditions of Lemma 4 in the particular case where

$$p=\tfrac{1}{2}\{1+(m-1)\alpha\} \quad \text{and} \quad q=\alpha,$$

for some $\alpha>-1/m$. Then setting

$$\eta=x(t+1)^{-\frac{1}{2}\{1+(m-1)\alpha\}}, \qquad \tau=\log(t+1), \tag{17}$$

we find that

$$u=(t+1)^{\alpha}v \tag{18}$$

is a positive solution of (1) in the set $D=\{(x,t)\colon 0<x<H(t+1)^{(\frac{1}{2})\{1+(m-1)\alpha\}},\ 0<t\leqq \exp(T)-1\}$, which satisfies

$$u(x,t)\leqq M(t+1)^{\alpha} \quad \text{for all } (x,t)\in\bar{D}, \tag{19}$$

$$u(0,t)\geqq A(t+1)^{\alpha} \quad \text{for all } t\in[0,\exp(T)-1], \tag{20}$$

and

$$u(x,0)=v(x,0) \quad \text{for all } x\in[0,H].$$

Conversely, if $u(x,t)\in C^{2,1}(\bar{D})$ is a positive solution of (1) in $\bar{D}$ such that $u_x\in C^{2,1}(D)$, and such that (19) and (20) hold for some positive A and M, where also $|(u^{m-1})_x(x,0)|\leqq C_1$ for all $x\in[0,H]$, then using the transformation (17), (18) we can construct a positive solution of (13) in R satisfying the conditions of Lemma 4 in the particular case where $p=\frac{1}{2}\{1+(m-1)\alpha\}$ and $q=\alpha$.

Similarly, if we let

$$p=\tfrac{1}{2}\alpha(m-1) \quad \text{and } q=\alpha,$$

for some $\alpha>0$, by means of the transformation

$$\eta=x\exp\{-\tfrac{1}{2}\alpha(m-1)t\}, \qquad \tau=t, \qquad u=\exp(\alpha t)v,$$

we can find a correspondence between solutions of (1) and (13).

This allows us to reformulate Lemma 4 in terms of solutions of equation (1). The outcome is the Hölder continuity estimate which we require.

LEMMA 5. *Assume $\alpha>-1/m$, and denote by D the set $\{(x,t)\colon 0<x<X(t+1)^{\{1+(m-1)\alpha\}/2},\ 0<t\leqq T\}$ and by D^* the set $\{(x,t)\colon 0<x<\frac{1}{2}X(t+1)^{\{1+(m-1)\alpha\}/2},\ 0<t\leqq T\}$. Let $u\in C^{2,1}(\bar{D})$ be a positive classical solution of (1) in $\bar{D}$ such that $u_x\in C^{2,1}(D)$. Assume also that there exist positive constants A, M and C_1 such that*

$$u(x,t)\leqq M(t+1)^{\alpha} \quad \textit{for all } (x,t)\in\bar{D},$$

$$u(0,t)\geqq A(t+1)^{\alpha} \quad \textit{for all } t\in[0,T],$$

and

$$|(u^{m-1})_x(x,0)|\leqq C_1 \quad \textit{for all } x\in[0,X].$$

Then

$$|u(x_1,t)-u(x_2,t)|\leqq K(t+1)^{\alpha-\frac{1}{2}\nu\{1+(m-1)\alpha\}}|x_1-x_2|^{\nu}$$

for all (x_1,t), $(x_2,t)\in\overline{D^*}$. Here K is a constant which only depends on m, α, A, M, C_1 and X.

LEMMA 6. *Assume $\alpha>0$, and denote by D the set $\{(x,t)\colon 0<x<X\exp\{\frac{1}{2}\alpha(m-1)t\},\ 0<t<T\}$ and by D^* the set $\{(x,t)\colon 0<x<\frac{1}{2}X\exp\{\frac{1}{2}\alpha(m-1)t\}$,*

$0<t\leqq T\}$. Let $u\in C^{2,1}(\bar{D})$ be a positive classical solution of (1) *in $\bar{D}$ such that $u_x\in C^{2,1}(D)$. Assume also that there are positive constants A, M and C_1 such that*

$$u(x,t)\leqq M\exp(\alpha t)\quad \textit{for all } (x,t)\in\bar{D},$$

$$u(0,t)\geqq A\exp(\alpha t)\quad \textit{for all } t\in[0,T],$$

and

$$|(u^{m-1})_x(x,0)|\leqq C_1\quad \textit{for all } x\in[0,X].$$

Then

$$|u(x_1,t)-u(x_2,t)|\leqq K\exp\{\tfrac{1}{2}\alpha[2-\nu(m-1)]t\}|x_1-x_2|^{\nu}$$

for all (x_1,t), $(x_2,t)\in\overline{D^*}$. Here K is a constant which only depends on m, α, A, M, C_1 and X.

4. The main result. Consider now a weak solution $u(x,t)$ of problem (1)–(3) in S, whose data satisfies assumptions A1 and A2. Assume also that

$$A(t+1)^{\alpha}\leqq\psi(t)\leqq B(t+1)^{\alpha}\quad \text{for all } t\geqq 0, \tag{21}$$

for some $\alpha>-1/m$ and some $B\geqq A>0$. Then it is possible to find a $U\geqq B$ such that the weak solution with compact support of problem (6a), (8), $f(\eta;U)$, is greater than or equal to $u_0(\eta)$ on $[0,\infty)$. We denote by $u(x,t;U)$ the similarity solution of equation (1) in S, which is derived from $f(\eta;U)$ by using transformation (5a).

Now,

$$u(0,t;U)=U(t+1)^{\alpha}\geqq B(t+1)^{\alpha}\geqq\psi(t)\quad \text{for all } t\geqq 0,$$

and

$$u(x,0;U)=f(x;U)\geqq u_0(x)\quad \text{for all } x\geqq 0.$$

Thus, by the maximum principle for weak solutions of problem (1)–(3) it follows that $u(x,t;U)\geqq u(x,t)$ everywhere in S. Specifically this means that there exists a positive constant M_0 such that

$$u(x,t)\leqq M_0(t+1)^{\alpha}\quad \text{for all } (x,t)\in S,$$

and that u is identically zero in $E^*=\{(x,t)\in S\colon a(U)(t+1)^{\{1+(m-1)\alpha\}/2}\leqq x\}$, in which, as in § 1, $a(U)$ is defined as $\sup\{\eta\colon f(\eta;U)>0\}$.

Let $X=2a(U)$, and let $D=\{(x,t)\colon 0<x<X(t+1)^{\{1+(m-1)\alpha\}/2},\ 0<t\leqq T\}$, for some $T>0$. Denote by $\{u_n(x,t)\}_{n=1}^{\infty}$ the decreasing sequence of positive classical solutions of (1) in the cylinders Q_T^n from which u may be constructed as a pointwise limit. We can assume that (i) $u_n\in C^{2,1}(\bar{Q}_T^n)$, (ii) $(u_n)_x\in C^{2,1}(Q_T^n)$ and (iii) $|(u_n^{m-1})_x(x,0)|\leqq C_1$ for all $x\in(0,n)$, where C_1 is a constant which is independent of n. Now by Dini's Theorem $u_n\to u$ as $n\to\infty$ uniformly on $\bar{D}$. Thus given any constant $M>M_0$, we can choose a number N so large that u_n is defined on $\bar{d}$ and, (iv)

$$u_n(x,t)\leqq M(t+1)^{\alpha}\quad \text{for all } (x,t)\in\bar{D},$$

for all $n\geqq N$.

The conditions on the decreasing sequence $\{u_n\}$ which we have indexed (i)–(iv) together with (21) provide sufficient evidence to show that for all $n\geqq N$, u_n satisfies the requirements of Lemma 5. Thus applying this lemma, for all $n\geqq N$,

$$|u_n(x_1,t)-u_n(x_2,t)|\leqq K(t+1)^{\alpha-\frac{1}{2}\nu\{1+(m-1)\alpha\}}|x_1-x_2|^{\nu}$$

for all (x_1, t), $(x_2, t) \in \overline{S_T \backslash E^*}$. Here K depends on m, α, A, M, C_1 and X, i.e. on m, α, A, B and u_0, but not on n. Because $u_n \to u$ as $n \to \infty$ pointwise in S_T, because $u \equiv 0$ in E^*, and because $T > 0$ was arbitrary, we have therefore proved the first of the following two lemmas. The proof of the second is similar.

LEMMA 7. *Let $u(x, t)$ be a weak solution of problem* (1)–(3) *in S, and suppose that for some $\alpha > -1/m$ there are positive constants A and B such that*

$$A(t+1)^\alpha \leqq \psi(t) \leqq B(t+1)^\alpha \quad \text{for } t \geqq 0.$$

Then

$$|u(x_1, t) - u(x_2, t)| \leqq K(t+1)^{\alpha - \frac{1}{2}\nu\{1+(m-1)\alpha\}} |x_1 - x_2|^\nu,$$

where

$$\nu = \min\{1, 1/(m-1)\},$$

for all (x_1, t), $(x_2, t) \in \bar{S}$. Here K is a constant which only depends on m, α, A, B and u_0.

LEMMA 8. *Let $u(x, t)$ be a weak solution of problem* (1)–(3) *in S, and suppose that for some $\alpha > 0$ there are positive constants A and B such that*

$$A \exp(\alpha t) \leqq \psi(t) \leqq B \exp(\alpha t) \quad \text{for } t \geqq 0.$$

Then

$$|u(x_1, t) - u(x_2, t)| \leqq K \exp\{\tfrac{1}{2}\alpha[2 - \nu(m-1)]t\} |x_1 - x_2|^\nu,$$

where

$$\nu = \min\{1, 1/(m-1)\},$$

for all (x_1, t), $(x_2, t) \in \bar{S}$. Here K is a constant which only depends on m, α, A, B and u_0.

We observe now that if $y_1(x)$ and $y_2(x)$ are two functions satisfying a Hölder condition with exponent γ and coefficients K_1 and K_2 respectively in some set Ω, then $|y_1 - y_2|(x)$ satisfies a Hölder condition with exponent γ and coefficient $(K_1 + K_2)$ in Ω. Using this observation and those of § 2, Theorems 1 and 2 become a simple consequence of Lemmas 2, 3, 7 and 8.

REFERENCES

[1] D. G. ARONSON, *Regularity properties of flows through porous media*, SIAM J. Appl. Math., 17 (1969), pp. 461–467.

[2] B. H. GILDING AND L. A. PELETIER, *On a class of similarity solutions of the porous media equation*, J. Math. Anal. Appl., 55 (1976), pp. 351–364.

[3] ———, *The Cauchy problem for an equation in the theory of infiltration*, Arch. Rational Mech. Anal., 61 (1976), pp. 127–140.

[4] M. MUSKAT, *The Flow of Homogeneous Fluids Through Porous Media*, McGraw-Hill, New York, 1937.

[5] O. A. OLEINIK, A. S. KALASHNIKOV AND CHZHOU YUI-LIN, *The Cauchy problem and boundary problems for equations of the type of unsteady filtration*, Izv. Akad. Nauk SSSR Ser. Mat., 22 (1958), pp. 667–704.

[6] O. E. OLEINIK AND S. N. KRUZHKOV, *Quasi-linear second-order parabolic equations with many independent variables*, Uspehi Mat. Nauk, 16 (1961), pp. 115–155 = Russian Math. Surveys, 16 (1961), pp. 105–146.

[7] L. A. PELETIER, *Asymptotic behavior of solutions of the porous media equation*, SIAM J. Appl. Math., 21 (1971), pp. 542–551.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0003 $01.00/0

ON THE BILATERAL PREDICTION ERROR MATRIX OF A MULTIVARIATE STATIONARY STOCHASTIC PROCESS*

A. G. MIAMEE† AND H. SALEHI‡

Abstract. A formula for the bilateral prediction error matrix of a multivariate stationary stochastic process in terms of the spectral density and its generalized inverse is given.

1. Introduction. An important problem in prediction theory of stationary stochastic processes is to obtain a formula for the bilateral prediction error matrix. This matrix represents the error arising when a stationary stochastic process (S.S.P.) is predicted on its combined past and future, and is related to the minimality problem of such processes. A. N. Kolmogorov in his celebrated paper [1] gave a complete solution to this problem for a univariate S.S.P. More precisely for a univariate S.S.P. X_n, $-\infty < n < \infty$, with spectral density F'_X he obtained the expression

$$\sigma_X^2 = \left[\frac{1}{2\pi}\int_0^{2\pi} (1/F'_X(\theta))\,d\theta\right]^{-1} \tag{1}$$

for its bilateral predictor error σ_X^2. In [4], P. Masani considered the minimality question for multivariate stationary stochastic processes, and therein gave a spectral characterization for minimal full rank processes. As a result he obtained the formula

$$\Sigma_X = \left[\frac{1}{2\pi}\int_0^{2\pi} (F'_X(\theta))^{-1}\,d\theta\right]^{-1} \tag{2}$$

expressing the bilateral predictor error matrix Σ_X of a minimal full rank S.S.P. X_n, $-\infty < n < \infty$, in terms of the inverse of its spectral density matrix F'_X, thereby extending Kolmogorov's result to the minimal full rank multivariate processes. The question of minimality was also the subject of a paper [6] of H. Salehi where he extended the above formulas to the multivariate processes which are not necessarily of full rank. He showed that

$$\Sigma_X = \left[\frac{1}{2\pi}\int_0^{2\pi} J\left(\frac{dF}{d\mu}\right)^{\#} J^*\,d\mu\right]^{\#} \tag{3}$$

where μ is a measure with respect to (w.r.t.) which F is absolutely continuous, J is the matrix representing the projection operator onto the range of Σ_X, and $A^{\#}$ stands for the generalized inverse of matrix A. However for practical purposes the appearance of this matrix J in the right hand side of (3) is not convenient because of its dependence on Σ_X itself. Therefore for computational purposes it is desirable to improve relation (3) and obtain an expression for Σ_X in terms of the spectral density alone. The aim of this paper is to establish such a formula. More precisely we will prove

$$\Sigma_X = \left[\frac{1}{2\pi}\int_0^{2\pi} (F'_X(\theta))^{\#}\,d\theta\right]^{\#} \tag{4}$$

for a minimal (not necessarily full rank) stationary stochastic process X_n, $-\infty < \infty$, under certain regularity conditions.

* Received by the editors May 27, 1977, and in revised form September 29, 1977.
† Department of Mathematics, Arya-Mehr University of Technology, Isfahan, Iran.
‡ Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824.
The work of this author was supported in part by the National Science Foundation under Grant MCS 77-02661.

The questions of minimality and interpolation for multivariate stationary processes were also studied by H. Salehi and J. K. Scheidt in [7], where the idea of Wold–Cramer concordance w.r.t. the past and future was introduced and some sufficient conditions for such a concordance were given. Recently these ideas have been pursued by A. Makagon and A. Weron in [2], where some important results such as a characterization for concordance in terms of the generalized inverse of the spectral density were obtained. To establish our results we will use their characterization of concordance w.r.t. J_0, the family of the complements of singletons in the set of integers.

2. Preliminaries. In this section we shall set down notations and preliminaries which will be needed in this paper. These notations and definitions are standard and can be found in [8].

Let $\mathscr{H}$ be a complex Hilbert space, $q \geqq 1$, and $\mathscr{H}^q$ be the Cartesian product of $\mathscr{H}$ with itself q times, i.e., the set of all column vectors $X = (X^1, X^2, \cdots, X^q)^T$ such that $X^i \in \mathscr{H}$. To make H^q serviceable in prediction theory, we endow it with a Gramian structure. For X and $Y \in \mathscr{H}^q$ the $q \times q$ matrix

$$(X, Y) = [(X^i, X^j)]_{i,j=1}^q$$

is called the Gramian of X and Y. It is easy to verify that

$$(X, X) \geqq 0; \qquad (X, X) = 0 \Leftrightarrow X = 0;$$

$$\left(\sum_{i=1}^m A_i X_i, \sum_{j=1}^n B_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n A_i (X_i, X_j) B_j^*,$$

for any $X, Y, X_i, Y_j \in \mathscr{H}^q$ and any $q \times q$ matrices A_i, B_j. We say that X is orthogonal to Y in $\mathscr{H}^q$ if $(X, Y) = 0$. It is well known that $\mathscr{H}^q$ is a Hilbert space under the inner product

$$((X, Y)) = \text{trace}\,(X, Y) = \sum_{j=1}^q (X^j, Y^j).$$

A closed subset $\bar{\mathscr{M}}$ of $\mathscr{H}^q$ is called a subspace if it is a manifold, i.e., $AX + BY \in \bar{\mathscr{M}}$ whenever $X, Y \in \bar{\mathscr{M}}$ and A, B are $q \times q$ matrices. It is easy to see that $\bar{\mathscr{M}}$ is a subspace of $\mathscr{H}^q$ if and only if $\bar{\mathscr{M}} = \mathscr{M}^q$ for some subspace $\mathscr{M}$ of $\mathscr{H}$. For any X in $\mathscr{H}^q$, $(X \mid \bar{\mathscr{M}})$ denotes the projection of X onto $\bar{\mathscr{M}}$. We note that the kth coordinate of $(X \mid \bar{\mathscr{M}})$ is $(X^k \mid \mathscr{M})$, where $\bar{\mathscr{M}} = \mathscr{M}^q$ and $(X^k \mid \mathscr{M})$ is the projection of X^k onto $\mathscr{M}$.

DEFINITION 2.1. A sequence X_n, $-\infty < n < \infty$, of elements of $\mathscr{H}^q$ is called a *q-variate S.S.P.* if the Gramian (X_m, X_n) depends only on $m - n$.

It is well known that given any q-variate S.S.P., say X_n, there exists a $q \times q$ nonegative matrix valued measure F, called its spectral measure on $[0, 2\pi]$ such that

$$(X_m, X_n) = \frac{1}{2\pi} \int_0^{2\pi} e^{-i(m-n)\theta}\, dF_X(\theta).$$

We denote the Radon–Nikodym derivative of F_X^a, the absolutely continuous component of F_X, w.r.t. Lebesgue measure $d\theta$ by F'_X or f_X and call it the spectral density of the S.S.P. X_n, $-\infty < n < \infty$.

DEFINITION 2.2. A stationary stochastic process X_n, $-\infty < n < \infty$, is called *minimal* if for each integer n, $X_n \notin \bar{\mathscr{M}}(n) =$ the subspace of $\mathscr{H}^q$ generated by X_k, $k \neq n$; or equivalently $X_n^i \notin \mathscr{M}(n) =$ the subspace of $\mathscr{H}$ generated by X_k^j, $k \neq n$, $1 \leqq j \leqq q$; for some i, $1 \leqq i \leqq q$. [Note that $\bar{\mathscr{M}}(n) = (\mathscr{M}(n))^q$.]

The bilateral innovation process $\hat{X}_n$ of the S.S.P. X_n is defined by $\hat{X}_n = X_n - (X_n | \bar{\mathcal{M}}(n))$, $-\infty < n < \infty$. The bilateral predictor error matrix of X_n is defined to be $\Sigma_X = (\hat{X}_0, \hat{X}_0)$. Clearly an S.S.P. X_n is minimal if and only if $\Sigma_X \neq 0$. An S.S.P. X_n is said to be minimal full rank if $\Sigma\ x$ is invertible.

The following interesting result is due to P. Masani [4]. It gives a characterization for a minimal full rank q-variate S.S.P.

THEOREM 2.3. *Let X_n, $-\infty < n < \infty$, be a q-variate S.S.P. with spectral measure F_X, and the bilateral predictor error matrix Σ_X. Then X_n is minimal full rank if and only if $(F'_X(\theta))^{-1}$ exists a.e. and is summable w.r.t. $d\theta$. In this case we have*

$$\Sigma_X = \left[\frac{1}{2\pi}\int_0^{2\pi} (F'_X(\theta))^{-1}\, d\theta\right]^{-1}. \tag{5}$$

DEFINITION 2.4. For any $q \times q$ matrix A there exists a unique $q \times q$ matrix $A^{\#}$ satisfying

$$AA^{\#}A = A, \quad A^{\#}AA^{\#} = A^{\#}, \quad (A^{\#}A)^* = A^{\#}A, \quad (AA^{\#})^* = AA^{\#}.$$

This matrix $A^{\#}$ is called the generalized inverse of A [5] and has the following further properties

$$\mathcal{N}^{\perp}(A) = \mathcal{R}(A^{\#}), \qquad \mathcal{R}^{\perp}(A) = \mathcal{N}(A^{\#}),$$

where for each matrix B, $\mathcal{R}(B)$ and $\mathcal{N}(B)$ stand for the range and null space of B, respectively.

DEFINITION 2.5. Let J_0 denote the family of all complements of singletons $\{n\}$ in the set of integers. An S.S.P. X_n is said to be *J_0-regular* if $\bigcap_n \bar{\mathcal{M}}(n) = 0$, and it is called *$J_0$-singular* if for some n, $\bar{\mathcal{M}}(n) = \mathcal{H}(\infty) =$ the subspace of $\mathcal{H}^q$ generated by X_n, $-\infty < n < \infty$.

It is well known [2], [7] that any q-variate S.S.P. X_n, w.r.t. past and future, admits a unique Wold–decomposition

$$X_n = Y_n + Z_n, \qquad -\infty < n < \infty,$$

where Y_n and Z_n are orthogonal processes such that $\bar{\mathcal{M}}_Y(n), \bar{\mathcal{M}}_Z(n) \subseteq \bar{\mathcal{M}}_X(n)$; Y_n is J_0-regular and Z_n is J_0-singular. This gives the decomposition $F_X = F_Y + F_Z$. There is also the usual Cramer decomposition $F_X = F^{\mathrm{a}} + F^{\mathrm{s}}$ of F_X as the sum of its absolutely continuous component F^{a} and its singular component F^{s}. We say that the Wold and Cramer decompositions are concordant whenever $F^{\mathrm{a}} = F_Y$ and $F^{\mathrm{s}} = F_Z$.

In the proof of our main results we use the following interesting theorems of [2]. For ease of reference and convenience of readers we state them below.

THEOREM 2.6. *Let X_n, $-\infty < n < \infty$, be a q-variate stationary stochastic process with spectral measure F_X. Then Wold and Cramer decomposition are concordant if and only if $\mathcal{R}(F'_X(\theta)) =$ constant a.e. $d\theta$ and $(F'_X(\theta))^{\#}$ is integrable w.r.t. $d\theta$.*

THEOREM 2.7. *Let X_n, $-\infty < n < \infty$, be a q-variate stationary stochastic process with spectral measure F_X. The process X_n is J_0-regular if and only if*

(a) *F_X is absolutely continuous w.r.t. $d\theta$,*

(b) *$\mathcal{R}(F'_X(\theta))$ is constant a.e. $d\theta$,*

(c) *$(F'_X(\theta))^{\#}$ is integrable w.r.t. $d\theta$.*

3. Main Results. In this section we give explicit formulas for evaluating the bilateral predictor error matrix Σ_X of a (not necessarily full rank) stationary stochastic process X_n, $-\infty<n<\infty$, in terms of its spectral measure F_X. These results extend Theorem 2.3 due to Masani to the nonfull rank case. These also improve Theorem 3 [6] of Salehi under some mild conditions. In the proof of our results we will use Theorems 2.3, 2.6 and 2.7 mentioned in § 2.

THEOREM 3.1. *Let X_n, $-\infty<n<\infty$, be a q-variate J_0-regular S.S.P. Then*

$$\Sigma_X=\left[\frac{1}{2\pi}\int_0^{2\pi}(F'_X(\theta))^{\#}\,d\theta\right]^{\#}. \tag{6}$$

Proof. By Theorem 2.7, F_X is absolutely continuous w.r.t. $d\theta$ and $\mathcal{R}(F'_X(\theta))=\mathcal{R}(f_X(\theta))$ is a.e. a constant subspace, say $\mathcal{R}$. Let $h_1,\cdots,h_p;h_{p+1},\cdots,h_q$ be an orthonormal basis for the q-dimensional complex Euclidean space C^q such that $\mathcal{R}=\mathfrak{S}\{h_1,h_2,\cdots,h_p\}$ and $\mathcal{N}=\mathcal{R}^\perp=\mathcal{N}(F'_X)=\mathfrak{S}\{h_{p+1},h_{p+2},\cdots,h_q\}$. Let $e_1,e_2,\cdots e_q$ be the standard basis of C^q. We define the unitary operator U on C^q by $Uh_i=e_i$, $1\leqq i\leqq q$. Letting $\mathcal{R}_1=\mathfrak{S}\{e_1,e_2,\cdots,e_p\}$ then $\mathcal{R}_1^\perp=\mathfrak{S}\{e_{p\geqq 1},e_{p+2},\cdots,e_q\}$. Clearly U maps $\mathcal{R}$ onto $\mathcal{R}_1$ and $\mathcal{R}^\perp$ onto $\mathcal{R}_1^\perp$ and U^* maps $\mathcal{R}_1$ onto $\mathcal{R}$ and $\mathcal{R}_1^\perp$ onto $\mathcal{R}^\perp$. As usual we will identify any linear operator onto C^q with its matrix w.r.t. the standard basis. By the choice of U we have

$$Uf_XU^*=\left[\begin{array}{c|c} g & 0\\ \hline 0 & 0\end{array}\right]$$

where g is a $p\times p$ nonnegative matrix valued function whose rank $=p$ a.e. Let $Y_n=UX_n$, $-\infty<n<\infty$. For the q-variate process Y_n, $-\infty<n<\infty$, we have

$$\begin{aligned}(Y_m,Y_n)&=(UX_m,UX_n)=U(X_n)U^*\\ &=U\left(\frac{1}{2\pi}\int_0^{2\pi}e^{-i(m-n)\theta}f_x(\theta)\,d\theta\right)U^*\\ &=\frac{1}{2\pi}\int_0^{2\pi}e^{-i(m-n)\theta}Uf_X(\theta)U^*\,d\theta\\ &=\frac{1}{2\pi}\int_0^{2\pi}e^{-i(m-n)\theta}\left[\begin{array}{c|c} g(\theta) & 0\\ \hline 0 & 0\end{array}\right]d\theta.\end{aligned}$$

This shows that the component Y_n^k of Y_n is zero whenever $1+p\leqq k\leqq q$ and $-\infty<n<\infty$. The p-variate S.S.P. $Z_n=(Y_n^1,\cdots,Y_n^p)^T$, $-\infty<n<\infty$, has spectral density $f_Z=g$. Noting that U takes $\mathcal{R}$ onto $\mathcal{R}_1$ and $\mathcal{R}^\perp$ onto $\mathcal{R}_1^\perp$ one can see that

$$\left[\begin{array}{c|c} g^{-1} & 0\\ \hline 0 & 0\end{array}\right]=\left[\begin{array}{c|c} g & 0\\ \hline 0 & 0\end{array}\right]^{\#}=(Uf_XU^*)^{\#}=Uf_X^{\#}U^*. \tag{7}$$

Now since X_n is J_0-regular by Theorem 2.7, $(f_X(\theta))^{\#}$ is integrable and (7) implies that $f_Z^{-1}=g^{-1}$ is integrable w.r.t. $d\theta$; and hence by theorem 2.3, Z_n,$-\infty<n<\infty$, is minimal full rank. Hence by (5) we have

$$\Sigma_Z=\left[\frac{1}{2\pi}\int_0^{2\pi}(f_Z(\theta))^{-1}\,d\theta\right]^{-1}.$$

Now

$$
\begin{aligned}
U\Sigma_X U^* &= U(\hat{X}_0, \hat{X}_0)U^* = (U\hat{X}_0, u\hat{X}_0) \\
&= (\widehat{UX}_0, \widehat{UX}_0) = (\hat{Y}_0, \hat{Y}_0) \\
&= \left[\begin{array}{c|c} (\hat{Z}_0, \hat{Z}_0) & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} \Sigma_Z & 0 \\ \hline 0 & 0 \end{array}\right] \\
&= \left[\begin{array}{c|c} ((1/(2\pi))\int_0^{2\pi}(f_Z(\theta))^{-1}\,d\theta)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} (1/(2\pi))\int_0^{2\pi}(f_Z(\theta))^{-1}\,d\theta & 0 \\ \hline 0 & 0 \end{array}\right]^{\#} \\
&= \left[\frac{1}{2\pi}\int_0^{2\pi}\left[\begin{array}{c|c} f_Z(\theta) & 0 \\ \hline 0 & 0 \end{array}\right]^{\#} d\theta\right]^{\#} \\
&= \left[\frac{1}{2\pi}\int_0^{2\pi}(Uf_X(\theta)U^*)^{\#}\,d\theta\right]^{\#} = U\left[\frac{1}{2\pi}\int_0^{2\pi}(f_X(\theta))^{\#}\,d\theta\right]^{\#} U^*;
\end{aligned}
$$

the third equality above follows from the fact that $\mathcal{M}_X(0) = \mathcal{M}_Y(0)$ and the rest of them can be easily verified. This gives the desired result.

Consider the Wold decomposition $X_n = Y_n + Z_n$, $-\infty < n < \infty$, of an S.S.P. X_n with respect to J_0. Then $\Sigma_X = \Sigma_Y$ and since the S.S.P. Y_n, $-\infty < n < \infty$, is regular applying Theorem 3.1 to the process Y_n we get

$$\Sigma_X = \Sigma_Y = \left[\frac{1}{2\pi}\int_0^{2\pi}(f_Y(\theta))^{\#}\,d\theta\right]^{\#}.$$

This proves the following theorem.

THEOREM 3.2. *Let X_n, $-\infty < n < \infty$, be an S.S.P. Let Y_n be the J_0-regular component of X_n in its Wold decomposition having spectral density f_Y. Then*

$$\Sigma_X = \left[\frac{1}{2\pi}\int_0^{2\pi}(f_Y(\theta))^{\#}\,d\theta\right]^{\#}.$$

Theorem 3.2 expresses Σ_X in terms of the spectral density of the regular component of X_n. A more useful formula linking Σ_X to F_X, is given in the following theorem.

THEOREM 3.3. *Let X_n, $-\infty < n < \infty$, be a q-variate S.S.P. with spectral measure F_X. Suppose F'_X has constant range a.e. and $F'^{\#}_X$ is integrable w.r.t. $d\theta$. Then*

$$\Sigma_X = \left[\frac{1}{2\pi}\int_0^{2\pi}(F'_X(\theta))^{\#}\,d\theta\right]^{\#}.$$

Proof. By Theorem 2.6, Wold–Cramer concordance holds and hence $F'_X = F'_Y$. Now one can apply Theorem 3.2 to complete the proof.

Remark 3.4. We would like to mention that the following questions are still open: What are necessary and sufficient conditions in terms of the spectral measure F_X in order that the S.S.P. X_n be minimal? A second question which seems to shed some light on the first question is to find a formula for evaluating Σ_X in terms of F_X. As the following examples show one should not expect that formula (6) holds for the general case.

Example 1. Let X_n, $-\infty < n < \infty$ be a bivariate S.S.P. whose spectral measure F_X is a.c. and has density

$$f_X(\theta) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & e^{i\theta} \\ e^{-i\theta} & 1 \end{bmatrix}.$$

It is easy to see that $\Sigma_X = 0$, range of f is not constant a.e., $(f(\theta))^{\#} = f(\theta)$ is integrable and

$$\frac{1}{2\pi}\int_0^{2\pi} (f(\theta))^{\#}\, d\theta = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence $\Sigma_X \neq [(1/(2\pi)) \int_0^{2\pi} (f(\theta))^{\#}\, d\theta]^{\#}$.

Example 2. Take a bivariate S.S.P. X_n, $-\infty < n < \infty$, whose spectral measure F_X is a.c. with respect to Lebesgue measure and its density is given by

$$f_X(\theta) = \begin{bmatrix} 1 & \varphi(\theta) \\ \varphi(\theta) & 1 \end{bmatrix},$$

where

$$\varphi(\theta) = \begin{cases} \sqrt{1 - \frac{1}{2} e^{-1/\theta}}, & 0 < \theta \leqq 2\pi, \\ 1, & \theta = 0. \end{cases}$$

This density is studied in [3]. One can see that: Rank of $f_x(\theta) = 2$ a.e., $(f_X(\theta))^{\#}$ is not integrable, Σ_X has rank one. So X_n is a minimal process for which formula (6) cannot hold.

REFERENCES

[1] A. N. KOLMOGOROV, *Stationary sequences in Hilbert space*, Bolletin Moskovskogo Gosudarstvenogo Universiteta. Matematika, 2, no. 6 (1941), pp. 1–40. (In Russian.)

[2] A. MAKAGON AND A. WERON, *Wold–Cramer concordance theorem for interpolation of q-variate stationary processes over locally compact abelian groups*, J. Multivariate Anal., 6 (1976), pp. 123–137.

[3] P. MASANI, *Cramer's theorem on monotone matrix-valued functions and the Wold-decomposition*, Probability and Statistics, U. Grenander, ed., Almquist and Wiksell, Stockholm, 1959, pp. 175–189.

[4] ———, *The prediction theory of multivariate stochastic processes, III*, Acta. Math., 104 (1960), pp. 141–162.

[5] J. B. ROBERTSON AND M. ROSENBERG, *The decomposition of matrix-valued measures*, Michigan Math. J., 15 (1968), pp. 353–368.

[6] H. SALEHI, *Application of Hellinger intergrals to q-variate stationary stochastic processes*, Ark. Mat., 7 (1967), pp. 305–311.

[7] H. SALEHI AND J. K. SCHEIDT, *Interpolation of q-variate stationary stochastic processes over a locally compact abelian group*, J. Multivariate Anal., 2 (1972), pp. 307–331.

[8] N. WEINER AND P. MASANI, *The prediction theory of multivariate stochastic processes*, Acta Math., 98 (1957), pp. 111–150.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0004 $01.00/0

INVERTIBILITY OF LIPSCHITZ CONTINUOUS MAPPINGS AND ITS APPLICATION TO ELECTRICAL NETWORK EQUATIONS*

YOSHIAKI KAWAMURA†

Abstract. Conditions for unique solvability of nonlinear simultaneous equations satisfying Lipschitz conditions and an application to nonlinear network equations are proposed. It is shown that the global invertibility of a Lipschitz continuous mapping f of $\mathbb{R}^n$ into itself and the Lipschitz continuity of f^{-1} are verified by investigating the positivity of some principal minors of the Jacobian matrix of f in spite of the existence of nondifferentiable points of f. This is a generalization of previous works, especially Fujisawa and Kuh's theorem, obtained for continuously differentiable or piecewise-linear mappings. This generalization is given by employing the Lebesgue integration of the Jacobian matrix over an open interval of $\mathbb{R}^n$.

This result is applied to network equations, both resistive and dynamical; especially to the latter, the Lipschitz continuity of such inverse mappings is of great importance to guarantee the uniqueness of the solutions. In addition, network examples for which the above result is useful are given, and it is demonstrated that simple conditions for the unique solvability are obtained in terms of differential coefficients of network element characteristics.

1. Introduction. The equations describing nonlinear resistive networks or memoryless systems are expressed in the form $f_E(x, u) = 0$ or $f_E(x) = y$ when some unknown signals are denoted by $x = (x_1, \cdots, x_n)$ and known signals by $u = (u_1, \cdots, u_m)$ or $y = (y_1, \cdots, y_n)$. In addition, these equations arise in deriving state equations of dynamical systems as RLC networks. The problem, whether these equations are uniquely solvable for all input signals, depends on the global invertibility of mappings determined by f_E. The necessary and sufficient condition for a global diffeomorphism of $\mathbb{R}^n$ onto itself, which is a basic application of the well-known local inverse mapping theorem and is referred to as Palais' theorem by network theorists [15], has played an important role in this area of network analysis. However, the same proposition is not sufficient except for continuously differentiable mapping. The piecewise-linear mapping, which is continuous and is composed of a finite number of regions with linear characteristics, is a typical one of this case. In 1972, Fujisawa and Kuh showed the following useful theorem [8]: A piecewise-linear mapping is a global homeomorphism if the leading principal minors associated with the first k rows and the first k columns of matrices describing the mapping do not vanish and have the same sign through all regions for each k. It is known [9] that a similar result is valid for the continuously differentiable mapping and is a fairly general sufficient condition containing most of earlier works [14], [16], [18] related to positive definiteness and class P matrices.

In this paper, we prove that similar sufficient conditions are also valid for the mapping satisfying a local Lipschitz condition. This class of mappings is general enough to include both continuously differentiable and piecewise-linear ones. Furthermore, this class is of great importance due to the fact that state equations of dynamic networks are given by using inverse mappings of resistive equations and that the Lipschitz condition is the most standard restriction for unique solvability of differential equations. Haneda has recently studied the invertibility for this class by introducing a local functional to express the monotonicity of mappings [10]. We see that many of his results are very much simplified and generalized in terms of the Jacobian matrix.

* Received by the editors February 19, 1976, and in final revised form August 6, 1977.

† Department of Electronics, College of Engineering, University of Osaka Prefectur, Sakai, Osaka, Japan.

2. Properties of the Jacobian matrix. We study the signals in $\mathbb{R}^n$ with the Euclidean norm. However, the main results are commonly verified even if the l_1 or l_∞ norm is used because the Lipschitz condition, the Jacobian matrix and the homeomorphism are common conceptions for these norms.

We say that a mapping f is LC (Lipschitz continuous) if f satisfies a Lipschitz condition. It is essential that a mapping, although being LC, is not always differentiable at every point of its domain. We will show in the following, however, such inconvenient points are negligible. The foundation is given by the next property of Rademacher.

Rademacher's theorem [7]: Let f be a mapping of an open set $B \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and suppose that f is LC. Then f is differentiable (totally differentiable) at almost all points in B.

In the above theorem, "almost all points" are, by definition, such that the Lebesgue measure of the total exceptional points in $\mathbb{R}^n$ is equal to zero. For example, a piecewise-linear mapping is differentiable at all points except the boundary points of all linear regions and the total of the boundary points has measure zero.

Since a Jacobian matrix J of f is determined at every differentiable point, it follows that $J(x)$ is defined at almost all $x = (x_1, \cdots, x_n)$ in B. However, it should be noted that the above theorem does not hold for the points of a line segment for $n \geqq 2$ because the line segment is not open in $\mathbb{R}^n$ and has measure zero. In other words, it is possible that f is not differentiable at every point of a line segment. Therefore, to employ relations obtained from total differentiability, we introduce an integration of the differential coefficient over an open interval of $\mathbb{R}^n$ instead of a line segment to evaluate $f(x^a)-f(x^b)$, where x^a and x^b are points of the domain. Such problems are essentially reduced to the evaluation of $f_1(x_1^b, \bar{x}^a)-f_1(x_1^a, \bar{x}^a)$ by a coordinate transformation, where f_1 is a mapping of a subset of $\mathbb{R}^n$ into $\mathbb{R}^1$, $x_1^a \in \mathbb{R}^1$, $x_1^b \in \mathbb{R}^1$ and $\bar{x}^a = (x_2^a, x_3^a, \cdots, x_n^a) \in \mathbb{R}^{n-1}$.

LEMMA 1. *Let V_δ be the interval defined by $V_\delta = \{x \mid x_1^a < x_1 < x_1^b,\ |x_2 - x_2^a| < \delta/2, \cdots, |x_n - x_n^a| < \delta/2\}$, where $x_1^a \leqq x_1^b$ and $\delta > 0$. Suppose that f_1 is LC on V_δ when δ is sufficiently small. Then*

$$f_1(x_1^b, \bar{x}^a)-f_1(x_1^a, \bar{x}^a) = \lim_{\delta\to 0} \frac{1}{\delta^{n-1}} \int_{V_\delta} \frac{\partial f_1(x)}{\partial x_1}\, d\mu_n,$$

where μ_n represents Lebesgue measure in $\mathbb{R}^n$.

Proof. Let δ_0 be a positive number such that f_1 is LC on V_{δ_0}. By referring to the definition [17, Chap. 3–74, p. 217] we know that the Lipschitz continuous function in one variable is also absolutely continuous. Hence $f_1(x_1, \bar{x})$ is absolutely continuous in x_1 when $\bar{x} = (x_2, \cdots, x_n)$ is fixed. From a fundamental property of absolutely continuous functions [17, Chap. 3–74, Thm. 3, p. 219], we know that $f_1(x_1, \bar{x})$ is differentiable at almost all x_1 for fixed $\bar{x}$, and that

$$\int_{x_1^a}^{x_1^b} \frac{\partial f_1(x_1, \bar{x})}{\partial x_1}\, dx_1 = f_1(x_1^b, \bar{x})-f_1(x_1^a, \bar{x}). \tag{1}$$

Let us prove that $\partial f_1(x)/\partial x_1$ is integrable over V_δ. Let $\{h_\nu(x)\}$ be the sequence of continuous functions on V_{δ_0} defined by

$$h_\nu(x) = \begin{cases} \{f_1(x_1+\theta_\nu, \bar{x})-f_1(x_1, \bar{x})\}/\theta_\nu & \text{if } x_1+\theta_\nu < x_1^b, \\ \{f_1(x_1^b-0, \bar{x})-f_1(x_1, \bar{x})\}/\theta_\nu & \text{if } x_1+\theta_\nu \geqq x_1^b, \end{cases}$$

where $\{\theta_\nu\}$ is a sequence of positive numbers such that $\theta_\nu \to 0$. Then it follows by continuity that $\{h_\nu(x)\}$ is a sequence of measurable functions [12, Chap. 2–42, p. 115] and that $h_\nu(x)=\partial f_1(x)/\partial x_1$ at almost all x in V_{δ_0} if $\nu\to\infty$. Hence $\partial f_1(x)/\partial x_1$ is also measurable [12, Chap. 2–9, Thm. 1, p. 54]. Since a bounded measurable function is integrable [17, Chap. 3–11, Thm. 4, p. 65] and $\partial f_1(x)/\partial x_1$ is bounded by the Lipschitz coefficient, we can conclude that it is integrable over V_δ for $\delta<\delta_0$. Hence the integral I_δ is representable, by Fubini's theorem [12, Chap. 3–16, p. 87], as

$$I_\delta=\int_{V_\delta}\frac{\partial f_1(x)}{\partial x_1}\,d\mu_n=\int_{U_\delta}\left\{\int_{x_1^a}^{x_1^b}\frac{\partial f_1(x_1,\bar{x})}{\partial x_1}\,dx_1\right\}d\mu_{n-1},\tag{2}$$

where U_δ is the interval of $\mathbb{R}^{n-1}$ defined by $\{\bar{x}|(x_1,\bar{x})\in V_\delta\}$ for $\delta<\delta_0$ and $x_1^a<x_1<x_1^b$.

Suppose that $x\in V_\delta\subset V_{\delta_0}$. Then $\bar{x}\in U_\delta$. Since $\|\bar{x}-\bar{x}^a\|<\sqrt{n-1}\,\delta/2$ and f_1 is LC, we have

$$|\{f_1(x_1^b,\bar{x})-f_1(x_1^a,\bar{x})\}-\{f_1(x_1^b,\bar{x}^a)-f_1(x_1^a,\bar{x}^a)\}|<\sqrt{n-1}\,\eta_1\delta\tag{3}$$

where η_1 is the Lipschitz coefficient of f_1 on V_{δ_0}. Substituting (1) and (3) into (2), we get

$$|I_\delta-\delta^{n-1}\{f_1(x_1^b,\bar{x}^a)-f_1(x_1^a,\bar{x}^a)\}|<\sqrt{n-1}\,\eta_1\delta^n.\tag{4}$$

Dividing both sides of (4) by δ^{n-1} and letting $\delta\to0$, we obtain Lemma 1. Q.E.D.

The Jacobian matrix of the inverse mapping is determined as follows.

LEMMA 2. *Let f be a mapping of an open set $B\subset\mathbb{R}^n$ into $\mathbb{R}^n$. Suppose that the inverse f^{-1} on $f(B)$ exists, and that both f and f^{-1} are LC. If f is differentiable at a point $P_d\in B$, and $J(P_d)$ is its Jacobian matrix, then f^{-1} is differentiable at $f(P_d)$ and its Jacobian matrix agrees with $[J(P_d)]^{-1}$.*

Proof. Let $y=f(x)$. Since f is differentiable at P_d, the increment at P_d may be expressed as

$$\Delta y=J(P_d)\Delta x+\xi(\Delta x).\tag{5}$$

where

$$\frac{\|\xi(\Delta x)\|}{\|\Delta x\|}\to0\quad\text{if } \|\Delta x\|\to0.\tag{6}$$

From the Lipschitz continuity of f^{-1},

$$\frac{\|\Delta x\|}{\|\Delta y\|}<\infty.\tag{7}$$

From (6) and (7), there is a neighborhood B' of $f(P_d)$ such that $\|\Delta y\|>\|\xi(\Delta x)\|$, if $\Delta y+f(P_d)\subset B'$. Therefore, $\|\Delta y-\xi(\Delta x)\|=0$ implies $\|\Delta x\|=\|\Delta y\|=0$ in this neighborhood. This means that the matrix $J(P_d)$ is regular in (5). From (7) and the existence of $[J(P_d)]^{-1}$, the equations (5) and (6) are rewritten as follows:

$$\begin{aligned}&\Delta x=[J(P_d)]^{-1}\Delta y+[J(P_d)]^{-1}\xi(\Delta x),\\&\|[J(P_d)]^{-1}\xi(\Delta x)\|/\|\Delta y\|\to0\quad\text{if } \|\Delta y\|\to0.\end{aligned}\tag{8}$$

Since (8) holds in the neighborhood of $f(P_d)$, this proves Lemma 2. Q.E.D.

3. Local invertibility. From the results of the preceding section, it seems possible to extend many of relations obtained for continuously differentiable mappings in terms of Jacobian matrices. In this section, we prove a local property of a homeomorphism which is an extension of Fujisawa and Kuh's theorem.

THEOREM 1. *Let f be a mapping of a neighborhood B of a point* $P_0 \in \mathbb{R}^n$ *into* $\mathbb{R}^n$. *Suppose that f is LC and let* $D_k(x)$ *be the determinant of the matrix consisting of the first k rows and the first k columns of the Jacobian matrix* $J(x)$ *of f. If* $D_k(x)$ *maintains its sign and there exists a positive number* ε *such that*

$$|D_k(x)| \geqq \varepsilon$$

at almost all points in B for each $k = 1, \cdots, n$, *then the inverse mapping* f^{-1} *is uniquely determined in a neighborhood of* $f(P_0)$ *and* f^{-1} *is LC.*

Proof. We prove Theorem 1 by induction. Let $y = f(x)$ and define $f_{(k)}: B \to \mathbb{R}^n$ by $f_{(k)}(x) = (f_1(x), \cdots, f_k(x), x_{k+1}, \cdots, x_n)$ for $k = 1, \cdots, n$, where $f_1(x), \cdots, f_k(x)$ are the first k elements of $f(x)$. The identity mapping $f_{(0)}$ clearly satisfies the statement of Theorem 1. We investigate $f_{(l)}(x)$ under the assumption that $f_{(l-1)}(x)$ satisfies the statement.

For simplicity, we denote $(x_1, \cdots, x_{l-1})$, $(x_{l+1}, \cdots, x_n)$ and $(y_1, \cdots, y_{l-1})$ by $\hat{x}$, $\bar{x}$ and $\hat{y}$, respectively Then

$$(\hat{y}, x_l, \bar{x}) = f_{(l-1)}(\hat{x}, x_l, \bar{x}), \tag{9}$$

$$(\hat{y}, y_l, \bar{x}) = f_{(l)}(\hat{x}, x_l, \bar{x}) \tag{10}$$

for each $x = (\hat{x}, x_l, \bar{x}) \in B$. Since $f_{(l-1)}$ is clearly LC on B and with the leading minors $D_1(x), \cdots, D_{l-1}(x), 1, \cdots, 1$, it follows, from the assumption of induction, that there exists the inverse $f_{(l-1)}^{-1}$ being LC on a neighborhood of $f_{(l-1)}(P_0)$. Without loss of generality, we assume that this neighborhood is an open interval V_0 such that $f_{(l-1)}(P_0) \in V_0 \subset f_{(l-1)}(B)$. Denote $f_{(l)} \circ f_{(l-1)}^{-1}$ by ϕ. Then ϕ is LC on V_0 and

$$(\hat{y}, y_l, \bar{x}) = \phi(\hat{y}, x_l, \bar{x}). \tag{11}$$

Let us prove the existence of $f_{(l)}^{-1} = f_{(l-1)}^{-1} \circ \phi^{-1}$ by deriving the existence of ϕ^{-1}. Let $P_a = (\hat{y}^a, x_l^a, \bar{x}^a)$ and $P_b = (\hat{y}^b, x_l^b, \bar{x}^b)$ be arbitrary points in V_0, where $x_l^a \leqq x_l^b$ is assumed without loss of generality. Denote corresponding points by $P_1 = (\hat{y}^a, x_l^b, \bar{x}^a)$ and $Q_2 = (\hat{y}^a, \phi_l(P_b), \bar{x}^a)$ as illustrated in Fig. 1, where ϕ_l is the lth element of ϕ. Since P_a, P_b and P_1 belong to V_0, it follows that

$$\begin{aligned} \|\phi(P_b) - \phi(P_a)\|^2 &= \|\phi(P_b) - Q_2\|^2 + \|\{Q_2 - \phi(P_1)\} + \{\phi(P_1) - \phi(P_a)\}\|^2 \\ &= \|P_b - P_1\|^2 + |\{\phi_l(P_b) - \phi_l(P_1)\} + \{\phi_l(P_1) - \phi_l(P_a)\}|^2. \end{aligned} \tag{12}$$

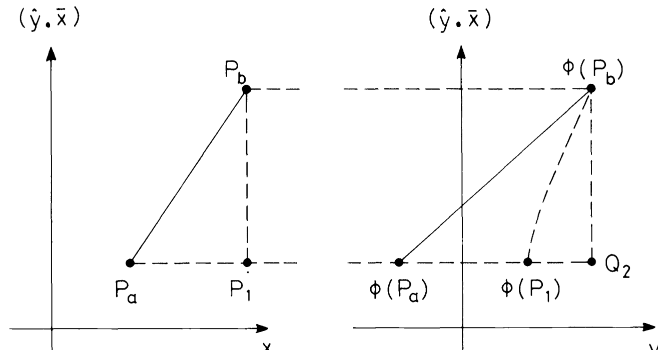

FIG. 1. *The relation of the points in the domain and range of* ϕ.

Now, we use the integration over an open interval shown in Lemma 1 to evaluate $|\phi_l(P_1)-\phi_l(P_a)|$ in (12). Let V_δ be defined by $V_\delta=\{(\hat{y}, x_l, \bar{x}) \mid |y_1-y_1^a|<\delta/2, \cdots, |y_{l-1}-y_{l-1}^a|<\delta/2, x_l^a<x_l<x_l^b, |x_{l+1}-x_{l+1}^a|<\delta/2, \cdots, |x_n-x_n^a|<\delta/2\}$. Since $V_\delta\subset V_0$ for a sufficiently small δ and ϕ is LC on V_δ, then

$$\phi_l(P_1)-\phi_l(P_a)=\lim_{\delta\to 0}\frac{1}{\delta^{n-1}}\int_{V_\delta}\frac{\partial\phi_l(\hat{y}, x_l, \bar{x})}{\partial x_l}\,d\mu_n. \tag{13}$$

Since $f_{(l-1)}^{-1}$ and ϕ are LC, it follows from Rademacher's theorem that both of $f_{(l-1)}^{-1}$ and ϕ are differentiable at almost all points in V_0. Now, let P_d be any one of such points. Then $f_{(l-1)}$ is differentiable at $f_{(l-1)}^{-1}(P_d)$ by Lemma 2, and $f_{(l)}=\phi\circ f_{(l-1)}$ is differentiable at the same point. Since y_l is a function of $\hat{y}$, x_l and $\bar{x}$ as is shown in (11), we can assume $d\hat{y}=0$, $d\bar{x}=0$ in the differential at P_d of (11), and in this case we know that $\partial\phi_l/\partial x_l$ at P_d agrees with dy_l/dx_l. On the other hand, the above differential dy_l, dx_l must satisfy the equation

$$\begin{pmatrix} 0 \\ dy_l \end{pmatrix}=J_l(Q_d)\begin{pmatrix} d\hat{x} \\ dx_l \end{pmatrix} \tag{14}$$

given from the differential at Q_d for the first l equations of (10) under the assumption that $d\hat{y}=0$, $d\bar{x}=0$, where $Q_d=f_{(l-1)}^{-1}(P_d)$, $J_l(Q_d)$ is the submatrix of $J(Q_d)$ corresponding to $D_l(Q_d)$ and $d\hat{x}$ is the differential given by $f_{(l-1)}^{-1}$. Solve (14) for dx_l in terms of dy_l by making use of Cramer's rule, and obtain dy_l/dx_l. Then we get

$$\frac{\partial\phi_l(P_d)}{\partial x_l}=\frac{D_l(Q_d)}{D_{l-1}(Q_d)}, \tag{15}$$

where $D_{l-1}(Q_d)=1$ if $l=1$.

Let η be the Lipschitz coefficient of f on B. Since $f_{(l)}$ is differentiable at $Q_d\in B$, then

$$D_l(Q_d)\geqq\varepsilon<0 \quad\text{or}\quad D_l(Q_d)\leqq-\varepsilon<0, \tag{16}$$

$$0<D_{l-1}(Q_d)\leqq(l-1)!\eta^{l-1} \quad\text{or}\quad 0>D_{l-1}(Q_d)\geqq-(l-1)!\eta^{l-1} \tag{17}$$

from the assumption of Theorem 1 and the fact that all elements of J_{l-1} are less than η. From (15)–(17) and the definition of P_d, we have

$$\frac{\partial\phi_l(x)}{\partial x_l}\geqq\zeta \quad\text{or}\quad \frac{\partial\phi_l(x)}{\partial x_l}\leqq-\zeta \quad\text{where}\quad \zeta=\frac{\varepsilon}{(l-1)!\eta^{l-1}}>0 \tag{18}$$

at almost all $x\in V_\delta\subset V_0$. Substituting (18) into (13), we have

$$|\phi_l(P_1)-\phi_l(P_a)|\geqq\zeta\|P_1-P_a\|. \tag{19}$$

Here, we rewrite (12) as follows:

$$\begin{aligned}\|\phi(P_b)-\phi(P_a)\|^2&=\|P_b-P_1\|^2+(\sigma_1\|P_b-P_1\|+\sigma_2\|P_1-P_a\|)^2\\&\geqq\|P_b-P_1\|^2;\end{aligned} \tag{20}$$

then from (19) and the Lipschitz continuity

$$|\sigma_1|<\eta_\phi, \qquad \sigma_2\geqq\zeta \tag{21}$$

should be satisfied, where η_ϕ is the Lipschitz coefficient of ϕ_l on V_0. Furthermore,

(20) is also transformed by making use of (21) as

$$\|\phi(P_d)-\phi(P_a)\|^2=\left\{\frac{\sigma_2^2}{1+\sigma_1^2}\right\}\|P_1-P_a\|^2+\left(\frac{\sigma_1\sigma_2}{\sqrt{1+\sigma_1^2}}\|P_1-P_a\|+\sqrt{1+\sigma_1^2}\,\|P_b-P_1\|\right)^2 \tag{22}$$
$$\geqq\{\zeta^2/(1+\eta_\phi^2)\}\|P_1-P_a\|^2.$$

Adding (20) and (22) multiplied by $(1+\eta_\phi^2)/\zeta^2$, we have

$$\|\phi(P_b)-\phi(P_a)\|\geqq(\zeta/\sqrt{1+\zeta^2+\eta_\phi^2})\|P_b-P_a\|. \tag{23}$$

Since P_a and P_b are arbitrary points in V_0, (23) indicates that ϕ^{-1} is uniquely determined and that ϕ^{-1} is LC. Recalling that $f_{(l-1)}^{-1}$ is LC on V_0 and noting that $\phi(V_0)$ is a neighborhood of $f_{(l)}(P_0)=\phi\circ f_{(l-1)}(P_0)$, we conclude that $f_{(l)}^{-1}=f_{(l-1)}^{-1}\circ\phi^{-1}$ is determined and is LC on the neighborhood. Thus $f_{(l)}^{-1}$ satisfies the statement of Theorem 1, and the proof of the theorem is completed. Q.E.D.

4. Global invertibility. We show some global properties derived from Theorem 1. We say that a mapping f defined on $\mathbb{R}^n$ is LLC (locally Lipschitz continuous) if f is LC on a neighborhood of each $x\in\mathbb{R}^n$.

THEOREM 2. *A mapping $f:\mathbb{R}^n\to\mathbb{R}^n$ is a homeomorphism of $\mathbb{R}^n$ onto itself and f^{-1} is LLC if*

(i) *f is LLC,*

(ii) *there exist a neighborhood $B(z)$ and a positive number $\varepsilon(z)$ for each $z\in\mathbb{R}^n$ such that $D_k(x)$ maintains its sign and*

$$|D_k(x)|\geqq\varepsilon(z)$$

at almost all $x\in B(z)$ for each $k=1,\cdots,n$,

(iii) *$\|f(x)\|\to\infty$ if $\|x\|\to\infty$.*

Proof. It is known [19] that a mapping f is a homeomorphism of $\mathbb{R}^n$ onto itself if and only if f is

(a) a local homeomorphism,

(b) a proper map.

Clearly, (a) is obtained from the assumption (ii) of Theorem 2 and Theorem 1, (b) is obtained from (iii) of Theorem 2 [19], and f^{-1} is LLC from Theorem 1. Q.E.D.

Remark 1. Fujisawa and Kuh have expressed their condition for invertibility of piecewise-linear mappings in terms of certain minors instead of the leading minors of the above theorems [8]. The results of this paper can be extended in like manner as follows. In general, the next proposition is clearly valid: A mapping $f^*:\mathbb{R}^n\to\mathbb{R}^n$ is a homeomorphism of $\mathbb{R}^n$ onto itself if there exist two homeomorphisms ψ_1 and ψ_2 of $\mathbb{R}^n$ onto itself and if (i)–(iii) of Theorem 2 are satisfied by letting $f=\psi_1\circ f^*\circ\psi_2$. As a special case, ψ_1 and ψ_2 may be regular matrices describing interchanges of the orders and directions of the coordinate axes. Therefore, in Theorem 2 and its corollaries, $D_1(x),\cdots,D_n(x)$ can be replaced with $D_1^*(x),\cdots,D_n^*(x)$ defined as follows: Let $(i_1,\cdots,i_k)$ and $(j_1,\cdots,j_k)$ be two permutations of $(1,\cdots,n)$, let $D_k^*(x)$ be the determinant of the $k\times k$ submatrix consisting of the i_1th, $\cdots$, i_kth rows and the j_1th, $\cdots$, j_kth columns of the Jacobian matrix $J(x)$.

On the other hand, if all leading minors $D_1(x),\cdots,D_n(x)$ are positive, it can be considered to be the canonical case of the theorems. Then, the condition (ii) of Theorem 2 is clearly related to the class of constant square matrices whose leading principal minors are positive. (Note that matrices are not assumed to be symmetric.) This class clearly contains the class P introduced by Fiedler and Pták [6] as a generalization of positive definiteness because the class P requires the positivity of all principal minors. To simplify the understanding, we only consider the canonical case in Remark 2 and 3.

Remark 2. In Theorem 2, the words "almost all x" can be replaced with "all differentiable x". Although the theorem shows a sufficient condition for global invertibility, it is not necessary [8, Example 1]. On the other hand, if f is continuously differentiable, $D_n(x) \geqq \varepsilon(z)$ alone is sufficient in (ii) of Theorem 2. However it should be noted that this condition is not sufficient for mappings being LC. A counterexample, in which $f: \mathbb{R}^2 \to \mathbb{R}^2$ is piecewise-linear and a local homeomorphism at every point except the origin but is not one-to-one at every point except the origin, is seen in [8, Example 2]. Yet a little doubt about the necessity of other leading minors $D_1(x), \cdots, D_{n-1}(x)$ may remain from the following facts: 1) f of the example is transformed into a mapping of class C^1 with isolated singular points of the Jacobian matrix by a sufficiently small modification of f, that is, smoothing the edges; 2) for the mapping of class C^1 with isolated singular points to be a homeomorphism, assumptions that $D_n(x) \neq 0$ except the singular points and (ii) of Theorem 2 are sufficient when $n \neq 2$ [1]. In the case of the Lipschitz continuous mapping, however, the conditions of the other leading minors can not be omitted even if $n \geqq 3$. An illustrative example $y = f(x)$ is easily obtained by adding $y_3 = x_3, \cdots, y_n = x_n$ to the above example of $n = 2$, where f must possess connected singular points if smoothed.

Remark 3. Recently, Haneda has shown conditions for a homeomorphism of the Lipschitz continuous mapping by making use of the local μ-functional [10], and his result is a generalization of the monotone operator by Minty [13] et al. in finite dimensional cases. In his result, the positivity in terms of the μ-functional is essentially used instead of condition (ii) of Theorem 2. We can now translate his conditions in terms of the Jacobian matrix. It is known that the positivity of matrices in terms of the μ-functional is related to positive definiteness, column-sum dominance and row-sum dominance for the l_2, l_1 and l_∞ norm, respectively [4], [10]. Haneda has given a further condition corresponding to an arbitrary norm. However the positivity of matrices in terms of the μ-functional implies the positivity of real parts of all eigenvalues [4, the property 4), p. 480], and such matrices belong to the class P [6, Thm. 1.1]. Therefore, recalling Remark 1, we see that the conditions by Haneda can be essentially derived from Theorem 2. On the other hand, it is notable that some matrices with the positive leading minors never satisfy the positivity in terms of μ-functional. For example, the 2×2 matrix $\{J_{ij}\}$ with $J_{11} = J_{21} = 1$, $J_{12} = -6$ and $J_{22} = -4$ has positive leading minors and has a negative eigenvalue. Therefore, there are mappings which satisfy the conditions of Theorem 2, but do not satisfy the monotonicity by μ-functional. These mappings often arise for networks with active elements such as transistors and tunnel diodes, and an example is shown in § 6.

From Theorem 2, we will derive some practical conditions. The next is an extension of the ratio condition introduced by Fujisawa and Kuh [9].

COROLLARY 1. *Suppose that $f: \mathbb{R}^n \to \mathbb{R}^n$ is LLC, each of $D_1(x)$, $D_2(x)/D_1(x), \cdots, D_n(x)/D_{n-1}(x)$ maintains its sign and there exists a positive number ε such that*

$$|D_1(x)| \geqq \varepsilon, \quad \left|\frac{D_2(x)}{D_1(x)}\right| \geqq \varepsilon, \quad \cdots, \quad \left|\frac{D_n(x)}{D_{n-1}(x)}\right| \geqq \varepsilon$$

at almost all $x \in \mathbb{R}^n$. Then f is a homeomorphism of $\mathbb{R}^n$ onto itself and f^{-1} is LLC.

Proof. Assumptions (i), (ii) of Theorem 2 are easily obtained from those of Corollary 1. Therefore, we derive (iii) by induction. Notations used here are the same as defined in the proof of Theorem 1. It is easy to verify (iii) for $f_{(0)}$. Now assume that the statement of Theorem 2 is satisfied for $f_{(l-1)}$. Then the size of the interval V_0, where (19) holds, can be determined arbitrarily. We see that ζ of (19) is bounded regardless of the size of V_0 by referring to (15), (18) and the assumption of Corollary 1. Let one of P_a or P_1 be the origin of $\mathbb{R}^n$ and let $\|P_1 - P_a\| = |x_l|$. Then we have

$$(24) \qquad \|f_{(l)}(x_l)\| \to \infty \quad \text{if } |x_l| \to \infty$$

from (19) and the fact that $\|f_{(l)}(x_l)\| \geqq \|\phi_l(P_1) - \phi_l(P_1)\|$. On the other hand, we have

$$(25) \qquad \|f_{(l-1)}(x_l)\| \to \infty \quad \text{if } \|\hat{x}\| \to \infty$$

from the assumption of induction. By definition of notations, (iii) of Theorem 2 for $f_{(l)}$ is verified from (24) and (25), and the proof is completed by making use of the theorem. Q.E.D.

A global Lipschitz condition holds in the following case.

COROLLARY 2. *Suppose that f: $\mathbb{R}^n \to \mathbb{R}^n$ is LC, $D_k(x)$ maintains its sign and there exists a positive number ε such that*

$$|D_k(x)| \geqq \varepsilon$$

at almost all $x \in \mathbb{R}^n$ for each $k = 1, \cdots, n$. Then f is a homeomorphism of $\mathbb{R}^n$ onto itself and f^{-1} is LC.

Proof. Since all elements of J and its leading minors are bounded by the Lipschitz condition of f, assumptions of Corollary 1 are satisfied from those of Corollary 2 and, furthermore, all elements of J^{-1} are bounded. Hence we conclude that f^{-1} with bounded partial derivatives on $\mathbb{R}^n$ is LC by referring to Lemma 1 and Lemma 2. Q.E.D.

The authors have started by extending the uniformly positive definiteness to the mapping being LC [11]. On the other hand, it is known that the uniformly positive definiteness implies the ratio condition [9]. Similarly, we obtain the extension as a corollary of Theorem 2. Note again that the following Jacobian matrix is not assumed to be symmetric. Recalling Lemma 1, we see that this result is related to the monotone operator by Minty [13].

COROLLARY 3. *Suppose that f: $\mathbb{R}^n \to \mathbb{R}^n$ is LC and that there exists a positive constant ε such that*

$$\langle z, J(x)z \rangle \geqq \varepsilon \langle z, z \rangle$$

for all $z \in \mathbb{R}^n$ and for almost all $x \in \mathbb{R}^n$, where $\langle \cdot, \cdot \rangle$ indicates the inner product. Then f is a homeomorphism of $\mathbb{R}^n$ onto itself and f^{-1} is LC.

5. Application to resistive network equations. We consider a network consisting of $\lambda + m$ branches where λ branches are independent signal (current or voltage) sources and m branches are resistive elements. The network equation is determined by the interconnection of the branches and branch characteristics, and is expressed as

$$(26) \qquad f_E(x, u) = 0,$$

where $u \in \mathbb{R}^m$ is the set of independent signals of the sources, $x \in \mathbb{R}^{2\lambda+m}$ is the set of all branch currents and voltages with the exception of u, and f_E: $\mathbb{R}^{2\lambda+2m} \to \mathbb{R}^{2\lambda+m}$ represents $\lambda + m$ Kirchhoff relations and λ resistive branch characteristics. To clarify the existence of a unique solution and the dependence on source signals, we apply Corollary 2. Since it is possible to eliminate some unknown signals by substitutions in many cases, we replace $\lambda + m$ with an arbitrary positive integer n and consider f_E to be an arbitrary Lipschitz continuous mapping.

COROLLARY 4. *Suppose that* $f_E: \mathbb{R}^{n+m} \to \mathbb{R}^n$ *is LC,* $D_k(x, u)$ *maintains its sign and there exists a positive number* ε *such that*

$$|D_k(x, u)| \geqq \varepsilon$$

at almost all $(x, u) \in \mathbb{R}^{n+m}$ *for each* $k = 1, \cdots, n$, *where* $D_k(x, u)$ *are the leading minors of the Jacobian matrix* $\partial f_E(x, u)/\partial x$. *Then,* (26) *has a unique solution* $x \in \mathbb{R}^n$ *for each* $u \in \mathbb{R}^m$, *and the solution is expressed as*

$$x = g_E(u)$$

by making use of a certain mapping $g_E: \mathbb{R}^m \to \mathbb{R}^n$ *being LC.*

Proof. This is easily verified by applying Corollary 2 to the mapping $f: (x, u) \mapsto (f_E(x, u), u)$. Q.E.D.

Let us continue in more detail. Usually, Kirchhoff's law associated with the resistive branches of the network and characteristics of the resistive branches are written respectively as follows:

$$\begin{bmatrix} 1_t & 0 & 0 & -w^T \\ 0 & 1_l & w & 0 \end{bmatrix} \begin{bmatrix} i_t \\ v_l \\ v_t \\ i_l \end{bmatrix} = \begin{bmatrix} j \\ e \end{bmatrix}, \qquad f_R(i_t, v_l, v_t, i_l) = 0. \tag{27}$$

In (27), i and v denote currents and voltages of the resistive branches respectively, and they are classified with the subscripts t and l indicating their association with a tree or its co-tree respectively of the network graph; the notations 1_t and 1_l denote unit matrices, and w (superscript T indicates its transpose) denotes a matrix describing the interconnection of the resistive branches (precisely speaking, $[1_l \quad w]$ and $[-w^T \quad 1_t]$ are the fundamental loop matrix and the fundamental cutset matrix respectively [5, Chap. 11]); j and e denote current sources and voltage sources respectively; $f_R: \mathbb{R}^{2\lambda} \to \mathbb{R}^\lambda$ is the mapping representing characteristics of the resistive branches and is assumed to be LC.

Now we study (27) directly without considering the elimination process of unknown variables. Then the preceding theorems and corollaries are applicable by investigating the Jacobian matrix

$$\frac{\partial f_E(x, u)}{\partial x} = \begin{bmatrix} 1_t & 0 & 0 & -w^T \\ 0 & 1_l & w & 0 \\ \dfrac{\partial f_R(x)}{\partial i_t} & \dfrac{\partial f_R(x)}{\partial v_l} & \dfrac{\partial f_R(x)}{\partial v_t} & \dfrac{\partial f_R(x)}{\partial i_l} \end{bmatrix} \tag{28}$$

where $x = (i_t, v_l, v_t, i_l)$ and $u = (j, e)$. Interchanges of the coordinate axes related to Remark 1 are sometimes required. However, the total number of possible interchanges is not usually large because the majority of elements of the Jacobian matrix (28) for practical networks are zero elements. For many networks, preceding theorems are directly applicable provided that the elements of (v_t, i_l) and those of f_R correspond to λ branches in the same order.

6. Application to dynamic network equations. We study this problem basically from the viewpoint of the simultaneous differential equations in the form

$$f_{DE}(x_1, \dot{x}_2, x_2, u, t) = 0 \tag{29}$$

defined on a time interval $T = [t_0, t_1]$, where $f_{DE}: \mathbb{R}^{n_1+2n_2+m} \times T \to \mathbb{R}^{n_1+n_2}$; $x_1 \in \mathbb{R}^{n_1}$ and

$x_2 \in \mathbb{R}^{n_2}$ accompanied by $\dot{x}_2 = dx_2/dt$ are unknown vector-valued function on T; $u \in \mathbb{R}^m$ is a known vector-valued function on T (n_1, n_2 and m are allowable to be zero as long as $n_1 + n_2 \geqq 1$). This is a more primitive form of system equations than the normal form. For example, equations of RLC networks (networks composed of resistors, inductors, capacitors and independent signal sources) in this form are usually obtained by letting x_2 be the set of the capacitor charges and inductor fluxes and x_1 be the set of all unknown network signals with the exception of x_2 and $\dot{x}_2$. Although further considerations on the determination of x_1, x_2 are interesting, they are outside of the purpose of this paper.

Now, consider x_2, u, t to be fixed points and x_1, $\dot{x}_2$ unknowns in (29). Then (29) is considered to be a set of resistive equations. When this set of the equations has a unique solution for each (x_2, u, t) the solution may be expressed as

$$x_1 = g_1(x_2, u, t), \tag{30}$$

$$\dot{x}_2 = g_2(x_2, u, t). \tag{31}$$

Clearly, (31) is a set of differential equations in the normal form. In other words, the existence of the solution shown in (30), (31) means that the system (29) has a state equation and that x_2 is suitable as a set of state-variables. By the well-known uniqueness theorem of the normal form [3], the following is easily obtained, where a global Lipschitz condition guarantees the uniqueness of the solution on the whole interval T.

LEMMA 3. *Suppose that* (29) *has a unique solution* $(x_1, \dot{x}_2)$ *in* $\mathbb{R}^{n_1+n_2}$ *for each fixed* $(x_2, u, t) \in \mathbb{R}^{n_2+m} \times T$ *and that* g_2 *is continuous on the domain and satisfies a Lipschitz condition with respect to* x_2 *in the domain. Then the simultaneous differential equations* (29) *have a unique solution* (x_1, x_2) *on* T *for each continuous function* u *on* T *and for each initial value* $x_2(t_0)$ *in* $\mathbb{R}^{n_2}$.

The following result is obtained by applying Corollary 4 to Lemma 3.

THEOREM 3. *Suppose that* $f_{DE}: \mathbb{R}^{n_1+2n_2+m} \times T \to \mathbb{R}^{n_1+n_2}$ *is continuous on the domain and satisfies a Lipschitz condition with respect to* $(x_1, \dot{x}_2, x_2)$ *in the domain. Let* $D_k(x_1, \dot{x}_2, x_2, u, t)$ *be the determinant of the matrix consisting of the first* k *rows and the first* k *columns of the Jacobian matrix* $\partial f_{DE}(x_1, \dot{x}_2, x_2, u, t)/\partial(x_1, \dot{x}_2)$. *If* $D_k(x_1, \dot{x}_2, x_2, u, t)$ *maintains its sign and there exists a positive number* ε *such that*

$$|D_k(x_1, \dot{x}_2, x_2, u, t)| \geqq \varepsilon$$

at almost all $(x_1, \dot{x}_2, x_2, u, t) \in \mathbb{R}^{n_1+2n_2+m} \times T$ *for each* $k = 1, \cdots, n$, *then simultaneous differential equations* (29) *have a unique solution* (x_1, x_2) *on* T *for each continuous function* u *on* T *and for each* $x_2(t_0) \in \mathbb{R}^{n_2}$.

Proof. Consider u, t to be fixed points in (29) and apply Corollary 4 to the equation. By the assumptions of Theorem 3, there exists a unique solution $(x_1, \dot{x}_2)$ shown in (30) and (31), and g_1, g_2 satisfy Lipschitz conditions with respect to x_2 for each (u, t). Therefore, by letting $y = f_{DE}(x_1, \dot{x}_2, x_2, u, t)$, their differential coefficients satisfy

$$\frac{\partial(x_1, \dot{x}_2)}{\partial x_2} = \left[\frac{\partial y}{\partial(x_1, \dot{x}_2)}\right]^{-1} \left[\frac{\partial y}{\partial x_2}\right] \tag{32}$$

at all differentiable points. Since $1/D_{n_1+n_2}$ and all elements of $\partial y/\partial(x_1, x_2)$ are bounded regardless of x_2, u, t, all elements of $[\partial y/\partial(x_1, x_2)]^{-1}$ are bounded as well as $\partial y/\partial x_2$. By Lemma 2, it follows that Lipschitz coefficients of g_1, g_2 with respect to x_2 are bounded regardless of u, t. If we denote matrices describing ratios of increments in a similar manner to (32), we have

$$\frac{\Delta(x_1, \dot{x}_2)}{\Delta u} = \left[\frac{\Delta y}{\Delta(x_1, \dot{x}_2)}\right]^{-1}\left[\frac{\Delta y}{\Delta u}\right], \qquad \frac{\Delta(x_1, \dot{x}_2)}{\Delta t} = \left[\frac{\Delta y}{\Delta(x_1, \dot{x}_2)}\right]^{-1}\left[\frac{\Delta y}{\Delta t}\right] \tag{33}$$

correspondingly, where (x_2, t) or (x_2, u) is fixed. Since f_{DE} is continuous and each element of $[\Delta y/\Delta(x_1, \dot{x}_2)]^{-1}$ is bounded by the maximum value of the elements of $[\partial y/\partial(x_1, \dot{x}_2)]^{-1}$, it follows from (33) that g_1, g_2 are continuous in u and t. Thus the conditions of Lemma 3 are satisfied and the proof is completed by making use of the lemma. Q.E.D.

Now, we consider the network example shown in Fig. 2 which contains a current source j_1, a capacitor S, a resistor R, an inductor Γ and, moreover, an ideal diode such that the voltage v_D across it is zero if forward-biased and the current i_D is zero if backward-biased as shown in Fig. 3(a). In Fig. 2, q_s and ϕ_r represent the charge and flux, respectively. Let us define f_D: $\mathbb{R}^2 \to \mathbb{R}^1$ by

$$f_D(i_D, v_D) = \begin{cases} i_D & \text{if } i_D \leqq -v_D, \\ -v_D & \text{if } i_D > -v_D. \end{cases} \tag{34}$$

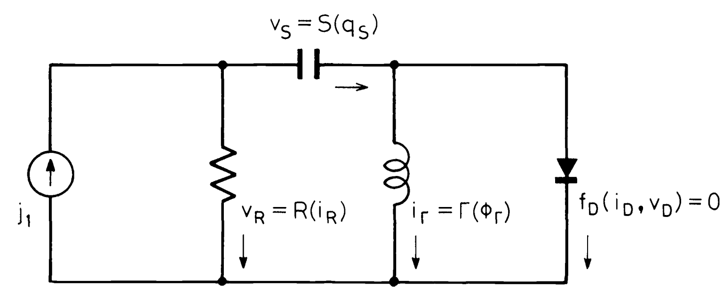

FIG. 2. *Example of a nonlinear RLC network.*

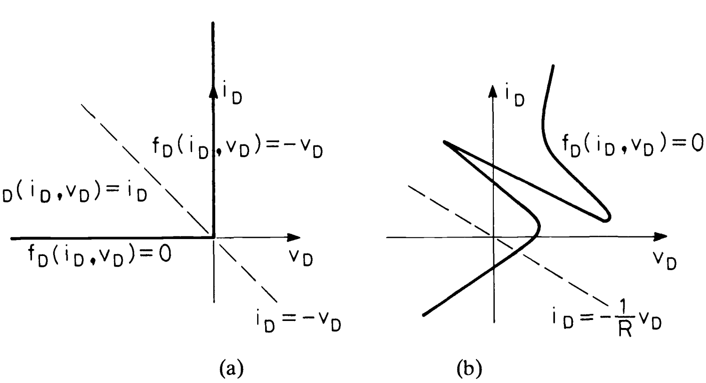

FIG. 3. *Voltage-current characteristics of* f_D: (a) *the ideal diode,* (b) *an example which guarantees the unique solvability.*

Then f_D is LC and the characteristic of the ideal diode is expressed as

$$f_D(i_D, v_D) = 0. \tag{35}$$

From Kirchhoff's law and branch characteristics, the equations of this network are expressed as (35) and as follows:

$$
\begin{aligned}
&i_S - i_D - i_r = 0, \qquad i_R + i_D + i_\Gamma = j_1, \qquad v_D + v_S - v_R = 0, \\
&v_\Gamma + v_S - v_R = 0, \qquad v_S = S(q_S), \qquad v_R = R(i_R), \qquad i_\Gamma = \Gamma(\phi_\Gamma),
\end{aligned} \tag{36}
$$

where

$$
i_S = \dot{q}_S, \qquad v_\Gamma = \dot{\phi}_\Gamma. \tag{37}
$$

Even if S, R and Γ are continuously differentiable, the whole equations are not so because of the existence of the diode, and the preceding results become useful. The Jacobian matrix of Theorem 3 arranged in a similar manner to (28) is

$$
\frac{\partial f_{DE}(i_S, i_R, v_D, v_\Gamma, v_S, v_R, i_D, i_\Gamma, q_S, \phi_\Gamma, j_1)}{\partial(i_S, i_R, v_D, v_\Gamma, v_S, v_R, i_D, i_\Gamma)}
= \begin{bmatrix}
1 & & & & & & -1 & -1 \\
 & 1 & & & & & 1 & 1 \\
 & & 1 & & 1 & -1 & & \\
 & & & 1 & 1 & -1 & & \\
 & & & & 1 & & & \\
 & -\dfrac{dR(i_R)}{di_R} & & & & 1 & & \\
 & & \dfrac{\partial f_D(i_D, v_D)}{\partial v_D} & & & & \dfrac{\partial f_D(i_D, v_D)}{\partial i_D} & \\
 & & & & & & & 1
\end{bmatrix}. \tag{38}
$$

Note that the differential coefficients about f_D are ± 1 or 0. The leading minors are

$$
D_1 \sim D_8 = 1 \quad \text{if } i_D \leqq -v_D,
$$

$$
D_1 \sim D_6 = 1, \qquad D_7 = D_8 = \frac{dR(i_R)}{di_R} \quad \text{if } i_D > -v_D.
$$

Thus we conclude, by Theorem 3, that if there exists a positive constant ε such that $dR(i_R)/di_R \geqq \varepsilon$ at each differentiable i_R, then the simultaneous equations (35)–(37) have a unique solution for each continuous input j_1 and for each $q_S(t_0)$, $\phi_\Gamma(t_0)$.

The ideal diode is a typical element which is neither current-controlled nor voltage-controlled, that is, neither current nor voltage of the diode can be considered to be dependent on the other. Chua and Rohrer have studied another formulation of networks with such branches [2], but for their examples the unique solvability is not taken into account. We show one more example with unique solvability. Suppose that the diode is replaced by a resistor with more general characteristic and that R is a linear resistor in Fig. 2. Then the above leading minors become

$$
D_1 \sim D_6 = 1, \qquad D_7 = D_8 = \frac{\partial f_D(i_D, v_D)}{\partial i_D} - R\frac{\partial f_D(i_D, v_D)}{\partial v_D}
$$

from (38). Hence the conditions of Theorem 3 are satisfied if there exists a positive constant ε such that

$$
\frac{\partial f_D(i_D, v_D)}{\partial i_D} - R\frac{\partial f_D(i_D, v_D)}{\partial v_D} \geqq \varepsilon
$$

at all differentiable (i_D, v_D). Thus it can be easily verified that a suitable f_D with characteristic shown in Fig. 3(b) satisfies this condition and this network has a state equation with unique solvability. In such cases, some real parts of eigenvalues of the Jacobian matrix become both positive and negative (for example, consider the case that $R = \partial f_D/\partial i_D = 1$, $\partial f_D/\partial v_D = 0$ at a point and $R = -\partial f_D/\partial i_D = 1$, $\partial f_D/\partial v_D = -2$ at another point in this network), and the results of this paper exhibit its usefulness as mentioned in Remark 3.

Acknowledgment. The author would like to thank Prof. K. Miyakoshi of Osaka Electro-Communication University and Dr. S. Minamoto of University of Osaka Pref. for the guidance and encouragement. He is also grateful to Dr. T. Matsumote of Waseda University for valuable discussion.

REFERENCES

[1] L. O. CHUA AND Y. F. LAM, *Global homeomorphism of vector-valued functions*, J. Math. Anal. Appl., 39 (1972), pp. 600–624.
[2] L. O. CHUA AND R. A. ROHRER, *On the dynamic equations of a class of nonlinear RLC networks*, IEEE Trans. Circuit Theory, CT-12 (1965), pp. 475–489.
[3] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
[4] C. A. DESOER AND H. HANEDA, *The measure of a matrix as a tool to analyze computer algorithms for circuit analysis*, IEEE Trans. Circuit Theory, CT-19 (1972), pp. 480–486.
[5] C. A. DESOER AND E. S. KUH, *Basic Circuit Theory*, McGraw-Hill, New York, 1969.
[6] M. FIDLER AND V. PTÁK, *Some generalization of positive definiteness and monotonicity*, Numer. Math., 9 (1966), pp. 163–172.
[7] A. FRIEDMAN, *Differential Games*, John Wiley, New York, 1971.
[8] T. FUJISAWA AND E. S. KUH, *Piecewise-linear theory of nonlinear networks*, SIAM J. Appl. Math., 22 (1972), pp. 307–328.
[9] ———, *Some results on existence and uniqueness of solutions of nonlinear networks*, IEEE Trans. Circuit Theory, CT-18 (1971), pp. 501–506.
[10] H. HANEDA, *Generalization of monotonicity applied to the DC analysis of nondifferentiable resistive networks*, IEEE Trans. Circuit and Systems, CAS-21 (1974), pp. 406–412.
[11] Y. KAWAMURA, S. MINAMOTO AND K. MIYAKOSHI, *Expansion of uniformly positive definiteness on nonlinear network equation*, Trans. IECE Japan (A), 57-A (1974), pp. 332–333.
[12] A. N. KOLMOGOROV AND S. V. FOMIN, *Measure, Lebesgue Integrals and Hilbert Space*, Academic Press, New York, 1961.
[13] G. J. MINTY, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.
[14] T. OHTSUKI AND H. WATANABE, *State-variable analysis of RLC networks containing nonlinear coupling elements*, IEEE Trans. Circuit Theory, CT-16 (1969), pp. 26–38.
[15] R. S. PALAIS, *Natural operations on differential forms*, Trans. Amer. Math. Soc., 92 (1959), pp. 125–141.
[16] I. W. SANDBERG AND A. N. WILLSON JR., *Some theorems on properties of DC equations of nonlinear networks*, Bell System Tech. J., 48 (1969), pp. 1–34.
[17] V. I. SMIRNOV, *A Course of Higher Mathematics* 5, Pergamon Press, New York, 1964.
[18] A. N. WILLSON JR., ED., *Nonlinear Networks: Theory and Analysis*, IEEE Press, New York, 1974.
[19] F. F. WU AND C. A. DESOER, *Global inverse function theorem*, IEEE Trans. Circuit Theory, CT-19 (1972), pp. 199–201.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0005 $01.00/0

A NOTE ON RESTRICTED PSEUDOINVERSES*

JOACHIM HARTUNG†

Abstract. For the restricted pseudoinverse of a linear Hilbert space operator, introduced by Minamide, Nakamura (1970) and further considered for instance by Holmes (1972), representations are derived that extend results of Den Broeder and Charnes (1962) and Morozov (1969) for nonrestricted pseudoinverses. This representation permits computation with ordinary methods of inversion. In often occurring problems of optimal control and best approximation, the operators to be inverted then are positive definite symmetric matrices.

Let H_1, H_2 and H_3 be real Hilbert spaces, where $\langle\cdot\,,\cdot\rangle$ denotes the inner product, which then defines the norm $\|\cdot\|=\langle\cdot\,,\cdot\rangle^{1/2}$. Let $T\colon H_1\to H_2$, $S\colon H_1\to H_3$ be continuous linear operators, and $(T, S)\colon H_1\to H_2\times H_3$ defined by $(T, S)\colon x\mapsto(Tx, Sx)$ be the product transformation, where $H_2\times H_3$ is a product Hilbert space equipped with the induced inner product. Denote $R(T)$ the range, $N(T)$ the null space, T^* the adjoint, T^{-1} the inverse, and T^+ the pseudoinverse of T.

The pseudoinverse $T^+\colon H_2\to H_1$ is uniquely determined, provided it exists, by producing for every $y\in H_2$ the "best approximate solution" $x(y)=T^+y$ of the linear equation

$$(1)\qquad Tx=y,\qquad x\in H_1,$$

i.e. T^+y is the element of minimum norm which gives a minimum value for the discrepancy $\|Tx-y\|$, $x\in H_1$. Let

$$(2)\qquad i(T):=\inf\left\{\frac{\|Tx\|}{\|x\|}\,\middle|\,x\neq 0,\ x\in N(T)^\perp\right\},$$

where $N(T)^\perp$ is the orthogonal complement of $N(T)$ in the domain of definition of T, then (e.g. Petryshyn [6]):

$$(3)\qquad i(T)>0\Leftrightarrow R(T)\text{ is closed}\Rightarrow T^+\text{ exists and }\|T^+\|=i(T)^{-1}<+\infty.$$

Restricting in (1) the x to satisfy $Sx=0$, $x\in H_1$, we come to the concept of restricted pseudoinverses, introduced by Minamide, Nakamura [4] and further considered by Holmes [3]. Since the null space $N(S)$ of S is a closed linear subspace, we can regard $N(S)$ as a Hilbert space $H_{N(S)}$. Let us consider the restriction of T to $H_{N(S)}$ and denote this transformation by T_S. If $R(T_S)$ is closed, the pseudoinverse T_S^+ from H_2 into $H_{N(S)}$ exists and is well-defined.

DEFINITION. The pseudoinverse T_S^+, regarded as a linear transformation from H_2 into H_1, is called the $(N(S)\text{-})$ *restricted pseudoinverse of* T.

LEMMA 1. *If* (T, S) *has closed range, then*

$$T_S^+\ \textit{exists}\quad\textit{and}\quad\|T_S^+\|\leqq\|(T, S)^+\|.$$

* Received by the editors February 14, 1977, and in revised form September 27, 1977.
† Abteilung für Angewandte Statistik, Universität Bonn, Bonn, West Germany.

Proof. Note that

$$\begin{aligned} i(T_S) &= \inf\left\{\frac{\|T_S x\|}{\|x\|}\,\middle|\, x \neq 0,\ x \in N(T_S)^\perp\right\} \\ &= \inf\left\{\frac{\|Tx\|}{\|x\|}\,\middle|\, x \neq 0,\ x \in N(S),\ x \in (N(T)\cap N(S))^\perp\right\} \\ &\geqq \inf\left\{\frac{\|(T,S)x\|}{\|x\|}\,\middle|\, x \neq 0,\ x \in N((T,S))^\perp\right\} \\ &= i((T,S)), \end{aligned}$$

hence (3) gives the lemma.

We give in Theorem 1 a representation of T_S^+, which permits the application of ordinary methods of inversion for self-adjoint positive definite operators. In the case, often occurring in optimal control problems, that $H_2 \times H_3$ are finite dimensional, the assumptions are satisfied, and the operators to be inverted then are positive definite symmetric matrices.

Let us consider the linear equation $Tx = y$, $x \in H_1$, $y \in R(T)$ given. If $R(T)$ is finite dimensional as for instance in the fixed endpoint quadratic regulator problem, it may be convenient to change over to the dual problem, find a solution z_0 of $TT^*z = y$, $z \in H_2$, because then TT^* is a matrix. Now $x_0 = T^*z_0$ is the minimum norm solution of $Tx = y$. If TT^* is invertible, then we have $x_0 = T^+y = T^*(TT^*)^{-1}y$, otherwise we consider the perturbed system $TT^*z + rI_2z = y$, where $r > 0$ and I_2 is the identity on H_2. $(TT^* + rI_2)$ is invertible and we get (cf. Corollary 1 after Theorem 1 below) that $\lim_{r\to+0} T^*(TT^* + rI_2)^{-1}y$ exists and is equal to T^+y. An analogous formula for T_S^+y is given now by the following theorem.

THEOREM 1. *If (T, S) has closed range, then we have the representation*

$$T_S^+ = \lim_{r\to+0} (r^2T, rS)^*((rT, S)(rT, S)^* + r^3I)^{-1}_{|H_2\times\{0\}} \tag{4}$$

(in the sense of pointwise convergence on H_2), where $r \in \mathbb{R}$, and I denotes the identity on $H_2 \times H_3$.

Proof. Let $r \in \mathbb{R}$, $0 < r \leqq 1$. Then

$$\begin{aligned} i((rT,S)) &= \inf\{\|(rT,S)x\| \mid \|x\| = 1,\ x \in N((rT,S))^\perp\} \\ &= \inf\{(\|rTx\|^2 + \|Sx\|^2)^{1/2} \mid \|x\| = 1,\ x \in N((T,S))^\perp\} \\ &= \inf\{(\|rTx\|^2 + \|rSx\|^2 + \|(1-r^2)^{1/2}Sx\|^2)^{1/2} \mid \|x\| = 1,\ x \in N((T,S))^\perp\} \\ &\geqq \inf\{(\|rTx\|^2 + \|rSx\|^2)^{1/2} \mid \|x\| = 1,\ x \in N((T,S))^\perp\} \\ &= \inf\{\|(rT,rS)x\| \mid \|x\| = 1,\ x \in N((T,S))^\perp\} \\ &= r\cdot i((T,S)), \end{aligned}$$

and by (3) $i((T,S)) > 0$, $(rT,S)^+$ exists and

$$\|(rT,S)^+\| \leqq \frac{1}{r}\, i((T,S))^{-1}. \tag{5}$$

Let $y_0 \in H_2$ be arbitrarily chosen, but fixed. Then for $(y, z) \in H_2 \times H_3$ the equation

$$(rT,S)(rT,S)^*(y,z) + r^3(y,z) = (ry_0, 0) \tag{6}$$

has the unique solution

$$(y, z)_r = ((rT, S)(rT, S)^* + r^3 I)^{-1}(ry_0, 0) =: Q_r^{-1}(ry_0, 0). \tag{7}$$

Pre-multiplying in (6) with $(rT, S)^*$ and putting

$$x = (rT, S)^*(y, z)$$

we get

$$(rT, S)^*(rT, S)x + r^3 x = (rT, S)^*(ry_0, 0), \tag{8}$$

which is uniquely solved by

$$x_r = (rT, S)^*(y, z)_r = (rT, S)^* Q_r^{-1}(ry_0, 0). \tag{9}$$

Now the solution of (8) is the solution of the variational problem

$$\text{minimize } \{\|(ry_0, 0) - (rT, S)x\|^2 + r^3\|x\|^2 | x \in H_1\}. \tag{10}$$

Denoting the solution of (10) by x_r, we have the estimate

$$\begin{aligned} r^3\|x_r\|^2 &\leqq \|(ry_0, 0) - (rT, S)x_r\|^2 + r^3\|x_r\|^2 - \min_{x \in H_1} \|(ry_0, 0) - (rT, S)x\|^2 \\ &\leqq \|(ry_0, 0) - (rT, S)(rT, S)^+(ry_0, 0)\|^2 + r^3\|(rT, S)^+(ry_0, 0)\|^2 \\ &\quad - \|(ry_0, 0) - (rT, S)(rT, S)^+(ry_0, 0)\|^2 \\ &\leqq r^3\|(rT, S)^+(ry_0, 0)\|^2, \end{aligned} \tag{11}$$

which gives, with (5),

$$\|x_r\| \leqq \|(rT, S)^+(ry_0, 0)\| \leqq i((T, S))^{-1}\|y_0\|, \tag{12}$$

and so the sequence $\{x_r\}_{r \to +0}$ is uniformly bounded. Let $\{x_{r_k}\}_{k \in N} \subset \{x_r\}_{r \to +0}$ be a subsequence converging weakly to an $\hat{x} \in H_1$:

$$x_{r_k} \rightharpoonup \hat{x}, \quad \text{for } r_k \to +0, \quad k \to \infty.$$

Putting

$$f_r(x) := \|(ry_0, 0) - (rT, S)x\|^2 = \|Sx\|^2 + r^2\|y_0 - Tx\|^2$$

and

$$g_r(x) := f_r(x) + r^3\|x\|^2,$$

we have for all $x \in N(S)$

$$\begin{aligned} r_k^3\|x_{r_k}\|^2 + r_k^2\|y_0 - Tx_{r_k}\|^2 &= g_{r_k}(x_{r_k}) - \|Sx_{r_k}\|^2 \\ &\leqq g_{r_k}(x_{r_k}) \leqq g_{r_k}(x) \\ &= r_k^2\|y_0 - Tx\|^2 + r_k^3\|x\|^2, \end{aligned} \tag{13}$$

which gives $\lim_{k \to \infty} g_{r_k}(x_{r_k}) = 0$, and so

$$\lim_{k \to \infty} \|Sx_{r_k}\| = 0. \tag{14}$$

Finally, dividing (12) by r_k^2, we have

$$\liminf_{k \to \infty} \|y_0 - Tx_{r_k}\| \leqq \|y_0 - Tx\|, \quad \text{for all } x \in N(S). \tag{15}$$

Since the occurring functions as convex and continuous functions are lower semicontinuous in the weak topology, it follows from (14) and (15)

$$\|y_0 - T\hat{x}\| \leqq \|y_0 - Tx\|, \quad \text{for all } x \in N(S),$$

$$\|S\hat{x}\| \leqq 0, \quad \text{i.e.} \quad \hat{x} \in N(S),$$

hence

$$\hat{x} \in T_S^+ y_0 + N(T_S), \tag{16}$$

the solution set of min $\{\|y_0 - Tx\| \mid x \in N(S)\}$. By (9), $x_{r_k} \in R((T, S)^*)$, which is assumed to be closed, and thus is also weakly closed, and so $\hat{x} \in R((T, S)^*) = N((T, S))^\perp$. Now $N(T_S) = N((T, S))$ and (16) gives

$$\hat{x} = T_S^+ y_0. \tag{17}$$

Because of (12) this yields, that the whole sequence $\{x_r\}_{r \to +0}$ converges weakly to $T_S^+ y_0$,

$$x_r \rightharpoonup T_S^+ y_0, \quad \text{for } r \to +0. \tag{18}$$

It remains to show the strong convergence. Similarly to (13) we have the estimations

$$\begin{aligned} r_k^2(\|y_0 - Tx_{r_k}\|^2 + \|Sx_{r_k}\|^2) &\leqq g_{r_k}(x_{r_k}) - (1 - r^2)\|Sx_{r_k}\|^2 \\ &\leqq g_{r_k}(x_{r_k}) \\ &\leqq r_k^2(\|y_0 - T\hat{x}\|^2 + \|S\hat{x}\|^2) + r_k^3\|\hat{x}\|^2, \end{aligned} \tag{19}$$

and

$$\begin{aligned} \|y_0 - T\hat{x}\|^2 + \|S\hat{x}\|^2 &\leqq \liminf_{k \to \infty} \{\|y_0 - Tx_{r_k}\|^2 + \|Sx_{r_k}\|^2\} \\ &\leqq \limsup_{k \to \infty} \{\|y_0 - Tx_{r_k}\|^2 + \|Sx_{r_k}\|^2\} \\ &\leqq \|y_0 - T\hat{x}\|^2 + \|S\hat{x}\|^2, \end{aligned} \tag{20}$$

thus

$$\lim_{k \to \infty} \{\|y_0 - Tx_{r_k}\|^2 + \|Sx_{r_k}\|^2\} = \|y_0 - T\hat{x}\|^2 + \|S\hat{x}\|^2. \tag{21}$$

Now splitting up

$$\|y_0 - Tx_{r_k}\|^2 + \|Sx_{r_k}\|^2 =: \langle x_{r_k}, (T^*T + S^*S)x_{r_k} \rangle + h(x_{r_k}),$$

where h is an affine functional on H_1, for which we have

$$h(x_{r_k}) \to h(\hat{x}), \quad \text{for } k \to \infty,$$

we get by (21)

$$\langle x_{r_k}, (T^*T + S^*S)x_{r_k} \rangle \to \langle \hat{x}, (T^*T + S^*S)\hat{x} \rangle, \quad \text{for } k \to \infty. \tag{22}$$

$x_{r_k}, \hat{x} \in R((T, S)^*)$ and so on $R((T, S)^*)$ we define the norm

$$|||\cdot||| := \langle \cdot\,, (T^*T + S^*S) \cdot \rangle^{1/2},$$

which is equivalent to

$$\|\cdot\|_{R((T,S)^*)},$$

since $i(T^*T+S^*S)>0$ if $i((T,S))>0$, and this is by (3) our assumption. Of course

$$x_{r_k}+\hat{x}\rightharpoonup 2\hat{x},\ \liminf_{k\to\infty} |||x_{r_k}+\hat{x}|||^2 \geqq |||2\hat{x}|||^2,$$

and the parallelogram law

$$|||x_{r_k}-\hat{x}|||^2 = 2|||x_{r_k}|||^2+2|||\hat{x}|||^2-|||x_{r_k}+\hat{x}|||^2$$

yields, with (22),

$$\lim_{k\to\infty} |||x_{r_k}-\hat{x}||| = 0.$$

The equivalence of the norms gives with (17), the strong convergence of $\{x_r\}_{r\to+0}$ to $T_S^+y_0$.

Thus with (7) and (9) we have

$$T_S^+y_0 = \lim_{r\to+0} (rT,S)^*((rT,S)(rT,S)^*+r^3I)^{-1}(ry_0,0), \tag{23}$$

and since (23) holds for every $y_0\in H_2$, we have proved the theorem.

If $S\equiv 0$, we get by (23) the following

COROLLARY 1. *Let T have closed range, then*

$$T^+ = \lim_{r\to+0} T^*(TT^*+rI_2)^{-1}, \qquad (r\in\mathbb{R}), \tag{24}$$

(in the sense of pointwise convergence on H_2) where I_2 is the identity on H_2.

The formula (24) (for square matrices T), tracing back to Den Broeder and Charnes [2], is cited as Theorem 5 by Ben-Israel and Charnes [1].

Estimating the rate of convergence, respectively deriving an error bound, we show now, that in (4) the operators also converge in norm.

THEOREM 2. *Let S and (T,S) have closed range, then*

$$\begin{aligned}\|T_S^+ &- (r^2T, rS)^*((rT,S)(rT,S)^*+r^3I)^{-1}_{|H_2\times\{0\}}\| \\ &\leqq r\left\{\frac{r}{1-r^2}\cdot\|(T,S)\|\cdot i((T,S))^{-1}\cdot i(S)^{-2}\cdot(\|T^*T\|\cdot i((T,S))^{-1}+\|T\|)\right. \\ &\quad\left. + i((T,S))^{-3}\cdot\left(1+\frac{r^2}{1-r^2}\cdot i(S)^{-1}\cdot\|(T,S)\|\right)\right\}, \qquad (0<r<1).\end{aligned} \tag{25}$$

Proof. Let

$$P_r := (rT,S)^*(rT,S)+r^3I_1,$$

where I_1 is the identity on H_1; then by (7), (8), (9), for all $y\in H_2$,

$$P_r^{-1}r^2T^*y = (rT,S)^*Q_r^{-1}(ry,0). \tag{26}$$

Let for a fixed $y\in H_2$ and $q,r\in\mathbb{R}$, $0<q,\ r<1$,

$$x_r = P_r^{-1}r^2T^*y, \quad\text{and}\quad x_q = P_q^{-1}q^2T^*y;$$

then by the identities

$$(P_r-P_q)x_q = (r^2-q^2)T^*Tx_q+(r^3-q^3)x_q,$$

$$P_rx_q = (P_r-P_q)x_q+q^2T^*y$$

we have

$$
\begin{aligned}
P_r(x_q - x_r) &= (P_r - P_q)x_q + q^2T^*y - r^2T^*y \\
&= q^2T^*y - q^2T^*Tx_q - q^3x_q - r^2T^*y + r^2T^*Tx_q + r^3x_q \qquad (27) \\
&= S^*Sx_q + r^2(T^*Tx_q - T^*y) + r^3x_q.
\end{aligned}
$$

By Theorem 1 and (26), x_q converges to $x_0 := T_S^+y$, as $q \to +0$, and (27) yields

$$P_r(x_0 - x_r) = S^*Sx_0 + r^2(T^*Tx_0 - T^*y) + r^3x_0. \qquad (28)$$

Now $x_0 \in N(S)$, that is,

$$Sx_0 = 0, \qquad (29)$$

and $\|Tx_0 - y\|^2 = \min_{x\in N(S)} \|Tx - y\|^2$, which implies that

$$T^*Tx_0 - T^*y \in N(S)^\perp.$$

Because $R(S)$ is assumed to be closed we have $N(S)^\perp = R(S^*)$ and S^+ exists. Now $z_0 := (S^*)^+(T^*Tx_0 - T^*y)$ satisfies the equation

$$S^*z_0 = T^*Tx_0 - T^*y \qquad (30)$$

and, by Lemma 1 and (3), has the property

$$\|z_0\| \leqq i(S)^{-1}(\|T^*T\| \cdot i((T, S))^{-1} + \|T\|)\|y\|. \qquad (31)$$

Since $x_0 \in R((T, S)^*)$, $(\hat{y}, \hat{z}) := ((T, S)^*)^+x_0$ satisfies

$$x_0 = (T, S)^*(\hat{y}, \hat{z}), \qquad (32)$$

and

$$\|(\hat{y}, \hat{z})\| \leqq i((T, S))^{-2}\|y\|. \qquad (33)$$

Using (29), (30) and (32), then (28) becomes

$$P_r(x_0 - x_r) = r^2S^*z_0 + r^3(T, S)^*(\hat{y}, \hat{z}). \qquad (34)$$

Now

$$
\begin{aligned}
\|P_r^{-1}(T, S)^*\| &= \sup\left\{\frac{\|P_r^{-1}(T, S)^*(y, z)\|}{\|(y, z)\|}\,\middle|\,(y, z) \in H_2 \times H_3\right\} \\
&= \sup\left\{\frac{\|P_r^{-1}T^*y + P_r^{-1}S^*z\|}{\|(y, z)\|}\,\middle|\,(y, z) \in H_2 \times H_3\right\} \\
&\leqq \sup\left\{\frac{\|P_r^{-1}T^*y\|}{\|(y, z)\|}\,\middle|\,(y, z) \in H_2 \times H_3\right\} \\
&\qquad + \sup\left\{\frac{\|P_r^{-1}S^*z\|}{\|(y, z)\|}\,\middle|\,(y, z) \in H_2 \times H_3\right\} \\
&\leqq \|P_r^{-1}T^*\| + \|P_r^{-1}S^*\|,
\end{aligned}
$$

From (34) we get

$$\|x_0 - x_r\| \leqq r^2\|P_r^{-1}S^*\| \cdot \|z_0\| + r^3(\|P_r^{-1}T^*\| + \|P_r^{-1}S^*\|) \cdot \|(\hat{y}, \hat{z})\|. \qquad (35)$$

By (12), $\|x_r\| \leqq i((T, S))^{-1} \cdot \|y\|$ and since $x_r = P_r^{-1}r^2T^*y$ it follows that

$$\|P_r^{-1}T^*\| \leqq r^{-2}i((T, S))^{-1}. \qquad (36)$$

Let $z \in H_3$ be fixed, then $v_r := P_r^{-1}S^*(1-r^2)z$ minimizes

$$(1-r^2)\|Sv-z\|^2+r^2\|Sv\|^2+r^2\|Tv\|^2+r^3\|v\|^2 \quad \text{over } H_1,$$

and we get

$$\begin{aligned} r^2(\|Sv_r\|^2+\|Tv_r\|^2)+(1-r^2)\|Sv_r-z\|^2 & \\ \leqq r^2(\|SS^+z\|^2+\|TS^+z\|^2)+(1-r^2)\|SS^+z-z\|^2 & \\ \leqq r^2(\|SS^+z\|^2+\|TS^+z\|^2)+(1-r^2)\|Sv_r-z\|^2, & \end{aligned}$$

which gives

$$\|(T,S)v_r\| \leqq \|(T,S)S^+z\|. \tag{37}$$

Since $P_rv_r = S^*(1-r^2)z$, we have

$$r^3v_r = (1-r^2)S^*z-(rT,S)^*(rT,S)v_r,$$

and thus $v_r \in R((T,S)^*) = N((T,S))^\perp$; hence for $v_r \neq 0$

$$\begin{aligned} \frac{\|(T,S)v_r\|}{\|v_r\|} &\geqq \inf\left\{\frac{\|(T,S)x\|}{\|x\|}\,\middle|\, x \in N((T,S))^\perp, x \neq 0\right\} \\ &\geqq i((T,S)), \end{aligned}$$

which with (37) gives the estimate. Also for $v_r = 0$,

$$\|v_r\| \leqq i((T,S))^{-1}\cdot\|(T,S)\|\cdot i(S)^{-1}\cdot\|z\|,$$

or

$$\|P_r^{-1}S^*\| \leqq \frac{1}{1-r^2}i((T,S))^{-1}\cdot i(S)^{-1}\cdot\|(T,S)\|. \tag{38}$$

Now (35) yields with (33), (36) and (38)

$$\begin{aligned} \|x_0-x_r\| \leqq r\Big\{&\frac{r}{1-r^2}i((T,S))^{-1}i(S)^{-1}\|(T,S)\|\cdot\|z_0\| \\ &+\left[i((T,S))^{-3}+\frac{r^2}{1-r^2}i((T,S))^{-3}\cdot i(S)^{-1}\|(T,S)\|\right]\|y\|\Big\}. \end{aligned} \tag{39}$$

Since $x_0 = T_S^+y$, and $x_r = (rT,S)^*Q_r^{-1}(ry,0)$ by (26), therefore (39) together with (31) gives the proposition (25).

Implicitly we have proved the following

COROLLARY 2. *If S and (T,S) have closed range, then*

$$T_S^+ = \lim_{r\to+0}((rT,S)^*(rT,S)+r^3I_1)^{-1}r^2T^*, \qquad (r \in \mathbb{R}), \tag{40}$$

where I_1 is the identity on H_1, and the estimation (25) *holds.*

For $S \equiv 0$ we get from (40)

$$T^+ = \lim_{r\to+0}(T^*T+rI_1)^{-1}T^* =: \lim_{r\to+0}T_r \tag{41}$$

and from (25)

$$\|T^+-T_r\| \leqq r\|T^+\|^3. \tag{42}$$

(41) and (42) were given by Morozov [5]. (40) is of particular interest for approximation problems over subspaces. Here often the space H_1 is finite dimensional and so the operators in (40) to be inverted then are positive definite symmetric matrices.

REFERENCES

[1] A. BEN-ISRAEL AND A. CHARNES, *Contributions to the theory of generalized inverses*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 667–699.

[2] G. G. DEN BROEDER, JR. AND A. CHARNES, *Contributions to the theory of a generalized inverse for matrices*, ONR Research Memo. No. 39, Northerwestern University, The Technological Institute, Evanston, Illinois, 1962.

[3] R. B. HOLMES, *A Course on Optimization and Best Approximation*, § 35: Generalized inverses, Lect. Not. Math., 257, Springer-Verlag, Berlin-Heidelberg-New York, 1972.

[4] N. MINAMIDE AND K. NAKAMURA, *A restricted pseudoinverse and its application to constrained minima*, SIAM J. Appl. Math., 19 (1970), pp. 167–177.

[5] V. A. MOROZOV, *Pseudo-Solutions*, U.S.S.R. Computational Math. and Math. Phys., 9 (1969), 6, pp. 196–203.

[6] W. PETRYSHYN, *On generalized inverses and on the uniform convergence of $(I-\beta K)^n$ with application to iterative methods*, J. Math. Anal. Appl., 18 (1967), pp. 417–439.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0006 $01.00/0

L^2 SOLUTIONS OF VOLTERRA INTEGRAL EQUATIONS*

T. KIFFE† AND M. STECHER†

Abstract. The existence of a unique $L^2[0, T; H]$ solution of the equation $u(t)+\int_0^t a(t-s)g(u(s))\,ds \ni f(t)$ is shown for any $L^2[0, T; H]$ function $f(t)$ where g is any maximal monotone operator satisfying a linear growth condition.

1. Introduction. We are interested in the question of the existence and uniqueness of solutions to the nonlinear Volterra equation

$$u(t)+\int_0^t a(t-s)g(u(s))\,ds \ni f(t), \qquad 0 \leqq t \leqq T. \tag{1.1}$$

The unknown function $u(t)$ may be scalar or vector valued, i.e., $u:[0, T] \to \mathbb{R}^n$, $(n \geqq 1)$. In fact our theorems will only assume that the range of u lies in some real Hilbert space.

In a recent paper by Levin [5], some a priori bounds for solutions to (1.1) are obtained where it is assumed that the range of u is contained in $\mathbb{R}^n$ $(1 \leqq n < \infty)$ and that the nonlinearity g is, among other things, assumed to be a continuous map from $\mathbb{R}^n$ to $\mathbb{R}^n$. These bounds are then used along with a local existence result of Nohel [7] to prove global existence. In the one dimensional case Levin [5, Thm. 1′] is able to deduce some a priori bounds by only assuming that g is nondecreasing and that f is differentiable. He is not, however, able to show existence in this case since he does not have a local existence result when g is not continuous.

One purpose of this paper is to obtain an existence result without imposing any continuity assumptions on g or differentiability conditions on f. The techniques which we use are those developed by Barbu [1], and Londen [6] who have proven existence and uniqueness theorems for (1.1) in a Hilbert space setting.

More precisely Barbu has shown that if the nonlinearity g is the subdifferential of a lower semicontinuous proper convex function, that $a(t)$ is a smooth real valued function of positive type, and that the forcing term $f(t)$ has an L^2 derivative, then (1.1) has a unique solution. Londen, using a different technique, has been able to prove that (1.1) has a unique solution without assuming $a(t)$ is of positive type. In fact Londen only assumes $a(0)>0$, and a' of bounded variation, but he too assumes $f(t)$ has an L^2 derivative and that g is a subdifferential. This last condition on g has been removed by Gripenberg [4]. In a recent paper [3] Crandall and Nohel were able to extend these existence results to an arbitrary Banach space with g an m-accretive operator.

The authors of this paper, using the techniques developed by Londen and Barbu, have been able to show that (1.1) has a unique solution for arbitrary f in L^2 and g maximal monotone. This generality however is obtained by drastically restricting the growth of g. This restriction rules out the possibility of g being a nonlinear partial differential operator, which the works of Barbu and Londen permit. It does however give new existence results in the finite dimensional case since we need not impose any continuity conditions on g.

The outline of this paper is as follows: Section 2 consists of notations, definitions, and the statements of our results whose proofs appear in §§ 3 and 4. Some examples are presented in § 5 which illustrate the applicability of our theorems.

---

* Received by the editors June 23, 1977, and in revised form September 13, 1977.

† Department of Mathematics, Texas A & M University, College Station, Texas 77843.

2. Statement and discussion of results. Throughout this paper the functions a and f will satisfy $a:R^1 \to R^1$, $f:R^1 \to H$ and g will denote a nonlinear possibly multiple-valued maximal monotone operator with its domain $D(g)$ and range $R(g)$ contained in a real Hilbert space H whose norm and inner product will be denoted by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ respectively. By a solution of (1.1) on the interval $[0, T]$ we mean a pair of functions $u, w:[0, T] \to H$ which satisfy

$$u, w \in L^2[0, T; H], \qquad w(t) \in g(u(t)) \quad \text{a.e.,} \tag{2.1}$$

$$u(t) + \int_0^t a(t-s)w(s)\,ds = f(t) \quad \text{a.e.,} \quad 0 \leqq t \leqq T. \tag{2.2}$$

$L^2[0, T; H]$ denotes the Hilbert space consisting of all functions $u:[0, T] \to H$ which satisfy $\int_0^T |u(t)|^2\,dt < \infty$.

We recall that a function $g: D(g) \subset H \to H$ is monotone if $\langle x_1 - x_2, y_1 - y_2 \rangle \geqq 0$ for all $y_1 \in g(x_1)$, $y_2 \in g(x_2)$, $x_1 \in D(g)$, $x_2 \in D(g)$, and g is maximal monotone if g has no proper monotone extension. The following closed graph property of maximal monotone operators is essential for the proof of the existence of solutions of (1.1).

PROPOSITION [2, Prop. 2.5]. *Suppose $\{x_n\}$ and $\{y_n\}$ are sequences in H which satisfy $x_n \in D(g)$ and $y_n \in g(x_n)$ for all n. If $x_n \rightharpoonup x$, $y_n \rightharpoonup y$ and* $\limsup_{n\to\infty} \langle x_n, y_n \rangle \leqq \langle x, y \rangle$, *then $x \in D(g)$ and $y \in g(x)$.* We use $\rightharpoonup$ to denote weak convergence and $\to$ strong convergence in H.

If g is maximal monotone it is well known that $(I + \lambda g)^{-1}$ is a contraction defined on all of H for each $\lambda > 0$ where I is the identity operator. We will let $J_\lambda = (I + \lambda g)^{-1}$ and $g_\lambda = \lambda^{-1}(I - J_\lambda)$, the Yosida approximation of g. In this paper we will need the facts that $g_\lambda(x) \in g(J_\lambda x)$ for all $x \in H$ and that g_λ is a Lipschitz function with constant λ^{-1}. We will denote by G the usual extension of g to $L^2[0, T; H]$, i.e., $v \in Gu$ if and only if $v \in L^2[0, T; H)$ and $v(t) \in g(u(t))$ a.e. on $[0, T]$. Then G is a maximal monotone operator on $L^2[0, T; H]$, and G_λ, the Yosida approximation of G, is just the extension of g_λ to $L^2[0, T; H]$.

In addition to the above preliminaries we will assume that $D(g) = H$ and there exist constants c_1 and c_2 such that either

$$|y| \leqq c_1 + c_2|x|, \quad \text{for all } y \in g(x),\ x \in H \tag{2.3}$$

or

$$|y| \leqq c_1|x|, \qquad c_2|x|^2 \leqq \langle x, y \rangle \quad \text{for all } y \in g(x),\ x \in H. \tag{2.4}$$

THEOREM 1. *Suppose* (2.3) *is satisfied and*

i) $a(0) > 0$, $a \in \mathrm{AC}[0, T]$, $a' \in \mathrm{BV}[0, T]$,

ii) $f \in L^2[0, T; H]$.

Then there exist unique functions u, w satisfying (2.1) *and* (2.2).

THEOREM 2. *Let the hypotheses of Theorem* 1 *be satisfied and suppose*

iii) *$\{f_n\}$ is a sequence in $L^2[0, T; H]$ such that $f_n \to f$ in $L^2[0, T; H]$. Then if u_n, w_n, u, and w satisfy* (1.1) *for f_n and f respectively, we have $u_n \to u$ and $w_n \to w$ in $L^2[0, T; H]$.*

In Theorems 1 and 2 we impose the same conditions on the kernel function $a(t)$ as Londen [6]. Our proofs of these theorems rely heavily on the ideas developed by Londen in [6, Thm. 1]. The essential difference between our approach to (1.1) and that used by Londen and Barbu is that we do not differentiate (1.1).

We can remove the differentiability condition on a and allow $a(0)=\infty$ and still prove existence for (1.1) if we require $a(t)$ to be of positive type on $[0, T]$.

DEFINITION. A function $a \in L^1(0, T)$ is of *positive type* if for every function $u \in L^2(0, T)$ we have

$$\int_0^t u(s) \int_0^s a(s-\tau)u(\tau)\,d\tau\,ds \geqq 0 \quad \text{for all } 0 \leqq t \leqq T. \tag{2.5}$$

The following characterization is due to Nohel and Shea [8, Thm. 2].

THEOREM. *Let $a(t)\,e^{-\sigma t} \in L^1(0, \infty)$ for all $\sigma > 0$ and let*

$$\hat{a}(s) = \int_0^\infty e^{-st} a(t)\,dt \quad \textit{for } \operatorname{Re} s > 0.$$

The following are equivalent,

iv) $a(t)$ *is of positive type on* $[0, \infty)$,

v) $\operatorname{Re} \hat{a}(s) \geqq 0$ *for* $\operatorname{Re} s > 0$,

vi) $U(i\tau) \geqq 0 (-\infty < \tau < \infty)$ *where* $U(i\tau) = \liminf_{\substack{s \to i\tau \\ \operatorname{Re} s > 0}} \operatorname{Re} \hat{a}(s)$.

Functions of the form $a(t) = t^{-\alpha}\,e^{-\beta t} \cos \gamma t (0 \leqq \alpha < 1, \beta \geqq 0, \gamma$ real) are of positive type.

THEOREM 3. *Let* (2.3) *and* ii) *be satisfied and suppose that $a(t)$ is of positive type on* $[0, T]$. *Then there exist functions $u(t)$ and $w(t)$ satisfying* (2.1) *and* (2.2). *If g is strictly monotone*, i.e., $\langle y_1 - y_2, x_1 - x_2 \rangle = 0$ *implies* $x_1 = x_2$, $(y_i \in g(x_i))$, *then $u(t)$ is unique.*

Our next result gives an asymptotic property of solutions of (1.1) if we impose further restrictions on g.

THEOREM 4. *Let* (2.4) *be satisfied. Suppose $a(t)$ is of positive type on* $[0, \infty)$ *and* $f \in L^2[0, \infty; H]$. *Then there exists a unique function u and a function w, both in* $L^2[0, \infty; H]$, *which satisfy* (2.1) *and* (2.2). The proofs of Theorems 3 and 4 will use techniques developed by Barbu [1].

The following lemma which Londen proved in [6] is stated to aid in the exposition of our results.

LEMMA. *Let $a(t)$ satisfy* i) *of Theorem* 1. *Let $\{u_n\}$ be a sequence in* $L^2[0, T; H]$ *such that $\|u_n\|_{L^2[0,T;H]}$ are uniformly bounded. Then either $\int_0^t a(t-s)u_n(s)\,ds$ converges uniformly on $[0, T]$ to zero as n tends to infinity, or there is a $\hat{t}$, $0 < \hat{t} < T$, such that*

$$\limsup_{n \to \infty} \int_0^{\hat{t}} \left\langle \int_0^\tau a(\tau - s)u_n(s)\,ds, u_n(\tau) \right\rangle d\tau > 0. \tag{2.6}$$

3. Proofs of Theorems 1 and 2. To begin the proof of Theorem 1 let $u_\lambda(t)$ denote the unique $L^2[0, T; H]$ solution of

$$u_\lambda(t) + \int_0^t a(t-s) g_\lambda(u_\lambda(s))\,ds = f(t), \qquad 0 \leqq t \leqq T, \quad \lambda > 0. \tag{3.1}$$

Since the Yosida approximates g_λ of g are Lipschitz continuous the existence and uniqueness of the u_λ is immediate. (By $\|h\|_{L^p[0,T;H]}$ we will mean the usual L^p norm of h as a mapping from $[0, T]$ into H.) Pick γ small enough so that $c_2\|a\|_{L^1[0,\gamma;\mathbb{R}]} \leqq \frac{1}{2}$, where c_2 is the constant in (2.3). We then have from (3.1),

$$\begin{aligned}\|u_\lambda\|^2_{L^2[0,\gamma;H]} &\leqq \|a\|_{L^1}\|g_\lambda(u_\lambda)\|_{L^2[0,\gamma]}\|u_\lambda\|_{L^2[0,\gamma]}+\|f\|_{L^2[0,\gamma]}\|u_\lambda\|_{L^2[0,\gamma]}\\ &\leqq \|a\|_{L^1}(c_1\sqrt{\gamma}+c_2\|u_\lambda\|_{L^2})\|u_\lambda\|_{L^2}+\|f\|_{L^2}\|u_\lambda\|_{L^2}\\ &=(\|a\|_{L^1}c_1\sqrt{\gamma}+\|f\|_{L^2})\|u_\lambda\|_{L^2}+c_2\|a\|_{L^1}\|u_\lambda\|^2_{L^2}.\end{aligned} \tag{3.2}$$

Thus we have

$$\|u_\lambda\|_{L^2[0,\gamma;H]}\leqq 2(c_1\sqrt{\gamma}\|a\|_{L^1[0,\gamma;\mathbb{R}]}+\|f\|_{L^2[0,\gamma;H]}). \tag{3.3}$$

Since the length of the interval on which we have a bound for $\|u_\lambda\|$ is determined by $a(t)$ and c_2 and not the norm of f we are able to translate and obtain a bound for $\|u_\lambda\|$ on $[0, T]$, i.e., there is an $M>0$ such that

$$\|u_\lambda\|_{L^2[0,T;H]}\leqq M, \qquad \lambda>0. \tag{3.4}$$

The constant M depends on $\|a\|_{L^1[0,\gamma;\mathbb{R}]}$, c_2, $\|f\|_{L^2[0,T;H]}$ and how many integer multiples of γ are less than or equal to T. From the fact that $|g_\lambda(u_\lambda(t))|\leqq|g(u_\lambda(t))|$ and (2.3) we see that the norms of $g_\lambda(u_\lambda)$ are also uniformly bounded in λ. We may therefore extract subsequences u_{λ_n} and $g_{\lambda_n}(u_{\lambda_n})$ such that

$$u_{\lambda_n}\rightharpoonup u, \qquad g_{\lambda_n}(u_{\lambda_n})\rightharpoonup w \tag{3.5}$$

in $L^2[0, T; H]$. Clearly u and w satisfy (2.2). To establish existence it only remains to show that $w(t)\in g(u(t))$ a.e. As in Londen's paper we claim that $\int_0^t a(t-s)g_{\lambda_n}(u_{\lambda_n}(s))\,ds$ converges uniformly on $[0, T]$ to $\int_0^t a(t-s)w(s)\,ds$. Suppose not; then there are subsequences $g_{\lambda_{n_k}}$ and $g_{\lambda_{m_k}}$ denoted by g_{n_k} and g_{m_k} respectively such that $\int_0^t a(t-s)[g_{n_k}(u_{n_k}(s))-g_{m_k}(u_{m_k}(s))]\,ds$ does not converge uniformly to zero. Putting g_{n_k} and then g_{m_k} into (3.1), subtracting the two equations, multiplying by $g_{n_k}-g_{m_k}$, and then integrating from 0 to $\hat{t}$, we have,

$$\begin{aligned}&\int_0^{\hat{t}}\langle u_{n_k}(t)-u_{m_k}(t),\, g_{n_k}(u_{n_k}(t))-g_{m_k}(u_{m_k}(t))\rangle\,dt\\ &\quad=-\int_0^{\hat{t}}\left\langle\int_0^t a(t-s)[g_{n_k}(s)-g_{m_k}(s)]\,ds,\, g_{n_k}(t)-g_{m_k}(t)\right\rangle dt,\end{aligned} \tag{3.6}$$

where $\hat{t}$ is as in the Lemma. Using the fact that $u_{n_k}=\lambda_{n_k}g_{n_k}(u_{n_k})+J_{\lambda_{n_k}}u_{n_k}$, $g_{n_k}(u_{n_k})\in g(J_{\lambda_{n_k}}(u_{n_k}))$, and that g is monotone, we derive from (3.6) the following inequality:

$$\begin{aligned}&\int_0^{\hat{t}}\langle\lambda_{n_k}g_{n_k}-\lambda_{m_k}g_{m_k},\, g_{n_k}-g_{m_k}\rangle\,dt\\ &\quad\leqq-\int_0^{\hat{t}}\left\langle\int_0^t a(t-s)[g_{n_k}(u_{n_k}(s))-g_{m_k}(u_{m_k}(s))]\,ds,\, g_{n_k}(u_{n_k}(t))-g_{m_k}(u_{m_k}(t))\right\rangle dt.\end{aligned} \tag{3.7}$$

Letting $k\to\infty$ we get from the lemma that $0<0$. This contradiction implies the uniform convergence of the integral terms $\int_0^t a(t-s)g_{\lambda_n}(u_{\lambda_n}(s))\,ds$ on $[0, T]$. This result along with (3.1) implies that $u_{\lambda_n}\to u$. Hence we have by [2, Prop. 2.5], applied to $L^2[0, T; H]$ that $u(t)\in D(g)$ and $w(t)\in g(u(t))$ a.e. To see that u and w are unique, suppose $\hat{u}$ and $\hat{w}$ is also a solution. This implies

$$u(t)-\hat{u}(t)+\int_0^t a(t-s)[w(s)-\hat{w}(s)]\,ds=0. \tag{3.8}$$

Multiplying (3.8) by $w-\hat{w}$ we conclude, by arguing as above, that the integral term must be zero on $[0, T]$. This of course gives $u(t)=\hat{u}(t)$ a.e. and $w(t)=\hat{w}(t)$ a.e.

To prove Theorem 2 let u_n, w_n and u, w be solutions of (1.1) for f_n and f respectively. We first note that $\|u_n\|_{L^2[0,T;H]}$ are bounded independent of n. This follows as in the derivation of (3.4). We also have

$$u_n(t)-u(t)+\int_0^t a(t-s)[w_n(s)-w(s)]\,ds=f_n(t)-f(t). \tag{3.9}$$

We claim as in Theorem 1 that the integral term converges uniformly to zero on the interval $[0, T]$. To see this one argues as in Theorem 1 using the assumption that $f_n\to f$. Clearly the uniform convergence of the integral term implies that $u_n\to u$. To see that $w_n \rightharpoonup w$ we note that any subsequence of the w_n must have a weakly convergent subsequence w_{n_k}. Moreover the w_{n_k} and u_{n_k} must then converge to the unique solution of (1.1). Thus $w_{n_k}\rightharpoonup w$, which implies $w_n\rightharpoonup w$.

4. Proof of Theorems 3 and 4. Choose $\gamma>0$ as in the proof of Theorem 1. Again letting u_λ denote the $L^2[0, T; H]$ solution of (3.1) we have that $\{u_\lambda\}$ and $\{g_\lambda(u_\lambda)\}$ are uniformly bounded in $L^2[0,\gamma; H]$. Hence we can extract sequences $\{u_{\lambda_n}\}$ and $\{g_{\lambda_n}(u_{\lambda_n})\}$ which satisfy (3.5) for some $u, w\in L^2[0,\gamma; H]$. To show that $u(t)\in D(g)$ and $w(t)\in g(u(t))$ a.e. $0\leqq t\leqq\gamma$ we follow Barbu [1, Thm. 1]. Write (3.1) with $\lambda=\lambda_n$ and $\lambda=\lambda_m$, subtract, multiply by $g_{\lambda_n}(u_{\lambda_n})-g_{\lambda_m}(u_{\lambda_m})$ and integrate. Since $a(t)$ is of positive type we get

$$\int_0^\lambda \langle g_{\lambda_n}(u_{\lambda_n}(t))-g_{\lambda_m}(u_{\lambda_m}(t)),\, u_{\lambda_n}(t)-u_{\lambda_m}(t)\rangle\,dt\leqq 0. \tag{4.1}$$

Expanding (4.1) we have

$$\begin{aligned}&\int_0^\gamma \langle g_{\lambda_n}(u_{\lambda_n}(t)), u_{\lambda_n}(t)\rangle dt\\ &\leqq -\int_0^\gamma \langle g_{\lambda_m}(u_{\lambda_m}(t)), u_{\lambda_m}(t)\rangle\,dt+\int_0^\gamma \langle g_{\lambda_n}(u_{\lambda_n}(t)), u_{\lambda_m}(t)\rangle\,dt\\ &\qquad+\int_0^\gamma \langle g_{\lambda_m}(u_{\lambda_m}(t)), u_{\lambda_n}(t)\rangle\,dt.\end{aligned} \tag{4.2}$$

Fix λ_m. Then, by (3.5), we have

$$\begin{aligned}&\limsup_{\lambda_n\to 0}\int_0^\gamma \langle g_{\lambda_n}(u_{\lambda_n}(t)), u_{\lambda n}(t)\rangle\,dt\\ &\leqq -\int_0^\gamma \langle g_{\lambda_m}(u_{\lambda_m}(t)), u_{\lambda_m}(t)\rangle\,dt+\int_0^\gamma \langle w(t), u_{\lambda_m}(t)\rangle\,dt\\ &\qquad+\int_0^\gamma \langle g_{\lambda_m}(u_{\lambda_m}(t)), u(t)\rangle\,dt.\end{aligned} \tag{4.3}$$

Letting $\lambda_m\to 0$ in (4.3) we have

$$\limsup_{\lambda_n\to 0}\int_0^\gamma \langle g_{\lambda_n}(u_{\lambda_n}(t)), u_{\lambda_n}(t)\rangle\,dt\leqq\int_0^\gamma \langle w(t), u(t)\rangle\,dt. \tag{4.4}$$

Combining (3.5) and (4.4) with [2, Prop. 2.5] applied to $L^2[0,\gamma; H]$ we get $u(t)\in D(g)$ and $w(t)\in g(u(t))$. The fact that $u(t)$ and $w(t)$ are solutions of (1.1) on $[0,\gamma]$ follows from exactly the same argument used in Theorem 1. If g is strictly monotone, the uniqueness of $u(t)$ is an immediate consequence of the assumption that $a(t)$ is of positive type. The continuation arguments of Barbu and Londen can now be applied to extend $u(t)$ and $w(t)$ to the whole interval $[0, T]$. This completes the proof of Theorem 3.

To prove Theorem 4 multiply (3.1) by $g_\lambda(u_\lambda(t))$ and integrate. By (2.4) and Young's inequality we have

$$c\int_0^T |u_\lambda(t)|^2\,dt \leqq \frac{1}{2\alpha}\int_0^T |f(t)|^2\,dt + \frac{\alpha}{2}\int_0^T |g_\lambda(u_\lambda(t))|^2\,dt, \qquad 0<\lambda\leqq 1. \tag{4.5}$$

We note that (2.4) holds for $g_\lambda (0<\lambda \leqq 1)$ with c_2 replaced by $c = c_2/(1+c_1)^2$. Choose α so that $c_1\alpha < c$. By (2.4) we have that there is a constant M such that

$$\int_0^T |u_\lambda(t)|^2\,dt \leqq M\int_0^T |f(t)|^2\,dt. \tag{4.6}$$

Since T was arbitrary, and $f\in L^2[0, \infty; H]$ we have

$$\sup_{0<\lambda\leqq 1}\int_0^\infty |u_\lambda(t)|^2\,dt < \infty. \tag{4.7}$$

Combining (2.4) and (4.7) we have

$$\sup_{0<\lambda\leqq 1}\int_0^\infty |g_\lambda(u_\lambda(t))|^2\,dt < \infty. \tag{4.8}$$

From (4.7) and (4.8) we can conclude that there is a sequence $\{\lambda_n\}$ converging to zero and functions $u, w \in L^2[0, \infty; H]$ such that

$$\begin{aligned} u_{\lambda_n} &\rightharpoonup u && \text{in } L^2[0, \infty; H] \\ g_{\lambda_n}(u_{\lambda_n}) &\rightharpoonup w && \text{in } L^2[0, \infty; H]. \end{aligned} \tag{4.9}$$

By the same argument used in Theorem 3 (lines (4.1)–(4.4) with γ replaced by arbitrary T) we get $u(t)\in D(g)$ a.e. and $w(t)\in g(u(t))$ a.e. for $0\leqq t<\infty$. The fact that $u(t)$ and $w(t)$ are solutions of (1.1) follows as before. To prove that $u(t)$ is unique, let u_1, w_1, u_2, w_2 be solutions of (1.1). Subtract these equations, multiply by $w_1 - w_2$, integrate and use (2.4). We have

$$c_2\int_0^\infty |u_1(t)-u_2(t)|^2\,dt \leqq 0 \tag{4.10}$$

so $u_1 = u_2$ a.e. This completes the proof of Theorem 4.

5. Examples. Our first example demonstrates that if we remove the linear growth condition, then (1.1) will not have a solution, in the standard sense, for arbitrary $f\in L^2[0, T; H]$. This example seems to indicate that this definition of a solution is too narrow.

Example 1. Let $H=\mathbb{R}$, $g(u)=|u|u$, $f(t)=t^{-1/4}$, $a(t)=1$. Clearly $f\in L^2[0, T]$. Suppose (1.1) had a solution. That is, there exists a function $u(t)$ such that

$$\begin{gathered} u\in L^2[0, T], \qquad g(u)=|u|u \quad \in L^2[0, T], \quad \text{and} \\ u(t)+\int_0^t |u(s)|u(s)\,ds = t^{-1/4}. \end{gathered} \tag{5.1}$$

Since $g(u) \in L^2$, we have $u \in L^4$ which implies that $f = t^{-1/4} \in L^4$ which is absurd. We also note that if $g(u) = |u|^\alpha u$, $\alpha > 0$, then (1.1) will not have a solution for an arbitrary function f in $L^2[0, T]$.

We next give a simple example of a discontinuous g for which our theorems apply.

Example 2. Let $H = \mathbb{R}$, $a(t) = t^{-1/2}$, $\hat{g}(u) = [u]$, where $[\cdot]$ denotes the greatest integer function. The function $\hat{g}$ has linear growth and is monotone but it is not maximal in the class of monotone operators. Thus let $g(u)$ be its maximal extension, i.e.,

$$g(u) = \begin{cases} \hat{g}(u), & u \notin N, \\ [n-1, n], & u = n \in N, \end{cases} \tag{5.2}$$

where N is the set of integers. Then for any L^2 function $f(t)$ equation (5.3) has a solution.

$$u(t) + \int_0^t (t-s)^{-1/2}[u(s)]\, ds \ni f(t). \tag{5.3}$$

It is understood that if $u(t_0) = n \in N$ for some t_0 then $[u(t_0)]$ may be any number between $n-1$ and n. Since $g(u)$ is not strictly monotone (cf. Theorem 3) we are not able to claim that (5.3) has a unique solution.

REFERENCES

[1] V. BARBU, *Nonlinear Volterra equations in a Hilbert space*, this Journal, 6 (1975) pp. 728–741.

[2] H. BREZIS, *Operateurs maximaux monotones et semi-groups de contractions dans espaces de Hilbert*, Math. Studies 5, North-Holland, Amsterdam, 1973.

[3] M. G. CRANDALL AND J. A. NOHEL, *An abstract functional differential equation and a related nonlinear Volterra equation*, Israel J. Math., to appear.

[4] G. GRIPENBERG, *An existence result for a nonlinear Volterra integral equation in Hilbert space*, this Journal, to appear.

[5] J. J. LEVIN, *Some a priori bounds for nonlinear Volterra equations*, this Journal, 7(1976), pp. 872–897.

[6] S. O. LONDEN, *On an integral equation in a Hilbert space*, this Journal, 8 (1977), pp. 950–970.

[7] J. A. NOHEL, *Some problems in nonlinear Volterra integral equations*, Bull. Amer. Math. Soc., 68 (1962), pp. 323–330.

[8] J. A. NOHEL AND D. F. SHEA, *Frequency domain methods for Volterra equations*, Advances in Math., 22 (1976), pp. 278–303.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0007 $01.00/0

ON THE INVERSE PROBLEM FOR SELFADJOINT OPERATORS DEFINED ON CERTAIN RIGGED HILBERT SPACES*

JOHN B. BUTLER, JR.†

Abstract. We consider the inverse problem for a selfadjoint extension H^1 of a differential operator $L^1=(i(d/dx))^n+q(x)$ defined on the rigged Hilbert space $\Phi\subseteq\mathfrak{h}\subseteq\Phi'$ where $\mathfrak{h}=L_2(l_2:R)$, $\Phi=C_0^\infty(l_2:R)$ and $q(x)$ is a selfadjoint operator valued function of x which is piecewise continuous in the l_2 norm. It is shown using the methods of I. M. Gel'fand, B. M. Levitan that $q(x)$ may be reconstructed if we are given (n, n) matrices $V^1(l\pm i0)$ whose elements are operator valued functions of l which are associated with the spectral density measure for H^1 and the scattering operator corresponding to H^1. The results extend to operators of the form $\tilde{H}^1=T^n+Q$, T, Q selfadjoint, defined on a rigged Hilbert space $\tilde{\Phi}\subseteq\tilde{\mathfrak{h}}\subseteq\tilde{\Phi}'$ in case $\|T\phi\|\geqq\|\phi\|$, $\phi\in\tilde{\Phi}$, since under conditions given by Ju. M. Berezanskii such operators have a representation as extensions of L^1 on a space of distributions $W^{-(m,p)}(l_2:R)$ containing $L_2(l_2:R)$.

1. Introduction. Let T, Q be operators defined on a rigged Hilbert space $\tilde{\Phi}\subseteq\tilde{\mathfrak{h}}\subseteq\tilde{\Phi}'$ where T is selfadjoint, $\|Tu\|\geqq\|u\|$, $u\in\tilde{\Phi}$, and Q is bounded selfadjoint. In this paper we shall consider the problem of determination of the operator Q assuming that the spectral density measure of the operator (T^n+Q) has been given. Ju. M. Berezanskii has given sufficient conditions such that an operator T with the above properties is similar to an extension of the operator $(i(d/dx))$ on a representation space $W^{-(m,p)}(l_2:R)$ of distributions over $L_2(l_2:R)$ [2, pp. 64, 663]. Assuming such a representation to hold we consider the inverse problem for an operator H^1 which is a selfadjoint extension of a differential operator of the form $L^1=(i(d/dx))^n+q(x)$ defined on the rigged Hilbert space $\Phi\subseteq\mathfrak{h}\subseteq\Phi'$ where $\mathfrak{h}=L_2(l_2:R)$, $\Phi=C_0^\infty(l_2:R)$ and $q(x)$ is an operator valued function of x which is piecewise continuous in the l_2 norm. The operator $q(x)$ in the representation space corresponds to a bounded operator Q in the original space when bounded in the $L_2(l_2:R)$ norm. It is shown that the procedure employed by I. M. Gel'fand, B. M. Levitan to solve the inverse problem for differential operators in $L_2[0,\infty)$ may be extended to the solution of the above problem. For $n>2$ we also use a generalization of the Gel'fand–Levitan algorithm due to L. A. Sahnovič, and certain results obtained by the writer for operators defined on $L_2(R)$ [3], [4], [13], [14]. The operator $q(x)$ is reconstructed assuming that one is given (n, n) matrices $V^1(l\pm i0)$ whose elements are operator valued functions of l which are related to the spectral density measure for H^1 and to the scattering matrix for H^1.

In § 2 of the paper the spectral theory of the operator H^1 is outlined and the matrices $V^1(l\pm i0)$ are defined. Sufficient conditions such that the inverse procedure can be applied and a description of the inverse procedure are given in § 3. The sufficient conditions are analogous to those employed by I. M. Gel'fand, B. M. Levitan in their paper dealing with differential operators of second order on $L_2(0,\infty)$ [7]. The main result of the paper is contained in formula (3.2) which states the relationship between $q(x)$ and the generalized translation kernel $k^1(x, y)$. As is usual in inverse theory $q(x)$ is obtained as a generalized multiplication operator. Formula (3.2) is a generalization of the formula given by I. Kay, H. E. Moses for differential operators on $L_2(0,\infty)$, $n=2$, and by L. A. Sahnovič for differential operators on $L_2(0,\infty)$, $n>2$ [9], [13], [14]. Proof that the inverse procedure is effective under the stated conditions is given

* Received by the editors July 5, 1973, and in final revised form February 2, 1978. This paper extends results presented to the American Mathematical Society, Abstract 71T-B213, Notices of the American Mathematical Society, Vol. 18, No. 4 Issue 130 (1971) 643.

† Department of Mathematics, Portland State University, Portland, Oregon 97207.

in Theorems 1 and 2. In § 4 of the paper we present an example in which $n=4$ and $q(x)$ has a diagonal form. Further references to the inverse theory for differential operators on $L_2(0,\infty)$ and $L_2(R)$ spaces are given in [1], [5], [11], [15].

2. Spectral theory of differential operators on $L_2(l_2:R)$. Let $\mathfrak{h}=L_2(l_2:R)$ denote the Hilbert space of vector functions $u(x)=(u_1(x), u_2(x),\cdots)$ on l_2 with inner product (u, v) defined by

$$(u, v)=\int_{\infty}^{\infty}(u(x), v(x))_{l_2}\,dx. \tag{2.1}$$

$(u, v)_{l_2}$ denotes the inner product on l_2, $\|u\|_{l_2}$ the vector or operator norms on l_2 and $\|u\|$ the norm in $\mathfrak{h}$ [12]. Let $C_0^\infty(l_2:R)$ denote the set of infinitely differentiable vector functions in $L_2(l_2:R)$ which vanish outside a compact set and let $\Phi=D(l_2:R)$ be the nuclear space consisting of vectors in $C_0^\infty(l_2:R)$ with topology generated by the norm $\|u\|$ together with the sequence of Sobolev norms $\{\|u\|_n\}$ where $\|u\|_n^2=(u,u)_n$ and $(u,v)_n=\sum_{j=1}^{\infty}2^{nj}\sum_{k=0}^{n}\int_{-\infty}^{\infty}p^n u_j^{(k)}(x)p^n v_j^{(k)}(x)\,dx$, $n=1,2,\cdots$, $u, v\in C_0^\infty(l_2:R)$, $p=\exp(x^2)$ [2, p. 59], [8, pp. 82, 108]. Let $W_2^{(n,p)}(l_2:R)$ be the Sobolev spaces obtained by completion of $C_0^\infty(l_2:R)$ with respect to the norms $\|u\|_n$ and let $W_2^{-(n,p)}(l_2:R)$ be the duals of these spaces [2, Chap. 1], [5, Chap. 1]. The triple of spaces $\Phi\subseteq\mathfrak{h}\subseteq\Phi'$ forms a rigged Hilbert space, Φ is the intersection of the spaces $\mathfrak{h}$ and $W_2^{(n,p)}$, and Φ' is the union of the $W_2^{-(n,p)}$. For $\phi\in\Phi$, $\phi'\in\Phi'$ write $\phi'(\phi)=\langle\phi',\phi\rangle$. Nuclear kernels $P\in\Phi'\otimes\Phi'$ are linear mappings from Φ into Φ' such that the bilinear form $B(\phi,\psi)=\langle P(\phi),\psi\rangle$ is continuous for $\phi,\psi\in\Phi$. If $P\in\Phi'\otimes\Phi'$ and A_1, A_2 are selfadjoint operators on $\mathfrak{h}$ whose domains contain Φ, $i=1,2$ then $(A_1\otimes A_2)P$ is the linear operator from Φ to Φ' such that $\langle(A_1\otimes A_2)P(\phi),\psi\rangle=\langle P(A_2\phi), A_1\psi\rangle$, $\phi,\psi\in\Phi$. For A_1, A_2 powers of $(i(d/dx))$ we shall write $(A_1\otimes A_2)P=(i(d/dx))^r\otimes(i(d/dx))^sP=i^{r+s}(\partial^{r+s}/\partial x^r\,\partial y^s)P=i^{r+s}D_x^rD_y^sP$. A kernel $P\in\Phi'\otimes\Phi'$ is said to $*$-commute with an operator A iff $(A\otimes I)P=(I\otimes A)P$. $\mathfrak{h}\otimes\mathfrak{h}$ is the set of those nuclear kernels which can be extended to Hilbert–Schmidt operators from $\mathfrak{h}$ into $\mathfrak{h}$. If $P\in\mathfrak{h}\otimes\mathfrak{h}$ is a kernel such that $D_x^iD_y^jP\in\mathfrak{h}\otimes\mathfrak{h}$, $i,j=0,1$ then P is an integral operator of Fredholm type: There exists a linear mapping $P(x,y)$ from l_2 into l_2, $P(x,y)\in C(l_2:R\times R)$ such that

$$\langle P(\phi),\psi\rangle=\left(\psi(x),\int_{\infty}^{\infty}P(x,y)\phi(y)\,dy\right),\qquad \phi,\psi\in\Phi$$

[2, p. 48, 8, p. 11].

For such kernels we identify P and $P(x,y)$. Let δ, η denote the kernels associated with the delta distribution and the Heaviside distribution. These kernels by definition satisfy $\langle\delta(\phi),\psi\rangle=(\psi(x),\phi(x))$, $\langle\eta(\phi),\psi\rangle=(\psi(x),\int_{-\infty}^{\infty}\chi_{[x,\infty)}(y)\phi(y)\,dy)$ for $\phi,\psi\in\Phi$ where $\chi_{[x,\infty)}(y)$ is the characteristic function of $[x,\infty)$. We also use the notations δ^ν, η^ν for kernels such that $\langle\delta^\nu(\phi),\psi\rangle=(\psi(x),\ \phi(-x))$ and $\langle\eta^\nu(\phi),\psi\rangle=(\psi(x),\int_{\infty}^{\infty}\chi_{[-x,\infty)}(y)\phi(y)\,dy)$ [6, p. 71]. If $P\in\mathfrak{h}\otimes\mathfrak{h}$ is an integral operator with kernel function $P(x,y)\in C(l_2:R\times R)$ then $(\delta\cdot P)$, $(\eta\cdot P)$ are defined to be kernels such that $\langle(\delta\cdot P)(\phi),\psi\rangle=(\psi(x),\ P(x,x)\phi(x))$ and $\langle(\eta\cdot P)(\phi),\psi\rangle=(\psi(x),\int_{\infty}^{\infty}\chi_{[x,\infty)}(y)P(x,y)\phi(y)\,dy)$, $\phi,\psi\in\Phi$. Kernels $(\delta^\nu\cdot P)$, $(\eta\,\nu\cdot P)$ are defined similarly. Using convolution one may define a dot product of nuclear kernels in general following the methods of [6, p. 73] but we do not require this degree of generality here.

The operator (iD) acts as a right and left derivation with respect to the dot products $(\delta\cdot P)$, $(\eta\cdot P)$ defined in the last paragraph and it follows directly from the definitions that the following lemma is valid:

LEMMA 1. *If* $D_x^j D_y^k P \in \mathfrak{h}\otimes\mathfrak{h}$, $j, k = 0, 1, 2$ *then*

(i) $D_x\eta = -\delta$, $D_y\eta = \delta$, $D_x\delta = -D_y\delta$,

(ii) $D_x(\eta \cdot P) = -\delta \cdot P + \eta \cdot D_xP$, $D_y(\eta \cdot P) = \delta \cdot P + \eta \cdot D_yP$,

(iii) $(D_x + D_y)(\delta \cdot P) = \delta \cdot (D_x + D_y)P$.

Also (i)′ $D_x\eta^\nu = \delta^\nu$, $D_y\eta^\nu = \delta^\nu$, $D_x\delta^\nu = D_y\delta^\nu$,

(ii)′ $D_x(\eta^\nu \cdot P) = \delta^\nu \cdot P + \eta^\nu \cdot D_xP$, $D_y(\eta^\nu \cdot P) = \delta^\nu \cdot P + \eta^\nu \cdot D_yP$,

(iii)′ $(D_x - D_y)(\delta^\nu \cdot P) = \delta^\nu \cdot (D_x + D_y)P$.

Extending this to higher powers leads to:

LEMMA 2. *Let* $P \in \mathfrak{h}\otimes\mathfrak{h}$ *be a nuclear kernel such that* $D_x^j D_y^k P \in \mathfrak{h}\otimes\mathfrak{h}$, $j, k = 0, 1, \cdots, m$ *and* $(\delta \cdot D_y^j P) = 0$, $j = 0, 1, \cdots, m-3$. *Then the following are valid*:

(a) $\delta \cdot D_x^j D_y^k P = 0$, $j + k = 0, 1, \cdots, m-3$;

(b) $\delta \cdot D_x^{j+1} D_y^k P = -\delta \cdot D_x^j D_y^{k+1} P$, $j + k = m-3$, $j, k = 0, 1, \cdots, m-3$;

(c) $\delta \cdot (D_x^{j+2} D_y^k + D_x^{j+1} D_y^{k+1})P = -\delta \cdot (D_x^{j+1} D_y^{k+1} P + D_x^j D_y^{k+2} P)$, $\quad j + k = m-3$, $j, k = 0, 1, 2, \cdots, m-3$.

Also if $\delta^\nu \cdot D_y^j P = 0$, $j = 0, 1, \cdots, m-3$, *then*

(a)′ $\delta^\nu \cdot D_x^j D_y^k P = 0$, $j + k = 0, 1, \cdots, m-3$;

(b)′ $\delta^\nu \cdot D_x^{j+1} D_y^k P = -\delta^\nu \cdot D_x^j D_y^{k+1} P$, $j + k = m-3$;

(c)′ $\delta^\nu \cdot (D_x^{j+2} D_y^k + D_x^{j+1} D_y^{k+1})P = \delta^\nu \cdot (D_x^{j+1} D_y^{k+1} + D_x^j D_y^{k+2})P$, $j + k = m-3$; $j, k = 0, 1, 2, \cdots, m-3$.

Proof. The proof of (a) is by a recursive argument. Suppose that $\delta \cdot D_x^j D_y^k P = 0$, $j + k = 0, 1, \cdots, r$, $0 \leqq r < m-3$. Then by (iii) of Lemma 1 for $j + k = r$, $(D_x + D_y)(\delta \cdot D_x^j D_y^k P) = \delta \cdot (D_x^{j+1} D_y^k + D_x^j D_y^{k+1})P = 0$. Since by hypothesis $(\delta \cdot D_y^{r+1} P) = 0$ we obtain $(\delta \cdot D_y^{r+1} P) = -\delta \cdot D_x D_y^r P = \cdots = (-1)^r \delta \cdot D_x^{r+1} P = 0$ and $\delta \cdot D_x^j D_y^k P = 0$, $j + k = r + 1$. Therefore (a) holds. Conclusion (b) follows from (a) using (iii) of Lemma 1 and (c) follows from (b) in the same way. Conclusions (a)′, (b)′, (c)′ are obtained by the same reasoning using (i)′, (ii)′, (iii)′ of Lemma 1 in place of (i), (ii), (iii).

Let $q(x)$ be an operator valued function of x from l_2 into l_2 which is piecewise continuous in the l_2 norm with respect to x, $-\infty < x < \infty$. We shall consider operators H^p, $p = 0, 1$ on $\mathfrak{h}$ where H^0 is a selfadjoint extension of $L^0 = (i(d/dx))^n$ and H^1 is a selfadjoint extension of $L^1 = (i(d/dx))^n + q(x)$. Let $s_j^p(x, \lambda)$, $j = 1, \cdots, n$, $p = 0, 1$ be linearly independent operator valued functions of x from l_2 into l_2 satisfying the equations $L^p y = \lambda y$ normalized so that $D^{j-1} s_k^p(0, l) = \delta_{jk} I$, $j, k = 1, \cdots, n$, where I is the identity mapping on l_2 [12], [14]. The functions $D_x^j s_k^p(x, \lambda)$ are continuous in x in the l_2 norm, $j = 0, 1, \cdots, n$ and they are entire in λ. The resolvent operators $R^p(\lambda) = (H^p - \lambda I)^{-1}$ are integral operators with Carleman kernels of the form

$$G^p(x, y, \lambda) = \sum_{j,k=1}^{n} M^{pjk}(\lambda) s_j^p(x, \lambda) s_k^p(y, \lambda), \qquad x \geqq y \tag{2.2}$$

with $G^p(x, y, \lambda)$ equal to the conjugate of the right side of (2.2) with (x, y) interchanged and λ replaced by $\bar{\lambda}$ when $x < y$. The kernels $G^0(x, y, \lambda)$ are given explicitly by

$$G^0(x, y, \lambda) = \sum_{j=1}^{v} \frac{\beta_j}{nw^{n-1}} \exp \beta_j w(x - y) \tag{2.3}$$

when $n = 2v$ is even and

$$G^0(x, y, \lambda) = \frac{\beta_1}{2nw^{n-1}} \exp \beta_1 w(x - y) + \sum_{j=2}^{v} \frac{\beta_j}{nw^{n-1}} \exp \beta_j w(x - y) \tag{2.4}$$

for $n = 2v - 1$ odd, $x < y$, with $w = \lambda^{1/n} = |\lambda|^{1/n} \exp(i(\theta/n))$, $\theta = \arg \lambda$, $0 \leqq \theta < 2\pi$ and

$\beta_j = -ie_j$ where e_j are nth roots of unity ordered so that $0 \leqq \arg e_1 < \cdots < \arg e_n < 2\pi$. The elements $M^{pjk}(\lambda)$ of the matrices $M^p(\lambda)$ have limits along the reals except at poles of $R^p(\lambda)$ which will be denoted by $M^{pjk}(l \pm i0)$, $l \in R$. We introduce operator valued matrices $V^p_{jk}(l \pm i0)$ to be solutions of the system of equations, $j = 1, \cdots, n$,

$$\begin{aligned}\sum_{k=1}^{n} V^p_{jk}(l \pm i0) s^q_k(x, l) \\ = s^p_j(x, l) + (-1)^q \int_{-\infty}^{\infty} G^q(x, y, l \pm i0) q(y) s^p_j(y, l)\, dy.\end{aligned} \tag{2.5}$$

Since $q(x)$ is piecewise continuous these equations have solutions except at poles of $R^p(\lambda)$, $-\infty < x < \infty$. It follows from the normalization condition $D^{j-1} s^p_k(0, l) = \delta_{jk} I$ applied to (2.5) that $V^p(l + i0)$ satisfy

$$\begin{aligned}V^p(l \pm i0) = I + (-1)^q [M^q(l \pm i0) F^p_1(l) \\ + M^q(l \mp i0)^* F^p_2(l)] \quad \text{for } p + q = 1,\end{aligned} \tag{2.6}$$

$p, q = 0, 1$ where $F^p_j(l)$, $j = 1, 2$ are matrices whose elements are operators from l_2 into l_2 defined by $F^{pjk}_1(l) = \int_{-\infty}^{0} s^q_k(x, l) q(x) s^p_j(x, l)\, dx$, $F^{pjk}_2(l) = \int_0^{\infty} s^q_j(x, l) q(x) s^p_k(x, l)\, dx$.

We shall suppose throughout that the resolvents $R^p(\lambda)$ have no poles along the real axis. In this case it follows from the spectral theory of ordinary differential operators that there exists measures $\mu^p(\Delta)$ and matrices $c^p(l)$ integrable with respect to μ^p such that for $\Delta \subseteq R$

$$\int_\Delta c^p(l)\, d\mu^p(l) = \lim_{\delta \downarrow 0} \frac{1}{\pi} \int_\Delta \operatorname{Im} M^p(l + i\delta)\, dl. \tag{2.7}$$

Moreover for Δ finite the spectral measures $E^p(\Delta)$ associated with H^p are integral operators with kernels $E^p(x, y, \Delta) \in C(l_2 : R \times R)$ given by

$$E^p(x, y, \Delta) = \int_\Delta \sum_{j,k=1}^{n} c^p_{jk}(l) s^p_j(x, l) s^p_k(y, l)\, d\mu^p(l), \tag{2.8}$$

and also by (2.5), (2.7)

$$\int_\Delta c^p(l)\, d\mu^p(l) I = \int_\Delta V^p(l \pm i0) c^q(l) V^p(l \mp i0)^T\, d\mu^q(l) \quad \text{for } \Delta \subseteq R \tag{2.9}$$

[4, p. 12]. The matrices $V^p(l \pm i0)$ satisfy $V^p(l \pm i0) V^q(l \pm i0) = I$, $p + q = 1$, $p = 0, 1$ and the matrices of the scattering operators S^p associated with H^p, $p = 0, 1$ may be given in terms of $V^p(l + i0)$ by the equation

$$S^p(l) = V^q(l + i0) V^p(l + i0), \qquad [4, \text{p. } 11]. \tag{2.10}$$

We shall refer to the matrix measures $\int_\Delta c^p(l)\, d\mu^p(l)$ as the spectral density measures of H^p in the following section. Since $c^0(l)$, $\mu^0(l)$ are explicitly known from the formula for $G^0(x, y, \lambda)$ it follows that the spectral density measure for H^1 is known when the matrices $V^1(l \pm i0)$ have been given.

3. The inverse problem. It was shown in § 2 that the spectral density measure corresponding to $L^1 = (i(d/dx))^n + q(x)$ depends on (n, n) matrices $V^1(l \pm i0)$ whose elements are operator valued functions of l, $-\infty < l < \infty$. In this section we suppose that the matrices $V^1(l \pm i0)$ have been given and consider the question of reconstruction of $q(x)$ from $V^1(l \pm i0)$. Following the procedure of I. M. Gel'fand, B. M.

Levitan we first form a nuclear kernel $\Omega^0(x, y) \in \mathfrak{h} \otimes \mathfrak{h}$ which is $*$-commutative with $L^0 = (i(d/dx))$. Next a generalized translation kernel $k^1(x, y) \in \mathfrak{h} \otimes \mathfrak{h}$ is determined by solving the integral equation.

$$k^1(x, y) + \Omega^0(x, y) + \int_{-y}^{x} k^1(x, t)\Omega^0(t, y)\, dt = 0, \tag{3.1}$$

$-x \leqq y \leqq x$, $0 \leqq x < \infty$ and finally $q(x)$ is obtained from the kernel $k^1(x, y)$ by means of the fórmula

$$q(x) = -(i)^n n\delta \cdot (D_x^{n-1} + D_x^{n-2} D_y) k^1(x, y). \tag{3.2}$$

We first consider operators W_c^p, $p = 0, 1$ on $\mathfrak{h}$ to $\mathfrak{h}$ whose kernels are given by the formal expression

$$W_c^p(x, y) = \int_{-\infty}^{\infty} \sum_{j,k=1}^{n} c_{jk}^q(l) s_j^p(x, l) s_k^p(y, l)\, d\mu^q(l). \tag{3.3}$$

In physics applications these operators W_c^p are called weight operators [9]. Since in general the integrals (3.3) are not convergent we assume following I. M. Gel'fand, B. M. Levitan that W_c^p have the form $W_c^p = \delta + \Omega^p$ where $\Omega^p \in \mathfrak{h} \otimes \mathfrak{h}$ are defined in terms of kernels F^p by setting $\Omega^p = D_x D_y F^p$, $p = 0, 1$ and F^p is given by

$$F^p(x, y) = \int_{-\infty}^{\infty} \sum_{j,k=1}^{n} \sigma_{jk}^q(l) \int_0^x s_j^p(t, l)\, dt \int_0^y s_k^p(t, l)\, dt\, d\mu^p(l) \tag{3.4}$$

with $\sigma^q(l) = V^q(l + i0) c^p(l) V^q(l + i0)^* - c^p(l)$.

Our results are based on the following assumptions regarding the kernels F^0 and $\Omega^0 = D_x D_y F^0$: For some integer $m \geqq 2$:

(i) $D_x^j D_y^k F^0 \in \mathfrak{h} \otimes \mathfrak{h}$ and these operators are integral operators whose kernels are continuous in x, y for $x + y > 0$, $0 \leqq x < \infty$, $j, k = 0, 1, \cdots, m+1$.

(ii) $\delta^\nu \cdot D_y^j \Omega^0 = 0$, $j = 0, 1, \cdots, m-2$, $\delta^\nu \cdot D_y^{m-1} \Omega^0 = \alpha \neq 0$, where α is an operator from l_2 to l_2 independent of x, y.

(iii) $\eta^\nu \cdot D_y^j \Omega^0 = 0$, $j = 0, 1, \cdots, m-2$.

(iv) $(I + \Omega^0)$ is a positive definite operator.

These assumptions are generalizations of those employed by I. M. Gel'fand, B. M. Levitan for differential operators of second order on $L_2(0, \infty)$ and of A. R. Sims for differential operators of second order on $L_2(R)$ [3], [7], [14], [15]. An operator $U^p = I + K^p$ on $\mathfrak{h}$ to $\mathfrak{h}$ is said to be a *generalized translation operator* for H^p iff $U^p U^q = I$ and

$$I = U^p W_c^q (U^p)^* \tag{3.5}$$

for $p + q = 1$, $p, q = 0, 1$. We shall also refer to K^p as generalized translation operators. Writing (3.5) in terms of K^p leads to the equations

$$K^p + \Omega^q + K^p \Omega^q = (K^q)^*, \tag{3.6}$$

and also

$$K^p + K^q + K^p K^q = 0 \tag{3.7}$$

$p + q = 1$, $p, q = 0, 1$ [9, § 7].

THEOREM 1. *Let* $\Omega^0 = D_x D_y F^0 \in \mathfrak{h} \otimes \mathfrak{h}$ *be a nuclear kernel* $*$*-commuting with* $(i(d/dx))$ *such that assumptions* (i), (ii), (iii), (iv) *hold. Then there exist kernels* $K^p \in \mathfrak{h} \otimes \mathfrak{h}$ *satisfying* (3.6), (3.7). *These kernels have the following properties*:

a) K^p, $p=0,1$ *are kernels of integral operators with kernel functions* $k^p(x,y)$ *of the form* $k^p(x,y)=(\chi_{[-x,\infty)}(y)-\chi_{[x,\infty)}(y))P^p(x,y)$, $p=0,1$ *where* $P^p(x,y)\in C^m(l_2:R\times R)$ *and* $P^1(x,y)$ *is the solution of* (3.1).

b) $D_x^jD_y^kK^1\in\mathfrak{h}\otimes\mathfrak{h}$, $j+k=0,1,\cdots,m$ *and*

$$\delta^\nu\cdot D_y^jK^1=0,\qquad j=0,1,\cdots,m-2,$$
$$\delta^\nu\cdot D_y^{m-1}K^1=-\alpha. \tag{3.8}$$

Proof. By assumption (i) $D_x^jD_y^k\Omega^0$ are Hilbert–Schmidt operators $j,k=1,\cdots,m$ and these operators are integral operators with kernels continuous in x,y for $x+y>0$, $0\leqq x<\infty$.

By assumption (iii) and the definition of η^ν in § 2, $\chi_{[-x,\infty)}(y)D_y^j\Omega^0(x,y)=D_y^j\Omega^0(x,y)$, $j=0,1,\cdots,m-2$, so that these kernels vanish for $y<-x$, $0\leqq x<\infty$.

Now fixing x, $0\leqq x<\infty$ consider the Fredholm equation (3.1) as a function of y for $-x\leqq y\leqq x$. By assumption (iv), (3.1) has a unique solution $P^1(x,y)$ which is continuous in y, $-x\leqq y\leqq x$. Also by a lemma of I. M. Gel'fand, B. M. Levitan the function $P^1(x,y)$ is continuous in x, $0\leqq x<\infty$, [7, p. 273]. Let $P^1(x,y)$ be extended continuously to all $(x,y)\in R\times R$ such that $P^1(x,y)=0$, $-\infty<x\leqq 0$ and define $k^1(x,y)=(\chi_{[-x,\infty)}(y)-\chi_{[x,\infty)}(y))P^1(x,y)$. Let K^1 be the integral operator whose kernel is $k^1(x,y)$. Next let $P^0(x,y)$ be the solution of the integral equation

$$P^1(x,y)+P^0(x,y)+\int_{-x}^{x}P^1(x,t)P^0(t,y)\,dt=0, \tag{3.9}$$

for $-x\leqq y\leqq x$, $0\leqq x<\infty$. Equation (3.9) is an equation of extended Volterra type and has again a unique continuous solution $P^0(x,y)$ which we extend to be zero for $-\infty<x\leqq 0$, $-\infty<y<\infty$. Let K^0 be the integral operator with kernel $k^0(x,y)$ defined by $k^0(x,y)=(\chi_{[-x,\infty)}(y)-\chi_{[x,\infty)}(y))P^0(x,y)$. By construction $K^p\in\mathfrak{h}\otimes\mathfrak{h}$, $p=0,1$. The operator K^0 has adjoint $(K^0)^*$ with kernel $k^{0*}(x,y)=k^0(y,x)$. Since $k^0(y,x)=0$, $y\leqq x$ and $K^1(x,y)$ satisfies (3.1) it follows that K^p, $p=0,1$ satisfy (3.6), (3.7).

Applying the operators $D_x^jD_y^k$ to integral equations (3.1) and (3.9) leads to a sequence of integral equations for the derivatives $D_x^jD_y^kP^p(x,y)$, $j+k=0,1,\cdots,m$ [3, p. 155]. It follows from these integral equations that $P^p(x,y)\in C^m(l_2:R\times R)$,

$$D_x^jD_y^kk^1(x,y)=(\chi_{[-y,\infty]}(x)-\chi_{[x,\infty]}(x))D_x^jD_y^kP^1(x,y),$$

except possibly on the lines $y=x$, $y=-x$, that the operator K^1 satisfies $D_x^jD_y^kK^1\in\mathfrak{h}\otimes\mathfrak{h}$, $j+k=0,1,\cdots,m$, and that (3.8) holds.

Given operators K^p, $p=0,1$ defined in Theorem 1 define $s_j^1(x,\lambda)$ by the equations

$$s_j^1(x,\lambda)=s_j^0(x,\lambda)+\int_{-x}^{x}k^1(x,y)s_j^0(y,\lambda)\,dy,\qquad j=1,\cdots,n, \tag{3.10}$$

where $k^1(x,y)$ is the kernel of the integral operator K^1. The $s_j^1(x,\lambda)$ are operator valued functions of x from l_2 into l_2 which are entire in l and $D_x^js_k^1(x,l)\in C(l_2:R)$, $j=0,1,\cdots,n$.

LEMMA 3. *Let* $\Omega^0=D_xD_yF^0\in\mathfrak{h}\otimes\mathfrak{h}$ *be a nuclear kernel $*$-commuting with* $(i(d/dx))$ *such that assumptions* (i), (ii), (iii), (iv) *hold for* $m\geqq n$. *Let* K^p, $p=0,1$, *be the integral operators defined in Theorem* 1 *and let* $s_j^1(x,\lambda)$ *be given in terms of* $s_j^0(x,\lambda)$ *by*

(3.10), $j=1, \cdots, n$. *Then the following estimate holds*:

$$\int_{-\infty}^{\infty} \sum_{j,k=1}^{n} c_{jk}^{1}(l) \int_{\alpha-\varepsilon}^{\alpha+\varepsilon} s_j^1(t, l)\, dt \int_{\alpha-\varepsilon}^{\alpha+\varepsilon} s_k^1(t, l)\, dt\, d\mu^1(l) = 2\varepsilon I + 0(\varepsilon^2) \tag{3.11}$$

as $\varepsilon \downarrow 0$.

Proof. The proof of this lemma will be omitted as it employs the same steps as those used in a corresponding argument given by the writer for differential operators on $L_2(R)$ [3, p. 149].

The purpose of Lemma 3 is to ensure that during the course of proof of Theorem 2 that (2.8) defines the kernel of a spectral measure when $p=1$ and $s_j^1(x, \lambda)$ are defined by (3.10).

THEOREM 2. *Let* $\Omega^0 = D_xD_yF^0 \in \mathfrak{h}\otimes\mathfrak{h}$ *be a nuclear kernel* $*$*-commuting with* $(i(d/dx))$ *such that assumptions* (i), (ii), (iii), (iv) *hold and let* K^p, $p=0, 1$, *be the kernels obtained by solving* (3.1). *Let n be the least integer*, $2 \leqq n \leqq m$ *such that* $\delta \cdot D_y^j K^1 = 0$, $0 \leqq j < n-2$, *and* $\delta \cdot D_y^{n-2}K^1 \neq 0$. *Then* K^p, $p=0, 1$, *are generalized translation operators for* H^p, $p=0, 1$, *where* H^p *are selfadjoint extensions of differential operators* $L^0 = (i(d/dx))^n$, $L^1 = L^1 = (i(d/dx))^n + q(x)$ *and the operator* $q(x)$ *is given in terms of* $k^1(x, y)$ *by the formula* (3.2).

Proof. According to Theorem 1, K^1 is an integral operator of the form $K^1 = (\eta^\nu - \eta) \cdot P^1$ with kernel function $k^1(x, y) = (\chi_{[-x,\infty)}(y) - \chi_{[x,\infty]}(y))P^1(x, y)$, $P^1(x, y) \in C(l_2 : R \times R)$. k^0 has the same form. The hypotheses $\delta \cdot D_y^j K^1 = 0$, $j=0, 1, \cdots, n-3$, and (3.8) imply using Lemma 2 that $\delta \cdot D_x^j D_y^k K^1 = \delta^\nu \cdot D_x^j D_y^k K^1 = 0$, $j+k=0, 1, \cdots, n-3$, and also $\delta \cdot D_x^{j+1} D_y^k K^1 = -\delta \cdot D_x^j D_y^{k+1} K^1$, $j+k=n-3$ and $\delta^\nu \cdot D_x^j D_y^k K^1 = \pm\alpha$, $j+k=n-2$, $n=m$. Also the conclusions (c), (c)′ of Lemma 2 hold with K^1 in place of P and n in place of m.

Let $E^1(x, y, \Delta) \in \Phi' \otimes \Phi'$ be the kernel defined by (2.8) where $s_j^1(x, l)$ are given in terms of $s_j^0(x, l)$ by (3.10), $\Delta \subseteq R$. By considering piecewise constant vector functions on l_2 one may show following the argument given in [7, § 5] that (3.11) implies that Parseval's formula in the following form is valid for all $u, v \in \mathfrak{h}$:

$$\int_{-\infty}^{\infty} \sum_{j,k=1}^{n} c_{jk}^1(l)(s_j^1, v)(s_k^1, u)\, d\mu^1(l) = (u, v). \tag{3.12}$$

Parseval's formula (3.12) implies that $E^1(x, y, \Delta)$ are kernels of bounded, selfadjoint, idempotent operators $E^1(\Delta)$, $\Delta \subseteq R$ and $E^1(\Delta')E^1(\Delta'') = E^1(\Delta' \cap \Delta'')$ for all Borel sets Δ', $\Delta'' \subseteq R$. Also $E^1([-n, n])$ converges weakly to I as $n \to \infty$. Therefore the kernels $E^1(x, y, \Delta)$ define a spectral measure on R. Let H^1 denote the selfadjoint operator whose spectral measure is $E^1(\Delta)$, $\Delta \subseteq R$. Since for $p=0, 1$, $E^p([-n, n])$ both converge weakly to I as $n \to \infty$ it follows referring to the definition of W_c^p that (3.5) holds and therefore K^p are generalized translation operators for H^p, $p=0, 1$.

It remains to check that H^1 is an extension of $L^1 = (i(d/dx))^n + q(x)$ where $q(x)$ is given by (3.2). Let $V^p(l \pm i0 : \Delta)$ be matrices equal to $V^p(l \pm i0)$ for $l \in \Delta$ and zero otherwise where Δ is a Borel set in R and let $W_c^p(\Delta)$ be defined as in (3.3) except that the integration is over the set Δ. Let $M_\pm^p(\Delta)$ be operators from $\mathfrak{h}$ into the range of $E^p(\Delta)$ defined in terms of $V^p(l \pm i0 : \Delta)$ by $M_\pm^p(\Delta)\phi = \psi$ iff

$$\left(\sum_{k=1}^{n} V_{jk}^q(l \pm i0 : \Delta) s_k^p(x, l), \phi(x)\right) = (s_j^p(x, l), \psi(x)),$$

where the equality holds for almost all l with respect to $\mu^p(l)$, $p+q=1$, $p, q=0, 1$. From the definitions of $M_\pm^p(\Delta)$, $E^p(\Delta)$, U^p and the properties of $V^p(l\pm i0)$ it follows that $M_\pm^p(\Delta)$ has a partial inverse $M_\pm^p(\Delta)^{-1}$ from $\mathfrak{h}$ into the range of $E^p(\Delta)$ and

$$M_\pm^p(\Delta)M_\pm^p(\Delta)^{-1}=E^p(\Delta),$$

$$(U^pM_\pm^q(\Delta))^*=M_\pm^q(\Delta)^{-1}U^q, \qquad p+q=1, \quad p, q=0, 1, \quad \text{and also}$$

$$E^p(\Delta)=U^pW_c^q(\Delta)(U^p)^*,$$

$$W_c^p(\Delta)=M_\pm^p(\Delta)M_\pm^p(\Delta)^* \qquad \text{(compare with [4, p. 7]).}$$

From these relations we obtain

$$\begin{aligned} E^1(\Delta)&=U^1M_\pm^0(\Delta)M_\pm^0(\Delta)^*(U^1)^*=U^1M_\pm^0(\Delta)(U^1M_\pm^0(\Delta))^*\\ &=U^1M_\pm^0(\Delta)M_\pm^0(\Delta)^{-1}U^0=U^1E^0(\Delta)U^0, \qquad \Delta\subseteq R, \end{aligned}$$

and therefore $H^1=U^1H^0U^0$. If Q is such that $H^1=H^0+Q$ then $Q=(U^1H^0-H^0U^1)U^0$ or equivalently

$$Q=(K^1H^0-H^0K^1)(I+K^0). \tag{3.13}$$

Next by the properties of K^1 given above

$$D_y^{n-2}K^1=D_y^{n-2}((\eta^\nu-\eta)\cdot P^1)=(\eta^\nu-\eta)\cdot D_y^{n-2}K^1, \tag{3.14}$$

$$D_x^{n-2}K^1=D_x^{n-2}((\eta^\nu-\eta)\cdot P^1)=(\eta^\nu-\eta)\cdot D_x^{n-2}K^1, \tag{3.15}$$

and using (ii), (ii)′ of Lemma 1, we obtain

$$D_y^{n-1}K^1=-\delta\cdot D_x^{n-2}K^1+(\eta^\nu-\eta)\cdot D_y^{n-1}K^1, \tag{3.16}$$

$$D_x^{n-1}K^1=\delta\cdot D_x^{n-2}K^1+(\eta^\nu-\eta)\cdot D_x^{n-1}K^1. \tag{3.17}$$

Multiplying (3.16), (3.17) respectively by D_y, D_y and forming the differences using $\delta\cdot D_y^{n-2}K^1=(-1)^n\delta\cdot D_x^{n-2}K^1$, $\delta^\nu\cdot D_y^{n-1}K^1=(-1)^n\delta^\nu\cdot D_x^{n-1}K^1=\text{const.}$, and (iii) of Lemma 1 leads to the equation

$$(-D_y)^nK^1-D_x^nK^1=-\delta\cdot r+(\eta^\nu-\eta)\cdot((-D_y)^nK^1-D_x^nK^1) \tag{3.18}$$

where $r=(2D_x^{n-1}+D_x^{n-2}D_y+(-1)^nD_y^{n-1})K^1$. Since $(K^1H^0-H^0K^1)=((-iD_y)^n-(iD_x)^n)K^1$ formula (3.18) may be written

$$K^1H^0-H^0K^1=-(i)^n\delta\cdot r+(\eta^\nu-\eta)\cdot(K^1H^0-H^0K^1), \tag{3.19}$$

and by (3.13), (3.19) and properties of the kernels K^0, K^1 we find that

$$Q=-(i)^n\delta\cdot r+(\eta^\nu-\eta)\cdot h, \tag{3.20}$$

where $h=\{K^1H^0-H^0K^1-(i)^n(\delta\cdot r)K^0+(K^1H-H^0K^1)K^0\}$. Next using the relations (c) of Lemma 2 which hold for K^1 in place of P and n in place of m it turns out that

$$\delta\cdot r=n\delta\cdot(D_x^{n-1}+D_x^{n-2}D_y)K^1 \tag{3.21}$$

or alternatively

$$\delta\cdot r=(-1)^nn\delta\cdot(D_xD_y^{n-2}+D_y^{n-1})K^1. \tag{3.22}$$

By (3.21), (3.22) $(i^n\delta\cdot r)^*=i^n(\delta\cdot r)$ and since Q is selfadjacent it follows from (3.20) that $((\eta^\nu-\eta)\cdot h)^*=(\eta^\nu-\eta)\cdot h$. Because by definition the kernels $(\eta^u-\eta)$ vanish when $-\infty < x < y$ a kernel of the form $(\eta^{\nu} - \eta) \cdot h$ can only be selfadjoint when it is zero and consequently $Q = -(i)^n \delta \cdot r$ and formula (3.2) follows from (3.21).

The above proof follows closely the argument given by I. Kay, H. E. Moses for differential operators of second order defined on $L_2(0, \infty)$ [9, pp. 288, 289]. During the course of proof of Theorem 2 it was shown that $(\eta^{\nu} - \eta) \cdot h = 0$ which implies $h(x, y) = 0$, $-x < y < x$, $0 \leqq x < \infty$. Referring to the definition of h, we find $K^1 H^0 - H^0 K^1 = QK^1$, from which it follows that

$$(3.23) \qquad (H^p \otimes I)K^p = (I \otimes H^q)K^p, \qquad p + q = 1, \quad p = 0, 1.$$

This anti-commuting property of the kernels K^p with H^p is characteristic of these generalized translation kernels.

Let $\mathfrak{h}' \otimes \mathfrak{h}'$ denote one of the spaces $W^{-(n,p)}(l_2 : R) \otimes W^{-(n,p)}(l_2 : R)$ for some fixed n. Theorems 1 and 2 remain valid if in the hypotheses and conclusions one replaces the space $\mathfrak{h} \otimes \mathfrak{h}$ of Hilbert–Schmidt kernels by a larger space $\mathcal{B}$ consisting of kernels which are in $\mathfrak{h}' \otimes \mathfrak{h}'$ and which extend to Carleman integral operators on $\mathfrak{h}$. The example presented in § 4 involves kernels which are initially in a space of type $\mathcal{B}$. By modifying the example by truncating the kernels we obtain an example involving kernels in the space $\mathfrak{h} \otimes \mathfrak{h}$.

4. An example. We shall present an example in which $n = 4$ and the operator $q(x)$ has a diagonal form. Let α be an operator from l_2 into l_2 defined by $\alpha u(x) = (\alpha_1 u_1(x), \alpha_2 u_2(x), \cdots)$ where $(\alpha_1, \alpha_2, \cdots) \in l_2$; let $f_i(l) = 1/\prod_{j=1}^{6} (l + \beta_{ji} i)$, $\beta_{ji} > 0$ and choose a kernel F^0 on l_2 into l_2 with diagonal components F_i^0 given by

$$(4.1) \qquad F_i^0(x, y) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_0^x e^{-ilt} \, dt \int_0^y e^{-ilt} \, dt \, f_i(l) \, d l \alpha_i.$$

Correspondingly Ω^0 has components

$$(4.2) \qquad \Omega_i^0(x, y) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-il(x+y)} f_i(l) \, d l \alpha_i.$$

For $x + y < 0$, $D_y^k \Omega^0(x, y) = 0$, $k = 0, 1, 2, 3, 4$. Also for $x + y > 0$ we find by integration

$$(4.3) \qquad \Omega_i^0(x, y) = \sum_{j=1}^{6} e^{-\beta_{ji}(x+y)} \alpha_i.$$

Choose the constants β_{ji} such that

$$D_y^j \Omega_i^0(x, y)|_{y=-x} = 0, \qquad j = 0, 1, 2, \qquad D_y^3 \Omega_i^0(x, y)|_{y=-x} = \alpha_i.$$

F^0, Ω^0 are Carleman kernels in the space $\mathfrak{h}' \otimes \mathfrak{h}'$, $\mathfrak{h}' = W^{-(1,p)}(l_2 : R)$ and Ω^0 satisfies assumptions (ii), (iii), (iv) of § 3. We shall suppose these kernels to be modified outside a bounded rectangle $[-a, a] \times [-a, a]$ so that they are continuously differentiable and satisfy assumptions (ii), (iii), (iv) of § 3 and such that they vanish identically outside some bounded measurable set. Then these modified kernels F^0, Ω^0 are in $\mathfrak{h} \otimes \mathfrak{h}$ and satisfy assumptions (i), (ii), (iii), (iv) of Theorems 1 and 2 with $m = 4$.

Assuming the operator $k^1(x, y)$ to have a diagonal form $k^1(x, y)u(y) = (k_1^1(x, y)u_1(x), k_2^1(x, y)u_2(x), \cdots)$ we try a solution of (3.1) with components

$$(4.4) \qquad \begin{aligned} k_i^1(x, y) &= \sum_{j=1}^{12} C_{ji}(x) e^{p_{ji} y}, \\ p_{j+6,i} = -p_{j,i}, &\qquad C_{j+6,i}(x) = E_{ji} C_{ji}(x), \qquad j = 1, 2, \cdots, 6. \end{aligned}$$

Substituting $k_i^1(x, y)$ into (3.1) and equating terms we find that p_{ji} must satisfy the equation

$$1-\prod_{k=1}^{6}\frac{\alpha_i}{(p_{ji}-\beta_{ki})}\prod_{k=1}^{6}\frac{\alpha_i}{(p_{ji}+\beta_{ki})}=0; \tag{4.5}$$

$E_{ji}=\prod_{k=1}^{6}\alpha_i/(p_{ji}-\beta_{jk})$ and $C_{ji}(x)$ are solutions of the system

$$1+\sum_{j=1}^{12}C_{ji}(x)\frac{e^{p_{ji}x}}{p_{ji}-\beta_{ki}}=0, \qquad k=1,2,\cdots,6.$$

For fixed i, the six constants β_{ji} satisfy four conditions stated above. Now choose β_{ji} such that the additional conditions $\sum_{k=1}^{12}C_{ki}(x)p_{ki}^j e^{p_{ki}x}=0$, $j=0, 1$ are satisfied [3, p. 147]. Then $k^1(x, y)$ is a solution of (3.1) such that $D_y^j k^1(x, y)|_{y=x}=0$, $j=0, 1$. $k^1(x, y)$ satisfies the conditions stated in Theorem 2 with $n=4$ and K^1 is a generalized translation kernel for a differential operator $L^1=(i(d/dx))^4+q(x)$ where $q(x)$ is given by (3.2),

$$q_i(x)=-4\sum_{k=1}^{12}(C_{ki}^{(2)}(x)p_{ki}+C_{ki}^{(3)}(x))\exp(p_{ki}x) \tag{4.6}$$

for $x\in[-a, a]$ and $q_i(x)=0$ outside some bounded interval. Starting with $k^1(x, y)$ one may construct generalized eigenfunctions $s_j^1(x, \lambda)$, $j=1, 2, 3, 4$ using (3.10) and then the matrices $V^p(l\pm i0)$ may be computed using (2.3), (2.4), (2.5). Note that the operator $q(x)$ given by (4.6) has a diagonal form given by $q(x)u(x)=(q_1(x)u_1(x), q_2(x)u_2(x), \cdots)$ since F^0 has this form. A nondiagonal $q(x)$ can be obtained by taking α to be nondiagonal. The method of construction of the above example is similar to the method that was used for constructing examples for the theory concerning differential operators on the space $L_2(R)$ by A. R. Sims for $n=2$ and by the writer for $n=4$ [3], [15]. Recently a method for constructing examples for inverse spectral theory of ordinary differential operators defined on a bounded interval, $n=4$, has been given in [10].

REFERENCES

[1] Z. S. AGRANOVICH AND V. A. MARCHENKO, *The Inverse Problem of Scattering Theory*, Gordon and Breach, New York, 1963.

[2] JU. M. BEREZANSKII, *Expansions in Eigenfunctions of Selfadjoint Operators*, Translations of Mathematical Monographs, vol. 17, American Mathematical Society, Providence, RI, 1968.

[3] J. B. BUTLER JR., *On the inverse problem for ordinary differential operators of even order*, J. Math. Anal. Appl., 26 (1969), pp. 142–158.

[4a] ———, *On smooth perturbations of selfadjoint operators defined on rigged Hilbert space*, Math. Nachr., 53 (1972), pp. 1–12.

[4b] ———, *Errata*, Ibid., 67 (1975), pp. 337–338.

[5] K. CHADAN AND P. C. SABATIER, *Inverse Problems in Quantum Scattering Theory*, Springer-Verlag, New York, 1977.

[6] R. E. EDWARDS, *Fourier Series, a Modern Introduction*, vol. 2, Holt, Rinehart and Winston, New York, 1967.

[7] I. M. GEL'FAND AND B. M. LEVITAN, *On the determination of a differential equation from its spectral function*, Isv. Akad. Nauk SSSR Ser. Mat., 15 (1951), pp. 309–360; transl. in Amer. Math. Soc. Transl., Ser. 2, 1 (1955), pp. 253–304.

[8] I. M. GEL'FAND AND N. YA. VILENKIN, *Generalized Functions*, vol. 4, Academic Press, New York, 1964.

[9] I. KAY AND H. E. MOSES, *The determination of the scattering potential from the spectral measure function: III, Calculation of the scattering potential from the scattering operator for the one dimensional Schrodinger equation*, Nuovo Cimento, 3 (1956), pp. 276–304.

[10] J. R. McLaughlin, *An inverse eigenvalue problem of order four*, this Journal, 7 (1976), pp. 646–661.
[11] V. A. Marchenko, *Spectral Theory of Sturm–Liouville Operators*, Naukova Dumka, Kiev, 1972. (In Russian.)
[12] H. H. Murtazin, *On the continuous spectra of differential operators of second degree*, Dokl. Akad. Nauk, SSSR, 212 (1973), pp. 1301–1304. (In Russian.)
[13] L. A. Sahnovič, *Differential operators of degree $n > 2$ with analytic coefficients*, Mat. Sb., 88 (1958), pp. 61–76.
[14] ———, *The method of operator translations for equations of higher degree*, Ibid., 55, No. 3 (1961), pp. 347–360.
[15] A. R. Sims, *Certain aspects of the inverse scattering problem*, J. Soc. Indust. Appl. Math., 4 (1957), pp. 183–205.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0008 $01.00/0

COMPLETELY CONVEX FUNCTIONS AND CONVERGENCE*

DALE H. MUGLER†

Abstract. A function $f(x)$ is completely convex (c.c.) on [0, 1] if $(-1)^k f^{(2k)}(x) \geqq 0$ for $k \geqq 0$ and all x in [0, 1]. This paper studies the convergence of the partial sums of the Maclaurin series of the function; in particular, how quickly the partial sums turn into a c.c. function. It is shown that no matter where the series is truncated, the resulting partial sum is a completely convex function in at least the interval $[0, \sqrt{10}/5]$.

1. Introduction. Functions whose derivatives have prescribed signs have been studied at least since 1914 when S. Bernstein showed that a function $f(x)$ which satisfies $f^{(k)}(x) \geqq 0$ for $k \geqq 0$ and all x in an interval is actually analytic in a disc containing that interval. Different types of conditions on the signs of derivatives have since been studied in relation to the region of analyticity of the function. See [3] for a general survey of those results.

A function $f(x)$ is called a completely convex (c.c.) function on an interval I if $f(x)$ has derivatives of all orders on that interval and if $(-1)^k f^{(2k)}(x) \geqq 0$ for x in I, $k \geqq 0$. D. V. Widder [9, p. 178] showed that any completely convex function on [0, 1] is actually an entire function of exponential type π. Further, every c.c. function on [0, 1] is of the form $g(x) = f(x) + c \sin(\pi x)$ for a "minimal" c.c. function $f(x)$ and nonnegative constant c.

Completely convex functions have been linked to positive harmonic functions in a region of the complex plane in an earlier report [5], and they have also been considered as a special case of solutions to certain Sturm–Liouville differential equations [4]. Further information about them thus does seem in order.

In this paper, we look at the question of how quickly a sequence approaching a c.c. function becomes completely convex. We consider in particular the Maclaurin series for the function, and how quickly the partial sums of that series become completely convex. One way to approach this might be to try to find a partial sum of high enough degree so that the sum itself always became a c.c. function. A second way is to consider the intervals in which each partial sum has the property.

At first we pay special attention to the c.c. function $f(x) = \sin(\pi x)$. We show in Theorem 1 the remarkable condition that all the resulting partial sums are completely convex in the interval $[0, \sqrt{6}/\pi]$. However, no matter how high the degree of the sum, the interval does not increase; for any partial sum $s_n(x)$ always has some even derivative $(-1)^k s_n^{(2k)}(x)$ nonnegative in only that interval. It is thus the second way that is further explored to see how quickly this c.c. property is attained.

In section three we consider a class of c.c. polynomials made up of what are essentially the Lidstone polynomials. The polynomials considered were recently shown [1] to be, along with $\sin(\pi x)$, the functions which lie on the extreme rays of the convex cone of c.c. functions on [0, 1]. We find that there are several different intervals where their partial sums are completely convex. However, one such interval is found to be minimal.

The main theorem is presented in § 4, where we use the results of the previous sections to show that every partial sum of a c.c. function on [0, 1] is itself a c.c. function in at least the interval $[0, \sqrt{10}/5]$. ($\sqrt{10}/5 \simeq .63$).

As a matter of convenience in stating some of the relationships in § 3, we will follow [1] and use the notation $\Lambda_n(x)$ to refer to those unique polynomials that satisfy

* Received by the editors September 1, 1976, and in final revised form November 10, 1977.
† Department of Mathematics, University of Santa Clara, Santa Clara, California 95053.

$\Lambda_0(x)=x$, $\Lambda_n''(x)=-\Lambda_{n-1}(x)$, and $\Lambda_n(0)=\Lambda_n(1)=0$ for $n\geqq 1$. This terminology differs slightly from the usual expressions for Lidstone polynomials, for example, [8], where $(-1)^n\Lambda_n(x)$ would be the nth Lidstone polynomial.

The other class of related polynomials will be denoted by $\Lambda_n^*(x)$, where $\Lambda_n^*(x)=\Lambda_n(1-x)$. Often both of these classes of polynomials share the same properties, so we will use $A_n(x)$ to represent either of the polynomials $\Lambda_n(x)$ or $\Lambda_n^*(x)$. For example, a property that is satisfied by the $A_n(x)$ will be satisfied by both of these polynomials.

The partial sum of order m of the polynomial $A_n(x)$ will be denoted $A_{n,m}(x)$. $A_n(x)$ is a polynomial of degree $2n+1$, so $A_{n,m}(x)=A_n(x)$ if $m\geqq 2n+1$. It follows from the definition of these polynomials that

$$A_{n,m}^{(2j)}(x)\doteq(-1)^jA_{n-j,m-2j}(x), \tag{1}$$

where $a\doteq b$ will mean that a is a positive multiple of b.

2. Partial sums.

THEOREM 1. *The partial sums of the Maclaurin series for* $\sin(\pi x)$ *are completely convex functions in the interval* $[0,\sqrt{6}/\pi]$.

Proof. We consider $s_n(x)=\sum_{k=1}^n(-1)^{k+1}(\pi x)^{2k-1}/(2k-1)!$ for $n\geqq 1$, so that n stands for the number of terms in the series. An alternate method might be to look at the partial sum of degree n corresponding to a sum whose highest power is n, but in that case all sums of even degree would be equal to the preceeding one of odd degree. Thus there is no loss of generality here.

We must first show that each $s_n(x)$ is nonnegative on the interval. Note first that the terms in the series for $s_n(x)$ alternate in sign, beginning with the positive function $s_1(x)=\pi x$. Considering the terms of the series in pairs from the first, a general pair is of the form $(\pi x)^{2k-1}/(2k-1)!-(\pi x)^{2k+1}/(2k+1)!$. Thus each *pair* is nonnegative if $(\pi x)^{2k-1}/(2k-1)!\geqq(\pi x)^{2k+1}/(2k+1)!$ or

$$(2k+1)(2k)/\pi^2\geqq x^2. \tag{2}$$

The left side of (2) is smallest when $k=1$, corresponding to $s_2(x)=\pi x-(\pi x)^3/3!$. In this case, (2) becomes $6/\pi^2\geqq x^2$, or $\sqrt{6}/\pi\geqq x$. Thus $s_2(x)$ is nonnegative only in the interval $[0,\sqrt{6}/\pi]$.

However, more terms increase the interval on which the sum is nonnegative. If n is even, $s_n(x)$ consists of pairs of terms each one of which is nonnegative in at least the interval $[0,\sqrt{6}/\pi]$. If n is odd, then $s_n(x)$ consists of pairs of terms along with a final positive term. In any case, each $s_n(x)$ is nonnegative in $[0,\sqrt{6}/\pi]$.

To consider the derivatives, note that

$$\begin{aligned} s_n''(x) &= \sum_{k=1}^n(-1)^{k+1}\frac{d^2}{dx^2}\{(\pi x)^{2k-1}/(2k-1)!\} \\ &= \sum_{k=2}^n(-1)^{k+1}\pi^2(2k-1)(2k-2)(\pi x)^{2k-3}/(2k-1)! \\ &= \pi^2\sum_{k=2}^n(-1)^{k+1}(\pi x)^{2k-3}/(2k-3)! = -\pi^2 s_{n-1}(x). \end{aligned}$$

Calculation of further derivatives proceeds as above, giving

$$s_n^{(2k)}(x)=(-1)^k\pi^{2k}s_{n-k}(x) \tag{3}$$

for k less than n. But since we have already shown each partial sum to be nonnegative in $[0, \sqrt{6}/\pi]$, (3) shows that each partial sum is actually completely convex in this interval.

3. The extremal polynomials.

LEMMA 1. *The $A_{n,3}(x)$ are nonnegative in an interval of the form $[0, c]$, for $c<1$ and $n \geqq 2$, and the smallest such interval corresponds to $\Lambda^*_{2,3}(x)$.*

Proof. The form of each of these sums is $A_{n,3}(x)=A'_n(0)x+A_n^{(3)}(0)x^3/6=x\{A'_n(0)+A_n^{(3)}(0)x^2/6\}$. Thus to consider the lemma, we need to consider the interval for which $A'_n(0)+A_n^{(3)}(0)x^2/6$ is nonnegative or

$$x^2 \leqq \frac{6A'_n(0)}{-A_n^{(3)}(0)}. \tag{4}$$

The Fourier series representation of these polynomials can be found in [7], [2], or [1], and has one of two forms: Either

$$\text{(a)} \qquad \Lambda^*_n(x)=\frac{2}{\pi^{2n+1}}\sum_{k=1}^{\infty}\frac{\sin(k\pi x)}{k^{2n+1}}$$

or

$$\text{(b)} \qquad \Lambda_n(x)=\frac{2}{\pi^{2n+1}}\sum_{k=1}^{\infty}(-1)^{k+1}\frac{\sin(k\pi x)}{k^{2n+1}}$$

for $n \geqq 0$ and $0 \leqq x \leqq 1$. The series converge uniformly for $n \geqq 2$, so the values of $A'_n(0)$ and $A_n^{(3)}(0)$ can then be found. Using these representations, we find that inequality (4) has one of the following two forms: If $A_n(x)=\Lambda^*_n(x)$, then (4) becomes

$$x^2 \leqq \frac{6}{\pi^2}\,\frac{\zeta(2n)}{\zeta(2n-2)} \tag{5}$$

where the zeta function is $\zeta(s)=\sum_{k=1}^{\infty}(1/k^s)$, and if $A_n(x)=\Lambda_n(x)$, (4) becomes

$$x^2 \leqq \frac{6}{\pi^2}\,\frac{\eta(2n)}{\eta(2n-2)} \tag{6}$$

where $\eta(s)=\sum_{k=1}^{\infty}((-1)^{k-1}/k^s)$.

To consider (5), $\zeta(2n)/\zeta(2n-2)$ is always less than 1. Also, as n increases this ratio increases monotonically. This can be seen as follows: If $f(x)=\zeta(2x)/\zeta(2x-2)$, then $f'(x)=2[\zeta(2x-2)\zeta'(2x)-\zeta(2x)\zeta'(2x-2)]/[\zeta(2x-2)]^2$. Thus $f'(x)>0$ if $\zeta(2x-2)\zeta'(2x)>\zeta(2x)\zeta'(2x-2)$ or

$$\zeta'(2x)/\zeta(2x)>\zeta'(2x-2)/\zeta(2x-2). \tag{7}$$

However, $\zeta'(u)/\zeta(u)=-\sum_{n=1}^{\infty}\Lambda(n)/n^u$ [6, p. 1], where here $\Lambda(n)=\log p$ if n is a power of the prime p, and otherwise $\Lambda(n)=0$. Note that $\Lambda(n)$ in this series is not related to the polynomial $\Lambda_n(x)$. From this form, one can see that $\zeta'(u)/\zeta(u)$ is an increasing function of u. Therefore (7) holds and $f(x)$ is seen to be monotonically increasing.

The smallest interval corresponds to $n=2$, where $\Lambda^*_{2,3}(x)=(8x-20x^3)/360$. The corresponding inequality is $x^2 \leqq 6\zeta(4)/[\pi^2\zeta(2)]=.4$, or $x \leqq \sqrt{10}/5 \simeq .632$.

For the second case (6), it can be shown by using properties of alternating series that $\eta(2n)/\eta(2n-2) \geqq 1$ for $n \geqq 2$. Thus the inequality will certainly be satisfied if $x^2 \leqq 6/\pi^2$ or $x \leqq \sqrt{6}/\pi \simeq .78$. The interval where $\Lambda_{n,3}(x)$ is nonnegative is at least $0 \leqq x \leqq \sqrt{6}/\pi$, and this interval includes the one corresponding to $\Lambda^*_{2,3}(x)$.

DEFINITION. Let I be the interval $[0, \sqrt{10}/5]$, the interval where $\Lambda_{2,3}^*(x)$ is nonnegative.

LEMMA 2. *The $A_{n,m}(x)$ are nonnegative for $x \in I$ and nonnegative integers n and m.*

Proof. This lemma will be proved in three cases. The first case will deal with the partial sums of order $4j+1$, the second with m of form $4j+3$, and the third for the even ordered sums. In each case, the equation

$$\int_0^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\,dt = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k \tag{8}$$

will be used.

Case 1. *$A_{n,4j+1}(x) \geqq A_n(x)$ for $j \geqq 0$ and $x \in [0, 1]$.*

Proof. Applying (8), we find

$$\int_0^x \frac{(x-t)^{4j+1}}{(4j+1)!} A_n^{(4j+2)}(t)\,dt = A_n(x) - A_{n,4j+1}(x).$$

But the polynomial $A_n(x)$ is completely convex, so that $A_n^{(4j+2)}(x) \leqq 0$. Thus the integral is less than or equal to zero, and equivalently $A_n(x) - A_{n,4j+1}(x) \leqq 0$.

Case 2. *$A_{n,4j+3}(x) \geqq A_{n,4j-1}(x)$ for $j \geqq 1$, $n \geqq 3$, and $x \in I$.*

Proof. We use (8) applied to $A_{n,4j+3}(x)$:

$$\int_0^x \frac{(x-t)^{4j-1}}{(4j-1)!} A_{n,4j+3}^{(4j)}(t)\,dt = A_{n,4j+3}(x) - A_{n,4j-1}(x).$$

Now property (1) of these partial sums applies so that $A_{n,4j+3}^{(4j)}(x) \doteq (-1)^{2j} A_{n-2j,4j+3-4j}(x) = A_{n-2j,3}(x)$. Using Lemma 1, we can see that the integral on the left above is positive. Thus $A_{n,4j+3}(x) - A_{n,4j-1}(x) \geqq 0$.

Case 3. *$A_{n,m}(x)$ is nonnegative in I for m even.*

Proof. First note that there are no strictly even partial sums for $\Lambda_n(x)$, $n \geqq 0$. This follows from the fact that $A_n^{(2k)}(0) = (-1)^k A_{n-k}(0)$ and the original condition that $\Lambda_j(0) = 0$ for $j \geqq 0$.

However, $\Lambda_0^*(0) = 1$, so that $\Lambda_n^{*(2n)}(0) = (-1)^n$. This will make the only even partial sum for $\Lambda_n^*(x)$ to be of degree $2n$, since for other positive n: $\Lambda_n^*(0) = \Lambda_n(1) = 0$. Thus Case 3 will be proved if it is shown that $\Lambda_{n,2n}^*(x)$ is nonnegative in I for any n.

If $n = 1$, $\Lambda_{1,2}^*(x) = (2x - 3x^2)/6$. This is nonnegative in the interval $[0, 2/3]$ which does contain I.

If $n \geqq 2$, we note first that (1) gives $\Lambda_{n,2n}^{*(2n-2)}(x) \doteq (-1)^{n-1} \Lambda_{1,2}^*(x)$. Applying (8) to $\Lambda_{n,2n}^*(x)$, we see

$$\int_0^x \frac{(x-t)^{2n-3}}{(2n-3)!} \Lambda_{n,2n}^{*(2n-2)}(t)\,dt = \Lambda_{n,2n}^*(x) - \Lambda_{n,2n-3}^*(x).$$

Using property (1), we find the integral on the left is a positive multiple of $\Lambda_{1,2}^*(x)$ if n is odd. Thus in this case $\Lambda_{n,2n}^*(x) \geqq \Lambda_{n,2n-3}^*(x)$ in I. But the nonnegativity of $\Lambda_{n,2n-3}^*(x)$ follows from Cases 1 and 2, since $2n-3$ is odd.

Finally, if n is even, $\Lambda_{n,2n}^*(x) = \Lambda_{n,2n-1}^*(x) + \Lambda_n^{*(2n)}(0) x^{2n}/(2n)!$. But since $\Lambda_n^*(x)$ is completely convex, $\Lambda_n^{*(2n)}(0)$ is nonnegative. The nonnegativity of the sum $\Lambda_{n,2n-1}^*(x)$ follows from the previous cases, so $\Lambda_{n,2n}^*(x)$ is the sum of two nonnegative functions in I.

This completes the proof of Case 3. In Case 1 we have shown the partial sums of order $4j+1$ to be bounded below by the completely convex, and thus nonnegative, function $A_n(x)$. For each n, it follows from Case 2 that all partial sums of order $4j+3$

are bounded below by the sum of degree three. Since this was shown by Lemma 1 to be nonnegative in the interval I, the proof of Lemma 2 is complete.

Combining now the conclusion of Lemma 2 with the general property (1), it follows that we have proved:

THEOREM 2. *The polynomials $A_{n,m}(x)$ are all completely convex functions in at least the interval I.*

4. Main result.

THEOREM 3. *If the Maclaurin series for a completely convex function is truncated after any term, the resulting polynomial is still a completely convex function in at least the interval $I=[0, \sqrt{10}/5]$.*

Proof. It follows from [8] that we can write every completely convex function $f(x)$ in the form

$$f(x)=\sum_{n=0}^{\infty}\{a_n\Lambda_n(x)+b_n\Lambda_n^*(x)\}+c\sin(\pi x)$$

where the a_n, b_n, and c are nonnegative constants for $n\geqq 0$. The series converges uniformly on compact sets, and can be differentiated term-by-term.

One could find the Maclaurin series for $f(x)$ by substituting the Maclaurin series for the terms $\Lambda_n(x)$, $\Lambda_n^*(x)$, and $\sin(\pi x)$, $n\geqq 0$. In fact, if the Maclaurin series for $f(x)$ is truncated after the term containing the multiple x^m, the resulting polynomial will consist of $\sum_{n=0}^{\infty}\{a_n\Lambda_{n,m}(x)+b_n\Lambda_{n,m}^*(x)\}$ and a partial sum of the Maclaurin series for $\sin(\pi x)$. But each of these terms has already been shown to be completely convex in at least the interval I, so the result also holds for the truncated Maclaurin expansion for $f(x)$.

Acknowledgment. The author wishes to thank the referees for their encouraging and helpful remarks.

REFERENCES

[1] C. BERG, *Representation of completely convex functions by the extreme-point method*, Enseignement Math., 23 (1977), pp. 181–190.

[2] R. P. BOAS, JR., *Representations for completely convex functions*, Amer. J. Math., 81 (1959), pp. 709–714.

[3] ———, *Signs of derivatives and analytic behavior*, Amer. Math. Monthly, 78 (1971), pp. 1085–1093.

[4] J. D. BUCKHOLTZ AND J. K. SHAW, *Generalized completely convex functions and Sturm–Liouville operators*, this Journal, 6 (1975), pp. 812–828.

[5] D. H. MUGLER, *Completely convex and positive harmonic functions*, this Journal, 6 (1975), pp. 681–688.

[6] E. C. TITCHMARSH, *The Zeta-Function of Riemann*, Stechert-Harner Service Agency, New York, 1964.

[7] J. M. WHITTAKER, *On Lidstone's series and two-point expansions of analytic functions*, Proc. London Math. Soc., 36 (1934), pp. 451–469.

[8] D. V. WIDDER, *Completely convex functions and Lidstone series*, Trans. Amer. Math. Soc., 51 (1942), pp. 387–398.

[9] ———, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1941.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0009 $01.00/0

WEAKLY PERTURBED FORMS AND THE SPECTRAL PROPERTIES OF ASSOCIATED EIGENVALUE PROBLEMS*

WILLIAM KOLATA†

Abstract. A bounded sesquilinear form, $a(\cdot,\cdot)$, defined on a Hilbert space X is called a weak perturbation of a Hermitian form, $h(\cdot,\cdot)$, if an operator associated with $a(\cdot,\cdot)$ is a weak perturbation, in the sense of Kreĭn, of a selfadjoint operator associated with $h(\cdot,\cdot)$. Under suitable hypotheses on the forms $a(\cdot,\cdot)$ and $h(\cdot,\cdot)$, it is shown that $a(\cdot,\cdot)$ is a weak perturbation of $h(\cdot,\cdot)$. In this case, results of Keldyš and Kreĭn on the spectral properties of weakly perturbed selfadjoint operators are applied to the eigenvalue problem, $a(u, v)=\lambda(u, v)$, where $(\cdot,\cdot)$ is the inner product on X. In particular, the completeness in X of the space of generalized eigenvectors and various results on the localization and distribution of eigenvalues are demonstrated. Applications of these results to three different kinds of elliptic differential eigenvalue problems are included. Among these applications is a problem from the theory of hydrodynamic stability and a problem in which the eigenvalue parameter appears in a boundary condition.

Introduction. The formulation of elliptic boundary value problems in terms of sesquilinear forms is a standard and very flexible technique for studying the solutions to such problems. It has proved useful in both source and eigenvalue problems. Usually, in the study of the spectral properties of linear, elliptic eigenvalue problems, it is the differential operators and their resolvents that play the central role (cf. [1], [5], [6]). The associated forms are used to prove statements about differential operators. In this paper however the focus is on eigenvalue problems posed in terms of forms, and the main theorem concerns the spectral properties of such problems. Elliptic eigenvalue problems appear as examples of the application of this theorem. The advantage of this approach is that it can be applied to a wide variety of elliptic eigenvalue problems, including problems with eigenvalue parameters in the boundary condition (see § 3). Moreover, in many cases, the hypotheses are easy to verify.

The forms that are studied in this paper are perturbations of bounded, Hermitian, positive definite forms, $h(\cdot,\cdot)$, defined on a Hilbert space V by bounded but not necessarily Hermitian forms, $k(\cdot,\cdot)$ also defined on V. The perturbation, $a(\cdot,\cdot)=h(\cdot,\cdot)+k(\cdot,\cdot)$, will be called a weak perturbation if a compact operator associated with $a(\cdot,\cdot)$ is a weak perturbation, in the sense of Kreĭn [7], of a selfadjoint operator associated with $h(\cdot,\cdot)$. In this case it is possible to apply theorems of Keldyš and Kreĭn to deduce conclusions about the location and distribution of the eigenvalues and the completeness of the space of principal vectors (generalized eigenfunctions) of the problem: $a(u, v)=\lambda(u, v)$ for all v in V, where $(\cdot,\cdot)$ is the inner product on a Hilbert space X in which V is compactly imbedded.

1. Weakly perturbed forms. Let X be a complex Hilbert space with inner product $(\cdot,\cdot)$ and norm $|\cdot|$, and let V be a Hilbert space with norm $\|\cdot\|$ that is compactly imbedded in X, i.e., there is an injective map of V into X that is continuous, compact, and has dense range. Identify V with its range in X and assume without loss of generality that $V\subset X$ and for all $v\in V$, $|v|\leqq\|v\|$.

Let $h(\cdot,\cdot)$ and $k(\cdot,\cdot)$ be bounded sesquilinear forms on V, and set $a(\cdot,\cdot)=h(\cdot,\cdot)+k(\cdot,\cdot)$. The adjoint form $a^*(\cdot,\cdot)$ is defined by $a^*(u, v)=\overline{a(v, u)}$. Assume

* Received by the editors August 11, 1977.

† Department of Mathematics, Statistics and Computer Science, American University, Washington, D.C. 20016. The results in this paper form part of a Ph.D. thesis written at the University of Maryland, College Park, under the direction of Professor John E. Osborn.

$h(\cdot,\cdot)$ is Hermitian, and suppose that there are positive constants α_0, α such that

$$h(u,u)\geqq \alpha_0\|u\|^2 \quad \text{for all } u\in V; \tag{1.1}$$

$$\operatorname{Re} a(u,u)\geqq \alpha\|u\|^2 \quad \text{for all } u\in V. \tag{1.2}$$

Note that both $a(\cdot,\cdot)$ and $h(\cdot,\cdot)$ can be considered as densely defined forms on X.

The following definitions and results from [8] play a fundamental role in this section. Suppose $t(\cdot,\cdot)$ is a sesquilinear form defined on a subspace $D(t)$ of X. The form t is closed if for every sequence $\{u_n\}$ in $D(t)$ such that $|u_n-u|$ and $t(u_n-u_m, u_n-u_m)$ converge to 0 as n, m approach ∞, it follows that $u\in D(t)$ and $t(u_n-u, u_n-u)$ converges to 0 as n approaches ∞. The form t is called sectorial with vertex γ_0 and semiangle θ_0 if the convex set of complex numbers $\{t(u,u)\colon u\in D(t);\ |u|=1\}$ is contained in the sector $|\arg(z-\gamma_0)|\leqq\theta_0<\pi/2$. Suppose $\mathcal{T}$ is an operator on X with domain $D(\mathcal{T})$. $\mathcal{T}$ is accretive if the convex set of complex numbers $\{(\mathcal{T}u,u)\colon u\in D(\mathcal{T});\ |u|=1\}$ is contained in the right half-plane. $\mathcal{T}$ is m-accretive if for every complex λ with $\operatorname{Re}\lambda>0$, $(\mathcal{T}+\lambda)^{-1}$ is a bounded operator on X with $\|(\mathcal{T}+\lambda)^{-1}\|\leqq(\operatorname{Re}\lambda)^{-1}$. If for some constant α, $\mathcal{T}+\alpha I$ is accretive (m-accretive) then $\mathcal{T}$ is called quasi-accretive (quasi-m-accretive).

LEMMA 1. *An m-accretive (quasi-m-accretive) operator $\mathcal{T}$ has no proper accretive (quasi-accretive) extension.*

Proof. See [8, Chap. V, § 3.10]. □

If the set $\{(\mathcal{T}u,u)\colon u\in D(t), |u|=1\}$ is contained in a sector, $|\arg(z-\gamma_0)|\leqq\theta_0<\pi/2$, $\mathcal{T}$ is called sectorial; if $\mathcal{T}$ is sectorial and quasi-m-accretive, it is called m-sectorial.

THEOREM 2. *Let $t(\cdot,\cdot)$ be a densely defined, closed, sectorial form on X. Then there exists an m-sectorial operator $\mathcal{T}$ on X such that*

(1.3) *$D(\mathcal{T})\subset D(t)$, and for every $u\in D(\mathcal{T})$ and $v\in D(t)$, $t(u,v)=(\mathcal{T}u,v)$;*

(1.4) *if $u\in D(t)$, $w\in X$, and $t(u,v)=(w,v)$ for all $v\in V$, then $u\in D(\mathcal{T})$ and $\mathcal{T}u=w$.*

Proof. See [8, Chap. VI, Thm. 2.1]. □

Theorem 2 can be applied to the forms a, a^*, and h provided they are shown to be closed and sectorial on X. That they are closed follows immediately from (1.1) and (1.2). Because $h(\cdot,\cdot)$ is Hermitian and satisfies (1.1) it is sectorial with vertex 0 and semiangle 0. Assumption (1.2) allows the estimate $|\operatorname{Im} a(u,u)|=|\operatorname{Im} k(u,u)|\leqq C\|u\|^2\leqq(C/\alpha)\operatorname{Re} a(u,u)$ from which it follows that $a(\cdot,\cdot)$ and $a^*(\cdot,\cdot)$ are sectorial with vertex 0 and semiangle $\tan^{-1}(C/\alpha)$. Consequently, by Theorem 2, there are m-sectorial operators $\mathcal{A}$, $\mathcal{A}^*$, $\mathcal{H}$ defined on X satisfying (1.3) and (1.4) with respect to the forms a, a^*, and h respectively. It is easily seen that $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$ and $\mathcal{H}$ is self-adjoint. For example, the adjoint of $\mathcal{A}$ is an extension of $\mathcal{A}^*$ and is accretive, but $\mathcal{A}^*$ cannot have a proper accretive extension and so is the adjoint of $\mathcal{A}$.

The Lax–Milgram lemma implies that there is a bounded linear map T from X into V such that $a(Tf,v)=(f,v)$ for all $f\in X$ and $v\in V$. Similarly there is a map H that satisfies $h(Hf,v)=(f,v)$ for all $f\in X$ and $v\in V$. Because V is compactly imbedded in X, the maps T and H are compact when considered as operators on X. The adjoint of T on X, T^*, satisfies $a^*(T^*f,v)=(f,v)$ and H is self-adjoint on X. A straightforward verification shows that $T=\mathcal{A}^{-1}$, $T^*=(\mathcal{A}^*)^{-1}$, and $H=(\mathcal{H}^{-1})$.

Again by the Lax–Milgram lemma, there is a bounded operator, J, on V such that $h(u,Jv)=k(u,v)$ for all $u,v\in V$. The operator J will be said to have the extension property if the restriction of J to $D(\mathcal{A}^*)$ has an extension to a compact operator on X such that -1 is not an eigenvalue of this operator.

THEOREM 3. *Suppose* (1.1) *and* (1.2) *hold, and the operator J satisfies the extension property. Then $a(\cdot, \cdot)$ is a weak perturbation of $h(\cdot, \cdot)$. That is, there exists a compact operator S on X such that $T = H(I+S)$. Moreover, -1 is not an eigenvalue of S and so $-1 \notin \sigma(S)$.*

Proof. Let $\bar{J}$ denote the restriction of J to $D(\mathscr{A}^*)$. For any $v \in D(\mathscr{A}^*)$ and any $u \in V$, $(u, \mathscr{A}^* v) = a(u, v) = h(u, v) + k(u, v) = h(u, (I+\bar{J})v)$. By Theorem 2, (1.4), $(I+\bar{J})v \in D(\mathscr{H})$ and $\mathscr{H}(I+\bar{J})v = \mathscr{A}^* v$. Thus, $\mathscr{H}(I+\bar{J})$ is an extension of $\mathscr{A}^*$ and consequently so is $\mathscr{H}(I+\tilde{J})$, where $\tilde{J}$ is the compact extension of $\bar{J}$ to X.

For any $f \in X$, $f = \mathscr{A}^* T^* f = \mathscr{H}(I+\bar{J})T^* f = \mathscr{H}(I+\tilde{J})T^* f$. Since $-1 \notin \sigma(\tilde{J})$, it follows that $(I+\tilde{J})^{-1} Hf = (I+\tilde{J})^{-1} H(\mathscr{H}(I+\tilde{J})T^* f) = T^* f$. Thus, $(I+\tilde{J})^{-1} H = T^*$, and $H(I+\tilde{J}^*)^{-1} = T$. Set $S = (I+\tilde{J}^*)^{-1} - I = -(I+\tilde{J}^*)^{-1}\tilde{J}^*$. Then $T = H(I+S)$, S is compact, and since $I + S = (I+\tilde{J}^*)^{-1}$, $-1 \notin \sigma(S)$. □

If the form $k(\cdot, \cdot)$ satisfies certain bounds, it is easy to see that the operator J has the extension property.

LEMMA 4. *Assume that* (1.1) *and* (1.2) *hold, and that*

$$|k(u, v)| \leqq C\|u\|\,|v| \quad \textit{for all } u, v \in V, \tag{1.5}$$

where C is a positive constant. Then J has the extension property.

Proof. By (1.1) and (1.5), $\alpha_0 \|Jv\|^2 \leqq h(Jv, Jv) = |k(Jv, v)| \leqq C\|Jv\|\,|v|$, where $v \in V$. The operator J can be extended by continuity to a map $\tilde{J}$ that is bounded from X into V. Because V is compactly imbedded in X, $\tilde{J}$ is compact when considered as an operator on X. Thus, J, and so $\bar{J}$, has a compact extension $\tilde{J}$ to X.

As in the proof of Theorem 3, $\mathscr{H}(I+\tilde{J})$ is an extension of $\mathscr{A}^*$. Suppose $v \in D(\mathscr{H}(I+\tilde{J}))$; then $(I+\tilde{J})v \in D(\mathscr{H}) \subset V$. Since $\tilde{J}v \in V$, v is also in V. Thus $(v, \mathscr{H}(I+\tilde{J})v) = h(v, (I+\tilde{J})v) = h(v, v) + k(v, v) = a(v, v)$. By (1.2), it follows that $\mathscr{H}(I+\tilde{J})$ is accretive. But $\mathscr{A}^*$ is m-accretive, and so $\mathscr{A}^* = \mathscr{H}(I+\tilde{J})$. Since $\mathscr{A}^*$ is injective, -1 cannot be an eigenvalue of $\tilde{J}$. □

Remark 5. If (1.1) and (1.5) hold, then for a suitably large constant γ_0 the form $a(u, v) + \gamma_0(u, v)$ satisfies (1.2). The effect of adding the term $\gamma_0(u, v)$ to the original eigenvalue problem $a(u, v) = \lambda(u, v)$ is to shift the eigenvalues by γ_0 and leave the principal vectors unchanged. It follows that without loss of generality it is enough to have (1.1) and (1.5).

2. The theorems of Keldyš and Kreĭn. A complex number λ is an eigenvalue of the form $a(\cdot, \cdot)$ on X if there exists a nonzero vector u^1 in V such that $a(u^1, v) = \lambda(u^1, v)$ for all v in V. The vector u^1 is called an eigenvector or principal vector of order 1 with respect to λ. A nonzero vector u^j in V is a principal vector of order j, j an integer >1, with respect to λ if $a(u^j, v) = \lambda(u^j, v) - \lambda a(u^{j-1}, v)$ for all v in V, where u^{j-1} is a principal vector of order $j-1$. It is easy to see that λ is an eigenvalue of $a(\cdot, \cdot)$ on X if and only if $\mu = \lambda^{-1}$ is an eigenvalue of T. Moreover, u^j is a principal vector of order j with respect to λ if and only if u^j is a principal vector of T with respect to μ of order j, i.e., $Tu^j = \mu u^j + u^{j-1}$.

The operator T is compact on X. By the Riesz–Schauder theory of compact operators, $\sigma(T)$ is a countable or finite set of complex numbers with cluster point possible only at 0. Each nonzero member of $\sigma(T)$ is an eigenvalue of T, and in this case, by (1.2), 0 is not an eigenvalue of T. For $\mu \in \sigma(T)$, $\mu \neq 0$, there exists a smallest positive integer $\alpha = \alpha(\mu)$, called the ascent of $\mu - T$, such that $\mathrm{N}((\mu - T)^\alpha) = \mathrm{N}((\mu - T)^{\alpha+1})$, where N denotes the null space. The subspace $\mathrm{N}((\mu - T)^\alpha)$ is finite dimensional and its members are the principal vectors of T with respect to μ. The order of a principal vector u is the smallest positive integer j such that $u \in \mathrm{N}((\mu - T)^j)$.

The dimension of $N((\mu - T)^{\alpha})$ is called the algebraic multiplicity of μ and the dimension of $N(\mu - T)$ is called the geometric multiplicity of μ. If T is self-adjoint or normal, the algebraic and geometric multiplicities are equal.

Let $\operatorname{sp}(T)$ denote the linear span of the principal vectors of T with respect to each of the eigenvalues. The space of principal vectors of T is called complete in X if $\operatorname{sp}(T)$ is dense in X.

Set $K = (T^*T)^{1/2}$ where T^* is the adjoint of T on X. The operator K is nonnegative, self-adjoint, and compact. Let $s_1 \geqq \cdots \geqq s_j \geqq s_{j+1} \geqq \cdots$ denote the eigenvalues of K counted according to multiplicity. The numbers s_j are called the singular values of T. If $\sum_{j=1}^{\infty} s_j^p < \infty$ for p with $0 < p < \infty$, T is said to be of class $\mathscr{C}_p$.

THEOREM 6 (Keldyš). *Suppose H is a self-adjoint operator of class $\mathscr{C}_p$ on X for some $p > 0$ and S is compact on X. Let $T = H(I+S)$. If T is injective, then the space of principal vectors of T, as well as the space of principal vectors of T^*, is complete in X. For any $\varepsilon > 0$, all eigenvalues of T, except for possibly finitely many of them, lie in the sectors $-\varepsilon < \arg \lambda < \varepsilon$, $\pi - \varepsilon < \arg \lambda < \pi + \varepsilon$. If H has finitely many negative (positive) eigenvalues, then T has no more than a finite number of eigenvalues in the sector $\pi - \varepsilon < \arg \lambda < \pi + \varepsilon$ $(-\varepsilon < \arg \lambda < \varepsilon)$.*

Proof. See [7, Chap. V, Thm. 8.1 and Remark 8.1]. $\square$

For a compact operator T, and an $r > 0$, let $n(r; T)$ denote the number of characteristic values of T (the reciprocals of the eigenvalues) in the disc $|z| \leqq r$, counted according to (algebraic) multiplicity.

THEOREM 7 (Keldyš). *Let S be a compact operator with $-1 \notin \sigma(S)$, and let H be a positive, compact operator. Set $T = H(I+S)$. Suppose there is a nondecreasing function φ defined on $[0, \infty)$ such that for some $\gamma > 0$,*

$$\frac{\varphi(s)}{\varphi(r)} \leqq \left(\frac{s}{r}\right)^{\gamma} \quad \text{for all sufficiently large } r \text{ with } r < s; \tag{2.1}$$

$$\lim_{r \to \infty} (n(r; H)/\varphi(r)) = 1. \tag{2.2}$$

Then,

$$\lim_{r \to \infty} \frac{n(r; T)}{n(r; H)} = 1.$$

Proof. See [7, Chap. V, Thm. 11.1]. $\square$

For a compact, self-adjoint operator A, let $\mu_1(A)$, $\mu_2(A)$, $\cdots$ denote the eigenvalues of A counted according to multiplicity and ordered in decreasing size of their absolute values.

THEOREM 8 (Kreĭn). *Let T be a self-adjoint operator of the form $T = H(I+S)$, where H is a compact, nonnegative operator, and S is a compact operator with $-1 \notin \sigma(S)$. Then, T has at most finitely many negative eigenvalues, and*

$$\lim_{n \to \infty} \frac{\mu_n(T)}{\mu_n(H)} = 1.$$

Proof. See [7, Chap. V, Thm. 11.4]. $\square$

These theorems can be immediately applied to the operators considered in Theorem 3.

THEOREM 9. *Suppose the hypotheses of Theorem 3 hold, and let T, H, and S be the operators in the conclusion.*

i) *Suppose H is of class $\mathscr{C}_p$ for some $p>0$. Then the principal vectors of T and T^* are complete in X. For any $\varepsilon>0$, all but finitely many eigenvalues of T lie in the sector $-\varepsilon<\arg\lambda<\varepsilon$.*

ii) *If a nondecreasing function φ exists on $[0,\infty)$ satisfying* (2.1) *and* (2.2) *for some $\gamma>0$, then*

$$\lim_{r\to\infty}\frac{n(r;T)}{n(r;H)}=1.$$

iii) *If T is self-adjoint, then T has positive eigenvalues and*

$$\lim_{n\to\infty}\frac{\mu_n(T)}{\mu_n(H)}=1.$$

Proof. The proof follows immediately from Theorems 3, 6, 7, and 8. □

By Theorem 9 and the remarks at the beginning of this section, the conclusions i)–iii) hold for the eigenvalues and principal vectors of the problems $a(u,v)=\lambda(u,v)$. In particular, by iii) if $a(\cdot,\cdot)$ is Hermitian, then T is self-adjoint and the large eigenvalues of $a(u,v)=\lambda(u,v)$ differ from the large eigenvalues of $h(u,v)=\zeta(u,v)$ with small relative error.

3. Examples. Several applications of the results from § 2 are outlined in this section. These applications were chosen to illustrate the flexibility of this approach to studying the spectral properties of elliptic differential eigenvalue problems.

The first example involves an elliptic partial differential operator of order $2m$ with Dirichlet boundary conditions. The spectral properties of these operators were studied by Browder and Agmon (cf. [5], [1], [3]). The second example includes two Steklov type eigenvalue problems. In the first problem, the eigenvalue parameter appears in a boundary condition and in the differential operator. In the second problem, the eigenvalue parameter appears only in the boundary condition. The spectral properties of problems of the latter type on Riemannian manifolds were studied by Koževnikov [10] who used the theory of pseudodifferential operators. In both cases, if the problem is self-adjoint, it is shown here that lower order perturbations in the boundary conditions do not disturb the large eigenvalues very much. The final example involves a system of ordinary differential operators arising in the linearized theory of hydrodynamic stability. This is the general Taylor problem, the spectral properties of which were studied by Di Prima and Habetler [6]. In this example it is particularly easy to apply the results of § 2, and the result is a simpler proof of the completeness of the principal vectors of the general Taylor problem than is in [6].

The following notation and fundamental facts will be used. Let Ω be a bounded domain in $\mathbb{R}^n$ with boundary Γ. For simplicity, Γ will be assumed smooth. Let $dx=dx_1\cdots dx_n$ denote Lebesque measure on Ω, and ds surface measure on Γ. Suppose that $E(\bar\Omega)$ is the space of complex valued C^∞ functions on Ω with C^∞ extensions to Γ, and $\mathscr{D}(\bar\Omega)$ is the subspace of $E(\bar\Omega)$ formed by the functions with compact support in Ω. For any real s, let $H^s(\Omega)$, $H^s(\Gamma)$ denote the Sobolev spaces of order s on Ω and Γ respectively [11]. For any positive integer k, let $H_0^k(\Omega)$ denote the subspace of $H^k(\Omega)$ formed by functions satisfying

$$u=\frac{\partial u}{\partial n}=\cdots=\frac{\partial^{k-1}u}{\partial n}=0$$

on Γ in the sense of trace [11], [4]. Denote by $\|\cdot\|_s$, $(\cdot,\cdot)_s$ the norm and inner product of $H^s(\Omega)$ and by $\langle\cdot\rangle_s$, $\langle\cdot,\cdot\rangle_s$ the norm and inner product of $H^s(\Gamma)$.

For $s_1 > s_2$, $H^{s_1}(\Omega)$ is compactly imbedded in $H^{s_2}(\Omega)$. A similar statement holds for $H^{s_1}(\Gamma)$ and $H^{s_2}(\Gamma)$. For any positive integer k, $H_0^k(\Omega)$ is compactly imbedded in $H^0(\Omega)$, [11].

As the first example, consider the eigenvalue problem

$$
\begin{aligned}
&Au = \lambda u \quad \text{in } \Omega, \\
&u = \frac{\partial u}{\partial n} = \cdots \frac{\partial^{m-1} u}{\partial n^{m-1}} = 0 \quad \text{on } \Gamma,
\end{aligned}
\tag{3.1}
$$

where A is the uniformly strongly elliptic operator of order $2m$ defined by the expression $Au = (-1)^{|\beta|} D^\beta (a^{\alpha\beta} D^\alpha u) + cu$, where α and β are multi-indices, and $0 < |\alpha|, |\beta| \leqq m$. Assume that $a^{\alpha\beta}, c \in E(\bar{\Omega})$, $a^{\alpha\beta}$ is real valued and $a^{\alpha\beta} = a^{\beta\alpha}$ for $|\alpha| = |\beta| = m$, and there is an $\alpha_0 > 0$ with $a^{\alpha\beta}\xi^\alpha\xi^\beta > \alpha_0|\xi|^2$ for any $\xi \in \mathbb{R}^n$. The summation convention is assumed.

Problem (3.1) can be put in the context of § 1 by posing the problem in terms of forms. This will be called a variational formulation of the problem. Set $V = H_0^m(\Omega)$, $X = H^0(\Omega)$, $a(u, v) = \int_\Omega a^{\alpha\beta} D^\alpha u D^\beta \bar{v} + cu\bar{v}\, dx$, $0 < |\alpha|, |\beta| \leqq m$, and $h(u, v) = \int_\Omega a^{\alpha\beta} D^\alpha u D^\beta \bar{v} + (\operatorname{Re} c) u\bar{v}\, dx$, $|\alpha| = |\beta| = m$. The forms $a(\cdot,\cdot)$ and $h(\cdot,\cdot)$ are bounded on V, and $h(\cdot,\cdot)$ is Hermitian. Moreover, it follows from Gårding's estimate (cf. [2]) that (1.1) and (1.2) hold provided $\operatorname{Re} c > 0$ and is large enough. It can be assumed without loss of generality that this is true. For if not, add a suitably large positive constant γ_0 to c. This has the effect of shifting the eigenvalues of (3.1) by γ_0 and leaving the principal vectors unchanged. The results of § 2 for the unshifted problem are then easy to reinterpret.

Recall that the operator $\mathscr{A}^*$ is defined by $a(u, v) = (u, \mathscr{A}^* v)$ for all $u \in V$. It follows from regularity theorems (cf. [11]) that $D(\mathscr{A}^*) = H^{2m}(\Omega) \cap H_0^m(\Omega)$ and $\mathscr{A}^* = A'$ (the formal adjoint of A). Thus, for $v \in H^{2m}(\Omega) \cap H_0^m(\Omega)$, $h(u, \bar{J}v) = k(u, v) = \int_\Omega (u(-1)^{|\alpha|} D^\alpha (a^{\alpha\beta} D^\beta \bar{v}) + i(\operatorname{Im} c) u\bar{v})\, dx$, where $0 < |\alpha| + |\beta| < 2m$. Thus $\bar{J}$ can be immediately extended to any $v \in H^{2m}(\Omega)$. Call this extension $\hat{J}$. It follows from the estimates of [12], that for any integer k such that $1 \leqq k$, $\|\hat{J}v\|_k \leqq C_k \|v\|_{k-1}$ for some positive constant C_k. But then $\hat{J}$ can be extended by continuity to a bounded map $\tilde{J}$ from $X = H^0(\Omega)$ into $H^1(\Omega)$ that is bounded from $H^{k-1}(\Omega)$ into $H^k(\Omega)$, for $1 \leqq k \leqq 2m$. The map $\tilde{J}$ is compact when considered as an operator on $H^0(\Omega)$.

Suppose -1 were an eigenvalue of $\tilde{J}$ with eigenvector $v \in X$. Then $\tilde{J}v = -v$ and consequently $v \in H^{2m}(\Omega) \cap H_0^m(\Omega)$. Thus $v \in \mathscr{D}(\mathscr{A}^*)$, $\tilde{J}v = \bar{J}v = -v$, and $\mathscr{A}^* v = \mathscr{H}(I + \bar{J})v = 0$. But this contradicts the fact that $\mathscr{A}^*$ is injective. Consequently J satisfies the extension property and Theorem 9 applies.

The operator H, defined on $H^0(\Omega)$ by $h(Hf, v) = (f, v)_0$ for all $v \in V$, is of class $\mathscr{C}_{2k}$ if $2mk > n/2$, and $n(r, H) = c_0 r^{n/2m} + o(r^{n/2m})$ as r approaches ∞ (cf. [2]). Thus the principal vectors of (3.1) are complete in $H^0(\Omega)$. The eigenvalues of (3.1) (or a suitable translate of them) are, except for possibly finitely many of them, in the sector $-\varepsilon < \arg \lambda < \varepsilon$, and the number of eigenvalues of (3.1) in the circle of radius r is $c_0 r^{n/2m} + o(r^{n/2m})$.

As a second example, two Stekov type boundary value problems will be considered:

$$
Lu = \lambda u \quad \text{in } \Omega, \qquad \frac{\partial u}{\partial \nu} + \beta u = \lambda u \quad \text{on } \Gamma,
\tag{3.2a}
$$

and

$$Lu=0 \quad \text{in } \Omega, \qquad \frac{\partial u}{\partial \nu}+\beta u=\lambda u \quad \text{on } \Gamma. \tag{3.2b}$$

The operator L is the uniformly strongly elliptic operator of order 2 in Ω given by the expression $Lu=-(a^{ij}u_{,i})_{,j}+b^i u_{,i}+cu$, and $\partial u/\partial\nu \equiv a^{ij}u_{,i}n_j$. In this example $\cdot_{,i} \equiv \partial\cdot/\partial x_i$. Assume that $a^{ij}, b^i, c\in E(\bar\Omega)$, $\beta\in C^{\infty}(\Gamma)$, a^{ij} is real valued, $a^{ij}=a^{ji}$, and there is a constant $\alpha>0$ with $a^{ij}\xi^i\xi^j \geqq \alpha|\xi|^2$ for all $\xi\in\mathbb{R}^n$. Let

$$c_0=\min_{\substack{x\in\bar\Omega\\ i=1,\cdots,n}} |b^i(x)|, \quad \text{and} \quad \beta_0=\min_{s\in\Gamma} \operatorname{Re}\beta(s).$$

It can be assumed without loss of generality in (3.2a) that $c_0\geqq b^2/(2\alpha)+\alpha/2$ and $\beta_0>0$ (otherwise add a suitably large positive constant γ to c and β). Suppose that in (3.2b), $c_0\geqq b^2/(2\alpha)+\alpha/2$, and without loss of generality assume $\beta_0>0$.

To get an appropriate variational formulation of (3.2a), let X be the completion of $E(\bar\Omega)$ with respect to the norm $|\cdot|=(\|\cdot\|_0^2+\langle\cdot\rangle_0^2)^{1/2}$, $V=H^1(\Omega)$, and set

$$a(u,v)=\int_\Omega (a^{ij}u_{,i}\bar v_{,j}+b^i u_{,i}\bar v+cu\bar v)\,dx+\int_\Gamma \beta u\bar v\,ds$$

and

$$h(u,v)=\int_\Omega (a^{ij}u_{,i}\bar v_{,j}+(\operatorname{Re} c)u\bar v)\,dx+\int_\Gamma (\operatorname{Re}\beta)u\bar v\,ds.$$

The inner product on X is $(\cdot,\cdot)=(\cdot,\cdot)_0+\langle\cdot,\cdot\rangle_0$.

It follows from the trace theorem that $|\cdot|\leqq C\|\cdot\|_{s+1/2}$ for $s>0$, [11]. Consequently, $V=H^1(\Omega)$ is compactly imbedded in X and since $\|\cdot\|_0\leqq|\cdot|$, X is contained in $H^0(\Omega)$. The forms $a(\cdot,\cdot)$ and $h(\cdot,\cdot)$ are bounded on V and satisfy (1.1) and (1.2). Furthermore, the form

$$k(u,v)=\int_\Omega (b^i u_{,i}+(\operatorname{Im} c)u\bar v)\,dx+\int_\Gamma (\operatorname{Im}\beta)u\bar v\,ds$$

satisfies (1.5), $|k(u,v)|\leqq C\|u\|_1|v|$, as straightforward estimation shows. Thus, Theorems 3 and 9 apply.

If $k+1/2>n/2$, the operator H in this problem is of class $\mathscr{C}_{2k}$ (cf. [9]). Thus, the principal vectors of (3.2a) are complete in X and all but finitely many eigenvalues are in $-\varepsilon<\arg\lambda<\varepsilon$ for any $\varepsilon>0$. Suppose $b^i=0$ and c, β are real valued; then $a(\cdot,\cdot)=h(\cdot,\cdot)$ is Hermitian. If β is perturbed by some real valued function β_0 defined on Γ, then by Theorem 9 part iii), it follows that the large eigenvalues of the perturbed problem differ from those of the unperturbed problem with small relative error. Roughly speaking, lower order perturbations in the boundary condition do not alter the large eigenvalues much.

A variational formulation of (3.2b) is given by

$$a(u,v)=\lambda\langle u,v\rangle_0 \quad \text{for all } u,v\in H^1(\Omega). \tag{3.3}$$

If U is the map from $H^0(\Gamma)$ into $H^1(\Omega)$ defined by: $a(Uf,v)=\langle f,v\rangle_0$ for $f\in H^0(\Gamma)$ and $v\in H^1(\Omega)$, then the eigenvalues of (3.3) are the characteristic values of the compact operator $U\tau$ on $H^1(\Omega)$, where τ is the trace mapping from $H^1(\Omega)$ into $H^{1/2}(\Gamma)$. This variational formulation of (3.2b) does not fit into the context of § 1. However, an auxillary variational problem can be defined on $H^{1/2}(\Gamma)$ that fits into the context of § 1, has the same eigenvalues as (3.3), and has principal vectors that are the traces of the principal vectors of (3.3) on $H^{1/2}(\Gamma)$.

For $\varphi \in H^{1/2}(\Gamma)$, let $\hat{\varphi}$ be the weak solution of the Dirichlet problem $L\hat{\varphi}=0$ in Ω, $\hat{\varphi}=\varphi$ on Γ. The solution $\hat{\varphi}$ is unique and is contained in $H^1(\Omega)$. If $\varphi \in H^{\alpha+1/2}(\Gamma)$ for $\alpha \geqq 0$, then $\hat{\varphi} \in H^{\alpha+1}(\Omega)$ and $\|\hat{\varphi}\|_{\alpha+1} \leqq C\langle\varphi\rangle_{\alpha+1/2}$. By the trace theorem, $(1/C)\langle\varphi\rangle_{1/2} \leqq \|\varphi\|_1$ (cf. [11]). For $\varphi, \psi \in H^{1/2}(\Gamma)$ let $a\langle\varphi, \psi\rangle \equiv a(\hat{\varphi}, \hat{\psi})$, $h\langle\varphi, \psi\rangle \equiv h(\hat{\varphi}, \hat{\psi})$, and $k\langle\varphi, \psi\rangle = k(\hat{\varphi}, \hat{\psi})$, where $a(\cdot, \cdot)$, $h(\cdot, \cdot)$, and $k(\cdot, \cdot)$ are defined as in (3.2a). It is easily seen that these are well-defined bounded forms on $H^{1/2}(\Gamma)$, and $a\langle\cdot, \cdot\rangle$ and $h\langle\cdot, \cdot\rangle$ satisfy (1.1) and (1.2). Let $V=H^{1/2}(\Gamma)$, $X=H^0(\Gamma)$, and consider the variational problem

$$a\langle\varphi, \psi\rangle = \lambda\langle\varphi, \psi\rangle_0 \quad \text{for all } \psi \in V. \tag{3.4}$$

Problem (3.4) fits into the context of Theorem 9 (see below). Moreover, if $\hat{T}$ is the operator defined by $a\langle\hat{T}f, \psi\rangle = \langle f, \psi\rangle_0$ for f in X and all ψ in V, then it is easily seen that $\hat{T}=\tau U$ (cf. [9]). From this it follows that the eigenvalues of (3.3) and (3.4) are identical, and the principal vectors of (3.4) are the traces of the principal vectors of (3.3) on $H^{1/2}(\Gamma)$.

The form $k\langle\cdot, \cdot\rangle$ satisfies (1.5). To see this, note that $|k\langle\varphi, \psi\rangle| = |k(\hat{\varphi}, \hat{\psi})| \leqq C\{\|\hat{\varphi}\|_1\|\hat{\psi}\|_0 + \langle\varphi\rangle_0\langle\psi\rangle_0\} \leqq C'\{\langle\varphi\rangle_{1/2}\|\hat{\psi}\|_0 + \langle\varphi\rangle_0\langle\psi\rangle_0\}$. But by the estimates in [12]; [13]; $\|\hat{\psi}\|_0 \leqq \|\hat{\psi}\|_{1/2} \leqq C''\langle\psi\rangle_0$, and this implies that (1.5) holds for $k\langle\cdot, \cdot\rangle$.

If U_* is the map from $X=H^0(\Gamma)$ into $H^1(\Omega)$ defined by $a(u, U_*f) = \langle u, f\rangle_0$ for f in X and all u in $H^1(\Omega)$, then $\hat{T}^*$, the adjoint of $\hat{T}$ on X, is equal to τU_*. Thus, $\hat{K}^{2k} = (\hat{T}^*\hat{T})^k = (\tau U_* \tau U)^k$. Using the regularity properties of U_* and U, it can be shown that if $2k+\frac{1}{2} > n/2$, $\hat{K}^{2k}$ is of class $\mathscr{C}_2$ and so $\hat{T}$ is of class $\mathscr{C}_{4k}$ (cf. [9]). Since $\hat{T} = \hat{H}(I+\hat{S})$ and $-1 \notin \sigma(\hat{S})$, $\hat{H} = (I+\hat{S})^{-1}\hat{T}$, and thus is itself of class $\mathscr{C}_{4k}$. From Theorem 9, it follows that all but finitely many eigenvalues of (3.2b) are in the sector $-\varepsilon < \arg\lambda < \varepsilon$ for any $\varepsilon > 0$, and the traces of the principal vectors of (3.2b) on $H^{1/2}(\Gamma)$ are complete in $H^0(\Gamma)$.

As the final example, consider the general Taylor problem:

$$\begin{pmatrix} (DD^*-a^2)^2 & a\sqrt{T}f(x) \\ a\sqrt{T}g(x) & -(DD^*-a^2) \end{pmatrix} \underset{\sim}{u} = \lambda \begin{pmatrix} -(DD^*-a^2) & 0 \\ 0 & \nu \end{pmatrix} \underset{\sim}{u}, \tag{3.5}$$

where $D=d/dx$, $D^*=D+\delta/(1+\delta x)$, $0 \leqq \delta \leqq 1$, a, T, and ν are positive constants, $f(x)$, (gx) are in $L^\infty(I)$, and $I=(0, 1)$. Equation (3.5) is assumed to hold for all $x \in I$ with boundary conditions $u_1 = Du_1 = u_2 = 0$ at 0 and at 1, where $\underset{\sim}{u} = (u_1, u_2)^T$.

Let $V = H_0^2(I) \oplus H_0^1(I)$. For $\underset{\sim}{u}, \underset{\sim}{v} \in V$, set

$$h(\underset{\sim}{u}, \underset{\sim}{v}) = \int_0^1 (1+\delta x)\{D^2u_1D^2\bar{v}_1 + 2a^2D^*u_1D^*\bar{v}_1 + a^4u_1\bar{v}_1 + D^*u_2D^*\bar{v}_2 + a^2u_2\bar{v}_2\}\,dx,$$

and

$$\begin{aligned} a(\underset{\sim}{u}, \underset{\sim}{v}) &= h(\underset{\sim}{u}, \underset{\sim}{v}) + k(\underset{\sim}{u}, \underset{\sim}{v}) \\ &= h(\underset{\sim}{u}, \underset{\sim}{v}) + \delta\int_0^1 D^*u_1D(\Delta\bar{v}_1)\,dx + 2\delta\int_0^1 D(\Delta u_1)D\bar{v}_1\,dx \\ &\quad + \int_0^1 (1+\delta x)a\sqrt{T}(fu_2\bar{v}_1 + gu_1\bar{v}_2)\,dx, \end{aligned}$$

where $\Delta = \Delta(x) = \delta/(1+\delta x)$. Let X be the completion of $\mathscr{D}(\bar{I}) \times E(\bar{I})$ with respect to the inner product

$$(\underset{\sim}{u}, \underset{\sim}{v}) = \int_0^1 (1+\delta x)\{D^* u_1 D^* \bar{v}_1 + a^2 u_1 \bar{v}_1 + \nu u_2 \bar{v}_2\}\, dx.$$

Then (3.5) has the variational formulation: $a(\underset{\sim}{u}, \underset{\sim}{v}) = \lambda(\underset{\sim}{u}, \underset{\sim}{v})$ for all $\underset{\sim}{v} \in V$.

Straightforward estimation shows that the inner product $(\cdot, \cdot)$ is equivalent to the standard inner product on $H_0^1(I) \oplus H^0(I)$. Thus, $X \simeq H_0^1(I) \oplus H^0(I)$, and so $V = H_0^2(I) \oplus H_0^1(I)$ is compactly imbedded in X. Furthermore, there is a suitably large positive constant γ_0 such that $a(\underset{\sim}{u}, \underset{\sim}{v}) + \gamma_0(\underset{\sim}{u}, \underset{\sim}{v}) = (h(\underset{\sim}{u}, \underset{\sim}{v}) + \gamma_0(\underset{\sim}{u}, \underset{\sim}{v})) + k(\underset{\sim}{u}, \underset{\sim}{v})$ satisfies (1.1) and (1.2). Finally, it is easily seen that $k(\underset{\sim}{u}, \underset{\sim}{v})$ satisfies (1.5). The operator $\underset{\sim}{H}$, defined by $h(\underset{\sim}{H}\underset{\sim}{f}, \underset{\sim}{v}) + \gamma_0(\underset{\sim}{H}\underset{\sim}{f}, \underset{\sim}{v}) = (\underset{\sim}{f}, \underset{\sim}{v})$ for $\underset{\sim}{f}$ in X and all $\underset{\sim}{v}$ in V, is of class $\mathscr{C}_2$ on X. In fact $\underset{\sim}{H}\underset{\sim}{f} = (H_1 f_1, H_2 f_2)^T$, where H_1 is of class $\mathscr{C}_2$ on $H_0^1(I)$ and H_2 is of class $\mathscr{C}_2$ on $H^0(I)$ (cf. [9]). By Theorem 9, the principal vectors of (3.5) are complete in $H_0^1(I) \oplus H^0(I)$, and all but finitely many of the eigenvalues of (3.5) lie in the sector $-\varepsilon < \arg(\lambda + \gamma_0) < \varepsilon$ for any $\varepsilon > 0$.

REFERENCES

[1] S. AGMON, *On the eigenfunctions and on the eigenvalues of general elliptic boundary value problems*, Comm. Pure Appl. Math., 15 (1962), pp. 119–147.

[2] ———, *Lectures on Elliptic Boundary Value Problems*, Van Nostrand, New York, 1965.

[3] ———, *Asymptotic formulas with remainder estimates for eigenvalues of elliptic operators*, Arch. Rational Mech. Anal., 28 (1968), pp. 165–183.

[4] J-P. AUBIN, *Approximation of Elliptic Boundary Value Problems*, Wiley-Interscience, New York, 1972.

[5] F. BROWDER, *On the eigenfunctions and eigenvalues of the general elliptic differential operator*, Proc. Nat. Acad. Sci. U.S.A., 39 (1959), pp. 433–439.

[6] R. C. DI PRIMA AND G. J. HABETLER, *A completeness theorem for nonselfadjoint eigenvalue problems in hydrodynamic stability*, Arch. Rational Mech. Anal., 34 (1969), pp. 218–227.

[7] I. C. GOHBERG AND M. G. KREĬN, *Introduction to the Theory of Linear Nonselfadjoint Operators*, American Mathematical Society, Providenc, RI, 1969.

[8] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.

[9] W. G. KOLATA, *Spectral approximation and spectral properties of variationally posed, nonselfadjoint problems*, Ph.D. thesis, Univ. of Maryland, College Park, 1976.

[10] A. N. KOŽEVNIKOV, *On the asymptotics of eigenvalues and on completeness of the root vectors of an operator generated by a boundary value problem with a spectral parameter in a boundary condition*, Soviet Math. Dokl., 12 (1971), pp. 1564–1568.

[11] J. L. LIONS AND E. MAGENES, *Problèmes aux Limites non Homogènes et Applications*, vol. I, Dunod, Paris, 1968.

[12] M. SCHECHTER, *On L^p estimates and regularity I*, Amer. J. Math., (1963), pp. 1–13.

[13] ———, *On L^p estimates and regularity II*, Math. Scand., 13 (1963), pp. 47–69.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0010 $01.00/0

SINGULARLY PERTURBED LINEAR STOCHASTIC ORDINARY DIFFERENTIAL EQUATIONS*

G. BLANKENSHIP† AND S. SACHS‡

Abstract. Singularly perturbed linear differential equations with random forcing functions have recently been studied as models of control and filtering systems. The analysis in these studies has been somewhat formal, and important properties of the boundary layer behavior have been neglected as a consequence. In the present paper we examine the asymptotic analysis of systems of this type using some limit theorems from the theory of stochastic processes. We show that a natural separation of time scales occurs between the "outer" and "boundary layer" solutions and their respective stochastic fluctuations. A total of four basic time scales is necessary for a complete description of the solution. The separation of scales is characterized by a parameter ε related to the time constant of the parasitics (fast subsystem) and the correlation time (inverse of bandwidth) of the stochastic fluctuations. We show that certain diffusion processes may be identified as the natural limits as $\varepsilon \to 0$ of the "outer solution" and the "boundary layer correction" of the original system.

1. Introduction and problem statement. Singularly perturbed deterministic ordinary differential equations have received considerable attention over the past decade as models of engineering systems (see [1], [2] for a survey of results and applications). Typically, systems of the form

$$\frac{d}{dt}x(t,\varepsilon)=A_1x(t,\varepsilon)+A_2y(t,\varepsilon)+f(t,\varepsilon) \tag{1.1a}$$

$$\varepsilon\frac{d}{dt}y(t,\varepsilon)=A_3x(t,\varepsilon)+A_4y(t,\varepsilon)+g(t,\varepsilon) \tag{1.1b}$$

$$x(0,\varepsilon)=x_0\in R^n,\quad y(0,\varepsilon)=y_0\in R^m,\quad 0\leqq t\leqq T,$$

have been studied. We identify the "outer solution" $(\bar{x}(t,\varepsilon),\bar{y}(t,\varepsilon))$ of (1.1) as the solution of

$$\frac{d}{dt}\bar{x}(t,\varepsilon)=(A_1-A_2A_4^{-1}A_3)\bar{x}(t,\varepsilon)+f(t,\varepsilon)-A_2A_4^{-1}g(t,\varepsilon) \tag{1.2a}$$

$$\varepsilon\frac{d}{dt}\bar{y}(t,\varepsilon)=A_3\bar{x}(t,\varepsilon)+A_4\bar{y}(t,\varepsilon)+g(t,\varepsilon)$$

$$\bar{x}(0,\varepsilon)=x_0,\qquad \bar{y}(0,\varepsilon)=y_0+A_4^{-1}A_3x_0 \tag{1.2b}$$

and the "boundary layer corrections" as

$$\begin{aligned}X(\tau,\varepsilon)&=x(\varepsilon\tau,\varepsilon)-\bar{x}(\varepsilon\tau,\varepsilon)\\ Y(\tau,\varepsilon)&=y(\varepsilon\tau,\varepsilon)-\bar{y}(\varepsilon\tau,\varepsilon)\end{aligned} \tag{1.3}$$

in the fast time scale

$$\tau=t/\varepsilon. \tag{1.4}$$

Under simple smoothness conditions on f and g, one may prove

$$\lim_{\varepsilon\to 0}(x(t,\varepsilon),y(t,\varepsilon))=(x^0(t),y^0(t)),\qquad 0<t\leqq T, \tag{1.5}$$

---

* Received by the editors March 11, 1977, and in revised form October 17, 1977. This work was supported in part by the National Science Foundation under Grant ENG75-08613.

† Division of Electric Energy Systems, Department of Energy Washington, D.C. 20545. On leave from Case Western Reserve University, Cleveland, Ohio 44106.

‡ Systems Engineering Department, Case Western Reserve University, Cleveland, Ohio 44106.

where $(x^0(t), y^0(t))$ satisfies the "reduced system"

$$\begin{aligned}\frac{d}{dt}x^0(t) &= (A_1 - A_2A_4^{-1}A_3)x^0(t) + f(t, 0) - A_2A_4^{-1}g(t, 0)\\ y^0(t) &= -A_4^{-1}(A_3x^0(t) + g(t, 0)); \qquad x^0(0) = x_0\end{aligned} \tag{1.6}$$

and $\lim_{\varepsilon \downarrow 0} (X(\tau, \varepsilon), Y(\tau, \varepsilon)) = (0, Y^0(\tau))$ where

$$\begin{aligned}\frac{d}{d\tau}Y^0(\tau) &= A_4Y^0(\tau) + \gamma(\tau), \qquad \tau \geqq 0\\ Y^0(0) &= y_0 + A_4^{-1}A_3x_0\end{aligned} \tag{1.7}$$

and

$$\gamma(\tau) = \lim_{\varepsilon \to 0} g(\varepsilon\tau, \varepsilon). \tag{1.8}$$

On the closed interval $[0, T]$ we have the uniform approximations

$$\begin{aligned}x(t, \varepsilon) &= x^0(t) + O(\varepsilon), \qquad 0 \leqq t \leqq \tau,\\ y(t, \varepsilon) &= y^0(t) + Y^0(t/\varepsilon) + O(\varepsilon), \qquad 0 \leqq t \leqq T.\end{aligned} \tag{1.9}$$

The asymptotic analysis of (1.1) as $\varepsilon \to 0$ has two basic features:

(i) *Order reduction* characterized by the difference in dimension between that $(n + m)$ of system (1.1) and that (n) of the reduced system (1.2),

(ii) *Separation of time scales* of order $1/\varepsilon$ between t, the natural time scale of the outer solution, and τ, the natural time scale of the boundary layer correction.

(For an elaboration and details of the analysis summarized above see [1] and the references therein.)

Recently, the analysis of (1.1) has been extended to the case when the functions $f(t, \varepsilon)$, $g(t, \varepsilon)$ are stochastic processes. In each of the papers [3]–[11] it has been assumed that $f(t, \varepsilon)$, $g(t, \varepsilon)$ are "white Gaussian noises" independent of ε. With this hypothesis the outer solution $(x^0(t), y^0(t))$ will be a Gauss–Markov process and the limit (1.5) is a limit of Gaussian processes. However, because $g(t)$ is a white noise, the analysis of the boundary layer behavior (1.3), (1.6) is much more delicate than before, and it is at this point that the formal analysis in [3]–[11] is weakest.

In this paper we reformulate the model (1.1) in the case when the forcing functions are stochastic processes in a way which clarifies the relationships among the time scales of the slow (1.1a) and the fast (1.1b) subsystems and their stochastic elements. Specifically, we consider the system

$$\frac{d}{dt}x = A_1(t)x + A_2(t)y + \frac{1}{\varepsilon}f(t/\varepsilon^2) + \frac{1}{\varepsilon}g(t/\varepsilon^3) \tag{1.10a}$$

$$\varepsilon\frac{d}{dt}y = A_3(t)x + A_4(t)y + \frac{1}{\varepsilon}h(t/\varepsilon^2) + \frac{1}{\varepsilon}j(t/\varepsilon^3) \tag{1.10b}$$

$$x(0) = x_0 \in R^n, \quad y(0) = y_0 \in R^m, \quad 0 \leqq t \leqq T.$$

Here $A_i(t)$, $i = 1, \cdots, 4$ are matrices of appropriate dimensions, and we assume that the stochastic processes $f(s)$, $g(s)$, $h(s)$, $j(s)$ are independent, zero mean processes which are rapidly mixing in a sense made precise below (equation (2.2)).

To illustrate the asymptotic behavior of the stochastic elements, suppose that the processes f, g, h, j are bandlimited. That is, we define

$$\Phi(s) = Ef(t)f^T(t+s)$$
$$S(\omega) = \int_0^\infty e^{i\omega t}\Phi(t)\,dt \tag{1.11}$$

and suppose $S(\omega) = 0$, $|\omega| \geqq \omega_0 > 0$. Consider the process

$$f^\varepsilon(t) = \frac{1}{\varepsilon} f(t/\varepsilon^2); \tag{1.12}$$

then

$$\begin{aligned}\Phi^\varepsilon(s) &= Ef^\varepsilon(t)(f^\varepsilon(t+s))^T \\ &= \frac{1}{\varepsilon^2}\Phi(s/\varepsilon^2) \\ &\xrightarrow[\varepsilon\downarrow 0]{} \left(\int_{-\infty}^{\infty}\Phi(u)\,du\right)\delta(s)\end{aligned} \tag{1.13}$$

and

$$\begin{aligned}S^\varepsilon(\omega) &= \int_0^\infty e^{i\omega t}\Phi^\varepsilon(t)\,dt \\ &= S(\varepsilon^2\omega) \\ &= 0, \qquad |\omega| \geqq \omega_0/\varepsilon^2.\end{aligned} \tag{1.14}$$

That is, we may regard $1/\varepsilon^2$ as the normalized bandwidth of process $f^\varepsilon(t)$, and, as the bandwidth approaches infinity, $f^\varepsilon(t)$ approaches a white noise. (The noise $(1/\varepsilon)g(t/\varepsilon^3) \to 0$ as $\varepsilon \to 0$ and is introduced in (1.10a) for symmetry.[1])

The ε on the left side of (1.10b) summarizes the effect of "parasitic elements" (see [2]), defining the length of time ($O(\varepsilon)$) over which those elements have a significant effect on the system behavior. In writing the model (1.10) we have presupposed a normalization of the time t so that ε also parameterizes the stochastic fluctuations. That is, suppose $\mu > 0$ is identified as the normalized time constant of the parasitic elements in the usual way [2] and $1/\rho^2$ is identified as the normalized bandwidth of the disturbances as above; i.e., in (1.10), $f^\rho(t) = (1/\rho)f(t/\rho^2)$, etc. Suppose there exist $a, b, \varepsilon > 0$ so that

$$\rho^2 = a^2\varepsilon^2, \qquad \mu = b\varepsilon. \tag{1.15}$$

If we define the new time scale $s = t/b$ and the new variables

$$\begin{aligned}&\tilde{A}_i(s) = bA_i(bs), \quad i = 1, 2, \qquad \tilde{A}_i(s) = A_i(bs), \qquad i = 3, 4. \\ &\tilde{f}^\varepsilon(s) = \frac{b}{a\varepsilon} f\left(\frac{bs}{a^2\varepsilon^2}\right), \qquad \tilde{g}^\varepsilon(s) = \frac{b}{a\varepsilon} g\left(\frac{bs}{a^3\varepsilon^3}\right) \\ &\tilde{h}^\varepsilon(s) = \frac{1}{a\varepsilon} h\left(\frac{bs}{a^2\varepsilon^2}\right), \qquad \tilde{j}^\varepsilon(s) = \frac{1}{a\varepsilon} j\left(\frac{bs}{a^3\varepsilon^3}\right) \\ &\tilde{x}(s) = x(bs), \qquad \tilde{y}(s) = y(bs),\end{aligned} \tag{1.16}$$

[1] Convergence is weak convergence in the set of probability measures on $C([0, T])$. See Lemma 3.1 below.

then the normalized system (1.10) results. Thus, (1.10) implicitly assumes a relationship of the form (1.15) between the scalings of the parasitic elements and the stochastic fluctuations.

The asymptotic analysis of (1.10) proceeds as follows. We identify

$$\dot{F}(t)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}f(t/\varepsilon^2),\qquad \dot{H}(t)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}h(t/\varepsilon^2) \tag{1.17}$$

as the white noise limits of $f^\varepsilon(t)$ and $h^\varepsilon(t)$ in the sense of (1.13), (1.14). Then the pair $(x^0(t), y^0(t))$ defined by

$$\begin{aligned}\frac{d}{dt}x^0(t)&=[A_1(t)-A_2(t)A_4^{-1}(t)A_3(t)]x^0(t)-\sqrt{2}A_2(t)A_4^{-1}(t)\dot{H}(t)+\sqrt{2}\dot{F}(t),\\ x^0(0)&=x_0,\quad 0\leqq t\leqq T,\\ y^0(t)&=-A_4^{-1}(t)(A_3(t)x^0(t)+\sqrt{2}\dot{H}(t))\end{aligned} \tag{1.18}$$

is a logical candidate for the limiting solution of (1.10) as $\varepsilon\to 0$. The limit $Y^0(\tau)=\lim_{\varepsilon\to 0}Y^\varepsilon(\tau)$ of

$$\begin{aligned}\frac{d}{d\tau}Y^\varepsilon(\tau)&=A_4(0)Y^\varepsilon(\tau)+\frac{1}{\varepsilon}j(\tau/\varepsilon^2)\\ Y^\varepsilon(0)&=y_0-y^0(0),\qquad \tau\geqq 0,\end{aligned} \tag{1.19}$$

is a logical candidate for the "boundary layer correction." We may anticipate that

$$\begin{aligned}\frac{d}{d\tau}Y^0(\tau)&=A_4(0)Y^0(\tau)+\sqrt{2}\dot{J}(\tau)\\ Y^0(0)&=y_0-y^0(0),\qquad \tau\geqq 0,\end{aligned} \tag{1.20}$$

where $\dot{J}(\tau)=\lim_{\varepsilon\downarrow 0}(1/\varepsilon)j(\tau/\varepsilon^2)$ is a white noise.

In § 2 we give conditions on $A_i(t)$, $f(t),\cdots$, so that these conclusions hold. A precise statement of the results is given in the theorem and its proof.

2. Formulation and statement of the theorem. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $f(s)$, $g(s)$, $h(s)$, $j(s)$, $s\geqq 0$ be random processes defined on $(\Omega, \mathcal{F}, P)$. We assume

(A1) The processes $f(s),\cdots,j(s)$; $s\geqq 0$ are independent.

(A2) Let $k(s)$ stand for any one of $(f(s),\cdots,j(s)$; *then* $k(s)$ is a zero mean, stationary, ergodic process with autocorrelation

$$R(t)=Ek(s)k^T(s+t) \tag{2.1}$$

which satisfies

$$\int_0^\infty t\|R(t)\|\,dt<\infty \tag{2.2}$$

where $\|\cdot\|$ is an appropriate matrix norm. Also, $k(s)$ has bounded trajectories which are right continuous and have left-hand limits at every point.

(A3) The matrices $A_i(t)$, $i=1,\cdots,4$, are continuously differentiable on $0\leqq t\leqq T$ and independent of ε. Also, all the eigenvalues of $A_4(t)$ have negative real parts for $0\leqq t\leqq T$.

LEMMA 2.1. *Let $k(s)$, $s \geqq 0$ satisfy* (A2) *and let*

$$K^\varepsilon(t) = \frac{1}{\varepsilon} \int_0^t k(s/\varepsilon^2)\, ds, \qquad 0 \leqq t \leqq T. \tag{2.3}$$

Then $K^\varepsilon(t)$ converges weakly[2] *in $C[0, T]$ to $K^0(t)$, a Wiener process with zero mean and covariance*

$$EK^0(t)(K^0(s))^T = \left(\int_0^\infty 2R(u)\, du \right) \min(t, s). \tag{2.4}$$

Proof. See Khas'minskii [12, Lemmas 3.1, 3.2, 3.3], where it is also shown that there is a $c > 0$, independent of ε, so that

$$E|K^\varepsilon(t+\delta) - K^\varepsilon(t)|^4 \leqq c\delta^2, \qquad \sup_{0 \leqq t \leqq T} E|K^\varepsilon(t)|^2 \leqq c. \tag{2.5}$$

LEMMA 2.2. *Let $A(t)$ be an $n \times n$ matrix-valued, continuous function on $[0, T]$ and let $k(s)$, $s \geqq 0$, satisfy* (A2). *Then the process $x^\varepsilon(s)$ defined by*

$$\begin{aligned} \frac{d}{ds} x^\varepsilon(s) &= A(s) x^\varepsilon(s) + \frac{1}{\varepsilon} k(s/\varepsilon^2) \\ x^\varepsilon(0) &= x_0, \qquad 0 \leqq s \leqq T, \end{aligned} \tag{2.6}$$

approaches weakly in $C[0, T]$ the solution $x^0(s)$ of

$$\begin{aligned} dx^0(s) &= A(s) x^0(s)\, ds + \sqrt{2}\, dK^0(s) \\ x^0(0) &= x_0, \qquad 0 \leqq s \leqq T, \end{aligned}$$

where $K^0(s)$ is the Wiener process identified in Lemma 2.1. *Moreover, for any smooth function $f: R^n \to R$, $s \geqq t$.*

$$|E[f(x^\varepsilon(s)) | x^\varepsilon(t) = x, \mathscr{F}_0^{t/\varepsilon^2}] - E[f(x^0(s)) | x^0(t) = x]| \leqq \varepsilon c(f, T, x) \tag{2.7}$$

for some constant $c > 0$, dependent on f, T, x, but independent of ε. Here $\mathscr{F}_0^t$ is the σ-algebra generated by $k(r)$, $0 \leqq r \leqq t$.

Proof. See Khas'minskii [13], Papanicolaou and Kohler [14], and [15].

Remark 2.1 [14]. The result (2.7) is not implied by weak convergence; in fact, it plays an important role in our proof of weak convergence. As a consequence of (2.7), moments of any order of $x^\varepsilon(s)$ converge to the corresponding moments of any order of $x^0(s)$. The $O(\varepsilon)$ estimate on the right of (2.7) is consistent with the central limit theorem and is the best possible. It was not achieved by Khas'minskii in [12], [13].

Remark 2.2. Suppose that the process $k(s)$, $s \geqq 0$, is Markov with state space $S \subset R^n$ and generator Q. Then the pair $(x^\varepsilon(s), k^\varepsilon(s))$ is Markov with state space $R^n \times (S \cdot 1/\varepsilon)$ and generator

$$L_t^\varepsilon = (A(t)x)^T \frac{\partial}{\partial x} + \frac{1}{\varepsilon} k^T \frac{\partial}{\partial x} + \frac{1}{\varepsilon^2} Q. \tag{2.8}$$

As $\varepsilon \to 0$, $x^\varepsilon(s)$ approaches a diffusion Markov process $x^0(s)$ with state space R^n and generator

$$\bar{L}_t = (A(t)x)^T \frac{\partial}{\partial x} + \operatorname{trace}\left[\int_0^\infty R(u)\, du \frac{\partial^2}{\partial x^2} \right] \tag{2.9}$$

[2] That is, weak convergence of the associated probability distributions. See [12], [17] for details.

(here $\partial/\partial x = \text{gradiant}$, $\partial^2/\partial x^2 = \text{Laplacian}$). Thus, the limiting behavior of (2.6) as $\varepsilon \to 0$ is singular in the sense that an order reduction occurs. This is parallel to the observation (i) made in § 1 on order reduction in singularly perturbed deterministic problems. Note also that the natural time scale of the disturbance process $k^\varepsilon(s)$ is s/ε^2 relative to the natural time scale s of the system variable $x^\varepsilon(s)$. Thus, the system (2.6) exhibits a separation of time scales analogous to observation (ii) in § 1.

Let $f(t)$, $g(t)$, $h(t)$, $j(t)$, $t \geqq 0$, be the processes in (1.10) and let

$$\begin{aligned} \Phi(t) &= Ef(t)f^T(0), \qquad \Gamma(t) = Eg(t)g^T(0) \\ \Pi(t) &= Eh(t)h^T(0), \qquad \Lambda(t) = Ej(t)j^T(0) \end{aligned} \tag{2.10}$$

and let $F(t)$, $G(t)$, $H(t)$, $J(t)$, $t \geqq 0$, be the limiting Wiener processes associated with $f^\varepsilon(t), \cdots, j^\varepsilon(t)$ in the sense of Lemma 2.1.

Associated with (1.10) we define an "outer solution" $(\bar{x}^\varepsilon(s), \bar{y}^\varepsilon(s))$ by

$$\begin{aligned} \frac{d\bar{x}^\varepsilon(s)}{ds} &= [A_1(s) - A_2(s)A_4^{-1}(s)A_3(s)]\bar{x}^\varepsilon(s) + \frac{1}{\varepsilon}f(s/\varepsilon^2) - A_2(s)A_4^{-1}(s)\frac{1}{\varepsilon}h(s/\varepsilon^2) \\ \varepsilon\frac{d\bar{y}^\varepsilon(s)}{ds} &= A_3(s)\bar{x}^\varepsilon(s) + A_4(s)\bar{y}^\varepsilon(s) + \frac{1}{\varepsilon}h(s/\varepsilon^2) \\ \bar{x}^\varepsilon(0) &= x_0, \quad \bar{y}^\varepsilon(0) = -A_4^{-1}(0)A_3(0)x_0, \quad 0 \leqq s \leqq T, \end{aligned} \tag{2.11}$$

and the "boundary layer corrections"

$$\begin{aligned} X^\varepsilon(\sigma) &= x^\varepsilon(\varepsilon\sigma) - \bar{x}^\varepsilon(\varepsilon\sigma) \\ Y^\varepsilon(\sigma) &= y^\varepsilon(\varepsilon\sigma) - \bar{y}^\varepsilon(\varepsilon\sigma). \end{aligned} \tag{2.12}$$

THEOREM. *Under assumptions* (A1)–(A3) *the system* (1.10) has a unique solution $x^\varepsilon(t)$, $y^\varepsilon(t)$, $0 \leqq t \leqq T$ *which satisfies*

$$\lim_{\varepsilon \downarrow 0} x^\varepsilon(s) = x^0(s) \tag{2.13}$$

$$\lim_{\varepsilon \downarrow 0} \left[\int_0^s y^\varepsilon(t)\,dt - w^0(s)\right] = 0 \tag{2.14}$$

where the limits are taken weakly in $C[0, T]$ *and* $x^0(s)$, $w^0(s)$ *satisfy*

$$\begin{aligned} dx^0(s) &= (A_1(s) - A_2(s)A_4^{-1}(s)A_3(s))x^0(s)\,ds - \sqrt{2}A_2(s)A_4^{-1}(s)\,dH(s) + \sqrt{2}\,dF(s) \\ x^0(0) &= x_0, \qquad 0 \leqq s \leqq T. \end{aligned} \tag{2.15}$$

$$\begin{aligned} dw^0(s) &= -A_4^{-1}(s)[A_3(s)x^0(s)\,ds + \sqrt{2}\,dH(s)] \\ w^0(0) &= 0, \qquad 0 \leqq s \leqq T. \end{aligned} \tag{2.16}$$

The boundary layers $X^\varepsilon(\sigma)$, $Y^\varepsilon(\sigma)$ *satisfy*

$$\lim_{\varepsilon \downarrow 0} X^\varepsilon(\sigma) = 0, \qquad \lim_{\varepsilon \downarrow 0} Y^\varepsilon(\sigma) = Y^0(\sigma) \tag{2.17}$$

(*weakly in* $C[0, T_1]$ *for any fixed* $T_1 > 0$) *where*

$$\begin{aligned} dY^0(\sigma) &= A_4(0)Y^0(\sigma)\,d\sigma + \sqrt{2}\,dJ(\sigma) \\ Y^0(0) &= y_0 + A_4^{-1}(0)A_3(0)x_0, \qquad 0 \leqq \sigma \leqq T_1. \end{aligned} \tag{2.18}$$

Remark 2.3. Note that the limiting behavior of the system (1.10) is singular in two senses—first, in the sense of an order reduction as in deterministic problems (Remark 1.1), and second, in the stochastic sense (of Remark 2.2) of a decrease in the size of the past information σ-algebras (or state in the Markov case).

Remark 2.4. Evidently, the process $g(t)$ plays no role in the limiting behavior. The process $(1/\varepsilon)j(s/\varepsilon^3)$ acts as a wide band noise disturbance to the "fast subsystem" which describes the boundary layer correction $Y^\varepsilon(\sigma)$. Its presence completes the symmetry between (2.15) and (2.18). Note that the system (1.10) has four basic time scales: (i) s, the natural scale of the dominant system state $x^\varepsilon(s)$; (ii) $\sigma = s/\varepsilon$, the fast time scale of the boundary layer correction $Y^\varepsilon(\sigma)$; (iii) $\tau = s/\varepsilon^2$, the time scale of the disturbances to the outer solution $(\bar{x}^\varepsilon(s), \bar{y}^\varepsilon(s))$; and (iv) $\sigma/\varepsilon^2 = s/\varepsilon^3$, the time scale of the disturbances to the boundary layer correction $Y^\varepsilon(\sigma)$. Or, in another sense there are two pairs of time scales $(s, s/\varepsilon^2)$, $(\sigma, \sigma/\varepsilon^2)$. In a stochastic system with l small parameters $\varepsilon_1, \cdots, \varepsilon_l$ (see for example, [16, § 3], and [10, Chap. 4]) there would be $2l$ pairs of time scales.

Remark 2.5. The network of Fig. 1 has the description

$$\begin{aligned} \frac{d}{dt}p &= A_1p + A_2q + \varepsilon\beta(t) + bu(t) \\ \varepsilon\frac{d}{dt}q &= A_3p + A_4q + \varepsilon\gamma(t) \end{aligned} \tag{2.19}$$

where $a = M^2 - L_1L_2$, $p(t) = [x(t), y(t)]^T$, $q(t) = [z(t), w(t)]^T$, and

$$A_1 = \frac{1}{a}\begin{pmatrix} L_1R_2 & -R_1M \\ -R_2M & R_1L_2 \end{pmatrix}, \quad A_2 = \frac{1}{a}\begin{pmatrix} -M & L_1 \\ L_2 & -M \end{pmatrix}, \quad b = \frac{1}{a}\begin{pmatrix} M - L_1 \\ M - L_2 \end{pmatrix}$$

$$\begin{aligned} &A_3 = \begin{pmatrix} 0 & C_1^{-1} \\ C_2^{-1} & 0 \end{pmatrix}, \qquad A_4 = -\begin{pmatrix} 1/(R_3C_1) & 0 \\ 0 & 1/(R_4C_2) \end{pmatrix} \\ &\beta(t) = \begin{pmatrix} \beta_1(t) \\ \beta_2(t) \end{pmatrix}, \qquad \beta_1 = \frac{1}{a}(L_1\eta_2 - M\eta_1) \\ &\beta_2 = \frac{1}{a}(L_2\eta_1 - M\eta_2) \qquad \gamma_1 = -\nu_1/C_1, \qquad \gamma_2 = -\nu_2/C_2. \end{aligned} \tag{2.20}$$

Using a result of Chang [19, p. 522], we introduce the variables

$$\tilde{q}(t) = q + T(t, \varepsilon)p, \qquad p = \tilde{p} + \varepsilon S(t, \varepsilon)\tilde{q} \tag{2.21}$$

where T, S satisfy

$$\begin{aligned} \frac{dT}{dt} &= \varepsilon^{-1}A_4T - TA_1 + TA_{12}T - \varepsilon^{-1}A_3 \\ \frac{dS}{dt} &= (A_1 - A_2T)S - \varepsilon^{-1}S(A_4 + \varepsilon TA_2) - \varepsilon^{-1}A_3. \end{aligned} \tag{2.22}$$

Existence of solutions $S(t, \varepsilon)$, $T(t, \varepsilon)$ uniformly bounded on $[0, T]\times[0, \varepsilon_0]$, for some $\varepsilon_0 > 0$, is guaranteed by the simple structure of $A_1, \cdots, A_4$. Moreover for $0 < t < T$,

$$\lim_{\varepsilon\downarrow 0} T(t, \varepsilon) = A_4^{-1}A_3, \qquad \lim_{\varepsilon\downarrow 0} S(t, \varepsilon) = -A_2A_4^{-1} \tag{2.23}$$

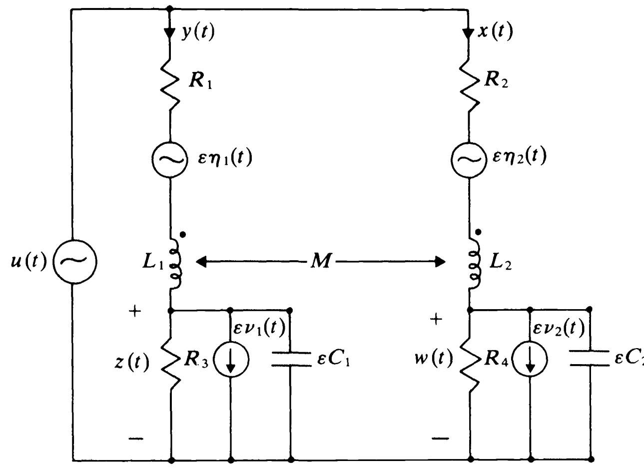

FIG. 1. *Network containing parasitic and fluctuating elements.*

In terms of the variables $\tilde{p}$, $\tilde{q}$, (2.19) becomes (assume $u(t)=0$ for simplicity)

(2.24)
$$\frac{d\tilde{p}}{dt}=\tilde{A}_1(t,\varepsilon)\tilde{p}+\varepsilon\tilde{\beta}(t,\varepsilon)$$
$$\varepsilon\frac{d\tilde{q}}{dt}=\tilde{A}_4(t,\varepsilon)\tilde{q}+\varepsilon\tilde{\gamma}(t,\varepsilon)$$

where

(2.25)
$$\tilde{A}_1=A_1-A_2T$$
$$\tilde{A}_4=A_4+\varepsilon TA_2$$

and $\tilde{\beta}(t,\varepsilon)$, $\tilde{\gamma}(t,\varepsilon)$, are obvious. To apply the results of the Theorem, we assume

(2.26)
$$\tilde{A}_i(t,\varepsilon)=A_{i0}(t)+\varepsilon^2\bar{A}_i(t,\varepsilon),\qquad i=1,4,$$

where $\bar{A}_i(t,\varepsilon)$ is regular in ε. Let $\varphi_1(t)$ be the fundamental matrix associated with A_{10} on $[0, T]$, and $\Phi_4(t)$ with A_{40}. Introducing the variables

(2.27)
$$\hat{p}(t)=\Phi_1^{-1}(t)\tilde{p}(t),\qquad \hat{q}(t)=\Phi_4^{-1}(t)\tilde{q}(t)$$
$$\hat{\beta}(t,\varepsilon)=\Phi_1^{-1}(t)\tilde{\beta}(t,\varepsilon),\qquad \hat{\gamma}(t,\varepsilon)=\Phi_4^{-1}(t)\tilde{\gamma}(t,\varepsilon)$$
$$\hat{A}_i(t,\varepsilon)=\Phi_i^{-1}(t)\bar{A}_i\Phi_i(t),\qquad i=1,4,$$

we have

(2.28)
$$\frac{d\hat{p}}{dt}=\varepsilon^2\hat{A}_1(t,\varepsilon)\hat{p}+\varepsilon\hat{\beta}(t,\varepsilon)$$
$$\varepsilon\frac{d\hat{q}}{dt}=\varepsilon^2\hat{A}_4(t,\varepsilon)\hat{q}+\varepsilon\hat{\gamma}(t,\varepsilon).$$

Assuming that $\hat{\beta}(t)$ and $\hat{\gamma}(t)$ have zero mean, that

(2.29)
$$\hat{\beta}(t)=f(t)+g(t/\varepsilon)$$
$$\hat{\gamma}(t)=h(t)+j(t/\varepsilon),$$

and introducing the slow time variable $s=\varepsilon^2 t$ in (2.28) and the variables

$$x^\varepsilon(s)=\hat{p}(s/\varepsilon^2), \qquad y^\varepsilon(s)=\hat{q}(s/\varepsilon^2)$$
$$A_i(s)=\hat{A}_i(s/\varepsilon^2), \qquad i=1,4, \tag{2.30}$$

we can convert (2.28) into the form (1.10). The hypotheses (2.26) and (2.29) are easy to interpret in terms of the circuit in Fig. 1. Note that the small fluctuation noise sources $\varepsilon\beta(t)$, $\varepsilon\gamma(t)$ in Fig. 1 affect the system dynamics at times t greater than ε^{-2}, and that the effect is a diffusion process.

3. Proof of theorem.

LEMMA 3.1. *Let $g(t)$ satisfy* (A2) *and let*

$$G^\varepsilon(s)=\frac{1}{\varepsilon}\int_0^s g(t/\varepsilon^3)\,ds; \tag{3.1}$$

then $\lim_{\varepsilon\downarrow 0} G^\varepsilon(s)=0$ *in* $C[0,T]$.

Proof of lemma. Let $\Gamma(t)$ be the autocorrelation function of $g(t)$. Then

$$\begin{aligned} E|G^\varepsilon(s)| &\leqq \left(\frac{1}{\varepsilon^2}\int_0^s\int_0^s \operatorname{tr}\Gamma\left(\frac{t-u}{\varepsilon^3}\right)dt\,du\right)^{1/2} \\ &= \left(\varepsilon^4\int_0^{s/\varepsilon^3}\int_0^{s/\varepsilon^3}\operatorname{tr}\Gamma(t-u)\,dt\,du\right)^{1/2} \\ &\leqq \left(\varepsilon sn\int_0^\infty \|\Gamma(t)\|\,dt\right)^{1/2}. \end{aligned} \tag{3.2}$$

Here tr is trace and n is the dimension of $g(t)$. Thus, $\lim_{\varepsilon\downarrow 0} E|G^\varepsilon(s)|=0$. Also, using (2.2) and an integral estimate of Khas'minskii [12, Lemma 2.1], we have

$$\begin{aligned} E|G_i^\varepsilon(s+\delta)-G_i^\varepsilon(s)|^4 &= \varepsilon^8\int_{s/\varepsilon^3}^{(s+\delta)\varepsilon^3}\cdots\int_{s/\varepsilon^3}^{(s+\delta)/\varepsilon^3} Eg_i(t_1)\cdots g_i(t_4)\,dt_1\cdots dt_4 \\ &\leqq \varepsilon^2\delta^2 c \end{aligned} \tag{3.3}$$

for some constant c depending on Γ and T, and each component $G_i^\varepsilon(t)$ of $G^\varepsilon(t)$. Since $G^\varepsilon(0)=0$, we conclude that $E|G^\varepsilon(s)|^4<\infty$, and this together with (3.2) and (3.3) implies weak convergence [17]. Q.E.D. (Lemma 3.1.)

LEMMA 3.2. *Let $A_i(s)$, $i=1,\cdots,4$, and f, g, h satisfy* (A1)–(A3) *and let $\bar{x}^\varepsilon(s)$, the "outer solution," satisfy*

$$\begin{aligned} \frac{d}{ds}\bar{x}^\varepsilon(s) &= [A_1(s)-A_2(s)A_4^{-1}(s)A_3(s)]\bar{x}^\varepsilon(s)-\frac{1}{\varepsilon}A_2(s)A_4^{-1}(s)h(s/\varepsilon^2)+\frac{1}{\varepsilon}f(s/\varepsilon^2) \\ \bar{x}^\varepsilon(0) &= x_0, \qquad 0\leqq s\leqq T; \end{aligned} \tag{3.4}$$

then with $x^0(s)$ given by (2.15) *we have*

$$\lim_{\varepsilon\downarrow 0}\bar{x}^\varepsilon(s)=x^0(s). \tag{3.5}$$

Proof of Lemma 3.2. Let

$$F^\varepsilon(s)=\frac{1}{\varepsilon}\int_0^s f(t/\varepsilon^2)\,dt, \qquad \tilde{H}^\varepsilon(s)=-\frac{1}{\varepsilon}\int_0^s A_2(t)A_4^{-1}(t)h(t/\varepsilon^2)\,dt; \tag{3.6}$$

then there exists a continuous reciprocal kernel $L(s, t)$ so that

$$\bar{x}^{\varepsilon}(s)=\tilde{H}^{\varepsilon}(s)+F^{\varepsilon}(s)+\int_0^s L(s, t)[\tilde{H}^{\varepsilon}(t)+F^{\varepsilon}(t)]\,dt+\bar{x}_0(s) \tag{3.7}$$

where $\bar{x}_0(s)$ is the initial condition response. To see this, use the variation of constants formula for the solution of (3.4) and integrate by parts. Since $\bar{x}^{\varepsilon}(s)$ in (3.7) is a continuous functional of $(F^{\varepsilon}, \tilde{H}^{\varepsilon})$, and $(F^{\varepsilon}, \tilde{H}^{\varepsilon})$ converges weakly to $(F^0, \tilde{H}^0)$ for some Brownian motions $F^0(s), \tilde{H}^0(s)$ (by Lemmas 2.1, 3.1), $(x^0, F^0, \tilde{H}^0)$ is a limit process for $(x^{\varepsilon}, F^{\varepsilon}, \tilde{H}^{\varepsilon})$ in the sense of weak convergence and it must satisfy (2.15). Uniqueness of solutions of (2.15) gives the desired conclusion. Q.E.D. (Lemma 3.2.)

Remark 3.1. It is a simple matter to extend the moment estimate (2.7) to show that $\bar{x}^{\varepsilon}(s)$ has moments of order four bounded on $0\leqq s\leqq T$.

Now by the assumption (A3) there exists $a>0$ such that the real part of every eigenvalue of $A_4(t)$ has absolute value $\geqq 2a$. Therefore by [18, Lemma 1], the linear equation

$$\varepsilon\frac{dp}{dt}=A_4(t)p, \qquad 0\leqq t\leqq T, \tag{3.8}$$

has a fundamental matrix $P^{\varepsilon}(t)$ satisfying

$$\|P^{\varepsilon}(t)(P^{\varepsilon}(s))^{-1}\|\leqq \exp(-a(t-s)/\varepsilon), \qquad 0\leqq s\leqq t\leqq T, \tag{3.9}$$

for some $c>0$ independent of ε.

LEMMA 3.3. *Under assumptions* (A1)–(A3)

$$\lim_{\varepsilon\downarrow 0} x^{\varepsilon}(s)=x^0(s) \tag{3.10}$$

weakly in $C[0, T]$.

Proof. With $\bar{x}^{\varepsilon}(s)$ from (3.4) let $z^{\varepsilon}(s)=x^{\varepsilon}(s)-\bar{x}^{\varepsilon}(s)$; then

$$\begin{gathered}\frac{d}{ds}z^{\varepsilon}(s)=A_1(s)z^{\varepsilon}(s)+A_2(s)y^{\varepsilon}(s)+\frac{1}{\varepsilon}g(s/\varepsilon^3)+A_2(s)A_4^{-1}(s)\left[A_3(s)\bar{x}^{\varepsilon}(s)+\frac{1}{\varepsilon}h(s/\varepsilon^2)\right]\\ z^{\varepsilon}(0)=0, \qquad 0\leqq s\leqq T.\end{gathered} \tag{3.11}$$

Now from (1.10b)

$$\begin{aligned}y^{\varepsilon}(s)=\varepsilon A_4^{-1}(s)\frac{d}{ds}y^{\varepsilon}(s)-A_4^{-1}(s)A_3(s)x^{\varepsilon}(s)\\ -A_4^{-1}(s)\frac{1}{\varepsilon}h(s/\varepsilon^2)-A_4^{-1}(s)\frac{1}{\varepsilon}j(s/\varepsilon^3).\end{aligned} \tag{3.12}$$

Let $A(t)=A_1(t)-A_2(t)A_4^{-1}(t)A_3(t)$; then using (3.11), (3.12) we have

$$\begin{aligned}\frac{d}{ds}z^{\varepsilon}(s)=A(s)z^{\varepsilon}(s)+\varepsilon A_2(s)A_4^{-1}(s)\frac{d}{ds}y^{\varepsilon}(s)\\ -A_2(s)A_4^{-1}(s)\frac{1}{\varepsilon}j(s/\varepsilon^3)+\frac{1}{\varepsilon}g(s/\varepsilon^3)z^{\varepsilon}(0)=0\end{aligned} \tag{3.13}$$

or

$$
\begin{aligned}
z^\varepsilon(s) = &\int_0^s A(t) z^\varepsilon(t)\,dt + \varepsilon A_2(s) A_4^{-1}(s) y^\varepsilon(s) \\
&- \varepsilon A_2(0) A_4^{-1}(0) y_0 - \varepsilon \int_0^s \frac{d}{dt}(A_2 A_4^{-1})(t) y^\varepsilon(t)\,dt \\
&- \frac{1}{\varepsilon}\int_0^s A_2(t) A_4^{-1}(t) j(t/\varepsilon^3)\,dt + \frac{1}{\varepsilon}\int_0^s g(t/\varepsilon^3)\,dt.
\end{aligned} \tag{3.14}
$$

Let $P^\varepsilon(t)$ be the fundamental matrix associated with $A_4(t)/\varepsilon$ via (3.8) and let $P^\varepsilon(t, r) = P^\varepsilon(t)(P^\varepsilon(r))^{-1}$. Then

$$
\begin{aligned}
y^\varepsilon(t) = &P^\varepsilon(t) y_0 + \frac{1}{\varepsilon}\int_0^t P^\varepsilon(t, r) A_3(r) z^\varepsilon(r)\,dr + \frac{1}{\varepsilon}\int_0^t P^\varepsilon(t, r) A_3(r) \bar{x}^\varepsilon(r)\,dr \\
&+ \frac{1}{\varepsilon}\int_0^t P^\varepsilon(t, r)\frac{1}{\varepsilon} h(r/\varepsilon^2)\,dr + \frac{1}{\varepsilon}\int_0^t P^\varepsilon(t, r)\frac{1}{\varepsilon} j(r/\varepsilon^3)\,dr.
\end{aligned} \tag{3.15}
$$

Using (3.14), (3.15), we wish to show

$$
E|z^\varepsilon(t+\delta) - z^\varepsilon(\delta)|^4 \leqq c\delta^2 \tag{3.16}
$$

$$
E|z^\varepsilon(t)|^4 \leqq c \tag{3.17}
$$

$$
\lim_{\varepsilon \downarrow 0} E|z^\varepsilon(t)| = 0 \tag{3.18}
$$

which will establish the weak convergence of $z^\varepsilon(s)$ to zero, and so, (3.10).

Using the boundedness of $A_i(t)$, etc., and the identity

$$
\begin{aligned}
\int_0^t P^\varepsilon(t, r) N(t)\,dr = &-\varepsilon[A_4^{-1}(t) N(t) - P^\varepsilon(t) A_4^{-1}(0) N(0)] \\
&+ \varepsilon \int_0^t P^\varepsilon(t, r) \frac{d}{dr}(A_4^{-1}(r) N(r))\,dr
\end{aligned} \tag{3.19}
$$

and introducing

$$
\begin{aligned}
\tilde{G}^\varepsilon(t) &= \int_0^t \frac{1}{\varepsilon} g(r/\varepsilon^3)\,dr \\
\hat{H}^\varepsilon(t) &= \int_0^t P^\varepsilon(t, r)\frac{1}{\varepsilon} h(r/\varepsilon^2)\,dr \\
\hat{J}^\varepsilon(t) &= \int_0^t P^\varepsilon(t, r)\frac{1}{\varepsilon} j(r/\varepsilon^3)\,dr
\end{aligned} \tag{3.20}
$$

then we find there is a positive constant K so that

$$
|y^\varepsilon(t)| \leqq K\left(1 + \frac{1}{\varepsilon}\int_0^t |z^\varepsilon(r)|\,dr + |\bar{x}^\varepsilon(t)| + \int_0^t |\bar{x}^\varepsilon(r)|\,dr\right) + \frac{1}{\varepsilon}|\hat{H}^\varepsilon(t)| + \frac{1}{\varepsilon}|\hat{J}^\varepsilon(t)|. \tag{3.21}
$$

From (3.14) and (3.21) we have

$$
\begin{aligned}
|z^\varepsilon(s)| \leqq &K\int_0^s |z^\varepsilon(t)|\,dt + \varepsilon K\left(1 + |\bar{x}^\varepsilon(s)| + \int_0^s |\bar{x}^\varepsilon(t)|\,dt\right) \\
&+ K\left(|\hat{H}^\varepsilon(s)| + \int_0^s |\hat{H}^\varepsilon(t)|\,dt\right) + K\left(|\hat{J}^\varepsilon(s)| + \int_0^s |\hat{J}^\varepsilon(t)|\,dt\right) \\
&+ |\tilde{J}^\varepsilon(s)| + |\tilde{G}^\varepsilon(s)|
\end{aligned} \tag{3.22}
$$

for some constant K and

$$\tilde{J}^\varepsilon(s)=\frac{1}{\varepsilon}\int_0^s A_2(t)A_4^{-1}(t)j(t/\varepsilon^3)\,dt. \tag{3.23}$$

If $\Lambda(t)=Ej(t)j^T(0)$, then

$$\begin{aligned}E|\tilde{J}^\varepsilon(s)|&\leqq\Big(\varepsilon^4\int_0^{s/\varepsilon^3}\int_0^{s/\varepsilon^3}\operatorname{trace}[A_2(\varepsilon^3 t)A_4^{-1}(\varepsilon^3 t)\\&\qquad\cdot\Lambda(t-\tau)A_2^T(\varepsilon^3\tau)(A_4^{-1}(\varepsilon^3\tau))^T]\,dt\,d\tau\Big)^{1/2}\\&\leqq\Big(\varepsilon^4 k\int_0^{s/\varepsilon^3}\int_0^{s/\varepsilon^3}\|\Lambda(t-\tau)\|\,d\tau\,dt\Big)^{1/2}\\&\leqq\Big(\varepsilon k'\int_0^\infty\|\Lambda(t)\|\,dt\Big)^{1/2}\end{aligned} \tag{3.24}$$

for some k, k' positive. In a similar way we can show that $E|\tilde{G}^\varepsilon(s)|$, $E|\hat{H}^\varepsilon(s)|$, and $E|\hat{J}^\varepsilon(s)|$ are $O(\sqrt{\varepsilon})$.

Taking expectations in (3.22), using (3.24), the fact that $\bar{x}^\varepsilon(t)$ has a bounded fourth moment, and Gronwall's inequality yields

$$\sup_{0\leqq t\leqq T}E|z^\varepsilon(t)|\leqq\sqrt{\varepsilon}c \tag{3.25}$$

for some $c>0$, which establishes (3.18).

Also, using Khas'minskii's integral estimate [14], we easily find ($\tilde{G}^\varepsilon$, $\hat{J}^\varepsilon$, and $\hat{H}^\varepsilon$ are handled similarly)

$$\sup_{0\leqq s\leqq T}E|\tilde{J}^\varepsilon(s)|^4\leqq\varepsilon^2 k \tag{3.26}$$

for some $k>0$. Taking fourth powers in (3.22), we have

$$E|z^\varepsilon(s)|^4\leqq k_1\int_0^s E|z^\varepsilon(t)|^4\,dt+k_2\varepsilon^4\cdot E|\bar{x}^\varepsilon(s)|^4+k_3 e^4\int_0^s E|\bar{x}^\varepsilon(t)|^4\,dt+\varepsilon^2 k_4 \tag{3.27}$$

for some $k_i>0$, $i=1,\cdots,4$. Gronwall's inequality yields (3.17).

Finally, from (3.14) (3.21) we have

$$\begin{aligned}|z^\varepsilon(s+\delta)-z^\varepsilon(s)|&\leqq k_1\int_s^{s+\delta}|z^\varepsilon(t)|\,dt+\varepsilon k_2\int_s^{s+\delta}|\bar{x}^\varepsilon(t)|\,dt\\&\quad+K\Big(|\hat{H}^\varepsilon(s+\delta)-\hat{H}^\varepsilon(s)|+\int_s^{s+\delta}|\hat{H}^\varepsilon(t)|\,dt\Big)\\&\quad+K\Big(|\hat{J}^\varepsilon(s+\delta)-\hat{J}^\varepsilon(s)|+\int_s^{s+\delta}|\hat{J}^\varepsilon(t)|\,dt\Big)\\&\quad+|\tilde{J}^\varepsilon(s+\delta)-\tilde{J}^\varepsilon(s)|+|\tilde{G}^\varepsilon(s+\delta)-\tilde{G}^\varepsilon(s)|\end{aligned} \tag{3.28}$$

for some $k_1, k_2>0$. Now for each component $\tilde{j}_i^\varepsilon(s)$ of $\tilde{J}^\varepsilon(s)$

$$\begin{aligned}E|\tilde{j}_i^\varepsilon(s+\delta)-\tilde{j}_i^\varepsilon(s)|^4&\leqq k_3\varepsilon^8\int_{s/\varepsilon^3}^{(s+\delta)/\varepsilon^3}\cdots\int_{s/\varepsilon^3}^{(s+\delta)/\varepsilon^3}Ej_i(t_1)\cdots j_i(t_4)\,dt_1\cdots dt_4\\&\leqq k_3\varepsilon^2\delta^2\end{aligned} \tag{3.29}$$

for some $k_3>0$; $\tilde{G}^\varepsilon(t)$, $\hat{H}^\varepsilon(t)$, and $\hat{J}^\varepsilon(t)$ are treated similarly. Taking fourth powers in (3.28), and using (3.29) and the bound on the fourth moments of $\bar{x}^\varepsilon(t)$, we find

$$E|z^\varepsilon(s+\delta)-z^\varepsilon(s)|^4 \leqq c\delta^2\varepsilon^2 \tag{3.30}$$

which proves (3.16) and (3.10). Q.E.D. (Lemma 3.3.)

LEMMA 3.4. *Let* $x^0(s)$, $w^0(s)$ *satisfy* (2.15), (2.16), *respectively*; *then for* $y^\varepsilon(s)$ *given by* (1.10b) *we have*

$$\lim_{\varepsilon\downarrow 0}\left[\int_0^s y^\varepsilon(t)\,dt - w^0(s)\right]=0. \tag{3.31}$$

Proof. Let the "outer solution" $\bar{y}^\varepsilon(s)$ be defined by

$$\begin{gathered}\varepsilon\frac{d}{ds}\bar{y}^\varepsilon(s)=A_3(s)\bar{x}^\varepsilon(s)+A_4(s)\bar{y}^\varepsilon(s)+\frac{1}{\varepsilon}h(s/\varepsilon^2)\\ \bar{y}^\varepsilon(0)=-A_4^{-1}(0)A_3(0)x_0\end{gathered} \tag{3.32}$$

and let $q^\varepsilon(s)=y^\varepsilon(s)-\bar{y}^\varepsilon(s)$. Then

$$\begin{gathered}\varepsilon\frac{d}{ds}q^\varepsilon(s)=A_4(s)q^\varepsilon(s)+A_3(s)z^\varepsilon(s)+\frac{1}{\varepsilon}j(s/\varepsilon^3)\\ q^\varepsilon(0)=y_0+A_4^{-1}(0)A_3(0)x_0\end{gathered} \tag{3.33}$$

and it is clear from (3.25) (3.28) that

$$\lim_{\varepsilon\downarrow 0}\int_0^s q^\varepsilon(t)\,dt=0. \tag{3.34}$$

If we rewrite (3.32) as

$$\int_0^s \bar{y}^\varepsilon(t)\,dt=\varepsilon\int_0^s A_4^{-1}(t)\frac{d}{dt}\bar{y}^\varepsilon(t)\,dt+w^\varepsilon(s) \tag{3.35}$$

where

$$w^\varepsilon(s)\equiv-\int_0^s A_4^{-1}(t)\Big(A_3(t)\bar{x}^\varepsilon(t)+\frac{1}{\varepsilon}h(t/\varepsilon^2)\Big)\,dt, \tag{3.36}$$

a simple modification of the argument of Lemma 3.3 (integrating the first term by parts, substituting the variations of constants formula for $\bar{y}^\varepsilon(t)$ in the result, etc.) leads to the conclusion

$$\lim_{\varepsilon\downarrow 0}\int_0^s \bar{y}^\varepsilon(t)\,dt=w^0(s). \tag{3.37}$$

Combined with (3.34), this yields (3.31) (or (2.14)) as desired. Q.E.D. (Lemma 3.4.)

It remains to establish the limit (2.17)

$$\lim_{\varepsilon\downarrow 0}Y^\varepsilon(\tau)=Y^0(\tau) \tag{3.38}$$

for the "boundary layer" defined by (2.11)–(2.12);

$$Y^\varepsilon(\tau)=y^\varepsilon(\varepsilon\tau)-\bar{y}^\varepsilon(\varepsilon\tau). \tag{3.39}$$

Evidently,

$$
(3.40)\qquad
\begin{aligned}
&\frac{dY^\varepsilon(\tau)}{d\tau}=A_4(\varepsilon\tau)Y^\varepsilon(\tau)+\frac{1}{\varepsilon}j(\tau/\varepsilon^2)+A_3(\varepsilon\tau)z^\varepsilon(\varepsilon\tau)\\
&Y^\varepsilon(0)=y_0-(-A_4^{-1}(0)A_3(0)x_0),\qquad 0\leqq\tau\leqq T_1.
\end{aligned}
$$

We may identify $X^\varepsilon(\tau)\equiv z^\varepsilon(\varepsilon\tau)=x^\varepsilon(\varepsilon\tau)-\bar{x}^\varepsilon(\varepsilon\tau)$ as the "boundary layer" associated with $x^\varepsilon(t)$. From (1.10a) and (3.4)

$$
(3.41)\qquad
\begin{aligned}
\frac{dX^\varepsilon(\tau)}{d\tau}&=\varepsilon A_1(\varepsilon\tau)X^\varepsilon(\tau)+\varepsilon A_2(\varepsilon\tau)A_4^{-1}(\varepsilon\tau)A_3(\varepsilon\tau)\bar{x}^\varepsilon(\varepsilon\tau)\\
&\quad+\varepsilon A_2(\varepsilon\tau)\bar{y}^\varepsilon(\varepsilon\tau)+\varepsilon A_2(\varepsilon\tau)Y^\varepsilon(\tau)\\
&\quad+A_4^{-1}(\varepsilon\tau)h(\tau/\varepsilon)+g(\tau/\varepsilon^2),\qquad X^\varepsilon(0)=0,
\end{aligned}
$$

and it is a simple matter to show that

$$
(3.42)\qquad \lim_{\varepsilon\downarrow 0}X^\varepsilon(\tau)=0,
$$

provided $Y^\varepsilon(\tau)$ is well behaved in ε. It is a simple matter to check that $Y^\varepsilon(\tau)$ is indeed well behaved and satisfies (3.38).

This concludes the proof of the theorem. Q.E.D.

REFERENCES

[1] P. KOKOTOVIC, R. O'MALLEY AND P. SANNUTI, *Singular perturbations and order reduction in control theory—an overview*, Automatica, 12 (1976), pp. 123–132.

[2] P. KOKOTOVIC, *A control engineer's introduction to singular perturbation*, Singular Perturbations: Order Reduction in Control System Design, ASME Publications, New York, 1972, pp. 1–12.

[3] Y.-C. HO, *Linear stochastic singular control problems*, J. Optimization Theory Appl., 9 (1972), pp. 24–31.

[4] A. HADDAD AND P. KOKOTOVIC, *On a singular perturbation problem in linear filtering theory*, Proc. 1971 Princeton Conf. on Information Sciences and Systems, pp. 263–266.

[5] ———, *On singular perturbations in linear filtering and smoothing*, Proc. 1974 Symp. Nonlinear Estimation and its Applications, San Diego, CA, pp. 96–103.

[6] ———, *On singularly perturbed stochastic linear regulators*, Proc. 1976 Allerton Conf., Allerton, IL.

[7] H. RAUCH, *Order reduction in estimation with singular perturbation*, Proc. Fourth Symp. on Nonlinear Estimation and its Applications, San Diego, CA, 1973.

[8] ———, *Application of singular perturbation to optimal estimation*, Proc. Eleven Allerton Conf. on Circuit and System Theory, Allerton, IL, 1973, pp. 718–728.

[9] A. HADDAD, *Linear filtering of singularly perturbed systems*, IEEE, Trans. Automatic Control, AC-21 (1976), pp. 515–519.

[10] D. TENEKIS, *Singularly perturbed stochastic control problems*, M.S. thesis, Dept. of Elec. Engrg. Mass, Inst. of Tech., Cambridge, MA, 1976.

[11] N. SANDELL, *Singularly perturbed linear stochastic regulator problems*, Proc. 1976 Allerton Conf., Allerton, IL.

[12] R. KHAS'MINSKII, *On stochastic processes defined by differential equations with a small parameter*, Theor. Probability Appl., 11 (1966), pp. 211–228.

[13] ———, *A limit theorem for the solutions of differential equations with random right-hand sides*, Ibid., 11 (1966), pp. 390–406.

[14] G. PAPANICOLAOU AND W. KOHLER, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math., 27 (1974), pp. 641–668.

[15] G. PAPANICOLAOU AND S. VARAHDAN, *A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations*, Ibid., 26 (1973), pp. 497–524.

[16] F. HOPPENSTEADT, *Properties of solutions of ordinary differential equations with small parameters*, Ibid., 24 (1971) pp. 804–840.

[17] Y. PROKHOROV, *Convergence of stochastic processes and limit theorems in probability theory*, Theor. Probability Appl., 1 (1956), pp. 156–214.
[18] K. CHANG AND W. COPPEL, *Singular perturbations of initial value problems over a finite interval*, Arch. Rational Mech. Anal., 32 (1969), pp. 268–280.
[19] K. CHANG, *Singular perturbation of a general boundary value problem*, SIAM J. Appl. Math., 3 (1972), pp. 520–526.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0011 $01.00/0

A SATURATION THEOREM FOR MODIFIED BERNSTEIN POLYNOMIALS IN L_p-SPACES*

C. P. MAY†

Abstract. A global saturation theorem for the modified Bernstein polynomials $P_n(f, t)$ is achieved, where

$$P_n(f, t) = \sum p_{nk}(t)(n+1)\int_{k/(n+1)}^{(k+1)/(n+1)} f(u)\,du, \qquad f \in L_p[0, 1], \quad p \geqq 1.$$

1. Introduction.

The Bernstein polynomials $B_n(f, t)$ as approximation operators on $C[0, 1]$ have been studied extensively. In particular, the local saturation theorem for Bernstein polynomials was first proved by K. de Leeuw [2]. He proved that $B_n(f, t)$ converges to $f(t)$ at the rate $O(1/n)$ if and only if $f'' \in L_\infty$ locally. This result was later refined by G. G. Lorentz [9] who showed that

$$|f(t) - B_n(f, t)| \leqq Mt(1-t)/(2n), \qquad 0 \leqq t \leqq 1, \quad n = 1, 2, \cdots,$$

is equivalent to $f' \in \mathrm{Lip}_M\, 1$. Note that the rate of convergence of $B_n(f, t) \to f(t)$ is faster near the endpoints 0 and 1.

In order to generalize the Bernstein polynomial to be an approximation operator for $L_p[0, 1]$ functions, a modification was made by Kantorovitch ([6], see also [8, p. 30]). The modified operator is defined as

$$P_n(f, t) = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k}(n+1)\int_{k/(n+1)}^{(k+1)/(n+1)} f(u)\,du. \tag{1.1}$$

In the paper [3], Ditzian and the author obtained a local saturation result for $P_n(f, t)$ on $L_p[0, 1]$, $1 \leqq p < +\infty$:

THEOREM 1.1. *For $f \in L_p[0, 1]$, $1 \leqq p < \infty$ and $0 < a < a_1 < b_1 < b < 1$:*

(i) *$\|P_n(f, t) - f(t)\|_{L_p[a,b]} = O(1/n)$, $n \to \infty$, implies $f' \in A.C.[a, b]$ and $f'' \in L_p[a, b]$ for $1 < p < \infty$, or $f \in A.C.[a, b]$ and $f' \in B.V.[a, b]$ for $p = 1$. The converse implication holds if the norm is $L_p[a_1, b_1]$.*

(ii) *$\|P_n(f, t) - f(t)\|_{L_p[a,b]} = o(1/n)$ implies $t(1-t)f'(t)$ is constant in $[a, b]$. Conversely, if $t(1-t)f'(t)$ is constant in $[a, b]$, then $\|P_n(f, t) - f(t)\|_{L_p[a_1,b_1]} = o(1/n)$ as $n \to \infty$.*

In order to extend Theorem 1.1 to the entire interval $[0, 1]$, we will make appropriate modifications near the endpoints as in the case with Bernstein polynomials. Let

$$\psi_n(t) = P_n\left(\frac{(u-t)^2}{2}, t\right) = \frac{p(t)}{2(n+1)} - \frac{p(t)}{(n+1)^2} + \frac{1}{6(n+1)^2}, \qquad p(t) = t(1-t). \tag{1.2}$$

Following Lorentz, it would be desirable to prove a global saturation theorem for $P_n(f, t)$ on $[0, 1]$ with the saturation order $\psi_n(t)$. This can be done if we allow a correction term. If

$$\phi_n(t) = \frac{n+1}{n} P_n(u-t, t) = (1-2t)/(2n) \tag{1.3}$$

then we have the following saturation result.

---

* Received by the editors December 9, 1975, and in final revised form October 10, 1977.

† Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada. This research was supported in part by the National Research Council of Canada under Grant A4048.

THEOREM 1.2. *Let* $f \in L_p[0, 1]$, $1 \leqq p < +\infty$. *Then*

1) $\|\psi_n^{-1}(t)[P_n(f, t) - f(t) - \phi_n(t)P'_n(f, t)]\|_p = O(1)$ *if and only if* $f' \in A.C.[0, 1]$ *and* $f'' \in L_p[0, 1]$ *for* $1 < p < \infty$, *or* $f \in A.C.[0, 1]$ *and* $f' \in B.V.[0, 1]$ *for* $p = 1$.

2) $\|\psi_n^{-1}(t)[P_n(f, t) - f(t) - \phi_n(t)P'_n(f, t)]\|_p = o(1)$ *if and only if* f *is linear.*

Here, and in the sequel, $\|\cdot\|_p$ denotes the $L_p[0, 1]$ norm.

We remark that the above result agrees with the result of Lorentz [9] for the case of Bernstein polynomials—in this case, $B_n(u - t, t) = 0$ and $B_n((u - t)^2, t) = t(1 - t)n^{-1}$.

We are indebted to the referee for calling our attention to the recent result of V. Maier [10] and its consequent extension to $p > 1$ by S. Riemenschneider [11]. Maier shows that $\|P_n f - f\|_1 = O(n^{-1})$ if and only if

$$f(x) = c + \int_\xi^x \frac{h(u)}{u(1-u)}\,du, \qquad h \in \text{B.V.}, \quad h(0) = h(1) = 0.$$

Maier's result is interesting since the correction factor is excluded from it, which is a saturation theorem in the classical sense. On the other hand, his saturation order is slower than ours, and avoiding the correction factor leads to a less natural saturation class. In order to achieve the approximation order $\psi_n(t)$, one has to consider the correction factor $\phi_n P'_n$, which comprises the most difficult part in our estimation.

The proof of the indirect part of Theorem 1.2 is given in § 3. The proof of the direct part is divided into two sections. In § 4 we prove the L_∞ case and in § 5 the L_1 case. Hence, by the Riesz–Thorin theorem, the direct part holds for all p. In § 2, we prove some lemmas.

2. Some lemmas. Let $B_n(f, t) = \sum_{k=0}^n p_{nk}(t)f(k/n)$ be the Bernstein polynomial operator where

$$p_{nk}(t) = \binom{n}{k} t^k (1-t)^{n-k}. \tag{2.1}$$

The following relation between P_n and B_n and recursion relation for P_n will be useful throughout the paper.

LEMMA 2.1. *For any* $m \geqq 0$, *the following relations hold*:

$$P_n((u-t)^m, t) = \frac{n+1}{(m+1)p(t)} B_{n+1}((u-t)^{m+2}, t) \tag{2.2}$$

and

$$P_n((u-t)^m, t) = \frac{(m-1)}{(n+1)} p(t) P_n((u-t)^{m-2}, t) + \frac{m}{(m+1)(n+1)} \frac{d}{dt}\{p(t)P_n((u-t)^{m-1}, t)\}. \tag{2.3}$$

Further,

$$\|n\psi_n^{-1}(t)P_n((u-t)^4, t)\|_\infty = O(1), \quad \text{as } n \to \infty, \tag{2.4}$$

and

$$\|P_n(|u-t|^m, t)\|_\infty = O(n^{-1}), \qquad n \to \infty, \quad \text{for } m \geqq 2. \tag{2.5}$$

Proof. Let $W(n, t, u)$ be the kernel of $B_n(\cdot, t)$, i.e.

$$B_n(f, t) = \int_0^1 W(n, t, u) f(u)\,du, \qquad f \in C[0, 1], \tag{2.6}$$

where the kernel $W(n, t, u)$ is given in terms of the point measure δ by

(2.7) $$W(n, t, u) = \sum_{k=0}^{n} p_{nk}(t)\delta\left(u - \frac{k}{n}\right).$$

It is well known that

(2.8) $$\frac{\partial}{\partial t} W(n, t, u) = [n/p(t)]\, W(n, t, u)(u - t).$$

If $F(u) = \int_0^u f(x)\,dx$, $f \in L_1[0, 1]$, following [8, p. 30], we have

(2.9) $$P_n(f, t) = \int_0^1 \left[\frac{\partial}{\partial t} W(n+1, t, u)\right] F(u)\,du = B'_{n+1}(F, t).$$

Using the linearity of P_n and F, we set $f(x) = (x - t)^m$ in (2.9) and combine with (2.8) to obtain (2.2).

The relation (2.3) follows from (2.2) and the recursion relation

(2.10) $$B_n((u-t)^{m+1}, t) = mn^{-1}p(t)B_n((u-t)^{m-1}, t) + p(t)n^{-1}\frac{d}{dt}B_n((u-t)^m, t).$$

Equation (2.4) follows from (2.3). □

Next we give a technical lemma useful in the proof of the direct theorem when $p = \infty$.

LEMMA 2.2. *Let $W(n, t, u)$ be defined by* (2.7). *We have*

(2.11) $$\left|\int_0^1 \left[\frac{\partial^2}{\partial t^2} W(n, t, u)\right](u-t)^4\,du\right|^{1/2} = O(\sqrt{p(t)/n} + 1/n), \quad \text{as } n \to \infty.$$

Proof. By direct calculation, we have

$$\int_0^1 \left[\frac{\partial^2}{\partial t^2} W(n, t, u)\right](u-t)^4\,du = \frac{12p}{n} + \frac{14}{n^2}(1-6p) + \frac{2}{n^3}(36p - 7)$$

where $p \equiv p(t) = t(1-t)$. Now (2.11) follows immediately. □

Our last lemma will be useful in the proof of the direct theorem when $p = 1$.

LEMMA 2.3. *Let m and r be nonnegative integers. Then*

(2.12) $$\int_0^v t^r(1-t)^m\,dt \leqq \frac{(r+2)v^{r+1}(1-v)^{m+1}}{(r+1)[r+2-v(m+r+2)]}$$

holds for any $v \in (0, (r+2)/(m+r+2))$.

Proof. The following simple proof is due to A. Meir. Let

$$I_r = \int_0^v t^r(1-t)^{s-r}\,dt;$$

then by partial integration

$$I_{r-1} = \frac{v^r(1-v)^{s-r+1}}{r} + \frac{s-r+1}{r} I_r.$$

The inequality (2.12) is trivial when $r = s = m + r$. Assume that

(2.13) $$I_r \leqq \frac{v^{r+1}(1-v)^{s-r+1}}{r+1}\,\frac{r+2}{r+2-v(s+2)}$$

and that $v<(r+1)/[r+1-v(s+2)]$; note that

$$\frac{r+2}{r+2-v(s+2)}<\frac{r+1}{r+1-v(s+2)},$$

and so by (2.12),

$$I_{r-1}\leqq\frac{v^r(1-v)^{s-r+2}}{r}\,\frac{r+1}{r+1-v(s+2)}.$$

Therefore, by reverse induction, (2.13) is proved for $r\leqq s$ provided that $r+2>v(s+2)$. Set $s=m+r$ and the lemma follows. □

3. The indirect theorem. The method of proof for the inverse theorem is essentially the method used by Lorentz [9] for Bernstein polynomials. We note that a slight modification of these arguments proves the inverse part of the aforementioned theorem of Maier for all p, $1\leqq p<\infty$.

We begin with two lemmas.

LEMMA 3.1. *Let*

$$K(n,t,u)=\sum_{k=0}^{n}p_{nk}(t)(n+1)\chi_{[k/(n+1),(k+1)/(n+1)]}(u) \tag{3.1}$$

be the kernel of the operator P_n *and let*

$$F_{mn}(u)=(n+1)\int_0^1 K(n,t,u)(t-u)^m\,dt.$$

Then, for $f\in L_1[0,1]$,

$$\langle f,F_{0n}\rangle=(n+1)\int_0^1 f(u)\,du \tag{3.2}$$

and

$$\lim_{n\to\infty}\langle f,F_{mn}\rangle=\begin{cases}\frac{1}{2}\langle f,p'\rangle, & m=1,\\ \langle f,p\rangle, & m=2,\quad p\equiv p(t).\\ 0, & m\geqq 3,\end{cases}$$

Proof. Since $\int_0^1\binom{n}{k}t^k(1-t)^{n-k}\,dt=1/(n+1)$, (3.2) follows.

For $m\geqq 1$, it is sufficient to show that the sequence of functionals is uniformly bounded, and that the lemma holds for C^∞ functions.

For $k/(n+1)\leqq u<(k+1)/(n+1)$, let

$$\Gamma(n,k,m,u)=\int_0^1 t^k(1-t)^{n-k}(t-u)^m\,dt;$$

then

$$\Gamma(n,k,m+1,u)=\Gamma(n+1,k+1,m,u)-u\Gamma(n,k,m,u). \tag{3.3}$$

For any $n\geqq k$,

$$\left|(n+1)^2\binom{n}{k}\Gamma(n,k,l,u)\right|=(n+1)\left|\frac{k+1}{n+2}-u\right|\leqq 1.$$

Hence, by (2.18), for any $m>0$ and $n \geqq k$,

$$\left|(n+1)^2\binom{n}{k}\Gamma(n, k, m+1, u)\right| \leqq M(m+1),$$

and

$$|\langle f, F_{mn}\rangle| = (n+1)^2\left|\left\langle f(\cdot), \sum_{k=0}^{n}\binom{n}{k}\Gamma(n, k, m, \cdot)\chi_{[k/(n+1),(k+1)/(n+1)]}(\cdot)\right\rangle\right| \leqq M(m)\|f\|_1.$$

Now assume that $f \in C^\infty$. Using Taylor's formula and Fubini's theorem we find.

$$\left|\langle f, F_{mn}\rangle - (n+1)\int_0^1 f(t)P_n((t-u)^m, t)\,dt + (n+1)\int_0^1 f'(t)P_n((t-u)^{m+1}, t)\,dt\right|$$

$$\leqq \frac{(n+1)}{2}\|f''\|_\infty\left|\int_0^1 P_n(|t-u|^{m+2}, t)\,dt\right|.$$

The result follows directly from (1.2), (1.3) and Lemma 2.1. □

For simplicity, let

(3.4) $$S_n(f, t) = P_n(f, t) - f(t) - \phi_n(t)P_n'(f, t).$$

The previous lemma can be used to show that if $\psi_n^{-1}(t)S_n(f, t)$ is a bounded sequence in L_p, then it converges weakly to the differential operator $D^2 = d^2/dt^2$.

LEMMA 3.2. *Let* $f \in L_p[0, 1]$, $1 \leqq p < +\infty$, *and* $g \in C_0^\infty$, *i.e.*, g *has compact support in* (0, 1). *If*

(3.5) $$\|\psi_n^{-1}(t)S_n(f, t)\|_p = O(1),$$

then

(3.6) $$\lim_{n\to\infty}\langle \psi_n^{-1}(t)S_n(f, t), g(t)\rangle = \langle f, g''\rangle.$$

Proof. Let $q(t) = g(t)/p(t)$ for $t \in (0, 1)$ and $q(0) = q(1) = 0$. Then $q(t) \in C_0^\infty$ and has the same support as g.

We next replace $\psi_n^{-1}(t)$ by its dominant term:

(3.7) $$\langle \psi_n^{-1}(t)S_n(f, t), g(t)\rangle = \langle S_n(f, t), 2nq(t)\rangle + o(1)$$

since $h_n(t) = \psi_n^{-1}(t) - 2n/p(t)$ is bounded in n and t on the support of g and $S_n(f, t)$ converges to zero in $L_p[0, 1]$ by (3.5).

In order to estimate the remaining term in (3.7), we use (3.4) and linearity.

First, use integration by parts,

$$\langle -\phi_n(t)P_n'(f, t), 2nq(t)\rangle = \langle P_n(f, t), [(1-2t)q(t)]'\rangle.$$

Since $P_n(f, t)$ converges to f in $L_p[0, 1]$ and $q \in C^\infty$, this yields

$$\langle -\phi_n(t)P_n'(f, t), 2nq(t)\rangle = \langle f(t), (p'(t)q(t))'\rangle + o(1).$$

Second, apply Taylor's expansion to $q(t)$, Fubini's theorem and Lemma 3.1, to obtain

(3.8) $$\begin{aligned} 2(n+1)\langle P_n(f, t), q(t)\rangle &= 2\sum_{m=0}^{3}\frac{1}{m!}\langle f(u)q^{(m)}(u), F_{mn}(u)\rangle \\ &\quad + 2(n+1)\int_0^1 f(u)\int_0^1 K(n, t, u)\frac{q^{(4)}(\xi)}{4!}(t-u)^4\,dt\,du \\ &= \langle fq', p'\rangle + \langle fq'', p\rangle + 2(n+1)\langle f, q\rangle + o(1). \end{aligned}$$

Combining the above estimates, we have

$$(3.9)\qquad \begin{aligned}\langle\psi_n^{-1}(t)S_n(f,t), g(t)\rangle &= \langle f, q'p'\rangle + \langle f, pq''\rangle + \langle f, (p'q)'\rangle + o(1)\\ &= \langle f, (pq)''\rangle + o(1)\\ &= \langle f, g''\rangle + o(1).\end{aligned}$$

Now we turn to the proof of the indirect part of Theorem 1.2. Hence, let $f \in L_p[0,1]$, $1 \leqq p < +\infty$, and assume that $\{\psi_n^{-1}(t)S_n(f,t)\}$ is a bounded sequence in $L_p[0,1]$. For $p>1$, using w*-compactness, there exists an $h_1 \in L_p[0,1]$ and a subnet $\{n_i\}$ such that $\psi_{n_i}^{-1}(t)S_{n_i}(f,t)$ converges to $h_1(t)$ in the w*-topology. In particular, for $g \in C_0^\infty$, supp $g \subset (0,1)$,

$$(3.10)\qquad \langle\psi_{n_i}^{-1}(t)S_{n_i}(f,t), g(t)\rangle \to \langle h_1, g\rangle.$$

Identifying each $f \in L_1[0,1]$ with $\tilde{f}(t) = \int_0^t f(u)\,du$, and an element in B.V.$[0,1]$ with $\|f\|_{L_1[0,1]} = \|\tilde{f}\|_{\mathrm{B.V.}[0,1]}$, again by w*-compactness, we have

$$(3.11)\qquad \langle[\psi_{n_i}^{-1}(t)S_{n_i}(f,t)]^{\sim}, g'(t)\rangle = -\langle\psi_{n_i}^{-1}(t)S_{n_i}(f,t), g(t)\rangle \to \langle h_2, g'\rangle$$

for some subnet $\{n_i\}$ and some $h_2 \in$ B.V.$[0,1]$.

By Lemma 3.2, we can conclude that for every $g \in C_0^\infty$, $\langle f, g''\rangle = \langle h_1, g\rangle$ holds in the case $1<p<+\infty$, and $\langle f, g''\rangle = -\langle h_2, g'\rangle$ holds for $p=1$. Hence $f'' \in L_p$ for $1<p<+\infty$, and $f' \in$ B.V.$[0,1]$ for $p=1$.

Under the condition that $\psi_n^{-1}(t)S_n(f,t)$ is $o(1)$ in $L_p[0,1]$, $1 \leqq p < +\infty$, we have

$$(3.12)\qquad \lim_{n\to\infty} |\langle\psi_n^{-1}(t)S_n(f,t), g(t)\rangle| \leqq \lim_{n\to\infty} \|g\|_\infty \|\psi_n^{-1}(t)S_n(f,t)\|_p = 0.$$

Therefore, $f''=0$ and f is linear.

4. The direct theorem for $p=\infty$. It is easy to check that the operator $P_n(f,t) - \phi_n(t)P_n'(f,t)$ preserves linear functions. Thus we only need to prove the first part of Theorem 2.1.

Let $P_n^*(f,t)$ be the integral operator

$$(4.1)\qquad P_n^*(f,t) = \int_0^1 K^*(n,t,u) f(u)\,du,$$

where

$$(4.2)\quad K^*(n,t,u) = \sum_{k=0}^{n} \binom{n}{k} (k-nt) t^{k-1} (1-t)^{n-k-1} (n+1) \chi_{[k/(n+1),(k+1)/(n+1)]}(u).$$

Observe that when f is independent of t, $P_n^*(f,t) = P_n'(f,t) = B_{n+1}''(F,t)$ for $F(u) = \int_0^u f(s)\,ds$. We introduce $P_n^*(f,t)$ so there is no confusion as to what is being differentiated in what follows. For brevity, let

$$L_n(f,t) = P_n(f,t) - \phi_n(t)P_n^*(f,t).$$

Suppose that $f'' \in L_\infty[0,1]$, then by Taylor's formula $f(u) = f(0) + uf'(0) + \int_0^1 (u-s)_+ f''(s)\,ds$. Since $L_n(f,t)$ preserves linear terms, we may write

$$(4.3)\qquad L_n(f,t) - f(t) = \int_0^1 L_n(\phi(\cdot,s,t),t) f''(s)\,ds$$

where

$$\phi(u,s,t)=(u-s)_+-(t-s)_+-(t-s)_+^0(u-t)$$

(4.4)
$$=\begin{cases}(u-s)_+, & t<s,\\ (s-u)_+, & t\geqq s.\end{cases}$$

From (4.3) it follows that

(4.5) $$\|\psi_n^{-1}(t)\{L_n(f,t)-f(t)\}\|_\infty\leqq\|f''\|_\infty\left\|\psi_n^{-1}(t)\int_0^1|L_n(\phi(\cdot,s,t),t)|\,ds\right\|_\infty.$$

Now $P_n(f,t)$ is a positive operator. Consequently,

(4.6)
$$\psi_n^{-1}(t)\int_0^1|P_n(\phi(\cdot,s,t),t)\|\,ds\leqq\psi_n^{-1}(t)P_n\left(\int_0^1|\phi(\cdot,s,t)|\,ds,t\right)$$
$$\leqq\psi_n^{-1}(t)P_n\left(\frac{(\cdot-t)^2}{2},t\right)=1.$$

On the other hand $P_n^*(f,t)$ is not a positive operator, but it does inherit some properties from $B_{n+1}(F,t)$. Because of the monotonicity preserving properties of $B_{n+1}(F,t)$, $P_n^*(f,t)\geqq 0$ if $f'\geqq 0$. Therefore, for fixed t_0, from (4.4) we derive

$$\int_0^1|P_n^*(\phi(\cdot,s,t_0),t)|\,ds=-\int_0^{t_0}P_n^*((s-\cdot)_+,t)\,ds+\int_{t_0}^1P_n^*((\cdot-s)_+,t)\,ds$$
$$=P_n^*\left(\frac{-((t_0-\cdot)_+)^2}{2}+\frac{((\cdot-t_0)_+)^2}{2},t\right)$$
$$=P_n^*\left(\int_{t_0}^{\cdot}|s-t_0|\,ds,t\right).$$

The linear operators $A_n(f,t)$ defined for $f\in C[0,1]$ by $A_n(f,t)=(d^2/dt^2)B_{n+1}(F_2,t)$ where $F_2(x)=\int_0^x\int_0^u f(s)\,ds\,du$ is a positive operator by the monotonicity properties of B_{n+1}. Hence,

$$\left|P_n^*\left(\int_{t_0}^{\cdot}|s-t_0|\,ds,t\right)\right|=A_n(|\cdot-t_0|,t)\leqq[A_n(1,t)]^{1/2}[A_n((\cdot-t_0)^2,t)]^{1/2}$$
$$=O\left(\frac{d^2}{dt^2}B_{n+1}((\cdot-t_0)^4,t)\right)^{1/2}.$$

By Lemma 2.2, this last expression is $O(\sqrt{p(t_0)/n}+1/n)$ when $t=t_0$. Thus, it follows that

(4.7)
$$\left|\phi_n(t_0)\int_0^1|P_n^*(\phi(\cdot,s,t_0),t_0)|\,ds\right|=O\left(\frac{\sqrt{t_0(1-t_0)}}{n^{3/2}}+\frac{1}{n^2}\right)$$
$$=O(\psi_n(t_0)).$$

The theorem follows from (4.3), (4.6) and (4.7).

5. The direct theorem for $p=1$. In order to complete the proof for the direct theorem, by the Riesz–Thorin theorem, it remains to prove that if $f\in L_1[0,1]$ and $f'\in$ B.V.[0, 1], then $\{\psi_n^{-1}S_n(f,t)\}$ is a bounded sequence in $L_1[0,1]$. In this case, we write

$$f(u)=f(0)+f'(0)u+\int_0^1(u-s)_+\,df'(s).$$

We may proceed as in § 4 through (4.3) while the estimates analogous to (4.6) and (4.7) require a different method. Hence, we have

$$L_n(f,t)-f(t)=\int_0^1 L_n(\phi(\cdot,s,t),t)\,df'(s) \tag{5.1}$$

where $\phi(\cdot,s,t)$ is given in (4.4). First, we have

$$\begin{aligned}\left|\int_0^1 P_n(\phi(\cdot,s,t),t)\,df'(s)\right| &\leqq \int_0^t\int_0^1 K(n,t,u)\phi(u,s,t)\,du\,|df'(s)|\\ &\quad+\int_t^1\int_0^1 K(n,t,u)\phi(u,s,t)\,du\,|df'(s)|\\ &=\int_0^t\int_0^s K(n,t,u)(s-u)\,du\,|df'(s)|\\ &\quad+\int_t^1\int_s^1 K(n,t,u)(u-s)\,du\,|df'(s)|.\end{aligned}$$

Using Fubini's theorem, we have

$$\begin{aligned}&\left\|\psi_n^{-1}(t)\int_0^1 P_n(\phi(\cdot,s,t),t)\,df'(s)\right\|_1\\ &\leqq\int_0^1\left\{\int_s^1\psi_n^{-1}(t)\int_0^s K(n,t,u)(s-u)\,du\,dt\right\}|df'(s)|\\ &\quad+\int_0^1\left\{\int_0^s\psi_n^{-1}(t)\int_s^1 K(n,t,u)(u-s)\,du\,dt\right\}|df'(s)|\\ &=\int_0^1[F(s)+G(s)]|df'(s)|.\end{aligned}$$

By the symmetry of the various quantities, $G(s)=F(1-s)$. Thus, it is sufficient to show that $G(s)$ is bounded.

Recall the definition of $K(n,t,u)$, e.g., (3.1). Let $k(s)$ be the unique integer such that $k(s)/(n+1)<s\leqq(k(s)+1)/(n+1)$. Hence,

$$\begin{aligned}G(s)&\leqq\frac{1}{2}\int_0^s(n+1)\psi_n^{-1}(t)\binom{n}{k(s)}t^{k(s)}(1-t)^{n-k(s)}\left(\frac{k(s)+1}{n+1}-s\right)^2dt\\ &\quad+\frac{1}{2}\sum_{k=k(s)+1}^{n-1}\binom{n}{k}\int_0^s t^k(1-t)^{n-k}\psi_n^{-1}(t)\,dt\,\frac{2k+1-2s(n+1)}{n+1}\\ &\quad+\int_0^s t^n\psi_n^{-1}(t)\,dt(1-s).\end{aligned} \tag{5.2}$$

Throughout the rest of the paper we require the estimate

$$\psi_n^{-1}(t)\leqq\begin{cases}6(n+1)^2, & n\geqq1,\\ 3(n+1)/p(t), & n\geqq5.\end{cases} \tag{5.3}$$

In the last two terms of (5.2), we use $\psi_n^{-1}(t) \leqq 3(n+1)/p(t)$. Clearly, the last term is bounded. The middle term on the right in (5.2) is bounded by a constant multiple of

$$\begin{aligned} &\sum_{k=k(s)+1}^{n-1} \binom{n}{k} \int_0^s t^{k-1}(1-t)^{n-k-1}\, dt[k+1-s(n+1)] \\ (5.4) \qquad &\leqq \sum_{k=k(s)+1}^{n-1} \binom{n}{k} \frac{(k+1)s^k(1-s)^{n-k}}{k[k+1-sn]}[k+1-s(n+1)] = O(1) \end{aligned}$$

by Lemma 2.3. If $k(s) > 0$, then the first term on the right in (5.2) may also be included in (5.4) (with an additional factor $[k(s)+1-s(n+1)]$ which is bounded by one).

If $k(s) = 0$, then we take $\psi_n^{-1}(t) \leqq 6(n+1)^2$ and the first term in (5.2) is majorized by

$$(n+1)\int_0^s (1-t)^n\, dt = O(1).$$

The boundedness of $G(s)$ follows, and consequently, so does the estimate corresponding to (4.6).

It remains to estimate

$$(5.5) \qquad \left\| \psi_n^{-1}(t)\phi_n(t) \int_0^1 P_n^*(\phi(\cdot, s, t), t)\, df'(s) \right\|_1 .$$

Let $K^*(n, t, u)$ be as in (4.2) and $|K^*|(n, t, u)$ be the kernel formed from $K^*(n, t, u)$ by replacing $(k-nt)$ with $|k-nt|$. Proceeding as before, (5.5) is majorized by

$$\begin{aligned} &\int_0^1 \left\{ \int_s^1 \psi_n^{-1}(t)|\phi_n(t)| \int_0^s |K^*|(n, t, u)(s-u)\, du\, dt \right\} |df'(s)| \\ &+ \int_0^1 \left\{ \int_0^s \psi_n^{-1}(t)|\phi_n(t)| \int_s^1 |K^*|(n, t, u)(u-s)\, du\, dt \right\} |df'(s)| \\ &= \int_0^1 \{F^*(s) + G^*(s)\} |df'(s)|. \end{aligned}$$

Again, $F^*(1-s) = G^*(s)$. Defining $k(s)$ as before and noticing that $k - nt \geqq 0$ for $k \geqq k(s)+1$, we have that $G^*(s)$ is bounded by

$$\begin{aligned} &\binom{n}{k(s)} \int_0^s |k(s)-nt| t^{k(s)-1}(1-t)^{n-k(s)-1} \psi_n^{-1}(t)|\phi_n(t)|\, dt (n+1)\left(\frac{k(s)+1}{n+1} - s\right)^2 \\ &+ \binom{n}{k(s)+1} \int_0^s (k(s)+1-nt) t^{k(s)}(1-t)^{n-k(s)-2} |\psi_n^{-1}(t)\phi_n(t)|\, dt \frac{3}{n+1} \\ (5.6) \qquad &+ n \sum_{k=k(s)+2}^{n-1} \binom{n}{k} \int_0^s t^{k-1}(1-t)^{n-k-1}\left(\frac{k}{n} - t\right) |\psi_n^{-1}(t)\phi_n(t)|\, dt \frac{2k+1-2s(n+1)}{n+1} \\ &+ n \int_0^s t^{n-1} |\psi_n^{-1}(t)\phi_n(t)|\, dt (1-s) \\ &= \sum_{j=1}^4 G_j^*(s). \end{aligned}$$

Using $\psi_n^{-1}(t) \leqq 3(n+1)/p(t)$, we find the bound for G_4^* is immediate.

Using the equality $|\psi_n^{-1}(t)\phi_n(t)|=O(1)/p(t)$, observing that $(k/n-t)/(1-t)$ is decreasing, and then applying Lemma 2.3, we can estimate G_3^* as

$$G_3^*(s)=O(1)\sum_{k=k(s)+2}^{n-1}\binom{n}{k}\frac{k}{n}\int_0^s t^{k-2}(1-t)^{n-k-1}\,dt[2k+1-2s(n+1)]$$

$$=O(1)\sum_{k=k(s)+2}^{n-1}\binom{n-1}{k-1}s^{k-1}(1-s)^{n-k}\frac{2k+1-2s(n+1)}{k-s(n-1)}.$$

Note that $[2k+1-2s(n+1)]/[k-s(n-1)]\leqq 2+|1-4s|/[k-s(n-1)]$ and $[k-s(n-1)]^{-1}\leqq 1$ for $s\leqq[k(s)+1]/(n+1)$. Hence, $G_3^*(s)=O(1)$.

In estimating G_1^* and G_2^*, we use $|\psi_n^{-1}(t)\phi_n(t)|=O(1)(n+1)$. Since $(k-nt)t^{k-1}(1-t)^{n-k-1}$ is the derivative of $t^k(1-t)^{n-k}$, we have

$$G_2^*(s)=O(1)\binom{n}{k(s)+1}s^{k(s)+1}(1-s)^{n-k(s)-1}=O(1).$$

Similarly, for $G_1^*(s)$, break the integration over $(0,k(s)/n)$ and $(k(s)/n,s)$, since $((k(s)+1)/(n+1)-s)^2=O((n+1)^{-2})$, as for $G_2^*(s)$, we obtain

$$G_1^*(s)=O(1)\binom{n}{k(s)}\left(\frac{k(s)}{n}\right)^{k(s)}\left(1-\frac{k(s)}{n}\right)^{n-k(s)}$$

$$=O(1).$$

Consequently, $G^*(s)$ and $F^*(s)$ are bounded. The boundedness of $\|\psi_n^{-1}(s)S_n(f,t)\|_1$ follows and the proof of Theorem 1.2 is complete. □

Acknowledgment. The author would like to express his sincere thanks to Professor S. Riemenschneider for his valuable comments, and to the referee who pointed out several simplifications to our proofs.

REFERENCES

[1] R. BOJANIC AND O. SHISHA, *Degree of L_1 approximation to integrable functions by modified Bernstein polynomials*, J. Approximation Theory, 13 (1975), pp. 66–72.

[2] K. DE LEEUW, *On the degree of approximation by Bernstein polynomials*, J. Anal. Math., 7 (1959), pp. 89–104.

[3] Z. DITZIAN AND C. P. MAY, *L_p saturation and inverse theorems for modified Bernstein polynomials*, Indiana Univ. Math. J., 25 (1976), pp. 733–751.

[4] N. DUNFORD AND J. SCHWARTZ, *Linear Operators, Part I: General Theory*, Interscience, New York, 1958.

[5] W. HOEFFDING, *The L_1 norm of the approximation error for Bernstein-type polynomials*, J. Approximation Theory, 4 (1971), pp. 347–356.

[6] L. V. KANTOROVITCH, *Sur certain développements suivant les polynômes de la forme de S. Bernstein I, II*, C.R. Acad. Sci. USSR, 20 (1930), pp. 563–568, pp. 595–600.

[7] G. G. LORENTZ, *Approximation of Functions*, Holt, Richart and Winston, New York, 1966.

[8] ———, *Bernstein Polynomials*, University of Toronto Press, Toronto, 1953.

[9] ———, *Inequalities and saturation class for Bernstein polynomials*, Proceedings of Conference at Oberwolfach, on Approximation Theory, 1963, Biskhauser Verlag, Basel, 1964, pp. 200–207.

[10] V. MAIER, *Güte-und Saturationsaussagen für die L_1-Approximation durch Spezielle Folgen Linearer Positiver Operatoren*, doctoral dissertation, Universität Dortmund, Germany, 1976.

[11] S. D. RIEMENSCHNEIDER, *The L_p-saturation of the Kantorovitch–Bernstein polynomials*, to appear.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0012 $01.00/0

ERROR BOUNDS IN THE METHOD OF AVERAGING BASED ON PROPERTIES OF AVERAGE MOTION*

M. VITINS†

Abstract. In applying asymptotic methods such as the averaging method, upper bounds for the deviation of the approximate solution from the exact solution can be derived. In this paper the dependence of such bounds on properties of the average system is discussed. A simple example illustrates the usefulness of the resulting bounds.

1. Introduction. We consider the initial value problem for the vector differential equation

$$\dot{\mathbf{x}} = \varepsilon \mathbf{X}(t, \mathbf{x}), \qquad \mathbf{x}(0) = \mathbf{x}_0, \tag{1}$$

where the dot denotes differentiation with respect to the independent variable t, $\varepsilon > 0$ is a small parameter, $\mathbf{x}_0, \mathbf{x}, \mathbf{X}$ are n-dimensional vectors and $\mathbf{x}_0, \mathbf{x}$ lie in a convex domain D of the n-dimensional Euclidean space R_n.

Let $\mathbf{X}$ be periodic in t with period T. The corresponding average system is given by the autonomous differential equation

$$\dot{\bar{\mathbf{x}}} = \varepsilon \bar{\mathbf{X}}(\bar{\mathbf{x}}), \qquad \bar{\mathbf{x}}(0) = \mathbf{x}_0, \tag{2}$$

where

$$\bar{\mathbf{X}}(\bar{\mathbf{x}}) = \frac{1}{T} \int_0^T \mathbf{X}(t, \bar{\mathbf{x}})\, dt. \tag{3}$$

The solution $\bar{\mathbf{x}}(t)$ of (2) is a first approximation for the solution $\mathbf{x}(t)$ of (1).

In this paper we investigate the behavior of upper bounds for the deviation $|\mathbf{x}(t) - \bar{\mathbf{x}}(t)|$. Bounds have been presented in [1] for more general systems than (1), the right-hand sides of which need not be periodic in t. As is pointed out in [3], [5] the assumption of periodicity with respect to t enables us to obtain simple and more useful estimates for the deviation. In this paper we go one step further and adopt general assumptions concerning the average system and its solution. In particular, an estimation for the deviation is found which varies linearly in t at a rate which is intimately connected with the stability of average motion. If the average motion is stable, the rate of growth of the upper bound will be small. Otherwise, in the case of unstable average motion, it will be comparatively large.

2. Dependence of error estimate on constants of the differential equations.

THEOREM 1. *Assumptions*:

(i) *The functions* $\mathbf{X}(t, \mathbf{x})$, $\bar{\mathbf{X}}(\mathbf{x})$ *satisfy the basic assumption*

$$\begin{aligned} &|\mathbf{X}(t, \mathbf{x})| \leqq M, && |\bar{\mathbf{X}}(\mathbf{x})| \leqq \bar{M}, \\ &\left|\frac{\partial \mathbf{X}(t, \mathbf{x})}{\partial x_k}\right| \leqq \alpha, && \left|\frac{\partial \bar{\mathbf{X}}(\mathbf{x})}{\partial x_k}\right| \leqq \bar{\alpha}, \qquad (k = 1, \cdots, n) \end{aligned} \tag{4.1}$$

* Received by the editors August 29, 1977, and in revised form December 5, 1977.

† Automatic Control Group, Brown Boveri Research Center, CH-5405 Baden, Switzerland. This work was carried out during the author's stay at the Seminar für Angewandte Mathematik der ETH, Zürich.

where $|\mathbf{X}|$ *denotes the norm* $|X_1|+|X_2|+\cdots+|X_n|$ *and where* $\mathbf{x}\in D$ (D = *convex*) *and* $t\in[0, T]$. *The function* $\mathbf{X}(t, \mathbf{x})$ *is continuous and* T-*periodic in* t *and* $\bar{\mathbf{X}}(\mathbf{x})$ *is defined by* (3), *with* $\bar{\mathbf{x}}$ *replaced by* $\mathbf{x}$.

(ii) *The solutions of*

$$\dot{\mathbf{x}}=\varepsilon\mathbf{X}(t, \mathbf{x}), \qquad \dot{\bar{\mathbf{x}}}=\varepsilon\bar{\mathbf{X}}(\bar{\mathbf{x}}), \tag{4.2}$$

with the initial conditions $\mathbf{x}(0)=\bar{\mathbf{x}}(0)=\mathbf{x}_0\in D$ *exist and are unique in a time interval* $t\in[0, L/\varepsilon]$, L *being a given constant.* (*This assumption is a consequence of assumption* (i) *if* $L<d/M$, *where* d *is the distance of* $\mathbf{x}_0$ *to the boundary of* D.)

(iii) *The solution* $\bar{\mathbf{x}}(t)$ *and its neighborhood of radius* ρ *is assumed to lie in the domain* D. *We require that* ρ *satisfy the inequality*

$$\rho>\varepsilon_0\eta_1(L), \tag{4.3}$$

where ε_0 *is some sufficiently small positive number and where*

$$\begin{aligned} \eta_1(L)&=\rho_1+\rho_2(L),\\ \rho_1&=T\cdot(M+\bar{M}),\\ \rho_2(L)&=\begin{cases}\dfrac{c}{\bar{\alpha}}[e^{\bar{\alpha}L}-1] & \text{for } \bar{\alpha}\neq 0,\\ cL & \text{for } \bar{\alpha}=0,\end{cases}\\ c&=T\cdot\{\bar{\alpha}(M+\bar{M})+M(\alpha+\bar{\alpha})\}.\end{aligned} \tag{4.4}$$

Conclusion: *The following inequality is valid for* $t\in[0, L/\varepsilon]$ *and for* $0<\varepsilon<\varepsilon_0$:

$$|\mathbf{x}(t)-\bar{\mathbf{x}}(t)|\leqq\varepsilon\eta_1(L). \tag{5}$$

Comment 1. For a given domain D and value L, the size of the estimate (5) depends to a great extent on the value of the Lipschitz constant $\bar{\alpha}$ of the *average* system. Since $\alpha\geqq\bar{\alpha}$ we expect classical estimates, such as [1], [3], [5], which are based on the Lipschitz constant α of the original system, to be somewhat weaker than (5). In particular, if $\bar{\alpha}$ vanishes altogether the bound (5) is linear in L, that is to say in time. Note that the solution is, in this case, given by

$$\bar{\mathbf{x}}(t)=\mathbf{Y}\cdot t+\mathbf{x}_0, \tag{6}$$

where $\mathbf{Y}$ is a constant n-vector, independent of the initial conditions. This solution is stable in the sense of Lyapunov.

Comment 2. If L is large, the bound (5) may be expected to be unsatisfactory in certain applications where it is known that the solutions $\mathbf{x}(t)$ and $\bar{\mathbf{x}}(t)$ remain bounded for all times. In the next section we derive a bound which takes the stability of the average solution $\bar{\mathbf{x}}(t)$ into account in order to obtain a sharper bound than (5), especially for large L.

Proof of Theorem 1. Let us define the auxiliary variable $\mathbf{z}(t)$

$$\mathbf{z}(t)=\mathbf{x}(t)-\varepsilon\mathbf{u}(t, \mathbf{x}(t)), \tag{7}$$

where

$$\mathbf{u}(t, \mathbf{x})=\int_0^t[\mathbf{X}(\sigma, \mathbf{x})-\bar{\mathbf{X}}(\mathbf{x})]\,d\sigma. \tag{8}$$

For $t=0$ the functions $\mathbf{z}(t)$ and $\mathbf{x}(t)$ lie in D, since $\mathbf{z}(0)=\mathbf{x}(0)=\mathbf{x}_0\in D$. For reasons of continuity these functions will remain in the domain D at least for times smaller than some time t^*, $t^*\leqq L/\varepsilon$.

After differentiating (7) and subtracting (2) we obtain

$$\dot{\mathbf{z}}-\dot{\bar{\mathbf{x}}}=\varepsilon[(\bar{\mathbf{X}}(\mathbf{z})-\bar{\mathbf{X}}(\bar{\mathbf{x}}))+(\bar{\mathbf{X}}(\mathbf{x})-\bar{\mathbf{X}}(\mathbf{z}))]+\varepsilon^2\mathbf{u}_{\mathbf{x}}(t,\mathbf{x})\mathbf{X}(t,\mathbf{x}), \tag{9}$$

where the compensative terms $\pm\bar{\mathbf{X}}(\mathbf{z})$ have been inserted for later convenience and where $\mathbf{u}_{\mathbf{x}}(t,\mathbf{x})$ denotes the Jacobian of the map $\mathbf{x}\to\mathbf{u}(t,\mathbf{x})$. By applying the assumptions (i), (ii) and (iii) and elementary properties of vector calculus we obtain the differential inequality

$$|\dot{\mathbf{z}}-\dot{\bar{\mathbf{x}}}|\leqq\varepsilon\bar{\alpha}|\mathbf{z}(t)-\bar{\mathbf{x}}(t)|+\varepsilon^2c, \tag{10}$$

from which follows that

$$|\mathbf{z}(t)-\bar{\mathbf{x}}(t)|\leqq\varepsilon\rho_2(\varepsilon t) \tag{11}$$

for $t\in[0,t^*]$, where the quantities c and $\rho_2(\varepsilon t)$ are defined in (4.4).

The norm we are aiming at can be written as follows

$$|\mathbf{x}(t)-\bar{\mathbf{x}}(t)|\leqq|\mathbf{z}(t)-\bar{\mathbf{x}}(t)|+\varepsilon|\mathbf{u}(t,\mathbf{x}(t))|. \tag{12}$$

Due to (11), (8) and the basic assumption (i) we obtain

$$|\mathbf{x}(t)-\bar{\mathbf{x}}(t)|\leqq\varepsilon(\rho_2(\varepsilon t)+\rho_1)=\varepsilon\eta_1(\varepsilon t) \tag{13}$$

for $t\in[0,t^*]$, where ρ_1 and η_1 are defined in (4.4). The upper bound (13) is a monotonically increasing function of εt and remains less than ρ for all values of t which do not exceed $t^*=L/\varepsilon$ by virtue of assumption (iii), provided that $0<\varepsilon<\varepsilon_0$. Hence, for $t\in[0,L/\varepsilon]$, the functions $\mathbf{x}(t)$ and $\mathbf{z}(t)$ remain in the ρ-neighborhood of $\bar{\mathbf{x}}(t)$ and thus in the domain D.

3. Dependence of error on stability of average motion. In this section we assume that the average system (2) can be solved analytically or at least that the stability of its solution can be discussed. This situation occurs very often in the applications. It is therefore of interest to study the behavior of the error bounds in terms of properties of the average motion such as its stability with the goal to obtain sharper bounds.

THEOREM 2. *Assumptions*:

(i) *The assumptions* (i), (ii) *of Theorem* 1 *are to hold true.*

(ii) *Denote the solution of the average system* (2) *which satisfies the initial condition* $\bar{\mathbf{x}}(0)=\mathbf{x}_0$ *by*

$$\bar{\mathbf{x}}(t)=\mathbf{S}(\tau,\mathbf{x}_0),\qquad \tau=\varepsilon t. \tag{14}$$

We require that the derivative of $\mathbf{S}(\tau,\mathbf{y})$ *with respect to each component* y_k *of* $\mathbf{y}$ *satisfies the inequality*

$$\left|\frac{\partial\mathbf{S}(\tau,\mathbf{y})}{\partial y_k}\right|\leqq\lambda,\qquad (k=1,2,\cdots,n)\quad \textit{for } \mathbf{y}\in D \textit{ and } \tau\in[-L,L]. \tag{15}$$

(iii) *The average solution* $\bar{\mathbf{x}}(t)$ *and its* ρ*-neighborhood are assumed to lie in D, where*

$$\rho>\varepsilon_0\max(\eta_2(L),\rho_3(L)), \tag{16.1}$$

where ε_0 is some sufficiently small positive number and where

$$\eta_2(L) = T(M+\bar{M})[1+\gamma L],$$

$$\rho_3(L) = T(M+\bar{M})\frac{\gamma L}{\lambda}, \tag{16.2}$$

and

$$\gamma = \lambda^2(\alpha+\bar{\alpha})\left[\frac{\bar{\alpha}}{\alpha+\bar{\alpha}}+\frac{M}{M+\bar{M}}\right]. \tag{16.3}$$

Conclusion: *For $t \in [0, L/\varepsilon]$ and for $0<\varepsilon<\varepsilon_0$ the inequality*

$$|\mathbf{x}(t)-\bar{\mathbf{x}}(t)| \leqq \varepsilon\eta_2(L) \tag{17}$$

is valid.

Comment 1. In contrast to the foregoing estimate (5) the formula (17) varies *linearly* in L for given values of M, $\bar{M}$, α, $\bar{\alpha}$ and λ.

Comment 2. The size of λ plays an important role in the estimate (17). According to (15) λ is a measure for the sensitivity of the solution $\mathbf{S}(\tau, \mathbf{x}_0)$ for varying initial conditions throughout the time interval $\tau \in [-L, L]$. Thus λ indicates the degree of stability of average motion. For unstable solutions we expect λ to be large, whereas for stable solutions the value of λ may be comparatively small. Due to this meaning of λ, we need not necessarily know the solution $\mathbf{S}(\tau, \mathbf{x}_0)$ explicitly. It is sufficient to determine the factor λ of instability in order to evaluate the bound (17). The existence of the constant λ, such that assumption (15) is valid for some L, is a consequence of the basic assumption (4.1) (cf. [4]).

Comment 3. It is of interest to write (15) in the form of a Lipschitz condition

$$|\mathbf{S}(\tau, \mathbf{y}_2)-\mathbf{S}(\tau, \mathbf{y}_1)| \leqq \lambda|\mathbf{y}_2-\mathbf{y}_1|, \qquad \mathbf{y}_1, \mathbf{y}_2 \in D, \quad \tau \in [-L, L]. \tag{18}$$

Let $\mathbf{S}(\tau, \mathbf{y}_1)$ be a reference solution with the initial condition $\mathbf{y}_1$ and let $\mathbf{S}(\tau, \mathbf{y}_2)$ be a neighboring solution with initial condition $\mathbf{y}_2$. The reference solution is said to be Lyapunov stable if a constant λ, satisfying (18), exists for $\tau \in [0, \infty)$. It must be pointed out that a reference solution which is not Lyapunov stable may still satisfy (18) for finite intervals of time τ. In general the size of λ will indicate the degree of instability, as previously mentioned.

Comment 4. There exists a connection between this theorem and the work of Kirchgraber [2].

Proof of Theorem 2. Before proceeding to the main part of the proof, we recall that the average system (2) is autonomous; hence the "group"-property

$$\mathbf{S}(\tau_2, \mathbf{S}(\tau_1, \mathbf{y})) = \mathbf{S}(\tau_1+\tau_2, \mathbf{y})$$

holds true. In particular, for $\tau_1 = \tau$, $\tau_2 = -\tau$

$$\mathbf{S}(-\tau, \mathbf{S}(\tau, \mathbf{y})) = \mathbf{S}(0, \mathbf{y}) = \mathbf{y}.$$

Differentiating this relation by the chain rule we find

$$\mathbf{S}_{\mathbf{z}}(-\tau, \mathbf{z})\mathbf{S}_{\mathbf{y}}(\tau, \mathbf{y}) = E,$$

where $\mathbf{z} = \mathbf{S}(\tau, \mathbf{y})$, E denotes the $n \times n$ unit matrix and where $\mathbf{S}_{\mathbf{y}}$ is the Jacobian matrix of the map $\mathbf{y} \to \mathbf{S}(\tau, \mathbf{y})$. Hence the inverse of the Jacobian matrix is given by

$$[\mathbf{S}_{\mathbf{y}}(\tau, \mathbf{y})]^{-1} = \mathbf{S}_{\mathbf{z}}(-\tau, \mathbf{z}). \tag{19}$$

Thus by virtue of assumption (15) the kth column vector $\mathbf{s}_k$ $(k=1,\cdots,n)$ of the inverse Jacobian matrix (19) satisfies the inequality

$$|\mathbf{s}_k|\leqq\lambda, \tag{20}$$

provided $\mathbf{z}\in D$, $\tau\in[-L,L]$.

We now proceed with the proof of Theorem 2. As in the previous theorem we introduce the auxiliary variable $\mathbf{z}(t)$, defined in (7).

Let us now define a second auxiliary variable $\mathbf{y}(t)$ by the relation

$$\mathbf{S}(\tau,\mathbf{y}(t))=\mathbf{z}(t). \tag{21}$$

It satisfies the differential system

$$\varepsilon\bar{\mathbf{X}}(\mathbf{S}(\tau,\mathbf{y}))+\mathbf{S}_{\mathbf{y}}(\tau,\mathbf{y})\dot{\mathbf{y}}=\dot{\mathbf{z}}. \tag{22}$$

Thus we have

$$\dot{\mathbf{y}}=[\mathbf{S}_{\mathbf{y}}(\tau,\mathbf{y})]^{-1}\{\varepsilon[\bar{\mathbf{X}}(\mathbf{x})-\bar{\mathbf{X}}(\mathbf{z})]-\varepsilon^2\mathbf{u}_{\mathbf{x}}(t,\mathbf{x})\mathbf{X}(t,\mathbf{x})\}. \tag{23}$$

Provided $\mathbf{x}(t)$ and $\mathbf{z}(t)$ are in the domain D, we obtain the estimates

$$\begin{aligned}|\bar{\mathbf{X}}(\mathbf{x})-\bar{\mathbf{X}}(\mathbf{z})|&\leqq\varepsilon\bar{\alpha}|\mathbf{u}(t,\mathbf{x})|\\&\leqq\varepsilon\bar{\alpha}T(M+\bar{M}),\end{aligned}$$

$$|\mathbf{u}_{\mathbf{x}}(t,\mathbf{x})\mathbf{X}(t,\mathbf{x})|\leqq MT(\alpha+\bar{\alpha}), \tag{24}$$

and hence the differential inequality

$$|\dot{\mathbf{y}}|\leqq c_1, \tag{25}$$

where

$$c_1=\varepsilon^2\lambda T[\bar{\alpha}(M+\bar{M})+M(\alpha+\bar{\alpha})]. \tag{26}$$

Equation (25) yields the estimate

$$|\mathbf{y}(t)-\mathbf{x}_0|\leqq c_1t=\varepsilon\rho_3(\tau), \tag{27}$$

where ρ_3 is defined in (16.2).

The desired norm may be written as follows

$$|\mathbf{x}(t)-\bar{\mathbf{x}}(t)|\leqq|\mathbf{z}(t)-\bar{\mathbf{x}}(t)|+\varepsilon|\mathbf{u}(t,\mathbf{x})|. \tag{28}$$

According to (15), (21), (27) we have

$$\begin{aligned}|\mathbf{z}(t)-\bar{\mathbf{x}}(t)|&\leqq|\mathbf{S}(\tau,\mathbf{y}(t))-\mathbf{S}(\tau,\mathbf{x}_0)|\\&\leqq\lambda|\mathbf{y}(t)-\mathbf{x}_0|\\&\leqq\lambda\varepsilon\rho_3(\tau).\end{aligned} \tag{29}$$

Thus (28) yields the final estimate

$$\begin{aligned}|\mathbf{x}(t)-\bar{\mathbf{x}}(t)|&\leqq\varepsilon^2T(M+\bar{M})\gamma t+\varepsilon T(M+\bar{M})\\&\leqq\varepsilon\eta_2(\tau)\end{aligned} \tag{30}$$

provided $\mathbf{x}(t)$, $\mathbf{z}(t)$ and $\mathbf{y}(t)$ are still in the domain D.

Since $\mathbf{x}(0)$, $\mathbf{z}(0)$ and $\mathbf{y}(0)$ coincide with $\bar{\mathbf{x}}(0)=\mathbf{x}_0$ and thus lie in D, these functions can only leave D after some time has elapsed. They will certainly still be in the ρ-neighborhood of $\bar{\mathbf{x}}(t)$ as long as the deviations $|\mathbf{x}(t)-\bar{\mathbf{x}}(t)|$, $|\mathbf{z}(t)-\bar{\mathbf{x}}(t)|$ and $|\mathbf{y}(t)-\bar{\mathbf{x}}(t)|$

remain less than ρ. A comparison of the bounds (30), (29), (27) with assumption (iii), Eq. (16.1), shows that the bounds are valid for any t less than L/ε, which proves the theorem.

4. Example. Consider for illustration the following linear differential equation for the scalar function $v(t)$

$$\ddot{v}(t)+v(t)=\varepsilon v(t), \qquad v(0)=0, \quad \dot{v}(0)=1. \tag{31}$$

The transformation

$$\begin{aligned} v(t, x_1, x_2)&=x_1 \cos t+x_2 \sin t, \\ \dot{v}(t, x_1, x_2)&=-x_1 \sin t+x_2 \cos t, \end{aligned} \tag{32}$$

carries this equation over to the standard form

$$\begin{aligned} \dot{x}_1&=-\varepsilon[x_1 \cos t \sin t+x_2 \sin^2 t], \qquad & x_1(0)=0, \\ \dot{x}_1&=\varepsilon[x_1 \cos^2 t+x_2 \cos t \sin t], \qquad & x_2(0)=1. \end{aligned} \tag{33}$$

Let us assume that $L=6\pi$ and $\varepsilon=10^{-6}$. Theorem 1 typically gives the estimate $\varepsilon\eta_1(L)=2.07$, while Theorem 2 yields the improved estimate $\varepsilon\eta_2(L)=1.33\ 10^{-3}$. The second estimate comes significantly closer to the actual deviation $4.04\ 10^{-6}$ between the averaged solution $\bar{x}(t)$ and the original solution $x(t)$.

REFERENCES

[1] N. N. BOGOLIUBOV AND Y. A. MITROPOLSKI, *Asymptotic Methods in the Theory of Nonlinear Oscillations*, Hindustan Publishing Corp., Delhi, India, 1961. (English translation.)

[2] U. KIRCHGRABER, *Error bounds for perturbation methods*, Celestial Mech. 14 (1976), pp. 351–362.

[3] V. V. LARICHEVA, *On averaging a certain class of nonlinear differential equations*, Differential Equations, 2 (1966), no. 2, pp. 169–173.

[4] F. J. MURRAY AND K. S. MILLER, *Existence Theorems*, New York University Press, New York, 1964.

[5] L. M. PERKO, *Higher order averaging and related methods for perturbed periodic and quasi-periodic systems*, SIAM J. Appl. Math., 17 (1968), pp. 698–724.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0013 $01.00/0

NONLINEAR ACCRETIVE MAPPINGS IN BANACH SPACES: THE SOLVABILITY AND A SOLUTION ALGORITHM*

YUZO OHTA†

Abstract. This paper deals with a nonlinear equation $Fx = u$, where F is a mapping of a Banach space $\mathbf{X}$ into itself, and is not necessarily differentiable, and u is a given arbitrary point in $\mathbf{X}$. The λ-functional is introduced for the concrete and practical characterization of accretive mappings in Banach spaces. In terms of the λ-functional, conditions on F which guarantee the solvability of $Fx = u$ are derived. These results are applied to investigate the solvability of an equation $\psi(x, v) = u$. An iterative method to solve $Fx = u$ is also proposed and a detailed study of the convergence of this method is given.

1. Introduction. The analyses of a wide class of nonlinear networks and feedback systems amount to the investigation of a nonlinear equation $Fx = u$, where F is a continuous mapping of a Banach space $\mathbf{X}$ into itself, and u is an arbitrary point in $\mathbf{X}$.

With respect to this equation, we first study the solvability. This problem has been extensively investigated. These results are classified as follows: (i) F is continuously differentiable [7], [10]–[12], [16], [17], [19], and (ii) F is not necessarily differentiable [2], [3], [13], [18]. It is noted that the works, except [10], [11], [17] in (i), are based on the accretive mapping theory.

In the first part of this paper, we introduce the λ-functional for the concrete and practical characterization of accretive mappings in Banach spaces. And, in terms of the λ-functional, we derive conditions on F which guarantee the solvability of $Fx = u$. Moreover, we apply these results to investigate the solvability of $\psi(x, v) = u$, where ψ is a continuous mapping of a Banach space $\mathbf{X}\times\mathbf{X}'$ into $\mathbf{X}$, and $u \in \mathbf{X}$ and $v \in \mathbf{X}'$ are arbitrary points.

Next, we are concerned with solution methods of $Fx = u$. Sandberg [22] presented one such method, which globally converges to the unique solution of $Fx = u$ in a Hilbert space if F is Lipschitzian and uniformly monotone. But it is not applicable in Banach spaces. So we propose an iterative method which is a modification of Sandberg's method, and investigate the behavior of it.

The format of this paper is the following: In § 2 we introduce the λ-functional. Several properties of it are given. The relation between the λ-functional and the duality map [14], [15] is clarified. The main results of this paper are § 3 and § 4, where the problem of the solvability of $Fx = u$ is treated and a detailed study of the convergence of an iterative method is given.

2. Preliminaries.

2.1. Notation. In this paper we denote the field of real (complex) numbers by $\mathbf{R}$ ($\mathbf{C}$), the set of all nonnegative real numbers by $\mathbf{R}_+$, the set of all nonnegative integers by $\mathbf{Z}_+$, the d-dimensional real (complex) space by $\mathbf{R}^d$ ($\mathbf{C}^d$), a Banach space by $\mathbf{X}$, the dual space of $\mathbf{X}$ by $\mathbf{X}^*$, the norm on $\mathbf{X}$ by $|\cdot|$, the induced norm on $\mathbf{X}^*$ by $\|\cdot\|$, the inner product on a Hilbert space $\mathbf{H}$ by $(\cdot,\cdot)$, the open ball in $\mathbf{X}$ with the center $z \in \mathbf{X}$ and the radius $r > 0$ by $B[z; r]$, the closure of a set $A \subset \mathbf{X}$ by $\bar{A}$, the identity mapping by I, and the zero element in $\mathbf{X}$ by θ.

2.2. The λ-functional. In this section, we introduce the λ-functional and list its useful properties. The λ-functional is a modification of the μ-functional [5]–[7], [11] and the local μ-functional [13], and is closely related to the concept of the duality map [14], [15].

* Received by the editors April 23, 1976, and in final revised form May 2, 1978.

† Department of Electrical Engineering, Fukui University, Bunkyo, Fukui, 910, Japan.

DEFINITION 1. Given $x, y \in \mathbf{X}$, $x \neq y$, the λ-functional for a mapping $F: \mathbf{X} \to \mathbf{X}$ is defined by

$$\lambda(F; x, y) = \lim_{\alpha \downarrow 0} \frac{\gamma(I + \alpha F; x, y) - 1}{\alpha}, \tag{1}$$

where $\gamma(\cdot\,; \cdot\,, \cdot)$ is given by

$$\gamma(F; x, y) = |Fx - Fy|/|x - y|. \tag{2}$$

Remark 1. Since the function $f(\alpha) = \gamma(I + \alpha F; x, y)$, $\alpha > 0$, is a convex function, the function $\alpha \mapsto (f(\alpha) - f(0))/\alpha$ is nondecreasing [9, p. 162]. Moreover $f(\alpha)$, $\alpha > 0$, is bounded from below by $-\gamma(F; x, y)$. So

$$\lambda(F; x, y) = \lim_{\alpha \downarrow 0} \frac{f(\alpha) - f(0)}{\alpha} = \inf_{\alpha > 0} \frac{\gamma(I + \alpha F; x, y) - 1}{\alpha} \tag{3}$$

exists, and, hence, the λ-functional is well defined.

LEMMA 1. *Let $F, G: \mathbf{X} \to \mathbf{X}$. The λ-functional possesses the following properties.*

(a) $\lambda(I; x, y) = 1, \qquad \lambda(-I; x, y) = -1;$

(b) $-\gamma(F; x, y) \leqq -\lambda(-F; x, y) \leqq \lambda(F; x, y) \leqq \gamma(F; x, y);$

(c) $\lambda(\beta F; x, y) = \beta \lambda(F; x, y)$ *for all* $\beta \geqq 0;$

(d) $\lambda(\beta I + F; x, y) = \beta + \lambda(F; x, y)$ *for all* $\beta \in \mathbf{R};$

(e) $\lambda(F; x, y) - \lambda(-G; x, y) \leqq \lambda(F + G; x, y) \leqq \lambda(F; x, y) + \lambda(G; x, y);$

(f) $|Fx - Fy| \geqq \lambda(F; x, y)|x - y|.$

Proof. The properties, except (f), are of the same form as those of the μ-functional [5]–[7]. Since these properties can be shown similarly, we omit them and we shall only establish (f). By (3), for each $\alpha > 0$, we have

$$\begin{aligned} \frac{|Fx - Fy|}{|x - y|} &= \frac{|x - y + \alpha(Fx - Fy) - (x - y)|}{\alpha |x - y} \\ &\geqq \frac{\gamma(I + \alpha F; x, y) - 1}{\alpha} \geqq \lambda(F; x, y). \end{aligned}$$

This implies (f).

DEFINITION 2. Given $F: \mathbf{X} \to \mathbf{X}$,

(i) F is *accretive* if $\lambda(F; x, y) \geqq 0$ for all $x, y \in \mathbf{X}$, $x \neq y$;

(ii) F is *strongly accretive* if $-\lambda(-F; x, y) \geqq 0$ for all $x, y \in \mathbf{X}$, $x \neq y$.

Remark 2. In view of the property (c) of Lemma 1, F is accretive if F is strongly accretive.

DEFINITION 3 [14], [15]. For each $u \in \mathbf{X}$, we define the *duality map* $F_D(u)$ *of* u by $F_D(u) = \{f \in \mathbf{X}^* \mid \|f\|^2 = |u|^2 = (f, u) \triangleq f(u)\}$.

LEMMA 2. *Given $F: \mathbf{X} \to \mathbf{X}$, $x, y \in \mathbf{X}$, $x \neq y$,*

(i) *there exists $f^* \in F_D(x - y)$ such that*

$$\lambda(F; x, y)|x - y|^2 = \max \{\mathrm{Re}\,(Fx - Fy, f) \mid f \in F_D(x - y)\} = \mathrm{Re}\,(Fx - Fy, f^*),$$

and

(ii) *there exists $f_* \in F_D(x - y)$ such that*

$$-\lambda(-F; x, y)|x - y|^2 = \min \{\mathrm{Re}\,(Fx - Fy, f) \mid f \in F_D(x - y)\} = \mathrm{Re}\,(Fx - Fy, f_*).$$

Proof. Let $M=\{m=\alpha(x-y)|\alpha\in\mathbf{R}\}$. If $Fx-Fy\in M$, i.e., $Fx-Fy=k(x-y)$ for some $k\in\mathbf{R}$, then the assertion is a direct consequence of the Hahn–Banach theorem [23, p. 102] and Definition 1. So we shall assume $Fx-Fy\notin M$ in the following. By the Hahn–Banach theorem, there exists $f\in F_D(x-y)$ such that $f(\beta(x-y)+\alpha(Fx-Fy))=\beta|x-y|^2+\alpha c$, where c is given by

$$c=\inf_{\beta\in\mathbf{R}}[|\beta(x-y)+Fx-Fy|-\beta|x-y|]|x-y|$$

$$=\inf_{\alpha>0}\frac{|(I+\alpha F)x-(I+\alpha F)y|-|x-y|}{\alpha|x-y|}|x-y|^2=\lambda(F;x,y)|x-y|^2.$$

On the other hand, for any $\alpha>0$ and $f\in F_D(x-y)$ we have

$$|(I+\alpha F)x-(I+\alpha F)y|\cdot|x-y|\geqq\operatorname{Re}((I+\alpha F)x-(I+\alpha F)y,f)$$
$$=|x-y|^2+\alpha\cdot\operatorname{Re}(Fx-Fy,f).$$

This implies

$$\lambda(F;x,y)|x-y|^2\geqq\operatorname{Re}(Fx-Fy,f)\quad\text{for all } f\in F_D(x-y).$$

So we have (i). From (i), we have (ii).

Remark 3. In view of Lemma 2, F is accretive if and only if, for each $x,y\in\mathbf{X}$, $x\neq y$, there exists $f\in F_D(x-y)$ such that $\operatorname{Re}(Fx-Fy,f)\geqq 0$.

COROLLARY 1. *Let* $\mathbf{X}$ *be a Hilbert space* $\mathbf{H}$. *Then we have*

$$-\lambda(-F;x,y)=\lambda(F;x,y)=\operatorname{Re}(Fx-Fy,x-y)/(x-y,x-y)\quad\text{for all } x\neq y.$$

LEMMA 3 [15]. *If* $\mathbf{X}^*$ *is strictly convex, then the duality map* $F_D(\cdot)$ *is a single-valued mapping from* $\mathbf{X}$ *into* $\mathbf{X}^*$.

From Lemmas 2 and 3, we have the following corollary.

COROLLARY 2. *If* $\mathbf{X}^*$ *is strictly convex, then* $\lambda(F;x,y)=-\lambda(-F;x,y)$ *for all* $x,y\in\mathbf{X}$, $x\neq y$.

Remark 4. If $\mathbf{X}^*$ is strictly convex, then F is strongly accretive if and only if F is accretive.

3. The solvability. In this section, we shall study the solvability of a nonlinear equation:

$$Fx=u, \tag{4}$$

where F is a continuous mapping of $\mathbf{X}$ into itself, and u is an arbitrary point in $\mathbf{X}$.

Our problem is to find conditions on F such that for any input $u\in\mathbf{X}$ the equation (4) has a unique solution $x^*(u)\in\mathbf{X}$ and that $u\mapsto x^*$ is continuous.

3.1. The homeomorphism of a mapping. In this section, we shall derive a condition for F to be a homeomorphism of $\mathbf{X}$ onto itself under the assumption that F is a local homeomorphism. The local homeomorphism is defined as follows: For each $z\in\mathbf{X}$ there exist open neighborhoods $U(z)$ of z and $V(Fz)$ of Fz, respectively, such that the restriction of F to U, $F_U: U\to V$, is a homeomorphism of U onto V.

We begin with the following definition.

DEFINITION 4. A mapping $F:\mathbf{X}\to\mathbf{X}$ *has the continuation property* for a given continuous function $q:[0,1]\to\mathbf{X}$ if the existence of a continuous function $p:[0,a)\to\mathbf{X}$, $a\in(0,1]$, such that $F[p(t)]=q(t)$ for all $t\in[0,a)$ implies that $\lim_{t\uparrow a}p(t)=p(a)$ exists with $F[p(a)]=q(a)$.

LEMMA 4 [20, p. 135]. *Let $F: \mathbf{X} \to \mathbf{X}$ be a local homeomorphism. Then F is a homeomorphism of $\mathbf{X}$ onto itself if and only if F has the continuation property for each linear function $q(t) = (1-t)y^0 + ty^1$, $t \in [0, 1]$, where $y^0, y^1 \in \mathbf{X}$ are arbitrary.*

The continuation property is an operational condition that may be difficult to verify in concrete situations. Therefore, we give a specific condition which implies the continuation property.

DEFINITION 5. We define P_∞ by $P_\infty \triangleq \{m: \mathbf{R}_+ \to \mathbf{R}_+ | m(\cdot)$ is continuous and $m(s) > 0$ for all $s \geqq 0$, and $\int_0^\infty m(s)\,ds = \infty\}$.

LEMMA 5. *Let $F: \mathbf{X} \to \mathbf{X}$ be a local homeomorphism. Suppose that there exists $m \in P_\infty$ such that the following condition holds: For each $z \in \mathbf{X}$, there is $r(z) > 0$ such that for each $x, y \in B[z; r]$,*

$$|Fx - Fy| \geqq m(\max\{|x|, |y|\})|x - y|. \tag{5}$$

Then F is a homeomorphism of $\mathbf{X}$ onto itself.

Proof. By means of Lemma 4, it suffices to show that F has the continuation property for any linear function:

$$q(t) = (1-t)y^0 + ty^1, \qquad t \in [0, 1]; \quad y^0, y^1 \in \mathbf{X}. \tag{6}$$

Suppose that for some continuous function $p: [0, a) \to \mathbf{X}$, $a \in (0, 1]$, we have

$$F[p(t)] = q(t) \quad \text{for all } t \in [0, a). \tag{7}$$

Claim 1. $|p(t)|$ is bounded on $[0, a)$.

Given $t_1 \in [0, a)$. Since $p(\cdot)$ is continuous, there exist $t_2 \in (t_1, a)$ and $r(p(t_1)) > 0$ so that $p(t) \in B[p(t_1); r(p(t_1))]$ for all $t \in [t_1, t_2]$, where $r(p(t_1))$ is a constant such that (5) holds for $z = p(t_1)$.

By means of (5)–(7), for any $t + \xi$, $t \in [t_1, t_2]$,

$$\begin{aligned} |\xi(y^0 - y^1)| &= |q(t+\xi) - q(t)| \\ &= |F[p(t+\xi)] - F[p(t)]| \geqq m(\max\{|p(t+\xi)|, |p(t)|\})|p(t+\xi) - p(t)|. \end{aligned} \tag{8}$$

So we have

$$\begin{aligned} |p(t)|^{\boldsymbol{\cdot}}_+ &= \lim_{h \downarrow 0} \sup_{\xi \in (0,h)} \frac{1}{\xi}[|p(t+\xi)| - |p(t)|] \\ &\leqq \lim_{h \downarrow 0} \sup_{\xi \in (0,h)} \frac{|y^0 - y^1|}{m(\max\{|p(t+\xi)|, |p(t)|\})} \\ &= \frac{|y^0 - y^1|}{m(|p(t)|)} \quad \text{for all } t \in [0, a). \end{aligned}$$

Set

$$v(t) = |p(t)|, \quad b(t) = |y^0 - y^1|, \quad \text{and} \quad M(\alpha) = \int_{v(0)}^{\alpha} m(s)\,ds.$$

Then using Redheffer's differential inequality (see the Appendix) we get

$$|p(t)| \leqq M^{-1}(|y^0 - y^1|t) \leqq M^{-1}(|y^0 - y^1|a) \quad \text{for all } t \in [0, a). \tag{9}$$

Claim 2. Let $\{t_k\}$ be an arbitrary monotone increasing sequence such that $\lim_{k\to\infty} t_k = a$. Then $\lim_{k\to\infty} p(t_k) = x$ so that $Fx = q(a)$.

Let r_k be a positive constant such that (5) holds for $z = p(t_k)$. Clearly there exists $K \in \mathbf{Z}_+$ such that $p(t_i) \in B[p(t_k); r_k]$ for all $i, k \geqq K$. By (8) and (9), for some $\varepsilon > 0$, we have

$$|p(t_i) - p(t_k)| \leqq (|y^0 - y^1|/\varepsilon)|t_i - t_k| \quad \text{for all } i, k \geqq K.$$

This implies that $\{p(t_k)\}$ is a Cauchy sequence, and, hence, it converges to some $x \in \mathbf{X}$. If $\lim_{k\to\infty} p(t_k) = x$, then, by the continuity of F, $Fx = q(a)$.

Claim 3. *If* $\lim_{k\to\infty} p(t_k) = x$ *and if* $Fx = q(a)$, *then* $\lim_{t\uparrow a} p(t) = x$.

Let U and V be open neighborhoods of x and Fx, respectively, such that the restriction of F to U is a homeomorphism of U onto V. Clearly, there exists $t' \in [0, a)$ so that $p(t_k) \in U$ for $t_k \in [t', a)$ and $q(t) \in V$ for $t \in [t', a)$. Therefore, the continuous function $\hat{p}(t) = F_U^{-1}[q(t)]$, $t \in [t', a)$, satisfies $\hat{p}(t_k) = p(t_k)$ for all $t_k \in [t', a)$. Then the set $J \triangleq \{t \in [t', a) | \hat{p}(s) = p(s) \text{ for all } s \in [t', t]\}$ is not empty, and, hence, $\bar{t} = \sup\{t | t \in J\}$ is well defined. Now, if $\bar{t} < a$, there exists a sequence $\{\tau_l\} \subset (\bar{t}, a)$ with $\lim_{l\to\infty} \tau_l = \bar{t}$ such that

$$p(\tau_l) \neq \hat{p}(\tau_l) \quad \text{for all } \tau_l \in (\bar{t}, a). \tag{10}$$

Since $p(t) = \hat{p}(t) \in U$ for all $t \in [t', \bar{t})$ and $p(\cdot)$ is continuous, there is a $\tau' \in (\bar{t}, a)$ so that $p(\tau) \in U$ for all $\tau \in [\bar{t}, \tau')$. This leads that (10) contradicts (7), because F_U is a homeomorphism of U onto V. So we have $\bar{t} = a$, and, hence, $p(t) = F_U^{-1}[q(t)]$ for all $t \in [t', a)$. Then we have

$$x = \lim_{k\to\infty} F_U^{-1}[q(t_k)] = \lim_{t\uparrow a} F_U^{-1}[q(t)] = \lim_{t\uparrow a} p(t).$$

Remark 5. Browder [3, p. 61] shows, in a different approach, that $F: \mathbf{X} \to \mathbf{X}$ is a homeomorphism of $\mathbf{X}$ onto itself if (5) holds for some nonincreasing function $m \in P_\infty$, and if F is a local homeomorphism.

3.2. Sufficient conditions for solvability. Lemma 5 in the previous section is highly useful to investigate properties of F. However, it is not much of practical value in determining whether or not a given mapping has such properties. In this respect, it is important to have various sufficient conditions under which the homeomorphism of F is guaranteed.

DEFINITION 6. A mapping F of a Banach space $\mathbf{X}$ into a Banach space $\mathbf{X}'$ is said to be *locally Lipschitzian* if for each $z \in \mathbf{X}$ there exist $\delta(z) > 0$ and $l(z) \geqq 0$ such that for each $r \in (0, \delta)$,

$$|Fx - Fy| \leqq l|x - y| \quad \text{for all } x, y \in \bar{B}[z; r]. \tag{11}$$

The *local Lipschitz seminorm* $\|F; z, r\|$ of F on $\bar{B}[z; r]$ is defined to be the least constant l in (11).

In the following, we consider the condition for F to be a local homeomorphism.

LEMMA 6 [20, p. 122]. *Let* $F: \mathbf{X} \to \mathbf{X}$ *be continuous. If, for some* $z \in \mathbf{X}$, *there exist positive numbers* k, l_0, $r_0 > 0$ *such that*

$$|Fx - Fy - k(x - y)| \leqq l_0|x - y| \quad \text{for all } x, y \in \bar{B}[z; r_0],$$

where $l_0 < k$, *then the restriction of* F *to* $\bar{B} = \bar{B}[z; r_0]$ *is a homeomorphism of* $\bar{B}$ *onto* $F(\bar{B})$. *Moreover,* $F(\bar{B})$ *contains the open ball* $B[Fz; \sigma]$, *where* $\sigma = (k - l_0)r_0$.

LEMMA 7. *Let* $F: \mathbf{X} \to \mathbf{X}$ *be locally Lipschitzian. If, for each* $z \in \mathbf{X}$, *there exist* $r(z) > 0$ *and* $m(z) > 0$ *such that*

$$\lambda(F; x, y) \geqq m > 0 \quad \textit{for all } x, y \in \bar{B}[z; r], x \neq y, \tag{12}$$

then $F: \mathbf{X} \to \mathbf{X}$ *is a local homeomorphism.*

Proof. We shall show that for each $z \in \mathbf{X}$ there exist open neighborhoods $U(z)$ of z and $V(Fz)$ of Fz, respectively, such that the restriction of F to U is a homeomorphism of U onto V. Fix an arbitrary point $z \in \mathbf{X}$, and set $r_0 = \min\{r(z), \delta(z)\}$, where $\delta(z)$ is given in Definition 6.

Claim 1. If $F(\bar{B}[z; r_0])$ contains an open ball V, then the restriction of F to $U = F^{-1}(V) \cap \bar{B}[z; r_0]$ is a homeomorphism of U onto V. Moreover, U is an open set.

Using (f) of Lemma 1 and (22) we have

$$|Fx - Fy| \geqq m|x - y| \quad \text{for all } x, y \in \bar{B}_0 = \bar{B}[z; r_0].$$

This implies that the restriction of F to $\bar{B}_0$ is a homeomorphism of $\bar{B}_0$ onto $F(\bar{B}_0)$. Hence, if we set $U = F^{-1}(V) \cap \bar{B}_0$, then U is an open set and the restriction of F to U is a homeomorphism of U onto itself.

Claim 2. Let $l_0 = \|F; z, r_0\| > 0$. Given $\beta > 1$, and set $k = l_0 + \beta$. Then $F(\bar{B}[z; r_0])$ contains the open ball $V = B[Fz; r_1]$, where $r_1 = [m/(m+k)]r_0$.

From Definition 6, we have

$$|(kI+F)x - (kI+F)y - k(x-y)| = |Fx - Fy| \leqq l_0|x - y| \quad \text{for all } x, y \in \bar{B}[z; r_0].$$

By Lemma 6, this implies that the restriction of $(kI+F)$ to $\bar{B}_0$ is a homeomorphism of $\bar{B}_0$ onto $(kI+F)(\bar{B}_0)$, and that $(kI+F)(\bar{B}_0)$ contains the open ball $B[(kI+F)z; \beta r_0]$. Fix an arbitrary point $y^1 \in V = B[Fz; r_1]$. Let $r_2 = r_0/(k+m) < r_0$. Then for each $x \in \bar{B}_2 = \bar{B}[z; r_2]$, we have

$$|kx + y^1 - (kI+F)z| < kr_2 + r_1 = r_0.$$

This implies that $kx + y^1 \in B[(kI+F)z; r_0] \subset B[(kI+F)z; \beta r_0] \subset (kI+F)(\bar{B}_0)$ for all $x \in \bar{B}_2 \subset \bar{B}_0$, and, hence the mapping $H: \bar{B}_2 \to \bar{B}_0$, $Hx = (kI+F)^{-1}_{\bar{B}_0}(kx + y^1)$, is well defined. Using (f) and (d) of Lemma 1, and (12) we get

$$\begin{aligned} |k(x-y)| = |(kx+y^1) - (ky+y^1)| &= |[(kI+F)_{\bar{B}_0}Hx] - [(kI+F)_{\bar{B}_0}Hy]| \\ &\geqq (k+m)|Hx - Hy| \quad \text{for all } x, y \in \bar{B}_2. \end{aligned}$$

Hence we have

$$|Hx - Hy| \leqq [k/(k+m)]|x - y| \quad \text{for all } x, y \in \bar{B}_2. \tag{13}$$

Similarly, we have

$$\begin{aligned} |Hx - z| &\leqq |Hx - Hz| + |(kI+F)^{-1}_{\bar{B}_0}(kz + y^1) - (kI+F)^{-1}_{\bar{B}_0}(kz + Fz)| \\ &\leqq [k/(k+m)]|x - z| + [1/(k+m)]|y^1 - Fz| \\ &\leqq [k/(k+m)]r_2 + [1/(k+m)]r_1 \\ &= r_2 \quad \text{for all } x \in \bar{B}_2. \end{aligned} \tag{14}$$

Equations (13) and (14) imply, respectively, that H is strictly contractive on $\bar{B}_2$, and that H maps $\bar{B}_2$ into itself, and, hence there exists a fixed point $x \in \bar{B}_2 \subset \bar{B}_0$ such that $x = Hx = (kI+F)_{\bar{B}_0}(kx + y^1)$, i.e., $Fx = y^1$. This implies $y^1 \in F(\bar{B}_2) \subset F(\bar{B}_0)$. Since $y^1 \in V$ is arbitrary, $F(\bar{B}_0)$ contains the open ball V.

Remark 6. Browderr [3, p. 32] proved an analogous result with a different approach.

Now we are in position to give the main results of our paper using the λ-functional.

THEOREM 1. *Let $F:\mathbf{X}\to\mathbf{X}$ be locally Lipschitzian. Suppose that there exists $m\in P_\infty$ such that the following condition holds: For each $z\in\mathbf{X}$, there exists $r(z)>0$ such that*

$$\lambda(F;x,y)\geqq m(\max\{|x|,|y|\})\quad \text{for all } x,y\in\bar{B}[z;r],\quad x\neq y. \tag{15}$$

Then F is a homeomorphism of $\mathbf{X}$ onto itself.

Proof. By Lemma 7 and (15), F is a local homeomorphism. On the other hand, (15) implies (5) by (f) of Lemma 1. The conclusion follows from Lemma 5.

Remark 7. With a different approach Browder [3, p. 62] shows a similar result: if F is a homeomorphism of $\mathbf{X}$ onto itself, F is locally Lipschitzian, and if there exist a nonincreasing function $m\in P_\infty$ and $c\in(0,1]$ such that each point $z\in\mathbf{X}$ has a neighborhood $B[z;r(z)]$, $r(z)>0$, such that for each $\alpha,\beta\geqq 0$, and $x,y\in B[z;r(z)]$,

$$|\alpha(x-y)+\beta(Fx-Fy)|\geqq[\alpha c+\beta m(\max\{|x|,|y|\})]|x-y|.$$

By (f), (d), and (c) of Lemma 1, (15) implies the above inequality for $c=1$. So Theorem 1 is an oblique generalization of Bowder's result in the sense that $m\in P_\infty$ is not necessarily nonincreasing.

Applying Theorem 1 to the solvability of the equation $\psi(x,v)=u$, we have the following result.

THEOREM 2. *Let $\psi:\mathbf{X}\times\mathbf{X}'\to\mathbf{X}$ be locally Lipschitzian, where the norm $|\cdot|^*$ of the Banach space $\mathbf{X}\times\mathbf{X}'$ is given by $|(x,v)|^*=|x|+|v|'$. Suppose that there exist $m\in P_\infty$ and $k\geqq 0$ such that for each $v\in\mathbf{X}'$,*

$$\lambda(\psi(\cdot,v);x,y)\geqq m(\max\{|x|,|y|\}+k|v|')\quad \textit{for all } x,y\in\bar{B}[z;r],\quad x\neq y. \tag{16}$$

Then there exists a unique continuous mapping $\phi:\mathbf{X}\times\mathbf{X}'\to\mathbf{X}$ such that

$$\psi(\phi(u,v),v)=u\quad \textit{for all } (u,v)\in\mathbf{X}\times\mathbf{X}'.$$

Proof. Define a mapping Φ: $\mathbf{X}\times\mathbf{X}'\to\mathbf{X}\times\mathbf{X}'$ by

$$\Phi(x,v)=(\psi(x,v),v). \tag{17}$$

Claim 1. $\Phi:\mathbf{X}\times\mathbf{X}'\to\mathbf{X}\times\mathbf{X}'$ is bijective, and, hence, $\Phi^{-1}:\mathbf{X}\times\mathbf{X}'\to\mathbf{X}\times\mathbf{X}'$ is well defined.

Consider a equation

$$\Phi(x,v)=(u,w), \tag{18}$$

where $(u,w)\in\mathbf{X}\times\mathbf{X}'$ is an arbitrary point. Clearly, $v=w$, and, hence, we have an equation

$$\psi(x,w)=u. \tag{19}$$

Define a mapping $F_w:\mathbf{X}\to\mathbf{X}$ by $F_w(x)=\psi(x,w)$. By the assumption, F_w is locally Lipschitzian, and $\hat{m}(\alpha)=m(\alpha+k|w|')$ is in class P_∞. So (19) has a unique solution $x^*=F_w^{-1}(u)$ by Theorem 1. Hence for each $(u,w)\in\mathbf{X}\times\mathbf{X}'$, (18) has a unique solution $(F_w^{-1}(u),w)$, and, hence, Φ is bijective.

Claim 2. Φ^{-1} is continuous.

Given arbitrary point $(z,v)\in\mathbf{X}\times\mathbf{X}'$, and let $r_0=\min\{r(z),\delta((z,v))\}$, where $\delta((z,v))$ is given in Definition 6. Let $\beta>2l_0$, where $l_0=\|\Phi;(z,v),r_0\|$, and define a linear mapping $A:\mathbf{X}\times\mathbf{X}'\to\mathbf{X}\times\mathbf{X}'$ by $A(x,v)=(x,\beta v)$. By (16) and (f) and (b) of

Lemma 1, we have for each $(x, w), (\tilde{x}, \tilde{w}) \in \bar{B}[(z, v); r_0]$, $x \neq x$,

$$\begin{aligned}
&\lambda(A\Phi; (x, w), (\tilde{x}, \tilde{w}))\\
&\quad = \lim_{\alpha \downarrow 0} \frac{|x + \alpha\psi(x, w) - [\tilde{x} + \alpha\psi(\tilde{x}, \tilde{w})]| + (1 + \alpha\beta)|w - \tilde{w}|' - (|x - \tilde{x}| + |w - \tilde{w}|')}{\alpha(|x - \tilde{x}| + |w - \tilde{w}|')}\\
&\quad \geqq \frac{\lambda(\psi(\cdot\,, w); x, \tilde{x})|x - \tilde{x}| - |\psi(\tilde{x}, w) - \psi(\tilde{x}, \tilde{w})| + \beta|w - \tilde{w}|'}{|x - \tilde{x}| + |w - \tilde{w}|'}\\
&\quad \geqq \lambda(\psi(\cdot\,, w); x, \tilde{x}) + (\beta - 2l_0)|w - \tilde{w}|'/(|x - \tilde{x}| + |w - \tilde{w}|')\\
&\quad \geqq \lambda(\psi(\cdot\,, w); x, \tilde{x})\\
&\quad \geqq m(\max\{|x|, |\tilde{x}|\} + k|w|').
\end{aligned}$$

Similarly we have for each $(x, w), (\tilde{x}, \tilde{w}) \in \bar{B}[(z, v); r_0]$, $x = \tilde{x}$, $w \neq \tilde{w}$,

$$\lambda(A\Phi; (x, w), (\tilde{x}, \tilde{w})) \geqq \beta.$$

Consequently, for some $m_0 > 0$, we have

$$\lambda(A\Phi; (x, w), (\tilde{x}, \tilde{w})) \geqq m_0 > 0 \quad \text{for all } (x, w), (\tilde{x}, \tilde{w}) \in \bar{B}[z; r_0], \quad (x, w) \neq (\tilde{x}, \tilde{w}).$$

From (f) of Lemma 1, this implies

$$\text{(20)} \qquad \begin{aligned} &|(A\Phi)(x, w) - (A\Phi)(\tilde{x}, \tilde{w})|^* \geqq m_0|(x, w) - (\tilde{x}, \tilde{w})|^*\\ &\qquad\qquad \text{for all } (x, w), (\tilde{x}, \tilde{w}) \in \bar{B}[z; r_0], \quad (x, w) \neq (\tilde{x}, \tilde{w}). \end{aligned}$$

On the other hand, we have

$$\text{(21)} \qquad |Ay - A\tilde{y}|^* \leqq (1 + \beta)|y - \tilde{y}|^* \quad \text{for all } y, \tilde{y} \in \mathbf{X} \times \mathbf{X}'.$$

By (20) and (21), we get

$$|\Phi(x, w) - \Phi(\tilde{x}, \tilde{w})|^* \geqq [m_0/(1 + \beta)]|(x, w) - (\tilde{x}, \tilde{w})|^* \quad \text{for all } (x, w), (\tilde{x}, \tilde{w}) \in \bar{B}[z; r_0].$$

Hence Φ^{-1} is continuous at (z, v).

Claim 3. There exists a continuous mapping $\phi: \mathbf{X} \times \mathbf{X}' \to \mathbf{X}$ such that

$$\text{(22)} \qquad \psi(\phi(u, v), v) = u \quad \text{for all } (u, v) \in \mathbf{X} \times \mathbf{X}'.$$

Define a mapping $P: \mathbf{X} \times \mathbf{X}' \to \mathbf{X}$ by $P(x, v) = x$. Clearly P is continuous. From above arguments, the mapping $\phi: \mathbf{X} \times \mathbf{X}' \to \mathbf{X}$ defined by $\phi = P\Phi^{-1}$ is continuous and satisfies (22).

Remark 8. Theorem 2 is an oblique generalization of the implicit function theorem.

For finite dimensional spaces, we have stronger results from Lemma 5 and the domain-invariance theorem (see the Appendix).

COROLLARY 3. *Let $F: \mathbf{R}^d \to \mathbf{R}^d$ be continuous. Suppose that there exists $m \in P_\infty$ such that one of the following conditions holds:*

(i) *For each $z \in \mathbf{R}^d$, there exists $r(z) > 0$ such that*

$$\text{(23)} \qquad |Fx - Fy| \geqq m(\max\{|x|, |y|\})|x - y| \quad \textit{for all } x, y \in \bar{B}[z; r].$$

(ii) *For each $z \in \mathbf{R}^d$, there exists $r(z) > 0$ such that*

$$\text{(24)} \qquad \lambda(F; x, y) \geqq m(\max\{|x|, |y|\}) \quad \textit{for all } x, y \in \bar{B}[z; r], \quad x \neq y.$$

Then F is a homeomorphism of $\mathbf{R}^d$ onto itself.

COROLLARY 4. *Let $\psi: \mathbf{R}^d \times \mathbf{R}^r \to \mathbf{R}^d$ be continuous. Suppose that there exist $m \in P_\infty$ and $k \geqq 0$ satisfying the following condition: For each $z \in \mathbf{R}^d$, there exists $r(z) > 0$ such that for each $v \in \mathbf{R}^r$,*

$$\lambda(\psi(\cdot, v); x, y) \geqq m(\max\{|x|, |y|\} + k|v|') \quad \text{for all } x, y \in \bar{B}[z; r], \quad x \neq y. \tag{25}$$

Then there exists a unique continuous mapping $\phi: \mathbf{R}^d \times \mathbf{R}^r \to \mathbf{R}^d$ such that

$$\psi(\phi(u, v), v) = u \quad \text{for all } (u, v) \in \mathbf{R}^d \times \mathbf{R}^r.$$

4. Solution method. In this section, we investigate the behavior of iterative methods to solve (4): $Fx = u$. Sandberg [22] has presented an iterative method which globally converges to the unique solution x^* of (4) if F is Lipschitzian in a Hilbert space $\mathbf{H}$ and if it is uniformly monotone, i.e., for some $m > 0$, $F - mI$ is monotone (accretive).

Sandberg's method is given by

$$x^{i+1} = Gx^i, \qquad i = 0, 1, \cdots; \qquad Gx = x + \alpha_0(u - Fx). \tag{26}$$

Suppose that $F: \mathbf{X} \to \mathbf{X}$ is Lipschitzian in $\mathbf{X}$ and there exists $m > 0$ such that

$$\lim_{\alpha \downarrow 0} \frac{1 - \|I - \alpha F; \theta, r\|}{\alpha} \geqq m > 0 \quad \text{for all } r > 0. \tag{27}$$

Then, by (27) and the continuity of $\alpha \mapsto [1 - \|I - \alpha F; \theta, r\|]/\alpha$, $\alpha > 0$, there exist $\alpha_0 > 0$ and $m_0 > 0$ such that $|(I - \alpha_0 F)x - (I - \alpha_0 F)y| \leqq (1 - \alpha_0 m_0)|x - y|$ for all $x, y \in \mathbf{X}$. This implies that the mapping G, which is given in (26), is strictly contractive on $\mathbf{X}$. This is the crucial point of Sandberg's method. And Sandberg has shown that if $\mathbf{X} = \mathbf{H}$, we can choose $\alpha_0 = m/l_0$, where l_0 is a Lipschitz constant of F on $\mathbf{X}$. However, in general Banach spaces, it is not easy to find such an α_0. So in the following we consider a modification of Sandberg's method.

Now we consider an iterative method:

$$x^{i+1} = Hx^i, \qquad i = 0, 1, \cdots; \qquad x^0 = z^0, \tag{28}$$

where H is given by

$$Hx = T^{-1}(u + kx), \qquad Tx = (kI + F)x. \tag{29}$$

As shown in the proof of Lemma 7, the mapping H is strictly contractive on $\mathbf{X}$ if there exists $m > 0$ such that $\lambda(F; x, y) \geqq m > 0$ for all $x, y \in X$, $x \neq y$. Note that, at each step of the iteration (28), we have to solve the equation $Tx^{i+1} = u + kx^i$ for x^{i+1}. Suppose we apply Sandberg's iteration, and we have

$$y_i^{j+1} = G_i y_i^j, \qquad j = 0, 1, \cdots; \qquad y_i^0 = x^i, \qquad i \in \mathbf{Z}_+, \tag{30}$$

where G_i is given by

$$G_i y = y + (1/k)(u + kx^i - Ty) = (1/k)(u + kx^1 - Fy). \tag{31}$$

In practice, we have to truncate the iteration (30). That is, instead of solving $Tx^{i+1} = u + kx^i$ exactly and thus obtaining the sequence $\{x^i\}$, we truncate the procedure: this will give us a sequence $\{\tilde{x}^i\}$. More specifically, during the ith step we consider solving the equation

$$T\tilde{y}_i^* = u + k\tilde{x}^i \tag{32}$$

for $\tilde{y}_i^*$ by the iteration method:

$$\tilde{y}_i^{j+1} = \tilde{G}_i \tilde{y}_i^j, \qquad j = 0, 1, \cdots; \qquad \tilde{y}_i^0 = \tilde{x}^i, \tag{33}$$

where $\tilde{G}_i$ is given by

$$\tilde{G}_i y = y + (1/k)(u + k\tilde{x}^i - Ty) = (1/k)(u + k\tilde{x}^i - Fy). \tag{34}$$

We terminate the iteration when we have an iteration, say $\tilde{y}_i^{N(i)}$, such that for some $\varepsilon > 0$,

$$|\tilde{y}_i^j - \tilde{y}_i^*| \leqq \varepsilon \quad \text{for all } j \geqq N(i), \tag{35}$$

where $\tilde{y}_i^*$ is the exact solution of (32). In Theorem 3 (below) we shall consider the behavior of the iteration:

$$\begin{aligned} &\tilde{x}^{i+1} = \tilde{y}_i^{N(i)}; \qquad \tilde{y}_i^{j+1} = \tilde{G}_i \tilde{y}_i^j, \qquad j = 0, 1, \cdots, N(i)-1; \\ &\tilde{y}_i^0 = \tilde{x}^i, \qquad i = 0, 1, \cdots; \tilde{x}^0 = z^0. \end{aligned} \tag{36}$$

DEFINITION 7. A mapping $F: \mathbf{X} \to \mathbf{X}$ is in class $\mathscr{F}(\mathbf{X}, \mathbf{X})$ if F is Lipschitzian in each bounded closed ball, i.e., for each $z \in \mathbf{X}$ and $r > 0$, there exists $l(z; r)$ such that $|Fx - Fy| \leqq l|x - y|$ for all $x, y \in \bar{B}[z; r]$.

Remark 9. If $F \in \mathscr{F}(\mathbf{X}, \mathbf{X})$, then F is locally Lipschitzian, and the local Lipschitz seminorm $\|F; z, r\|$ is well defined for each bounded closed ball $\bar{B}[z; r]$.

Now we set the following assumption.

Assumption 1. The mapping $F: \mathbf{X} \to \mathbf{X}$ is in class $\mathscr{F}(\mathbf{X}, \mathbf{X})$, and there exists $m > 0$ such that

$$\lambda(F; x, y) \geqq m > 0 \quad \text{for all } x, y \in \mathbf{X}, x \neq y. \tag{37}$$

LEMMA 8. *Let Assumption* 1 *be satisfied. Then for any positive constant k, the mapping T, which is given by* (29), *is a homeomorphism of* $\mathbf{X}$ *onto itself. Moreover for any* $z^0 \in \mathbf{X}$, *the sequence* $\{x^1\}$, *which is generated by the iteration* (28), *converges to the unique solution* x^* *of* (4).

Proof. By (37) and (d) of Lemma 1, we have

$$\lambda(T; x, y) = k + \lambda(F; x, y) \geqq k + m > 0 \quad \text{for all } x, y \in \mathbf{X}, \quad x \neq y. \tag{38}$$

Therefore, T is a homeomorphism of $\mathbf{X}$ onto itself by Theorem 1.

By (f) of Lemma 1 and (38), we get

$$|Tx - Ty| \geqq (k + m)|x - y| \quad \text{for all } x, y \in \mathbf{X}. \tag{39}$$

So we have

$$|Hx - Hy| \leqq [k/(k + m)]|x - y \quad \text{for all } x, y \in \mathbf{X}. \tag{40}$$

By the principle of contraction-mapping, (40) implies that H has the unique fixed point x^*, $x^* = Hx^*$, i.e., x^* is the unique solution of (4), and that for any $z^0 \in \mathbf{X}$, $\{x^i\}$ converges to x^*.

LEMMA 9. *Let Assumption* 1 *be satisfied. Set*

$$r^* = |u - Fz^0|/m; \tag{41}$$

$$l_0 = \|F; z, 2r^*\|; \tag{42}$$

$$k = l_0 + \beta, \qquad \beta > 0; \tag{43}$$

$$\gamma = l_0/k < 1. \tag{44}$$

Let $\{x^i\}$ *be the sequence generated by the iteration* (28); *then for each* $i \in \mathbf{Z}_+$, *the sequence* $\{y_i^j\}$, *which is generated by the iteration* (30), *converges to the unique fixed point* x^{i+1} *of* G_i, *and is contained in* $\bar{B}_i = \bar{B}[x^{i+1}; \alpha\delta^i r^*] \subset \bar{B}^* = \bar{B}[z^0; 2r^*]$, *where* α *and*

δ are given by

$$\alpha = m/(m+k); \tag{45}$$

$$\delta = k/(m+k). \tag{46}$$

Proof. Since T is a homeomorphism of X onto itself (see Lemma 8), for each $i \in \mathbf{Z}_+$, x^{i+1} is the unique fixed point of G_i.

Claim 1. $\bar{B}_i \subset \bar{B}^*$ for all $i \in \mathbf{Z}_+$.

By (28), (46), and (40), we get

$$|x^{i+1} - x^i| \leqq \delta|x^i - x^{i-1}| \leqq \delta^i |x^1 - x^0| \quad \text{for all } i \in \mathbf{Z}_+. \tag{47}$$

From (28) and (29),

$$Tx^1 - Tx^0 = u - Fz^0. \tag{48}$$

By (48), (39), (41) and (45), we have

$$|x^1 - x^0| \leqq |u - Fz^0|/(k+m) = \alpha r^*. \tag{49}$$

From (47) and (49), we get

$$|x^{i+1} - x^i| \leqq \alpha\delta^i r^* \quad \text{for all } i \in \mathbf{Z}_+. \tag{50}$$

From (50), (45), and (46), we have

$$\begin{aligned} |x^{i+1} - z^0| &\leqq |x^{i+1} - x^i| + |x^i - x^{i-1}| + \cdots + |x^1 - x^0| \\ &\leqq \alpha(\delta^i + \cdots + 1) r^* \leqq [\alpha/(1-\delta)] r^* = r^* \quad \text{for all } i \in \mathbf{Z}_+. \end{aligned} \tag{51}$$

This implies that for each $i \in \mathbf{Z}_+$, $\bar{B}_i$ is contained in $\bar{B}^*$.

Claim 2. For each $i \in \mathbf{Z}_+$, $\{y_i^j\}$ converges to x^{i+1} and is contained in $\bar{B}_i$.

Given $i \in \mathbf{Z}_+$. Since $\bar{B}_i \subset \bar{B}^*$, we have

$$|Fx - Fx^{i+1}| \leqq l_0 |x - x^{i+1}| \quad \text{for all } x \in \bar{B}_i.$$

Hence, from (30), (31), (43), and (44), we get

$$\begin{aligned} |G_i x - x^{i+1}| &= |G_i x - G_i x^{i+1}| \\ &= (1/k)|Fx - Fx^{i+1}| \leqq \gamma |x - x^{i+1}| \quad \text{for all } x \in \bar{B}_i. \end{aligned} \tag{52}$$

On the other hand, (50) implies $x^i \in \bar{B}_i$. Hence, (52) implies that $\{y_i^j\}$ converges to x^{i+1}, and that it is contained in $\bar{B}_i$.

Finally we investigate the behavior of the sequence $\{\tilde{x}^i\}$ which is generated by the iteration (36). Note that we have three sequences in mind: $\{x^i\}$, $\{\tilde{y}_i^*\}$, and $\{\tilde{x}^i\}$, where $\{x^i\}$ and $\{\tilde{y}_i^*\}$ are the sequences generated by the iterations (28) and (32), respectively.

THEOREM 3. *Let Assumption* 1 *be satisfied. Let r^*, l_0, k, γ, α, and δ be positive constants which are defined by* (41)–(46), *and let ε be a positive constant such that*

$$\varepsilon \leqq mr^*/(3k+m). \tag{53}$$

For each $i \in \mathbf{Z}_+$, choose $N(i) \in \mathbf{Z}_+$ sufficiently large and

$$\gamma^{N(i)}(|u - F\tilde{x}^i|m/) \leqq \varepsilon. \tag{54}$$

Let $\{\tilde{x}^i\}$ be the sequence generated by the iteration (36); *then we have an estimate*:

$$|\tilde{x}^i - x^*| \leqq (\varepsilon/\alpha) + \delta^i r^*, \tag{55}$$

where x^* is the unique solution of (4).

Proof. Since T is a homeomorphism of $\mathbf{X}$ onto itself, for each $i \in \mathbf{Z}_+$, (32) has the unique solution $\tilde{y}_i^*$ which is the unique fixed point of $\tilde{G}_i$. From (28), (29), (32), and (50), we have

$$\begin{aligned}(k+m)|x^{i+1}-\tilde{y}_i^*| &\leqq |Tx^{i+1}-T\tilde{y}_i^*| \\ &= |(u+kx^i)-(u+k\tilde{x}^i)| = k|x^i-\tilde{x}^i| \quad \text{for all } i \in \mathbf{Z}_+.\end{aligned}$$

This implies

$$|x^{i+1}-\tilde{y}_i^*| \leqq \delta|x^i-\tilde{x}^i| \quad \text{for all } i \in \mathbf{Z}_+. \tag{56}$$

Claim 1. For each $i \in \mathbf{Z}_+$, $\bar{B}[\tilde{y}_i^*; r_i] \subset \bar{B}^*$, where $r_i = |\tilde{y}_i^* - \tilde{x}^i|$.

The proof of Claim 1 is done by induction. If $i=0$, then the assertion is obvious, since $\bar{B}[\tilde{y}_0^*; r_0] = \bar{B}[x^1; |x^1-x^0|] \subset \bar{B}_0 \subset \bar{B}^*$ (see Lemma 9). Suppose that for some $n \in \mathbf{Z}_+$, $\bar{B}[\tilde{y}_i^*; r_i] \subset \bar{B}^*$ for all $i \leqq n$. Since $\tilde{y}_i^*$ is the unique fixed point of $\tilde{G}_i$, the assumption of the induction implies (see (52))

$$|\tilde{y}_i^* - \tilde{G}_i y| \leqq \gamma|\tilde{y}_i^* - y| \quad \text{for all } y \in \bar{B}[\tilde{y}_i^*; r_i]. \tag{57}$$

Hence, by (36) and (54), we have

$$\begin{aligned}|\tilde{y}_i^* - \tilde{x}^{i+1}| &\leqq \gamma^{N(i)}|\tilde{y}_i^* - \tilde{x}^i| \\ &\leqq \gamma^{N(i)}(|u-F\tilde{x}^i|/m) \leqq \varepsilon \quad \text{for all } i \leqq n.\end{aligned} \tag{58}$$

From (56) and (58) we get

$$|x^{i+1}-\tilde{x}^{i+1}| \leqq |x^{i+1}-\tilde{y}_i^*| + |\tilde{y}_i^* - \tilde{x}^{i+1}| \leqq \delta|x^i-\tilde{x}^i| + \varepsilon \quad \text{for all } i \leqq n. \tag{59}$$

So we obtain by (59) and (58)

$$\begin{aligned}|x^{i+1}-\tilde{x}^{i+1}| &\leqq (1+\delta+\cdots+\delta^{i-1})\varepsilon + \delta^i|x^1-\tilde{x}^1| \\ &= (1+\delta+\cdots+\delta^{i-1})\varepsilon + \delta^i|\tilde{y}_0^* - \tilde{x}^1| \leqq \varepsilon/(1-\delta) \\ &= \varepsilon/\alpha \quad \text{for all } i \leqq n.\end{aligned} \tag{60}$$

By (50), (56) and (60) we have

$$\begin{aligned}r_{n+1} &= |\tilde{y}_{n+1}^* - \tilde{x}^{n+1}| \\ &\leqq |\tilde{y}_{n+1}^* - x^{n+2}| + |x^{n+2}-x^{n+1}| + |x^{n+1}-\tilde{x}^{n+1}| \\ &\leqq (1+\delta)|x^{n+1}-\tilde{x}^{n+1}| + \alpha\delta^{n+1}r^* \leqq (1+\delta)(\varepsilon/\alpha) + \alpha\delta^{n+1}r^*.\end{aligned} \tag{61}$$

From (50), (56), and (60) we get

$$\begin{aligned}|\tilde{y}_{n+1}^* - z^0| &\leqq |\tilde{y}_{n+1}^* - x^{n+2}| + |x^{n+2}-x^{n+1}| + \cdots + |x^1-x^0| \\ &\leqq \delta|x^{n+1}-\tilde{x}^{n+1}| + \alpha r^*(1-\delta^{n+2})/(1-\delta) \\ &\leqq \delta\varepsilon/\alpha + [\alpha/(1-\delta)]r^*(1-\delta^{n+2}).\end{aligned} \tag{62}$$

By (61), (62), (45), (46), and (53) we obtain

$$\begin{aligned}|\tilde{y}_{n+1}^* - z^0| + r_{n+1} &\leqq (1+2\delta)\varepsilon/\alpha + [1-(k-m)\delta^{n+1}/(k+m)]r^* \\ &\leqq r^* + (1+2\delta)\varepsilon/\alpha \leqq 2r^*.\end{aligned} \tag{63}$$

This implies that $\bar{B}[\tilde{y}_{n+1}^*; r_{n+1}] \subset \bar{B}^*$.

Claim 2. Equation (55) holds for all $i \in \mathbf{Z}_+$.

From Claim 1, we have (see (60))

$$|x^i - \tilde{x}^i| \leqq \varepsilon/\alpha. \tag{64}$$

From (40) and (46) we get

(65) $$|x^i - x^*| \leqq \delta^i |x^0 - x^*| \quad \text{for all } i \in \mathbf{Z}_+.$$

Since x^* is the solution of (4), we have

(66) $$Fx^* - Fx^0 = u - Fz^0.$$

By (f) if Lemma 1, (37), and (60) we have

(67) $$|x^0 - x^*| \leqq |u - Fz^0|/m = r^*.$$

From (64), (65), and (67) we obtain

$$|\tilde{x}^i - x^*| \leqq |\tilde{x}^i - x^i| + |x^i - x^*| \leqq (\varepsilon/\alpha) + \delta^i r^* \quad \text{for all } i \in \mathbf{Z}_+.$$

5. Applications. In this section we shall present some applications of the results of the preceding sections. In § 5.1, we shall consider a nonlinear feedback system $\underset{\sim}{S}$ shown in Fig. 1, and give a condition for the continuity of $\underset{\sim}{S}$. In § 5.2, we shall consider a nonlinear resistive network $\underset{\sim}{N}$ shown in Fig. 2, and give conditions such that for each $u \in \mathbf{R}^d$, $\underset{\sim}{N}$ has a unique dc operating point.

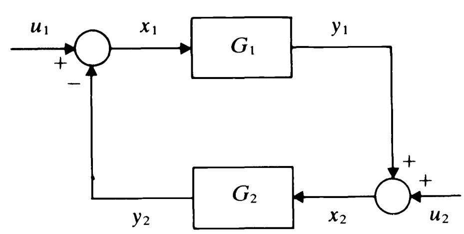

FIG. 1. *The feedback system $\underset{\sim}{S}$.*

5.1. Continuity of a feedback system. Consider a feedback system $\underset{\sim}{S}$ shown in Fig. 1. Let G_1 and G_2 denote operators (mappings) of $\mathbf{X}$ into itself, and let $u_1, u_2 \in \mathbf{X}$ be given. It is convenient to introduce the product space $\mathbf{X} \times \mathbf{X}$, with $|(x_1, x_2)| = |x_1| + |x_2|$. G denotes the operator from $\mathbf{X} \times \mathbf{X}$ into itself defined by $Gx = (G_2 x_2, -G_1 x_1)$ for $(x_1, x_2) \in \mathbf{X} \times \mathbf{X}$. The element $u = (u_1, u_2) \in \mathbf{X} \times \mathbf{X}$ will be referred to as the *input* of system $\underset{\sim}{S}$.

DEFINITION 8. The feedback system $\underset{\sim}{S}$ is continuous if i) for any $u \in \mathbf{X} \times \mathbf{X}$, there exists a solution $x \in \mathbf{X} \times \mathbf{X}$ of the equation $(I + G)x = u$, and (ii) there exists $C \geqq 0$ such that $|x - x'| \leqq C|u - u'|$ for all $u, u' \in \mathbf{X} \times \mathbf{X}$, where x and x' are the solutions of $(I + G)x = u$ and $(I + G)x' = u'$, respectively.

THEOREM 4. *Let G_1 and G_2 be locally Lipschitzian. Assume that there exist real numbers m, ε, and δ such that*

(68) $$\lambda(G_2; x, y) \geqq m > 0 \quad \textit{for all } x, y \in \mathbf{X}, \quad x \neq y,$$

(69) $$-\lambda(G_2^{-1}(-I); x, y) \geqq \varepsilon \quad \textit{for all } x, y \in \mathbf{X}, \quad x \neq y,$$

(70) $$\lambda(G_1; x, y) \geqq \delta \quad \textit{for all } x, y \in \mathbf{X}, \quad x \neq y,$$

and

(71) $$\varepsilon + \delta > 0,$$

where G_2^{-1} denotes the inverse operator of G_2. Then $\underset{\sim}{S}$ is continuous.

Proof. From Fig. 1, we have

(72) $$x_1 = u_1 - y_2,$$

(73) $$x_2 = u_2 + y_1,$$

(74) $$y_1 = G_1 x_1,$$

(75) $$y_2 = G_2 x_2.$$

By Theorem 1 and (68), the operator G_2 is invertible, so we obtain the following equation by (72)–(75):

(76) $$G_2^{-1}(-I)(x_1 - u_1) - G_1 x_1 = u_2.$$

Define operators $\hat{G}_2$, G': $\mathbf{X} \to \mathbf{X}$ by

(77) $$\hat{G}_2 x = G_2^{-1}(-I)(x - u_1),$$

(78) $$G'x = -\hat{G}_2 x + G_1 x.$$

Then we have by (69)–(71), (77), (78), and (e) of Lemma 1,

(79) $$-\lambda(\hat{G}_2; x, y) = -\lambda(G_2^{-1}(-I); x - u_1, y - u_1) \geqq \varepsilon \quad \text{for all } x, y \in \mathbf{X},\ x \neq y,$$

(80) $$\lambda(G'; x, y) \geqq -\lambda(\hat{G}_2; x, y) + \lambda(G_1; x, y) \geqq \varepsilon + \delta > 0 \quad \text{for all } x, y \in \mathbf{X},\ x \neq y.$$

By (68) and (f) of Lemma 1, G_2^{-1} is Lipschitzian, and, hence, G' is locally Lipschitzian. So Theorem 1 and (80) imply that for any $u \in \mathbf{X} \times \mathbf{X}$, (76) has the unique solution x_1, and, hence, there exists a unique solution $x \in \mathbf{X} \times \mathbf{X}$ of the equation $(I + G)x = u$. Let $x = (x_1, x_2)$ and $x' = (x_1', x_2')$ be, respectively, the solutions of $(I + G)x = u$ and $(I + G)x' = u'$, $u = (u_1, u_2)$, $u' = (u_1', u_2') \varepsilon \mathbf{X} \times \mathbf{X}$. By (76)–(78) we have

(81) $$G'x_1' = u_2' + G_2^{-1}(u_1 - x_1') - G_2^{-1}(u_1' - x_1').$$

By (f) of Lemma 1, (76)–(78), (80), and (81) we get

(82) $$|x_1 - x_1'| \leqq [1/(\varepsilon + \delta)](|u_2 - u_2'| + |G_2^{-1}(u_1 - x_1') - G_2^{-1}(u_1' - x_1')|).$$

By (f) of Lemma 1 and (68) we have

(83) $$|G_2^{-1}(u_1 - x_1') - G_2^{-1}(u_1' - x_1')| \leqq (1/m)|u_1 - u_1'|.$$

From (82) and (83) we get

(84) $$|x_1 - x_1'| \leqq [1/(\varepsilon + \delta)][|u_2 - u_2'| + (1/m)|u_1 - u_1'|].$$

By (72), (75), (83), and (84) we obtain

(85) $$\begin{aligned} |x_2 - x_2'| &= |G_2^{-1}(u_1 - x_1) - G_2^{-1}(u_1' - x_1')| \\ &\leqq (1/m)(|u_1 - u_1'| + |x_1 - x_1'|) \\ &\leqq (1/m)[1 + 1/m(\varepsilon + \delta)]|u_1 - u_1'| + [1/m(\varepsilon + \delta)]|u_2 - u_2'|. \end{aligned}$$

Set $C = \max\{(1/m)[1 + 1/m(\varepsilon + \delta)], [1/m(\varepsilon + \delta)]\}$. Then we have from (84) and (85)

$$|x - x'| \leqq C|u - u'|.$$

COROLLARY 5 (Passivity theorem [8, p. 184]). *Let $\mathbf{X} = \mathbf{H}$ be a real Hilbert space. Assume that G_1 is locally Lipschitzian, and that G_2 is Lipschitzian in $\mathbf{H}$. Further, assume that G_1 is monotone (accretive), i.e.,*

(86) $$\lambda(G_1; x, y) \geqq 0 \quad \textit{for all } x, y \in \mathbf{H}, \quad x \neq y,$$

and that there exists $m > 0$ *such that*

(87) $$\lambda(G_2; x, y) \geqq m > 0 \quad \textit{for all } x, y \in \mathbf{H}, \quad x \neq y.$$

Then $\underset{\sim}{S}$ *is continuous.*

Proof. By Theorem 1 and (87), G_2 is invertible. Set $x = -G_2 v$, $y = -G_2 w$. From Corollary 1 and (87) we have

$$\begin{aligned}-\lambda(G_2^{-1}(-I); x, y) &= -(G_2^{-1}(-x) - G_2^{-1}(-y), x-y)/(x-y, x-y)\\ &= -(v-w, -G_2 v + G_2 w)/(G_2 v - G_2 w, G_2 v - G_2 w) \geqq (m/l^2)\\ &> 0,\end{aligned}$$

where l is a Lipschitz constant of G_2. Hence (86) and (87) imply (68)–(71). The conclusion follows from Theorem 4.

Remark 10. The condition (86) can be replaced by the following weaker condition: There exists δ such that $\delta + (m/l^2) > 0$.

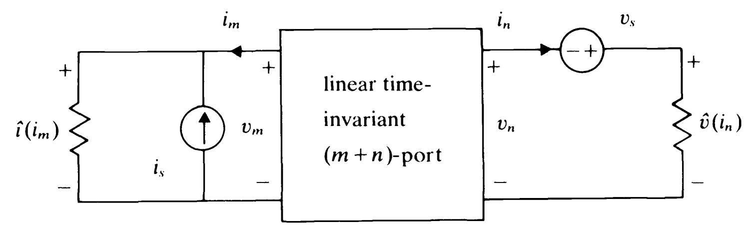

FIG. 2. *Resultant network* $\underset{\sim}{N}$.

5.2. Nonlinear resistive networks. Consider a network consisting of voltage-controlled nonlinear resistors, current-controlled nonlinear resistors, independent sources, linear dependent sources, and linear resistors. We assume that the network has a unique dc operating point if we replace all the voltage-controlled and current-controlled nonlinear resistors by arbitrary independent voltage and current sources, respectively. Then by the Thévenin–Norton equivalent transformation, we have a network $\underset{\sim}{N}$ as shown in Fig. 2. $\underset{\sim}{N}$ has a linear time invariant $(m+n)$-port with a hybrid matrix H, i.e.,

(88) $$\begin{pmatrix} i_m \\ v_n \end{pmatrix} = -H\begin{pmatrix} v_m \\ i_n \end{pmatrix},$$

m numbers of independent current sources $i_s \in \mathbf{R}^m$, and m numbers of voltage-controlled nonlinear resistors (which may be coupled) $i = \hat{\imath}(v)$, where $\hat{\imath}: \mathbf{R}^m \to \mathbf{R}^m$ is continuous, are connected to the m-port terminal in parallel: and n numbers of independent voltage sources $v_s \in \mathbf{R}^n$ and n numbers of current-controlled nonlinear resistors (which may be coupled) $v = \hat{v}(i)$, where $\hat{v}: \mathbf{R}^n \to \mathbf{R}^n$ is continuous, are connected to the n-port terminal in series.

Applying Kirchoff's laws to $\underset{\sim}{N}$, we have

(89) $$\hat{\imath}(v_m) = i_s + i_m,$$

(90) $$\hat{v}(i_n) = v_s + v_n.$$

From (88)–(90), the dc equation of $\underset{\sim}{N}$ is given by

(91) $$H\begin{pmatrix} v_m \\ i_n \end{pmatrix} + \begin{pmatrix} \hat{\imath}(v_m) \\ \hat{v}(i_n) \end{pmatrix} = \begin{pmatrix} i_s \\ v_s \end{pmatrix}.$$

For related discussion to obtain dc equations of the form (91) the reader is referred to [4].

THEOREM 5. *Let* $\hat{\imath}:\mathbf{R}^m \to \mathbf{R}^m$ *and* $\hat{v}:\mathbf{R}^n \to \mathbf{R}^n$ *be continuous. Define* $G:\mathbf{R}^{m+n} \to \mathbf{R}^{m+n}$ *by*

$$Gx = \begin{pmatrix} \hat{\imath}(v_m) \\ \hat{v}(i_n) \end{pmatrix},$$

where $x = (v_m, i_n) \in \mathbf{R}^{m+n}$.

If one of the following three conditions is satisfied, then for any given independent sources $(i_s, v_s) \in \mathbf{R}^{m+n}$, $\underline{N}$ *has a unique dc operating point* $(v_m^*, i_n^*) \in \mathbf{R}^{m+n}$. *Furthermore,* $(i_s, v_s) \mapsto (v_m^*, i_n^*)$ *is continuous.*

(I) $\lambda(G; x, y) \geqq 0$ *for all* $x, y \in \mathbf{R}^{m+n}$, $x \neq y$, *and for some* $m > 0$, $-\mu(-H) \geqq m > 0$, *where* $\mu(\cdot)$ *denotes the measure of a matrix which is defined by*

$$\mu(H) = \lim_{\alpha \downarrow 0} \frac{\|I + \alpha H\| - 1}{\alpha}.$$

(II) *For some* $m \in P_\infty$, $\lambda(G; x, y) \geqq m(\max\{|x|, |y|\})$ *for all* $x, y \in \mathbf{R}^{m+n}$, $x \neq y$, *and* $-\mu(-H) \geqq 0$.

(III) *For some* $m \in P_\infty$, $\lambda(G + H; x, y) \geqq m(\max\{|x|, |y|\})$ *for all* $x, y \in \mathbf{R}^{m+n}$, $x \neq y$.

Proof. By the definitions of the λ-functional and the measure of a matrix we have

$$(92) \qquad -\lambda(-H; x, y) \geqq -\mu(-H) \quad \text{for all } x, y \in \mathbf{R}^{m+n}, \quad x \neq y.$$

From (92) and (e) of Lemma 1, both (I) and (II) imply (III). The conclusion follows from Corollary 3.

6. Conclusion. In this paper we have first introduced the λ-functional, and showed several useful properties. We clarified the relation between the λ-functional and the duality map. We have given sufficient conditions for the solvability of $Fx = u$ where F is not necessarily differentiable. We have also shown sufficient conditions for the solvability of $\psi(x, v) = u$. Moreover we have proposed an iterative method to solve $Fx = u$, which globally converges to the unique solution under the accretive property of F. It should be stressed that the results not only generalize a number of previous results but also strengthen them in a single framework of the λ-functional.

Acknowledgments. The author would like to thank the reviewers for their constructive comments which are reflected in Remarks 1, 5, 6, and 7 in this paper. The author also wishes to thank Professor S. Kodama, Osaka University, Associate Professor H. Haneda, Kobe University, Dr. H. Maeda, and Dr. S. Kumagai, Osaka University, for their helpful suggestions and fruitful discussions.

Appendix.

DOMAIN-INVARIANCE THEOREM [1]. *Let* $F:\mathbf{R}^d \to \mathbf{R}^d$ *be continuous and injective. Then for each open subset* $D \subset \mathbf{R}^d$, $F(D)$ *is also open.*

REDHEFFER'S DIFFERENTIAL INEQUALITY [12], [21]. *Let* $b, m:\mathbf{R}_+ \to \mathbf{R}$ *be continuous. Consider the following differential inequality:*

$$\dot{v}_+ \leqq b(t)/m(v); \qquad v(0) \leqq c.$$

Let $B(t) = \int_0^t b(s)\,ds$ *and* $M(\alpha) = \int_{v(0)}^{\alpha} m(s)\,ds$. *If* $M(c) + B(t) \in \text{Domain}(M^{-1})$, *and for any* $\xi \geqq v(0)$, $m(\xi) > 0$, *then* $v(t) \leqq M^{-1}[M(c) + B(t)]$ *for all* $t \in \mathbf{R}_+$.

REFERENCES

[1] L. Bers, *Introduction to Topology*, New York University, 1945–1955.

[2] F. E. Browder, *The solvability of non-linear functional equations*, Duke Math. J., 30 (1963), pp. 557–566.

[3] ———, *Nonlinear operators in Banach spaces*, Proc. Symp. Pure Math., 18, part 2, American Mathematical Society, Providence, RI, 1976.

[4] L. O. Chua, *Efficient computer algorithms for piecewise-linear analysis of restive nonlinear networks*, IEEE Trans. Circuit Theory, CT-18 (1971), pp. 73–85.

[5] W. A. Coppel, *Stability and Asymptotic Behavior of Differential Equations*, Heath, Boston, 1965.

[6] G. Dahlquist, *Stability and error bounds in the numerical integration of ordinary differential equations*, Trans. Roy. Inst. Tech. Stockholm (Sweden), 130, 1959.

[7] C. A. Desoer and H. Haneda, *The measure of a matrix as a tool to analyze computer algorithms for circuit analysis*, IEEE Trans. Circuit Theory, CT-19 (1972), pp. 480–486.

[8] C. A. Desoer and M. Vidyasager, *Feedback Systems: Input-Output Properties*, Academic Press, New York, 1975.

[9] J. Dieudonné, *Foundation of Modern Analysis*, Academic Press, New York, 1960.

[10] J. Hadamard, *Sur les transformation ponctuelles*, Bull. Soc. Math. France, 34 (1906), pp. 71–84.

[11] H. Haneda, *μ-functional on bounded linear operators*, Mem. of Faculty of Engrg. Kobe Univ., 19 (1973), pp. 15–25.

[12] ———, *Analysis of dc equations characterized k-times* ($1 \leqq k \leqq \infty$) *continuously differentiable mappings*, Paper of Tech. Group on Circuit and System Theory IECE Japan, CT72-39, 1972.

[13] ———, *A generalization of monotonicity applied to the dc analysis of non-differentiable resistive networks*, IEEE Trans. Circuit and Systems, CAS-21 (1974), pp. 406–412.

[14] T. Kato, *Nonlinear semigroup and evolution equations*, J. Math. Soc. Japan, 19 (1974), pp. 508–520.

[15] ———, *Accretive operators and nonlinear evolution equations in Banach spaces*, Nonlinear Functional Analysis, Proc. Symp. Pure Math., 18, Part 1, American Mathematical Society, Providence, RI, 1970, pp. 138–161.

[16] E. S. Kuh and I. N. Hajj, *Nonlinear circuit theory: Resistive networks*, Proc. IEEE, 59 (1971), pp. 340–355.

[17] G. Meyer, *On solving nonlinear equations with a one-parameter operator imbedding*, SIAM J. Numer. Anal., 5 (1968), pp. 739–752.

[18] G. J. Minty, *Monotone (nonlinear) operators in Hilbert-space*, Duke Math. J., 29 (1962), pp. 341–346.

[19] T. Ohtsuki and H. Watanabe, *State-variable analysis of RLC networks containing nonlinear coupling elements*, IEEE Trans. Circuit Theory, CT-18 (1970), pp. 146–152.

[20] J. M. Ortega and W. C. Reinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[21] R. M. Redheffer, *Differential and integral inequalities*, Proc. Amer. Math. Soc., 15 (1964), pp. 715–716.

[22] I. W. Sandberg, *On the properties of some systems that distort signals I*, Bell System Tech. J., 42 (1963), pp. 2033–2047.

[23] K. Yosida, *Functional Analaysis*, Springer-Verlag, New York, 1964.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0014 $01.00/0

OPTIMAL LINEARIZATION OF EQUATIONS INVOLVING MONOTONE OPERATORS*

VACLAV DOLEZAL†

Abstract. In this paper we consider linearizations of a nonlinear equation in a Hilbert space which are close, in a certain sense, to an optimal linearization. We focus our attention on equations involving monotone operators, particularly on those with Hammerstein-type operators. Finally, we discuss a construction of a linearizing operator in the case when a Nemytskii operator is involved.

Introduction. Let T be a nonlinear operator with $T0=0$ and suppose that we wish to solve the equation $Tx=y$ for many different elements y which satisfy the inequality $\|y\|\leqq r$. Since solving a nonlinear equation is usually a difficult task, we may attempt to replace T by a linear operator T_0, and take the solution x_0 of $T_0x_0=y$ for an approximation to the exact solution x of $Tx=y$. Naturally, we would like to choose T_0 so that the over-all relative error, i.e., the supremum of all relative errors $\|x-x_0\|\cdot\|y\|^{-1}$ taken over all y's with $\|y\|\leqq r$, is made as small as possible.

An ideal situation would occur if there exists a linear operator T_* which would minimize the over-all relative error. Unfortunately, such an optimal operator T_* need not exist. Also, even if T_* exists, it would be hard to construct it.

On the other hand, looking at the problem from a practical point of view, we may satisfy ourselves with a partial solution: we can try to find a linear operator T_0 which is sufficiently close to T in a certain sense, and hope that the over-all relative error corresponding to T_0 will be small. It turns out that this is true for monotone and Hammerstein-type operators. As we shall see below, quite reasonable estimates can be given for $\|T^{-1}y-T_0^{-1}y\|\cdot\|y\|^{-1}$. Moreover, in particular cases it is usually not difficult to find such a T_0, and consequently, to replace the equation $Tx=y$ by a linear equation $T_0x_0=y$ without committing an unacceptable error.

Results. In order to formulate the ideas indicated above precisely, let us introduce some notations and concepts.

Let L be a Banach space; if $r>0$, let

$$B_r=\{x: x\in L, \|x\|\leqq r\}. \tag{1}$$

Let $\mathfrak{S}(L)$ be the collection of all nonempty subsets of L. Let $D\subset L$, $D\neq\emptyset$; if $T: D\to\mathfrak{S}(L)$, let $R(T)=\bigcup_{x\in D}Tx$. If, in particular, Tx is a singleton for each $x\in D$, T will be called an operator.

If $T: D\to\mathfrak{S}(L)$, we define the mapping $T^{-1}: R(T)\to\mathfrak{S}(D)$ by

$$T^{-1}y=\{x: x\in D,\ y\in Tx\}.$$

A mapping $T: D\to\mathfrak{S}(L)$ will be called simple on D, if

$$x_1, x_2\in D,\ x_1\neq x_2\Rightarrow (Tx_1)\cap(Tx_2)=\emptyset.$$

It is easy to see that:

T is simple on $D\Leftrightarrow T^{-1}$ is an operator on $R(T)$.

Finally, let $\mathfrak{L}$ be the set of all linear, bounded operators which are 1-to-1 from L onto L. (Note that, by the open mapping theorem, $A\in\mathfrak{L}\Rightarrow A^{-1}\in\mathfrak{L}$.)

---

* Received by the editors April 19, 1977, and in revised form September 19, 1977.

† Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York 11794. This research was supported by the National Science Foundation under Grant MPS7505268.

DEFINITION. Let $D \subset L$, $D \neq \emptyset$, $0 \in D$, let $T: D \to \mathfrak{S}(L)$ be simple on D, and let $r > 0$. If

(a) $R(T) \supset B_r$,

(b) there exists $K > 0$ such that the operator $T^{-1}: R(T) \to D$ satisfies the inequality

$$\|T^{-1}y\| \leqq K\|y\| \tag{2}$$

for all $y \in B_r$, we let

$$\lambda_r(T) = \inf_{T_\alpha \in \mathfrak{L}} \sup_{y \in B'_r} \|T^{-1}y - T_\alpha^{-1}y\| \cdot \|y\|^{-1}, \tag{3}$$

where $B'_r = B_r - \{0\}$.

Observe that this definition of the "minimal over-all relative error" $\lambda_r(T)$ is meaningful. Indeed, for any $T_\alpha \in \mathfrak{L}$ and $y \in B'_r$ we have $\|T^{-1}y - T_\alpha^{-1}y\| \cdot \|y\|^{-1} \leqq K + \|T_\alpha^{-1}\|$. Hence, $\lambda_r(T) \leqq K + 1$. Also, $\lambda_r(T)$ is a certain measure of nonlinearity of T with respect to B_r, since $\lambda_r(T) = 0$ whenever T is linear.

For our purposes, the meaning of $\lambda_r(T)$ is clarified by the following obvious proposition:

PROPOSITION. (i) *Let $\varepsilon > 0$; then there exists $T_0 \in \mathfrak{L}$ such that, for any $y \in B_r$ and x, $x_0 \in L$ with $y \in Tx$, $y = T_0x_0$, we have*

$$\|x - x_0\| \leqq (\lambda_r(T) + \varepsilon)\|y\|. \tag{4}$$

(ii) *Let $T_0 \in \mathfrak{L}$; then for any $\varepsilon > 0$ there exists $y' \in B_r$ such that*

$$\|x - x_0\| \geqq (\lambda_r(T) - \varepsilon)\|y'\| \tag{5}$$

for x, $x_0 \in L$ satisfying $y' \in Tx$, $y' = T_0x_0$.

(iii) *If, for some $T_0 \in \mathfrak{L}$ and $a > 0$, we have $\|T^{-1}y - T_0^{-1}y\| \leqq a\|y\|$ for all $y \in B_r$, then $\lambda_r(T) \leqq a$.*

As mentioned in the Introduction, an ideal case occurs if there exists $T_* \in \mathfrak{L}$ such that

$$\sup_{y \in B'_r} \|T^{-1}y - T_*^{-1}y\| \cdot \|y\|^{-1} = \lambda_r(T). \tag{6}$$

The question whether T_* exists or not can be reduced to a standard problem of approximation theory. Indeed, to see this, consider the following:

Let $\Omega \subset L$ be nonempty, and let $\mathscr{C}_0(\Omega)$ be the set of all operators $A: \Omega \to L$ such that $\|Ax\| \leqq C_A\|x\|$ for all $x \in \Omega$ with some $C_A > 0$. Note that if $A: L \to L$ is a linear bounded operator, then the restriction of A to Ω is in $\mathscr{C}_0(\Omega)$. Clearly, if we define the sum and scalar multiple in $\mathscr{C}_0(\Omega)$ pointwise and put

$$\|A\|_0 = \sup_{\substack{x \in \Omega \\ x \neq 0}} \|Ax\| \cdot \|x\|^{-1}, \tag{7}$$

then $\mathscr{C}_0(\Omega)$ becomes a Banach space.

Referring to (3) and (6) we see easily that an ideal case occurs iff there exists an $F_* \in \mathfrak{L}$ such that, in $\mathscr{C}(B_r)$, we have

$$\|T^{-1} - F_*\|_0 = \inf_{F_\alpha \in \mathfrak{L}} \|T^{-1} - F_\alpha\|_0, \tag{8}$$

i.e., F_* is a best approximation to $T^{-1} \in \mathscr{C}(B_r)$ in $\mathfrak{L}$, [1]. In this case, $T_* = F_*^{-1}$.

With this degree of generality we are unable to guarantee the existence of F_*. (Note that, in general, $\mathscr{C}_0(B_r)$ is not reflexive, and $\mathfrak{L}$ is neither convex nor closed in $\mathscr{C}_0(B_r)$.) A partial answer to the existence problem can be given for the rather uninteresting case in which L is finite dimensional. Indeed, here the subspace $\mathfrak{L}'$ of all linear (and consequently, bounded) operators $F_\alpha: L \to L$ is finite dimensional. Thus, there exists (see [1, p. 20]) $F_*^0 \in \mathfrak{L}'$ such that $\|T^{-1} - F_*^0\|_0 = \inf_{F_\alpha \in \mathfrak{L}'} \|T^{-1} - F_\alpha\|_0$. Now, if F_*^0 is invertible, it is the operator F_* we seek. (This would happen, if, for example, L is a Hilbert space, T^{-1} is strongly monotone and $\|T^{-1} - F_*^0\|_0$ is sufficiently small.)

Due to these facts, let us now turn to the alternative approach indicated in Introduction, i.e., examine operators $T_0 \in \mathfrak{L}$, for which $\|T^{-1} - T_0^{-1}\|_0$ is close to $\lambda_r(T)$.

Our first step in this direction is the following, rather trivial proposition.

THEOREM 1. *Let $D \subset L$, $D \neq \varnothing$, $0 \in D$, let $T: D \to \mathfrak{S}(L)$ be simple on D and satisfy conditions* (a), (b) *in the Definition, and let $r > 0$. Moreover, assume that there exists $T_0 \in \mathfrak{L}$ and $a > 0$ such that*

$$\|z - T_0 x\| \leqq a\|x\| \tag{9}$$

for all $x \in D \cap B_r$ and $z \in Tx$, where $R = Kr$. Then

$$\|T^{-1}y - T_0^{-1}y\| \leqq aK\|T_0^{-1}\| \cdot \|y\| \tag{10}$$

for all $y \in B_r$, i.e., $\lambda_r(T) \leqq aK\|T_0^{-1}\|$.

Proof. Choose $y \in B_r$ and put $x = T^{-1}y$. Then $y \in Tx$, and by (2), $\|x\| = \|T^{-1}y\| \leqq Kr = R$, i.e., $x \in D \cap B_R$. Thus, we have by (9),

$$\|T^{-1}y - T_0^{-1}y\| = \|T_0^{-1}(T_0x - y)\| \leqq \|T_0^{-1}\| \cdot \|T_0x - y\| \leqq \|T_0^{-1}\|a\|x\| \leqq aK\|T_0^{-1}\| \cdot \|y\|.$$

The inequality for $\lambda_r(T)$ follows from proposition (iii).

Much more efficient results can be obtained for monotone operators. Thus, from now on, let $L = H$ with H being a real Hilbert space.

THEOREM 2. *Let $D \subset H$, $D \neq \varnothing$, $0 \in D$, and let $r > 0$. Let $T: D \to \mathfrak{S}(H)$ be such that*

(i) *$R(T) \supset B_r$,*

(ii) *with some $b > 0$,*

$$\langle y_1 - y_2, x_1 - x_2 \rangle \geqq b\|x_1 - x_2\|^2 \tag{11}$$

for all $x_1, x_2 \in D$ and $y_1 \in Tx_1$, $y_2 \in Tx_2$. Furthermore, assume that

(iii) *$D \cap B_R$* is dense in *B_R with $R = b^{-1}r$,*

(iv) *there exists a linear bounded operator $T_0: H \to H$ and a constant a with $0 < a < b$ such that*

$$\|z - T_0 x\| \leqq a\|x\| \tag{12}$$

for all $x \in D \cap B_R$ and $z \in Tx$.

Then T is simple on D, $T_0 \in \mathfrak{L}$ and

$$\|T^{-1}y - T_0^{-1}y\| \leqq M\|y\| \tag{13}$$

for all $y \in B_r$, where $M = ab^{-1}(b-a)^{-1}$, i.e., $\lambda_r(T) \leqq M$.

Proof. Simplicity of T follows immediately from (11). Thus, $T^{-1}: R(T) \to D$ is an operator, and (11) implies that

$$\|T^{-1}x_1 - T^{-1}x_2\| \leqq b^{-1}\|x_1 - x_2\| \tag{14}$$

for all $x_1, x_2 \in R(T)$.

Next, the fact that $0 \in D$ and (12) show that $T0 = \{0\}$. Thus, $T^{-1}0 = 0$, and by (14),

(15) $$\|T^{-1}y\| \leqq b^{-1}\|y\|$$

for all $y \in R(T)$. Due to (i), (15) holds for every $y \in B_r$. Also, (15) shows that

(16) $$T^{-1}B_r \subset D \cap B_R.$$

On the other hand, if $x \in D$ and $y \in Tx$, we have by (11)

(17) $$\langle y, x\rangle \geqq b\|x\|^2.$$

Now, if $x \in D \cap B_R$, then for any $y \in Tx$,

(18) $$\langle T_0x, x\rangle = \langle y, x\rangle - \langle y - T_0x, x\rangle;$$

also, by (12),

$$\langle y - T_0x, x\rangle \leqq \|y - T_0x\| \cdot \|x\| \leqq a\|x\|^2.$$

Hence, (17), (18) yield

(19) $$\langle T_0x, x\rangle \geqq (b-a)\|x\|^2.$$

However, by (iii) and continuity of T_0 it follows that (19) holds in B_r, and consequently, by linearity of T_0, everywhere in H. Since $b - a > 0$, (19) shows that T_0 is a maximal monotone, coercive operator, so that $T_0H = H$ (see [2]). Hence, $T_0 \in \mathfrak{L}$. Furthermore, (19) gives, by the Schwarz inequality, $\|T_0x\| \geqq (b-a)\|x\|$; thus,

(20) $$\|T_0^{-1}\| \leqq (b-a)^{-1}.$$

Applying now Theorem 1 to our T_0 and using (20), (15), we find inequality (13) follows.

Remark 1. If T satisfies condition (ii) and is maximal monotone, then (i) is automatically met. Indeed, for each $x \in D$, let T^0x be the point in Tx having minimal norm. (Such point always exists, see [3, p. 104]). Then, by (17), $\|T^0x\| \geqq b\|x\|$ for each $x \in D$. Consequently, by Theorem 5 in [3], $R(T) = H$.

Condition (iii) in Theorem 2 can be modified so that we can drop the assumption on boundedness of T_0. Actually, we have:

THEOREM 2a. *The Theorem* 2 *remains true, if conditions* (iii) *and* (iv) *are replaced by the following assumptions*:

(iii)* $B_R \subset D$ *with* $R = b^{-1}r$,

(iv)* *there exists a linear operator* $T_0: H \to H$ *and a constant* a *with* $0 < a < b$ *such that* (12) *holds for all* $x \in B_r$ *and* $z \in Tx$.

Proof. If $x \in B_R$, we conclude as before that (19) holds. Thus, by linearity of T_0, (19) holds on the entire space H. Since T_0, being a linear operator, is hemicontinuous on H, it is maximal monotone and coercive by (19). Hence, $T_0H = H$, [2], i.e., T_0 is 1-to-1 onto. Moreover, by (19) the inverse T_0^{-1} is bounded, and we have (20). Finally, by the open mapping theorem, T_0 itself is bounded, and consequently, $T_0 \in \mathfrak{L}$. The rest of the proof is the same as before.

Let us now turn our attention to Hammerstein-type operators. To simplify the formulation of further results, let us introduce the following notation:

If α is a real number, let $\mathcal{M}(\alpha)$ be the set of all operators $T: H \to H$ such that

(21) $$\langle Tx_1 - Tx_2, x_1 - x_2\rangle \geqq \alpha\|x_1 - x_2\|^2$$

for all $x_1, x_2 \in H$. Similarly, if $\alpha \geqq 0$, let Lip (α) be the set of all operators $T: H \to H$

such that

$$\|Tx_1 - Tx_2\| \leqq \alpha \|x_1 - x_2\| \tag{22}$$

for all $x_1, x_2 \in H$.

If $T \in \mathrm{Lip}\,(\alpha)$, we put

$$\|T\|^* = \sup_{\substack{x_1, x_2 \in H \\ x_1 \neq x_2}} \|Tx_1 - Tx_2\| \cdot \|x_1 - x_2\|^{-1}. \tag{23}$$

Clearly, if $T_1, T_2 \in \mathrm{Lip}\,(\alpha)$, then

$$\|T_1 + T_2\|^* \leqq \|T_1\|^* + \|T_2\|^*, \qquad \|T_1 T_2\|^* \leqq \|T_1\|^* \|T_2\|^*.$$

If, in particular, T is linear and bounded, then $\|T\| = \|T\|^*$.

LEMMA 1. *Let $T \in \mathcal{M}(\alpha)$ with $\alpha > 0$ and let $T \in \mathrm{Lip}\,(\beta)$ with $\beta > 0$. Then T is 1-to-1 onto H, and we have*

$$T^{-1} \in \mathcal{M}(\alpha\beta^{-2}), \qquad T^{-1} \in \mathrm{Lip}\,(\alpha^{-1}). \tag{24}$$

Proof. Our assumptions show that T is maximal monotone and coercive. Hence, $TH = H$. Since $T \in \mathcal{M}(\alpha)$ implies that T is 1-to-1, T^{-1} exists. Relations (24) follows readily from our assumptions by using Schwarz inequality.

THEOREM 3. *Let $A \in \mathcal{M}(0)$ be linear and bounded, and let $B \in \mathcal{M}(b)$, $b > 0$, $B \in \mathrm{Lip}\,(c)$. If $T = I + AB$, then T is 1-to-1 onto H and*

$$\|T^{-1}\|^* \leqq b^{-2}c^2. \tag{25}$$

Furthermore, let $r > 0$ and assume that there exists a linear operator $B_0 : H \to H$ and a constant a with $0 < a < b$ such that

$$\|Bx - B_0 x\| \leqq a\|x\| \tag{26}$$

for all $x \in B_R$ with $R = b^{-2}c^2 r$. Then B_0 is bounded, $\|B_0\| \leqq a + c$, and the operator $T_0 = I + AB_0$ has the following properties:

(i) *T_0 is 1-to-1 onto H and*

$$\|T_0^{-1}\| \leqq (b-a)^{-2}\|B_0\|^2; \tag{27}$$

(ii) *For every $y \in B_r$,*

$$\|T^{-1}y - T_0^{-1}y\| \leqq K\|y\| \tag{28}$$

with

$$K = \frac{ac^2\|A\|\,\|B_0\|^2}{b^2(b-a)^2}, \tag{29}$$

i.e., $\lambda_r(T) \leqq K$.

Proof. The conditions $B \in \mathcal{M}(b)$, $b > 0$, $B \in \mathrm{Lip}\,(c)$ imply by Lemma 1 that B is invertible, and $B^{-1} \in \mathcal{M}(bc^{-2})$, $B^{-1} \in \mathrm{Lip}\,(b^{-1})$. Since $A \in \mathcal{M}(0)$, it follows that $B^{-1} + A \in \mathcal{M}(bc^{-2})$ and $B^{-1} + A \in \mathrm{Lip}\,(b^{-1} + \|A\|)$. Consequently, again by Lemma 1, $B^{-1} + A$ is invertible and $(B^{-1} + A)^{-1} \in \mathrm{Lip}\,(b^{-1}c^2)$. Next, $T = (B^{-1} + A)B$, and consequently, T is invertible. Since $T^{-1} = B^{-1}(B^{-1} + A)^{-1}$, we have

$$\|T^{-1}\|^* \leqq \|B^{-1}\|^* \|(B^{-1} + A)^{-1}\|^* \leqq b^{-2}c^2. \tag{30}$$

Now, assume that (26) holds. Then, first of all, $B0 = 0$. Also, (26) implies by condition $B \in \mathrm{Lip}\,(c)$ that $\|B_0 x\| \leqq (a + c)\|x\|$ for all $x \in B_R$; hence, due to linearity, B_0 is bounded and $\|B_0\| \leqq a + c$.

Using the same argument as in the proof of Theorem 2 we conclude that $\langle B_0x, x\rangle \geqq (b-a)\|x\|^2$ for all $x \in B_R$. Hence, by linearity, $B_0 \in \mathcal{M}(b-a)$, $B_0 \in \mathrm{Lip}\,(\|B_0\|)$. Applying the above results concerning T to T_0, we see readily that T_0 is invertible, and by (30),

$$(31) \qquad \|T_0^{-1}\|^* = \|T_0^{-1}\| \leqq (b-a)^{-2}\|B_0\|^2.$$

Finally, let $y \in B_r$; then $\|T^{-1}y\| \leqq \|T^{-1}\|^*\|y\| \leqq b^{-2}c^2r = R$, and we have by (26),

$$\begin{aligned}\|T^{-1}y - T_0^{-1}y\| &= \|T_0^{-1}(T_0 - T)T^{-1}y\| \\ &= \|T_0^{-1}A(B_0 - B)T^{-1}y\| \leqq (b-a)^{-2}\|B_0\|^2\|A\|ab^{-2}c^2\|y\|,\end{aligned}$$

which confirms (28), (29). Hence, the proof.

THEOREM 4. *Let $B \in \mathcal{M}(0)$ be linear and bounded, and let $A \in \mathcal{M}(b)$, $b > 0$, $A \in \mathrm{Lip}\,(c)$. If $T = I + AB$, then T is 1-to-1 onto H and*

$$(32) \qquad \|T^{-1}\|^* \leqq b^{-2}c^2.$$

Furthermore, let $r > 0$ and assume that there exists a linear operator $A_0: H \to H$ and a constant a with $0 < a < b$ such that

$$(33) \qquad \|Ax - A_0x\| \leqq a\|x\|$$

for all $x \in B_R$ with $R = \|B\|b^{-2}c^2r$. Then A_0 is bounded, $\|A_0\| \leqq a + c$, and the operator $T_0 = I + A_0B$ has the following properties:

(i) *T_0 is 1-to-1 onto H and*

$$(34) \qquad \|T_0^{-1}\| \leqq (b-a)^{-2}\|A_0\|^2.$$

(ii) *For every $y \in B_r$,*

$$(35) \qquad \|T^{-1}y - T_0^{-1}y\| \leqq K\|y\|$$

with

$$(36) \qquad K = \frac{ac^2\|A_0\|^2\|B\|}{b^2(b-a)^2},$$

i.e., $\lambda_r(T) \leqq K$.

Proof. Using the same argument as in the proof of Theorem 3 we confirm easily that all claims up to (i) are true. As for (ii), note that if $y \in B_r$, then by (32) $\|BT^{-1}y\| \leqq \|B\|b^{-2}c^2r = R$. Thus, we have by (35), $\|T^{-1}y - T_0^{-1}y\| = \|T_0^{-1}(A_0 - A)BT^{-1}y\| \leqq (b-a)^{-2}\|A_0\|^2a\|B\|b^{-2}c^2\|y\|$ which concludes the proof.

In order to prove certain "duals" of Theorems 3 and 4, we will need the following result:

LEMMA 2. *Let $A: H \to H$ be a linear operator with $A \in \mathcal{M}(b)$, $b > 0$, and let $B: H \to H$ be a hemicontinuous operator with $B \in \mathcal{M}(-\mu)$, $\mu \geqq 0$. If $b - \mu\|A\|^2 > 0$, then $T = I + AB$ is 1-to-1 onto H, and*

$$(37) \qquad \|T^{-1}\|^* \leqq \|A\|(b - \mu\|A\|^2)^{-1}.$$

Proof. First observe that A is bounded. Indeed, since A is linear, it is hemicontinuous, and consequently, by assumption $A \in \mathcal{M}(b)$, $b > 0$, A is maximal monotone and coercive. Hence, A is 1-to-1 onto H, and $\|A^{-1}\| \leqq b^{-1}$. Thus, by the open mapping theorem, A itself is bounded.

Next, consider the operator $C = TA^* = A^* + ABA^*$, where A^* signifies the adjoint of A. Clearly, C is hemicontinuous. Indeed, A^* is continuous, and for any x_0, $w \in H$ and number sequence $t_n \to 0$ we have $BA^*(x_0+t_nw)=B(A^*x_0+t_nA^*w)\overset{w}{\to} BA^*x_0$, so that, by boundedness of A, $ABA^*(x_0+t_nw)\overset{w}{\to} ABA^*x_0$. Moreover, for any $x_1, x_2 \in H$,

$$\begin{aligned}\langle Cx_1-Cx_2, x_1-x_2\rangle &= \langle A^*(x_1-x_2), x_1-x_2\rangle+\langle ABA^*x_1-ABA^*x_2, x_1-x_2\rangle \\ &= \langle A(x_1-x_2), x_1-x_2\rangle+\langle BA^*x_1-BA^*x_2, A^*x_1-A^*x_2\rangle.\end{aligned} \tag{38}$$

However,

$$\langle BA^*x_1-BA^*x_2, A^*x_1-A^*x_2\rangle \geqq -\mu\|A^*x_1-A^*x_2\|^2=-\mu\|A^*(x_1-x_2)\|^2,$$

and $\|A^*(x_1-x_2)\|\leqq\|A^*\|\,\|x_1-x_2\|=\|A\|\,\|x_1-x_2\|$. Hence, by (38) and since $A\in\mathcal{M}(b)$,

$$\langle Cx_1-Cx_2, x_1-x_2\rangle \geqq (b-\mu\|A\|^2)\|x_1-x_2\|^2. \tag{39}$$

Thus, if $b-\mu\|A\|^2>0$, (39) and hemicontinuity of C show that C is a 1-to-1, maximal monotone and coercive operator. Hence, C^{-1} exists, and by (39),

$$\|C^{-1}\|^* \leqq (b-\mu\|A\|^2)^{-1}.$$

Now, since A^* is invertible and $\|A^{*-1}\|\leqq b^{-1}$, the relation $T=CA^{*-1}$ implies that T is also invertible. Thus, due to $T^{-1}=A^*C^{-1}$,

$$\|T^{-1}\|^*\leqq\|A^*\|\,\|C^{-1}\|^*\leqq\|A\|(b-\mu\|A\|^2)^{-1}.$$

This completes the proof.

THEOREM 5. *Let $A:H\to H$ be a linear operator with $A\in\mathcal{M}(b)$, $b>0$, and let $B:H\to H$ be a hemicontinuous operator with $B\in\mathcal{M}(0)$. Let $r>0$ and assume that there exist a linear operator $B_0:H\to H$ and $a>0$ such that*

$$\|Bx-B_0x\|\leqq a\|x\| \tag{40}$$

for all $x\in B_R$ with $R=\|A\|b^{-1}r$. Let $T=I+AB$ and $T_0=I+AB_0$; if $b-a\|A\|^2>0$, then both T and T_0 are 1-*to*-1 *onto H, and*

$$\|T^{-1}y-T_0^{-1}y\|\leqq K\|y\|$$

for all $y\in B_r$ with

$$K=ab^{-1}\|A\|^3(b-a\|A\|^2)^{-1}, \tag{41}$$

i.e., $\lambda_r(T)\leqq K$.

Proof. By Lemma 2, T is invertible and $\|T^{-1}\|^*\leqq\|A\|b^{-1}$. Moreover, from (40) and because $B\in\mathcal{M}(0)$ it follows as before that $\langle B_0x, x\rangle\geqq -a\|x\|^2$ for all $x\in B_R$. Due to the linearity of B_0, this inequality is satisfied for each $x\in H$, i.e., $B_0\in\mathcal{M}(-a)$. Also, since B_0 is linear, it is hemicontinuous. Thus, if $b-a\|A\|^2>0$, Lemma 2 shows that T_0 is invertible and $\|T_0^{-1}\|^*\leqq\|A\|(b-a\|A\|^2)^{-1}$.

Next, if $y\in B_r$, then $\|T^{-1}y\|\leqq\|T^{-1}\|^*\|y\|\leqq\|A\|b^{-1}r=R$ (note that (40) yields $B0=0\Rightarrow T0=0\Rightarrow T^{-1}0=0$), and we have

$$\|T^{-1}y-T_0^{-1}y\|=\|T_0^{-1}A(B_0-B)T^{-1}y\|\leqq\|A\|(b-a\|A\|^2)^{-1}a\|A\|^2b^{-1}\|y\|$$

which is what we wanted to show.

LEMMA 3. *Let $B:H\to H$ be a linear operator with $B\in\mathcal{M}(b)$, $b>0$, and let $A:H\to H$ be a hemicontinuous operator with $A\in\mathcal{M}(-\mu)$, $\mu\geqq 0$. If $b-\mu\|B\|^2>0$, then $T=I+AB$ is* 1-*to*-1 *onto H, and*

$$\|T^{-1}\|^*\leqq\|B\|(b-\mu\|B\|^2)^{-1}. \tag{42}$$

Proof. As in the proof of Lemma 2 it follows that B is an invertible bounded operator and $\|B^{-1}\| \leqq b^{-1}$. Moreover, letting $C = B^*T = B^* + B^*AB$, we confirm easily as before that C is hemicontinuous and satisfies the condition

$$\langle Cx_1 - Cx_2, x_1 - x_2 \rangle \geqq (b - \mu\|B\|^2)\|x_1 - x_2\|^2$$

for all $x_1, x_2 \in H$. Thus, C is invertible and

$$\|C^{-1}\|^* \leqq (b - \mu\|B\|^2)^{-1}. \tag{43}$$

Now, since B^* is invertible and $\|B^{*-1}\| \leqq b^{-1}$, we have $T = B^{*-1}C$. Consequently, $T^{-1} = C^{-1}B^*$, so that $\|T^{-1}\|^* \leqq \|C^{-1}\|^*\|B^*\|$, from which (42) follows immediately.

THEOREM 6. *Let $B: H \to H$ be a linear operator with $B \in \mathcal{M}(b)$, $b > 0$, and let $A: H \to H$ be a hemicontinuous operator with $A \in \mathcal{M}(0)$. Let $r > 0$ and assume that there exist a linear operator $A_0: H \to H$ and $a > 0$ such that*

$$\|Ax - A_0x\| \leqq a\|x\| \tag{44}$$

for all $x \in B_r$ with $R = \|B\|^2 b^{-1} r$. Let $T = I + AB$ and $T_0 = I + A_0B$; if $b - a\|B\|^2 > 0$, then both T and T_0 are 1-to-1 onto H, and

$$\|T^{-1}y - T_0^{-1}y\| \leqq K\|y\| \tag{45}$$

for all $y \in B_r$ with

$$K = ab^{-1}\|B\|^3(b - a\|B\|^2)^{-1}, \tag{46}$$

i.e., $\lambda_r(T) \leqq K$.

Proof. By Lemma 3, T is invertible and $\|T^{-1}\|^* \leqq \|B\|b^{-1}$. From (44) and $A \in \mathcal{M}(0)$ it follows that $A_0 \in \mathcal{M}(-a)$. Since A_0 is hemicontinuous, Lemma 3 shows that T_0^{-1} exists and $\|T_0^{-1}\|^* \leqq \|B\|(b - a\|B\|^2)^{-1}$. Now, if $y \in B_r$, then $\|BT^{-1}y\| \leqq \|B\|^2 b^{-1} r = R$, and we have $\|T^{-1}y - T_0^{-1}y\| = \|T_0^{-1}(A_0 - A)BT^{-1}y\| \leqq \|B\|(b - a\|B\|^2)^{-1} a\|B\| \cdot \|B\|b^{-1}\|y\|$ which proves (45) and (46).

In Theorems 2a and 3–6 we assumed existence of a linear operator $T_0: H \to H$ (or A_0, B_0) such that the norm $\|T - T_0\|_0$ in $\mathscr{C}_0(B_R)$ is sufficiently small. Thus, from a practical point of view it is important to know how to construct such T_0 for a given T. A quite simple construction of T_0 can be divsed if $H = L_2^n(a, b)$, (n-tuple Cartesian product of $L_2(a, b)$ with itself having the obvious inner product) and T has the form $T = T_1 + T_2F$, where T_1, T_2 are linear and F is a Nemytskii operator. We will look for T_0 in the form $T_0 = T_1 + T_2F_0$, where F_0 is generated by a matrix and $\|F - F_0\|_0$ is sufficiently small. In order to discuss the construction of F_0, we will need the following proposition:

LEMMA 4. *Let $f \in \mathscr{C}_0(R^n)$ be continuous and let the operator F be defined on $L_2^n(a, b)$ by $(Fx)(t) = f(x(t))$, $t \in (a, b)$. Then $F \in \mathscr{C}_0(L_2^n(a, b))$ and $\|f\|_0 = \|F\|_0$.*

Proof. For any $\xi \in R^n$ we have $|f(\xi)| \leqq \|f\|_0|\xi|$. (Here, $|\cdot|$ signifies the Euclidean norm in R^n.) Thus, if $x \in L_2^n(a, b)$, Fx is measurable and

$$\|Fx\|^2 = \int_a^b |f(x(t))|^2\,dt \leqq \|f\|_0^2 \int_0^b |x(t)|^2\,dt = \|f\|_0^2\|x\|^2,$$

i.e., $F \in \mathscr{C}_0(L_2^n(a, b))$ and $\|F\|_0 \leqq \|f\|_0$.

On the other hand, for each $\varepsilon > 0$ there exists $\xi_0 \in R^n$, $\xi_0 \neq 0$ such that $\|f\|_0 - \varepsilon \leqq |f(\xi_0)| \cdot |\xi_0|^{-1}$. Choose α, β with $\alpha < \beta$ so that $[\alpha, \beta] \subset (a, b)$ and define $x_0(t)$ by $x_0(t) = \xi_0$ on $[\alpha, \beta]$, $x_0(t) = 0$ elsewhere in (a, b). Then $x_0 \in L_2^n(a, b)$, $x_0 \neq 0$, and the above inequality yields

$$(\|f\|_0-\varepsilon)^2\|x_0\|^2=\int_\alpha^\beta (\|f\|_0-\varepsilon)^2|\xi_0|^2\,dt \leqq \int_\alpha^\beta |f(\xi_0)|^2\,dt=\|Fx_0\|^2,$$

i.e., $\|f\|_0-\varepsilon \leqq \|Fx_0\|\cdot\|x_0\|^{-1}\leqq\|F\|_0$. Hence, $\|f\|_0=\|F\|_0$.

To state the theorem, let us introduce the following notation:

Let G be the space of all $n\times n$ matrices with constant entries. If $\nu\in G$, let $m_\nu\in\mathscr{C}_0(R^n)$ be defined by $m_\nu\xi=\nu\cdot\xi$, and $F_\nu\in\mathscr{C}_0(L_2^n(a,b))$ by $(F_\nu x)(t)=\nu\cdot x(t)$.

THEOREM 7. *Let $f\in\mathscr{C}_0(R^n)$ be continuous; then there exists $\nu_0\in G$ such that in $\mathscr{C}_0(R^n)$ we have*

$$\|f-m_{\nu_0}\|_0=\inf_{\nu\in G}\|f-m_\nu\|_0=\lambda. \tag{47}$$

Furthermore, if $F\in\mathscr{C}_0(L_2^n(a,b))$ is defined by $(Fx)(t)=f(x(t))$, then in $\mathscr{C}_0(L_2^n(a,b))$ we have

$$\|F-F_{\nu_0}\|_0=\inf_{\nu\in G}\|F-F_\nu\|_0=\lambda. \tag{48}$$

Proof. Since G is finite dimensional, a $\nu_0\in G$ satisfying (47) exists [1]. However, Lemma 4 shows that, for any $\nu\in G$, $\|f-m_\nu\|_0=\|F-F_\nu\|_0$. Hence, (48) follows.

A minimizing matrix ν_0 can easily be found, provided f is "diagonal". Actually, we have the following, fairly obvious result:

THEOREM 8. *Let $f\in\mathscr{C}(R^n)$ have the form $f(\xi)=[f_1(\xi_1),f_2(\xi_2),\cdots,f_n(\xi_n)]^T$, where $f_i\in\mathscr{C}_0(R^1)$, $i=1,2,\cdots,n$. Then there exists a unique $\nu_0\in G$ that satisfies* (47). *Moreover, $\nu_0=\operatorname{diag}(d_1,d_2,\cdots,d_n)$, where*

$$d_i=\tfrac12(S_i+I_i),\qquad S_i=\sup_{R_0^1}\xi^{-1}f_i(\xi),$$

$$I_i=\inf_{R_0^1}\xi^{-1}f_i(\xi),\qquad R_0^1=R^1-\{0\},$$

$$i=1,2,\cdots,n,\quad\text{and}\quad \lambda=\tfrac12\max_i(S_i-I_i).$$

In conclusion, let us consider two simple examples dealing with vector integral equations.

Example 1. Let L_2 and L_1 be the real space $L_2(-\infty,\infty)$ and $L_1(-\infty,\infty)$, respectively, and let us make the following assumptions:

(i) Let $k(t)$ be an $n\times n$ matrix each element of which is in $L_2\cap L_1$, let $K(iw)$ be the Fourier transform of $k(t)$, and let $M(iw)=\frac12(K(iw)+\overline{K(iw)}^T)$ be positive semidefinite for every $w\in R^1$, i.e., $\bar\xi^TM(iw)\xi\geqq0$ for each complex ξ.

(ii) Let $f:R^n\to R^n$ be such that $f(0)=0$ and

$$(f(\xi_1)-f(\xi_2))^T(\xi_1-\xi_2)\geqq b|\xi_1-\xi_2|^2,\qquad |f(\xi_1)-f(\xi_2)|\leqq c|\xi_1-\xi_2| \tag{49}$$

for all $\xi_1,\xi_2\in R^n$ and some $b>0$ and $c>0$.

(iii) Let C be a constant $n\times n$ matrix such that

$$|f(\xi)-C\xi|\leqq a|\xi| \tag{50}$$

for all $\xi\in R^n$ and some $a>0$ with $a<b$.

We are going to show that, for any $g \in L_2^n$,

(a) the equation

$$x(t)+\int_{-\infty}^{\infty} k(t-\tau)f(x(\tau))\,d\tau = g(t), \qquad t \in R^1 \tag{51}$$

possesses a unique solution x in L_2^n and x depends continuously on g,

(b) the equation

$$x_0(t)+\int_{-\infty}^{\infty} k(t-\tau)Cx_0(\tau)\,d\tau = g(t), \qquad t \in R^1 \tag{52}$$

possesses a unique solution x_0 in L_2^n, and

$$\|x-x_0\| \leqq a\mathscr{H}|C|^2c^2b^{-2}(b-a)^{-2}\|g\|, \tag{53}$$

where $\mathscr{H}=(\sum_{ij}\|k_{ij}\|_{L_1}^2)^{1/2}$ and $|C|$ is the norm of the matrix C associated with the Euclidean vector norm.

Indeed, let the operators A, B be defined on L_2^n by

$$(Ax)(t)=\int_{-\infty}^{\infty} k(t-\tau)x(\tau)\,d\tau, \qquad (Bx)(t)=f(x(t)). \tag{54}$$

It is well known [4] that if $k_0 \in L_2 \cap L_1$ and $(A_0z)(t)=\int_{-\infty}^{\infty} k_0(t-\tau)z(\tau)\,d\tau$, $z \in L_2$, then A_0 is a linear bounded operator from L_2 into itself, and $\|A_0\| \leqq \|k_0\|_{L_1}$. Consequently, our A is linear and bounded, because for any $x \in L_2^n$ we have

$$\begin{aligned}\|Ax\|^2 &= \sum_i \|\sum_j k_{ij} * x_j\|^2 \leqq \sum_i \Big(\sum_j \|k_{ij} * x_j\|\Big)^2 \\ &\leqq \sum_i \Big(\sum_j \|k_{ij}\|_{L_1} \cdot \|x_j\|\Big)^2 \leqq \sum_i \Big(\sum_j \|k_{ij}\|_{L_1}^2\Big)\Big(\sum_j \|x_j\|^2\Big) = \mathscr{H}^2\|x\|^2.\end{aligned}$$

Hence, $\|A\| \leqq \mathscr{H}$.

Next, choose $x \in L_2^n$, and let $\hat{x}$ and $(\widehat{Ax})$ be the Fourier–Plancherel transform of x and Ax, respectively [4]. Then $(\widehat{Ax})=K\hat{x}$, and Paraseval's equality yields

$$\langle Ax, x\rangle = \langle(\widehat{Ax}), \hat{x}\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\overline{K\hat{x}})^T\hat{x}\,dw = (2\pi)^{-1}\int_{-\infty}^{\infty} \bar{\hat{x}}^T\bar{K}^T\hat{x}\,dw.$$

Since $\langle Ax, x\rangle$ is real, we also have

$$\langle Ax, x\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{x}^TK^T\bar{\hat{x}}\,dw = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{\hat{x}}^TK\hat{x}\,dw;$$

hence, $\langle Ax, x\rangle = (2\pi)^{-1}\int_{-\infty}^{\infty} \bar{\hat{x}}^TM\hat{x}\,dw \geqq 0$ by (i), i.e., $A \in \mathscr{M}(0)$.

On the other hand, from assumption (ii) it follows easily that B maps L_2^n into itself, and that $B \in \mathscr{M}(b)$, $B \in \operatorname{Lip}(c)$. Thus, by Theorem 3, equation (51) possesses a unique solution x in L_2^n, and x depends continuously on g.

Moreover, (50) shows that $\|f-C\|_0 \leqq a$ in $\mathscr{C}_0(R^n)$; thus, by Lemma 4, $\|B-C\|_0 \leqq a$ in $\mathscr{C}_0(L_2^n)$, i.e., $\|Bx-Cx\| \leqq a\|x\|$ for all $x \in L_2^n$. Also, $\|C\|=|C|$. Invoking again Theorem 3 we see that equation (52) has a unique solution x_0 in L_2^n, and that estimate (53) holds.

Example 2. Let now L_2^0 and L_1^0 be the real space $L_2(0, T)$ and $L_1(0, T)$, respectively, and let $k^0(t)$ be an $n \times n$ matrix each element of which is in $L_2^0 \cap L_1^0$. Define matrix $k(t)$ by $k(t)=k^0(t)$ for $t \in [0, T]$, $k(t)=0$ elsewhere, and assume that $k(t)$

satisfies the condition (i) in Example 1. (Note that now $K(iw)=\int_0^T k^0(t)\,e^{iwt}\,dt$.) Moreover, let f and C satisfy the assumptions (ii) and (iii).

We are going to prove that, for any $g \in L_2^{0n}$, the equations

$$x(t)+\int_0^t k^0(t-\tau)f(x(\tau))\,d\tau = g(t), \qquad t\in[0, T], \tag{55}$$

and

$$x_0(t)+\int_k^t k^0(t-\tau)Cx_0(\tau)\,d\tau = g(t), \qquad t\in[0, T], \tag{56}$$

possess a unique solution x and x_0 in L_2^{0n}, respectively, and that inequality (53) is true.

To do this, we can use results established in Example 1. Indeed, if $x^0 \in L_2^0$, let $x\in L_2$ be defined by $x(t)=x^0(t)$ on $[0, T]$, $x(t)=0$ elsewhere. Similarly, if $m^0\in L_2^0\cap L_1^0$, construct $m\in L_2\cap L_1$.

Defining operators $\bar{A}^0 : L_2^0 \to L_2^0$ and $\bar{A}: L_1 \to L_2$ by

$$(\bar{A}^0x)(t)=\int_0^t m^0(t-\tau)x(\tau)\,d\tau, \qquad (\bar{A}z)(t)=\int_{-\infty}^{\infty} m(t-\tau)z(\tau)\,d\tau, \tag{57}$$

we confirm easily that, for any $x^0\in L_2^0$ and the corresponding $x\in L_2$, we have $(\bar{A}x)(t)=(\bar{A}^0x^0)(t)$ for every $t\in[0, T]$. Consequently, $\|\bar{A}^0x^0\|\leqq\|\bar{A}x\|\leqq\|m\|_{L_1}\cdot\|x\|=\|m^0\|_{L_1^0}\cdot\|x^0\|$. From this it follows as in Example 1 that the operator $(A^0x)(t)=\int_0^t k^0(t-\tau)x(\tau)\,d\tau$ is bounded on L_2^{0n}, and $\|A^0\|\leqq\mathcal{K}$.

Moreover, for any $x^0\in L_2^0$ we clearly have $\langle\bar{A}^0x^0, x^0\rangle=\langle\bar{A}x, x\rangle$. Consequently, under the assumptions made, $\langle A^0z, z\rangle\geqq 0$ for each $z\in L_2^{0n}$, i.e., $A^0\in\mathcal{M}(0)$.

Defining B on L_2^{0n} by (54), we have as before $B\in\mathcal{M}(b)$ and $B\in\mathrm{Lip}\,(c)$. Hence, operators A^0, B meet the assumptions of Theorem 3 and our claim follows.

Note that the bound for $\|x-x_0\|$ given by (53) is sharper than that obtained by applying the Bellman–Gronwall lemma.

REFERENCES

[1] E. W. CHENEY, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.

[2] R. T. ROCKAFELLAR, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., 149 (1970), pp. 75–88.

[3] H. BREZIS, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, Contributions to Nonlinear Functional Analysis, E. Zarantonello, ed., Academic Press, New York, 1971, pp. 101–156.

[4] S. BOCHNER AND K. CHANDRASEKHARAN, *Fourier Transforms*, Annals of Mathematics Studies 19, Princeton University Press, Princeton, New Jersey, 1948.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0015 $01.00/0

ABSTRACT LINEAR AND NONLINEAR VOLTERRA EQUATIONS PRESERVING POSITIVITY*

PH. CLÉMENT† AND J. A. NOHEL‡

Abstract. Let X be a real or complex Banach space. We study the Volterra equation

$$u(t)+\int_0^t a(t-s)Au(s)\,ds=f(t) \qquad (0\leqq t\leqq T,\ T>0), \tag{v}$$

where a is a given kernel, A is a bounded or unbounded linear operator from X to X, and f is a given function with values in X. (Of particular importance is the case $f=u_0+a*g$, $u_0\in X$, $g\in L^1(0,T;X)$, where $*$ denotes the convolution). We establish existence, uniqueness, continuity results and sufficient conditions involving a, A, f which insure that solutions of (v) are positive by using certain representation formulas for solutions of (v). We also discuss the positivity of solutions of (v) when A is a nonlinear (m-accretive) operator and we discuss several examples when A is a partial differential operator.

1. Introduction and principal results. Let X be a real or complex Banach space. We study the linear Volterra equation

$$u(t)+a*Au(t)=f(t) \qquad (0\leqq t\leqq T;\ T>0) \tag{1.1}$$

where $a*Au(t)=\int_0^t a(t-s)Au(s)\,ds$, a is a given real kernel, A is a bounded or unbounded linear operator from X to X and f is a given function with values in X.

An important and perhaps the most useful special case of (1.1) for certain applications is the linear equation

$$u(t)+a*Au(t)=u_0+a*g(t) \qquad (0\leqq t\leqq T;\ T>0) \tag{1.1a}$$

where $u_0\in X$ and the given function $g\in L^1(0,T;X)$. We will establish conditions on the kernel a and the operator A which insure that the respective solutions operators for (1.1) and (1.1a) preserve a convex cone in X (see Theorems 3 and 4). One motivation for studying this property of solutions of (1.1) is that if in (1.1) $a\equiv 1$ and $Au\equiv -\nabla^2 u$ in a bounded region $\Omega\subseteq\mathbb{R}^n$ with Dirichlet boundary conditions on the smooth boundary Γ of Ω, (1.1) is equivalent to the heat equation; thus our theory for (1.1) is a natural extension of classical results on the positivity of solutions of the heat equation. We then consider in § 3 the same question for a nonlinear problem of the form (1.1) in which A is an m-accretive operator. Finally, in § 4 we discuss three examples to illustrate the theory.

We will suppose throughout that the following assumptions are satisfied:

(H_1) $A:D(A)\subseteq X\to X$ and $-A$ generates a linear continuous contraction semigroup on X, which we shall denote by $e^{-\omega A}$ ($\omega\geqq 0$),

(H_2) $a\in L^1(0,T;\mathbb{R})$, and

(H_3) $f\in L^1(0,T;X)$.

* Received by the editors February 9, 1977, and in final revised form November 17, 1977.

† Mathematics Research Center, University of Wisconsin—Madison, Madison, Wisconsin. Now at University of Technology, Delft, The Netherlands. The work of this author was sponsored by the United States Army under Contract DAAG29-75-C-0024 and by the Swiss National Foundation.

‡ Department of Mathematics, University of Wisconsin—Madison, Madison, Wisconsin 53706. The work of this author was sponsored by the United States Army under Contract DAAG29-75-C-0024 and under Grant DAAG29-77-G-0004, and by the National Science Foundation under Grant MCS75-21868.

DEFINITION 1. We say that u: $[0, T]\to X$ is a *strong solution of* (1.1) if $u\in L^1(0, T; X)$, $u(t)\in D(A)$ a.e. on $[0, T]$, $Au\in L^1(0, T; X)$, and u satisfies (1.1) a.e. on $[0, T]$.

We denote the norm in X by $\|\cdot\|$. If B is a linear unbounded operator on X, we use the notation $X_B = D(B)$; if $u\in X_B$, its graph norm is denoted by $\|u\|_{X_B} = \|u\|+\|Bu\|$. Of particular interest are the spaces X_A and X_{A^2} where A satisfies (H_1). Recall that the space X_A is dense in X and X_{A^2} is dense in X_A; see [20, Thm 2.9, p. 8]. If u is a strong solution of (1.1), Definition 1 states that $u\in L^1(0, T; X_A)$.

To discuss solutions of (1.1) and (1.1a) we make use of the operators R and S defined respectively by the equations

$$(R)\qquad u(t)+a*Au(t)=a(t)x \qquad (x\in X_A;\ 0\leqq t\leqq T),$$

$$(S)\qquad u(t)+a*Au(t)=x \qquad (x\in X_A;\ 0\leqq t\leqq T).$$

It follows that under the assumptions of Theorem 1 below, equations (R) and (S) each have a unique strong solution which we write respectively as $R(t)x$ and $S(t)x$. While the operators R and S are so defined for $x\in X_A$, Theorem 1 together with a density argument shows that R and S can be extended uniquely as bounded operators in $L^1(0, T; X)$ and $L^\infty(0, T; X)$ respectively.

Much of the analysis will rest on properties of solutions of the two scalar equations corresponding to the abstract equations (R), (S) respectively. The first is the (resolvent) equation associated with the kernel a:

$$(r_\lambda)\qquad r(t)+\lambda a*r(t)=a(t) \qquad (0\leqq t\leqq T;\ \lambda\geqq 0).$$

It is well known [19] that if a satisfies assumption (H_2), then for every $\lambda\geqq 0$ equation (r_λ) has a unique solution $r(t,\lambda)\in L^1(0, T; \mathbb{R})$. We shall make use of the assumption (see also Proposition 1(i) below):

(H_4) For every $\lambda\geqq 0$ the solution $r(t,\lambda)$ of (r_λ) satisfies $r(t,\lambda)\geqq 0$ a.e. on $[0, T]$.

The scalar equation corresponding to (S) is

$$(s_\lambda)\qquad s(t)+\lambda a*s(t)=1 \qquad (0\leqq t\leqq T;\ \lambda\geqq 0).$$

If a satisfies assumption (H_2) it is clear that (s_λ) has a unique solution $s(t,\lambda)\in L^1(0, T{:}\ \mathbb{R})$. But as is readily verified,

$$s(t,\lambda)=1-\lambda\int_0^t r(\sigma,\lambda)\,d\sigma \qquad (0\leqq t\leqq T)$$

(recall [19] that $x+\lambda a*x=f$ and $f\in L^1(0, T; \mathbb{R})$ imply $x=f-\lambda r(\cdot,\lambda)*f$; here take $x=s$, $f\equiv 1$). Thus the unique solution $s(t,\lambda)$ of (s_λ) is absolutely continuous on $[0, T]$. We shall also use the assumption (see Proposition 1(ii) below):

(H_5) For every $\lambda\geqq 0$ the solution $s(t,\lambda)$ of (s_λ) satisfies $s(t,\lambda)\geqq 0$ on $[0, T]$.

Remark 1.1. It should be observed that assumption (H_4) implies that $a(t)\geqq 0$, so that $\int_0^t r(\sigma,\lambda)\,d\sigma\leqq\int_0^t a(\sigma)\,d\sigma$, $0\leqq t\leqq T$. If both (H_4) and (H_5) are satisfied on $[0, T]$ for every $T>0$, then $0\leqq\int_0^\infty r(t,\lambda)\,dt\leqq 1/\lambda$ $(\lambda>0)$; in particular, $r(t,\lambda)\in L^1(0,\infty)$, $\lambda>0$.

Our main result for the linear case is:

THEOREM 1. (i) *Let the assumptions* (H_1), (H_2), (H_4) *be satisfied. Then for every* $x\in X_A$, *the equation* (R) *has a unique strong solution which we denote by* $R(t)x$, $0\leqq t\leqq T$. *Moreover, for almost every* $t\in[0, T]$, *there exists a positive measure* μ_t *on* $\mathbb{R}^+$, *depending only on the kernel a, such that*

$$R(t)x = \int_0^\infty e^{-\omega A}x \, d\mu_t(\omega), \qquad a(t) = \int_0^\infty d\mu_t(\omega), \qquad t \in [0, T] \text{ a.e.} \tag{1.2}$$

and the following estimates are satisfied:

$$\|Rx\|_{L^1(0,T;Y)} \leqq \|a\|_{L^1(0,T;\mathbb{R})} \|x\|_Y, \tag{1.3}$$

where $Y = X$ or X_A or X_{A^2} and

$$\|R * v\|_{L^p(0,T;Y)} \leqq \|a\|_{L^1(0,T;\mathbb{R})} \|v\|_{L^p(0,T;Y)} \qquad (1 \leqq p \leqq \infty). \tag{1.4}$$

(ii) *In addition, let assumption* (H_5) *be satisfied. Then for every $x \in X_A$ the equation* (S) *has a unique strong solution which we denote by $S(t)x$, $0 \leqq t \leqq T$. Moreover, for every $t \in [0, T]$, there exists a probability measure ν_t on $\mathbb{R}^+$ depending only on the kernel a, such that*

$$S(t)x = \int_0^\infty e^{-\omega A}x \, d\nu_t(\omega) \qquad (t \in [0, T]), \tag{1.5}$$

and the following estimates hold:

$$\|S(t)x\|_Y \leqq \|x\|_Y, \tag{1.6}$$

$$\|S * v\|_{L^\infty(0,T;Y)} \leqq \|v\|_{L^1(0,T;Y)}, \tag{1.7}$$

where $Y = X$ or X_A or X_{A^2}.

Remark 1.2. If $a \equiv 1$, then $R(t) = S(t) = e^{-tA}$ and $\mu_t = \nu_t =$ the Dirac measure at t.

Assumptions (H_4) and (H_5) require some clarification.

PROPOSITION 1. (i) *Let* (H_2) *be satisfied and let $a \in C(0, T)$ and $a(t) > 0$. If* $\log a(t)$ *is convex on* $(0, T)$ *then* (H_4) *is satisfied on* $[0, T]$.

(*ii*) *Let* (H_2) *be satisfied and let $a(t)$ be nonnegative and nonincreasing on* $(0, T)$. *Then* (H_5) *is satisfied on* $[0, T]$.

While the content of Proposition 1 is implicitly contained in the literature (see [8], [9], [15] and [17]), we give the proof in Appendix A. In the literature the results are for t on the infinite interval and under slightly stronger assumptions.

Remark 1.3. If $a(t)$ satisfies (H_2) and is completely monotonic on $(0, T)$ (i.e. $(-1)^k a^{(k)}(t) \geqq 0 (0 < t < T;\ k = 0, 1, \cdots)$), then a satisfies (H_4) and (H_5); see [8], [17], [22].

Remark 1.4. We also note that, if $a(t) = e^t$, then (H_4) is satisfied but not (H_5). However, (H_5) does not imply (H_4). To see this, take

$$a(t) = \begin{cases} 1 & \text{if } 0 \leqq t \leqq 1, \\ 0 & \text{if } t > 1. \end{cases}$$

Then by Proposition 1 (ii), (H_5) is satisfied. But for $\lambda = 1$, as shown by Levin [15, example following Thm 1.4], $r(t, 1) < 0$ for some $1 < t < 2$.

Theorem 1 is used to deduce the following results about solutions of equations (1.1) and (1.1a).

THEOREM 2. (i) *Let the assumptions* (H_1), (H_2), (H_4) *and* $g \in L^1(0, T; X_A)$ *be satisfied. Then the equation*

$$u(t) + a * Au(t) = a * g(t) \qquad (0 \leqq t \leqq T) \tag{1.8}$$

has a unique strong solution u given by

$$u = R * g, \tag{1.9}$$

where R is the solution of equation (R) given by (1.2), *and* (*by* (1.3))

$$\|u\|_{L^1(0,T;X)} \leqq \|a\|_{L^1(0,T;\mathbb{R})} \|g\|_{L^1(0,T;X)}. \tag{1.10}$$

(ii) *Let the assumptions* (H_1), (H_2), (H_4), (H_5) *and*

(H_6) $f = f_1 + f_2$ *where* $f_1 \in L^p(0, T; X_{A^2})$ *for some* $1 \leqq p \leqq \infty$ *and* $f_2 \in W^{1,1}(0, T; X_A)$,

where $W^{1,1}$ *is the usual Sobolev space, be satisfied. Then equation* (1.1) *has a unique strong solution* $u = u_1 + u_2$ *where*

$$u_1(t) = f_1(t) - R * Af_1(t) \quad \text{a.e. on } [0, T], \tag{1.11}$$

and

$$u_2(t) = S(t)f_2(0) + S * f_2'(t), \quad t \in [0, T], \tag{1.12}$$

where S is the solution of equation (S) given by (1.5); *moreover, there is a constant* $c = c(T) > 0$, *depending only on T and a but not A, such that*

$$\|u\|_{L^1(0,T;X)} \leqq c\{\|f_1\|_{L^1(0,T;X_A)} + \|f_2\|_{W^{1,1}(0,T;X)}\}. \tag{1.13}$$

Remark 2.1. If A is any bounded linear operator, then $X = X_A = X_{A^2}$ and the existence and uniqueness of solutions of (1.1), with only $a \in L^1(0, T; \mathbb{R})$, $f \in L^1(0, T; X)$ is well-known. In the case when A is not bounded, existence and uniqueness results for solutions of (1.1) have been obtained by Friedman and Shinbrot [10], even for the case $A(t)$, where $A(t)$ generates an analytic semi-group, under different conditions both for the kernel and the function f with, however, different objectives than ours. Miller [18] has studied abstract Volterra integrodifferential equations.

Remark 2.2. Formula (1.11) is well-known when A is a bounded operator; formula (1.12) has also been employed in [9], [10], where S is called a fundamental solution of (1.1).

Remark 2.3. In the unbounded case we may define a weak solution of (1.1) as follows: there exist sequences $\{u_n\}$, $\{f_n\}$ where each $f_n \in L^1(0, T; X)$ and each u_n is a strong solution of (1.1) with $f = f_n$ such that $f_n \to f$ and $u_n \to u$ in $L^1(0, T; X)$. From (1.13) it follows that if $f \in L^1(0, T; X_A) + W^{1,1}(0, T; X)$, then equation (1.1) possesses a unique weak solution. (Note that $L^1(0, T; X_{A^2})$ is dense in $L^1(0, T; X_A)$ with respect to the norm in $L^1(0, T; X)$; similarly for $W^{1,1}$ $(0, T; X_A)$ in $W^{1,1}$ $(0, T; X)$). A similar remark applies to (1.8).

Remark 2.4. If $f_1 = 0$, then conclusion (1.13) can be strengthened to:

$$\|u\|_{L^\infty(0,T;X)} \leqq c\|f_2\|_{W^{1,1}(0,T;X)} \tag{1.14}$$

by using (1.6), (1.7), (1.12). Moreover, $S(t)x \in W^{1,1}(0, T; X)$ if $x \in X_A$ (for the proof see the end of § 2); thus, if $f_1 = 0$, it follows that

$$\|u\|_{W^{1,1}(0,T;X)} \leqq c\|f_2\|_{L^1(0,T;X_A)}.$$

Remark 2.5. Since the kernel is real, the case when X is a real Banach space can be treated as a special case of the complex case: If $\tilde{X} = X + iX$, the operator $\tilde{A}(x + iy) = Ax + iAy$ satisfies (H_1) whenever A satisfies (H_1). Therefore, we can restrict ourselves to the complex case.

Remark 2.6. If $a(t)=\delta(t)$ where $\delta(t)$ is the Dirac measure, then (1.1) reduces to $u(t)+Au(t)=f(t)$, and

$$S(t)=(I+A)^{-1}=\int_0^{\infty} e^{-\omega A}e^{-\omega}\,d\omega \qquad [26;\text{ p. } 240]. \tag{1.15}$$

The kernel $a(t)=\delta(t)$ does not satisfy (H_2). However, $\delta(t)$ can be approximated by kernels $a_\sigma(t)=(1/\sigma)e^{-t/\sigma}$ ($\sigma\to 0^+$); each a_σ satisfies (H_2), (H_4), (H_5) so that $a(t)=\delta(t)$ is a limiting case of our theory and the corresponding measures $\nu_t^{(\sigma)}$ approach the measures ν_t in (1.15) of density $e^{-\omega}$, independent of t, as $\sigma\to 0^+$.

Remark 2.7. Grimmer and Miller [12] obtain continuity results, as well as existence and uniqueness, in different spaces for the abstract Volterra equation

$$x(t)=f(t)+\int_0^t B(t-s)x(s)\,ds \qquad (t\geqq 0),$$

where the kernel B is operator valued and possibly unbounded, via semigroup theory by different methods and with different objectives.

By (1.2) and (1.5), $R(t)$ and $S(t)$ are respectively positive and convex "combinations" of contraction semigroups $e^{-\omega A}$. From this observation we obtain the following applications of Theorems 1 and 2 which we state as Theorems 3 and 4.

THEOREM 3. *Let* (H_1), (H_2), (H_4) *be satisfied. Let P be a closed convex cone in X, such that*

$$(1+\lambda A)^{-1}P\subseteq P \quad \textit{for every } \lambda\geqq 0. \tag{1.16}$$

Then

$$R(t)P\subseteq P \quad \textit{a.e. on } [0,T]. \tag{1.17}$$

Moreover, if in equation (1.8) *$g(t)\in P$ a.e., then the solution u of* (1.8) *lies in P a.e. on* $[0,T]$. *If in* (1.1) $f\in L^1(0,T;X_A)$ *and* $Af(t)\in P$ *a.e. on* $[0,T]$, *then*

$$f(t)\in u(t)+P \quad \textit{a.e. on } [0,T], \tag{1.18}$$

where u is the (weak) *solution of* (1.1); *in particular, if P is a positive cone in X* (i.e. P is a closed convex cone with $P\cap -P=\{0\}$; this induces the natural ordering: $x\leqq y \Leftrightarrow y-x\in P$), *the last statement is equivalent to the "maximum principle"*:

$$u(t)\leqq f(t) \quad \textit{a.e. on } [0,T]. \tag{1.19}$$

The proof of (1.17) in Theorem 3 is an immediate consequence of formula (1.2) for the operator R, together with the standard fact that assumption (1.16) implies that $e^{-\omega A}$ maps P into P for every $\omega\in\mathbb{R}^+$. With (1.17) established, the remaining conclusions of Theorem 3 follow from the representation formula (1.11).

Remark 3.1. If one studies equation (1.8) in the scalar case, one takes $A=\lambda\geqq 0$ to satisfy (H_1). If (H_2) is satisfied and if $P=\mathbb{R}^+$, then the condition (H_4) is necessary and sufficient in order to guarantee that the solution u of (1.8) satisfies $u(t)\geqq 0$ for every $g\geqq 0$. Thus one cannot hope to improve on condition (H_4) in the abstract case.

THEOREM 4. *Let* (H_1), (H_2), (H_4), (H_5) *be satisfied. Let P be a closed convex cone in X satisfying* (1.16). *Then*

$$S(t)P\subseteq P \quad \textit{for } 0\leqq t\leqq T. \tag{1.20}$$

(i) *Moreover, if $u_0\in P$ and if $g(t)\in P$ a.e. in equation* (1.1a), *then the solution u of* (1.1a) *lies in P for almost every $t\in[0,T]$.*

(ii) *If in* (1.1), $f \in W^{1,1}(0, T; X)$ *where* $f(0) \in P$ *and* $f'(t) \in P$ *a.e. on* $[0, T]$, *then the (weak) solution* u *of* (1.1) *lies in* P *for every* $t \in [0, T]$. (The last assertion holds for any closed convex set P in X).

(iii) *Moreover, if* X *is a real Hilbert space, and if the function* $\varphi: X \to [0, \infty]$ *is convex, lower semicontinuous, proper and satisfies*

$$\varphi(I+\lambda A)^{-1}x) \leqq \varphi(x) \quad \textit{for every } \lambda \geqq 0 \textit{ and every } x \in X, \tag{1.21}$$

then

$$\varphi(S(t)x) \leqq \varphi(x) \quad \textit{for every } t \in [0, T] \textit{ and every } x \in X. \tag{1.22}$$

The proof of (1.20) in Theorem 4 follows from formula (1.5) for the operator S, together with the observation that assumption (1.16) implies that $e^{-\omega A}$ maps P into P for every $\omega \in \mathbb{R}^+$. Then conclusion (i) of Theorem 4 follows from (1.9), (1.12) with $f(t) \equiv u_0$, and the fact that the operators R and S each map P into P. Similarly, conclusion (ii) follows from (1.12). To establish (iii) recall that assumption (1.21) implies that

$$\varphi_\lambda(e^{-\omega A}x) \leqq \varphi_\lambda(x) \quad \text{for every } \omega \geqq 0, \quad \lambda > 0, \quad x \in X,$$

where φ_λ is the Yosida approximation of φ [3, Prop. 2.11]. Then (1.22) follows from (1.5), Jensen's inequality and $\sup_{\lambda>0} \varphi_\lambda(x) = \varphi(x)$ [3, Prop. 2.11].

Remark 4.1. Conclusion (ii) of Theorem 4 is an abstraction of a result of Levin [15; Lemma 1.3] in $\mathbb{R}^+$. His result is

Let $a \in L^1_{\text{loc}}(0, \infty)$, $a(t)$ nonnegative nonincreasing on $(0, \infty)$. Let $f \in C[0, \infty]$ be nonnegative and nondecreasing on $[0, \infty)$. Then the solution x of the equation

$$x(t) + a * x(t) = f(t) \qquad (0 \leqq t < \infty)$$

satisfies $0 \leqq x(t) \leqq f(t)$.

This result is also an immediate consequence of Proposition 1(ii) and of the formula

$$x(t) = S(t)f(0) + \int_0^t S(t-\sigma)\, df(\sigma).$$

Levin's proof in [15] is different; he improves his result by a smoothing argument which permits him to remove the assumption $f \in C[0, \infty)$. This is also evident from the preceding formula. We also note that Levin's result can be obtained by approximating a and f by smooth functions and by applying a general result for functional differential equations due to Seifert [23, Cor. 1 and Thm. 3].

In Theorem 4(ii) both assumptions (H_4) and (H_5) are used. It is of interest to note that in the abstract case the assumption (H_5) (which is satisfied when a is positive and nonincreasing) is not sufficient to insure that S maps P into P when condition (1.16) is satisfied. To see this we consider the following example in $\mathbb{R}^2$.

Let

$$a(t) = \begin{cases} 1 & \text{if } 0 \leqq t < 1, \\ 0 & \text{if } t \geqq 1, \end{cases} \tag{1.23}$$

and consider for $\alpha > 0$ the operator A_α defined by

$$A_\alpha = U^T \Lambda_\alpha U, \quad \text{where } \Lambda_\alpha = \begin{pmatrix} \log 2 & 0 \\ 0 & \alpha + \log 2 \end{pmatrix}, \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \tag{1.24}$$

For every $\alpha > 0$, the real matrix A_α is symmetric and positive definite. Thus $-A_\alpha$

generates a contraction semigroup on $\mathbb{R}^2$, with the usual Euclidean norm. If P is the cone $\mathbb{R}_+^2=\{(x,y)\in\mathbb{R}^2: x\geqq 0, y\geqq 0\}$, then it is easily checked that $(I+\lambda A_\alpha)^{-1}P\subseteq P$ for every $\alpha>0$, $\lambda>0$, so that (1.16) is satisfied.

Corresponding to the kernel a defined by (1.23), the function $s(t,\lambda)$ of (H_5) is

$$s(t,\lambda)=\begin{cases} e^{-\lambda t} & \text{if } 0\leqq t<1,\\ e^{-\lambda t}+\lambda(t-1)e^{-\lambda(t-1)} & \text{if } 1\leqq t\leqq 2,\end{cases} \tag{1.25}$$

and clearly (H_5) is satisfied on the interval $0\leqq t\leqq 2$.

We next compute the operator S_α corresponding to A_α. Consider equation

$$u+a*A_\alpha u=x, \qquad x\in\mathbb{R}^2. \tag{1.26}$$

By setting $v=Uu$, $y=Ux$ equation (1.26) is transformed to the equivalent diagonal form

$$v+a*\Lambda_\alpha v=y, \tag{1.27}$$

which by the definition of $s(t,\lambda)$ in (H_5) gives the solution

$$v(t)=\begin{pmatrix} v_1(t)\\ v_2(t)\end{pmatrix}=\begin{pmatrix} s(t,\log 2)y_1\\ s(t,\alpha+\log 2)y_2\end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix} s(t,\log 2)[x_1+x_2]\\ s(t,\alpha+\log 2)[-x_1+x_2]\end{pmatrix}.$$

Thus the solution of (1.26) is

$$u(t)=\frac{1}{2}\begin{pmatrix} s(t,\log 2)[x_1+x_2]-s(t,\alpha+\log 2)[-x_1+x_2]\\ s(t,\log 2)[x_1+x_2]+s[t,\alpha+\log 2)[-x_1+x_2]\end{pmatrix}$$

and the operator $S_\alpha(t)$ is

$$S_\alpha(t)=\frac{1}{2}\begin{pmatrix} s(t,\log 2)+s(t,\alpha+\log 2) & s(t,\log 2)-s(t,\alpha+\log 2)\\ s(t,\log 2)-s(t,\alpha+\log 2) & s(t,\log 2)+s(t,\alpha+\log 2)\end{pmatrix}.$$

To show that (H_5) is not sufficient to prove that S_α maps $P=\mathbb{R}_2^+$ into P, it is sufficient to have $s(t,\log 2)-s(t,\alpha+\log 2)<0$ for some $t>0$ and for some $\alpha>0$. Observe that from (1.25)

$$\frac{\partial s}{\partial\lambda}(t,\lambda)=-e^{-\lambda(t-1)}[\lambda(t-1)^2-(t-1)+te^{-\lambda}] \tag{1.28}$$

for $1\leqq t\leqq 2$, $\lambda>0$. Thus $(\partial s/\partial\lambda)(2,\log 2)>0$, so that there exists $\alpha>0$ such that $s(2,\log 2)-s(2,\alpha+\log 2)<0$, which establishes the claim.

We note the above argument also shows that (H_5) does not imply that $s(t,\lambda)$ is completely monotonic in λ. (See remarks following Lemma 2.1 below.)

2. Proof of Theorems 1 and 2. We will prove Theorems 1 and 2 in two main steps. We first consider the case when A is a bounded operator. In this case, by Remarks 2.1 and 2.2, it suffices to prove the representation formulas (1.2) and (1.5); for, having these, one immediately has the estimates (1.3), (1.4), (1.6), (1.7) as well as the conclusions of Theorem 2. We then consider the case when A is an unbounded operator as a limiting situation of the bounded case using the Yosida approximation of A. The case where A is bounded is further divided into two parts:

(i) *Scalar case.* We require the following preliminary result.

LEMMA 2.1. *If $a(t)$ satisfies assumptions* (H_2), (H_4), *then $r(t,\lambda)$, defined in* (H_4), *is completely monotonic in λ for $0\leqq\lambda<\infty$ for $t\in[0,T]$ a.e. If, moreover, $a(t)$ satisfies* (H_5), *then $s(t,\lambda)$ is completely monotonic in λ for $0\leqq\lambda<\infty$ for every $t\in[0,T]$.*

Proof of Lemma 2.1. We consider the equations

$$r(t, \lambda)+\lambda a * r(t, \lambda)=a(t), \tag{2.1}$$

$$s(t, \lambda)+\lambda a * s(t, \lambda)=1 \tag{2.2}$$

of assumptions (H_4) and (H_5) respectively with λ complex rather than $\lambda \geqq 0$. Let E denote the space $L^1(0, T; \mathbb{C})$ or $C(0, T; \mathbb{C})$. Define the operator $K: E \to E$ by $Kx(t)= a * x(t)$ $(x \in E)$. K is a bounded linear operator with spectrum $\sigma(K)=\{0\}$. Thus for $u \in E$, the function v defined by $v(\lambda)=(I+\lambda K)^{-1}u$, $\lambda \in \mathbb{C}$, is an entire function of λ with values in E. By differentiation and induction one has the formula:

$$(-1)^n \frac{d^n}{d\lambda^n} v(\lambda)=n! K_\lambda^n v(\lambda), \qquad n=0, 1, 2, \cdots, \tag{2.3}$$

where the operator K_λ is defined by

$$K_\lambda = K(I+\lambda K)^{-1}. \tag{2.4}$$

We claim that

$$K_\lambda x(t)=\int_0^t r(t-s, \lambda)x(s)\, ds \qquad (x \in E). \tag{2.5}$$

To prove (2.5) take the convolution product of both sides of (2.1) by $x \in L^1(0, T; \mathbb{C})$, obtaining

$$r(t, \lambda) * x(t)+\lambda a * r(t, \lambda) * x(t)=a * x(t).$$

Thus $u_\lambda(t)=r(t, \lambda) * x(t)$ satisfies the equation

$$u_\lambda(t)+\lambda a * u_\lambda(t)=a * x(t);$$

by uniqueness of solutions of this scalar equation and by the definition of K_λ in (2.4) this shows that $u_\lambda(t)=K_\lambda x(t)$ and proves (2.5).

For $\lambda \geqq 0$, assumption (H_4) implies that the operators K_λ map the set of nonnegative real functions in E into itself. To prove the first assertion of Lemma 2.1, consider $v_a(\lambda)=(I+\lambda K)^{-1}a$; then $v_a(\lambda)(t)=r(t, \lambda)$ a.e. in $[0, T]$, $r(t, \lambda)\geqq 0$ by (H_4), and by (2.3) and (2.5), $(-1)^n(\partial^n/\partial\lambda^n)r(t, \lambda)\geqq 0$ a.e. in $[0, T]$ for $0<\lambda<\infty$. To prove the second assertion of Lemma 2.1, take $v_1(\lambda)=(I+\lambda A)^{-1}1$; then $v(\lambda)(t)=s(t, \lambda)\geqq 0$ by (H_5), and complete the proof as above. This completes the proof of Lemma 2.1.

It should be noted that the second assertion of Lemma 2.1 is stated by Friedman [9, Lemma 2.7] under only the hypothesis that $a \geqq 0$ and nonincreasing. However, his proof also uses (H_4). (He should also require (H_4) for his Theorem 5.2, p. 144.) To see that (H_5) is not sufficient for the complete monotonicity of $s(t, \lambda)$ with respect to λ, we consider again the kernel a defined in (1.23). The corresponding function $s(t, \lambda)$ is given by (1.25) and $s(t, \lambda)\geqq 0$, for $0 \leqq t \leqq 2$. However, as seen from (1.28),

$$\frac{ds}{d\lambda}(2, \log 2)>0.$$

We shall next obtain representations of the entire functions $r(t, \lambda), s(t, \lambda)$ for $\operatorname{Re} \lambda \geqq 0$.

By Lemma 2.1 and Bernstein's theorem [25, pp. 175–176], there exists a positive finite measure μ_t on $\mathbb{R}^+$ ((fixed hereafter) such that

$$
(2.6)\qquad \begin{aligned} r(t,\lambda) &= \int_0^\infty e^{-\omega\lambda}\, d\mu_t(\omega) \quad (\operatorname{Re}\lambda > 0;\ t\in[0,T]\text{a.e.}),\\ a(t) &= \int_0^\infty d\mu_t(\omega) \qquad (t\in[0,T]\text{a.e.}). \end{aligned}
$$

Similarly, using $s(0,\lambda)=1$, there exists a unique probability measure ν_t on $\mathbb{R}^+$ such that

$$
(2.7)\qquad s(t,\lambda) = \int_0^\infty e^{-\omega\lambda}\, d\nu_t(\omega) \qquad (\operatorname{Re}\lambda \geqq 0;\ t\in[0,T]).
$$

Thus (2.6) and (2.7) correspond to formulas (1.2) and (1.5) in the scalar case.

(ii) *A Bounded Operator Satisfying* (H_1). By a standard argument equations (R) and (S) possess for every $x\in X$ a unique solution which we denote by $R(t)x$ and $S(t)x$ respectively. We first prove the representation formulas (1.2) and (1.5) for the operators A_ε defined by

$$
(2.8)\qquad A_\varepsilon = \varepsilon I + A \qquad (1>\varepsilon>0).
$$

Define the operators R_ε and S_ε by the formulas

$$
(2.9)\qquad R_\varepsilon(t)x = \frac{1}{2i\pi}\int_{C_\varepsilon} r(t,\lambda)(\lambda I - A_\varepsilon)^{-1}x\, d\lambda \qquad (0\leqq t\leqq T),
$$

$$
(2.10)\qquad S_\varepsilon(t)x = \frac{1}{2i\pi}\int_{C_\varepsilon} s(t,\lambda)(\lambda I - A_\varepsilon)^{-1}x\, d\lambda \qquad (0\leqq t\leqq T),
$$

where $x\in X$, $r(t,\lambda)$, $s(t,\lambda)$ are defined by (2.1) and (2.2) respectively for $\lambda\in\mathbb{C}$. C_ε is a closed contour in the complex λ plane, oriented counterclockwise, consisting of a finite number of rectifiable Jordan arcs and such that $C_\varepsilon=\partial U_\varepsilon$, where U_ε is an open set containing the spectrum of A_ε. The integrals in (2.9), (2.10) are the usual Dunford integrals [26, p. 225]. It is shown by Friedman [9, Thm. 3.1] that $S_\varepsilon(t)x$ defined by (2.10) is the unique solution of equation (S) with A replaced by A_ε. An entirely analogous argument shows that $R_\varepsilon(t)x$ defined by (2.9) is the unique solution of equation (R) with A replaced by A_ε.

We next observe that the spectrum $\sigma(A_\varepsilon)$ is contained in the half plane $\operatorname{Re}\lambda\geqq\varepsilon$, and, if $\varepsilon<1$, in the ball of radius $1+\|A\|$. Thus we may choose C_ε to be the rectangle bounded by the segments joining the points $(\varepsilon/2-i(2+\|A\|))$, $((2+\|A\|)(i-i))$, $((2+\|A\|)(1+i))$, $(\varepsilon/2+i(2+\|A\|))$ oriented counterclockwise. Using the representation (2.6) in (2.9) under assumption (H_4) and the representation (2.7) in (2.10) under assumptions (H_4), (H_5) we obtain

$$
(2.11)\qquad R_\varepsilon(t)x = \int_0^\infty e^{-\omega A}x\, d\mu_t(\omega)(x\in X) \quad \text{a.e. on } [0,T],
$$

$$
(2.12)\qquad S_\varepsilon(t)x = \int_0^\infty e^{-\omega A}\varepsilon x\, d\nu_t(\omega) \qquad (x\in X).
$$

The proofs of (2.11), (2.12) follow from a theorem on the Dunford integral [26, p.

226], together with Fubini's theorem and the definition of the operator $e^{-\omega A_\varepsilon}$ by

$$e^{-\omega A_\varepsilon}x = \frac{1}{2\pi i}\int_{C_\varepsilon} e^{-\omega\lambda}(\lambda I - A_\varepsilon)^{-1}x\,d\lambda \qquad (x \in X).$$

Thus formulas (2.11), (2.12) establish (1.2) and (1.5) respectively with $A = A_\varepsilon$. We next let $\varepsilon \to 0^+$. We first show that

$$R_\varepsilon(t)x \to z(t) = \int_0^\infty e^{-\omega A}x\,d\mu_t(\omega) \quad \text{in } L^1(0, T; X). \tag{2.13}$$

We then show that $z(t)$ is the unique solution of equation (R). Substituting (2.8) in (2.11) we have

$$\left\|R_\varepsilon(t)x - \int_0^\infty e^{-\omega A}x\,d\mu_t(\omega)\right\| = \left\|\int_0^\infty (e^{-\varepsilon\omega} - 1)\,e^{-\omega A}\,d\mu_t(\omega)\right\|.$$

Therefore, by a simple application of Lebesgue's dominated convergence theorem

$$\lim_{\varepsilon\to 0^+}\left\|R_\varepsilon(t)x - \int_0^\infty e^{-\omega A}x\,d\mu(\omega)\right\| = 0 \quad \text{a.e. on } [0, T].$$

Moreover, since $e^{-\omega A}$ is a contraction semigroup, we have

$$\|R_\varepsilon(t)x\| \leqq \int_0^\infty \|e^{-\varepsilon\omega}e^{-\omega A}x\|\,d\mu_t(\omega) \leqq \|x\|a(t) \quad \text{a.e.} \tag{2.14}$$

Since $a \in L^1(0, T)$, another application of Lebesgue's theorem establishes (2.13).

We next show that the function z defined in (2.13) is the unique solution of equation (R). We know that $R_\varepsilon(t)x$ is the unique solution of the equation

$$u_\varepsilon(t) + a * Au_\varepsilon(t) + \varepsilon a * u_\varepsilon(t) = a(t) \quad \text{a.e.} \tag{R_ε}$$

Observe that by (2.14)

$$\|u_\varepsilon\|_{L^1(0,T;X)} \leqq \|x\|\int_0^T a(t)\,dt.$$

Consequently $\varepsilon a * u_\varepsilon \to 0$ in $L^1(0, T; X)$ as $\varepsilon \to 0^+$. Since $u_\varepsilon \to z$ in $L_1(0, T; X)$ as $\varepsilon \to 0^+$, one has that $z(t)$ satisfies equation (R) a.e. on $[0, T]$. By uniqueness, $z(t) = R(t)x$, establishing (1.2). An entirely similar argument with $L^1(0, T; X)$ replaced by $C(0, T; X)$ and assuming (H_5) establishes (1.5).

***A* an unbounded operator satisfying (H_2).** Define

$$J_\lambda = (I + \lambda A)^{-1} \qquad (\lambda \geqq 0)$$

and the Yosida approximation A_λ of A by

$$A_\lambda = \frac{1}{\lambda}(I - J_\lambda) \qquad (\lambda > 0);$$

recall that by (H_1) J_λ is a contraction on X for every $\lambda \geqq 0$ and that [26, Cor. 2, p. 241]

$$A_\lambda x = J_\lambda A_\lambda x = AJ_\lambda x \qquad (x \in X_A). \tag{2.15}$$

From linear semigroup theory we also need the fact that [26]

$$\lim_{\lambda\downarrow 0^+}\|e^{-\omega A_\lambda}x - e^{-\omega A}x\| = 0 \qquad (x \in X), \tag{2.16}$$

uniformly in ω on compact subsets of $\mathbb{R}^+$. Consider the equations

$$(R_\lambda) \qquad u(t)+a * A_\lambda u(t)=a(t)x \qquad (x \in X; 0 \leqq t \leqq T),$$

$$(S_\lambda) \qquad u(t)+a * A_\lambda u(t)=x \qquad (x \in X; 0 \leqq t \leqq T),$$

obtained from equations (R) and (S) respectively by replacing A by the bounded operator A_λ, $\lambda > 0$. By standard results the equations (R_λ) and (S_λ) have for each $\lambda > 0$ unique solutions which we denote by $u_\lambda(t)=R_\lambda(t)x$, $v_\lambda(t)=S_\lambda(t)x$ respectively and $R_\lambda(\cdot)x \in L^1(0, T; X)$, $S_\lambda(\cdot)x \in C([0, T]; X)$. Applying the result of the bounded case (ii) above (formulas (1.2), (1.5) with A replaced by the bounded operator A_λ) we have for $x \in X$

$$(2.17) \qquad R_\lambda(t)x = \int_0^\infty e^{-\omega A_\lambda} x \, d\mu_t(\omega) \qquad (\lambda > 0; t \in [0, T] \text{ a.e.}),$$

$$(2.18) \qquad S_\lambda(t)x = \int_0^\infty e^{-\omega A_\lambda} x \, d\nu_t(\omega) \qquad (\lambda > 0; t \in [0, T]).$$

Our first objective is to show that $R_\lambda(t)x$ and $S_\lambda(t)x$ converge to $R(t)x$ and $S(t)x$ of Theorem 1 as $\lambda \downarrow 0^+$, respectively. We carry out the proof in detail for $R(\cdot)x$, followed by the proofs of all conclusions of Theorems 1 and 2 pertaining to $R(\cdot)x$. We then sketch the proofs of the remaining conclusions pertaining to S.

We first consider $R_\lambda(t)x$. Let $\lambda > \mu > 0$; from (2.6), (2.17) we have

$$(2.19) \qquad \begin{aligned} \|R_\lambda(t)x - R_\mu(t)x\| &\leqq \int_0^\infty \|e^{-\omega A_\lambda}x - e^{-\omega A_\mu}x\| \, d\mu_t(\omega) \\ &\leqq 2\|x\| a(t) \quad (x \in X; t \in [0, T] \text{ a.e.}). \end{aligned}$$

Thus by assumption (H_2) $\|R_\lambda(t)x - R_\mu(t)x\|$ is bounded by a L^1 function on $[0, T]$. We next show that

$$(2.20) \qquad \lim_{\lambda,\mu \to 0^+} \|R_\lambda(t)x - R_\mu(t)x\| = 0 \qquad (x \in X; t \in [0, T] \text{ a.e.}).$$

By (2.19) it suffices to show that

$$(2.21) \qquad \lim_{\lambda,\mu \to 0^+} \int_0^\infty \|e^{-\omega A_\lambda}x - e^{-\omega A_\mu}x\| \, d\mu_t(\omega) = 0 \qquad (x \in X; t \in [0, T] \text{ a.e.}).$$

Since the measure $\mu_t(\omega)$ is finite for almost all $t \in [0, T]$, it suffices to show that $\|e^{-\omega A_\lambda}x - e^{-\omega A_\mu}x\|$ is bounded uniformly in $\omega \in \mathbb{R}^+$ and tends to zero a.e. in $\mathbb{R}^+$ as $\lambda, \mu \to 0^+$; this enables us to apply Lebesgue's theorem in $L^1(\mathbb{R}^+, \mu_t)$. But $\|e^{-\omega A_\lambda}x - e^{-\omega A_\mu}x\| \leqq 2\|x\|$ uniformly in $\omega \in \mathbb{R}^+$, and the second statement is immediate from (2.16).

Using (2.19) and (2.20), together with Lebesgue's theorem in $L^1(0, T; X)$, we have that $R_\lambda(\cdot)x$ is a Cauchy sequence in $L^1(0, T; X)$. For $x \in X$ define

$$R(t)x = \lim_{\lambda \downarrow 0^+} R_\lambda(t)x \quad \text{in } L^1(0, T; X),$$

and by the above argument we conclude that

$$(2.22) \qquad R(t)x = \int_0^\infty e^{-\omega A} x \, d\mu_t(\omega) \qquad (x \in X; t \in [0, T] \text{ a.e.}).$$

Our next objective is to show that $u(t)=R(t)x$ is a strong solution of equation (R) for $x\in X_A$. By the preceding argument we know that $u=R(\cdot)x\in L^1(0,T;X)$, and we must verify the remaining properties of a strong solution of equation (R) on $[0,T]$ (See Definition 1). Thus we consider (2.22) with $x\in X_A$ and show first that $u(t)=R(t)x\in D(A)$ a.e. on $[0,T]$ and that $AR(\cdot)x\in L^1(0,T;X)$. The mapping $\omega\to e^{-\omega A}x$ is continuous for $\omega\in\mathbb{R}^+$ with the values in X_A and is bounded in X_A. Since the measure μ_t is finite, $R(t)x\overset{\text{a.e.}}{=}\int_0^\infty e^{-\omega A}x\,d\mu_t(\omega)$ is the usual Bochner integral with values in $X_A=D(A)$. For the same reason and because A is a bounded operator in X_A, we have

$$A\int_0^\infty e^{-\omega A}x\,d\mu_t(\omega)=\int_0^\infty A\,e^{-\omega A}x\,d\mu_t(\omega).$$

Thus

$$\begin{aligned}AR(t)x\overset{\text{a.e.}}{=}A\int_0^\infty e^{-\omega A}x\,d\mu_t(\omega)&=\int_0^\infty A\,e^{-\omega A}x\,d\mu_t(\omega)\\&=\int_0^\infty e^{-\omega A}Ax\,d\mu_t(\omega)\overset{\text{a.e.}}{=}R(t)Ax,\end{aligned}$$

where we have used $Ae^{-\omega A}x=e^{-\omega A}Ax$ for $x\in X_A$, as well as (2.22) with x replaced by Ax. Since $Ax\in X$, $R(\cdot)Ax\in L^1(0,T;X)$ by the definition of $R(\cdot)x$. It remains to show that $u=R(\cdot)x$ satisfies equation (R). Since $u_\lambda=R_\lambda(\cdot)x$ satisfies equation (R_λ), and since $u_\lambda\to u=R(\cdot)x$ in $L^1(0,T;X)$ as $\lambda\downarrow 0^+$, it suffices, by assumption (H_2), to show that $A_\lambda u_\lambda\to Au$ in $L^1(0,T;X)$ as $\lambda\downarrow 0^+$. But since $A_\lambda R_\lambda(\cdot)x=R_\lambda(\cdot)A_\lambda x$ and $AR(\cdot)x=R(\cdot)Ax$, this is true if we show that $R_\lambda(\cdot)A_\lambda x\to R(\cdot)Ax$ in $L^1(0,T;X)$ as $\lambda\downarrow 0^+$; the latter follows from conclusion (2.23) of the following lemma with $w_\lambda=A_\lambda x$ and $w=Ax$. This completes the proof that $R(\cdot)x$ is a strong solution of equation (R).

LEMMA 2.2. *Let* (H_1), (H_2), (H_4) *be satisfied. Let* $R_\lambda(\cdot)x$ *be the unique solution of equation* (R_λ). *As above let* $R(\cdot)x=\lim_{\lambda\downarrow 0^+}R_\lambda(\cdot)x$ *in* $L^1(0,T;X)$. *Let* w_λ, $w\in X$ *and* $w_\lambda\to w$ *in* X *as* $\lambda\downarrow 0^+$; *let* z_λ, $z\in L^1(0,T;X)$ *and* $z_\lambda\to z$ *in* $L^1(0,T;X)$ *as* $\lambda\downarrow 0^+$. *Then*

$$\lim_{\lambda\downarrow 0^+}\|R_\lambda(\cdot)w_\lambda-R(\cdot)w\|_{L^1(0,T;X)}=0\quad\textit{and}\quad\|Rw\|_{L^1(0,T;X)}\leqq\|a\|_{L^1(0,T;\mathbb{R})}\|w\|_X,\tag{2.23}$$

$$\textit{for almost all } t\in[0,T],\ R(t-s)z(s)\in L^1(0,T;X),\tag{2.24}$$

$$\begin{gathered}\int_0^t R(t-s)z(s)\,ds=R*z(t)\in L^1(0,T;X)\quad\textit{and}\\ \|R*z\|_{L^1(0,T;X)}\leqq\|a\|_{L^1(0,T;\mathbb{R})}\|z\|_{L^1(0,T;X)},\end{gathered}\tag{2.25}$$

$$\lim_{\lambda\downarrow 0^+}\|R_\lambda*z-R*z\|_{L^1(0,T;X)}=0.\tag{2.26}$$

Proof. To prove the first statement in (2.23) we use Lebesgue's theorem in $L^1(0,T;X)$. From (2.17) with x replaced by w_λ we obtain the estimate

$$\|R_\lambda(t)w\|\leqq\|w_\lambda\|\int_0^\infty d\mu_t(\omega)\leqq Ka(t)\quad\text{on }[0,T]\quad\text{a.e.},$$

where $K>0$ is a constant independent of λ; thus by (H_2) $\|R_\lambda(\cdot)w_\lambda\|$ is bounded by a

L^1 function on $[0, T]$. To complete this part of the proof it suffices to show

$$\|R_\lambda(t)w_\lambda - R(t)w\| \to 0 \quad \text{a.e.} \quad \text{on } [0, T] \quad (\lambda \downarrow 0^+). \tag{2.27}$$

To do this we again use (2.17), (2.22), and Lebesgue's theorem in $L^1(\mathbb{R}^+; \mu_t)$. We have

$$\|e^{-\omega A_\lambda} w_\lambda\| \leqq \|w_\lambda\| \leqq K,$$

so that the finiteness of the measure μ_t implies that $\|e^{-\omega A_\lambda} w_\lambda\|$ is bounded by a $L^1(\mathbb{R}^+, \mu_t)$ function. Moreover, by (2.16)

$$\begin{aligned}\|e^{-\omega A_\lambda} w_\lambda - e^{-\omega A} w\| &\leqq \|e^{-\omega A_\lambda} w_\lambda - e^{-\omega A_\lambda} w\| + \|e^{-\omega A_\lambda} w - e^{-\omega A} w\| \\ &\leqq \|w_\lambda - w\| + o(1) \qquad (\lambda \downarrow 0^+).\end{aligned}$$

This together with Lebesgue's theorem, proves (2.27) and the first statement in (2.23). To prove the second statement in (2.23) we estimate using (2.22):

$$\begin{aligned}\|Rw\|_{L^1(0,T;X)} &= \int_0^T \left\| \int_0^\infty e^{-\omega A} w \, d\mu_t(\omega) \right\| dt \\ &\leqq \int_0^T \int_0^\infty \|e^{-\omega A} w\| d\mu_t(\omega) \leqq \int_0^T \int_0^\infty d\mu_t(\omega)\, dt \|w\| \\ &= \left(\int_0^T a(t)\, dt \right) \|w\|,\end{aligned}$$

which completes the proof.

To prove (2.24), (2.25) we observe that they both hold when R is replaced by R_λ; for, $R_\lambda(\cdot)x \in L^1(0, T; X)$, which implies that $R_\lambda(t-s)x$ is measurable in (t, s) for $0 \leqq s \leqq t \leqq T$ with values in X. In particular, for almost all $t \in [0, T]$, $R_\lambda(t-s)z(s)$ is measurable in s on $0 \leqq s \leqq t$. Letting $\lambda \downarrow 0^+$ and using the definition it follows that $R(t-s)z(s)$, $z \in L^1(0, T; X)$, is measurable in (t, s) on $0 \leqq s \leqq t \leqq T$ with values in X, as well as that $R(t-s)z(s)$ is measurable in s on $0 \leqq s \leqq t$, $t \in [0, T]$ a.e. The validity of (2.24) and of the first statement in (2.25) is implied by the second statement of (2.25). To prove the latter we estimate using (2.22) and the L^1 convolution inequality:

$$\begin{aligned}\int_0^T \left\| \int_0^t R(t-s)z(s)\, ds \right\| dt &\leqq \int_0^T \int_0^t \int_0^\infty \|e^{-\omega A} z(s)\|\, d\mu_{t-s}(\omega)\, ds\, dt \\ &\leqq \int_0^T \int_0^t \int_0^\infty \|z(s)\|\, d\mu_{t-s}(\omega)\, ds\, dt \\ &= \int_0^T \int_0^t \|z(s)\| a(t-s)\, ds\, dt \leqq \|a\|_{L^1(0,T;\mathbb{R})} \|z\|_{L^1(0T;X)},\end{aligned}$$

which is the desired result. We omit the proof of (2.26) since it is similar to the proof of the first assertion in (2.23). This completes the proof of Lemma 2.2.

We next establish the uniqueness of strong solutions of equation (R) as a particular case of uniqueness of strong solutions of (1.1).

LEMMA 2.3. *Let* (H_1), (H_2), (H_4) *be satisfied and let* $u \in L^1(0, T; X_A)$ *be a strong solution of the equation*

$$u + a * Au = 0.$$

Then $u = 0$.

Proof of Lemma 2.3. For any $\lambda > 0$ we have from the given equation and from (2.15) that

$$J_\lambda u + a * A_\lambda u = 0,$$

or equivalently

$$u + a * A_\lambda u = u - J_\lambda u.$$

By using the fact that A_λ is a bounded operator, together with the representation formula (1.11) where A is replaced by the bounded operator A_λ, f_1 is replaced by $u - J_\lambda u$, and R is replaced by R_λ, and the uniqueness of solutions of (1.1) in the bounded case, we obtain

$$u = u - J_\lambda u - R_\lambda * A_\lambda (u - J_\lambda u). \tag{2.28}$$

We wish to show that $u - J_\lambda u$ and $A_\lambda(u - J_\lambda u)$ each tend to zero as $\lambda \to 0^+$ in $L^1(0, T; X)$ for $u \in L^1(0, T; X_A)$. We have

$$\int_0^T \|u - J_\lambda u\|(t)\, dt = \int_0^T \lambda \|A_\lambda u\|(t)\, dt \leqq \lambda \int_0^T \|Au\|(t)\, dt,$$

which tends to zero as $\lambda \to 0^+$, where (2.15) is used to estimate $\|A_\lambda u\| \leqq \|Au\|$, $u \in L^1(0, T; X_A)$. Also

$$\|A_\lambda (u - J_\lambda u)\|(t) = \|A_\lambda \lambda A_\lambda u\|(t) = \|\lambda A_\lambda J_\lambda v\|(t),$$

where $v = Au$; thus by the contraction property of J_λ

$$\|A_\lambda (u - J_\lambda u)\|(t) = \|J_\lambda v - J_\lambda (J_\lambda v)\|(t) \leqq \|v - J_\lambda v\|(t).$$

But

$$\|v - J_\lambda v\|(t) \leqq 2\|v\|(t) = 2\|Au\|(t) \in L^1(0, T);$$

moreover,

$$\|v - J_\lambda v\|(t) \to 0 \quad \text{a.e.} \quad \text{on } [0, \mathrm{T}],$$

and therefore, by Lebesgue's theorem, $A_\lambda(u - J_\lambda u) \to 0$ as $\lambda \to 0^+$ in $L^1(0, T; X)$ for $u \in L^1(0, T; X_A)$. Letting $\lambda \to 0^+$ in (2.28) and using the above facts together with (2.26) of Lemma 2.2, we obtain $u = 0$. This completes the proof of Lemma 2.3.

Up to this point we have established Theorem 1(i) with the estimate (1.3) for $Y = X$ (done in proving (2.23)), and the estimate (1.4) for $Y = X$, $p = 1$ (done in proving (2.25)). To prove (1.3) with $Y = X_A$ we use $RAx = ARx$ $(x \in X_A)$ (see paragraph following (2.22)), so that by definition of norm in X_A and by two applications of (1.3) with $Y = X$, we have

$$\begin{aligned}\|Rx\|_{L^1(0,T;X_A)} &= \|Rx\|_{L^1(0,T;X)} + \|ARx\|_{L^1(0,T;X)}\\ &= \|Rx\|_{L^1(0,T;X)} + \|RAx\|_{L^1(0,T;X)} \leqq \|a\|_{L^1(0,T;\mathbb{R})}[\|x\| + \|Ax\|]\\ &= \|a\|_{L^1(0,T;\mathbb{R})}\|x\|_{X_A},\end{aligned}$$

which is the desired result. The proofs of (1.3) with $Y = X_{A^2}$ and of (1.4) for $p = 1$ and $Y = X_A$ or X_{A^2} are similar and we omit them.

For (1.4) we indicate the proof of the case $Y = X$ and $1 < p < \infty$ (the case $p = \infty$ is immediate). Let q be the conjugate exponent of p; let $h \in L^q(0, T; \mathbb{R})$, and consider the following calculation:

$$\int_0^T |h(t)| \left\| \int_0^t R(t-s)z(s)\,ds \right\| dt$$
$$\leqq \int_0^T |h(t)| \int_0^t \|R(s)z(t-s)\|\,ds\,dt$$
$$= \int_0^T |h(t)| \int_0^t \left\| \int_0^\infty e^{-\omega A} z(t-s)\,d\mu_s(\omega) \right\| ds\,dt$$
$$\leqq \int_0^T |h(t)| \int_0^t a(s)\|z(t-s)\|\,ds\,dt \leqq \int_0^T a(s) \int_0^T |h(t)| \|z(t-s)\|\,dt\,ds$$
$$\leqq \int_0^T a(s) \left(\int_0^T |h(t)|^q\,dt \right)^{1/q} \left(\int_0^T \|z(t-s)\|^p\,dt \right)^{1/p} ds$$
$$\leqq \left(\int_0^T a(s)\,ds \right) \|h\|_{L^q(0,T;\mathbb{R})} \|z\|_{L^p(0,T;X)} < \infty$$

where we used (2.22), (H_1), Fubini's theorem, and Hölder's inequality. By a standard argument (see [14, p. 398]), this implies the result. The proof for the case $Y = X_A$ or X_{A^2} is similar. This completes the proof of Theorem 1(i).

We next prove Theorem 2(i). We first show that $u = R * g \in L^1(0, T; X_A)$ we use $ARx = RAx$ $(x \in X_A)$ proved in Theorem 1 to conclude that $Au = AR * g = RA * g = R * Ag \in L^1(0, T; X)$. To show that $u = R * g$, $g \in L^1(0, T; X_A)$, satisfies (1.8) we substitute and use equation (R):

$$R * g + a * AR * g = (R + a * AR) * g = a * g.$$

The uniqueness of strong solutions of (1.8) on $[0, T]$ follows from Lemma 2.3. The estimate (1.10) is a particular case of (1.4). This completes the proof of Theorem 2(i).

We next prove assertion (1.11) of Theorem 2(ii). We define u_1 by (1.11) and we show that u_1 is a strong solution of (1.11) on $[0, T]$ with $f = f_1$. Since $f_1 \in L^p(0, T; X_{A^2})$, $Af_1 \in L^p(0, T; X_A)$, and from (1.4), $R * Af_1 \in L^p(0, T; X_A)$. Thus $u_1 \in L^p(0, T; X_A)$ (hence in $L^1(0, T: X_A)$). To see that u_1 satisfies (1.1), substitute and use equation (R):

$$u_1 + a * Au_1(t) = f_1(t) - R * Af_1(t) + a * A(f_1(t) - R * Af_1(t))$$
$$= f_1(t) + [aI - R - a * AR] * Af_1(t) = f_1(t), \qquad t \in [0, T] \quad \text{a.e.},$$

where I is the identity operator in X. The uniqueness of strong solutions of (1.1) with $f = f_1$ follows from Lemma 2.3, and the estimate (1.3) with $f_2 = 0$ follows from (1.4).

We next sketch the proof Theorem 1(ii). Returning to the approximating equation (S_λ) and its unique solution $v_\lambda = S_\lambda(\cdot)x$, one shows as before but for every $t \in [0, T]$, that $S_\lambda(t)x$ is a Cauchy sequence in X for $x \in X$. We define

$$S(t)x = \lim_{\lambda \downarrow 0^+} S_\lambda(t)x \qquad (t \in [0, T]; x \in X).$$

From (2.18), $\|S_\lambda(t)x\| \leqq \|x\|$ $(0 \leqq t \leqq T;\ \lambda > 0)$, and therefore, in particular $S(\cdot)x \in L^1(0, T: X)$. As was done for $R(\cdot)x$, one then proves (1.5),

$$AS(t)x = S(t)Ax \qquad (x \in X_A, 0 \leqq t \leqq T), \tag{2.29}$$

and one shows that $S(\cdot)x$ is a strong solution of equation (S) on $[0, T]$ for $x \in X_A$.

To complete the proof of Theorem 1(ii) it remains to prove (1.6), (1.7). The estimate (1.6) is immediate from (1.5) when $Y = X$; to prove it for $Y = X_A$ or X_{A^2} one uses (2.29). By using (2.29) it is sufficient to prove the estimate (1.7) for $Y = X$. To do this we estimate using (1.5), (H_1):

$$\begin{aligned}\|S * v\|_{L^\infty(0,T;X)} &= \operatorname*{ess\,sup}_{0\leqq t\leqq T}\left\|\int_0^t S(t-\tau)v(t)\,dt\right\| = \operatorname*{ess\,sup}_{0\leqq t\leqq T}\left\|\int_0^t\int_0^\infty e^{-\omega A}v(t)\,d\nu_{t-\tau}\,d\tau\right\| \\ &\leqq \operatorname*{ess\,sup}_{0\leqq t\leqq T}\int_0^t\int_0^\infty \|e^{-\omega A}v(t)\|\,d\nu_{t-\tau}\,d\tau \\ &\leqq \int_0^T \|v(\tau)\|\,d\tau = \|v\|_{L^1(0,T;X)}.\end{aligned}$$

This completes the proof of Theorem 1(ii).

To prove assertion (1.12) of Theorem 2(ii), we define u_2 by (1.12) and we show that u_2 is a strong solution of (1.1) on $[0, T]$ with $f = f_2$. To show that $u_2 \in L^1(0, T; X_A)$ is immediate from (2.29), (1.6), (1.7) and $f_2 \in W^{1,1}(0, T; X_A)$. To see that u_2 satisfies (1.1) substitute and use equation (S):

$$\begin{aligned}u_2(t)+a * Au_2(t) &= S(t)f_2(0)+S * f_2'(t)+a * A(S(t)f_2(0)+S * f_2'(t)) \\ &= (S(t)+a * AS(t))f_2(0)+(S+a * AS) * f_2'(t) \\ &= f_2(0)+I * f_2'(t) = f_2(0)+\int_0^t f_2'(\tau)\,d\tau = f_2(t), \qquad t\in[0, T].\end{aligned}$$

The uniqueness of strong solutions of (1.1) with $f = f_2$ follows from Lemma 2.3, and the estimate (1.13) with $f_1 = 0$ follows from (1.6), (1.7), (1.12). This completes the proof of Theorem 2(ii).

Finally, we prove that $S(t)x \in W^{1,1}(0, T; X)$ if $x \in X_A$. To see this we start from the relation

$$S_\lambda(t)x = x - A_\lambda\int_0^t R_\lambda(\tau)x\,d\tau = x - \int_0^t R_\lambda(\tau)A_\lambda x\,d\tau, \tag{2.30}$$

for $x \in X$ and $0 \leqq t \leqq T$, where $S_\lambda(\cdot)x$, $R_\lambda(\cdot)x$ are the unique solutions of the approximating equations (S_λ), (R_λ) respectively; (2.30) is proved by direct substitution into equation (S_λ), and by using equation (R_λ) with x replaced by $A_\lambda x$, as well as $A_\lambda R_\lambda(\cdot)x = R_\lambda(\cdot)A_\lambda x$, and Fubini's theorem. Letting $\lambda \downarrow 0$ in (2.30), using Lemma 2.2 and the definitions of $S(\cdot)x$, $R(\cdot)x$ we obtain

$$S(t)x = x - \int_0^t R(\tau)Ax\,d\tau, \qquad x\in X_A, \quad t\in[0, T]. \tag{2.31}$$

Since $R(\cdot)Ax \in L^1(0, T; X)$, the conclusion $S(t)x \in W^{1,1}(0, T; X)$, $x \in X_A$, is immediate from (2.31).

3. *A* a nonlinear operator. In this section we give a nonlinear analogue of Theorems 3 and 4. Let X be a real Banach space and let $P \subseteq X$ be a closed convex cone. Let $A\colon D(A) \subseteq X \to 2^X$ be a given, possibly multivalued, m-accretive operator [6, p. 139] satisfying the condition

$$(I+\lambda A)^{-1}P \subseteq P \qquad (\lambda > 0). \tag{3.1}$$

Let a satisfy (H_2) and (H_4) and let f satisfy (H_3). Consider the equation

$$u(t)+a * Au(t) \ni f(t), \qquad t \in [0, T], \tag{3.2}$$

where $T>0$. We say that $u \in L^1(0, T; X)$ is a solution of (3.2) on $[0, T]$ if there exists $w \in L^1(0, T; X)$, where $w(t) \in Au(t)$ a.e., such that $u(t)+a * w(t)=f(t)$ a.e. for $t \in [0, T]$.

THEOREM 5. *Let* (H_2), (H_3), (H_4) *be satisfied. Then for every* $\lambda>0$ *the approximating equation*

$$u_\lambda(t)+a * A_\lambda u_\lambda(t)=f(t), \qquad t \in [0, T] \quad a.e., \tag{3.3}$$

where A_λ *is the Yosida approximation of* A, *has a unique strong solution on* $[0, T]$. *In addition, let* f *be such that*

(H_7) *for every* $\lambda>0$, *the unique solution* v *of the linear equation*

$$v(t)+\lambda a * v(t)=f(t), \qquad t \in [0, T] \quad a.e. \tag{3.4}$$

satisfies $v(t) \in P$ *a.e. on* $[0, T]$.

Then if (3.1) *is satisfied* $u_\lambda(t) \in P$ *a.e. on* $[0, T]$. *Consequently, if* u *is a solution of equation* (3.2) *such that* $u = \text{weak} \lim_{\lambda\to 0} u_\lambda$ *in* $L^1(0, T; X)$, *then* $u(t) \in P$ *a.e. on* $[0, T]$.

Remark 5.1. Under the assumptions of Theorem 5 it follows from Theorems 3 and 4 with $A=\lambda I$ that if $f(t)=a * g(t)$, $g \in L^1(0, T; X)$, then (H_7) is satisfied if $g(t) \in P$ a.e. on $[0, T]$. If $f(t)=u_0+a * g(t)$, where $u_0 \in P$ and g is as above, then (H_7) is satisfied provided that (H_5) holds. If $f \in W^{1,1}(0, T; X)$, then (H_7) is satisfied provided that (H_5) holds and that $f(0) \in P$ and $f'(t) \in P$ a.e. on $[0, T]$.

Remark 5.2. If A is linear and satisfies (H_1), equation (3.2) is (1.1); it was shown in § 2 that the unique solution u_λ of (3.4) converges to u, the unique solution of (1.1), under the assumptions of Theorem 2.

Remark 5.3. Crandall and Nohel have recently shown [7, Thm. 4] that equation (3.2) has a unique solution u on $[0, T]$, and u is the limit of solutions u_λ of the approximating equation (3.3) as $\lambda \downarrow 0$ under the conditions that A is m-accretive, a is absolutely continuous on $[0, T]$, $a' \in BV[0, T]$, $a[0]>0$, $f \in W^{1,1}(0, T; X)$, and $f(0) \in \overline{D(A)}$. If in addition $f' \in BV([0, T]; X)$ and $f(0) \in D(A)$, then u is Lipschitz continuous; and if X is reflexive u is a strong solution on $[0, T]$. Thus Theorem 5 is applicable if, e.g., a is positive decreasing, $\log a$ is convex on $[0, T]$, and $a(0)>0$. The results of [7] generalize and simplify considerably the known existence and uniqueness theory for equation (3.2) when $X=H$ is a real Hilbert space and A is a subdifferential of a convex function on H obtained by Barbu [1, Thm. 1] and Londen [16], and for a maximal monotone A on H as obtained by Gripenberg [13]. It should be noted that in [1, Thm. 1], [16], [13], [7] the assumption $0<a(0)<\infty$ is needed, while in [1, Thm. 3] as well as in Theorem 5 above $a(0^+)=\infty$ is permitted.

Remark 5.4. If one is interested in applying Theorem 5 to the limiting equation (3.2), rather than the approximating equation (3.3), it is clear from the proof of Theorem 5 that it suffices to require conditions (3.1) and (H_7) to hold only for small $\lambda>0$.

Proof of Theorem 5. Consider the equation (3.3) written in the equivalent form

$$u_\lambda+\frac{1}{\lambda} a * u_\lambda=f+\frac{1}{\lambda} a * J_\lambda u_\lambda. \tag{3.5}$$

Define $f_\lambda \in L^1(0, T; X)$ to be the unique solution of (3.4) with λ replaced by $1/\lambda$. By

(H_7) $f_\lambda(t)\in P$ a.e. on $[0, T]$. It is easily checked by use of

$$r\left(t,\frac{1}{\lambda}\right)+\frac{1}{\lambda}\int_0^t a(t-\sigma)r\left(\sigma,\frac{1}{\lambda}\right)d\sigma=a(t) \quad\text{and}\quad f_\lambda(t)=f(t)-\frac{1}{\lambda}\int_0^t r\left(t-\sigma,\frac{1}{\lambda}\right)f(\sigma)\,d\sigma$$

that equation (3.5) is equivalent to the equation

$$v_\lambda=F_\lambda(v_\lambda),\qquad v_\lambda=u_\lambda-f_\lambda, \tag{3.6}$$

where

$$F_\lambda(z)(t)=\frac{1}{\lambda}\int_0^t r\left(t-\sigma,\frac{1}{\lambda}\right)J_\lambda(z+f_\lambda)(\sigma)\,d\sigma. \tag{3.7}$$

Observe that F_λ maps $L^1(0, T; X)$ into itself. We prove that some iterate of F_λ is a strict contraction in $L^1(0, T; X)$. Indeed, from (3.7), (H_2) and the contraction property of J_λ (recall A is m-accretive) one has

$$\|F_\lambda(u)(t)-F_\lambda(v)(t)\|\leqq\frac{1}{\lambda}\int_0^t\|u(s)-v(s)\|\,ds. \tag{3.8}$$

Define $b_\lambda(t)=(1/\lambda)r(t, 1/\lambda)$ and $b_\lambda^n(t)=b_\lambda * b_\lambda * \cdots * b_\lambda(t)$, where the convolution is taken n times. Iterating (3.8) n times, we obtain

$$\|F_\lambda^n(u)(t)-F_\lambda^n(v)(t)\|\leqq b_\lambda^n * \|u-v\|(t). \tag{3.9}$$

For any fixed λ choose n_λ so large that $\int_0^t b_\lambda^{n_\lambda}(\sigma)\,d\sigma=K_\lambda<1$; then integrating (3.9) we obtain

$$\|F_\lambda^{n_\lambda}(u)-F_\lambda^{n_\lambda}(v)\|_{L^1(0,T;X)}\leqq K_\lambda\|u-v\|_{L^1(0,T;X)}. \tag{3.10}$$

Thus (3.6) has a unique solution $v_\lambda\in L^1(0, T; X)$ given by

$$v_\lambda=\lim_{n\to\infty}F_\lambda^n(u_0),\quad\text{for any } u_0\in L^1(0, T; X),$$

and (3.3) has the unique solution $u_\lambda=v_\lambda+f_\lambda$. In particular if $u_0(t)\in P$ a.e. on $[0, T]$ and if assumptions (H_4) and (H_7) are satisfied, then by (3.1) and (3.7) $F_\lambda(u_0)(t)\in P$ a.e. on $[0, T]$ and the same holds for $F_\lambda^n(u_0)(t)$ for every n. Consequently the unique solution of (3.3) $u_\lambda(t)\in P$ a.e. on $[0, T]$. This completes the proof of Theorem 5.

Remark 5.5. From the proof of Theorem 5 it is clear that Theorem 5 provides an alternative, and in fact simpler, treatment of Theorems 3 and 4 in the linear case. However, in the linear case Theorems 1 and 2 provide explicit representations for the operators R and S and hence more information about the solution. Moreover, the method of proof of Theorem 5 can be used to analyze more general situations. For example, let X be the product of n Banach spaces $X_1, X_2, \cdots, X_n$, and interpret equation (3.2) as a system of n equations with $u(t)$, $f(t)\in X$ for $t\in[0, T]$ and the kernel a being an $n\times n$ matrix satisfying (H_2) componentwise, and such that the associated matrix resolvent $r(t,\lambda)\geqq 0$ componentwise (analogue of (H_4)). Let P be a closed convex cone in X and let A be an m-accretive operator on X for a suitable norm satisfying (3.1). If f satisfies (H_3) and (H_7), then the conclusions of Theorem 5 hold.

Remark 5.6. The proof of Theorem 5 is in the same spirit as the proof of Theorem 1 of Weis [24] for the equation

$$x(t)=f(t)+\int_0^t a(t-s)g(s, x(s))\,ds$$

where x, f, g have values in $\mathbb{R}^n$ and a is an $n \times n$ matrix $\in L^1_{\text{loc}}(0, \infty)$ and where g has "separated structure" in the sense that $g(t, x) = \text{col}(g_i(t, x_i))$, $i = 1, \cdots, n$, where each g_i is locally Lipschitz with respect to x_i uniformly for t bounded. Weis gives a condition which corresponds to (H_4) and (H_7) which insures that the solution $x(t) \geqq 0$ for as long as it exists.

4. Examples.

Example 1. This example is an application of Theorem 5. Consider the equation

$$u(t, x) + a * (-\nabla^2 u(t, x) + \beta(u(t, x))) \ni f(t, x), \tag{4.1}$$

$0 < t < \infty$, $x \in \Omega$, a bounded open set in $\mathbb{R}^n$ with smooth boundary Γ with u satisfying Dirichlet boundary conditions on Γ. β is a maximal monotone graph on $\mathbb{R} \times \mathbb{R}$ with $0 \in \beta(0)$. For simplicity we assume that the kernel a is completely monotonic on $[0, \infty)$; thus (see Remark 1.3) assumptions (H_2), (H_4), (H_5) are satisfied on $[0, T]$ for every $T > 0$. We assume $f \in W^{1,2}_{\text{loc}}(0, \infty; X)$, $X = L^2(\Omega)$. To see that equation (4.1) is a particular case of (3.2) define

$$Au = -\nabla^2 u + \beta(u); \qquad D(A) = \{u \in W^{2,2}(\Omega) \cap W^{1,2}_0(\Omega) \colon \beta(u) \in L^2(\Omega)\}. \tag{4.2}$$

As is known (see Brézis [4, Cor. 13]) where $\beta(u)$ in (4.2) is $u + \gamma(u)$ in [4] and $\beta(u)$ in [4] is ϕ if $u \neq 0$ and $\mathbb{R}$ if $u = 0$, A is the subdifferential of the convex, lower semicontinuous (l.s.c.) proper function $\varphi\colon L^2(\Omega) \to (-\infty, +\infty]$ defined by

$$\varphi(u) = \begin{cases} \frac{1}{2}\int_\Omega |\operatorname{grad} u|^2\, dx + \int_\Omega j(u)\, dx & \text{if } u \in W^{1,2}_0(\Omega), \quad j(u) \in L^1(\Omega), \\ +\infty & \text{otherwise,} \end{cases}$$

where j is the unique, convex, l.s.c., proper function mapping $\mathbb{R}$ into $(-\infty, +\infty]$ such that $j(0) = 0$ and $\beta = \partial j$. Thus A is maximal monotone on the Hilbert space $L^2(\Omega)$ and hence A is m-accretive. Thus (4.1) with the boundary condition $u = 0$ on Γ is a particular case of (3.2). Let $f \in W^{1,2}_{\text{loc}}(0, \infty; X)$; in particular, $f \in C([0, \infty); X)$ and $f(0)$ is well defined as an element of $L^2(\Omega)$. We assume that $f(0) \in W^{1,2}_0(\Omega)$ and $\int_\Omega j(f(0))\, dx < \infty$. These assumptions on f imply that (H_3), (H_6) are satisfied. It is now easily checked that all the assumptions Londen [16, Thm. 1] or Barbu [1, Thm. 1] are satisfied and therefore, (4.1) possesses a unique solution u on $[0, T]$ for every $T > 0$ in the sense of the definition given following equation (3.2) above. Moreover, $u = \lim_{\lambda \to 0^+} u_\lambda$ in $L^1(0, T; X)$ (even in $L^2(0, T; X)$) for every $T > 0$, where u_λ is the unique solution of the approximating equation (3.3). We shall apply Theorem 5 with $P = L^2_+(\Omega)$. It is well known that the operator A defined by (4.2) satisfies condition (3.1). Therefore, if we require that condition (H_7) is satisfied—this will be the case. For example, if $f(0) \in P$ and $f'(t) \in P$ a.e. on $[0, \infty)$ (see Remark 5.1), then the solution $u(t)$ of (4.1) is nonnegative a.e. on $(0, \infty)$.

Example 2. This example is an application of Theorem 4(iii). Let Ω be a bounded open set in $\mathbb{R}^n$ with smooth boundary Γ. On Ω we consider the linear second order differential operator

$$Au = -\sum_{i,j=1}^n \frac{\partial}{\partial x_j}\left(a_{ij}\frac{\partial u}{\partial x_i}\right) + \sum_{i,j=1}^n \frac{\partial}{\partial x_i}(a_i u) + Cu$$

where $a_{ij}, a_i \in C^1(\bar{\Omega})$, $C \in L^\infty(\Omega)$,

$$C \geqq 0, \qquad C + \sum_i \frac{\partial a_i}{\partial x_i} \geqq 0 \quad \text{a.e.},$$

and for some positive constant α

$$\sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j \geqq \alpha|\xi|^2 \quad \text{a.e.,} \qquad \xi \in \mathbb{R}^n.$$

We define $D(A) = W^{2,2}(\Omega) \cap W_0^{1,2}(\Omega)$. It is known (see [5]) that A satisfies (H_1) with $X = L^2(\Omega)$. Consider the equation

$$u(t) + a * \cdot Au(t) = u_0, \qquad t \in [0, T], \tag{4.3}$$

where $u_0 \in L^2(\Omega)$ and where a satisfies assumptions (H_2), (H_4), (H_5) on $[0, T]$. Equation (4.3) has a unique weak solution u (see Remark 2.3); moreover, if $u_0 \in D(A)$, then the solution u is strong. Let j be a convex, l.s.c., proper function: $\mathbb{R} \to [0, \infty]$ with $0 \in \partial j(0)$, and we fix $j(0) = 0$. Define $\varphi: X \to [0, \infty]$ by

$$(v) = \begin{cases} \int_\Omega j(v)\, dx & \text{if } j(v) \in L^1(\Omega), \\ +\infty & \text{otherwise.} \end{cases}$$

Then by [5, Lemma 2] we have $(A_\lambda x, y) \geqq 0$ for every $[x, y] \in \partial\varphi$ and for all $\lambda > 0$. Moreover, by [3, Thm. 4.4], (1.21) is satisfied. Consequently, by Theorem 4(iii), if $j(u_0) \in L^1(\Omega)$, one has

$$\int_\Omega j(u(t))(x)\, dx \leqq \int_\Omega j(u_0)(x)\, dx, \qquad t \in [0, T].$$

In particular, if $j(u) = |u|^p$, $1 \leqq p < \infty$, one obtains

$$\|u(t)\|_{L^p(\Omega)} \leqq \|u_0\|_{L^p(\Omega)} \tag{4.4}$$

if $u_0 \in L^p(\Omega)$. Note that the case $p = \infty$ can be obtained by passing to the limit. Inequality (4.4) can be obtained directly from Theorem 1, inequality (1.6), if one uses the known fact that A satisfies (H_1) with $X = L^p(\Omega)$, $1 \leqq p < \infty$; see [5, Thm. 8 and remarks preceding].

Example 3. This example is an application of the linear theory developed in Theorems 1–4 to a nonlinear problem. Let Ω be a bounded open set in $\mathbb{R}^n$ with smooth boundary Γ. Let $\gamma: \mathbb{R} \to \mathbb{R}$, $\gamma(0) = 0$, γ continuous and nondecreasing. Assume that the nonlinear elliptic equation

$$-\nabla^2 u = \gamma(u), \qquad u|_\Gamma = 0 \tag{4.5}$$

has a nontrivial, positive solution $u_\infty \in L^\infty(\Omega)$. Let a satisfy (H_2), (H_4), (H_5) for every $T > 0$, and consider the nonlinear integral equation

$$\begin{aligned} &u(t) + a * (-\nabla^2 u - \gamma(u))(t) = u_0 \qquad (0 < t < \infty), \\ &u_0 \in L^\infty(\Omega), \qquad u = 0 \quad \text{on } \Gamma. \end{aligned} \tag{4.6}$$

Let $Au = -\nabla^2 u$ with $D(A) = \{u \in W_0^{1,2}(\Omega) \cap W^{2,2}(\Omega)\}$. Let $X = L^2(\Omega)$. Then A satisfies (H_1). If u is a solution of (4.6) in the sense that $g = \gamma(u) \in L^\infty(0, \infty; X)$ and u is a weak solution (in the sense of Remark 2.3) of the equation

$$u(t) + a * Au(t) = u_0 + a * g(t), \qquad t \in [0, T] \quad \text{a.e.,} \quad \forall T > 0,$$

then by Theorem 2 it is easily shown that u satisfies the nonlinear functional equation

$$u(t) = F_{u_0}(u)(t) \qquad (0 \leqq t < \infty), \tag{4.7}$$

where

$$F_{u_0}(u)(t)=S(t)u_0+R*\gamma(u)(t). \tag{4.8}$$

We prove the following result about solutions of (4.7), (4.8).

THEOREM 6. *Let* (H_2), (H_4), (H_5) *be satisfied for every* $T>0$. *For every* $u_0 \in X$ *satisfying* $0 \leqq u_0 \leqq u_\infty$, *the equation* (4.7), (4.8) *has a positive maximal solution* $u_M \in L^\infty(0,\infty;M)$ *and a positive minimal solution* $u_m \in L^\infty(0,\infty;X)$, *such that if* $u \in L^\infty(0,\infty;X)$ *is any solution of* (4.7), (4.8), *then*

$$0 \leqq u_m(t) \leqq u(t) \leqq u_M(t) \leqq u_\infty \quad a.e. \text{ on } (0,\infty). \tag{4.9}$$

Remark 6.1. If $u \in L^\infty(0,\infty;X)$ is a solution of (4.7), (4.8), then it is easily checked that u is a solution of (4.6) in the sense defined above. Note that if the solution $u \in L^\infty(0,\infty;X)$ satisfies the estimate (4.9), then $u \in L^\infty(0,\infty;L^\infty(\Omega))$, and thus $\gamma(u) \in L^\infty(0,\infty;X)$, as well as $\gamma(u) \in L^1(0,T;X)$ for every $T>0$. These observations are needed for the definition of weak solution.

Remark 6.2. Theorem 6 also holds if the requirement γ nondecreasing is replaced by $\rho u+\gamma(u)$ nondecreasing for some $\rho>0$. To see this replace $-\nabla^2 u$ by $-\nabla^2 u+\rho u$ and replace $\gamma(u)$ by $\rho u+\gamma(u)$ in (4.5), (4.6) and apply the above analysis.

Remark 6.3. Comparing equations (4.1) of Example 1 and (4.6) and taking $f(t)\equiv u_0$ in (4.1), $u_0 \in L^2_+(\Omega)$ we note that if β is single valued and continuous, equations (4.1) and (4.6) differ only by the sign of the nonlinearity. For equation (4.1) one has existence and uniqueness of solutions of $(0,\infty)$ for every $u_0 \in L^2_+(\Omega)$. By contrast, for equation (4.6) it is known that if equation (4.5) has $u=0$ as the only nonnegative solution, then equation (4.6) can have a positive solution only on a finite interval $(0,T)$. From example, if $n=3$ and $\gamma(u)=u^5$, it follows from [21, Remark 3.2] that if Ω is star shaped, then (4.5) has $u=0$ as the only nonnegative solution. Taking $a(t)\equiv 1$, applying [11, Thm. 2.6 and Remark 2.7], and assuming that $u_0 \geqq 0$ and that

$$\int_\Omega u_0(x)\phi_0(x)\,dx \geqq \lambda_0^{1/4},$$

where λ_0 is the smallest eigenvalue and the corresponding unique eigenfunction $\phi_0>0$ in Ω:

$$-\nabla^2\phi_0=\lambda_0\phi_0 \quad \text{in } \Omega, \qquad \phi_0|_\Gamma=0,$$

then the unique nonnegative solution u of (4.6) exists only on a finite interval.

Proof of Theorem 6. Let $E=L^\infty(0,\infty;X)$ with the usual ordering (i.e., $u, v \in E$, $u \leqq v \Leftrightarrow \tilde{u}(t,x) \leqq \tilde{v}(t,x)$ a.e. in $(t,x) \in (0,\infty)\times\Omega$, where $\tilde{u}$ and $\tilde{v}$ are elements of the equivalence classes u and v respectively). In E let I denote the interval $[0,u_\infty]$ in the sense of order in E. It can be shown that I is a complete lattice with respect to this ordering. For every $u_0 \in I$ we define the function $\tilde{F}_{u_0}$ by

$$\tilde{F}_{u_0}(u)(t)=S(t)u_0+R*\tilde{\gamma}(u)(t)$$

where

$$\tilde{\gamma}(u)=\begin{cases}\gamma(u) & \text{if } u \leqq \|u_\infty\|_{L^\infty(\Omega)},\\ \|u_\infty\|_{L^\infty(\Omega)} & \text{otherwise,}\end{cases}$$

in place of the function F_{u_0} defined by (4.8). Then $\tilde{F}_{u_0}$ satisfies

$$\tilde{F}_{u_0}: I \to I \tag{4.10}$$

and

(4.11) $\tilde{F}_{u_0}$ is monotone ($u, v \in I$ and $u \leqq v \Rightarrow \tilde{F}_{u_0}(u) \leqq \tilde{F}_{u_0}(v)$).

Let $u \in I$. Then, by Theorems 3 and 4, $\tilde{F}_{u_0}(u) \geqq 0$. Moreover, by the fact that

$$u_\infty = Su_\infty + R * \gamma(u_\infty),$$

we have

$$\tilde{F}_{u_0}(u) = Su_0 + R * \tilde{\gamma}(u) \leqq Su_0 + R * \gamma(u_\infty) = S(u_0 - u_\infty) + u_\infty \leqq u_\infty,$$

which proves (4.10). Clearly, (4.11) is evident from Theorem 3. By [2, Thm. 11, p. 115], the operator $\tilde{F}_{u_0}$ has a least and a greatest fixed point in I, which correspond respectively to the solutions u_m and u_M, since $u_m \leqq u_M \leqq u_\infty$ and therefore, $\tilde{\gamma}(u_m) = \gamma(u_m)$, $\tilde{\gamma}(u_M) = \gamma(u_M)$, and so $\tilde{F}_{u_0}(u_m) = F_{u_0}(u_m)$, $\tilde{F}_{u_0}(u_M) = F_{u_0}(u_M)$. This completes the proof of Theorem 6.

Appendix A. An assumption which has been used frequently in the literature concerning the kernel a is

$$(\mathrm{A}_1) \qquad a(t) \in C(0, T), \quad a(t) > 0, \quad t \in (0, T),$$
$$\text{and } \frac{a(t)}{a(t+\sigma)} \text{ nonincreasing as a function of } t \text{ for each } \sigma > 0,\ 0 < t + \sigma < T;$$

see Friedman [8], Levin [15], Miller [17], [19]. We shall prove that condition (A_1) is equivalent to the condition

$$(\mathrm{A}_2) \qquad a(t) \in C(0, T), \quad a(t) > 0, \quad t \in (0, T), \quad \text{and} \quad \log a(t) \text{ convex on } (0, T).$$

Moreover, we first prove a preliminary result.

LEMMA 1. *Let assumption* (A_2) *be satisfied. Then for every* $\nu > 0$, *there exists a function* a_ν *satisfying* (A_2) *and* $a_\nu \in C^1[0, T]$, *and* $a_\nu(t) \uparrow q(t)$ as $\nu \downarrow 0^+$ *for* $t \in (0, T)$.

Proof. Define $b: \mathbb{R} \to (-\infty, +\infty]$ by

$$b(t) = \begin{cases} \log a(t) & \text{if } t \in (0, T), \\ \lim\limits_{t \to 0^+} \log a(t) & \text{if } t = 0, \\ \lim\limits_{t \to T^-} \log a(t) & \text{if } t = T, \\ +\infty & \text{if } t \notin [0, T]. \end{cases}$$

Observe that $a(t) > 0$ on $(0, T)$ and the definition of convexity of $\log a(t)$ on $(0, T)$ excludes $a(0^+) = 0$ and $a(T^-) = 0$. Thus b is convex, l.s.c. and proper on $\mathbb{R}$. Define b_ν, $\nu > 0$, to be the Yosida approximation of b; then (see [3, Prop. 2.11])

$$b_\nu(t) = \min_{y \in \mathbb{R}} \left\{ \frac{1}{2\nu} |y - t|^2 + b(y) \right\}, \qquad t \in \mathbb{R},$$

and $b_\nu \in C^1(\mathbb{R})$, b'_ν satisfies a Lipschitz condition on $\mathbb{R}$ with constant $1/\nu$; moreover $b_\nu(t) \uparrow b(t)$ as $\nu \downarrow 0^+$, $t \in \mathbb{R}$. Define $a_\nu = e^{b_\nu}$ and the result follows. This completes the proof of Lemma 1. Using Lemma 1 we shall prove

LEMMA 2. *The conditions* (A_1) *and* (A_2) *are equivalent.*

Proof. That $(\mathrm{A}_1) \Rightarrow (\mathrm{A}_2)$ follows from

$$\frac{a(t)}{a(t+\sigma)} \geqq \frac{a(t+\tau)}{a(t+\tau+\sigma)} \qquad (0 < t < t+\sigma,\ t+\tau < t+\sigma+\tau < T);$$

using $a(t)>0$ and putting $\sigma=\tau$ we obtain

$$a(t)a(t+2\tau)\geqq a^2(t+\tau).$$

Thus putting $t_1=t$, $t_2=t+2\tau$ we have

$$\log a\left(\frac{t_1+t_2}{2}\right)\leqq \tfrac{1}{2}\log a(t_1)+\tfrac{1}{2}\log a(t_2).$$

We note that in [15, calculation following Theorem 1.3] it is only shown that $(A_1)\Rightarrow a(t)$ convex, with the additional assumption that a is nonincreasing, which is not used. Of course, $\log a(t)$ convex implies $a(t)$ convex.

To prove that $(A_2)\Rightarrow(A_1)$, it is sufficient by Lemma 1 to prove $(A_2)\Rightarrow(A_1)$ with the additional assumption $a\in C'[0, T]$. Then $\log a(t)$ convex implies

$$\frac{a'(t)}{a(t)}\leqq\frac{a'(t+\sigma)}{a(t+\sigma)}\qquad (0<t<t+\sigma<T).$$

Using $a(t)>0$ we then have

$$\frac{d}{dt}\frac{a(t)}{a(t+\sigma)}=\frac{a(t+\sigma)a'(t)-a'(t+\sigma)a(t)}{a^2(t+\sigma)}\leqq 0,$$

which completes the proof of Lemma 2.

Proof of Proposition 1. By Lemma 2 it is sufficient to prove Proposition 1(i) under assumptions (A_1) and (H_2). If, in addition, $a<C[0, T]$, Proposition 1(i) follows directly from [17, Thm. 1] with $h=f=a$ and $g(x, t)=x$.

Let a satisfy assumptions (A_1) and (H_2). Consider the functions a_ν of Lemma 1. Then by the above remark, the functions $r_\nu(t, \lambda)\geqq 0$, $t\in[0, T]$, for every $\lambda>0$, $\nu>0$, where $r_\nu(t, \lambda)$ is the resolvent kernel associated with $\lambda a_\nu(t)$. The functions a_ν converge to a in $L^1(0, T)$ as $\nu\downarrow 0^+$, since $a_\nu(t)\uparrow a(t)$ as $\nu\downarrow 0^+$ and $a\in L^1(0, T)$. Therefore, it is easily checked that the functions $r_\nu(\cdot, \lambda)$ converge to $r(\cdot, \lambda)$ in $L^1(0, T)$, where $r(t, \lambda)$ is the resolvent kernel corresponding to $\lambda a(t)$, and $r(t, \lambda)\geqq 0$ on $[0, T]$ a.e. This completes the proof of part (i).

Part (ii) is proved in [9, Lemma 2.5] with $h=\lambda a$ (see also [15, Lemma 1.3] with $f\equiv 1$), where the proof is carried out on $(0, \infty)$; this can be applied by extending $a(t)$ as a constant on $[T, \infty]$. This completes the proof of Proposition 1.

Acknowledgments. Example 3 was proposed to us by Professor L. A. Peletier. We are grateful to Professor M. G. Crandall for discussing Example 3 with us. We also gratefully acknowledge several helpful comments of the referee; in particular, our original proof of Theorems 1 and 2 in the unbounded case of A was incomplete.

REFERENCES

[1] V. BARBU, *Nonlinear Volterra equations in a Hilbert space*, this Journal, 6 (1975), pp. 728–741.

[2] G. BIRKHOFF, *Lattice Theory*, Colloquium Publications, 25, American Mathematical Society, Providence, RI, 1973.

[3] H. BRÉZIS, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland, Amsterdam, 1973.

[4] ———, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, pp. 101–156.

[5] H. BRÉZIS AND W. STRAUSS, *Semilinear elliptic equations in L^1*, J. Math. Soc. Japan, 25 (1973), pp. 565–590.

[6] M. G. CRANDALL, *An introduction to evolution governed by accretic operators*, Dynamical Systems, vol. 1, Academic Press, New York, 1976, pp. 131–165.
[7] M. G. CRANDALL AND J. A. NOHEL, *An abstract functional differential equation and a related nonlinear Volterra equation*, Israel J. Math 29 (1978), pp. 313–328.
[8] A. FRIEDMAN, *On integral equations of Volterra type*, J. Analyse Math., 11 (1963), pp. 381–413.
[9] ———, *Monotonicity of solutions of Volterra integral equations in Banach space*, Trans. Amer. Math. Soc., 138 (1969), pp. 129–148.
[10] A. FRIEDMAN AND M. SHINBROT, *Volterra integral equations in Banach space*, Ibid., 126 (1967), pp. 131–179.
[11] H. FUJITA, *On some nonexistence and nonuniqueness theorems for nonlinear parabolic equations*, Proceedings of Symposia in Pure Mathematics, vol. 18, American Mathematical Society, Providence, RI, 1970, pp. 105–113.
[12] R. C. GRIMMER AND R. K. MILLER, *Existence, uniqueness, and continuity for integral equations in Banach space*, J. Math. Anal. Appl., 57 (1977), pp. 429–447.
[13] G. GRIPENBERG, *An existence result for a nonlinear Volterra integral equation in Hilbert space*, this Journal, to appear.
[14] E. HEWITT AND K. STROMBERG, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965.
[15] J. J. LEVIN, *Resolvents and bounds for linear and nonlinear Volterra equations*, Trans. Amer. Math. Soc., 228 (1977), pp. 207–222.
[16] S. O. LONDEN, *On an integral equation in a Hilbert space*, this Journal, 8 (1977), pp. 950–970.
[17] R. K. MILLER, *On Volterra integral equations with nonnegative integrable resolvents*, J. Math. Anal. Appl., 22 (1968), pp. 319–340.
[18] ———, *Volterra integral equations in a Banach space*, Funkcial. Ekvac., 18 (1975), pp. 163–193.
[19] ———, *Nonlinear Volterra Integral Equations*, W. A. Benjamin, Menlo Park, CA, 1971.
[20] A. PAZY, *Semigroups of linear operators and applications to partial differential equations*, Lecture notes, Univ. of Maryland, College Park, 1974.
[21] P. H. RABINOWITZ, *Variational methods for nonlinear eigenvalue problems*, Eigenvalues of Non-Linear Problems, C.I.M.E., Edizioni Cremonese, Rome, 1974, pp. 141–195.
[22] G. E. H. REUTER, *Über eine Volterrasche Integralgleichung mit total-monotoneum Kern*, Arch. Math., 7 (1956), pp. 59–66.
[23] G. SEIFERT, *Positively invariant closed sets for systems of delay differential equations*, J. Differential Equations, 22 (1976), pp. 292–304.
[24] D. G. WEIS, *Asymptotic behaviour of some nonlinear Volterra integral equations*, J. Math. Anal. Appl., 49 (1975), pp. 59–87.
[25] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1941.
[26] K. YOSIDA, *Functional Analysis*, Springer-Verlag, Berlin, 1966.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0016 $01.00/0

NONLINEAR BUCKLING AND STABILITY OF CYLINDRICAL PANELS*

GEORGE H. KNIGHTLY† AND D. SATHER‡

Abstract. The problem of the buckling of a simply supported thin, elastic, cylindrical panel subjected to a uniform axial compression is studied by constructive methods. In the case of "narrow" panels we obtain near the classical buckling load λ_1 a description of *all* stable and unstable buckled states which branch from the unbuckled state at λ_1. The method also yields useful information on the asymptotic form of the buckled states near λ_1, the exchange of stabilities at λ_1, and whether the buckling at the ends of a narrow panel is "inward or outward".

1. Introduction. A typical situation in buckling problems for thin, elastic shells subjected to some sort of conservative uniform load σ is that the shell possesses an "unbuckled" equilibrium state U for all $\sigma > 0$. The state U is usually stable for small values of σ; however, as σ increases beyond some critical value σ_c, the shell may suddenly deform with "large" displacements into a new "buckled" equilibrium state; the unbuckled state U may or may not be stable for $\sigma > \sigma_c$. Even though the value of σ_c may be significantly less than the classical buckling load σ_1 of the problem (i.e., the smallest (positive) eigenvalue of the appropriate linearized problem), a description of the possible stable and unstable buckled states of the shell for σ near σ_1 is often useful because it sheds some light on the triggering of the buckling phenomenon; for example, if there are no stable buckled states near U for $\sigma < \sigma_1$, such a result is widely accepted as a theoretical explanation of the occurrence of snap-buckling for thin, elastic shells. Since the initial behavior of the buckled states at σ_1 also provides in some cases a theoretical explanation of the imperfection sensitivity of a shell (e.g., see [6], [7]), a complete buckling and stability analysis near the classical buckling load σ_1 may be of considerable theoretical interest for shell buckling problems even though, in practice, the shell undergoes snap-buckling for loads σ significantly less than σ_1.

In the present paper we study the buckling of a thin, elastic cylindrical panel which is subjected to a uniform axial compression and is simply supported at its edges. Although the approach used is a constructive one and may be used to study panels of any width (and even circular cylindrical shells under general boundary conditions), we deal principally here with "narrow" panels (see Definition 3 in § 2) for which the results are more complete. In the case of narrow panels, we obtain near the classical buckling load λ_1 a description of *all* stable and unstable buckled states which branch from the unbuckled state at λ_1; in particular, for $\lambda_1 - \delta < \lambda < \lambda_1$, we show rigorously that there are no stable buckled states which branch from the unbuckled state at λ_1 and depend continuously on the load parameter λ. In addition, our approach yields useful information on the asymptotic form of the buckled states near λ_1, the exchange of stabilities at λ_1, and whether the buckling at the ends of the panel is "inward" or "outward."

The model used here for a simply supported cylindrical panel is obtained from the classical von Kármán–Donnell equations and consists of a coupled pair of nonlinear fourth-order partial differential equations together with simply supported boundary conditions. By making use of generalized (or weak) solutions of the problem, we

* Received by the editors July 28, 1977, and in revised form April 17, 1978.

† Department of Mathematics, University of Massachusetts, Amherst, Massachusetts 01002. The work of this author was supported in part by the National Science Foundation under Grant No. MPS 75-07579.

‡ Department of Mathematics, University of Colorado, Boulder, Colorado 80302. The work of this author was supported in part by the National Science Foundation under Grant No. MPS 73-08948 and in part by the Council on Research and Creative Work of the University of Colorado.

reformulate the problem in an appropriate real Hilbert space $\mathscr{H}$ so that the problem of determining the buckled states of the axially-compressed panel is equivalent to one of finding the nontrivial solutions of a single operator equation of the form

$$(*) \qquad w-\lambda Aw+\alpha^2A^2w+\alpha Q(w)+C(w)=0, \quad w\in\mathscr{H}.$$

Here A: $\mathscr{H}\to\mathscr{H}$ is a linear, selfadjoint, positive, compact operator, Q: $\mathscr{H}\to\mathscr{H}$ and C: $\mathscr{H}\to\mathscr{H}$ are continuous, homogeneous, polynomial operators of degree two and three, respectively, Q is the gradient of the functional $\tau(w)=\frac{1}{3}(Q(w), w)$, and C is the gradient of $\sigma(w)=\frac{1}{4}(C(w), w)$; the function w is a measure of the radial deflection of the panel from its unbuckled state, the parameter λ is a measure of the uniform compressive load, and the (fixed) geometrical parameter α is proportional to $(\mathrm{Rh})^{-1}$, where R denotes the radius of the undeformed cylindrical panel and h denotes the (uniform) thickness of the panel.

The approach used to study equation (*) is based upon the Lyapunov–Schmidt method rather than topological methods (e.g., see the variational approaches in [1], [9]) because we are interested here not only in the existence of nontrivial solution branches of (*), but also in the form of the solution branches near λ_1 and the stability properties of the various solution branches. The concept of stability alluded to here (and throughout the paper) is that of "linearized stability", i.e., a solution w of (*) is stable at λ if the Fréchet derivative of the left hand side of (*) at (w, λ) has only positive eigenvalues whereas w is unstable at λ if some eigenvalue is negative.

The nonlinear buckling of cylindrical panels was first studied by Koiter [6] under boundary conditions different from those considered here. Koiter obtained a number of interesting results for cylindrical panels and, in particular, he showed for the first time that a certain class of "narrow" panels exhibits the stable behavior of a flat plate while yet another class has the unstable behavior of a shell. On the other hand, for simply supported narrow cylindrical panels, we show rigorously that an even more complicated situation exists, depending on whether the smallest eigenvalue λ_1 of the linearized problem has multiplicity $k=1$ or $k=2$ (the only possible multiplicities for the buckling load of a narrow panel).

For example, if $k=1$ and the load is near λ_1, the panel in some cases exhibits throughout the entire range of narrow panels the stable behavior of a flat plate with an exchange of stabilities at λ_1, while in other cases the panel exhibits in part of the range of narrow panels the unstable behavior of a cylinder (see Theorem 3 and Remark 1).

On the other hand, if $k=2$ and the load λ is near λ_1, we show that there are no stable states near $w=0$ for $\lambda<\lambda_1$ but there always exists a unique stable buckled state near $w=0$ for $\lambda>\lambda_1$. More precisely, if $k=2$, we show that there exists a positive constant δ such that, for $0<|\lambda-\lambda_1|<\delta$, there are exactly three buckled states which branch from the unbuckled state at λ_1 and depend continuously on the load parameter λ. All three of these buckled states are unstable for $\lambda<\lambda_1$ while the unbuckled state is stable for $\lambda<\lambda_1$. As λ crosses λ_1 the unbuckled state loses its stability to one of the buckled states so that, for $\lambda_1<\lambda<\lambda_1+\delta$, there is one stable buckled state, two unstable buckled states, and one unstable unbuckled state; thus, under the assumption that the model being used here adequately describes the buckling of a cylindrical panel, the theoretical possibility of an exchange of stabilities at λ_1 always exists for narrow panels when $k=2$. Furthermore, when the length of the panel greatly exceeds its width and $\lambda_1<\lambda<\lambda_1+\delta$, the three buckled states have the qualitative shapes shown in Fig. 1 (in each case the x axis is the cylindrical axis, the broken line represents a generator of the unbuckled cylindrical panel, and the solid curve indicates

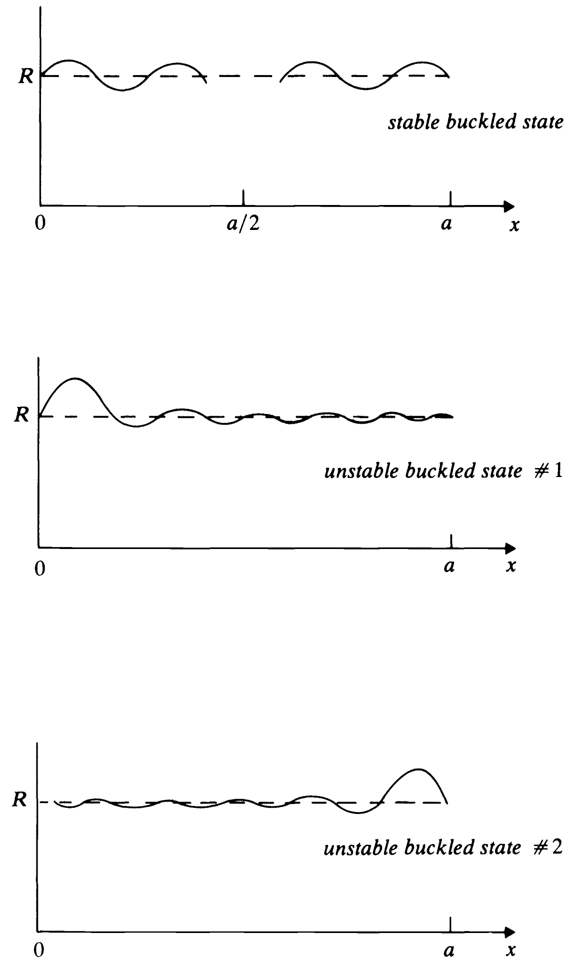

FIG. 1

the shape of a generator ($y = \text{constant}$) of the buckled panel). Note that in each case the panel buckles outward at both ends $x = 0$ and $x = a$. The indicated buckling pattern is symmetric about $x = a/2$ for the stable state, whereas the unstable buckled states are distinguished by a concentration of the dimpling at one of the ends and relatively little buckling at the other end.

Since the classical buckling load λ_1 has only multiplicity $k = 1$ or $k = 2$ for narrow panels (see § 2), the indicated results provide an essentially complete branching and stability analysis near λ_1 for narrow panels.

The main results of the present paper were announced in [11]. Other related branching and stability results for flat plates and spherical shells are described in [5] and [11]. The reader is also referred to a recent paper of Mallet-Paret [8] in which a different approach is used to study the existence of buckled states of cylindrical panels.

2. Formulation of the problem. The model used here for the problem of an axially compressed cylindrical panel of (circumferential) width b is that of the von Kármán–Donnell equations (e.g., see [7, pp. 268–269]). If x denotes the axial coordinate and y denotes the circumferential coordinate of the panel, and if a uniform compressive load is applied normal to the edges of the panel at $x = 0$ and $x = a$, then the governing equations may be written as

$$\Delta^2 f = -\tfrac{1}{2}[w, w] + \alpha w_{xx}, \tag{2.1a}$$

$$\Delta^2 w = [w, f] - \lambda w_{xx} - \alpha f_{xx} \quad \text{in } \Omega, \tag{2.1b}$$

where $\Omega = \{(x, y): 0 < x < a, 0 < y < b\}$, Δ^2 denotes the biharmonic operator and

$$[u, v] = u_{xx}v_{yy} + u_{yy}v_{xx} - 2u_{xy}v_{xy}. \tag{2.2}$$

Here, after appropriate scaling, w represents the radial component of deflection of the panel, f is an "excess" stress function, and λ measures the compressive axial load on the panel, while the (fixed) geometrical parameter $\alpha = [12(1-\nu^2)]^{1/2}(Rh)^{-1}$ is proportional to the curvature R^{-1}, R is the (constant) radius of the undeformed panel, h its thickness and ν is Poisson's ratio. In addition we shall impose the "simply supported" conditions on the boundary $\partial\Omega$,

$$w = f = \Delta w = \Delta f = 0 \quad \text{on } \partial\Omega. \tag{2.3}$$

DEFINITION 1. A *classical solution of Problem CP* is a pair of functions w, f that belong to $C^4(\Omega) \cap C^2(\bar{\Omega})$ and satisfy (2.1) and (2.3) pointwise.

Next we shall obtain a Hilbert space formulation of a generalized solution of Problem CP. To do this we make use of the Hilbert space $\mathcal{W}$ defined in [3] as the closure, in the norm $\|\cdot\|_{2,2}$ of the Sobolev space $W_{2,2}(\Omega)$, of the set of smooth functions defined on $\bar{\Omega}$ and vanishing on $\partial\Omega$. From the Sobolev embedding theorem it follows that the functions in $\mathcal{W}$ are continuous on $\bar{\Omega}$ and zero on $\partial\Omega$. In this paper we use an equivalent inner product on $\mathcal{W}$ given by

$$(u, v) = \int_\Omega \Delta u \, \Delta v,$$

with corresponding norm denoted by $\|\cdot\|$.

Let φ, ψ be smooth functions on $\mathcal{W}$. Multiplying equation (2.1a) by φ and (2.1b) by ψ and integrating by parts over Ω yields

$$(f, \varphi) = -\tfrac{1}{2}b(w, w; \varphi) - \alpha c(w; \varphi), \tag{2.4a}$$

$$(w, \psi) = b(w, f; \psi) + \lambda c(w; \psi) + \alpha c(f; \psi), \tag{2.4b}$$

where

$$b(u, v; \varphi) = \int_\Omega [(u_{xy}v_y - u_{yy}v_x)\varphi_x + (u_{xy}v_x - u_{xx}v_y)\varphi_y] \tag{2.5}$$

and

$$c(u; \varphi) = \int_\Omega u_x \varphi_x. \tag{2.6}$$

DEFINITION 2. A *generalized solution of Problem CP* is a pair of functions w, f in $\mathcal{W}$ satisfying (2.4) for all φ, ψ in $\mathcal{W}$.

As in [3] we observe that, for fixed $u, v \in \mathcal{W}$, $b(u, v; \varphi)$ and $c(u; \varphi)$ are, in φ, bounded linear functionals on $\mathcal{W}$. Invoking the Riesz representation theorem we may recast equations (2.4) as the following pair of uncoupled operator equations on $\mathcal{W}$ (note that (2.8) is the basic equation (*) discussed in the Introduction):

$$f = -\tfrac{1}{2}B(w, w) - \alpha A w, \tag{2.7}$$

$$L_\lambda w + \alpha Q(w) + C(w) = 0. \tag{2.8}$$

Here $B: \mathscr{W} \times \mathscr{W} \to \mathscr{W}$ is a bounded bilinear operator such that, for all $\varphi \in \mathscr{W}$,

$$(B(u, v), \varphi) = b(u, v; \varphi), \qquad u, v \in \mathscr{W}, \tag{2.9}$$

$$(Au, \varphi) = c(u; \varphi), \qquad u \in \mathscr{W}. \tag{2.10}$$

Then L_λ, Q and C are mappings of $\mathscr{W}$ to itself, defined by

$$L_\lambda w = w - \lambda A w + \alpha^2 A^2 w, \tag{2.11}$$

$$Q(w) = B(w, Aw) + \tfrac{1}{2} AB(w, w) \tag{2.12}$$

and

$$C(w) = \tfrac{1}{2} B(w, B(w, w)). \tag{2.13}$$

The preceeding steps reveal that solving (2.7), (2.8) is equivalent to finding a generalized solution of Problem CP. A slight modification of the proof of Theorem B.1 in [3] shows that the following theorem is valid (see also [2]).

THEOREM 1. *Every classical solution of Problem CP is a generalized solution. Every generalized solution is a classical solution in Ω and up to $\partial'\Omega$, where $\partial'\Omega$ denotes the boundary of Ω with its corners deleted.*

In view of the above observations we have reduced Problem CP to a study of equation (*).

The following two lemmas are used below in the discussion of the linear eigenvalue problem associated with equation (*).

LEMMA 1. *A is a bounded, linear, compact, selfadjoint, positive operator. Its characteristic values μ_{pq} are given by*

$$\mu_{pq} = \left(\pi/a\right)^2 \left(p + \frac{q^2 a^2}{pb^2}\right)^2, \qquad p, q = 1, 2, 3, \cdots, \tag{2.14}$$

and the normalized eigenfunction corresponding to μ_{pq} is

$$w_{pq} = C_{pq} \sin \frac{p\pi x}{a} \sin \frac{q\pi y}{b}, \quad \text{where } C_{pq}^{-1} = \frac{\pi^2(b^2p^2 + a^2q^2)}{(2(ab)^{3/2})}.$$

The sequence $\{w_{pq}\}$ is complete in $\mathscr{W}$.

Proof. The listed properties of A follow readily from the definitions (2.10) and (2.6) (for the compactness, see [2]). The completeness of the sequence $\{w_{pq}\}$ in $\mathscr{W}$ follows from its completeness in $\mathscr{L}^2(\Omega)$; if $f \in \mathscr{W}$ and $(f, w_{pq}) = 0$, $p, q = 1, 2, \cdots$, then

$$0 = (f, w_{pq}) = \mu_{pq}(f, Aw_{pq}) = \mu_{pq} c(f; w_{pq}) = -\mu_{pq} \int_\Omega f(w_{pq})_{xx}$$

$$= \mu_{pq}(p\pi/a)^2 \int_\Omega f w_{pq},$$

for $p, q = 1, 2, \cdots$, which implies that $f = 0$.

As in Lemma 4 of [4], the selfadjoint property of A together with the factorizations

$$L_\lambda = (I - \mu_+ A)(I - \mu_- A) = (I - \mu_- A)(I - \mu_+ A),$$

imply the next lemma.

LEMMA 2. (i) *L_λ is selfadjoint.*

(ii) *For fixed $\alpha > 0$, λ_0 is an eigenvalue of L_λ if and only if at least one of the two numbers $\mu_+ = \frac{1}{2}[\lambda_0 + (\lambda_0^2 - 4\alpha^2)^{1/2}]$ or $\mu_- = \frac{1}{2}[\lambda_0 - (\lambda_0^2 - 4\alpha^2)^{1/2}]$ is a characteristic value* *of A, i.e. if and only if*

$$\lambda_0 = \mu_0 + \alpha^2 \mu_0^{-1} \tag{2.15}$$

holds for some characteristic value, μ_0, of A.

(iii) *If λ_0 is an eigenvalue of L_λ, then the corresponding eigenfunctions of A (i.e. at μ_+, μ_- or both) are also eigenfunctions of L_{λ_0} and span the null space $\mathcal{N}(L_{\lambda_0})$ of L_{λ_0}.*

As a consequence of part (ii) of Lemma 2 we see that an eigenvalue λ_0 of L_λ may have multiplicity $k>1$ in either of two ways: (i) if (2.15) holds for some characteristic value μ_0 of A having multiplicity exceeding one, or (ii) if α^2 equals the product of two distinct characteristic values of A. The first of these is the only way to achieve multiplicity $k>1$ if the panel is required to be "narrow" in the following sense.

DEFINITION 3. For fixed $\alpha>0$, the cylindrical panel is *narrow* if

$$b\sqrt{\alpha}/(2\pi) \leqq 1. \tag{2.16}$$

We note that the condition (2.16) is equivalent to the one given by Koiter [6] in his discussion of a narrow panel. From (2.14) one easily obtains a lower bound for the characteristic values of A: $\mu_{pq} \geqq \mu_{p1} \geqq (2\pi/b)^2$. Thus, if the panel is narrow, no characteristic value of A is smaller than α and α^2 cannot equal the product of distinct characteristic values of A. For a narrow panel we also see that the relation (2.15) generates the eigenvalues of L_λ from the characteristic values of A with ordering by magnitude preserved. Since it follows from (2.14) that the smallest characteristic value of A is either simple or, if the ratio a^2/b^2 is the product of *successive* integers, double, and since L_λ reduces to $I-\lambda A$ when $\alpha=0$, which is actually the case of a rectangular plate [3], we see that the multiplicity of the buckling load for a narrow panel is the same (either one or two) as that for a rectangular plate having the same dimensions.

The following lemma, which lists some useful properties of the nonlinear operators B, Q and C, is easily proved by combining Lemma 1 of [3] with the definitions (2.5), (2.9), (2.12) and (2.13) of the present paper.

LEMMA 3. (i) *If u, v, w lie in $\mathcal{W}$, then the form $(B(u,v), w)$ is symmetric in u, v, w.*

(ii) *Q is a continuous, homogeneous polynomial operator of degree two and the gradient of the real-valued functional*

$$\tau(w) = \tfrac{1}{3}(Q(w), w), \qquad w \in \mathcal{W}. \tag{2.17}$$

For each $w \in \mathcal{W}$ the operator Q has a differential $Q'(w)$, which satisfies

$$Q'(w)u = B(w, Au) + B(u, Aw) + AB(w, u), \tag{2.18}$$

for all $u \in \mathcal{W}$, and is Lipschitz continuous in w.

(iii) *C is a continuous, homogeneous, polynomial operator of degree three and the gradient of the functional*

$$\sigma(w) = \tfrac{1}{4}(C(w), w), \qquad w \in \mathcal{W}. \tag{2.19}$$

For each $w \in \mathcal{W}$ the operator C has a differential $C'(w)$, which satisfies

$$C'(w)u = \tfrac{1}{2}B(u, B(w, w)) + B(w, B(u, w)), \tag{2.20}$$

for all $u \in \mathcal{W}$, and is Lipschitz continuous in w.

As a final preliminary remark, we state the following lemma which will be useful in evaluating the coefficients of the branching equations derived in § 3.

LEMMA 4. (i) *Let $\{w_{pq}\}$ be the eigenfunctions of A given in Lemma 1 and set $u_{pq} = (\mu_{pq})^{1/2} w_{pq}$, for $p, q = 1, 2, \cdots$. Then $\{u_{pq}\}$ is orthonormal with respect to A in the* *sense that*

(2.21) $$(Au_{pq}, u_{mn}) = \delta_{pm}\delta_{qn}, \qquad p, q, m, n = 1, 2, \cdots.$$

(ii) *In addition,* $(Q(u_{pq}), u_{mn}) = 0$ *unless m and n are both odd, in which case it equals*

$$\frac{64}{\pi}\left(\frac{1}{\mu_{pq}} + \frac{1}{2\mu_{mn}}\right)b^{-2}\left(\frac{a}{b}\right)^{1/2}\frac{q^2}{m}\,\frac{2q^2m^2 + 2p^2n^2 - m^2n^2}{mn(4p^2 - m^2)(4q^2 - n^2)}.$$

Proof. (i) is an immediate consequence of Lemma 1, since $\|w_{pq}\| = 1$. To establish (ii), note that (2.12), Lemma 3 and Lemma 1 imply that

$$(Q(u_{pq}), u_{mn}) = [\mu_{pq}^{-1} + (2\mu_{mn})^{-1}](B(u_{pq}, u_{pq}), u_{mn}).$$

From (2.9), (2.5) and an integration by parts, we find that

$$(B(u_{pq}, u_{pq}), u_{mn}) = \int_\Omega [u_{pq}, u_{pq}]u_{mn},$$

and (ii) follows by direct computation.

3. The branching and stability results. Let λ_0 be any eigenvalue of L_λ of multiplicity k so that the null space $\mathcal{N} \equiv \mathcal{N}(L_{\lambda_0})$ of L_{λ_0} is k-dimensional. As in [4], [11] we use the Lyapunov–Schmidt technique to reduce the problem of finding solutions of equation (*) in $\mathcal{W}$ to a problem in $\mathbb{R}^k$ (Euclidean k-space). If S denotes the orthogonal projection of $\mathcal{W}$ onto $\mathcal{N}$, and $\mathcal{N}^\perp$ denotes the orthogonal complement of $\mathcal{N}$ in $\mathcal{W}$, then the elements w of $\mathcal{W}$ can be resolved into components, $w = v + V$, with $v = Sw$ in $\mathcal{N}$ and $V = (I - S)w$ in $\mathcal{N}^\perp$. Moreover, if a basis $\{v_1, \cdots, v_k\}$ for $\mathcal{N}$ is introduced such that

(3.1) $$(Av_i, v_j) = \delta_{ij}, \qquad i, j = 1, \cdots, k,$$

(δ_{ij} is the Kronecker delta), then an element $v \in \mathcal{N}$ can be represented as $v = \sum_{i=1}^k \xi_i v_i$ for suitable $\xi = (\xi_1, \cdots, \xi_k)$ in $\mathbb{R}^k$. Equation (*) can then be decomposed into its projections onto $\mathcal{N}$ and $\mathcal{N}^\perp$ and, for sufficiently small (fixed) ξ and $\eta = \alpha^{-1}(\lambda - \lambda_0)$, say $|\xi| < \rho_0$ and $|\eta| < \eta_0$, the resulting equation on $\mathcal{N}^\perp$ can be solved for V by means of the contraction mapping principle. The solution $V = V(\xi, \eta)$ is analytic in ξ and η and satisfies, for some constant $K > 0$,

(3.2) $$\|V(\xi, \eta)\| \leq K|\xi|^2, \qquad |\xi| < \rho_0, \quad |\eta| < \eta_0.$$

The projection of (*) onto $\mathcal{N}$ then becomes

(3.3) $$0 = -(\lambda - \lambda_0)\xi_i + \alpha\left(Q\left(\sum_{j=1}^k \xi_j v_j + V\right), v_i\right) + \left(C\left(\sum_{j=1}^k \xi_j v_j + V\right), v_i\right) \qquad (i = 1, \cdots, k),$$

where $V = V(\xi, \eta)$ is determined as in the above.

If we now set

(3.4) $$\beta = (\beta_1, \cdots, \beta_k) \equiv \eta^{-1}(\xi_1, \cdots, \xi_k), \qquad \eta = \alpha^{-1}(\lambda - \lambda_0) \neq 0,$$

in (3.3) and cancel a factor $\alpha\eta^2$, the resulting system is

(3.5) $$0 = -\beta_i + \left(Q\left(\sum_{j=1}^k \beta_j v_j + \eta^{-1}V\right) + \eta\alpha^{-1}C\left(\sum_{j=1}^k \beta_j v_j + \eta^{-1}V\right), v_i\right) \qquad (i = 1, \cdots, k).$$

Formally setting $\eta = 0$ in (3.5) and using (3.2) we obtain the so-called reduced branching system (in $\mathbb{R}^k$)

$$0 = -\beta_i + \left(Q\left(\sum_{j=1}^{k} \beta_j v_j \right), v_i \right) \equiv F_i(\beta) \qquad (i = 1, \cdots, k) \tag{3.6}$$

which plays a major role in the subsequent analysis.

Suppose that $\beta^* \neq 0$ is a solution of (3.6) at which the Jacobian $J = \partial(F_1, \cdots, F_k)/\partial(\beta_1, \cdots, \beta_k)$ is different from zero. Then we conclude from the implicit function theorem that, for $|\eta|$ sufficiently small, the system (3.5) has a nontrivial solution $\beta(\eta) = \beta^* + \tilde{\beta}(\eta)$, where $\tilde{\beta}$ is analytic in η and satisfies $\lim_{\eta \to 0} \tilde{\beta}(\eta) = 0$. Then $\xi(\eta) = \eta\beta(\eta)$ satisfies (3.3) and, in turn, generates a solution w^* of (*) of the form

$$w^*(\lambda) = \alpha^{-1}(\lambda - \lambda_0)v^* + V^*,$$

where $v^* = \sum_{i=1}^{k} \beta_i^* v_i$, and V^* is analytic in λ and satisfies $\lim_{\lambda \to \lambda_0} (\lambda - \lambda_0)^{-1} V^* = 0$. Thus, the problem of determining nontrivial solutions of equation (*) near $w = 0$ is closely related to finding nontrivial solutions of the reduced branching system (3.6).

By relating solutions of (3.6) to extreme values of the functional τ in (2.17) on the ellipse $\mathscr{E} = \{u \in \mathscr{N}: (Au, u) = 1\}$, the following general results can be established for cylindrical panels (for a proof see [12, Thm. 4 and Thm. 5]).

THEOREM 2. *Suppose that $\mathscr{N} \equiv \mathscr{N}(L_{\lambda_1})$ is k-dimensional $(k \geqq 2)$ where, for fixed $\alpha > 0$, $\lambda_1 = \lambda_1(\alpha)$ denotes the smallest (positive) eigenvalue of L_λ. If the functional τ restricted to $\mathscr{E}$ has a positive relative minimum at u^* then equation (*) has a nontrivial solution branch $w^*(\lambda)$ of the form*

$$w^*(\lambda) = (\alpha a)^{-1}(\lambda - \lambda_1)u^* + U^*, \qquad 0 < |\lambda - \lambda_1| < \delta, \tag{3.7}$$

where $a = (Q(u^*), u^*)$ and U^* is analytic in λ and satisfies $\lim_{\lambda \to \lambda_1} (\lambda - \lambda_1)^{-1} U^* = 0$; for such a minimum the resultant solution $w^*(\lambda)$ is stable for $\lambda_1 < \lambda < \lambda_1 + \delta$ and unstable for $\lambda_1 - \delta < \lambda < \lambda_1$. If the functional τ restricted to $\mathscr{E}$ has a positive relative maximum at u^*, and if, in addition, $a = (Q(u^*), u^*)$ is not an eigenvalue of $SQ'(u^*)$, then equation (*) has a nontrivial solution branch $w^*(\lambda)$ of the form (3.7) which is unstable for $0 < |\lambda - \lambda_1| < \delta$.

Let us remark that for a relative maximum the condition in Theorem 2 on the eigenvalues of $SQ'(u^*)$ implies that the relevant Jacobian J in the above is different from zero. On the other hand, for a positive relative minimum the condition $J \neq 0$ (at the nontrivial solution β^* of (3.6) generated by u^*) necessarily holds (see [12]). Thus, whatever the dimension of the null space $\mathscr{N}$, the theoretical possibility of an exchange of stabilities at $\lambda = \lambda_1$ exists whenever the functional $\tau|_{\mathscr{E}}$ has a positive relative minimum on $\mathscr{E}$.

From this point on we restrict our attention to panels which are "narrow" in the sense of Definition 3 (see, however, Remark 3 below).

Before stating the results for narrow panels, let us note that in each case considered in Theorem 3 and Theorem 4 below, the reduced branching system (3.6) is solved completely and its Jacobian is shown to be different from zero at each of its solutions. Since in each case the system (3.6) has only a finite number of solutions, one can then show, using an argument involving a Newton polygon, as in the proof of Theorem 4.3 in [10], that there exist positive numbers ρ and δ such that, if λ_1 denotes the smallest (positive) eigenvalue of L_λ, the only nontrivial solutions (w, λ) of (*) in the set

$$\mathscr{S}(\lambda_1, \rho, \delta) = \{(w, \lambda): \|w\| < \rho, |\lambda - \lambda_1| < \delta\}$$

lie on the branches obtained by the implicit function theorem from the nontrivial solutions of the reduced branching system.

THEOREM 3. *Suppose that* λ_1, *the smallest eigenvalue of* L_λ *for a narrow panel, is simple with eigenfunction* $u_p \equiv u_{p1}$.

(i) *If* p is odd *then there are positive numbers* δ *and* ρ *such that the set of nontrivial solutions of* (*) *in* $\mathcal{S}(\lambda_1, \rho, \delta)$ *consists of a single buckled state* $(w^*(\lambda), \lambda)$ *that exists for* $0<|\lambda-\lambda_1|<\delta$ *and has the form*

$$w^*(\lambda)=(\alpha a_p)^{-1}(\lambda-\lambda_1)u_p+\Phi, \tag{3.8}$$

where $a_p=(Q(u_p), u_p)$ *and* Φ *is an analytic function of* λ *satisfying* $\lim_{\lambda\to\lambda_1}(\lambda-\lambda_1)^{-1}\Phi(\lambda)=0$. *Moreover,* $w^*(\lambda)$ *is stable for* $\lambda_1<\lambda<\lambda_1+\delta$ *and unstable for* $\lambda_1-\delta<\lambda<\lambda_1$.

(ii) *Suppose that* p is even *and that the number* γ *defined in* (3.17) *below is negative. Then there are positive numbers* δ *and* ρ *such that the set of nontrivial solutions of* (*) *in* $\mathcal{S}(\lambda_1, \rho, \delta)$ *consists of exactly two branches of buckled states,* $(w^+(\lambda), \lambda)$ *and* $(w^-(\lambda), \lambda)$, *given by*

$$w^\pm=\pm|\gamma|^{-1/2}|\lambda-\lambda_1|^{1/2}u_p+\Phi^\pm, \qquad \lambda_1-\delta<\lambda<\lambda_1, \tag{3.9}$$

where the $\Phi^\pm$ *are analytic functions of* $|\lambda-\lambda_1|^{1/2}$ *satisfying* $\lim_{\lambda\to\lambda_1}|\lambda-\lambda_1|^{-1/2}\Phi^\pm=0$. *Both states* $w^\pm$ *are unstable for* $\lambda_1-\delta<\lambda<\lambda_1$. *If* $\gamma>0$, *again there exist positive numbers* δ *and* ρ *such that the solutions of* (*) *in* $\mathcal{S}(\lambda_1, \rho, \delta)$ *consist of a pair of buckled states of the form* (3.9), *except that in this case the branches* $(w^\pm(\lambda), \lambda)$ *are defined for* $\lambda_1<\lambda<\lambda_1+\delta$ *and are stable.*

Remark 1. It is shown in the Appendix that for p even and for ratios $a/b>(2)^{1/2}$, there are numbers θ_1 and θ_2, $0<\theta_2<\theta_1<1$, such that if the curvature parameter $\theta\equiv b\sqrt{\alpha}/(2\pi)$ satisfies $\theta_1\leqq\theta\leqq 1$ then $\gamma<0$ whereas if $0<\theta\leqq\theta_2$ then $\gamma>0$. Hence all of the cases described in Theorem 3 actually occur within the range of narrow panels. For example, if $a/b=4$ then $\theta_1\leqq 0.72$ and $\theta_2\geqq 0.27$ so that $\gamma<0$ when $0.72\leqq\theta\leqq 1$ and $\gamma>0$ when $0<\theta\leqq 0.27$.

Proof of Theorem 2. (i) If p is odd the branching analysis proceeds along standard lines as described at the beginning of this section and the stability analysis is also obtained from known results (e.g., see [13]). In fact, the reduced branching system (3.6) is now a single equation

$$F(\beta)\equiv-\beta+\beta^2 a_p=0, \tag{3.10}$$

where according to Lemma 4, $a_p\neq 0$ when p is odd. Therefore, $\beta^*=a_p^{-1}$ is a nontrivial solution of (3.10) with $F'(\beta^*)\neq 0$ and the result follows. (ii) on the other hand, $a_p=0$ when p is even so that "higher order" terms must be retained in carrying out the branching analysis. For the case being considered the branching system (3.3) reduces to a single equation

$$0=-\eta\xi+(Q(\xi u_p+V(\xi,\eta))+\alpha^{-1}C(\xi u_p+V(\xi,\eta)), u_p), \tag{3.11}$$

where $\xi\in\mathbb{R}^1$ and $V\in\mathcal{N}^\perp$ satisfies the inequality (3.2) and the equation ($T=I-S$ denotes the orthogonal projection of $\mathcal{W}$ onto $\mathcal{N}^\perp$)

$$L_{\lambda_1}V-\alpha\eta AV+T[\alpha Q(\xi u_p+V)+C(\xi u_p+V)]=0. \tag{3.12}$$

Since L_{λ_1} has a bounded inverse on $\mathcal{N}^\perp$, V has the form

$$V=\xi^2 u+U \tag{3.13}$$

with u and U in $\mathcal{N}^\perp$ satisfying $U = O(|\xi|^3 + |\eta||\xi|^2)$ and

$$L_{\lambda_1} u = -\alpha TQ(u_p) = -\alpha Q(u_p), \tag{3.14}$$

where in the last equality we have used $TQ(u_p) = Q(u_p)$, a consequence of the condition $(Q(u_p), u_p) = 0$. Using (3.13) and the identity

$$Q(w+h) = Q(w) + Q'(h)w + Q(h), \qquad w, h \in \mathcal{W}, \tag{3.15}$$

we find that the branching equation (3.11) becomes

$$0 = -\eta\xi + \gamma\xi^3 + s(\xi, \eta), \tag{3.16}$$

where

$$\gamma = (Q'(u_p)u + \alpha^{-1}C(u_p), u_p) \tag{3.17}$$

and

$$s(\xi, \eta) = (\xi Q'(u_p)U + Q(V), u_p) + \alpha^{-1}(C(\xi u_p + V) - C(\xi u_p), u_p)$$

satisfies

$$\lim_{\xi \to 0} \xi^{-3} s(\xi, \eta) = O(\eta) \quad \text{for } |\eta| < \eta_0. \tag{3.18}$$

Instead of (3.4) the form of (3.16) suggests, for $\eta < 0$, the substitution

$$\xi = |\eta|^{1/2}\beta, \quad \text{for } \beta \in \mathbb{R}^1. \tag{3.19}$$

Using (3.19) and dividing by $|\eta|^{3/2}$ in (3.16) yields

$$0 = \beta + \gamma\beta^3 + |\eta|^{-3/2} s(|\eta|^{1/2}\beta, \eta). \tag{3.20}$$

Now letting $\eta \to 0^-$ in (3.20) and using (3.18) we obtain as the reduced branching equation the simple cubic equation

$$0 = \beta + \gamma\beta^3. \tag{3.21}$$

If $\gamma < 0$, equation (3.21) has two nontrivial solutions $\beta = \pm|\gamma|^{-1/2}$ yielding two distinct solutions of (3.20) for $-\eta_0 < \eta < 0$. These solutions, in turn, generate the desired branches (3.9); that these solutions are unstable again follows from standard arguments as in case (i). If $\gamma > 0$ then equation (3.21) has no nontrivial real solutions so that no branching takes place "to the left" in this situation. In an analogous way one sees when $0 < \eta < \eta_0$ that the substitution (3.19) leads to a reduced branching equation

$$0 = -\beta + \gamma\beta^3 \tag{3.22}$$

so that no branching "to the right" occurs when $\gamma < 0$ while two stable branches are obtained, corresponding to $\beta = \pm\gamma^{-1/2}$, when $\gamma > 0$.

Theorem 3 deals with the case in which the smallest eigenvalue of the narrow panel is simple. Now we discuss the only remaining possibility for narrow panels, namely, that λ_1 has multiplicity two; since a^2/b^2 then is necessarily a product of sucessive integers, we assume that $a^2/b^2 = m(m+1)$ so that λ_1 has eigenfunctions $u_m \equiv u_{m1}$ and u_{m+1}.

THEOREM 4. *If the panel is narrow and $a^2/b^2 = m(m+1)$ for some positive integer m, then there exist positive numbers δ and ρ such that the set of nontrivial solutions of* (*) *in $\mathcal{S}(\lambda_1, \rho, \delta)$ consists of exactly three buckled states which exist for $0 < |\lambda - \lambda_1| < \delta$.*

When m is odd *the solutions have the form*

$$\varphi = a_m \eta u_m + \Phi$$

and

$$\varphi^{\pm} = b_m \eta (u_m \pm c_m u_{m+1}) + \Phi^{\pm}, \qquad \eta = \alpha^{-1}(\lambda - \lambda_1),$$

where the constants a_m, b_m, c_m *are defined below and* Φ, $\Phi^{\pm}$ *are analytic functions of order* $O(\eta^2)$ *as* $\eta \to 0$.

When m is even *the solutions have the form*

$$\psi = d_m \eta u_{m+1} + \Psi$$

and

$$\psi^{\pm} = e_m \eta (u_{m+1} \pm f_m u_m) + \Psi^{\pm},$$

where the constants d_m, e_m, f_m *are defined below and* Ψ, $\Psi^{\pm}$ *are analytic functions of order* $O(\eta^2)$ *as* $\eta \to 0$.

Moreover, the solutions φ *and* ψ *are stable for* $\lambda_1 < \lambda < \lambda_1 + \delta$ *and unstable for* $\lambda_1 - \delta < \lambda < \lambda_1$, *whereas all of the solutions* $\varphi^{\pm}$, $\psi^{\pm}$ *are unstable for* $0 < |\lambda - \lambda_1| < \delta$.

Remark 2. If m is odd and large (i.e., if a/b is large) in the context of Theorem 4, then from the equality in (3.28) below we see that c/e is close to unity, hence so is the coefficient $c_m \equiv (2-(c/e))^{1/2}$. It follows that for η small the initial shapes of $\varphi^{\pm}$ are approximately those of $u_m \pm u_{m+1}$, i.e., of $\sin(\pi y/b)[\sin(m\pi x/a) \pm \sin((m+1)\pi x/a)]$. Similarly, if m is even and large, the shapes of $\psi^{\pm}$ are approximately those of $\sin(\pi y/b)[\sin((m+1)\pi x/a) \pm \sin(m\pi x/a)]$. Thus, for fixed y in $0<y<b$ and for fixed λ in $\lambda_1 < \lambda < \lambda_1 + \delta$, the buckled states φ^+ and ψ^+ have approximately the form shown in Fig. 1 as unstable state #1 while φ^- and ψ^- correspond to unstable state #2; the stable state φ (or ψ) has approximately the form $\sin(\pi y/b)\sin(m\pi x/a)$ (m odd) which is shown in Fig. 1 as the stable buckled state.

Proof of Theorem 3. Setting $\mu_p = \mu_{p1}$ for $p = 1, 2, \cdots$, we see from (2.14) that $\mu_m = \mu_{m+1} = [(2m+1)\pi/a]^2$ is the smallest characteristic value of A and has multiplicity two. Then the smallest eigenvalue of L_λ is $\lambda_1 = \mu_m + \alpha^2 \mu_m^{-1}$, and the null space $\mathcal{N} \equiv \mathcal{N}(L_{\lambda_1})$ is spanned by the eigenfunctions $v_1 = u_m$ and $v_2 = u_{m+1}$, which satisfy (3.1). In this case, the system (3.6) becomes

$$\begin{aligned} 0 &= -\beta_1 + c\beta_1^2 + 2d\beta_1\beta_2 + e\beta_2^2, \\ 0 &= -\beta_2 + d\beta_1^2 + 2e\beta_1\beta_2 + f\beta_2^2, \end{aligned} \tag{3.23}$$

where $c = (Q(v_1), v_1)$, $d = (Q(v_1), v_2) = \frac{1}{2}(Q'(v_1)v_2, v_1)$, $e = (Q(v_2), v_1)$ and $f = (Q(v_2), v_2)$. From Lemma 4, we see that $d = f = 0$ when m is odd and $c = e = 0$ when m is even, so that (3.23) reduces to

$$\begin{aligned} 0 &= -\beta_1 + c\beta_1^2 + e\beta_2^2, \\ 0 &= -\beta_2 + 2e\beta_1\beta_2 \qquad (m \text{ odd}), \end{aligned} \tag{3.24}$$

or

$$\begin{aligned} 0 &= -\beta_1 + 2d\beta_1\beta_2, \\ 0 &= -\beta_2 + d\beta_1^2 + f\beta_2^2 \qquad (m \text{ even}). \end{aligned} \tag{3.25}$$

If the inequality

$$0 < c < 2e \tag{3.26}$$

holds, then (3.24) has exactly three distinct nontrivial solutions, $(a_m, 0)$, $b_m(1, \pm c_m)$, where $a_m = c^{-1}$, $b_m = (2e)^{-1}$ and $c_m = (2-(c/e))^{1/2}$. The condition (3.26) also ensures that the Jacobian, J, of (3.24) satisfies $J \neq 0$ at each of these solutions. Thus, when m is odd, the existence of the three specified nontrivial solutions (φ and $\varphi^{\pm}$) and no others in a suitable set $\mathcal{S}(\lambda_1, \rho, \delta)$, follows as in the discussion just before Theorem 3. To obtain the stability results when m is odd, we consider the functional which is the restriction of τ to $\mathscr{E} = \{\beta_1 v_1 + \beta_2 v_2 \colon \beta_1^2 + \beta_2^2 = 1\}$, namely, $t(\beta_1, \beta_2) = \frac{1}{3}[c\beta_1^3 + 3e\beta_1\beta_2^2]$ for $\beta_1^2 + \beta_2^2 = 1$. When normalized to $\beta_1^2 + \beta_2^2 = 1$ the nontrivial solutions of (3.24) become $(1, 0)$, $(1, \pm c_m)/(1 + c_m^2)^{1/2}$ and are the points at which $t(\beta_1, \beta_2)$ has positive extrema. A calculation using (3.26) shows that t has a positive minimum at $(1, 0)$ and positive maxima at the two remaining points. Since the solution φ corresponds to the minimum point its stability properties are a consequence of Theorem 4 in [12], while the instability of the solutions $\varphi^{\pm}$ are obtained using Theorem 5 in [12] (the condition required in Theorem 5 of [12], that $(Q(w), w)$ not be an eigenvalue of $SQ'(w)$ when w maximizes t, is equivalent here to $J \neq 0$). The results for the case m even are obtained in a similar way if

$$0 < f < 2d \tag{3.27}$$

holds and the three nontrivial solutions of (3.25) are denoted by $(0, d_m)$, $e_m(\pm f_m, 1)$ where $d_m = f^{-1}$, $e_m = (2d)^{-1}$ and $f_m = (2-(f/d))^{1/2}$. We complete the proof of Theorem 4 by verifying (3.26); similar steps will yield (3.27) when m is even. Since $c = (Q(u_m), u_m)$ and $e = (Q(u_{m+1}), u_m)$, when m is odd and $a^2/b^2 = m(m+1)$, Lemma 4 yields

$$c = \frac{32(a/b)^{1/2}}{\pi b^2 \mu_m m^2} > 0, \qquad e = \frac{32(a/b)^{1/2}(3m^2+4m+2)}{\pi b^2 \mu_m m^2 (3m^2+8m+4)}$$

so that, for $m \geqq 1$,

$$2e/c = 2(3m^2+4m+2)/(3m^2+8m+4) \tag{3.28}$$

from which it follows that $2e/c > 1$.

Remark 3. Similar methods to those used in the proof of Theorem 4 may apply at other eigenvalues of a narrow panel (and even at some eigenvalues of a panel which is not narrow) to give the existence of unstable buckled states. For example, if, for some λ_0, L_{λ_0} has a two-dimensional null space spanned by u_{pq} and u_{mn}, where at least one of the products mn or pq is an odd number, then the methods of [12] yield the existence of at least one buckled state which branches from the trivial solution at $\lambda = \lambda_0$ and is unstable for λ on both sides of λ_0.

Appendix. We establish in this appendix the properties stated in Remark 1 for the coefficient γ defined by (3.17). Using the symmetry properties of the operator B (see Lemma 3) and the Euler identity $Q'(w)w = 2Q(w)$ (see (2.18) with $u = w$), we may rewrite (3.17) as

$$\gamma = 2(Q(u_p), u) + \alpha^{-1}(C(u_p), u_p). \tag{A.1}$$

Expanding $u = \sum_{m,n=1}^{\infty} \Gamma_{mn} w_{mn}$ and $Q(u_p) = \sum_{m,n=1}^{\infty} Q_{mn} w_{mn}$ in terms of the eigenfunctions of A (see Lemma 1), we see that $\Gamma_{mn} = (u, w_{mn})$, $Q_{mn} = (Q(u_p), w_{mn})$ and, since u and $Q(u_p)$ belong to $\mathcal{N}^{\perp}$ in the case being considered, $\Gamma_{p1} = Q_{p1} = 0$; in fact, by Lemma 4 we have $Q_{mn} = 0$ unless both m and n are odd. A simple calculation gives

$$L_{\lambda_1} u = \sum_{m,n=1}^{\infty} \Gamma_{mn} \Lambda_{mn}^{-1} w_{mn}, \tag{A.2}$$

where

$$\Lambda_{mn} = \left(1 - \frac{\lambda_1}{\mu_{mn}} + \frac{\alpha^2}{\mu^2_{mn}}\right)^{-1}, \quad \text{for } (m, n) \neq (p, 1). \tag{A.3}$$

Substituting these expansions into (3.14) and equating coefficients yields

$$\Gamma_{mn} = -\alpha Q_{mn} \Lambda_{mn}, \quad \text{for } (m, n) \neq (p, 1), \tag{A.4}$$

so that

$$(Q(u_p), u) = \sum_{m,n \text{ odd}} Q_{mn} \Gamma_{mn} = -\sum_{m,n \text{ odd}} \alpha Q^2_{mn} \Lambda_{mn}, \tag{A.5}$$

where the sums are extended over the set of pairs (m, n) of odd positive integers. Similarly, if we expand $B(u_p, u_p)$ as $B(u_p, u_p) = \sum_{m,n \text{ odd}} B_{mn} w_{mn}$, where $B_{mn} = (B(u_p, u_p), w_{mn})$, we see that

$$(C(u_p), u_p) = \frac{1}{2} \|B(u_p, u_p)\|^2 = \frac{1}{2} \sum_{m,n \text{ odd}} B^2_{mn} \tag{A.6}$$

and, from (2.12),

$$\begin{aligned} Q_{mn} &= (Q(u_p), w_{mn}) = (B(u_p, Au_p) + \tfrac{1}{2} AB(u_p, u_p), w_{mn}) \\ &= \left(\frac{1}{\mu_p} + \frac{1}{2\mu_{mn}}\right) B_{mn}. \end{aligned} \tag{A.7}$$

Using (A.5), (A.6) and (A.7), we can now express the constant γ in (A.1) as the sum

$$\gamma = (2\alpha)^{-1} \sum_{m,n \text{odd}} B^2_{mn} (1 - \sigma_{mn}), \tag{A.8}$$

where

$$\sigma_{mn} = 4\alpha^2 \left(\frac{1}{\mu_p} + \frac{1}{2\mu_{mn}}\right)^2 \Lambda_{mn}. \tag{A.9}$$

If we let s denote the constant $\alpha \mu_p^{-1}$ and define

$$h(r) = \frac{(r + 2s)^2}{(1 - rs)[1 - (r/s)]}, \qquad r \geqq 0, \quad r \neq s, \quad r \neq s^{-1}, \tag{A.10}$$

then (A.3), (A.9) and the relation $\lambda_1 = \mu_p + \alpha^2 \mu_p^{-1}$ imply that $\sigma_{mn} = h(\alpha \mu_{mn}^{-1})$. From (A.8) we see that $\gamma < 0$ if $h(r) > 1$ whenever $r = \alpha \mu_{mn}^{-1}$ and we seek first to ensure this condition by suitably restricting the parameter $\theta = b\sqrt{\alpha}/(2\pi)$.

Since we are assuming that μ_p is the characteristic value of A which minimizes $\lambda(\mu) = \mu + \alpha^2 \mu^{-1}$, and since $\lambda(\mu_p) = \lambda(\alpha^2 \mu_p^{-1})$, no characteristic value of A lies between μ_p and $\alpha^2 \mu_p^{-1}$. Furthermore, the restriction to a narrow panel implies that all other characteristic values of A exceed μ_p. Consequently, $r < \min(s, s^{-1})$ when $r = \alpha \mu_{mn}^{-1}$. On the interval $0 \leqq r < \min(s, s^{-1})$, $h(r)$ increases from $4s^2$ at $r = 0$ to a vertical asymptote at $r = \min(s, s^{-1})$ so that $h(r) > 1$ if $4s^2 \geqq 1$, i.e., if $\mu_p \leqq 2\alpha$. In terms of θ these conditions become $\theta_1 \leqq \theta \leqq 1$, where

$$\theta_1 = \theta_1(p) \equiv b[\mu_p/(8\pi^2)]^{1/2} = \frac{1}{2\sqrt{2}} \frac{b}{a} \left(p + \frac{a^2}{pb^2}\right). \tag{A.11}$$

Since p here is the positive integer minimizing $\theta_1(q)$, p lies in the unit interval $(t, t+1)$, where $t = \frac{1}{2}(-1 + \sqrt{1 + 4(a^2/b^2)})$ is such that $\theta_1(t) = \theta_1(t+1)$. Note that, since p is even, we must have $t>1$ so that $a/b>(2)^{1/2}$ in this appendix. An elementary calculation then yields

$$2^{-1/2}\leqq \theta_1(p)\leqq \theta_1(t)=2^{-1/2}[1+b^2/(4a^2)]^{1/2},$$

so that $\theta_1(p)<3/4$ when $(a/b)>(2)^{1/2}$. Thus, the condition $\theta_1(p)\leqq\theta\leqq 1$ is not void for the class of narrow panels. For example, if $a/b=4$ then $\theta_1(t)\leqq 0.72$ so that $\gamma<0$ when $0.72\leqq\theta\leqq 1$.

Next we observe that $\gamma>0$ if $h(r)\leqq 1$ whenever $r=\alpha\mu_{mn}^{-1}$ with $(m, n)\neq(p, 1)$. Again we may restrict r to the interval $0\leqq r<\min(s, s^{-1})$, on which h is increasing, so that $h(r)\leqq h(s_1)$, where $s_1=\alpha/\mu_{m_1n_1}$ and $\mu_{m_1n_1}$ is the second (in order of increasing magnitude) characteristic value of A. An elementary analysis of the quantity $(m+a^2n^2/(b^2m))$, for $a/b\geqq 1$ and positive integer values of m and n, reveals that $n_1=1$ and (by comparing the values of $(m+a^2/(b^2m))$ at $m=p+1$ and $m=p-1$) $m_1=p+1$ [resp., $m_1=p-1$] when $t<p\leqq(1+a^2/b^2)^{1/2}$ [resp., $(1+a^2/b^2)^{1/2}\leqq p<t+1$] with t defined as in the above. The remainder of the discussion is given only for the case $m_1=p+1$, since the case $m_1=p-1$ is similar. First of all, from (A.10), it is easy to show that the condition $h(r)\leqq 1$ is equivalent to

$$5rs+4s^2\leqq 1-r/s. \tag{A.12}$$

If we set $r=s_1=\alpha/\mu_{p+1}$, $s=\alpha/\mu_p$ and introduce the parameter θ, then (A.12) may be expressed as $\theta^4\leqq\theta_2^4$, where

$$\theta_2^4=\{1-(\mu_p/\mu_{p+1})\}\left\{16\frac{a^4}{b^4}\left(p+\frac{a^2}{b^2p}\right)^{-2}\left[4\left(p+\frac{a^2}{b^2p}\right)^{-2}+5\left((p+1)+\frac{a^2}{b^2(p+1)}\right)^{-2}\right]\right\}^{-1}. \tag{A.13}$$

Since μ_p is a simple characteristic value of A, $\mu_p<\mu_{p+1}$ so that $\theta_2>0$ and the condition $0<\theta\leqq\theta_2$ is not void; for example, if $a/b=4$ then $p=4$, $m_1=5$ and $\theta_2>0.27$, so that $\gamma>0$ when $0\leqq\theta\leqq 0.27$. Finally, let us note that, if θ_1 is defined as in (A.11), the condition $\theta_2<\theta_1$ is also satisfied so that $0<\theta_2<\theta_1<1$ for all p being considered. It is possible that a more refined analysis would yield a positive constant θ^* such that $\gamma>0$ when $0<\theta<\theta^*$ and $\gamma<0$ when $\theta^*<\theta\leqq 1$.

REFERENCES

[1] M. S. BERGER, *The existence of equilibrium states of thin elastic shells (I)*, Indiana Univ. Math. J., 20 (1971), pp. 591–602.

[2] M. S. BERGER, AND P. C. FIFE, *von Kármán's equations and the buckling of a thin elastic plate II. Plate with general edge conditions*, Comm. Pure Appl. Math., 21 (1968), pp. 227–241.

[3] G. H. KNIGHTLY AND D. SATHER, *Nonlinear buckled states of rectangular plates*, Arch. Rational Mech. Anal., 54 (1974), pp. 356–372.

[4] ———, *Nonlinear axisymmetric buckled states of shallow spherical caps*, this Journal, 6 (1975), pp. 913–924.

[5] ———, *Existence and stability of axisymmetric buckled states of spherical shells*, Arch. Rational Mech. Anal., 63 (1977), pp. 305–319.

[6] W. T. KOITER, *Buckling and postbuckling of a cylindrical panel under axial compression*, NLR Report S 476, Amsterdam, 1956; see also: Reports and Transactions, National Aero. Res. Inst. 20.

[7] ———, *The effect of axisymmetric imperfections on the buckling of cylindrical shells under axial compression*, Proc. Kon. Ned. Akad. Wet. Amsterdam Ser. B, 66 (1963), pp. 265–279.

[8] J. MALLET-PARET, *Buckling of plates with small cylindrical imperfections*, preprint, 1976.

[9] P. H. RABINOWITZ, *A note on topological degree for potential operators*, J. Math. Anal. Appl., 51 (1975), pp. 483–492.

[10] D. SATHER, *Branching of solutions of nonlinear equations*, Rocky Mountain J. Math., 3 (1973), pp. 203–250.

[11] ———, *Branching and stability for nonlinear shells*, Proceedings of the IUTAM/IMU Symposium on Applications of Methods of Functional Analysis to Problems of Mechanics (Marseille, 1975), Lecture Notes in Mathematics, No. 503, Springer-Verlag, New York, 1976.

[12] ———, *Bifurcation and stability for a class of shells*, preprint, 1976, Arch. Rational Mech. Anal., 63 (1977), pp. 295–304.

[13] D. H. SATTINGER, *Stability of bifurcating solutions by Leray–Schauder degree*, Arch. Rational Mech. Anal., 43 (1971), pp. 154–166.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0017 $01.00/0

SOLVABILITY OF A CLASS OF HILBERT NETWORKS*

VACLAV DOLEZAL†

Abstract. In this paper we give sufficient conditions for solvability of a Hilbert network some elements of which are described by monotone operators defined only on subsets of the underlying Hilbert space. As a special case we consider a *finite* nonlinear LRC-network, whose inductors are linear, time-varying.

1. Introduction. A Hilbert network $\hat{\mathcal{N}}$ is called solvable, if for any excitation by EMF and/or current sources there exists in $\hat{\mathcal{N}}$ a current distribution obeying Kirchhoff laws. The question, whether a given $\hat{\mathcal{N}}$ is solvable or not, is of crucial importance. Known effective results concerning solvability [1, Thms. 4, 5] make the assumption that the operators describing network elements are defined on the entire underlying Hilbert space $\mathcal{H}$. Consequently, these results cannot be applied, if the network contains inductors, because in this case the definition domain of operators describing inductors has to be restricted to the space of absolutely continuous functions or the like.

This paper attempts to fill this gap. It turns out that, if monotone operators are involved, a surprisingly simple, yet powerful, theorem can be proved. This result has the following interesting interpretation: If all elements in a network $\hat{\mathcal{N}}$ which are described by operators defined on the entire space are removed and a unit resistor is inserted in every branch, and if the network $\hat{\mathcal{N}}_1$ thus obtained is solvable, then $\hat{\mathcal{N}}$ is also solvable.

Based on this, we then give solvability conditions for an L-proper LT-network—a concept slightly more general than a finite nonlinear LRC-network whose inductors are linear, time-varying.

2. Results. To avoid unnecessary repetition of definitions and restatement of basic theorems, we refer the reader to the survey paper [1]. We will consistently use the notation introduced there.

THEOREM 1. *Let H be a real Hilbert space, let G be a locally finite oriental graph having $c_2 \leqq \aleph_0$ branches, and let $D \subset H^{c_2}$, $N_{\hat{a}} \cap D \neq \emptyset$. Assume that*

(i) *$\hat{Z}_1: D \to H^{c_2}$ is a monotone operator,*

(ii) *$\hat{Z}_2: H^{c_2} \to H^{c_2}$ is a hemicontinuous operator such that*

$$\langle \hat{Z}_2 x_1 - \hat{Z}_2 x_2, x_1 - x_2 \rangle \geqq c\|x_1 - x_2\|^p \tag{1}$$

for all $x_1, x_2 \in H^{c_2}$ with some fixed $c > 0$ and $p > 1$,

(iii) *the Hilbert network $\hat{\mathcal{N}} = (\hat{Z}_1 + I, G)$ possesses a solution for any $e \in H^{c_2}$ (I is the identity operator on H^{c_2}).*

Then, for any $e \in H^{c_2}$, the network $\hat{\mathcal{N}} = (\hat{Z}_1 + \hat{Z}_2, G)$ possesses a unique solution in D corresponding to e. Moreover, the admittance $A: H^{c_2} \to D$ of $\hat{\mathcal{N}}$ satisfies the inequality

$$\|Ae_1 - Ae_2\| \leqq c^{-1/(p-1)}\|\hat{X}^*(e_1 - e_2)\|^{1/(p-1)} \tag{2}$$

for all $e_1, e_2 \in H^{c_2}$.

Proof. Let the subset $F \subset H^{c_0}$ be defined by $F = \hat{X}^*(N_{\hat{a}} \cap D)$, and define the operator $W_1: F \to H^{c_0}$ by $W_1 = \hat{X}^* \hat{Z}_1 \hat{X}$. Clearly, W_1 is monotone on F, since for any $z_1, z_2 \in F$ we have

$$\langle W_1 z_1 - W_1 z_2, z_1 - z_2 \rangle = \langle \hat{Z}_1 \hat{X} z_1 - \hat{Z}_1 \hat{X} z_2, \hat{X} z_1 - \hat{X} z_2 \rangle \geqq 0. \tag{3}$$

---

* Received by the editors May 10, 1977, and in revised form March 14, 1978.

† Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York 11794. This research was supported by the National Science Foundation under Grant MPS7505268.

Moreover, the assumption (iii) implies by Theorem 2 in [1] that

$$\hat{X}^*(\hat{Z}_1+I)\hat{X}F = H^{c_0}, \qquad \text{i.e.,} \quad (W_1+I)F = H^{c_0}.$$

Consequently, Theorem 2 in [2] shows that W_1 is maximal monotone.

Next, let W_2: $H^{c_0} \to H^{c_0}$ be defined by $W_2 = \hat{X}^*\hat{Z}_2\hat{X}$. Then (1) yields for any $z_1, z_2 \in H^{c_0}$,

$$(4) \qquad \langle W_2z_1 - W_2z_2, z_1 - z_2\rangle = \langle \hat{Z}_2\hat{X}z_1 - \hat{Z}_2\hat{X}z_2, \hat{X}z_1 - \hat{X}z_2\rangle$$
$$\geqq c\|\hat{X}(z_1-z_2)\|^p = c\|z_1-z_2\|^p.$$

(See [1, Prop. 1].) Thus, W_2 is monotone on H^{c_0}. Since $\hat{X}^*$, $\hat{X}$ are linear bounded operators, it follows that W_2 is hemicontinuous on H^{c_0}. Hence, W_2 is maximal monotone.

On the other hand, since $F \cap \operatorname{Int} H^{c_0} = F \neq \emptyset$ and both W_1 and W_2 are maximal monotone, it follows by Theorem 1 in [3] that $W = (W_1 + W_2)$: $F \to H^{c_0}$ is also maximal monotone.

Moreover, (3) and (4) imply that

$$(5) \qquad \langle Wz_1 - Wz_2, z_1 - z_2\rangle \geqq c\|z_1 - z_2\|^p$$

for all $z_1, z_2 \in F$. Thus, for any $z \in F$ we have by the Schwartz inequality

$$(6) \qquad \|Wz\| \geqq c\|z\|^{p-1} - \|W0\|.$$

Hence, if $z_n \in F$ and $\|z_n\| \to \infty$ as $n \to \infty$, we have $\|Wz_n\| \to \infty$. Consequently, by Theorem 5 in [2], W is surjective, i.e., $WF = H^{c_0}$. However, this means (Theorem 2 in [1]) that, for any $e \in H^{c_2}$, $\hat{\mathcal{N}}$ possesses a solution corresponding to e.

Finally, (5) shows that W is 1-to-1, and that the inverse W^{-1}: $H^{c_0} \to F$ satisfies the condition

$$(7) \qquad \|W^{-1}x_1 - W^{-1}x_2\| \leqq c^{-1/(p-1)}\|x_1 - x_2\|^{1/(p-1)}$$

for all $x_1, x_2 \in H^{c_0}$. Since the admittance A: $H^{c_2} \to D$ of $\hat{\mathcal{N}}$ is given by $A = \hat{X}W^{-1}\hat{X}^*$ and $\hat{X}$ is a norm-preserving isomorphism [1 Prop. 1], inequality (2) follows and, hence, the proof.

Remark 1. The above proof suggests that (i), (ii) can be replaced by the following, weaker assumptions:

(i)* W_1 is monotone,

(ii)* W_2 is hemicontinuous and satisfies (4) for all $z_1, z_2 \in H^{c_0}$.

Let us now consider the generalization of an LRC-network we mentioned in the introduction.

From now on, let H be the real space $L_2[0, \tau]$, $0 < \tau < \infty$, (we will also write L_2 for brevity), and let

$$(8) \qquad K = \{x : x \text{ absolutely continuous on } [0, \tau],\ x' \in L_2\}.$$

If a is a real number, let $K_a = \{x : x \in K, x(0) = a\}$.

Let G be a finite oriented graph having $c_2 < \aleph_0$ branches, let $1 \leqq k \leqq c_2$ and let d be the incidence matrix of G, [1].

A vector $j = [j_1, j_2, \cdots, j_k]^T \in R^k$ will be called admissible, if there exist numbers $j_{k+1}, j_{k+2}, \cdots, j_{c_2}$ such that the c_2-vector $j_0 = [j_1, j_2, \cdots, j_{c_2}]$ satisfies the equation

$$(9) \qquad d^Tj_0 = 0.$$

Let $\tilde{L}(\xi, t)$: $R^{c_2} \times [0, \tau] \to R^{c_2}$ be such that, for any $x \in K^k \times L_2^{c_2-k} \subset L_2^{c_2}$ we have $\tilde{L}(x(t), t) \in K^{c_2}$.

If $j \in R^k$ is admissible, let $D_j = K_{j_1} \times K_{j_2} \times \cdots \times K_{j_k} \times L_2^{c_2-k}$ and let the (inductance) operator $\hat{L}_j\colon D_j \to L_2^{c_2}$ be defined by

(10) $$(\hat{L}_j x)(t) = \{\tilde{L}(x(t), t)\}'.$$

Finally, let $\hat{T}\colon L_2^{c_2} \to L_2^{c_2}$ be an operator. Then the network $\hat{\mathcal{N}} = (\hat{L}_j + \hat{T}, G)$ will be called an LT-network having initial condition j.

Observe that a general LRC-network is a special case of an LT-network. Indeed, let $\tilde{R}(\xi, t), \tilde{S}(\xi, t)\colon R^{c_2} \times [0, \tau] \to R^{c_2}$ be functions such that $\tilde{R}(x(t), t), \tilde{S}(\int_0^t x(\sigma)\, d\sigma, t) \in L_2^{c_2}$ whenever $x \in L_2^{c_2}$.

If we define $\hat{T}$ by

(11) $$(\hat{T}x)(t) = \tilde{R}(x(t), t) + \tilde{S}\left(\int_0^t x(\sigma)\, d\sigma, t\right),$$

then clearly $(\hat{L}_j + \hat{T}, G)$ will be a model of a nonlinear LRC-network. Observe that if $i \in D_j$ is a solution of $(\hat{L}_j + \hat{T}, G)$ corresponding to some $e \in L_2^{c_2}$, then each component i_m, $m = 1, 2, \cdots, k$ of i is absolutely continuous and satisfies the condition $i_m(0) = j_m$.

In order to define an L-proper LT-network, let us introduce the following concepts:

A nonzero vector $\xi = [\xi_n] \in R^{c_2}$ will be called a loop, if $d^T\xi = 0$ and each element ξ_n attains one of the values $-1, 0, 1$. We will say that a loop $\xi = [\xi_n]$ contains (does not contain) a branch b_m of G, if $\xi_m \neq 0$, $(\xi_m = 0)$.

Definition. Let G be a finite oriented graph having c_2 branches, let $1 \leqq k \leqq c_2$, and let the following conditions be satisfied:

(i) there exist loops $\xi^1, \xi^2, \cdots, \xi^k$ such that, for each $m = 1, 2, \cdots, k$, the loop ξ^m contains b_m and does not contain any other branch in the set $\{b_1, b_2, \cdots, b_k\}$.

(ii) there exists a $k \times k$ matrix $l(t)$ having a continuous derivative $l'(t)$ on $[0, \tau]$ with $l(t)$ and $l'(t)$ being positive definite and positive semidefinite for each $t \in [0, \tau]$, respectively, such that

(12) $$\tilde{L}(\xi, t) = L(t)\xi,$$

where the $c_2 \times c_2$ matrix $L(t)$ is given by

(13) $$L(t) = \left[\begin{array}{c|c} l(t) & 0 \\ \hline 0 & 0 \end{array}\right].$$

If $\hat{L}_j\colon D_j \to L_2^{c_2}$ is given by (10), then the LT-network $\hat{\mathcal{N}} = (\hat{L}_j + \hat{T}, G)$ will be called L-proper.

Clearly, $\hat{\mathcal{N}}$ being L-proper means that all inductors in $\hat{\mathcal{N}}$ (and possible mutual couplings) are passive, linear, time-varying and nondecreasing, and are confined only to the branches $b_1, b_2, \cdots, b_k$.

Also, observe that if $\hat{\mathcal{N}}$ is L-proper, then any $j \in R^k$ is admissible. Indeed, let $\xi^1, \xi^2, \xi^3, \cdots, \xi^k$ be the loops existing by (i). Select vectors $\xi^{k+1}, \xi^{k+2}, \cdots, \xi^{c_0} \in R^{c_2}$ so that $\{\xi^1, \xi^2, \cdots, \xi^{c_0}\}$ is a basis in the solution space $N_{\tilde{a}}$ of $d^T\xi = 0$ (obviously, $k \leqq c_0 = \dim N_{\tilde{a}}$), and let $Y = [\xi^1 \mid \xi^2 \mid \xi^3 \mid \cdots \mid \xi^{c_0}]$. Then Y has the form

(14) $$Y = \left[\begin{array}{c|c} I & Y_1 \\ \hline Y_2 & Y_3 \end{array}\right],$$

where matrices I, Y_1, Y_2, Y_3 are of type $k \times k$, $k \times (c_0 - k)$, $(c_2 - k) \times k$ and $(c_2 - k) \times (c_0 - k)$, respectively. Now, a vector $\xi \in R^{c_2}$ satisfies $d^T\xi = 0$, iff $\xi = Yw$ for some $w \in R^{c_0}$. Thus, if $j \in R^k$ is given, we can put $w = [j^T \mid 0, 0, \cdots, 0]^T \in R^{c_0}$ and let

$j_0 = Yw$. Then $d^T j_0 = 0$, and the first k components of j_0 coincide with those of j, i.e., j is admissible.

We will need the following result [4], [5]:

LEMMA 1. *Assume that*

(i) *$A(t)$ is a symmetric $n \times n$ matrix having a continuous derivative $A'(t)$ on* $[0, \tau]$ *such that* rank $A(t) = r$ *for each $t \in [0, \tau]$ with a fixed $r \leqq n$;*

(ii) *$H(t, \sigma)$ is a symmetric $n \times n$ matrix defined and continuous on $\Omega = \{(t, \sigma): 0 \leqq \sigma \leqq t \leqq \tau\}$ which has a continuous derivative $\partial/\partial t H(t, \sigma)$ on Ω;*

(iii) *$A(t) + H(t, t)$ is positive definite for every $t \in [0, \tau]$;*

(iv) *$f \in K^n$;*

(v) *there exists $\xi \in R^n$ such that $A(0)\xi = f(0)$.*

Then there exists a unique $x \in L_2^n$ such that

$$A(t)x(t) + \int_0^t H(t, \sigma)x(\sigma)\, d\sigma = f(t), \qquad t \in [0, \tau]. \tag{15}$$

(Note that in [4] and [5] this result is stated for L_1^n rather than for L_2^n. However, retracing the proof given in [5] we can easily confirm that the assertion holds for L_2^n, too.)

THEOREM 2. *Let G be a finite oriented graph having c_2 branches, let $1 \leqq k \leqq c_2$, and let $j \in R^k$. Let $\hat{T}: L_2^{c_2} \to L_2^{c_2}$ be a hemicontinuous operator such that, for some $c > 0$ and $p > 1$, we have*

$$\langle \hat{T}x_1 - \hat{T}x_2, x_1 - x_2 \rangle \geqq c\|x_1 - x_2\|^p \tag{16}$$

for all $x_1, x_2 \in L_2^{c_2}$.

Furthermore, let $\hat{\mathcal{N}} = (\hat{L}_j + \hat{T}, G)$ be an L-proper LT-network, where $\hat{L}_j: D_j \to L_2^{c_2}$ is defined by (9), (10), (12) *and* (13).

Then for any $e \in L_2^{c_2}$, $\hat{\mathcal{N}}$ possesses a unique solution $i \in D_j$ corresponding to e, i.e., if $i = [i_m]$, i_m is absolutely continuous, $i'_m \in L_2$, $i_m(0) = j_m$ for $m = 1, 2, \cdots, k$, and $i_m \in L_2$ for $m = k+1, k+2, \cdots, c_2$.

Moreover, the admittance A of $\hat{\mathcal{N}}$ satisfies the inequality

$$\|Ae_1 - Ae_2\| \leqq c^{-1/(p-1)} \|\hat{X}^*(e_1 - e_2)\|^{1/(p-1)} \tag{17}$$

for all $e_1, e_2 \in L_2^{c_2}$.

Also, (16) *can be replaced by the weaker condition*

$$\langle Wz_1 - Wz_2, z_1 - z_2 \rangle \geqq c\|z_1 - z_2\|^p \tag{16*}$$

for all $z_1, z_2 \in L_2^{c_0}$, where $W: L_2^{c_0} \to L_2^{c_0}$ is defined by $W = \hat{X}^* \hat{T} \hat{X}$.

Proof. Referring to Theorem 1, it suffices to show that (I) $\hat{L}_j$ is monotone on D_j, (II) the network $\hat{\mathcal{N}} = (\hat{L}_j + I, G)$ has a solution for any $e \in L_2^{c_2}$.

As for (I), observe this fact: if $x \in D_j$ and $\bar{x}$ is a k-vector containing the first k components of x, then $\bar{x} \in K^k$ and we have by (13), $x^T\{Lx\}' = \bar{x}^T\{l\bar{x}\}'$. Thus, choosing $x_1, x_2 \in D_j$, we have by (10), (12),

$$\begin{aligned} J = \langle \hat{L}_j x_1 - \hat{L}_j x_2, x_1 - x_2 \rangle &= \int_0^\tau (x_1 - x_2)^T \{L(x_1 - x_2)\}'\, d\sigma \\ &= \int_0^\tau (\bar{x}_1 - \bar{x}_2)^T \{l(\bar{x}_1 - \bar{x}_2)\}'\, d\sigma \\ &= [(\bar{x}_1 - \bar{x}_2)^T l(\bar{x}_1 - \bar{x}_2)]_0^\tau - \int_0^\tau (\bar{x}'_1 - \bar{x}'_2)^T l(\bar{x}_1 - \bar{x}_2)\, d\sigma. \end{aligned}$$

On the other hand,

$$J=\int_0^\tau (\bar{x}_1-\bar{x}_2)^T l'(\bar{x}_1-\bar{x}_2)\,d\sigma+\int_0^\tau (\bar{x}_1-\bar{x}_2)^T l(\bar{x}_1'-\bar{x}_2')\,d\sigma.$$

Since $\bar{x}_1(0)=\bar{x}_2(0)=j$ and l is symmetric, it follows that

$$J=\tfrac{1}{2}(\bar{x}_1-\bar{x}_2)^T(\tau)l(\tau)(x_1-x_2)(\tau)+\frac{1}{2}\int_0^\tau (\bar{x}_1-\bar{x}_2)^T l'(\bar{x}_1-\bar{x}_2)\,d\sigma.$$

Hence, by (ii) in the Definition, $J\geqq 0$, i.e., $\hat{L}_j$ is monotone on D_j.

To prove (II), choose a $c_2\times c_0$ matrix X whose columns constitute an orthonormal basis in the solution space of $d^T\xi=0$, $\xi\in R^{c_2}$. First, we are going to show that

$$\text{rank}\, X^T L(t) X = k \tag{18}$$

for each $t\in[0,\tau]$. Indeed, if Y is the matrix defined by (14), then there exists a nonsingular $c_0\times c_0$ matrix Q such that $X=YQ$. Thus, $X^TLX=Q^TY^TLYQ$, so that rank $X^TLX=$ rank Y^TLY. On the other hand, an easy calculation shows that

$$Y^TLY=\left[\begin{array}{c|c} l & lY_1 \\ \hline Y_1^Tl & Y_1^TlY_1\end{array}\right].$$

Since l is nonsingular for each $t\in[0,\tau]$, it follows that rank $Y^TLY\geqq k$. Also, since rank $Y=c_0\geqq k$, we have rank $Y^TLY\leqq \min[c_0, k, c_0]=k$, which proves our claim.

Next, choose a $j\in R^k$, construct $j_0\in R^{c_2}$ with $d^Tj_0=0$ and find $w_0\in R^{c_0}$ such that $j_0=Xw_0$. (See the paragraph preceding Lemma 1.) Also, pick $e\in L_2^{c_2}$ and consider the equation

$$X^TL(t)Xw(t)+\int_0^t w(\sigma)\,d\sigma=X^T\Big(L(0)j_0+\int_0^t e(\sigma)\,d\sigma\Big). \tag{19}$$

It is clear that the right hand side $f(t)$ of (19) is in K^{c_0}. Also, putting $A(t)=X^TL(t)X$, $H(t,\sigma)=I$, we see easily that the assumptions (i) through (iv) in Lemma 1 are satisfied. Moreover, $f(0)=X^TL(0)j_0=X^TL(0)Xw_0$, so that (v) is satisfied for $\xi=w_0$, too. Hence, there exists a unique $w\in L_2^{c_2}$ satisfying (19) in $[0,\tau]$.

Now, put $i=Xw\in L_2^{c_2}$. Since $X^TX=I$, (19) can be written as

$$X^TL(t)i+X^T\int_0^t i(\sigma)\,d\sigma=X^T\Big(L(0)j_0+\int_0^t e(\sigma)\,d\sigma\Big). \tag{20}$$

Multiplying (20) by a nonsingular $c_0\times c_0$ matrix Q' such that $Q'X^T=Y^T$ with Y being defined by (14), we obtain an equation like (20), where X^T is replaced by Y^T. Define $i_1\in L_2^k$ and $i_2\in L_2^{c_0-k}$ by $i=[i_1^T\,|\,i_2^T]^T$, and $e_1\in L_2^k$, $e_2\in L_2^{c_0-k}$ by $e=[e_1^T\,|\,e_2^T]^T$. Since

$$Y^TL(t)i=\left[\begin{array}{c} li_1 \\ \hline Y_1^Tli_1\end{array}\right],\qquad Y^Ti=\left[\begin{array}{c} i_1+Y_2^Ti_2 \\ \hline Y_i^Ti_1+Y_3^Ti_2\end{array}\right],$$

and

$$Y^TL(0)j_0=\left[\begin{array}{c} l(0)j \\ \hline Y_1^Tl(0)j\end{array}\right],$$

our equation can be split into the following two equations:

$$li_1+\int_0^t (i_1+Y_2^Ti_2)\,d\sigma=l(0)j+\int_0^t (e_1+Y_2^Te_2)\,d\sigma, \tag{21}$$

$$Y_1^Tli_1+\int_0^t (Y_1^Ti_1+Y_3^Ti_2)\,d\sigma=Y_1^Tl(0)j+\int_0^t (Y_1^Te_1+Y_3^Te_2)\,d\sigma. \tag{22}$$

However, because $l(t)$ is nonsingular and continuously differentiable on $[0, \tau]$, (21) shows that $i_1 \in K^k$. Moreover, setting $t=0$ in (21), we get $l(0)i_1(0) = l(0)j$, i.e., $i_1(0) = j$. Hence, we have shown that $i \in D_j$.

Furthermore, since $L(t)i = \left[\frac{li_1}{0}\right]$, we see that $L(t)i$ is differentiable almost everywhere. Thus, (20) yields

$$X^T\{L(t)i\}' + X^T i = X^T e(t), \tag{23}$$

i.e.,

$$\hat{X}^*(\hat{L}_j + I)i = \hat{X}^* e. \tag{24}$$

These results show that i is the solution of $\hat{\mathcal{N}}_1 = (\hat{L}_j + I, G)$ corresponding to $e \in L_2^{c_2}$. Indeed, since $\hat{X}\hat{X}^*$ is the orthogonal projection from $L_2^{c_2}$ onto $N_{\hat{a}}$ (see [1]), (24) implies that $\hat{X}\hat{X}^*\{(\hat{L}_j + I)i - e\} = 0$, i.e.,

$$(\hat{L}_j + I)i - e \in N_{\hat{a}}^{\perp}. \tag{25}$$

Moreover, since $i = \hat{X}w \in N_{\hat{a}}$, we have

$$i \in N_{\hat{a}} \cap D_j. \tag{26}$$

However, relations (25) and (26) define a solution; thus, $\hat{\mathcal{N}}_1$ is solvable for any $e \in L_2^{c_2}$, and our proof is completed. The last assertion follows from Remark 1.

Remark 2. The solvability of $\hat{\mathcal{N}}_1$ may also be proved by using the state variable technique. Actually, it can be easily seen that condition (i) in the Definition means that $\hat{\mathcal{N}}_1$ does not contain any "inductor-only cut-sets" [6, p. 199], and consequently, the current distribution in $\hat{\mathcal{N}}_1$ can be described by a canonic system of differential equations.

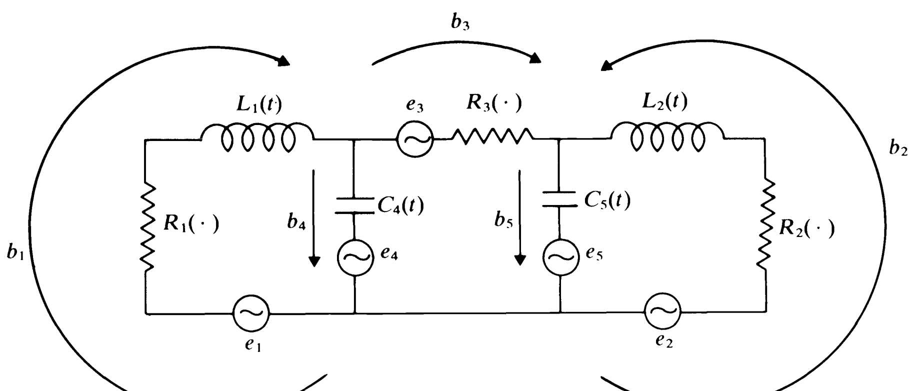

FIG. 1

To illustrate the application of Theorem 2, let us present a simple example. Consider the LRC-network $\mathcal{N}$ given in Fig. 1. Assume that

(i) the inductors in branches b_1 and b_2 are *linear, time-varying*, described by functions $L_1(t)$ and $L_2(t)$, respectively, which are continuously differentiable, positive and nondecreasing on the interval $[0, \tau]$, $(\tau < \infty)$,

(ii) the resistors in branches b_1, b_2, b_3 are *nonlinear, time-invariant*, described by continuous functions $R_m: R^1 \to R^1$, $m = 1, 2, 3$ such that

$$[R_m(\xi) - R_m(\eta)](\xi - \eta) \geqq c(\xi - \eta)^2 \tag{27}$$

and

$$|R_m(\xi)| \leqq a + b|\xi| \tag{28}$$

for all $\xi, \eta \in R^1$ and some $c > 0$, $a, b \geqq 0$,

(iii) the capacitors in branches b_4 and b_5 are *linear, time-varying*, described by functions $C_4(t)$ and $C_5(t)$, respectively, which are continuously differentiable, positive and nondecreasing on $[0, \tau]$.

Using the notation introduced above, (refer to (13), (12) and (11)), we have for $\hat{\mathcal{N}}$,

$$\tilde{R}(\xi, t) = R(\cdot)\xi, \qquad \tilde{S}(\xi, t) = S(t)\xi, \tag{29}$$

where

$$R(\cdot) = \operatorname{diag}(R_1(\cdot), R_2(\cdot), R_3(\cdot), 0, 0), \tag{30}$$

$$S(t) = \operatorname{diag}(0, 0, 0, C_4^{-1}(t), C_5^{-1}(t)), \tag{31}$$

and

$$L(t) = \operatorname{diag}(L_1(t), L_2(t), 0, 0, 0). \tag{32}$$

From Fig. 1 and (32) it is immediately apparent that our network is L-proper.

Moreover, continuity of R_m and (28) show that, for $m = 1, 2, 3$, $R_m(\cdot)$ is a continuous operator from L_2 into itself. Thus, $R(\cdot): L_2^5 \to L_2^5$ is continuous.

Next, recalling the construction of the matrix Y (see (14)), we see readily that

$$Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}. \tag{33}$$

Thus, if $w = [w_1, w_2, w_3]^T \in R^3$, a simple calculation confirms that $Y^T R(\cdot) Y w = [R_1(w_1), R_2(w_2), R_3(w_3)]^T$. From this it follows by (27) that, for any $u, v \in R^3$,

$$(u - v)^T (Y^T R(\cdot) Y u - Y^T R(\cdot) Y v) \geqq c|u - v|^2. \tag{34}$$

(Here, $|\cdot|$ signifies the Euclidean norm in R^3.) Now, if X is any 5×3 matrix whose columns constitute an orthonormal basis in the column space of Y, and if $W_R = \hat{X}^* R(\cdot) \hat{X}$, we can easily verify by using (34) that

$$\langle W_R z_1 - W_R z_2, z_1 - z_2 \rangle \geqq c' \|z_1 - z_2\|^2 \tag{35}$$

for all $z_1, z_2 \in L_2^3$ and some $c' > 0$. (Note that $X = YQ$ with Q nonsingular.)

Finally, letting $(\hat{S}x)(t) = S(t) \int_0^t x(\sigma)\, d\sigma$, we see by (iii) that $\hat{S}$ is continuous from L_2^5 into itself. Moreover, if $W_C = \hat{X}^* \hat{S} \hat{X}$, it follows that W_C is monotone on L_2^3.

Summarizing these results we conclude that the operator $W = W_R + W_C = \hat{X}^* \hat{T} \hat{X}$ (with $\hat{T}$ being defined by (11)) fulfils the condition (16)*.

Hence, by Theorem 2, for any $e_1, \cdots, e_5 \in L_2$ and numbers j_1, j_2 there exists a unique current distribution $i_1, \cdots, i_5$ in our network such that i_1, i_2 are absolutely continuous on $[0, \tau]$, $i_1(0) = j_1$, $i_2(0) = j_2$ and $i_1', i_2', i_3, i_4, i_5 \in L_2$. Moreover, the i_m's depend continuously on the e_n's.

REFERENCES

[1] V. DOLEZAL, *Basic properties of Hilbert networks*, IEEE Trans. Circuit and Systems, CAS-23 (1976), pp. 490–497.

[2] H. BREZIS, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, Contributions to Nonlinear Functional Analysis, E. Zarantonello, ed., Academic Press, 1971, pp. 101–156.

[3] R. T. ROCKAFELLAR, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., 149 (1970), pp. 75–88.

[4] L. WEISS AND P. L. FALB, *Dolezal's theorem, linear algebra with continuously parametrized elements, and time-varying systems*, Math. Systems Theory, 3 (1969), pp. 67–75.

[5] V. DOLEZAL, *Some properties of non-canonic systems of linear integro-differential equations*, Časopis Pěst. Mat., 89 (1964), pp. 470–491.

[6] B. PEIKARI, *Fundamentals of Network Analysis and Synthesis*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0018 $01.00/0

AN ABSTRACT SECOND ORDER SEMILINEAR VOLTERRA INTEGRODIFFERENTIAL EQUATION*

C. C. TRAVIS† AND G. F. WEBB†

Abstract. The theory of strongly continuous cosine families is used to obtain existence results for semilinear second order Volterra integrodifferential equations in Banach spaces. The results are applied to examples of integro-partial differential equations which have nonlinearities involving the highest order spatial derivatives.

1. Introduction. We treat the existence question for the abstract semilinear second order Volterra integrodifferential initial value problem,

$$
\begin{aligned}
&u''(t)=Au(t)+\int_0^t g(t,s,u(s),u'(s))\,ds+f(t), \qquad t\in R,\\
(1.1)\qquad &u(0)=x\in X, \qquad u'(0)=y\in X.
\end{aligned}
$$

In (1.1) A is the infinitesimal generator of a strongly continuous cosine family $C(t)$, $t\in R$, of bounded linear operators in the Banach space X, g is a nonlinear unbounded operator from $R\times R\times X\times X$ to X, and f is a function from R to X.

The approach we take is to extend the methods of [16] from the first order to the second order case by employing the theory of strongly continuous cosine families in Banach spaces. In many cases it is advantageous to treat second order abstract differential equations directly rather than to convert them to first order systems. A useful machinery for the study of abstract second order equations is the theory of strongly continuous cosine families. We will make use of some of the basic ideas from cosine family theory and we refer the reader to [14] for a discussion of the results we will use.

In § 2 we treat the case that g is continuous (but not Lipschitz continuous) and A^{-1} is compact, in § 3 we treat the case that g is Lipschitz continuous, in § 4 we give some simple examples to illustrate our results, and in § 5 we compare our results to those of earlier researchers.

2. Existence of solutions in the non-Lipschitz case. We make the following assumptions on the linear operator A:

(2.1) A is the infinitesimal generator of a strongly continuous cosine family $C(t)$, $t\in R$, of bounded linear operators in the Banach space X.

We define the associated sine family $S(t)$, $t\in R$, by $S(t)x=\int_0^t C(s)x\,ds$, $x\in X$, $t\in R$. It is known (see (2.11) and (2.12) of [14]) that (2.1) implies

(2.2) there exist constants $M\geqq 1$ and $\omega\geqq 0$ such that

$$|C(t)|\leqq Me^{\omega|t|},\ t\in R \text{ and } |S(t)-S(\hat t)|\leqq M\left|\int_{\hat t}^t e^{\omega|s|}\,ds\right|,\ t,\hat t\in R.$$

For a strongly continuous cosine family we define $E=\{x\in X: C(t)x$ is once continuously differentiable on $R\}$. We remark that $S(t)X\subset E$ for $t\in R$, $S(t)E\subset D(A)$ for $t\in R$, $d/dt\,C(t)x=AS(t)x$ for $x\in E$ and $t\in R$, and $d^2/dt^2C(t)x=AC(t)x=C(t)Ax$ for $x\in D(A)$ and $t\in R$ (see (2.17), (2.18), and (2.19) of [14]).

* Received by the editors June 7, 1977, and in revised form December 7, 1977.

† Department of Mathematics, Vanderbilt University, Nashville, Tennessee 37235.

It is proved in [7, § 6] that for $0 \leqq \alpha \leqq 1$ the fractional powers $(-A)^\alpha$ exist as closed linear operators in X, $D((-A)^\alpha) \subset D((-A)^B)$ for $0 \leqq B \leqq \alpha \leqq 1$, and $(-A)^\alpha(-A)^B = (-A)^{\alpha+B}$ for $0 \leqq \alpha + B \leqq 1$ (at least after a suitable translation $A - cI$ of A). We assume in addition

(2.3) for $0 < \alpha \leqq 1$, $(-A)^\alpha$ maps onto X and is 1–1, so that $D((-A)^\alpha)$ is a Banach space when endowed with the norm $\|x\|_\alpha = \|(-A)^\alpha x\|$, $x \in D((-A)^\alpha)$. We denote this Banach space by X_α. We further assume that A^{-1} is compact.

We require the following lemmas:

LEMMA 2.1. *Let $C(t)$, $t \in R$, be a strongly continuous cosine family in X satisfying* (2.3). *The following are true*:

(2.4) *For $0 < \alpha < 1$, $(-A)^{-\alpha}$ is compact if and only if A^{-1} is compact,*

$$\text{for } 0 < \alpha < 1 \text{ and } t \in R,\ (-A)^{-\alpha}C(t) = C(t)(-A)^{-\alpha} \tag{2.5}$$

$$\text{and } (-A)^{-\alpha}S(t) = S(t)(-A)^{-\alpha}.$$

Proof. If $(-A)^{-\alpha}$ is compact for some α, $0 < \alpha < 1$, then $A^{-1} = -(-A)^{-\alpha}(-A)^{\alpha-1}$ is compact being the composition of a compact operator with a bounded operator. Conversely, if A^{-1} is compact, it follows from the resolvent identity that $(\lambda I - A)^{-1}$ is compact for every λ in the resolvent set of A. In [7], it is established that for $0 < \alpha < 1$,

$$(-A)^{-\alpha} = \frac{\sin \pi\alpha}{\pi} \int_0^\infty s^{-\alpha}(sI - A)^{-1}\, ds \tag{2.6}$$

exists in the uniform operator topology. Considering the integral as a limit of Riemann sums, we notice that $(-A)^{-\alpha}$ is the limit in norm of compact operators and therefore compact. (2.5) follows from (2.6) and the fact that $C(t)$ and $S(t)$ commute with $(sI - A)^{-1}$.

LEMMA 2.2. *Let* (2.1) *hold, let v: $R \to X$ such that v is continuously differentiable, and let $q(t) = \int_0^t S(t-s)v(s)\, ds$. Then*

(2.7) *q is twice continuously differentiable and for $t \in R$, $q(t) \in D(A)$, $q'(t) = \int_0^t C(t-s)v(s)\, ds$, and $q''(t) = \int_0^t C(t-s)v'(s)\, ds + C(t)v(0) = Aq(t) + v(t)$;*

$$\text{for } 0 \leqq \alpha \leqq 1 \text{ and } t \in R,\ (-A)^{\alpha-1}q'(t) \in E. \tag{2.8}$$

Proof. Statement (2.7) is established by Proposition 2.4 of [14]. To prove (2.8), observe that

$$C(r)(-A)^{\alpha-1}q'(t) = 2^{-1}\left[\int_0^t C(r+t-s)(-A)^{\alpha-1}v(s)\, ds + \int_0^t C(r-t+s)(-A)^{\alpha-1}v(s)\, ds\right]$$

$$= 2^{-1}\left[\int_r^{r+t} C(s)(-A)^{\alpha-1}v(r+t-s)\, ds\right.$$

$$\left. + \int_{r-t}^{r} C(s)(-A)^{\alpha-1}v(s-r+t)\, ds\right].$$

Thus

$$\frac{d}{dr}C(r)(-A)^{\alpha-1}q'(t)$$
$$=2^{-1}[C(r+t)(-A)^{\alpha-1}v(0)-C(r-t)(-A)^{\alpha-1}v(0)]$$
$$+2^{-1}\left[\int_0^t C(r+t-s)(-A)^{\alpha-1}v'(s)\,ds-\int_0^t C(r-t-s)(-A)^{\alpha-1}v'(s)\,ds\right],$$

is a continuous function of r.

We make the following assumptions on the functions g and f:

(2.9) $g: R\times R\times D\to X$ is continuous, where D is an open subset of X_α, for some $\alpha\in[0,1)$;

(2.10) $g_1: R\times R\times D\to X$ is continuous, where g_1 denotes the derivative of g with respect to its first variable;

(2.11) $f: R\to X$ is continuously differentiable.

PROPOSITION 2.1. *Let* (2.1), (2.3), (2.9), (2.10), *and* (2.11) *hold. Let* $x\in D$ *and* $(-A)^{\alpha-1}y\in E$. *There exists* $T>0$ *and a continuous function* $u: [-T,T]\to X_\alpha$ *satisfying*

$$\begin{aligned}u(t)&=C(t)x+S(t)y+\int_0^t S(t-s)\int_0^s g(s,r,u(r))\,dr\,ds\\&\quad+\int_0^t S(t-s)f(s)\,ds. \qquad t\in[-T,T].\end{aligned}\tag{2.12}$$

If, in addition, $x\in D(A)$ *and* $y\in E$, *then the solution* u *of* (2.12) *is twice continuously differentiable,* $u(t)\in D(A)$ *for* $t\in[-T,T]$, *and* u *satisfies*

$$\begin{aligned}&u''(t)=Au(t)+\int_0^t g(t,s,u(s))\,ds+f(t), \qquad t\in[-T,T],\\&u(0)=x, \qquad u'(0)=y.\end{aligned}\tag{2.13}$$

(We remark that a solution of (2.12) is called a mild solution of (2.13)).

Proof. For $\gamma>0$ let $N_\gamma(x)=\{x_1\in X_\alpha:\|x-x_1\|_\alpha<\gamma\}$. Let $\phi(t)\overset{\text{def}}{=}C(t)x+S(t)y+\int_0^t S(t-s)f(s)\,ds$ and observe that $\phi: R\to X_\alpha$ is continuous by virtue of Lemma 2.1. Now choose $\gamma>0$ and $T>0$ such that

$$N_\gamma(x)\subset D;\tag{2.14}$$

for $r,s\in[-T,T]$ and $x_1\in N_\gamma(x)$,

$$\|g(r,s,x_1)\|\leqq 1 \quad\text{and}\quad \|g_1(r,s,x_1)\|\leqq 1;\tag{2.15}$$

for $t\in[-T,T]$,

$$\|\phi(t)-x\|_\alpha<\gamma/2;\tag{2.16}$$

for $t\in[-T,T]$ and $x_1,x_2,x_3\in N_\gamma(x)$,

$$\begin{aligned}&\Big\|(-A)^{\alpha-1}\Big[-\int_0^t C(t-s)\Big(g(s,s,x_1)\\&\qquad+\int_0^s g_1(s,r,x_2)\,dr\Big)\,ds+\int_0^t g(t,s,x_3)\,ds\Big]\Big\|<\frac{\gamma}{2}.\end{aligned}\tag{2.17}$$

Let K be the closed bounded convex subset of $C \stackrel{\text{def}}{=} C([-T, T]; X_\alpha)$, with norm $\|\cdot\|_C$, defined by

$$K = \{\eta \in C : \|\eta - \phi\|_C \leqq \gamma/2\}.$$

Notice that for $\eta \in K$ and $t \in [-T, T]$, $\eta(t) \in D$, since $\|\eta(t) - x\|_\alpha \leqq \|\eta - \phi\|_C + \|\phi(t) - x\|_\alpha \leqq \gamma/2 + \gamma/2$. Define the transformation G on K by

$$(G\eta)(t) = \phi(t) + \int_0^t S(t-s) \int_0^s g(s, r, \eta(r))\, dr\, ds, \qquad t \in [-T, T].$$

Using (2.3) and (2.17) we see that for $t \in [-T, T]$

$$\|(G\eta)(t) - \phi(t)\|_\alpha = \left\|(-A)^{\alpha-1}\left[-\int_0^t C(t-s)\Big(g(s, s, \eta(s)) + \int_0^s g_1(s, r, \eta(r))\, dr\Big)\, ds + \int_0^t g(t, s, \eta(s))\, ds\right]\right\| < \frac{\gamma}{2}.$$

Further, $G\eta$ is continuous as a function from $[-T, T]$ to X_α, and thus G maps K into K.

We next show that G is continuous. By (2.9) and (2.10), given $\varepsilon > 0$ there exists $\delta > 0$ such that for $\eta_1, \eta_2 \in K$, $\|\eta_1 - \eta_2\|_C < \delta$, and $s \in [-T, T]$, we have

$$\sup_{-T \leqq r \leqq T} \|g(s, r, \eta_1(r)) - g(s, r, \eta_2(r))\| < \varepsilon,$$

$$\sup_{-T \leqq r \leqq T} \|g_1(s, r, \eta_1(r)) - g_1(s, r, \eta_2'(r))\| < \varepsilon.$$

Thus, for $\eta_1, \eta_2 \in K$, $t \in [-T, T]$,

$$\begin{aligned}
&\|(G\eta_1)(t) - (G\eta_2)(t)\|_\alpha \\
&= \left\|(-A)^{\alpha-1}\left[-\int_0^t C(t-s)\Big(g(s, s, \eta_1(s)) - g(s, s, \eta_2(s)) + \int_0^s g_1(s, r, \eta_1(r))\, dr - \int_0^s g_1(s, r, \eta_2(r))\, dr\Big)\, ds + \int_0^t (g(t, s, \eta_1(s)) - g(t, s, \eta_2(s)))\, ds\right]\right\| \\
&\leqq |(-A)^{\alpha-1}|\left[\left|\int_0^t M e^{\omega|t-s|}\Big(\varepsilon + \Big|\int_0^s \varepsilon\, dr\Big|\Big)\, ds\right| + \left|\int_0^t \varepsilon\, ds\right|\right]
\end{aligned}$$

and the continuity of G follows immediately.

We next show that the set $\{G\eta : \eta \in K\}$ is equicontinuous as a collection of functions in

C. For $\eta \in K$ and $-T \leqq t \leqq \hat{t} \leqq T$,

$$\begin{aligned}
&\|(G\eta)(t)-(G\eta)(\hat{t})\|_\alpha \\
&\quad \leqq \|(C(t)-C(\hat{t}))(-A)^\alpha x\| + \|A(S(t)-S(\hat{t}))(-A)^{\alpha-1}y\| \\
&\qquad + \left\|(-A)^{\alpha-1}\left[\int_0^t C(t-s)\left(g(s,s,\eta(s))+\int_0^s g_1(s,r,\eta(r))\,dr\right)ds\right.\right. \\
&\qquad\qquad \left.\left. -\int_0^{\hat{t}} C(\hat{t}-s)\left(g(s,s,\eta(s))+\int_0^s g_1(s,r,\eta(r))\,dr\right)ds\right]\right\| \\
&\qquad + \left\|(-A)^{\alpha-1}\left[\int_0^t g(t,s,\eta(s))\,ds-\int_0^{\hat{t}} g(\hat{t},s,\eta(s))\,ds\right]\right\| \qquad (2.18) \\
&\qquad + \left\|(-A)^{\alpha-1}\left[\int_0^t C(t-s)f'(s)\,ds-\int_0^{\hat{t}} C(\hat{t}-s)f'(s)\,ds\right]\right\| \\
&\qquad + \|(-A)^{\alpha-1}(C(t)-C(\hat{t}))f(0)\| + \|(-A)^{\alpha-1}(f(t)-f(\hat{t}))\|.
\end{aligned}$$

Now notice that

$$\begin{aligned}
&\left\|(-A)^{\alpha-1}\left[\int_0^t C(t-s)\left(g(s,s,\eta(s))+\int_0^s g_1(s,r,\eta(r))\,dr\right)ds\right.\right. \\
&\qquad \left.\left. -\int_0^{\hat{t}} C(\hat{t}-s)(g\left(s,s,\eta(s))+\int_0^s g_1(s,r,\eta(r))\,dr\right)ds\right]\right\| \\
&\quad \leqq \left\|\int_0^t (C(t-s)-C(\hat{t}-s))(-A)^{\alpha-1}\right. \qquad (2.19) \\
&\qquad \left. \cdot\left[g(s,s,\eta(s))+\int_0^s g_1(s,r,\eta(r))\,dr\right]ds\right\| + |(-A)^{\alpha-1}| \\
&\qquad \cdot\left\|\int_t^{\hat{t}} C(\hat{t}-s)\left[g(s,s,\eta(s))+\int_0^s g_1(s,r,\eta(r))\,dr\right]ds\right\| \to 0 \\
&\qquad\qquad \text{as } |t-\hat{t}|\to 0 \text{ uniformly for } \eta\in K
\end{aligned}$$

by virtue of (2.15), the fact that $(-A)^{\alpha-1}$ is compact from X to X, and the fact that $C(t)$ is uniformly continuous in finite t-intervals on compact subsets of X. Further,

$$\begin{aligned}
&\left\|(-A)^{\alpha-1}\left[\int_0^t g(t,s,\eta(s))\,ds-\int_0^{\hat{t}} g(\hat{t},s,\eta(s))\,ds\right]\right\| \\
&\quad \leqq |(-A)^{\alpha-1}|\left[\left\|\int_0^t\int_t^{\hat{t}} g_1(r,s,\eta(s))\,dr\,ds\right\| + \left\|\int_t^{\hat{t}} g(\hat{t},s,\eta(s))\,ds\right\|\right] \qquad (2.20) \\
&\quad = O(|t-\hat{t}|) \quad \text{uniformly for } \eta\in K
\end{aligned}$$

by virtue of (2.15). Finally,

$$\begin{aligned}
&\left\|(-A)^{\alpha-1}\left[\int_0^t C(t-s)f'(s)\,ds-\int_0^{\hat{t}} C(\hat{t}-s)f'(s)\,ds\right]\right\| \\
&\quad \leqq |(-A)^{\alpha-1}|\left[\left\|\int_0^t (C(t-s)-C(\hat{t}-s))f'(s)\,ds\right\| + \left\|\int_t^{\hat{t}} C(\hat{t}-s)f'(s)\,ds\right\|\right] \qquad (2.21) \\
&\quad \to 0 \quad \text{as } |t-\hat{t}|\to 0 \quad \text{uniformly for } \eta\in K,
\end{aligned}$$

by virtue of (2.11) and the uniform continuity of $C(t)$ in finite t-intervals on compact subsets of X. The claimed equicontinuity of $\{G\eta: \eta \in K\}$ now follows.

Lastly, we show that for each fixed $t \in [-T, T]$ the set $\{(G\eta)(t): \eta \in K\}$ is precompact in X_α. Since $(-A)^{-\beta}: X \to X_\alpha$ is compact where $\alpha < \beta$, it suffices to show that $\{(-A)^\beta((G\eta)(t) - \phi(t)): \eta \in K\}$ is bounded in X for $\alpha < \beta \leqq 1$. By (2.3)

$$\|(-A)^\beta (G\eta - \phi)(t)\|$$

$$\leqq \left\|(-A)^{\beta-1} \int_0^t C(t-s)\left[g(s, s, \eta(s)) + \int_0^s g_1(s, r, \eta(r))\, dr\right] ds\right\|$$

$$+ \left\|(-A)^{\beta-1} \int_0^t g(t, s, \eta(s))\, ds\right\|$$

and the boundedness claim then follows from (2.15).

By Schauder's fixed point theorem, G has a fixed point in K, which is a solution of (2.12). If $x \in D(A)$ and $y \in E$, then the solution of (2.12) is a solution of (2.13) by Proposition 2.4 of [14].

PROPOSITION 2.2. *Let* (2.1), (2.3), (2.9), (2.10), *and* (2.11) *hold and in addition let g and g_1 map closed, bounded sets in $R \times R \times D$ into bounded sets in X. If $x \in D(A)$, $(-A)^{\alpha-1} y \in E$, and u is a solution of* (2.12) *noncontinuable to the right on* $[0, d)$, *then either $d = +\infty$ or given any closed, bounded set U in D, there is a sequence $t_k \to d^-$ such that $u(t_k) \notin U$. An analogous result holds for a solution noncontinuable to the left.*

Proof. Assume that $d < \infty$ and the conclusion of the proposition is false. Then there is a closed bounded set U in D such that $u(t) \in U$ for $t_1 \leqq t < d$, where $0 \leqq t_1 < d$. Arguing as in (2.18), (2.19), (2.20), and (2.21) we see that for $t_1 < t < \hat{t} < d$,

$$\|u(t) - u(\hat{t})\|_\alpha$$

$$\leqq \|(C(t) - C(\hat{t}))(-A)^\alpha x\| + \|A(S(t) - S(\hat{t}))(-A)^{\alpha-1} y\|$$

$$+ \left\|\int_0^t (C(t-s) - C(\hat{t}-s))(-A)^{\alpha-1}\left[g(s, s, u(s)) + \int_0^s g_1(s, r, u(r))\, dr\right] ds\right\|$$

$$+ |(-A)^{\alpha-1}| \left[\left\|\int_t^{\hat{t}} C(\hat{t}-s)\left[g(s, s, u(s)) + \int_0^s g_1(s, r, u(r))\, dr\right] ds\right\|\right.$$

$$+ \left\|\int_0^t \int_t^{\hat{t}} g_1(r, s, u(s))\, dr\, ds\right\| + \left\|\int_t^{\hat{t}} g(\hat{t}, s, u(s))\, ds\right\|$$

$$+ \left\|\int_0^t (C(t-s) - C(\hat{t}-s)) f'(s)\, ds\right\| + \left\|\int_t^{\hat{t}} C(\hat{t}-s) f'(s)\, ds\right\|$$

$$\left. + \|(C(t) - C(\hat{t})) f(0)\| + \|f(t) - f(\hat{t})\|\right].$$

Since g and g_1 are bounded on $[0, d] \times [0, d] \times U$, $(-A)^{\alpha-1}$ is compact, and $C(t)$ is uniformly continuous in finite t-intervals on compact subsets of X, we see that $\lim_{t \to d^-} \|u(t) - \hat{x}\|_\alpha = 0$, where $\hat{x} = C(d)x + S(d)y + \int_0^d S(d-s) \int_0^s g(s, r, u(r))\, dr + \int_0^d S(d-s) f(s)\, ds$. Since $u'(t) = S(t)Ax + C(t)y + \int_0^t C(t-s) \int_0^s g(s, r, u(r))\, dr\, ds + \int_0^t C(t-s) f(s)\, ds$, $0 \leqq t < d$, we see that $\lim_{x \to d^-} \|u'(t) - \hat{y}\|_\alpha = 0$, where $\hat{y} = S(d)Ax + C(d)y + \int_0^d C(d-s) \int_0^s g(s, r, u(r))\, dr\, ds + \int_0^d C(d-s) f(s)\, ds$. By (2.8) we have

$(-A)^{\alpha-1}\hat{y}\in E$. Thus, we can apply Proposition 2.1 to find a solution for $0\leqq t\leqq d_1$ to the equation

$$\begin{aligned}v(t)=C(t)\hat{x}+S(t)\hat{y}+\int_0^t S(t-s)\int_0^s g(s+d,r+d,v(r))\,dr\,ds\\ +\int_0^t S(t-s)\Big[\int_{-d}^0 g(s+d,r+d,u(r+d))\,dr+f(s+d)\Big]\,ds.\end{aligned}\tag{2.22}$$

Extend u to $[0,d+d_1]$ by defining $u(t)=v(t-d)$ for $d\leqq t\leqq d+d_1$. Then, for $d\leqq t\leqq d+d_1$,

$$\begin{aligned}u(t)=C(t-d)\hat{x}+S(t-d)\hat{y}+\int_0^{t-d} S(t-d-s)\int_0^s g(s+d,r+d,u(r+d))\,dr\,ds\\ +\int_0^{t-d} S(t-d-s)\Big[\int_{-d}^0 g(s+d,r+d,u(r+d))\,ds+f(s+d)\Big]\,ds.\end{aligned}\tag{2.23}$$

Using (2.19) (2.20), and the identities (2.9) and (2.23) of [14] one sees that $u(t)$ satisfies (2.12), for $0\leqq t\leqq d+d_1$. But this contradicts the noncontinuability assumption and the proof is complete.

COROLLARY 2.1. *Let the hypothesis of Proposition* 2.2 *hold and, in addition, let* $D=D((-A)^\alpha)$. *If* $x\in D(A)$, $(-A)^{\alpha-1}y\in E$, *and u is a solution of* (2.12) *noncontinuable to the right on* $[0,d)$, *then either* $d=+\infty$ *or* $\overline{\lim}_{t\to d^-}\|u(t)\|_\alpha=+\infty$. *An analogous result holds for a solution noncontinuable to the left.*

3. Existence of solutions in the Lipschitz case. For an operator A as in (2.1), we let X_A denote the Banach space which is $D(A)$ with graph norm $\|x\|_A=\|x\|+\|Ax\|$, $x\in D(A)$. We make the following assumptions on g:

(3.1) $g: R\times R\times D_1\times D_2\to X$ is continuous and continuously differentiable with respect to its first variable, where D_1 is an open subset of X_A and D_2 is an open subset of X;

(3.2) g and g_1 are locally Lipschitz continuous in the following sense: for each $(x,y)\in D_1\times D_2$ there exist a neighborhood $D_{x,y}$ of (x,y) such that $D_{x,y}\subset D_1\times D_2$ and constants a and b such that for $s,r\in R$ and $(x_1,y_1),(x_2,y_2)\in D_{x,y}$,

$$\|g(s,r,x_1,y_1)-g(s,r,x_2,y_2)\|\leqq a(\|x_1-x_2\|_A+\|y_1-y_2\|),$$

$$\|g_1(s,r,x_1,y_1)-g_1(s,r,x_2,y_2)\|\leqq b(\|x_1-x_2\|_A+\|y_1-y_2\|).$$

PROPOSITION 3.1. *Let* (2.1), (2.11), (3.1), *and* (3.2) *hold. For each* $(x,y)\in D_1\times D_2$ *such that* $y\in E$ *there exists* $T>0$ *and a unique function* $u:[-T,T]\to X$ *such that u is continuous from* $[-T,T]$ *to* X_A, *u is twice continuously differentiable from* $[-T,T]$ *to* X, *and u satisfies*

$$\begin{aligned}&u''(t)=Au(t)+\int_0^t g(t,s,u(s),u'(s))\,ds+f(t),\qquad t\in[-T,T],\\ &u(0)=x,\qquad u'(0)=y.\end{aligned}\tag{3.3}$$

Proof. Let $(x,y)\in D_1\times D_2$, let $D_{x,y}$ be a neighborhood about (x,y) as in (3.2), and let N be a neighborhood about (x,y) such that $\bar{N}\subset D_{x,y}$. Let $T>0$ (we will specify T later) and let C_1 be the complete metric space of continuous functions u from $[-T,T]$ to X_A, which are continuously differentiable from [−T, T] to X, *which satisfy* $(u(t),u'(t))\in$

$\bar{N}$ for $-T \leqq t \leqq T$, and which have metric $\rho(u, v) \stackrel{\text{def}}{=} \sup_{-T\leqq t\leqq T}(\|u(t)-v(t)\|_A + \|u'(t)-v'(t)\|)$. Define the transformation K on C_1 by

$$(3.4)\qquad \begin{aligned}(Ku)(t) = C(t)x + S(t)y + \int_0^t S(t-s)\int_0^s g(s, r, u(r), u'(r))\,dr\,ds \\ + \int_0^t S(t-s)f(s)\,ds, \qquad u \in C_1, \quad t \in [-T, T].\end{aligned}$$

By virtue of the hypothesis we have placed on g, the function $v(s) \stackrel{\text{def}}{=} \int_0^s g(s, r, u(r), u'(r))\,dr$, $-T \leqq s \leqq T$ is continuously differentiable from $[-T, T]$ to X and $v'(s) = g(s, s, u(s), u'(s)) + \int_0^s g_1(s, r, u(r), u'(r))\,dr$. By (2.7), for $u \in C_1$, $t \in [-T, T]$, we have that $(Ku)(t) \in D(A)$ and

$$(3.5)\qquad \begin{aligned}(AKu)(t) = C(t)Ax + AS(t)y + \int_0^t C(t-s)\Big[g(s, s, u(s), u'(s)) \\ + \int_0^s g_1(s, r, u(r), u'(r))\,dr\Big]\,ds - \int_0^t g(t, s, u(s), u'(s))\,ds \\ + \int_0^t C(t-s)f'(s)\,ds + C(t)f(0) - f(t).\end{aligned}$$

Thus, for $u \in C_1$, Ku is continuous from $[-T, T]$ to X_A. Further,

$$(3.6)\qquad \begin{aligned}(Ku)'(t) = S(t)Ax + C(t)y + \int_0^t C(t-s)\int_0^s g(s, r, u(r), u'(r))\,dr\,ds \\ + \int_0^t C(t-s)f(s)\,ds, \qquad u \in C_1, \quad t \in [-T, T].\end{aligned}$$

Thus, for $u \in C_1$, Ku is continuously differentiable from $[-T, T]$ to X. If T is chosen sufficiently small, then K maps C_1 into C_1. Further if T is chosen sufficiently small, then K is a contraction on C_1, since

$$(3.7)\qquad \begin{aligned}&\|(Ku)(t) - (Kv)(t)\| \\ &\leqq \left|\int_0^t M\left|\int_0^{t-s} e^{\omega|\sigma|}\,d\sigma\right| \left|\int_0^s a(\|u(r)-v(r)\|_A + \|u'(r)-v'(r)\|)\,dr\right|\,ds\right|,\end{aligned}$$

$$(3.8)\qquad \begin{aligned}&\|(AKu)(t) - (AKv)(t)\| \\ &\leqq \left|\int_0^t Me^{\omega|t-s|}\Big[a(\|u(s)-v(s)\|_A + \|u'(s)-v'(s)\|) \right. \\ &\qquad \left. + \left|\int_0^s b(\|u(t)-v(t)\|_A + \|u'(r)-v'(r)\|)\,dr\right|\Big]\,ds\right| \\ &\quad + \left|\int_0^t a(\|u(s)-v(s)\|_A + \|u'(s)-v'(s)\|)\,ds\right|,.\end{aligned}$$

and

$$(3.9)\qquad \|(Ku)'(t) - (Kv)'(t)\| \leqq \left|\int_0^t Me^{\omega|t-s|}\left|\int_0^s a(\|u(r)-v(r)\|_A + \|u'(r)-v'(r)\|)\,dr\right|\,ds\right|.$$

By the contraction mapping theorem there exists a unique $u \in C_1$ such that $Ku = u$. From (2.7) and (3.4) we see that $u: [-T, T] \to X_A$ is continuous, $u: [-T, T] \to X$ is twice continuously differentiable, and u satisfies (3.3) uniquely.

PROPOSITION 3.2. *Let* (2.1), (2.8), (3.1) *hold and, in addition, let g satify the following*:

(3.10) *g and g_1 are uniformly continuous on bounded sets of $R \times R \times D_1 \times D_2$;*

(3.11) *for every bounded set W in $R \times R \times D_1 \times D_2$ there exists a constant L such that if $(s, r, x_1, y_1), (s, r, x_2, y_2) \in W$, then*

$$\|g(s, r, x_1, y_1) - g(s, r, x_2, y_2)\| \leqq L(\|x_1 - x_2\|_A + \|y_1 - y_2\|),$$

$$\|g_1(s, r, x_1, y_1) - g_1(s, r, x_2, y_2)\| \leqq L(\|x_1 - x_2\|_A + \|y_1 - y_2\|).$$

If $(x, y) \in D_1 \times D_2$, $y \in E$, and u is a solution of (3.3) *noncontinuable to the right on* $[0, d)$, *then either $d = +\infty$ or given any closed bounded set U in $D_1 \times D_2$ there is a sequence $t_k \to d^-$ such that $(u(t_k), u'(t_k)) \notin U$. An analogous result holds for a solution noncontinuable to the left.*

Proof. Assume $d < \infty$ and the conclusion of the proposition is false. Then there is a closed bounded set U in $D_1 \times D_2$ such that $(u(t), u'(t)) \in U$ for $t_1 \leqq t < d$, where $0 \leqq t_1 < d$. By Proposition 2.4 of [14] $u(t)$ must satisfy

$$u(t) = C(t)x + S(t)y + \int_0^t S(t-s) \int_0^s g(s, r, u(r), u'(r))\, dr\, ds$$
$$+ \int_0^t S(t-s) f(s)\, ds, \qquad 0 \leqq t < d.$$

Thus, for $t_1 < t < t + h < d$,

$$\begin{aligned} &\|u(t+h) - u(t)\| \\ &\quad \leqq \|(C(t+h) - C(t))x\| + \|(S(t+h) - S(t))y\| \\ &\qquad + \left\| \int_{-h}^0 S(t-s) \int_0^{s+h} g(s+h, r, u(r), u'(r))\, dr\, ds \right\| \\ &\qquad + \left\| \int_0^t S(t-s) \left[\int_0^{s+h} g(s+h, r, u(r), u'(r))\, dr - \int_0^s g(s, r, u(r), u'(r))\, dr \right] ds \right\| \\ &\qquad + \left\| \int_{-h}^0 S(t-s) f(s+h)\, ds \right\| + \left\| \int_0^t S(t-s)[f(s+h) - f(s)]\, ds \right\|. \end{aligned} \tag{3.12}$$

There exists a constant L as in (3.11) such that

$$\begin{aligned} &\left\| \int_0^t S(t-s) \left[\int_0^{s+h} g(s+h, r, u(r), u'(r))\, dr - \int_0^s g(s, r, u(r), u'(r))\, dr \right] ds \right\| \\ &= \left\| \int_0^t S(t-s) \left[\int_{-h}^0 g(s+h, r+h, u(r+h), u'(r+h))\, dr \right.\right. \\ &\qquad + \int_0^s [g(s+h, r+h, u(r+h), u'(r+h)) - g(s, r, u(r+h), u'(r+h))]\, dr \\ &\qquad \left.\left. + \int_0^s [g(s, r, u(r+h), u'(r+h)) - g(s, r, u(r), u'(r))]\, dr \right] ds \right\| \end{aligned} \tag{3.13}$$

$$\leqq \int_0^t |S(t-s)| \int_{-h}^0 \|g(s+h, r+h, u(r+h), u'(r+h))\| \, dr \, ds$$

$$+ \int_0^t |S(t-s)| \int_0^s \|g(s+h, r+h, u(r+h), u'(r+h))$$

$$- g(s, r, u(r+h), u'(r+h))\| \, dr \, ds$$

$$+ \int_0^t \int_r^t |S(t-s)| L(\|u(r+h)-u(r)\|_A + \|u'(r+h)-u'(r)\|) \, ds \, dr.$$

Using (3.10), (3.11), (3.12), (3.13), and the fact that g is bounded on $[0, d]\times[0, d]\times U$, we see that there exists a constant M_1 and a continuous function N_1: $R^+ \to R^+$ satisfying $N_1(0)=0$ such that

$$\|u(t+h)-u(t)\| \leqq N_1(h)+M_1 \int_0^t (\|u(s+h)-u(s)\|_A + \|u'(s+h)-u'(s)\|) \, ds, \qquad t_1<t<t+h<d. \tag{3.14}$$

In a similar fashion one shows that there exists a constant M_2 and a continuous function N_2: $R^+ \to R^+$ satisfying $N_2(0)=0$ such that

$$\|Au(t+h)-Au(t)\| \leqq N_2(h)+M_2 \int_0^t (\|u(s+h)-u(s)\|_A + \|u'(s+h)-u'(s)\|) \, ds, \qquad t_1<t<t+h<d, \tag{3.15}$$

and a continuous function N_3: $R^+ \to R^+$ satisfying $N_3(0)=0$ such that

$$\|u'(t+h)-u'(t)\| \leqq N_3(h)+M_3 \int_0^t (\|u(s+h)-u(s)\|_A + \|u'(s+h)-u'(s)\|) \, ds, \qquad t_1<t<t+h<d. \tag{3.16}$$

Using (3.14), (3.15), (3.16), and Gronwall's lemma we have that there exists a constant M_4 and a continuous function N_4: $R^+ \to R^+$ satisfying $N_4(0)=0$ such that for $t_1<t<t+h<d$,

$$\|u(t+h)-u(t)\|_A + \|u'(t+h)-u'(t)\| \leqq N_4(h) \exp(M_4 t).$$

Thus, $\lim_{t\to d^-}(u(t), u'(t))$ exists in $\bar{U}=U\subset D_1\times D_2$ and using Proposition 3.1 and the argument of Proposition 2.2 u can be continued past d, contradicting the noncontinuability hypothesis.

COROLLARY 3.1. *Let the hypothesis of Proposition* 3.2 *hold and, in addition, let* $D_1\times D_2 = X_A\times X$. *If* $x\in D(A)$, $y\in E$, *and u is a solution of* (3.3) *noncontinuable to the right on* $[0, d)$, *then either* $d=+\infty$ *or* $\overline{\lim}_{t\to d^-}(\|u(t)\|_A + \|u'(t)\|) = +\infty$. *An analogous result holds for a solution noncontinuable to the left.*

4. Examples. We first consider the following integro-partial differential equation:
Example 4.1.

$$w_{tt}(x,t)=w_{xx}(x,t)+\int_0^t \sigma(t,s,w(x,s))\,ds+h(x,t), \qquad 0<x<\pi, \quad t\in R,$$

$$(4.1)\quad w(0,t)=w(\pi,t)=0, \qquad t\in R,$$

$$w(x,0)=w_0(x), \qquad w_t(x,0)=w_1(x), \qquad 0<x<\pi.$$

Let $\sigma: R\times R\times R\to R$ be continuous, and also continuously differentiable with respect to its first variable. Let $h: R\times R\to R$ be continuous, and continuously differentiable with respect to its second variable. We shall demonstrate that equation (4.1) satisfies the hypothesis of Proposition 2.1, and hence establish the local existence of a solution to this integro-partial differential equation.

Let $X=L^2[0,\pi]$ and let $A: X\to X$ be defined by

$$(4.2)\qquad Az=z'', \qquad D(A)=\{z\in X: z, z' \text{ are absolutely continuous}, \; z''\in X, z(0)=z(\pi)=0\}.$$

Then

$$Az=\sum_{n=1}^{\infty} -n^2(z,z_n)z_n, \qquad z\in D(A),$$

where $z_n(s)=\sqrt{2/\pi}\sin ns$, $n=1,2,\cdots$, is the orthonormal set of eigenvalues of A. It is easily shown that A is the infinitesimal generator of a strongly continuous cosine family $C(t)$, $t\in R$, in X given by

$$C(t)z=\sum_{n=1}^{\infty}\cos nt(z,z_n)z_n, \qquad z\in X,$$

and that the associated sine family is given by

$$S(t)z=\sum_{n=1}^{\infty}(\sin nt/n)(z,z_n)z_n, \qquad z\in X.$$

If we choose $\alpha=\frac{1}{2}$, then A satisfies (2.3) since

$$(-A)^{1/2}z=\sum_{n=1}^{\infty} n(z,z_n)z_n, \qquad z\in D((-A)^{1/2})$$

and

$$(-A)^{-1/2}z=\sum_{n=1}^{\infty}(1/n)(z,z_n)z_n, \qquad z\in X.$$

The compactness of A^{-1} follows from Lemma 2.1, and the fact that the eigenvalues of $(-A)^{-1/2}$ are $\lambda_n=1/n$, $n=1,2,\cdots$.

Let $g: R\times R\times X_{1/2}\to X$ be defined by $(g(t,s,z))(x)=\sigma(t,s,z(x))$, $z\in X_{1/2}$, $x\in[0,\pi]$, and let $f: R\to X$ be defined by $(f(t))(x)=h(x,t)$, $x\in[0,\pi]$. With this choice of A, g and f, (2.13) is the abstract formulation of (4.1).

We claim that g and f satisfy the hypothesis of Proposition 2.1. We will show that (2.9) is satisfied. First note that $z\in D((-A)^{1/2})$ if and only if z is absolutely continuous, $z'\in X$, and $z(0)=z(\pi)=0$. We also have that $\|z\|_{1/2}=\|z'\|$. Now, let $(t_1,s_1,z_1)\in R\times R\times X_{1/2}$ and let $\varepsilon>0$. There exists $\delta>0$ such that if $t,s\in R$, $x\in[0,\pi]$, $p\in R$, and $|t_1-t|<\delta$, $|s_1-s|<\delta$, $|z_1(x)-p|<\delta$, then $|\sigma(t_1,s_1,z_1(x))-\sigma(t,s,p)|<\varepsilon$. Let $z\in X_{1/2}$

such that $\|z_1 - z\|_{1/2} < \delta/\sqrt{\pi}$. Then, for $x \in [0, \pi]$, $|z_1(x) - z(x)| \leqq \int_0^x |z_1'(\tau) - z'(\tau)|\, d\tau \leqq \sqrt{\pi}\|z_1' - z'\| = \sqrt{\pi}\|z_1 - z\|_{1/2}$. Thus, for $|t_1 - t| < \delta$, $|s_1 - s| < \delta$, and $\|z_1 - z\|_{1/2} < \delta/\sqrt{\pi}$, $\|g(t_1, s_1, z_1) - g(t, s, z)\|^2 = \int_0^\pi |\sigma(t_1, s_1, z_1(x)) - \sigma(t, s, z(x))|^2\, dx \leqq \pi\varepsilon^2$. This establishes the continuity of g. Conditions (2.10) and (2.11) can be established analogously.

Example 4.2. Consider the integro-partial differential equation

$$w_{tt}(x, t) = w_{xx}(x, t) + \int_0^t \sigma(t, s, w_{xx}(x, s), w_t(x, s))\, ds + h(x, t), \qquad 0 < x < \pi,$$

$$(4.3)\quad w(0, t) = w(\pi, t) = 0, \qquad t \in R,$$

$$w(x, 0) = w_0(x), \qquad w_t(x, 0) = w_1(x), \qquad 0 < x < \pi.$$

Let $\sigma: R \times R \times R \times R \to R$ such that σ is uniformly continuous on bounded sets, σ is differentiable with respect to its first variable, σ_1 is uniformly continuous on bounded sets, and σ is Lipschitz continuous in the following sense: there exists a constant L such that for $t, s, p_1, q_1, p_2, q_2 \in R$, $|\sigma(t, s, p_1, q_1) - \sigma(t, s, p_2, q_2)| \leqq L(|p_1 - p_2| + |q_1 - q_2|)$. Let h, X, A, and f be as in Example 4.1 and let $g: R \times R \times X_A \times X \to X$ be defined by $(g(t, s, z_1, z_2))(x) = \sigma(t, s, z_1''(x), z_2(x))$. Then, (3.3) is the abstract formulation of (4.2). It is readily verified that the hypothesis of Proposition 3.2 is satisfied.

5. Comparison with earlier results. In recent years there has been considerable effort in the investigation of abstract Volterra integrodifferential equations. Much of this effort was inspired by hereditary partial integrodifferential equations which serve as models for various problems in continuum mechanics. Most of the results obtained thus far have been for equations of parabolic type, that is, equation (2.13) with u'' replaced by u'. In addition, most of the results in the literature are restricted to equations of convolution type, that is, equation (2.13) with $g(t, s, u(s))$ having the form $a(t-s)Au(s)$ where A is a mapping from X to X. We remark that in the results of §§ 2 and 3 the nonlinear term g need not have this special form.

In our references we have listed some of the contributions to the theory of abstract integrodifferential equations. The results of [1], [2], [3], [4], [8], [9] and [17] treat nonlinear parabolic equations of convolution type. In [5] and [13] hyperbolic linear equations are treated, that is, equation (2.13) with g linear. In [10] a one-dimensional nonlinear hyperbolic equation of convolution type is considered and in [11] a hyperbolic equation of convolution type which is nonlinear in the partial differential equation part and linear in the hereditary part is treated. In [12] a linear parabolic equation is studied using techniques related to [16]. Lastly, in [18] hyperbolic linear equations are treated using the theory of strongly continuous cosine families.

REFERENCES

[1] V. BARBU, *Integro-differential equations in Hilbert spaces*, Anal. Sti. Univ. "AL. I. Cuza", Din Iaşi, 19 (1973), pp. 365–383.

[2] ———, *Nonlinear Volterra equations in a Hilbert space*, this Journal, 6 (1975), pp. 728–741.

[3] ———, *Nonlinear Semigroups and Differential Equations in Banach spaces*, Noordhof, Leyden, the Netherlands, 1976.

[4] M. G. CRANDALL, S. O. LONDEN AND J. A. NOHEL, *An abstract nonlinear Volterra integrodifferential equation*, MRC Technical Summary Report # 1684, The University of Wisconsin, Madison, 1976.

[5] C. M. DAFERMOS, *An abstract Volterra equation with applications to linear viscoelasticity*, J. Differential Equations, 7 (1970), pp. 554–569.

[6] ———, *Asymptotic stability in viscoelasticity*, Arch. Rational Mech. Anal., 37 (1970), pp. 297–308.

[7] H. O. FATTORINI, *Ordinary differential equations in linear topological spaces, I*, J. Differential Equations, 5 (1968), pp. 72–105.

[8] S.-O. LONDEN, *An existence result on a Volterra equation in a Banach space*, Trans. Amer. Math. Soc., to appear.
[9] ———, *On an integral equation in a Hilbert space*, this Journal, 8 (1977), pp. 950–971.
[10] R. C. MACCAMY, *A model for one-dimensional, nonlinear viscoelasticity*, Quart. Appl. Math., 35 (1977), pp. 21–33.
[11] ———, *An integro-differential equation with applications in heat flow*, Ibid., 35 (1977), pp. 1–19.
[12] R. K. MILLER, *Volterra integral equations in a Banach space*, Funkcial. Ekvac., 18 (1975), pp. 163–194.
[13] M. SLEMROD, *A hereditary partial differential equation with applications in the theory of simple fluids*, Arch. Rational Mech. Anal., to appear.
[14] C. C. TRAVIS AND G. F. WEBB, *Cosine families and abstract nonlinear second order differential equations*, Acta Math., to appear.
[15] ———, *Compactness, regularity, and uniform continuity properties of strongly continuous cosine families*, Houston J. Math., 3 (1977), pp. 555–567.
[16] G. F. WEBB, *An abstract semilinear Volterra integrodifferential equation*, Proc. Amer. Math. Soc., 69 (1978), pp. 255–260.
[17] M. G. CRANDALL AND J. A. NOHEL, *An abstract functional differential equation and a related nonlinear Volterra equation*, Israel J. Math., to appear.
[18] R. K. MILLER AND R. L. WHEELER, *Well-posedness and stability of linear Volterra integrodifferential equations in abstract spaces*, to appear.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0019 $01.00/0

ADDITION FORMULAS FOR JACOBI, GEGENBAUER, LAGUERRE, AND HYPERBOLIC BESSEL FUNCTIONS OF THE SECOND KIND*

LOYAL DURAND†

Abstract. We present a set of addition formulas for the Jacobi, Laguerre, Gegenbauer, and hyperbolic Bessel functions of the second kind, $Q_\nu^{(\alpha,\beta)}$, N_ν^α, D_ν^α, and K_ν. These addition formulas are analogues of Koornwinder's addition formulas for $P_n^{(\alpha,\beta)}$ and L_n^α, and of Gegenbauer's addition formulas for C_n^α and J_n. The addition formulas are derived from a set of product formulas for the functions of the second kind derived previously by the author, and, conversely, can be integrated to give the product formulas.

1. Introduction. A number of addition formulas are known for the classical orthogonal functions (see, for example, [1]–[3]). The most important of these are probably the symmetrical addition formulas for the Bessel functions [4, §§ 11.4, 11.41] and the Gegenbauer polynomials (or Legendre functions) [5, § 3.15.1 (9)], and the recently-discovered addition formulas for the Jacobi [6]–[8] and Laguerre [9] polynomials,

$$\frac{J_\nu([x^2+y^2+2xy\cos\phi]^{1/2})}{(x^2+y^2+2xy\cos\phi)^{\nu/2}} = 2^\nu\Gamma(\nu)\sum_{m=0}^{\infty}(\nu+m)\frac{J_{\nu+m}(x)}{x^\nu}\frac{J_{\nu+m}(y)}{y^\nu}C_m^\nu(\cos\phi), \tag{1}$$

$$\begin{aligned} &C_n^\alpha(\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos\phi)\\ &=\frac{\Gamma(2\alpha-1)}{[\Gamma(\alpha)]^2}\sum_{l=0}^{n}4^l(2l+2\alpha-1)\frac{\Gamma(n-l+1)[\Gamma(\alpha+l)]^2}{\Gamma(n+2\alpha+l)}\\ &\qquad\cdot(\sin\theta_1\sin\theta_2)^l C_{n-l}^{\alpha+l}(\cos\theta_1)C_{n-l}^{\alpha+l}(\cos\theta_2)C_l^{\alpha-1/2}(\cos\phi), \end{aligned} \tag{2}$$

$$\begin{aligned} P_n^{(\alpha,\beta)}(\cos 2\Theta) &= \sum_{k=0}^{n}\sum_{l=0}^{k}c_{k,l}(n,\alpha,\beta)(\sin\theta_1\sin\theta_2)^{k+l}\\ &\quad\cdot(\cos\theta_1\cos\theta_2)^{k-l}P_{n-k}^{(\alpha+k+l,\beta+k-l)}(\cos(2\theta_1))P_{n-k}^{(\alpha+k+l,\beta+k-l)}(\cos(2\theta_2))\\ &\quad\cdot(\cos\psi)^{k-l}P_l^{(\alpha-\beta-1,\beta+k-l)}(\cos(2\psi))C_{k-l}^{\beta}(\cos\phi), \end{aligned} \tag{3a}$$

$$\begin{aligned} \cos(2\Theta) &= \cos(2\theta_1)\cos(2\theta_2)+\sin(2\theta_1)\sin(2\theta_2)\cos\psi\cos\phi\\ &\quad -2\sin^2\theta_1\sin^2\theta_2\sin^2\psi, \end{aligned} \tag{3b}$$

$$\begin{aligned} c_{k,l}(n,\alpha,\beta) &= (\alpha+k+l)(\beta+k-l)\Gamma(\beta)\\ &\quad\cdot\frac{\Gamma(n-k+1)\Gamma(n+\alpha+\beta+k+1)\Gamma(\alpha+k)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)\Gamma(n+\alpha+l+1)\Gamma(\beta+k+1)\Gamma(n+\beta-l+1)}, \end{aligned} \tag{3c}$$

and

$$\begin{aligned} &\exp(-\sqrt{xy}\cos\phi\, e^{-i\psi})L_n^\alpha(x+y+2\sqrt{xy}\cos\phi\cos\psi)\\ &=\sum_{k=0}^{\infty}\sum_{l=0}^{n}(\alpha+k+l)\frac{\Gamma(\alpha+k)\Gamma(n-l+1)}{\Gamma(k+1)\Gamma(n+\alpha+k+1)}(-1)^{k+l}\\ &\qquad\cdot(xy)^{(1/2)(k+l)}L_{n-l}^{\alpha+k+l}(x)L_{n-l}^{\alpha+k+l}(y)P_l^{(\alpha-1,k-l)}(\cos(2\phi))(\cos\phi)^{k-l}e^{-i(k-l)\psi}. \end{aligned} \tag{4}$$

---

* Received by the editors March 18, 1976, and in revised form August 31, 1977.

† Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico. Permanent address: Department of Physics, University of Wisconsin—Madison, Madison, Wisconsin 53706. This work was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the U.S. Energy Research and Development Administration.

These addition formulas have played an important role in the theory of the corresponding functions. For example, Koornwinder [6] integrated (3) to obtain a product formula for the Jacobi polynomials, and then obtained a Laplace-type integral representation for those functions by considering an appropriate limit of the product formula. The product formula itself can be used to establish the positivity of the convolution structure for Jacobi series and expansions in Jacobi functions [10]. The author has modified the product formulas corresponding to (1)–(3) to obtain product formulas for functions of the second kind. These lead immediately to generalizations of Nicholson's formula for Bessel functions [4, §§ 13.73, 13.74], and to Nicholson-type integrals for sums of squares of Jacobi, Gegenbauer, Laguerre, and Hermite functions of the first and second kind [11], [12]. Moreover, the relations in (1), (2), and (4) appear frequently as given, or in various limits, in problems in physics and applied mathematics.

The addition formulas (1)–(4) have a deep connection with theory of Lie groups, and can be interpreted for special values of the indices as the addition formulas for spherical functions on appropriate homogeneous spaces. For example, Gegenbauer's addition formula (2) can be derived for n and 2α integers by noting that Gegenbauer polynomials are the spherical functions on the sphere $S^{2\alpha+1}$ which are invariant under the subgroup $SO(2\alpha+1)$ of $SO(2\alpha+2)$ [5, Chap. 11], [1, Chap. 9]. Gegenbauer's addition formula for Bessel functions, (2), can be derived for $\nu=\alpha-\frac{1}{2}$ by considering the Euclidean group in $2\alpha+1$ dimensions, $E(2\alpha+1)$ [1, Chap. 11], or equivalently, as a limit of (2) by considering the contraction of $SO(2\alpha+2)$ to $E(2\alpha+1)$. The addition formula (3) for the Jacobi polynomials can be derived for 2α, 2β, and n integers by considering the Jacobi polynomials as intertwining functions on $SO(2\alpha+2\beta+4)$ with respect to the subgroups $SO(2\alpha+2\beta+3)$ and $SO(2\alpha+2)\times SO(2\beta+2)$ [7]. Finally, the addition formula (4) for the Laguerre polynomials is a limiting case of the addition formula for the spherical functions (the so-called disk polynomials) on the homogeneous space $U(\alpha+2)/U(\alpha+1)$ [9]. It can also be interpreted for $\alpha=0$ as the addition formula for the Heisenberg group [1, Chap. 8, § 5], [2, §§ 4.11–4.16], [13], [14, Chap. 13].

The situation is rather different for the Jacobi, Gegenbauer, and Laguerre functions of the second kind, $Q_n^{(\alpha,\beta)}(z)$, $D_n^\alpha(z)$, and $N_n^\alpha(z)$, and the modified or hyperbolic Bessel function of the second kind (MacDonald function), $K_\nu(z)$ These functions cannot be interpreted directly as spherical functions on the homogeneous spaces noted above. However, the Lie algebras of the underlying groups can be realized in terms of differential operators acting on these functions.[1] One might expect, then, that the group structure would persist, and that the functions of the second kind would satisfy addition formulas similar in structure to (1)–(4), but with functions of the second kind appearing throughout. These addition formulas would presumably lead to product formulas for the functions of the second kind, e.g., for $Q_n^{(\alpha,\beta)}(x)Q_n^{(\alpha,\beta)}(y)$. As noted above, product formulas of this type have recently been discovered for general Jacobi, Gegenbauer, Laguerre, and Bessel functions [11], [12]. Some addition formulas are known for the functions of the second kind [1, Chap. 5, § 5.6; Chap. 7, § 6.4], [3, eqs. (63), (83), (89)], [4, § 11.41 (8)], [15, eqs. (8.6), (9.5), (10.6)], [16], [17], but with the exception of some special cases, are not of the form sought.

[1] For example, the Lie algebra of $SO(3)$ can be implemented by differential operators on either the familiar set of functions $e^{im\phi}P_l^m(\cos\theta)$, or on the set of Legendre functions of the second kind, $e^{im\phi}Q_l^m(\cos\theta)$. The differential recurrence relations satisfied by the Legendre functions reflect the action of the Lie algebra [1, Chap. 3, § 4.4]. These relations are the same for the P's and Q's.

We wish to point out here that the product formulas of [11] and [12] can be inverted to give addition formulas for the Jacobi, Gegenbauer, Laguerre, and hyperbolic Bessel functions of the second kind. These addition formulas are similar in structure to (1)–(4). We intend to give a group-theoretical derivation of these results elsewhere, but will use purely analytical methods in the present paper.

In the next section, we summarize some results on pairs of integral transforms which we need in our derivations. The product formulas and the corresponding addition formulas for the functions of the second kind are presented in the third section, and some comments on the results are given in the final section.

2. Transformation formulas.

2.1. Generalized Mehler transformation. The generalized Mehler transform [1, Chap. 10, § 4.5] of a function $f(t)$ may be written in terms of Gegenbauer functions $C_\lambda^\alpha(z)$ [5, § 3.15] as

$$\tilde{f}(\lambda)=\int_0^\infty \frac{C_{i\lambda-\alpha}^\alpha(\cosh t)}{\sin[\pi(i\lambda-\alpha)]} f(t)(\sinh t)^{2\alpha}\,dt, \qquad 0\leqq\lambda<\infty, \quad \alpha>0. \tag{5}$$

The function $C_{i\lambda-\alpha}^\alpha(z)/\sin[\pi(i\lambda-\alpha)]$ is a symmetric function of λ. Thus, $\tilde{f}$ is also symmetric, $\tilde{f}(-\lambda)=\tilde{f}(\lambda)$.

The inverse Mehler transform is given by

$$f(t)=\int_0^\infty \frac{C_{i\lambda-\alpha}^\alpha(\cosh t)}{\sin[\pi(i\lambda-\alpha)]}\tilde{f}(\lambda)r(\lambda,\alpha)\,d\lambda, \qquad 0\leqq t<\infty, \quad \alpha>0, \tag{6}$$

$$r(\lambda,\alpha)=2^{2\alpha-1}\frac{\lambda\sinh(\pi\lambda)[\Gamma(\alpha)]^2}{\Gamma(i\lambda+\alpha)\Gamma(-i\lambda+\alpha)}. \tag{7}$$

The pair of generalized Mehler transforms (5) and (6) gives a bijection of the spaces of L^2 functions defined relative to the weights $(\sinh t)^{2\alpha}$ and $r(\lambda,\alpha)$. These transforms may be considered also as special cases of the Fourier–Jacobi transforms studied recently by Flensted-Jensen [18], Flensted-Jensen and Koornwinder [10], and Koornwinder [19]. See also Braaksma and Meulenbeld [20].

It is convenient to express the function $C_{i\lambda-\alpha}^\alpha(z)$ in (6) in terms of Gegenbauer functions of the second kind,

$$\frac{C_{i\lambda-\alpha}^\alpha(z)}{\sin[\pi(i\lambda-\alpha)]}=e^{-i\pi\alpha}\frac{1}{i\sinh(\pi\lambda)}[D_{i\lambda-\alpha}^\alpha(z)-D_{-i\lambda-\alpha}^\alpha(z)], \tag{8}$$

where [11]

$$D_\lambda^\alpha(z)=e^{i\pi\alpha}\frac{\Gamma(\lambda+2\alpha)}{\Gamma(\alpha)\Gamma(\lambda+\alpha+1)}(2z)^{-\lambda-2\alpha}\,{}_2F_1(\tfrac{1}{2}\lambda+\alpha, \tfrac{1}{2}\lambda+\alpha+\tfrac{1}{2}; \lambda+\alpha+1; z^{-2}). \tag{9}$$

If we make this substitution and use the symmetry of $\tilde{f}(\lambda)$, we find that (6) can be rewritten as

$$f(t)=-i\,e^{-i\pi\alpha}\int_{-\infty}^\infty \tilde{f}(\lambda)D_{i\lambda-\alpha}^\alpha(\cosh t)r(\lambda,\alpha)[\sinh(\pi\lambda)]^{-1}\,d\lambda, \qquad 0\leqq t<\infty, \quad \alpha>0. \tag{10}$$

2.2. Fourier–Jacobi transformation. The general Fourier–Jacobi transform of a function $f(t)$ considered in [10], [18], and [19] will be defined in terms of the standard Jacobi functions $P_\lambda^{(\alpha,\beta)}(z)$ [5, § 10.8 (16)] as[2]

$$\tilde{f}(\lambda)=\frac{\Gamma(i\lambda/2-\alpha/2-\beta/2+1/2)}{\Gamma(i\lambda/2-\alpha/2-\beta/2+1/2)} \tag{11}$$
$$\cdot\int_0^\infty f(t)P^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(\cosh(2t))(\sinh t)^{2\alpha+1}(\cosh t)^{2\beta+1}\,dt,$$
$$0\leqq\lambda<\infty,\qquad \alpha\geqq\beta>-\tfrac{1}{2}.$$

It follows from the symmetry relation

$$\frac{\Gamma(i\lambda/2-\alpha/2-\beta/2+1/2)}{\Gamma(i\lambda/2+\alpha/2-\beta/2+1/2)}P^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(z) \tag{12}$$
$$=\frac{\Gamma(-i\lambda/2-\alpha/2-\beta/2+1/2)}{\Gamma(i\lambda/2+\alpha/2-\beta/2+1/2)}P^{(\alpha,\beta)}_{-i\lambda/2-\alpha/2-\beta/2-1/2}(z)$$

that $\tilde{f}(\lambda)$ is a symmetric function of λ, $\tilde{f}(-\lambda)=\tilde{f}(\lambda)$.

The inverse Fourier–Jacobi transform is given by

$$f(t)=\frac{1}{2\pi}\int_0^\infty \tilde{f}(\lambda)P^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(\cosh(2t))\frac{\lambda\sinh(\pi\lambda)}{\sin[\pi(i\lambda/2-\alpha/2-\beta/2-1/2)]} \tag{13}$$
$$\cdot\Gamma\left(i\frac{\lambda}{2}+\frac{\alpha}{2}+\frac{\beta}{2}+\frac{1}{2}\right)\Gamma\left(-i\frac{\lambda}{2}+\frac{\alpha}{2}-\frac{\beta}{2}+\frac{1}{2}\right)d\lambda,\qquad 0\leqq t<\infty,\quad \alpha\geqq\beta>-\tfrac{1}{2}.$$

We can re-express this transform in terms of Jacobi functions of the second kind, $Q_\lambda^{(\alpha,\beta)}(z)$ [5, § 10.8 (18)], by using the relation

$$\frac{\sinh(\pi\lambda)}{\sin[\pi(i\lambda/2-\alpha/2-\beta/2-1/2)]}\Gamma\left(i\frac{\lambda}{2}+\frac{\alpha}{2}+\frac{\beta}{2}+\frac{1}{2}\right)\Gamma\left(-i\frac{\lambda}{2}+\frac{\alpha}{2}-\frac{\beta}{2}+\frac{1}{2}\right)$$
$$\cdot P^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(z)$$
$$=-2i\frac{\Gamma(-i\lambda/2+\alpha/2+\beta/2+1/2)}{\Gamma(-i\lambda/2-\alpha/2+\beta/2+1/2)}Q^{(\alpha,\beta)}_{-i\lambda/2-\alpha/2-\beta/2-1/2}(z) \tag{14}$$
$$+2i\frac{\Gamma(i\lambda/2+\alpha/2+\beta/2+1/2)}{\Gamma(i\lambda/2-\alpha/2+\beta/2+1/2)}Q^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(z).$$

This gives the symmetrical expression

$$f(t)=\frac{i}{\pi}\int_{-\infty}^\infty\frac{\Gamma(i\lambda/2+\alpha/2+\beta/2+1/2)}{\Gamma(i\lambda/2-\alpha/2+\beta/2+1/2)}Q^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(\cosh(2t))\tilde{f}(\lambda)\lambda\,d\lambda, \tag{15}$$
$$0\leqq t<\infty,\quad \alpha\geqq\beta>-\tfrac{1}{2}.$$

The pair of Fourier–Jacobi transforms (11) and (13) or (15) gives a bijection of the spaces of functions which are L^2 with respect to the indicated weights.

[2] A somewhat different definition, with the Γ-functions distributed differently between $\tilde{f}(\lambda)$ and $f(t)$, is used by Flensted-Jensen and Koornwinder [10]. The definition used here appears to be the most symmetrical possible. Compare, for example, (11) and (15).

3. Addition formulas for functions of the second kind.

3.1. Hyperbolic Bessel functions. A product formula for the hyperbolic Bessel functions $K_\nu(z)$ was given in [11] and [12],

$$\text{(16)} \quad \frac{K_{\nu+m}(x)}{x^\nu}\frac{K_{\nu+m}(y)}{y^\nu} = 2^{\nu-1}\frac{\Gamma(\nu)\Gamma(m+1)}{\Gamma(m+2\nu)}\int_0^\infty \omega^{-\nu}K_\nu(\omega)C_m^\nu(\cosh t)(\sinh t)^{2\nu}\,dt,$$

$$\text{(17)} \quad \omega = [x^2+y^2+2xy\cosh t]^{1/2}, \qquad \operatorname{Re}\nu > -\frac{1}{2}, \quad |\arg\omega| < \frac{\pi}{2}.$$

This product formula can be regarded for ν real as a generalized Mehler transform of the function $\omega^{-\nu}K_\nu(\omega)$, (5). Thus, for $m = i\lambda-\nu$, λ real, $\nu > -\frac{1}{2}$,

$$\text{(18)} \quad \begin{aligned}&\int_0^\infty \omega^{-\nu}K_\nu(\omega)\frac{C_{i\lambda-\nu}^\nu(\cosh t)}{\sin[\pi(i\lambda-\nu)]}(\sinh t)^{2\nu}\,dt\\ &\qquad = -\frac{2^{-\nu+1}}{\pi\Gamma(\nu)}\Gamma(i\lambda+\nu)\Gamma(-i\lambda+\nu)\frac{K_{i\lambda}(x)}{x^\nu}\frac{K_{i\lambda}(y)}{y^\nu}.\end{aligned}$$

The inverse of (18) follows from (6) or (10).

$$\text{(19a)} \quad \omega^{-\nu}K_\nu(\omega) = -\frac{2^\nu}{\pi}\Gamma(\nu)\int_0^\infty \frac{K_{i\lambda}(x)}{x^\nu}\frac{K_{i\lambda}(y)}{y^\nu}\frac{C_{i\lambda-\nu}^\nu(\cosh t)}{\sin[\pi(i\lambda-\nu)]}\sinh(\pi\lambda)\lambda\,d\lambda$$

$$\text{(19b)} \quad = -\frac{2^\nu}{i\pi}\Gamma(\nu)\,e^{-i\pi\nu}\int_{-\infty}^\infty \frac{K_{i\lambda}(x)}{x^\nu}\frac{K_{i\lambda}(y)}{y^\nu}D_{i\lambda-\nu}^\nu(\cosh t)\lambda\,d\lambda,$$

$$\nu > -\frac{1}{2}, \quad |\arg x| < \frac{\pi}{2}, \quad |\arg y| < \frac{\pi}{2}, \quad |\arg x| + |\arg y| + |\operatorname{Im} t| < \pi.$$

This result can be extended at once by analytic continuation to complex ν with $\operatorname{Re}\nu > -\frac{1}{2}$. The convergence of the integrals in (18), (19a), and (19b), and the L^2 property of the Bessel functions, can be checked using the asymptotic estimates of the Bessel and Gegenbauer functions given in the Appendix in equations (A.1) to (A.4).

Equation (19) gives an addition formula for the hyperbolic Bessel functions of the second kind. The similarity of this result to (1) can be made more apparent by rewriting (19b) as

$$\text{(20)} \quad \omega^{-\nu}K_\nu(\omega) = \frac{2^\nu}{i\pi}\Gamma(\nu)\,e^{-i\pi\nu}\int_{-i\infty-\nu}^{i\infty-\nu}\frac{K_{\nu+m}(x)}{x^\nu}\frac{K_{\nu+m}(y)}{y^\nu}D_m^\nu(\cosh t)(\nu+m)\,dm.$$

The sum in (1) has been replaced by an integral in (20), and the ordinary Bessel functions and Gegenbauer polynomials have been replaced by functions of the second kind.

If we let $\nu \to 0$ in (19b) and use the limit

$$\text{(21)} \quad \lim_{\nu\to 0}\Gamma(\nu)D_{i\lambda-\nu}^\nu(\cosh t) = \frac{e^{i\lambda t}}{i\lambda},$$

we find that

$$\text{(22)} \quad K_0(\omega) = \frac{1}{\pi}\int_{-\infty}^\infty K_{i\lambda}(x)K_{i\lambda}(y)\,e^{i\lambda t}\,d\lambda.$$

This is just the Fourier transform of Nicholson's integral for the product of two K's [4, § 13.72 (1)],

$$K_\mu(x)K_\mu(y)=\int_0^\infty K_0(\omega)\cosh(\mu t)\,dt, \qquad \mu=i\lambda. \tag{23}$$

It appears also as a special case of a formula of Vilenkin [1, Chap. 5, § 5.6 (2)].

3.2. Gegenbauer functions of the second kind. The product formula for the Gegenbauer functions of the second kind was given in [11] and [12],

$$\begin{aligned}(24)\quad &(x^2-1)^{m/2}(y^2-1)^{m/2}D_{\nu-m}^{\alpha+m}(x)D_{\nu-m}^{\alpha+m}(y)\\ &\quad=C^{-1}(\nu,\alpha,m)\int_0^\infty D_\nu^\alpha(Z)C_m^{\alpha-1/2}(\cosh t)(\sinh t)^{2\alpha-1}\,dt,\\ &\quad\operatorname{Re}\alpha>0,\quad \operatorname{Re}(\nu-m+1)>0,\quad \operatorname{Re}(m+a-\tfrac12)\geqq 0,\\ &\quad|\arg(x\pm1)|<\pi,\quad |\arg(y\pm1)|<\pi,\quad |\arg(\sqrt{x^2-1}\sqrt{y^2-1})|<\pi.\end{aligned}$$

Here

$$Z=xy+\sqrt{x^2-1}\sqrt{y^2-1}\cosh t, \tag{25}$$

and

$$C(\nu,\alpha,m)=e^{-i\pi(\alpha+2m)}2^{2\alpha+2m-1}\frac{[\Gamma(\alpha+m)]^2\Gamma(2\alpha+m-1)\Gamma(\nu-m+1)}{\Gamma(2\alpha-1)\Gamma(m+1)\Gamma(\nu+2\alpha+m)}. \tag{26}$$

The substitution $m=i\lambda-\alpha+\frac12$, λ and α real, $\alpha>0$, brings (24) into the form of a generalized Mehler transform, (5),

$$\begin{aligned}&\int_0^\infty D_\nu^\alpha(Z)C_{i\lambda-\alpha+1/2}^{\alpha-1/2}(\cosh t)[\sin(\pi(i\lambda-\alpha+\tfrac12))]^{-1}(\sinh t)^{2\alpha-1}\,dt\\ (27)\quad &\quad=\frac{C(\nu,\alpha,i\lambda-\alpha+\frac12)}{\sin[\pi(i\lambda-\alpha+\frac12)]}[(x^2-1)(y^2-1)]^{(1/2)(i\lambda-\alpha+1/2)}\\ &\quad\cdot D_{\nu-i\lambda+\alpha-1/2}^{i\lambda+1/2}(x)D_{\nu-i\lambda+\alpha-1/2}^{i\lambda+1/2}(y).\end{aligned}$$

The inverse Mehler transform of (27) gives the addition formula,

$$\begin{aligned}&D_\nu^\alpha(Z)=\sqrt{\pi}\frac{\Gamma(\alpha-\frac12)}{\Gamma(\alpha)}e^{i\pi\alpha}\int_0^\infty\frac{\Gamma(i\lambda+\frac12)\Gamma(\nu-i\lambda+\alpha+\frac12)}{\Gamma(-i\lambda+\frac12)\Gamma(\nu+i\lambda+\alpha+\frac12)}e^{2\pi\lambda}\tanh(\pi\lambda)\\ (28a)\quad &\quad\cdot 2^{2i\lambda}[(x^2-1)(y^2-1)]^{(1/2)(i\lambda-\alpha+1/2)}D_{\nu-i\lambda+\alpha-1/2}^{i\lambda+1/2}(x)D_{\nu-i\lambda+\alpha-1/2}^{i\lambda+1/2}(y)\\ &\quad\cdot\frac{C_{i\lambda-\alpha+1/2}^{\alpha-1/2}(\cosh t)}{\sin[\pi(i\lambda-\alpha+\frac12)]}\lambda\,d\lambda\\ &=\sqrt{\pi}\frac{\Gamma(\alpha-\frac12)}{\Gamma(\alpha)}\int_{-\infty}^\infty\frac{\Gamma(i\lambda+\frac12)\Gamma(\nu-i\lambda+\alpha+\frac12)}{\Gamma(-i\lambda+\frac12)\Gamma(\nu+i\lambda+\alpha+\frac12)}\frac{e^{2\pi\lambda}}{\cosh(\pi\lambda)}\\ (28b)\quad &\quad\cdot 2^{2i\lambda}[(x^2-1)(y^2-1)]^{(1/2)(i\lambda-\alpha+1/2)}D_{\nu-i\lambda+\alpha-1/2}^{i\lambda+1/2}(x)\\ &\quad\cdot D_{\nu-i\lambda+\alpha-1/2}^{i\lambda+1/2}(y)D_{i\lambda-\alpha+1/2}^{\alpha-1/2}(\cosh t)\lambda\,d\lambda,\\ &\quad\alpha>0,\quad \operatorname{Re}(\nu+\alpha+\tfrac12)>0,\quad |\arg(x\pm1)|<\pi,\quad |\arg(y\pm1)|<\pi,\\ &\quad\frac12\left|\arg\left(\frac{x+1}{x-1}\right)\right|+\frac12\left|\arg\left(\frac{y+1}{y-1}\right)\right|+|\operatorname{Im}t|<\pi.\end{aligned}$$

This result extends by analytic continuation to complex α with $\operatorname{Re}\alpha>0$, $\operatorname{Re}(\nu+\alpha+\frac{1}{2})>0$. The convergence of the integrals in (24), (27), and (28), and the L^2 convergence of the functions involved in the Mehler transforms can be checked by using the asymptotic estimates of the Gegenbauer functions given in the Appendix, (A.4)–(A.6).

The similarity of this addition formula to the standard addition formula for the Gegenbauer functions, (2), can be made more apparent by replacing λ in (28) by $l=i\lambda-\alpha+\frac{1}{2}$, and rewriting (28b) as

$$D_\nu^\alpha(Z)=\frac{\Gamma(2\alpha-1)}{[\Gamma(\alpha)]^2}\int_{-i\infty-\alpha+1/2}^{i\infty-\alpha+1/2}dl(2l+2\alpha-1)4^l\frac{\Gamma(\nu-l+1)[\Gamma(\alpha+l)]^2}{\Gamma(\nu+2\alpha+l)}$$

$$\cdot e^{-2\pi i(\alpha+l)}[(x^2-1)(y^2-1)]^{l/2}D_{\nu-l}^{\alpha+l}(x)D_{\nu-l}^{\alpha+l}(y)D_l^{\alpha-1/2}(\cosh t), \tag{29}$$

$$\operatorname{Re}\alpha>0,\quad \operatorname{Re}(\nu+\alpha+\tfrac{1}{2})>0.$$

In the special case $\alpha=\frac{1}{2}$, (29) reduces to a formula of Vilenkin [16, eq. 2]. The addition formula (20) for the hyperbolic Bessel functions $K_\nu(z)$ can be obtained immediately from (29) by using the confluent limit of the D's [11],

$$\lim_{\lambda\to\infty}\lambda^{-2\alpha+1}e^{-i\pi\alpha}D_\lambda^\alpha\left(1+\frac{z^2}{2\lambda^2}\right)=\frac{1}{\sqrt{\pi}\Gamma(\alpha)}(2z)^{-\alpha+1/2}K_{\alpha-1/2}(z). \tag{30}$$

3.3. Jacobi functions of the second kind. The product formula for the Jacobi functions of the second kind was given in [12],

$$[(x-1)(y-1)]^{l+(1/2)m}[(x+1)(y+1)]^{(1/2)m}Q_{\nu-l-m}^{(\alpha+2l+m,\beta+m)}(x)Q_{\nu-l-m}^{(\alpha+2l+m,\beta+m)}(y)$$

$$=N_{\nu,l,m}^{\alpha,\beta}\int_0^\infty du\int_0^\infty dt\,Q_\nu^{(\alpha,\beta)}(Z)P_l^{(\alpha-\beta-1,\beta+m)}(\cosh(2u)) \tag{31}$$

$$\cdot C_m^\beta(\cosh t)(\sinh u)^{2\alpha-2\beta-1}(\cosh u)^{2\beta+m+1}(\sinh t)^{2\beta},$$

where the argument of $Q_\nu^\alpha(Z)$ is

$$Z=xy+\sqrt{x^2-1}\sqrt{y^2-1}\cosh t\cosh u+\tfrac{1}{2}(x-1)(y-1)\sinh^2 u, \tag{32}$$

and the normalization factor $N_{\nu,l,m}^{\alpha,\beta}$ is

$$N_{\nu,l,m}^{\alpha,\beta}=2^{2\beta+2l+2m}\frac{\Gamma(\beta)\Gamma(\nu+\beta-l+1)\Gamma(\nu+\alpha+l+1)}{\Gamma(\nu-l-m+1)\Gamma(\nu+\beta+1)\Gamma(\nu+\alpha+\beta+l+m+1)}$$

$$\cdot\frac{\Gamma(\nu+\alpha+\beta+1)\Gamma(m+1)\Gamma(l+1)}{\Gamma(m+2\beta)\Gamma(l+\alpha-\beta)}. \tag{33}$$

The product formula (31) holds for general complex values of ν, l, m, α, β for which $\operatorname{Re}\alpha>\operatorname{Re}\beta>-\frac{1}{2}$, $\operatorname{Re}(m+\beta)\geqq 0$, $\operatorname{Re}(l+\alpha/2+m/2)\geqq 0$, $\operatorname{Re}(\nu-l-m+1)>0$, and $\operatorname{Re}(\nu+\alpha-\beta-m+1)>0$, and for complex values of x and y with $|\arg(x\pm 1)|<\pi$, $|\arg(y\pm 1)|<\pi$, $|\arg(x-1)\cdot(y-1)|<\pi$, and $|\arg\sqrt{x^2-1}\sqrt{y^2-1}|<\pi$. (The limit $\alpha=\beta$ exists, and gives the product formula stated earlier for Gegenbauer functions.)

It is convenient as the first step in the derivation of an addition formula from (31) to let $m=m'-\beta$ and $l=\frac{1}{2}(i\lambda-m'-\alpha+\beta)$, λ real, and rewrite (31) as

$$[(x-1)(y-1)]^{(i\lambda-\alpha)/2}[(x+1)(y+1)]^{(m'-\beta)/2}Q_{\nu-(i\lambda+m'-\alpha-\beta)/2}^{(i\lambda,m')}(x)Q_{\nu-(i\lambda+m'-\alpha-\beta)/2}^{(i\lambda,m')}(y)$$

$$=N'^{\alpha,\beta}_{\nu,\lambda,m'}\int_0^\infty du\int_0^\infty dt\,Q_\nu^{(\alpha,\beta)}(Z)P_{(i\lambda-m'-\alpha+\beta)/2}^{(\alpha-\beta-1,m')}(\cosh(2u)) \tag{34}$$

$$\cdot C_{m'-\beta}^\beta(\cosh t)(\sinh u)^{2\alpha-2\beta-1}(\cosh u)^{m'+\beta+1}(\sinh t)^{2\beta}.$$

For α, β, and m' real with $\alpha-\beta-1 \geqq m' > -\frac{1}{2}$, (34) can be regarded as the Fourier–Jacobi transform of the function

$$(\cosh u)^{\beta-m'}\int_0^\infty dt\, Q_\nu^{(\alpha,\beta)}(Z)C_{m'-\beta}^{\beta}(\cosh t)(\sinh t)^{2\beta}.$$

The transform can be inverted by using (15), with the result

$$\begin{aligned}
&\int_0^\infty dt\, Q_\nu^{(\alpha,\beta)}(Z)C_{m'-\beta}^{\beta}(\cosh t)(\sinh t)^{2\beta}\\
&\quad=\frac{i}{\pi}\int_{-\infty}^{\infty} d\lambda\, \lambda \frac{\Gamma(\frac{1}{2}(i\lambda-m'-\alpha+\beta)+1)\Gamma(\frac{1}{2}(i\lambda+m'+\alpha-\beta))}{\Gamma(\frac{1}{2}(i\lambda-m'+\alpha-\beta))\Gamma(\frac{1}{2}(i\lambda+m'-\alpha+\beta)+1)}[N'^{\alpha,\beta}_{\nu,\lambda,m'}]^{-1}\\
&\qquad\cdot[(x-1)(y-1)]^{(i\lambda-\alpha)/2}[(x+1)(y+1)]^{(m'-\beta)/2}Q^{(i\lambda,m')}_{\nu-(i\lambda+m'-\alpha-\beta)/2}(x)\\
&\qquad\cdot Q^{(i\lambda,m')}_{\nu-(i\lambda+m'-\alpha-\beta)/2}(y)Q^{(\alpha-\beta-1,m')}_{(i\lambda-m'-\alpha+\beta)/2}(\cosh(2u))(\cosh u)^{m'-\beta}.
\end{aligned}\tag{35}$$

The convergence of the integrals and the behavior of the transforms for $\lambda\to\pm\infty$ can be checked by using the asymptotic estimates (A.4) and (A.7)–(A.10).

It can be shown that the result in (35), derived for m' real, $m'\geqq 0$, can be extended by analytic continuation to complex m' with Re $m'\geqq 0$. The substitution $m'=i\mu$, μ real, then brings the left hand side of (35) into the form of a generalized Mehler transform. This can be inverted by using (10). The result is a remarkable addition formula for the Jacobi functions of the second kind,

$$\begin{aligned}
Q_\nu^{(\alpha,\beta)}(Z)&=\int_{-\infty}^{\infty}d\mu\int_{-\infty}^{\infty}d\lambda\, W(\lambda,\mu)[(x-1)(y-1)]^{(i\lambda-\alpha)/2}[(x+1)(y+1)]^{(i\mu-\beta)/2}\\
&\quad\cdot Q^{(i\lambda,i\mu)}_{\nu-(i\lambda+i\mu-\alpha-\beta)/2}(x)Q^{(i\lambda,i\mu)}_{\nu-(i\lambda+i\mu-\alpha-\beta)/2}(y)\\
&\quad\cdot Q^{(\alpha-\beta-1,i\mu)}_{(i\lambda-i\mu-\alpha+\beta)/2}(\cosh(2u))D^{\beta}_{i\mu-\beta}(\cosh t)(\cosh u)^{i\mu-\beta},
\end{aligned}\tag{36}$$

where the weight function $W(\lambda,\mu)$ is

$$\begin{aligned}
W(\lambda,\mu)&=-\frac{1}{\pi^2}e^{-i\pi\beta}2^{-i\lambda-i\mu+\alpha+\beta-1}\frac{\lambda\mu\Gamma(\beta)\Gamma(\nu+\beta+1)}{\Gamma(\nu+\alpha+\beta+1)}\\
&\quad\cdot\frac{\Gamma(\frac{1}{2}(i\lambda+i\mu+\alpha-\beta))\Gamma(\frac{1}{2}(-i\lambda-i\mu+\alpha+\beta)+\nu+1)\Gamma(\frac{1}{2}(i\lambda+i\mu+\alpha+\beta)+\nu+1)}{\Gamma(\frac{1}{2}(i\lambda+i\mu-\alpha+\beta)+1)\Gamma(\frac{1}{2}(-i\lambda+i\mu+\alpha+\beta)+\nu+1)\Gamma(\frac{1}{2}(i\lambda-i\mu+\alpha+\beta)+\nu+1)}.
\end{aligned}\tag{37}$$

This result holds for $|\arg(x\pm1)|<\pi$, $|\arg(y\pm1)|<\pi$, $|\arg\sqrt{x^2-1}\sqrt{y^2-1}|<\pi$, $|\arg(x-1)(y-1)|<\pi$, Re $(\nu+\alpha+1)>0$, Re $(\nu+\alpha/2+\beta/2+1)>0$, and $\alpha>\beta>-\frac{1}{2}$. The last condition can be relaxed by analytic continuation in α and β to Re $\alpha>$ Re $\beta>-\frac{1}{2}$.

Finally, if we let $k=\frac{1}{2}(i\lambda+i\mu-\alpha-\beta)$, $l=\frac{1}{2}(i\lambda-i\mu-\alpha+\beta)$, $x=\cosh(2\theta_1)$, $y=\cosh(2\theta_2)$, $t=\phi$, and $u=\psi$, we can write (36) in a form analogous to (3),

$$\begin{aligned}
&Q_\nu^{(\alpha,\beta)}(\cosh(2\theta_1)\cosh(2\theta_2)+\sinh(2\theta_1)\sinh(2\theta_2)\cosh\phi\cosh\psi\\
&\qquad\qquad+2\sinh^2\theta_1\sinh^2\theta_2\sinh^2\psi)\\
&\quad=\frac{1}{\pi^2}e^{-i\pi\beta}\int_{-i\infty-(\alpha+\beta)/2}^{i\infty-(\alpha+\beta)/2}dk\int_{-i\infty-(\alpha-\beta)/2}^{i\infty-(\alpha-\beta)/2}dl\, c_{k,l}(\nu,\alpha,\beta)(\sinh\theta_1\sinh\theta_2)^{k+l}\\
&\qquad\cdot(\cosh\theta_1\cosh\theta_2)^{k-l}Q_{\nu-k}^{(\alpha+k+l,\beta+k-l)}(\cosh(2\theta_1))Q_{\nu-k}^{(\alpha+k+l,\beta+k-l)}(\cosh(2\theta_2))\\
&\qquad\cdot Q_l^{(\alpha-\beta-1,\beta+k-l)}(\cosh(2\psi))D^{\beta}_{k-l}(\cosh\phi),
\end{aligned}\tag{38}$$

where $c_{k,l}(\nu,\alpha,\beta)$ is defined in (3).

3.4. Laguerre functions of the second kind. The Laguerre functions of the second kind will be defined as in [12],

$$N_\nu^\alpha(z)=\tfrac{1}{2}\Gamma(\nu+\alpha+1)\,e^z\Psi(\nu+\alpha+1,\alpha+1,-z), \tag{39}$$

$0<\arg z<2\pi$, $-z=e^{-i\pi}z$, where $\Psi(a,c,z)$ is the usual confluent hypergeometric function of the second kind [5, § 6.7]. A product formula for these functions can be derived starting with Koornwinder's addition formula for the Laguerre polynomials [9]. The result is as follows [21],

$$\begin{aligned}(xy)^{(k+l)/2}&N_{\nu-l}^{\alpha+k+l}(x)N_{\nu-l}^{\alpha+k+l}(y)\\ &=c_{\nu,\alpha,k,l}\int_0^\infty d\phi\int_{-\infty}^\infty d\psi\,\exp\left[-\sqrt{xy}(\cosh\phi)\,e^\psi-(k-l)\psi\right]\\ &\quad\cdot N_\nu^\alpha(x+y+2\sqrt{xy}\cosh\phi\cosh\psi)P_l^{(\alpha-1,k-l)}(\cosh(2\phi))\\ &\quad\cdot(\sinh\phi)^{2\alpha-1}(\cosh\phi)^{k-l+1},\end{aligned} \tag{40}$$

where

$$c_{\nu,\alpha,k,l}=\frac{\Gamma(l+1)\Gamma(\nu+\alpha+k+1)}{\Gamma(\nu-l+1)\Gamma(\alpha+l)}e^{i\pi(k+l)}. \tag{41}$$

This product formula holds for $\operatorname{Re}\alpha>0$, $\operatorname{Re}(\alpha+k+l)\geqq 0$, $\operatorname{Re}(\nu-\alpha-k-l+1)>0$, and $\operatorname{Re}(\nu+\alpha+k-l+1)>0$, and for $0<\arg x<2\pi$, $0<\arg y<2\pi$, and $\pi/2<\arg<\sqrt{xy}<3\pi/2$. A special case corresponding to $k=l=0$ was given in [12].

The derivation of the addition formula which corresponds to the product formula (40) is similar to the derivation used in the case of the Jacobi functions. Let $k-l=m$ and $l=\frac{1}{2}(i\lambda-m-\alpha)$. Then for λ, α, and m real, $\alpha-1\geqq m>-\frac{1}{2}$, (40) has the form of a Fourier–Jacobi transform,

$$\begin{aligned}(xy)^{i\lambda-\alpha}&N_{\nu-(i\lambda-m-\alpha)/2}^{i\lambda}(x)N_{\nu-(i\lambda-m-\alpha)/2}^{i\lambda}(y)\\ &=c'_{\nu,\alpha,\lambda,m}\int_0^\infty d\phi\int_{-\infty}^\infty d\psi\,\exp\left[-\sqrt{xy}(\cosh\phi)\,e^\psi-m\psi\right]\\ &\quad\cdot N_\nu^\alpha(x+y+2\sqrt{xy}\cosh\phi\cosh\psi)P_{(i\lambda-m-\alpha)/2}^{(\alpha-1,m)}(\cosh(2\phi))\\ &\quad\cdot(\sinh\phi)^{2\alpha-1}(\cosh\phi)^{m+1}.\end{aligned} \tag{42}$$

The convergence of the integrals and the properties of the functions for $\lambda\to\pm\infty$ can be checked by using (A.7), (A.11), and (A.12). The transform can be inverted by using (15), with the result

$$\begin{aligned}\int_{-\infty}^\infty &d\psi\,\exp\left[-\sqrt{xy}(\cosh\phi)\,e^\psi-m\psi\right]N_\nu^\alpha(x+y+2\sqrt{xy}\cosh\phi\cosh\psi)\\ &=\frac{i}{\pi}\int_{-\infty}^\infty d\lambda\,\lambda\,\frac{\Gamma(\frac{1}{2}(i\lambda-\alpha-m)+1)\Gamma(\frac{1}{2}(i\lambda+\alpha+m))}{\Gamma(\frac{1}{2}(i\lambda+\alpha-m))\Gamma(\frac{1}{2}(i\lambda-\alpha+m)+1)}\frac{1}{c'_{\nu,\alpha,\lambda,m}}\\ &\quad\cdot(xy)^{(i\lambda-\alpha)/2}N_{\nu-(i\lambda-m-\alpha)/2}^{i\lambda}(x)N_{\nu-(i\lambda-m-\alpha)/2}^{i\lambda}(y)\\ &\quad\cdot Q_{(i\lambda-m-\alpha)/2}^{(\alpha-1,m)}(\cosh(2\phi))(\cosh\phi)^m.\end{aligned} \tag{43}$$

The convergence of the integrals can be checked by using the asymptotic estimates (A.9), (A.11), and (A.12).

Although (43) was derived for m real, it can be continued to complex m with $\operatorname{Re} m \geqq 0$. The substitution $m = i\mu$, μ real, then brings (43) into the form of a Fourier transform. This can be inverted to give the addition formula

$$\begin{aligned}
&\exp(-\sqrt{xy}(\cosh\phi)\,e^{\psi})N_\nu^\alpha(x+y+2\sqrt{xy}\cosh\phi\cosh\psi)\\
(44)\qquad &= \int_{-\infty}^{\infty} dm \int_{-\infty}^{\infty} d\lambda\, q(\lambda,\mu)(xy)^{(i\lambda-\alpha)/2} N^{i\lambda}_{\nu-(i\lambda-i\mu-\alpha)/2}(x) N^{i\lambda}_{\nu-(i\lambda-i\mu-\alpha)/2}(y)\\
&\quad \cdot Q^{(\alpha-1,i\mu)}_{(i\lambda-i\mu-\alpha)/2}(\cosh(2\phi))(\cosh\phi)^{i\mu}\, e^{i\mu\psi},\\
&\alpha > \tfrac{1}{2},\quad \operatorname{Re}(\nu+1) > 0,\quad 0 < \arg x < 2\pi,\quad 0 < \arg y < 2\pi,\\
&\pi/2 < \arg\sqrt{xy} < 3\pi/2.
\end{aligned}$$

The weight function $q(\lambda, \mu)$ is given by

$$(45)\qquad q(\lambda,\mu) = \frac{i\lambda}{2\pi^2}\,\frac{\Gamma(\nu-\frac{1}{2}(i\lambda-i\mu-\alpha)+1)\Gamma(\frac{1}{2}(i\lambda+i\mu+\alpha))}{\Gamma(\nu+\frac{1}{2}(i\lambda+i\mu+\alpha)+1)\Gamma(\frac{1}{2}(i\lambda+i\mu-\alpha)+1)}\, e^{\pi\lambda-i\pi\alpha}.$$

Finally, if we let $k = \frac{1}{2}(i\lambda + i\mu - \alpha)$ and $l = \frac{1}{2}(i\lambda - i\mu - \alpha)$, we can put (44) in a form similar to (4),

$$\begin{aligned}
&\exp(-\sqrt{xy}(\cosh\phi)\,e^{\psi})N_\nu^\alpha(x+y+2\sqrt{xy}\cosh\phi\cosh\psi)\\
(46)\qquad &= \frac{1}{\pi^2}\int_{-i\infty-\alpha/2}^{i\infty-\alpha/2} dk \int_{-i\infty-\alpha/2}^{i\infty-\alpha/2} dl\,(\alpha+k+l)\frac{\Gamma(\nu-l+1)\Gamma(\alpha+k)}{\Gamma(\nu+\alpha+k+1)\Gamma(k+1)}\, e^{-i\pi(k+l)}\\
&\quad \cdot (xy)^{(k+l)/2} N^{\alpha+k+l}_{\nu-l}(x) N^{\alpha+k+l}_{\nu-l}(y) Q_l^{(\alpha-1,k-l)}(\cosh(2\phi))(\cosh\phi)^{k-l} e^{(k-l)\psi}.
\end{aligned}$$

4. Remarks.

1. Equations (19b), (28b), (36) and (44) give the addition theorems for the functions K_ν, D_ν^α, $Q_\nu^{(\alpha,\beta)}$, and N_ν^α in their natural form. The close connection of these formulas with the addition formulas for the functions of the first kind, (1)–(4), is evident from a comparison of the latter with the alternative forms of the addition formulas for the functions of the second kind given in (20), (29), (38), and (46).

2. The replacement of ordinary angles and sums over integer-valued indices in (1)–(4) by hyperbolic angles and integrals over complex indices in (20), (29), (38), and (46) corresponds to a change from the compact to noncompact forms of the underlying groups. The new addition formulas can in fact be interpreted in terms of the representation of noncompact groups by integral transforms. This will be discussed in detail elsewhere. For some group-theoretical background for the noncompact cases, see, for example, [1, Chaps. 5, 9], [22], [23], [24], [25], [26]. We note here only that the special case of the addition formula (20) for the hyperbolic Bessel functions given in (23) was obtained by Vilenkin [1, Chap. 5, § 5.6 (2), $\lambda = 0$] from the representation of the group of motions of the pseudo-Euclidean plane by integral transforms.

3. The product formulas (16), (24), (31), and (40), although originally derived analytically [11], [12], [21], can be regarded on a deeper level as direct consequences of the addition formulas for the functions of the second kind. As noted in [11], [12], and [21], the product formulas lead immediately to Nicholson-type integrals for the sum of squares of the functions of the first and second kind, which may in turn be used to derive a number of useful bounds, asymptotic expansions, and monotonicity properties for the corresponding functions [11].

Appendix. In this section, we will collect a number of asymptotic formulas for the Bessel, Gegenbauer, Jacobi, and Laguerre functions which are useful in checking the regions of validity of our product and addition formulas. We give references to known results, and indicate the method of derivation of other results.

a. *Hyperbolic Bessel functions.*

(A.1) $$K_{i\lambda}(x)\sim(2\pi/|\lambda|)^{1/2}e^{-(\pi/2)|\lambda|}\cos\left(|\lambda|\ln(2|\lambda|/x)-|\lambda|-\frac{\pi}{4}\right),\qquad \lambda\to\pm\infty,\quad x\text{ fixed}.$$

(Derived from the integral representation [4, § 6.22 (11)] for $K_{i\lambda}(x)$ using the method of steepest descents.)

(A.2) $$K_{i\lambda}(x)\sim\left(\frac{\pi}{2x}\right)^{1/2}e^{-x},\qquad |x|\to\infty,\quad |\arg x|<\frac{3\pi}{2},\qquad [4,\ \S\ 7.23\ (1)].$$

b. *Gegenbauer functions.*

(A.3) $$D^{\nu}_{i\lambda-\nu}(\cosh t)\sim -i\exp(3\pi i\nu/2)2^{-\nu}[\Gamma(\nu)]^{-1}(\sinh t)^{-\nu}\lambda^{\nu-1}e^{-i\lambda t},$$
$$|\lambda|\to\infty,\quad -\frac{3\pi}{2}<\arg\lambda<\frac{\pi}{2},\quad \operatorname{Re}t>0,\quad |\operatorname{Im}t|<\pi.\qquad [15,\text{ eq. }6.3].$$

(A.4) $$C^{\nu}_{i\lambda-\nu}(\cosh t)\sim -[\pi\Gamma(\nu)]^{-1}\sin[\pi(i\lambda-\nu)]e^{-\nu t}\cdot[\Gamma(i\lambda+\nu)\Gamma(-i\lambda)e^{-i\lambda t}+\Gamma(-i\lambda+\nu)\Gamma(i\lambda)e^{i\lambda t}],$$
$$\operatorname{Re}t\to\infty,\qquad |\operatorname{Im}t|<\pi.\qquad [15,\text{ eq. }2.6].$$

(A.5) $$D^{\alpha}_{\nu}(Z)\sim e^{i\pi\alpha}\frac{\Gamma(\nu+2\alpha)}{\Gamma(\alpha)\Gamma(\nu+\alpha+1)}(2Z)^{-\nu-2\alpha},$$
$$|Z|\to\infty,\quad |\arg(Z\pm1)|<\pi.\qquad \text{(From (9) or [15, eq. 2.3].)}$$

(A.6) $$D^{\alpha+i\lambda}_{\nu-i\lambda}(x)\sim(\pi|\lambda|)^{-1/2}2^{-i\lambda-\alpha-1/2}e^{i\pi(\alpha+i\lambda\mp1/4)}\cdot[(x-1)^{-i\lambda-\alpha+1/2}+e^{\pm i\pi(\nu+\alpha+1/2)}(x+1)^{-i\lambda-\alpha+1/2}],$$
$$\lambda\to\pm\infty,\quad |\arg(x-1)|<\pi/2.$$

(Derived from the integral representation [15, eq. 1.5] for $D^{\alpha}_{\lambda}(z)$ by the method of steepest descents.)

(A.7) $$P^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(\cosh(2t))\sim 2^{\alpha+\beta+1}e^{-(\alpha+\beta+1)t}\cdot\left[\frac{\Gamma(i\lambda)}{\Gamma(i\lambda/2-\alpha/2-\beta/2+1/2)\Gamma(i\lambda/2+\alpha/2+\beta/2+1/2)}e^{i\lambda t}+\frac{1}{\pi}\sin\left(\pi\left(\frac{i\lambda}{2}-\frac{\alpha}{2}-\frac{\beta}{2}-\frac{1}{2}\right)\right)\frac{\Gamma(-i\lambda)\Gamma(i\lambda/2+\alpha/2-\beta/2+1/2)}{\Gamma(-i\lambda/2+\alpha/2-\beta/2+1/2)}e^{-i\lambda t}\right],$$
$$\operatorname{Re}t\to\infty,\qquad |\operatorname{Im}t|<\pi.\qquad [5,\ \S\S\ 10.8\ (16)\text{ and }2.10\ (11)].$$

(A.8) $$Q^{(\alpha,\beta)}_{\nu}(Z)\sim 2^{\nu+\alpha+\beta}\frac{\Gamma(\nu+\alpha+1)\Gamma(\nu+\beta+1)}{\Gamma(2\nu+\alpha+\beta+2)}Z^{-\nu-\alpha-\beta-1},$$
$$|Z|\to\infty,\qquad |\arg(Z\pm1)|<\pi.\qquad [5,\ \S\ 10.8\ (18)].$$

$$Q^{(\alpha,\beta)}_{i\lambda/2-\alpha/2-\beta/2-1/2}(\cosh(2t)) \sim \left(\frac{\pi}{2i\lambda}\right)^{1/2} (\sinh t)^{-\alpha-1/2} (\cosh t)^{-\beta-1/2} e^{-i\lambda t},$$

(A.9)

$$|\lambda| \to \infty, \quad -\frac{3\pi}{2} < \arg\lambda < \frac{\pi}{2}, \quad \operatorname{Re} t > 0, \quad |\operatorname{Im} t| < \pi. \qquad [5, \S\, 2.3.2\,(16)].$$

(A.10)

$$\begin{aligned} Q^{(i\lambda,m)}_{-i\lambda/2+\nu}(x) \sim {} & -\left(\frac{\pi}{2i\lambda}\right)^{1/2} e^{-\pi|\lambda|/2} \left(\frac{x+1}{2}\right)^{-m/2-1/4} \\ & \cdot \left\{ e^{\pm i\pi(\nu-1/4)} \left[\sqrt{\frac{x+1}{2}} + 1\right]^{-i\lambda} \right. \\ & \left. + e^{\mp i\pi(\nu+m-1/4)} \left[\sqrt{\frac{x+1}{2}} - 1\right]^{-i\lambda} \right\}, \quad \lambda \to \pm\infty, \quad |\arg(x \pm 1)| < \pi. \end{aligned}$$

(Derived by the method of steepest descents from the integral representation for $Q^{(i\lambda,m)}_{-i\lambda/2+\nu}(x)$,

$$\begin{aligned} Q^{(i\lambda,m)}_{-i\lambda/2+\nu}(x) = {} & \frac{1}{2}\left(\frac{x+1}{2}\right)^{-i\lambda/2-\nu-m-1} [1 - e^{\pi\lambda + 2\pi i(\nu+m)}]^{-1} \\ & \cdot \int_{(\infty,1-)} dt\, t^{i\lambda/2-\nu-1} (t-1)^{-i\lambda/2+\nu+m} \left(t - \frac{2}{x+1}\right)^{-i\lambda/2-\nu-m-1}, \end{aligned}$$

$\operatorname{Re}\nu > -1$, $\operatorname{Re}(\nu+m) > -1$, $|\arg(x-1)| < \pi$.)

(A.11)

$$N^{\alpha}_{\nu}(Z) \sim \tfrac{1}{2}\Gamma(\nu+\alpha+1)\, e^{Z} Z^{-\nu-\alpha-1},$$

$$|Z| \to \infty, \qquad 0 < \arg Z < 2\pi. \qquad \text{(Equation (38) and [5, § 6.13.1 (1)].)}$$

(A.12)

$$N_{\nu-i\lambda/2}(x) \sim [\tfrac{1}{2}\Gamma(i\lambda)\, e^{-\pi\lambda} x^{-i\lambda} - 2^{-i\lambda-1/2}(\pi/|\lambda|)^{1/2} e^{\pm i\pi(\nu-1/4)-\pi|\lambda|/2}]\, e^{x/2},$$

$$\lambda \to \pm\infty, \qquad 0 < \arg x < 2\pi.$$

(Derived by the method of steepest descents from the integral representation for $N^{i\lambda}_{\nu-i\lambda/2}(x)$,

$$\begin{aligned} N^{i\lambda}_{\nu-i\lambda/2}(x) = {} & -\frac{1}{4i} e^{x-i\pi\nu+\pi\lambda/2} \csc\left[\pi\left(\nu + \frac{i\lambda}{2}\right)\right] \\ & \cdot \int_{(\infty e^{i\phi},0-)} e^{xt} t^{\nu+i\lambda/2} (t+1)^{-\nu+i\lambda/2-1}\, dt, \end{aligned}$$

$$\frac{\pi}{2} < \phi + \arg x < \frac{3\pi}{2}, \quad |\phi| < \pi, \quad 0 < \arg x < 2\pi.)$$

Acknowledgment. The author would like to thank the faculty of the School of Natural Sciences of the Institute for Advanced Study, Princeton, New Jersey, for the hospitality accorded him during the fall of 1975 when the work on Bessel and Gegenbauer functions was done.

REFERENCES

[1] N. JA. VILENKIN, *Special functions and the theory of group representations*, American Mathematical Society, Providence, RI, 1968.

[2] WILLARD MILLER, JR., *Lie Theory and Special Functions*, Academic Press, New York, 1968.

[3] P. HENRICI, *Addition theorems for general Legendre and Gegenbauer functions*, J. Rational Mech. Anal., (Indiana Univ. Math. J.), 4 (1955), pp. 983–1018.

[4] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, Cambridge, England, 1966.

[5] A. ERDÉLYI, W. MAGNUS, F. OBERHETTINGER AND F. G. TRICOMI, *Higher Transcendental Functions*, vols. I, II, McGraw-Hill, New York, 1953.

[6] T. H. KOORNWINDER, *The addition formula for Jacobi polynomials, I. Summary of results*, Indag. Math., 34 (1972), pp. 188–191.

[7] ———, *The addition formula for Jacobi polynomials and spherical harmonics*, SIAM J. Appl. Math., 25 (1973), pp. 236–246.

[8] ———, *Jacobi polynomials, III. An analytic proof of the addition formula*, this Journal, 6 (1975), pp. 533–543.

[9] ———, *The addition formula for Laguerre polynomials*, this Journal, 8 (1977), pp. 535–540.

[10] M. FLENSTED-JENSEN AND T. KOORNWINDER, *The convolution structure for Jacobi function expansions*, Ark. Mat., 11 (1973), pp. 245–262.

[11] L. DURAND, *Nicholson-type integrals for products of Gegenbauer functions and related topics*, Theory and Applications of Special Functions, R. Askey, ed., Academic Press, New York, 1975, pp. 353–374.

[12] ———, *Product formulas and Nicholson-type integrals for Jacobi functions, I. Summary of results*, this Journal, 9 (1978), pp. 76–86.

[13] J. PEETRE, *The Weyl transform and Laguerre polynomials*, Mathematische (Catania), 27 (1972), pp. 301–323.

[14] J. D. TALMAN, *Special Functions, A Group Theoretic Approach*, W. A. Benjamin, New York, 1968.

[15] L. DURAND, P. M. FISHBANE AND L. M. SIMMONS, JR., *Expansion formulas and addition theorems for Gegenbauer functions*, J. Mathematical Phys., 17 (1976), pp. 1933–1948.

[16] N. JA. VILENKIN, *Functional composition theorems for the hypergeometric function*, Mat. Sb., 65 (1964), pp. 28–46. (In Russian.)

[17] ———, *The hypergeometric function and representations of the group of real second order matrices*, Amer. Math. Soc. Transl., 60 (1966), pp. 159–184.

[18] M. FLENSTED-JENSEN, *Paley–Wiener type theorems for a differential operator connected with symmetric spaces*, Ark. Mat., 10 (1972), pp. 143–162.

[19] T. H. KOORNWINDER, *A new proof of a Paley–Wiener type theorem for the Jacobi transform*, Ibid., 13, (1975), pp. 145–159.

[20] B. L. J. BRAAKSMA AND B. MEULENBELD, *Integral transforms with generalized Legendre functions as kernels*, Compositio Math., 18 (1967), pp. 235–287.

[21] L. DURAND, *A product formula and Nicholson-type integral for Laguerre functions*, in preparation.

[22] M. FLENSTED-JENSEN, *The spherical functions on the universal covering of $SU(n-1, 1)/SU(n-1)$*, Preprint Series 1973, no. 1, Københavns Universitet, Matematisk Institut, Copenhagen, Denmark.

[23] R. S. STRICHARTZ, *Fourier transforms and non-compact rotation groups*, Indiana Univ. Math. J., 24 (1974), pp. 499–526.

[24] ———, *Harmonic analysis on hyperboloids*, J. Functional Analysis, 12 (1973), pp. 241–383.

[25] N. LIMIĆ, J. NIEDERLE AND R. RACZKA, *Continuous degenerate representations of noncompact rotation groups*, J. Mathematical Phys., 7 (1966), pp. 2026–2035.

[26] ———, *Eigenfunction expansions associated with the second-order invariant operator on hyperboloids and cones*, Ibid., 8 (1967), pp. 1074–1093.

SIAM J. MATH. ANAL.
Vol. 10, No. 2, March 1979

0036-1410/79/1002-0020 $01.00/0

AN ELEMENTARY PROOF OF DUNKL'S ADDITION THEOREM FOR KRAWTCHOUK POLYNOMIALS*

MIZAN RAHMAN†

Abstract. A product formula for the Krawtchouk polynomials is obtained by using a formula due to Gasper. Hypergeometric series techniques are then applied to obtain an addition theorem for these polynomials which was previously obtained by Dunkl by group-theoretic methods.

1. Introduction. The Krawtchouk polynomials $K_n(x;\, p, N)$ are orthogonal with respect to the binomial distribution $b(x;\, N, p) = \binom{N}{x} p^x (1-p)^{N-x}$, $0<p<1$, on the discrete set $x = 0, 1, \cdots, N$, and have a hypergeometric series representation [3], [4]:

$$K_n(x;\, p, N) = {}_2F_1(-n, -x;\, -N;\, p^{-1}), \qquad n, x = 0, 1, \cdots, N. \tag{1.1}$$

For $p = k/(k+1)$, $k = 1, 2, \cdots$, they also have a geometrical interpretation as spherical functions on the N-fold product of the k-simplex. It is this second interpretation that led Dunkl [2] to derive an addition theorem for the Krawtchouk polynomials by using group theoretic techniques. The approach of considering certain types of orthogonal polynomials as spherical functions on homogeneous spaces seems to be a natural one when one is trying to find an addition theorem. Koornwinder found his addition theorem for Jacobi polynomials [7]–[9] by applying group theoretic methods. But he also found an alternative proof of the same theorem by using only analytic methods. One would like to see if Dunkl's theorem could likewise be proved by using hypergeometric series methods alone. The present paper provides just such a proof.

Our starting point is Gasper's formula for the product of two Krawtchouk polynomials [4, eq. (4.4)]:

$$\begin{aligned} &K_n(x;\, p, N) K_n(y;\, p, N) \\ &\quad = (1-p^{-1})^n \sum_{k=0}^{n} \sum_{l=0}^{n-k} \frac{(-n)_{k+l}(-x)_k(-y)_k(x-N)_l(y-N)_l}{k!\, l!\, (-N)_{k+l}(-N)_{k+l}} p^{-k}(1-p)^{-l}. \end{aligned} \tag{1.2}$$

In § 2 we use this formula to obtain a product formula for the Krawtchouk polynomials and in § 3 we derive the full addition theorem by considering a certain set of discrete orthogonal polynomials in two variables.

2. A product formula. For the sake of definiteness let us suppose $0 \leqq x \leqq y \leqq N$. Then we observe that

$$\begin{aligned} &(-x)_k(-y)_k(x-N)_l(y-N)_l/(-N)_{k+l} \\ &\quad = \frac{(-y)_x}{(-N)_x} \sum_{s=0}^{x} \frac{(-x)_s(y-N)_s}{s!\,(1-x+y)_s} (s-x)_k(y+s-N)_l. \end{aligned} \tag{2.1}$$

This follows as a consequence of Chu–Vandermonde's theorem [1]:

$${}_2F_1(-n, b;\, c;\, 1) = (c-b)_n/(c)_n. \tag{2.2}$$

---

* Received by the editors August 26, 1977, and in revised form September 28, 1977.

† Department of Mathematics, Carleton University, Ottawa, Ontario, Canada K1S 5B6. This work was supported by the National Research Council of Canada under Grant A6197.

Combining (1.2) and (2.1) we have

$$K_n(x;\, p, N)K_n(y;\, p, N) = (1-p^{-1})^n \frac{(-y)_x}{(-N)_x} \sum_{s=0}^{x} \frac{(-x)_s(y-N)_s}{s!\,(1-x+y)_s} W(n, s;\, x, y), \tag{2.3}$$

where

$$\begin{aligned} W(n, s;\, x, y) &= \sum_{k=0}^{n} \sum_{l=0}^{n-k} \frac{(-n)_{k+l}(s-x)_k(y+s-N)_l}{k!\,l!\,(-N)_{k+l}} p^{-k}(1-p)^{-l} \\ &= \sum_{k=0}^{\min(n,x-s)} \frac{(s-x)_k}{k!}(-p)^{-k} \sum_{l=k}^{n} \frac{(-n)_l(1-p)^{k-l}}{l!\,(-N)_l}(y+s-N)_{l-k}(-l)_k. \end{aligned} \tag{2.4}$$

If $x-s \leqq n$ and $s \leqq \min(x, N-y)$, then, since $(-l)_k = 0$ for $0 \leqq l \leqq k-1$ and $(y+s-N)_{l-k} = (-1)^{k-l}/(N+1-y-s)_{k-l}$ we may write

$$W(n, s;\, x, y) = \sum_{k=0}^{x-s} \frac{(s-x)_k p^{-k}}{k!} \sum_{l=0}^{n} \frac{(-n)_l(y+s-N)_l}{l!(-N)_l}(1-p)^{k-l} \frac{(-l)_k}{(N+1-y-s-l)_k}. \tag{2.5}$$

On the other hand if $x-s \geqq n$ the k-series in (2.4) terminates at $k=n$; however, the terms in the l-series vanish if k is allowed to assume values between $n+1$ and $x-s$ because of the factor $(-n)_l(-l)_k$. Thus, (2.5) is a representation of $W(n, s;\, x, y)$ in either case.

Using (2.2) for the last factor on the right of (2.5) and then replacing k by $x-s-k$ we obtain

$$\begin{aligned} W(n, s;\, x, y) &= (1-p^{-1})^{x-s} \sum_{l=0}^{n} \frac{(-n)_l(y+s-N)_l}{l!(-N)_l}(1-p)^{-l} \sum_{k=0}^{x-s} \frac{(s-x)_k}{k!}(p^{-1}-1)^{-k} \\ &\quad \cdot \sum_{m=0}^{x-s-k} \frac{(s+k-x)_m(N+1-y-s)_m}{m!\,(N+1-y-s-l)_m} \\ &= (1-p^{-1})^{x-s} \sum_{l=0}^{n} \frac{(-n)_l(y+s-N)_l}{l!(-N)_l}(1-p)^{-l} \sum_{m=0}^{x-s} \frac{(s-x)_m(N+1-y-s)_m}{m!\,(N+1-y-s-l)_m} \\ &\quad \cdot \sum_{k=0}^{x-s-m} \frac{(s+m-x)_k}{k!}(p^{-1}-1)^{-k}. \end{aligned}$$

But

$$\sum_{k=0}^{x-s-m} \frac{(s+m-x)_k}{k!}\left(\frac{p}{1-p}\right)^k = \left(\frac{2p-1}{p-1}\right)^{x-s-m},$$

and

$$\frac{(y+s-N)_l(N+1-y-s)_m}{(N+1-y-s-l)_m} = (y+s-N-m)_l.$$

Hence

$$\begin{aligned} W(n, s;\, x, y) &= \left(\frac{2p-1}{p}\right)^{x-s} \sum_{m=0}^{x-s} \frac{(s-x)_m}{m!}\left(\frac{p-1}{2p-1}\right)^m \sum_{l=0}^{n} \frac{(-n)_l(y+s-N-m)_l}{l!(-N)_l}(1-p)^{-l} \\ &= \left(\frac{2p-1}{p}\right)^{x-s} \sum_{m=0}^{x-s} \frac{(s-x)_m}{m!}\left(\frac{p-1}{2p-1}\right)^m {}_2F_1\left(-n, y+s-N-m;\, -N;\, \frac{1}{1-p}\right) \\ &= \left(\frac{1-p}{p}\right)^{x-s} \sum_{r=0}^{x-s} \frac{(s-x)_r}{r!}\left(\frac{2p-1}{p-1}\right)^r {}_2F_1(-n, 2s+y+r-x \\ &\qquad\qquad -N;\, -N;\, (1-p)^{-1}) \\ &= \left(\frac{1-p}{p}\right)^{x-s}\left(\frac{p}{p-1}\right)^n \sum_{r=0}^{x-s} \frac{(s-x)_r}{r!}\left(\frac{2p-1}{p-1}\right)^r K_n(2s+y-x+r;\, p, N). \end{aligned} \tag{2.6}$$

In deriving the second-to-last line above we have replaced the summation variable m by $r = x - s - m$. The last line follows from the transformation property of the hypergeometric function

$$ {}_2F_1(a, b;\ c;\ x) = (1-x)^{-a}{}_2F_1\left(a, c-b;\ c;\ \frac{x}{x-1}\right). \tag{2.7}$$

Using (2.6) in (2.3) we finally obtain the product formula in the form

$$\begin{aligned} K_n(x;\ p, N)K_n(y;\ p, N) = \sum_{s=0}^{x} \binom{x}{s} \frac{(y-N)_s(-y)_{x-s}}{(-N)_x} \sum_{r=0}^{x-s} \binom{x-s}{r} \\ \cdot \left(\frac{2p-1}{p}\right)^r \left(\frac{1-p}{p}\right)^{x-s-r} K_n(2s+y-x+r;\ p, N). \end{aligned} \tag{2.8}$$

A more suggestive form of this formula is

$$\begin{aligned} K_n(x;\ p, N)K_n(y;\ p, N) = \sum_{s=0}^{x} \sum_{r=0}^{x-s} \rho(s;\ y-N-1, -y-1, x) b\left(r;\ x-s, \frac{2p-1}{p}\right) \\ \cdot K_n(2s+y-x+r;\ p, N) \end{aligned} \tag{2.9}$$

where

$$\rho(s;\ \alpha, \beta, M) = \binom{M}{s} \frac{(\alpha+1)_s(\beta+1)_{M-s}}{(\alpha+\beta+2)_M} \tag{2.10}$$

is the hypergeometric distribution.

3. The addition theorem. For $0 \leqq l \leqq k \leqq x$, $0 \leqq s \leqq \min(x, N-y)$, $0 \leqq r \leqq x-s$, consider the bivariate polynomials

$$R_{k,l}(r, s;\ x, y) = {}_3F_2\left[\begin{matrix} l-k, k-N-1, -s \\ y-N, l-x \end{matrix}\right] {}_2F_1\left(-l, -r;\ s-x;\ \frac{p}{2p-1}\right). \tag{3.1}$$

Obviously

$${}_3F_2\left[\begin{matrix} l-k, k-N-1, -s \\ y-N, l-x \end{matrix}\right] = \begin{cases} Q_{k-l}(s;\ y-N-1, l-y-1, x-l) & \text{if } x-l \leqq N-y, \\ Q_{k-l}(s;\ l-x-1, x-N-1, N-y) & \text{if } N-y < x-l. \end{cases} \tag{3.2}$$

Let us first suppose $x - l \leqq N - y \leqq N - x$. The two hypergeometric functions on the right of (3.1) satisfy the orthogonality relations [4], [5]

$$\begin{aligned} \sum_{r=0}^{x-s} b\left(r;\ x-s, \frac{2p-1}{p}\right) {}_2F_1\left(-l, -r;\ s-x;\ \frac{p}{2p-1}\right) {}_2F_1\left(-l', -r;\ s-x;\ \frac{p}{2p-1}\right) \\ = \left(\frac{1-p}{2p-1}\right)^l \binom{x-s}{l}^{-1} \delta_{l,l'}; \end{aligned} \tag{3.3}$$

$$\begin{aligned} \sum_{s=0}^{x-l} \rho(s;\ y-N-1, l-y-1, x-l) Q_m(s;\ y-N-1, l-y-1, x-l) \\ \cdot Q_{m'}(s;\ y-N-1, l-y-1, x-l) = \pi_m^{-1}\delta_{m,m'}, \end{aligned} \tag{3.4}$$

where

$$\pi_m = \binom{x-l}{m} \frac{(y-N)_m(l-N-1)_m}{(x-N)_m(l-y)_m} \cdot \frac{2m+l-N-1}{l-N-1}. \tag{3.5}$$

Hence

$$\begin{aligned}&\sum_{s=0}^{x-l}\sum_{r=0}^{x-s}\rho(s;\ y-N-1, l-y-1, x-l)b\left(r;\ x-s, \frac{2p-1}{p}\right)\\ &\qquad\cdot\binom{x-s}{l}R_{k,l}(r,s;\ x,y)R_{k',l'}(r,s;\ x,y)\\ &\qquad=\left(\frac{1-p}{2p-1}\right)^{l}\pi_{k-l}^{-1}\delta_{l,l'}\delta_{k-l,k'-l'}.\end{aligned}\tag{3.6}$$

Let us now seek an expansion in the form

$$K_n(2s+r+y+x;\ p,N)=\sum_{k'}\sum_{l'}A_{k',l}(x,y)\binom{x-s}{l'}R_{k',l'}(r,s;\ x,y).\tag{3.7}$$

The range of the summation variables k', l' will be determined shortly. We now multiply both sides of (3.7) by $\rho(s;\ y-N-1, l-y-1, x-l)b(r;\ x-s, (2p-1)/p)R_{k,l}(r,s;\ x,y)$ and sum over r and s, $0\leqq r\leqq x-s$, $0\leqq s\leqq x-l$. Using (3.6) we obtain

$$\begin{aligned}&\left(\frac{1-p}{2p-1}\right)^{l}\pi_{k-l}^{-1}A_{k,l}(x,y)\\ &\quad=\sum_{s=0}^{x-l}\sum_{r=0}^{x-s}\rho(s;\ y-N-1, l-y-1, x-l)b\left(r;\ x-s,\frac{2p-1}{p}\right)R_{k,l}(r,s;\ x,y)\\ &\qquad\cdot K_n(2s+r+y-x;\ p,N).\end{aligned}\tag{3.8}$$

Applying the transformation formula (2.7) to the Krawtchouk polynomial above we get

$$\begin{aligned}&\left(\frac{1-p}{2p-1}\right)^{l}(1-p^{-1})^{-n}\pi_{k-l}^{-1}A_{k,l}(x,y)\\ &\quad=\sum_{s=0}^{x-l}\rho(s;\ y-N-1, l-y-1, x-l)\,{}_3F_2\left[\begin{matrix}l-k, k-N-1, -s\\ y-N, l-x\end{matrix}\right]\\ &\qquad\cdot\sum_{r=0}^{x-s}b\left(r;\ x-s,\frac{2p-1}{p}\right)K_l\left(r;\ \frac{2p-1}{p}, x-s\right){}_2F_1(-n, 2s+r+y-x-N;\\ &\qquad\qquad -N;\ (1-p)^{-1}).\end{aligned}\tag{3.9}$$

The summation over r can be carried out by using the Rodrigues' formula for Krawtchouk polynomials [3, p. 224]:

$$\begin{aligned}&b\left(r;\ x-s,\frac{2p-1}{p}\right)K_l\left(r;\ \frac{2p-1}{p}, x-s\right)\\ &\quad=(x-s-l)!\,((1-p)/p)^{x-s}\Delta_r^{l}\left[\left(\frac{2p-1}{1-p}\right)^{r-l}\Big/\ (r-l)!(x-s-r)!\right],\end{aligned}\tag{3.10}$$

where Δ_r is the difference-operator defined by

$$\Delta_r f(r)\equiv f(r+1)-f(r),\quad\text{and}\quad \Delta_r^l f(r)=\Delta_r(\Delta_r^{l-1}f(r)).$$

Hence, for $l \leqq q \leqq n$, we get, on summing by parts l times,

$$\sum_{r=0}^{x-s} b\left(r;\ x-s, \frac{2p-1}{p}\right) K_l\left(r;\ \frac{2p-1}{p},\ x-s\right)(2s+y-x-N+r)_q$$
$$= (-q)_l (x-s-l)!\,((1-p)/p)^{x-s} \sum_{r=0}^{x-s-l} \frac{(2s+y-x-N+r+l)_{q-l}}{r!\,(x-s-l-r)!}\left(\frac{2p-1}{1-p}\right)^r.$$

Replacing r by $x-s-l-r$, applying (2.7) once again and simplifying we obtain

$$\left(\frac{1-p}{2p-1}\right)^l (1-p^{-1})^{-n-l} \pi_{k-l}^{-1} A_{k,l}(x,y)$$
$$= \sum_{q=0}^{n-l} \frac{(-n)_{q+l}(1-p)^{-q-l}}{q!\,(-N)_{q+l}} \sum_{s=0}^{x-l} \rho(s;\ y-N-1, l-y-1, x-l)$$
$$\cdot\ {}_3F_2\left[\begin{matrix} l-k, k-N-1, -s \\ y-N, l-x \end{matrix}\right] \tag{3.11}$$
$$\cdot\ (s+y-N)_q\ {}_2F_1(s+l-x, -q;\ N+1-s-y-q;\ p^{-1}-1).$$

In order to perform the partial summation over the s variable we need to use the Rodrigues' formula for Hahn polynomials [3], [5]:

$$\rho(s;\ y-N-1, l-y-1, x-l)\ {}_3F_2\left[\begin{matrix} l-k, k-N-1, -s \\ y-N, l-x \end{matrix}\right]$$
$$= \frac{\binom{k-y-1}{k-l}}{\binom{x-l}{k-l}\binom{x-N-1}{x-l}} \Delta_s^{k-l}\left[\binom{s+y-N-1}{k-l+y-N-1}\binom{x-y-1+k-l-s}{k-y-1}\right].$$

Summing by parts $k-l$ times and expressing the binomial coefficients in terms of Pochhammer products (e.g. $\binom{n}{r} = (-n)_r(-1)^r/r!$), we have

$$\left(\frac{1-p}{2p-1}\right)^l (1-p^{-1})^{-n-l} \pi_{k-l}^{-1} A_{k,l}(x,y)$$
$$= \frac{(-N)_l(-y)_x}{(-N)_x(-y)_l}(-1)^{k-l} \sum_{k=0}^{n-l} \frac{(-n)_{q+l}(1-p)^{-q-l}}{q!\,(-N)_{q+l}} \sum_{s=0}^{x-k} \frac{(k-x)_s(k-l+y-N)_s}{s!\,(1+y-x)_s} \tag{3.12}$$
$$\cdot\ \Delta_s^{k-l}[(s+y-N)_q\ {}_2F_1(s+l-x, -q;\ N+1-s-y-q;\ p^{-1}-1)].$$

Now

$$\Delta_s^{k-l}[(s+y-N)_q\ {}_2F_1(s+l-x, -q;\ N+1-s-y-q;\ p^{-1}-1)]$$
$$= \sum_{m=0}^{q} \binom{q}{m}(p^{-1}-1)^m \Delta_s^{k-l}[(s+y-N)_{q-m}(s+l-x)_m]$$
$$= \sum_{m=0}^{q} \binom{q}{m}(p^{-1}-1)^m \sum_{r=0}^{k-l} \binom{k-l}{r} \Delta_s^r (s+y-N)_{q-m} \Delta_s^{k-l-r}(s+r+l-x)_m \tag{3.13}$$
$$= \sum_{m=0}^{q}\sum_{r=0}^{k-l} \binom{k-l}{r} \frac{q!\,(s+r+y-N)_{q-m-r}(s+k-x)_{m-k+l+r}}{(q-m-r)!\,(m-k+l+r)!}(p^{-1}-1)^m,$$

provided $k-l-r \leqq m \leqq q-r$; otherwise it is zero.

Note that we may write (3.12) in the form

$$\left(\frac{1-p}{2p-1}\right)^l (1-p^{-1})^{-n-l}\pi_{k-l}^{-1}A_{k,l}(x,y)$$

$$= \frac{(l-y)_{k-l}(x-N)_{k-l}}{(l-N)_{2k-2l}} \sum_{s=0}^{x-k} \rho(s;\ y-k-N+2k-l-1, k-y-1, x-k) \tag{3.14}$$

$$\cdot\ V_{k,l}(n,s;\ x,y)$$

where

$$V_{k,l}(n,s;\ x,y) = \sum_{q=0}^{n-l} \frac{(-n)_{q+l}(1-p)^{-q-l}}{q!(-N)_{q+l}} \Delta_s^{k-l}[(s+y-N)_q\ {}_2F_1(s+l-x,-q;\ N+1-s-y-q;\ p^{-1}-1)].$$

With the aid of (3.13) and transforming the variables q, m to $q+k-l$ and $m+k-l-r$ respectively we find

$$V_{k,l}(n,s;\ x,y) = (p^{-1}-1)^{k-l} \sum_{q=0}^{n-k} \frac{(-n)_{q+k}(1-p)^{-q-k}}{q!(-N)_{q+k}} \sum_{r=0}^{k-l} \binom{k-l}{r} (p^{-1}-1)^{-r}$$

$$\cdot (s+r+y-N)_q\ {}_2F_1(s+k-x,-q;\ N+1-s-y-q;\ p^{-1}-1). \tag{3.15}$$

$$= (p^{-1}-1)^{k-l}(2-p^{-1})^{x-k-s} \sum_{q=0}^{n-k} \frac{(-n)_{q+k}(1-p)^{-q-k}}{q!(-N)_{q+k}} \sum_{r=0}^{k-l} \binom{k-l}{r} (p^{-1}-1)^{-r}$$

$$\cdot\ {}_2F_1\left(s+k-x, N+1-s-y-r; N+1-s-y-r-q;\ \frac{p-1}{2p-1}\right).$$

In deriving the last line we have again used (2.7).

Expanding this last hypergeometric function and using the identity

$$\frac{(N+1-s-y-r)_m(s+r+y-N)_q}{(N+1-s-y-r-q)_m} = (s+r+y-N-m)_q$$

we obtain, after some simplification,

$$V_{k,l}(n,s;\ x,y)$$

$$= (p^{-1}-1)^{k-l}(2-p^{-1})^{x-k-s}(1-p)^{-k}\frac{(-n)_k}{(-N)_k} \sum_{m=0}^{x-k-s} \frac{(s+k-x)_m}{m!}\left(\frac{p-1}{2p-1}\right)^m$$

$$\cdot\ (1-p^{-1})^{s+y-N-m} \tag{3.16}$$

$$\cdot \sum_{r=0}^{k-l} \binom{k-l}{r} (-1)^r\ {}_2F_1(n-N, s+r+y-N-m;\ k-N;\ p^{-1}).$$

But

$$\sum_{r=0}^{k-l}\binom{k-l}{r}(-1)^r{}_2F_1(n-N,s+r+y-N+m;\ p^{-1})$$

$$=\sum_{q=0}^{N-n}\frac{(n-N)_q(s+y-N-m)_q}{q!\,(k-N)_q}p^{-q}\sum_{r=0}^{k-l}\frac{(l-k)_r(s+y-N-m+q)_r}{r!\,(s+y-N-m)_r}$$

$$(3.17)\quad =\sum_{q=0}^{N-n}\frac{(n-N)_q(s+y-N-m)_q(-q)_{k-l}}{q!\,(k-N)_q(s+y-N-m)_{k-l}}p^{-q},\quad \text{by (2.2)}$$

$$=(-p)^{l-k}\frac{(n-N)_{k-l}}{(k-N)_{k-l}}{}_2F_1(n-N+k-l,s+y-N-m+k-l;\ 2k-N-l;\ p^{-1})$$

$$=(-p)^{l-k}\frac{(n-N)_{k-l}}{(k-N)_{k-l}}(1-p^{-1})^{N-y+l-s+m-n}{}_2F_1(k-n,k-s-y+m;\ \cdot 2k-N-l;\ p^{-1}).$$

The last line follows as a consequence of another transformation formula for the hypergeometric function:

$${}_2F_1(a,b;\ c;\ z)=(1-z)^{c-a-b}{}_2F_1(c-a,c-b;\ c;\ z).$$

Substituting (3.17) in (3.16) and replacing the summation variable m by $x-k-s-m$ we finally obtain

$$V_{k,l}(n,s;\ x,y)$$

$$(3.18)\quad =\frac{(-n)_k(n-N)_{k-l}}{(-N)_{2k-l}}(-1)^k(1-p^{-1})^{-n}p^{l-2k}\sum_{m=0}^{x-k-s}b\left(m;\ x-k-s,\frac{2p-1}{p}\right)\cdot K_{n-k}(2s+y-x+m;\ p,N-2k+l).$$

Using this in (3.14) we have

$$\frac{p^{2k}}{(2p-1)^l}\pi_{k-l}^{-1}A_{k,l}(x,y)$$

$$=\frac{(-n)_k(l-y)_{k-l}(x-N)_{k-l}(n-N)_{k-l}}{(-N)_{2k-l}(l-N)_{2k-2l}}(-1)^{k+l}\sum_{s=0}^{x-k}\rho(s;\ y-k-N+2k-l-1,k-y-1,x-k)$$

$$\cdot\sum_{r=0}^{x-k-s}b\left(r;\ x-k-s,\frac{2p-1}{p}\right)K_{n-k}(2s+y-x+r;\ p,N-2k+l)$$

$$=\frac{(-n)_k(l-y)_{k-l}(x-N)_{k-l}(n-N)_{k-l}}{(-N)_{2k-1}(l-N)_{2k-2l}}(-1)^{k+l}K_{n-k}(x-k;\ p,N-2k+l)\cdot K_{n-k}(y-k;\ p,N-2k+l)$$

by (2.9).

Finally (3.5) and some simplification yield

$$(3.19)\quad A_{k,l}(x,y)=\binom{x-l}{k-l}\frac{(-n)_k(y-N)_{k-l}(l-N-1)_{k-l}}{(-N)_{2k-l}(l-N-1)_{2k-2l}}(-1)^{k+l}\cdot K_{n-k}(x-k;\ p,N-2k+l)K_{n-k}(y-k;\ p,N-2k+l).$$

The factor $(y-N)_{k-l}$ on the right-hand side implies that $l \geqq k+y-N$ if $k+y-N$ happens to be negative. Thus we have the addition formula

$$
\begin{aligned}
&K_n(2s+r+y-x;\ p, N)\\
&= \sum_{k=0}^{\min(n,x)} \sum_{l=\max(0,k+y-N)}^{k} \binom{x-l}{k-l}\binom{x-s}{l} \frac{(-n)_k(y-N)_{k-l}(l-N-1)_{k-l}}{(-N)_{2k-l}(l-N-1)_{2k-2l}}(-1)^{k+l}\\
&\quad \cdot K_{n-k}(x-k;\ p, N-2k+l)K_{n-k}(y-k;\ p, N-2k+l)(2p-1)^l p^{-2k}\\
&\quad \cdot {}_2F_1\left(-l,-r;\ s-x;\ \frac{p}{2p-1}\right) {}_3F_2\left[\begin{matrix} l-k, k-N-1, -s\\ y-N, l-x\end{matrix}\right],
\end{aligned}
\tag{3.20}
$$

where $0 \leqq x \leqq y \leqq N$, $0 \leqq s \leqq \min(x, N-y)$, $0 \leqq r \leqq x-s$ and $0 \leqq n \leqq N$.

REFERENCES

[1] R. ASKEY, *Orthogonal Polynomials and Special Functions*, Regional Conference Series in Applied Mathematics, vol. 21, Society for Industrial and Applied Mathematics, Philadelphia, 1975.

[2] CHARLES F. DUNKL, *A Krawtchouk polynomial addition theorem and Wreath products of symmetric groups*, Indiana Univ. Math. J., 25 (1976), pp. 335–358.

[3] A. ERDÉLYI, W. MAGNUS, F. OBERHETTINGER AND F. G. TRICOMI, EDS., *Higher Transcendental Functions*, vol. II, McGraw-Hill, New York, 1953.

[4] G. GASPER, *Non-negativity of a discrete Poisson kernel for the Hahn polynomials*, J. Math. Anal. Appl., 42 (1973), pp. 438–451.

[5] S. KARLIN AND J. MCGREGOR, *The Hahn polynomials, formulas and an application*, Scripta Math., 26 (1961), pp. 33–46.

[6] T. H. KOORNWINDER, *The addition formula for Jacobi polynomials, I. Summary of results*, Indag. Math., 34 (1972), pp. 188–191.

[7] ———, *The addition formula for Jacobi polynomials, II. The Laplace type integral representation and the product formula*, Math. Centrum Amsterdam, Afd. Toegepaste Wiskunde, Rep. TW 133, 1972.

[8] ———, *The addition formula for Jacobi polynomials, III. Completion of the proof*, Math. Centrum Amsterdam, Afd. Toegepaste Wiskunde, Rep. TW 135, 1972.

[9] ——, *The addition formula for Jacobi polynomials and spherical harmonics*, SIAM J. Appl. Math., 25 (1973), pp. 236–246.

[10] ———, *Jacobi polynomials, II. An analytic proof of the product formula*, this Journal, 5 (1974), pp. 125–137.

[11] ———, *Jacobi Polynomials, III. An analytic proof of the addition formula*, this Journal, 6 (1975), pp. 533–543.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0001 $01.00/0

ON ELLIPTIC SINGULAR PERTURBATION PROBLEMS WITH TURNING POINTS*

S. KAMIN†

Abstract. The boundary value problem for the elliptic equation $\varepsilon\Delta u + \sum b_i u_{x_i} = 0$ is considered in the case that the characteristic curves of the reduced equation enter the domain and have a singular point inside (turning point). Assume that there exists a potential function $\psi(x)$ such that $b_i = \psi_{x_i}$ $(i = 1, 2, \cdots n)$. It is proved that if $\varepsilon \to 0$ then the solutions $u_\varepsilon(x)$ converge to a constant, a formula for which was derived by Matkowsky and Schuss using formal asymptotic expansion.

1. In this paper we consider solutions of the first boundary-value problem for the equation

$$L_\varepsilon u = \varepsilon\Delta u + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} = 0, \qquad x = (x_1, \cdots x_n) \in R^n \tag{1.1}$$

in a compact domain $D \subset R^n$ with smooth boundary Γ. On Γ we prescribe

$$u|_\Gamma = \varphi(x) \tag{1.2}$$

where $\varphi(x)$ is a smooth bounded function. We study the asymptotic behavior of the solutions u_ε of (1.1), (1.2) as $\varepsilon \to 0$.

It is known (see Levinson [4], Vishik and Lyusternik [10], Eckhaus [1]) that the behavior of the characteristics of the reduced equation

$$\sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} = 0 \tag{1.3}$$

is of decisive importance in this connection.

Let the positive direction along the characteristics coincide with the direction of the vector $\bar{b} = (b_1, \cdots, b_n)$. In this paper we consider the case in which the vector $\bar{b}$ points into the domain and $x = 0 \in D$ is a singular point of the field of characteristic curves (Fig. 1). The point 0 is a turning point.

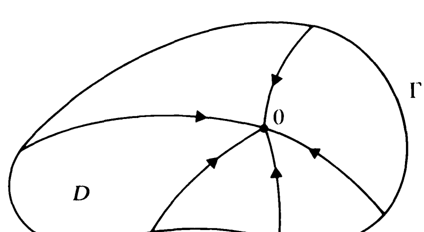

FIG. 1

Problems of this type were studied using probabilistic methods by Ventcel and Freidlin [9] and Friedman [2], who obtained some partial results. In recent papers of Grasman and Matkowsky [3], Matkowsky and Schuss [6], [7], a method was proposed to obtain the leading term in the asymptotic expansion of the solution of (1.1), (1.2). We shall prove their basic formula for the case of potential fields.

* Received by the Editors June 7, 1977, and in revised form November 14, 1977.
† Department of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel.

We shall make the following assumptions.

1) The functions $b_i(x)$ are smooth in x and the field of directions defined by the vector $\bar{b}=(b_1, \cdots, b_n)$ is a potential field, i.e., there exists a function $\psi(x)$ such that

$$b_i=\frac{\partial\psi}{\partial x_i}, \qquad 1 \leqq i \leqq n. \tag{1.4}$$

2) Everywhere in D

$$\sum_{i=1}^{n} b_i x_i \leqq 0 \tag{1.5}$$

while

$$\sum_{i=1}^{n} b_i^2(0)=0 \tag{1.6}$$

and in some neighborhood D^0 of the boundary

$$\sum_{i=1}^{n} b_i x_i \leqq -\lambda \sum_{i=1}^{n} x_i^2, \qquad \lambda=\text{const.}>0. \tag{1.7}$$

3) Let $\bar{\nu}=(\nu_1, \cdots, \nu_n)$ be the exterior normal to Γ. We assume that

$$(\bar{b}, \bar{\nu})=\sum_{i=1}^{n} b_i \nu_i<0 \quad \text{on } \Gamma. \tag{1.8}$$

In particular, if we consider the equation

$$\varepsilon \Delta u-\sum_{i=1}^{n} x_i \frac{\partial u}{\partial x_i}=0 \tag{1.9}$$

then conditions (1.4)–(1.7) are satisfied with $\psi(x)=-r^2/2$ $(r^2=\sum x_i^2)$.

Note that if the direction of the characteristics is opposite to that shown in Fig. 1, the Vishik–Lyusternik (V.–L.) method is applicable, $u_\varepsilon(x) \rightarrow V(x)$, $\forall x \neq 0$, where $V(x)$ is constant along the characteristics and takes the values $\varphi(x)$ on Γ.

The main result of the present paper is a proof of the following theorem.

THEOREM. *Assume that conditions* (1.4)–(1.8) *hold, and let $u_\varepsilon(x)$ be a solution of the boundary-value problem* (1.1), (1.2). *Then $u_\varepsilon(x) \rightarrow C_0$ uniformly on any compact subset of D and*

$$C_0=\lim_{\varepsilon \rightarrow 0} \frac{\int_\Gamma (\bar{b}, \bar{\nu}) \varphi \exp(\psi \varepsilon^{-1})\, ds}{\int_\Gamma (\bar{b}, \bar{\nu}) \exp(\psi \varepsilon^{-1})\, ds}. \tag{1.10}$$

The proof is based on several lemmas, to be proved below. Formula (1.10) was first derived by Matkowsky and Schuss [6], [7] using formal methods.

Examples of the application of formula (1.10) may be found in [3] and [8]. As indicated in these papers, the integrals in (1.10) are Laplace-type integrals and thus the major contributions come from the points at which $\psi(x)$ achieves its maximum on Γ.

For example, for equation (1.9) with $n=2$, $\psi(x)$ achieves its maximum on Γ at the points of contact of Γ with the largest circle centered at the origin and inscribed in D. If contact occurs along an arc (or a union of arcs) S of Γ, then

$$C_0=\frac{\int_S \varphi(x(s))\, ds}{\int_S ds}.$$

If contact occurs at distinct points (x_1^i, x_2^i) $(i=1,\cdots,m)$ on the boundary, only the l points of highest contact are considered. In that case,

$$C_0=\frac{\sum_{i=1}^l d_i\varphi(x_1^i, x_2^i)}{\sum_{i=1}^l d_i}.$$

The coefficients d_i are computed explicitly in [3] and [7].

In all formulations and proofs in the sequel, we shall use the following functions of the "boundary-layer" type, constructed by the V.–L. method. Let m be a fixed integer. Applying the iterative process described in [10], we see that there is a smooth function $v_\varepsilon(x)$, different from zero only in some strip D^1 around Γ, and having the form

$$v_\varepsilon(x)=h_\varepsilon(x)\,e^{-g(x)/\varepsilon} \tag{1.11}$$

where

$$h_\varepsilon(x)|_\Gamma=\varphi(x), \qquad h_\varepsilon(x)=0 \quad \text{outside } D^1, \tag{1.12}$$

$$g(x)|_\Gamma=0, \qquad \frac{\partial g}{\partial\nu}\bigg|_\Gamma=(\bar b, \bar\nu), \qquad g(x)>0 \quad \text{in } D^1. \tag{1.13}$$

Moreover,

$$L_\varepsilon v_\varepsilon=\varepsilon^m\rho(x,\varepsilon), \qquad |\rho|\leqq B, \tag{1.14}$$

$$\left|\frac{\partial h}{\partial\nu}\right|\leqq C, \tag{1.15}$$

B and C depend only on the smoothness of the boundary and the coefficients of the equation. Similarly, there is a function $v_\varepsilon^{(1)}(x)$ of the form

$$v_\varepsilon^{(1)}=h_\varepsilon^{(1)}(x)\,e^{-g(x)/\varepsilon} \tag{1.16}$$

where

$$h_\varepsilon^{(1)}(x)|_\Gamma=1, \qquad h_\varepsilon^{(1)}(x)=0 \quad \text{outside } D^1, \tag{1.17}$$

$$L_\varepsilon v_\varepsilon^{(1)}=\varepsilon^m\rho^{(1)}, \qquad |\rho^{(1)}|\leqq B^{(1)}, \tag{1.18}$$

$$\left|\frac{\partial h_\varepsilon^{(1)}}{\partial\nu}\right|\leqq C^{(1)}. \tag{1.19}$$

2. LEMMA 1. *Let u_ε be a solution of the equation*

$$\varepsilon\Delta u+\sum_{i=1}^n b_i\frac{\partial u}{\partial x_i}=0, \qquad |u|\leqq M, \tag{2.1}$$

in D and assume that (1.6) *is true.*

Then for $|x|\leqq\sqrt{\varepsilon}$,

$$|u_\varepsilon(x)-u_\varepsilon(0)|\leqq M_1\left(\frac{|x|}{\sqrt{\varepsilon}}\right), \tag{2.2}$$

where M_1 depends on n, M and the coefficients of the equation.

Proof. Let us introduce new variables

$$y_i=\frac{x_i}{\sqrt{\varepsilon}}, \qquad (i=1,2,\cdots,n). \tag{2.3}$$

From (2.1), we have

$$\sum \frac{\partial^2 u}{\partial y_i^2} + \sum \beta_i(y, \varepsilon)\frac{\partial u}{\partial y_i} = 0 \tag{2.4}$$

where $\beta_i(y, \varepsilon) = b_i(y\sqrt{\varepsilon})/\sqrt{\varepsilon}$. It is clear that the ball $|x| \leqq \sqrt{\varepsilon}$ is stretched into the ball $|y| \leqq 1$. Let $S_1 = \{y : |y| \leqq 1\}$, $S_2 = \{y : |y| \leqq 2\}$. From the smoothness assumptions concerning $b_i(x)$ and from (1.6) it follows that $|b_i(y\sqrt{\varepsilon})| \leqq \beta\sqrt{\varepsilon}|y|$ for some $\beta > 0$ and so

$$|\beta_i(y, \varepsilon)| \leqq \beta|y|. \tag{2.5}$$

The inequality (2.5) means that in S_2 the coefficients $\beta_i(y, \varepsilon)$ are bounded uniformly with respect to ε. The same is also true for the derivatives of β_i. Now, applying Schauder estimates to the solution of (2.4) in S_2, we get

$$\left|\frac{\partial u}{\partial y_i}\right| \leqq M_1 \quad \text{for } y \in S_1$$

and hence

$$\left|\frac{\partial u}{\partial x_i}\right| \leqq \frac{M_1}{\sqrt{\varepsilon}} \quad \text{for } |x| \leqq \sqrt{\varepsilon}. \tag{2.6}$$

From (2.6) follows (2.2).

LEMMA 2. *Assume that conditions* (1.4)–(1.8) *hold,* $v_\varepsilon(x) = h_\varepsilon e^{-g/\varepsilon}$ *and* $v_\varepsilon^{(1)}(x) = h_\varepsilon^{(1)} e^{-g/\varepsilon}$ *are functions of the "boundary-layer" type satisfying conditions* (1.11)–(1.15) *and* (1.16)–(1.19) *for* $m = n$. *Then the solution* $u_\varepsilon(x)$ *of problem* (1.1), (1.2) *admits the following asymptotic expansion in* D:

$$u_\varepsilon(x) = u_\varepsilon(0) + v_\varepsilon(x) - u_\varepsilon(0)v_\varepsilon^{(1)}(x) + o(1) \qquad (\varepsilon \to 0) \tag{2.7}$$

where $o(1)$ *is uniform in* D *as* $\varepsilon \to 0$.

Proof. By Lemma 1, for $|x| \leqq \varepsilon$,

$$|u_\varepsilon(x) - u_\varepsilon(0)| \leqq M_1\varepsilon^{1/2}, \tag{2.8}$$

i.e., expansion (2.7) is valid in the ball $|x| \leqq \varepsilon$. Consequently, it will suffice to verify the validity of (2.7) in the domain

$$D_\varepsilon = \{x \in D, |x| > \varepsilon\}.$$

Note that by the maximum principle

$$|u_\varepsilon(x)| \leqq M = \max_{x \in 1} |\varphi(x)|.$$

Set

$$w(x, \varepsilon) = \varepsilon\left(1 - \frac{\varepsilon^{n-2}}{r^{n-2}}\right) \quad \text{for } n \geqq 3, \tag{2.9}$$

$$w(x, \varepsilon) = \varepsilon\left(1 - \frac{\ln r}{\ln \varepsilon}\right) \quad \text{for } n = 2, \tag{2.10}$$

where $r=|x|$, $x\in D_\varepsilon$. Clearly,

$$w(x,\varepsilon)|_{r=\varepsilon}=0,\qquad w(x,\varepsilon)=O(\varepsilon)\qquad(\varepsilon\to 0) \tag{2.11}$$

$$L_\varepsilon w=\varepsilon\Delta w+\sum b_i\frac{\partial w}{\partial x_i}=\frac{1}{r}\frac{dw}{dr}\sum b_ix_i. \tag{2.12}$$

By (2.9) and (2.10),

$$\frac{dw}{dr}=(n-2)\frac{\varepsilon^{n-1}}{r^{n-1}}\quad\text{for } n\geqq 3, \tag{2.13}$$

$$\frac{dw}{dr}=-\frac{\varepsilon}{\ln\varepsilon}\frac{1}{r},\qquad n=2. \tag{2.14}$$

Using conditions (1.5) and (1.7), we obtain from (2.12)–(2.14)

$$L_\varepsilon w\leqq 0\quad\text{in } D_\varepsilon,\qquad n\geqq 2, \tag{2.15}$$

while in D^0 (cf. (1.7))

$$L_\varepsilon w\leqq -\lambda r^2\frac{1}{r}(n-2)\frac{\varepsilon^{n-1}}{r^{n-1}}=-\varepsilon^{n-1}\lambda(n-2)r^{2-n}\quad\text{for } n\geqq 3, \tag{2.16}$$

$$L_\varepsilon w\leqq\frac{\lambda\varepsilon}{\ln\varepsilon}\quad\text{for } n=2. \tag{2.17}$$

Now let $v_\varepsilon(x)$ and $v_\varepsilon^{(1)}(x)$ be the functions indicated in the statement of the lemma. We may assume here that condition (1.7) is satisfied in D^1, i.e., $D^1\subset D^0$. Then

$$v_\varepsilon(x)=v_\varepsilon^{(1)}(x)=0\quad\text{in } D_\varepsilon\backslash D^0 \tag{2.18}$$

and by (1.14) and (1.18)

$$L_\varepsilon v_\varepsilon=O(\varepsilon^n),\qquad L_\varepsilon v_\varepsilon^{(1)}=O(\varepsilon^n)\qquad(\varepsilon\to 0)\quad\text{in } D^0. \tag{2.19}$$

Denote

$$z_+(x)=v_\varepsilon(x)-u_\varepsilon(0)v_\varepsilon^{(1)}(x)+w(x,\varepsilon).$$

By virtue of (2.9), (2.10) and (2.18),

$$z_+(x)|_{r=\varepsilon}=0. \tag{2.20}$$

In addition, by (1.12) *and* (1.17),

$$z_+(x)|_\Gamma=\varphi(x)-u_\varepsilon(0)+w(x,\varepsilon)|_\Gamma$$

whence, in view of (2.11),

$$z_+(x)|_\Gamma=\varphi(x)-u_\varepsilon(0)+O(\varepsilon),\qquad(\varepsilon\to 0). \tag{2.21}$$

Next, it follows from (2.15)–(2.19) that, in $D_\varepsilon\backslash D^0$,

$$L_\varepsilon z_+\leqq 0 \tag{2.22}$$

and in D^0

$$L_\varepsilon z_+\leqq -\lambda\varepsilon^{n-1}(n-2)r^{2-n}+O(\varepsilon^n),\qquad n\geqq 3,$$

$$L_\varepsilon z_+\leqq\frac{\lambda\varepsilon}{\ln\varepsilon}+O(\varepsilon^2),\qquad n=2.$$

From the last two inequalities we see that, if ε_0 is sufficiently small,

$$L_\varepsilon z_+ \leqq 0 \quad \text{in } D^0, \quad \forall \varepsilon \leqq \varepsilon_0. \tag{2.23}$$

Combining (2.22) and (2.23), we find that

$$L_\varepsilon z_+ \leqq 0, \quad \forall \varepsilon \leqq \varepsilon_0 \text{ in } D_\varepsilon \tag{2.24}$$

hence

$$L_\varepsilon[u_\varepsilon(x) - u_\varepsilon(0) - z_+(x, \varepsilon)] \geqq 0 \tag{2.25}$$

in D_ε. It follows from (2.20), (2.21) and (2.8) that for $|x| = \varepsilon$,

$$u_\varepsilon(x) - u_\varepsilon(0) - z_+(x, \varepsilon) \leqq M_1 \varepsilon^{1/2}, \tag{2.26}$$

and for $x \in \Gamma$,

$$u_\varepsilon(x) - u_\varepsilon(0) - z_+(x, \varepsilon) = O(\varepsilon) \qquad (\varepsilon \to 0). \tag{2.27}$$

By the maximum principle, see [8], it follows from (2.25)–(2.27) that in D_ε

$$u_\varepsilon(x) - u_\varepsilon(0) - z_+(x, \varepsilon) \leqq O(\varepsilon^{1/2})$$

that is,

$$u_\varepsilon(x) - u_\varepsilon(0) - v_\varepsilon(x) + u_\varepsilon(0) v_\varepsilon^{(1)}(x) \leqq O(\varepsilon^{1/2}). \tag{2.28}$$

Applying similar reasoning to the function

$$z_-(x) = v_\varepsilon(x) - u_\varepsilon(x) - u_\varepsilon(0) v_\varepsilon^{(1)}(x) - w(x, \varepsilon)$$

instead of z_+, we obtain

$$u_\varepsilon(x) - u_\varepsilon(0) - v_\varepsilon(x) + u_\varepsilon(0) v_\varepsilon^{(1)}(x) \geqq O(\varepsilon^{1/2}). \tag{2.29}$$

The conclusion of Lemma 2 now follows from (2.28) and (2.29).

LEMMA 3. *Let*

$$L_\varepsilon u = f_1(x, \varepsilon), \quad u|_\Gamma = 0, \quad |f_1| < K_1, \quad |u| < K_2$$

in D, where K_1 and K_2 are constants independent of ε. Then

$$\left|\frac{\partial u}{\partial \nu}\right|_\Gamma \leqq \frac{K_3}{\varepsilon}. \tag{2.30}$$

The proof of this lemma is quite simple: one chooses a suitable barrier function and applies the maximum principle. For example, a good choice for the barrier function is

$$\tilde{u}(x) = W(x) + (K_2 + \kappa)[1 - h_\varepsilon^{(1)}(x)\, e^{-2g/\varepsilon}],$$

where $h_\varepsilon^{(1)}(x)$ and $g(x)$ are functions with properties (1.13), (1.17), (1.18), constructed by the V.–L. method, $W(x)$ a function satisfying the equation

$$\sum b_i \frac{\partial W}{\partial x_i} = -K_1 - 1,$$

and the condition $W|_\Gamma = 0$, $\kappa = \max_{x \in D^1} |W(x)|$. It is readily seen that in the strip D^1, if ε is sufficiently small,

$$L_\varepsilon(\tilde{u} \pm u) \leqq 0$$

and on the boundary of D^1,

$$\tilde{u} \pm u \geqq 0.$$

It follows from the last two inequalities by the maximum principle [8] that

$$\tilde{u} \pm u \geqq 0 \quad \text{in } D^1,$$

and so

$$\frac{\partial(\tilde{u} \pm u)}{\partial \nu} \leqq 0, \qquad \left|\frac{\partial u}{\partial u}\right| \leqq -\frac{\partial \tilde{u}}{\partial \nu}. \tag{2.31}$$

Inequality (2.30) now follows from (2.31) with the help of (1.13) and (1.19).

3. Proof of the theorem. Let

$$Z_\varepsilon(x) = u_\varepsilon(x) - u_\varepsilon(0) - v_\varepsilon(x) - u_\varepsilon(0) v_\varepsilon^{(1)}(x) \tag{3.1}$$

where $v_\varepsilon(x)$ and $v_\varepsilon^{(1)}(x)$ are the functions defined in the conditions of Lemma 2. By Lemma 2 and (2.19),

$$Z_\varepsilon(x) \to 0, \qquad L_\varepsilon Z_\varepsilon = O(\varepsilon^n) \quad \text{as } \varepsilon \to 0.$$

In addition,

$$Z_\varepsilon|_\Gamma = 0.$$

Set $\alpha(\varepsilon) = \max_D \{|Z_\varepsilon|, |L_\varepsilon Z_\varepsilon|\}$. Applying Lemma 3 to the function $Z_\varepsilon(x)/\alpha(\varepsilon)$, we obtain

$$\varepsilon \left|\frac{\partial Z_\varepsilon}{\partial \nu}\right|\Bigg|_\Gamma \leqq K\alpha(\varepsilon) \to 0 \quad \text{as } \varepsilon \to 0. \tag{3.2}$$

Using (1.4), we readily check that

$$L_\varepsilon^* e^{\psi/\varepsilon} = 0 \tag{3.3}$$

where $L_\varepsilon^* f = \varepsilon \Delta f - \sum (\partial b_i f)/(\partial x_i)$ (cf. [7] and [5]).

Integrating by parts in (3.3), we obtain

$$\begin{aligned} &\int_D (e^{\psi/\varepsilon} L_\varepsilon u_\varepsilon - u_\varepsilon L_\varepsilon^* e^{\psi/\varepsilon})\, dx \\ &\qquad = \varepsilon \int_\Gamma \left(e^{\psi/\varepsilon} \frac{\partial u_\varepsilon}{\partial \nu} - u_\varepsilon \frac{\partial e^{\psi/\varepsilon}}{\partial \nu}\right) ds \\ &\qquad\quad + \int_\Gamma u_\varepsilon e^{\psi/\varepsilon} (\bar{b}, \bar{\nu})\, ds = 0. \end{aligned} \tag{3.4}$$

It follows from (3.1), (3.2), (1.15) and (1.19) that

$$\begin{aligned} \varepsilon \frac{\partial u_\varepsilon}{\partial \nu} &= \varepsilon\left(\frac{\partial h}{\partial \nu} - u_\varepsilon(0) \frac{\partial h^{(1)}}{\partial \nu}\right) - \varepsilon(h - u_\varepsilon(0) h^{(1)}) \frac{1}{\varepsilon} \frac{\partial g}{\partial \nu} \\ &\quad + \varepsilon \frac{\partial Z_\varepsilon}{\partial \nu} = (-h + u_\varepsilon(0) h^{(1)}) \frac{\partial g}{\partial \nu} + o(1). \end{aligned}$$

Hence, by (3.4), we have

$$\int_\Gamma e^{\psi/\varepsilon}\left[\frac{\partial g}{\partial \nu}(u_\varepsilon(0)h^{(1)}-h)-u_\varepsilon\frac{\partial \psi}{\partial \nu}+u_\varepsilon(\bar{b},\bar{\nu})+o(1)\right]ds=0,$$

and so, in view of the fact that

$$h|_\Gamma=\varphi(x),\quad h^{(1)}|_\Gamma=1,\quad \frac{\partial \psi}{\partial \nu}=(\bar{b},\bar{\nu}),\quad \frac{\partial g}{\partial \nu}=(\bar{b},\bar{\nu}),$$

$$\int_\Gamma e^{\psi/\varepsilon}(\bar{b},\bar{\nu})(u_\varepsilon(0)-\varphi(x))\,ds=o(1)\int_\Gamma e^{\psi/\varepsilon}\,ds\qquad(\varepsilon\to 0). \tag{3.5}$$

Dividing both sides of (3.5) by $\int e^{\psi/\varepsilon}\,ds$, we see that as $\varepsilon\to 0$,

$$u_\varepsilon(0)\frac{\int e^{\psi/\varepsilon}(\bar{b},\bar{\nu})\,ds}{\int e^{\psi/\varepsilon}\,ds}-\frac{\int \varphi\, e^{\psi/\varepsilon}(\bar{b},\bar{\nu})\,ds}{\int e^{\psi/\varepsilon}\,ds}\to 0$$

and consequently,

$$u_\varepsilon(0)-\frac{\int \varphi\, e^{\psi/\varepsilon}(\bar{b},\bar{\nu})\,ds}{\int e^{\psi/\varepsilon}(\bar{b},\bar{\nu})\,ds}\to 0$$

completing the proof of the theorem.

Remark. Assume that

i) conditions (1.6) and (1.8) hold,

ii) condition (1.7) is satisfied in some neighborhood D_0 of the origin and every characteristic of (1.3) enters D_0. Then by some slight modifications in the proof given above Lemma 2 can be generalized to the equation

$$\varepsilon\Big(\sum_{i,j}(a_{ij}u_{x_i})_{x_j}+\sum_i a_iu_{x_i}\Big)+\sum_i b_iu_{x_i}=0. \tag{3.6}$$

But in order to proceed to the proof of the Theorem for equation (3.6), one needs some additional considerations. We shall present that in the next paper. Nevertheless the proof of formula (1.10) follows in the same way as above if $a_i=0$ in (3.6) and if there exists a function $\psi(x)$ such that $b_i=\sum a_{ij}\psi_{x_j}$.

REFERENCES

[1] W. ECKHAUS, *Matched Asymptotic Expansions and Singular Perturbations*, North-Holland, Amsterdam, 1973.

[2] A. FRIEDMAN, *Stochastic Differential Equations*, Academic Press, New York, 1975.

[3] J. GRASMAN AND B. MATKOWSKY, *A variational approach to boundary value problems and resonance for ordinary and partial differential equations*, SIAM, J. Appl. Math. 32 (1977), pp. 588–597.

[4] N. LEVINSON, *The first boundary value problem for* $\varepsilon\Delta u+A(x,y)u_x+B(x,y)u_y+C(x,y)u=D(x,y)$ *for small* ε, Ann. of Math. (2), 51 (1950), pp. 428–445.

[5] D. LUDWIG, *Persistence of dynamical systems under random perturbations*, SIAM Rev., 17 (1975), pp. 605–640.

[6] B. J. MATKOWSKY AND Z. SCHUSS, *On the problem of exit*, Bull. Amer. Math. Soc. 82 (1976), pp. 321–324.

[7] ———, *The exit problem for randomly perturbed dynamical systems*, SIAM J. Appl. Math., 33 (1977), pp. 365–382.

[8] M. H. PROTTER AND H. F. WEINBERGER, *Maximum Principles in Differential Equation*, Prentice-Hall, Englewood Cliffs, NJ, 1967.

[9] A. D. VENTCEL AND M. I. FREIDLIN, *On small perturbations of dynamical systems*, Uspehi Mat. Nauk, 25 (1970), no. 1 (151), pp. 3–55 = Russian Math. Surveys, no. 1, 25 (1970), pp. 1–56.

[10] M. I. VISIK AND L. A. LYUSTERNIK, *Regular degeneration and boundary layers for linear differential equations with small parameters*, Uspehi Mat. Nauk, 12 (1957) = Amer. Mat. Soc. Translations Series 2, 20 (1962), pp. 239–364.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0002 $01.00/0

CRITERIA OF LIMIT CIRCLE TYPE FOR NONLINEAR DIFFERENTIAL EQUATIONS*

PAUL W. SPIKES†

Abstract. Criteria which ensure that all solutions of a perturbed second order differential equation satisfy certain integrability properties are obtained. For special cases of the equation these criteria reduce to sufficient conditions for all solutions to be L^p-bounded, $p>0$. In particular, we obtain limit circle criteria for the perturbed linear equation. Examples are included to illustrate the results.

1. Introduction. In this paper we are concerned with establishing criteria which guarantee that all solutions of the nonlinear differential equation

$$(*)\qquad (a(t)x')'+q(t)f(x)=r(t,x)$$

satisfy the integrability condition

$$(**)\qquad \int^{\infty} x(v)f(x(v))\,dv<\infty,$$

provided $xf(x)>0$ for $x\neq 0$.

Since such criteria reduce to sufficient conditions for L^{2n}-boundedness of all solutions of $(*)$ when $f(x)\equiv x^{2n-1}$, n a positive integer, results of this type are known for various special cases of $(*)$. In particular, the problem of finding sufficient conditions for L^2-boundedness of solutions of $(*)$ for the case $f(x)\equiv x$ has been investigated extensively. For example see [2], [4]–[7], [9]–[12], [16]–[21] and the references contained therein. Our Theorem 4 gives a new result for this case.

When $f(x)\not\equiv x$ much less is known regarding integrability properties of the solutions of $(*)$. Furthermore, such results are mostly of the limit-point type—sufficient conditions for $(*)$ to have one or more solutions that are not L^p-bounded. We cite as examples the work of Atkinson [1], Burlak [3], Hallam [8], and Suyemoto and Waltman [15]. Exept for the result indicated by Atkinson [1, p. 311] for the equation

$$x''+t^{\gamma}x^{2n-1}=0$$

(γ a constant), the author knows of no criteria other than those in [14] which imply $(**)$ when $f(x)\not\equiv x$. In addition to other differences, the criteria established in this paper do not require the strong monotonicity condition imposed on the product aq in [14].

We also obtain, as a corollary to our main result, sufficient conditions to ensure that all solutions of $(*)$ tend to zero as $t\to\infty$.

2. Integrability and asymptotic properties. Consider the equation

$$(1)\qquad (a(t)x')'+q(t)f(x)=r(t,x)$$

where $a,q\colon[t_0,\infty)\to R$, $f\colon R\to R$, and $r\colon[t_0,\infty)\times R\to R$ are continuous, $a',q'\in AC_{\mathrm{loc}}[t_0,\infty)$, $a'',q''\in L^2_{\mathrm{loc}}[t_0,\infty)$, $a(t)>0$, $q(t)>0$, and $xf(x)>0$ for $x\neq 0$.

We define $F(x)=\int_0^x f(s)\,ds$, and for any function g we let $g(t)_+=\max\{g(t),0\}$ and $g(t)_-=\max\{-g(t),0\}$. The following conditions will also be utilized as needed.

---

* Received by the editors June 7, 1977.

† Department of Mathematics, Mississippi State University, Mississippi State, Mississippi 39762. This work was supported by the Mississippi State University Biological and Physical Sciences Research Institute.

Assume that there exist positive constants A, B, C, and $b \leqq 2$, a nonnegative function $h \in C[t_0, \infty)$, and a positive function $H \in C'[t_0, \infty)$ such that

$$H'(t)_+/H(t) \leqq (2-b)(a(t)q(t))'_+/(2a(t)q(t)), \tag{2}$$

$$H'(t)_+/H(t) \leqq bB(a(t)q(t))'_+/(4a(t)q(t)), \tag{3}$$

$$BF(x) \leqq xf(x) \leqq CF(x), \tag{4}$$

$$|r(t, x)| \leqq h(t)(F(x))^{1/2}, \tag{5}$$

and

$$x^2 \leqq 2F(x) + 2A. \tag{6}$$

Let

$$Q(t) = (2-b)[(a(t)q(t))']^2/(4a^{3/2}(t)q^{5/2}(t)) + [(a(t)q(t))'/(a^{1/2}(t)q^{3/2}(t))]'$$

and assume that

$$\int_{t_0}^{\infty} |Q(v)|H(v)\,dv < \infty, \tag{7}$$

$$\int_{t_0}^{\infty} [(a(v)q(v))'_-/(a(v)q(v))]\,dv < \infty, \tag{8}$$

$$\int_{t_0}^{\infty} [h(v)/(a(v)q(v))^{1/2}]\,dv < \infty, \tag{9}$$

$$\int_{t_0}^{\infty} |Q(v)|\,dv < \infty, \tag{10}$$

and

$$\int_{t_0}^{\infty} [1/H(v)]\,dv < \infty. \tag{11}$$

We note that since condition (5) will be required in the hypothesis of all our results, it follows from [13, Thm. 1] that all solutions of (1) exist on $[t_0, \infty)$. We now give our principal result—sufficient conditions for every solution $x(t)$ of (1) to satisfy

$$\int_{t_0}^{\infty} x(v)f(x(v))\,dv < \infty. \tag{I}$$

THEOREM 1. *If conditions* (2)–(11) *hold, then every solution* $x(t)$ *of* (1) *satisfies* (I).

Proof. Proceeding by techniques similar to those used by Burton and Patula in [4, Thm. 1], let $y(s) = x(t)$ where $s = \int_{t_0}^{t} [q(v)/a(v)]^{1/2}\,dv$. Define

$$R(t) = (a(t)q(t))'/(4a^{1/2}(t)q^{3/2}(t))$$

and let "$\cdot$" $= d/ds$. Then (1) becomes

$$\ddot{y} + 2R(t)\dot{y} + f(y) = r(t, y)/q(t).$$

Now consider the equivalent system

$$\begin{aligned} \dot{y} &= z - bR(t)y \\ \dot{z} &= (b-2)R(t)z + b[(2-b)R^2(t) + \dot{R}(t)]y - f(y) + r(t, y)/q(t), \end{aligned}$$

and define V and W by

$$V(s)=z^2(s)/2+F(y(s))$$

and

$$W(s)=V(s)H(t).$$

Then

$$\begin{aligned}\dot V(s)&=z(s)\dot z(s)+\dot y(s)f(y(s))\\&=(b-2)R(t)z^2(s)+b[(2-b)R^2(t)+\dot R(t)]y(s)z(s)\\&\quad -z(s)f(y(s))+z(s)r(t,y(s))/q(t)\\&=(b-2)R(t)z^2(s)-bR(t)y(s)f(y(s))+b[(2-b)R^2(t)+\dot R(t)]y(s)z(s)\\&\quad +z(s)r(t,y(s))/q(t)\end{aligned}$$

and hence

$$\begin{aligned}\dot W(s)=[(b-2)R(t)H(t)+\dot H(t)/2]z^2(s)\\ -bR(t)y(s)f(y(s))H(t)+\dot H(t)F(y(s))\\ +G(t)y(s)z(s)+H(t)z(s)r(t,y(s))/q(t)\end{aligned}$$

where $G(t)=b[(2-b)R^2(t)+\dot R(t)]H(t)$. From (2) we have

$$\begin{aligned}&(b-2)R(t)H(t)+\dot H(t)/2\\&\quad=(b-2)H(t)R(t)_+-(b-2)H(t)R(t)_-\\&\qquad+[H'(t)_+-H'(t)_-]a^{1/2}(t)/(2q^{1/2}(t))\\&\quad\leqq a^{1/2}(t)H(t)[H'(t)_+/H(t)-(2-b)(a(t)q(t))'_+/(2a(t)q(t))]/(2q^{1/2}(t))\\&\qquad+(2-b)H(t)R(t)_-\\&\quad\leqq(2-b)H(t)R(t)_-.\end{aligned}$$

Also it follows from (3) and (4) that

$$\begin{aligned}&\dot H(t)F(y(s))-bR(t)y(s)f(y(s))H(t)\\&\quad\leqq F(y(s))a^{1/2}(t)H'(t)_+/q^{1/2}(t)-by(s)f(y(s))H(t)R(t)_+\\&\qquad+by(s)f(y(s))H(t)R(t)_-\\&\quad\leqq a^{1/2}(t)H(t)y(s)f(y(s))[H'(t)_+/H(t)-bB(a(t)q(t))'_+/(4a(t)q(t))]/(Bq^{1/2}(t))\\&\qquad+bCH(t)F(y(s))R(t)_-\\&\quad\leqq bCH(t)F(y(s))R(t)_-,\end{aligned}$$

and therefore

$$\begin{aligned}\dot W(s)\leqq(2-b)H(t)z^2(s)R(t)_-+bCH(t)F(y(s))R(t)_-\\+G(t)y(s)z(s)+H(t)z(s)r(t,y(s))/q(t).\end{aligned}$$

Putting $D=\max\{2(2-b),bC\}$ and using (5) we obtain

$$\dot W(s)\leqq DW(s)R(t)_-+|G(t)||y(s)z(s)|+H(t)|z(s)|h(t)(F(y(s)))^{1/2}/q(t).$$

Then (6), together with the inequalities

$$|z(s)y(s)| \leqq [z^2(s)+y^2(s)]/2 \quad \text{and} \quad |z(s)|(F(y(s)))^{1/2} \leqq V(s),$$

implies

$$\begin{aligned}\dot{W}(s) &\leqq DW(s)R(t)_- + |G(t)|\,W(s)/H(t) + h(t)W(s)/q(t) + A|G(t)| \\ &= A|G(t)| + P(t)W(s)\end{aligned}$$

where $P(t) = DR(t)_- + |G(t)|/H(t) + h(t)/q(t)$. Integrating we have

$$W(s) \leqq W(s_0) + A\int_{s_0}^{s} |G(k(u))|\,du + \int_{s_0}^{s} P(k(u))W(u)\,du.$$

Now

$$\int_{s_0}^{s} |G(k(u))|\,du = (b/4)\int_{t_1}^{t} |Q(v)|H(v)\,dv,$$

so (7) implies that there exists a constant $L_1 > 0$ such that

$$W(s) \leqq L_1 + \int_{s_0}^{s} P(k(u))W(u)\,du.$$

It then follows from Gronwall's inequality that

$$W(s) \leqq L_1 \exp\left(\int_{s_0}^{s} P(k(u))\,du\right).$$

Since

$$\begin{aligned}\int_{s_0}^{s} P(k(u))\,du &= (D/4)\int_{t_1}^{t} [(a(v)q(v))'_-/(a(v)q(v))]\,dv \\ &\quad + (b/4)\int_{t_1}^{t} |Q(v)|\,dv + \int_{t_1}^{t} [h(v)/(a(v)q(v))^{1/2}]\,dv,\end{aligned}$$

then (8)–(10) imply that $W(s)$ is bounded. Hence there exists a constant $M > 0$ such that $F(x(t))H(t) \leqq M$ for $t \geqq t_0$. To complete the proof, we use (4) and (11) to obtain

$$\begin{aligned}\int_{t_0}^{\infty} x(v)f(x(v))\,dv &\leqq CM\int_{t_0}^{\infty} [1/H(v)]\,dv \\ &< \infty.\end{aligned}$$

Remark. Sufficient conditions for all solutions of (1) to satisfy (I) have also been obtained in [14, Thms. 5–7], but the hypotheses of these theorems differ in several respects from those of Theorem 1. In particular all the results in [14] require $(a(t)q(t))' \geqq 0$, $H \in C^2[t_0, \infty)$, $H'(t) \geqq 0$, and $H(t) \to \infty$ as $t \to \infty$ which are considerably more restrictive than the requirements imposed on these quantities in Theorem 1. For example, Theorem 1 shows that every solution of the equation

$$x'' + q(t)x^3 = r(t, x), \qquad t > 2,$$

where $q(t) = t^2\{1 + (4/5)[\sin(\ln t) - \cos(\ln t)/2] + 1/t^3\}$ and $r(t, x) = x^2(\ln t) \cdot [\tanh(t^2x)]/[t(1+\sin^2 x)]$ is $L^4(2, \infty)$. This can be verified by taking $b = 2/3$, $A = 1$, $B = C = 4$, $h(t) = (2\ln t)/t$, and $H(t) = q^{2/3}(t)$ in Theorem 1. Notice that none of Theorems 5–7 in [14] apply to this example since $q'(t) = 2t[1 + \sin(\ln t)] - 1/t^2$ is negative for arbitrarily large values of t.

The next result has simpler hypotheses than Theorem 1, but stronger conditions are placed on the product aq.

COROLLARY 2. *Let $Q(t)$ be defined as in Theorem* 1, *and let conditions* (4)–(6) *and* (8)–(9) *hold. If there is a constant $0<p<1$ such that*

$$\int_{t_0}^{\infty} [1/(a(v)q(v))^p]\, dv < \infty, \tag{12}$$

$$2p \leqq 2-b \quad \text{and} \quad 4p \leqq bB, \tag{13}$$

and

$$\int_{t_0}^{\infty} |Q(v)|(a(v)q(v))^p\, dv < \infty, \tag{14}$$

then every solution of (1) *satisfies* (I).

Proof. Let $H(t)=(a(t)q(t))^p$. Then (13) implies (2) and (3), (14) is equivalent to (7), (14) and (8) together imply (10), and (12) is equivalent to (11). Thus all the hypotheses of Theorem 1 are satisfied and the conclusion follows.

As an example to which Corollary 2 applies, consider the equation

$$x'' + t^{\gamma} x^{2n-1} = 0, \qquad t \geqq t_0 > 0, \tag{15}$$

where n is a positive integer and γ is a constant satisfying $\gamma > 1 + 1/n$. By taking $A=n$, $B=C=2n$, $p=[n\gamma+n-1]/(2n\gamma)$ and $b=[n\gamma+1-n]/(n\gamma)$, it is easy to see that Corollary 2 implies that every solution of (15) is in $L^{2n}(t_0,\infty)$. This example is of particular interest in view of the fact that Atkinson [1, p. 311] has pointed out that solutions of (15) are in $L^{2n}(t_0,\infty)$ only if $\gamma > 1+1/n$.

By replacing condition (11) in the hypotheses of Theorem 1 by

$$H(t) \to \infty \quad \text{as } t \to \infty, \tag{11'}$$

we obtain the following partial asymptotic stability result for (1).

COROLLARY 3. *If* (2)–(10) *and* (11′) *are satisfied, then every solution $x(t)$ of* (1) *satisfies $x(t)\to 0$ as $t\to\infty$.*

Proof. Let $x(t)$ be a solution of (1). From the proof of Theorem 1 we have $F(x(t)) \leqq M/H(t)$ for some constant $M>0$. Hence (11′) implies that $F(x(t))\to 0$ as $t\to\infty$, which in turn implies that $x(t)\to 0$ as $t\to\infty$.

For the special case

$$(a(t)x')' + q(t)x = r(t,x) \tag{1'}$$

of (1), notice that $F(x)=x^2/2$. Thus if (5) holds, then $|r(t,x)| \leqq |x|h(t)$. Moreover, it is not difficult to show that if (5) holds and $x(t)$ is a nontrivial solution of (1′), then $x(t)$ and $x'(t)$ are not zero simultaneously. These observations enable us to use (with simple modifications) the proof of Burton and Patula [4, Thm. 1] to obtain Theorem 4 below for solutions of (1′). For this purpose we define

$$S(t) = |[(a(t)q(t))']^2/(4a^{3/2}(t)q^{5/2}(t)) + [(a(t)q(t))'/(a^{1/2}(t)q^{3/2}(t))]'| \\ + 4h(t)/(a(t)q(t))^{1/2}.$$

THEOREM 4. *If in addition to* (5),

$$\int_{t_0}^{\infty} [1/(a(v)q(v))^{1/2}]\left[\exp \int_{t_0}^{v} (S(u)/4)\, du\right] dv < \infty \tag{16}$$

holds, then (I) *holds for every solution of* (1′).

Proof. Let $x(t)$ be a nontrivial solution of (1'). Proceeding exactly as in the proof of Theorem 1, with $b=1$ and $H(t)\equiv 1$, we have $W(s)=V(s)=[z^2(s)+y^2(s)]/2$ and

$$\dot V(s)=-R(t)[z^2(s)+y^2(s)]+[R^2(t)+\dot R(t)]z(s)y(s)+z(s)r(t,y(s))/q(t).$$

But $|z(s)y(s)|\leqq V(s)$ and $|z(s)||r(t,y(s))|\leqq h(t)V(s)/q(t)$, so

$$\dot V(s)\leqq[-2R(t)+|R^2(t)+\dot R(t)|+h(t)/q(t)]V(s).$$

Noticing that $x(t)$ and $x'(t)$ not simultaneously zero implies $V(s)>0$, we complete the proof by an argument similar to that used in the proof of Theorem 1 in [4].

Remark. If $a(t)\equiv 1$ and $r(t,x)\equiv 0$, then Theorem 4 reduces to Theorem 1 in [4]. Notice also that Theorem 4 cannot be deduced from either Theorem 1 or Corollary 2.

Results of type (I) have been obtained by a number of authors for equation (1'), particularly when $r(t,x)\equiv -h(t)x$. These results were obtained in almost all instances by requiring somewhat rigid growth conditions on the function h. Theorem 4 differs from these results in that the growth condition is placed on the quotient $h/(aq)^{1/2}$, and hence applies in cases not covered by previous results. We illustrate this with the example

$$(t^6x')'+[6t^4+(t^\gamma \sin t)/(1+x^2)]x=0 \tag{17}$$

where γ is a constant and $t>1$. Theorem 4 applies to (17) for any $\gamma<4$, whereas Theorem 1 in [2] applies only in case $\gamma<3$ and Theorem B in [18] only in case $\gamma<0$.

REFERENCES

[1] F. V. ATKINSON, *Nonlinear extensions of limit-point criteria*, Math. Z., 130 (1973), pp. 297–312.

[2] J. S. BRADLEY, *Comparison theorems for the square integrability of solutions of* $(r(t)y')'+q(t)y=f(t,y)$, Glasgow Math. J., 13 (1972), pp. 75–79.

[3] J. BURLAK, *On the non-existence of* L^2*-solutions of nonlinear differential equations*, Proc. Edinburgh Math. Soc., 14 (1965), pp. 257–268.

[4] T. A. BURTON AND W. T. PATULA, *Limit circle results for second order equations*, Monatsh. Math., 81 (1976), pp. 185–194.

[5] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators; Part II: Spectral Theory*, Interscience, New York, 1963.

[6] M. S. P. EASTHAM, *Limit-circle differential expressions of the second order with an oscillating coefficient*, Quart. J. Math. Oxford Ser. (2), 24 (1973), pp. 257–263.

[7] W. N. EVERITT, *On the limit-circle classification of second-order differential expressions*, Ibid., 23 (1972), pp. 193–196.

[8] T. G. HALLAM, *On the non-existence of* L^p*-solutions of certain nonlinear differential equations*, Glasgow Math. J., 8 (1967), pp. 133–138.

[9] D. B. HINTON, *Limit point-limit circle criteria for* $(py')'+qy=\lambda ky$, Lecture Notes in Mathematics 415, Springer, New York, 1974, pp. 173–183.

[10] I. W. KNOWLES, *On a limit-circle criterion for second-order differential operators*, Quart J. Math. Oxford Ser. (2), 24 (1973), pp. 451–455.

[11] ———, *On second-order differential operators of limit-circle type*, Lecture Notes in Mathematics 415, Springer, New York, 1974, pp. 184–187.

[12] W. T. PATULA AND J. S. W. WONG, *An* L^p*-analogue of the Weyl alternative*, Math. Ann., 197 (1972), pp. 9–28.

[13] P. W. SPIKES, *Some stability type results for a nonlinear differential equation*, Rend. Mat. (6), 9 (1976), pp. 259–271.

[14] ———, *On the integrability of solutions of perturbed non-linear differential equations*, Proc. Roy. Soc. Edinburgh Sect. A, 77 (1977), pp. 309–318.

[15] L. SUYEMOTO AND P. WALTMAN, *Extension of a theorem of A. Wintner*, Proc. Amer. Math. Soc., 14 (1963), pp. 970–971.

[16] H. WEYL, *Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörige Entwicklung willkürlicher Funktionen*, Math. Ann., 68 (1910), pp. 220–269.

[17] J. S. W. WONG, *Remarks on the limit-circle classification of second-order differential operators*, Quart. J. Math. Oxford Ser. (2). 24 (1973), pp. 423–425.

[18] ———, *Square integrable solutions of L^p perturbations of second order linear differential equations*, Lecture Notes in Mathematics 415, Springer, New York, 1974, pp. 282–292.

[19] ———, *Square integrable solutions of perturbed linear differential equations*, Proc. Roy. Soc. Edinburgh Sect. A, 73 (1974/75), pp. 251–254.

[20] A. ZETTL, *Square integrable solutions of $Ly = f(t, y)$*, Proc. Amer. Math. Soc., 26 (1970), pp. 635–639.

[21] ———, *Perturbation of the limit circle case*, Quart. J. Math. Oxford Ser. (2), 26 (1975), pp. 355–360.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0003 $01.00/0

CONSTANT LIMIT OF A SEQUENCE OF ITERATES*

W. F. SMYTH†

Abstract Let N be the set of all points on the complex sphere σ at which a sequence F of iterates of a meromorphic function f is normal. N is shown to be exactly those points in some neighborhood of which F breaks down uniformly into a finite number of fixed subsequences. Consider any domain $D \subseteq N$ together with all its images under f: $F(D)=\{D, f(D), f_2(D), \cdots\}$. Then, corresponding to $F(D)$, there exists a fixed finite integer $r>0$ such that F breaks down everywhere in $F(D)$ into r infinite subsequences $F_i=\{f_{nr+i} \mid n=0, 1, \cdots\}$, $i=0, \cdots, r-1$; and, if $f_r(z)$ is not identically z in $F(D)$, each subsequence f_i converges in each domain of $F(D)$ to a distinct constant limit point drawn from a set of distinct values

$$\{\alpha_0, \cdots, \alpha_{r-1}\} \subset F(D), \quad \text{where } \alpha_{i+1 (\text{mod } r)}=f(\alpha_i), \quad i=0, \cdots, r-1.$$

Introduction. The study of the global convergence properties of iterations of meromorphic, especially rational and entire, functions, as established by Julia [1] and Fatou [2], [3], [4], is little referred to in recent mathematical literature, even though the theory yields elegant, powerful, and suggestive results and leaves questions of considerable practical interest unresolved. It is an area which might usefully be surveyed and organized into an integrated theory: building on previous efforts in this direction by Cremer [5] and Montel [6]; incorporating later results by, for example, Rådström [7] and Talanov [8]; and making reference to cases of practical interest such as Brooker's example [9], [10]. Our objective in this paper is a more modest one: to settle one of the unresolved questions in this field: under what circumstances subsequences of the sequence of iterates of a meromorphic function converge to a constant limit function. It has long been suspected [5], [6] that the limit function was constant under quite general conditions, but no such result has ever been proved: a number of deceptive "counterexamples" [11], [12], [13] have perhaps served as red herrings to distract researchers attempting to formulate a constant limit theorem in appropriate form. As we shall see in this paper, the chief difficulty lies in specifying the circumstances under which the sequence of iterates of a single point gives rise to an infinity of limit points; the resolution of this difficulty provides us with a new criterion of normality: a sequence of iterates is normal at a point z if and only if there exists some neighborhood of z in which the sequence breaks down uniformly into a finite number of regular subsequences. We develop our results in three main sections:

(1) Some basic properties are established for a sequence F of iterates of a function f single-valued and continuous on a point set S of the complex sphere[1] σ. If $f(S) \subseteq S$, $f_{-1}(S) \subseteq S$, and every accumulation point of F is a point of S, it is shown in particular that the sequence $F(z)$ of iterates of a point $z \in S$ has an infinity of distinct accumulation points if and only if $F(z)$ gives rise to no fixpoints; and that $F(z)$ has a finite number r of accumulation points if and only if $F(z)$ breaks down into exactly r subsequences $\{f_{nr+i} \mid n=0, 1, \cdots\}$, $i=0, \cdots, r-1$, each convergent to a distinct one of the accumulation points of $F(z)$.

(2) The concept of a normal family of functions is defined, and the results of the previous section are applied to establish a new criterion of normality and to show that for a meromorphic function f the regular breakdown of F into subsequences holds uniformly in any domain in which F is normal.

* Received by the editors September 9, 1977.
† 84 Dupont Street, Toronto, Ontario, Canada M5R 1V2.
[1] The sphere and the chordal metric χ are used throughout, in order that convergence at infinity should be well defined. For "continuous" and "convergent" and "differentiable", therefore, read "spherically continuous", "spherically convergent", etc.

(3) An elementary argument then establishes that for a meromorphic function the limit function of any subsequence is in fact constant under quite general conditions.

In a final section we discuss some examples, counterexamples, and exceptions.

1. Properties of a sequence of iterates. Let f be a function defined on a point set $S \subseteq \sigma$. For $z \in S$ the terms of a sequence

$$F(z)=\{f_n(z)\}=\{f_n(z)\,|\,n=0,1,\cdots\}$$

may be specified recursively as follows:

$$f_0(z)=z; \qquad f_{n+1}(z)=f\cdot f_n(z), \qquad n=0,1,\cdots,$$

where if $f_n(z)\notin S$, $f_{n+1}(z)$ and all subsequent terms of $F(z)$ are undefined. $F(z)$ is called the *sequence of iterates* of f at z, and $f_n(z)$ the nth *consequent* of z under f. If for every nonnegative integer n, $f_n(z)\in S$, then $F(z)$ determines at least one *accumulation point* (and perhaps uncountably many of them); if f is single-valued in S, we may regard an accumulation point as the uniquely-determined *limit point* of a Cauchy sequence which is a *subsequence* of $F(z)$; a subsequence $F_k(z)$ is denoted by

$$F_k(z)=\{f_{n_i}(z)\}=\{f_{n_i}(z)\,|\,n_{i+1}>n_i,\ i=0,1,\cdots\},$$

where the n_i are nonnegative integers, k is in general any real number, and it is understood that $n_i=n_i(k)$. The sequences $F(z)$ and $F_k(z)$ will sometimes be called, and treated as, sets of points.

The *inverse function* of f is denoted f_{-1}. For $z\in S$ and integral $n\geqq 0$, $\{f_{-n}(z)\}$ is the (possibly empty) set of points ζ such that $f_n(\zeta)=z$. An individual or representative element of $\{f_{-n}(z)\}$ is denoted $f_{-n}(z)$ and called an nth *antecedent* of z under f. If f is single-valued in S, if $f_n(z)\in S$ for sufficiently large positive or negative values of the integer n, and if $s\geqq t\geqq 0$, then

(a) $f_s\cdot f_t(z)=f_{s+t}(z)=f_t\cdot f_s(z)$,

(b) $f_s\cdot f_{-t}(z)=f_{s-t}(z)\in\{f_{-t}\cdot f_s(z)\}$,

(c) $\{f_t\cdot f_{-s}(z)\}=\{f_{t-s}(z)\}\subseteq\{f_{-s}\cdot f_t(z)\}$.

In this section we restrict ourselves to functions f and point sets S such that for $z\in S$

(1) $f(z)$ is single-valued and continuous;

(2) $f(z)\in S$ and $\{f_{-1}(z)\}\subset S$;

(3) every accumulation point of $F(z)$ is a point of S.

We shall see in the next section that these are some of the important conditions satisfied not only by the set N of points at which the sequence F of iterates of a mermorphic function is normal, but also by the set E of points at which F is not normal.

LEMMA 1. *Denote by* $\alpha=\lim_{i\to\infty} f_{n_i}(z)$ *an accumulation point of* $F(z)$, $z\in S$. *Then* $f(\alpha)$ *and* $f_{-1}(\alpha)$ *are also accumulation points of* $F(z)$, *where*

$$f(\alpha)=\lim f_{n_i+1}(z), \qquad f_{-1}(\alpha)=\lim_{j\to\infty} f_{m_j}(z), \quad \textit{and}$$
$$\{f_{m_j+1}(z)\}\subseteq\{f_{n_i}(z)\}.$$

Proof. By (2), $F(z)$ is an infinite sequence, hence possesses an accumulation point.

Therefore α and $\{f_{n_i}(z)\}$ exist. By (1) and (3), f is continuous at α. Then given $\delta > 0$, we may find $I = I(\delta)$ such that for all $i > I$,

$$\chi[f(\alpha), f \cdot f_{n_i}(z)] < \delta.$$

Hence by (a), $f(\alpha)$ is the (unique) limit point of the subsequence $\{f_{n_i+1}(z)\} \subseteq F(z)$, and therefore an accumulation point of $F(z)$.

Next, let β be an accumulation point of the infinite subsequence $\{f_{n_i-1}(z)\} \subseteq F(z)$. Then by (1) and (3), f is continuous at β. Then given $\delta > 0$, we may find $I' = I'(\delta)$ such that for all $i > I'$,

$$\chi[f(\beta), f \cdot f_{m_j}(z)] < \delta,$$

where $\{f_{m_j}(z)\} \subseteq \{f_{n_i-1}(z)\}$ is the subsequence with limit point β. But by (a) this tells us that an infinite subsequence $\{f_{m_j+1}(z)\} \subseteq \{f_{n_i}(z)\}$ has limit point $f(\beta)$. Hence $f(\beta) = \alpha$, so that $\beta = f_{-1}(\alpha)$.

We note that perhaps $f(\alpha) = \alpha$; in this case α is called a *fixpoint* of f. In case no fixpoints occur, we see that Lemma 1 provides us with the means of constructing an infinite sequence of accumulation points of $F(z)$ from a given accumulation point $\alpha = \lim f_{n_i}(z)$. For we may choose a specific $\beta = f_{-1}(\alpha) = \lim \mathrm{f}_{\mathrm{m}_j}(z)$; then replace $\{f_{n_i}(z)\}$ by its subsequence $\{f_{m_j+1}(z)\}$, so that $\alpha = \lim f_{m_j+1}(z)$; then include $f(\alpha) = \lim f_{m_j+2}(z)$. Repetition of this process yields the sequence of accumulation points $\{f_{\pm n}(\alpha)\} = \{\cdots, f_{-1}(\alpha), \alpha, f(\alpha), \cdots\}$.

LEMMA 2 *If for $z \in S$, α is the limit point of the two subsequences $\{f_{n_i}(z)\}$ and $\{f_{n_i+1}(z)\}$, then $\alpha = \lim f_n(z)$ is the only accumulation point of $F(z)$.*

Proof. Since $\alpha = \lim f_{n_i}(z)$, from Lemma 1 it follows that $f(\alpha) = \lim f_{n_i+1}(z)$, and we conclude that $f(\alpha) = \alpha$. Moreover, we may combine the terms of the two sequences $\{f_{n_i}(z)\}$ and $\{f_{n_i+1}(z)\}$, eliminating duplicates if necessary, into a new subsequence $\{f_{n_i,n_i+1}(z)\}$ which will also have limit point α (any infinite subsequence of $\{f_{n_i,n_i+1}(z)\}$ will necessarily contain an infinite number of terms from either of the constituent subsequences, or both, and so can have only α as a limit point). Now consider $\{f_{n_i+2}(z)\}$. By Lemma 1, this subsequence has limit point $f(\alpha) = \alpha$. Then we may join $\{f_{n_i+2}(z)\}$ to the combined subsequence $\{f_{n_i,n_i+1}(z)\}$ to form $\{f_{n_i,n_i+1,n_i+2}(z)\}$, which as before we conclude has α as its unique limit point. We see that corresponding to every integer $j \geqq 0$, we may construct a subsequence $\{f_{n_i,\cdots,n_i+j}(z)\}$ with limit point α. Every term of $F(z)$ is contained in this subsequence. Hence $\alpha = \lim f_n(z)$.

LEMMA 3. *For $z \in S$ and any integer $r \geqq 1$, $f_r(z)$ satisfies conditions* (1), (2), *and* (3).

Proof. The lemma may be proved by induction as a straightforward consequence of conditions (1)–(3).

From Lemma 3 it follows that we may apply Lemmas 1 and 2 to the sequence $\{f_{nr}(z)\}$.

LEMMA 4. *If for $z \in S$, α is the limit point of the two subsequences $\{f_{n_i}(z)\}$ and $\{f_{n_i+r}(z)\}$ for some least integer $r \geqq 1$, then there exist exactly r distinct accumulation points $f_m(\beta) = \lim f_{nr+m}(z)$, $m = 0, \cdots, r \times 1$, of $F(z)$, where for some particular value $0 \leqq M \leqq r-1$, $f_M(\beta) = \alpha$.*

Proof. Lemmas 1 and 3 enable us to conclude that $f_r(\alpha) = \alpha = \lim f_{n_i+r}(z)$, and the argument of Lemma 2 may be applied to construct a subsequence $\{f_{n_i,\cdots,n_i+jr}(z)\} \subseteq F(z)$ with limit point α. Then the subsequence

$$\{f_{n_1+jr}(z) \mid j = 0, 1, \cdots\} \subseteq \{f_{n_i,\cdots,n_i+jr}(z)\}$$

also has limit point α. This new subsequence consists of every rth term of $F(z)$

beginning with n_1. Repeated use of Lemma 1 leads to the construction of a set of r subsequences as follows:

$$\begin{array}{ll} \{f_{jr+n_1}(z)\} & \text{with limit point } \alpha, \\ \{f_{jr+n_1+1}(z)\} & \text{with limit point } f(\alpha), \\ \vdots & \vdots \\ \{f_{jr+n_1+r-1}(z)\} & \text{with limit point } f_{r-1}(\alpha). \end{array}$$

The union of these r subsequences exhausts $F(z)$ except for n_1 initial terms which may be assigned arbitrarily, in particular so that the subsequences become $\{f_{jr+R_m}(z) \mid R_m = (n_1+m) \bmod r\}$ with respective limit points $f_m(\alpha), m=0, \cdots, r-1$. These subsequences are just a renumbering, but in the same cyclic order, of the subsequences $\{f_{nr+m}(z)\}$, $m=0, \cdots, r-1$, given in the statement of the lemma. There are no accumulation points of $F(z)$ other than the $f_m(\alpha)$, because as in Lemma 2 any infinite subsequence of $F(z)$ must contain an infinity of points drawn from at least one of the r subsequences $\{f_{nr+m}(z)\}$.

It remains to show that for $0 \leqq m' < m \leqq r-1, f_{m'}(\alpha) \neq f_m(\alpha)$. But this follows from the fact that r is the least integer such that $\alpha = \lim f_{n_i+r}(z)$; for otherwise, setting $K = m - m'$, we would necessarily have $\alpha = f_K(\alpha)$. This completes the proof.

A point $f_m(\alpha)$, $m=0, \cdots, r-1$, is called a *fixpoint of order* r of $F(z)$; and the ordered set $\{\alpha, f(\alpha), \cdots, f_{r-1}(\alpha)\}$ a *cycle of order* r of $F(z)$. From the proof of Lemma 4, we see that if a fixpoint of order r exists, then a cycle of order r exists, and $F(z)$ breaks down into exactly r subsequences whose limit points are the r fixpoints. We have then

LEMMA 5. *The sequence $F(z)$ of iterates at a point $z \in S$ gives rise to no fixpoints if and only if $F(z)$ has an infinite number of accumulation points.*

Proof. The result is an immediate consequence of Lemmas 1 and 4.

It is tempting here to try to make a much stronger statement:

CONJECTURE. *The sequence $F(z)$ of iterates at a point $z \in S$ gives rise to no fixpoints if and only if the closure of $F(z)$ is a perfect set.*

For sufficiency is a consequence of Lemma 5; and the proof of necessity leads, by application of Lemmas 1 and 5, to the construction of an infinite sequence of sequences

$$F(z) \to F(\alpha_1) \to F(\alpha_2) \to \cdots,$$

where

(a) each term of each sequence $F(\alpha_i)$ is the limit point of a subsequence of *each* preceding sequence;

(b) each α_i stands for at least one, but perhaps uncountably many, distinct accumulation points.

It appears, however, that proof of this result is stymied by lack of information about the neighborhood of the point z, which could be an isolated point in S. On the other hand, no counterexample is known.

2. Sequence of iterates of a meromorphic function. We recall the definition of normality:

DEFINITION. A family F of functions f meromorphic in a domain[2] D is said to be *normal in D* if every infinite sequence $\{f_n(z)\} \subset F$ contains a subsequence which converges uniformly on every compact subset of D.

[2] A domain is an open arcwise-connected set.

$F(z)$ is *normal at* z if it is normal in a neighborhood of z.

For a sequence of iterates F of a meromorphic function f, we find [5], [7] that we may identify a set N of domains in which F is normal, and a set E of points at which no subsequence of F is normal. In the case of a rational function, $N \cup E = \sigma$; we shall consider this case first, then extend our results to general meromorphic functions. For rational f, then, generalized versions of conditions (1)–(3) specified in the previous section apply separately to both $S = N$ and $S = E$ (for proofs, see [1], [5]):

(1) f is meromorphic in S;

(2) $f(S) \subseteq S$ and $f_{-1}(S) \subseteq S$;

(3) for $z \in S$, every accumulation point of $F(z)$ is a point of S. Hence the results of the previous section apply to every point of both N and E.

Before proceeding, we state without proof two basic results [5], [7] of the theory of iterations which apply to any meromorphic function f:

(a) Denote by α_0 an accumulation point of $F(z)$, where z_0 is a point of a domain $D \subseteq N$. Then for every $z \in D$, hence every $z \in F(D)$, we may choose from $F(z)$ a subsequence which converges uniformly everywhere in $F(D)$ to a meromorphic limit function $g(z)$ such that $g(z_0) = \alpha_0$. For $z \in F(D)$, $g(z) \in F(D)$.

(b) Every point of E is an accumulation point of antecedents of every point of σ, with at most two exceptions (that is, for $z \in E$ and a neighborhood $R(z), F(R)$ includes every point of σ except at most two). This result is an immediate consequence of Montel's theorem [14].

Next we establish for meromorphic functions the conjecture stated in the previous section:

LEMMA 6. *The sequence $F(z)$ of iterates of a rational (meromorphic) function f at a point z has an infinite number of accumulation points if and only if the closure of $F(z)$ is a perfect set.*

Proof. Suppose the closure of $F(z)$ is perfect. Then $F(z)$ must have an infinite number of accumulation points.

Next suppose $F(z)$ has an infinity of accumulation points and let one of them be $\alpha = \lim \zeta_i = \lim f_{n_i}(z)$. Recall from Lemma 1 that therefore every $f_{\pm n}(\alpha)$ is also an accumulation point of $F(z)$. Corresponding to each $\delta_i = \chi(\alpha, \zeta_i)$, define a sequence $\{R_i(\alpha)\}$ of neighborhoods $R_i(\alpha) = R(\alpha, \delta_i)$ with boundary point ζ_i. Since f is meromorphic, hence also every f_{n_i}, we may denote by f_{-n_i} that branch of $\{f_{-n_i}\}$ such that $f_{-n_i}(\zeta_i) = z$, and then define a sequence $\{f_{-n_i}[R_i(\alpha)]\}$ of open regions, each of which has z as a boundary point and $f_{-n_i}(\alpha)$ as an interior point, and such that $\lim_{i\to\infty} \chi[f_{-n_i}(\alpha), z] = 0$. Then z is a limit point of accumulation points of $F(z)$. Hence every neighborhood of z contains terms of $F(z)$, so that z is an accumulation point of $F(z)$. Then by Lemma 1, so is every $f_{\pm n}(z)$ an accumulation point of $F(z)$, and therefore the closure of $F(z)$ is a perfect set.

We may now state and prove

THEOREM 1. *Let $F(z_0)$ be the sequence of iterates of a rational (meromorphic) function f at a point z_0. Then the following three statements are equivalent*:

(A) *$F(z_0)$ gives rise to no fixpoints.*

(B) *$F(z_0)$ has an infinite number of accumulation points.*

(C) *The closure of $F(z_0)$ is a perfect set.*

If any of these statements is true, then z_0 is a point of E, the set of points at which $F(z_0)$ is not normal.

Proof. The equivalence of (A), (B), and (C) has been established by Lemmas 5 and 6. Suppose these statements are true, and assume that z_0 belongs to a domain $D \subseteq N$. Then by basic result (a), for every $z \in D$, there exists a subsequence of $F(z)$

which converges everywhere in D to a meromorphic limit function $g(z)$ such that $g(z_0)=\alpha_0$. There are an uncountable number of distinct limit points α_0, hence an uncountable number of distinct subsequences. Thus at every point $z\in D$, $F(z)$ has an infinite number of accumulation points, and therefore by Lemma 6 the closure of $F(z)$, $\overline{F(z)}$, is a perfect set. The totality of accumulation points of the sequence of iterates in D is thus represented by a union $\bigcup_{z\in D}\overline{F(z)}$ which contains every point in D. But this union is a perfect set, hence closed, and must therefore contain the boundary points of D, which are points of E. These points of E must then be accumulation points of $F(z)$ for some $z\subseteq D\subseteq N$, contradicting condition (3). Hence $z_0\in E$, as required.

Theorem 1 lays the basis for a classification of points into *ordinary* (those whose sequence of iterates given rise to fixpoints) and *special*. We have shown that N consists only of ordinary points; it is not known whether or not E necessarily contains special points, but it may do so: for $f(z)=z^2$, points $e^{i\pi/k}$ are special if the real number $k\neq 0$ is irrational, ordinary otherwise. Using Lemma 4 and Theorem 1, we may now state our criterion of normality:

THEOREM 2. *Let $F(z)$ be the sequence of iterates of a rational (meromorphic) function f at any point z. Then F is normal at a given point z_0 if and only if there exist a finite integer $r>0$ and a neighborhood $R(z_0)$ such that for every $z\in R(z_0)$, $F(z)$ has exactly r distinct accumulation points*

$$f_i(\alpha)=\lim f_{nr+i}(z),\qquad i=0,\cdots,r-1.$$

Proof. Suppose F is normal at z_0. Since N is by definition open, by Theorem 1 there exists a neighborhood $R(z_0)$ containing only ordinary points. Then by Lemma 4, corresponding to every $z\in R(z_0)$ there exists the required r-way breakdown of $F(z)$, and by basic result (a) the same value r and the same breakdown holds uniformly throughout $R(z_0)$. This proves necessity.

Now suppose F is not normal at z_0. Then by basic result (b), for every $R(z_0)$, $F(R)$ includes every point except at most two. Hence $\overline{F(R)}=\sigma$, and there are certainly more than a finite number r of accumulation points of $F(z)$, $z\in R(z_0)$. This proves sufficiency.

We conclude this section by indicating how these results are extended to general meromorphic functions. For this purpose we exhibit in table form the interesting classification due to Rådström [7]. Essentially the classification is based on the set X of points z at which $F(z)$ has only a finite number of terms; there turn out to be only four possibilities, depending on whether X contains 0, 1, or 2 points, or is dense-in-itself. For classes I, II, and III, $E\cup X$ always exists and is perfect; for class IV, $E=\varnothing$. For class I, we have already seen that $X=\varnothing$; the example of Lattès [15],

$$f(z)=(z^2+1)^2/(4z(z^2-1)),$$

shows that it is even possible that $N=\varnothing$.

Observe first that Theorems 1 and 2 hold not only for class I functions, but for class IV as well, since conditions (1)–(3) are satisfied by both N and $E=\varnothing$. For classes II and III, N of course satisfies conditions (1)–(3); E satisfies condition (1) trivially; and Rådström shows that points of X have no consequents or antecedents in E, so that E also satisfies condition (2). For classes II and III, then, we need only examine the applicability of condition (3) to E.

Consider then a subsequence $\{f_{n_i}(z)\}$ with limit point ∞. The subsequence $\{f_{n_i+1}(z)\}$ must also have a limit point β, which we may suppose belongs to E. But since condition (3) is satisfied for β, Lemma 1 applies, and we find ∞ must be an antecedent of β, which is impossible since ∞ is an essential singularity. Hence ∞ is the only

TABLE 1

Rådström's classification of meromorphic functions according to the set X.

Class	Description of $f(z)$	X	E	N	Remarks
I	Rational	$\varnothing$	Perfect		Possibly $N=\varnothing$
II	Nonpolynomial entire	$\{\infty\}$	$E\cup X$ perfect		∞ is an essential singularity
IIIa	$z^{-p}e^{A(z)}$: integer $p>0$, A nonconstant entire	$\{0,\infty\}$	$E\cup X$ perfect	$\sigma-(E\cup X)$	0 is a pole, ∞ an essential singularity, and $f(0)=\infty$
IIIb	$z^{p}e^{A(z)+B(1/z)}$: integer p, A and B nonconstant entire				0 and ∞ are both essential singularities (if f belongs to class IIIa, f_2 belongs to IIIb)
IV	Other	Dense-in-itself	$\varnothing$	interior of $\sigma-X$	X not necessarily closed

accumulation point of $F(z)$. Similarly for class IIIa we find that if $0=\lim f_{n_i}(z)$, then in fact 0 is the only accumulation point of $F(z)$. Hence if for $z\in E$ $F(z)$ has an accumulation point $\alpha\in X$, α is the unique limit point of $F(z)$. Then this case is not relevant either to Theorem 1 or Theorem 2, and may be treated as a single exceptional case in the proof of the Lemmas. Our main results apply therefore to all meromorphic functions.

3. Constant limit function. In practice, attention focuses on the domains in which F is normal, because it is the attractive fixpoints in these domains to which convergence takes place, and which are as a rule solutions of the given computational problem. We have already seen in Theorem 2 that in every domain D of N, every point z gives rise to a finite number r of accumulation points, which are limit points of regular subsequences of $F(z)$, and which are the values at z of r meromorphic limit functions g. We shall now show that these limit functions are, in cases of practical interest, constant. In other words, with trivial exceptions the sequence F converges everywhere in N to fixpoints.

THEOREM 3. *Let $F(z)$ be the sequence of iterates of a meromorphic function f at points $z\in F(D)$, where $F(D)$ is the set of consequents under f of a domain $D\subseteq N$, the set of points at which F is normal. Then corresponding to $F(D)\subseteq N$, there exists a fixed finite integer $r>0$ such that, if $f_r(z)$ is not identically z in $F(D)$, $F(z)$ converges everywhere in $F(D)$ to exactly r distinct accumulation points α_i, $i=0,\cdots,r-1$, where $\alpha_{i+1 (\mathrm{mod}\, r)}=f(\alpha_i)$, $i=0,\cdots,r-1$, and each α_i is a limit point of a subsequence $\{f_{nr+j}(z)\}\subseteq F(z)$, $j=0,\cdots,r-1$.*

Proof. Lemma 4 and Theorem 2, together with basic result (a), have already made clear that at each point $z\in F(D)$, convergence to exactly r meromorphic limit functions $g_{(i)}(z)$, $i=0,\cdots,r-1$, takes place, according to a breakdown of $F(z)$ into subsequences $\{f_{nr+i}(z)\}$, and that at some particular point z_0, $g_{(i)}(z_0)=\alpha_i$, $i=0,\cdots,r-1$. We need to show that unless $f_r(z)\equiv z$ for every $z\in F(D)$, then $g_{(i)}(z)$ can take only constant values drawn from the set $\{\alpha_0,\cdots,\alpha_{r-1}\}$.

Consider any domain $D^*\subseteq F(D)$ such that $g_{(0)}(z)=\lim f_{nr}(z)$, $z\in D^*$. Note that by the uniform convergence of $f_{nr}(z)$, we may ensure that $\chi[g_{(0)}(z), f_{nr}(z)]<\delta$ simply by choosing $n>N(\delta)$, uniformly for every $z\in D^*$. Then given $\varepsilon>0$, by the continuity of f in domains of $F(D)$, we may make

$$\chi[f_i\cdot g_{(0)}(z), f_i\cdot f_{nr}(z)]<\varepsilon$$

by choosing $n > N[\delta(\varepsilon)]$ uniformly for all $z \in D^*$, so that $\chi[g_{(0)}(z), f_{nr}(z)] < \delta(\varepsilon)$. Then the subsequence $\{f_{nr+i}(z)\}$ converges uniformly everywhere in D^* to $f_i \cdot g_{(0)}(z) = g_{(i)}(z)$, $i = 0, \cdots, r-1$. For $i = r$, $\{f_r \cdot f_{nr}(z)\}$ has the same limit function as $\{f_{nr}(z)\}$, so that

$$g_{(o)}(z) = f_r \cdot g_{(0)}(z),$$

for every $z \in D^*$. This means that every point $\zeta = g_{(0)}(z)$ in $g_{(0)}(D^*) \subseteq F(D) \subseteq N$ is a fixpoint of f_r. Since f_r is meromorphic, this can only be true if for every $z \in D^*$

(a) $g_{(0)}(z)$ is constant, or

(b) $f_r(z) \equiv z$.

From possibility (a) we conclude that $g_{(i)}(z)$, $i = 0, \cdots, r-1$, is constant for every $z \in F(D)$; and that $g_{(0)}(z) = \alpha_j$ for some particular value $0 \leqq j \leqq r-1$, $g_{(i)}(z) = f_i(\alpha_j)$ for $i = 1, \cdots, r-1$. From possibility (b) we conclude that $f_r(z) \equiv z$ everywhere in $F(D)$.

4. Counterexamples and exceptions. In a well-known paper [11], Schröder presented an apparent counterexample to Theorem 3. He studied the bilinear form

$$f(z) = (az+b)/(cz+d),$$

and showed that if ζ_1, ζ_2 are the roots of the quadratic equation $f(z) = z$, then

$$g(z) = \lim f_n(z) = \begin{cases} \dfrac{(\zeta_1 z + b/c)}{(z-\zeta_2)} & \text{if } |\zeta_1 + d/c| > |\zeta_2 + d/c|, \\ \dfrac{(\zeta_2 z + b/c)}{(z-\zeta_1)} & \text{if } |\zeta_2 + d/c| > |\zeta_1 + d/c|. \end{cases}$$

He claimed that therefore $g(z)$ was nonconstant. However, substituting $\zeta_1\zeta_2 = -b/c$ in the expressions for $g(z)$ yields

$$g(z) = \zeta_1 \text{ or } \zeta_2,$$

and we see that the limit function is actually constant.

We may nevertheless make use of special cases of the bilinear form to generate some true exceptions to Theorem 3, and to demonstrate that the class of functions f excluded by the condition,

$$f_r(z) \neq z \text{ identically},$$

is not trivial. We begin with the examples

$$f(z) = \omega_p z^{\omega_q}, \qquad (a + \omega_r z)/(1 + d\omega_r z),$$

where a, d are arbitrary constants, $p, q \geqq 1$ are integers such that $pq = r$, and ω_i stands for an ith root of unity. For each of these examples $f_r(z) = z$ everywhere on σ.

In 1815 Babbage [12] was led by certain construction problems in geometry to consider the equation $f_r(z) = z$ (sometimes called "Babbage's equation"). He showed that given a particular solution $f(z)$ such as either of the two examples above, then

$$f^*(z) = h_{-1} \cdot f \cdot h(z)$$

is also a solution for any homeomorphism h. Thus, for example,

$$f^*(z) = h_{-1}[(a + \omega_r h(z))/(1 + d\omega_r h(z))]$$

is always a solution, for every a, d and homeomorphism h.

Seventy years later, Serret, attempting to classify the polynomial equations which are solvable by radicals, showed [13] that if f was a rational function, then solutions of

$f_r(z) = z$ could be represented by

$$f(z) = (1/r) \sum_{l \leqq i \leqq r} \omega_r^i [A_i(z)]^{1/r},$$

where as above ω_r is an rth root of unity, and the $A_i(z)$, $i = 1, \cdots, r$, are certain rational functions whose coefficients are expressible in terms of the coefficients of an arbitrary rth order polynomial. We see then that in general solutions f of $f_r(z) = z$ may even be algebraic.

Acknowledgment. I should like to thank James L. Howland, University of Ottawa, for his assistance and support in the early stages of this research, and for introducing me to the constant limit problem.

REFERENCES

[1] GASTON JULIA, *Mémoire sur l'itération des fonctions rationnelles*, J. Math. Pures Appl., 8-1 (1918), pp. 47–245.
[2] P. FATOU, *Sur les équations fonctionnelles I*, Bull. Soc. Math. France, 47 (1919), pp. 161–271.
[3] ———, *Sur les équations fonctionnelles II*, Ibid., 48 (1920), pp. 33–94.
[4] ———, *Sur les équations fonctionnelles III*, Ibid., 48 (1920), pp. 208–314.
[5] HUBERT CREMER, *Über die Iteration rationaler Funktionen*, Jahresbericht Deutschen Math.-Vereinigung, 33-1 Abt. Heft 9/12 (1925), pp. 185–210.
[6] PAUL MONTEL, *L'itération*, Univ. Nac. La Plata Publ. Fac. Ci. Fisicomat. Rev., 3 (1940), pp. 201–211.
[7] H. RÅDSTRÖM, *On the iteration of analytic functions*, Math. Scand., 1 (1953), pp. 85–92.
[8] D. I. TALANOV, *On certain problems of the theory of iteration of rational functions*, Proc. Acad. Sci. U.S.S.R., 93 (1953), pp. 413–416.
[9] R. A. BROOKER, *The solution of algebraic equations on the EDSAC*, Proc. Cambridge Philos. Soc., 48 (1952), pp. 255–270.
[10] W. F. SMYTH, *The construction of rational iterating functions*, Math. Comput., (July 1978).
[11] ERNST SCHRÖDER, *Über iterirte Funktionen*, Math. Ann., 3 (1871), pp. 296–322.
[12] CHARLES BABBAGE, *An essay towards the calculus of functions*, Trans. Roy. Soc. London (1815), pp. 389–423.
[13] J. A. SERRET, *Cours d'algèbre superiéure II*, Gauthier-Villars, Paris, 1885.
[14] EINAR HILLE, *Analytic function theory II*, Gin, Boston, 1962.
[15] LATTÈS, C. R. Acad. Sci. Française, 166 (1918), p. 28.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0004 $01.00/0

AN ASYMPTOTIC FORMULA FOR THE DERIVATIVES OF ORTHOGONAL POLYNOMIALS*

PAUL G. NEVAI†

Abstract. An asymptotic formula is found for the derivatives of orthogonal polynomials on the unit circle. The condition on the distribution function is essentially local and the result is stronger than those known before.

Let σ be a bounded nondecreasing function on $[0, 2\pi]$ taking infinitely many values. Then there exists a unique sequence of polynomials $\{\varphi_n(d\sigma)\}_{n=0}^{\infty}$ such that $\varphi_n(d\sigma, z)=\alpha_n(d\sigma)z^n+\cdots$, $\alpha_n(d\sigma)>0$ and

$$\frac{1}{2\pi}\int_0^{2\pi} \varphi_n(d\sigma, z)\,\overline{\varphi_m(d\sigma, z)}\,d\sigma(\theta)=\delta_{nm} \qquad (z=e^{i\theta}).$$

One of the basic problems in the theory of orthogonal polynomials on the unit circle is to find asymptotic expressions for $\varphi_n(d\sigma, z)$ as $n\to\infty$. There is an extensive literature dealing with this question. (See e.g. [2], [3] and [7].) In order to obtain asymptotics one has to assume that σ behaves nicely in a certain sense. Usually there are two kinds of assumptions: globally σ must satisfy a growth condition and locally (near θ, $z=e^{i\theta}$) σ has to be smooth. The weakest condition under which one can prove asymptotics for $\varphi_n(d\sigma, z)$ belongs to G. Freud [2].

Let the Szegö function $D(d\sigma, z)$ corresponding to σ be defined by

$$D(d\sigma, z)=\exp\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\log\sigma'(t)\frac{1+z\,e^{-it}}{1-z\,e^{-it}}\,dt\right\} \qquad (|z|<1).$$

If

$$\int_{-\pi}^{\pi}\log\sigma'(t)\,dt>-\infty \tag{1}$$

then $D(d\sigma)\in H_2(|z|<1)$, $D(d\sigma, z)\neq 0$ for $|z|<1$, $D(d\sigma, 0)>0$ and

$$\lim_{r\to 1^-} D(d\sigma, r\,e^{it})=D(d\sigma, e^{it})$$

exists and $|D(d\sigma, e^{it})|^2=\sigma'(t)$ for almost every $t\in[-\pi, \pi]$.

Using the notion of Szegö's function we can formulate Freud's result. Assume that (1) is satisfied and in a neighborhood of θ ($z=e^{i\theta}$) σ is absolutely continuous with $0<m\leqq\sigma'(t)\leqq M<\infty$ for $|\theta-t|$ small and

$$\int_{|\theta-t|\,\text{small}}\left(\frac{\sigma'(\theta)-\sigma'(t)}{\theta-t}\right)^2 dt<\infty. \tag{2}$$

Then

$$\lim_{n\to\infty}\left[\varphi_n(d\sigma, e^{i\theta})-e^{in\theta}\,\overline{D(d\sigma, e^{i\theta})^{-1}}\right]=0. \tag{3}$$

In the end of L. Geronimus' book [3] fourteen other conditions are given all of which imply the asymptotic relation (3). The problem of finding asymptotics for

* Received by the editors August 15, 1977, and in final revised form December 8, 1977. This work was supported by the United States Army under Contract DAAG29-75-C-0024 and by the National Science Foundation under Grant MPS75-06687 #3.

† Department of Mathematics, Ohio State University, Columbus, Ohio 43210.

$\varphi_n^{(k)}(d\sigma, z)(k=1,2,\cdots)$ seems to be more difficult. There are only a very few papers investigating the relationship

$$\lim_{n\to\infty} [n^{-k}\varphi_n^{(k)}(d\sigma, z) - z^{n-k}\overline{D(d\sigma, z)^{-1}}] = 0 \qquad (z = e^{i\theta}). \tag{4}$$

(See [4], [5] and [6].) In all these papers it is assumed that σ satisfies some very restrictive conditions. In particular, σ has to be absolutely continuous and $(\sigma')^{-1} \in L^2$. In [4] and [5] the authors apply strong methods of approximation theory. The purpose of this paper is to show that (4) can be proved under Freud's conditions. Instead of approximation theory our approach is based on the following two observations. First, it is easy to prove (4) provided that σ is very nice locally. Second, the weak asymptotics

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} |\varphi_n(d\sigma, e^{i\theta}) - e^{in\theta}\overline{D(d\sigma, e^{i\theta})^{-1}}|^2\, d\sigma(\theta) = 0 \tag{5}$$

always holds whenever (1) is satisfied. (See [2, § V.4].)

In the following Δ will denote a closed interval in $[-2\pi, 2\pi]$, Δ^0 is its interior and $\widehat{\Delta}$ is the corresponding arc on the unit circle. If P is a polynomial then $\bar{P}$ denotes the polynomial whose coefficients are the complex conjugates of the corresponding coefficients of P. The polynomial $\varphi_n^*(d\sigma)$ is defined by

$$\varphi_n^*(d\sigma, z) = z^n\bar{\varphi}_n(d\sigma, z^{-1}).$$

We have therefore by (5)

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} |\varphi_n^*(d\sigma, e^{i\theta}) - D(d\sigma, e^{i\theta})^{-1}|^2\, d\sigma(\theta) = 0 \tag{6}$$

provided that (1) holds.

LEMMA. *Let σ be such that* (1) *is satisfied. Let Δ and $\Delta_1 \subset \Delta^0$ be given. Suppose that σ is absolutely continuous on Δ and $\sigma'(t) = 1$ for $t \in \Delta$. Then for every fixed $k = 0, 1, \cdots$, $n^{-k}|\varphi_n^{(k)}(d\sigma, z)|$ and $n^{-k}|\varphi_n^{*(k)}(d\sigma, z)|$ are uniformly bounded for $z \in \widehat{\Delta_1}$. Further* (4) *holds uniformly for $z \in \widehat{\Delta_1}$.*

Proof. Fix Δ_2 so that $\Delta_1 \subset \Delta_2^0 \subset \Delta_2 \subset \Delta^0$. By a result of L. Geronimus [3]

$$\varphi_n^*(d\sigma, e^{i\theta}) = D(d\sigma, e^{i\theta})^{-1} + o(1) \tag{7}$$

uniformly for $\theta \in \Delta_2$. Because of the assumptions $D(d\sigma, e^{i\theta})^{-1}$ is continuous on Δ_2. We have

$$\left|\frac{d^k}{d\theta^k}\varphi_n^*(d\sigma, e^{i\theta})\right| \leqq \left|\frac{d^k}{d\theta^k}[\varphi_n^*(d\sigma, e^{i\theta}) - \varphi_{[\sqrt{n}]}^*(d\sigma, e^{i\theta})]\right| + \left|\frac{d^k}{d\theta^k}\varphi_{[\sqrt{n}]}^*(d\sigma, e^{i\theta})\right|.$$

Therefore by the local version of Bernstein's inequality (see [1, p. 896])

$$\max_{\theta\in\Delta_1}\left|\frac{d^k}{d\theta^k}\varphi_n^*(d\sigma, e^{i\theta})\right|$$

$$\leqq \text{const}\,[n^k \cdot \max_{\theta\in\Delta_2}|\varphi_n^*(d\sigma, e^{i\theta}) - \varphi_{[\sqrt{n}]}^*(d\sigma, e^{i\theta})| + n^{k/2}\max_{\theta\in\Delta_2}|\varphi_{[\sqrt{n}]}^*(d\sigma, e^{i\theta})|].$$

Consequently for $k = 1, 2, \cdots$

$$\left|\frac{d^k}{d\theta^k}\varphi_n^*(d\sigma, e^{i\theta})\right| = o(n^k) \tag{8}$$

uniformly for $\theta \in \Delta_1$. Hence $|\varphi_n^{*(k)}(d\sigma, z)| = o(n^k)$ uniformly for $z \in \widehat{\Delta}_1$ if $k \geqq 1$ is fixed. Now we have

$$\varphi_n(d\sigma, e^{i\theta}) = e^{in\theta}\, \overline{\varphi_n^*(d\sigma, e^{i\theta})}.$$

Differentiating this identity and using (8) we obtain for $k = 1, 2, \cdots$,

$$\frac{d^k}{d\theta^k} \varphi_n(d\sigma, e^{i\theta}) = (in)^k e^{in\theta}\, \overline{\varphi_n^*(d\sigma, e^{i\theta})} + o(n^k) \tag{9}$$

uniformly for $\theta \in \Delta_1$. This is also valid when $k = 0$. Replacing differentiation in θ by differentiation in z and using the fact that (9) is valid for every k we obtain (4).

THEOREM. *Let σ satisfy* (1) *and let $z = e^{i\theta}$ be fixed. Assume that σ is absolutely continuous near θ, $0 < m \leqq \sigma'(t) \leqq M < \infty$ for $|\theta - t|$ small and* (2) *holds. Then for every fixed $k = 1, 2, \cdots$ the asymptotic relation* (4) *holds true.*

Proof. Pick up a sufficiently small neighborhood Δ of θ and define σ_1 by

$$d\sigma_1(t) = \begin{cases} d\sigma(t) & \text{for } t \notin \Delta, \\ dt & \text{for } t \in \Delta. \end{cases}$$

Let the function g be defined by

$$g(t) = \begin{cases} 1 & \text{for } t \notin \Delta, \\ \sigma'(t) & \text{for } t \in \Delta. \end{cases}$$

If Δ is small enough then $d\sigma = g\, d\sigma_1$, $0 < m_1 \leqq g(t) < M_1 < \infty$ and

$$\int_{-\pi}^{\pi} \left(\frac{g(\theta) - g(t)}{\theta - t} \right)^2 dt < \infty.$$

Note that σ_1 satisfies the conditions of the lemma. Let us expand $\varphi_n(d\sigma)$ into Fourier series in $\varphi_l(d\sigma_1)$. We have

$$\varphi_n(d\sigma, z) = \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(d\sigma, e^{it}) K_n(d\sigma_1, z, e^{it})\, d\sigma_1(t) \tag{10}$$

where

$$K_n(d\sigma_1, z, y) = \sum_{l=0}^{n} \varphi_l(d\sigma_1, z)\, \overline{\varphi_l(d\sigma_1, y)}.$$

Differentiating (10) by z and using the fact that $\varphi_n(d\sigma)$ is orthogonal with respect to $g\, d\sigma_1$ we obtain

$$\begin{aligned} \varphi_n^{(k)}(d\sigma, z) = {} & \frac{1}{2\pi} \int_0^{\pi} \varphi_n(d\sigma, e^{it})\, \overline{\varphi_n(d\sigma_1, e^{it})}\, d\sigma_1(t) \varphi_n^{(k)}(d\sigma_1, z) \\ & + \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(d\sigma, e^{it}) \frac{d^k}{dz^k} K_{n-1}(d\sigma_1, z, e^{it}) \left[1 - \frac{g(t)}{g(\theta)} \right] d\sigma_1(t). \end{aligned}$$

Using the Christoffel–Darboux formula we obtain

$$(1 - z\bar{y}) K_{n-1}(d\sigma_1, z, y) = \varphi_n^*(d\sigma_1, z)\, \overline{\varphi_n^*(d\sigma_1, y)} - \varphi_n(d\sigma_1, z)\, \overline{\varphi_n(d\sigma_1, y)}.$$

(See [2].) We get

$$\begin{aligned} (1 - z\bar{y}) & \frac{d^k}{dz^k} K_{n-1}(d\sigma_1, z, y) - k\bar{y} \frac{d^{k-1}}{dz^{k-1}} K_{n-1}(d\sigma_1, z, y) \\ & = \varphi_n^{*(k)}(d\sigma_1, z)\, \overline{\varphi_n^*(d\sigma_1, y)} - \varphi_n^{(k)}(d\sigma_1, z)\, \overline{\varphi_n(d\sigma_1, y)}. \end{aligned}$$

Therefore $\varphi_n^{(k)}(d\sigma, z)$ $(z = e^{i\theta})$ can be written as

$$\varphi_n^{(k)}(d\sigma, z) = A + B + C$$

where

$$A = \varphi_n^{(k)}(d\sigma_1, z) \cdot \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(d\sigma, e^{it}) \overline{\varphi_n(d\sigma_1, e^{it})} \cdot \left[1 - \frac{1 - g(t)/g(\theta)}{1 - e^{i(\theta - t)}}\right] d\sigma_1(t),$$

$$B = \varphi_n^{*(k)}(d\sigma_1, z) \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(d\sigma, e^{it}) \overline{\varphi_n^*(d\sigma_1, e^{it})} \cdot \frac{1 - g(t)/g(\theta)}{1 - e^{i(\theta - t)}} d\sigma_1(t)$$

and

$$C = k \cdot \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(d\sigma, e^{it}) \frac{d^{k-1}}{dz^{k-1}} K_{n-1}(d\sigma_1, z, e^{it}) \cdot e^{-it} \cdot \frac{1 - g(t)/g(\theta)}{1 - e^{i(\theta - t)}} d\sigma_1(t).$$

We will estimate A, B and C separately. First we consider A. We will show that the integral in A converges as $n \to \infty$. Write σ and σ_1 as

$$\sigma = \sigma^a + \sigma^s + \sigma^j, \qquad \sigma_1 = \sigma_1^a + \sigma_1^s + \sigma_1^j$$

where a, s and j refer to the absolutely continuous, singular and jump components respectively. It is clear from the construction that $\sigma^s = \sigma_1^s$, $\sigma^j = \sigma_1^j$ and $d\sigma^a = g d\sigma_1^a$. It follows from (5) that

$$\lim_{n \to \infty} \frac{1}{2\pi} \int_0^{2\pi} |\varphi_n(d\sigma_*, e^{it})|^2 \sigma_*'(t)\, dt = 1$$

(σ_* is either σ or σ_1). Therefore

$$\lim_{n \to \infty} \frac{1}{2\pi} \int_0^{2\pi} |\varphi_n(d\sigma_*, e^{it})|^2 d[\sigma_*^s(t) + \sigma_*^j(t)] = 0$$

must hold. Since the function

$$f(t) = 1 - \frac{1 - g(t)/g(\theta)}{1 - e^{i(\theta - t)}}$$

is uniformly bounded on the support of $d(\sigma_1^s + \sigma_1^j)$ we obtain

$$\lim_{n \to \infty} \frac{1}{2\pi} \int_0^{2\pi} \varphi_n(d\sigma, e^{it}) \overline{\varphi_n(d\sigma_1, e^{it})} f(t)\, d[\sigma_1^s(t) + \sigma_1^j(t)] = 0.$$

Now fix $\varepsilon > 0$ and choose $\delta > 0$ so that $\theta \pm \delta \in \Delta_1 \subset \Delta^0$ and

$$\frac{1}{2\pi} \int_{\theta - \delta}^{\theta + \delta} |f(t)|^2 g(t)^{-1}\, dt < \varepsilon.$$

We have

$$\frac{1}{2\pi} \int_{\theta - \delta}^{\theta + \delta} |\varphi_n(d\sigma, e^{it}) \varphi_n(d\sigma_1, e^{it}) f(t)|\, d\sigma_1^a(t)$$

$$\leqq \left[\frac{1}{2\pi} \int_{\theta - \delta}^{\theta + \delta} |\varphi_n(d\sigma, e^{it})|^2 g(t)\, d\sigma_1^a(t)\right]^{1/2} \cdot \left[\frac{1}{2\pi} \int_{\theta - \delta}^{\theta + \delta} |f(t)|^2 g(t)^{-1}\, d\sigma_1^a(t)\right]^{1/2}$$

$$\cdot \max_{t \in \Delta_1} |\varphi_n(d\sigma_1, e^{it})|$$

which by the lemma is $O(\sqrt{\varepsilon})$. Using (5) we obtain that

$$\lim_{n\to\infty}\frac{1}{2\pi}\int_{|\theta-t|>\delta}\varphi_n(d\sigma,e^{it})\overline{\varphi_n(d\sigma_1,e^{it})}f(t)\,d\sigma_1^a(t)$$
$$=\frac{1}{2\pi}\int_{|\theta-t|>\delta}\overline{D(d\sigma,e^{it})^{-1}}\,D(d\sigma_1,e^{it})^{-1}f(t)\,d\sigma_1^a(t).$$

Letting $\varepsilon\to 0$ we get that

$$\lim_{n\to\infty}\frac{1}{2\pi}\int_0^{2\pi}\varphi_n(d\sigma,e^{it})\overline{\varphi_n(d\sigma_1,e^{it})}f(t)\,d\sigma_1(t)=\rho$$

exists and equals

$$\frac{1}{2\pi}\int_0^{2\pi}\overline{D(d\sigma,e^{it})^{-1}}\,D(d\sigma_1,e^{it})^{-1}f(t)\,d\sigma_1^a(t).$$

Because $D(d\sigma_1)^{-1}$ is bounded and $D(d\sigma)^{-1}\in L^2(d\sigma_1^a)$, and $f\in L^2(d\sigma_1^a)$, ρ is finite and independent of k. Applying the lemma we finally obtain

$$A=\rho\varphi_n^{(k)}(d\sigma_1,z)+o(n^k)\tag{11}$$

as $n\to\infty$ where the number ρ does not depend on k. The expression B can be estimated in a similar way. The only difference is that this time we have to apply both (5) and (6). We write

$$\frac{1}{2\pi}\int_0^{2\pi}\varphi_n(d\sigma,e^{it})\overline{\varphi_n^*(d\sigma_1,e^{it})}f(t)\,d\sigma_1(t)$$
$$=\frac{1}{2\pi}\int_0^{2\pi}\varphi_n(d\sigma,e^{it})[\overline{\varphi_n^*(d\sigma_1,e^{it})}-\overline{D(d\sigma_1,e^{it})^{-1}}]f(t)\,d\sigma_1(t)$$
$$+\frac{1}{2\pi}\int_0^{2\pi}\varphi_n(d\sigma,e^{it})\overline{D(d\sigma_1,e^{it})^{-1}}f(t)\,d\sigma_1(t).$$

Repeating the above argument and using (6) we see that the first integral on the right side converges to 0 as $n\to\infty$. The second integral is a Fourier coefficient of a function belonging to $L^2(d\sigma)$. Thus the second integral also converges to 0 as $n\to\infty$. Using again the lemma we obtain

$$B=o(n^k)\tag{12}$$

as $n\to\infty$. In order to show that

$$C=o(n^k)\tag{13}$$

as $n\to\infty$ we use Cauchy's inequality. We get

$$|C|^2\leqq k^2\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{d^{k-1}}{dz^{k-1}}K_{n-1}(d\sigma_1,z,e^{it})\right|^2\left|\frac{1-g(t)/g(\theta)}{1-e^{i(\theta-t)}}\right|^2g(t)^{-1}\,d\sigma_1(t).$$

By the lemma,

$$\left|\frac{d^{k-1}}{dz^{k-1}}K_{n-1}(d\sigma_1,z,e^{it})\right|^2\leqq \text{const}\cdot n^{2k}$$

for $|\theta - t|$ small and by the conditions

$$\lim_{\delta\to 0} \frac{1}{2\pi} \int_{|t-\theta|<\delta} \left| \frac{1-g(t)/g(\theta)}{1-e^{i(\theta-t)}} \right|^2 g(t)^{-1}\, d\sigma_1(t) = 0.$$

Therefore we have to estimate

$$\frac{1}{2\pi} \int_{|t-\theta|\geqq\delta} \left| \frac{d^{k-1}}{dz^{k-1}} K_{n-1}(d\sigma_1, z, e^{it}) \right|^2 \left| \frac{1-g(t)/g(\theta)}{1-e^{i(\theta-t)}} \right|^2 g(t)^{-1}\, d\sigma_1(t)$$

for fixed $\delta > 0$. But this is less than

$$\text{const} \cdot \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{d^{k-1}}{dz^{k-1}} K_{n-1}(d\sigma_1, z, e^{it}) \right|^2 d\sigma_1(t) = \text{const} \cdot \sum_{j=0}^{n-1} |\varphi_j^{(k-1)}(d\sigma_1, z)|^2$$

which is $O(n^{2k-1})$ by the lemma. Hence we have proved (13). From (11)–(13) we obtain

$$n^{-k}\varphi_n^{(k)}(d\sigma, z) = \rho n^{-k}\varphi_n^{(k)}(d\sigma_1, z) + o(1) \qquad (n\to\infty)$$

for $k = 0, 1, \cdots$ fixed where ρ is independent of k. By the lemma

$$n^{-k}\varphi_n^{(k)}(d\sigma_1, z) = z^{-k}\varphi_n(d\sigma_1, z) + o(1)$$

as $n \to \infty$. Therefore

$$n^{-k}\varphi_n^{(k)}(d\sigma, z) = z^{-k}\varphi_n(d\sigma, z) + o(1)$$

and the theorem follows from Freud's result which was formulated in the beginning.

REFERENCES

[1] N. K. BARI, *Trigonometric Series*, Moscow, 1961.
[2] G. FREUD, *Orthogonal Polynomials*, Pergamon Press, New York, 1971.
[3] L. GERONIMUS, *Orthogonal Polynomials*, Consultants Bureau, New York, 1961.
[4] B. L. GOLINSKII, *The asymptotic representation at a point of the derivatives of orthonormal polynomials*, Math. Notes, 19 (5/6) (1976), pp. 397–404.
[5] C. HÖRUP, *An asymptotic formula for the derivatives of orthogonal polynomials on the unit circle*, Math. Scand., 20 (1967), pp. 32–40.
[6] S. Z. RAFAL'SON, *On an asymptotic formula for orthogonal polynomials*, Soviet Math. Dokl., 7 (1966), No. 6, pp. 1561–1564.
[7] G. SZEGÖ, *Orthogonal Polynomials*, American Mathematical Society, New York, 1967.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0005 $01.00/0

A COMBINATORIAL APPROACH TO SOME POSITIVITY PROBLEMS*

MOURAD E. H. ISMAIL† AND M. V. TAMHANKAR‡

Abstract. We give a purely combinatorial proof of the nonnegativity of the integrals

$$\int_0^\infty e^{-x}x^\alpha L_n^\alpha(\lambda x)L_m^\alpha((1-\lambda)x)L_k^\alpha(x)\,dx \quad \text{and} \quad \int_0^\infty e^{-2x}x^\alpha L_n^\alpha(x)L_m^\alpha(x)L_k^\alpha(x)\,dx,$$

where α, k, m, n are nonnegative integers and $\lambda \in [0, 1]$. We also state the combinatorial equivalent of a conjecture of H. Lewy. This is done by applying the Master theorem of MacMahon.

1. Introduction. In their study of discretization of the time dependent wave equation in two dimensions, K. O. Friedrichs and H. Lewy needed the positivity of the coefficients $A(k, m, n)$ in

$$\{(1-r)(1-s)+(1-r)(1-t)+(1-s)(1-t)\}^{-1} = \sum_{k,m,n=0}^{\infty} A(k, m, n)r^k s^m t^n \tag{1.1}$$

in order to show that the finite difference approximations to the solution do converge to a solution of the wave equation. The question of the positivity of A's is very simple to state and surprisingly turned out to be difficult to prove. G. Szegö [13] solved this problem using Sonine type integrals for Bessel functions. Szegö was also able to generalize this problem in several directions. He also observed that, Askey and Gasper [2], the coefficients $A(k, m, n)$ can be expressed in terms of the simple Laguerre polynomials $\{L_n(x)\}_{n=0}^\infty$ as

$$A(k, m, n) = \int_0^\infty e^{-3x}L_k(x)L_m(x)L_n(x)\,dx. \tag{1.2}$$

The generalized Laguerre polynomials $\{L_n^\alpha(x)\}_{n=0}^\infty$ have the generating function, Szegö [14]

$$\sum_{n=0}^\infty L_n^\alpha(x)t^n = (1-t)^{-\alpha-1}\exp\left(\frac{-xt}{1-t}\right), \tag{1.3}$$

and satisfy the orthogonality relation, [11],

$$\int_0^\infty e^{-x}x^\alpha L_m^\alpha(x)L_n^\alpha(x)\,dx = \frac{\Gamma(\alpha+1+n)}{n!}\delta_{mn}, \qquad \alpha > -1. \tag{1.4}$$

The simple Laguerre polynomial $L_n(x)$ is $L_n^0(x)$. The Laguerre polynomials are often called Rook polynomials; see Riordan [12]. Askey and Gasper [2] observed that, due to the orthogonality relation (1.4), the coefficients $A(k, m, n)$ are linearization coefficients in

$$e^{-2x}L_k(x)\,e^{-2x}L_m(x) = \sum_{n=0}^\infty A(k, m, n)\,e^{-2x}L_n(x),$$

and proceeded to study the numbers

$$A^\alpha(\mu; k, m, n) = \int_0^\infty e^{-\mu x}x^\alpha L_k^\alpha(x)L_m^\alpha(x)L_n^\alpha(x)\,dx, \qquad \alpha > -1, \quad \mu > 0. \tag{1.5}$$

---

* Received by the editors July 27, 1977, and in final revised form December 7, 1977. This work was supported by grants from the National Research Council of Canada and from McMaster University Research Board.

† Department of Applied Mathematics, MacMaster University, Hamilton, Ontario, Canada L8S 4K1.

‡ Department of Statistics, University of California at Riverside, Riverside, California 92502.

Because of the well-known formula, Rainville [11, p. 209],

$$L_n^\alpha(xy)=\sum_{k=0}^{n}\frac{(1+\alpha)_n(1-y)^{n-k}y^k}{(n-k)!(1+\alpha)_k}L_k^\alpha(x) \tag{1.6}$$

the nonnegativity of $A^\alpha(\mu; k, m, n)$ for some μ implies the nonnegativity of $A^\alpha(\nu; k, m, n)$ for all $\nu \geqq \mu > 0$. The following are three results on the sign behavior of these coefficients.

THEOREM 1.1 (Askey and Gasper [3]). *The numbers* $A^\alpha(2, k, m, n)$ *are nonnegative for all* $k, m, n=0, 1, \cdots$, *if and only if* $\alpha>(-5+\sqrt{17})/2$.

THEOREM 1.2 (Debbi and Gillis [7]). *The numbers* $(-1)^kA(\frac{3}{2}, k, n, n)$ *are nonnegative for* $n=0, 1, \cdots$ *and* $k=0, 1, \cdots, n$.

Quite recently T. Koornwinder [9] proved

THEOREM 1.3. *Let*

$$B^\alpha(n, m, k)=\int_0^\infty e^{-x}x^\alpha L_n^\alpha(\lambda x)L_m^\alpha((1-\lambda)x)L_k^\alpha(x)\,dx.$$

Then $B^\alpha(n, m, k)\geqq 0$ *for* $\alpha \geqq 0$ *and* $0\leqq \lambda \leqq 1$. *Furthermore if* $0<\lambda<1$ *then* $B^\alpha(0, m, n)>0$.

Theorem 1.3 was iterated by Askey, Ismail and Koornwinder [5, Thm. 1]. Using the orthogonality relation (1.4) one can easily see that

$$L_n^\alpha(\lambda x)L_m^\alpha((1-\lambda)x)=\sum_{k=0}^{m+n}\frac{k!B^\alpha(n, m, k)}{\Gamma(\alpha+1+k)}L_k^\alpha(\alpha).$$

Going back to the Friedricks and Lewy conjecture, immediately after Szegö [13] settled it, Kaluza [8] published an "elementary" proof of that conjecture and obtained some monotonicity properties for the coefficients $A(k, m, n)$. Although Kaluza's proof is elementary, in the sense that it uses only elementary series manipulations, it depends on a very tricky way of combining various powers of r, s and t. It is probably fair to say that the impression one forms by reading Kaluza's paper is that Kaluza's method is too tricky to work in any other problem. One purpose of the present work is to give Kaluza's paper another chance by systematizing his method. While doing so we shall give combinatorial proofs of Theorems 1.1 and 1.3 when α is a nonnegative integer. Our proofs use a powerful tool given by MacMahon, which he calls the Master theorem; see MacMahon [10, pp. 93–98].

THEOREM 1.4 (the Master theorem). *Set*

$$V_n=\det(I_n-A_nX_n),$$

where I_n *is the* $n\times n$ *identity matrix,* $A_n=(a_{ij})_{i,j=1}^n$ *and* $X_n=(x_{ij})_{i,j=1}^n$, $x_{ij}=x_i\delta_{ij}$. *Then the coefficient of* $x_1^{k_1}\cdots x_n^{k_n}$ *in the power series* $1/V_n$ *is the same as the coefficient of the same term in the expansion of*

$$(a_{11}x_1+\cdots+a_{1n}x_n)^{k_1}\cdots(a_{n1}x_1+\cdots+a_{nn}x_n)^{k_n}.$$

In § 2 we start by proving Koornwinder's theorem, Theorem 1.3 for $\alpha=0, 1, 2, \cdots$. The proof we give is purely combinatorial and uses only the Master theorem and the binomial theorem. This is followed by a similar proof of Theorem 1.1 for $\alpha=0, 1, 2, \cdots$. In § 3 we mention an interpretation of the numbers $A(2; k, m, n)$ and $B^0(k, m, n)$. In § 4 we discuss an open problem of Lewy and point out its combinatorial equivalent. Although we do not offer a solution to this open problem, we hope to eventually get a solution by this approach. The present work is a continuation of the work started in Askey and Ismail [4], and in Askey, Ismail and Koornwinder [5].

2. Combinatorial results and proofs. We first prove Theorem 1.3 then proceed to the proof of Theorem 1.1. In both proofs, we first obtain a generating function for the triple sequence under consideration by using the generating function (1.3). Then we apply the Master theorem to transform the problem to a combinatorial one. Finally using elementary manipulations we handle the combinatorial problem.

Proof of Theorem 1.3 *for* $\alpha = 0, 1, 2, \cdots$. A generating function for $B^\alpha(k, m, n)$ can be obtained as follows

$$\sum_{k,m,n=0}^{\infty} B^\alpha(k, m, n)r^k s^m t^n = \int_0^\infty e^{-x}x^\alpha\left\{\sum_{k,m,n=0}^{\infty} r^k L_k^\alpha(\lambda x)s^m L_m^\alpha((1-\lambda)x)t^n L_n^\alpha(x)\right\} dx$$
$$= \int_0^\infty e^{-x}x^\alpha\left\{\sum_{k=0}^{\infty} L_k^\alpha(\lambda x)r^k\right\} \cdot\left\{\sum_{m=0}^{\infty} L_m^\alpha((1-\lambda)x)s^m\right\}\left\{\sum_{n=0}^{\infty} L_n^\alpha(x)t^n\right\} dx.$$

Using the generating function (1.3) we get

$$\sum_{k,m,n=0}^{\infty} B^\alpha(k, m, n)r^k s^m t^n = \{(1-r)(1-s)\cdot(1-t)\}^{\alpha-1}\int_0^\infty x^\alpha \exp\left\{-x\left(1+\frac{\lambda r}{1-r}+\frac{(1-\lambda)s}{1-s}+\frac{t}{1-t}\right)\right\} dx$$
$$= \Gamma(\alpha+1)\left\{(1-r)(1-s)(1-t)\left(1+\frac{\lambda r}{1-r}+\frac{(1-\lambda)s}{1-s}+\frac{t}{1-t}\right)\right\}^{-\alpha-1}$$
$$= \Gamma(\alpha+1)\{1-(1-\lambda)r-\lambda s-\lambda rt-(1-\lambda)st+rst\}^{-\alpha-1}.$$

Because α is a nonnegative integer, it then suffices to show that the rational function $\{1-(1-\lambda)r-\lambda s-\lambda rt-(1-\lambda)st+rst\}^{-1}$ has nonnegative power series coefficients. It is not too difficult to see that

$$V_3 = 1-(1-\lambda)x_1-\lambda x_2-\lambda x_1x_3-(1-\lambda)x_2x_3+x_1x_2x_3$$

if we choose the corresponding matrix A_3 of the Master theorem as

$$A_3 = \begin{pmatrix} 1-\lambda & -\sqrt{\lambda(1-\lambda)} & -\sqrt{\lambda} \\ -\sqrt{\lambda(1-\lambda)} & \lambda & -\sqrt{1-\lambda} \\ -\sqrt{\lambda} & -\sqrt{1-\lambda} & 0 \end{pmatrix}.$$

So, it remains to show that the coefficient of $r^k s^m t^n$ in $[(1-\lambda)r-\sqrt{\lambda(1-\lambda)}s-\sqrt{\lambda}t]^k[-\sqrt{\lambda(1-\lambda)}r+\lambda s-\sqrt{1-\lambda}t]^m[-\sqrt{\lambda}r-\sqrt{1-\lambda}s]^n$ is nonnegative. This is a combinatorial problem and can be proved as follows:

$$[(1-\lambda)r-\sqrt{\lambda(1-\lambda)}s-\sqrt{\lambda}t]^k[-\sqrt{\lambda(1-\lambda)}r+\lambda s-\sqrt{1-\lambda}t]^m[-\sqrt{\lambda}r-\sqrt{1-\lambda}s]^n$$
$$= (-1)^n\sum_{i,j}\binom{k}{i}[(1-\lambda)r-\sqrt{\lambda(1-\lambda)}s]^i(-\sqrt{\lambda}t)^{k-i}\binom{m}{j} \cdot[-\sqrt{\lambda(1-\lambda)}r+\lambda s]^j(-\sqrt{1-\lambda}t)^{m-j}[\sqrt{\lambda}r+\sqrt{1-\lambda}s]^n$$
$$= (-1)^{k+m+n}\sum_{i,j}(-1)^i\binom{k}{i} \cdot\binom{m}{j}[\sqrt{1-\lambda}r-\sqrt{\lambda}s]^{i+j}[\sqrt{\lambda}r+\sqrt{1-\lambda}s]^n t^{k+m-i-j}\lambda^{(k+j-i)/2}(1-\lambda)^{(m+i-j)/2}$$
$$= (-1)^{k+m+n}\sum_{i,j,p,q}(-1)^{j+p}\binom{k}{i}\binom{m}{j}\binom{i+j}{p} \cdot\binom{n}{q}r^{p+q}s^{i+j+n-p-q}t^{k+m-i-j}\lambda^{(q+2j+k-p)/2}(1-\lambda)^{(p+n-q+m+i-j)/2}.$$

Here we follow the standard convention that the bionomial coefficient $\binom{a}{b}$ is zero if either b or $a-b$ is a negative integer. From the above computations we see that the coefficient of $r^k s^m t^n$ can be obtained by letting $p+q=k$, $i+j+n-p-q=m$ and $k+m-i-j=n$. Therefore for $k+m\geqq n$

$$\begin{aligned} B^0(k,m,n) &= \sum_{i,p}(-1)^{i+p}\binom{k}{i}\binom{m}{k+m-n-i}\binom{k+m-n}{p} \\ &\quad\cdot\binom{n}{k-p}\lambda^{2k+m-n-i-p}(1-\lambda)^{n-k+p+i} \\ &= \lambda^{2k+m-n}(1-\lambda)^{n-k}\frac{(k+m-n)!n!}{k!m!}\left\{\sum_i(-1)^i[(1-\lambda)/\lambda]^i\binom{k}{i}\binom{m}{n-k+i}\right\}^2 \\ &\geqq 0. \end{aligned}$$

Clearly $B^0(0,m,n)>0$ if $0<\lambda<1$. This completes the proof.

Note that in the process of proving Theorem 1.3 we actually proved that the coefficient of $r^k s^m t^n$ in $[(1-\lambda)r-\sqrt{\lambda(1-\lambda)}s-\sqrt{\lambda}t]^k[-\sqrt{\lambda(1-\lambda)}r+\lambda s-\sqrt{1-\lambda}t]^m\cdot[-\sqrt{\lambda}r-\sqrt{1-\lambda}s]^n$ is $B^0(k,m,n)$. Now we prove Theorem 1.1.

Proof of Theorem 1.1 *for* $\alpha=0,1,\cdots$. First we compute a generating function for $A^\alpha(2,k,m,n)$. Straightforward manipulations similar to the first steps in the proof of Theorem 1.3 yield, see also Askey and Ismail [4],

$$\begin{aligned} \sum_{k,m,n=0}^{\infty} A^\alpha(2;k,m,n)r^k s^m t^n &= \int_0^\infty e^{-2x}x^\alpha\left\{\sum_{k=0}^\infty L_k^\alpha(x)r^k\right\}\left\{\sum_{m=0}^\infty L_m^\alpha(x)s^m\right\} \\ &\quad\cdot\left\{\sum_{n=0}^\infty L_n^\alpha(x)t^n\right\}dx \\ &= \left\{(1-r)(1-s)(1-t)\left(2+\frac{r}{1-r}+\frac{s}{1-s}+\frac{t}{1-t}\right)\right\}^{-\alpha-1}; \end{aligned}$$

hence it suffices to consider the case $\alpha=0$. Thus

$$2\sum_{k,m,n=0}^{\infty}A^0(2,k,m,n)r^k s^m t^n=\{1-\tfrac{1}{2}(r+s+t)+\tfrac{1}{2}rst\}^{-1}.$$

Next we look for a matrix A_3 such that the corresponding determinant V_3 of the Master theorem is

$$V_3=1-\tfrac{1}{2}(r+s+t)+\tfrac{1}{2}rst.$$

The matrix A_3 is certainly not unique, see Askey and Ismail [4], and

$$A_3=\frac{1}{2}\begin{pmatrix}1&-1&-1\\-1&1&-1\\-1&-1&1\end{pmatrix},$$

is one such matrix. Hence $A^0(2,k,m,n)$ is a positive multiple $(=2^{-k-m-n})$ of the coefficient of $r^k s^m t^n$ in $(r-s-t)^k(-r+s-t)^m(-r-s+t)^n$. The last step in the proof is to establish the nonnegativity of the coefficient of $r^k s^m t^n$ in $(r-s-t)^k(-r+s-t)^m(-r-s+t)^n$. Let $C(k,m,n)$ be the coefficient of $r^k s^m t^n$ in $(r-s-t)^k(-r+s-t)^m(-r-s)^n$. Clearly $(r-s-t)^k(-r+s-t)^m(-r-s+t)^n=\sum_{j=0}^n\binom{n}{j}t^{n-j}(r-s-t)^k(-r+s-t)^m(-r-s)^j$. Therefore the nonnegativity of $C(k,m,n)$ will imply the nonnegativity of $A^0(2,k,m,n)$.

The nonnegativity of $C(k, m, n)$ can be established exactly as in the last part of Theorem 1.3. Indeed it is the special case $\lambda = \frac{1}{2}$ of Theorem 1.3 and the proof is complete.

With reference to the remark made in the Introduction about Kaluza's proof for Theorem 1.3, it should be noted that the use of the Master theorem and the manipulations in the proof of the Theorem 1.1 capture and systematize the essential idea of Kaluza's proof.

3. Weighted permutation problems. The combinatorial approach to the problems discussed in the preceding sections reveals their combinatorial nature. The numbers $A(\mu; k, m, n)$ can be interpreted as the number of certain weighted permutations. Let us consider the following derangement problem. We have three boxes containing k, m and n objects, respectively. We rearrange the objects in such a way that the number of objects in each box remains unchanged and no object remains in its own box. One can easily see that there is a one-to-one correspondence between derangements and ways of obtaining the monomial $x^k y^m z^n$ in $(y+z)^k(x+z)^m(x+y)^n$. Indeed the coefficient of $x^k y^m z^n$ in $(y+z)^k(x+z)^m(x+y)^n$ is the number of derangements. Askey, Ismail and Rashed [6] used this idea and the generating function (1.3) to show that

$$\mathscr{D}(n_1, n_2, \cdots, n_k) = (-1)^k \int_0^\infty e^{-x} L_{n_1}(x) \cdots L_{n_k}(x)\, dx,$$

where $\mathscr{D}(n_1, n_2, \cdots, n_k)$ is the number of derangements of objects in boxes with n_j objects in the jth box, $j = 1, \cdots, k$. These derangements are really weighted permutations. In this case a permutation has weight zero if an object remains in its original box, otherwise the permutation has weight 1. As another example we consider the numbers of Askey and Gasper

$$A^0(2; k, m, n) = \int_0^\infty e^{-2x} L_k(x) L_m(x) L_n(x)\, dx.$$

We have seen, in § 2, that $A^0(2; k, m, n)$ is a positive multiple of the coefficient of $x^k y^m z^n$ in $(x-y-z)^k(-x+y-z)^m(-x-y+z)^n$. The coefficient of $x^k y^m z^n$ in $(x-y-z)^k(-x+y-z)^m(-x-y+z)^n$ has the following interpretation. Consider three boxes with k objects in the first box, m objects in the second box and n objects in the third box. Rearrange the objects so that the number of objects in each box remains unchanged. To each permutation attach the weight

$$(-1)^{\text{number of objects changing boxes}}.$$

It is easy to see that the coefficient of $x^k y^m z^n$ in $(x-y-z)^k(-x+y-z)^m(-x-y+z)^n$ equals the number of these weighted permutations. Another combinatorial interpretation of $A(2; k, m, n)$, in terms of distances in a Hamming scheme, can be found in Askey, Ismail and Koornwinder [5].

As Askey, Ismail and Koornwinder pointed out in [5], the numbers

$$B^0(k, m, n) = \int_0^\infty e^{-x} L_k(\lambda x) L_m((1-\lambda)x) L_n(x)\, dx,$$

have a combinatorial interpretation. We have three boxes containing k, m and n objects, respectively. As we have seen, in § 2, $B^0(k, m, n)$ equals the coefficient of $x^k y^m z^n$ in

$$[(1-\lambda)x - \sqrt{\lambda(1-\lambda)}y - \sqrt{\lambda}z]^k[-\sqrt{\lambda(1-\lambda)}x + \lambda y - \sqrt{1-\lambda}z]^m[-\sqrt{\lambda}x - \sqrt{1-\lambda}y]^n.$$

Thus the third box is a derangement box in the sense that all the objects originally occupying it must move to other boxes. The reader can easily figure out the weights.

For more examples and related results we refer the interested reader to [4] and [5].

4. A problem of H. Lewy. The finite difference approximations to the time dependent wave equation in three (space) dimensions lead to the following four variable analogue of Friedrichs and Lewy's problem.

Conjecture 1 (H. Lewy). If

$$\begin{aligned}&[(1-r)(1-s)(1-t)(1-u)\{(1-r)(1-s)+(1-r)(1-t)+(1-r)(1-u)\\ &\qquad +(1-s)(1-t)+(1-s)(1-u)+(1-t)(1-u)\}]^{-1}\\ &\qquad\qquad = \sum_{k,l,m,n=0}^{\infty} E(k,l,m,n) r^k s^l t^m u^n,\end{aligned} \tag{4.1}$$

then $E(k, l, m, n) \geqq 0$.

The early coefficients in the power series expansion of

$$\{(1-r)(1-s)+(1-r)(1-t)+(1-r)(1-u)+(1-s)(1-t)+(1-s)(1-u)+(1-t)(1-u)\}^{-1}$$

are positive but the later coefficients do change sign, because Huygen's principle holds in three-space. The factor $\{(1-r)(1-s)(1-t)(1-u)\}^{-1}$ is an averaging factor that makes the early positive terms count more than the later terms. This fascinating problem of Lewy was mentioned in Askey [1] and in Askey and Gasper [2]. The following is a combinatorial equivalent of Lewy's conjecture.

Conjecture 2. Let

$$\begin{pmatrix} y_1\\ y_2\\ y_3\\ y_4\end{pmatrix} = \begin{pmatrix} \frac12 & 1 & \sqrt{12}i & -12\\ \frac1{12} & \frac12 & 1 & -\sqrt{12}i\\ -\dfrac{i\sqrt{12}}{144} & \frac1{12} & \frac12 & 1\\ -\frac1{144} & \dfrac{i\sqrt{12}}{144} & \frac1{12} & \frac12\end{pmatrix}\begin{pmatrix} r\\ s\\ t\\ u\end{pmatrix}$$

and let $G(k_1, k_2, k_3, k_4)$ be the coefficient of $r^{k_1}s^{k_2}t^{k_3}u^{k_4}$ in $y_1^{k_1}y_2^{k_2}y_3^{k_3}y_4^{k_4}$. Then

$$\sum_{\substack{0\leqq k_j\leqq l_j\\ 1\leqq j\leqq 4}} G(k_1,k_2,k_3,k_4)\geqq 0 \quad \text{for all } l_j,\ 0\leqq l_j,\ 1\leqq j\leqq 4.$$

The equivalence of conjectures 1 and 2 follows from the MacNahon's Master theorem, since

$$\begin{vmatrix} 1-\frac12 r & -s & -\sqrt{12}it & 12u\\ -\frac1{12}r & 1-\frac12 s & -t & \sqrt{12}iu\\ \dfrac{ir}{12\sqrt{12}} & -\frac1{12}s & 1-\frac12 t & -u\\ \dfrac{r}{144} & \dfrac{-is}{12\sqrt{12}} & -\frac1{12}t & 1-\frac12 u\end{vmatrix}$$

$$\begin{aligned}&= 1-\tfrac12(r+s+t+u)+\tfrac16(rs+rt+ru+st+su+tu)\\ &= \tfrac16[(1-r)(1-s)+(1-r)(1-t)+(1-r)(1-u)+(1-s)(1-t)+(1-s)(1-u)\\ &\qquad +(1-t)(1-u)].\end{aligned}$$

We conclude this section by proving that the 4×4 matrix appearing in Conjecture 2 is essentially unique. This is not the case in all the previously mentioned problems; see Askey and Ismail [4]. Let

$$\begin{vmatrix} 1-ar & -es & -ht & -ju \\ -e'r & 1-bs & -ft & -ku \\ -h'r & -f's & 1-ct & -gu \\ -j'r & -k's & -g't & 1-du \end{vmatrix} = 1-\tfrac{1}{2}(r+s+t+u) + \tfrac{1}{6}(rs+rt+ru+st+su+tu).$$

If we let $r=1$, $s=t=u=0$ we see that $a=\frac{1}{12}$. Similarly $b=c=d=\frac{1}{2}$. Now let $r=s=1$, $t=u=0$. This proves $ee'=\frac{1}{2}$ and similarly $ff'=gg'=hh'=kk'=jj'=\frac{1}{12}$. Thus

$$\begin{vmatrix} 1-\frac{1}{2}r & -es & -ht & -ju \\ -\dfrac{r}{12e} & 1-\frac{1}{2}s & -ft & -ku \\ -\dfrac{r}{12h} & -\dfrac{s}{12f} & 1-\frac{1}{2}t & -gu \\ -\dfrac{r}{12j} & -\dfrac{s}{12k} & \dfrac{-t}{12g} & 1-\frac{1}{2}u \end{vmatrix} = 1-\tfrac{1}{2}(r+s+t+u) + \tfrac{1}{6}(rs+rt+ru+st+su+tu).$$

The substitution $r=s=t=2$, $u=0$ implies $12e^2f^2+h^2=0$; hence $h=\lambda ef$ where $\lambda=\pm\sqrt{12}i$. Similarly $k=\mu fg$, $j=vhg$ where $\mu=\pm\sqrt{12}i$, $\nu=\pm\sqrt{12}i$. Clearly, by rescaling, we may take $e=f=g=1$. If in the resulting identity we let $r=s=t=u=2$ we arrive at

$$1=\frac{1}{9}-\frac{\lambda}{9\mu}-\frac{\lambda\nu}{108}-\frac{\mu}{9\lambda}-\frac{4}{3\lambda\nu}-\frac{(\mu-\nu)^2}{9\mu\nu}.$$

Assuming that $\mu=\delta\lambda$, $\nu=\varepsilon\lambda$, where $\delta=\pm 1$, $\varepsilon=\pm 1$, we can rewrite the above equation as

$$1=\frac{1}{9}-\frac{\delta}{9}+\frac{\varepsilon}{9}-\frac{\delta}{9}+\frac{\varepsilon}{9}-\frac{(\delta-\varepsilon)^2}{9\delta\varepsilon},$$

that is

$$6=2\varepsilon-2\delta-\frac{\delta}{\varepsilon}-\frac{\varepsilon}{\delta},$$

which holds only when $\delta=-1$ and $\varepsilon=1$, since the only admissible values for ε and δ are ± 1. Thus there are only two matrices corresponding to $\lambda=\pm\sqrt{12}i$. The entries in these two matrices are complex conjugates, which is expected because their determinants are real. This shows that, up to a complex conjugation, the matrix appearing in Conjecture 2 is unique.

REFERENCES

[1] R. ASKEY, *Orthogonal Polynomials and Special Functions*, CBMS, vol. 21, Society for Industrial and Applied Mathematics, Philadelphia, 1975.

[2] R. ASKEY AND G. GASPER, *Certain rational functions whose power series have positive coefficients*, Amer. Math. Monthly, 79 (1972), pp. 327–341.

[3] ———, *Convolution structures for Laguerre polynomials*, J. Analyse Math., 31 (1977), pp. 48–68.

[4] R. ASKEY AND M. E. H. ISMAIL, *Permutation problems and special functions*, Canadian J. Math., 28 (1976), pp. 853–874.

[5] R. ASKEY, M. E. H. ISMAIL AND T. KOORNWINDER, *Weighted permutation problems and Laguerre polynomials*, J. Combinatorial Theory ser. A (1978), to appear.

[6] R. ASKEY, M. E. H. ISMAIL AND T. RASHED, *A derangement problem*, MRC Technical Summary Report #1522, Math. Res. Center, Univ. of Wisconsin, Madison, 1975.

[7] D. DEBBI AND J. GILLES, *An inequality relating to triple product integrals of Laguerre functions*, Israel J. Math., 18 (1974), pp. 45–52.

[8] TH. KALUZA, *Elementarer Beweis einer Vermutung von K. Friedricks und H. Lewy*, Math. Z., 37 (1933), pp. 689–697.

[9] T. KOORNWINDER, *Positivity proofs for linearization and connection coefficients of orthogonal polynomials satisfying an addition theorem*, J. London Math. Soc., (1978), to appear.

[10] P. A. MACMAHON, *Combinatory Analysis*, vol. 1, Chelsea, New York, 1960.

[11] E. D. RAINVILLE, *Special Functions*, Macmillan, New York, 1960.

[12] J. RIORDAN, *An Introduction to Combinatorial Analysis*, Wiley, New York, 1958.

[13] G. SZEGÖ, *Über gewisse Potenzreihen mit lauter positiven Koeffizienten*, Math. Z., 37 (1933), pp. 674–688.

[14] ———, *Orthogonal Polynomials*, Colloquium Publications, vol. 23, fourth ed., American Mathematical Society, Providence, RI, 1975.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0006 $01.00/0

ON THE DIFFUSION OF IMMISCIBLE FLUIDS IN POROUS MEDIA*

C. J. VAN DUYN†

Abstract. From the mathematical formulation of the diffusion of two immiscible fluids, we arrive at a nonlinear two-sided degenerate parabolic equation. Existence, uniqueness and a weak maximum principle are proved for the Cauchy problem in the half plane $x \in \mathbb{R}$, $t > 0$. Furthermore, it is shown that the solutions of a class of Cauchy problems converge towards a similarity solution as $t \to \infty$ and the rate of convergence is discussed.

1. Introduction. We shall discuss some mathematical aspects which arise in the study of the diffusion of two immiscible fluids, e.g. water and oil, in a porous medium, under the following preliminary assumptions: the flow is laminar and one-dimensional, the medium is isotropic and homogeneous and the fluids are incompressible. Let S_i, $i = 1, 2$, denote the saturation of the fluids and $k_i(S_i)$ and v_i their coefficients of conductivity and velocities, respectively.

Then, the equation which describes the saturation of one of the fluids as a function of position and time, is found by combining the continuity equation

$$\frac{\partial S_i}{\partial t} + \frac{\partial v_i}{\partial x} = 0, \qquad i = 1, 2, \tag{1}$$

Darcy's law

$$v_i = -k_i(S_i)\frac{\partial \Phi_i}{\partial x}, \qquad i = 1, 2, \tag{2}$$

and the complementary conditions

$$S_1 + S_2 = 1, \tag{3}$$

and

$$p_2 - p_1 = p_c(S_1). \tag{4}$$

Here Φ_i, which depends linearly on the pressure p_i, is the piezometric head of the fluids and $p_c(S_1)$ is the capillary pressure; see Bear [3] and Morel-Seytoux [10]. Condition (3) expresses the assumption of only two phases, while the last condition indicates the pressure difference across the interface between the two fluids. Combining (1) to (4) we obtain an equation which consists of three parts: a flow term, a diffusion term caused by the capillary pressure in the pores and a gravity term [3, p. 468].

In this paper, we shall assume the flow term and the gravity term to be negligible and confine ourselves to the diffusion process. Putting $S_1(x, t) \equiv u(x, t)$, we obtain

$$u_t = (D(u)u_x)_x, \tag{5}$$

where the subscripts denote partial differentiation and

$$D(u) = -\frac{k_1(u)k_2(1-u)}{k_1(u)+k_2(1-u)} \cdot \frac{dp_c(u)}{du}.$$

From the properties of the medium and the fluids, [3, p. 451], we know that the diffusion coefficient is not uniquely defined. This is caused by the effect of hysteresis in

* Received by the editors June 3, 1977, and in revised form January 3, 1978.
† Department of Mathematics, Delft University of Technology, Delft, the Netherlands.

the $p_c(u)$-curve: the difference between drainage and imbibition. However, in this paper we shall neglect this effect. Further, it follows from physical consideration that $dp_c(s)/ds<0$ for all $s\in[0,1]$ and that the functions $k_1(s)$ and $k_2(s)$ are positive for $s>0$ with $k_1(0)=k_2(0)=0$.

Thus we shall consider $D(s)$ to be uniquely defined on $[0,1]$, such that $D(s)>0$ for $s\in(0,1)$ and $D(0)=D(1)=0$. However, this means that equation (5) is degenerate parabolic. In a neighborhood of a point (x,t) where $0<u<1$ it is parabolic, but near points where $u=0$ or $u=1$, it is not. These degeneracies can cause perturbations to have a finite speed of propagation. Moreover, at the interface, where the regions in which $u=1$ and $u\in(0,1)$ meet, u_x may not be continuous. A similar discontinuity can occur at the transition from a region where $u\in(0,1)$ to one where $u=0$. It is therefore necessary to generalize the notion of a solution of (5).

We shall discuss (5) in the strip $S_T=(-\infty,\infty)\times(0,T]$, where T is some fixed positive number, which may eventually tend to infinity. Along the lower boundary we prescribe

$$u(x,0)=u_0(x),\qquad -\infty<x<\infty, \tag{6}$$

where $u_0(x)$ is a given function, which is defined, real and continuous on $(-\infty,\infty)$ such that $u_0\in[0,1]$. Set

$$\phi(u)=\int_0^u D(s)\,ds.$$

Then, following Oleinik, Kalashnikov and Yui-Lin [11], we shall say that a function $u(x,t)$, defined on $\bar S_T$, is a weak solution of the Cauchy problem (5), (6) if:

(i) u is real and continuous in $\bar S_T$, with values in the interval $[0,1]$;

(ii) $\phi(u)$ possesses a bounded generalized derivative with respect to x in S_T;

(iii) u satisfies the identity:

$$\iint_{S_T}\{\zeta_x(\phi(u))_x-\zeta_t u\}\,dx\,dt=\int_{\mathbb{R}}\zeta(x,0)u_0(x)\,dx, \tag{7}$$

for all $\zeta\in C^1(\bar S_T)$ which vanish for large $|x|$ and for $t=T$. In § 2 we shall establish existence and uniqueness of a weak solution of the Cauchy problem, and we shall prove a weak maximum principle.

Equation (5) has a similarity solution, which tends to one as $x\to-\infty$ and which vanishes as $x\to+\infty$. Let $\eta=x(t+1)^{-1/2}$. Then, if we look for solutions of the form $u(x,t)=f(\eta)$, we find that f must satisfy the equation

$$(D(f)f')'+\tfrac12\eta f'=0\quad\text{on }\mathbb{R}, \tag{8}$$

where a prime denotes differentiation with respect to η. At the boundaries we require

$$f(-\infty)=1,\qquad f(+\infty)=0. \tag{9}$$

Van Duyn and Peletier [4] showed that problem (8), (9) has a unique solution which is monotonically decreasing. Moreover, three cases can be distinguished.

$$\text{A.}\quad \int_0\frac{D(s)}{s}\,ds<\infty,\qquad \int^1\frac{D(s)}{1-s}\,ds<\infty. \tag{10}$$

Then there exist numbers a^-, a^+ with $-\infty < a^- < 0 < a^+ < \infty$ such that

$$f(\eta) = 1 \quad \text{for } \eta \in (-\infty, a^-],$$
$$0 < f(\eta) < 1 \quad \text{for } \eta \in (a^-, a^+), \tag{11}$$
$$f(\eta) = 0 \quad \text{for } \eta \in [a^+, \infty).$$

$$\text{B.} \quad \int_0 \frac{D(s)}{s} = \infty, \qquad \int^1 \frac{D(s)}{1-s} = \infty. \tag{12}$$

Then $f(\eta) \in (0, 1)$ for any $\eta \in \mathbb{R}$.

C. Two combinations of the previous cases.

In this case there exists only one number $a \in \mathbb{R}$, and either $f(\eta) = 1$ for $\eta \in (-\infty, a]$ and $f(\eta) \in (0, 1)$ for $\eta \in (a, \infty)$ with $a < 0$, or $f(\eta) \in (0, 1)$ for $\eta \in (-\infty, a)$ and $f(\eta) = 0$ for $\eta \in [a, \infty)$ with $a > 0$.

In § 3 we shall show that in all three cases, convergence of a solution of (5), (6) towards the similarity solution will occur as $t \to \infty$, if $u_0(x)$ has the same asymptotic behavior as $f(x)$ for large values of $|x|$.

We shall obtain two types of estimates. An integral estimate of the form

$$\int_{\mathbb{R}} |\tilde{u}(\eta, t) - f(\eta)| \, d\eta = O(t^{-1/2}) \quad \text{as } t \to \infty,$$

where we have set $\tilde{u}(\eta, t) \equiv u(x, t)$, and secondly, pointwise estimates whose form depends on the three above mentioned cases. We find

$$\text{A.} \quad \left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} \frac{D(s)}{s(1-s)} \, ds \right| = O(t^{-1/2});$$

$$\text{B.} \quad \left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} D(s) \, ds \right| = O(t^{-1/2});$$

$$\text{C.} \quad \text{Either } \left| \int_{u(x,t)}^{f(x(t-1)^{-1/2})} \frac{D(s)}{1-s} \, ds \right| = O(t^{-1/2}),$$

$$\text{or} \quad \left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} \frac{D(s)}{s} \, ds \right| = O(t^{-1/2});$$

and all these estimates hold uniformly in $x \in \mathbb{R}$.

2. Uniqueness and existence of the weak solution. We start with the following uniqueness result:

THEOREM 1. *There exists at most one weak solution of the Cauchy problem* (5), (6).

Proof. Since the proof is identical to that of Theorem 1 of [11], we shall omit it here.

To establish the existence of a weak solution of (5), (6), we use a constructive method, which is based on the one Oleinik, Kalashnikov and Yui-Lin [11] introduced in proving the existence of a weak solution of equation (5), in which $D(s) \geqq 0$ and only vanishes at $s = 0$. First we write

$$v(x, t) = \phi(u(x, t)), \qquad v_0(x) = \phi(u_0(x)). \tag{13}$$

Then problem (5), (6) becomes

(14) $$v_t = a(v)v_{xx} \quad \text{in } S_T,$$

(15) $$v(x, 0) = v_0(x) \quad \text{on } \mathbb{R}.$$

In [7], [11] the existence of a weak solution of problem (14), (15) was proved by approximating the initial function v_0 by a sequence of C^∞ functions, defined on expanding intervals, which are bounded away from zero. They were obtained by adding to v_0 a small quantity and smoothing the result. A corresponding sequence of smooth solutions, defined on an expanding sequence of cylinders, was then shown to converge to a function v which had all the properties required of a weak solution.

In our problem, the function $D(s)$ vanishes for $s = 1$, as well as for $s = 0$. It is therefore necessary to construct a sequence which is bounded away from both zero and one. Moreover, whereas in [7], [11] the convergence followed at once from a monotonicity argument, we now have to rely on the Ascoli–Arzela theorem. This required an equicontinuity property, which we shall establish first.

About the initial value function u_0, we shall assume that $u_0 \in [0, 1]$ and that it is such that $v_0(x)$ is uniformly Lipschitz continuous on $(-\infty, \infty)$. In view of (13), this implies that there exist positive constants L_0, M_0 such that

$$0 = \phi(0) \leqq v_0(x) \leqq \phi(1) = M_0,$$

and

(16) $$|v_0(x) - v_0(y)| \leqq L_0|x - y| \quad \text{for any } x, y \in \mathbb{R}.$$

Then (16) enables us to construct a sequence of functions $\{v_{0,n}(x)\}_{n=1}^\infty$, whose properties are given in the following lemma:

LEMMA 1. *Let $v_0(x)$ satisfy (16). Then there exists a sequence of functions $\{v_{0,n}(x)\}_{n=1}^\infty$ with $v_{0,n}(x) \in C^\infty([-n, n])$ such that:*

(i) *$v_{0,n} \to v_0$ as $n \to \infty$, uniformly on bounded intervals;*
(ii) *$M_0 2^{-n-1} \leqq v_{0,n} \leqq M_0(1 - 2^{-n-1})$ for all $n \geqq 1$;*
(iii) *$v_{0,n}(x) = M_0(1 - 2^{-n-1})$ on $[-n, -n+1] \cup [n-1, n]$ for all $n \geqq 1$;*
(iv) *$|v_{0,n}'| \leqq L$ for all $n \geqq 1$, where $L \geqq L_0$.*

Proof. Let $\rho \in C_0^\infty(\mathbb{R})$, $\rho \geqq 0$, $\int_{\mathbb{R}} \rho(x)\,dx = 1$ and set, as usual, for any $\varepsilon > 0$, $\rho_\varepsilon(x) = \varepsilon^{-1}\rho(x\varepsilon^{-1})$.

Next define $\omega_\varepsilon \in C^\infty(\mathbb{R})$ by

(17) $$\omega_\varepsilon(x) = \int_{\mathbb{R}} v_0(y)\rho_\varepsilon(x - y)\,dy,$$

and consider a sequence $\{\varepsilon_k\}_{k=1}^\infty$, with $\varepsilon_k \to 0$ as $k \to \infty$, and the sequence of functions $\{h_k(x)\}_{k=1}^\infty$, defined by

(18) $$h_k(x) = \omega_{\varepsilon_k}(x) - \{\omega_{\varepsilon_k}(x) - \tfrac{1}{2}M_0\}2^{-k}.$$

Then it is easy to see that:

(i) $h_k \in C^\infty(\mathbb{R})$ for all $k \geqq 1$;
(ii) $h_k \to v_0$ as $k \to \infty$, uniformly on $\mathbb{R}$;
(iii) $M_0 2^{-k-1} \leqq h_k \leqq M_0(1 - 2^{-k-1})$ for all $k \geqq 1$;
(iv) $|h_k'| \leqq L_0$ for all $k \geqq 1$.

To complete the proof, we define the elements of the sequence $\{v_{0,n}\}_{n=1}^\infty$ by:

(i) $v_{0,n}(x) = h_n(x)$ for $|x| \leqq n - 2$ and for all $n \geqq 3$;
(ii) $v_{0,n}(x) = M_0(1 - 2^{-n-1})$ for $n - 1 \leqq |x| \leqq n$ and for all $n \geqq 1$;
(iii) $M_0 2^{-n-1} \leqq v_{0,n}(x) \leqq M_0(1 - 2^{-n-1})$ and $|v_{0,n}'(x)| \leqq L$ for $n - 2 \leqq |x| \leqq n - 1$ and for all $n \geqq 2$, where $L \geqq L_0$.

In what follows, we shall assume that $D \in C([0,1]) \cap C^{1+\alpha}(0,1)$ for some $\alpha \in (0,1]$.[1] Since $\phi'(s)=D(s)>0$ on $(0,1)$, it follows that there exists a continuous function ψ, with ψ: $R[\phi] \to [0,1]$, such that $\psi(\phi(u))=u$: i.e. $\psi=\phi^{-1}$. Now, for a given $n \geqq 1$, we consider the following mixed initial boundary value problem:

$$v_t = a(v)v_{xx} \quad \text{in } Q_n=(-n,n)\times(0,T], \tag{19}$$

$$v(\pm n, t) = M_0(1-2^{-n-1}) \quad \text{for } t \in [0,T], \tag{20}$$

$$v(x,0)=v_{0,n}(x) \quad \text{for } x \in [-n,n], \tag{21}$$

where $a(s)=D(\psi(s))$.

LEMMA 2. *Problem* (19)–(21) *has a unique classical solution* $v_n(x,t)$ *in* $\bar{Q}_n$, *i.e.* $v_n \in C^{2+\alpha}(\bar{Q}_n)$, *such that*[2]

$$M_0 2^{-n-1} \leqq v_n(x,t) \leqq M_0(1-2^{-n-1}) \quad \textit{for all } (x,t)\in \bar{Q}_n, \tag{22}$$

and

$$|v_{n_x}(x,t)| \leqq \max\{L, M_0\} \quad \textit{for all } (x,t) \in \bar{Q}_n. \tag{23}$$

Proof. In view of the regularity of D, it can easily be seen that $a \in C([0,M_0]) \cap C^{1+\alpha}(0,M_0)$. Now, let $\tilde{a}(s)$ be a $C^{1+\alpha}(-\infty,\infty)$ function such that $\tilde{a}(s)=a(s)$ for $s \in [M_0 2^{-n-1}, M_0(1-2^{-n-1})]$ and $\tilde{a}(s) \in [\frac{1}{2}m, 2M]$ for $s \in (-\infty,\infty)$, where $m, M = \min, \max\{a(s): s \in [M_0 2^{-n-1}, M_0(1-2^{-n-1})]\}$. Then we consider the equation

$$v_t = \tilde{a}(v)v_{xx} \quad \text{in } Q_n, \tag{24}$$

together with the data (20) and (21).

Since (24) is uniformly parabolic, it follows from [9, p. 564] that problem (24), (20) and (21) has a unique solution $v_n \in C^{2+\alpha}(\bar{Q}_n)$. Moreover, by the maximum principle it follows that $M_0 2^{-n-1} \leqq v_n(x,t) \leqq M_0(1-2^{-n-1})$ in $\bar{Q}_n$, so v_n is a solution of (19)–(21) as well. Writing equation (19) as

$$v_t = c(x,t)v_{xx}, \tag{25}$$

with $c(x,t)=a(v(x,t))$, an elementary computation shows that c, $\partial c/\partial x \in C^{\alpha^2}(\bar{Q}_n)$. Therefore, by a result of the linear theory [5, p. 72], we find that $v_{n_x} \in C^{2,1}(Q_n)$. Since it is clear that we have $v_{n_x} \in C(\bar{Q}_n)$, we can apply a standard barrier-function argument [11, p. 675] to equation (25) in order to obtain the gradient estimate (23).

Thus we have constructed a sequence of smooth functions $\{v_n(x,t)\}$, each satisfying (19)–(21), which are bounded and Lipschitz continuous with respect to x in the corresponding cylinder $\bar{Q}_n$, with a constant which does not depend on n.

Furthermore, each v_n satisfies the linear, uniformly parabolic, equation (25). This enables us to apply a theorem of Gilding [6] about the Hölder continuity of solutions of parabolic equations with respect to t. We obtain

$$|v_n(x,t)-v_n(x,t_0)| \leqq C|t-t_0|^{1/2}, \tag{26}$$

for all $x \in [-n,n]$, t, $t_0 \in [0,T]$ and $|t-t_0| \leqq 1$. Here, the constant C depends only on $\max\{L, M_0\}$ and $\sup_{s\in(0,1)} D(s)$.

Therefore, if we fix $I \geqq 1$, we know that the set $\{v_n(x,t)\}_{n=I}^{\infty}$ is bounded and equicontinuous on $\bar{Q}_I$. Hence, by the Ascoli–Arzela theorem, there exists a continu-

[1] Here $C^{1+\alpha}(0,1)$ denotes the set of functions $u \in C^1(0,1)$ with Hölder continuous first derivatives (exponent α), [9, p. 8].

[2] For the definition of $C^{2+\alpha}(\bar{Q}_n)$, see [9, p. 7].

ous function $\bar{v}_I(x, t)$ and a convergent subsequence $\{v_{n_j}(x, t)\}$, with $n_j \geqq I$, such that $v_{n_j}(x, t) \to \bar{v}_I(x, t)$ as $n_j \to \infty$, uniformly on $\bar{Q}_I$. Then, by a diagonal process, it follows that there exists a function $v(x, t)$, defined on $\bar{S}_T$, and a convergent subsequence, denoted by $\{v_k(x, t)\}$ such that $v_k(x, t) \to v(x, t)$ as $k \to \infty$, pointwise on $\bar{S}_T$. Since this convergence is uniform on any bounded subset of $\bar{S}_T$, the limit function v is continuous on $\bar{S}_T$. The continuity of v also follows from (23) and (26). We have

$$|v(x, t) - v(x_0, t_0)| \leqq B\{|x - x_0| + |t - t_0|^{1/2}\},$$

for all (x, t), $(x_0, t_0) \in \bar{S}_T$ with $|t - t_0| \leqq 1$. Here $B = \max\{L, M_0, C\}$. Bounds on v follow from (22). We find

$$0 \leqq v(x, t) \leqq M_0 \quad \text{in } \bar{S}_T.$$

In order to return to the original dependent variable u, we define the sequence $\{u_k(x, t)\}$ and the function $u(x, t)$ by:

$$\{u_k(x, t)\} = \{\psi(v_k(x, t))\} \quad \text{and} \quad u(x, t) = \psi(v(x, t)).$$

Then, by the continuity of ψ, we have $u_k(x, t) \to u(x, t)$ pointwise on $\bar{S}_T$ as $k \to \infty$, and this convergence is uniform on any bounded subset of $\bar{S}_T$. Since v is continuous and $v \in [0, M_0]$, it follows that u is continuous and $u \in [0, 1]$.

It now remains to prove that u is the desired weak solution: i.e. we must prove that $v = \phi(u)$ has a bounded generalized derivative with respect to x in S_T and that u satisfies (7).

Let $\zeta \in C^1(\bar{S}_T)$ be an admissible test function, and let $K \geqq 1$ be such that supp $\zeta \subset Q_K$.

Then, for $k \geqq K$, it follows from (23) that $\{v_{k_x}\} = \{(\phi(u_k))_x\}$ is a bounded sequence in $L^2(Q_K)$. Hence, there exists a subsequence $\{(\phi(u_{k_l}))_x\}$, with $k_l \geqq K$ which converges weakly to a bounded function $p \in L^2(Q_K)$. Thus we have

$$(\Psi, (\phi(u_{k_l}))_x) \to (\Psi, p) \quad \text{as } k_l \to \infty,$$

for all $\Psi \in C_0^1(Q_K)$, where $(\cdot, \cdot)$ denotes the inner product on $L^2(Q_K)$. However, since

$$(\Psi, (\phi(u_{k_l}))_x) = -(\Psi_x, \phi(u_{k_l})),$$

and because $u_{k_l} \to u$ as $k_l \to \infty$, uniformly on $\bar{Q}_K$, we obtain

$$(\Psi, p) = -(\Psi_x, \phi(u)),$$

for all $\Psi \in C_0^1(Q_K)$. Hence p is the generalized derivative of $\phi(u)$.

Because the functions $\phi(u_{k_l})$ are classical solutions of (19), we have

$$\iint_{Q_K} \{\zeta_x \phi_x(u_{k_l}) - \zeta_t u_{k_l}\}\, dx\, dt = \int_{-K}^{+K} \zeta(x, 0) u_{0,k_l}(x)\, dx.$$

Now, letting $k_l \to \infty$, we obtain (7). Because ζ has been chosen arbitrarily, we may conclude that u is indeed a weak solution.

Summarizing, we have the following result:

THEOREM 2. *Let u_0 be a continuous function on $\mathbb{R}$, such that $u_0 \in [0, 1]$ and $\phi(u_0(x))$ is uniformly Lipschitz continuous on $\mathbb{R}$. Further let $D \in C([0, 1]) \cap C^{1+\alpha}(0, 1)$ for some $\alpha \in (0, 1]$. Then there exists a weak solution of the Cauchy problem* (5), (6).

As a supplement to the existence proof we have the following regularity result:

THEOREM 3. *Let u be a weak solution of the Cauchy-problem* (5), (6) *and let (x_0, t_0) be a point in S_T where $u(x_0, t_0) \in (0, 1)$. Then there exists a neighborhood N of (x_0, t_0) such that $u \in C^{2,1}(N)$.*

Proof. Let (x_0, t_0) be a point in S_T where $0 < u(x_0, t_0) < 1$. Then, if $\{u_k\}$ is the convergent sequence constructed in the existence proof, there exists a positive number δ and a neighborhood N_0 of (x_0, t_0) such that $\delta < u_k(x, t) < 1-\delta$ in N_0 for all $k \geqq K$. Here K has been chosen such that $N_0 \subset Q_k$ for $k \geqq K$.

Next we write equation (5) as $u_t = (c(x, t)u_x)_x$, where $c(x, t) = D(u(x, t))$. Now, since every u_k for $k \geqq K$ satisfies this equation in N_0 and because $c(x, t) \in (c', c'')$, where $c', c'' = \inf, \sup \{D(s): s \in (\delta, 1-\delta)\}$, we can apply a theorem of [9, p. 204]. We obtain for a neighborhood $N_1 \subset N_0$ of (x_0, t_0): $u_k \in C^\beta(\bar{N}_1)$ for $k \geqq K$ and $\beta \in (0, 1]$, where the Hölder coefficient and exponent do not depend on k. But then $c \in C^\beta(\bar{N}_1)$ and it follows from the linear theory [5, p. 72] applied to equation (25) that v_k, v_{k_x}, $v_{k_{xx}}$ and v_{k_t} exist and are Hölder continuous in N_1 with exponent β for all $k \geqq K$.

A second application of the linear theory [5, p. 64] to equation (25) in a neighborhood $N_2 \subset N_1$ gives: $v_k \in C^{2+\beta}(\bar{N}_2)$ for all $k \geqq K$, where $\|v_k\|_{2+\beta}$ is uniformly bounded with respect to k. Hence, $v \in C^{2+\beta}(\bar{N}_2)$ and therefore $u \in C^{2,1}(\bar{N}_2)$.

Because of the particular choice of the initial function (21), it is an easy task to prove a maximum principle for the weak formulation of the Cauchy problem (5), (6). This maximum principle will prove to be useful in determining the asymptotic behavior, as $t \to \infty$, of solutions of (5), (6).

THEOREM 4 (Maximum principle.) *Let u_1 and u_2 be two weak solutions of the Cauchy problem* (5), (6) *with initial values u_{01} and u_{02}, such that $u_{01}, u_{02} \in [0, 1]$ and $\phi(u_{01}(x))$, $\phi(u_{02}(x))$ are uniformly Lipschitz continuous on $\mathbb{R}$. Suppose $u_{01}(x) \geqq u_{02}(x)$ for all $x \in \mathbb{R}$. Then $u_1(x, t) \geqq u_2(x, t)$ for all $(x, t) \in \bar{S}_T$.*

Proof. Because $u_{01}(x) \geqq u_{02}(x)$, we have $v_{01}(x) \geqq v_{02}(x)$ and, in view of (17), $\omega_{1\varepsilon}(x) \geqq \omega_{2\varepsilon}(x)$ for $x \in \mathbb{R}$. Now, since $H(\omega_{\varepsilon_k}) = \omega_{\varepsilon_k} - \{\omega_{\varepsilon_k} - \frac{1}{2}M_0\}2^{-k}$ is a monotonically strictly increasing function of ω_{ε_k}, it follows from (18) that $h_{1k}(x) \geqq h_{2k}(x)$ for all $x \in \mathbb{R}$. Therefore, we can construct the functions $v_{01,n}$ and $v_{02,n}$ such that $v_{01,n}(x) \geqq v_{02,n}(x)$ for all $x \in [-n, n]$ and for all $n \geqq 1$. Applying this to the mixed problems (19)–(21) and using the maximum principle for this classical problem, we obtain $v_{1,n}(x, t) \geqq v_{2,n}(x, t)$ for all $(x, t) \in \bar{Q}_n$ and for all $n \geqq 1$. Therefore $v_1(x, t) \geqq v_2(x, t)$ in $\bar{S}_T$ and hence $u_1(x, t) \geqq u_2(x, t)$ in $\bar{S}_T$.

3. Asymptotic behavior. We start this section with a theorem which is closely related to the phenomenon of mass-conservation. We shall follow a method which has been introduced by Gilding [8], in proving mass-conservation for solutions of the equation $u_t = (u^m)_{xx} + (u^n)_x$.

THEOREM 5. *Let u_1 and u_2 be weak solutions of the Cauchy problem* (5), (6), *with initial values u_{01} and u_{02}. If*

$$\int_{-\infty}^{+\infty} |u_{01}(x) - u_{02}(x)|\, dx < \infty,$$

then

$$\int_{-\infty}^{+\infty} \{u_1(x, t) - u_2(x, t)\}\, dx = \int_{-\infty}^{+\infty} \{u_{01}(x) - u_{02}(x)\}\, dx \quad \textit{for all } t \in [0, T].$$

Proof. Define the new initial data

$$u_0^+(0) = \max\{u_{01}(x), u_{02}(x)\}, \tag{27}$$

$$u_0^-(x) = \min\{u_{01}(x), u_{02}(x)\}. \tag{28}$$

Clearly, u_0^+ and u_0^- satisfy the conditions from Theorem 2. Let $u^+(x, t)$ and $u^-(x, t)$ denote the weak solutions of (5), with initial data u_0^+ and u_0^-, respectively. Then, since $u_0^+(x) \geqq u_0^-(x)$ on $\mathbb{R}$, we obtain from the maximum principle

$$u^+(x, t) \geqq u^-(x, t) \quad \text{in } \bar{S}_T.$$

Next, let $\{\alpha_i(x)\}_{i=1}^{\infty}$ be a sequence of functions in $C^{\infty}(\mathbb{R})$, such that

$$\alpha_i = 1 \quad \text{for } |x| \leqq i, \qquad 0 \leqq \alpha_i \leqq 1 \quad \text{for } i \leqq |x| \leqq i+1,$$
$$\alpha_i = 0 \quad \text{for } |x| \geqq i+1,$$

where $\alpha_i''(x)$ is uniformly bounded with respect to $i \geqq 1$ and $x \in \mathbb{R}$, and for fixed $t_0 \in (0, T]$, let $\{\beta_j(t)\}_{j=1}^{\infty}$ be a sequence of functions in $C^{\infty}([0, T])$, such that

$$\beta_j(t) = \left(\int_{j(t-t_0)}^{j(T-t_0)} B(s)\, ds\right) \Big/ \left(\int_{-\infty}^{+\infty} B(s)\, ds\right),$$

where

$$B(s) = \exp\{-1/(1-s^2)\} \quad \text{for } |s| < 1,$$
$$B(s) = 0 \qquad\qquad \text{for } |s| \geqq 1.$$

Since $\alpha_i(x) \cdot \beta_j(t)$ vanishes for $t = T$ and for large values of $|x|$, it is clear that $\zeta(x, t) = \alpha_i(x) \cdot \beta_j(t)$ is an admissible test function for all $i \geqq 1$ and for all $j \geqq 1$. Then, we substitute ζ, u^+, u_0^+ and u^-, u_0^- into the integral identity (7) and subtract. This yields

$$\tag{29} \begin{aligned} &\iint_{S_T} \alpha_i'\beta_j\{\phi_x(u^+) - \phi_x(u^-)\}\, dx\, dt \\ &= \iint_{S_T} \alpha_i\beta_j'\{u^+ - u^-\}\, dx\, dt + \int_{-\infty}^{+\infty} \alpha_i(x)\beta_j(0)\{u_0^+(x) - u_0^-(x)\}\, dx. \end{aligned}$$

Now, by the definition (ii) of a weak solution, we can write the left-hand side of (29) as

$$-\iint_{S_T} \alpha_i''\beta_j\{\phi(u^+) - \phi(u^-)\}\, dx\, dt.$$

Then, letting $j \to \infty$, we apply the dominated convergence theorem and find

$$\tag{30} \begin{aligned} &\int_{-\infty}^{+\infty} \alpha_i(x)\{u^+(x, t_0) - u^-(x, t_0)\}\, dx - \int_{-\infty}^{+\infty} \alpha_i(x)\{u_0^+(x) - u_0^-(x)\}\, dx \\ &= \iint_{S_{t_0,i}} \alpha_i''(x)\{\phi(u^+) - \phi(u^-)\}\, dx\, dt, \end{aligned}$$

where

$$S_{t_0,i} = ((-i-1, -i) \cup (i, i+1)) \times (0, t_0].$$

Now, the right hand side of (30) can be estimated by

$$\left|\iint_{S_{t_0,i}} \alpha''\{\phi(u^+) - \phi(u^-)\}\, dx\, dt\right|$$
$$\leqq \sup_{x \in (i, i+1)} |\alpha_i''(x)| \cdot \sup_{s \in (0,1)} D(s) \cdot \iint_{S_{t_0,i}} |u^+ - u^-|\, dx\, dt \leqq A < \infty,$$

because u^+ and u^- are bounded by definition and the measure of $S_{t_0,i}$ is bounded.

Therefore, from (30), we obtain:

$$\int_{-\infty}^{+\infty} \{u^+(x, t_0) - u^-(x, t_0)\}\, dx < \infty \quad \text{if and only if} \quad \int_{-\infty}^{+\infty} \{u_0^+(x) - u_0^-(x)\}\, dx < \infty,$$

for any $t_0 \in (0, T]$. Let $\int_{-\infty}^{+\infty} |u_{01}(x) - u_{02}(x)|\, dx < \infty$. Then, since $u_0^+ - u_0^- = |u_{01} - u_{02}|$, we can integrate (30) with respect to t_0. We find

$$\iint_{S_T} \alpha_i(x)\{u^+(x, t_0) - u^-(x, t_0)\}\, dx\, dt_0 - T \int_{-\infty}^{+\infty} \{u_0^+(x) - u_0^-(x)\}\, dx \leqq A \cdot T,$$

for all $i \geqq 1$. Therefore $u^+ - u^- \in L^1(S_T)$. However, this means that

$$\iint_{S_{t_{0,i}}} \{u^+ - u^-\}\, dx\, dt \to 0 \quad \text{as } i \to \infty, \text{ for all } t_0 (0, T].$$

So from (30), letting $i \to \infty$, we obtain

$$\int_{-\infty}^{+\infty} \{u^+(x, t) - u^-(x, t)\}\, dx = \int_{-\infty}^{+\infty} \{u_0^+(x) - u_0^-(x)\}\, dx \tag{31}$$

for any $t \in [0, T]$.

Returning to the original dependent variables, we can repeat this procedure to obtain (30), where the functions u^+ and u^- are now replaced by u_1 and u_2. Then, the right-hand side can be estimated by

$$\left| \iint_{S_{t_{0,i}}} \alpha_i''\{\phi(u_1) - \phi(u_2)\}\, dx\, dt \right| \leqq \iint_{S_{t_{0,i}}} |\alpha_i''| \cdot \{\phi(u^+) - \phi(u^-)\}\, dx\, dt,$$

which tends to zero as $i \to \infty$, for all $t_0 \in (0, T]$, and the proof is complete.

Convergence, for $t \to \infty$, of a solution u of the Cauchy problem (5), (6) towards the solution f of problem (8), (9) can now easily be proved. From (31) and the maximum principle, it follows that

$$\int_{-\infty}^{+\infty} |u_1(x, t) - u_2(x, t)|\, dx \leqq \int_{-\infty}^{+\infty} |u_{01}(x) - u_{02}(x)|\, dx, \tag{32}$$

for all $t \in [0, T]$.

Now let $u_1(x, t) \equiv \tilde{u}(\eta, t)$ be the weak solution of (5), (6) with initial value $u_0(x)$, and let $u_2(x, t) = f(\eta)$ be the weak solution of (5), (6) with initial value $f(x)$, which, by a uniqueness argument, is the solution of (8), (9). Then, if

$$\int_{-\infty}^{+\infty} |u_0(x) - f(x)|\, dx = K < \infty,$$

it follows from (32), letting T tend to infinity, that

$$\int_{-\infty}^{+\infty} |\tilde{u}(\eta, t) - f(\eta)|\, d\eta \leqq K(t+1)^{-1/2} \quad \text{for all } t \in [0, \infty).$$

In order to prove the pointwise estimates, we start with the following observation. Let the initial value u_0 satisfy the conditions mentioned in Theorem 2, and let there exist numbers b^+ and b^-, with $-\infty < b^- < b^+ < \infty$, such that $u_0(x) = 1$ for $x \in -\infty, b^-]$, $u_0(x) \in [0, 1]$ for $x \in (b^-, b^+)$ and $u_0(x) = 0$ for $x \in [b^+, \infty)$. Further, let the diffusion coefficient D be of type A. Then it is clear that there exists a number $c \in (0, \infty)$, such that $f(x+c) \leqq u_0(x) \leqq f(x-c)$ on $\mathbb{R}$. Hence, by the maximum principle

$$f((x+c)(t+1)^{-1/2}) \leqq u(x,t) \leqq f((x-c)(t+1)^{-1/2}) \quad \text{for all } (x,t) \in \bar{S}_T.$$

Therefore, we can estimate

$$\tag{33} \begin{aligned} \left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} \frac{D(s)}{s(1-s)} ds \right| &\leqq \int_{f((x+c)(t+1)^{-1/2})}^{f((x-c)(t+1)^{-1/2})} \frac{D(s)}{s(1-s)} ds \\ &= \int_{f((x+c)(t+1)^{-1/2})}^{f((x-c)(t+1)^{-1/2})} \frac{D(s)}{s} ds + \int_{f((x+c)(t+1)^{-1/2})}^{f((x-c)(t+1)^{-1/2})} \frac{D(s)}{1-s} ds, \end{aligned}$$

for all $(x,t) \in \bar{S}_T$. Next, we use an argument of [1, p. 372]. We know from [4], that the solution of (8), (9), with D of type A, satisfies $D(f(\eta)) \cdot f'(\eta) \to 0$ as $\eta \uparrow a^+$ and $\eta \downarrow a^-$, and $f(\eta) \to 0$ as $\eta \uparrow a^+$, $f(\eta) \to 1$ as $\eta \downarrow a^-$. Therefore, if we integrate equation (8) from a^- to η, where $\eta \in (a^-, a^+)$, we find

$$D(f(\eta))f'(\eta) = -\frac{1}{2} \int_{a^-}^{\eta} \xi f'(\xi)\, d\xi.$$

So

$$\begin{aligned} D(f(\eta))|f'(\eta)| = \frac{1}{2} \left| \int_{a^-}^{\eta} \xi f'(\xi)\, d\xi \right| &\leqq \frac{1}{2} \int_{a^-}^{\eta} |\xi| \cdot |f'(\xi)|\, d\xi \\ &\leqq \tfrac{1}{2} \max\{|a^-|, a^+\} \cdot \{1 - f(\eta)\}, \end{aligned}$$

since $f'(\xi)$ is negative on (a^-, a^+), and

$$-\frac{D(f'(\eta))f'(\eta)}{1-f(\eta)} \leqq \tfrac{1}{2} \max\{|a^-|, a^+\}.$$

Because of (10) we may integrate this expression from $\eta = a_1$ to $\eta = a_2$, with $a_2 > a_1$ and $a_1, a_2 \in (a^-, a^+)$, to obtain

$$\tag{34} \int_{f(a_2)}^{f(a_1)} \frac{D(s)}{1-s} ds \leqq \tfrac{1}{2} \max\{|a^-|, a^+\}(a_2 - a_1).$$

Now, letting $a_2 \uparrow a^+$ and $a_1 \downarrow a^-$, we can see that (34) holds for all $a_1, a_2 \in \mathbb{R}$. In the same way we can find, see [1, p. 372]

$$\tag{35} \int_{f(a_2)}^{f(a_1)} \frac{D(s)}{s} ds \leqq \tfrac{1}{2} \max\{|a^-|, a^+\}(a_2 - a_1) \quad \text{for all } a_1, a_2 \in \mathbb{R}.$$

Thus, combining (33), (34) and (35), we find

$$\left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} \frac{D(s)}{s(1-s)} ds \right| \leqq K' \cdot (t+1)^{-1/2} \quad \text{for all } (x,t) \in \bar{S}_T,$$

where $K' = 2c \cdot \max\{|a^-|, a^+\}$.

Summarizing, we have proved the following convergence theorem:

THEOREM 6. *Let u be a weak solution of the Cauchy problem* (5), (6) *and let D be of type* A. *Further, suppose there exists a number $c \in (0, \infty)$ such that $f(x+c) \leqq u_0(x) \leqq f(x-c)$. Then, there exists a constant K', depending on the data of the problem, such that*

$$\left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} \frac{D(s)}{s(1-s)} ds \right| \leqq K'(t+1)^{-1/2},$$

holds for any $t \in [0, \infty)$ and uniformly in $x \in \mathbb{R}$.

To conclude this section, we shall assume that $D(s)$ is of type B. About u_0 we shall assume that $u_0 \in (0, 1)$ and that there exists a constant $c_0 \in (0, \infty)$ such that $f(x+c_0) \leqq u_0(x) \leqq f(x-c_0)$. Then by the maximum principle

$$f((x+c_0)(t+1)^{-1/2}) \leqq u(x,t) \leqq f((x-c_0)(t+1)^{-1/2}) \quad \text{in } \bar{S}_T.$$

Now, we estimate

$$\tag{36} \begin{aligned} \left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} D(s)\,ds \right| &\leqq \int_{f((x+c_0)(t+1)^{-1/2})}^{f((x-c_0)(t+1)^{-1/2})} D(s)\,ds \\ &= F((x-c_0)(t+1)^{-1/2}) - F((x+c_0)(t+1)^{-1/2}), \end{aligned}$$

for all $(x,t) \in \bar{S}_T$, where $F(\eta) = \int_0^{f(\eta)} D(s)\,ds$.

Then, using some results obtained in [4], we know that $F'(\eta) \in C(\mathbb{R})$ and that $|F'(\eta)| \leqq |F'(0)|$ for all $\eta \in \mathbb{R}$. Moreover, for $|F'(0)|$, which is physically the flux at $x = 0$, bounds are known in terms of integrals over the diffusion coefficient. Therefore, from (36) we find

$$\tag{37} \left| \int_{u(x,t)}^{f(x(t+1)^{-1/2})} D(s)\,ds \right| \leqq K'' \cdot (t+1)^{-1/2} \quad \text{for all } (x,t) \in \bar{S}_T,$$

where $K'' = 2c_0|F'(0)|$.

Thus, we obtained the following result:

THEOREM 7. *Let u be a weak solution of the Cauchy problem* (5), (6), *and let D be of type* B. *Further, suppose there exists a number $c_0 \in (0, \infty)$, such that $f(x+c_0) \leqq u_0(x) \leqq f(x-c_0)$. Then, there exists a constant K'', depending on the data of the problem, such that*

$$\left| \int_{u(x,t)}^{f(x(t+1)-1/2)} D(s)\,ds \right| \leqq K''(t+1)^{-1/2},$$

holds for any $t \in [0, \infty)$ and uniformly in $x \in \mathbb{R}$.

Obviously, if D is of type C, one uses either (34) or (35).

Acknowledgment. The author is greatly indebted to Professor L. A. Peletier for his assistance with this work.

REFERENCES

[1] F. V. ATKINSON AND L. A. PELETIER, *Similarity profiles of flows through porous media*, Arch. Rational Mech. Anal., 42 (1971), pp. 369–379.

[2] ———, *Similarity solution of the nonlinear diffusion equation*, Ibid., 54 (1974), pp. 373–392.

[3] J. BEAR, *Dynamics of Fluids in Porous Media*, American Elsevier, New York, 1972.

[4] C. J. VAN DUYN AND L. A. PELETIER, *A class of similarity solutions of the nonlinear diffusion equation*, Nonlinear Analysis, Theory, Methods and Applications, 1 (1977), pp. 223–233.

[5] A. FRIEDMAN, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, NJ., 1964.

[6] B. H. GILDING, *Hölder continuity of solutions of parabolic equations*, J. London Math. Soc., 13 (1976), pp. 103–106.

[7] B. H. GILDING AND L. A. PELETIER, *The Cauchy problem for an equation in the theory of infiltration*, Arch. Rational Mech. Anal., 61, (1976), pp. 127–140.

[8] B. H. GILDING, *Properties of solutions of an equation in the theory of infiltration*, Arch. Rational Mech. Anal., to appear.

[9] O. A. LADYZHENSKAYA, V. A. SOLONNIKOV AND N. N. URAL'CEVA, *Linear and Quasilinear Equations of Parabolic Type*, Translations of Mathematical Monographs, vol. 23, American Mathematical Society, Providence, RI, 1968.

[10] H. J. MOREL-SEYTOUX, *Introduction to flow of immiscible liquids in porous media*, Flow Through Prorous Media, R. J. M. De Wiest, ed., Academic Press, New York, 1969, pp. 455–516.

[11] O. A. OLEINIK, A. S. KALASHNIKOV AND CHZHOU YUI-LIN, *The Cauchy problem and boundary problems for equations of the type of unsteady filtration*, Izv. Akad. Nauk. SSSR Ser. Mat., 22 (1958), pp. 667–704.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0007 $01.00/0

ON WEAK EQUIVALENCE OF LINEAR SYSTEMS AND FINITE STATE SYSTEMS*

C. T. MULLIS† AND R. A. ROBERTS†

Abstract. It is shown that finite state machines exist which are weakly equivalent to linear systems, for some nontrivial definitions of weak equivalence. Two systems, one linear with state space $\mathbb{R}^n$, the other with a finite state space, operate on the same stationary uncorrelated input sequence u. The two systems have real valued output sequences y and $\hat{y}$. Notions of weak equivalence are formulated which involve sets of mixed second moments of the input and two outputs. "Power spectrum equivalence" requires that $E(y_t y_{t+\tau}) = E(\hat{y}_t \hat{y}_{t+\tau})$ for all τ. "Cross-correlation equivalence" requires that $E(u_t y_{t+\tau}) = E(u_t \hat{y}_{t+\tau})$ for all τ. The interdependence of these and other notions of weak equivalence are studied. The existence of weakly equivalent finite state systems is constructively demonstrated for a standard class of linear systems.

1. Introduction. Let us agree to call the system

$$x_{t+1} = Ax_t + bu_t, \qquad y_t = cx_t, \qquad -\infty < t < \infty, \tag{1}$$

where $x_t \in \mathbb{R}^n$, and $u_t, y_t \in \mathbb{R}$ a *linear system* although it is more properly called a discrete-time, finite dimensional, non-time-varying linear system. The vector x_t is the state, u_t the input, and y_t the output. We assume that the system (1) is *stable*, (i.e. its eigenvalues satisfy $|\lambda| < 1$), *controllable*, (i.e. $\rho A^k b = 0$ for all $k \geqq 0$ implies $\rho = 0$), and *observable* ($cA^k x = 0$ for all $k \geqq 0$ implies $x = 0$). The system

$$\phi_{t+1} = f(\phi_t, u_t), \qquad \hat{y}_t = g(\phi_t), \qquad -\infty < t < \infty, \tag{2}$$

is a finite-state system if $\phi_t \in \Phi = \{1, 2, \cdots, m\}$, i.e. the state space is a finite set even though the input u_t and output $\hat{y}_t$ are real valued.

We shall consider the question "when are the systems (1) and (2) in some sense equivalent?" The motivation for this question lies in the fact that physical simulations of the ubiquitous linear system are performed on digital machines, which are in reality finite state systems. The approximation is generated by discretizing the linear system, and, usually, the discretization error is negligible. This requires however that the number m of elements of Φ be quite large. There are two questions which occur about such an approximation.

First, how good an approximation can be obtained if m is constrained, i.e. what is the tradeoff between number of states and output error variance $E(y_t - \hat{y}_t)^2$? This question is the more practical one and is considered in [2] and [6]. In fact the problem is closely related to the classical problem of discretization error in numerical analysis, such as may be found in [3], [4].

The second question is more theoretical. What sort of notions of system equivalence can be formulated so that the systems (1) and (2) are *precisely* equivalent? We shall give some nontrivial formulations in this paper. These involve a "white noise" test input, and mixed second moments involving the sequences $\{u_t\}$, $\{y_t\}$ and $\{\hat{y}_t\}$ which are then stationary second-order random processes.

It is obvious that the error $y_t - \hat{y}_t$ depends on the nature of the common input sequence $\{u_t\}$. Since Φ is finite, one can choose $\{u_t\}$ to generate an arbitrary error. In order to generate a reasonable problem, therefore, one must choose a reasonable class of inputs. There are three more or less standard test inputs for linear systems; a unit

* Received by the editors November 18, 1975, and in final revised form January 3, 1978. This research was supported by the National Science Foundation under Grant GK-43138 and the Army Research Office under Grant DAHCO4-75-G-0153.

† Department of Electrical Engineering, University of Colorado, Boulder, Colorado 80309.

pulse, sinusoids and white noise. The unit pulse response and the frequency response each characterize the input-output relation for a linear system. However, this is no longer true for nonlinear systems. A white noise input, on the other hand, can characterize a wide class of nonlinear systems and was proposed by Wiener [8] for this purpose. We shall thus take $\{u_t\}$ to be an independent and identically distributed random sequence with $E(u_t)=0$ and $E(u_t^2)=1$. The distribution function of u_t is

$$F(u)=Pr\{u_t \leqq u\}. \tag{3}$$

The cross-correlation sequence $E[u_t y_{t+\tau}]$ of the input and output sequences of (1) is known to be the unit pulse response of the system, under the conditions we have imposed on $\{u_t\}$. This fact provides an identification technique which has in fact been realized in hardware. This suggests a notion of weak system equivalence. We shall call the systems (1) and (2) *cross-correlation equivalent* if $E[u_t y_{t+\tau}] \equiv E[u_t \hat{y}_{t+\tau}]$.

The autocorrelation sequence $E[y_t y_{t+\tau}]$ determines the power spectrum of the output process. In many situations, such as signal generation, the output power spectrum is the only important property of the system. We shall call the systems (1) and (2) *power spectrum equivalent* if $E[y_t y_{t+\tau}] \equiv E[\hat{y}_t \hat{y}_{t+\tau}]$. The problem of constructing Markov chains with a real valued output having a specified power spectrum has some precedence in the signal processing literature [5], [7].[1]

We shall constructively demonstrate the existence of cross-correlation and power spectrum equivalent finite-state systems in this paper.

2. Pulse response and autocorrelation sequences. The common input process $\{u_t\}$ for the systems (1) and (2) is independent and identically distributed. Therefore, since the right-hand sides of (1) and (2) do not depend explicitly on time, the state trajectories $\{x_t\}$ and $\{\phi_t\}$ are stationary Markov processes. So in fact, is $\{x_t, \phi_t\}$. We shall briefly discuss these processes and consider all mixed second moments involving elements of the input sequence and the output sequences $\{y_t\}$ and $\{\hat{y}_t\}$.

The sequence $\{x_t\}$ is a stationary, second-order, $\mathbb{R}^n$-valued process with mean

$$E(x_t)=0 \tag{4}$$

and covariance

$$E[x_t x_{t+\tau}^T]=K(A^\tau)^T, \qquad \tau \geqq 0. \tag{5}$$

Here

$$K=AKA^T+bb^T=\sum_{k=0}^{\infty}(A^k b)(A^k b)^T. \tag{6}$$

(Since (A, b) is stable and controllable, the sum in (6) exists and K is positive definite.) The scalar output sequence satisfies

$$y_t=cx_t=\sum_{\tau=1}^{\infty} h_\tau u_{t-\tau}, \tag{7}$$

where

$$h_\tau \triangleq cA^{\tau-1}b=E[u_t y_{t+\tau}], \qquad \tau \geqq 1. \tag{8}$$

[1] See also a recent paper by R. Brockett, *Stationary covariance generation with finite state Markov processes*, 1977 Joint Automatic Control Conference.

The sequence $\{h_t\}$ is called the *pulse response* for the linear system. The autocorrelation sequence for the stationary process $\{y_t\}$ is

$$r_\tau \triangleq E[y_t y_{t+\tau}] = cK(cA^\tau)^T = \sum_{t=1}^{\infty} h_t h_{t+\tau} \tag{9}$$

for $\tau \geqq 0$.

The sequence $\{\phi_t\}$ is a stationary Markov chain taking values in $\Phi \triangleq \{1, 2, \cdots, m\}$. The next-state function f and distribution function F determine an $m \times m$ transition matrix Q and an associated $m \times m$ matrix R via

$$Q(\phi, \phi') = \int_I dF(u), \qquad I = \{u : f(\phi, u) = \phi'\}, \tag{10}$$

$$R(\phi, \phi') = \int_I u\, dF(u), \qquad I = \{u : f(\phi, u) = \phi'\}. \tag{11}$$

(We will usually adhere to the convention that components of vectors and matrices indexed by Φ be identified by arguments. Subscripts will usually denote time.) We have

$$Pr\{\phi_{t+1} = \phi' | \phi_t = \phi\} = Q(\phi, \phi'), \tag{12}$$

$$E\{u_t | \phi_t = \phi, \phi_{t+1} = \phi'\} = \frac{R(\phi, \phi')}{Q(\phi, \phi')}. \tag{13}$$

Let $\mathbf{1}$ denote the m-dimensional column vector whose components are all unity. Then

$$Q\mathbf{1} = \mathbf{1} \tag{14}$$

$$R\mathbf{1} = 0 \tag{15}$$

(since the input is zero-mean). The finite system (2) (or the matrix Q) is *ergodic* if for some t, every component of Q^t is positive. (Note that ergodicity implies a "controllability" property, since for any pair of states in Φ there will be a trajectory of positive probability which joins them.) If Q is ergodic, then $Q^t \to \mathbf{1}p$ where p is a probability vector. The vector p is the unique solution to $p\mathbf{1} = 1$, $pQ = p$, and satisfies

$$Pr\{\phi_t = \phi\} = p(\phi) > 0 \qquad \text{for all } \phi. \tag{16}$$

Let g be the output map in (2); alternately, the column vector whose ϕth component is $g(\phi)$. Suppose that the system (2) is ergodic. Then the output process $\{\hat{y}_t\}$ has autocorrelation sequence

$$\begin{aligned} \hat{r}_\tau \triangleq E[\hat{y}_t \hat{y}_{t+\tau}] &= E[g(\phi_t) g(\phi_{t+\tau})] \\ &= \sum_{\phi, \phi'} p(\phi) Q^\tau(\phi, \phi') g(\phi) g(\phi') \\ &= g^T D_p Q^\tau g, \qquad \tau \geqq 0, \end{aligned} \tag{17}$$

where

$$D_p = \operatorname{diag}\{p(1), \cdots, p(m)\}. \tag{18}$$

(We use this notation throughout. Note that if Q is ergodic, D_p is nonsingular.)

The state of a deterministic system characterizes an equivalence class of past input sequences, motivating the use of the word "memory." Let us consider the memory (in the mean) that the finite-state machine (2) has of its past input.

$$
\begin{aligned}
E[u_{t-\tau}|\phi_t] &= E[E[u_{t-\tau}|\phi_{t-\tau}=\phi, \phi_{t-\tau+1}=\phi', \phi_t]|\phi_t] \\
&= \sum_{\phi,\phi'} \left[\frac{R(\phi,\phi')}{Q(\phi,\phi')}\right]\cdot\left[\frac{p(\phi)Q(\phi,\phi')Q^{\tau-1}(\phi',\phi_t)}{p(\phi_t)}\right] \\
&= [pRQ^{\tau-1}D_p^{-1}](\phi_t), \qquad \tau>0.
\end{aligned} \tag{19}
$$

Using this result, we may compute the cross-correlation of the input and output sequences for the finite-state machine. In analogy with the pulse response of the linear system, define for $\tau>0$

$$
\begin{aligned}
\hat{h}_\tau \triangleq E[u_t\hat{y}_{t+\tau}] &= E[E[u_t g(\phi_{t+\tau})|\phi_{t+\tau}]] \\
&= \sum_{\phi\in\Phi} p(\phi)g(\phi)[pRQ^{\tau-1}D_p^{-1}](\phi) \\
&= pRQ^{\tau-1}g.
\end{aligned} \tag{20}
$$

We shall call the sequence $\{\hat{h}_\tau\}$ the *statistical pulse response* for the finite-state machine.

One can express the expectation of the product of any two elements drawn from the sequences $\{u_t\}$, $\{y_t\}$, $\{\hat{y}_t\}$ in terms of the sequences $\{h_\tau\}$, $\{\hat{h}_\tau\}$, $\{\hat{r}_\tau\}$. The sequence $\{r_\tau\}$ is expressible in terms of $\{h_\tau\}$ (equation (9)). The only such expectation that has not been discussed is

$$
\begin{aligned}
E[y_t\hat{y}_{t+\tau}] &= E\left[\left(\sum_{s=1}^{\infty} h_s u_{t-s}\right)\hat{y}_{t+\tau}\right] \\
&= \sum_{s=1}^{\infty} h_s\hat{h}_{\tau+s},
\end{aligned} \tag{21}
$$

where $\hat{h}_\tau=0$ for $\tau\leqq 0$.

3. Weak system equivalence. Two systems are generally considered equivalent if they can't be distinguished externally; that is, identical inputs produce identical outputs. We will call the systems (1) and (2) *externally equivalent* if

$$
E(y_t-\hat{y}_t)^2=0. \tag{22}
$$

It is unreasonable to expect that external equivalence would be possible unless the system (1) is somehow trivial. Therefore it is of interest to consider weaker notions of equivalence. The following are of special interest in the theory of linear systems. We shall call the systems (1) and (2)

(i) *cross-correlation equivalent* if $h_\tau=\hat{h}_\tau$ for all $\tau>0$;

(ii) *power spectrum equivalent* if $r_\tau=\hat{r}_\tau$ for all $\tau\geqq 0$.

Each of these notions of equivalence involve the systems (1) and (2) and the distribution function F.

One can compute the mean squared error $E(y_t-\hat{y}_t)^2$ from the pulse response and autocorrelation sequences as follows. Using (9), (17), and (21), we have

$$
\begin{aligned}
E(y_t-\hat{y}_t)^2 &= r_0-2\sum_{\tau=1}^{\infty} h_\tau\hat{h}_\tau+\hat{r}_0 \\
&= \left[\sum_{\tau=1}^{\infty}(h_\tau-\hat{h}_\tau)^2\right]+\left[\hat{r}_0-\sum_{\tau=1}^{\infty}\hat{h}_\tau^2\right].
\end{aligned} \tag{23}
$$

The two bracketed error terms in (23) have the following interpretation. Define an "equivalent linear" system for the finite state machine to be the *linear* system defined by

$$\tilde{y}_t = \sum_{\tau=1}^{\infty} \hat{h}_\tau u_{t-\tau}.$$

Then the first term in (23), namely

$$\|h - \hat{h}\|^2 = E(y_t - \tilde{y}_t)^2$$

is the mean squared error between the two linear systems. The second term

$$\hat{r}_0 - \|\hat{h}\|^2 = E(\hat{y}_t - \tilde{y}_t)^2$$

is the mean squared error between the finite-state machine and its equivalent linear system. This can serve as a measure of nonlinearity, for the finite state system.

For a given next-state map f, the output map g which minimizes $E(y_t - \hat{y}_t)^2$ must agree with the conditional expectation, i.e.,

$$\begin{aligned} g(\phi) &= E[y_t|\phi_t = \phi] \\ &= \sum_{\tau=1}^{\infty} h_\tau E[u_{t-\tau}|\phi_t = \phi] \qquad (24)\\ &= \sum_{\tau=1}^{\infty} h_\tau (pRQ^{\tau-1}D_p^{-1})(\phi). \end{aligned}$$

We will call this map the *minimum variance output map*. For machines with minimum variance output maps,

$$\begin{aligned} \hat{r}_\tau &= g^T D_p Q^\tau g \\ &= \sum_{t=1}^{\infty} h_t (pRQ^{t+\tau-1} g) = \sum_{t=1}^{\infty} h_t \hat{h}_{t+\tau}. \qquad (25) \end{aligned}$$

It follows that $E[\hat{y}_t \hat{y}_{t+\tau}] = E[y_t \hat{y}_{t+\tau}]$ (see (21)).

If the systems (1) and (2) are externally equivalent, then they are cross-correlation equivalent and power spectrum equivalent. Furthermore (24) holds. The following proposition considers the converse.

PROPOSITION 1. *Suppose that the finite-state system* (2) *is ergodic. The following three statements are incompatible in the sense that if any two hold, then the systems* (1) *and* (2) *must be externally equivalent.*

(a) $\hat{h}_\tau = h_\tau$, $\tau \geqq 0$ (*Cross-correlation equivalence*);

(b) $\hat{r}_\tau = r_\tau$, $\tau \geqq 0$ (*Power spectrum equivalence*);

(c) $\hat{y}_t = E(y_t|\phi_t)$ (*Minimum variance output map*).

Proof. If (c) holds, then (25) holds and for $\tau = 0$, we have

$$\hat{r}_0 = \sum_{t=1}^{\infty} h_t \hat{h}_t.$$

Consequently (23) reduces to

$$E(y_t - \hat{y}_t)^2 = r_0 - \hat{r}_0.$$

If (a) holds, then (23) reduces to

$$E(y_t - \hat{y}_t)^2 = \hat{r}_0 - r_0.$$

If (b) holds, then $r_0 = \hat{r}_0$. Consequently, if any two of the three hold, then $E(y_t - \hat{y}_t)^2 = 0$. Q.E.D.

Is external equivalence possible? Given the disparity between the state spaces of the systems (1) and (2), it is obvious that if they are externally equivalent, then the linear system must trivialize. The following proposition characterizes this situation.

PROPOSITION 2. *There exists a finite state machine externally equivalent to the system* (1) *if and only if* $\{\tau: h_\tau \neq 0\}$ *is finite, and there is a finite set* U *for which* $Pr\{u_t \in U\} = 1$.

Let U be the support of $Pr\{\cdot\}$. Suppose that $n = \max\{\tau: h_\tau \neq 0\} < \infty$ and that U is finite. We may then take $\Phi = U^n$, with obvious choices of f and g to construct a finite-state machine externally equivalent to the system (1).

On the other hand suppose that $E(y_t - \hat{y}_t)^2 = 0$ for some finite-state system. Then with probability one, y_t can take on only finitely many values. Consider the input-output map (7). We must have $h_k \neq 0$ for some $k > 0$ since $n \geqq 1$ and the system (1) is a minimal realization. Assuming U is infinite means that u_{t-k} can take on infinitely many values, leading to a contradiction. Therefore U is finite. But U must contain more than one element since u_t has mean zero and variance one. If the set $\{\tau: h_\tau \neq 0\}$ were infinite, then the sequence $\{h_0, h_1, \cdots\}$ would contain infinitely many values (since $h \in \ell_2$). Therefore, the set of possible right hand sides of (7) is infinite; again a contradiction. Q.E.D.

External equivalence is possible only in trivial situations. However, we shall show (§ 6) that if (A, b) is stable and controllable and $F(\cdot)$ is continuous, then there exist finite-state machines which are cross-correlation equivalent to (1) and finite-state machines which are power spectrum equivalent to (1). The construction of these machines is via an "internal" approximation, wherein the *state* x_t of (1) is estimated, rather than merely the output y_t.

4. Internal system equivalence. Let $\hat{x}: \Phi \to \mathbb{R}^n$ be a map which assigns a point in $\mathbb{R}^n$ to each element ϕ of the state space of (2), and satisfies

$$c\hat{x}(\phi) = g(\phi) \quad \text{for all } \phi \in \Phi. \tag{26}$$

We will think of $\hat{x}(\phi_t)$ as an "estimate" for x_t, and will call a machine described by the maps $(f, \hat{x})$ an *internal approximation* of the system (1). Let $\hat{X}$ be the $n \times m$ matrix whose ϕth column is $\hat{x}(\phi)$. Then the vector g with components $\{g(\phi)\}$ is $g = (c\hat{X})^T$.

The idea of estimating the state of the linear system suggests further notions of system equivalence. We will now list some of these and partially catalogue the dependence of the weak equivalences in the list. If all hold, then the systems must be *internally* equivalent, i.e. the covariance of the vector $x_t - \hat{x}(\phi_t)$ must vanish. In the following, we assume that the system (2) is ergodic.

(E1) $$E[\hat{x}(\phi_t)] = E[x_t] = 0, \quad \text{or} \quad \sum_{\phi\in\Phi} p(\phi)\hat{x}(\phi) = \hat{X}p^T = 0.$$

(E2) $$E[\hat{x}(\phi_t)\hat{x}(\phi_t)^T] = E[x_t x_t^T] = K, \quad \text{or} \quad \sum_{\phi\in\Phi} p(\phi)\hat{x}(\phi)\hat{x}(\phi)^T = \hat{X}D_p\hat{X}^T = K.$$

(E3) $$\hat{x}(\phi_t) = E[x_t|\phi_t] \quad \text{(minimum variance condition), or}$$
$$\hat{X} = A\hat{X}(D_pQD_p^{-1}) + bpRD_p^{-1} = \sum_{\tau=0}^{\infty} A^\tau bpRQ^\tau D_p^{-1}.$$

(E4) $$E[\hat{x}(\phi_{t+1})|\phi_t] = A\hat{x}(\phi_t), \quad \text{or} \quad A\hat{X} = \hat{X}Q^T.$$

(E5) $$E[u_t\hat{x}(\phi_{t+1})] = E[u_t x_{t+1}], \quad \text{or} \quad \hat{X}(pR)^T = b.$$

PROPOSITION 3. *If* (A, b) *is stable and controllable, and* Q *is ergodic, then*

(i) E3 $\Rightarrow$ E1;

(ii) E4 $\Rightarrow$ E1;

(iii) E4, E5 $\Rightarrow$ *internal cross correlation equivalence, i.e.*

$$E[u_t\hat{x}(\phi_{t+\tau})] = E(u_t x_{t+\tau}), \qquad \tau > 0;$$

(iv) E2, E4 $\Rightarrow$ *internal power spectrum equivalence, i.e.*

$$E[\hat{x}(\phi_t)\hat{x}(\phi_{t+\tau})^T] = E[x_t x_{t+\tau}^T], \qquad \tau \geqq 0;$$

(v) E2, E3 $\Rightarrow$ *internal equivalence, i.e.*

$$E\{[x_t - \hat{x}(\phi_t)][x_t - \hat{x}(\phi_t)]^T\} = 0.$$

(vi) E3, E4, E5 $\Rightarrow$ *internal equivalence.*

Proof. (i) $E[\hat{x}(\phi_t)] = E[E[x_t|\phi_t]] = E[x_t]$.

(ii) Take the expectation of E4 to get $(A-I)E[\hat{x}(\phi_t)] = 0$. Since A is stable, 1 is not an eigenvalue of A, and $A-I$ is invertible.

(iii) $E[u_t x_{t+\tau}] = A^{\tau-1}b = A^{\tau-1}\hat{X}(pR)^T = \hat{X}(pRQ^{\tau-1})^T = E[u_t\hat{x}(\phi_{t+\tau})]$.

(iv) $K(A^T)^\tau = \hat{X}D_p\hat{X}^T(A^T)^\tau = \hat{X}D_pQ^\tau\hat{X}^T = \sum_{\phi,\phi'}[p(\phi)Q^\tau(\phi,\phi')]\hat{x}(\phi)\hat{x}(\phi')^T$.

(v) Given E2,

$$\operatorname{cov}(x_t - \hat{x}(\phi_t)) = K - E[x_t\hat{x}(\phi_t)^T] - E[\hat{x}(\phi_t)x_t^T] + K.$$

But from E3

$$\begin{aligned} E[x_t x(\hat{\phi}_t)^T] &= E[E[x_t|\phi_t]\hat{x}(\phi_t)^T] \\ &= E[\hat{x}(\phi_t)\hat{x}(\phi_t)^T] \\ &= K. \end{aligned}$$

(vi) In view of (v) it suffices to show that E3, E4, E5 $\Rightarrow$ E2. Thus

$$\begin{aligned} \hat{X}D_p\hat{X}^T &= A\hat{X}D_pQ\hat{X}^T + bpR\hat{X}^T \\ &= A(\hat{X}D_p\hat{X}^T)A^T + bb^T \end{aligned}$$

Since A is stable, the solution K to (6) is unique, and therefore E2 holds. Q.E.D.

5. A Markov chain approximation. In the preceding section certain notions of weak equivalence were introduced. In § 6 it is shown that the consideration of these equivalence relations is not futile; for any linear system and suitably nontrivial input there exist finite state machines which are either cross-correlation or power spectrum equivalent. In order to construct such finite state machines, one is led to consider the problem which is posed and solved in the present section.

Consider an ergodic finite Markov chain with state space $\Phi = \{1, 2, \cdots, m\}$, transition matrix Q defined on $\Phi \times \Phi$, and trajectories $\{\phi_t\}$. These trajectories satisfy the Markov property and the fundamental conditional probabilities (12). Let $\hat{x}: \Phi \to \mathbb{R}^n$ be a given mapping from the state space of the chain into the state space of the linear system (1). The vector valued sequence $\{\hat{x}(\phi_t)\}$ then depends *only* on the trajectory of the Markov chain.

The problem posed in this section addresses the following question. *Can the Markov chain and estimator* $\hat{x}$ *be constructed so that any sample sequence* $\hat{x}(\phi_t)$ *is also a possible trajectory of the linear system* (1)?

It should be clear that not every vector sequence is such a possible trajectory. There must exist a scalar input sequence $\tilde{u}_t$ (for every sample sequence ϕ_t) for which

(27) $$\hat{x}(\phi_{t+1}) = A\hat{x}(\phi_t) + b\tilde{u}_t, \quad \text{for all } t.$$

In other words, the vector $\hat{x}(\phi_{t+1}) - A\hat{x}(\phi_t)$ must be a scalar multiple of the vector b. If this is the case, then the random sequence $\tilde{u}_t$ is uniquely determined by (27). This pseudo "input" is actually a function of the trajectory $\{\phi_t\}$. (Therefore, the usual cause and effect relationship of input and state are reversed.) For the linear system, the present state x_t and input u_t are independent random variables. We shall require a weaker version for $\tilde{u}_t$, namely

(28) $$E[\tilde{u}_t|\phi_t] = 0.$$

Finally, in order to avoid trivialities we shall also require that

(29) $$\sigma^2 \triangleq E(\tilde{u}_t^2) > 0.$$

An algebraic version of this problem is as follows:

Problem statement. Given (A, b) stable and controllable, find an integer m, an ergodic transition matrix Q, a state estimator function $\hat{x}$ and a real matrix v, for which

(30) $$Q(i,j)[\hat{x}(j) - A\hat{x}(i) - bv(i,j)] = 0, \qquad 1 \leqq i, j \leqq m,$$

(31) $$\sum_{j=1}^{m} Q(i,j) \cdot v(i,j) = 0, \qquad 1 \leqq i \leqq m,$$

(32) $$\hat{x}(i) \neq 0 \quad \text{for some } i.$$

The correspondence of this problem with the original question is established by setting $\tilde{u}_t = v(\phi_t, \phi_{t+1})$. Note that given $(Q, \hat{x})$ the matrix v is essentially determined by (30); $v(k, j)$ is arbitrary if $Q(k, j) = 0$. It is perhaps not obvious that (30)–(32) guarantee that

(33) $$\sigma^2 = \sum_{i,j=1}^{m} p(i)Q(i,j)[v(i,j)]^2 > 0$$

(where p is the stationary measure for Q). We will demonstrate this below. Consider, for the present, certain properties of solutions.

Let $\hat{X}$ be the $n \times m$ matrix whose ϕth column is $\hat{x}(\phi)$. Let $\tilde{R}$ be the $m \times m$ matrix with components

(34) $$\tilde{R}(i,j) = Q(i,j) \cdot v(i,j)$$

in analogy with the matrix R in (11). We may then rewrite (31) as

(35) $$\tilde{R}\mathbf{1} = 0$$

which is analogous to (15). Multiply (30) by $p(i)$ and sum over i to get

(36) $$\hat{X}D_p = A\hat{X}D_pQ + bp\tilde{R}.$$

This is analogous to (E3). Now sum (30) over j, using (34) to get

(37) $$A\hat{X} = \hat{X}Q^T$$

which is (E4). Multiply (30) by $p(i)v(i,j)$ and sum over i, j to get

(38) $$\hat{X}(p\tilde{R})^T = \sigma^2 b$$

which is similar to (E5). Define

(39) $$\hat{K} = \hat{X} D_p \hat{X}^T = E[\hat{x}(\phi_t)\hat{x}(\phi_t)^T].$$

Equations (36), (37) and (38) may be combined to get

(40) $$\hat{K} = A\hat{K}A^T + \sigma^2 bb^T.$$

Since the only solution to the homogeneous equation $Y = AYA^T$ is the trivial solution $Y = 0$, (6) and (40) imply that

(41) $$\hat{K} = \sigma^2 K.$$

It follows that there can be but two possibilities. Either $\hat{X} = 0$, $\sigma^2 = 0$ and $\hat{K} = 0$, or $\hat{X} \neq 0$, $\sigma^2 > 0$ and $\hat{K}$ is positive definite. Thus if (32) holds, then $\sigma^2 > 0$ as claimed.

The random sequence $\{\tilde{u}_t\}$ is uncorrelated. To see this, write

(42) $$\begin{aligned} E[\tilde{u}_t\tilde{u}_{t+\tau}] &= E[E[\tilde{u}_t\tilde{u}_{t+\tau} | \phi_t, \phi_{t+1}, \phi_{t+\tau}, \phi_{t+\tau+1}]] \\ &= p\tilde{R}Q^{\tau-1}\tilde{R}\mathbf{1} \\ &= 0 \quad \text{for } \tau > 0, \qquad \text{(by (35))}. \end{aligned}$$

(This is a consequence of the Markov property.) Finally, consider the covariance function for $\{\hat{x}(\phi_t)\}$. Using (37) and (41) we find that

(43) $$\begin{aligned} E[\hat{x}(\phi_t)\hat{x}(\phi_{t+\tau})^T] &= \hat{X} D_p Q^\tau \hat{X}^T \\ &= (\hat{X} D_p \hat{X}^T)(A^T)^\tau \\ &= \sigma^2 K (A^T)^\tau. \end{aligned}$$

Thus, any special approximation for which $\sigma^2 = 1$ is (internally) power spectrum equivalent to the system (1).

Construction. In order to find a special approximation, one must first construct a finite set of vectors $\{\hat{x}(\phi)\}$ such that for each ϕ, there is a ϕ' for which $\hat{x}(\phi') - A\hat{x}(\phi) = bv(\phi, \phi')$. Furthermore the set of all such $v(\phi, \phi')$ must contain zero in its convex hull. This set will be constructed using a special basis for $\mathbb{R}^n$. Given such a set, one may construct a transition matrix Q such that (30) and (31) hold. It then becomes necessary to verify that the transition matrix is ergodic.

Consider the following special basis for $\mathbb{R}^n$. Let

$$a(z) = z^n + a_1 z^{n-1} + \cdots + a_n = \det(zI - A)$$

and let

$$\psi_1 = b, \qquad \psi_{k+1} = A\psi_k + a_k b \quad \text{for } k = 1, \cdots, n-1.$$

Since (A, b) is controllable, these vectors are linearly independent. By the Cayley–Hamilton theorem,

(44) $$A\psi_n + a_n b = a(A)b = 0.$$

Consider the lattice L of vectors whose components with respect to the basis $\{\psi_k\}$ are integer valued, i.e.

$$L = \left\{ \sum_{k=1}^{n} \xi_k \psi_k : \xi_k \in \mathbb{Z}, k = 1, 2, \cdots, n \right\}.$$

If $x \in L$, then

$$
\begin{aligned}
Ax + bu &= A \sum_{k=1}^{n} \xi_k \psi_k + bu \\
&= \sum_{k=1}^{n} \xi_k (A\psi_k + a_k b) + b\left(u - \sum_{k=1}^{n} a_k \xi_k\right) \qquad (45)\\
&= \left(u - \sum_{k=1}^{n} a_k \xi_k\right)\psi_1 + \sum_{k=2}^{n} \xi_{k-1}\psi_k.
\end{aligned}
$$

Therefore, $Ax + bu$ is also an element of L, provided that

$$u - \sum_{k=1}^{n} a_k \xi_k \quad \text{is an integer.}$$

(Note that (44) was used in (45).)

Denote the set of points reachable from the origin via the system (1) with inputs bounded by 1, in time t by

$$\mathscr{A}(t) = \left\{ \sum_{k=1}^{t} A^{k-1} b u_k,\ |u_k| \leqq 1 \ \text{ for } k = 1, \cdots, t \right\}.$$

Let

$$\mathscr{A}(\infty) = \bigcup_{t>0} \mathscr{A}(t).$$

Since the eigenvalues of A satisfy $|\lambda| < 1$, $\mathscr{A}(\infty)$ is bounded. Impose the norm

$$\|x\|_\infty = \max\{|\xi_j| : j = 1, \cdots, n\} \quad \text{where } x = \sum_{j=1}^{n} \xi_i \psi_j,$$

and choose M satisfying $x \in \mathscr{A}(\infty) \Rightarrow \|x\|_\infty < M$.

LEMMA. *For any $x \in L$, there is a trajectory of the system* (1) *which begins at x, remains always in L, and passes through the origin in finite time, with $|u_t| \leqq M(|a_1| + \cdots + |a_n|)$.*

Proof. If $x_t = \sum_{j=1}^{n} \xi_{jt} \psi_j \in L$, and $u_t - \sum_{j=1}^{n} a_j \xi_{jt} \in \mathbb{Z}$, then $x_{t+1} \in L$. Construct $\{u_t\}$ as follows. If $\|x_t\|_\infty > M$, then choose u_t so that $u_t - \sum_{j=1}^{n} a_j \xi_{jt}$ is an integer but $|u_t| \leqq \frac{1}{2}$. If $\|x_t\|_\infty \leqq M$, then set $u_t - \sum_{j=1}^{n} a_j \xi_{jt} = 0$. We will still have

$$|u_t| = \left| \sum_{j=1}^{n} a_j \xi_{jt} \right| \leqq M \sum_{j=1}^{n} |a_j|.$$

We cannot have $\|x_t\|_\infty > M$ for all t, since for each t we would have $x_t = A^t x + z_t$, where $z_t = \sum_{k=1}^{t} A^{k-1} b u_k$, $|u_k| \leqq \frac{1}{2}$. Therefore, since $2z_t \in \mathscr{A}(t)$,

$$\|x_t\|_\infty \leqq \|A^t x\|_\infty + M/2.$$

But $A^t x \to 0$. Thus there exists

$$\tau = \min\{t : \|x_t\|_\infty < M\} < \infty.$$

By construction, the ψ_1 component of $x_{\tau+1}$ will be zero, the ψ_1 and ψ_2 components of $x_{\tau+2}$ will be zero, etc. (see (45)). In fact $x_t = 0$ for $t \geqq \tau + n$. Q.E.D.

We may now construct Q and $\hat{x}$. Let

$$U = \max\left\{1, M \sum_{j=1}^{n} |a_j|\right\}. \qquad (46)$$

Define L_0 to be the set of all lattice points in L which are reachable via a trajectory of (1) which begins at the origin, remains always in L, and satisfies $|u_t| \leqq U$ for all t. The set L_0 is finite since all permissible trajectories are bounded by UM. Let m be the number of points in L_0 and define $\hat{x}$ so that

$$\{\hat{x}(1), \cdots, \hat{x}(m)\} = L_0. \tag{47}$$

For a given $\phi \in \Phi = \{1, 2, \cdots, m\}$ consider the possible solutions (ϕ', u) to the relations

$$\hat{x}(\phi') = A\hat{x}(\varphi) + bu \in L_0, \qquad |u| \leqq U. \tag{48}$$

Let J be the greatest integer in U. By the definition (46), $J \geqq 1$. By the construction of L_0 there are at least J solutions to (48) with $u < 0$ and at least J solutions with $u > 0$. Define $v(\phi, \phi') = u$ whenever (48) holds, arbitrary otherwise. Let Q be any transition matrix satisfying $Q(\phi, \phi') > 0$ if and only if a solution (ϕ', u) to (48) exists, and satisfying (31). This is possible because for each there are at least $2J$ solutions to (48) (and thus at least $2J$ positive elements in the ϕth row of Q). Furthermore, since the numbers $v(\phi, \phi')$ associated with the solutions contain at least J positive and J negative values, the nonzero elements of the ϕth row of Q can be chosen so that (31) holds. Because $Q(\phi, \phi') > 0$ implies that (48) holds with $u = v(\phi, \phi')$, equation (30) holds. Finally (32) must hold trivially since the vector b must be in L_0, and also therefore one of the $\hat{x}(\phi)$.

Therefore we have constructed a Markov chain with transition matrix Q, a state estimator $\hat{X}$, and a matrix v satisfying all the conditions of the problem except possibly ergodicity.

Is Q ergodic? The origin is an element of L_0, and we may take $\hat{x}(1) = 0$. By construction, $Q(1, 1) > 0$ since $(\phi', u) = (1, 0)$ is a solution to (48) with $\phi = 1$. Furthermore, by the definition of L_0, and the fact that $Q(\phi, \phi')$ is positive whenever a solution to (48) exists, there must be a path of positive probability from the origin ($\phi = 1$) to any other state ϕ'. But by the Lemma, there is also a path of positive probability from ϕ' back to 1. Let T be sufficiently large so that for any ϕ' there are paths of length $\geqq T$ from ϕ' to 1 and 1 to ϕ' of positive probability. Then for any pair (ϕ, ϕ') there is a path of positive probability of length $2T$ which begins at ϕ, proceeds to 1, remains there for a time and then proceeds to ϕ' at time $2T$. Since every element of Q^{2T} is positive, Q is ergodic. We have proved the following proposition.

PROPOSITION 4. *There exists a solution to the Markov chain approximation problem.*

Example. The construction of $(Q, \hat{x})$ is no doubt grossly inefficient. There may be much smaller (in the size of Φ) solutions. As an example, consider first order linear systems ($n = 1$) with $0 < A < 1$. Let $b = 1$. We have $L = \mathbb{Z}$, and $\mathcal{A}(\infty)$ is the open interval $(-M, M)$ with $M = 1/(1 - A)$. Then U is given by

$$U = \begin{cases} 1, & \text{if } A < \frac{1}{2}, \\ A/(1-A) & \text{if } A \geqq \frac{1}{2}. \end{cases}$$

Some analysis shows that

$$L_0 = \{k \in \mathbb{Z}, -\ell \leqq k \leqq \ell\}$$

where ℓ is the integer satisfying

$$U^2 - 1 < \ell \leqq U^2.$$

For the case $A<\frac{1}{2}$, the construction yields the following chain (which is also a solution for $\frac{1}{2}\leqq A<1$, even though the construction will produce larger chains for this range of values of A).

$$\hat{X}=[-1,0,1] \quad \text{(the points in } L_0\text{)},$$

$$[v_{ij}]=\begin{bmatrix} A-1 & A & ? \\ -1 & 0 & 1 \\ ? & -A & 1-A \end{bmatrix} \quad \text{(to satisfy (48))},$$

$$Q=\begin{bmatrix} A & 1-A & 0 \\ q & 1-2q & q \\ 0 & 1-A & A \end{bmatrix}, \qquad 0<q<\tfrac{1}{2}.$$

This chain is ergodic with stationary measure

$$p=[\pi, 1-2\pi, \pi], \qquad \pi=q/(1-A+2q).$$

One may readily check to see that (30), (31), and (32) are satisfied.

6. Cross-correlation and power spectrum equivalence. In § 5 we constructed a Markov chain $\{\phi_t\}$ with a vector-valued output $\{\hat{x}(\phi_t)\}$ and a scalar-valued output $\hat{u}_t=v(\phi_t,\phi_{t+1})$ so as to satisfy (27)–(29). In order to construct an honest finite state system (2) we must reverse the cause and effect relationship between $\{\phi_t\}$ and $\{\tilde{u}_t\}$, and relate $\tilde{u}_t$ to the actual input u_t. In other words we must construct functions f and w so that if

$$\phi_{t+1}=f(\phi_t,u_t), \tag{49}$$

$$\tilde{u}_t=w(\phi_t,u_t) \tag{50}$$

then (27)–(29) hold. Furthermore we must have

$$Pr\{\phi_{t+1}|\phi_t\}=Q(\phi_t,\phi_{t+1}) \tag{51}$$

where $(Q,\hat{x})$ is the solution to the special approximation problem. We will assume that the distribution function F is continuous.

Define the functions $f\colon \Phi\times\mathbb{R}\to\Phi$ and $w\colon\Phi\times\mathbb{R}\to\mathbb{R}$ as follows. Let v be the matrix which appears in (30). For a given $\phi\in\Phi$, let $(\theta_1,\cdots,\theta_m)$ be a permutation of Φ for which

$$v(\phi,\theta_1)\leqq v(\phi,\theta_2)\leqq\cdots\leqq v(\phi,\theta_m). \tag{52}$$

Since F is continuous, we may choose numbers $\{\alpha_0,\cdots,\alpha_{m+1}\}$ satisfying

$$-\infty=\alpha_0\leqq\alpha_1\leqq\cdots\leqq\alpha_m\leqq\alpha_{m+1}=\infty$$

$$F(\alpha_k)-F(\alpha_{k-1})=Q(\phi,\theta_k), \qquad 1\leqq k\leqq m.$$

Then f and w are such that

$$\alpha_{k-1}<u\leqq\alpha_k\Rightarrow\begin{cases} f(\phi,u)=\theta_k,\\ w(\phi,u)=v(\phi,\theta_k).\end{cases} \tag{53}$$

By construction, the Markov chain defined by (49) will satisfy (51), and if $\tilde{u}_t$ is given by (50), then (27)–(29) will hold. Furthermore

$$E[\tilde{u}_t u_t]>0. \tag{54}$$

This positive correlation is due to the agreement in the ordering of the $\{\alpha_k\}$ with the order in (52). In other words $u' > u \Rightarrow w(\phi, u') \geqq w(\phi, u)$ with strict inequality on a set of positive probability.

PROPOSITION 5. *If* (A, b) *is stable and controllable, and* F *is continuous, then there exists a finite state system which is internally cross-correlation equivalent to the linear system* (1).

Proof. Let $(Q, \hat{x})$ be a solution to the special approximation problem. Construct the maps f and w to satisfy (53) for each ϕ. Let $\tilde{u}_t$ satisfy (50) where $\{\phi_t\}$ is generated by the finite state machine (49). Let $\eta = E[u_t\tilde{u}_t] > 0$. Define a "scaled" estimator

$$\hat{x}'(\phi) = \eta^{-1}\hat{x}(\phi).$$

Since (27) and (28) hold, we have

$$E[\hat{x}'(\phi_{t+1})|\phi_t] = A\hat{x}'(\phi_t)$$

and

$$\begin{aligned} E[u_t\hat{x}'(\phi_{t+1})] &= \eta^{-1}E[u_t\hat{x}(\phi_{t+1})] \\ &= \frac{E[u_t(A\hat{x}(\phi_t) + b\tilde{u}_t)]}{E[u_t\tilde{u}_t]} \\ &= b. \end{aligned}$$

Therefore, for the internal approximation $(f, \hat{X}')$, the weak equivalences (E4) and (E5) hold and by Proposition 3, we have internal cross-correlation equivalence. Q.E.D.

We could modify this construction to obtain internal power spectrum equivalence, but it is not really necessary since in this case the input is of no relevance. As we mentioned below equation (43), we need only construct a special approximation for which

$$\sigma^2 = E[\tilde{u}_t^2] = 1.$$

This is easily done by the obvious scaling $\hat{x}'(\phi) = \hat{x}(\phi)/\sigma^2$ of a given special approximation. Therefore we have the following.

PROPOSITION 6. *If* (A, b) *is stable and controllable, then there exists a finite state system which is internally power spectrum equivalent to the linear system* (1).

We have constructively demonstrated the existence of *internal* cross-correlation equivalent finite state systems and *internal* power spectrum equivalent finite state systems when (A, b) is stable and controllable and F is continuous. The existence of cross-correlation or power spectrum equivalent finite state systems follows by simply defining $\hat{y}_t = g(\phi_t) = c\hat{x}(\phi_t)$.

Conclusions. The motivation for considering finite state machines (with real-valued inputs and outputs) that are equivalent (in some sense) to a given linear system arises from the widespread use of digital systems for the simulation of linear systems. Fidelity in this context is usually phrased in terms of quantities such as the output mean square error $E(y_t - \hat{y}_t)^2$. This error variance can be zero only for certain trivial cases. A different problem is formulated here. Namely, what kinds of equivalence relations are there for which it is possible that a linear system and a finite state machine can be equivalent?

Two nontrivial equivalence relations have been offered: cross-correlation, and power spectrum equivalence. These relations involve two systems and a white noise input process. It was shown that for any stable linear system, if the distribution function F is continuous, then there are finite state machines which are equivalent to the linear system in either sense. (Not simultaneously, however, since Proposition 1 showed that this would imply external equivalence, and Proposition 2 showed that this could happen only in trivial cases.)

As a corollary, it follows that any power spectrum obtained by shaping white noise with a finite order linear filter can also be obtained with a finite Markov chain. This has been an unproved conjecture.

REFERENCES

[1] R. KALMAN, M. ARBIB AND P. FALB, *Topics in Mathematical Systems Theory*, McGraw-Hill, New York, 1969.
[2] J. KOPLOWITZ, *Estimation and Approximation with a Finite Statistic*. Ph.D. dissertation, Department of Electrical Engineering, University of Colorado, Boulder, CO, 1973.
[3] B. LIU, *Effect of finite word length on the accuracy of digital filters—A review*, IEEE Trans. Circuits and Systems, CT 18 (1971), pp. 670–677.
[4] J. MAX, *Quantization for minimum distortion*, IRE Trans. Information Theorem, IT 4 (1960), pp. 7–12.
[5] C. T. MULLIS AND K. STEIGLITZ, *Circulant Markov chains as digital signal sources*, IEEE Trans. Audio and Electroacoustics, AU 20 (1972), pp. 246–249.
[6] C. T. MULLIS AND R. A. ROBERTS, *Synthesis of minimum roundoff noise fixed point digital filters*, IEEE Trans. Circuits and Systems, CAS-23 (1976), pp. 551–562.
[7] R. SITTLER, *System analysis of discrete Markov processes*, IRE Trans. Circuit Th., CT 13 (1956), pp. 257–266.
[8] N. WIENER, *Nonlinear Problems in Random Theory*, M.I.T. Press, Cambridge, 1958.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0008 $01.00/0

CONVERGENCE OF 'PERIODIC IN THE LIMIT' OPERATOR CONTINUED FRACTIONS*

ROBERT C. BUSBY† AND WYMAN FAIR†

Abstract. By straightforward analysis, we prove a convergence theorem regarding continued fractions whose entries are members of a Banach algebra. If the entries converge fast enough to a certain fixed element, convergence is assured. Under suitable other restrictions, a stronger theorem results.

1. Introduction. It was shown by Fair [1] that if A is an element of a Banach algebra $\mathcal{A}$ (by which we will always mean a Banach algebra with norm one identity I), and the spectrum of A does not intersect $(-\infty, -\frac{1}{4}]$; then the continued fraction (c.f.)

$$\frac{I}{I+}\frac{A}{I+}\frac{A}{I+\cdots} \tag{1.1}$$

converges to an element K of $\mathcal{A}$ (see (1.4) below) where AK is a root of the polynomial X^2+X-A. In this paper we investigate the c.f.

$$\frac{I}{I+}\frac{A_1}{I+}\frac{A_2}{I+\cdots} \tag{1.2}$$

in what Khovanskii [2] calls the 'periodic in the limit' case, i.e. when the A_n converge suitably to an element A of $\mathcal{A}$. The c.f. (1.2) can be regarded as a 'perturbed' version of (1.1). We will usually write $A_n = A + E_n$ and look for 'smallness' conditions on the perturbations E_n which guarantee convergence.

In § 2 we will prove our most general result. We say that the c.f. (1.2) is 'tail-end convergent' if there is an $n \geqq 1$ such that the c.f.

$$\frac{I}{I+}\frac{A_n}{I+}\frac{A_{n+1}}{I+\cdots} \tag{1.3}$$

converges. We will prove that if the E_n form a series which converges in norm in $\mathcal{A}$ with sufficient rapidity, then (1.2) will be tail-end convergent. This result requires a hypothesis on A which in general is slightly stronger than that required of A in (1.1).

In § 3 we consider the case where the A_n (equivalently $\mathcal{A}$) are commutative. We do not, however, require any other conditions (such as normality or positivity) on the A_n. We show here that if $\lim_{n\to\infty} A_n = A$ then (1.2) is tail-end convergent. In fact we actually show that if the E_n are uniformly 'small' and converge to zero in norm, then (1.2) itself converges, from which the result previously stated is immediate.

This result is proven by modifying the argument of Khovanskii [2, pp. 62–65] in which he proves a similar result for the case where the A_n are complex numbers. Actually Khovanskii proves (by a method essentially due to Pringsheim [4] that when $\lim_{n\to\infty} A_n = A$, (1.2) will be tail-end convergent, except for particular A which occur as unspecified zeros of certain polynomials. We refine his argument very slightly to remove this restriction; then we rearrange and modify it so that is may be applied to elements of a commutative Banach algebra. At various steps we provide lemmas and detours needed for these 'operator modifications', but we will refer directly to Khovanskii whenever possible to avoid needless duplication of his arguments. We conclude this section with some notation and preliminary results.

* Received by the editors August 18, 1977.
† Department of Mathematics, Drexel University, Philadelphia, Pennsylvania 19104.

Let A belong to a Banach algebra $\mathscr{A}$, and let $\sigma(A)$ be the spectrum of A in $\mathscr{A}$. If $f(z)$ is an analytic function in some domain D containing $\sigma(A)$, we define $f(A)$ in the well-known fashion using a 'Cauchy integral formula' (see Naimark [3, ch. III]); then $f \to f(A)$ is an algebra homomorphism (with some care taken with domains) and $\sigma(f(A)) = f(\sigma(A))$. If $B \in \mathscr{A}$ and $\sigma(B) \cap (-\infty, 0] = \emptyset$, then we will always let $\sqrt{B}$ be $f(B)$, where f is the unique branch of $\sqrt{z}$ defined in the cut plane $\mathbb{C} \dashv (-\infty, 0]$ and positive on positive real numbers.

Consider the c.f. (1.1) and define elements P_n and Q_n ($n \geqq 0$) in $\mathscr{A}$ inductively by the formulas:

$$
\begin{aligned}
&P_0 = P_1 = Q_0 = I, \qquad Q_1 = I + A, \\
&P_{n+1} = P_n + AP_{n-1} \qquad (n \geqq 1), \\
&Q_{n+1} = Q_n + AQ_{n-1} \qquad (n \geqq 1).
\end{aligned} \tag{1.4}
$$

It can easily be shown by induction that if the Q_n are invertible, then

$$Q_n^{-1} P_b = \frac{I}{I+} \frac{A}{I+} \cdots \frac{A}{+I}$$

the nth 'convergent' of the c.f.: thus:

(1.5) We say (1.1) converges if all Q_n are invertible and the sequence $Q_n^{-1} P_n$ converges in $\mathscr{A}$. If so, the corresponding limit will be called the value of (1.1).

Similarly consider (1.2) and define R_n and $S_n (n \geqq 0)$ inductively by:

$$
\begin{aligned}
&R_0 = R_1 = S_0 = I, \qquad S_1 = I + A_1 \\
&R_{n+1} = R_n + A_{n+1} R_{n-1} \qquad (n \geqq 1), \\
&S_{n+1} = S_n + A_{n+1} S_{n-1} \qquad (n \geqq 1).
\end{aligned} \tag{1.6}
$$

Again if all S_n are invertible then

$$S_n^{-1} R_n (n \geqq 1) = \frac{I}{I+} \frac{A_1}{I+} \cdots \frac{A_n}{+I}.$$

(1.7) We say (1.2) converges if all S_n are invertible and the sequence $S_n^{-1} R_n$ converges in $\mathscr{A}$.

Finally if $\mathscr{A}$ is a Banach algebra, we let $\mathscr{A}_2$ be the Banach space $\mathscr{A} \times \mathscr{A}$ with norm given by:

$$\left\| \begin{bmatrix} U \\ V \end{bmatrix} \right\| = \max (\|U\|, \|V\|) \tag{1.8}$$

and let $\mathscr{A}_{2,2}$ be the Banach algebra of 2×2 matrices with entries from $\mathscr{A}$ and ordinary matrix multiplications (using the product in $\mathscr{A}$), and where $[A_{ij}] \in \mathscr{A}_{2,2}$ is given the operator norm relative to the obvious action of $\mathscr{A}_{2,2}$ on $\mathscr{A}_2$, i.e.

$$\|[A_{ij}]\| = \sup_{\|U\|, \|V\| \leqq 1} \{\|A_{11} U + A_{12} V\|, \|A_{21} U + A_{22} V\|\}. \tag{1.9}$$

Since we may choose U or V to be independently equal to I or 0, we see that:

$$\|[A_{ij}]\| \geqq \max_{1 \leqq i,j \leqq 2} \{\|A_{ij}\|\}. \tag{1.10}$$

2. The main result. We begin with some general lemmas, in all of which $\mathscr{A}$ is a fixed Banach algebra over the complex numbers.

LEMMA 1. *Let* $\alpha=[A_{ij}]$ *and* $\beta=[B_{ij}]$ *be in* $\mathscr{A}_{2,2}$. *Then* $\|A_{ij}-B_{ij}\|\leqq\|\alpha-\beta\|$, $i,j=1,2$.

Proof. For $i=1,2$, let

$$\varepsilon_i=\begin{bmatrix}\delta_{1i}I & 0\\ 0 & \delta_{2i}I\end{bmatrix}$$

in $\mathscr{A}_{2,2}$ where δ_{ij} is the usual Kronecker delta. Then by (1.9) we easily see that $\|\varepsilon_i\|=1$, $i=1,2$. If $1\leqq i,j\leqq 2$,

$$\begin{aligned}\|A_{ij}-B_{ij}\|&\leqq\left\|(A_{ij}-B_{ij})\begin{bmatrix}\delta_{1i}\delta_{1j} & \delta_{1i}\delta_{2j}\\ \delta_{2i}\delta_{1j} & \delta_{2i}\delta_{2j}\end{bmatrix}\right\|\\ &=\|\varepsilon_i(\alpha-\beta)\varepsilon_j\| \qquad \text{(by (1.10))}\\ &\leqq\|\alpha-\beta\|.\end{aligned}$$

LEMMA 2. *Let* $(B_n)_{1\leqq n<\infty}$ *be a sequence in* $\mathscr{A}$ *and let* $\varepsilon>0$ *be given. Then*:

(a) *If* $\sum_{n=1}^{\infty}\|B_n\|\equiv S<\infty$, *the infinite product* $\prod_{n=1}^{\infty}(I+B_n)$ *converges to an element P in* $\mathscr{A}$, *and*

(b) $S<\ln(1+\varepsilon)\Rightarrow\|I-P\|<\varepsilon$.

Proof. Define B_0 to be 0, and for each integer $n\geqq 0$ let $P_n=\prod_{s=0}^{n}(I+B_s)$. Then if $n>m\geqq 0$,

$$P_n-P_m=B_{m+1}+\cdots+B_n+\Big(\sum_{m+1\leqq i,j\leqq n}B_iB_j\Big)+\cdots+B_{m+1}\cdots B_n,$$

and so

$$\|P_n-P_m\|\leqq\|B_{m+1}\|+\cdots+\|B_n\|+\Big(\sum_{m+1\leqq i,j\leqq n}\|B_i\|\|B_j\|\Big)+\cdots+\|B_{m+1}\|\cdots\|B_n\|$$

$$=\Big(\prod_{k=m+1}^{n}(1+\|B_k\|)\Big)-1\leqq\exp\Big(\sum_{k=m+1}^{n}\|B_k\|\Big)-1$$

(since $1+a\leqq e^a$ when $a\geqq 0$). Since $\sum_{n=1}^{\infty}\|B_n\|<\infty$, the above computation shows that $\{P_n\}_{0\leqq n<\infty}$ is Cauchy in $\mathscr{A}$ and so converges in $\mathscr{A}$ to an element P. This proves (a).

If we let $m=0$ and allow n to pass to infinity in the above inequality, we have $\|P-I\|\leqq\exp(\sum_{k=1}^{\infty}\|B_k\|)-1<\varepsilon$ and this proves (b).

LEMMA 3. *If A is an invertible element of* $\mathscr{A}$ *and for some* $B\in\mathscr{A}$, $\|B-A\|<\|A^{-1}\|^{-1}$, *then B has an inverse* (*in fact* $B^{-1}=A^{-1}(\sum_{n=0}^{\infty}[(A-B)A^{-1}]^n)$).

Proof. See Rickart [5, Thm. 1.4.6].

We now refer to the notation and remarks of § 1 concerning the c.f. (1.2). We suppose that we have $A\in\mathscr{A}$ such that $\sigma(A)\cap(-\infty,-\frac{1}{4}]=\varnothing$ and A invertible. We write $A_n=A+E_n$, $n\geqq 1$. Let $C=\begin{bmatrix}I & A\\ I & 0\end{bmatrix}$ in $\mathscr{A}_{2,2}$, and let P_n, Q_n, R_n, S_n be as in (1.4) and (1.6). Then from (1.4) it can easily be seen that:

$$\begin{bmatrix}Q_n\\ Q_{n-1}\end{bmatrix}=C^{n+1}\begin{bmatrix}I\\ 0\end{bmatrix}\quad\text{in }\mathscr{A}_2,\qquad(n\geqq 1), \tag{2.1}$$

and

$$C^{n+1} = \begin{bmatrix} Q_n & P_n A \\ P_n & P_{n-1} A \end{bmatrix}. \tag{2.2}$$

Then the discussion following (1.1), and (1.5) imply that:

$$Q_n^{-1} C^{n+1} = \begin{bmatrix} I & Q_n^{-1} P_n A \\ Q_n^{-1} P_n & (Q_n^{-1} P_n)(Q_{n-1}^{-1} P_{n-1}) A \end{bmatrix} \tag{2.3}$$

and

$$\lim_{n \to \infty} Q_n^{-1} C^{n+1} = \begin{bmatrix} I & KA \\ K & K^2 A \end{bmatrix}, \tag{2.4}$$

where KA is a root of $X^2 + X - A$. Similarly by using (1.6) we can see that:

$$\begin{bmatrix} S_n \\ S_{n+1} \end{bmatrix} = C \begin{bmatrix} S_{n-1} \\ S_{n-2} \end{bmatrix} + F_n \begin{bmatrix} S_{n-1} \\ S_{n-2} \end{bmatrix} \qquad (n \geqq 2), \tag{2.5}$$

where

$$F_n = \begin{bmatrix} 0 & E_n \\ 0 & 0 \end{bmatrix} \qquad (n \geqq 1) \text{ in } \mathscr{A}_{2,2}. \tag{2.6}$$

By induction,

$$\begin{bmatrix} S_n \\ S_{n-1} \end{bmatrix} = \left[\prod_{k=0}^{n} (C + F_k) \right] \begin{bmatrix} I \\ 0 \end{bmatrix},$$

where we let F_0 be the zero matrix. Since C is invertible (in fact $C^{-1} = A^{-1} \begin{bmatrix} 0 & -A \\ -I & I \end{bmatrix}$) we can show that

$$\prod_{k=0}^{n} (C + F_k) = C^{n+1} \prod_{k=0}^{n-1} (I + C^{-n+k-1} F_{n-k} C^{n-k}).$$

Thus

$$\begin{bmatrix} S_n \\ S_{n-1} \end{bmatrix} = C^{n+1} Z_n \begin{bmatrix} I \\ 0 \end{bmatrix} \qquad (n \geqq 1) \text{ in } \mathscr{A}_2 \tag{2.7}$$

where

$$Z_n = \prod_{k=0}^{n-1} (I + C^{-n+k-1} F_{n-k} C^{n-k}) = \prod_{k=1}^{n} (I + C^{-k-1} F_k C^k). \tag{2.8}$$

In a completely analogous way we can show the following:

$$\begin{bmatrix} R_n \\ R_{n-1} \end{bmatrix} = C^n Y_n \begin{bmatrix} I \\ 0 \end{bmatrix} \qquad (n \geqq 2), \tag{2.9}$$

where

$$Y_n = \prod_{k=0}^{n-2} (I + C^{-n+k} F_{n-k} C^{n-k-1}) = \prod_{k=2}^{n} (I + C^{-k} F_k C^{k-1}). \tag{2.10}$$

Now, keeping all of the above notation, we can state the main result of this section.

THEOREM 1. *Let $A_n = A + E_n$ in $\mathcal{A}$, with A invertible and $\sigma(A) \cap (-\infty, -\frac{1}{4}] = \varnothing$. Suppose $\sum_{n=1}^{\infty} \|E_n\|\gamma^n < \infty$ where*

$$\gamma_n = \|C\|\|C^{-1}\|.$$

Then the c.f. (1.2) *is tail-end convergent.*

Proof. Suppose first that we can find a real number ε such that:

1) $0 < \varepsilon < \frac{1}{2}$;

2) $\varepsilon/(1-\varepsilon) < \inf_n \|Q_n^{-1}P_nA\|^{-1}$;

3) $\sum_{n=1}^{\infty} \|E_n\|\gamma^n < \ln(1+\varepsilon)(\|C^{-1}\|^{-1})$.

Note that there are numbers ε satisfying 1) and 2) since $Q_n^{-1}P_nA$ is invertible for all n and converges to the invertible operator KA.

We must show that with these assumptions, each S_n is invertible and the sequence $(S_n^{-1}R_n)1 \leqq n < \infty$ converges in $\mathcal{A}$. By (2.7)

$$\begin{bmatrix} S_n \\ S_{n-1} \end{bmatrix} = C^{n+1}Z_n\begin{bmatrix} I \\ 0 \end{bmatrix}$$

with Z_n a partial product of $\prod_{k=1}^{\infty} (I + C^{-k-1}F_kC^k)$. By (2.6) and (1.9), $\|F_k\| = \|E_k\|$ for all k. Then

$$\sum_{k=1}^{\infty} \|C^{-k-1}F_kC^k\| \leqq \sum_{k=1}^{\infty} \|C^{-1}\|\|E_k\|\gamma^k < \ln(1+\varepsilon)$$

and so Lemma 2 tells us that Z_n converges in $\mathcal{A}_{2,2}$ to an element $Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$ and $\|I - Z\| < \varepsilon$. It is also clear that $\|I - Z_n\| < \varepsilon$ for all n. By (2.3) and (2.7), $Q_n^{-1}S_n = Z_n^{11} + Q_n^{-1}P_nAZ_n^{21}$, which by (2.4) converges to $Z_{11} + KAZ_{21}$. Now by Lemma 1, we have:

$$(2.11)\qquad \begin{aligned} &\text{(a)}\quad \|Z_n^{11} - I\| < \varepsilon \text{ for all } n, \|Z_{11} - I\| < \varepsilon, \\ &\text{(b)}\quad \|Z_n^{21}\| < \varepsilon \text{ for all } n, \|Z_{21}\| < \varepsilon. \end{aligned}$$

Since $\varepsilon < 1$, (2.11) and Lemma 3 tell us that Z_n^{11} is invertible with inverse $I + (I - Z_n^{11}) + \cdots$, thus $\|(Z_n^{11})^{-1}\| \leqq \sum_{k=0}^{\infty} \|I - Z_n^{11}\|^k = 1/(1-\varepsilon)$. Hence

$$\|Z_n^{11} - Q_n^{-1}S_n\| = \|Q_n^{-1}P_nAZ_n^{21}\| < \|Q_n^{-1}P_nA\|\varepsilon < 1 - \varepsilon$$

(by hypothesis 2) $\leqq \|(Z_n^{11})^{-1}\|^{-1}$, and so by Lemma 3, $Q_n^{-1}S_n$ (hence S_n) is invertible for each n. The hypotheses also imply that $\varepsilon/(1-\varepsilon) < \|KA\|^{-1}$ and so by repeating the above arguments, we see that $\lim_{n\to\infty} Q_n^{-1}S_n) \equiv Z_{11} + KAZ_{21}$ is also invertible. Thus $\lim_{n\to\infty} S_n^{-1}Q_n$ exists.

One can show as with (2.3) and (2.4) that $Q_n^{-1}C^n$ converges to $\begin{bmatrix} K & K^2A \\ K^2 & K^3A \end{bmatrix}$ as n goes to infinity. Thus by (2.9), $Q_n^{-1}R_n$ will converge if Y_n does; since $\sum_{k=2}^{\infty} \|C^{-k}F_kC^{k-1}\| \leqq \|C^{-1}\| \sum_{k=2}^{\infty} \|E_k\|\gamma^k < \infty$, (2.10) and Lemma 2(a) yield the convergence of Y_n. Thus $S_n^{-1}R_n = (S_n^{-1}Q_n)(Q_n^{-1}R_n)$ exists for all n and converges in $\mathcal{A}$.

Now suppose the hypotheses of the theorem hold. Since $\gamma \geqq 1$, the sequence γ^n is monotone increasing, and $\sum_{n=1}^{\infty} \|E_n\| < \infty$. Choose n so large that

$$\sum_{n=N}^{\infty} \|E_n\|\gamma^n < \ln(1+\varepsilon)(\|C^{-1}\|^{-1})$$

where ε is any number satisfying conditions 1) and 2) stated at the beginning of the proof. Then

$$\sum_{n=1}^{\infty} \|E_{N-1+n}\|\gamma^n \leqq \sum_{n=1}^{\infty} \|E_{N-1+n}\|\gamma^{N-1+n}$$

$$= \sum_{n=N}^{\infty} \|E_n\|\gamma^n < \ln(1+\varepsilon)(\|C^{-1}\|^{-1}),$$

and so by the first part of this proof, the c.f. (1.3) converges. By definition this means that (1.2) is tail-end convergent, and we are done.

It may be that γ will be difficult to find for certain algebras $\mathscr{A}$. When $\mathscr{A}$ is an algebra of operators on a Hilbert space H, we can estimate γ directly in terms of A as follows:

PROPOSITION 1. *Suppose $A_i = A + E_i$ for all i, where A and the E_i are bounded linear operators on a Hilbert space H.*

Then (1.2) will be tail-end convergent provided $\sum_{n=1}^{\infty} \|E_n\|\mu^n < \infty$ where

$$\mu^2 = 4\left[\frac{1+\frac{1}{2}\|A\|^2 + \frac{1}{2}\sqrt{4+\|A\|^4}}{1+\frac{1}{2}\|A^{-1}\|^{-2} - \frac{1}{2}\sqrt{4+\|A^{-1}\|^{-4}}}\right].$$

In particular such convergence holds if $\overline{\lim}_{n\to\infty} \sqrt[n]{\|E_n\|} < \mu^{-1}$.

Proof. We have regarded $\mathscr{A}_{2,2}$ as an algebra of operators on $\mathscr{A}_2$ and used the corresponding norm, $\|\cdot\|$ (see (1.9)). We may also, in this case, regard $\mathscr{A}_{2,2}$ as an algebra of operators on $H \oplus H$. If we denote the corresponding operator norm by $|||\cdot|||$ then we have,

$$|||[A_{ij}]_{1\leqq i,j\leqq 2}||| = \sup_{\|\xi\|^2+\|\eta\|^2\leqq 1} \sqrt{\|A_{11}\xi + A_{12}\eta\|^2 + \|A_{21}\xi + A_{22}\eta\|^2} \tag{2.12}$$

where ξ and η are vectors in H.

Now let ξ, η, α, and β range over the unit ball in H, E and F range independently over the operators on H of norm less than or equal to one on H. Comparing (2.12) with (1.9), we see that:

$$\begin{aligned}
\|[A_{ij}]\| &= \sup_{E,F,\xi,\eta} \{\|(A_{11}E + A_{12}F)\xi\|, \|(A_{21}E + A_{22}F)\eta\|\} \\
&\leqq \sup_{\xi,\eta,\alpha,\beta} \{\|A_{11}\xi + A_{12}\eta\|, \|A_{21}\alpha + A_{22}\beta\|\} \\
&\leqq \sup_{\xi,\eta} \sqrt{\|A_{11}\xi + A_{12}\eta\|^2 + \|A_{21}\xi + A_{22}\eta\|^2} \\
&\leqq \sup_{(\xi/\sqrt{2})^2+(\eta/\sqrt{2})^2\leqq 1} \sqrt{2}\sqrt{\left\|A_{11}\left(\frac{\xi}{\sqrt{2}}\right) + A_{12}\left(\frac{\eta}{\sqrt{2}}\right)\right\|^2 + \left\|A_{21}\left(\frac{\xi}{\sqrt{2}}\right) + A_{22}\left(\frac{\eta}{\sqrt{2}}\right)\right\|^2} \\
&= \sqrt{2}|||[A_{ij}]|||.
\end{aligned}$$

We may now use standard Hilbert space results to compute norms, $|||\cdot|||$. Note that CC^* (C^* is the adjoint operator to C on $H \oplus H$) $= \begin{bmatrix} (I+AA^*) & I \\ I & I \end{bmatrix}$ is a positive self adjoint operator on $H \oplus H$, and so $(|||C|||)(|||C^{-1}|||)^2 = |||CC^*|||\,|||(C^{-1})^*C^{-1}||| = |||CC^*|||\,|||(CC^*)^{-1}||| = \mu_0/\lambda_0$, where μ_0 and λ_0 are respectively the maximum and minimum (positive real) numbers in the spectrum of CC^*. Since the operator entries of CC^* commute, it is routine to show that

$$\lambda \in \sigma(CC^*) \leftrightarrow \det\left\{\lambda\begin{bmatrix} I & 0 \\ 0 & I\end{bmatrix} - CC^*\right\} \equiv (\lambda^2-2)I - (\lambda-1)AA^*$$

is a singular operator on H.

Clearly $\lambda = 1$ does not satisfy this conditon so we see that $\lambda \in \sigma(CC^*) \leftrightarrow \lambda(\lambda-2)/(\lambda-1) \in \sigma(AA^*)$.

Now the maximum and minimum numbers in the spectrum of the positive self adjoint operator AA^* are $\|AA^*\| \equiv \|A\|^2$, and $\|(AA^*)^{-1}\|^{-1} = \|A^{-1}\|^{-2}$, respectively. Thus it is easy to see that μ_0 must be the largest root of the quadratic equation

$$\frac{\lambda(\lambda-2)}{(\lambda-1)} = \|A\|^2, \quad \text{namely} \quad 1 + \tfrac{1}{2}\|A\|^2 + \tfrac{1}{2}\sqrt{4 + \|A\|^4},$$

and λ_0 must be the smallest root of the equation $\lambda(\lambda-2)/(\lambda-1) = \|A^{-1}\|^{-2}$, namely $1 + \frac{1}{2}\|A^{-1}\|^{-2} - \frac{1}{2}\sqrt{4 + \|A^{-1}\|^{-4}}$. Then $\gamma = \|C\|\|C^{-1}\| \leqq 2|||C|||\,|||C^{-1}||| \equiv \mu$, and the result follows from Theorem 1 and the above computations.

Note that different specific estimates can be made for other types of Banach algebras $\mathscr{A}$. Also one could modify the norm definition for operators in $\mathscr{A}_{2,2}$, given in Theorem 1 (with some care) and achieve slightly modified results.

3. The commutative case. In all of this section, $\mathscr{A}$ will denote a fixed commutative Banach algebra with identity I. Let $A \in \mathscr{A}$. Then we say:

(3.1) $\qquad A$ is proper if $\sigma(A) \cap (-\infty, -\frac{1}{4}] = \varnothing$.

If A is proper in $\mathscr{A}$, we define elements $U(A)$ and $U'(A)$ in $\mathscr{A}$ by:

(3.2) $\quad U(A) = \frac{1}{2}I + \frac{1}{2}\sqrt{I + 4A}$; $U'(A) = I - U(A)$, where $\sqrt{(\cdot)}$ is the canonical branch of square root defined in the Introduction. It is clear that the spectrum of $U(A)$ does not contain zero when A is proper, and so $U(A)$ is invertible in $\mathscr{A}$.

(3.3) $\qquad$ Define $V(A)$, for A proper, to be $U'(A)U(A)^{-1}$.

By the spectral mapping theorem [5 Thm. 3.5.1] we see that the spectral radius $r(V(A))$ $(\sup\{\sigma(V(A))\})$ of the element $V(A)$ is less than one. In general, however, this will not extend to the norm. We will say:

(3.4) $\qquad A$ is strongly proper if A is proper and $\|V(A)\| < 1$.

Note that if $r(V(A)) = \|V(A)\|$ or if $\mathscr{A}$ is a commutative C^*-algebra (e.g. operators on a Hilbert space or matrices) and A is normal in $\mathscr{A}$, then proper implies strongly proper. In particular this is so if $\mathscr{A}$ is the complex numbers.

LEMMA 4. *Suppose A is strongly proper and invertible in $\mathscr{A}$. Then for every ε, $0 < \varepsilon < 1$, there is a $\delta \equiv \delta(\varepsilon) > 0$ such that if $(A_n)_{1 \leqq n \leqq \infty}$ is a sequence in $\mathscr{A}$ with*:

(α) $\lim_{n\to\infty} A_n = A$ *and*

(β) $\|A - A_n\| \leqq \delta$,

then there exists a sequence $(U_n)_{0 \leqq n < \infty}$ in $\mathscr{A}$ with:

(1) $U_n(I - U_{n+1}) = -A_{n+1}$;

(2) $\lim_{n\to\infty} U_n = U \equiv U(A)$, *and*

(3) $\|I - U_nU^{-1}\| < \varepsilon$; $n = 0, 1, 2, \cdots$.

Proof. This proof is simply a sight refinement and reorganization of computations given by Khovanskii in [2, pp. 62–65]. Let $M = \|V(A)\| < 1$, and $N = M^{1/3}$. Since

$0<N<1$, we can choose an integer $k \geqq 1$ such that $N^{k+1}(1-N)<\varepsilon$. Let δ be $N^k(1-N)^2\|A^{-1}\|^{-1}$, and suppose $(A_n)_{1\leqq n<\infty}$ is a sequence in $\mathscr{A}$ satisfying (α) and (β) above. First we note that if $U_n \in \mathscr{A}$ satisfies (3) above, then by Lemma 3, U_nU^{-1} (and so U_n) is invertible. Moreover if we let $W_n=(U_n+A_{n+1})U_n^{-1}$, we can show, precisely as in Khovanskii [2, pp. 62, 63] that

$$\begin{aligned} U-W_n &= (UU_n-U_n-A_{n+1})U_n^{-1} \\ &= U'[(I-U_nU^{-1})-(I-A_{n+1}A^{-1})][I-(I-U_nU^{-1})]^{-1}, \end{aligned}$$

i.e.,

$$(I-W_nU^{-1})=U'U^{-1}[(I-U_nU^{-1})-(I-A_{n+1}A^{-1})][I-(I-U_nU^{-1})]^{-1}. \tag{3.5}$$

Now $\|I-U_nU^{-1}\|<\varepsilon<1-N$, which implies that $\|[I-(I-U_nU^{-1})]^{-1}\| \leqq \sum_{k=0}^{\infty}\|I-U_nU^{-1}\|^k \leqq 1/N$, and so by computation (3.5), we have:

$$\begin{aligned} \|I-W_nU^{-1}\| &< \left\|\frac{V(A)}{N}\right\|(\|I-U_nU^{-1}\|+\|I-A_{n+1}A^{-1}\|) \\ &= N^2(\|I-U_nU^{-1}\|+\|I-A_{n+1}A^{-1}\|). \end{aligned} \tag{3.6}$$

Now by (α) and (β) above, there must exist a sequence $0=n_1<n_2<\cdots$ of integers such that if ν is an integer, $\nu \geqq n_\lambda$, then $\|A-A_{\nu+1}\|<N^{k+\lambda-1}(1-N)^2\|A^{-1}\|^{-1}$. Thus

$$\nu \geqq n_\lambda \Rightarrow \|I-A_{\nu+1}A^{-1}\|<N^{k+\lambda-1}(1-N)^2. \tag{3.7}$$

We choose a, $0<a<1$, such that $1-a<N^{k+1}(1-N)$ and define U_0 to be aU. Since U_0 satisfies (3) above, we may let $U_1=W_0\equiv(U_0+A_1)U_0^{-1}$. Then clearly $U_0(I-U_1)=-A_1$, and by (3.6) and (3.7) we see that $\|I-U_1U^{-1}\| \leqq N^2(N^{k+1}(1-N)+N^k(1-N)^2)=N^{k+2}(1-N)<\varepsilon$. Now if $2\leqq \mu \leqq n_2$ and we have defined U_s for $s<\mu$ satisfying (1) and (3) above, we may let $U_\mu=W_{\mu-1}$. Exactly as above we see that $U_{\mu-1}(I-U_\mu)=-A_\mu$, and that $\|I-U_\mu U^{-1}\|<N^{k+2}(1-N)<\varepsilon$. When we let $U_{n_2+1}=W_{n_2}$, we see by (3.6) and (3.7) that $\|I-U_{n_2+1}U^{-1}\|<N^2N^{k+2}(1-N)+N^2N^{k+1}(1-N)^2=N^{k+3}(1-N)$, and (1) continues to hold. In the same way we can define $U_{n_2+\mu}$ for $\mu=2,\cdots,(n_3-n_2)$, and show that $\|I-U_{n_2+\mu}U^{-1}\|<N^{k+3}(1-N)$. Continuing this whole process by induction, we finally define the sequence U_n for all n.

Since always, $U_{n+1}=W_n$, (1) above holds, and the continuation of the above estimates shows that:

$$\|I-U_{n_\lambda+\mu}U^{-1}\| \leqq N^{\lambda+k+1}(1-N) \quad \text{for } \mu=1,2,\cdots,(n_{\lambda+1}-n_\lambda), \quad \lambda=0,1,2,\cdots, \tag{3.8}$$

where in order to include the first term, we set $n_0=-1$.

From (3.8) it is clear that (2) and (3) hold, and we are done. The above computations are quite similar to those in Khovanskii, and in spirit originated in the paper by Pringsheim [4].

We will also need the following result:

LEMMA 5. *Let $B\in\mathscr{A}$ be an invertible element with $\|B\|<1$. Then there exists a $\delta(B)>0$ such that if $(B_n)_{1\leqq n<\infty}$ is a sequence in $\mathscr{A}$ with:*

(a) *$\lim_{n\to\infty}B_n=B$ and*

(b) *$\|B_n-B\|\leqq\delta(B)$ for all n, then*

(1) *for each n, $\sigma_n=I+B_1+B_1B_2+\cdots+B_1B_2\cdots B_n$ is invertible in $\mathscr{A}$ and*

(2) *σ_n converges in $\mathcal{A}$ to an invertible element.*

Proof. Since $\|B\|<1$, $2(1+\|B\|)^{-1}>1$, and so we may choose $\delta\equiv\delta(B)$ so small that $(1-(\delta+\|B\|))^{-1}\leqq 2(1-\|B\|)^{-1}(1+\|B\|)^{-1}$.

For each $n\geqq 1$, let $D_n=B_n-B$. Then

$$\begin{aligned}\left\|\sigma_n-\sum_{k=0}^n B^k\right\| &= \|D_1+[(D_1+D_2)B+D_1D_2]+\cdots+[(D_1+\cdots+D_n)B^{n-1}\\ &\qquad +(D_1D_2+D_1D_3+\cdots+D_{n-1}D_n)B^{n-2}+\cdots D_1D_2\cdots D_n]\|\\ &\leqq \delta+[2\delta\|B\|+\delta^2]+\cdots+\left[n\delta\|B\|^{n-1}+\binom{n}{2}\delta^2\|B\|^{n-2}+\cdots+\delta^n\right]\\ &= \sum_{k=0}^n(\delta+\|B\|)^k-\sum_{k=0}^n\|B\|^k.\end{aligned}$$

Thus

$$\left\|\sigma_n-\sum_{k=0}^n B^k\right\|<\left[\frac{1-(\delta+\|B\|)^{n+1}}{1-(\delta+\|B\|)}\right]-\left[\frac{1-\|B\|^{n+1}}{1-\|B\|}\right]. \tag{3.9}$$

Now since $\|B\|<1, I-B$ is invertible, and $I-B^{n+1}$ is invertible. Then since $\sum_{k=0}^n B^k=(I-B^{n+1})(I-B)^{-1}$, it follows that $\sum_{k=0}^n B^k$ is invertible and its inverse is $(I-B)(I-B^{n+1})^{-1}=(I-B)\sum_{k=0}^\infty B^{k(n+1)}$. Thus $\|(\sum_{k=0}^n B^k)^{-1}\|<\|I-B\|\sum_{k=0}^\infty\|B\|^{k(n+1)}$ $=\|I-B\|(1-\|B\|^{n+1})^{-1}$ and so

$$\left\|\left(\sum_{k=0}^n B^k\right)^{-1}\right\|^{-1}\geqq\frac{1-\|B\|^{n+1}}{\|I-B\|}\geqq\frac{1-\|B\|^{n+1}}{1+\|B\|}. \tag{3.10}$$

Now using (3.9), (3.10) and the definition of δ we see that

$$\begin{aligned}\left\|\sigma_n-\sum_{k=0}^n B^k\right\| &\leqq\frac{2(1-(\delta+\|B\|)^{n+1})}{(1-\|B\|)(1+\|B\|)}-\frac{(1-\|B\|^{n+1})}{(1-\|B\|)}\\ &\leqq\frac{1-\|B\|^{n+1}}{1-\|B\|}\left[\frac{2}{1+\|B\|}-1\right]=\frac{1-\|B\|^{n+1}}{1+\|B\|}\\ &\leqq\left\|\left(\sum_{k=0}^n B^k\right)^{-1}\right\|^{-1}.\end{aligned}$$

Thus by Lemma 3, σ_n is invertible. Now it is clear that σ_n is a Cauchy sequence and thus converges to an element σ of $\mathcal{A}$. By extending (3.9) and (3.10) to the limit, we see that

$$\begin{aligned}\left\|\sigma-\sum_{k=0}^\infty B^k\right\| &\leqq\frac{1}{1-(\delta+\|B\|)}-\frac{1}{1-\|B\|}\\ &\leqq\frac{1}{1-\|B\|}\left[\frac{2}{1+\|B\|}-1\right]\\ &=\frac{1}{1+\|B\|}\leqq\left\|\left(\sum_{k=0}^\infty B^k\right)^{-1}\right\|^{-1},\end{aligned}$$

and as before σ must be invertible. This completes the proof.

We note in passing that even if $(b_n)_{1\leqq n<\infty}$ is a sequence of complex numbers converging to zero and $|b_n|<1$ for all n, in fact even if $|b_n|\leqq\frac{1}{2}$ for all n, the sum

$b_1+b_1b_2+\cdots$ need not be nonzero. In fact, as Professor C. Rorres has pointed out to the authors, one may take $b_1=1$, $b_k=-2^{1-k}(k>1)$.

We now put the various pieces together to prove the main result of this section.

THEOREM 2. *Suppose A is a strongly proper, invertible element of $\mathcal{A}$. Let $(A_n)_{1\leqq n<\infty}$ be a sequence in $\mathcal{A}$ converging to A.*

Then the continued fraction (1.2) *is tail-end convergent.*

Proof. Let $U=U(A)$ be defined as before, $U'=I-U$, $V=U'U^{-1}$. Since $\|V\|<1$ by hypothesis, we may choose $\delta(V)$ as in Lemma 5. Then select ε so small that $\|U^{-1}\|\varepsilon(1-\varepsilon)^{-1}<\min(1,\delta(V))$ and choose $\delta\equiv\delta(\varepsilon)$ as in Lemma 4. Let us suppose, first of all that $\|A-A_n\|\leqq\delta$ for all $n\geqq 1$, and choose the sequence $(U_n)_{0\leqq n<\infty}$ in accordance with Lemma 4. For each n, let $U'_n=I-U_n$, and note that Lemma 4, (1) becomes:

$$U_nU'_{n+1}=-A_{n+1}\qquad(n\geqq 0).\tag{3.11}$$

We now recall equations (1.6). Note that if we define $R_{-1}=0$ and $S_{-1}=I$, the initial values in (1.6) tell us that the inductive equation continues to hold for $n=0$. Thus if we let D_n, $n\geqq -1$, denote either R_n or S_n, then we see that $D_{n+1}=D_n+A_{n+1}D_{n-1}(n\geqq 0)$. By combining this with (3.11) we get: $D_n-(U_n+U'_n)D_{n-1}+U_{n-1}U'_nD_{n-2}=0$, $n\geqq 1$ or

$$(D_n-U_nD_{n-1})=U'_n(D_{n-1}-U_{n-1}D_{n-2}),\qquad n\geqq 1.\tag{3.12}$$

Let n run through the integers from 1 to k and add the resulting equations, getting:

$$\begin{aligned}D_k-U_kD_{k-1}&=U'_k\cdots U'_1(D_0-U_0D_{-1})\\&=(U_1U_2\cdots U_k)(D_0-U_0D_{-1})(U'_1U_1^{-1})(U'_2U_2^{-1})\cdots(U'_nU_n^{-1}).\end{aligned}\tag{3.13}$$

If we now let k run from 1 to n and adjoin a simple identity, we get the system:

$$\begin{aligned}(D_n-U_nD_{n-1})&=(U_1\cdots U_n)(D_0-U_0D_{-1})(U'_1U_1^{-1})\cdots(U'_nU_n^{-1})\\(D_{n-1}-U_{n-1}D_{n-2})&=(U_1\cdots U_{n-1})(D_0-U_0D_{-1})(U'_1U_1^{-1})\cdots(U'_{n-1}U_{n-1}^{-1})\\&\vdots\\(D_1-U_1D_0)&=U_1(D_0-U_0D_{-1})(U'_1U_1^{-1})\\(D_0-U_0D_{-1})&=(D_0-U_0D_{-1}).\end{aligned}\tag{3.14}$$

Multiply these equations respectively by $I, U_n,\cdots,(U_nU_{n-1}\cdots U_2)$, $(U_n\cdots U_1)$ and add getting:

$$\begin{aligned}&D_n-(U_n\cdots U_0)D_{-1}=(U_1U_2\cdots U_n)(D_0-U_0D_{-1})\sigma_n,\\&\sigma_n=I+(U'_1U_1^{-1})+(U'_1U^{-1})(U'_2U_2^{-1})+\cdots+(U'_1U_1^{-1})\cdots(U'_nU_n^{-1}).\end{aligned}\tag{3.15}$$

If we let $D_n=S_n$, we see that $S_n=(U_n\cdots U_0)+(U_1U_2\cdots U_n)U'_0\sigma_n$, or:

$$\begin{aligned}S_n=(U_n\cdots U_0)[I&+(U'_0U_0^{-1})+(U'_0U_0^{-1})+\cdots\\&+(U'_0U_0^{-1})\cdots(U'_nU_n^{-1})]\equiv(U_n\cdots U_0)\mu_n.\end{aligned}\tag{3.16}$$

Similarly if we let $D_n=R_n$ we have:

$$R_n=(U_n\cdots U_1)\sigma_n.\tag{3.17}$$

We know by Lemma 4 that $\|I-U_nU^{-1}\|<\varepsilon$ for all $n\geqq 0$. If we let $\gamma_n=I-U_nU^{-1}$, then $U_n=(I-\gamma_n)U$, and so $(U'_nU_n^{-1}-U'U^{-1})=U_n^{-1}-U^{-1}=U^{-1}[(I-\gamma_n)^{-1}-I]=U^{-1}(\sum_{k=1}^{\infty}(\gamma_n)^k)$ (since $\|\gamma\|<1$). Hence if $n\geqq 0$, $\|U'_nU_n^{-1}-V\|\leqq\|U^{-1}\|\sum_{k=1}^{\infty}\varepsilon^k=\|U^{-1}\|\varepsilon/(1-\varepsilon)<\delta(V)$ (by assumption).

Then Lemma 5 tells us that each σ_n and μ_n above are invertible and converge respectively to invertible elements σ and μ of $\mathscr{A}$. Thus $S_n^{-1}R_n=U_0^{-1}\mu_n^{-1}\sigma_n$ exists for all n and converges to $U_0^{-1}\mu^{-1}\sigma$, which means (1.2) converges. Finally if the conditions of the theorem hold, there must be an N such that $\|A-A_n\|\leqq\delta$ for $n\geqq N$. Applying all the above to the c.f. (1.3), we see that (1.2) is tail-end convergent, and we are done.

REFERENCES

[1] W. FAIR, *Noncommutative continued fractions*, this Journal, 2 (1971), pp. 226–232.
[2] A. N. KHOVANSKII, *The Application of Continued Fractions and Their Generalizations to Problems in Approximation Theory*, P. Noordhoff, Groningen, the Netherlands, 1963.
[3] M. A. NAIMARK, *Normed Rings*, P. Noordhoff, Groningen, the Netherlands, 1960.
[4] A. PRINGSHEIM, *Ueber Konvergenz and functionen-theoretischen Charakter gewisser Limitar-periodischer Kettenbrüche*, Sitzungsber. der Math. phys. Klasse der Kgl. Bayer. Akad. Wiss. München, 6 (1910), pp. 1–52.
[5] C. E. RICKART, *General Theory of Banach Algebras*, Van Nostrand, Princeton, NJ, 1960.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0009 $01.00/0

LINEARIZATION OF MONOTONE HILBERT NETWORKS*

VACLAV DOLEZAL†

Abstract. In this paper we consider a linearization of a nonlinear monotone Hilbert network in a neighborhood of some operating point. Three theorems are established which give estimates for the difference between the exact and approximate current distribution. Two examples of specific networks are considered.

Introduction. The problem dealt with in this paper resembles the small-signal analysis of nonlinear networks [1] and [2], but differs from it in several aspects. To compare the present approach, let us recall the following facts. In the standard small-signal analysis we assume that the network under consideration is described by a canonic system of differential equations

$$(*) \qquad \dot{x} = f(x, u, t), \quad x(0) = 0,$$

where x is a state variable vector and u represents the signals. If the solution $x_0(t)$ of $(*)$ with $u = u_0(t)$ is known, we seek the solution $x_v(t)$ of $(*)$ with $u = u_0(t) + v(t)$ in the form $x_v(t) = x_0(t) + \xi_v(t)$. Assuming that $\|v\|$ is small, we then take $x_0(t) + \xi_v^*(t)$ as an approximation to $x_v(t)$, where $\xi_v^*(t)$ is the solution of the linear equation

$$\binom{*}{*} \qquad \dot{\xi}_v^* = A(t)\xi_v^* + B(t)v(t), \qquad \xi_v^*(0) = 0.$$

Here, $A(t)$ is the Jacobian of f with respect to x evaluated at (x_0, u_0), and the meaning of $B(t)$ is analogous. Thus, in the vicinity of (x_0, u_0) we replace $f(x, u, t)$ by a strictly local approximation $l(x, u, t) = A(x - x_0) + B(u - u_0) + f(x_0, u_0, t)$ that is linear in x and u.

Now, suppose that, for a given $r > 0$, we define the "overall relative error" λ_r by

$$\lambda_r = \sup_{\|v\| \leqq r} \|\xi_v - \xi_v^*\| \cdot \|v\|^{-1}.$$

From the construction of ξ_v^* it follows that λ_r decreases if r does. However, if r is not "too small", then clearly λ_r can be diminished, if we replace A and B in $\binom{*}{*}$ by some more suitable matrices A_r and B_r such that $A_r(x - x_0) + B_r(u - u_0) + f(x_0, u_0, t)$ is a better approximation of $f(x, u, t)$ on some neighborhood $\Omega_r(x_0, u_0)$ (depending on r) than $l(x, u, t)$ is. In other words, the small-signal analysis fails to give good results, if $\|v\|$ is not sufficiently small.

The present approach, apart from its more general setting, introduces more flexibility into the problem. To explain the underlying idea, consider a Hilbert network $\hat{\mathcal{N}} = (\hat{Z}, G)$ [4]. Assume that we know the current distribution i_0 in $\hat{\mathcal{N}}$ corresponding to some EMF vector e_0 (operating point), and that we seek distributions i_{e^*}, which correspond to excitations $e_0 + e^*$, where the e^*'s satisfy the inequality $\|e^*\| \leqq r$ with some given $r > 0$. To find approximations to the i_{e^*}'s, we linearize the network $\hat{\mathcal{N}}$ in a vicinity of i_0. More specifically, defining the operator $\hat{Z}'$ by $\hat{Z}'\xi = \hat{Z}(i_0 + \xi) - \hat{Z}i_0$, we assume that we can find a linear operator $\hat{Z}_0$ which satisfies the inequality $\|\hat{Z}'\xi - \hat{Z}_0\xi\| \leqq a\|\xi\|$ on a certain ball centered at the origin, where $a > 0$ is

* Received by the editors October 26, 1976, and in revised form May 10, 1977.

† Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York 11794. This research was supported by the National Science Foundation under Grant MPS7505268.

not too large. If $i^*_{e^*}$ is the solution of the *linear* network $\hat{\mathcal{N}}_0=(\hat{\mathcal{Z}}_0, G)$ corresponding to e^*, we will take $i_0+i^*_{e^*}$ for an approximation to the solution i_{e^*} of $\hat{\mathcal{N}}$.

The overall relative error

$$(+)\qquad \lambda_r=\sup_{\|e^*\|\leqq r}\|i_{e^*}-(i_0+i^*_{e^*})\|\cdot\|e^*\|^{-1}$$

depends, of course, on the choice of $\hat{\mathcal{Z}}_0$, and consequently, on the constant a. However, in specific cases of networks it is usually not difficult to construct $\hat{\mathcal{Z}}_0$ so that, for a given $r>0$, a is small enough. (Refer to the examples at the end of the paper). In this context, let us emphasize that taking the Fréchet derivative of $\hat{\mathcal{Z}}$ at i_0 for $\hat{\mathcal{Z}}_0$ (a counterpart of the small-signal analysis) need not lead to smallest values of a.

It turns out that if $\hat{\mathcal{Z}}$ is strongly monotone, we can give a simple upper bound for λ_r, which is roughly proportional to a for a small. The theorems that follow deal with various cases of assumptions imposed on the definition domain of $\hat{\mathcal{Z}}$. Note also that $\hat{\mathcal{Z}}$ need not be single-valued.

Results. Let us begin with some preliminary considerations. The concepts "set mapping", "simple", $\mathfrak{S}(\mathcal{H})$, etc. have the same meaning as in [4]. Also, if $h>0$, let $B_h=\{x: x\in\mathcal{H}, \|x\|\leqq h\}$.

LEMMA 1. *Let $\mathcal{H}$ be a real Hilbert space, let $D\neq\{0\}$ be a linear subspace of $\mathcal{H}$, and let $r>0$. Assume that*

(i) *$T: D\to\mathfrak{S}(\mathcal{H})$ is a set mapping such that*

$$(1)\qquad \langle y_1-y_2, x_1-x_2\rangle\geqq b\|x_1-x_2\|^2$$

for all $x_i\in D$, $y_i\in Tx_i$, $i=1,2$, and some fixed $b>0$,

(ii) *there exists a* linear *operator $T_0: D\to\mathcal{H}$ and a constant a with $0<a<b$ such that*

$$(2)\qquad \|y-T_0x\|\leqq a\|x\|$$

for all $x\in D\cap B_R$ with $R=b^{-1}r$ and all $y\in Tx$,

(iii) *$T_0D=\mathcal{H}$.*

Then T is simple on D, and T_0 possesses a bounded inverse $T_0^{-1}:\mathcal{H}\to D$ with $\|T_0^{-1}\|\leqq (b-a)^{-1}$. Moreover, if $y\in(TD)^0\cap B_r$ and $x=T^-y$, $x_0=T_0^{-1}y$, then

$$(3)\qquad \|x-x_0\|\leqq ab^{-1}(b-a)^{-1}\|y\|.$$

Proof. Simplicity of T follows immediately from (1). Thus, $T^-:(TD)^0\to D$ is an operator, and (1) implies that

$$(4)\qquad \|T^-y_1-T^-y_2\|\leqq b^{-1}\|y_1-y_2\|$$

for all $y_1, y_2\in(TD)^0$. Moreover, since $0\in D$, (2) shows that $T0=\{0\}$. Hence, $T^-0=0$ and (4) yields

$$(5)\qquad \|T^-y\|\leqq b^{-1}\|y\|$$

for all $y\in(TD)^0$.

On the other hand, if $x\in D$ and $y\in Tx$, we have by (1),

$$(6)\qquad \langle y, x\rangle\geqq b\|x\|^2.$$

Now, if $x\in D\cap B_R$, then, for any $y\in Tx$,

$$(7)\qquad \langle T_0x, x\rangle=\langle y, x\rangle-\langle y-Tx_0, x\rangle.$$

Also, by (2),

$$\langle y - T_0x, x\rangle \leqq \|y - T_0x\| \cdot \|x\| \leqq a\|x\|^2.$$

Hence, (6) and (7) yield

$$\langle T_0x, x\rangle \geqq (b-a)\|x\|^2. \tag{8}$$

However, since T_0 is linear, (8) holds for every $x \in D$.

Moreover, since $b-a>0$, (8) shows that T_0 is one to one on D. Hence, the inverse $T_0^{-1}: \mathscr{H} \to D$ exists. Thus, by (8) and Schwarz inequality, $\|T_0^{-1}y\| \leqq (b-a)^{-1}\|y\|$ for each $y \in \mathscr{H}$, i.e., T_0^{-1} is bounded and

$$\|T_0^{-1}\| \leqq (b-a)^{-1}. \tag{9}$$

Next, let $y \in (TD)^0 \cap B_r$ and $x = T^-y$, $x_0 = T_0^{-1}y$. Then $x \in D$, and by (5), $\|x\| \leqq b^{-1}\|y\| \leqq b^{-1}r = R$, i.e., $x \in D \cap B_R$. Moreover, we have by (2) and (9),

$$\begin{aligned}\|x - x_0\| &= \|T_0^{-1}(T_0x - T_0x_0)\| = \|T_0^{-1}(T_0x - y)\| \\ &\leqq \|T_0^{-1}\| \cdot \|T_0x - y\| \leqq (b-a)^{-1}a\|x\| \leqq (b-a)^{-1}ab^{-1}\|y\|\end{aligned}$$

and the proof is complete.

The assumption (iii) in Lemma 1 is sometimes inconvenient. Fortunately, it can be traded for another assumption. Indeed, we have the following modification of Lemma 1.

LEMMA 2. *Let $\mathscr{H}$ be a real Hilbert space, let D be a nonempty subset of $\mathscr{H}$ such that $0 \in D$, and let $r>0$. Assume that*

(i) *$T: D \to \mathfrak{S}(\mathscr{H})$ is a set mapping satisfying* (1) *with some fixed $b>0$,*

(ii) *there exists a linear bounded operator T_0: $\mathscr{H} \to \mathscr{H}$ and a constant a with $0<a<b$ such that* (2) *is satisfied for all $x \in D \cap B_R$ with $R = b^{-1}r$ and all $y \in Tx$,*

(iii) *$D \cap B_R$ is dense in B_R, i.e., $\overline{D \cap B_R} = B_R$.*

Then T is simple on D, and T_0 possesses a bounded inverse T_0^{-1}, $\mathscr{H} \to \mathscr{H}$ with $\|T_0^{-1}\| \leqq (b-a)^{-1}$. Moreover, if $y \in (TD)^0 \cap B_r$ and $x = T^-y$, $x_0 = T_0^{-1}y$, then (3) *holds.*

Proof. As before, it follows that (8) holds for each $x \in D \cap B_R$. However, (iii) and continuity of T_0 imply that (8) holds on B_R, and consequently, by linearity of T_0, everywhere on $\mathscr{H}$. Since $b-a>0$, (8) shows that T_0 is a maximal monotone, coercive operator [3], so that $T_0\mathscr{H} = \mathscr{H}$. The rest of the proof is the same as that of Lemma 1.

The assumption on boundedness of T_0 can be dropped, if we strengthen the above condition (iii). Actually, we have

LEMMA 3. *Lemma 2 remains true if conditions* (ii) *and* (iii) *are replaced by the following assumptions:*

(ii)* *$B_R \subset D$ for $R = b^{-1}r$,*

(iii)* *there exists a linear operator T_0: $\mathscr{H} \to \mathscr{H}$ and a constant a with $0<a<b$ such that* (2) *holds for all $x \in B_R$ and $y \in Tx$.*

If, in addition, $D = \mathscr{H}$ and T is a hemicontinuous operator, then T possesses an inverse $T^{-1}: \mathscr{H} \to \mathscr{H}$, and (3) *holds for any $y \in B_r$, where $x = T^{-1}y$, $x_0 = T_0^{-1}y$.*

Proof. We conclude as before that (8) holds for every $x \in B_R$. However, since T_0 is linear, (8) is satisfied on the entire space $\mathscr{H}$. Moreover, linearity of T_0 implies that T_0 is hemicontinuous on $\mathscr{H}$. Hence, T_0 is maximal monotone and coercive, so that $T_0\mathscr{H} = \mathscr{H}$ [3]. The rest of the proof is obvious.

As for the last assertion, note that our assumptions imply that T is maximal monotone and coercive by (1). Thus, $T\mathscr{H} = \mathscr{H} \supset B_r$, which proves our claim.

Observe that, under the assumptions of Lemma 3, T_0 is automatically bounded. Indeed, it follows that $T_0^{-1}:\mathcal{H}\to\mathcal{H}$ is bounded and (9) holds. Thus, by the open mapping theorem, T_0 itself is bounded.

Let us now apply the above results to Hilbert networks. To avoid restatement of basic definitions, we refer the reader to the survey paper [4]. We will consistently use the notation introduced there.

In order to simplify the formulation of the following theorems, we will assume without loss of generality that the operating point e_0 and the corresponding solution i_0 of $\hat{\mathcal{N}}=(\hat{Z},G)$ are zero. Actually, if i_0 corresponding to $e_0=0$ were not zero, then instead of $\hat{\mathcal{N}}$ we could consider the network $\hat{\mathcal{N}}'=(\hat{Z}',G)$ with $\hat{Z}'$ being defined on $D'=D-i_0$ by

$$\hat{Z}'x=\hat{Z}(i_0+x)-\hat{Z}i_0, \tag{10}$$

whose solution i_0' corresponding to $e_0'=0$ is zero.

THEOREM 1. *Let H be a real Hilbert space, let $\hat{\mathcal{N}}=(\hat{Z},G)$ be a Hilbert network with $\hat{Z}$ being a set mapping defined on a nontrivial linear subspace D of H^{c_2}, and let $r>0$. Furthermore, let $F=\hat{X}^*(N_{\hat{a}}\cap D)\subset H^{c_0}$, and let $W:F\to\mathfrak{S}(H^{c_0})$ be defined by $W=\hat{X}^*\hat{Z}\hat{X}$.*

Assume that

(i) *there exists $b>0$ such that*

$$\langle y_1-y_2,x_1-x_2\rangle\geqq b\|x_1-x_2\|^2 \tag{11}$$

for all $x_i\in F$, $y_i\in Wx_i$, $i=1,2$,

(ii) *there exists a* linear *operator $\hat{Z}_0:D\to H^{c_2}$ and a constant a with $0<a<b$ such that*

$$\|y-W_0x\|\leqq a\|x\| \tag{12}$$

for all $x\in F\cap B_R'$ and $y\in Wx$, where $W_0=\hat{X}^*\hat{Z}_0\hat{X}$, $B_R'=\{x:x\in H^{c_0},\|x\|\leqq R\}$ and $R=b^{-1}r$,

(iii) *the network $\hat{\mathcal{N}}_0=(\hat{Z}_0,G)$ possesses a solution for any $e\in H^{c_2}$.*

If $i\in H^{c_2}$ is the (unique) solution of $\hat{\mathcal{N}}$ corresponding to some $e^*\in H^{c_2}$ with $\|\hat{X}^*e^*\|\leqq r$, and i^* is the (unique) solution of $\hat{\mathcal{N}}_0$ corresponding to e^*, then

$$\|i-i^*\|\leqq ab^{-1}(b-a)^{-1}\|\hat{X}^*e^*\|. \tag{13}$$

Moreover, (12) *can be replaced by the stronger assumption that*

$$\|u-\hat{Z}_0v\|\leqq a\|v\| \tag{14}$$

for all $v\in D\cap B_R$ and $u\in\hat{Z}v$, where $B_R=\{v:v\in H^{c_2},\|v\|\leqq R\}$.

Proof. By Theorem 2 in [4], (iii) is equivalent to the conditon $W_0F=H^{c_0}$. Thus, by Lemma 1, W is simple on F, and W_0 has a bounded inverse $W_0^{-1}:H^{c_0}\to F$. Consequently, Theorem 2 in [4] shows that each solution of $\hat{\mathcal{N}}$ is determined uniquely, and the same is true for $\hat{\mathcal{N}}_0$.

Next, let $e^*\in H^{c_2}$ be such that $\hat{\mathcal{N}}$ possesses a solution i corresponding to e^*, and let $\|\hat{X}^*e^*\|\leqq r$. Then $\hat{X}^*e^*\in(WF)^0$ and we have by (3) in Lemma 1,

$$\|W^-\hat{X}^*e^*-W_0^{-1}\hat{X}^*e^*\|\leqq ab^{-1}(b-a)^{-1}\|\hat{X}^*e^*\|. \tag{15}$$

However, by Theorem 2 in [4],

$$i=\hat{X}W^-\hat{X}^*e^*,\qquad i^*=\hat{X}W_0^{-1}\hat{X}^*e^*. \tag{16}$$

Since $\hat{X}$ is a norm-preserving isomorphism ([4, Proposition 1]), the inequality (13) follows immediately from (15) and (16).

To prove the last proposition, assume that (14) holds. Since $\hat{X}H^{c_0} = N_{\hat{a}}$ and $\hat{X}$ is norm-preserving, it follows that $\hat{X}B'_R = N_{\hat{a}} \cap B_R$. Moreover, since $\hat{X}F = N_{\hat{a}} \cap D$, we have

$$\hat{X}(F \cap B'_R) = N_{\hat{a}} \cap B_R \cap D. \tag{17}$$

Now, let $x \in F \cap B'_R$, and let $y \in Wx = \hat{X}^*\hat{Z}\hat{X}x$. Then $\hat{X}x \in B_R \cap D$ by (17), $y = \hat{X}^*z$ for some $z \in \hat{Z}\hat{X}x$, and (14) yields

$$\begin{aligned}\|y - W_0x\| = \|\hat{X}^*z - \hat{X}^*\hat{Z}_0\hat{X}x\| &= \|\hat{X}^*(z - \hat{Z}_0\hat{X}x)\| \\ &\leqq \|z - \hat{Z}_0\hat{X}x\| \leqq a\|\hat{X}x\| = a\|x\|.\end{aligned}$$

Hence, (12) is satisfied and the proof is completed.

THEOREM 2. *Let H be a real Hilbert space, let $\hat{\mathcal{N}} = (\hat{Z}, G)$ be a Hilbert network with $\hat{Z}$ being a set mapping defined on a nonempty subset $D \subset H^{c_2}$ such that $0 \in D$, and let $r > 0$. Furthermore, let $F = \hat{X}^*(N_{\hat{a}} \cap D) \subset H^{c_0}$, and let $W: F \to \mathfrak{S}(H^{c_0})$ be defined by $W = \hat{X}^*\hat{Z}\hat{X}$.*

Assume that

(i) *there exists $b > 0$ such that* (11) *is satisfied for all $x_i \in F$, $y_i \in Wx_i$, $i = 1, 2$,*

(ii) *there exists a linear bounded operator $\hat{Z}_0: H^{c_2} \to H^{c_2}$ and a constant a with $0 < a < b$ such that* (12) *holds for all $x \in F \cap B'_R$ and $y \in Wx$, where $W_0 = \hat{X}^*\hat{Z}_0\hat{X}$, $B'_R = \{x : x \in H^{c_0}, \|x\| \leqq R\}$ and $R = b^{-1}r$.*

(iii) *$F \cap B'_R$ is dense in B'_R.*

If $i \in H^{c_2}$ is the (unique) solution of $\hat{\mathcal{N}}$ corresponding to some $e^* \in H^{c_2}$ with $\|\hat{X}^*e^*\| \leqq r$, and i^* is the (unique) solution of $\hat{\mathcal{N}}_0 = (\hat{Z}_0, G)$ corresponding to e^*, then inequality (13) *is satisfied.*

Moreover, in (ii) *the requirement that* (12) *holds can be replaced by the stronger assumption that* (14) *is fulfilled for all $v \in D \cap B_R$ and $u \in \hat{Z}v$, where $B_R = \{v : v \in H^{c_2}, \|v\| \leqq R\}$.*

Proof. If we recall Lemma 2 it follows that W is simple on F, and W_0 possesses a bounded inverse $W_0^{-1}: H^{c_0} \to H^{c_0}$. Thus, by Theorem 2 in [4], every solution of $\hat{\mathcal{N}}$ is determined uniquely, and, for any $e^* \in H^{c_2}$, $\hat{\mathcal{N}}_0$ has a unique solution corresponding to e^*.

The proof of inequality (13) follows the same pattern as in the proof of Theorem 1. As for the last assertion, we confirm as before that (14)$\Rightarrow$(12). (Note that (17) holds independently of whether D is a linear subspace or not).

THEOREM 3. *Theorem* 2 *remains true if its conditions* (ii) *and* (iii) *are replaced by the following assumptions*:

(ii)* *$B_R \subset D$ for $R = b^{-1}r$,*

(iii)* *there exists a* linear *operator $\hat{Z}_0: H^{c_2} \to H^{c_2}$ and a constant a with $0 < a < b$ such that* (12) *holds for all $x \in B'_R$ and $y \in Wx$, where $W_0 = \hat{X}^*\hat{Z}_0\hat{X}$.*

If, in addition, $D = H^{c_2}$ and W is a hemicontinuous operator on H^{c_0}, then $\hat{\mathcal{N}}$ possesses a unique solution for any $e^* \in H^{c_2}$, i.e., (13) *holds for any $e^* \in H^{c_2}$ with $\|\hat{X}^*e^*\| \leqq r$.*

Proof. If (ii)* holds, we have by (17), $\hat{X}(F \cap B'_R) = N_{\hat{a}} \cap B_R$. Since $\hat{X}B'_R = N_{\hat{a}} \cap B_R$ and $\hat{X}$ is one-to-one, it follows that $F \cap B'_R = B'_R$, and consequently, $B'_R \subset F$. Applying now Lemma 3 to W and W_0, we confirm our claims.

Finally, if $D = H^{c_2}$, then $F = H^{c_0}$. Hemicontinuity of W and (11) imply that W is a maximal monotone, coercive operator, and consequently, $WH^{c_0} = H^{c_0}$ [3]. Thus, by Theorem 2 in [4], $\hat{\mathcal{N}}$ has a solution for any $e^* \in H^{c_2}$.

Remark 1. As it was pointed out above, Theorems 1–3 deal with the particular case that the operating point e_0 and the corresponding solution i_0 of $\hat{\mathcal{N}}$ are zero. On the other hand, the conditions (11), (12) we used are invariant with respect to "shifting". More specifically, if the solution i_0 of $\hat{\mathcal{N}}$ corresponding to $e_0=0$ is not zero and $\hat{X}^*\hat{Z}\hat{X}$ satisfies (11), then we see easily that $\hat{X}^*\hat{Z}'\hat{X}$ also satisfies (11), where $\hat{Z}'$ is defined by (10). A similar statement is true for condition (12).

Let us now consider two examples of specific networks.

Example 1. Let H be the real space $L_2[0, \tau]$, (we will also write L_2 for brevity), and let

$$K_0=\{x : x \text{ absolutely continuous on } [0, \tau], x(0)=0, x' \in L_2\}.$$

Let G be a finite oriented graph having c_2 branches and c_1 nodes, and let d be the $c_2 \times c_1$ incidence matrix of G. Let X be a real $c_2 \times c_0$ matrix, whose columns constitute an orthonormal basis in the solution space of the equation $d^T \cdot \xi=0$, $\xi \in R^{c_2}$. Finally, let k be a fixed integer with $1 \leqq k \leqq c_2$.

Furthermore, assume that

(i) l is a real, symmetric, positive definite $k \times k$ matrix, and that the $c_2 \times c_2$ matrix L is defined by

$$L=\left[\begin{array}{c|c} l & 0 \\ \hline 0 & 0 \end{array}\right].$$

(ii) $R : R^{c_2} \to R^{c_2}$ is a continuous function such that

$$|R(\sigma)| \leqq \alpha+\beta|\sigma| \tag{18}$$

for all $\sigma \in R^{c_2}$ and some fixed $\alpha, \beta \geqq 0$, and

$$X^T[R(X\xi)-R(X\eta)](\xi-\eta) \geqq b|\xi-\eta|^2 \tag{19}$$

for all $\xi, \eta \in R^{c_0}$ and some fixed $b>0$ (here, $|\cdot|$ signifies the Euclidean norm).

(iii) S is a real, symmetric, positive semidefinite matrix.

Let $D=K_0^k \times L_2^{c_2-k} \subset L_2^{c_2}$, let $\hat{Z}$ be defined on D by

$$(\hat{Z}x)(t)=\{Lx(t)\}'+R(x(t))+S\int_0^t x(\sigma)\,d\sigma, \tag{20}$$

and consider the Hilbert network $\hat{\mathcal{N}}=(\hat{Z}, G)$.

Clearly, $\hat{\mathcal{N}}$ is a model of an LRC-network, whose resistors are nonlinear, time-invariant, and whose inductors and capacitors are constant. Note that the inductors (and possible mutual inductive couplings) are confined only to branches $b_1, b_2, \cdots, b_k$, and that the initial currents in these branches are assumed to be zero.

From the assumptions (i)–(iii) it follows easily that $\hat{Z}$ maps D into $L_2^{c_2}$, and that

$$\langle Wx_1-Wx_2, x_1-x_2\rangle \geqq b\|x_1-x_2\|^2$$

for all $x_1, x_2 \in F$, where $F=\hat{X}^*(N_{\hat{a}} \cap D)$ and $W=\hat{X}^*\hat{Z}\hat{X}$, i.e., $\hat{\mathcal{N}}$ satisfies the condition (i) in Theorem 1.

In addition to (i)–(iii), let us make the following assumptions:

(iv) There exists a $c_2 \times c_2$ matrix M and a constant a with $0<a<b$ such that

$$|R(\sigma)-M\sigma| \leqq a|\sigma| \tag{21}$$

for all $\sigma \in R^{c_2}$.

(v) If the operator $\hat{Z}_0$ is defined on D by

$$(\hat{Z}_0 x)(t) = \{Lx(t)\}' + Mx(t) + S\int_0^t x(\sigma)\,d\sigma, \tag{22}$$

then the linear network $\hat{\mathcal{N}}_0 = (\hat{Z}_0, G)$ (with constant elements) possesses a solution in D for any $e^* \in L_2^{c_2}$. (Note that we can give fairly simple conditions under which Hilbert networks of type like $\hat{\mathcal{N}}$ or $\hat{\mathcal{N}}_0$ have a solution for each $e^* \in L_2^{c_2}$. A detailed treatment can be found in [5].)

From (20), (22) it follows that $\hat{Z}x - \hat{Z}_0 x = R(x) - Mx$. A routine argument will convince us that, due to (21), we have $\|\hat{Z}x - \hat{Z}_0 x\| \leqq a\|x\|$ for all $x \in D$, i.e., condition (14) in Theorem 1 is met with any $r > 0$.

Thus, all assumptions of Theorem 1 are satisfied. Consequently, if $\hat{\mathcal{N}}$ has a solution i for some $e^* \in L_2^{c_2}$, and if i^* is the solution of the linear network $\hat{\mathcal{N}}_0$ corresponding to e^*, then by (13),

$$\|i - i^*\| \leqq ab^{-1}(b-a)^{-1}\|X^T e^*\|.$$

Example 2. Consider the infinite DC-network $\hat{\mathcal{N}}$ given in Fig. 1 whose graph G is formed by an infinite grid in the plane. We will assume that each branch of G contains the *same* nonlinear resistor described by a function $\rho: R^1 \to R^1$, whose graph may look like the one in Fig. 2. Our objective will be to establish a linearization $\hat{\mathcal{N}}_0$ of $\hat{\mathcal{N}}$ which will make the error in the solution as small as possible for all excitations e^*, whose norms do not exceed some given $r > 0$. More specifically, we will try to find a resistance ρ_0 such that the linear network $\hat{\mathcal{N}}_0$, obtained from $\hat{\mathcal{N}}$ by replacing each nonlinear resistor by a linear resistor with resistance ρ_0, will approximate the behavior of $\hat{\mathcal{N}}$.

To this end, assume that the function $\rho: R^1 \to R^1$ satisfies the following conditions:

(i) $\rho(0) = 0$,

(ii) $\beta(\xi_1 - \xi_2)^2 \geqq (\rho(\xi_1) - \rho(\xi_2))(\xi_1 - \xi_2) \geqq \alpha(\xi_1 - \xi_2)^2$ for all $\xi_1, \xi_2 \in R^1$ and some fixed α, β such that

$$3\alpha > \beta \geqq \alpha > 0. \tag{23}$$

Using the standard notation, in our case we have $H = R^1$, and consequently,

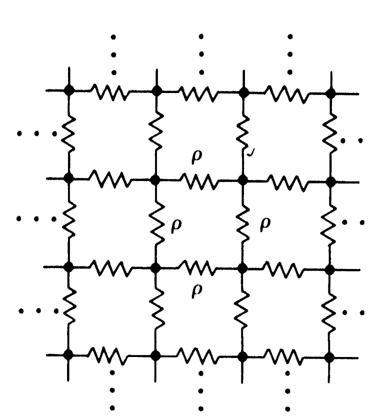

FIG. 1

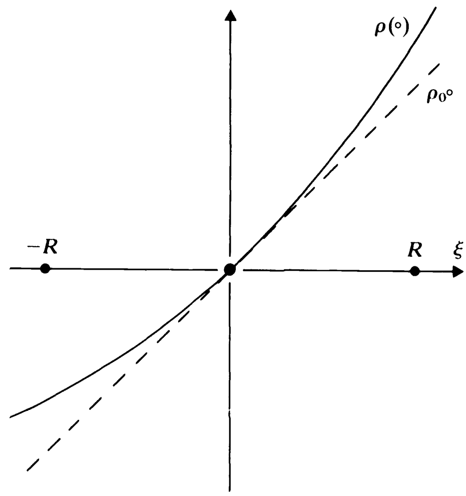

FIG. 2

$H^{c_2}=R^{\aleph_0}=l_2$. Moreover, the operator $\hat{Z}$ is defined on l_2 by

(24) $$\hat{Z}z=[\rho(z_1),\rho(z_2),\rho(z_3),\cdots]^T.$$

Thus, our network is modeled by the Hilbert network $\hat{\mathfrak{N}}=(\hat{Z},G)$.

From the assumptions (i), (ii) it follows readily that $\hat{Z}$ maps l_2 into itself, $\hat{Z}0=0$ and

(25) $$\langle \hat{Z}z_1-\hat{Z}z_2,z_1-z_2\rangle \geqq \alpha\|z_1-z_2\|^2,$$

(26) $$\|\hat{Z}z_1-\hat{Z}z_2\|\leqq\beta\|z_1-z_2\|$$

for all $z_1,z_2\in l_2$. However, (25) shows that the operator $W=\hat{X}^*\hat{Z}\hat{X}$ satisfies the condition (11) on $F=R^{c_0}=l_2$. Indeed, if $x_1,x_2\in l_2$, we have

$$\begin{aligned}\langle Wx_1-Wx_2,x_1-x_2\rangle&=\langle\hat{X}^*\hat{Z}\hat{X}x_1-\hat{X}^*\hat{Z}\hat{X}x_2,x_1-x_2\rangle\\&=\langle\hat{Z}\hat{X}x_1-\hat{Z}\hat{X}x_2,\hat{X}x_1-\hat{X}x_2\rangle\geqq\alpha\|\hat{X}(x_1-x_2)\|^2=\alpha\|x_1-x_2\|^2.\end{aligned}$$

Similarly, (26) implies that W is continuous on l_2.

Next, if ρ_0 is a real number, define the operator $\hat{Z}_0: l_2\to l_2$ by $\hat{Z}_0z=\rho_0 z$, and let $\hat{\mathcal{N}}_0=(\hat{Z}_0,G)$. Naturally, for our linearization we are going to choose ρ_0 so that the constant a in (12) (or in (14)) can be made as small as possible. To find such ρ_0, denote $R=\alpha^{-1}r$, $K_R=[-R,R]-\{0\}$, and let

(27) $$S_R=\sup_{\xi\in K_R}\xi^{-1}\rho(\xi),\qquad J_R=\inf_{\xi\in K_R}\xi^{-1}\rho(\xi).$$

A little thought will convince us that with

(28) $$\rho_0=\tfrac12(S_R+J_R),\qquad a_R=\tfrac12(S_R-J_R)$$

we have

(29) $$a_R=\sup_{\xi\in K_R}|\xi|^{-1}|\rho(\xi)-\rho_0\xi|=\inf_{\omega\in R^1}\sup_{\xi\in K_R}|\xi|^{-1}|\rho(\xi)-\omega\xi|.$$

Consequently, we take this ρ_0 for the resistance of resistors in our linear network $\hat{\mathcal{N}}_0$.

Now, (29) yields $|\rho(\xi)-\rho_0\xi|\leqq a_R|\xi|$ for every $\xi\in[-R,R]$. Thus, if $z=[z_j]\in l_2$, $\|z\|\leqq R$, then $|z_j|\leqq R$ for every j and we have

$$\|\hat{Z}z-\hat{Z}_0z\|=\sum_j|\rho(z_j)-\rho_0z_j|^2\leqq\sum_j a_R^2|z_j|^2=a_R^2\|z\|^2. \tag{30}$$

On the other hand, (i) and (ii) yield $\beta\geqq\xi^{-1}\rho(\xi)\geqq\alpha$ for every $\xi\neq0$; hence, by (27),

$$\beta\geqq S_R,\qquad J_R\geqq\alpha. \tag{31}$$

Consequently, by (28),

$$a_R\leqq\tfrac{1}{2}(\beta-\alpha), \tag{32}$$

so that

$$\alpha-a_R\geqq\tfrac{1}{2}(3\alpha-\beta)>0. \tag{33}$$

Hence, (30) and (33) show that the operator $\hat{Z}_0$ satisfies (14), and consequently, the assumption (iii)* in Theorem 3.

Finally, referring to Theorem 3, let $e^*\in l_2$ and $\|e^*\|\leqq r$. Then $\hat{\mathcal{N}}$ possesses a unique solution i corresponding to e^*, and $\hat{\mathcal{N}}_0$ possesses a unique solution i^* corresponding to e^*. Since $\|\hat{X}^*e^*\|\leqq\|\hat{X}^*\|\cdot\|e^*\|\leqq r$, we have by (13) and (28),

$$\|i-i^*\|\leqq\tfrac{1}{2}(S_R-J_R)\alpha^{-1}[\alpha-\tfrac{1}{2}(S_R-J_R)]^{-1}\|e^*\|. \tag{34}$$

Using (32) and (33), we also get

$$\|i-i^*\|\leqq\alpha^{-1}(\beta-\alpha)(3\alpha-\beta)^{-1}\|e^*\|. \tag{35}$$

Note that if we put $r=\infty$ and $\rho_0'=\frac{1}{2}(S_\infty+J_\infty)$, where S_∞, J_∞ are given by (27), (with K_R replaced by $R^1-\{0\}$), then (35) holds for every $e^*\in l_2$.

Observe also the following: In our considerations leading to inequalities (34) and (35) we have never used the fact that the graph G is a grid, i.e., our results do not depend on G. Consequently, (34) and (35) hold for any (finite or infinite) DC-network, each branch of which contains the same nonlinear resistor that satisfies the requirements (i) and (ii).

REFERENCES

[1] B. PEIKARI, *Fundamentals of Network Analysis and Synthesis*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[2] C. A. DESOER, *Small signal theory of nonlinear lumped networks*, Network Theory, R. Boite, ed., Gordon and Breach, New York, 1972, pp. 311–318.

[3] R. T. ROCKAFELLAR, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., 149 (1970), pp. 75–88.

[4] V. DOLEZAL, *Basic properties of Hilbert networks*, IEEE Trans. Circuits and Systems CAS-23 (1976), no. 8, pp. 490–497.

[5] ———, *Solvability of a class of Hilbert networks*, this Journal, to appear.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0010 $01.00/0

ON CHARACTERIZATION OF FUNCTIONS BY THEIR GAUSS–CHEBYSHEV QUADRATURES*

ITSHAK BOROSH† AND CHARLES K. CHUI‡

Abstract. If p is a polynomial, then all but a finite number of the Gauss–Chebyshev quadrature formulae for p are exact. The purpose of this paper is to establish a converse as well as some related results. In particular, a formula is given to recapture a function from its Gauss–Chebyshev quadratures.

1. Introduction. Throughout this paper, we only consider Gauss quadrature formulae with nodes at the zeros of the Chebyshev polynomials

$$T_n(x)=\cos(n\cos^{-1}x)=2^{n-1}x^n+\cdots,$$

given by

$$\int_{-1}^{1} f(x)(1-x^2)^{-1/2}\,dx = Q_n(f)+R_n(f), \tag{1}$$

where

$$Q_n(f)=\frac{\pi}{n}\sum_{j=1}^{n} f(x_{n,j}),\qquad x_{n,j}=\cos\frac{(2j-1)\pi}{2n}. \tag{2}$$

The Gauss–Chebyshev quadrature formula (1) is said to be exact for a function f if $R_n(f)=0$. It is well-known that if f is a polynomial of degree at most m, then $R_n(f)=0$ for all $n>m/2$. It is natural to ask if a converse would hold; namely, if $R_n(f)=0$ for all sufficiently large n, is it possible to conclude that f is a polynomial? It is clear, however, that if f is an odd function on $[-1,1]$, then $R_n(f)=0$ for all n. Hence, based on the conditions $R_n(f)=0$, we have no information on the odd part of f. This leads to the consideration of $R_n(f')$ as well.

DEFINITION 1. For $f\in C[-1,1]$, consider the Chebyshev series

$$f(x)=\tfrac{1}{2}a_0+\sum_{k=1}^{\infty} a_kT_k(x), \tag{3}$$

where

$$a_k=\frac{2}{\pi}\int_{-1}^{1} f(x)T_k(x)(1-x^2)^{-1/2}\,dx, \tag{3'}$$

$k=0,1,\cdots$. We say that f belongs to the class $\mathscr{C}$ if

$$\sum|a_k|<\infty \tag{4}$$

and

$$2^n\sum_{k=1}^{\infty}|a_{2^nk}|\to 0. \tag{4'}$$

* Received by the editors November 23, 1977.

† Department of Mathematics, Texas A & M University, Texas 77843. The work of this author is partially supported by a National Science Foundation grant.

‡ Department of Mathematics, Texas A & M University, Texas 77843. The work of this author is supported in part by a National Science Foundation grant and a grant from the U.S. Army Research Office.

It will be shown that $\mathscr{C}$ is a fairly large class of functions. In particular, it contains every function f with a Lipschitz continuous derivative (cf. § 5).

In this article, we will establish the following results:

THEOREM 1. *Let $f \in \mathscr{C}$ satisfy $Q_n(f) = 0$ for all $n = 1, 2, \cdots$. Then f is an odd function on $[-1, 1]$.*

THEOREM 1′. *Let f be differentiable on $[-1, 1]$ such that $f, f' \in \mathscr{C}$. If $Q_n(f) = Q_n(f') = 0$ for all $n = 1, 2, \cdots$, then f is the zero function.*

THEOREM 2. *Let $f \in \mathscr{C}$ be an even function on $[-1, 1]$. If $R_n(f) = 0$ for all $n > N$, then f is a polynomial with degree at most $2N$.*

THEOREM 2′. *Let f be differentiable on $[-1, 1]$ such that $f, f' \in \mathscr{C}$. If $R_n(f) = R_n(f') = 0$ for all $n > N$, then f is a polynomial with degree at most $2N + 1$.*

Two examples will be given to show that the above results are sharp in the sense that both (4) and (4′) are needed in the hypotheses.

In order to give a formula to recapture a function from its Gauss–Chebyshev quadratures $Q_n(f)$, we need the following definition.

DEFINITION 2. Let ν be a function defined on the set of positive integers N as follows:

$$\nu(1) = 1,$$
$$\sum_{d|n} \nu(d)(-1)^{n/d} = 0 \quad \text{if } n > 1. \tag{5}$$

Hence, $\nu(1) = 1$, $\nu(2) = 1$, $\nu(3) = -1$, $\nu(4) = 2$, etc.

THEOREM 3. *For $n = 0, 1, 2, \cdots$, let p_{2n} be even polynomials defined as follows:*

$$p_{2n}(x) = -\frac{1}{\pi} \sum_{k|n} \nu\left(\frac{n}{k}\right) T_{2k}(x). \tag{6}$$

Then

$$Q_m(p_{2n}) = \delta_{m,n}$$

for $m, n = 1, 2, \cdots$.

Theorem 3 shows that *all* the conditions $Q_n(f) = 0$, $n = 1, 2, \cdots$, in Theorem 1 are required to guarantee that f is an odd function. The following result shows that an even function can be reconstructed from its Gauss–Chebyshev quadratures.

THEOREM 4. *Let $\{\alpha_n\}$ be a sequence of real numbers converging to α_0 such that $\alpha_n - \alpha_0 = O(1/n^{1+\varepsilon})$ for some $\varepsilon > 0$. Then the polynomial series*

$$\alpha_0 + \sum_{n=1}^{\infty} (\alpha_n - \alpha_0) p_{2n}(x) \tag{7}$$

converges uniformly on $[-1, 1]$ to a function f. Moreover, $Q_n(f) = \alpha_n$ for $n = 1, 2, \cdots$.

It should be mentioned that somewhat similar but different results were obtained in [2], [3], [4], [7].

2. Preliminary lemmas. In order to establish the above results, we need several lemmas.

LEMMA 1. *For $n, k = 1, 2, \cdots$,*

$$Q_n(T_k) = \begin{cases} 0 & \text{if } 2n \nmid k, \\ (-1)^l \pi & \text{if } k = 2nl. \end{cases}$$

Proof. We note that

$$Q_n(T_k)=\frac{\pi}{n}\sum_{j=1}^{n} T_k(x_{n,j})=\frac{\pi}{n}\sum_{j=1}^{n}\cos\frac{k(2j-1)\pi}{2n}$$
$$=\operatorname{Re}\left\{\frac{\pi}{n}\sum_{j=1}^{n} e^{ijk\pi/n}\cdot e^{-ik\pi/(2n)}\right\}.$$

Hence, if $k=2nl$, then

$$Q_n(T_k)=\operatorname{Re}\left\{\frac{\pi}{n}\sum_{j=1}^{n} e^{i2\pi lj}\cdot e^{-i\pi l}\right\}=(-1)^l\pi,$$

and if $2n \nmid k$, then $e^{ik\pi/n}\neq 1$, so that

$$\sum_{j=1}^{n} e^{ijk\pi/n}=e^{ik\pi/n}\frac{1-e^{ik\pi}}{1-e^{ik\pi/n}}=\frac{1-(-1)^k}{e^{-ik\pi/n}-1},$$

and

$$Q_n(T_k)=\operatorname{Re}\left\{\frac{\pi}{n}\frac{1-(-1)^k}{e^{-ik\pi/(2n)}-e^{ik\pi/(2n)}}\right\}$$
$$=\operatorname{Re}\left\{\frac{\pi}{n}\frac{1-(-1)^k}{2\sin(k\pi/(2n))}i\right\}=0.$$

As a consequence, we see that if f has a uniformly convergent Chebyshev expansion

$$f(x)=\tfrac{1}{2}a_0+\sum_{k=1}^{\infty} a_kT_k(x)$$

on $[-1,1]$ where the coefficients are given by (3′) (for instance, if f is continuous and is of bounded variation on $[-1,1]$), then

$$(8)\qquad Q_n(f)=\frac{\pi}{2}a_0+\pi\sum_{k=1}^{\infty}(-1)^k a_{2nk},\qquad n=1,2,\cdots,$$

where we have convergence for each n. If, in addition, $Q_n(f)=0$ for infinitely many n, then $a_0=0$ since $Q_n(f)$ is a Riemann sum of the integral

$$a_0=\frac{2}{\pi}\int_{-1}^{1}(1-x^2)^{-1/2}f(x)\,dx.$$

That is, we have the following

LEMMA 2. *Let f have a uniformly convergent Chebyshev expansion* (3) *on* $[-1,1]$. *If $Q_n(f)=0$ for infinitely many n, then $a_0=0$.*

Hence, the conditions $Q_n(f)=0$, $n=1,2,\cdots$, give the following system of linear equations:

$$(9)\qquad C_n\equiv\sum_{k=1}^{\infty}(-1)^k c_{nk}=0,\qquad n=1,2,\cdots,$$

where $c_n=a_{2n}$.

LEMMA 3. *Let C_n be defined as in* (9) *with $\sum|c_k|<\infty$. Suppose that $C_n=0$ for $n=1,2,\cdots$. Then*

$$(10)\qquad c_i=2c_{2i}\quad \textit{for } i=1,2,\cdots.$$

Proof. Throughout this paper, $\mu(n)$ will denote, as usual, the Möbius function; i.e.,

$$\mu(n)=\begin{cases} 1 & \text{if } n=1, \\ (-1)^k & \text{if } n=q_1\cdots q_k, \\ 0 & \text{if } p^2|n \text{ for some } p>1, \end{cases}$$

where $q_1,\cdots,q_k$ are distinct primes. (a,b) will denote the greatest common divisor of the positive integers a and b, and $p_1=2, p_2=3,\cdots,p_k,\cdots$ the primes in increasing order. Let $P_k=p_1\cdots p_k$. For each i, which we fix, let $b_d=c_{di}$ and $B_d=C_{di}$. Then

$$0=\sum_{d|P_k}\mu(d)B_d=\left\{\sum_{m=1}^{\infty}\sum_{d|(P_k,m)}\mu(d)(-1)^{m/d}\right\}b_m. \tag{11}$$

Let $(P_k,m)=u$ and $m=un$, and consider the following two cases.

Case (1), *n odd*. In this case, $(-1)^{m/d}=(-1)^{u/d}$ and the coefficient of b_m in (11) becomes $\sum_{d|u}\mu(d)(-1)^{u/d}$. If u is odd, then

$$\sum_{d|u}\mu(d)(-1)^{u/d}=-\sum_{d|u}\mu(d)=\begin{cases}-1 & \text{if } u=1,\\ 0 & \text{if } u>1\end{cases}$$

(cf. [6]). If u is even, then $u=2v$ where v must be odd since $u|P_k$, and hence, we have

$$\begin{aligned}\sum_{d|u}\mu(d)(-1)^{u/d}&=\sum_{d|v}\mu(d)+\sum_{d|v}\mu(2d)(-1)^{v/d}\\ &=2\sum_{d|v}\mu(d)=\begin{cases}2 & \text{if } v=1,\\ 0 & \text{if } v>1.\end{cases}\end{aligned}$$

Case (2), *n even*. In this case, m is even so that u must be even and $4|m$. This implies that m/d is even whenever $d|u$. Hence

$$\sum_{d|u}\mu(d)(-1)^{m/d}=\sum_{d|u}\mu(d)=\begin{cases}1 & \text{if } u=1,\\ 0 & \text{if } u>1.\end{cases}$$

But u is even in this case. Hence, the coefficient of b_m is zero whenever n is even.

Combining the above two cases, we can conclude from (11) that

$$0=\sum_{d|P_k}\mu(d)B_d=-b_1+2b_2+R_k,$$

where $|R_k|\leqq 2\sum_{m>P_k}|b_m|\leqq 2\sum_{m=p_k}^{\infty}|c_k|$. Hence, $R_k\to 0$ as k tends to infinity. That is, $b_1=2b_2$, or $c_i=2c_{2i}$.

We also have the following

LEMMA 4. *Let C_n be defined as in* (9) *with $\sum|c_k|<\infty$. Suppose that $C_n=0$ for $n=1,2,\cdots$. Then*

$$c_1=c_2+\sum_{i=1}^{\infty}c_{4i}. \tag{12}$$

Proof. The proof of this lemma is quite similar to the previous one except that we now use $\mu(2d)$ instead. Indeed,

$$0=\sum_{d|P_k}\mu(2d)C_d=\sum_{m=1}^{\infty}\left\{\sum_{d|u}\mu(2d)(-1)^{m/d}\right\}c_m, \tag{13}$$

where $u=(P_k,m)$ and we set $m=nu$.

Case (1), *n odd*. In this case, if u is also odd then

$$\sum_{d|u} \mu(2d)(-1)^{m/d} = \sum_{d|u} \mu(d) = \begin{cases} 1 & \text{if } u = 1, \\ 0 & \text{if } u > 1, \end{cases}$$

and if u is even, then $u = 2v$ where v is odd and

$$\sum_{d|u} \mu(2d)(-1)^{m/d} = \begin{cases} -1 & \text{if } v = 1, \\ 0 & \text{if } v > 1. \end{cases}$$

Case (2), *n even*. In this case both m and u are even and m must be divisible by 4. Setting $u = 2v$, v odd, we have

$$\sum_{d|u} \mu(2d)(-1)^{m/d} = \sum_{d|v} \mu(2d) + \sum_{d|v} \mu(4d)$$

$$= \sum_{d|v} \mu(2)\mu(d) = \begin{cases} -1 & \text{if } v = 1, \\ 0 & \text{if } v > 1. \end{cases}$$

Combining the above two cases, we have, from (13),

$$0 = \sum_{d|P_k} \mu(2d) C_d = c_1 - c_2 - \sum_{1 \leqq l \leqq p_k/4} c_{4l} + \tilde{R}_k$$

where $|\tilde{R}_k| \leqq \sum_{j > p_k} |c_j| \to 0$ as $k \to \infty$. This completes the proof of the lemma.

Combining Lemmas 3 and 4, we have the following main lemma, which is essential in our proof of Theorem 1.

LEMMA 5. *Let C_n be as defined in* (9) *with $\sum |c_k| < \infty$ and*

$$\sum_{k=1}^{\infty} |c_{2^n k}| = o(2^{-n}). \tag{14}$$

Suppose that $C_n = 0$ for $n = 1, 2, \cdots$. Then $c_n = 0$ for $n = 1, 2, \cdots$.

Proof. By combining (10) with $i = 1$ with (12), we have

$$c_1 = 2 \sum_{k=1}^{\infty} c_{4k}.$$

Solving this equation simultaneously with $0 = 2C_4$ gives

$$c_1 = 4 \sum_{k=1}^{\infty} c_{8k}.$$

Again combining this equation with $0 = 4C_8$ yields

$$c_1 = 8 \sum_{k=1}^{\infty} c_{16k},$$

etc. Hence, continuing this process, we can conclude, using the condition (14), that $c_1 = 0$. Next, for each fixed i, let $b_k = c_{ki}$, $k = 1, 2, \cdots$. The same proof above gives $b_1 = c_i = 0$. This completes the proof of the lemma.

We remark that Lemma 5 is sharp in the sense that $o(2^{-n})$ in (14) cannot be replaced $O(2^{-n})$. Indeed, if $c_{2^n} = 2^{-n}$, $n = 0, 1, \cdots$ and $c_m = 0$ for $m \neq 2^n$, then it is easy to verify that $C_n = 0$ for $n = 1, 2, \cdots$, and at the same time,

$$\sum_{k=1}^{\infty} |c_{2^n k}| = 2 \cdot 2^{-n},$$

$n = 0, 1, 2, \cdots$.

The following result gives a complete description of the function ν.

LEMMA 6. *The function $\nu(n)$ is multiplicative and its values are given by*

$$\begin{cases} \nu(n)=\mu(n) & \text{if } n \text{ is odd},\\ \nu(2^\alpha)=2^{\alpha-1} & \text{for } \alpha=1,2,\cdots. \end{cases} \tag{15}$$

Proof. From (5), it is known that $\nu(n)$ is the convolutive inverse of the function $(-1)^{n+1}$. If $(m, n)=1$, then one of the integers $(m+1)$ and $(n+1)$ is even and therefore we have

$$(m+1)(n+1)\equiv o(\text{mod } 2).$$

This gives $mn+1\equiv(m+1)+(n+1)(\text{mod } 2)$ and $(-1)^{mn+1}=(-1)^{m+1}(-1)^{n+1}$. That is, the function $(-1)^{n+1}$, and therefore its convolutive inverse $\nu(n)$, are multiplicative. Next, if n is odd, then the equations in (5) are exactly the defining equations of the Möbius functions $\mu(n)$, so that $\nu(n)=\mu(n)$. If $n=2^\alpha$, $\alpha=1,2,\cdots$, then we have, from (5),

$$\nu(2^\alpha)=\sum_{\beta=0}^{\alpha-1}\nu(2^\beta).$$

This gives, by induction, that $\nu(2^\alpha)=2^{\alpha-1}$.

If n is any positive integer, then $n=2^\alpha n^*$ where n^* is odd, so that $\nu(n)=\nu(2^\alpha)\nu(n^*)=2^{\alpha-1}\mu(n^*)$.

We next give a bound for the polynomials $p_{2n}(x)$ defined by (6).

LEMMA 7. *Let n be a positive integer with $n=2^\alpha n^*$ where n^* is odd. Then*

$$|p_{2n}(x)|\leqq\frac{2^a}{\pi}d(n^*) \tag{16}$$

for all $x\in[-1,1]$.

Here, and throughout, $d(n)$ denotes, as usual, the number of divisors of n (cf. [6]).

Proof. Since the Chebyshev polynomials are bounded by one on $[-1,1]$, we have, by using Lemma 6,

$$\begin{aligned}\max_{-1\leqq x\leqq 1}|p_{2n}(x)|&\leqq\frac{1}{\pi}\sum_{d|n}|\nu(d)|\leqq\frac{1}{\pi}\sum_{\beta=0}^{\alpha}\sum_{d|n^*}|\nu(2^\beta d)|\\&=\frac{1}{\pi}\left\{1+\sum_{\beta=1}^{\alpha}2^{\beta-1}\right\}\sum_{d|n^*}|\nu(d)|\\&=\frac{2^\alpha}{\pi}\sum_{d|n^*}|\mu(d)|\\&\leqq\frac{2^\alpha}{\pi}d(n^*).\end{aligned}$$

This completes the proof of the lemma.

3. Proofs of the theorems. Theorems 1 and 1′ follow immediately from the lemmas in § 2. Indeed, if $f\in\mathscr{C}$ is represented by the Chebyshev series (3), then $Q_n(f)=0$ for $n=1,2,\cdots$ gives $a_0=0$ by Lemma 2, and

$$\sum_{k=1}^{\infty}(-1)^k c_{nk}=0$$

$n=1,2,\cdots$, with $c_n=a_{2n}$. Also, since $f\in\mathscr{C}$, the sequence $\{c_n\}$ satisfies the hypotheses of Lemma 5, and is therefore the zero sequence. That is, $a_{2n}=0$ for all n and f is an odd function, proving Theorem 1. If, in addition, $Q_n(f')=0$ for all n, then both f and f' are odd functions, so that $f\equiv 0$, proving Theorem 1′.

To prove Theorem 2, we assume Theorem 3 which will be proved afterwards. Set

$$g(x)=\int_{-1}^{1} f(t)(1-t^2)^{-1/2}\,dt-f(x).$$

Then, by hypothesis, g is an even function in C and satisfies $Q_n(g)=R_n(f)=0$ for all $n>N$. Let $p(x)=\sum_{k=1}^{N} Q_k(g)p_{2k}(x)$, where p_{2k} are the polynomials defined in Theorem 3. Then p is an even polynomial with degree at most $2N$ such that

$$Q_n(p)=Q_n(g)$$

for $n=1,2,\cdots$. Hence, $Q_n(p-g)=0$, $n=1,2,\cdots$, so that $g\equiv p$ by Theorem 1. That is, f is a polynomial with degree at most $2N$.

If the hypotheses of Theorem 2′ are satisfied, then the even part f_e of f is a polynomial with degree $2N$ and the odd part f_0 of f is such that f_0' is a polynomial with degree $2N$ also. Hence, f is a polynomial with degree at most $2N+1$.

To prove Theorem 3, we consider

$$Q_m(p_{2n})=-\frac{1}{\pi}\sum_{k|n}\nu\left(\frac{n}{k}\right)Q_m(T_{2k}).$$

By Lemma 1, we have $Q_m(T_{2k})=0$ unless $k=ml$, in which case, we have $Q_m(T_{2k})=(-1)^l\pi$. Hence, if $m\nmid n$, there is no k with $m|k$ and $k|n$, so that $Q_m(p_{2n})=0$. If, on the other hand, $m|n$, then $n=mu$ and

$$Q_m(p_{2n})=-\frac{1}{\pi}\sum_{l|u}\nu\left(\frac{u}{l}\right)(-1)^l\pi$$

which is $\delta_{m,n}$ by (5).

To prove Theorem 4, we use the usual notation: $2^{\alpha}||n$ if and only if $2^{\alpha}|n$ but $2^{\alpha+1}\nmid n$. Hence, by (16), we have

$$\left|\sum_{n=N_1}^{N_2}(\alpha_n-\alpha_0)p_{2n}(x)\right|\leqq\sum_{\alpha=0}^{\infty}\sum_{\substack{2^{\alpha}||n\\ N_1\leqq n\leqq N_2}}|\alpha_n-\alpha_0|\,|p_{2n}(x)|$$
$$\leqq\sum_{\alpha=0}^{\infty}\sum_{n^*=N_1/2^{\alpha}}^{N_2/2^{\alpha}}\frac{C2^{\alpha}d(n^*)}{(2^{\alpha}n^*)^{1+\varepsilon}}$$

where each n^* is odd and $n=2^{\alpha}n^*$, and this yields

$$\left|\sum_{n=N_1}^{N_2}(\alpha_n-\alpha_0)p_{2n}(x)\right|\leqq C\sum_{\alpha=0}^{\infty}2^{-\alpha\varepsilon}\sum_{n=N_1}^{\infty}\frac{d(n)}{n^{1+\varepsilon}}.$$

But for each $\delta>0$, $d(n)=o(n^{\delta})$ (cf. [6]). Therefore, for every $\eta>0$, and N_1 sufficiently large, we have

$$\left|\sum_{n=N_1}^{N_2}(\alpha_n-\alpha_0)p_{2n}(x)\right|\leqq C\sum_{\alpha=0}^{\infty}2^{-\alpha\varepsilon}\eta=\kappa\eta.$$

This proves that the series (7) converges uniformly on $[-1,1]$ to some continuous function f. It is clear that $Q_n(f)=\alpha_n$, $n=1,2,\cdots$, from Theorem 3.

4. Sharpness of the results. We will now give two examples to indicate the sharpness of Theorem 1. These examples can be easily modified to show that Theorems 1′, 2, and 2′ are also sharp in the same sense.

Example 1. Let f be an even function on $[-1, 1]$ defined by

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{2^n} T_{2^{n+1}}(x).$$

Since $|T_j(x)| \leqq 1$ on $[-1, 1]$ for all j, it is clear that the Chebyshev series converges uniformly and $f \in C[-1, 1]$. Furthermore, $Q_n(f) = 0$ for all $n = 1, 2, \cdots$. Note that f satisfies (4) but not (4′).

Example 2. Let g be an even function on $[-1, 1]$ defined by

$$g(x) = \sum_{n=1}^{\infty} \frac{\mu(2n)}{n} T_{2n}(x). \tag{17}$$

It will be shown below that this Chebyshev series also converges uniformly on $[-1, 1]$, so that $g \in C[-1, 1]$. Furthermore, for each n,

$$\begin{aligned} Q_n(g) &= \frac{\pi}{n} \sum_{k=1}^{\infty} (-1)^k \frac{\mu(2kn)}{k} \\ &= \frac{\pi}{n} \sum_{(k,2n)=1} (-1)^k \frac{\mu(2kn)}{k} \\ &= -\frac{\pi}{n} \mu(2n) \sum_{(k,2n)=1} \frac{\mu(k)}{k} = 0 \end{aligned}$$

where the last equality was proved in [1]. Note that g satisfies (4′) but not (4).

To prove the uniform convergence of the series, we recall a result of Davenport [5]:

$$\max_{\theta} \left| \sum_{k=1}^{n} \mu(2k) e^{i2k\theta} \right| \leqq Cn/(\log n)^2$$

where C is an absolute constant. Hence, by partial summation, we have

$$\begin{aligned} \left| \sum_{j=n+1}^{m} \frac{\mu(2j)}{j} T_{2j}(x) \right| &= \left| \sum_{j=n+1}^{m} \frac{\mu(2j)}{j} \cos 2j\theta \right| \\ &= \left| \operatorname{Re} \left\{ \frac{1}{m+1} \sum_{j=1}^{m} \mu(2j) e^{i2j\theta} - \frac{1}{n} \sum_{j=1}^{n} \mu(2j) e^{i2j\theta} \right. \right. \\ &\qquad \left. \left. + \sum_{j=n}^{m} \frac{1}{j(j+1)} \sum_{k=1}^{j} \mu(2k) e^{i2k\theta} \right\} \right| \\ &\leqq C \left\{ \frac{m}{(m+1)(\log m)^2} + \frac{n}{n(\log n)^2} \geqq \sum_{j=n}^{m} \frac{1}{j(\log j)^2} \right\} \end{aligned}$$

for all $x \in [-1, 1]$. Hence, the Chebyshev series converges uniformly on $[-1, 1]$ to g.

5. The class $\mathscr{C}$. We will show that the class $\mathscr{C}$ is fairly large.

DEFINITION. Let $\alpha > 0$. $\Omega_\alpha = \Omega_\alpha[-1, 1]$ will denote the class of all functions f on $[-1, 1]$ such that

$$\omega(\delta; f) = O\left\{ \frac{1}{(\log \delta)^\alpha} \right\},$$

where $\omega(\delta; f)$, $\delta > 0$, denotes the modulus of continuity of f.

Hence, $\Omega_\alpha \supset \Lambda_\beta$ for any $\beta > 0$, where Λ_β is the usual Lipschitz class (cf. [8]).

PROPOSITION 1. *Let f be differentiable on $[-1, 1]$ such that $f' \in \Omega_\alpha$ for some $\alpha > 1$. Then $f \in \mathscr{C}$.*

Proof. By using standard techniques, we have

$$\begin{aligned} a_n &= \frac{2}{\pi} \int_{-1}^{1} (1-x^2)^{-1/2} f(x) T_n(x)\, dx \\ &= \frac{2}{\pi} \int_0^{\pi} f(\cos t) \cos nt\, dt \\ &= \frac{-1}{\pi n} \int_{-\pi}^{\pi} f'(\cos t) \sin t \sin nt\, dt \\ &= \frac{-1}{\pi n} \int_{-\pi}^{\pi} f'\left(\cos\left(t+\frac{\pi}{n}\right)\right) \sin\left(t+\frac{\pi}{n}\right) \sin nt\, dt, \end{aligned}$$

so that

$$2\pi n a_n = \int_{-\pi}^{\pi} \left[f'(\cos t)\sin t - f'\left(\cos\left(t+\frac{\pi}{n}\right)\right)\sin\left(t+\frac{\pi}{n}\right)\right] \sin nt\, dt.$$

Hence,

$$\begin{aligned} |2\pi n a_n| &\leqq \int_{-\pi}^{\pi} \left|f'(\cos t)\sin t - f'\left(\cos\left(t+\frac{\pi}{n}\right)\right)\sin\left(t+\frac{\pi}{n}\right)\right| dt \\ &\leqq \int_{-\pi}^{\pi} |f'(\cos t)| \left|\sin t - \sin\left(t+\frac{\pi}{n}\right)\right| dt \\ &\qquad + \int_{-\pi}^{\pi} \left|f'\left(\cos\left(t+\frac{\pi}{n}\right)\right) - f'(\cos t)\right| \left|\sin\left(t+\frac{\pi}{n}\right)\right| dt. \end{aligned}$$

The first term on the right is of order $O(1/n)$, while the second term is equal to

$$\begin{aligned} &\int_{-\pi}^{\pi} \left|f'(\cos t) - f'\left(\cos\left(t-\frac{\pi}{n}\right)\right)\right| |\sin t|\, dt \\ &\quad= \int_0^{\pi} \left|f'(\cos t) - f'\left(\cos\left(t-\frac{\pi}{n}\right)\right)\right| \sin t\, dt \\ &\qquad + \int_0^{\pi} \left|f'(\cos t) - f'\left(\cos\left(t+\frac{\pi}{n}\right)\right)\right| \sin t\, dt \\ &\quad= \int_{-1}^{1} \left|f'(x) - f'\left(x\cos\frac{\pi}{n} + \sqrt{1-x^2}\sin\frac{\pi}{n}\right)\right| dx \\ &\qquad + \int_{-1}^{1} \left|f'(x) - f'\left(x\cos\frac{\pi}{n} - \sqrt{1-x^2}\sin\frac{\pi}{n}\right)\right| dx. \end{aligned}$$

Clearly, each of the terms on the right has the same order, so that it is sufficient to consider one of them. But

$$\int_{-1}^{1}\left|f'(x)-f'\left(x\cos\frac{\pi}{n}+\sqrt{1-x^2}\sin\frac{\pi}{n}\right)\right|dx$$

$$\leqq\int_{-1}^{1}\left|f'(x)-f'\left(x\cos\frac{\pi}{n}\right)\right|dx$$

$$+\int_{-1}^{1}\left|f'\left(x\cos\frac{\pi}{n}\right)-f'\left(x\cos\frac{\pi}{n}+\sqrt{1-x^2}\sin\frac{\pi}{n}\right)\right|dx$$

$$\leqq C_1\Big/\left|\log\left(1-\cos\frac{\pi}{n}\right)\right|^{\alpha}+C_2\Big/\left|\log\left(\sin\frac{\pi}{n}\right)\right|^{\alpha}=O\left(\frac{1}{(\log n)^{\alpha}}\right).$$

Therefore, we have

$$|a_n|=O\left(\frac{1}{n(\log n)^{\alpha}}\right),\qquad \alpha>1,$$

so that condition (4) is satisfied. Also,

$$N\sum_{k=1}^{\infty}|a_{Nk}|\leqq CN\sum_{k=1}^{\infty}\frac{1}{Nk(\log N+\log k)^{\alpha}}$$

$$=C\sum_{k=1}^{\infty}\frac{1}{k(\log N+\log k)^{\alpha}}\to 0$$

by the dominated convergence theorem. In particular, condition (4′) is satisfied and $f\in\mathscr{C}$.

6. Final remarks. The problems considered in this paper could be asked for any Gaussian quadrature formula, q_n. For instance, if f belongs to some class $\mathscr{D}$ of "sufficiently smooth" functions and $q_n(f)=q_n(f')=0$ for $n=1,2,\cdots$, is f necessarily the zero function? It is not difficult to show that the answer is affirmative if $\mathscr{D}$ is the class of all polynomials. However, the methods we use in this paper depend very heavily on the structure introduced by the Gauss–Chebyshev formulae and do not apply to the general case.

REFERENCES

[1] G. R. BLAKLEY, I. BOROSH AND C. K. CHUI, *A two-dimensional mean problem*, J. Approximation Theory, 22 (1978), pp. 11–26.

[2] C. H. CHING AND C. K. CHUI, *Analytic functions characterized by their names on an arc*, Trans. Amer. Math. Soc., 186 (1973), pp. 175–183.

[3] ———, *Recapturing a holomorphic function on an annulus from its mean boundary values*, Proc. Amer. Math. Soc., 41 (1973), pp. 120–126.

[4] C. K. CHUI AND P. W. SMITH, *Characterization of a function by a certain series it generates*, Pacific J. Math., 53 (1974), pp. 363–371.

[5] H. DAVENPORT, *On some infinite series involving arithmetical functions*, Quart. J. Math., 8 (1937), pp. 8–13 and pp. 313–320.

[6] G. H. HARDY AND E. M. WRIGHT, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, 1954.

[7] I. KÁTAI, *On a transformation of power series*, Anal. Math., 2 (1976), pp. 3–10.

[8] A. ZYGMUND, *Trigonometric Series*, vol. 1, 2nd ed., Cambridge University Press, New York, 1959.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0011 $01.00/0

SINGULAR PERTURBATION OF AUTONOMOUS LINEAR SYSTEMS*

STEPHEN L. CAMPBELL† AND NICHOLAS J. ROSE†

Abstract. Let $X_\varepsilon(t) = \exp((A + B/\varepsilon)t)$ where A, B are $n \times n$ matrices. It is shown that $X_\varepsilon(t)$ converges pointwise for $t > 0$ as $\varepsilon \to 0^+$ if and only if Index $B \leqq 1$ and the nonzero eigenvalues of B have negative real part. An explicit representation of the limit of $X_\varepsilon(t)$ is given. These results are applied to the singularly perturbed system $\dot{x} = A_1(\varepsilon)x + A_2(\varepsilon)y$, $\varepsilon\dot{y} = B_1(\varepsilon)x + B_2(\varepsilon)y$. This paper differs from earlier work both in the derivation of necessary and sufficient conditions and in the explicit forms for the limits.

1. Introduction. The motivation for this paper is the study of the singularly perturbed autonomous system of differential equations

$$\begin{aligned} \dot{x} &= A_1(\varepsilon)x + A_2(\varepsilon)y, \\ \varepsilon\dot{y} &= B_1(\varepsilon)x + B_2(\varepsilon)y \end{aligned} \tag{1}$$

where $A_i(\varepsilon)$, $B_i(\varepsilon)$ are matrices, x and y are column vectors and $\varepsilon > 0$. Such systems arise, for example, in the analysis of linear dynamical systems [8] where the parameter ε may represent various "small" quantities (mass, time constants, etc.). When these small quantities are neglected ($\varepsilon = 0$), we obtain the reduced system

$$\begin{aligned} \dot{x} &= A_1(0)x + A_2(0)y, \\ 0 &= B_1(0)x + B_2(0)y. \end{aligned} \tag{2}$$

Sufficient conditions are known [11] for solutions of (1) to converge as $\varepsilon \to 0^+$ to a solution of (2) for $t > 0$. We shall present both necessary and sufficient conditions for such convergence and in addition obtain an explicit formula for the limit.

Equation (1) may be rewritten

$$\frac{d}{dt}\begin{bmatrix} x \\ y \end{bmatrix} = (A(\varepsilon) + B(\varepsilon)/\varepsilon)\begin{bmatrix} x \\ y \end{bmatrix}, \qquad \varepsilon > 0 \tag{3}$$

where

$$A(\varepsilon) = \begin{bmatrix} A_1(\varepsilon) & A_2(\varepsilon) \\ 0 & 0 \end{bmatrix}, \qquad B(\varepsilon) = \begin{bmatrix} 0 & 0 \\ B_1(\varepsilon) & B_2(\varepsilon) \end{bmatrix}. \tag{4}$$

The fundamental solution of (3) is

$$X_\varepsilon(t) = e^{(A(\varepsilon) + B(\varepsilon)/\varepsilon)t}. \tag{5}$$

Thus our problem may be reformulated as follows: to determine necessary and sufficient conditions for (5) to converge as $\varepsilon \to 0^+$ for $t > 0$ and to find an explicit formula for the limit. In a later paper, we shall present an asymptotic expansion of (5).

The calculation of limits similar to (5) also occur in Ellis and Pinsky's study of the Navier–Stokes equation [5], [6]. In [3] these limits are analyzed using the techniques of this paper.

2. Terminology. All matrices are over the complex field. If A is an $n \times n$ matrix, then Index A is the least nonnegative integer k such that rank $(A^{k+1}) =$ rank (A^k) (by definition $0^k = I$, the identity matrix).

* Received by the editors February 28, 1977, and in final revised form February 13, 1978.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607.

If Index $A = k$, then it follows from the Jordan form of A that

$$A = T^{-1}\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix} T \tag{6}$$

where C is nonsingular and N is nilpotent of index k. If $k = 0$, N is absent in (6) and A is nonsingular; if $k = n$, C is absent and A is nilpotent. The Drazin inverse of A, denoted by A^D, is given by

$$A^D = T^{-1}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} T. \tag{7}$$

The properties $AA^D = A^DA$, $A^DAA^D = A^D$, $A^{l+1}A^D = A^l$ for $l \geqq \text{Index}\, A$ easily follow (these properties uniquely determine A^D). If A has index 1 ($N = 0$ in (6)), then A^D is sometimes denoted by $A^\#$ (the group inverse of A) and has the additional property that $A^\# AA^\# = A$. If Index $A = 1$, the number of linearly independent eigenvectors corresponding to the eigenvalue zero is equal to the algebraic multiplicity of the zero eigenvalue.

Recall that a matrix A is called *stable* if all eigenvalues have negative real part. A is stable if and only if $e^{At} \to 0$ as $t \to \infty$. We need a generalization of this concept: a matrix A is called *semistable* if A has index 0 or 1 and all nonzero eigenvalues have negative real part. It follows from (6) that A is semistable if and only if e^{At} converges as $t \to \infty$.

For vectors x, y we use the Euclidean inner product $\langle x, y\rangle = y^*x$. The norm of x is $\|x\| = \sqrt{\langle x, x\rangle}$; the norm of a matrix A is $\|A\| = \sup\{\|Ax\| : \|x\| = 1\}$; the numerical range of A is $W(A) = \{\langle Ax, x\rangle : \|x\| = 1\}$. The set of eigenvalues of A is denoted by $\sigma(A)$.

Re, Im refer to real and imaginary parts of complex numbers. If Σ is a set of complex numbers then $\operatorname{Re}\Sigma = \{\operatorname{Re}\lambda : \lambda \in \Sigma\}$. For sets of complex numbers Σ, Σ' we let $\rho(\Sigma, \Sigma') = \inf\{|\alpha - \alpha'| : \alpha \in \Sigma, \alpha' \in \Sigma'\}$.

The following inequality will be needed:

$$0 < \rho(\lambda, W(A)) \leqq \|(\lambda - A)^{-1}\| \quad \text{for } \lambda \notin W(A). \tag{8}$$

To prove (8), we note that $W(A)$ is a compact set. Thus

$$\begin{aligned} 0 < \rho(\lambda, W(A)) &= \inf\{|\lambda - \mu| : \mu \in W(A)\} \\ &= \inf\{|\langle \lambda x - Ax, x\rangle| : \|x\| = 1\} \end{aligned}$$

since the continuous function $f(\mu) = |\lambda - \mu|$ achieves its minimum on $W(A)$. Using the Schwarz inequality we get

$$\rho(\lambda, W(A)) \leqq \inf\{\|(\lambda - A)x\| : \|x\| = 1\}.$$

Inequality (8) now follows from Theorem 6.5.1 of [9].

3. Main result. For notational convenience, it is easiest to study (5) with A, B independent of ε. The more general case will follow quickly. Our main result is:

THEOREM 1. *Suppose that A, B are $n \times n$ matrices. Let $X_\varepsilon(t) = e^{(A+B/\varepsilon)t}$. Then $X_\varepsilon(t)$ converges pointwise as $\varepsilon \to 0^+$ for $t > 0$, if and only if B is semistable. If B is semistable, then*

$$\lim_{\varepsilon \to 0^+} e^{(A+B/\varepsilon)t} = e^{(I-BB^D)At}(I - BB^D). \tag{9}$$

The proof of Theorem 1 will be given in §§ 4 and 5.

Example. Let

$$A\begin{bmatrix} 0 & 0 \\ 0 & 2\pi i \end{bmatrix}, \qquad B=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then Index $B=2$ so that B is not semistable. By Theorem 1 the limit (9) will fail to exist for all $t \geqq 0$. However,

$$e^{(A+B/\varepsilon)t}=\begin{bmatrix} 1 & [e^{2\pi it}-1]/(2\pi i\varepsilon) \\ 0 & e^{2\pi it} \end{bmatrix}$$

so that (9) does hold if t is an integer. In studying (1), however, it is of interest to know when (9) exists for all $t>0$ and the semistability of B is necessary.

4. Sufficiency of semistability. In this section we prove

THEOREM 2. *If B is semistable,*

$$\lim e^{(A+B/\varepsilon)}=e^{(I-BB^D)A}(I-BB^D)=(I-BB^D)\,e^{A(I-BB^D)}. \tag{10}$$

Proof. Suppose B is semistable. If $B=0$, then (10) is immediate. If B is invertible, then (10) is known; however, it may also be proved by the same technique we use. Assume then that $0\in\sigma(B)$ and $B\neq 0$.

First we show that without loss of generality, we may assume that A and B in (10) are:

$$B=\begin{bmatrix} B_{11} & 0 \\ 0 & 0 \end{bmatrix}, \qquad A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{11}$$

where B_{11} is nonsingular and Re $W(B_{11})<-\beta$ for some $\beta>0$. To prove this, we note that if T is any nonsingular matrix (independent of ε), then

$$e^{(A+B/\varepsilon)}=T^{-1}\,e^{\{TAT^{-1}+(TBT^{-1})/\varepsilon\}}T.$$

It follows that a simultaneous similarity may be applied to A and B without affecting our results.

Assume then that B is already in Jordan form, $B=\operatorname{diag}(\tilde{B}_{11},0)$, where $\tilde{B}_{11}$ is nonsingular and has its "ones" on the subdiagonal. Apply a similarity to B using $T=\operatorname{diag}(\alpha,\alpha^2,\cdots,\alpha^n)$ where $\alpha>0$. Then B is similar to $\operatorname{diag}(B_{11},0)$ where B_{11} is the same as $\tilde{B}_{11}$ except that the subdiagonal of B_{11} contains an α in each place that $\tilde{B}_{11}$ has a 1. Since Re $\sigma(B_{11})<0$, it is easy to see that, by taking α sufficiently small, we may insure that Re $W(B_{11})<-\beta$ for some $\beta>0$.

To calculate the limit (10) we shall use the Cauchy integral formula for matrices

$$e^{(A+B/\varepsilon)}=\frac{1}{2\pi i}\int_{\mathscr{C}(\varepsilon)} e^{\lambda}(\lambda-A-B/\varepsilon)^{-1}\,d\lambda \tag{12}$$

where $\mathscr{C}(\varepsilon)$ is a contour containing $\sigma(A+B/\varepsilon)$ in its interior. Therefore, it is necessary to obtain information about $\sigma(A+B/\varepsilon)$. The needed information is contained in the following two lemmas.

LEMMA 1. *For $\varepsilon>0$ and A, B given by* (11),

$$\sigma(A+B/\varepsilon)\subseteq G_0\cup G_1(\varepsilon)$$

where

$$G_0=\{z:\|(z-A_{22})^{-1}\|^{-1}\leqq\|A_{21}\|\},$$
$$G_1(\varepsilon)=\{z:\|(z-A_{11}-B_{11}/\varepsilon)^{-1}\|^{-1}\leqq\|A_{12}\|\}.$$

Proof. This is just the Gerschgorin theorem for block matrices [7]. □

Let $\varepsilon_0>0$ be picked so that, with β determined above,

$$\frac{\beta}{\varepsilon_0}>\|A_{11}\|+\|A_{12}\|+\|A_{21}\|+\|A_{22}\|+3.$$

LEMMA 2. *For $0<\varepsilon\leqq\varepsilon_0$, there exist two circles $\mathscr{C}_0$ and $\mathscr{C}_1(\varepsilon)$ such that G_0 is contained in the interior of $\mathscr{C}_0$ and $G_1(\varepsilon)$ is contained in the interior of $\mathscr{C}_1(\varepsilon)$. Furthermore, $\rho(\mathscr{C}_0,\mathscr{C}_1(\varepsilon))\geqq 1$ and* $\mathrm{Re}\,\mathscr{C}_1(\varepsilon)<-\beta/\varepsilon$.

Proof. Since $\mathrm{Re}\,W(B_{11})<-\beta$ for some $\beta>0$, it follows that $\mathrm{Re}\,W(B_{11}/\varepsilon)<-\beta/\varepsilon$. Thus there exists $\gamma>0$ so that the circle $\tilde{\mathscr{C}}_1(\varepsilon)$ with center at $(-(\gamma+\beta)/\varepsilon,0)$ and radius γ/ε contains $W(B_{11}/\varepsilon)$ in its interior. Now using (8) and the triangle inequality it follows that for $\mu\in G_1(\varepsilon)$

$$\rho(\mu,W(B_{11}/\varepsilon))\leqq\|A_{11}\|+\|A_{12}\|.$$

Therefore the circle $\mathscr{C}_1(\varepsilon)$ with center at $(-(\gamma+\beta)/\varepsilon,0)$ and radius $r_1(\varepsilon)=\gamma/\varepsilon+\|A_{11}\|+\|A_{12}\|+1$ contains $G_1(\varepsilon)$ in its interior and for $0<\varepsilon\leqq\varepsilon_0$, $\mathrm{Re}\,\mathscr{C}_1(\varepsilon)<-\beta/\varepsilon$.

Similarly one can show that for $\mu\in G_0$, $|\mu|<\|A_{21}\|+\|A_{22}\|$. Therefore, the circle $\mathscr{C}_0$ with center at $(0,0)$ and radius $r_0=\|A_{12}\|+\|A_{22}\|+1$ contains G_0 in its interior. Also we have

$$\rho(\mathscr{C}_0,\mathscr{C}_1(\varepsilon))=(\gamma+\beta)/\varepsilon-r_1(\varepsilon)-r_0\geqq 1. \qquad \square$$

For the rest of the proof, assume that $0<\varepsilon\leqq\varepsilon_0$. We now proceed to establish (10). From (12) and Lemmas 1, 2 we have

$$e^{(A+B/\varepsilon)}=I_0(\varepsilon)+I_1(\varepsilon) \tag{13}$$

where

$$I_0(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_0}e^{\lambda}(\lambda-A-B/\varepsilon)^{-1}\,d\lambda, \tag{14}$$

$$I_1(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_1(\varepsilon)}e^{\lambda}(\lambda-A-B/\varepsilon)^{-1}\,d\lambda. \tag{15}$$

Consider (15). Let $\zeta=\varepsilon\lambda$ to obtain

$$I_1(\varepsilon)=\frac{1}{2\pi i}\int_{\hat{\mathscr{C}}_1(\varepsilon)}e^{\zeta/\varepsilon}(\zeta-\varepsilon A-B)^{-1}\,d\zeta \tag{16}$$

where $\hat{\mathscr{C}}_1(\varepsilon)$ is the circle with center at $(-(\gamma+\beta),0)$ and radius $\gamma+\varepsilon k$, $k=\|A_{11}\|+\|A_{12}\|+1$. Thus

$$\|I_1(\varepsilon)\|\leqq\frac{1}{2\pi}\cdot 2\pi(\gamma+\varepsilon k)\cdot e^{-\beta/\varepsilon}\cdot\sup\{\|(\zeta-\varepsilon A-B)^{-1}\|:\zeta\in\mathscr{C}_1^*(\varepsilon)\}. \tag{17}$$

Since $\|(\zeta-\varepsilon A-B)^{-1}\|\to\|(\zeta-B)\|^{-1}$ as $\varepsilon\to 0^+$, $\|(\zeta-\varepsilon A-B)^{-1}\|$ is continuous on the compact set $\hat{\mathscr{C}}_1(\varepsilon)\times[0,\varepsilon_0]$ and, therefore, attains its maximum on this set. Thus $\|I_1(\varepsilon)\|\leqq Me^{-\beta/\varepsilon}$ for some $M>0$ and $I_1(\varepsilon)\to 0$ as $\varepsilon\to 0^+$.

Finally we consider (14), which may be rewritten

$$I_0(\varepsilon)=e^{A+B/\varepsilon}P_0(\varepsilon) \tag{18}$$

where

$$P_0(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_0}(\lambda-A-B/\varepsilon)^{-1}\,d\lambda. \tag{19}$$

Since $P_0(\varepsilon)$ is a projection which commutes with $A+B/\varepsilon$, we have

$$I_0(\varepsilon)=e^{F_0(\varepsilon)}P_0(\varepsilon) \tag{20}$$

where

$$F_0(\varepsilon)=(A+B/\varepsilon)P_0(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_0}\lambda(\lambda-A-B/\varepsilon)^{-1}\,d\lambda. \tag{21}$$

In (19), let $\zeta=\varepsilon\lambda$ to obtain

$$P_0(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_0(\varepsilon)}(\zeta-\varepsilon A-B)^{-1}\,d\zeta=\frac{1}{2\pi i}\int_{\mathscr{C}_0(\varepsilon_0)}(\zeta-\varepsilon A-B)^{-1}\,d\zeta \tag{22}$$

where $\mathscr{C}_0(\varepsilon)$ is a circle with center at the origin and radius εr_0. Since $\mathscr{C}(\varepsilon_0)$ does not contain any of $W(B_{11})$ we have

$$\lim_{\varepsilon\to 0^+}P_0(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_0(\varepsilon_0)}(\zeta-B)^{-1}\,d\zeta. \tag{23}$$

It is easy to verify that if B has index 1, the following identity holds [12]

$$(\zeta-B)^{-1}=-(I-\zeta B^D)^{-1}B^D+\zeta^{-1}(I-BB^D). \tag{24}$$

Thus the integral in (23) may be evaluated using the residue theorem to obtain

$$\lim_{\varepsilon\to 0^+}P_0(\varepsilon)=(I-BB^D). \tag{25}$$

In (21), let $\zeta=\varepsilon\lambda$ to obtain,

$$\begin{aligned} F_0(\varepsilon)&=\frac{1}{2\pi i}\int_{\mathscr{C}_0(\varepsilon_0)}\frac{\zeta}{\varepsilon}(\zeta-\varepsilon A-B)^{-1}\,d\zeta\\ &=\frac{1}{2\pi i}\int_{\mathscr{C}_0(\varepsilon_0)}\frac{\zeta}{\varepsilon}\{(\zeta-\varepsilon A-B)^{-1}\varepsilon A(\zeta-B)^{-1}+(\zeta-B)^{-1}\}\,d\zeta.\end{aligned} \tag{26}$$

Thus

$$F_0(\varepsilon)=\frac{1}{2\pi i}\int_{\mathscr{C}_0}(\zeta-B)^{-1}A(\zeta-B)^{-1}\,d\zeta+\frac{1}{2\pi i}\int_{\mathscr{C}_0}\frac{\zeta}{\varepsilon}(\zeta-B)^{-1}\,d\zeta+O(\varepsilon).$$

From (24) we see that the integrand of the second integral is analytic inside $\mathscr{C}_0$ and the integral is zero. The first integral can be evaluated by residues using (24). Thus,

$$\lim_{\varepsilon\to 0^+}F_0(\varepsilon)=(I-BB^D)A(I-BB^D).$$

Therefore,

$$\lim_{\varepsilon\to 0^+}I(\varepsilon)=\lim_{\varepsilon\to 0^+}I_0(\varepsilon)=e^{(I-BB^D)A(I-BB^D)}(I-BB^D)=e^{(I-BB^D)A}(I-BB^D). \qquad \square$$

5. Necessity of semistability. This section will complete the proof of Theorem 1 by showing the necessity of semistability.

Completion of Proof of Theorem 1. Assume that $X_\varepsilon(t)$ has a pointwise limit as $\varepsilon \to 0^+$ for $t \geqq 0$. As observed in the proof of Theorem 2, we may assume that a similarity transformation has been performed on B so that

$$B = \begin{bmatrix} J_1 & 0 \cdots & & 0 \\ \vdots & & & \vdots \\ 0 & & J_r & 0 \\ 0 & \cdots 0 & & N \end{bmatrix}$$

where $\sigma(J_i) = \lambda_i$, $\sigma(N) = 0$, $\lambda_i \neq \lambda_j$ if $i \neq j$, $|\lambda_i| \neq 0$, $i = 1, \cdots, r$. In addition, we may assume that $W(J_i)$ is in the open left-half plane if $\operatorname{Re} \lambda_i < 0$, $W(J_i)$ is in the open right-half plane if $\operatorname{Re} \lambda_i > 0$ and $W(J_i) \cap W(J_j) = \phi$ if $i \neq j$, $i = 1, 2, \cdots, r$. Let $\lambda_0 = 0$ if the nilpotent block N is present. From [7] and the same reasoning used in the proof of Theorem 2, for $i = 1, 2, \cdots, r$, there exists $\lambda_\varepsilon(i) \in \sigma(A + B/\varepsilon)$ such that $|\lambda_\varepsilon(i) - \lambda_i/\varepsilon| \leqq K_i + \delta_i/\varepsilon$, where K_i and δ_i are constants.

First we show that $\operatorname{Re} \lambda \leqq 0$ for $\lambda \in \sigma(B)$. Suppose that $\lambda \in \sigma(B)$ and $\operatorname{Re} \lambda > 0$, then there exists $\lambda_\varepsilon \in \sigma(A + B/\varepsilon)$ such that $\operatorname{Re} \lambda_\varepsilon \to +\infty$. Let ϕ_ε be such that $(A + B/\varepsilon)\phi_\varepsilon = \lambda_\varepsilon \phi_\varepsilon$, $\|\phi_\varepsilon\| = 1$. Then $\|e^{(A+B/\varepsilon)}\phi_\varepsilon\| = \|e^{\lambda_\varepsilon}\phi_\varepsilon\| \to \infty$ which is a contradiction.

Some further calculations are necessary before we can rule out the possibility that $\operatorname{Re} \lambda_i = 0$, $i \neq 0$. Using [7], we can argue as in the proof of Theorem 2 that for ε less than some ε_0 there exist contours $\mathscr{C}_i(\varepsilon)$, $\mathscr{C}_0(\varepsilon)$ which do not intersect such that $W(J_i/\varepsilon_0) \subseteq$ Interior $(\mathscr{C}_i(\varepsilon))$, $W(N/\varepsilon) \subseteq$ Interior $(\mathscr{C}_0(\varepsilon))$, $\sigma(A + B/\varepsilon) \subseteq \cup_i$ Interior $(\mathscr{C}_i(\varepsilon))$, and $\rho(W(J_i/\varepsilon), \mathscr{C}_i(\varepsilon))$, $\rho(W(N/\varepsilon), \mathscr{C}_0(\varepsilon))$ are bounded independent of ε. Thus each $\mathscr{C}_i(\varepsilon)$ contains only those eigenvalues of $A + B/\varepsilon$ that are clustering 'near' those of J_i/ε. Then

$$X_\varepsilon(t) = \sum_{i=0}^{r} e^{F_i(\varepsilon)t} P_i(\varepsilon)$$

where

$$P_i(\varepsilon) = \frac{1}{2\pi i} \int_{\mathscr{C}_i(\varepsilon)} (\lambda - A - B/\varepsilon)^{-1}\, d\lambda,$$

$$F_i(\varepsilon) = (A + B/\varepsilon) P_0(\varepsilon) = \frac{1}{2\pi i} \int_{\mathscr{C}_i(\varepsilon)} \lambda (\lambda - A - B/\varepsilon)^{-1}\, d\lambda.$$

As in the proof of Theorem 2, it follows that $P_i(\varepsilon) \to I - (\lambda_i - B)^D(\lambda_i - B)$. Since $X_\varepsilon(t)$ has a limit for $t > 0$, so does $X_\varepsilon(t)P_i(\varepsilon)$. Thus $e^{F_i(\varepsilon)t}$ has a pointwise limit for $t > 0$ for each i.

We shall show that

$$F_i(\varepsilon) = \lambda_i P_i(\varepsilon)/\varepsilon + G_i(\varepsilon) + (\lambda_i - B)[I - (\lambda_i - B)^D(\lambda_i - B)]/\varepsilon, \tag{27}$$

where $G_i(\varepsilon)$ is continuous at zero.

Assume for the moment that (27) holds. We shall use this to show that we cannot have $\operatorname{Re} \lambda_i = 0$, $i \neq 0$, and that B has index 0 or 1.

Note that $P_i(\varepsilon)$ is indempotent and the last term in (27) is nilpotent, so that

$$\operatorname{Trace} F_i(\varepsilon) = \mu_i \lambda_i/\varepsilon + \operatorname{Trace} G_i(\varepsilon)$$

where μ_i is an integer. Assume that $\operatorname{Re}\lambda_i=0$ and $\operatorname{Im}\lambda_i\neq 0$. Then there exists a $\mu(\varepsilon)\in\sigma(F_i(\varepsilon))$ such that $\operatorname{Re}(\mu(\varepsilon))$ is bounded and $|\operatorname{Im}\mu(\varepsilon)|\to\infty$. Let $\phi(\varepsilon)$ be an eigenvector of $F_i(\varepsilon)$ corresponding to $\mu(\varepsilon)$ and assume $\|\phi(\varepsilon)\|=1$. Pick a subsequence $\phi(\varepsilon_k)$ so that $\phi(\varepsilon_k)$ converges; then

$$\langle e^{F_i(\varepsilon_k)t}\phi(\varepsilon_k),\phi(\varepsilon_k)\rangle=e^{\mu(\varepsilon_k)t}$$

converges for all $t>0$ which is a contradiction. Thus $\operatorname{Re}\lambda_i<0$, $i=1,\cdots,r$.

Now consider (27) for $i=0$. Since $\lambda_0=0$ we have $F_0(\varepsilon)=G_0(\varepsilon)+Q/\varepsilon$, where $Q=B(I-BB^D)$. Since $\operatorname{Trace}F_0(\varepsilon)=\operatorname{Trace}G_0(\varepsilon)$, we must have that $\sigma(F_0(\varepsilon))$ is bounded. Pick t_1 such that $|\operatorname{Im}\sigma((G_0(\varepsilon)+Q/\varepsilon)t_1)|\leqq\pi/2$. Let Ln be the principal branch of $\ln z$. Then if $e^{(G_0(\varepsilon)+Q/\varepsilon)t_1}\to\tilde{Q}$ we have

$$(G_0(\varepsilon)+Q/\varepsilon)t_1=\operatorname{Ln}e^{(G_0(\varepsilon)+Q/\varepsilon)t_1}\to\operatorname{Ln}\tilde{Q}.$$

However, if $Q\neq 0$, the left-hand side cannot possess a limit. Thus $Q=B(I-BB^D)=0$ and B has index zero or one.

Hence it suffices to show that (27) holds for Theorem 1 to be proven. To see (27), let $\hat{B}=B-\lambda_i$, and $\zeta=\varepsilon\lambda-\lambda_i$. Then,

$$\begin{aligned}(F_i(\varepsilon)-\lambda_i/\varepsilon)P_i(\varepsilon)&=\frac{1}{2\pi i}\int_{\mathscr{C}_i(\varepsilon)}(\lambda-\lambda_i/\varepsilon)(\lambda-A-B/\varepsilon)^{-1}\,d\lambda\\&=\frac{1}{2\pi i}\int_{\mathscr{C}_i(\varepsilon)}\frac{\zeta}{\varepsilon}(\zeta-\varepsilon A-\hat{B})^{-1}\,d\zeta,\end{aligned}$$

where $\mathscr{C}_i(\varepsilon)$ is a circle with center at the origin which does not contain any nonzero eigenvalues of $\hat{B}$. Hence as in the proof of Theorem 2:

$$F_i(\varepsilon)=\lambda_i/\varepsilon P_i(\varepsilon)+\frac{1}{2\pi i}\int_{\mathscr{C}_i}\zeta(\zeta-\varepsilon A-\hat{B})^{-1}A(\zeta-\hat{B})\,d\zeta+\frac{1}{2\pi i}\int_{\mathscr{C}_i}\frac{\zeta}{\varepsilon}(\zeta-\hat{B})^{-1}\,d\zeta,$$

The first integral is a continuous function of ε and the second can be evaluated to yield $\hat{B}(I-\hat{B}^D\hat{B})/\varepsilon$. Thus (27) follows. □

6. Generalizations. While A, B were assumed constant for notational convenience, the following more general theorem holds.

THEOREM 3. *Suppose that $A(\varepsilon)\to A_0$ as $\varepsilon\to 0^+$. Suppose that $B(\varepsilon)$ is right differentiable at $\varepsilon=0$. Let*

$$X_\varepsilon(t)=e^{(A(\varepsilon)+B(\varepsilon)/\varepsilon)t}.$$

Then $X_\varepsilon(t)$ converges pointwise for $0<t$ if and only if $B(0)$ is semistable. If $B(0)$ is semistable, then

(i) $\lim_{\varepsilon\to 0^+}X_\varepsilon(t)=e^{(I-B(0)B(0)^D)(A_0+B'(0))t}(I-B(0)B(0)^D)$,

(ii) *$X_\varepsilon(t)$ converges uniformly on compact subsets of* $(0,\infty)$,

(iii) *$X'_\varepsilon(t)$ converges uniformly on compact subsets of* $(0,\infty)$ *to*

$$(I-B(0)B(0)^D)(A_0+B'(0))\,e^{(I-B(0)B(0)^D)(A_0+B'(0))t}(I-B(0)B(0)^D).$$

Proof. It is clear from the proofs of Theorems 1, 2 that they go over immediately to $A(\varepsilon)\to A_0$. Suppose then that $B(\varepsilon)$ is differentiable at $\varepsilon=0$. Thus $B(\varepsilon)=B(0)+B'(0)\varepsilon+\phi(\varepsilon)$ where $\phi(\varepsilon)/\varepsilon\to 0$ as $\varepsilon\to 0$. Then, $A(\varepsilon)+B(\varepsilon)/\varepsilon=[A(\varepsilon)+B'(0)+\phi(\varepsilon)/\varepsilon]+B(0)/\varepsilon$ and $A(\varepsilon)+B'(0)+\phi(\varepsilon)/\varepsilon\to A(0)+B'(0)$ which is the case just discussed. There remains then only to verify (ii) and (iii). But if $B(0)$ is semistable, then,

$$X_\varepsilon(t)=e^{F_0(\varepsilon)t}P_0(\varepsilon)+e^{F_1(\varepsilon)t}P_1(\varepsilon)$$

where $\operatorname{Re}[\sigma(F_1(\varepsilon))\backslash\{0\}]<0$ and $F_0(\varepsilon)\to(I-B(0)^DB(0))A_0(I-B(0)^DB(0))$. Since $\|e^{F_1(\varepsilon)t}P_1(\varepsilon)\|=O(e^{-\beta t/\varepsilon})$ by (12), both (ii) and (iii) follow. □

7. An application. As an application of Theorem 3, consider the system

$$\begin{aligned}\dot{x}&=A_1(\varepsilon)x+A_2(\varepsilon)y,\\ \varepsilon\dot{y}&=B_1(\varepsilon)x+B_2(\varepsilon)y\end{aligned}\tag{28}$$

with the initial conditions $x(0)=x_0$, $y(0)=y_0$ and the corresponding reduced system

$$\dot{x}=A_1x+A_2y,\qquad 0=B_1x+B_2y\tag{29}$$

where $A_i(\varepsilon)\to A_i$, $B_i(\varepsilon)\to B_i$ for $i=1,2$ as $\varepsilon\to0^+$ and $B_1(\varepsilon)$, $B_2(\varepsilon)$ are right differentiable at $\varepsilon=0$.

THEOREM 4. *The solution* $[x_\varepsilon(t), y_\varepsilon(t)]$ *of* (28) *has a pointwise limit for* $t>0$ *for all* (x_0, y_0) *if and only if* $B=\begin{bmatrix}0&0\\B_1&B_2\end{bmatrix}$ *is semistable. B is semistable if and only if* B_2 *is semistable and* $B_2^DB_2B_1=B_1$. *If B is semistable, then* $[x_\varepsilon(t), y_\varepsilon(t)]$ *converges to a solution of the reduced problem* (29). *Let* $[x(t), y(t)]$ *be this limiting solution. Note that* $B^\#$ *then exists and equals* B^D. *Then*

$$\begin{bmatrix}x(t)\\y(t)\end{bmatrix}=e^{(I-B^\#B)\hat{A}(0)t}(I-B^\#B)\begin{bmatrix}x_0\\y_0\end{bmatrix}\tag{30}$$

where

$$\hat{A}(\varepsilon)=\begin{bmatrix}A_1(\varepsilon)&A_2(\varepsilon)\\B_1'(0)&B_2'(0)\end{bmatrix}.$$

If the $B_i(\varepsilon)$ *are constant, then* (30) *is*

$$\begin{bmatrix}x(t)\\y(t)\end{bmatrix}=\begin{bmatrix}e^{(A_1-A_2B_2^\#B_1)t}x(0)+\theta((A_1-A_2B_2^\#B_1)t)A_2(I-B_2B_2^\#)ty(0)\\-B_2^\#B_1x(t)+(I-B_2^\#B_2)y(0)\end{bmatrix}\tag{31}$$

where $\theta(z)=(e^z-1)/z$. *All limiting solutions satisfy the initial conditions*

$$B^\#B\begin{bmatrix}x(0)\\y(0)\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}.$$

Proof. Let

$$A(\varepsilon)=\begin{bmatrix}A_1(\varepsilon)&A_2(\varepsilon)\\0&0\end{bmatrix},\qquad B(\varepsilon)=\begin{bmatrix}0&0\\B_1(\varepsilon)&B_2(\varepsilon)\end{bmatrix}.$$

Then all solutions of (28) are of the form:

$$z_\varepsilon(t)=\begin{bmatrix}x_\varepsilon(t)\\y_\varepsilon(t)\end{bmatrix}=e^{(A(\varepsilon)+B(\varepsilon)/\varepsilon)t}\begin{bmatrix}x(0)\\y(0)\end{bmatrix}.$$

By Theorem 3, $z_\varepsilon(t)$ has a limit for all x_0, y_0 if and only if B is semistable. But $\sigma(B)=\sigma(B_2)\cup\{0\}$, and from [10],

$$\begin{bmatrix}0&0\\B_1&B_2\end{bmatrix}^D=\begin{bmatrix}0&0\\(B_2^D)^2B_1&B_2^D\end{bmatrix}.$$

Thus one can verify that B has index 1 if and only if B_2 has index 0 or 1 and $B_2^D B_2 B_1 = B_1$. Equation (30) follows from Theorem 3. We shall now show (31) holds. Suppose B has index 1 and the $B_i(\varepsilon)$ are constant. Then,

$$BB^{\#}=\begin{bmatrix} 0 & 0 \\ B_2^{\#}B_1 & B_2^{\#}B_2 \end{bmatrix}, \qquad A(I-BB^{\#})=\begin{bmatrix} A_1-A_2B_2^{\#}B_1 & A_2(I-B_2^{\#}B_2) \\ 0 & 0 \end{bmatrix}.$$

Hence,

$$e^{A(I-BB^{\#})t}=\begin{bmatrix} e^{(A_1-A_2B_2^{\#}B_1)t} & \phi(t)A_2(I-B_2^{\#}B_2)t \\ 0 & I \end{bmatrix}$$

where $\phi(t)=\theta((A_1-A_2B_2^{\#}B_2)t)$ and $\theta(z)$ has the power series expansion $(e^z-1)/z$. Thus,

$$(I-BB^{\#})\,e^{A(I-BB^{\#})t}=\begin{bmatrix} e^{(A_1-A_2B_2^{\#}B_1)t} & \phi(t)A_2[I-B_2^{\#}B_1]t \\ -B_2^{\#}B_1\,e^{(A_1-A_2B_2^{\#}B_1)t} & -B_2^{\#}B_1\phi(t)A_2(I-B_2^{\#}B_2)t+(I-B_2^{\#}B_2) \end{bmatrix}.$$

Note that given initial conditions x_0, y_0 for (28),

$$e^{(A(\varepsilon)+B(\varepsilon)/\varepsilon)t}BB^{\#}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \to 0 \quad \text{for } t>0$$

and

$$e^{(A(\varepsilon)+B(\varepsilon)/\varepsilon)t}(I-BB^{\#})\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \text{ converges.}$$

Hence to describe the limits of x_ε, y_ε, we may as well assume $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \in R(I-BB^{\#})$. Then, $B_2^{\#}B_2y_0=-B_2^{\#}B_1x_0$, x_0 arbitrary. Hence,

$$x(t)=e^{(A_1-A_2B_2^{\#}B_1)t}x_0+\phi(t)A_2t[(I-B_2^{\#}B_2)y_0],$$
$$y(t)=-B_2^{\#}B_1x(t)+(I-B_2^{\#}B_2)y_0$$

are the limiting solutions of (28). That (31) is a solution of (29) follows from the uniform convergence of $x_\varepsilon(t)$, $x'_\varepsilon(t)$, $y_\varepsilon(t)$, $y'_\varepsilon(t)$ on compact subsets of $(0,\infty)$. □

COROLLARY 1. *If B is semistable, then the solution of* (28) *can be written*

$$\begin{bmatrix} x_\varepsilon(t) \\ y_\varepsilon(t) \end{bmatrix} = e^{(I-BB^{\#})\hat{A}(0)t}(I-BB^{\#})\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + e^{(A(\varepsilon)+B(\varepsilon)/\varepsilon)t}BB^{\#}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + E(t)\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

where the first term is the solution of the reduced problem (29), *the second term is the "boundary layer correction" and $E(t)\to 0$ uniformly on* $[0, t_0]$ *for every $t_0>0$ as $\varepsilon \to 0^+$.*

It should be noted that previous work on (28)–(29) usually assumes that B_2 is invertible and has no purely imaginary eigenvalues. Finally from [4], we have

COROLLARY 2. *The fundamental matrix $Z_\varepsilon(t)$ of* (28) *converges pointwise for $t>0$ and solutions of the reduced problem* (29) *for consistent initial conditions are uniquely determined, if and only if, B_2 is stable.*

Acknowledgment. We wish to thank the referees and the editor for their many helpful comments.

REFERENCES

[1] S. L. CAMPBELL, *Optimal control of autonomous linear processes with singular matrices in the quadratic cost functional*, SIAM J. Control, 14 (1976), pp. 1092–1106.

[2] ———, *Linear systems of differential equations with singular coefficients*, this Journal, 8 (1977), pp. 1057–1066.

[3] ———, *On the limit of a product of matrix exponentials*, J. Linear and Multilinear Algebra, 6 (1978), pp. 55–59.

[4] S. L. CAMPBELL, C. D. MEYER, JR. AND N. J. ROSE, *Applications of the Drazin inverse to linear systems of differential equations*, SIAM J. Appl. Math., 31 (1976), pp. 411–425.

[5] R. S. ELLIS AND M. A. PINSKY, *Asymptotic nonuniqueness of the Navier Stokes equations in kinetic theory*, Bull. Amer. Math. Soc., 80 (1974), pp. 1160–1164.

[6] ———, *The projection of the Navier Stokes equations upon the Euler equations*, J. Math. Pures Appl., 54 (1975), pp. 157–182.

[7] A. FEINGOLD AND R. VARGA, *Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem*, Pacific J. Math., 12 (1962), pp. 1241–1250.

[8] P. V. KOKOTOVIC AND R. A. YACKEL, *Singular perturbation theory of linear state regulators*, IEEE Trans. Automatic Control, 17 (1972), pp. 29–37.

[9] P. LANCASTER, *Theory of Matrices*, Academic Press, New York, 1969.

[10] C. D. MEYER, JR. AND N. J. ROSE, *The index and the Drazin inverse of block triangular matrices*, SIAM J. Appl. Math., 33 (1976), pp. 1–7.

[11] R. E. O'MALLEY, JR., *Topics in singular perturbations*, Lectures on Ordinary Differential Equations, R. McKelvey, ed., Academic Press, New York, 1970, pp. 155–260.

[12] N. J. ROSE, *The Laurent expansion of a generalized resolvent with some applications*, SIAM J. Appl. Math., 9 (1978), pp. 751–758.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0012 $01.00/0

EXISTENCE FOR NONLINEAR VOLTERRA EQUATIONS IN HILBERT SPACES*

VIOREL BARBU†

Abstract. Existence results for Volterra equations of the form $Bu + a * Au \ni f$ are established. Here A and B are subdifferentials of convex functions on a real Hilbert space.

In particular, existence results for the nonlinear differential equation $(d/dt)Bu + Au \ni g$, $u(0) = u_0$ are derived.

1. Introduction. In this paper we are concerned with the nonlinear Volterra equation

$$Bu(t) + \int_0^t a(t-s)Au(s)\,ds \ni f(t), \qquad (0 < t < \infty). \tag{1.1}$$

Here the unknown u takes values in a real Hilbert space H, $a(t)$ is a scalar function and A, B are (possibly) multivalued maximal monotone operators acting in H.

When $a(t) \equiv 1$, equation (1.1) is formally equivalent to nonlinear differential equations of the form

$$\begin{gathered} \frac{d}{dt}Bu(t) + Au(t) \ni g(t), \qquad (0 < t < \infty) \\ u(0) = u_0 \end{gathered} \tag{1.2}$$

for which the standard existence theory of nonlinear differential equations of monotone type (see e.g. [2], [6], [8]) is not in general applicable.

The reader will certainly recognize that in applications, A and B will often take the form of a partial differential operator on a domain Ω of some Euclidean space (see Examples 1 and 2 below). Such equations arise very naturally in various problems of nonlinear viscoelasticity and heat conduction in material with memory.

Our approach when studying (1.1) consists in "regularizing" the equation by adding the term $\varepsilon u'$ to the left-hand side. After establishing a priori estimates independent of ε, existence results for equation (1.1) are obtained on letting $\varepsilon \to 0^+$. When one passes to the limit, the compactness assumptions and monotonicity of A and B play a crucial role.

The plan of the paper is the following: In § 2 the basic assumptions and the existence theorems are stated. The proofs are delivered in §§ 3, 4 and 5, respectively. Two examples illustrating the theory are presented in § 6.

To our knowledge, under the general conditions considered here, equation (1.1) has not been studied in literature. As for equation (1.2), there is an extensive literature treating the case in which A and B are linear but few results are known about the solvability of nonlinear equation (1.2), (see e.g. [4], [5], [7], [11], [12], [16]). Our results can be compared most closely with that of O. Grange and F. Mignot [12]. However, there is not a large overlap and the methods are quite different.

2. The main results. To begin with we list the notations and general assumptions which will be in effect in the sequel.

(a) H is a real Hilbert space and V, W are real reflexive Banach spaces dense in H. It is assumed that

$$V \subset W \subset H \subset W' \subset V' \tag{2.1}$$

* Received by the editors March 21, 1977, and in final revised form November 30, 1977.
† Faculty of Mathematics, Universitatea "Al. I. Cuza", Iaşi, Romania.

algebraically and topologically, where W' (respectively V') is the dual of W (resp. V) via the inner product $(\cdot,\cdot)$ of H. In other words W' (resp. V') is the completion of H under the norm $\|w^*\|_{W'} = \sup(|(w^*, w)|/\|w\|_W;\ w \in W)$ (resp., $\|v^*\|_{V'} = \sup(|v^*, v)|/\|v\|_V;\ v \in V)$). The norms in H, V and W are denoted by $|\cdot|$, $\|\cdot\|_V$ and $\|\cdot\|_W$, respectively.

(b) $A = \partial\psi$ and $B = \partial\varphi$ where $\psi: W \to]-\infty, +\infty]$ and $\varphi: V \to]-\infty, +\infty]$ are lower semicontinuous, proper convex functions. Here $\partial\psi$ and $\partial\varphi$ denote the subdifferentials of ψ and φ respectively (see e.g. [14, p. 59]).

In particular assumption (b) implies that the operators A and B are maximal monotone from W and V to W' and V' respectively. Let $A_H: H \to H$ be the operator defined by: $A_H u = Au \cap H$. We shall assume that

(c) the operator A_H is maximal monotone on H.

A sufficient condition for this is (see e.g. [2, p. 254])

$$\lim_{\|u\|_W \to \infty} \psi(u) = \infty. \tag{2.2}$$

Let $A_\lambda = \lambda^{-1}(I - (I + \lambda A_H)^{-1})$, $\lambda > 0$, be the Yosida approximation of A_H (see [6, p. 28] for properties of A_λ). The next assumption relating A and B is

(d) there exists a real number γ such that

$$(A_\lambda u, v) \geqq \gamma \quad \text{for } u \in H,\ v \in Bu \cap H \text{ and } \lambda > 0. \tag{2.3}$$

(e) The function φ is strictly convex and

$$\lim_{\|u\|_V \to \infty} \varphi(u)/\|u\|_V = \infty. \tag{2.4}$$

As regards the scalar kernel a we will require that

(f) $a \in C^1([0, \infty[)$, a' is locally absolutely continuous on $[0, \infty[$ and $a(0) > 0$.

The first existence result is

THEOREM 1. *Let the general assumption* (a)–(f) *be satisfied. Further, assume that A is single valued, everywhere defined on W and*

(2.5) *the injection of V into W is compact.*

Then for every $f \in W^{1,2}_{\text{loc}}([0, \infty[; H)$ equation (1.1) *has a solution u in the sense*

$$u \in C([0, \infty[; V_w) \cap C([0, \infty[; W), \tag{2.6}$$

$$Au \in L^\infty_{\text{loc}}([0, \infty[; W'), \qquad \int_0^t Au(s)\, ds \in L^\infty_{\text{loc}}([0, \infty[; H), \tag{2.7}$$

$$f - a * Au \in L^2_{\text{loc}}([0, \infty[; H) \tag{2.8}$$

$$(f - a * Au)(t) \in Bu(t) \quad a.e.\ t > 0. \tag{2.9}$$

Here $W^{k,p}_{\text{loc}}([0, \infty[; H)$, $1 \leqq p \leqq \infty$, denotes the space of all locally absolutely continuous functions $u(t)$ on $[0, \infty[$ with values in H such that $u^{(k)} \in L^p_{\text{loc}}([0, \infty[; H)$. We have denoted by $C([0, \infty[; V_w)$ the space of V-valued weakly continuous functions on $[0, \infty[$. Finally, $(a * Au)(t)$ stands for the convolution product $\int_0^t a(t-s)Au(s)\, ds$ which makes sense in virtue of (2.7). In particular it follows from (2.7) that $f - a * Au \in L^\infty_{\text{loc}}([0, \infty[; W')$ and $(f - a * Au)' \in L^\infty_{\text{loc}}([0, \infty[; W')$.

Under additional conditions relating A and B, the compactness condition (2.5) may be avoided. For example we have

THEOREM 2. *Let the general assumptions* (a)–(f) *be satisfied with $V = W = H$. Assume in addition that*:

(g) $D(B) \subset D(A)$ *and for each* $u_0 \in D(B)$ *there exist positive constants* r, α *and* β *such that*

$$|A^0 u| \leqq \alpha |B^0 u| + \beta \quad \text{for } u \in D(B), \quad |u - u_0| \leqq r. \tag{2.10}$$

(h) *For every* $\delta > 0$ *the level subset* $\{u \in H; \varphi(u) \leqq \delta\}$ *is compact in* H.

Then for every $f \in W^{1,2}_{\text{loc}}$ $([0, \infty[; H)$ *equation* (1.1) *has a solution* $u \in C([0, \infty[; H)$ *in the sense that there exist* v *and* w *in* L^2_{loc} $([0, \infty[; H)$ *such that*

$$w(t) \in Au(t), \qquad v(t) \in Bu(t) \quad \text{a.e. } t > 0 \tag{2.11}$$

$$v(t) + (a * w)(t) = f(t) \quad \text{a.e. } t > 0. \tag{2.12}$$

Here $D(A)$ and $D(B)$ denote the domains of A and B while A^0 and B^0 denote their minimal sections, i.e. $|A^0 u| = \inf\{|z|; z \in Au\}$.

Inasmuch as the constant kernel $a(t) \equiv 1$ satisfies condition (f), the above existence theorems are applicable to differential equations of the form (1.2). For example, Theorem 1 gives

COROLLARY 1. *Assume that* A *and* B *satisfy conditions of Theorem* 1. *Let* $g \in L^2_{\text{loc}}$ $([0, \infty[; H)$ *and* $u_0 \in H$ *be given such that*

$$g \in L^\infty_{\text{loc}} ([0, \infty[; W'), \qquad Bu_0 \cap K \neq \emptyset \tag{2.13}$$

and the injection of V *into* W *is compact.*

Then initial value problem (1.2) *has a solution* $u \in C([0, \infty[; V_w) \cap C([0, \infty[; W)$ *in the sense that there exists* $v \in L^2_{\text{loc}}$ $([0, \infty[; H)$ *such that*

$$v' \in L^\infty_{\text{loc}} ([0, \infty[; W'), \qquad v(t) \in Bu(t) \quad \text{a.e. } t > 0 \tag{2.14}$$

$$v'(t) + Au(t) = g(t) \quad \text{a.e. } t > 0 \qquad \left(' = \frac{d}{dt}\right). \tag{2.15}$$

An analogue of Theorem 2 may also be formulated for equation (1.2) but we do not give details.

It should be said that assumption (e) together with condition (2.5) imply that the operator B^{-1} is continuous from V' to W. Hence under assumptions of Theorem 1 the operator AB^{-1} is demicontinuous (i.e. strongly-weakly continuous) from V' to W'. This fact reveals that Theorems 1 and 2 require that in a certain sense the operator A be "dominated" by B.

If A and B are partial differential operators this means roughly that the order of B is bigger than that of A. A different situation arises in Theorem 3 below.

As before W is a real reflexive Banach space which is dense in H and

$$W \subset H \subset W'$$

algebraically and topologically. Furthermore,

(2.16) the injection of W into H is compact.

We shall assume that $A = \partial\psi \colon W \to W'$ and $B = \partial\varphi \colon H \to H$ where ψ and φ are lower semicontinuous, proper convex functions from W and H, respectively to $]-\infty, +\infty]$. Further, assume that

$$\lim_{\|u\|_W \to \infty} \psi(u)/\|u\|_W = \infty, \tag{2.17}$$

$$(A_\lambda u, v) \geqq 0 \quad \text{for all } \lambda > 0 \text{ and } v \in Bu, \tag{2.18}$$

(2.19) $D(A) = W$ and A maps bounded subsets of W into bounded subsets of W'.

As regards the kernel $a(t)$ we shall assume that

$$a \in C^2([0, \infty[), \qquad a(0) > 0, \quad a'(0) \leqq 0 \tag{2.20}$$

and a'' is locally absolutely continuous on $[0, \infty[$.

THEOREM 3. *Let the assumptions* (2.16)–(2.20) *be satisfied. Let f be given such that*

$$f \in W^{1,2}_{\text{loc}}([0, \infty[; H), \qquad f'' \in L^2_{\text{loc}}([0, \infty[; W') \tag{2.21}$$

$$f(0) \in B(W). \tag{2.22}$$

Then equation (1.1) *has a solution $u \in L^\infty_{\text{loc}}([0, \infty[; W)$ in the sense that there exist functions v and w satisfying*

$$w(t) \in Au(t), \qquad v(t) \in Bu(t), \quad \text{a.e. } t > 0, \tag{2.23}$$

$$v \in W^{1,\infty}_{\text{loc}}([0, \infty[; W') \cap C([0, \infty[; H_w) \tag{2.24}$$

$$\int_0^t w(s)\, ds \in L^\infty_{\text{loc}}([0, \infty[; H), \qquad w \in L^\infty_{\text{loc}}([0, \infty[; W'), \tag{2.25}$$

$$v(t) + (a * w)(t) = f(t) \quad \text{for every } t \geqq 0. \tag{2.26}$$

COROLLARY 2. *Let A and B satisfy assumptions* (2.16)–(2.19). *Let $g \in L^2_{\text{loc}}([0, \infty[; H)$ and $u_0 \in W$, $v_0 \in H$ be given such that*

$$g' \in L^2_{\text{loc}}([0, \infty[; W'), \qquad v_0 \in Bu_0. \tag{2.27}$$

Then the initial value problem (1.2) *has a solution $u \in L^\infty_{\text{loc}}[(0, \infty[; W)$ in the sense*:

$$v \in W^{1,\infty}_{\text{loc}}([0, \infty[; W') \cap C([0, \infty[; H_w) \tag{2.28}$$

$$\int_0^t w(s)\, ds \in L^\infty_{\text{loc}}([0, \infty[; H), \qquad w \in L^\infty_{\text{loc}}([0, \infty[; W') \tag{2.29}$$

$$v(t) \in Bu(t), \qquad w(t) \in Au(t) \quad \text{a.e. } t > 0 \tag{2.30}$$

$$v'(t) + w(t) = g(t) \quad \text{a.e. } t > 0 \tag{2.31}$$

$$v(0) = v_0. \tag{2.32}$$

It should be observed that (2.32) makes sense because by virtue of (2.28), $v \in C([0, \infty[; H_w)$, i.e., v is weakly continuous from $[0, \infty[$ to H.

Under appropriate conditions on A and B, equation (1.2) has been studied by O. Grange and F. Mignot [12] by using a discretization method. Corollary 2 extends their main results in the sense that B need not be the subdifferential of a positive homogeneous function as it is assumed in [12]. Also, more general operators A are allowed here.

3. Proof of Theorem 1. It suffices to prove the existence on each interval of the form $[0, T]$.

Let $u_0 \in D(\varphi) = \{x \in H; \varphi(x) < \infty\}$ (the effective domain of φ) be such that $Bu_0 \cap H \neq \emptyset$. For each $\varepsilon > 0$ consider the integro-differential equation

$$\varepsilon u' + Bu + a * Au \ni f, \qquad 0 < t < \infty \qquad \left(' = \frac{d}{dt}\right) \tag{3.1}$$

with the initial condition

$$u(0) = u_0. \tag{3.2}$$

This equation has been studied under appropriate conditions on A and B by the author in [1] (in this context see also Chap. IV in [2]) and recently by M. G. Crandall, S. O. Londen and J. A. Nohel [9] who allow more general conditions on the kernel a.

We shall denote again by B the operator $u \to Bu \cap H$. It is well known that condition (2.4) implies that this operator is maximal monotone from H into itself. As a matter of fact it is the subdifferential of convex function

$$\tilde{\varphi}(u)=\begin{cases}\varphi(u) & \text{if } u \in V\\ +\infty & \text{if } u \in H\setminus V.\end{cases}$$

Thus we are in the situation of Theorem 1 in [9] so we may infer that for each $\varepsilon>0$, equation (3.1), with the Cauchy condition (3.2) has a solution $u_\varepsilon \in C([0,\infty[;W)$ in the sense

$$u_\varepsilon' \in L^2_{\text{loc}}([0,\infty[;H), \qquad \int_0^t Au_\varepsilon(s)\,ds \in L^\infty_{\text{loc}}([0,\infty[;H), \tag{3.3}$$

$$v_\varepsilon \in L^2_{\text{loc}}([0,\infty[;H), \qquad v_\varepsilon(t) \in Bu_\varepsilon(t) \quad \text{a.e. } t>0, \tag{3.4}$$

$$\varepsilon u_\varepsilon'(t)+v_\varepsilon(t)+(a*Au_\varepsilon)(t)=f(t) \quad \text{a.e. } t>0. \tag{3.5}$$

Furthermore, u_ε may be obtained as limit for $\lambda \to 0^+$ of solutions $u_{\varepsilon\lambda}$ to the regularized equation

$$\varepsilon u_{\varepsilon\lambda}'+Bu_{\varepsilon\lambda}+a*A_\lambda u_{\varepsilon\lambda} \ni f \quad \text{a.e. } t>0 \tag{3.6}$$

satisfying initial condition $u_{\varepsilon\lambda}(0)=u_0$. More precisely, for every $T>0$ there exists a sequence (again denoted λ) tending to zero such that (see relations (2.32) in [9])

$$\begin{aligned}&u_{\varepsilon\lambda} \to u_\varepsilon \quad \text{in } C([0,T];W) \text{ and weak-star in } L^\infty(0,T;V)\\ &u_{\varepsilon\lambda}' \to u_\varepsilon' \quad \text{weakly in } L^2(0,T;H)\\ &v_{\varepsilon\lambda} \to v_\varepsilon \quad \text{weakly in } L^2(0,T;H)\\ &\int_0^t A_\lambda u_{\varepsilon\lambda}(s)\,ds \to \int_0^t Au_\varepsilon(s)\,ds \quad \text{weak-star in } L^\infty(0,T;H)\end{aligned} \tag{3.7}$$

where

$$\varepsilon u_{\varepsilon\lambda}'+v_{\varepsilon\lambda}+a*A_\lambda u_{\varepsilon\lambda}=f \quad \text{a.e. } t>0. \tag{3.8}$$

Next we shall derive bounds on solutions u_ε to problem (3.1), (3.2).

LEMMA 1. *Let $T>0$ be given. There exists a positive constant C independent of $\varepsilon \in]0,1]$ such that*

$$\int_0^T |v_\varepsilon(t)|^2\,dt+\left|\int_0^t Au_\varepsilon(s)\,ds\right|^2 \leqq C \quad \textit{for } 0<\varepsilon\leqq 1, \quad t\in[0,T], \tag{3.9}$$

$$\varepsilon\int_0^T |u_\varepsilon'(t)|^2\,dt+\|u_\varepsilon(t)\|_V^2+\|Au_\varepsilon(t)\|_{W'}^2 \leqq C \quad \textit{for } \varepsilon\in]0,1] \quad \textit{a.e. } t\in]0,T[. \tag{3.10}$$

Proof. Without any loss of generality we may assume that $\varphi \geqq 0$ on V and $\psi \geqq 0$ on W. For example this may be achieved replacing the operator A by $\tilde{A}u=Au-Au_1$ and B by $\tilde{B}u=Bu-v_1$ where $u_1\in W$ and $v_1 \in Bu_1$ are arbitrary but fixed.

Forming the inner product of (3.8) with $v_{\varepsilon\lambda}$, integrating over $[0,t]$ and using the chain rule (see e.g. [6, Lemma 3.3])

$$\frac{d}{dt}\tilde{\varphi}(u_{\varepsilon\lambda}(t))=(v_{\varepsilon\lambda}(t),u_{\varepsilon\lambda}'(t)) \quad \text{a.e. } t>0 \tag{3.11}$$

we get

$$\int_0^t |v_{\varepsilon\lambda}(s)|^2\, ds \leqq \varepsilon\varphi(u_0)+\int_0^t |f(s)|\,|v_{\varepsilon\lambda}(s)|\, ds \tag{3.12}$$
$$+\int_0^t |(a * A_\lambda u_{\varepsilon\lambda})(s)|\,|v_{\varepsilon\lambda}(s)|\, ds.$$

Next we multiply (3.8) by $A_\lambda u_{\varepsilon\lambda}$ and integrate over $[0, t]$. Using condition (d) we get after an integration by parts that (we drop the argument under integral sign)

$$\int_0^t (a * A_\lambda u_{\varepsilon\lambda}, A_\lambda u_{\varepsilon\lambda})ds = \varepsilon\psi(u_0)+|f(t)|\,|F_{\varepsilon\lambda}(t)| \tag{3.13}$$
$$+\int_0^t |f'|\,|F_{\varepsilon\lambda}|\, ds+\gamma t$$

which yields

$$\int_0^t (a * A_\lambda u_{\varepsilon\lambda}, A_\lambda u_{\varepsilon\lambda})\, ds \leqq C_1+C_2 \max\{|F_{\varepsilon\lambda}(s)|; 0\leqq s\leqq t\}.$$

Here $F_{\varepsilon\lambda}(t)=\int_0^t A_\lambda u_{\varepsilon\lambda}(s)\, ds$. Then by virtue of condition (f) we may conclude that (see [9, proposition (a)]):

$$|F_{\varepsilon\lambda}(t)|\leqq C \quad \text{for all } \varepsilon, \lambda>0, \quad t\in[0, T], \tag{3.14}$$

where C is independent of ε and λ.

Integrating by parts in $a * A_\lambda u_{\varepsilon\lambda}$ we get

$$(a * A_\lambda u_{\varepsilon\lambda})(t)=a(0)F_{\varepsilon\lambda}(t)+(a' * F_{\varepsilon\lambda})(t). \tag{3.15}$$

Substituting (3.15) in (3.12) and using (3.14) we get (3.9) as claimed.

To establish (3.10) we shall first prove that $u'_\varepsilon\in L^\infty(0, T; H)$ and

$$\varepsilon|u'_\varepsilon(t)|\leqq C \quad \text{for } \varepsilon>0 \text{ and a.e. } t\in]0, T[. \tag{3.16}$$

To this purpose consider t, h real numbers satisfying $0<t<t+h<T$. Then equation (3.8) and the monotonicity of B yield

$$\frac{\varepsilon}{2}\frac{d}{dt}|u_{\varepsilon\lambda}(t+h)-u_{\varepsilon\lambda}(t)|^2+\left(u_{\varepsilon\lambda}(t+h)-u_{\varepsilon\lambda}(t), \int_0^t (a(t+h-s)-a(t-s))A_\lambda u_{\varepsilon\lambda}(s)\, ds\right)$$
$$+\left(u_{\varepsilon\lambda}(t+h)-u_{\varepsilon\lambda}(t), \int_t^{t+h} a(t+h-s)A_\lambda u_{\varepsilon\lambda}(s)\, ds\right) \tag{3.17}$$
$$\leqq (f(t+h)-f(t), u_{\varepsilon\lambda}(t+h)-u_{\varepsilon\lambda}(t)).$$

Next we multiply (3.8) by $u_{\varepsilon\lambda}(t)-u_0$ to find

$$\frac{\varepsilon}{2}\frac{d}{dt}|u_{\varepsilon\lambda}(t)-u_0|^2\leqq |B^0u_0|\,|u_{\varepsilon\lambda}(t)-u_0|+|f(t)|\,|u_{\varepsilon\lambda}(t)-u_0|$$
$$+\left|\int_0^t a(t-s)A_\lambda u_{\varepsilon\lambda}(s)\, ds\right|\,|u_{\varepsilon\lambda}(t)-u_0|, \quad \text{a.e. } t\in]0, T[$$

and therefore

$$\limsup_{h\to 0} \varepsilon|u_{\varepsilon\lambda}(h)-u_0|/h\leqq |B^0u_0|+|f(0)|, \quad \text{for all } \varepsilon, \lambda>0. \tag{3.18}$$

Integrating (3.17) and using estimate (3.18) yields

$$\frac{\varepsilon}{2}|u'_{\varepsilon\lambda}(t)|^2+\int_0^t (a' * A_\lambda u_{\varepsilon\lambda}, u'_{\varepsilon\lambda})\,ds+a(0)\int_0^t (u'_{\varepsilon\lambda}, A_\lambda u_{\varepsilon\lambda})\,ds$$

$$(3.19)\qquad \leqq (|B^0u_0|+|f(0)|)^2/\varepsilon+\int_0^t (f', u'_{\varepsilon\lambda})\,ds, \quad \text{a.e. } t\in\,]0, T[.$$

Let $\psi_\lambda: H\to[0,+\infty[$ be such that $A_\lambda=\partial\psi_\lambda$. Then ψ_λ is Fréchet differentiable on H and (see e.g. [6, Proposition 2.11]), $\psi_\lambda(u)\leqq\psi(u)$ for all $\lambda>0$ and $u\in W$.

Also observe that absolute continuity of a' together with estimate (3.14) imply that $a' * A_\lambda u_{\varepsilon\lambda}$ is uniformly bounded on $[0, T]$. These facts and estimate (3.9) yield after some calculations involving Gronwalls's lemma that $\varepsilon|u'_{\varepsilon\lambda}(t)|\leqq C$ for all $\varepsilon, \lambda>0$ and a.e. $t\in\,]0, T[$ where C is independent of ε and λ. Remembering (3.7) we obtain (3.16) as desired.

Now we multiply equation (3.5) by $u_\varepsilon-u^0$ where u^0 is any fixed element of $D(\varphi)$ and use estimates (3.9) and (3.16) to find

$$(3.20)\qquad \varphi(u_\varepsilon(t))\leqq\varphi(u_0)+C\Big(|u_\varepsilon(t)|+\int_0^t |u_\varepsilon(s)|\,ds\Big),\qquad 0\leqq t\leqq T$$

(we notice that (3.9) also implies that $|(a * Au_\varepsilon)(t)|$ is uniformly bounded over $[0, T]$). On the other hand condition (2.4) implies that for every $\mu>0$ there exists θ real such that

$$\varphi(u)\geqq\mu\|u\|_V+\theta\quad \text{for all } u\in V.$$

From this and (3.20) we deduce that

$$(3.21)\qquad \|u_\varepsilon(t)\|_V\leqq C\quad \text{for } t\in[0, T] \text{ and all } \varepsilon\in\,]0, 1].$$

Since A is monotone from W to W' and defined on all of W, it is locally bounded on W so that estimate (3.21) together with assumption (2.5) imply that

$$(3.22)\qquad \|Au_\varepsilon(t)\|_{W'}\leqq C\quad \text{for all } \varepsilon\in\,]0, 1] \text{ and } t\in[0, T].$$

Finally, we take the inner product of (3.5) with u'_ε and integrate over $[0, T]$. Making use once again of equation (3.11) we get

$$\varepsilon\int_0^T |u'_\varepsilon|^2\,dt+\varphi(u_\varepsilon(T))+((a * Au_\varepsilon)(T), u_\varepsilon(T))$$

$$-a(0)\int_0^T (Au_\varepsilon, u_\varepsilon)\,dt-\int_0^T (a' * Au_\varepsilon, u_\varepsilon)\,dt$$

$$=(f(T), u_\varepsilon(T))-\int_0^T (f', u_\varepsilon)\,dt-\varphi(u_0)-(f(0), u_\varepsilon(0)).$$

On the other hand integrating by parts,

$$(a' * Au_\varepsilon)(t)=a'(0)F_\varepsilon(t)+(a'' * F_\varepsilon)(t)$$

we deduce by (3.9) that $|a' * Au_\varepsilon(t)|$ are uniformly bounded on $[0, T]$ because, by assumption (f), $a''\in L^1(0, T)$. These relations imply (3.10) as claimed.

We continue the proof of Theorem 1 by letting $\varepsilon\to 0^+$. First we notice that by virtue of (2.5) the injection of W' into V' is compact. On the other hand the family $\{F_\varepsilon\}\subset C([0, T]; V')$ is equibounded and equicontinuous (because by (3.10), $\{dF_\varepsilon/dt= Au_\varepsilon\}$ is bounded in $L^\infty(0, T; W')$). Then by the Arzela–Ascoli theorem we have the existence of a function $F \in C([0, T]; V')$, and a sequence $\varepsilon_n \to 0^+$ such that

$$F_{\varepsilon_n} \to F \quad \text{in } C([0, T]; V') \text{ and weak-star in } L^\infty(0, T; H). \tag{3.23}$$

Extracting a further subsequence if necessary, by virtue of Lemma 1 we may assume that there exist the functions $u \in L^\infty(0, T; V)$ and $v \in L^2(0, T; H)$ such that

$$\begin{aligned} &u_{\varepsilon_n} \to u \quad \text{weak-star in } L^\infty(0, T; V) \\ &v_{\varepsilon_n} \to v \quad \text{weakly in } L^2(0, T; H) \\ &\varepsilon_n u'_{\varepsilon_n} \to 0 \quad \text{in } L^2(0, T; H). \end{aligned} \tag{3.24}$$

We set $g_n = f - \varepsilon_n u'_{\varepsilon_n} - a(0)F_{\varepsilon_n} - a' * F_{\varepsilon_n} = f - \varepsilon_n u'_{\varepsilon_n} - a * Au_{\varepsilon_n}$. By (3.23) and (3.24) it follows that

$$g_n(t) \to f(t) - a(0)F(t) - (a' * F)(t), \quad \text{a.e. } t \in]0, T[\tag{3.25}$$

in the strong topology of V'.

It should be said that condition (e) part (2.4) implies that the range of B is all of V' (see e.g. [2, p. 56]), while the strict convexity of φ implies that B^{-1} is single valued on V'. Thus B^{-1} is demicontinuous from V' into V. Since the bounded subsets of V are compact into W we may therefore deduce from (3.23) that

$$u_{\varepsilon_n}(t) = B^{-1} g_n(t) \to u(t) \quad \text{a.e. } t \in]0, T[\tag{3.26}$$

in the strong topology of W. Hence

$$Au_{\varepsilon_n} \to Au \quad \text{in } L^\infty(0, T; W')$$

$(a * Au_{\varepsilon_n})(t) \to (a * Au)(t)$ weakly in W' for all $t \in [0, T]$ because A is demicontinuous from W to W'. Finally since by virtue of (2.4), B is maximal monotone on H, relations (3.25) and (3.26) imply by a standard argument that $v(t) \in Bu(t)$ a.e. $t \in]0, T[$. Summarizing at this point we have proved that $u(t)$ is a solution of equation (1.1). It remains to prove that $u \in C([0, T]; W) \cap C([0, T]; V_w)$. This follows by noticing that $(f - a * Au)(t)$ is V'-valued continuous and, as observed earlier, B^{-1} is demicontinuous from V' to V and continuous from V' to W. The proof is thereby complete.

4. Proof of Theorem 2. For the proof we need the following existence result for equation (3.1).

PROPOSITION 1. *Let the assumptions of Theorem 2 be satisfied. Then for each* $\varepsilon > 0$, $f \in W^{1,2}_{\mathrm{loc}}([0, \infty[; H)$ *and* $u_0 \in D(B) \cap D(\psi)$, *problem* (3.1), (3.2) *has a solution* u_ε satisfying $u_\varepsilon, u'_\varepsilon \in L^\infty_{\mathrm{loc}}([0, \infty[; H)$ *and there exist* $v_\varepsilon, w_\varepsilon \in L^2_{\mathrm{loc}}([0, \infty[; H)$ *such that* $v_\varepsilon(t) \in Bu_\varepsilon(t)$, $w_\varepsilon(t) \in Au_\varepsilon(t)$ *and*

$$\varepsilon u'_\varepsilon + v_\varepsilon + a * w_\varepsilon = f \quad \textit{a.e. } t > 0. \tag{4.1}$$

Proposition 1 has been proved in [9, Thm. 2] in the special case in which condition (2.10) holds on every bounded subset of $D(B)$. (On these lines more general results can be found in [3].)

Proof. Let $u_{\varepsilon\lambda}$ be the unique solution of approximating equation ($\lambda > 0$)

$$\begin{aligned} &\varepsilon u'_{\varepsilon\lambda} + Bu_{\varepsilon\lambda} + a * A_\lambda u_{\varepsilon\lambda} \ni f \quad \text{a.e. } t > 0 \\ &u_{\varepsilon\lambda}(0) = u_0 \end{aligned} \tag{4.2}$$

where A_λ denotes as usually the Yosida approximation of A which in our situation is maximal monotone on H. It is well known (see e.g. [9, Lemma 2.1]) that problem (4.2) has a unique solution $u_{\varepsilon\lambda} \in W^{1,2}_{\text{loc}}$ ($[0, \infty[; H$). We shall use the notations

$$F_{\varepsilon\lambda}(t) = \int_0^t A_\lambda u_{\varepsilon\lambda}\, ds, \qquad v_{\varepsilon\lambda} = f - \varepsilon u'_{\varepsilon\lambda} - a * A_\lambda u_{\varepsilon\lambda}.$$

Reasoning as in the proof of Lemma 1 we obtain the estimate (see (3.12) and (3.15))

$$\int_0^T |v_{\varepsilon\lambda}|^2\, dt + |F_{\varepsilon\lambda}(t)| \leqq C \quad \text{for } \varepsilon, \lambda > 0, \quad \in [0, T] \tag{4.3}$$

where $T > 0$ is fixed and C is independent of ε and λ. Furthermore, the inequality (3.19) yields

$$\varepsilon |u'_{\varepsilon\lambda}(t)| \leqq C \quad \text{for all } \varepsilon, \lambda > 0, \quad \text{a.e. } t \in]0, T[$$

and therefore

$$\varphi(u_{\varepsilon\lambda}(t)) \leqq C\Big(u_{\varepsilon\lambda}(t)| + \int_0^t |u_{\varepsilon\lambda}(s)|\, ds + 1\Big) \quad \text{for all } \varepsilon, \lambda > 0, \quad t \in [0, T]. \tag{4.4}$$

(We shall denote by C several positive constants independent of $\varepsilon \in]0, 1]$ and $\lambda > 0$.) As seen in the proof of Theorem 1, inequality (4.4) in conjunction with condition (2.4) implies that $|u_{\varepsilon\lambda}(t)|$ is uniformly bounded on $[0, T]$. Thus by virtue of condition (h), $\{u_{\varepsilon\lambda}(t); 0 \leqq t \leqq T, 0 < \varepsilon \leqq 1; \lambda > 0\}$ belongs to a compact subset of H. Then assumption (g) yields

$$|A^0 u_{\varepsilon\lambda}(t)| \leqq \alpha |v_{\varepsilon\lambda}(t)| + \beta \quad \text{for } \varepsilon, \lambda > 0, \quad t \in [0, T]$$

where α and β are independent of ε and λ. Recalling that $|A_\lambda u| \leqq |A^0 u|$ for all $u \in D(A)$, we see that it follows from (4.3) that

$$\int_0^T |A_\lambda u_{\varepsilon\lambda}|^2\, dt \leqq C \quad \text{for all } \varepsilon, \lambda > 0. \tag{4.5}$$

Next we multiply (4.2) by $u'_{\varepsilon\lambda}$ and integrate over $[0, t]$ to find

$$\begin{aligned} &\varepsilon \int_0^t |u'_{\varepsilon\lambda}|^2\, ds + \varphi(u_{\varepsilon\lambda}(t)) + (u_{\varepsilon\lambda}(t), (a * A_\lambda u_{\varepsilon\lambda})(t)) \\ &\quad - a(0) \int_0^t (u_{\varepsilon\lambda}, A_\lambda u_{\varepsilon\lambda})\, ds - \int_0^t (u_{\varepsilon\lambda}, a' * A_\lambda u_{\varepsilon\lambda})\, ds \\ &\quad = \varphi(u_0) + (f(t), u_{\varepsilon\lambda}(t)) - \int_0^t (f', u_{\varepsilon\lambda})\, ds - (f(0), u_0). \end{aligned}$$

By an elementary calculation involving estimate (4.5) we get

$$\varepsilon \int_0^T |u'_{\varepsilon\lambda}|^2\, dt \leqq C \quad \text{for all } \varepsilon, \lambda > 0. \tag{4.6}$$

Then by assumptions (e), (h) and the Arzela–Ascoli theorem we deduce from (4.5) and (4.6) the existence of functions $u_\varepsilon \in C([0, T]; H)$, v_ε, $w_\varepsilon \in L^2(0, T; H)$ with $u'_\varepsilon \in L^2(0, T; H)$ and a sequence $\lambda_n \to 0$ such that

$$\begin{aligned} u_{\varepsilon\lambda_n} &\to u_\varepsilon && \text{in } C([0, T]; H) \\ w_{\varepsilon\lambda_n} &\to w_\varepsilon && \text{weakly in } L^2(0, T; H) \\ v_{\varepsilon\lambda_n} &\to v_\varepsilon && \text{weakly in } L^2(0, T; H) \\ u'_{\varepsilon\lambda_n} &\to u'_\varepsilon && \text{weakly in } L^2(0, T; H). \end{aligned} \tag{4.7}$$

Since A and B are maximal monotone, we have

$$v_\varepsilon(t)\in Bu_\varepsilon(t), \qquad w_\varepsilon(t)\in Au_\varepsilon(t) \quad \text{a.e. } t\in]0, T[\tag{4.8}$$

and

$$\varepsilon u'_\varepsilon+v_\varepsilon+a*w_\varepsilon=f \quad \text{a.e. } t\in]0, T[. \tag{4.9}$$

Therefore u_ε is a solution to (3.1). It is instructive to notice that estimates (4.4), (4.5), (4.6) in conjunction with (4.7) give a bound independent of ε for u_ε. More precisely we have

$$\int_0^T (\varepsilon|u'_\varepsilon|^2+|v_\varepsilon|^2+|w_\varepsilon|^2)\,dt+|u_\varepsilon(t)|+\varphi(u_\varepsilon(t))\leqq C \quad \text{for } \varepsilon>0, \quad t\in[0, T] \tag{4.10}$$

where C is independent of ε. In particular, we deduce that

$$\varepsilon u'_\varepsilon\to 0 \quad \text{strongly in } L^2(0, T; H) \text{ for } \varepsilon\to 0. \tag{4.11}$$

Returning to the proof of the theorem we set

$$g_\varepsilon=f-a*Au_\varepsilon, \qquad q_\varepsilon=-\varepsilon u'_\varepsilon.$$

Then (4.1) may be rewritten as

$$u_\varepsilon(t)=B^{-1}(g_\varepsilon(t)+q_\varepsilon(t)) \quad \text{a.e. } t\in]0, T[. \tag{4.12}$$

Let $\tilde{u}_\varepsilon=B^{-1}g_\varepsilon$. By definition of B we have

$$\varphi(\tilde{u}_\varepsilon(t))\leqq\varphi(v_1)+(g_\varepsilon(t), \tilde{u}_\varepsilon(t)-v_1)$$

where v_1 is arbitrary but fixed in $D(\varphi)$. Invoking conditions (2.4) and (h) we deduce from the latter that $\{\tilde{u}_\varepsilon(t)\}$ belongs to a compact subset of H. Moreover, as is easily seen from estimate (4.10) and condition (f), $\{g_\varepsilon\}$ is an equibounded and equicontinuous family of H-valued functions on $[0, T]$. As observed earlier condition (e) implies that the operator B^{-1} is single valued and demicontinuous from H into itself. Moreover, condition (h) shows that B^{-1} maps bounded subsets of H into relatively compact subsets of H. Let us assume for the time being

LEMMA 2. *The operator B^{-1} is uniformly continuous on every bounded subset of H.*

Hence we may conclude that $\{\tilde{u}_\varepsilon\}$ is an equicontinuous and equibounded subset of $C([0, T]; H)$. Then again using the Ascoli theorem we have the existence $\varepsilon_n\to 0$ and $u\in C([0, T]; H)$ such that

$$\tilde{u}_{\varepsilon_n}\to u \quad \text{in } C([0, T]; H) \tag{4.13}$$

while by estimate (4.10) we have

$$\lim_{\varepsilon_n\to 0} q_{\varepsilon_n}(t)=0 \quad \text{a.e. } t\in]0, T[\tag{4.14}$$

in the strong topology of H. Since B^{-1} is continuous it follows from (4.11), (4.12) and (4.13) that

$$\lim_{\varepsilon_n\to 0} u_{\varepsilon_n}(t)=u(t) \quad \text{a.e. } t\in]0, T[. \tag{4.15}$$

Hence the Lebesgue dominated convergence theorem implies that

$$u_{\varepsilon_n}\to u \quad \text{in } L^2(0, T; H) \tag{4.16}$$

while by estimate (4.10) we may assume that

$$v_{\varepsilon_n} \to v \quad \text{weakly in } L^2(0, T; H)$$

$$w_{\varepsilon_n} \to w \quad \text{weakly in } L^2(0, T; H).$$

Since A and B are maximal monotone, just as above, we have

$$w(t) \in Au(t), \qquad v(t) \in Bu(t) \quad \text{a.e. } t \in]0, T[$$

which show that u is a solution to (1.1) on the interval $[0, T]$.

It remains to prove Lemma 2. Suppose that the sequences $\{u_n\}$ and $\{v_n\}$ exist in H such that $|u_n|+|v_n| \leqq M$, $|u_n - v_n| \to 0$ as $n \to \infty$ and $|B^{-1}u_n - B^{-1}v_n| \geqq \rho > 0$ for all n. We argue from this to a contradiction. Let $y_n = B^{-1}u_n$ and $z_n = B^{-1}v_n$. By definition of $\partial\varphi$ we have

$$\varphi(y_n) \leqq (u_n, y_n - u) + \varphi(u) \quad \text{for all } u \in H, \tag{4.17}$$

respectively,

$$\varphi(z_n) \leqq (v_n, z_n - u) + \varphi(u) \quad \text{for all } u \in H. \tag{4.18}$$

In particular, conditions (2.4) and (h) imply that $\{y_n\}$ and $\{z_n\}$ remain in a compact subset of H. Thus extracting subsequences, we have as $n_l \to \infty$

$$y_{n_k} \to y_1, \qquad z_{n_k} \to y_2 \quad \text{strongly in } H \tag{4.19}$$

$$u_{n_k} \to u_1, \qquad v_{n_k} \to u_1 \quad \text{weakly in } H. \tag{4.20}$$

Since φ is lower semicontinuous, it follows by (4.17) and (4.18) that

$$\varphi(y_1) \leqq (u_1, y_1 - u) + \varphi(u) \quad \text{for all } u \in H,$$

and

$$\varphi(y_2) \leqq (u_1, y_2 - u) + \varphi(u) \quad \text{for all } u \in H.$$

Hence $y_1 \in B^{-1}u_1$ and $y_2 \in B^{-1}u_1$. Since B^{-1} is single valued (because φ is strictly convex) we conclude that $y_1 = y_2$. On the other hand relations (4.19) imply that $|y_1 - y_2| \geqq \rho > 0$. The contradiction we have arrived at concludes the proof.

5. Proof of Theorem 3. Without any loss of generality we may assume that $0 \in A0$ and $0 \in B0$. This can be achieved by shifting the domains and ranges of A and B. As above we confine ourselves to proving the existence on an arbitrary interval $[0, T]$.

We apply Theorem 2 where B is replaced by $B + \varepsilon A_H + \varepsilon I$ ($\varepsilon > 0$). It should be said that condition (2.10) is satisfied in the present situation because condition (2.18) implies (see [6, Proposition (2.17)]),

$$\varepsilon |A_H^0 u| \leqq |(B + \varepsilon A_H)^0 u| \quad \text{for all } u \in D(A_H) \cap D(B). \tag{5.1}$$

(The notations are those used in § 4.)

Thus for each $\varepsilon > 0$ the equation

$$\varepsilon u_\varepsilon + \varepsilon A_H u_\varepsilon + Bu_\varepsilon + a * A_H u_\varepsilon \ni f \qquad (0 < t < T)$$

has at least one solution $u_\varepsilon \in C([0, T); H)$ in the sense that there exist v_ε and w_ε in $L^2(0, T; H)$ satisfying

$$v_\varepsilon(t) \in Bu_\varepsilon(t), \qquad w_\varepsilon(t) \in A_H u_\varepsilon(t) \quad \text{a.e. } t \in]0, T[$$

and

$$\varepsilon u_\varepsilon(t)+\varepsilon w_\varepsilon(t)+v_\varepsilon(t)+(a * w_\varepsilon)(t)=f(t) \quad \text{a.e. } t\in\,]0, T[. \tag{5.2}$$

Furthermore, it follows from the proof of Theorem 2 (see estimate (4.10)) that

$$\int_0^t |v_\varepsilon|^2\,ds+|F_\varepsilon(t)|\leqq C \quad \text{for } \varepsilon<0, \quad t\in[0, T] \tag{5.3}$$

$$\varepsilon\int_0^T (|u_\varepsilon|^2+|w_\varepsilon|^2)\,dt\leqq C \quad \text{for all } \varepsilon>0 \tag{5.4}$$

where $F_\varepsilon(t)=\int_0^t w_\varepsilon\,ds$ and C is independent of ε.

Making use of monotonicity of A and B it follows from (5.2) that

$$\varepsilon|u_\varepsilon(t+h)-u_\varepsilon(t)|^2+\left(u_\varepsilon(t+h)-u_\varepsilon(t), \int_0^t (a(t+h-s)-a(t-s))w_\varepsilon(s)\,ds\right)$$
$$+\left(u_\varepsilon(t+h)-u_\varepsilon(t), \int_t^{t+h} a(t+h-s)w_\varepsilon(s)\,ds\right)\leqq (f(t+h)-f(t), u_\varepsilon(t+h)-u_\varepsilon(t))$$

for $0\leqq t<t+h\leqq T$. Since $a'\in L^1(0, T)$ and $f\in W^{1,2}(0, T; H)$ the latter implies that u_ε is absolutely continuous from $[0, T]$ to H and $u_\varepsilon'\in L^2(0, T; H)$ (i.e., $u_\varepsilon\in W^{1,2}(0, T; H)$). Dividing by h^2 and letting h tend to zero we get

$$\varepsilon\int_0^t |u_\varepsilon'|^2\,ds+\int_0^t (u_\varepsilon', a' * w_\varepsilon)\,ds+a(0)\int_0^t (u_\varepsilon', w_\varepsilon)\,ds\leqq\int_0^t (f', u_\varepsilon')\,ds. \tag{5.5}$$

On the other hand, one has

$$(u_\varepsilon', w_\varepsilon)=(\psi(u_\varepsilon))' \quad \text{a.e. } t\in\,]0, T[$$

and

$$(a' * w_\varepsilon)(t)=a'(0)F_\varepsilon(t)+(a'' * F_\varepsilon)(t),$$

respectively,

$$(a'' * w_\varepsilon)(t)=a''(0)F_\varepsilon(t)+(a''' * F_\varepsilon)(t).$$

Substituting these equalities in (5.5) and recalling that $f''\in L^2(0, T; W')$, $a'''\in L^1(0, T)$ we obtain after rearrangements that

$$a(0)\psi(u_\varepsilon(t))\leqq a(0)\psi(u_\varepsilon(0))+a'(0)\int_0^t (u_\varepsilon, w_\varepsilon)\,ds+|f'(0)|\,|u_\varepsilon(0)|$$
$$+C\left(1+\|u_\varepsilon(t)\|_W+\left(\int_0^t \|u_\varepsilon\|_W^2\,ds\right)^{1/2}\right) \quad \text{for } \varepsilon>0.$$

(Here we have also used estimate (5.3)). Since $a'(0)\leqq 0$ and $(w_\varepsilon, u_\varepsilon)\geqq 0$ (because A is monotone) we find

$$\psi(u_\varepsilon(t))\leqq\psi(u_\varepsilon(0))+C\left(1+\|u_\varepsilon(t)\|_W+\left(\int_0^t \|u_\varepsilon\|_W^2\,ds\right)^{1/2}+|u_\varepsilon(0)|\right) \tag{5.6}$$

for all $\varepsilon>0$ and $t\in[0, T]$.

Now we shall find a bound for $\psi(u_\varepsilon(0))$. Let $u_0\in W=D(\psi)$ be such that $f(0)\in Bu_0$ (see condition (2.22)). Multiplying (5.2) by $u_\varepsilon(t)-u_0$ and using the monotonicity of B together with definition of $\partial\psi$, we find

$$\varepsilon(u_\varepsilon(t), u_\varepsilon(t)-u_0)+\varepsilon\psi(u_\varepsilon(t))+((a * w_\varepsilon)(t); u_\varepsilon(t)-u_0)$$
$$\leqq \varepsilon\psi(u_0)+(f(t)-f(0), u_\varepsilon(t)-u_0) \quad \text{for } t\in[0, T].$$

Letting $t\to 0^+$ yields

(5.7) $$\psi(u_\varepsilon(0))+\tfrac{1}{2}|u_\varepsilon(0)|^2 \leqq \psi(u_0)+\tfrac{1}{2}|u_0|^2 \quad \text{for all } \varepsilon<0.$$

Combining this inequality with (5.6) we get that

$$\psi(u_\varepsilon(t))\leqq C\left(1+\|u_\varepsilon(t)\|_W+\left(\int_0^t \|u_\varepsilon\|_W^2\, ds\right)^{1/2}\right),$$

which in conjunction with condition (2.17) implies that

(5.8) $$\|u_\varepsilon(t)\|_W \in C \quad \text{for all } \varepsilon<0 \text{ and } t\in[0, T]$$

(C is independent of ε) and therefore

(5.9) $$\|w_\varepsilon(t)\|_{W'}\leqq C \quad \text{for all } \varepsilon>0 \text{ and } t\in[0, T]$$

because A is bounded on bounded subsets. In particular it follows that $\{F_\varepsilon\}$ is an equibounded and equicontinuous subset of $C([0, T]; W')$. Since by virtue of (2.16) the injection of H into W' is compact, we may conclude by the Arzela–Ascoli theorem that $\{F_\varepsilon\}$ is compact in $C([0, T]; W')$. Thus there exist functions F, u, g, w and a sequence $\varepsilon_n\to 0$ such that

(5.10) $$F_{\varepsilon_n}\to F \quad \text{in } C([0, T]; W') \text{ and weak-star in } L^\infty(0, T; H)$$

(5.11) $$u_{\varepsilon_n}\to u \quad \text{weak-star in } L^\infty(0, T; W)$$

(5.12) $$\varepsilon_n u_{\varepsilon_n}, \varepsilon_n w_{\varepsilon_n}\to 0 \quad \text{in } L^2(0, T; H)$$

(5.13) $$w_{\varepsilon_n}\to g \quad \text{weak-star in } L^\infty(0, T; W')$$

(5.14) $$v_{\varepsilon_n}\to v \quad \text{weakly in } L^2(0, T; H).$$

It should be observed that estimate (5.3) together with (5.10) imply that $F(t)\in H$ and $|F(t)|\leqq C$ for every $t\in[0, T]$. Moreover, since $F\in C([0, T]; W')$ and the injection of H into W' is compact we deduce that F is weakly continuous from $[0, T]$ to H, i.e., $F\in C([0, T]; H_w)$. Remembering that

$$(a * w_\varepsilon)(t)=a(0)F_\varepsilon(t)+(a' * F_\varepsilon)(t)$$

we find that it follows from (5.2), (5.8), (5.9) and (5.10)

(5.15) $$v_{\varepsilon_n}\to v=f-a*g \quad \text{in } C([0, T]; W')$$

and

$$v_{\varepsilon_n}(t)\to v(t) \quad \text{a.e. } t\in\,]0, T[, \quad \text{weakly in } H.$$

In particular, we have

(5.16) $$v(t)+(a*g)(t)=f(t) \quad \text{a.e. } t\in\,]0, T[$$

and

(5.17) $$v(t)+a(0)F(t)+(a'*F)(t)=f(t) \quad \text{a.e. } t\in\,]0, T[.$$

Thus modifying the values of $v(t)$ on a subset of Lebesgue measure equal to zero we may assume that $v \in C([0, T]; H_w)$ and

$$\lim_{\varepsilon_n \to 0} v_{\varepsilon_n}(t) = v(t) \quad \text{weakly in } H \text{ for every } t \in [0, T]. \tag{5.18}$$

Now we shall prove that $v(t) \in Bu(t)$ a.e. $t \in]0, T[$. By monotonicity of B we have

$$\int_9^T (\tilde{v} - v_{\varepsilon_n}, \tilde{u} - u_{\varepsilon_n})\, dt \geqq 0$$

for all $\tilde{u}, \tilde{v} \in L^2(0, T; H)$ such that $\tilde{v}(t) \in B\tilde{u}(t)$ a.e. $t \in]0, T]$. Letting $\varepsilon_n \to 0$, we find by (5.11), (5.14) and (5.15) that

$$\int_0^T (\tilde{v} - v, \tilde{u} - u)\, dt \geqq 0. \tag{5.19}$$

Let $\tilde{B}$ be the realization of B in $L^2(0, T; H)$, i.e.,

$$\tilde{B}\tilde{u} = \{\tilde{w} \in L^2(0, T; H),\ \tilde{w}(t) \in B\tilde{u}(t), \text{ a.e. } t \in]0, T[\}.$$

The operator $\tilde{B}$ is maximal monotone in $L^2(0, T; H)$ because B is maximal monotone in H. Since inequality (5.19) holds for each $[\tilde{u}, \tilde{v}] \in \tilde{B}$ we infer that $v \in \tilde{B}u$, i.e.,

$$v(t) \in Bu(t) \quad \text{a.e. } t \in]0, T[. \tag{5.20}$$

To conclude the proof it remains to show that $g(t) \in Au(t)$ a.e. $t \in]0, T[$. Since A is maximal monotone from W to W' it suffices to prove that

$$\lim_{\varepsilon_n \to 0} \int_0^T (w_{\varepsilon_n}, u_{\varepsilon_n})\, dt \leqq \int_0^T (g, u)\, dt. \tag{5.21}$$

(Extracting a further subsequence if necessary, we may assume that $\int_0^T (w_{\varepsilon_n}, u_{\varepsilon_n})\, dt$ converges as $\varepsilon_n \to 0$.)

To this end we take the inner product of (5.2) with u'_ε and integrate over $[0, T]$. After some calculations we get

$$\begin{aligned} -a(0) \int_0^T (w_\varepsilon, u_\varepsilon)\, dt = {} & \varphi(u_\varepsilon(0)) - (f(0), u_\varepsilon(0)) \\ & + \frac{\varepsilon}{2}(|u_\varepsilon(0)|^2 - |u_\varepsilon(T)|^2) + \varepsilon(\psi(u_\varepsilon(0)) - \psi(u_\varepsilon(T))) \\ & - \int_0^T (f', u_\varepsilon)\, dt + \int_0^T (a' * w_\varepsilon, u_\varepsilon)\, dt \\ & \qquad - ((a * w_\varepsilon)(T), u_\varepsilon(T)) \\ & + (f(T), u_\varepsilon(T)) - \varphi(u_\varepsilon(T)). \end{aligned} \tag{5.22}$$

Let $\varphi^*: H \to]-\infty, +\infty]$ be the conjugate function of φ,

$$\varphi^*(y) = \sup\{(y, x) - \varphi(x);\ x \in H\}. \tag{5.23}$$

We have (see e.g. [14, p. 60]) $B^{-1} = \partial\varphi^*$ and

$$\varphi^*(y) = (y, x) - \varphi(x) \quad \text{for } y \in \partial\varphi(x). \tag{5.24}$$

By (5.23) and (5.24) it follows that

$$(f(0), u_\varepsilon(0)) - \varphi(u_\varepsilon(0)) \leqq \varphi^*(f(0))$$

and

$$(u_\varepsilon(T), v_\varepsilon(T)) - \varphi(u_\varepsilon(T)) = \varphi^*(v_\varepsilon(T)).$$

Taking into account that

$$v_\varepsilon(T) = f(T) - (a * w_\varepsilon)(T) - \varepsilon(u_\varepsilon(T) + w_\varepsilon(T))$$

we deduce from (5.11), (5.13), (5.14) and (5.22),

$$\tag{5.25} \begin{aligned} a(0) \lim_{\varepsilon_n \to 0} \int_0^T (w_{\varepsilon_n}, u_{\varepsilon_n})\,dt \leqq \varphi^*(f(0)) + \int_0^T (f', u)\,dt \\ - \int_0^T (a' * g, u)\,dt - \lim_{\varepsilon_n \to 0} \varphi^*(v_\varepsilon(T)). \end{aligned}$$

Since φ^* is convex and lower semicontinuous on H, it is weakly lower semicontinuous and therefore (5.18) yields

$$\lim_{\varepsilon_n \to 0} \varphi^*(v_{\varepsilon_n}(T)) \geqq \varphi^*(v(T)).$$

Substituting in (5.25) we get

$$\tag{5.26} \begin{aligned} a(0) \lim_{\varepsilon_n \to 0} \int_0^T (w_{\varepsilon_n}, u_{\varepsilon_n})\,dt \leqq \varphi^*(f(0)) - \varphi^*(v(T)) \\ + \int_0^T (f', u)\,dt - \int_0^T (a' * g, u)\,dt. \end{aligned}$$

Returning to equation (5.16), we observe that it implies $v' \in L^\infty(0, T; W')$ and

$$v' + a(0)g + a' * g = f' \quad \text{a.e. } t \in]0, T[.$$

Hence

$$\tag{5.27} a(0) \int_0^T (g, u)\,dt = \int_0^T (f', u)\,dt - \int_0^T (a' * g, u)\,dt - \int_0^T (v', u)\,dt$$

while

$$\tag{5.28} \int_0^T (v', u)\,dt = \varphi^*(v(T)) - \varphi^*(v(0)).$$

Comparison of (5.26), (5.27) and (5.28) yields (5.21) as desired. Before concluding the proof some words of explanation are in order concerning (5.28).

Since $u(t) \in B^{-1}(v(t)) = \partial\varphi^*(v(t))$ a.e. $t \in]0, T[$, we have

$$\tag{5.29} \varphi^*(v(t)) - \varphi^*(v(s)) \leqq (u(t), v(t) - v(s))$$

for all s, t in $[0, T]$. Since $v' \in L^\infty(0, T; W')$ and $u \in L^\infty(0, T; W)$ it follows from (5.29) that the function $t \to \varphi^*(v(t))$ is absolutely continuous on $[0, T]$ and

$$(\varphi^*(v(t)))' = (v'(t), u(t)) \quad \text{a.e. } t \in]0, T[$$

which implies (5.28) as claimed.

Thus the proof of Theorem 3 is complete.

6. Examples. In the sequel Ω will be a bounded and open subset in R^n with smooth boundary Γ. $W^{1,p}(\Omega)$ and $W_0^{1,p}(\Omega)$ will denote usual Sobolev spaces on Ω.

Example 1. Consider the integro-differential equation

$$-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\right)+\int_0^t a(t-s)g(u(s,x))\,ds=f(t,x) \tag{6.1}$$

in the domain $\{t\geqq 0, x\in\Omega\}$, together with the boundary-value conditions

$$u=0 \quad \text{on } \Gamma \tag{6.2}$$

and the initial condition

$$u(0,x)=u_0(x), \qquad x\in\Omega. \tag{6.3}$$

We require that $p\geqq 2$ and that g is a nondecreasing continuous function on $R=\,]-\infty,+\infty[$ satisfying

$$g(0)=0, \qquad |g(u)|\leqq C(|u|^{\alpha-1}+1) \quad \text{for } u\in R \tag{6.4}$$

where α satisfies

$$p\leqq\alpha<\infty \quad \text{if } p\geqq n; \qquad 2\leqq\alpha>np/(n-p) \quad \text{if } n>p. \tag{6.5}$$

Theorem 1 applies neatly to this situation if we make the choices:

$$H=L^2(\Omega), \quad V=W_0^{1,p}(\Omega), \quad W=L^\alpha(\Omega)$$

and $B\colon W_0^{1,p}(\Omega)\to W^{-1,p'}(\Omega)$, $A\colon L^\alpha(\Omega)\to L^{\alpha'}(\Omega)(1/p+1/p'=1, 1/\alpha+1/\alpha'=1)$ defined by

$$(Bu,v)=\int_\Omega\sum_{i=1}^{n}\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i}\,dx \quad \text{for } u,v\in W_0^{1,p}(\Omega) \tag{6.6}$$

$$(Au(x))=g(u(x)) \quad \text{a.e. } x\in\Omega \quad \text{for } u\in L^\alpha(\Omega). \tag{6.7}$$

It is well known (see e.g. [13, p. 185]) that under above conditions the operators A and B are monotone and demicontinuous from W and V to W' and V' respectively. Moreover, the restriction of A to $L^2(\Omega)$ is maximal monotone (see e.g. [2, p. 87]) and one has

$$(A_\lambda u, Bu)\geqq 0 \quad \text{for all } u\in D(B) \text{ and } \lambda>0.$$

The functions φ and ψ are given by

$$\varphi(u)=\left(\sum_{i=1}^{n}\int_\Omega\left|\frac{\partial u}{\partial x_i}\right|^p dx\right)\Big/p \quad \text{for } u\in W_0^{1,p}(\Omega)$$

and

$$\psi(u)=\int_\Omega\int_0^{u(x)} g(r)\,dr\,dx \quad \text{for } u\in L^\alpha(\Omega).$$

Finally, we observe that condition (6.5) in conjunction with the Sobolev imbedding theorem implies that the injection of V into W is compact. Thus if the kernel $a(t)$ satisfies condition (f) we may apply Theorem 1 to deduce that for each $f\in L^2_{\text{loc}}([0,\infty[;L^2(\Omega))$ with $\partial f/\partial t\in L^2_{\text{loc}}([0,\infty[;L^2(\Omega))$ problem (6.1), (6.2), (6.3) has at least one solution $u(t,x)$ satisfying

$$u\in C([0,\infty[;L^\alpha(\Omega))\cap C([0,\infty[;(W_0^{1,p}(\Omega))_w) \tag{6.8}$$

$$Bu\in L^2_{\text{loc}}([0,\infty[;L^2(\Omega)). \tag{6.9}$$

We do not know whether or not we have uniqueness in the above problem. This happens, for instance, if g is locally Lipschitzian on R.

Example 2. Relevant examples of nonlinear differential equations of the form (1.2) can be found in [5], [11]. One example, which comes essentially from [13, p. 451] is the degenerate parabolic equation

$$\frac{\partial}{\partial t}\beta(u(t,x)) - \sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(\left|\frac{\partial u(t,x)}{\partial x_i}\right|^{p-2}\frac{\partial u(t,x)}{\partial x_i}\right) = g(t,x) \tag{6.10}$$

in the domain $\{t \geqq 0, x \in \Omega\}$, with boundary value conditions

$$u(t,x) = 0 \quad \text{for } t \geqq 0, \quad x \in \Gamma \tag{6.11}$$

and initial conditions

$$v_0(x) \in \beta(u(o,x)), \qquad x \in \Omega. \tag{6.12}$$

Here β is a (possible multivalued) maximal monotone graph in $R \times R$ such that $0 \in \beta(0)$.

In Corollary 2 we make the following choices:

$$H = L^2(\Omega), \qquad W = W_0^{1,p}(\Omega), \qquad p \geqq 2 \tag{6.13}$$

$$Au = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\right) \quad \text{for } u \in W_0^{1,p}(\Omega). \tag{6.14}$$

$$Bu = \{w \in L^2(\Omega);\ w(x) \in \beta(u(x)) \text{ a.e. } x \in \Omega\}, \qquad u \in L^2(\Omega). \tag{6.15}$$

We recall that $B = \partial\varphi$ where $\varphi\colon L^2(\Omega) \to]-\infty, +\infty]$ is given by

$$\varphi(u) = \int_\Omega j(u(x))\,dx, \qquad u \in L^2(\Omega)$$

and $j\colon R \to]-\infty, +\infty]$ is uniquely determined (up to an additive constant) by condition $\beta = \partial j$.

It is obvious that conditions (2.16), (2.17) and (2.19) are satisfied. To verify (2.18) it suffices to show (see [6, Thm. 4.4)] that

$$(Au, B_\lambda u) \geqq 0 \quad \text{for all } u \in D(A_H) \text{ and } \lambda > 0$$

which follows immediately from the obvious relation

$$(B_\lambda u(x)) = \beta_\lambda(u(x)) \quad \text{a.e. } x \in \Omega, \quad u \in L^2(\Omega).$$

Hence Corollary 2 is applicable.

Let $g \in L^2_{\mathrm{loc}}([0,\infty[;\ L^2(\Omega))$ and $v_0 \in L^2(\Omega)$ be given such that

$$\partial g/\partial t \in L^2_{\mathrm{loc}}([0,\infty[;\ W^{-1,p'}(\Omega))$$

$$v_0(x) \in \beta(u_0(x)) \quad \text{a.e. } x \in \Omega, \quad u_0 \in W_0^{1,p}(\Omega).$$

Then there exists a solution u to problem (6.10)–(6.12) in the following sense

$$u \in L^\infty_{\mathrm{loc}}([0,\infty[;\ W_0^{1,p}(\Omega)) \tag{6.16}$$

$$v \in W^{1,\infty}_{\mathrm{loc}}([0,\infty[;\ W^{-1,p'}(\Omega)) \cap C([0,\infty[;\ L^2(\Omega)_w) \tag{6.17}$$

$$v(t,x) \in \beta(u(t,x)) \quad \text{a.e. } x \in \Omega, \quad t > 0 \tag{6.18}$$

$$\frac{\partial v}{\partial t} - \sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\right) = g, \qquad 0 < t < \infty, \quad x \in \Omega \tag{6.19}$$

$$v(0,x) = v_0(x) \quad \text{a.e. } x \in \Omega. \tag{6.20}$$

Remark. In the special case $\beta(r)=|r|^{\alpha-2}r$, $\alpha \geqq 2$, the above problem has been studied by Raviart [14] (see also [10], [12]).

Since our result allows multivalued functions β (in particular, we might take β as the subdifferential of indicator function of some closed, convex subset of R), equation (6.10) includes a large class of parabolic variational inequalities associated with the pseudo-Laplacian A. In particular, the classical Stefan problem can be reformulated as (6.10) where $p=2$ and β is suitably chosen (see e.g. J. L. Lions [13, p. 208]).

Acknowledgment. The author acknowledges gratefully the referee's remarks and suggestions on the paper, which helped him to improve its readability.

REFERENCES

[1] V. BARBU, *Integro-differential equations in Hilbert space*, An. Şti. Univ. "Al.I.Cuza" Iaşi Sect. I a Mat., 19 (1973), pp. 365–383.
[2] ———, *Nonlinear Semigroups and Differential Equations in Banach Spaces*, P. Noordhoff, Leyden, 1976.
[3] ———, *Nonlinear Volterra integro-differential equations in Hilbert space*, Conf. Sem. Matematico Univ. Bari, 1976.
[4] ———, *Existence theorems for a class of two point boundary value problems*, J. Differential Equations, 17 (1975), pp. 236–257.
[5] C. BARDOS AND H. BRÉZIS, *Sur une classe de problems d'evolution non lineaires*, Ibid., 6 (1969), pp. 343–345.
[6] H. BRÉZIS, *Opérateurs Maximaux Monotones et Semigroupes de Contractions dans les Espaces de Hilbert*, Math. Studies 5, North-Holland, Amsterdam, 1973.
[7] H. BRILL, *Sobolev equations in Banach space*, J. Differential Equations, 24 (1977), pp. 412–425.
[8] M. G. CRANDALL AND A. PAZY, *Semigroups of nonlinear contractions and dissipative sets*, J. Functional Analysis, 3 (1969), pp. 376–418.
[9] M. G. CRANDALL, S. O. LONDEN AND J. A. NOHEL, *An abstract nonlinear Volterra integro-differential equation*, J. Math. Anal. Appl., to appear.
[10] J. DUBINSKI, *Weak convergence in nonlinear elliptic and parabolic equations*, Mat. Sb., 67(109) 1965, pp. 609–642; English transl., Amer. Math. Soc. Transl., 67 (1968), no. 2, pp. 226–258.
[11] H. GAJEWSKI AND K. ZACHARIAS, *Über eine Klasse nichtlinearer Differential gleichungen im Hiltertraum*, J. Math. Anal. Appl., 34 (1973), no. 2, pp. 71–87.
[12] O. GRANGE AND F. MIGNOT, *Sur la résolution d'une équation et d'une inéquation paraboliques non linéaires*, J. Functional Analysis, 11 (1972), pp. 77–92.
[13] J. L. LIONS, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires*, Dunod, Paris, 1969.
[14] J. MOREAU, *Fonctionelles convexes*, Mimeographed lecture notes, College de France, Paris, 1966–1967.
[15] P. A. RAVIART, *Sur la résolution de certaines équations paraboliques non linéaires*, J. Functional Analysis, 5 (1970), pp. 299–328.
[16] R. E. SHOWALTER, *A nonlinear parabolic-Sobolev equation*, J. Math. Anal. Appl., 50 (1975), pp. 183–190.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0013 $01.00/0

ON THE SPEED OF PROPAGATION OF SOLUTIONS OF INTEGRODIFFERENTIAL EQUATIONS*

PAUL L. DAVIS†

Abstract. Linear Volterra integrodifferential equations of hyperbolic type from the electromagnetic theory of dielectrics, heat conduction theory of materials with memory and from viscoelasticity theory are examined. For a particular class of kernels solutions of initial value problems for these equations can all be split into a part propagating outwards and a part whose support remains within the support of the forcing term. The speed of the outward propagating part of each equation is exhibited.

1. Introduction. The notion of hyperbolicity was extended to integrodifferential equations of Volterra type in [1]. Equations arising in the electromagnetic theory of dielectrics and in heat conduction for materials with memory were examined in [2] and [3] respectively and sufficient conditions for hyperbolicity were given. An equation is said to be hyperbolic if solutions of an appropriate problem propagate with finite speed. In [2] and [3] we prove that this occurs without exhibiting the speeds of propagation. We now prove that for the class of kernels considered in [2] and [3] the solutions of these equations and an equation from viscoelasticity theory can be represented as the sum of a function that behaves as a solution of the wave equation and a function that propagates with zero speed. The latter function is said to be trivially hyperbolic. We exhibit the speeds of propagation of the first function. They are the speeds of propagation of the integrodifferential equation.

A summary of [1], [2], [3] is given in § 2. By carefully examining the roots of certain polynomial equations in two variables in § 3, we reach the desired characterization of solutions of these integrodifferential equations.

2. Background. Equations of the form

$$u(x, t)=\sum_{\nu=1}^{p}\int_0^t k_\nu(t-s)L_\nu u(x, s)\,ds+g(x, t) \tag{2.1}$$

when $x=(x_1, \cdots, x_n)$ and L_ν is a constant coefficient differential operator with respect to these variables are considered in [1]. A kernel k_ν is said to be in the set $\mathscr{A}$ if its Laplace transform $\tilde{k}_\nu$ is a rational function (degree of numerator less than that of denominator). In all that follows kernels are assumed to be in $\mathscr{A}$.

DEFINITION. The above equation is hyperbolic if for each $g\in C(R^n\times[0,\infty))$ that is infinitely differentiable in x for each $t>0$ and satisfies $g(x, t)=0$ when $|x_i|\geqq b_i$, $i=1, \cdots, n$, there exists a unique solution $u\in C(R^n\times[0,\infty))$ that is infinitely differentiable in x for each $t>0$ and has finite signal speed; i.e., there are $c_i\geqq 0$ such that u at time t, $t>0$, vanishes outside of $\{-b_i-c_it<x_i<b_i+c_it \mid i=1, \cdots, n\}$.

Denote the polynomial associated with L_ν by $Q_\nu(\zeta)$; that is,

$$L_\nu=Q_\nu\left(\frac{\partial}{\partial x_i}, \cdots, \frac{\partial}{\partial x_n}\right).$$

The analysis in [1] leads to sufficient conditions for hyperbolicity in terms of

$$M(w, \zeta)=\left(\sum_{\nu=1}^{p} Q_\nu(i\zeta)k_\nu(w)\right)\left(1-\sum_{\nu=1}^{p} Q_\nu(i\zeta)k_\nu(w)\right)^{-1}. \tag{2.2}$$

* Received by the editors June 6, 1977, and in revised form December 12, 1978.
† 47 Grandview Avenue, White Plains, New York 10605.

To show that a particular equation is hyperbolic one uses a representation of $\hat{u}(\zeta, t)$, the Fourier transform of u with respect to the space variables and the following theorem:

PALEY–WIENER THEOREM [12]. *Suppose $F \in L^2(R^n)$. F is the Fourier transform of a function vanishing outside of $\{-b_i \leqq x_i \leqq b_i,\ i = 1, \cdots, n\}$ if and only if F is the restriction to R^n of an entire function $F(\zeta)$, $\zeta \in C^n$, of exponential type; that is, for each $\varepsilon > 0$, there is $C_\varepsilon > 0$ such that*

$$|F(\zeta)| \leqq C_\varepsilon \exp \sum_{i=1}^{n} (b_i + \varepsilon)|\zeta_i|.$$

When $k_\nu \in \mathcal{A}$ for $\nu = 1, \cdots, p$, $M(x, \zeta)$ is a rational function w for each ζ with poles $w_i(\zeta)$, $i = 1, \cdots, s(\zeta)$. Let $C(w, \delta)$ be a circle of radius δ about w. The representation of $\hat{u}$ is as follows:

$$\hat{u}(\zeta, t) = \hat{g}(\zeta, t) + \int_0^t M(t-s, \zeta)\hat{g}(\zeta, s)\, ds \tag{2.3}$$

where

$$M(t, \zeta) = \frac{1}{2\pi i} \sum_{j=1}^{s(\zeta)} \exp(w_j(\zeta)t) \int_{C(0,\delta(\zeta,j))} \exp(tw)\tilde{M}(w + w_j(\zeta), \zeta)\, dw \tag{2.4}$$

and $\delta(\zeta, j)$ is a sufficiently small radius.

Maxwell's equations together with the constitutive relations

$$\mathbf{D}(t) = \mathbf{E}(t) + \int_{-\infty}^{t} \varphi(t-s)\mathbf{E}(s)\, ds, \qquad \mathbf{H}(t) = \mu^{-1}\mathbf{B}(t)$$

are considered in [2]. Such relations are also considered in [8], [9], [13]. $\mathbf{E} = (E_1, E_2, E_3)$ is the electric field, $\mathbf{D}$ the electric displacement, $\mathbf{H}$ the magnetic intensity and $\mathbf{B}$ the magnetic flux density. Such constitutive relations serve to model electromagnetic fields propagating in a dielectric. We assume that $\varepsilon\mu = c^{-2}$ where c is the speed of light in vacuum. The speeds of propagation of the solution of the integrodifferential equation associated with these constitutive relations are shown in § 3 to be $\pm c$. It is shown in [2] that each component of $\mathbf{E}$ satisfies the equation

$$\Delta u = \mu\left(\varepsilon u + \int_{-\infty}^{t} \varphi(t-s)u(x, s)\, ds\right)_{tt}.$$

Assuming that u is known for $t \leqq 0$, we are led to an equation of the form (2.1) by integrating twice with respect to t and denoting all known quantities by g. The equation is given in the beginning of the next section.

A hyperbolic theory of rigid conductors of heat composed of materials with memory [4], [5], [10], [11] is based on the constitutive relations

$$e(t) = c\theta(t) + \int_{-\infty}^{t} \alpha(t-s)\theta(s)\, ds$$

and

$$\mathbf{q}(t) = -\int_{-\infty}^{t} \varphi(t-s)\nabla\theta(s)\, ds$$

and the energy-balance law

$$\dot{e} = -\nabla \cdot \mathbf{q} + r.$$

θ denotes the departure of the temperature from its reference value, e the internal energy, $\mathbf{q}$ the heat flux and r is the (known) heat supply. When these relations are substituted into the energy-balance law, we have

$$\frac{\partial}{\partial t}\left(c\theta(x, t) + \int_{-\infty}^{t} \alpha(t-s)\theta(x, s)\, ds\right) = \int_{-\infty}^{t} \varphi(t-s)\Delta\theta(x, s)\, ds + r(x, t).$$

Assuming that θ is known for $t \leqq 0$, we arrive at an equation of the form (2.1) by integrating once with respect to t and denoting all known quantities by g. The resulting equation is (3.2) of § 3. The speeds of propagation of its solution are $\pm(\varphi(0)/c)^{1/2}$ when this quantity is real. A thorough analysis of the linear equations of rigid heat conductors for a wider class of kernels is given in [16].

A one dimensional equation of linear viscoelasticity is given by

$$u_{tt} = c^2\frac{\partial^2 u}{\partial x^2} + \int_{-\infty}^{t} \varphi(t-s)\frac{\partial^2 u}{\partial x^2}(x, s)\, ds,$$

where $u(x, t)$ is the displacement at time t of that point in the material body which is at position x at time $t = 0$. A general theory of linear viscoelasticity is given in [6]. Existence, uniqueness and stability results for the general equations of linear viscoelasticity are given in [14] and [15] along with an examination of the above one dimensional equation. If u is known for $t \leqq 0$, another equation of the form (2.1) can be constructed from the one dimensional equation. It is equation (3.3) of the next section. The speeds of propagation associated with this equation are $\pm c$.

3. Polynomial analysis. The equations discussed in § 2 are as follows:

Electromagnetic theory:

$$u(x, t) = c^2 \int_0^t (t-s)\Delta u(x, s)\, ds - \varepsilon^{-1} \int_0^t \varphi(t-s)u(x, s)\, ds + g(x, t) \tag{3.1}$$

Heat conduction:

$$u(x, t) = -c^{-1} \int_0^t \alpha(t-s)u(x, s)\, ds + c^{-1} \int_0^t \left(\int_0^{t-s} \varphi(\tau)\, d\tau\right)\Delta u(x, s)\, ds + g(x, t) \tag{3.2}$$

Viscoelasticity:

$$u(x, t) = c^2 \int_0^t (t-s)\frac{\partial^2}{\partial x^2}u(x, s)\, ds + \int_0^t \left(\int_0^{t-s} (t-s-\tau)\varphi(\tau)\, d\tau\right)\frac{\partial^2}{\partial x^2}u(x, s)\, ds + g(x, t). \tag{3.3}$$

Let M_i, $i = 1, 2, 3$, be the expression (2.2) associated with equations (3.1), (3.2), (3.3) respectively. We have

$$M_1(w, \zeta) = -\frac{c^2|\zeta|^2 + \varepsilon^{-1}\tilde{\varphi}(w)w^2}{w^2 + c^2|\zeta|^2 + \varepsilon^{-1}\tilde{\varphi}(w)w^2}, \qquad M_2(w, \zeta) = -\frac{w\tilde{\alpha}(w) + \tilde{\varphi}(w)|\zeta|^2}{cw + w\tilde{\alpha}(w) + \tilde{\varphi}(w)|\zeta|^2},$$

$$M_3(w, \zeta) = -\frac{c^2\zeta^2 + \zeta^2\tilde{\varphi}(w)}{w^2 + c^2\zeta^2 + \zeta^2\tilde{\varphi}(w)}.$$

Let

$$\tilde{\varphi}(w)=\left(\sum_{j=0}^{m-1} a_j w^j\right)\left(\sum_{j=0}^{m} b_j w^j\right)^{-1}$$

where the numerator and denominator have no common divisors and $b_m \neq 0$. That is, $\varphi \in \mathscr{A}$. Assume that $\alpha \in \mathscr{A}$ also so that

$$\tilde{\alpha}(w)=\left(\sum_{j=0}^{q-1} c_j w^j\right)\left(\sum_{j=0}^{q} d_j w^j\right)^{-1}$$

where $d_q \neq 0$. Substituting these into $\tilde{M}_i$, $i=1,2,3$ we are led to the study of the following polynomials in two variables for electromagnetic theory, heat conduction and viscoelasticity, respectively.

$$P_1(w,z)=b_m w^{m+2}+c^2z^2\sum_{k=0}^{m} b_{m-k}w^{m-k}+\sum_{k=2}^{m}(b_{k-2}+\varepsilon^{-1}a_{k-2})w^k,$$

$$P_2(w,z)=c\sum_{j=0}^{q}\sum_{k=0}^{m} d_j b_k w^{j+k+1}+\sum_{j=0}^{q-1}\sum_{k=0}^{m} c_j b_k w^{j+k+1}+z^2\sum_{j=0}^{q}\sum_{k=0}^{m-1} a_k d_j w^{j+k},$$

$$P_3(w,z)=\sum_{j=0}^{m} b_j w^{j+2}+c^2z^2\sum_{j=0}^{m} b_j w^j+z^2\sum_{j=0}^{m-1} a_j w^j.$$

The behavior of the roots $w(|\zeta|)$ of $P_k(w,|\zeta|)=0$ as $|\zeta|\to\infty$ can be investigated by examining the roots of

$$R_k(w,z)=w^s z^2 P_k(1/w,1/z)$$

as $z\to 0$ where s is the highest power of w to appear in $P_k(w,z)$.

THEOREM 1. *Only two roots, $w_{k1}(z)$ and $w_{k2}(z)$, of $R_k(w,z)=0$, $k=1$ and 3, approach zero as $z\to 0$. These roots can be written as*

$$w_{k1}(z)=iz/c+\sum_{j=1}^{\infty} g_j z^{j+1}, \qquad w_{k2}(z)=-iz/c+\sum_{j=1}^{\infty} g_j z^{j+1}.$$

(The g_j are generic constants.) *If $c\varphi(0)>0$, two roots of $R_2(w,z)=0$ approach zero as $z\to 0$ also. Moreover,*

$$w_{21}(z)=iz(c/\varphi(0))^{1/2}+\sum_{j=1}^{\infty} g_j z^{j+1}$$

and

$$w_{22}(z)=-iz(c/\varphi(0))^{1/2}+\sum_{j=1}^{\infty} g_j z^{j+1}.$$

Proof. Since $s=m+2$ for R_1 and R_3 and $s=m+q+1$ for R_2,

$$R_1(w,z)=b_m z^2+c^2\sum_{k=0}^{m} b_{m-k}w^{k+2}+z^2\sum_{k=2}^{m}(b_{k-2}+\varepsilon^{-1}a_{k-2})w^{m-k+2},$$

$$R_2(w,z)=cz^2\sum_{k=0}^{m} d_q b_k w^{m-k}+z^2\sum_{j=0}^{q-1}\sum_{k=0}^{m}(cd_jb_k+c_jb_k)w^{m+q-k-j}$$

$$+\sum_{j=0}^{q}\sum_{k=0}^{m-1} a_k d_j w^{m+q+1-k-j}$$

and

$$R_3(w, z) = z^2 \sum_{j=0}^{m} b_j w^{m-j} + c^2 \sum_{j=0}^{m} b_j w^{m-j+2} + \sum_{j=0}^{m-1} a_j w^{m-j+2}.$$

Since $R_1(w, 0)$ and $R_3(w, 0)$ have $c^2 b_m w^2$ as lowest order term and $R_2(w, 0)$ has $a_{m-1} d_q w^2$ as lowest order term, each of $R_k(w, z)$ has two roots approaching zero as z approaches zero. (It is shown in Lemma 3.1 of [3] that $\varphi(0) = a_{m-1}/b_m$ which is not zero.) Motivated by Newton diagrams [7], we let $z = t$ and $w = tu$. Consider $G_k(t, u) = R_k(tu, t)/t^2$. For $k = 1$ and 3, we have

$$G_k(t, u) = b_m(1 + c^2 u^2) + tH_k(t, u).$$

Moreover,

$$G_2(t, u) = a_{m-1} d_q (cb_m/a_{m-1} + u^2) + tH_2(t, u).$$

It follows that $G_k(0, \pm i/c) = 0$ and

$$\frac{\partial}{\partial u} G_k(0, \pm i/c) \neq 0$$

for $k = 1$ and 3. Since $a_{m-1}/b_m = \varphi(0)$, $G_2(0, \pm i(c/\varphi(0))^{1/2}) = 0$ and

$$\frac{\partial}{\partial u} G_2(0, \pm i(c/\varphi(0))^{1/2}) \neq 0.$$

Whenever $G_k(0, s) = 0$ and $(\partial/\partial u) G_k(0, s) \neq 0$ for some complex number s, the implicit function theorem implies the existence of $u(t)$ analytic in a neighborhood of the origin such that $u(0) = s$ and $G_k(t, u(t)) = 0$. Letting $w(z) = zu(z)$, we have $R_k(w(z), z) = 0$ where (with g_j as generic constants)

$$w(z) = sz + \sum_{j=1}^{\infty} g_j z^{j+1}.$$

The theorem now follows by choosing the appropriate values of s.

For $|\zeta|$ sufficiently large, the roots $W_{km}(|\zeta|)$ of $P_k(w, |\zeta|) = 0$ are given by

$$W_{km}(|\zeta|) = 1/w_{km}(1/|\zeta|),$$

where $R_k(w_{km}(z), z) = 0$.

COROLLARY. *Let $w_{k1}(z)$ and $w_{k2}(z)$ be the two roots approaching zero as z approaches zero. For $k = 1$ and 3,*

$$W_{k1}(|\zeta|) = ic|\zeta| + O(|\zeta|)$$

and

$$W_{k2}(|\zeta|) = -ic|\zeta| + O(|\zeta|).$$

Moreover,

$$W_{21}(|\zeta|) = i(\varphi(0)/c)^{1/2}|\zeta| + O(|\zeta|)$$

and

$$W_{22}(|\zeta|) = -i(\varphi(0)/c)^{1/2}|\zeta| + O(|\zeta|).$$

Proof. If $R_k(w(z), z) = 0$ where

$$w(z) = sz + \sum_{j=1}^{\infty} g_j z^{j+1},$$

then

$$
\begin{aligned}
W(|\zeta|) &= 1/w(1/|\zeta|)\\
&= \left((s/|\zeta|) + \sum_{j=1}^{\infty} g_j|\zeta|^{-j-1}\right)^{-1}\\
&= \frac{|\zeta|}{s}(1+Y(|\zeta|))^{-1}\\
&= \frac{|\zeta|}{s}\left(1-\frac{Y(|\zeta|)}{1+Y(|\zeta|)}\right)
\end{aligned}
$$

where $|\zeta|Y(|\zeta|)$ is bounded as $|\zeta|\to\infty$. The corollary follows when we choose s appropriately.

DEFINITION. A function $u(x, t)$ satisfying $u(x, 0)=0$ for $x_k\notin[-b_k, b_k]$, $k=1,\cdots,n$, is said to *propagate with speed* c_k *in the* x_k *direction* if $u(x,t)=0$ for $x_k\notin[-b_k-c_kt, b_k+c_kt]$, $t>0$.

THEOREM 2. *Assume $g(x, t)$ propagates with speed 0 in each direction and that it is infinitely differentiable. If $\varphi\in\mathscr{A}$ and $\alpha\in\mathscr{A}$, then solutions of (3.2) when $c\varphi(0)>0$, and of (3.1) and (3.3) exist, are unique and infinitely differentiable. Furthermore the solutions of (3.1) and (3.3) can each be written as the sum of three functions, one propagating with speed 0 in each direction (trivially hyperbolic), the other two propagating with speeds c and $-c$ in each direction. The same can be said of (3.2) except that the speeds of propagation are $\pm(\varphi(0)/c)^{1/2}$.*

Proof. Existence, uniqueness and differentiability results are contained in [2] and [3] for (3.1) and (3.2). These results for (3.3) follow in a similar manner. The Fourier transform of each of the hyperbolic parts can be represented using (2.3) and (2.4). Terms containing expressions of the form

$$
(2\pi i)^{-1}\exp(W_{km}(|\zeta|)t)\int_{C(0,\delta(\zeta,m))}\exp(tw)\tilde{M}_k(w+W_{km}(|\zeta|),\zeta)\,dw,
$$

$m=1, 2$, give rise to the nontrivial hyperbolic parts. The remaining terms represent the Fourier transform of the trivial hyperbolic part. The previous corollary together with the Paley–Wiener theorem are all that are needed to show that each of these parts has the desired properties.

REFERENCES

[1] P. L. DAVIS, *Hyperbolic integrodifferential equations*, Proc. Amer. Math. Soc., 47 (1975), pp. 155–160.
[2] ———, *Hyperbolic integrodifferential equations arising in the electromagnetic theory of dielectrics*, J. Differential Equations, 18 (1975), pp. 170–178.
[3] ———, *On the hyperbolicity of the equations of the linear theory of heat conduction for materials with memory*, SIAM J. Appl. Math., 30 (1976), pp. 75–80.
[4] J. M. FINN AND L. T. WHEELER, *Wave propagational aspects of the generalized theory of heat conduction*, Z. Angew. Math. Phys., 23 (1972), pp. 927–940.
[5] M. E. GURTIN AND A. G. PIPKIN, *A general theory of heat conduction with finite wave speeds*, Arch. Rational Mech. Anal., 31 (1968), pp. 113–126.
[6] M. E. GURTIN AND E. STERNBERG, *On the linear theory of viscoelasticity*, Ibid., 11 (1962), pp. 291–356.
[7] E. HILLE, *Analytic Function Theory*, vol. II, Ginn and Company, Boston, 1962.
[8] J. HOPKINSON, *The residual charge of the Leyden jar*, Philos. Trans. Roy. Soc. London, 167 (1877), pp. 599–626.
[9] L. D. LANDAU AND E. M. LIFSHITZ, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford, 1960.

[10] J. W. NUNZIATO, *On heat conduction in materials with memory*, Quart. Appl. Math., 29 (1971), pp. 187–204.

[11] ———, *On uniqueness in the linear theory of heat conduction with finite wave speeds*, SIAM J. Appl. Math., 25 (1973), pp. 1–4.

[12] E. M. STEIN AND G. WEISS, *Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, NJ, 1971.

[13] R. A. TOUPIN AND R. S. RIVLIN, *Linear functional electromagnetic constitutive relations and plane waves in a hemihedral isotropic material*, Arch. Rational Mech. Anal., 6 (1960), pp. 188–197.

[14] C. DAFERMOS, *An abstract Volterra equation with applications to linear viscoelasticity*, J. Differential Equations, 7 (1970), pp. 554–569.

[15] ———, *Asymptotic stability of viscoelasticity*, Arch. Rational Mech. Anal., 37 (1970), pp. 297–308.

[16] R. K. MILLER, *An integrodifferential equation for rigid heat conductors with memory*, submitted for publication.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0014 $01.00/0

GEGENBAUER TRANSFORMS VIA THE RADON TRANSFORM*

STANLEY R. DEANS†

Abstract. Use is made of the Radon transform on even dimensional spaces and Gegenbauer functions of the second kind to obtain a general Gegenbauer transform pair. In the two-dimensional limit the pair reduces to a Chebyshev transform pair.

1. Introduction. Gegenbauer polynomials of the first kind appear in a natural way when studying the Radon transform of functions which have certain spherical symmetry. We shall make use of this property of the Radon transform to obtain a new Gegenbauer transform pair. Although the final result does not contain Gegenbauer functions of the second kind, these functions are important in the derivation and their use here supplements the informative recent study of these particular special functions by Durand [1] and by Durand, Fishbane, and Simmons [2].

The work which follows serves a threefold purpose. First, we are able to demonstrate an important use of the Radon transform as a tool. Second, more insight is obtained regarding the use of Gegenbauer functions of the second kind. Useful material on these functions is contained in the Appendix. Finally, we derive a set of equations which constitute a Gegenbauer transform pair with a fundamental connection to the dimensionality of the space $\mathbb{R}^n$.

2. The Radon transform. Let $x=(x_1, x_2, \cdots, x_n)$ be a point in $\mathbb{R}^n$ $(n \geqq 2)$ and let $F \in \mathcal{S}$ be a function of the n real variables $x_1, x_2, \cdots, x_n$. The properties of the space $\mathcal{S}$, which consists of all rapidly decreasing C^∞ functions on $\mathbb{R}^n$, are developed by Schwartz [3]. The reason for working in a space with such nice properties will be clear when it becomes necessary to make changes in the order of integration and to perform repeated integrations by parts.

Given $F \in \mathcal{S}$, the Radon transform of F is given by [4],

$$f(\xi, p)=\int F(x)\delta(p-\xi\cdot x)\,dx, \tag{1}$$

where p is real, ξ is an arbitrary unit vector in $\mathbb{R}^n$, $\xi\cdot x=\sum_{\kappa=1}^n \xi_\kappa x_\kappa$, δ is the Dirac delta function, $dx=dx_1\,dx_2\cdots dx_n$, and the integral is over the entire space. It is important to observe that the symmetry condition

$$f(\xi,-p)=f(-\xi,p) \tag{2}$$

follows directly from the definition (1).

Following the initial work by Radon [5], many of the technical properties of the Radon transform were worked out by several authors [4], [6]–[10]. Among other things these authors develop a formal expression for inversion of the transform, valid for functions in $\mathcal{S}$, and it turns out that the inversion formula for even n is considerably more complicated than the formula for odd n. There is a Hilbert transform associated with the even case which remains unevaluated for the most general functions. Our concern here is with this even n case exclusively and involves defining F in such a fashion that it is possible to perform the Hilbert transform.

---

* Received by the editors January 27, 1978, and in revised form June 8, 1978. This work was partially supported by the U.S. Department of Energy.

† Department of Physics, University of South Florida, Tampa, Florida 33620, and Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720.

3. Decomposition of *F*. We consider those functions F which may be decomposed either as

$$F(x)=G_l(r)S_{lm}(\hat{x}), \tag{3}$$

or as a linear combination of terms of this form. Here, $x \in \mathbb{R}^n$, $r=|x|$, $\hat{x}=x/r$, and the doubly subscripted $S_{lm}(\hat{x})$ is a real spherical (or surface) harmonic [11], [12] of degree l which comes from an orthonormal set with $N(n, l)$ members. That is, members of the set $\{S_{l1}(\hat{x}), S_{l2}(\hat{x}), \cdots, S_{lN(n,l)}(\hat{x})\}$ satisfy

$$\int_\Omega S_{lm}(\hat{x})S_{l'm'}(\hat{x})\,d\hat{x}=\delta_{ll'}\delta_{mm'} \tag{4}$$

where $d\hat{x}$ is understood to be the surface element in hyperspherical polar coordinates, and $\int_\Omega$ designates an integral over the unit sphere. A very useful property of the $S_{lm}(\hat{x})$ is that they satisfy the symmetry condition

$$S_{lm}(-\hat{x})=(-1)^l S_{lm}(\hat{x}). \tag{5}$$

Further properties, including an explicit expression for $N(n, l)$, can be found in Hochstadt [12].

4. Radon transform of the decomposed function. When the Radon transform (1) is applied to (3) the result is

$$f(\xi, p)=\int G_l(r)S_{lm}(\hat{x})\delta(p-\xi\cdot x)\,dx. \tag{6}$$

Without loss of generality we may assume that $p \geqq 0$ since one may always calculate $f(\xi, -p)$ from (2). If we convert (6) to spherical coordinates $(dx \to r^{n-1}\,dr\,d\hat{x})$ and observe that $\delta(p-r\xi\cdot\hat{x})=\delta(p/(\xi\cdot\hat{x})-r)/|\xi\cdot\hat{x}|$ for purposes of doing the r integration we obtain

$$f(\xi, p)=\int_\Omega S_{lm}(\hat{x})\left(\frac{p}{\xi\cdot\hat{x}}\right)^{n-1} G_l(p/(\xi\cdot\hat{x}))\frac{d\hat{x}}{|\xi\cdot\hat{x}|}. \tag{7}$$

Application of the Hecke–Funk theorem [12] yields

$$f(\xi, p)=\frac{\omega_{n-1}S_{lm}(\xi)}{C_l^\nu(1)}\int_0^1 \left(\frac{p}{t}\right)^{n-1} G_l\left(\frac{p}{t}\right)C_l^\nu(t)(1-t^2)^{\nu-1/2}\frac{dt}{t}, \tag{8}$$

where $C_l^\nu(t)$ is a Gegenbauer polynomial of the first kind, ω_n is the surface area of a unit sphere in $\mathbb{R}^n$, and $\nu=\frac{1}{2}(n-2)$. The ratio $\omega_{n-1}/C_l^\nu(1)$ may be written in terms of Gamma functions,

$$M_l^\nu=\omega_{n-1}/C_l^\nu(1)=\frac{(4\pi)^\nu\Gamma(l+1)\Gamma(\nu)}{\Gamma(l+2\nu)}. \tag{9}$$

Equation (8) may be converted to the desired form by making the change of variables $r=p/t$,

$$f(\xi, p)=M_l^\nu S_{lm}(\xi)\int_p^\infty r^{2\nu}G_l(r)C_l^\nu\left(\frac{p}{r}\right)\left[1-\left(\frac{p}{r}\right)^2\right]^{\nu-1/2}dr. \tag{10}$$

It will be especially useful to write this equation as

$$f(\xi, p)=g_l(p)S_{lm}(\xi), \tag{11}$$

where

$$g_l(p) = M_l^\nu \int_p^\infty r^{2\nu} G_l(r) C_l^\nu\left(\frac{p}{r}\right)\left[1-\left(\frac{p}{r}\right)^2\right]^{\nu-1/2} dr. \tag{12}$$

The symmetry conditions on f and S_{lm} yield the defining equation for $g_l(-p)$,

$$g_l(-p) = (-1)^l g_l(p). \tag{13}$$

5. The inversion. We now turn our attention to inverting the Radon transform when F is given by (3) and f is given by (11). The inversion may be written as an integration over a unit sphere in ξ space [9]

$$F(x) = \int_\Omega f^*(\xi, \xi \cdot x)\, d\hat{\xi}, \tag{14}$$

where

$$f^*(\xi, \xi \cdot x) = \Upsilon f(\xi, p). \tag{15}$$

(Keep in mind that ξ is already a unit vector. We have used the notation $\hat{\xi}$ in (14) to emphasize that the integration is over the unit sphere.) For even n the operator Υ is defined by

$$f^*(\xi, t) = \frac{(-1)^{n/2}}{2(2\pi)^{n-1}} \mathrm{H}\left\{\left(\frac{\partial}{\partial p}\right)^{n-1} f(\xi, p)\right\}, \tag{16}$$

and H designates the Hilbert transform

$$\mathrm{H}\{q(p)\} = \frac{1}{\pi}\int_{-\infty}^\infty \frac{q(p)}{p-t}\, dp. \tag{17}$$

After inserting the decompositions for F and f in (14) we have

$$\begin{aligned} G_l(r) S_{lm}(\hat{x}) &= \int_\Omega S_{lm}(\xi) g_l^*(r\xi \cdot \hat{x})\, d\hat{\xi} \\ &= M_l^\nu S_{lm}(\hat{x}) \int_{-1}^1 g_l^*(rt) C_l^\nu(t)(1-t^2)^{\nu-1/2}\, dt, \end{aligned} \tag{18}$$

where the second step was obtained by applying the Hecke–Funke theorem again [12].

By inspection of (18) it is clear that

$$G_l(r) = M_l^\nu \int_{-1}^1 g_l^*(rt) C_l^\nu(t)(1-t^2)^{\nu-1/2}\, dt \tag{19}$$

and g_l^* must be calculated from $g_l^* = \Upsilon g_l$.

6. Improvement on the inversion formula. For even n $(2, 4, 6, \cdots)$ and $\nu = \frac{1}{2}(n-2)$ it is possible to modify (19) considerably by actually doing the t integration. Explicitly, $g_l^*(rt)$ is given by

$$g_l^*(rt) = \frac{(-1)^{n/2}}{2(2\pi)^{n-1}} \frac{1}{\pi} \int_{-\infty}^\infty g_l^{(n-1)}(p)(p-rt)^{-1}\, dp, \tag{20}$$

where $g_l^{(n-1)}(p) = (d/dp)^{n-1} g_l(p)$. If (20) is substituted into (19) and the order of integration reversed, the equation for $G_l(r)$ becomes

$$G_l(r) = \frac{(-1)^{n/2}}{2(2\pi)^{n-1}} \frac{M_l^\nu}{\pi r} \int_{-\infty}^\infty g_l^{(n-1)}(p) I_l^\nu\left(\frac{p}{r}\right) dp, \tag{21}$$

where

$$I_l^\nu\left(\frac{p}{r}\right)=\int_{-1}^{1} C_l^\nu(t)\left(\frac{p}{r}-t\right)^{-1}(1-t^2)^{\nu-1/2}\,dt. \tag{22}$$

At this point it is clearly desirable to require that $r>0$. The $r=0$ case may be done separately starting with (19). The integration in (22) may be taken over four separate regions, $\int_{-\infty}^{-r}+\int_{-r}^{0}+\int_0^r+\int_r^\infty$. If we observe that $I_l^\nu(-p/r)=(-1)^{l+1}I_l^\nu(p/r)$ then by a change of variable $p\to -p$ over the negative p region in (21) it follows that

$$G_l(r)=\frac{(-1)^{n/2}}{(2\pi)^{n-1}}\frac{M_l^\nu}{\pi r}\left\{\int_0^r g_l^{(n-1)}(p)I_l^\nu\left(\frac{p}{r}\right)dp+\int_r^\infty g_l^{(n-1)}(p)I_l^\nu\left(\frac{p}{r}\right)dp\right\}. \tag{23}$$

The reason for writing $G_l(r)$ in this form is to enable us to evaluate the I_l^ν integrals.

Notice that in (23) the $\int_0^r$ integral forces $p/r\leqq 1$ and the $\int_r^\infty$ integral forces $p/r\geqq 1$. For convenience we momentarily designate

$$\frac{p}{r}=x \quad \text{if } \frac{p}{r}<1,$$

$$\frac{p}{r}=z \quad \text{if } \frac{p}{r}>1.$$

(Unlike prior usage of x, here x is a real variable rather than a vector.) This establishes contact with the usage of x and z in the Appendix and in [1], [2] where the Gegenbauer functions of the second kind D_λ^α are discussed.

From (A.1) and (A.3) we immediately obtain

$$I_l^\nu(x)=\pi(1-x^2)^{\nu-1/2}D_l^\nu(x) \tag{24}$$

and

$$I_l^\nu(z)=2\pi e^{-i\pi\nu}(z^2-1)^{\nu-1/2}D_l^\nu(z). \tag{25}$$

These results, combined with (23) give

$$\begin{aligned}G_l(r)=\frac{M_l^\nu}{(2\pi)^{n-1}r}\bigg\{&(-1)^{n/2}\int_0^r g_l^{(n-1)}(p)D_l^\nu(x)(1-x^2)^{\nu-1/2}\,dp\\&-2\int_r^\infty g_l^{(n-1)}(p)D_l^\nu(z)(z^2-1)^{\nu-1/2}\,dp\bigg\}.\end{aligned} \tag{26}$$

By use of (A.4) and (A.14) the two integrals may be combined, and after some simplification we find

$$\begin{aligned}G_l(r)=&\frac{\Gamma(l+1)2^{-\nu}\pi^{-\nu-1}}{\Gamma(l+2\nu)r}\int_l^\infty g_l^{(n-1)}(p)E_{l-2\nu-1}^\nu\left(\frac{p}{r}\right)dp\\&-\frac{\Gamma(l+1)\Gamma(\nu)}{2\pi^{\nu+1}\Gamma(l+2\nu)r}\int_r^\infty g_l^{(n-1)}(p)C_l^\nu\left(\frac{p}{r}\right)\left[\left(\frac{p}{r}\right)^2-1\right]^{\nu-1/2}dp.\end{aligned} \tag{27}$$

The $\int_0^\infty$ integral can be shown to vanish. To see this, first perform $n-1$ integrations by parts to obtain an integral of the form

$$\int_0^\infty g_l(p)Q_{l-2}\left(\frac{p}{r}\right)dp,$$

where $Q_{l-2}(z)$ is a polynomial of degree $l-2$, and $Q_{l-2}(-z)=(-1)^lQ_{l-2}(z)$. (The integrated parts always vanish by symmetry.) Next, make use of (12) to replace $g_l(p)$. This leads to an integral of the form

$$\int_0^\infty dp\, Q_{l-2}\left(\frac{p}{r}\right) \int_p^\infty dt\, t^{2\nu} G_l(t) C_l^\nu\left(\frac{p}{t}\right)\left[1-\left(\frac{p}{t}\right)^2\right]^{\nu-1/2}.$$

A change in the order of integration over the indicated region of the pt plane leads to

$$\int_0^\infty dt\, t^{2\nu} G_l(t) \int_0^t dp\, Q_{l-2}\left(\frac{p}{r}\right) C_l^\nu\left(\frac{p}{t}\right)\left[1-\left(\frac{p}{t}\right)^2\right]^{\nu-1/2}.$$

Now the p integration can be shown to vanish. If the variable change $p = yt$ is made, and the symmetry of the functions in the integrand taken into account the p integral becomes (aside from a constant factor)

$$\int_{-1}^1 Q_{l-2}\left(\frac{ty}{r}\right) C_l^\nu(y)(1-y^2)^{\nu-1/2}\, dy.$$

Since Q_{l-2} is a polynomial of degree $l-2$ in y it follows by orthogonality that this integral vanishes.

Hence we finally have the desired result, which consists of the Gegenbauer transform pair,

$$G_l(r) = \frac{-\Gamma(l+1)\Gamma(\nu)}{2\pi^{\nu+1}\Gamma(l+2\nu)r} \int_r^\infty g_l^{(n-1)}(p) C_l^\nu\left(\frac{p}{r}\right)\left[\left(\frac{p}{r}\right)^2 - 1\right]^{\nu-1/2} dp \tag{28}$$

and

$$g_l(p) = \frac{(4\pi)^\nu \Gamma(l+1)\Gamma(\nu)}{\Gamma(l+2\nu)} \int_p^\infty r^{2\nu} G_l(r) C_l^\nu\left(\frac{p}{r}\right)\left[1-\left(\frac{p}{r}\right)^2\right]^{\nu-1/2} dr, \tag{29}$$

where $\nu = \frac{1}{2}(n-2)$ and the dimensionality n is even $(n = 2, 4, 6, \cdots)$.

7. Limiting case $n = 2$. It is especially interesting to examine the $n = 2$ limiting case of the above transform pair since that corresponds to the Radon transform on a plane. The result is straightforward if one first multiplies by ν/ν and then lets $\nu \to 0$. The result is the Chebyshev transform pair [13]

$$G_l(r) = \frac{-1}{\pi r} \int_r^\infty g_l'(p) T_l\left(\frac{p}{r}\right)\left[\left(\frac{p}{r}\right)^2 - 1\right]^{-1/2} dp \tag{30}$$

and

$$g_l(p) = 2 \int_p^\infty G_l(r) T_l\left(\frac{p}{r}\right)\left[1-\left(\frac{p}{r}\right)^2\right]^{-1/2} dr. \tag{31}$$

8. Closing remarks. The connection between Radon and Gegenbauer transforms was observed by Ludwig [9] in his discussion of the Radon transform on Euclidean space where it was pointed out that results about the Radon transform imply corresponding results for Gegenbauer transforms and conversely. (Apparently the Radon transform—Gegenbauer transform connection was known prior to Ludwig's work, but not published. See the remark in the introduction to Ludwig's paper.) For spaces of even dimension the inversion formula obtained in [9] involved both a Hilbert transform and an integration with a Gegenbauer polynomial. The work here shows that after the Hilbert transform is performed the resulting inversion formula (28) still may be expressed as one member of a Gebenbauer transform pair. Recent work with odd dimensional spaces [14] shows that with minor modification the inversion formula (28) also holds for odd n. The only modification required is that the overall minus sign in (28) must be changed to plus. Finally, it should be pointed out that the transform pair (28), (29) is not the only possible transform pair involving Gegenbauer polynomials, and the method here using the Radon transform is not the only way to obtain such pairs. For example, another transform pair was obtained by Higgins [15] using a very different approach, and a still different procedure developed by Sneddon [16] utilizing the Mellin transform could be adapted to solving the inversion problem discussed here.

Appendix. In this appendix we collect several formulas which are needed in the preceding work. Some of these are included for convenience and may be found in [2]. Others, notably those involving the Chebyshev functions, do not seem to be available in the standard sources. Our notation and conventions conform to that used by Durand, Fishbane, and Simmons [2], since their treatment of the Gegenbauer functions is the best available source for the type of results needed here. These authors derive many properties of the Gegenbauer functions of the first kind $C_\lambda^\alpha(z)$ and second kind $D_\lambda^\alpha(z)$ for general values of α, λ, and z. Our concern here is primarily with the restricted case where both α and λ are nonnegative integers (designated by writing $\alpha=\nu$ and $\lambda=l$) and z is real. We use x (in place of z) to emphasize that the argument lies on the interval $[-1,+1]$ or $[0,1]$ and z whenever the argument is complex or greater than unity.

For integral λ and $\operatorname{Re}\alpha>-\frac{1}{2}$, D_λ^α and C_λ^α are related by [2]

$$D_l^\alpha(z)=e^{i\pi\alpha}(z^2-1)^{1/2-\alpha}\frac{1}{2\pi}\int_{-1}^{1}C_l^\alpha(t)(z-t)^{-1}(1-t^2)^{\alpha-1/2}\,dt. \tag{A.1}$$

To obtain $D_l^\alpha(x)$ we make use of the general prescription

$$\begin{gathered}D_\lambda^\alpha(x)=\lim_{\varepsilon\to0^+}\frac{e^{-i\pi\alpha}}{i}[e^{i\pi\alpha}D_\lambda^\alpha(x+i\varepsilon)-e^{-i\pi\alpha}D_\lambda^\alpha(x-i\varepsilon)],\\ (z^2-1)\to\begin{cases}(1-x^2)\,e^{i\pi} & \text{for } (x+i\varepsilon),\\ (1-x^2)\,e^{-i\pi} & \text{for } (x-i\varepsilon).\end{cases}\end{gathered} \tag{A.2}$$

This yields

$$D_l^\alpha(x)=(1-x^2)^{1/2-\alpha}\frac{1}{\pi}\int_{-1}^{1}C_l^\alpha(t)(x-t)^{-1}(1-t^2)^{\alpha-1/2}\,dt. \tag{A.3}$$

For integral $\alpha=\nu$ the following relation holds

$$(z^2-1)^{\nu-1/2}D_l^\nu(z)=\tfrac{1}{2}(z^2-1)^{\nu-1/2}C_l^\nu(z)-\frac{2^{-\nu}}{\Gamma(\nu)}E_{l+2\nu-1}^\nu(z), \tag{A.4}$$

where $E_{l+2\nu-1}^\nu(z)$ is a polynomial of degree $l+2\nu-1$. These polynomials can be expressed in terms of associated Legendre functions of the first kind,

$$E_{l+2\nu-1}^\nu(z)=\left(\frac{\pi}{2}\right)^{1/2}(z^2-1)^{1/2(\nu-1/2)}P_{l+\nu-1/2}^{\nu-1/2}(z). \tag{A.5}$$

Explicitly, in terms of Chebyshev polynomials of the first kind T_l and second kind U_l (the argument may be either z or x),

$$\begin{aligned}E_{l-1}^0&=\frac{1}{l}U_{l-1},\\ E_{l+1}^1&=T_{l+1},\\ E_{l+3}^2&=\tfrac{1}{2}[(l+1)T_{l+3}-(l+3)T_{l+1}].\end{aligned} \tag{A.6}$$

In general,

$$E^{\nu}_{l+2\nu-1} = 2^{1-\nu} \sum_{k=0}^{\nu-1} (-1)^k \binom{\nu-1}{k} \frac{(l+\nu-k-1)!(l+2\nu-1)!}{l!\,(l+2\nu-k-1)!} T_{l+2\nu-2k-1}, \tag{A.7}$$

where we have used the standard symbols for factorials and binomial coefficients. The E's satisfy the recursion relation

$$E^{\nu}_{l+2\nu-1}(z) = (l+1)zE^{\nu-1}_{l+2\nu-2}(z) - (l+2\nu-2)E^{\nu-1}_{l+2\nu-3}(z) \tag{A.8}$$

and the symmetry property

$$E^{\nu}_{l+2\nu-1}(-z) = (-1)^{l+1}E^{\nu}_{l+2\nu-1}(z). \tag{A.9}$$

Explicit results for the functions D^{ν}_l may be written conveniently in terms of the function $V_l(z)$ where

$$V_l(z) = T_l(z) - (z^2-1)^{1/2}U_{l-1}(z), \tag{A.10}$$

with recursion relation

$$V_{l+2}(z) = 2zV_{l+1}(z) - V_l(z). \tag{A.11}$$

We have

$$\begin{aligned} D^0_l(z) &= \lim_{\alpha\to 0} \frac{1}{\alpha} D^{\alpha}_l(z) = \frac{1}{l} V_l(z), \\ D^1_l(z) &= -\tfrac{1}{2}(z^2-1)^{-1/2} V_{l+1}(z), \\ D^2_l(z) &= -\tfrac{1}{4}(z^2-1)^{-3/2}\, \tfrac{1}{2}[(l+1)V_{l+3}(z) - (l+3)V_{l+1}(z)]. \end{aligned} \tag{A.12}$$

In general,

$$\begin{aligned} D^{\nu}_l(z) = \frac{-(z^2-1)^{1/2-\nu}}{2^{2\nu-1}\Gamma(\nu)} \sum_{k=0}^{\nu-1} (-1)^k \binom{\nu-1}{k} \\ \frac{(l+\nu+k-1)!\,(l+2\nu-1)!}{l!\,(l+2\nu-k-1)!} V_{l+2\nu-2k-1}(z). \end{aligned} \tag{A.13}$$

By application of (A.2) and (A.4),

$$D^{\nu}_l(x) = \frac{(-1)^{\nu+1}(1-x^2)^{1/2-\nu}}{2^{\nu-1}\Gamma(\nu)} E^{\nu}_{l+2\nu-1}(x). \tag{A.14}$$

The Chebyshev expansion for the C^{ν}_l is given by

$$\begin{aligned} C^0_l(z) &= \lim_{\alpha\to 0} \frac{1}{\alpha} C^{\alpha}_l(z) = \frac{2}{l} T_l(z), \\ C^1_l(z) &= U_l(z), \\ C^2_l(z) &= \tfrac{1}{4}(z^2-1)^{-1}[(l+1)U_{l+2}(z) - (l+3)U_l(z)]. \end{aligned} \tag{A.15}$$

In general,

$$C^{\nu}_l(z) = \frac{(z^2-1)^{1-\nu}}{4^{\nu-1}\Gamma(\nu)} \sum_{k=0}^{\nu-1} (-1)^k \binom{\nu-1}{k} \frac{(l+\nu-k-1)!\,(l+2\nu-1)!}{l!\,(l+2\nu-k-1)!} U_{l+2\nu-2k-2}(z). \tag{A.16}$$

These results also hold for $z \to x$.

For future reference, we examine another form for D_l^ν which is especially valuable for large values of l or z. We first define $\zeta=(z^2-1)^{1/2}$ and observe that

$$(z+\zeta)^{-1}=z-\zeta \quad \text{and} \quad (z+\zeta)^2-1=2\zeta(z+\zeta).$$

In terms of these variables we have

$$\begin{aligned} D_l^0(z)&=\frac{1}{l}(z-\zeta)^{-l}, \\ D_l^1(z)&=\frac{-1}{2\zeta}(z+\zeta)^{-l-1}, \\ D_l^2(z)&=\frac{l}{4\zeta^2}(z+\zeta)^{-l-2}\left\{1+\frac{z+2\zeta}{l\zeta}\right\}. \end{aligned} \tag{A.17}$$

Higher terms have the form

$$D_l^\nu(z)=\frac{(-1)^\nu l^{\nu-1}(z+\zeta)^{-l-\nu}}{2^\nu\Gamma(\nu)\zeta^\nu}\left\{1+\frac{\nu(\nu-1)}{2l}\cdot\frac{z+2\zeta}{\zeta}+O(l^{-2})\right\}. \tag{A.18}$$

Or, in general,

$$\begin{aligned} D_l^\nu(z)=&\frac{(-1)^\nu l^{\nu-1}(z+\zeta)^{-l-\nu}}{2^\nu\Gamma(\nu)\zeta^\nu}(2l\zeta)^{1-\nu}\sum_{k=0}^{\nu-1}(-1)^k\binom{\nu-1}{k} \\ &\cdot\frac{(l+k)!\,(l+2\nu-1)!}{l!\,(l+\nu+k)!}(z+\zeta)^{\nu-1-2k}. \end{aligned} \tag{A.19}$$

Acknowledgments. It is a pleasure to thank the Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico, for their hospitality during the summer of 1977, when this work originated. Discussions with Prof. T. F. Budinger and G. T. Gullberg are gratefully acknowledged along with helpful suggestions from the referee.

REFERENCES

[1] L. DURAND, *Nicholson-type integrals for products of Gegenbauer functions and related topics*, Theory and Application of Special Functions, R. A. Askey, ed., Academic Press, New York, 1975, pp. 353–374.
[2] L. DURAND, P. M. FISHBANE AND L. M. SIMMONS, JR., *Expansion formulas and addition theorems for Gegenbauer functions*, J. Math. Phys., 17 (1976), pp. 1933–1948.
[3] L. SCHWARTZ, *Mathematics for the Physical Sciences*, Addison Wesley, Reading, MA, 1966.
[4] I. M. GEL'FAND, M. I. GRAEV AND N. YA. VILENKIN, *Geneneralized Functions*, vol. V, Academic Press, New York, 1966.
[5] J. RADON, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*, Ber. Verh. Sächs. Akad., 69 (1917), pp. 262–277.
[6] F. JOHN, *Bestimmung einer Funktion aus ihren Integralen über gewisse Manningfaltigkeiten*, Math. Ann., 109 (1934), pp. 488–520.
[7] ———, *Plane Waves and Spherical Means Applied to Differential Equations*, Interscience, New York, 1955.
[8] S. HELGASON, *The Radon transform on Euclidean spaces, compact two-point homogeneous spaces and Grassman manifolds*, Acta Math., 113 (1965), pp. 153–180.
[9] D. LUDWIG, *The Radon transform on Euclidean space*, Comm. Pure. Appl. Math., 19 (1966), pp. 49–81.
[10] K. T. SMITH, D. C. SOLOMON AND S. L. WAGNER, *Practical and mathematical aspects of the problem of reconstructing objects from their radiographs*, Bull. Amer. Math. Soc., 83 (1977), pp. 1227–1270.
[11] A. ERDÉLYI, W. MAGNUS, F. OBERHETTINGER AND F. G. TRICOMI, *Higher Transcendental Functions*, vols. I, II, McGraw-Hill, New York, 1953.

[12] H. HOCHSTADT, *The Functions of Mathematical Physics*, John Wiley, New York, 1971.
[13] S. R. DEANS, *The Radon transform: Some remarks and formulas for two dimensions*, Lawrence Berkeley Laboratory Report LBL-5691, 1977.
[14] ———, *A unified Radon inversion formula*, J. Math. Phys., 19 (1978), pp. 2346–2349.
[15] T. P. HIGGINS, *An inversion integral for a Gegenbauer transformation*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 886–893.
[16] I. N. SNEDDON, *A procedure for deriving inversion formulae for integral transform pairs of a general kind*, Glagow Math. J., 67 (1968), pp. 67–77.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0015 $01.00/0

ALGEBRAIC METHOD FOR SOLVING LINEAR PARTIAL DIFFERENTIAL EQUATIONS WITH VARIABLE COEFFICIENTS. PART I: BASIC THEORY*

SERGE VASILACH†

Abstract. The present paper is devoted to an algebraic method of solving linear partial differential equations whose coefficients are functions of the independent variables. The method, based on the composition product of tensor products of kernel-distributions, permits the transformation of partial differential equations with variable coefficients, into linear equations of composition, whose elements belong to some composition algebra of kernel-distributions.

1. The composition algebras.

1.1. Introduction. It is known that the symbolic methods based on the convolution product and on the integral transformations of distributions, such as Fourier transform (cf. L. Schwartz [1, Chap. VII]) or Laplace transform (cf. J. Leray [2]: Garnir [3], Silva [4]), permit the transformation of linear differential and partial differential equations with constant coefficients, into algebraic equations of *convolution*.

Similar results can be obtained in the case of functions of one variable, by means of Heaviside's operational calculus (cf. Heaviside, [5]), whose mathematical foundation has been given by Mikusinski, (cf. [6]).

The extension of this operational calculus to functions and distributions of several variables is given in our previous papers [7], [8], [9].

Unfortunately, all these methods can be applied only to linear equations with constant coefficients. The problem of construction of an algebraic method for solving linear equations with variable coefficients is quite different.

We construct such a method on the one hand, by means of the composition product to be defined in tensor products of locally convex spaces, (cf. our previous papers [10], [11]), and on the other hand, by introducing a new class of distributions (cf. Vasilach, [12]).

Our method permits us to transform linear equations with variable coefficients, in algebraic composition equations. For the application of our method to linear ordinary differential equations with variable coefficients cf. Vasilach [14].

The present article is devoted to the basic theory of the extension of this algebraic method to linear partial differential equations with variable coefficients.

1.2. Composition algebras in $\mathscr{D}'_{(-\Gamma x)(+\Gamma y)}$ and $\mathscr{D}'_{(+\Gamma x)(-\Gamma y)}$. For the general algebraic definition of the composition product of tensor products of bimodules and of locally convex spaces, see [10] and [11].

Let X^n (resp. Y^n) be a topological space isomorphic to the Euclidean space $\mathbb{R}^n$, $n \geqq 1$. Let (cf. [8, Chap. II, § 2]) $\mathscr{D}_{(+\Gamma x)}$ (resp. $\mathscr{D}_{(-\Gamma y)}$) be the locally convex space of indefinitely differentiable functions with support limited to the left (resp. to the right) for $x \in X^n$ (resp. $y \in Y^n$). Let, (cf. [8], loc. cit.), $\mathscr{D}'_{(-\Gamma x)}$ (resp. $\mathscr{D}'_{(+\Gamma y)}$) be the strong dual of $\mathscr{D}_{(+\Gamma x)}$ (resp. $\mathscr{D}_{(-\Gamma y)}$). Let (cf. [12, § 2, pp. 2–7]) $\mathscr{D}_{(+\Gamma x)(-\Gamma y)}$ be the locally convex space of indefinitely differentiable functions with support limited to the left for $x \in X^n$ and to the right for $y \in Y^n$.

Let [12, § 2, pp. 7–9] $\mathscr{D}'_{(-\Gamma x)(+\Gamma y)}$ be the strong dual of $\mathscr{D}_{(+\Gamma x)(-\Gamma y)}$. Then, $\mathscr{D}'_{(-\Gamma x)(+\Gamma y)}$ is the locally convex space of distributions with support limited to the right for $x \in X^n$ and to the left for $y \in Y^n$.

* Received by the editors December 11, 1975, and in final revised form January 3, 1978.
† Département Mathématiques, Université Laval, Québec 10, Canada.

It is known [12, § 3, No. 2 Thm. 1] that (kernel theorem):

$$\mathscr{D}'_{(-\Gamma x)(+\Gamma y)} = \mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}. \tag{1.1}$$

For $\mathscr{D}_{(-\Gamma x)}$ considered as subspace of $\mathscr{D}'_{(-\Gamma x)}$, endowed with the topology induced by the latter, one has $\mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)} \subset \mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$, and $\mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma)y)}$ is a composition algebra for the operation of $(\mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}) \times (\mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)})$ into $\mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$ defined by

$$(S, T) \mapsto S \circ T = \int_{\Xi} S(x, \xi), T(\xi, y)\, d\xi \tag{1.2}$$

$$= \langle S(x, \xi), T(\xi, y) \rangle_{\xi}$$

where Ξ^n is a topological space isomorphic with R^n, $n \geqq 1$.

Moreover, if we take $y \leqq \xi \leqq x \Leftrightarrow y_j \leqq \xi_j \leqq x_j$ for all $j \in \{1, 2, \cdots, n\}$, we have [13, pp. 848–849]:

$$S \circ T = \int_y^x S(x, \xi) T(\xi, y)\, d\xi. \tag{1.3}$$

1.3. The composition modules $\mathscr{D}'_{(-\Gamma x)} \hat{\Theta} \mathscr{D}'_{(+\Gamma y)}$ and $\mathscr{D}'_{(+\Gamma x)} \hat{\Theta} \mathscr{D}'_{(-\Gamma y)}$. We have here a similar definition as for the composition algebra $\mathscr{D}'_{(-\Gamma x)} \hat{\otimes} D_{(+\Gamma y)}$. Then, $\forall S \in \mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$, $\forall T \in \mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$ we have $T \circ S \in \mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$.

Therefore $\mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$ is a right *composition module* over the composition algebra $\mathscr{D}_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$.

Likewise, $\mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$ is a *left composition module* over the composition algebra $\mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}_{(+\Gamma y)}$. In the same way one can show that $\mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$, the strong dual of $\mathscr{D}_{(-\Gamma x)(+\Gamma y)}$ is a *right* (resp. *left*) *composition module* over the composition algebra $\mathscr{D}_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$ (resp. $\mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}_{(-\Gamma y)}$).

Remark 1.1. The relation (1.3) shows that the composition operation is an extension of the composition product of kernel functions such as defined in the theory of Volterra's integral equations.

Remark 1.2. For $x \in X^n$, $y \in Y^n$ let

$$\delta(x-y) \overset{\text{def}}{=} \delta(y-x) = \bigotimes_{j=1}^{n} \delta(x_j - y_j)$$

be the Dirac kernel. Then (*cf.* [14, remark of § 1.2]) $\delta(x-y)$ is the unit element of the composition product, i.e.,

$$S \circ \delta(x-y) = \delta(x-y) \circ S = S, \quad \text{for all } S \in \mathscr{D}'_{xy}. \tag{*}$$

Indeed, (*) is true $\forall S \in \mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$ since $\delta(x-y)$ belongs to the composition algebras $\mathscr{D}_{(-\Gamma x)} \hat{\otimes} \mathscr{D}'_{(+\Gamma y)}$ and $\mathscr{D}'_{(-\Gamma x)} \hat{\otimes} \mathscr{D}_{(+\Gamma y)}$.

On the other hand, $\mathscr{D}'_x \hat{\otimes} \mathscr{D}_y$ and $\mathscr{D}_x \hat{\otimes} \mathscr{D}'_y$ are composition algebras for the same composition operation (1.2) and $\mathscr{D}'_x \hat{\otimes} \mathscr{D}'_y = \mathscr{D}'_{xy}$ is a right (resp. left) *composition module* over the composition algebra $\mathscr{D}_x \hat{\otimes} \mathscr{D}'_y$ (resp. $\mathscr{D}'_x \otimes \mathscr{D}_y$), whence (*).

Now, let

$$\partial^{k_1+k_2+\cdots+k_n}_{x_1^{k_1}, x_2^{k_2}, \cdots, x_n^{k_n}} \delta(x-y) = \bigotimes_{1=j}^{n} \delta^{k_j}_{x_j} (x_j - y_j)$$

be a derivative of $\delta(x-y)$. Then $\delta(x-y) \in \mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}_{(-\Gamma y)}$ implies

$$\bigotimes_{k=1}^{n} \delta^{k_j}_{x_j} (x_j - y_j) \in \mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}_{(-\Gamma y)}.$$

Therefore

$$\bigotimes_{k=1}^{n} \delta_{x_j^{k_j}}^{k_j}(x_j - y_j) \circ T \in \mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$$

$\forall T \in \mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$, because $\mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$ is a composition module.

2. Composition product of kernel-functions.

2.1. Preliminaries. Let $(L^1_{\text{loc}})_x(L^\infty_{\text{loc}})_y$ be the vector space of functions f defined in $X^n \times Y^n$, such that for each fixed $y \in Y^n$, f is locally integrable with respect to Lebesgue measure on X^n, and for each fixed $x \in X^n$, f is a locally bounded and measurable function of $y \in Y^n$.

Then, for arbitrary elements f, g of $(L^1_{\text{loc}x}(L^\infty_{\text{loc}})_y$, the composition product

$$(f \circ g)(x, y) = \int_y^x f(x, \xi) g(\xi, y)\, d\xi \tag{2.1}$$

has meaning for $-\infty < a \leqq y \leqq \xi \leqq x \leqq b < +\infty \Leftrightarrow -\infty < a_j \leqq y_j \leqq \xi_j \leqq x_j \leqq b_j < +\infty$, $\forall j \in (1, 2, \cdots, n)$, $d\xi = d\xi_1\, d\xi_2\, d\xi_3 \cdots d\xi_n$ and $\int_y^x d\xi = \int_{y_1}^{x_1} d\xi_1 \int_{y_2}^{x_2} d\xi_2 \cdots \int_{y_n}^{x_n} d\xi_n$.

2.2. Heaviside's kernel.

DEFINITION 2.1. We will call *Heaviside's kernel in* $X^N \times Y^n$, the

$$Y(x, y) = \bigotimes_{j=1}^{n} Y(x_j - y_j) = \begin{cases} 1 & \text{for } x_j \geqq y_j,\ j \in (1, 2, \cdots, n) \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2}$$

2.3. Kernel-functions in $\mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$.

DEFINITION 2.2. For f an arbitrary element of $(L^1_{\text{loc}})_x(L^\infty_{\text{loc}})_y$ we call *kernel-function* the element of $\mathscr{D}'_{(+\Gamma x)} \hat{\otimes} \mathscr{D}'_{(-\Gamma y)}$ defined by

$$\{f\} = Y(x, y) f(x, y) = \begin{cases} f(x, y) & \text{for } x \geqq y, \\ 0 & \text{elsewhere.} \end{cases} \tag{2.3}$$

2.4. Composition product of kernel-functions. For each pair (f, g) of elements of $(L^1_{\text{loc}})_x(L^\infty_{\text{loc}})_y$ the composition product of the kernel-functions $\{f\}$ and $\{g\}$ is given by

$$\begin{aligned} \{f\} \circ \{g\} &= \int_{\Xi^n} Y(x, \xi) f(x, \xi) Y(\xi, y) g(\xi, y)\, d\xi \\ &= \begin{cases} \int_y^x f(x, \xi) g(\xi, y)\, d\xi & \text{for } x \geqq y \\ 0 & \text{elsewhere} \end{cases} \end{aligned} \tag{2.4}$$

in which $\int_y^x f(x, \xi) g(\xi, y)\, d\xi$ is the composition product (2.1).

Remark 2.1. For $x \in X^n$, $y \in Y^n$, let $\mathscr{C}^{(l,m)}_{(x,y)}$ be the vector space of continuously differentiable functions of order $\leqq 1$ (resp. $\leqq m$) with respect to $x \in X^n$ (resp. $y \in Y^n$).

Let us set

$$|l| = l_1 + l_2 + \cdots + l_n, \qquad |m| = m_1 + m_2 + \cdots + m_n,$$

$$\frac{\partial^{|l|+|m|} f(x, y)}{\partial x_1^{l_1}\, \partial x_2^{l_2} \cdots \partial x_n^{l_n}\, \partial y_1^{k_1}\, \partial y_2^{k_2} \cdots \partial y_n^{k_n}} = \frac{\partial^{|l|+|m|} f(x, y)}{\partial x^l\, \partial y^m}.$$

Then, $\mathscr{C}^{(l,m)}_{(x,y)}$, provided with the composition product (2.1), is a topological algebra for the topology defined by the sequence of norms:

$$\|f\|_\nu = \operatorname{Sup} \frac{\partial^{|l|+|m|} f}{\partial x^s \, \partial y^t} \tag{2.5}$$
$$|s| \leqq |l|, \qquad |t| \leqq |m|, \qquad (x, y) \in K_\nu \times L_\nu,$$

for K_ν (resp. L_ν) an exhaustive sequence of compact subsets of X^n (resp. Y^n).

Clearly, each element of $\mathscr{C}^{(l,m)}_{(x,y)}$ becomes a kernel-function in $\mathscr{D}'_{(+\Gamma x)(-\Gamma y)}$, if we multiply it by the Heaviside kernel.

3. Derivation operators of kernel-functions.

3.1. Derivatives in $\mathscr{C}^{(1,m)}_{xy}$. In order to simplify our text, we will suppose in the sequel that $n=2$ and will consider functions of two variables $(x, y) \in R^2$ and of two variables $\alpha, \beta \in R^2$, such that

$$-\infty < \alpha \leqq \xi \leqq x \quad \text{and} \quad -\infty < \beta \leqq \eta \leqq y.$$

Let $\mathscr{C}^{(l,m)(p,q)}$ be the vector space of functions continuously differentiable of order $\leqq 1$ (resp. $\leqq m$) with respect to x (resp. y), and of order $\leqq p$ (resp. $\leqq q$) with respect to α (resp. β).

Then, we denote by $\mathscr{D}'_{(+\Gamma xy)(-\Gamma_{\alpha\beta})}$ the corresponding space of distributions, and for f an arbitrary element of $\mathscr{C}^{(l,m)(p,q)}$, we denote its kernel function in $\mathscr{D}'_{(+\Gamma xy)(-\Gamma_{\alpha\beta})}$ by

$$\{f\} = Y(x-\alpha) \otimes Y(y-\beta) f(x, y, \alpha, \beta).$$

Under these conditions we have

$$\frac{\partial\{f\}}{\partial x} = \delta(x-\alpha) \otimes \{f(\alpha, y, \alpha, \beta)\} + \left\{\frac{\partial f}{\partial x}\right\} \tag{3.1}$$

where $\{f(\alpha, y, \alpha, \beta)\} = Y(y-\beta) f(\alpha, y, \alpha, \beta)$ is the kernel-function with respect to the pair of variables (y, β).

Likewise, we obtain:

$$\frac{\partial\{f\}}{\partial y} = \{f(x, \beta, \alpha, \beta)\} \otimes \delta(y-\beta) + \left\{\frac{\partial f}{\partial y}\right\} \tag{3.2}$$

where $\{f(x, \beta, \alpha, \beta)\} = Y(x-\alpha) f(x, \beta, \alpha, \beta)$ is the kernel-function with respect to the pair of variables (x, α).

From (3.1) we obtain by recurrence:

$$\frac{\partial^\mu\{f\}}{\partial x^\mu} = \left\{\frac{\partial^\mu f}{\partial x^\mu}\right\} + \sum_{j=0}^{\mu-1} \delta_x^{(\mu-j-1)}(x-\alpha) \otimes \left\{\frac{\partial^j f(\alpha, y, \alpha, \beta)}{\partial x^j}\right\}, \tag{3.3}$$

$$\frac{\partial^\nu\{f\}}{\partial y^\nu} = \left\{\frac{\partial^\nu f}{\partial y^\nu}\right\} + \sum_{k=0}^{\nu-1} \left\{\frac{\partial^k f(x, \beta, \alpha, \beta)}{\partial y^k}\right\} \otimes \delta_y^{(\nu-k-1)}(y-\beta). \tag{3.4}$$

On the other hand, from (3.4) we obtain:

$$\begin{aligned} \frac{\partial^{1+\nu}\{f\}}{\partial x\, \partial y^\nu} &= \frac{\partial}{\partial x}\left\{\frac{\partial^\nu f}{\partial y^\nu}\right\} + \sum_{k=0}^{\nu-1} \frac{\partial}{\partial x}\left\{\frac{\partial^k f(x, \beta, \alpha, \beta)}{\partial y^k}\right\} \otimes \delta_y^{(\nu-k-1)}(y-\beta) \\ &= \left\{\frac{\partial^{1+\nu} f}{\partial x\, \partial y^\nu}\right\} + \delta(x-\alpha) \otimes \left\{\frac{\partial^\nu f(\alpha, y, \alpha, \beta)}{\partial y^\nu}\right\} \\ &\quad + \sum_{k-0}^{\nu-1} \frac{\partial}{\partial x}\left\{\frac{\partial^k f(x, \beta, \alpha, \beta)}{\partial y^k}\right\} \otimes \delta_y^{(\nu-k-1)}(y-\beta). \end{aligned} \tag{3.5}$$

But for $\{\partial^\nu f(\alpha, y, \alpha, \beta)/\partial y^\nu\}$ considered as a kernel-function with respect to the pair of variables (y, β) we have [14, (3.5)]:

$$\left\{\frac{\partial^\nu f(\alpha, y, \alpha, \beta)}{\partial y^\nu}\right\}$$
$$= \frac{\partial^\nu\{f(\alpha, y, \alpha, \beta)\}}{\partial y^\nu} - \sum_{k=0}^{\nu-1} \delta_y^{(\nu-k-1)}(y-\beta)\frac{\partial^k f(\alpha, \beta, \alpha, \beta)}{\partial y^k}, \tag{3.6}$$

where we have set

$$\frac{\partial^k f(\alpha, \beta, \alpha, \beta)}{\partial y^k} = \left.\frac{\partial^k f(\alpha, y, \alpha, \beta)}{\partial y^k}\right|_{y=\beta}.$$

Therefore

$$\begin{aligned}\frac{\partial^{1+\nu}\{f\}}{\partial x\, \partial y^\nu} = \left\{\frac{\partial^{1+\nu} f}{\partial x\, \partial y^\nu}\right\} + \sum_{k=0}^{\nu-1} \frac{\partial}{\partial x}\left\{\frac{\partial^k f(x, \beta, \alpha, \beta)}{\partial y^k} \otimes \delta_y^{(\nu-k-1)}(y-\beta) + \delta(x-\alpha)\right\} \\ \otimes \left[\frac{\partial^\nu f(\alpha, y, \alpha, \beta)}{\partial y^\nu} - \sum_{k=0}^{\nu-1} \delta_y^{(\nu-k-1)}(y-\beta)\frac{\partial^k f(\alpha, \beta, \alpha, \beta)}{\partial y^k}\right]\end{aligned} \tag{3.7}$$

whence

$$\begin{aligned}\frac{\partial^{(1+\nu)}\{f\}}{\partial x\, \partial y^\nu} &= \left\{\frac{\partial^{1+\nu} f}{\partial x\, \partial y^\nu}\right\} + \delta(x-\alpha) \otimes \frac{\partial^\nu\{f(\alpha, y, \alpha, \beta)\}}{\partial y^\nu} \\ &\quad - \delta(x-\alpha) \otimes \sum_{k=0}^{\nu-1} \delta_y^{(\nu-k-1)}(y-\beta)\frac{\partial^k f(\alpha, \beta, \alpha, \beta)}{\partial y^k} \\ &\quad + \sum_{k=0}^{\nu-1} \frac{\partial}{\partial x}\left\{\frac{\partial^k f(x, \beta, \alpha, \beta)}{\partial y^k}\right\} \otimes \delta_y^{(\nu-k-1)}(y-\beta).\end{aligned} \tag{3.8}$$

Finally, from (3.8) we obtain by recurrence:

$$\begin{aligned}\frac{\partial^{\mu+\nu}\{f\}}{\partial x^\mu\, \partial y^\nu} &= \left\{\frac{\partial^{\mu+\nu} f}{\partial x^\mu\, \partial y^\nu}\right\} + \frac{\partial^\mu}{\partial x^\mu}\sum_{l=1}^{\nu-1} \left\{\frac{\partial^l f(x, \beta, \alpha, \beta)}{\partial x^l}\right\} \otimes \delta_y^{(\nu-l-1)}(y-\beta) \\ &\quad + \frac{\partial^\nu}{\partial y^\nu}\sum_{k=0}^{\mu-1} \delta_x^{(\mu-k-1)}(x-\alpha) \otimes \left\{\frac{\partial^k f(\alpha, y, \alpha, \beta)}{\partial x^k}\right\} \\ &\quad - \sum_{k=0}^{\mu-1}\sum_{l=0}^{\nu-1} \delta_x^{(\mu-k-1)}(x-\alpha) \otimes \delta_y^{(\nu-l-1)}(y-\beta)\frac{\partial^{k+1} f(\alpha, \beta, \alpha, \beta)}{\partial x^k\, \partial y^l}.\end{aligned} \tag{3.9}$$

The fundamental formula (3.9) gives us the relation between the derivative $\partial^{\mu+\nu}\{f\}/(\partial x^\mu\, \partial y^\nu)$ of the kernel function $\{f\}$ and the kernel function of the derivative $\partial^{\mu+\nu} f/(\partial x^\mu\, \partial y^\nu)$ of the function f.

Remark 3.1. In the same way we obtain the derivatives of the kernel-functions, with respect to the pair of variables (α, β) and, also, with respect to the variables x, y, α, β.

Thus, for example, we obtain, with respect to the variables α, β.

$$\begin{aligned}\frac{\partial\{f\}}{\partial\alpha}&=-\delta(x-\alpha)\otimes\{f(x,y,x,\beta)\}+\left\{\frac{\partial f}{\partial\alpha}\right\}\\ \frac{\partial\{f\}}{\partial\beta}&=-\{f(x,y,\alpha,y)\}\otimes\delta(y-\beta)+\left\{\frac{\partial f}{\partial\beta}\right\},\end{aligned}\tag{3.10}$$

whence by recurrence:

$$\frac{\partial^\mu\{f\}}{\partial\alpha^\mu}=\left\{\frac{\partial^\mu f}{\partial\alpha^\mu}\right\}+\sum_{k-0}^{\mu-1}(-1)^{\mu-k}\delta_x^{(\mu-k-1)}(x-\alpha)\otimes\left\{\frac{\partial^k f(x,y,x,\beta)}{\partial\alpha^k}\right\}\tag{3.11}$$

and

$$\frac{\partial^\nu\{f\}}{\partial\beta^\nu}=\left\{\frac{\partial^\nu f}{\partial\beta^\nu}\right\}+\sum_{l-0}^{\nu-1}(-1)^{\nu-l}\left\{\frac{\partial^l f(x,y,\alpha,y)}{\partial\beta^l}\right\}\otimes\delta_y^{(\nu-l-1)}(y-\beta).\tag{3.12}$$

From (3.12) we obtain

$$\frac{\partial^{1+\nu}\{f\}}{\partial\alpha\,\partial\beta^\nu}=\frac{\partial}{\partial\alpha}\left\{\frac{\partial^\nu f}{\partial\beta^\nu}\right\}+\sum_{l=0}^{\nu-1}(-1)^{\nu-l}\frac{\partial}{\partial\alpha}\left\{\frac{\partial^l f(x,y,\alpha,y)}{\partial\beta^l}\right\}\otimes\delta_y^{(\nu-l-1)}(y-\beta).\tag{3.13}$$

But

$$\frac{\partial}{\partial\alpha}\left\{\frac{\partial^\nu f}{\partial\beta^\nu}\right\}=-\delta_{(x-\alpha)}\otimes\left\{\frac{\partial^\nu f(x,y,x,\beta)}{\partial\beta^\nu}\right\}+\left\{\frac{\partial^{1+\nu}}{\partial\alpha\,\partial\beta^\nu}\right\}$$

and

$$\left\{\frac{\partial^\nu f(x,y,x,\beta)}{\partial\beta^\nu}\right\}=\frac{\partial^\nu\{f(x,y,x,\beta)\}}{\partial\beta^\nu}-\sum_{j=0}^{\nu-1}(-1)^{\nu-j}\frac{\partial^j f(x,y,x,y)}{\partial\beta^j}\,\delta_y^{(\nu-j-1)}(y-\beta)$$

whence for $\partial^{1+\nu}\{f\}/(\partial\alpha\,\partial\beta^\nu)$ we have the following expression:

$$\begin{aligned}\frac{\partial^{1+\nu}\{f\}}{\partial\alpha\,\partial\beta^\nu}&=\left\{\frac{\partial^{1+\nu}f}{\partial\alpha\,\partial\beta^\nu}\right\}-\delta(x-\alpha)\otimes\frac{\partial^\nu\{f(x,y,x,\beta)\}}{\partial\beta^\nu}\\&\quad+\delta(x-\alpha)\otimes\sum_{l-0}^{\nu-1}(-1)^{\nu-1}\,\delta_y^{(\nu-l-1)}(y-\beta)\frac{\partial^l f(x,y,x,y)}{\partial\beta^l}\\&\quad+\sum_{k=0}^{\nu-1}(-1)^{\nu-k}\frac{\partial}{\partial\alpha}\left\{\frac{\partial^k f(x,y,\alpha,y)}{\partial\beta^k}\right\}\otimes\delta_y^{(\nu-k-1)}(y-\beta).\end{aligned}\tag{3.14}$$

Finally, from (3.14) we obtain by recurrence the following fundamental formula:

$$\begin{aligned}\frac{\partial^{\mu+\nu}\{f\}}{\partial\alpha^\mu\,\partial\beta^\nu}&=\left\{\frac{\partial^{\mu+\nu}f}{\partial\alpha^\mu\,\partial\beta^\nu}\right\}+\frac{\partial^\mu}{\partial\alpha^\mu}\sum_{l=0}^{\nu-1}(-1)^{\nu-l}\left\{\frac{\partial^l f(x,y,\alpha,y)}{\partial\beta^l}\right\}\otimes\delta_y^{(\nu-l-1)}(y-\beta)\\&\quad+\frac{\partial^\nu}{\partial\beta^\nu}\sum_{k=0}^{\mu-1}(-1)^{\mu-k}\,\delta_x^{(\mu-k-1)}(x-\alpha)\otimes\left\{\frac{\partial^k f(x,y,x,\beta)}{\partial\alpha^k}\right\}\\&\quad-\sum_{k=0}^{\mu-1}\sum_{l=0}^{\nu-1}(-1)^{\mu-k}(-1)^{\nu-l}\,\delta_x^{(\mu-k-1)}(x-\alpha)\otimes\delta_y^{(\nu-l-1)}\,(y-\beta)\frac{\partial^{k+l}f(x,y,x,y)}{\partial\alpha^k\,\partial\beta^l}\end{aligned}\tag{3.15}$$

which gives us the relation between the kernel distributions $(\partial^{\mu+\nu}\{f\})/(\partial\alpha^\mu\,\partial\beta^\nu)$ and $\{\partial^{\mu+\nu}f/(\partial\alpha^\mu\,\partial\beta^\nu)\}$.

Remark 3.2. Suppose f independent of the variable $x \in X$ (resp. $y \in Y$). Then we set

$$f = 1_x \otimes g(y, \alpha, \beta), \qquad (\text{resp. } f = 1_y \otimes g(x, \alpha, \beta),$$

where 1_x (resp. 1_y) is the function of $x \in X$ (resp. $y \in Y$) equal to 1.

Using these identities we obtain by formula (3.3) (resp. (3.4))

$$\frac{\partial^\mu \{f\}}{\partial x^\mu} = \delta_x^{(\mu-1)}(x-\alpha) \otimes \{1_\alpha \otimes g(y, \alpha, \beta)\}$$

$$\left(\text{resp. } \frac{\partial^\nu \{f\}}{\partial y^\nu} = \{1_\beta \otimes g(x, \alpha, \beta)\} \otimes \delta_y^{(\nu-1)}(y-\beta)\right).$$

Similar formulas can be obtained for functions independent of α, β or independent of x, y, from (3.9) (resp. (3.15)).

Let $f(x, y, \alpha, \beta) = a(x, \alpha) \otimes b(y, \beta)$ and $g(x, y, \alpha, \beta) = c(x, \alpha) \otimes d(y, \beta)$ be elements of

$$(L^1_{\text{loc}})_x (L^\infty_{\text{loc}})_\alpha \otimes (L^1_{\text{loc}})_y (L^\infty_{\text{loc}})_\beta.$$

Then we have

$$\begin{aligned} \{f\} \circ \{g\} &= \{f \circ g\} \\ &= \{a \otimes b\} \circ \{c \otimes d\} \\ &= \left\{ \int_\alpha^x a(x, \xi) c(\xi, \alpha)\, d\xi \otimes \int_\beta^y b(y, \eta) d(\eta, \beta)\, d\eta \right\} \end{aligned} \tag{3.16}$$

which implies $f \circ g \in (L^1_{\text{loc}})_x (L^\infty_{\text{loc}})_\alpha \otimes (L^1_{\text{loc}})_y (L^\infty_{\text{loc}})_\beta$. Suppose now $f(x, y, \alpha, \beta) = Y(x-\alpha) \otimes Y(y-\beta) = Y(x, y, \alpha, \beta)$.

Then, it follows from (3.16) that the 2nd composition power of $Y(x, y, \alpha, \beta)$ is given by [14, (3.6)]:

$$\begin{aligned} \{Y(x, y, \alpha, \beta)\}^2 &= \{Y(x-\alpha)\}^2 \otimes \{Y(y-\beta)\}^2 \\ &= \left\{\frac{(x-\alpha)}{1!}\right\} \otimes \left\{\frac{(y-\beta)}{1!}\right\}. \end{aligned}$$

Whence, by recurrence, the following expression for the nth composition power of $Y(x, y, \alpha, \beta)$:

$$\{Y(x, y, \alpha, \beta)\}^n = \left\{ \frac{(x-\alpha)^{n-1}}{(n-1)!} \otimes \frac{(y-\beta)^{n-1}}{(n-1)!} \right\} \tag{3.17}$$

3.2. Unit element for composition of kernel-functions. We have:

$$\begin{aligned} \{f\} \circ \delta(x, y, \alpha, \beta) &= Y(x-\alpha) \otimes Y(y-\beta) f(x, y, \alpha, \beta) \circ \partial(x-\alpha) \otimes \partial(y-\beta) \\ &= \left\{ \int_\alpha^x \int_\beta^y f(x, y, \xi, \eta)\, \delta(\xi-\alpha)\, \delta(\eta-\beta)\, d\xi\, d\eta \right\} = \{f\}. \end{aligned}$$

Likewise

$$\delta(x, y, \alpha, \beta) \circ \{f\} = \{f\}.$$

Therefore, $\delta(x-\alpha) \otimes \delta(y-\beta)$ is the unit element for the composition of kernel-functions. In particular one has:

$$\begin{aligned} \delta(x, y, \alpha, \beta) \circ Y(x, y, \alpha, \beta) &= Y(x, y, \alpha, \beta) \circ \delta(x, y, \alpha, \beta) \\ &= Y(x, y, \alpha, \beta). \end{aligned}$$

PROPOSITION 3.1. *We have*

$$\delta''_{xy}(x, y, \alpha, \beta) \circ Y(x, y, \alpha, \beta) = \delta(x, y, \alpha, \beta)$$

(3.18) *and*

$$Y(x, y, \alpha, \beta) \circ \delta''_{xy}(x, y, \alpha, \beta) = \delta(x, y, \alpha, \beta).$$

Indeed, by virtue of Proposition 3.1 [14, § 3] and the formula (3.16) above, we have:

$$\begin{aligned}\delta''_{xy}(x, y, \alpha, \beta) &\circ Y(x, y, \alpha, \beta)\\ &= (\delta'_x(x-\alpha) \otimes \delta'_y(y-\beta)) \circ (Y(x-\alpha) \otimes Y(t-\beta))\\ &= (\delta'_x(x-\alpha) \circ Y(x-\alpha)) \otimes (\delta'_y(y-\beta) \circ Y(y-\beta))\\ &= \delta(x-\alpha) \otimes \delta(y-\beta).\end{aligned}$$

The second statement can easily be proved in a similar way. Therefore, we can write by definition:

$$(\delta_{xy})^{-1} = Y(x, y, \alpha, \beta) \tag{3.19}$$

and

$$\delta''_{xy}(x, y, \alpha, \beta) = \{Y(x, y, \alpha, \beta)\}^{-1} = \delta'_x(x-\alpha) \otimes \delta'_y(y-\beta) \tag{3.20}$$

whence

$$Y(x, y, \alpha, \beta) \circ \{Y(x, y, \alpha, \beta)\}^{-1} = \delta(x-\alpha) \otimes \delta(y-\beta). \tag{3.21}$$

Then we find by recurrence or by (3.17):

$$\begin{aligned}\delta^{(\mu+\nu)}_{xy}(x, y, \alpha, \beta) &= \delta^{(\mu)}_x(x-\alpha) \otimes \delta^{(\nu)}_y(y-\beta)\\ &= \{Y(x-\alpha)\}^{-\mu} \otimes \{Y(y-\beta)\}^{-\nu}\end{aligned} \tag{3.22}$$

and by (3.19):

$$\begin{aligned}\delta^{(-\mu-\nu)}_{xy}(x, y, \alpha, \beta) &\overset{\text{def}}{=} \{\delta'(x-\alpha)\}^{-\mu} \otimes \{\delta'(y-\beta)\}^{-\nu}\\ &= \{Y(x-\alpha)^{\mu} \otimes Y(y-\beta)\}^{\nu}\\ &= \left\{\frac{(x-\alpha)^{\mu-1}}{(\mu-1)!}\right\} \otimes \left\{\frac{(y-\beta)^{\nu-1}}{(\nu-1)!}\right\}.\end{aligned} \tag{3.23}$$

3.3. Derivation of the composition product of kernel-functions.

PROPOSITION 3.2. *We have*:

$$\begin{aligned}\delta'_x(x, y, \alpha, \beta) \circ \{f\} &= \frac{\partial\{f\}}{\partial x},\\ \delta'_y(x, y, \alpha, \beta) \circ \{f\} &= \frac{\partial\{f\}}{\partial y}.\end{aligned} \tag{3.24}$$

Indeed we have

$$\begin{aligned}\delta'_x(x, y, \alpha, \beta) \circ \{f\} &= (\delta'_y(x-\alpha) \otimes \delta(y-\beta)) \circ \{f\}\\ &= \left\{\int_\alpha^\infty \delta'_x(x-\xi) f(\xi, y, \alpha, \beta)\, d\xi\right\}\\ &= \delta(x-\alpha) \otimes \{f(x, y, \alpha, \beta)\} + \left\{\frac{\partial f}{\partial x}\right\} = \frac{\partial\{f\}}{\partial x}\end{aligned}$$

by virtue of (3.1) and the formula (3.15) of [14, § 3]. The proof for the second formula (3.24) is similar.

PROPOSITION 3.3. *We have*:

$$\{f\}\circ\delta'_x(x,y,\alpha,\beta)=-\frac{\partial\{f\}}{\partial\alpha}$$
$$\{f\}\circ\delta'_y(x,y,\alpha,\beta)=-\frac{\partial\{f\}}{\partial\beta}. \tag{3.25}$$

Indeed,

$$\{f\}\circ\delta'_x(x,y,\alpha,\beta)=\left\{\int_{-\infty}^{x}f(x,y,\xi,\beta)\,\delta'_\xi(\xi-\alpha)\,d\xi\right\}$$
$$=\delta(x-\alpha)\otimes\{f(x,y,\chi,\beta)\}-\left\{\frac{\partial f}{\partial\alpha}\right\}=\frac{\partial\{f\}}{\partial\alpha},$$

according to the formula (3.10).

The proof for the second formula (3.25) is similar. Likewise, (3.24) and (3.25) give us:

$$\delta_x^{(k)}(x,y,\alpha,\beta)\circ\{f\}=\frac{\partial^k\{f\}}{\partial x^k},\qquad \delta_y^{(l)}(x,y,\alpha,\beta)\circ\{f\}=\frac{\partial^l\{f\}}{\partial y^l} \tag{3.26}$$

$$\{f\}\circ\delta_x^{(k)}(x,y,\alpha,\beta)=(-1)^k\frac{\partial^k\{f\}}{\partial\alpha^k}$$
$$\{f\}\circ\delta_y^{(l)}(x,y,\alpha,\beta)=(-1)^l\frac{\partial^l\{f\}}{\partial\beta^l} \tag{3.27}$$

whence

$$\delta_x^{(k)}\circ\delta_y^{(l)}\circ\{f\}\circ\delta_x^{(p)}\circ\delta_y^{(q)}=(-1)^{p+q}\frac{\partial^{k+l+p+q}\{f\}}{\partial x^k\,\partial y^l\,\partial\alpha^p\,\partial\beta^q} \tag{3.28}$$

Moreover, the associativity of the composition product gives us:

$$\delta_x^{(k)}\circ\delta_y^{(l)}\circ\{f\}\circ\delta_x^{(p)}\circ\delta_y^{(q)}\circ\{g\}\circ\delta_x^{(s)}\circ\delta_y^{(t)}$$
$$=(-1)^{p+q+s+t}\frac{\partial^{k+l+p+q}\{f\}}{\partial x^k\,\partial y^l\,\partial\alpha^p\,\partial\beta^q}\circ\frac{\partial^{s+t}\{g\}}{\partial\alpha^s\,\partial\beta^t}$$
$$=\frac{\partial^{k+l}\{f\}}{\partial x^k\,\partial y^l}\circ(-1)^{s+t}\frac{\partial^{p+q+s+t}\{g\}}{\partial x^p\,\partial y^q\,\partial\alpha^s\,\partial\beta^t} \tag{3.29}$$

for all k, l, p, q, s, t, nonnegative integers.

3.4. Operators of multiplication in $\mathscr{D}'_{(+\Gamma x)(-\Gamma y)}$. Let $\mathscr{E}_{xy}$ be the locally convex space of indefinitely differentiable functions in $X^n\times Y^n$. Let T be an arbitrary element of $\mathscr{D}'_{(+\Gamma x)(-\Gamma y)}$; then $\forall u\in\mathscr{E}_{xy}$, the product $u\cdot T$ has meaning. Indeed, $\forall\varphi\in\mathscr{D}_{(-\Gamma x)(+\Gamma y)}$, one has $u\cdot\varphi\in\mathscr{D}_{(-\Gamma x)(+\Gamma y)}$, where $\langle u\cdot T,\varphi\rangle=\langle T,u\varphi\rangle$. We say that $u\in\mathscr{E}_{xy}$ is an operator of multiplication in $\mathscr{D}'_{(+\Gamma x)(-\Gamma y)}$. Therefore, $\mathscr{E}_{xy}$, and a fortiori, $\mathscr{D}_{(-\Gamma x)(+\Gamma y)}$ are spaces of multiplication operators in $\mathscr{D}'_{(-\Gamma x)(-\Gamma y)}$.

For $u\in\mathscr{E}_{xy}$ (resp. $u\in\mathscr{D}_{(-\Gamma x)(+\Gamma y)}$), the support of $u\cdot T$ is contained in the intersection of the support of u and the support of T; i.e. $u\cdot T\in\mathscr{D}'_{(+\Gamma x)(-\Gamma y)}$ (resp. $u\cdot T\in\mathscr{E}'_{xy}$) [12, § 3, Prop. 1].

Moreover, the bilinear mapping $(u, T) \leftrightarrow u \cdot T$, from $\mathcal{D}_{(+\Gamma x)(-\Gamma y)} \times \mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$ in $\mathcal{E}'_{xy}$ (resp. from $\mathcal{E}_{xy} \times \mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$ in $\mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$) is hypo-continuous [1, loc. cit., Prop. 2].

Remark 3.4. The bilinear mapping $(u, \varphi \leftrightarrow u \cdot \varphi$ from $\mathcal{E}_{xy} \times \mathcal{D}_{(-\Gamma x)(+\Gamma y)}$ in $\mathcal{D}_{(-\Gamma x)(+\Gamma y)}$ is also hypocontinuous [12, loc. cit., Remark 1].

PROPOSITION 3.4. i) *Let $a(x)$ (resp. $b(y)$), for $x \in X^n$* (resp. $y \in Y^n$), *be a multiplication operator in $\mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$. Then one has:*

$$(a(x) \otimes b(y))(S \circ T) = (a(x)S) \circ (b(y)T) \tag{3.30}$$

For $S \in \mathcal{D}'_{(+\Gamma x)} \hat{\otimes} \mathcal{D}_{(-\Gamma y)}$ and $T \in \mathcal{D}'_{(+\Gamma x)} \hat{\otimes} \mathcal{D}'_{(-\Gamma y)}$.

ii) *Let $a(x)$ (resp. $a(y)$), for $x \in X^n$ (resp. $y \in Y^n$) be a multiplication operator in $\mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$. Then one has:*

$$S \circ (a(x)T) = (Sa(y)) \circ T \tag{3.31}$$

for $S \in \mathcal{D}'_{(+\Gamma x)} \hat{\otimes} \mathcal{D}_{(-\Gamma y)}$ and $T \in \mathcal{D}'_{(+\Gamma x)} \hat{\otimes} \mathcal{D}'_{(-\Gamma y)}$.

Proof. It follows from the fact that $\mathcal{D}'_{(+\Gamma x)} \hat{\otimes} \mathcal{D}'_{(-\Gamma y)}$ is a composition bimodule (cf. (1.3)).

Remark 3.4. In particular for $\{f\}$, $\{g\}$ kernel functions we obtain

$$(a(x) \otimes b(y))(\{f\} \circ \{g\}) = \{a(x)f\} \circ \{b(y)g\}. \tag{3.32}$$

Now, let $a(x, y)$ (resp. $b(x, y)$) be a function such that $a(x, y) \cdot f(x, y, \alpha, \beta)$ (resp. $b(x, y) \cdot g(x, y, \alpha, \beta)$) belongs to the same space as f (resp. g). It is clear that

$$a(x, y) \cdot \{f\} = \{a(x, y) \cdot f\} \quad \text{and} \quad b(x, y) \cdot \{g\} = \{b(x, y) \cdot g\}. \tag{3.33}$$

Then, we have

$$\{a(x, y) \cdot f\} \circ \{b(x, y) \cdot g\} = \{-a(x, y) \cdot f(x, y, \alpha, \beta) b(\alpha, \beta)\} \circ \{g(x, y, \alpha, \beta)\}. \tag{3.34}$$

PROPOSITION 3.5. *Let $a(x, y)$ be an operator of multiplication in $\mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$. Then we have:*

$$\begin{aligned} &\{Y(x-\alpha)\}^p \otimes \{Y(y-\beta)\}^q \circ \{a(x, y)\, \delta_{xy}^{p+q}(x, y, \alpha, \beta)\} \\ &\qquad = (-1)^{p+q} \frac{\partial^{p+q}}{\partial \alpha^p\, \partial \beta^q} \left\{ \frac{(x-\alpha)^{p-1}}{(p-1)!} \otimes \frac{(y-\beta)^{q-1}}{(q-1)!}\, a(\alpha, \beta) \right\}. \end{aligned} \tag{3.35}$$

The proposition is an obvious consequence of (3.29) and (3.31).

3.5. Derivatives of the composition product in $\mathcal{D}'_{(+\Gamma x)} \otimes \mathcal{D}'_{(-\Gamma y)}$.

PROPOSITION 3.6. *Let X^n, Y^n be topological vector spaces isomorphic with R^n, $n \leqq 1$.*

For $x = (x_1, x_2, \cdots, x_n) \in (+\Gamma x)$; $y = (y_1, y_2, \cdots, y_n) \in (-\Gamma y)$ we have:

$$\delta_{x_j}^{(k)}(x-y) \circ S(x, y) = \frac{\partial^k S(x,)}{\partial x_j^k};^{1} \tag{3.36}$$

$$S(x, y) \circ \delta_{x_i}^{(l)}(x-y) = (-1)^l \frac{\partial^l S(x, y)}{\partial y_i^l}^{1} \tag{3.37}$$

where $\delta_{x_j}^{(k)}(x-y)$ (resp. $\delta_{x_i}^{(l)}(x-y)$) is the partial derivative of order k (resp. l) with respect to x_j (resp. x_i) of Dirac kernel $\delta(x-y)$ in $X^n \times Y^n$ and S is an arbitrary element of $\mathcal{D}'_{(+\Gamma x)(-\Gamma y)}$.

Proof. $\delta(x-y)$ being the unit element for the composition product (cf. Remark

[1] The composition product in the left hand side of (3.36) and (3.37) has meaning, because $\mathcal{D}'_{(+\Gamma x)} \hat{\otimes} \mathcal{D}_{(-\Gamma y)}$ is a composition bimodule (cf. § 1, (1.3)).

1.1), we can write:

$$\delta_{x_j}^{(k)}(x-y)\circ S(x,y)=\delta_{x_j}^{(k)}(x_j-y_j)\circ S(x,y).$$

On the other hand, we have $\delta_{x_j}^{(k)}(x_j-y_j)\circ S(x,y)=\{\int_{y_j}^{x_j}\delta_{x_j}^{(k)}(x_j-\xi_j)\cdot S(x_1,\cdots,x_{j-1},\xi_j,x_{j+1},\cdots,x_n,y)\,d\xi_j\}=\delta^{(k)}(x_j)*^{x_j}S(x,y)$ where $*^{x_j}$ denotes the *convolution operator* with respect to the variable x_j, between $\delta^{(k)}(x_j)$ and $S(x,y)$. But, one has:

$$\delta^{(k)}(x_j)*^{x_j}S(x,y)=\frac{\partial^k S(x,y)}{\partial x_j^k},$$

which proves (3.36).

Likewise we obtain

$$\begin{aligned} S(x,y)\circ\delta_{x_i}^{(1)}(x-y)&=S(x,y)\circ\delta_{x_i}^{(l)}(x_i-y_i)\\ &=\left\{\int_{y_i}^{x_i}S(x,y_1,\cdots,y_{i-1},\eta_i,y_{i+1},\cdots,y_n)\delta_{\eta_i}^{(l)}(\eta_i-y_i)\,d\eta_i\right\}\\ &=(-1)^l S(x,y)*^{y_i}\delta_{y_i}^{(l)}(y_i)=(-1)^l\frac{\partial^l S(x,y)}{\partial y_i^l}, \end{aligned}$$

whence (3.37).

Remark 3.5. The conjunction of (3.36) and (3.37) gives us the more general formula:

$$\delta_{x_j}^{(k)}(x-y)\circ S(x,y)\circ\delta_{x_i}^{(l)}(x-y)=(-1)^l\frac{\partial^{k+l}S(x,y)}{\partial x_j^k\,\partial y_i^l}. \tag{3.38}$$

In particular, in $\mathscr{D}'_{(-\Gamma xy)}\hat{\otimes}\mathscr{D}'_{(+\Gamma_{\alpha\beta})}$ we obtain the formulas:

$$\delta_x^{(k)}(x-\alpha)\otimes\delta_y^{(l)}(y-\beta)\circ S=\frac{\partial^{k+l}S(x,y,\alpha,\beta)}{\partial x^k\,\partial y^l}, \tag{3.39}$$

$$S\circ(\delta_x^{(p)}(x-\alpha)\otimes\delta_y^{(q)}(y-\beta))=(-1)^{p+q}\frac{\partial^{p+q}S(x,y,\alpha,\beta)}{\partial\alpha^p\,\partial\beta^q} \tag{3.40}$$

and

$$\begin{aligned}&(\delta_x^{(k)}(x-\alpha)\otimes\delta_y^{(l)}(y-\beta))\circ S\circ(\delta_x^{(p)}(x-\alpha)\otimes\delta_y^{(q)}(y-\beta))\\ &\qquad=(-1)^{p+q}\frac{\partial^{k+l+p+q}S(x,y,\alpha,\beta)}{\partial x^k\,\partial y^l\,\partial\alpha^p\,\partial\beta^q}.\end{aligned} \tag{3.41}$$

Now, the extension of the formula (3.35) is given by the

PROPOSITION 3.7. *Let $a(x,y)$ be an operator of multiplication in $\mathscr{D}'_{(-\Gamma xy)(+\Gamma_{\alpha\beta})}$. Then we have*:

$$\begin{aligned}&(\{Y(x-\alpha)\}^p\otimes\{Y(y-\beta)\}^q)\circ\left(a(x,y)\frac{\partial^{k+l}S(x,y,\alpha,\beta)}{\partial x^k\,\partial y^l}\right)\\ &\qquad=(-1)^{k+l}\frac{\partial^{k+1}}{\partial\alpha^k\,\partial\beta^l}\left\{\frac{(x-\alpha)^{p-1}}{(p-1)!}\otimes\frac{(y-\beta)^{q-1}}{(q-1)!}a(\alpha,\beta)\right\}\circ S.\end{aligned} \tag{3.42}$$

Proof. If we denote by A the left-hand side of (3.42) we obtain, by virtue of the formulas (3.39) and (3.40):

$$A=\left\{\frac{(x-\alpha)^{p-1}}{(p-1)!}\otimes\frac{(y-\beta)^{q-1}}{(q-1)!}\right\}\circ(a(x,y)(\delta_x^{(k)}(x-\alpha)\otimes\delta_y^{(l)}(y-\beta)))\circ S$$

$$= \left\{ \frac{(x-\alpha)^{p-1}}{(p-1)!} \otimes \frac{(y-\beta)^{q-1}}{(q-1)!} a(\alpha, \beta) \right\} \circ (\delta_x^{(k)}(x-\alpha) \otimes \delta_y^{(l)}(y-\beta)) \circ S$$

$$= \left[(-1)^{k+l} \frac{\partial^{k+l}}{\partial \alpha^k \, \partial \beta^l} \left\{ \frac{(x-\alpha)^{p-1}}{(p-1)!} \otimes \frac{(y-\beta)^{q-1}}{(q-1)!} a(\alpha, \beta) \right\} \right] \circ S.$$

4. Algebraic composition equation of a linear partial differential equation with variable coefficients.

4.1. Fundamental formulas of composition in $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

THEOREM 4.1. *Each linear partial differential operator with variable coefficients may be transformed in a composition product in* $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

Proof. Let Ω be a partial linear differential operator with variable coefficients, of order m, (resp. n) with respect to x (resp. y) acting on an element of $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$ defined by

$$\Omega(S) = \sum_{j=0}^{m} \sum_{k=0}^{n} a_{jk}(x, y) \frac{\partial^{j+k} S(x, y, \alpha, \beta)}{\partial x^j \, \partial y^k}. \tag{4.1}$$

We suppose that the coefficients $a_{jk}(x, y)$ are operators of multiplication in $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$, for any j and k belonging to the set $\{1, 2, 3, \cdots, n\}$. Then, according to (3.39) we can write:

$$\Omega(S) = \sum_{j=0}^{m} \sum_{k=0}^{n} a_{jk}(x, y)\{\delta_x^{(j)}(x-\alpha) \otimes \delta_y^{(k)}(y-\beta) \circ S\}$$

whence, by virtue of (3.36):

$$\Omega(S) = \left[\sum_{j=0}^{m} \sum_{k=0}^{n} a_{jk}(x, y)\{\delta_x^{(j)}(x-\alpha) \otimes \delta_y^{(k)}(y-\beta)\} \right] \circ S. \tag{4.2}$$

Clearly, $\Omega(S)$ is an element of $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

By composition to the left of both sides of (4.2) with $\{Y(x-\alpha)\}^m \otimes \{Y(y-\beta)\}^n$ we obtain:

$$\{Y(x-\alpha)\}^m \otimes \{Y(y-\beta)\}^n \circ \Omega(S)$$
$$= \left[\sum_{j=0}^{m} \sum_{k=0}^{n} \{Y(x-\alpha)^m \otimes Y(y-\beta)^n\} \circ (a_{jk}(x, y))(\delta_x^{(j)}(x-\alpha) \otimes \delta_y^{(k)}(y-\beta)) \right] \circ S(x, y, \alpha, \beta).$$

Now keeping in mind the formula (3.42) we obtain:

$$\begin{aligned} &\{Y(x-\alpha)^m \otimes Y(y-\beta)^n\} \circ \Omega(S) \\ &= \sum_{j=0}^{m} \sum_{k=0}^{n} (-1)^{j+k} \frac{\partial^{j+k}}{\partial \alpha^j \, \partial \beta^k} \left\{ \frac{(x-\alpha)^{m-1}}{(m-1)!} \otimes \frac{(y-\beta)^{n-1}}{(n-1)!} a_{jk}(\alpha, \beta) \right\} \circ S(x, y, \alpha, \beta). \end{aligned} \tag{4.3}$$

This fundamental formula (4.3) gives us the proof of our Theorem 4.1.

In particular, if we suppose $a_{mn}(x, y) \equiv 1$, we obtain:

$$\begin{aligned} &\{Y(x-\alpha)^m\} \otimes \{Y(y-\beta)^n\} \circ \Omega(S) \\ &= \left[\delta(x-\alpha) \otimes \delta(y-\beta) + \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} (-1)^{j-k} \cdot \frac{\partial^{j+k}}{\partial \alpha^j \, \partial \beta^k} \left\{ \frac{(x-\alpha)^{m-1}}{(m-1)!} \otimes \frac{(y-\beta)^{n-1}}{(n-1)!} a_{jk}(\alpha, \beta) \right\} \right] \\ &= \{\delta(x-\alpha) \otimes \delta(y-\beta) + H(x, y, \alpha, \beta)\} \circ S(x, y, \alpha, \beta), \end{aligned} \tag{4.4}$$

in which we have set:

$$(4.5)\quad \{H(x, y, \alpha, \beta)\} = +\sum_{j=0}^{m-1}\sum_{k=0}^{n-}(-1)^{j+k}\frac{\partial^{j+k}}{\partial\alpha^j\,\partial\beta^k}\left\{\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}a_{jk}(\alpha, \beta)\right\}.$$

Obviously, $\{H(x, y, \alpha, \beta)\}$ belongs to $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

4.2. Fundamental kernels. Consider in $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$ the equation

$$(4.6)\qquad \frac{\partial^{m+n}\{E\}}{\partial x^m\,\partial y^n}+\sum_{j=0}^{m-1}\sum_{k=0}^{n-1}a_{jk}(x, y)\frac{\partial^{j+k}\{E\}}{\partial x^j\,\partial y^k}=\delta(x-\alpha)\otimes\delta(y-\beta)$$

where $a_{jk}(x, y)$ are operators of multiplication in $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

Then, the fundamental formula (4.4) allows us to write (4.6) as follows:

$$(4.7)\quad \{\delta(x-\alpha)\otimes\delta(y-\beta)+H(x, y, \alpha, \beta)\}\circ\{E\}=\left\{\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}\right\},$$

in which $H(x, y, \alpha, \beta)$ is given by (4.5).

On the other hand, let $[\![(\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)})^{\mathbb{N}}]\!]$ be the ring of formal series with respect to the addition and the composition product, whose terms belong to $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

On the algebraic structures of algebraic formal operations see our previous papers [15] and [16].

More precisely, let $[\![S]\!]=\sum_{\nu\in\mathbb{N}}S_\nu(x, y, \alpha, \beta)$ and $[\![T]\!]=\sum_{\nu\in\mathbb{N}}T_\nu(x, y, \alpha, \beta)$ be arbitrary elements of $[\![(\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)})^{\mathbb{N}}]\!]$. Then, the addition and composition between $[\![S]\!]$ and $[\![T]\!]$ are given respectively by:

$$(4.8)\qquad [\![S]\!]+[\![T]\!]=[\![S+T]\!]=\sum_{\nu\in\mathbb{N}}(S_\nu(x, y, \alpha, \beta)+T_\nu(x, y, \alpha, \beta))$$

and

$$(4.9)\qquad [\![S]\!]\circ[\![T]\!]=[\![S\circ T]\!]=\sum_{\nu\in\mathbb{N}}\Big(\sum_{\nu=p+q}S_p\circ T_q\Big).$$

$[\![S+T]\!]$ and $[\![S\circ T]\!]$ belong to $[\![(\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)})^{\mathbb{N}}]\!]$.

Let us now return to the composition equation (4.7). By transfer of this equation to $[\![(\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)})^{\mathbb{N}}]\!]$, we obtain

$$(4.10)\qquad [\![\delta(x-\alpha)\otimes\delta(y-\beta)+H]\!]\circ[\![E]\!]=\left[\!\!\left[\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}\right]\!\!\right]$$

in which we have set $[\![\delta(x-\alpha)\otimes\delta(y-\beta)\circ H]\!]=\sum_{\nu\in\mathbb{N}}S_\nu$, with $S_0=\delta(x-\alpha)\otimes\delta(y-\beta)$; $S_1=H$, $S_\nu=0$ for $\nu\geqq 2$, and

$$(4.11)\qquad \left[\!\!\left[\left\{\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}\right\}\right]\!\!\right]=\sum_{\nu\in\mathbb{N}}T_\nu,$$

where

$$T_0=\left\{\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}\right\},$$

and $T_y=0$ for $\nu\geqq 1$. On the other hand, we have

$$\begin{aligned}
&\Big(\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H^\nu\}\Big)\circ[\![\delta(x-\alpha)\otimes\delta(y-\beta)\circ\{H\}]\!]\\
(4.12)\qquad &=[\![\delta(x-\alpha)\otimes\delta(y-\beta)]\!]\circ\Big(\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H^\nu\}\Big)\\
&=[\![\delta(x-\alpha)\otimes\delta(y-\beta)]\!]
\end{aligned}$$

where $[\![\delta(x-\alpha)\otimes\delta(y-\beta)]\!]$ is the unit element of the composition algebra $[\![(\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)})^{\mathbb{N}}]\!]$.

Therefore $\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H\}^\nu$ is the inverse element of $[\![\delta(x-\alpha)\otimes\delta(y-\beta)+\{H\}]\!]$ in this composition algebra. Then, by composition to the left of both sides of (4.10) with $\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H^\nu\}$, we obtain for the solution of (4.8) in $[\![(\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)})^{\mathbb{N}}]\!]$, the expression

$$[\![E]\!]=\Big(\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H^\nu\}\Big)\circ\left[\!\!\left[\left\{\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}\right\}\right]\!\!\right]. \tag{4.13}$$

Now, if the resolvent kernel $[\![\Gamma]\!]=\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H^\nu\}$ converges in $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$, then $[\![E]\!]$ belongs to this composition algebra and $[\![E]\!]$ may be considered as the solution of (4.6).

Now, we will show that the series $\sum_{\nu\in\mathbb{N}}(-1)^\nu\{H(x,y,\alpha,\beta)\}^\nu$ converges in the topology of $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$.

The proof is similar to that of the corresponding "resolvent kernel" for ordinary differential equations (cf. [14, § 4, No. 4.1]).

To do this, let $\mathscr{C}(\mathbb{R}^2_{xy}\times\mathbb{R}^2_{\alpha\beta})$ be the vector space of continuous functions in $(x,y)\in\mathbb{R}^2_{xy}$ and $(\alpha,\beta)\in\mathbb{R}^2_{\alpha\beta}$.

In [14, § 2, Prop. 2.2] we have proved that $\mathscr{C}(\mathbb{R}^2_{xy}\times\mathbb{R}^2_{\alpha\beta})$ is a topological composition algebra, for the topology defined by the sequence of norms:

$$\|f\|_p=\sup_{\substack{-\infty<a_p\leqq\alpha\leqq x\leqq b_p<\infty\\-\infty<c_p\leqq\beta\leqq y\leqq d_p<\infty}}|f(x,y,\alpha,\beta)|$$

where $[a_p,b_p]\times[c_p,d_p]=K_p$ are compact subsets of $\mathbb{R}^2_{xy}\times\mathbb{R}^2_{\alpha\beta}$ such that $K_p\subset\mathring{K}_{p+1}$.

On the other hand it is easy to show that we have:

$$\begin{aligned}&\{H(x,y,\alpha,\beta)\}\\&=\left\{\sum_{j=0}^{m-1}\sum_{k=0}^{n-1}(-1)^{j+k}\frac{\partial^{j+k}}{\partial\alpha^j\,\partial\beta^k}\left(\frac{(x-\alpha)^{m-1}}{(m-1)!}\frac{(y-\beta)^{n-1}}{(n-1)!}a_{jk}(\alpha,\beta)\right)\right\}\end{aligned}$$

i.e., $H(x,y,\alpha,\beta)$ is the kernel-function of the function:

$$\begin{aligned}&H(x,y,\alpha,\beta)\\&=\sum_{j=0}^{m-1}\sum_{k=0}^{n-1}(-1)^{j+k}\frac{\partial^{j+k}}{\partial\alpha^j\,\partial\beta^k}\left(\frac{(x-\alpha)^{m-1}}{(m-1)!}\frac{(y-\beta)^{n-1}}{(n-1)!}a_{jk}(\alpha,\beta)\right).\end{aligned} \tag{4.14}$$

Therefore, without loss of generality we can consider $H(x,y,\alpha,\beta)$ as an element of $\mathscr{C}(\mathbb{R}^2_{xy}\times\mathbb{R}^2_{\alpha\beta})$.

Then, we have

$$\begin{aligned}\{H(x,y,\alpha,\beta)\}^0&=\delta(x-\alpha)\otimes\delta(y-\beta),\\\{H(x,y,\alpha,\beta)\}^1&=\{H(x,y,\alpha,\beta)\}=\{H^{(1)}(x,y,\alpha,\beta)\},\\\{H(x,y,\alpha,\beta)\}^2&=\left\{\int_\alpha^x\int_\beta^y H(x,y,\xi,\eta)H(\xi,\eta,\alpha,\beta)\,d\xi\,d\eta\right\}\\&=\{H^{(2)}(x,y,\alpha,\beta)\},\\\cdots\quad\cdots&\quad\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots,\\\{H(x,y,\alpha,\beta)\}^\nu&=\left\{\int_\alpha^x\int_\beta^y H^{(\nu-1)}(x,y,\xi,\eta)H(\xi,\eta,\alpha,\beta)\,d\xi,\,d\eta\right\}\\&=\{H^{(\nu)}(x,y,\alpha,\beta)\},\end{aligned}$$

in which $H^{(\nu)}(x, y, \alpha, \beta)$ satisfies the inequalities

$$\|H^{(\nu)}\|_p \leqq \|H\|_p^\nu \frac{(b_p - a_p)^{\nu-1}}{(m-1)!} \frac{(d_p - c_p)^{\nu-1}}{(\nu-1)!} \tag{4.15}$$

$\forall p \in \{1, 2, 3, \cdots\}$, $-\infty < a_p \leqq \alpha \leqq x \leqq b_p < \infty$ and $-\infty < c_p \leqq \beta \leqq y \leqq d_p < \infty$.

Therefore $\sum_{\nu \geqq 1} (-1)^\nu H^{(\nu)}(x, y, \alpha, \beta)$ converges in $\mathscr{C}(\mathbb{R}^2_{xy} \times \mathbb{R}^2_{\alpha\beta})$, and then, the resolvent kernel $\{\Gamma(x, y, \alpha, \beta)\}$ is given by:

$$\{\Gamma(x, y, \alpha, \beta)\} = \delta(x-\alpha) \otimes \delta(y-\beta) + \left\{ \sum_{\nu \geqq 1} (-1)^\nu H^{(\nu)}(x, y, \alpha, \beta) \right\} \tag{4.16}$$

where $\{\sum_{\nu \geqq 1} (-1)^\nu H^{(\nu)}(x, y, \alpha, \beta)\}$ is the distribution of $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$ which corresponds to the function $\sum_{\nu \geqq 1} (-1)^\nu H^{(\nu)}(x, y, \alpha, \beta)$ of $\mathscr{C}(\mathbb{R}^2_{xy} \times \mathbb{R}^2_{\alpha\beta})$.

This result shows that the solution $[\![E]\!]$, given by (4.13) coincides with

$$\{E\} = \left\{ \delta(x-\alpha) \otimes \delta(y-\beta) + \sum_{\nu \geqq 1} (-1)^\nu H^{(\nu)}(x, y, \alpha, \beta) \right\} \circ \{Y(x-\alpha)^m \otimes Y(y-\beta)^n\}, \tag{4.17}$$

which is the required solution of (4.6).

We will call $\{E\}$ the *fundamental solution* of (4.6). Now, if $S(x, y, \alpha, \beta)$ is a given element of $\mathscr{D}'_{(+\Gamma xy)(-\Gamma\alpha\beta)}$ the solution of the equation

$$\frac{\partial^{m+n} T}{\partial x^m \, \partial y^n} + \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} a_{jk}(x, y) \frac{\partial^{j+k} T}{\partial x^j \, \partial y^k} = S \tag{4.18}$$

is given by the composition product:

$$T = E \circ S.$$

4.3. Example. Consider the equation

$$\frac{\partial^2 E}{\partial x \, \partial y} + a(x, y) E = \delta(x-\alpha) \otimes \delta(y-\beta). \tag{4.19}$$

The formula (4.14) gives us $\{H^{(1)}(x, y, \alpha, \beta)\} = \{a(\alpha, \beta)\}$ whence

$$\begin{aligned}
&\{H^{(2)}(x, y, \alpha, \beta)\} \\
&= \{a(\alpha, \beta)\}^2 \\
&= \left\{ a(\alpha, \beta) \int_\alpha^x \int_\beta^y a(\xi, \eta)\, d\xi\, d\eta \right\} \cdots \{H^{(\nu)}(x, y, \alpha, \beta)\} \\
&= \left\{ a(\alpha, \beta) \int_\alpha^x \int_\beta^y a(\xi_1, \eta_1)\, d\xi_1\, d\eta_1 \int_{\xi_1}^x \int_{\eta_1}^y a(\xi_2, \eta_2) \cdot d\xi_\eta\, d\eta_\eta \cdots \int_{\xi_{\nu-2}}^x \int_{\eta_{\nu-2}}^y \right. \\
&\qquad\qquad \left. a(\xi_{\nu-1}, \eta_{\nu-1})\, d\xi_{\nu-1}\, d\eta_{\nu-1} \right\} \cdots .
\end{aligned}$$

Then, according to the inequalities (4.15) we have:

$$\|H^{(\nu)}\|_p \leqq \|a\|_p^\nu \frac{(b_p - a_p)^{\nu-1}}{(\nu-1)!} \frac{(d_p - c_p)^{\nu-1}}{(\nu-1)!},$$

where

$$\|a\|_p = \sup_{\substack{a_p \leqq x \leqq b_p \\ c_p \leqq y \leqq d_p}} |a(x, y)| \quad \forall p \in \{1, 2, 3, \cdots\}.$$

Consequently, the fundamental solution of (4.19) is given by

$$\{E\} = Y(x-\alpha)\otimes Y(y-\beta) + \sum_{\omega \geqq 1} (-1)^{\nu}\{H^{(\nu)}(x, y, \alpha, \beta)\} \circ Y(x-\alpha) \otimes Y(y-\beta). \tag{4.20}$$

But, [14, § 3, formulas (3.8) and (3.9)], we have:

$$\begin{aligned} &H^{(\nu)}(x, y, \alpha, \beta) \circ (Y(x-\alpha)\otimes (y-\beta)) \\ &= \int_\alpha^x \int_\beta^y H^{(\nu)}(x, y, \xi, \eta)\, d\xi\, d\eta \\ &= \textit{primitive} \text{ of } H^{(\nu)}(x, y, \alpha, \beta) \text{ with respect to } (\alpha, \beta) \end{aligned} \tag{4.21}$$

and

$$\begin{aligned} &(Y(x-\alpha)\otimes Y(y\times\beta)) \circ H^{(\nu)}(x, y, \alpha, \beta) \\ &= \int_\alpha^x \int_\beta^y H^{(\nu)}(\xi, \eta, \alpha, \beta)\, d\xi\, d\eta \\ &= \textit{primitive} \text{ of } H^{(\nu)}(x, y, \alpha, \beta) \text{ with respect to } (x, y). \end{aligned} \tag{4.22}$$

Then, we obtain for $\{E\}$ the expression:

$$\{E\} = Y(x-\alpha)\otimes Y(y-\beta) + \left\{ \sum_{\nu \geqq 1} (-1)^{\nu} \int_\alpha^x \int_\beta^y H^{(\nu)}(x, y, \xi, \eta)\, d\xi\, d\eta \right\}, \tag{4.23}$$

in which we have

$$\begin{aligned} &\int_\alpha^x \int_\beta^y H^{(\nu)}(x, y, \xi, \eta)\, d\xi\, d\eta \\ &= \int_\alpha^x \int_\beta^y a(\xi, \eta)\, d\xi\, d\eta \int_\xi^x \int_\eta^y a(\xi_1, \eta_1)\, d\xi_1\, d\eta_1 \cdots \int_{\xi_{\nu-1}}^x \int_{\eta_{\nu-1}}^y a(\xi_{\nu-1}, \eta_{\nu-1})\, d\xi_{\nu-1}\, d\eta_{\nu-1}. \end{aligned}$$

In particular, if $a(x, y)$ is a constant function, our method implies (cf. [8, Chap. V, Sec. IV, No 2, formula (10)]):

$$\{E\} = \begin{cases} \{J_0(2\sqrt{a(x-\alpha)(y-\beta)})\} & \text{for } a > 0, \\ \{I_0(2\sqrt{|a|(x-\alpha)(y-\beta)})\} & \text{for } a < 0. \end{cases}$$

Our next paper will be devoted to the construction of formal fundamental solutions of linear partial differential equations in which $a_{mn}(x, y) = 0$, in particluar, to the construction of fundamental solutions for some linear partial differential equations with constant coefficients.

Other papers will be devoted to:

a) the study of some problems of convergence of formal solutions;

b) boundary value problems;

c) translation operators in $\mathscr{D}'_{(+\Gamma_{xy})(-\Gamma_{\alpha\beta})}$ and their applications to solving linear equations with finite differences and variable coefficients;

d) algebraic method for solving integro-differential equations.

REFERENCES

[1] L. SCHWARTZ, *Théorie des Distributions*, Hermann, Paris 1966.

[2] J. LERAY, *Hyperbolic differential equations*, Lectures, Institute for advanced Study, Princeton, NJ, 1954.

[3] H. GARNIR, *Sur la transformation de Laplace des distributions*, C.R. Acad. Sci. Paris, 234 (1952), pp. 1583–1585.

[4] SEBASTIAO E SILVA, *Les fonctions analytiques comme ultradistributions dans le calcul operationel*, Math. Ann., 136 (1958), pp. 58–96.

[5] O. HEAVISIDE, *Operators in physical mathematics*, Proc. Roy. Soc. A, 52 (1893), pp. 504–529; 52 (1894), pp. 105–143.

[6] J. MIKUSINSKI, *Sur le fondement du calcul operationnel*, Studia Math., 11 (1950), pp. 41–70.

[7] S. VASILACH, *Sur un calcul operationnel algébric pour fonctions de deux variables*, Rev. Math. Pures Appl., 2 (1957), pp. 181–238.

[8] ———, *Calcul opérationnel algébrique des distributions à support dans* $\mathbb{R}^n_+$, $n \geqq 1$, Ibid., I, 4 (1959), pp. 185–219; II, 5 (1960), pp. 495–531; III, 6 (1961), pp. 69–100; IV–V, 8 (1963), pp. 19–66.

[9] ———, *Solutions of the parabolic partial differential equations by means of the algebraic operational calculus*, Publ. Inst. Math. (Beograd), 27 (1972), pp. 127–141.

[10] ———, *Product of composition in tensor product of bimodules*, Atti Acad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., 52 (1972), pp. 277–286.

[11] ———, *Produit de composition dans des produits tensoriels*, C.R. Acad. Sci. Paris Sér. A, 284 (1967), pp. 100–103.

[12] ———, *Sur une classe d'espaces de distributions*, Boll. Un. Mat. Ital., 5 (1970), pp. 745–760.

[13] ———, *Sur une classe d'espaces de distributions*, C.R. Acad. Sci. Paris Sér. A, 289 (1968), pp. 847–849.

[14] ———, *Algebraic method for solving differential equations whose coefficients are functions of one variable*, this Journal, 6 (1975), pp. 295–311.

[15] ———, *Algebraic structures of algebraic formal operations*, Rev. Math. Pures Appl., Chap. I–III (1963), pp. 350–390; Chap. IV (1963), pp. 543–566.

[16] ———, *Algebraic structures of algebraic formal operations, part II: Applications*, Ibid., 9 (1964), pp. 279–301.

SIAM J. MATH. ANAL
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0016 $01.00/0

STRICTLY AND STRONGLY STRICTLY CAUSAL LINEAR OPERATORS*

AVRAHAM FEINTUCH†

Abstract. The relationship between the strictly causal operators and the Jacobson radical of the algebra of causal operators is exploited to study causal operators. The strongly strictly causal compact operators are characterized in terms of their spectra.

1. Introduction. A formal approach to causality was first considered by R. Saeks in [12] where he introduced the idea of a resolution space. By using this concept and the theory of Brodskii, Gohberg and Krein [4] on Cauchy integrals of operators on chains, Saeks was able to give a unified theory of causality and strict causality and apply these ideas in an abstract setting to problems on feedback, time invariance, and various other aspects of dynamical systems.

Our purpose here is to give a slightly more general formalism of causality which contains that of Saeks. This is done by noting the important link between the above mentioned work of Brodskii, Gohberg and Krein with the nest algebra theory of Ringrose [11]. Although this link has been made in [2] its relevance for causality has not, to our knowledge, been previously pointed out. This will allow us to characterize the strictly causal operators in terms of the Jacobson radical of the Banach algebra of causal operators, which in turn will make the actual identification of strictly causal operators simpler. It will also allow us to give a complete characterization of compact strictly causal operators, thus solving a problem raised in [12, p. 52] by Saeks. We extend these results to the class of strongly strictly causal operators and show that for compact operators, these two classes are the same.

In his work Saeks has pointed out the applicability of the Nagy-Foias theory of contractions [6] to causality, in particular [13, Thm. 7.3] and to systems theory in general. This has recently been followed up by a number of authors. A particular class of contractions which appear quite naturally in a causality structure (as seen in [8]) is the class of C_0 contractions.

In the last part of this paper we use our formalism to study strict and strong strict causality for C_0 contractions. Since these concepts were introduced to obtain results on stability of systems it is not surprising that they are closely related to spectral properties of operators.

2. Preliminaries and notation. $\mathscr{H}$ will represent a Hilbert space. If $\mathscr{N}$ is a family of subspaces of $\mathscr{H}$, $\bigvee_{N\in\mathscr{N}} N$ will represent the closed linear span of the subspaces of $\mathscr{N}$ and $\bigcap_{N\in\mathscr{N}} N$ is the intersection of the subspaces in $\mathscr{N}$. All subspaces will be assumed closed.

$\mathscr{B}(\mathscr{H})$ will denote the algebra of all bounded linear operators on $\mathscr{H}$. If $T\in\mathscr{B}(\mathscr{H})$ and M is an invariant subspace of $T(TM\subset M)$ we will write $M\in\operatorname{Lat} T$.

If $\mathscr{N}$ is a family of subspaces of $\mathscr{H}$, Alg $\mathscr{N}$ will denote the strongly closed algebra

$$\{T\in\mathscr{B}(\mathscr{H})|TN\subset N\ \forall N\in\mathscr{N}\},$$

and $\operatorname{Alg}\mathscr{N}^\perp=(\operatorname{Alg}\mathscr{N})^*=\{T\colon T^*\in\operatorname{Alg}\mathscr{N}\}$.

3. Nests, nest spaces and nest algebras. Let $\mathscr{H}$ be a Hilbert space. In order to simplify technical matters we will assume that $\mathscr{H}$ is separable although this assumption is not necessary.

* Received by the editors September 24, 1976, and in final revised form December 30, 1977.
† Department of Mathematics, Ben Gurion University, Beer Sheva, Israel.

DEFINITION 3.1. A family $\mathcal{N}$ of subspaces of $\mathcal{H}$ is a *nest* if it is totally ordered by inclusion. $\mathcal{N}$ is *complete* if

(i) $\{0\}, \mathcal{H} \in \mathcal{N}$;

(ii) for $\mathcal{N}_0 \subset \mathcal{N}$, $\bigcap_{N \in \mathcal{N}_0} N$ and $\bigvee_{N \in \mathcal{N}_0} N$ are both in $\mathcal{N}$.

N_- will denote the subspace $\bigvee\{M: M \in \mathcal{N}, M \subset N\}$ with $\{0\}_- = \{0\}$. If $N_- \neq N$, N_- is called the immediate predecessor of N.

DEFINITION 3.2. A nest $\mathcal{N}$ is *maximal* if

(i) $\mathcal{N}$ is complete.

(ii) for all $N \in \mathcal{N}$, $N\theta N_-$ has dimension not greater than 1.

If for all $N \in \mathcal{N}$, $N\theta N_- = \{0\}$ we say that $\mathcal{N}$ is *continuous*. Otherwise, $\mathcal{N}$ has *gaps*.

It should be mentioned that Ringrose's original definition of maximality is different than that given above. However he showed that the condition given above is equivalent [10].

We can now define the concept which will allow us to consider causality in an abstract setting.

DEFINITION 3.3. A *nest space* is a pair $(\mathcal{H}, \mathcal{N})$ consisting of a Hilbert space $\mathcal{H}$ and a maximal nest $\mathcal{N}$ of subspaces of $\mathcal{H}$.

Alg $\mathcal{N}$, the algebra of operators, leaving invariant the subspaces of $\mathcal{N}$, will be called the *nest algebra* for $\mathcal{N}$.

4. Causality and anti-causality.

DEFINITION 4.1. Let $(\mathcal{H}, \mathcal{N})$ be a nest space. A bounded linear operator T is *anti-causal* if $T \in \text{Alg}\,\mathcal{N}$ and *causal* if $T \in \text{Alg}\,\mathcal{N}^\perp$ (or equivalently, $T^* \in \text{Alg}\,\mathcal{N}$).

At this point it is worthwhile to see the relationship between the resolution space of Saeks and our nest space.

Let E be a spectral measure defined on the Borel sets in an ordered topological group G whose values are projections on $\mathcal{H}$. This defines a resolution of the identity via

$$E^t = E(-\infty, t), \qquad E_t = I - E^t = E(t, \infty).$$

Let H^t and H_t denote, respectively, the ranges of E^t and E_t. Then T is causal [12, p. 18] if for any $x, y \in \mathcal{H}$ and $t \in G$ such that $E^t x = E^t y$ then $E^t Tx = E^t Ty$. As is pointed out in the Theorem on that page, this is equivalent to T leaving invariant $\{H_t: t \in G\}$ or $T^* \in \text{Alg}\,\mathcal{N}$ with $\mathcal{N} = \{H^t: t \in G\}$. Since $\mathcal{N}$ is easily seen to be a maximal nest, Saek's definition is included in ours.

5. Causal invertibility. If $\mathcal{H}$ is finite dimensional and T is an invertible causal operator on $(\mathcal{H}, \mathcal{N})$ then T^{-1} is causal. This is generally not true if $\mathcal{H}$ is infinite dimensional as is seen in the following example.

Example 5.1. Let $\mathcal{H} = l^2(-\infty, \infty; C)$, the Hilbert space of complex sequences $\{a_n\}_{-\infty}^{\infty}$ satisfying $\sum_{n=-\infty}^{\infty} |a_n|^2 < \infty$. Let e_n $(-\infty < n < \infty)$ be the vector in $\mathcal{H}$ whose nth coordinate is 1 and all other coordinates 0. Let $M_k = \bigvee_{n=-\infty}^{k} e_n$. Then $\mathcal{N} = \{\{0\}, \mathcal{H}, M_k: -\infty < k < \infty\}$ is a maximal nest and $(\mathcal{H}, \mathcal{N})$ is a nest space. Define an operator W on $\mathcal{H}$ by:

$$We_n = e_{n+1}.$$

Then W is a unitary operator and is the well-known bilateral shift. Also $W \in \text{Alg}\,\mathcal{N}$ and is causal. It is easily seen that $W^* = W^{-1}$ is not causal.

The question—is the inverse of a causal operator causal?— is of major importance in the study of feedback stability of linear systems. We give, at least formally, a complete solution of this problem. This contains as a special case a theorem of Saeks [12, p. 20]. Practically, however, it seems quite difficult to use this theorem in all of its generality.

DEFINITION 5.2. The operator A on $\mathscr{H}$ is *strictly positive* if there exists a real number $\delta \geqq 0$ such that

$$\operatorname{Re}(Ax, x) \geqq \delta \|x\|^2$$

for all $x \in \mathscr{H}$.

In the engineering literature such operators are often called *strictly passive* or *dissipative* because of the relationship between positivity and passive systems.

THEOREM 5.3. *Let T be causal on $(\mathscr{H}, \mathscr{N})$. T^{-1} exists and is causal if and only if there exists a causal operator A with A^{-1} causal such that AT (or TA) is strictly positive.*

Proof. Necessity. Suppose T^{-1} exists and is causal. Just take $A = T^{-1}$. Then $AT = I$ which is strictly positive.

Sufficiency. Let E be a projection on some subspace of $\mathscr{N}^{\perp}$. It suffices to show that ETE acting on $E\mathscr{H}$ is invertible, for then by the invariance of $E\mathscr{H}$ we have

$$E\mathscr{H} = ETE\mathscr{H} = TE\mathscr{H}.$$

Thus $T^{-1}E\mathscr{H} = E\mathscr{H}$ and $E\mathscr{H}$ is invariant under T^{-1}. Since E is arbitrary, $T^{-1} \in \operatorname{Alg} \mathscr{N}^{\perp}$.

Since EAE is invertible by hypothesis, this is equivalent to showing that $ETAE = ETEAE$ is invertible. This is where we use the strict positivity. There exists $\delta > 0$ such that for $\|x\| = 1$, $x \in EH$,

$$\begin{aligned} \delta \leqq \operatorname{Re}(TAx, x) &\leqq |(TAx, x)| \\ &= |(ETAEx, x)| \\ &\leqq \|ETAEx\| \cdot \|x\| = \|ETAEx\|. \end{aligned}$$

Noting that by the coinvariance of EH for A^* and T^* we have

$$EA^*T^*E = EA^*ET^*E,$$

the same computation shows that

$$\|E(TA)^*Ex\| \geqq \delta.$$

By [1, p. 84, Cor. 4.9], this implies that $ETAE$ is invertible and thus that ETE is invertible. This completes the proof.

An operator T is *definite* if $(Tx, x) \neq 0$ for all $x \neq 0$ in $\mathscr{H}$.

COROLLARY 5.4. *Suppose T is invertible and causal. Then T^{-1} is causal if and only if there exists A, A^{-1} causal such that TA (or AT) is definite.*

COROLLARY 5.5 (Saeks [12, p. 20]). *Suppose T is invertible, causal and for some $n \geqq 1$, T^n is definite. Then T^{-1} is causal.*

Proof. By Corollary 5.4, $(T^n)^{-1}$ is causal (just take $A = I$). Then so is $T^{-1} = T^{n-1}(T^n)^{-1}$.

6. Memoryless operators.

DEFINITION 6.1. Let T be an operator on $(\mathscr{H}, \mathscr{N})$. T is *memoryless* if T is both causal and anti-causal.

If $\mathscr{A} = \operatorname{Alg} \mathscr{N} \cap \operatorname{Alg} \mathscr{N}^{\perp}$, then T is memoryless if and only if $T \in \mathscr{A}$. Also, if $\mathscr{E}$ denotes the family of orthogonal projections onto the subspaces of $\mathscr{N}$, then $\mathscr{A}$ is the commutant of $\mathscr{E}$. This easily implies that if T is an invertible memoryless operator so is T^{-1}.

We give a characterization of the memoryless operators. This proceeds in two stages. First we reduce the problem to unitary operators and then characterize the unitary memoryless operators.

THEOREM 6.2. *Let T be a causal operator on $(\mathcal{H}, \mathcal{N})$ with polar decomposition $T = UP$. Then T is memoryless if and only if U is.*

Proof. If $T = UP$ and U is memoryless then so is U^*. Since T is causal, so is $U^*T = P$. But P is self-adjoint and thus is memoryless. T is then a product of memoryless operators.

To see that if T is memoryless so is U we observe the fact that the algebra of memoryless operators $\mathcal{A}$ is the commutant of the self-adjoint set $\mathcal{E}$. Thus $\mathcal{A}$ is a Von Neumann algebra and contains each factor in the polar decomposition of its members.

Let $\mathcal{K}$ be a Hilbert space and consider the Hilbert space $\mathcal{H} = l^2(-\infty, \infty; \mathcal{K})$. The bilateral shift W on $\mathcal{H}$ is defined as in Example 5.1. The multiplicity of W is defined to be the dimension of $\mathcal{K}$. The next theorem is well known.

THEOREM 6.3. *Let U be a unitary operator on a Hilbert space $\mathcal{H}$. Then $\mathcal{H}$ decomposes into an orthogonal sum $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ which are invariant under U. $U|\mathcal{H}_0$ is a bilateral shift of some multiplicity and $U|\mathcal{H}_1$ has the property that if M is invariant under $U|\mathcal{H}_1$ then so is $M^\perp$.*

Proof. See [14] and (or) [6].

As a consequence of Theorem 6.3 and the invariant subspace structure of the bilateral shift we have

THEOREM 6.4. *A causal unitary operator U on $(\mathcal{H}, \mathcal{N})$ is memoryless if it doesn't contain a bilateral shift direct summand of finite multiplicity.*

7. Integral Representations. Let $\mathcal{E}$ be the set of projections onto the members of $\mathcal{N}$. By a partition $\mathcal{P}$ of $\mathcal{E}$ is meant a finite subset $\mathcal{P} = \{E_i : 0 \leqq i \leqq n\}$ of $\mathcal{E}$ such that

$$0 = E_0 < E_1 < \cdots < E_n = I.$$

ΔE_i will denote the projection $E_i - E_{i-1}$. A partition $\mathcal{P}_1$ is a refinement of $\mathcal{P}$ if $\mathcal{P} \subseteq \mathcal{P}_1$. Note that the partitions of $\mathcal{E}$ form a directed set under refinement.

A detailed theory of integration on nests is presented in [4] and [10]. Here we consider a special case.

Let T be a bounded operator on $\mathcal{H}$ and $\mathcal{P}$ any partition of $\mathcal{E}$. Then we form the sums:

$$\mathcal{L}_{\mathcal{P}}(T) = \sum_{i=1}^{n} E_{i-1} T \,\Delta E_i, \qquad \mathcal{U}_{\mathcal{P}}(T) = \sum_{i=1}^{n} E_i T \,\Delta E_i,$$

$$\mathcal{T}_{\mathcal{P}}(T) = \sum_{i=1}^{n} F_i T \,\Delta E_i, \qquad E_{i-1} \leqq F_i \leqq E_i, \qquad \mathcal{D}_{\mathcal{P}}(T) = \sum_{i=1}^{n} \Delta E_i T \,\Delta E_i.$$

If the above sums converge in the uniform topology on $\mathcal{B}(\mathcal{H})$ we write

$$\mathcal{L}(T) = (m) \int ET\,dE, \qquad \mathcal{U}(T) = (M) \int ET\,dE,$$

$$\mathcal{T}(T) = \int ET\,dE, \qquad \mathcal{D}(T) = \int dET\,dE.$$

$\mathcal{T}(T)$, $\mathcal{U}(T)$ and $\mathcal{L}(T)$ are called, respectively, the integral, upper integral and lower integral of triangular truncation of T with respect to $\mathcal{N}$. $\mathcal{D}(T)$ is called the integral of diagonal truncation. Corresponding definitions exist in the strong operator topology. For this case we will place an '$s-$' in front of the symbol (e.g. $s-\mathcal{D}(T)$).

When $\mathcal{N}$ is replaced by $\mathcal{N}^\perp$ we obtain the corresponding integrals $\mathcal{T}^+$, $\mathcal{U}^+$, $\mathcal{L}^+$ and $\mathcal{D}^+$. Simple computations give the following proposition.

PROPOSITION 7.1. (a) $\mathscr{D}^+(T)=\mathscr{D}(T)$;
(b) $\mathscr{T}^+(T)=\int dETE$;
(c) $\mathscr{U}^+(T)=T-\mathscr{L}(T)$;
(d) $\mathscr{L}^+(T)=T-\mathscr{U}(T)$.

The next definition is motivated by the finite dimensional case.

DEFINITION 7.2. T is *lower triangular* if $T=\mathscr{U}^+(T)$, *upper triangular* if $T=\mathscr{U}(T)$, *diagonal* if $T=\mathscr{D}(T)$.

The idea of the next theorem is due to Saeks in the resolution space formalism. The proof in the next space formalism is essentially the same. We will therefore prove only one part.

THEOREM 7.3. *Let T be an operator on $(\mathscr{H},\mathscr{N})$. Then*:
(i) *T is causal if and only if $T=\mathscr{U}^+(T)$*;
(ii) *T is anti-causal if and only if $T=\mathscr{U}(T)$.*
(iii) *T is memoryless if and only if $T=\mathscr{D}(T)$.*

Proof of (i). If T is causal, then $T(I-E)=(I-E)T(I-E)$ for all $E\in\mathscr{E}$. Let $\mathscr{P}$ be a partition

$$0=E_0<E_1<\cdots<E_n=I.$$

Then $(I-E_i)T\,\Delta E_i=T\,\Delta E_i$ for all i. Thus $\mathscr{U}^+_{\mathscr{P}}(T)=\sum_{i=1}^n(I-E_i)T\,\Delta E_i=\sum_{i=1}^n T\,\Delta E_i=T\sum_{i=1}^n\Delta E_i=T$. Then $\mathscr{U}^+(T)=T$.

Now suppose $T=\mathscr{U}^+(T)$. Then

$$\begin{aligned}T(I-E)&=\mathscr{U}^+(T)(I-E)\\&=\lim_{\mathscr{P}}\mathscr{U}^+_{\mathscr{P}}(T)(I-E)\\&=\lim\sum_{i=1}^n(I-E_{i-1})T\,\Delta E_i(I-E).\end{aligned}$$

Let k be the integer such that

$$E_{k-1}\leqq E\leqq E_k.$$

Then $\Delta E_i(I-E)=0$ for $i\leqq k+1$. Thus $T(I-E)=\lim_{\mathscr{P}}\sum_{i=1}^{k+1}(I-E_{i-1})T\,\Delta E_i$. Also $(I-E_{i-1})=(I-E)(I-E_{i-1})$ for $i\leqq k+1$. Then

$$\begin{aligned}T(I-E)&=\lim_{\mathscr{P}}\sum_{i=1}^{k+1}(I-E)(I-E_{i-1})T\,\Delta E_i\\&=(I-E)\lim_{\mathscr{P}}\sum_{i=1}^{k+1}(I-E_{i-1})T\,\Delta E_i\\&=(I-E)\lim_{\mathscr{P}}\sum_{i=1}^{n}(I-E_{i-1})T\,\Delta E_i(I-E)\\&=(I-E)T(I-E).\end{aligned}$$

This completes the proof.

Remark. It is not hard to show that Theorem 7.3 could be stated for strongly convergent integrals as well. This is in contrast to the results discussed in the coming sections where we will differentiate between the two topologies.

8. Radicals of nest algebras. At this point we diverge from causality to study the relationship between the Jacobson radical of the Banach algebra Alg $\mathscr{N}$ and the integral representations studied in § 7. Most of the results brought here first appeared in [2].

DEFINITION 8.1. Let $\mathscr{B}$ be a Banach Algebra. The Jacobson radical $\mathscr{R}$ of $\mathscr{B}$ is the intersection of the kernels of all irreducible representations of $\mathscr{B}$.

The next lemma lists the properties of $\mathscr{R}$ that we will use. These are well known and proofs can be found in [9, Thms. 2.3.2–2.3.5].

LEMMA 8.2. *Let $\mathscr{R}$ be the radical of the Banach algebra $\mathscr{B}$. Then*:

(i) *$\mathscr{R}$ is a norm-closed two-sided ideal of $\mathscr{B}$*;

(ii) *$\mathscr{R}$ contains every quasi-nilpotent one-sided ideal of $\mathscr{B}$*;

(iii) *$\mathscr{R}=\{R\in\mathscr{B}|RA$ is quasi-nilpotent for all $A\in\mathscr{B}\}=\{R\in\mathscr{B}|AR$ is quasi-nilpotent or all $A\in\mathscr{B}\}$.*

If $\mathscr{B}$ is a nest algebra $\operatorname{Alg}\mathscr{N}$ we have the following characterization of $\mathscr{R}$ due to Ringrose [11, Thm. 5.4]:

THEOREM 8.3. *$T\in\operatorname{Alg}\mathscr{N}$ belongs to $\mathscr{R}$ if and only if for any $\varepsilon>0$, there exists a partition $\mathscr{P}$ of $\mathscr{E}$ such that*

$$\|\Delta E_i T\,\Delta E_i\|<\varepsilon$$

for $1\leqq i\leqq n$.

This leads to the following corollary which is of central importance for what follows.

COROLLARY 8.4. $\mathfrak{R}=\{T\in\operatorname{Alg}\mathscr{N}:\mathscr{D}(T)=0\}$.

Proof. See [2].

This simple corollary was seen by Erdos and Longstaff to provide the link between the Ringrose theory and that of Brodski, Gohberg and Krein. This led them to the next theorem which is the main result of [2].

THEOREM 8.5. *$\mathfrak{T}(T)$ exists if and only if $T=T_1+T_2$ with T_1, $T_2^*\in\mathscr{R}$ where $\mathscr{R}$ is the radical of $\operatorname{Alg}\mathscr{N}$. T_1 and T_2 are uniquely given by*

$$T_1=\mathscr{T}(T),\qquad T_2=\mathscr{T}^+(T).$$

9. Strict causality and anti-causality. We have seen that the causality of an operator does not in general ensure the causality of its inverse. Since this property is of major importance in the study of stability of input-output feedback systems, a number of authors were motivated to strengthen the causality condition. The most natural way to do this in the context of the above discussion was given by Saeks.

DEFINITION 9.1. T is *strictly anti-causal* if $T=\mathscr{T}(T)=\int ET\,dE$ and *strictly causal* if $T=\mathscr{L}^+(T)$.

Physically, strict causality is a delay type condition. As a consequence of the above discussion we get a natural characterization of strict anti-causality.

THEOREM 9.2. *The following are equivalent*:

(1) *T is strictly anti-causal*;

(2) $T\in\mathscr{R}$;

(3) *T is anti-causal and $\mathscr{T}^+(T)=0$*;

(4) *T is anti-causal and $\mathscr{D}(T)=0$.*

Proof. (1)$\Rightarrow$(2). Suppose T is strictly anti-causal. The $\mathscr{T}(T)$ exists and equals T, by Theorem 8.5 $T\in\mathscr{R}$.

(2)$\Rightarrow$(3). If $T\in\mathscr{R}$, $\mathscr{T}(T)$ converges. Since $0\in\mathscr{R}$ and the decomposition $T=\mathscr{T}(T)+0$ is unique it follows that $\mathscr{T}^+(T)=0$. Since $T\in\mathscr{R}$ and $\mathscr{R}\subset\operatorname{Alg}\mathscr{N}$, T is anti-causal.

(3)$\Rightarrow$(4). If $\mathscr{T}^+(T)=0$, then $\mathscr{U}^+(T)=\mathscr{L}^+(T)=0$. Since $\mathscr{D}_{\mathscr{P}}(t)=\mathscr{U}^+_{\mathscr{P}}(T)-\mathscr{L}^+_{\mathscr{P}}(T)$ it follows that $\mathscr{D}(T)=0$.

(4)$\Rightarrow$(1). Since T is anti-causal, $T=\mathscr{U}(T)$. But $\mathscr{D}(T)=\mathscr{U}(T)-\mathscr{L}(T)$ implies $\mathscr{U}(T)=\mathscr{L}(T)=T$ and thus $\mathscr{T}(T)$ exists and equals T.

We have a dual result for strict causality.

THEOREM 9.3. *The following are equivalent*:

(1) *T is strictly causal*;

(2) $T^* \in \mathcal{R}$;

(3) *T is causal and $\mathcal{T}(T) = 0$*;

(4) *T is causal and $\mathcal{D}(T) = 0$.*

Remark 9.4. As immediate consequences of the identification of the strictly causal operators with the radical of Alg $\mathcal{N}$ we obtain the following:

(1) The strictly causal operators form a uniformly closed two-sided ideal in $\mathcal{N}$ [12, p. 37];

(2) if T is strictly causal, then T is quasi-nilpotent;

(3) if T is strictly causal, then $(I-T)^{-1}$ can be obtained by a Neumann series in T. (This was proved directly in [15].)

In [12] Saeks raised the question under what conditions would the resolvent of a causal operator be analytic in the open right half-plane. The motivation for this question is historical. For a time-invariant causal operator it is classical that this is the case. While it is not hard to see that in general this is not the case it follows from (2) of the above remark that for T strictly causal even more is true. In fact, $(\lambda I - T)^{-1}$ is analytic everywhere in the complex plane with a puncture at the origin.

The next definition is motivated by Theorem 9.3.

DEFINITION 9.5. The causal operator T is *strongly strictly causal* if $T = s - \mathcal{L}^+(T)$ or equivalently if $s - \mathcal{D}(T) = 0$.

While a simple characterization of the strictly causal operators was possible in terms of the radical of Alg $\mathcal{N}$, the problem of characterizing strongly strictly causal operators is much more difficult. This problem is made more acute by the fact that strong strict causality is quite common. For example, if K is a convolution operator on $L^2(-\infty, \infty)$ with kernel $K \in L'$, then K is causal if and only if $k(t) = 0$ for $t < 0$ and strongly strictly causal if and only if $k(t) = 0$ for $t \leqq 0$ (see [12, p. 33]). On the other hand K will rarely be strictly causal.

However, there is a relationship between strict and strong strict causality.

THEOREM 9.6. *Every strongly strictly causal operator is a strong limit of strictly causal operators.*

Proof. Suppose T is strongly strictly causal. If $\mathcal{P}$ is any partition

$$0 = E_0 < E_1 < \cdots < E_n = I$$

of $\mathcal{E}$, by causality

$$T = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \Delta E_2 T\, \Delta E_1 & 0 & 0 & & 0 \\ & \Delta E_3 T\, \Delta E_2 & 0 & \ddots & 0 \\ & \vdots & & \ddots & \vdots \\ \Delta E_n T\, \Delta E_1 & \Delta E_n T\, \Delta E_2 & & & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} \Delta E_1 T\, \Delta E_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & \Delta E_n T\, \Delta E_n \end{pmatrix}$$

We use matrix notation to make the argument more transparent. The first matrix corresponds to $T-\mathcal{U}_{\mathcal{P}}(T)$ and the second to $\mathcal{D}_{\mathcal{P}}(T)$. Note that $T-\mathcal{U}_{\mathcal{P}}(T)$ is nilpotent and that for any A causal, $[T-\mathcal{U}_{\mathcal{P}}(T)]A$ is nilpotent. By Lemma 8.2 (iii) $[T-\mathcal{U}_{\mathcal{P}}(T)]$ is strictly causal. Since T is strongly strictly causal, $\mathcal{D}_{\mathcal{P}}(T)$ converges strongly to zero and $[T-\mathcal{U}_{\mathcal{P}}(T)]$ converges strongly to T. This completes the proof.

10. Compact operators. The problem of characterizing all compact strictly causal operators was mentioned in [12, p. 52]. Our characterization of strictly causal operators as the radical of Alg $\mathcal{N}$ will make the problem quite transparent. In fact, a solution will follow immediately from known results. What is more surprising is that this also leads to a characterization of strongly strictly causal compact operators.

We begin with the following result proved in [2].

THEOREM 10.1. *Let T be a compact causal operator. Then T is strictly causal if and only if T is quasi-nilpotent.*

COROLLARY 10.2. *Let T be a strictly causal operator. Then T is compact if and only if* Im $T=(1/(2i))(T-T^*)$ *is compact.*

Proof. This follows immediately from Theorem 10.1 and the fact that if T is quasi-nilpotent and Im $T=(1/(2i))(T-T^*)$ is compact then so is T [10, p. 60].

DEFINITION 10.3. For each $N\in\mathcal{N}$ such that $N\neq N_-$ other than $\{0\}$ define α_N to be the number α such that $PTP=\alpha P$ where P is the projection onto $N\theta N_-$. Then $\{\alpha_N | N\in\mathcal{N}$ and $N\neq N_-\}$ is the set of *diagonal coefficients* of T relative to $\mathcal{N}$.

THEOREM 10.4 [10, p. 176]. *If T is compact then $\sigma(T)$ is the union of the set of diagonal coefficients of K relative to $\mathcal{N}$ and the point* 0.

This theorem allows us to characterize the strongly strictly causal compact operator.

THEOREM 10.5. *If T is a compact causal operator then the following are equivalent*:

(1) *T is quasi-nilpotent*;

(2) *T is strictly causal*;

(3) *T is strongly strictly causal.*

Proof. It suffices to show that (3)$\Rightarrow$(1). Suppose T is strongly strictly causal and $\alpha\neq 0$ is in $\sigma(T)$. Choose $\varepsilon>0$ such that $\varepsilon^2<|\alpha|^2$.

Since $\alpha\in\sigma(T)$ there exists, by the previous theorem, some $N\in\mathcal{N}$ such that $\alpha=\alpha_N$. Let E and E_- denote the projections on N and N_- respectively, and let x be a unit vector in the range of $E-E_-$. By strong strict causality there exists a partition $\mathcal{P}$ such that for any refinement $\mathcal{P}_1$ of $\mathcal{P}$,

$$\|\mathcal{D}_{\mathcal{P}_1}(T)x\|<\varepsilon.$$

Let $\mathcal{P}_1=\mathcal{P}\cup\{E, E_-\}$. If $\mathcal{P}_1$ is given by $0<E_1<\cdots<E_n=I$ with $E=E_j$, then, since $\Delta E_i x=0$ for $i\neq j$,

$$\begin{aligned}\|\mathcal{D}_{\mathcal{P}_1}(T)x\|^2&=\left\|\sum_{i=1}^{n}\Delta E_i T\,\Delta E_i x\right\|^2\\&=\|\alpha\,\Delta E_j x\|^2=\|\alpha x\|^2=|\alpha|^2>\varepsilon,\end{aligned}$$

which contradicts the strong strict causality of T.

11. C_0 contractions. Here we study strict and strong causality properties of C_0 contractions. The motivation for studying such operators in a causality setting was given in [8].

We describe these operators very briefly. For a complete and elegant treatment of this subject the reader is referred to [6].

H^∞ will denote the algebra of bounded analytic functions on the open unit disc with the usual norm. If T is a completely nonunitary contraction and $u \in H^\infty$, the Nagy–Foias functional calculus described in [6] allows us to define the operator $u(T)$.

DEFINITION 11.1. T is a C_0 *contraction* if there exists a nonzero function $u \in H^\infty$ such that $u(T)=0$.

THEOREM 11.2. *Suppose T is a causal, invertible C_0* contraction on $(\mathcal{H}, \mathcal{N})$. *Then T^{-1} is causal.*

Proof. Since T^{-1} is in the double commutant of T, it follows from [5] that T^{-1} is a strong limit of polynomials in T and I. Thus $T \in \operatorname{Alg} \mathcal{N}$ implies $T^{-1} \in \operatorname{Alg} \mathcal{N}$.

We now turn to strict and strong causality. Here we will add the natural restriction (see [6]) that $I - T^*T$ (or $I - TT^*$) is compact. We have seen that a necessary condition for an operator T to be strictly causal is $\sigma(T)=\{0\}$. Since, except for the trivial case, the continuous spectrum of a C_0 contraction is contained in the unit circle [6, p. 126], we will study $I-T$ instead of T.

THEOREM 11.3. *Suppose T is causal C_0 contraction on $(\mathcal{H}, \mathcal{N})$ such that $I-T^*T$ (or $I-TT^*$) is compact. Then $I-T$ is strictly causal if and only if $\sigma(T)=\{1\}$.*

Proof. Since $I-T^*T$ is compact, so is $T(I-T^*T)=T-TT^*T$. Thus we can write $T=U+K$ where K is compact. Since the essential spectrum of $T=$ essential spectrum of $U=\{1\}$, it follows from [3, p. 23] that $U=I+K_1$, where K_1 is compact. This implies that $I-T$ is compact.

Since $I-T$ is a compact causal and $\sigma(I-T)=\{0\}$, it follows from Theorem 10.1 that $I-T$ is strictly causal.

The converse is trivial.

To obtain information about strong strict causality we use a result similar to that of Theorem 10.4.

THEOREM 11.4 [7]. *Suppose T is a C_0 contraction such that $I-T^*T$ is compact. If T is causal on $(\mathcal{H}, \mathcal{N})$ then the point spectrum of T is identical to the set of diagonal coefficients of T.*

This allows us to prove the following:

THEOREM 11.5. *Let T be a causal C_0 contraction on $(\mathcal{H}, \mathcal{N})$ such that $I-T^*T$ is compact. Then if $I-T$ is strongly strictly causal,*

$$\sigma(T) \subset \{z \mid |z|=1\}.$$

Proof. Suppose $\alpha \in \sigma(T)$ such that $|\alpha| \neq 1$. By [6, p. 126], α is in the point spectrum of T and thus by Theorem 11.4 there exists $N \in \mathcal{N}$ such that $\alpha = \alpha_N$. We proceed as in the proof of Theorem 10.5.

Let E and E_- be the projections on N and N_- respectively and let x be a unit vector in the range of $E-E_-$. Choose $\varepsilon > 0$ such that $\varepsilon^2 < |1-\alpha|^2$. By strong strict causality, there exists a partition $\mathcal{P}$ such that for any refinement $\mathcal{P}_1$ of $\mathcal{P}$,

$$\|\mathcal{D}_{\mathcal{P}_1}(I-T)x\| < \varepsilon.$$

Let $\mathcal{P}_1 = \mathcal{P} \cup \{E, E_-\}$. If $\mathcal{P}_1$ is given by

$$0 < E_1 < \cdots < E_n = I$$

with $E = E_j$, the same computation as in Theorem 10.5 leads to

$$\|\mathcal{D}_{\mathcal{P}_1}(I-T)x\|^2 > \varepsilon^2,$$

which contradicts the fact that $I-T$ is strongly strictly causal.

Remark 11.6. It is not known whether the converse is true; i.e. if $\sigma(T) \subset \{z \mid |z|=1\}$ then T is strictly causal.

12. Application to feedback systems. A feedback system is defined by the equations

$$y = Ke + d, \qquad e = Fy + u,$$

where K and F are causal operators on a fixed nest space $(\mathcal{H}, \mathcal{N})$ and e, d, u, y are vectors in $\mathcal{H}$. Conceptually, u is viewed as the system input signal, d as a disturbance in the output; y is the output and e is the input to K.

Rewriting these equations as

$$(I - KF)y = Ku + d, \qquad (I - FK)e = u + Fd$$

it is seen that to solve for y or e, $(I - KF)$ or $(I - FK)$ must be invertible. It is easy to see that $(I - KF)^{-1}$ exists if and only if $(I - FK)^{-1}$ does [12, p. 64].

DEFINITION 12.1. A feedback system is *well posed and stable* if $(I - KF)^{-1}$ exists as a bounded causal operator on $(\mathcal{H}, \mathcal{N})$.

We note that while this is not the usual definition of well posedness and stability, this is equivalent and is used in most applications (see [12, pp. 66–75]).

As applications we present greatly simplified proofs of results of Saeks given in [12, pp. 78–80].

THEOREM 12.2. *For the feedback system defined on $(\mathcal{H}, \mathcal{N})$ by the equations $y = Ke + d$, $e = Fy + u$, suppose that $\int dE\, KF\, dE$ exists and has norm less than one. Then the feedback system is well posed and stable.*

Proof. We show that $(I - KF)^{-1}$ exists and is causal. By the additive decomposition theorem [12, p. 41],

$$KF = T_1 + T_2,$$

where T_1 is strictly causal and $T_2 = \int dE\, KF\, dE$ with $\|T_2\| < 1$. Thus $(I - T_2)^{-1}$ exists and can be written as a power series in T_2.

Then $I - KF = (I - T_2) - T_1$ and formally

$$(I - KF)^{-1} = [(I - T_2) - T_1]^{-1} = (I - T_2)^{-1}[I - T_1(I - T_2)^{-1}]^{-1}.$$

Thus to complete the proof it is enough to show the second term exists and is causal.

Since T_1 is strictly causal, so is $T_1(I - T_2)^{-1}$ and its spectrum consists only of zero. Thus if $S = T_1(I - T_2)^{-1}$, there exists n such that $\|S^n\| < 1$. This implies that the Neumann series $\sum_{i=0}^{\infty} S^i$ converges to $(I - S)^{-1}$. Thus $(I - S)^{-1}$ is causal and the proof is complete.

COROLLARY 12.3. *If either K or F is strictly causal, then the feedback system is well posed and stable.*

Proof. KF is then strictly causal and if $KF = T_1 + T_2$ as in the above theorem $T_2 = 0$. We now apply the previous theorem.

REFERENCES

[1] R. G. DOUGLAS, *Banach Algebra Techniques in Operator Theory*, Academic Press, New York–London, 1972.

[2] J. A. ERDOS AND W. E. LONGSTAFF, *The convergence of triangular integrals of operators on Hilbert space*, Indiana Univ. Math. J., 22 (1973), pp. 929–938.

[3] I. C. GOHBERG AND M. G. KREIN, *Introduction to the Theory of Linear Nonself-Adjoint Operators*, Translations of Mathematical Monographs 18, American Mathematical Society, Providence, RI, 1969.

[4] ———, *Theory and Applications of Volterra Operators in Hilbert Space*, Translations of Mathematical Monographs 24, American Mathematical Society, Providence, RI, 1970.

[5] B. Sz-Nagy and C. Foias, *Commutants and bicommutants of operators of class C_0*, Acta Sci. Math., 38 (1976), pp. 311–315.

[6] ———, *Harmonic Analysis of Operators on Hilbert Space*, North-Holland, American Elsevier, Amsterdam-New York, 1970.

[7] A. Feintuch, *Diagonal coefficients for C_0 operators*, Proc. London. Math. Soc., to appear.

[8] ———, *Causality and C_0 contractions*, Proceedings, 2nd Int'l Symp. on Operator Theory and Networks and Systems, Texas Tech Univ., Lubbock, TX, 1977.

[9] C. E. Rickart, *General Theory of Banach Algebras*, Van Nostrand, Princeton, NJ, 1960.

[10] J. R. Ringrose, *Compact Non-selfadjoint Operators*, Van Nostrand, Reinhold, New York, 1971.

[11] ———, *On some algebras of operators*, Proc. London Math. Soc., 3 (1965), pp. 61–83.

[12] R. Saeks, *Resolution Space, Operators and Systems*, Lecture Notes in Economics and Math. Sys., Springer-Verlag, Berlin, 1973.

[13] ———, *Causality in Hilbert Space*, SIAM Rev., 12 (1970), pp. 357–383.

[14] J. Wermer, *On invariant subspaces of normal operators*, Proc. Amer. Math. Soc., 3 (1952), pp. 270–277.

[15] R. M. DeSantis and W. A. Porter, *On the Generalization of the Volterra Principle of Inversion*, J. Math. Anal. Appl., 48 (1974), pp. 743–748.

SIAM J. MATH ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0017 $01.00/0

A COMPUTER METHOD FOR VERIFICATION OF ASYMPTOTICALLY STABLE PERIODIC ORBITS*

JOHN E. FRANKE† AND JAMES F. SELGRADE†

Abstract. This paper provides criteria for locating a periodic solution to an autonomous system of ordinary differential equations and for showing the solution is orbitally asymptotically stable. The numerical analysis and the computer program needed to establish these criteria for a specific 2-dimensional system of equations are discussed.

1. Introduction. The intent of this work is to show how techniques of global analysis and numerical analysis enable a computer to prove certain qualitative properties of solution curves to autonomous ordinary differential equations. General results in global analysis are discussed in §§ 2 and 3. These results provide criteria for locating a connected, attracting invariant set and for proving this set is a periodic orbit with one zero characteristic exponent and the remaining characteristic exponents having negative real parts. Such a periodic orbit is orbitally asymptotically stable.

In §§ 4 and 5 we develop the numerical details and computer programs needed to implement our method on an actual system of differential equations. For this demonstration we choose the following system on R^2:

$$\dot{x} = y, \qquad \dot{y} = \alpha(1-x^2-y^2)y - x. \tag{1}$$

Here α is a parameter which we take to be .05. This is a modification of the Van der Pol equation, chosen to reduce programming cost and to simplify some of the analysis. The periodic solution we find is circular of radius one centered at the origin. The existence of this solution and the result concerning its characteristic exponents may be obtained by classical methods, as explained in § 4. However, this simple example illustrates how to handle the numerical difficulties encountered in applying our computer method to more complicated systems of equations.

2. Background. Let $X \in R^n$ and F be a continuously differentiable function from R^n to R^n. An autonomous system of ordinary differential equations is denoted

$$\dot{X} = F(X) \tag{2}$$

where "·" represents differentiation with respect to $t \in R$. F is called a *vector field.* The unique solution to (2) with initial condition X is a parametrized curve $\phi_t(X)$, with t in an open interval of R containing zero, such that $\phi_0(X) = X$ and

$$\frac{d}{dt}\phi_t(X)|_{t=s} = F(\phi_s(X)).$$

This curve $\phi_t(X)$ will be referred to as the *orbit* through X. For each fixed t, ϕ_t is a differentiable function from an open subset of R^n into R^n. The parametrized family of differentiable functions ϕ_t is called the flow of F. See [9], [16] for details. If $X \in R^n$ and $F(X) = 0$, then $\phi_t(X) = X$ for all $t \in R$ and we call X a *rest point* of ϕ_t. If X is not a rest point and $\phi_T(X) = X$ for some $T > 0$ then X is called a *periodic point* of ϕ_t.

* Received by the editors November 23, 1977, and in revised form March 9, 1978.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607.

Equation (2) determines a system of ordinary differential equations on $R^n \times R^n$ given by:

(3) $$\dot{X} = F(X), \qquad \dot{V} = DF(X)V$$

where $(X, V) \in R^n \times R^n$ and $DF(X)$ is the derivative matrix of F at X. This system is called the *linearized equations*, [9], corresponding to (2). The solution curve to (3) with initial condition $(X, V) \in R^n \times R^n$ is $\Psi_t(X, V) \equiv (\phi_t(X), D\phi_t(X)V)$. Here $\phi_t(X)$ is the solution to (2) and $D\phi_t(X)$ is the derivative matrix of ϕ_t at X. Ψ_t is called the *tangent flow*. If X is a periodic point of ϕ_t then the second equation of (3) is linear with periodic coefficients. In this case, Floquet theory may be used to find characteristic exponents [9] if the periodic orbit is known explicitly. In order to obtain information about the characteristic exponents, we study the component of the solution $D\phi_t(X)V$ normal to the vector field at $\phi_t(X)$. We make the pertinent definitions in a setting more general than that of a periodic orbit.

A compact subset Λ of R^n is an *invariant set* of the flow ϕ_t if, for each $t \in R$, Λ is a subset of the domain of ϕ_t and $\phi_t(\Lambda) = \Lambda$. For each $X \in \Lambda$, the vector field at X determines two subspaces of $R^n \times R^n$ called the *tangent* and *normal* subspaces:

$$E_X \equiv \{(X, V) \in R^n \times R^n \colon V = \alpha F(X) \text{ for some } \alpha \in R\},$$

$$N_X \equiv \{(X, V) \in R^n \times R^n \colon \langle V, F(X) \rangle = 0\}.$$

Here "$\langle \cdot , \cdot \rangle$" denotes the usual inner product on R^n. If Λ contains no rest points then $E \equiv \bigcup_{X \in \Lambda} E_X$ is a continuous line bundle over Λ and $N \equiv \bigcup_{X \in \Lambda} N_X$ is its normal bundle. In general, E and N are only semicontinuous. Let $\theta_X \colon \{X\} \times R^n \to N_X$ be the orthogonal projection of $\{X\} \times R^n$ onto N_X. Explicitly, $\theta_X(X, V) = (X, \pi_X(V))$ where $\pi_X(V) = V$ if $F(X) = 0$ and otherwise

$$\pi_X(V) = V - (\langle V, F(X) \rangle F(X) / \langle F(X), F(X) \rangle).$$

Since we will be concerned with only the second coordinate of θ_X, we will refer to π_X as the *normal projection*.

A compact invariant set Λ is *quasi-hyperbolic* if
(i) E is a continuous subbundle of $\Lambda \times R^n$, and
(ii) the flow on the quotient bundle $(\Lambda \times R^n)/E$ induced by the tangent flow has no nonzero bounded orbits.

Conditions (i) and (ii) guarantee that each rest point in Λ is an open-closed subset of Λ and Λ has only a finite number of rest points. Condition (ii) says that the normal projection of the orbit of each $(X, V) \in \Lambda \times R^n - E$ is unbounded, i.e., $|\pi_{\phi_t(X)}(D\phi_t(X)V)|$ is unbounded where "$|\cdot|$" denotes the Euclidean norm on R^n. Quasi-hyperbolicity is a weakening of the classical notion of hyperbolicity [14].

Λ is *hyperbolic* if there exist continuous invariant subbundles E^s and E^u of $\Lambda \times R^n$ so that $\Lambda \times R^n = E^s \oplus E^u \oplus E$ and if there exist real numbers $\lambda \in (0, 1)$ and $c > 0$ such that
(i) if $(X, V) \in E^u$ and $t > 0$ then $|D\phi_t(X)V| > c\lambda^{-t}|V|$, and
(ii) if $(X, V) \in E^s$ and $t > 0$ then $|D\phi_t(X)V| < c\lambda^t|V|$.

If Λ is connected and the dimension of E^s is k, then stable manifold theory [13] asserts the existence of a $(k+1)$-dimensional submanifold W^s of R^n containing Λ such that for each $Y \in W^s$ there is an $X \in \Lambda$ and $\alpha \in (0, 1)$ so that $|\phi_t(Y) - \phi_t(X)| < c\alpha^t$ for all $t > 0$. Thus the distance between $\phi_t(Y)$ and Λ approaches zero as $t \to \infty$. Similarly there is an unstable manifold corresponding to E^u whose points approach Λ as $t \to -\infty$.

In the presence of chain recurrence, hyperbolicity and quasi-hyperbolicity are equivalent. Given $X, Y \in R^n$ and $\varepsilon, T > 0$, an (ε, T)-chain from X to Y is two finite sequences, of points and times,

$$\{X_0 = X, X_1, \cdots, X_n = Y; t_0, t_1, \cdots, t_{n-1}\}$$

such that $t_i \geqq T$ and $|\phi_{t_i}(X_i) - X_{i+1}| < \varepsilon$ for all i, $0 \leqq i < n-1$. A point $X \in \Lambda$ is *Λ-chain recurrent* if for every $\varepsilon, T > 0$ there is an (ε, T)-chain from X to X contained in Λ. If all points in Λ are Λ-chain recurrent then Λ is called *chain recurrent.* For example, rest points and periodic orbits are chain recurrent. The following are proved in [5]:

THEOREM 2.1. *Let Λ be a compact invariant set of the flow ϕ_t. If Λ is quasi-hyperbolic and chain recurrent, then Λ is hyperbolic.*

THEOREM 2.2. *Let Λ be a compact invariant set of ϕ_t. If Λ is hyperbolic and chain recurrent, then Λ is contained in the closure of the set of periodic points of ϕ_t.*

3. General theory. In this section we establish the theoretical results needed for the computer analysis.

If Λ is a subset of R^n and J a subset of R, define $\phi_J(\Lambda) \equiv \bigcup_{t \in J} \phi_t(\Lambda)$. Let Λ be a compact subset of the domain of ϕ_t for all $t \geqq 0$. Then the *ω-limit set* of Λ is

$$\omega(\Lambda) \equiv \bigcap_{t \geqq 0} \mathscr{C}\ell(\phi_{(t,\infty)}(\Lambda))$$

where $\mathscr{C}\ell$ denotes the toplogical closure. Intuitively, $\omega(\Lambda)$ is the positive end of the orbit of Λ. Similarly, the *α-limit set* of Λ is defined as the negative end of the orbit of Λ. A subset Λ of R^n is an *attractor* for ϕ_t if it has a compact neighborhood U such that $\Lambda = \omega(U)$. An attractor is compact and invariant.

PROPOSITION 3.1. *Let U be a nonempty, open, connected subset of R^n with compact closure. If there is a $T > 0$ such the $\mathscr{C}\ell U$ is contained in the domain of ϕ_T and $\phi_T(\mathscr{C}\ell U)$ is contained in U, then $\omega(\mathscr{C}\ell U)$ is a nonempty connected attractor contained in U.*

Proof. $\omega(\mathscr{C}\ell U)$ is connected since U is connected. What remains to be shown is that $\mathscr{C}\ell(\phi_{(t,\infty)}(\mathscr{C}\ell U))$ is a subset of U for large t.

Notice that $\phi_T(\mathscr{C}\ell U) = \mathscr{C}\ell\phi_T(U)$. Since $\mathscr{C}\ell\phi_T(U)$ is compact, there is a minimum distance d from $\mathscr{C}\ell\phi_T(U)$ to the complement of U. Because of the compactness of $\mathscr{C}\ell U$, there is a time S such that for each $Y \in \mathscr{C}\ell U$ we have

$$|Y - \phi_T(Y)| < d \quad \text{for all } t \in [0, S).$$

Let $X \in \mathscr{C}\ell U$. Since $\phi_T(X) \in \phi_T(\mathscr{C}\ell U)$, $\phi_{(T-S,T]}(X)$ is contained in U. Thus the orbit segment $\phi_{[0,T]}(X)$ intersects the complement of U for at most $T-S$ units of time. Now consider $\phi_T(\phi_{[O,T]}(X))$. Since $\phi_T(X)$ and $\phi_T(\phi_{[T-S,T]}(X))$ are contained in $\phi_T(\mathscr{C}\ell U)$, the orbit segment $\phi_{[T,2T]}(X)$ intersects the complement of U for at most $(T-S)-2S$ units of time. Continue to apply ϕ_T to the orbit segments of X until the multiple of S is larger than T, thus showing that segment $\phi_{[mT,(m+1)T]}(X)$ is a subset of U for some integer m. Since m is independent of $X \in \mathscr{C}\ell U$, we have that $\phi_{[mT,\infty)}(\mathscr{C}\ell U)$ is contained in U and so $\phi_{[(m+1)T,\infty)}(\mathscr{C}\ell U)$ is contained in $\phi_T(\mathscr{C}\ell U)$. Thus for $t = (m+1)T$ we have

$$\mathscr{C}\ell(\phi_{(t,\infty)}(\mathscr{C}\ell U)) \subset \phi_T(\mathscr{C}\ell U) \subset U.$$

COROLLARY 3.2. *Suppose U is an open annulus in R^2 with boundary ∂U and suppose the flow ϕ_t has a rest point within the inner bounding circle of the annulus. If there is a $T > 0$ such that $\phi_T(\partial U)$ is contained in U, then $\omega(\mathscr{C}\ell U)$ is a connected attractor contained in U.*

Proof. Using domain invariance and the Jordan curve theorem (see [3, chap. 17]), it follows that $\phi_T(\mathscr{C}\ell U)$ is a subset of U.

In § 4 we use the computer to verify the hypothesis of Corollary 3.2 pertaining to (1). We construct a grid of disks covering the two bounding circles of the annulus. Then the computer uses a numerical method to show that after a fixed time T the orbits through the centers of the disks are inside the annulus. In fact, each center is far enough inside to show that the entire disk is inside the annulus even compensating for the numerical error and the machine error. Thus the existence of an attractor Λ inside the annulus is established.

Now we discuss results which allow us to conclude that Λ is a periodic orbit whose nontrivial characteristic exponents have negative real parts. Henceforth the positional subscipt on π_X will be suppressed to simplify notation.

PROPOSITION 3.3. *Let Λ be a compact invariant set of ϕ_t. Suppose there is a $T>0$ so that for each $X\in\Lambda$ and $(X, W)\in N_X$ with $|W|=1$ we have*

$$|\pi(D\phi_T(X)W)|<1.$$

Then there exists a $\beta>1$ so that for all $(X, V)\in\Lambda\times R^n$

$$|\pi(D\phi_{-T}(X)V)|\geqq\beta|\pi(V)|.$$

Proof. If $(X, V)\in E_X$ then both sides of the concluding inequality are zero. This follows because E is invariant under the tangent flow.

If $(X, V)\notin E_X$ then $\pi(V)\in N_X$ is not zero. By compactness, there is a $\alpha\in(0, 1)$ so that for all $(X, W)\in N_X$ with $|W|=1$ we have

$$|\pi(D\phi_T(X)W)|\leqq\alpha.$$

From linearity, it follows that for all $(X, W)\in N_X$

$$|\pi(D\phi_T(X)W)|\leqq\alpha|W|.$$

Let $Y\equiv D\phi_{-T}(X)V$. Since $Y-\pi(Y)\in E_{\phi_{-T}(X)}$, $D\phi_T(\phi_{-T}(X))(Y-\pi(Y))$ belongs to E_X. Since $D\phi_T(\phi_{-T}(X))Y=V$, $\pi(V)=\pi(D\phi_T(\phi_{-T}(X))\pi(Y))$. Also since $|\pi(D\phi_T(\phi_{-T}(X))\pi(Y))|\leqq\alpha|\pi(Y)|$, we have

$$|\pi(V)|\leqq\alpha|\pi(D\phi_{-T}(X)V)|.$$

Taking $\beta=\alpha^{-1}$ completes the proof.

COROLLARY 3.4. *Let Λ be a compact invariant set of ϕ_t. Suppose there is a $T>0$ so that for each $X\in\Lambda$ and $(X, W)\in N_X$ with $|W|=1$ we have*

$$|\pi(D\phi_T(X)W)|<1.$$

Then $\{\pi(D\phi_t(X)V): t\leqq 0\}$ is unbounded for all $(X, V)\in\Lambda\times R^n-E$.

The conclusion of Corollary 3.4 plus a condition on the rest points is even stronger than quasi-hyperbolicity. Assuming the conclusion of Corollary 3.4 we show that Λ is a periodic orbit or rest point.

THEOREM 3.5. *Suppose Λ is a compact, connected, invariant set and suppose that for each $X\in\Lambda$ the set $\{\pi(D\phi_t(X)V): t\leqq 0\}$ is unbounded for all $(X, V)\in\Lambda\times R^n-E$. Then Λ is a periodic orbit or rest point whose nontrivial characteristic exponents have negative real parts.*

Proof. If $X\in\Lambda$ is a rest point then, by elementary linear theory [9], it is hyperbolic with n-dimensional stable manifold. Thus X has an open neighborhood with the orbit of each point in the neighborhood limiting on X as $t\to\infty$; and no orbit can limit on X as $t\to-\infty$.

If $X \in \Lambda$ is not a rest point, the α-limit set $\alpha(X)$ is a connected chain recurrent subset of Λ [6]. $\alpha(X)$ contains no rest points because the orbit of X would limit on such a point as $t \to -\infty$. Thus, by hypothesis, $\alpha(X)$ is quasi-hyperbolic; and Theorem 2.1 implies that $\alpha(X)$ is hyperbolic with n-dimensional stable manifold. So there is a neighborhood of $\alpha(X)$ such that the orbit of each point limits on $\alpha(X)$ as $t \to \infty$. Hence this neighborhood of $\alpha(X)$ contains no periodic points except those in $\alpha(X)$. Since Theorem 2.2 implies that $\alpha(X)$ is contained in the closure of the set of periodic points, $\alpha(X)$ contains a periodic point. This periodic orbit is hyperbolic with no nontrivial unstable manifold, so no orbits limit on this periodic orbit as $t \to -\infty$. In particular, the orbit of X cannot limit on it unless X is on the periodic orbit.

Thus every point in Λ is a periodic or rest point. But each such point has a neighborhood containing no other periodic or rest point. Hence, by connectedness, Λ is one periodic orbit or rest point. The condition on the characteristic exponents follows from Floquet theory.

In § 5 we use the computer to verify the hypothesis of Proposition 3.3 for the linearized equations corresponding to (1). Hence, Corollary 3.4 and Theorem 3.5 show that the invariant set within the annulus is a periodic orbit with characteristic exponents having negative real parts.

4. Finding an attractor. In §§ 4 and 5 we willustrate the numerical analysis, the specific error estimates, and the computer program needed to apply our method to show the existence of an asymptotically stable periodic orbit in (1). This result can be obtained by other methods. The existence of the periodic solution is obvious after changing the equations to polar coordinates and the Poincaré criterion [2] implies that one characteristic exponent has a negative real part.

Equation (1) has a rest point at the origin. So in this section we use the computer and Corollary 3.2 to show that the annulus $U(.99, 1.01)$ centered at the origin with inner radius .99 and outer radius 1.01 contains an attractor. Preliminary calculations indicate that the bounding circles of $U(.99, 1.01)$ move inside of $U(.99, 1.01)$ in time $T = .64$. First we show that certain orbits remain in $U(.94, 1.06)$ for .64 units of time. Many of the error bounds will be computed over $U(.94, 1.06)$, but care must be taken since this annulus is not convex.

Since 1.064 is an upper bound on the vector field in $U(.94, 1.06)$ a solution remaining in $U(.94, 1.06)$ for time .04 can travel a distance at most .043 from its initial position. Thus a solution with initial position in $U(.988, 1.012)$ remains in $U(.94, 1.06)$ for time .04.

LEMMA 4.1. *If $X_1 \in U(.987, 1.013)$ and $|X_1 - X_2| < .001$ then the line segment from $\phi_t(X_1)$ to $\phi_t(X_2)$ belongs to $U(.94, 1.06)$ for all $t \in [0, .04]$.*

Proof. The triangle inequality and the proceeding paragraph imply that

$$|X_1 - \phi_t(X_2)| \leqq |X_1 - X_2| + |X_2 - \phi_t(X_2)| < .001 + .043.$$

And $|X_1 - \phi_t(X_1)| < .043$. Therefore $\phi_t(X_1)$ and $\phi_t(X_2)$ belong to the disk of radius .044 centered at X_1. Since this disk is convex and is contained in $U(.94, 1.06)$, the desired result follows.

Let $\mathscr{D}(r)$ be the disk of radius r centered at the origin. A Lipschitz contant $L = 1.15$ for the vector field in the disk $\mathscr{D}(1.06)$ is computed as the maximum norm for DF over $\mathscr{D}(1.06)$. See [2] for the following result:

LEMMA 4.2. *If $\phi_t(X_1)$, $\phi_t(X_2) \in \mathscr{D}(1.06)$ for all $t \in [0, T]$, then*

$$|\phi_t(X_1) - \phi_t(X_2)| \leqq e^{Lt}|X_1 - X_2|.$$

In particular, for $T=.04$, we have

$$|\phi_t(X_1)-\phi_t(X_2)|\leqq e^{(1.15)(.04)}|X_1-X_2|<1.048|X_1-X_2|.$$

The previous lemmas will be used in investigating the truncation error due to the numerical method. An upper bound for the local error for 5th order Taylor series is $M_1h^6/6!$ where M_1 is the maximum over $X\in U(.94, 1.06)$ of the norm of the sixth derivative of $\phi_t(X)$ with respect to t. For (1), $M_1\leqq 83.85$ and so the local error is bounded by 4.77×10^{-10} for $h=.04$.

Let X^i denote the ith iterate of the Taylor series numerical method with step size .04 applied to (1). The next lemma states that keeping track of the first 16 numerical iterates can be used to show that the true solution remains in $U(.94, 1.06)$ for $0\leqq t\leqq .64$. To prove this, the cumulative error between each iterate and the true solution is calculated.

LEMMA 4.3. *Suppose* X^i, $i=0,\cdots,16$, *belong to* $U(.988, 1.012)$. *Then* $\phi_t(X^0)$ *belong to* $U(.94, 1.06)$ *for* $t\in[0, .64]$. *Also*

$$|X^i-\phi_{i(.04)}(X^0)|<(4.77\times 10^{-10})\sum_{j=0}^{i-1}(e^{.04L})^j$$

where $L=1.15$ *and so* $e^{.04L}<1.048$.

Proof. We induct on i, the number of iterates of the numerical method. For $i=1$, $\phi_t(X^0)$ belongs to $U(.94, 1.06)$ for $0\leqq t\leqq .04$ since $|X^0-\phi_t(X^0)|\leqq .043$. Thus our upper bound for the local error due to the numerical method can be used, i.e.,

$$|X^1-\phi_{.04}(X^0)|<4.77\times 10^{-10}<.001.$$

Since $X^1\in U(.988, 1.012)$, we have $\phi_{.04}(X^0)\in U(.987, 1.013)$.

By induction assume $\phi_t(X^0)\in U(.94, 1.06)$ for all t, $0\leqq t\leqq i(.04)$; and

$$|X^i-\phi_{i(.04)}(X^0)|<(4.77\times 10^{-10})\sum_{j=0}^{i-1}(1.048)^j.$$

We now show the result for $i+1\leqq 16$.

Since $(4.77)\times 10^{-10}\sum_{j=0}^{i-1}(1.048)^j\leqq 1.12\times 10^{-8}$, we have $|X^i-\phi_{i(.04)}(X^0)|<.001$ and X^i, $\phi_{i(.04)}(X^0)\in U(.987, 1.013)$. Thus Lemma 4.1 gives that $\phi_t(X^i)$, $\phi_{(i+t)(.04)}(X^0)$ $\in U(.94, 1.06)$ for all $t\in[0, .04]$. Also another step of the numerical method can be used to generate X^{i+1}. Using Lemma 4.2, we get:

$$\begin{aligned}|X^{i+1}-\phi_{(i+1)(.04)}(X^0)|&\leqq|X^{i+1}-\phi_{.04}(X^i)|+|\phi_{.04}(X^i)-\phi_{(i+1)(.04)}(X^0)|\\&<4.77\times 10^{-10}+(1.048)(4.77\times 10^{-10})\sum_{j=0}^{i-1}(1.048)^j\\&=(4.77\times 10^{-10})\sum_{j=0}^{i}(1.048)^j.\end{aligned}$$

Thus the result is established for $i+1$.

COROLLARY 4.4. *Suppose* X^i, $i=0,\cdots,16$, *belong to* $U(.988, 1.012)$ *then* $|X^i-\phi_{i(.04)}(X^0)|<1.12\times 10^{-8}$ *and so* $\phi_{i(.04)}(X^0)\in U(.987, 1.013)$. *In fact*, $\phi_t(X^0)\in U(.965, 1.035)$ *for all* $t\in[0, .64]$.

Proof. The first assertion follows from Lemma 4.3. Then since the orbit of $\phi_{i(.04)}(X^0)$ leaves and returns to $U(.987, 1.013)$ in time .04, the farthest it can travel in .02 units of time is $(.02)(1.064)<.022$. Hence $\phi_{t+i(.04)}(X^0)\in U(.965, 1.035)$ for $0\leqq t\leqq .04$ and $i=0,\cdots,15$.

R. Martin has shown us how we can improve on the Lipschitz constant $L = 1.15$ used in Lemma 4.2. First we need a result from linear algebra [12, p. 140]. Let "$\langle \cdot , \cdot \rangle$" denote the usual inner product on R^n.

THEOREM 4.5. *If A is a real $n \times n$ matrix with transpose A^* and $X \in R^n$, then $\langle AX, X\rangle \leqq K\langle X, X\rangle$ where K is any upper bound for the eigenvalues of $(A + A^*)/2$.*

LEMMA 4.6. *Let F be a continuously differentiable vector field on an open set $G \subset R^n$. Let $\phi_t(X)$ denote solutions to $\dot{X} = F(X)$. If $X, Y \in G$ such that the line segments connecting $\phi_t(X)$ and $\phi_t(Y)$ are in G for $t \in [0, T]$, then*

$$|\phi_t(X) - \phi_t(Y)| \leqq e^{Kt}|X - Y|$$

where $K = \sup_{Z \in G} \{\lambda(Z)$: λ is the largest eigenvalue of $(DF(Z) + DF(Z)^*)/2\}$.

Proof. Fix $X, Y \in G$. For each $t \in [0, T]$, define g_t: $[0, 1] \to R^n$ by

$$g_t(s) = s\phi_t(X) + (1 - s)\phi_t(Y).$$

By assumption the straight line image of g_t is in G. Using Theorem 4.5 and the fundamental theorem of calculus, we have

$$\begin{aligned}
&\langle \phi_t(X) - \phi_t(Y), F(\phi_t(X)) - F(\phi_t(Y))\rangle \\
&= \left\langle \phi_t(X) - \phi_t(Y), \int_0^1 DF(g_t(s))(\phi_t(X) - \phi_t(Y))\, ds \right\rangle \\
&= \int_0^1 \langle \phi_t(X) - \phi_t(Y), DF(g_t(s))(\phi_t(X) - \phi_t(Y))\rangle\, ds \\
&\leqq \int_0^1 K|\phi_t(X) - \phi_t(Y)|^2\, ds.
\end{aligned}$$

Now define h: $[0, T] \to R$ by $h(t) = |\phi_t(X) - \phi_t(Y)|^2$. Using the preceding inequality, we get

$$\begin{aligned}
\frac{dh}{dt} &= 2\langle \phi_t(X) - \phi_t(Y), F(\phi_t(X)) - F(\phi_t(Y))\rangle \\
&\leqq 2K|\phi_t(X) - \phi_t(Y)|^2 = 2Kh(t).
\end{aligned}$$

Thus $h(t) \leqq e^{2Kt}h(0)$ and so $|\phi_t(X) - \phi_t(Y)| \leqq e^{Kt}|X - Y|$.

To bound the largest eigenvalue of $(DF(Z) + DF(Z)^*)/2$ we use a theorem of Gershgorin [12, p. 146]. Let a_{ij} denote the ij component of $(DF(Z) + DF(Z)^*)/2$. Then

$$\lambda(Z) \leqq \max_j \left\{ a_{jj} + \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \right\}. \tag{4}$$

For (1), taking the supremum over $Z \in U(.94, 1.06)$, we conclude that $K < .03427$.

LEMMA 4.7. *Suppose $X^0 \in U(.988, 1.012)$ and $\phi_t(X^0) \in U(.965, 1.035)$ for all $t \in [0, .64]$. If $|X - X^0| < .0242$ and $|Y - X^0| < .0242$, then for $t \in [0, .64]$ the line segment from $\phi_t(X)$ to $\phi_t(Y)$ is in $U(.944, 1.06)$ and*

$$|\phi_t(X) - \phi_t(Y)| \leqq e^{Kt}|X - Y|$$

where K is as in Lemma 4.6.

Proof. For each X such that $|X - X^0| < .0242$, let $s > 0$ be the first time that the line segment from $\phi_s(X)$ to $\phi_s(X^0)$ intersects $\partial U(.94, 1.06)$. (Note s may be ∞.) If $s \geqq .64$ then Lemma 4.6 gives

$$|\phi_t(X) - \phi_t(X^0)| \leqq e^{Kt}|X - X^0| \quad \text{for all } t \in [0, .64].$$

Assume $s<.64$. Since line segments from $\phi_t(X)$ to $\phi_t(X^0)$ belong to $U(.94, 1.06)$ for all $t\leqq s$ and $\phi_s(X^0)\in U(.965, 1.035)$, then

$$.025\leqq|\phi_s(X)-\phi_s(X^0)|\leqq e^{Ks}|X-X^0|<e^{(.03427)(.64)}|X-X^0|<.0248.$$

This contradiction implies that $s\geqq .64$. Thus for $t\in[0, .64]$,

$$|\phi_t(X)-\phi_t(X^0)|\leqq e^{Kt}|X-X^0|<.0248.$$

This same inequality is true if Y replaces X. So both $\phi_t(X)$ and $\phi_t(Y)$ belong to a disk of radius .0248 centered at $\phi_t(X^0)$. This disk is contained in $U(.94, 1.06)$ so the line segment from $\phi_t(X)$ to $\phi_t(Y)$ is in $U(.94, 1.06)$. Therefore, Lemma 4.6 asserts that

$$|\phi_t(X)-\phi_t(Y)|\leqq e^{Kt}|X-Y|.$$

This completes the proof.

For $X^0\in U(.988, 1.012)$ and X such that $|X-X^0|<.01$, the triangle inequality implies

$$|X^{16}-\phi_{.64}(X)|\leqq|X^{16}-\phi_{.64}(X^0)|+|\phi_{.64}(X^0)-\phi_{.64}(X)|.$$

If $X^0, X^1, \cdots, X^{16}\in U(.988, 1.012)$ then Corollary 4.4 implies that $\phi_t(X^0)\in U(.965, 1.035)$ for all $t\in[0, .64]$ and so

$$|X^{16}-\phi_{.64}(X)|<1.12\times10^{-8}+|\phi_{.64}(X^0)-\phi_{.64}(X)|.$$

Using Lemma 4.7, we get

$$(5)\qquad |X^{16}-\phi_{.64}(X)|<1.12\times10^{-8}+e^{.64K}|X-X^0|<1.12\times10^{-8}+(1.023)|X-X^0|.$$

Equation (5) bounds the error between the 16th iterate starting at X^0 and the true position $\phi_{.64}(X)$ for X near X^0.

We use a PL/1 program on an I.B.M. 360 computer to show that $\partial U(.99, 1.01)$ gets mapped inside $U(.99, 1.01)$ by $\phi_{.64}$. Using double precision variables, we are able to bound the total roundoff error by 1×10^{-12}, which is insignificant when compared with the error in (5). We wish to thank S. Danielopoulos for his assistance in computing the roundoff error.

The program uses a fifth order Taylor method with 16 steps of size .04. The initial position is on $\partial U(.99, 1.01)$. Each iterate is checked to see that it is in $U(.988, 1.012)$. Hence Lemma 4.3 and Corollary 4.4 apply to show that $\phi_t(X^0)\in U(.965, 1.035)$ for all $t\in[0, .64]$. So, if $|X-X^0|<.01$, we have equation (5). Then X^{16} is checked to be in the interior of $U(.99, 1.01)$ by more than 1.12×10^{-8}. Let $\underset{\sim}{a}$ be the distance X^{16} is within $\partial U(.99, 1.01)$. By equation (5), $\phi_{.64}(X)$ is in the interior of $U(.99, 1.01)$ for all X such that

$$|X-X^0|<r\equiv\min\{(\underset{\sim}{a}-1.12\times10^{-8})/1.023, .01\}.$$

This shows that a disk of radius r centered at X^0 is taken inside $U(.99, 1.01)$ by $\phi_{.64}$. The next initial position is chosen to be a distance of 1.99 r along $\partial U(.99, 1.01)$ from X^0. A disk centered at this initial position (or grid point) is determined and the union of these two disks is shown to cover the arc in $\partial U(.99, 1.01)$ connecting the two initial positions. The program continues to choose grid points until the $\partial U(.99, 1.01)$ is covered. In fact, because of the symmetry in (1), we need only cover the portion of $\partial U(.99, 1.01)$ in the lower half-plane. Thus, with Corollary 3.2, we have established the following:

THEOREM 4.8. *Let $\phi_t(X)$ denote the solution curve to equation* (1) *with initial condition $X\in R^2$. Then $\phi_{.64}(\partial U(.99, 1.01))$ is contained in the interior of* $U(.99, 1.01)$; *and* $U(.99, 1.01)$ *contains a connected attractor.*

Also we have the following results which will be used in the next section.

LEMMA 4.9. *If* $X \in U(.98, 1.02)$ *then* $\phi_t(X) \in U(.94, 1.06)$ *for all* $t \in [0, .64]$.

Proof. Since $X \in U(.98, 1.02)$, X is within a distance of .01 from $\partial U(.99, 1.01)$. Each point of $\partial U(.99, 1.01)$ is within .01 of some grid point $X^0 \in \partial U(.99, 1.01)$ and so $|X - X^0| < .02$. The computer program and Corollary 4.4 give that $\phi_t(X^0) \in U(.965, 1.035)$ for all $t \in [0, .64]$. Thus Lemma 4.7 implies

$$|\phi_t(X) - \phi_t(X^0)| \leqq e^{Kt}|X - X^0| < .02046.$$

Therefore $\phi_t(X) \in U(.94, 1.06)$.

LEMMA 4.10. *For each* $Y \in U(.98, 1.02)$ *and* X *such that* $|Y - X| < .01$ *then the line segments from* $\phi_t(Y)$ *to* $\phi_t(X)$ *belong to* $U(.94, 1.06)$ *for all* $t \in [0, .64]$.

Proof. Since $Y \in U(.99, 1.01)$, Y is within $.01\sqrt{2}$ of some grid point X^0. Since $.01\sqrt{2} < .0142$ and $|X - Y| < .01$, we have $|X^0 - X| < .0242$. The computer program and Corollary 4.4 give that $\phi_t(X^0) \in U(.965, 1.035)$ for all $t \in [0, .64]$. Thus Lemma 4.7 implies the desired result.

5. Characteristic exponents. The previous section establishes the existence of an attractor Λ contained within $U(.99, 1.01)$. Now we intend to use Theorem 3.5 to show Λ is an asymptotically stable periodic orbit. Note that $F(X) \neq 0$ for all $X \in \Lambda$, so Λ contains no rest points. For each $X \in U(.99, 1.01)$ and (X, V) in the normal subspace at X with $|V| = 1$, we will show that the normal projection $|\pi(D\phi_{.64}(X)V)|$ is less than 1. Hence this is true for all $X \in \Lambda$; and Proposition 3.3 and Theorem 3.5 apply. The error bounds are computed in $U(.94, 1.06) \times \mathscr{D}(1.06)$.

The linearized system corresponding to (1) is:

$$\begin{aligned} \dot{x} &= y, \\ \dot{y} &= \alpha(1 - x^2 - y^2)y - x, \\ \dot{u} &= v, \\ \dot{v} &= (-2\alpha xy - 1)u + \alpha(1 - x^2 - y^2)v - 2\alpha y^2 v, \end{aligned} \tag{6}$$

where $(x, y, u, v) = (X, V) \in R^2 \times R^2$. To simplify notation let $W = (x, y, u, v) \in R^4$ and let $\Psi_t(W) = (\phi_t(X), D\phi_t(X)V)$ be the solution to (6) with initial condition W. If the convexity condition in Lemma 4.6 holds, we conclude that

$$|\Psi_t(W_1) - \Psi_t(W_2)| \leqq e^{Kt}|W_1 - W_2|$$

for all $t \in [0, .64]$. Here K is computed from the vector field of (6). Maximizing over $U(.94, 1.06) \times \mathscr{D}(1.06)$ we get $K = .3714$.

LEMMA 5.1. *If* $(X, V) \in U(.98, 1.02) \times \mathscr{D}(1.012)$, *then* $\Psi_t(X, V) \in U(.94, 1.06) \times \mathscr{D}(1.0426)$ *for all* $t \in [0, .08]$.

Proof. Consider the solution to (6) with initial condition $(X, 0)$. Note $\Psi_t(X, 0) = (\phi_t(X), 0)$ for all $t \geqq 0$. Since $X \in U(.98, 1.02)$, Lemma 4.9 asserts that $\phi_t(X) \in U(.94, 1.06)$ for $t \in [0, .08]$. The convexity condition in Lemma 4.6 is satisfied as long as $D\phi_t(X)V$ remains in the disk $\mathscr{D}(1.06)$. But the first time $D\phi_t(X)V$ meets the $\partial\mathscr{D}(1.06)$ is greater than .08 because

$$|D\phi_t(X)V| = |\Psi_t(X, V) - \Psi_t(X, 0)| \leqq e^{K(.08)}|V| < e^{(.3714)(.08)}(1.012) \leqq 1.0426.$$

So the assertion is proved.

The norm of the sixth derivative of $\Psi_t(X, V)$ with respect to t maximized over $U(.94, 1.06) \times \mathscr{D}(1.06)$ is 375.51. Hence the error due to a fifth order Taylor series approximation with step size $h = .08$ is

$$\frac{375.51(.08)^6}{6!} \leqq 1.37 \times 10^{-7}.$$

Let $W^i = (X^i, V^i)$ denote the ith iterate of the Taylor method with step size .08 (the step size is doubled to reduce cost) applied to (6). The following lemma uses numerical iterates to show the true solution remains in $U(.94, 1.06) \times \mathscr{D}(1.0426)$ for $t \in [0, .64]$ and computes the truncation error.

LEMMA 5.2. *Suppose* W^i, $i = 0, 1, \cdots, 8$, *belong to* $U(.988, 1.012) \times \mathscr{D}(1.01)$. *Then* $\Psi_t(W^0)$ *belongs to* $U(.94, 1.06) \times \mathscr{D}(1.0426)$ *for* $t \in [0, .64]$ *and for* $i = 1, 2, \cdots, 8$.

$$|W^i - \Psi_{i(.08)}(W^0)| \leqq (1.37 \times 10^{-7}) \sum_{j=0}^{i-1} (e^{.08K})^j < 1.219 \times 10^{-6}$$

where $K = .3714$.

Proof. We induct on i. For $i = 1$, note that Lemma 5.1 implies $\Psi_t(W^0) \in U(.94, 1.06) \times \mathscr{D}(1.0426)$ for $t \in [0, .08]$ since $W^0 \in U(.98, 1.02) \times \mathscr{D}(1.012)$. Therefore, we can use the local truncation error to get

$$|W^1 - \Psi_{.08}(W^0)| < 1.37 \times 10^{-7} < 1.219 \times 10^{-6}.$$

By induction assume $|W^i - \Psi_{i(.08)}(W^0)| < (1.37 \times 10^{-7}) \sum_{j=0}^{i-1} e^{.08Kj}$ and $\Psi_t(W^0) \in U(.94, 1.06) \times \mathscr{D}(1.0426)$ for all $0 \leqq t \leqq i(.08)$. We now prove the result for $i + 1 \leqq 8$. First we establish the convexity condition needed for Lemma 4.6.

Since $W^i = (X^i, V^i) \in U(.988, 1.012) \times \mathscr{D}(1.01)$, X^i is within a distance .01 of $\partial U(.99, 1.01)$. Thus there is a grid point $Z \in \partial U(.99, 1.01)$ so that $|X^i - Z| < .02$. The computer and Corollary 4.4 have shown that $\phi_t(Z) \in U(.965, 1.035)$ for $t \in [0, .64]$. Also

$$\begin{aligned} |Z - \phi_{i(.08)}(X^0)| &\leqq |Z - X^i| + |X^i - \phi_{i(.08)}(X^0)| \\ &< .02 + 1.219 \times 10^{-6} \\ &< .0242. \end{aligned}$$

So X^i and $\phi_{i(.08)}(X^0)$ are within a disk of radius .0242 of Z. Therefore, Lemma 4.7 implies that the line segments from $\phi_t(X^i)$ to $\phi_t(\phi_{i(.08)}(X^0))$ belong to $U(.94, 1.06)$ for $0 \leqq t \leqq .08$. Also, since $V^i \in \mathscr{D}(1.01)$ and $D\phi_{i(.08)}(X^0)V^0 \in \mathscr{D}(1.012)$, Lemma 5.1 asserts that $D\phi_t(X^i)V^i$ and $D\phi_{t+i(.08)}(X^0)V^0$ belong to the convex set $\mathscr{D}(1.0426)$ for $0 \leqq t \leqq .08$. We may conclude that the line segments from $\Psi_t(W^i)$ to $\Psi_t(\Psi_{i(.08)}(W^0))$ belong to $U(.94, 1.06) \times \mathscr{D}(1.0426)$ for all $t \in [0, .08]$. Hence Lemma 4.6 gives

$$\begin{aligned} |\Psi_{.08}(W^i) - \Psi_{(i+1)(.08)}(W^0)| &\leqq e^{K(.08)} |W^i - \Psi_{i(.08)}(W^0)| \\ &\leqq e^{K(.08)} (1.37 \times 10^{-7}) \sum_{j=0}^{i-1} e^{.08Kj}. \end{aligned}$$

Using the local truncation error, we see that $|W^{i+1} - \Psi_{.08}(W^i)| < 1.37 \times 10^{-7}$. By the triangle inequality

$$\begin{aligned} |W^{i+1} - \Psi_{(i+1)(.08)}(W^0)| &\leqq |W^{i+1} - \Psi_{.08}(W^i)| + |\Psi_{.08}(W^i) - \Psi_{(i+1)(.08)}(W^0)| \\ &< 1.37 \times 10^{-7} + (1.37 \times 10^{-7}) e^{K(.08)} \sum_{j=0}^{i-1} e^{.08Kj} \\ &= (1.37 \times 10^{-7}) \sum_{j=0}^{i} e^{.08Kj} \\ &< 1.219 \times 10^{-6}. \end{aligned}$$

This completes the proof.

LEMMA 5.3. *Let* $(X^0, V^0) \in U(.99, 1.01) \times \mathcal{D}(1)$ *and choose* (X, V) *such that* $|(X, V)-(X^0, V^0)| < .01$. *If* $\Psi_t(X^0, V^0) \in U(.94, 1.06) \times \mathcal{D}(1.0426)$ *for all* $t \in [0, .64]$ *then* $\Psi_t(X, V) \in U(.94, 1.06) \times \mathcal{D}(1.06)$ *and*

$$|\Psi_t(X, V)-\Psi_t(X^0, V^0)| \leqq e^{Kt}|(X, V)-(X^0, V^0)| < .01269.$$

Proof. By Lemma 4.10 the straight line connecting $\phi_t(X)$ and $\phi_t(X^0)$ are in $U(.94, 1.06)$ for all $t \in [0, .64]$. Thus the straight line connecting $\Psi_t(X, V)$ and $\Psi_t(X^0, V^0)$ will be in $U(.94, 1.06) \times \mathcal{D}(1.06)$ as long as $|D\phi_t(X)(V)| < 1.06$. Since $D\phi_t(X^0)(V^0) \in \mathcal{D}(1.0426)$ the above inequality will be true while $|D\phi_t(X)(V) - D\phi_t(X^0)(V^0)| < .0174$. Now

$$\begin{aligned}|\Psi_t(X, V)-\Psi_t(X^0, V^0)| &\leqq e^{Kt}|(X, V)-(X^0, V^0)| \\ &\leqq e^{.3714(.64)}(.01) \\ &\leqq .01269\end{aligned}$$

for $t \in [0, .64]$. Hence $\Psi_t(X, V) \in U(.94, 1.06) \times \mathcal{D}(1.06)$.

Lemmas 5.2 and 5.3 allow us to bound the distance between the eigth computer iterate of W^0 and $\Psi_{.64}(W)$ where W is within .01 of W^0. When W^i, $i = 0, 1, \cdots, 8$ are checked to be in $U(.988, 1.012) \times \mathcal{D}(1.01)$ then

$$\begin{aligned}|W^8 - \Psi_{.64}(W)| &\leqq |W^8 - \Psi_{.64}(W^0)| + |\Psi_{.64}(W^0) - \Psi_{.64}(W)| \\ &< 1.219 \times 10^{-6} + 1.269|W^0 - W| \\ &\leqq .012691219.\end{aligned} \tag{7}$$

Likewise the distance between the eighth computer iterate of X^0 and $\Psi_{.64}(X)$ where $|X - X^0| < .01$ can be bounded. So we analyze the errors for a Taylor method applied to (1) as was done for (5) except now the step size is .08 instead of .04. Since $M_1 = 83.85$ and $h = .08$, the local error $M_1 h^6/6!$ is bounded by 3.053×10^{-8}. For $K = .03427$

$$|X^8 - \phi_{.64}(X^0)| \leqq (3.053 \times 10^{-8}) \sum_{j=0}^{7} e^{.08Kj} < 2.466 \times 10^{-7}.$$

Thus

$$\begin{aligned}|X^8 - \phi_{.64}(X)| &\leqq |X^8 - \phi_{.64}(X^0)| + |\phi_{.64}(X^0) - \phi_{.64}(X)| \\ &< 2.466 \times 10^{-7} + e^{.64K}|X^0 - X| \\ &< 2.466 \times 10^{-7} + 1.023|X^0 - X|.\end{aligned} \tag{8}$$

To use Theorem 3.5, we must show that for each $X \in U(.99, 1.01)$ and $(X, V) \in N_X$ with $|V| = 1$, the normal projection of the orbit of V has shrunk in time $T = .64$, i.e.,

$$\pi_{\phi_{.64}(X)}(D\phi_{.64}(X)V)| < 1.$$

Thus the initial position $V^0 = (u^0, v^0)$ is the unit vector normal to the vector field

$$F(X^0) = (y^0, \alpha(1-(x^0)^2-(y^0)^2)y^0 - x^0).$$

For X near X^0, $|V - V^0|$ can be bounded in terms of $|X - X^0|$. Since $|V - V^0|$ is not changed in taking the normals, the total change comes in making them unit vectors. On $U(.98, 1.02)$, $|F(X)| > .978$. Hence

$$|V - V^0| < \frac{1}{.978}|F(X) - F(X^0)|.$$

Since $|X-X^0|<.01$ and $X^0\in U(.99, 1.01)$, we apply the mean value theorem to a disk of radius .01 inside $U(.98, 1.02)$ to see that

$$\begin{aligned}|F(x, y)-F(x^0, y^0)|&\\ &\leqq ((y-y^0)^2+[1.05202(x-x^0)+.10606(y-y^0)]^2)^{1/2}\\ &\leqq 1.104|X-X^0|.\end{aligned}$$

So

$$|V-V^0|\leqq 1.1286|X-X^0|$$

and

$$\begin{aligned}|(X, V)-(X^0, V^0)|&\leqq (|X-X^0|^2+(1.1286)^2|X-X^0|^2)^{1/2}\\ &\leqq 1.508|X-X^0|.\end{aligned}$$

Combining this with equation (7) gives

(9) $$|W^8-\Psi_{.64}(W)|<1.219\times 10^{-6}+1.914|X-X^0|.$$

Equation (9) implies that, if $W^8\in U(.99, 1{,}01)\times\mathscr{D}(1.01)$ and $|X-X^0|\leqq .01$, then $|W^8-\Psi_{.64}(W)|<.02$ and the straight line from W^8 to $\Psi_{.64}(W)$ is in $U(.97, 1.03)\times\mathscr{D}(1.03)$. Thus we may use the mean value theorem when computing the projection.

To compute the projection of a vector (u, v) onto the unit vector perpendicular to the vector field $(\dot{x}, \dot{y})=F(x, y)$, we use the function

$$\pi(x, y, u, v)\equiv\frac{-u\dot{y}+v\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}.$$

This abuse of the notation in § 3 is done to stress the positional coordinates (x, y). If there are changes $\Delta\dot{x}$, $\Delta\dot{y}$, Δu, and Δv, then the mean value theorem for a ball in $U(.97, 1.03)\times\mathscr{D}(1.03)$ can be used to bound the change in π, i.e.,

$$\Delta\pi\leqq\Delta\dot{x}\sup\frac{\partial\pi}{\partial\dot{x}}+\Delta\dot{y}\sup\frac{\partial\pi}{\partial\dot{y}}+\Delta u\sup\frac{\partial\pi}{\partial u}+\Delta v\sup\frac{\partial\pi}{\partial v}$$

where the suprema of the partials are taken over $U(.97, 1.03)\times\mathscr{D}(1.03)$. Evaluating the suprema gives

$$\Delta\pi\leqq 4.7\max\{\Delta\dot{x}, \Delta\dot{y}\}+2\max\{\Delta u, \Delta v\}.$$

Again using the mean value theorem for a ball in $U(.97, 1.03)\times\mathscr{D}(1.03)$, we get $\Delta\dot{x}=\Delta y$ and $\Delta\dot{y}\leqq 1.163\max\{\Delta x, \Delta y\}$. Thus we have

(10) $$\Delta\pi\leqq 5.467\max\{\Delta x, \Delta y\}+2\max\{\Delta u, \Delta v\}.$$

This projection map π is applied to the numerical solution W^8 producing an error of $|\pi(W^8)-\pi(\Psi_{.64}(W))|$. Since

$$\max\{\Delta u, \Delta v\}\leqq|W^8-\Psi_{.64}(W)|$$

and

$$\max\{\Delta x, \Delta y\}\leqq|X^8-\phi_{.64}(X)|,$$

we use inequalities (8) and (9) in (10) to get

$$\begin{aligned}|\pi(W^8)-\pi(\Psi_{.64}(W))|&<5.464(2.466\times 10^{-7}+1.023|X^0-X|)\\ &\quad+2(1.219\times 10^{-6}+1.914|X^0-X|).\end{aligned}$$

Thus

$$|\pi(W^8)-\pi(\Psi_{.64}(W))|<3.786\times 10^{-6}+9.418|X^0-X|. \tag{11}$$

The computer calculates $|\pi(W^8)|$ which is the length of the projection of V^8 onto the normal subspace at X^8. Since $|V^0|=1$, $\underline{b}\equiv 1-|\pi(W^8)|$ is the decrease in the length of the projection. If $\underline{b}$ is greater than 3.786×10^{-6}, then we know that the projection $|\pi(\Psi_{.64}(W^0))|$ is less than one. In fact, from (11) we know that the projection $|\pi(\Psi_{.64}(W))|$ has decreased for all unit vectors V in the normal subspaces at points in a disk of radius $(\underline{b}-3.786\times 10^{-6})/9.418$ centered at the initial point X^0, i.e., $|X-X^0|$ in (11) is taken to be the minimum of $(\underline{b}-3.786\times 10^{-6})/9.418$ and .01.

A second PL/1 program is employed to numerically approximate the solutions to (6) using a 5th order Taylor Series method with 8 steps of size .08. Each iterate W^i, $i=1,\cdots,8$, is checked to see that it belongs to $U(.988, 1.012)\times\mathscr{D}(1.01)$. Thus Lemma 5.2 applies, establishing the error bound for $\Psi_{.64}(W^0)$ and also the hypothesis of Lemma 5.3. Lemma 5.3 gives equation (7) which establishes the error bound for $\Psi_{.64}(W)$ where W is near W^0. Finally the normal projection of V^8 is computed and compared with the error in equation (11). Thus the normal projection for a neighborhood of W^0 is shown to shrink. Such a neighborhood is called a *valid neighborhood.* If we can show this for all points $(X, V)\in N_X$ where $X\in U(.99, 1.01)$ and $|V|=1$, then we can appeal to Theorem 3.5.

To cover the annulus with valid neighborhoods (actually only the lower half because of symmetry) we fix an angle and cover a radial segment with disks from the inner bounding circle to the outer bounding circle. Then the angle is changed slightly and the new radial segment is covered by disks overlapping the previous disks. We start on the x-axis, i.e., with the angle equal to zero. The initial position is $(x^0, y^0, u^0, v^0)=(.99, 0, 1, 0)$. After eight steps the normal projection has shrunk an amount $\underline{b}$. Thus the radius of the first valid neighborhood is $r=(\underline{b}-3.786\times 10^{-6})/9.418$. To cover the first

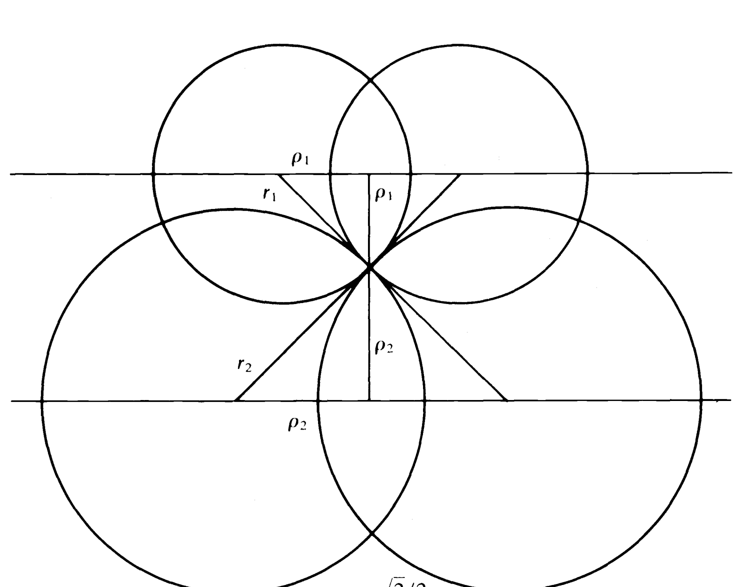

FIG. 1

radial segment we use disks with a constant radius $r_1 = .9r$. Each initial point is chosen a distance of $\sqrt{2}r_1$ from the previous initial point. The radius of the valid neighborhood for each initial point is computed and shown to be greater than r_1. Thus the r_1 disks are also valid neighborhoods. In this way the computer program passes across the annulus covering a radial segment. Nine tenths of the radius of the disk covering the outer boundary component is used as the radius r_2 for the disks of the next pass across the annulus. The angle for the second segment is $(.5\sqrt{2}r_1 + .5\sqrt{2}r_2)/1.01$ radians in the clockwise direction from the previous angle. With this choice of angle and radii the region between the two radial segments is covered. (See Fig. 1.) This procedure continues until half the annulus is covered.

The first PL/1 program, discussed in § 4, used 2 minutes and 14 seconds of computer time for a total cost of \$8. The second program ran for 4 minutes and 22 seconds with a cost of \$15.

6. Observations. This example was chosen to illustrate a method in a relatively simple situation. The fact that (1) is two dimensional not only allows the use of results concerning annuli (Cor. 3.2) but also forces the normal subspace to be one dimensional. Thus the shrinking of only one vector over each point in the annulus must be verified. Also, the normal projection of the orbit of each unit vector in the normal subspace is a nonincreasing function of time t. So it does not take long before shrinking is noticed. Although these simplifications make the computing less expensive, the basic method is applicable to the general n-dimensional situation.

Since a small neighborhood of a periodic orbit is not convex, such regions may need to be studied when finding a periodic orbit. Our discussion illustrates one approach to handling a nonconvex region numerically.

Equation (1) is similar to the Van der Pol equation

$$\dot{x} = y,$$
$$\dot{y} = \alpha(1 - x^2)y - x.$$

Our method can be used to prove the existence of a stable periodic orbit for this equation. However, in the Van der Pol equation the normal projection is increasing for certain ranges of time so it will take longer to notice that all projections have shrunk.

More interesting examples occur in three dimensions. B. Goodwin's [8] model for a cellular control process having negative feedback appears to exhibit limit cycle behavior. J. J. Tyson [15] shows the existence of at least one periodic orbit for certain parameter values. S. Hastings [10] proves this cycle is orbitally asymptotically stable when the repression is of sufficiently high order. We may be able to prove this stability for a wider range of parameter values. Also limit cycle behavior is suspected for the Field and Noyes' equations [4], [11]. This system may also lend itself to our analysis.

Acknowledgment. The authors would like to thank S. Danielopoulos and R. Martin for helpful conversations.

REFERENCES

[1] R. C. Churchill, J. Franke and J. Selgrade, *A geometric criterion for hyperbolicity of flows*, Proc. Amer. Math. Soc., 62 (1977), pp. 137–143.

[2] W. A. Coppel, *Stability and Asymptotic Behavior of Differential Equations*, Heath, Boston, 1965.

[3] J. Dungundji, *Topology*, Allyn and Bacon, Boston, 1966.

[4] R. J. Field and R. M. Noyes, *Ocillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction*, J. Chem. Phys., 60 (1974), pp. 1877–1884.

[5] J. FRANKE AND J. SELGRADE, *Hyperbolicity and chain recurrence*, J. Differential Equations, 26 (1977), pp. 27–36.

[6] ———, *Hyperbolicity and cycles*, Trans. Amer. Math. Soc., to appear.

[7] C. W. GEAR, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[8] B. C. GOODWIN, *Oscillatory behavior in enzymatic control processes*, Advances Enz. Reg., 3 (1965), pp. 425–438.

[9] P. HARTMAN, *Ordinary Differential Equations*, John Wiley, New York, 1964.

[10] S. P. HASTINGS, *On the uniqueness and global asymptotic stability of periodic solutions for a third order system*, Rocky Mountain J. Math., 7 (1977), pp. 513–538.

[11] S. P. HASTINGS AND J. D. MURRAY, *The existence of oscillatory solutions in the Field–Noyes model for the Belousov–Zhabotinskii reaction*, SIAM J. Appl. Math., 28 (1975), pp. 678–688.

[12] M. MARCUS AND H. MINC, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston, 1964.

[13] C. PUGH AND M. SHUB, *The Ω-stability theorem for flows*, Invent. Math., 11 (1970), pp. 150–158.

[14] S. SMALE, *Differentiable dynamical systems*, Bull. Amer. Math. Soc., 73 (1967), pp. 747–817.

[15] J. J. TYSON, *On the existence of oscillatory solutions in negative feedback cellular control processes*, J. Math. Biol., 1 (1975), pp. 311–315.

[16] F. WARNER, *Foundations of Differentiable Manifolds and Lie Groups*, Scott Foresman, Glenview, Il., 1971.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0018 $01.00/0

OPERATIONAL RULES*

RAIMOND A. STRUBLE†

Abstract. The concept of operational rules in the two-sided operational calculus is formalized and extensively illustrated. The paper continues the conceptual development, begun in an earlier one, which is based upon the Fourier transform from an operator ring K onto the arithmetical ring $\mathcal{K}$ of all functions which are measurable and finite-valued almost everywhere on the real line. The Fourier transform can be further extended so that it is a vector space isomorphism from the vector (nondirect) sum $\mathcal{S}=K+\mathcal{K}$ onto itself. This is an extension of the classical L_1 and L_2 Fourier transforms which is also compatible with the distributional Fourier transform. The concept of operational rules is extended to $\mathcal{S}$.

1. Introduction. In [24] (referred to hereafter as Part I), the Fourier transform has been extended to the ultimate setting in which it can be considered numerical-valued: as a ring isomorphism onto the arithmetical ring $\mathcal{K}$ of all measurable and finite-valued almost everywhere functions on the real line, under pointwise addition and multiplication. The "original space" is fabricated as a convolution-type ring K of two-sided operators analogous to those constructed by Mikusiński on the half-line [14]. The ring K contains (isomorphically) the subspaces of all classical L_1 and L_2 functions and all distributions whose Fourier transforms are regular. It thus provides a very suitable setting for development of the classical and distributional two-sided operational calculus. Basic definitions and theorems have been presented in Part I. In this second part, we continue the conceptual development with special emphasis placed upon operational rules. We are now able to formalize this important concept in a meaningful and comprehensive manner because of the immenseness of the function ring $\mathcal{K}$. Also we can now express all operators as functions (not necessarily algebraic) of the differentiation operator and all functions of the differentiation operator as operators. The work suggests, mainly through examples and illustrations, how the Fourier transform can be used to exploit the arithmetic and analysis of ordinary functions. A similar development could be given for functions and operators of several variables.

In § 2, we restate needed definitions from Part I. Some of the terminology and notation used is standard; some of it is not. The relationships between operators and distributions is explained in § 2 and § 3. General operational rules are defined and illustrated in § 4, and functions of operators are treated in § 5. These sections demonstrate the versatility of the numerical-valued Fourier transform and suggest many areas for investigation. In § 6 we extend the Fourier transform (even further) as a vector space isomorphism from the vector sum $\mathcal{S}=K+\mathcal{K}$ onto itself and apply this extended transform to obtain an interesting version of the Mikusiński field of one-sided operators.

The classical theory of Fourier transforms has a long history beginning, appropriately, with Fourier himself [4]. References [1], [16], [28] indicate the extent of the predistributional development. (See [8] for a brief treatment.) The present work (a sketch of the operational calculus on the real line) is perhaps best viewed as an addendum to that classical theory in that it is primarily concerned with numerical-valued Fourier transforms. However, it makes no contribution to hard analysis at all (the central theme of classical Fourier analysis), and is more suggestive of abstract, soft analysis.

* Received by the editors January 21, 1977, and in final revised form December 19, 1977.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607.

2. Notation and preliminary definitions. We refer the reader to [7] and [30] for standard concepts relating to distributions and their Fourier transforms. The notation and terminology we shall employ are as in Part I: R denotes the real line, C the complex plane, and $\mathscr{R}$ the real C-axis; the latter to be distinguished from R. $\mathscr{D}$ denotes the space of infinitely differentiable test functions $\phi(t)$ of the real variable $t \in R$, with compact supports, together with the standard topology given by Schwartz [19]. Z denotes the space of entire functions $\psi(z)$ of the complex variable $z = \omega + i\rho \in C$, which are the Fourier transforms of the elements of $\mathscr{D}$, together with the standard topology for which the Fourier transform from $\mathscr{D}$ onto Z,

$$\phi(t) \mapsto \tilde{\phi}(z) = \int_{-\infty}^{\infty} e^{-izt} \phi(t)\, dt, \tag{1}$$

and the inverse Fourier transform from Z onto $\mathscr{D}$,

$$\tilde{\phi}(z) \mapsto \phi(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \tilde{\phi}(\omega)\, d\omega, \qquad (\omega \in \mathscr{R}) \tag{2}$$

become topological vector space isomorphisms. We recall that Z is the collection of entire functions ψ which satisfy families of inequalities of the form

$$|z^k \psi(z)| \leqq c_k e^{a|\mathrm{Im}\, z|} \qquad (k = 1, 2, \cdots) \tag{3}$$

for positive constants c_k and a depending upon ψ.

The topological duals of these spaces are denoted, as usual, by $\mathscr{D}'$ and Z', and an element $f \in \mathscr{D}'$ is called a distribution, and an element $g \in Z'$ is called an ultradistribution. A distribution f (ultradistribution g) is said to be regular if $f(g)$ is a locally Lebesgue integrable function and $\langle f, \phi \rangle = \int_{-\infty}^{\infty} f(t)\phi(t)\, dt$ for all $\phi \in \mathscr{D}$ ($\langle g, \psi \rangle = \int_{-\infty}^{\infty} g(\omega)\psi(\omega)\, d\omega$ for all $\psi \in Z$). The latter integral is required to be absolutely convergent. (See [27] for some pertinent observations concerning regular ultradistributions.)

If f is a distribution, then $\tilde{f}$ will denote its Fourier transform (as an ultradistribution) which satisfies the Parseval relation

$$2\pi \langle f, \check{\phi} \rangle = \langle \tilde{f}, \tilde{\phi} \rangle \tag{4}$$

for all $\tilde{\phi} \in Z$, where $\check{\phi}(t) = \phi(-t)$. With this definition, the Fourier transform $f \mapsto \tilde{f}$ becomes a vector space topological isomorphism from $\mathscr{D}'$ onto Z' (with respect to their weak topologies, say), and satisfies

$$\widetilde{f * \phi} = \tilde{f}\tilde{\phi} \quad \text{(with } \tilde{\phi} \text{ as a multiplier on } Z'\text{)} \tag{5}$$

for all $f \in \mathscr{D}'$ and $\phi \in \mathscr{D}$, where $*$ denotes, as usual, convolution. Moreover, $\mathscr{D}$ is considered a subspace of $\mathscr{D}'$, and the Fourier transform of a function $\phi \in \mathscr{D}$, given by (1) is the same as the Fourier transform of the distribution $\phi = f \in \mathscr{D}'$, given by (4). The delta function will be denoted, as usual, by $\delta(t)$. Its Fourier transform is the constant function 1.

The following definitions and propositions are not standard, but have been given earlier in Part I, where additional results and discussion may be found.

DEFINITION 1. The collection of all (Lebesgue) measurable functions $\tilde{x}$ which are finite a.e. (almost everywhere) on $\mathscr{R}$ is denoted by $\mathscr{K}$, and will be considered a ring under pointwise addition and multiplication of functions a.e., $\tilde{x}$ is *unit* (is invertible) in $\mathscr{K}$ iff $\tilde{x}$ is nonzero a.e. on $\mathscr{R}$.

PROPOSITION 1. *Let $\tilde{x} \in \mathcal{K}$. Then there exists a unit $\tilde{y} \in \mathcal{K}$ such that both $\tilde{y}$ and the product $\tilde{x}\tilde{y}$ are bounded functions.*

Proof. Let $\tilde{y}(\omega) = 1/(1+|\tilde{x}(\omega)|)$.

Each of $\tilde{y}$ and $\tilde{x}\tilde{y}$ in this proposition is a regular ultradistribution and so $\tilde{x}$ can be expressed as the ratio $\tilde{x}\tilde{y}/\tilde{y}$ of two such ultradistributions. This observation we reformulate in the following proposition.

PROPOSITION 2. *If $\tilde{x} \in \mathcal{K}$, then $\tilde{x}$ may be expressed in the form $\tilde{x} = \tilde{w}/\tilde{y}$ with $\tilde{w} \in \mathcal{K}$ and $\tilde{y} \in \mathcal{K}$ both regular ultradistributions which are bounded. If $\tilde{v}$ is a regular ultradistribution, then $\tilde{v} \in \mathcal{K}$.*

By considering the inverse Fourier transforms of regular ultradistributions we can (using Proposition 2) form an isomorphic ring of fractions of the corresponding distributions.

DEFINITION 2. Let K denote the collection of all (formal) fractions g/f of distributions with $\tilde{g} \in \mathcal{K}$ and $\tilde{f} \in \mathcal{K}$ both regular ultradistributions which are bounded and $\tilde{f}$ a unit of $\mathcal{K}$. Such fractions are identified, added and multiplied, just as ordinary numerical fractions are, with the operations corresponding to addition for distributions and (extended) convolution for distributions defined by $f_1 * f_2 = f_3$ iff $\tilde{f}_3 = \tilde{f}_1\tilde{f}_2$. The mapping which sends the fraction $x = g/f \in K$ to the function $\tilde{x} = \tilde{g}/\tilde{f} \in \mathcal{K}$ is called the *Fourier transform*, and the elements of the ring K are called *operators*.

It turns out that distributions are not actually needed at all, just ordinary functions together with the classical (inverse) Fourier transform. This is because a function $\tilde{x}$ in $\mathcal{K}$ can also be expressed as the ratio of two functions each of which is abolutely integrable and has a classical Fourier inverse. (See [26].)

PROPOSITION 3. *The ring K of operators is isomorphic with a convolution ring of fractions of continuous functions on R, each of which is the classical inverse Fourier transform of a function on $\mathcal{R}$ which is absolutely integrable.*

Because of this last proposition it becomes important to further clarify the relationships between operators, distributions, and functions. This we undertake in the following section.

3. Identification and representation of distributions. It is suggested by Proposition 2 and Definition 2 that we identify distributions with operators having the same Fourier transforms. Thus whenever

$$\tilde{\phi} \mapsto \int_{-\infty}^{\infty} \tilde{f}(\omega)\tilde{\phi}(\omega)\, d\omega \quad \text{(absolute convergence)} \tag{6}$$

defines a continuous mapping from Z into C for some locally integrable function $\tilde{f}$, the corresponding distribution f, defined by $2\pi\langle f, \check{\phi}\rangle = \int_{-\infty}^{\infty} \tilde{f}(\omega)\tilde{\phi}(\omega)\, d\omega$, is identified with the operator x for which $\tilde{x} = \tilde{f}$. In this way the operator ring K is seen to contain (isomorphically) the subspace N of all distributions whose Fourier transforms are regular. This is a very large class of distributions and includes, of course, all distributions with compact support. Whenever a tempered [19] distribution in N is also regular, we shall say that the corresponding operator is a "function". Thus in particular, K contains the subspaces of all classical L_1 and L_2 functions wherein extended convolution becomes ordinary convolution (because of the classical convolution theorems).

Suppose now that $f \in N$, i.e. $f \in \mathcal{D}'$ and $\tilde{f}$ is regular. Then for each $\phi \in \mathcal{D}$, the distributional Fourier transform of the convolution $f * \phi$ is the product $\tilde{f}\tilde{\phi}$ of the distributional Fourier transform of f and the classical Fourier transform of ϕ. Normally, $\tilde{\phi}$ is only considered a multiplier in Z', but in this case $\tilde{f}$ is regular, and so this product is just that in the function ring $\mathscr{K}$. Thus the above identification of distributions and operators preserves all convolution properties in $\mathscr{D}'$.

On the other hand, there are distributions f with distributional Fourier transforms $\tilde{f}$ which are not regular, but for which the product $\tilde{f}\tilde{\phi}$ with some (nonzero) function $\tilde{\phi} \in Z$ becomes regular. For example, the distributional Fourier transform [30, pp. 190] of the Heaviside step function $H(t) = (1 + \operatorname{sign} t)/2$ is $pv\,1/(i\omega) + \pi\delta(\omega)$, and does not belong to $\mathscr{K}$, while the product of this singular ultradistribution with any $\tilde{\phi} \in Z$ for which $\tilde{\phi}(0) = \int_{-\infty}^{\infty} \phi(t)\,dt = 0$ is the regular ultradistribution $\tilde{\phi}(\omega)/(i\omega)$, which does belong to $\mathscr{K}$. Whenever this happens, we say that the distribution f is *represented* by the operator x for which $\tilde{x} = \tilde{f}\tilde{\phi}/\tilde{\phi}$. Thus $H(t)$ is represented by the operator $x = 1/s$, for which $\tilde{x}(\omega) = 1/(i\omega)$, and we say that $1/(i\omega)$ is the Fourier transform of $H(t)$ *in* $\mathscr{K}$. Similarly, $t^n H(t)$ is represented by the operator $n!/s^n$, whose Fourier transform is $(-i)^{n+1} n!/\omega^{n+1}$. According to this definition, different distributions can be represented by a single operator (for example $\frac{1}{2}$ sign t is represented by the same operator as is $H(t)$), but only one operator can represent any given distribution. This concept of representation is consistent with (and only a slight aberration of) that treated in several earlier papers. Operators which represent distributions are called neocontinuous, inasmuch as they inherit certain continuity properties from distributions. (For details see [12], [21], [23], [25].) The analysis of neocontinuous operators has not been vigorously pursued.

4. General operational rules. The operational calculus is initially concerned with the resolution of analytical problems by applying algebraic (and/or other simple) procedures in a suitable transformed space. For one-sided problems, i.e. problems relating to functions, distributions, operators, etc. on half-lines or half-spaces, the one-sided Laplace transformation is often used to convert analytical initial boundary-value problems into algebraic or functional equations in the space of holomorphic (analytic) functions in complex half-planes. For two-sided problems, the two-sided Laplace transformation [29] is often used to convert analytical problems into similar equations in spaces of holomorphic functions in vertical complex strips. In either of these situations, manipulative (and analytical) procedures in the transformed spaces are interpreted in the original spaces as operational rules, and it is always hoped (and sometimes shown) that inversion back to the original spaces is ultimately possible. The latter can be a serious stumbling block in the method and restricts applications to those particular instances where the transforms of the solutions are known (or assumed) to exist. Mikusiński's operator field may be viewed as a device for partially circumventing this last problem in the one-sided operational calculus. However, the simplicity stemming from working with ordinary functions is largely lost, and a new type of analysis for operators (sometimes as intractible as the original) must be substituted. Meaningful operational rules, on the other hand, are severely restricted due to the necessity of always working backward through convolution to obtain an interpretation.

One may similarly view the present extension of the numerical-valued Fourier transform as a device for fulfilling (historical) expectations of the two-sided operational calculus in the exploitation of the arithmetic and analysis of ordinary functions. Thus *any* procedure (whatsoever) leading from and to (nonpathological) functions on the real line $\mathscr{R}$ can be given an operator interpretation and can be viewed as an operational rule. These procedures could be the solving of linear or nonlinear integral equations, differential equations, functional equations, difference equations (for step functions), or the application of limiting processes or integral transforms, etc., in short, nearly any procedures of classical analysis of ordinary functions. We shall, therefore, formally *define* operational rules and then proceed to illustrate them with some familiar, and not so familiar, examples.

DEFINITION 3. Let $M: x \mapsto y$ be a mapping from a (nonempty) subset J of the operator ring K into K. Then the corresponding mapping $\tilde{M}: \tilde{x} \mapsto \tilde{y}$ from the subset $\tilde{J} = \{\tilde{x} | x \in J\}$ of the function ring $\mathcal{K}$ into $\mathcal{K}$ is called the *Fourier transform* of M. Similarly let $\tilde{M}: \tilde{x} \mapsto \tilde{y}$ be a mapping from a (nonempty) subset $\tilde{J}$ of $\mathcal{K}$ into $\mathcal{K}$. Then the corresponding mapping $M: x \to y$ from $J = \{x | \tilde{x} \in \tilde{J}\}$ into K is called the *inverse Fourier transform* of $\tilde{M}$. A pair of mappings $(M, \tilde{M})$ so related is called an **operational rule**.

We reserve the square bracket notation $y = M[x]$ and $\tilde{y} = \tilde{M}[\tilde{x}]$ to depict such mappings, and Definition 3 can then be re-expressed simply through the equation

$$\tilde{M}[\tilde{x}] = \widetilde{M[x]} \quad \text{for all } \tilde{x} \in \tilde{J} \quad (\text{or } x \in J). \tag{7}$$

As is customary, we shall call the sets J and $\tilde{J}$ the *domains* of the mappings, and we shall at times consider mappings which are either restrictions or extensions (relative to their domains) of given ones as the same mappings.

We have already encountered some operational rules in Part I. The exponential shifts $e^{-i\alpha t}$, the dilatations U_n and the algebraic derivative D are familiar mappings of operators to operators. Their Fourier transforms are the mappings from functions to functions given by $\tilde{x}(\omega) \mapsto \tilde{x}(\omega + \alpha)$, $\tilde{x}(\omega) \mapsto \tilde{x}(\omega/n)$ and $\tilde{x}(\omega) \mapsto d\tilde{x}/d\omega$, respectively. The domains of the first two kinds are all of $\mathcal{K}$, while the domain of the last includes all continuously differentiable functions. (See Part I.) The dilatations U_n illustrate the class of operational rules based upon "changes of variables" in the operator and function rings. The exponential shifts illustrate operational rules based upon "changes of variables" in the function ring (transformed space) and, at the same time, "multipliers" in the operator ring (untransformed space). The algebraic derivative illustrates operational rules based upon "multipliers" in the operator ring and somewhat more sophisticated procedures, like differentiation, in the function ring. Even more sophisticated are the procedures of (classical) fractional differentiation (and integration) [17] in the function ring. These procedures (mappings) lead to operational rules symbolized by fractional powers D^λ of the algebraic derivative in the operator ring. For many interesting and useful operational rules of the various types mentioned above, see [29].

Even more familiar (and fundamental) are the operational rules which say that addition and convolution in the operator ring correspond to addition and multiplication in the function ring Thus for example, the convolution mapping $x \mapsto y * x$ (for a fixed y and all $x \in K$) has for its Fourier transform the multiplication mapping $\tilde{x} \mapsto \tilde{y}\tilde{x}$. In particular, the differentiation operator s, whose Fourier transform is $i\omega$, leads to the familiar operational rule $x \mapsto s * x = x^{(1)}$ iff $\tilde{x}(\omega) \mapsto i\omega\tilde{x}(\omega)$.

As a less familiar (and yet well-known) example, let us consider a measurable function $\tilde{\gamma}(\omega, \xi)$ of two real variables which, for each fixed ω, is bounded and for which $\sup\{\tilde{\gamma}(\omega, \xi)| -\infty < \xi < \infty\}$ as a function of ω is absolutely integrable. Then for each $\tilde{\phi} \in L_1$, the function $\tilde{\psi}$ defined by

$$\tilde{\psi}(\omega) = \int_{-\infty}^{\infty} \tilde{\gamma}(\omega - \xi, \xi)\tilde{\phi}(\xi)\, d\xi \tag{8}$$

is also in L_1, and the mapping $\tilde{\Gamma}: \tilde{\phi} \mapsto \tilde{\psi}$ (defined at least on L_1) is the Fourier transform of the mapping Γ: $\phi \mapsto \psi$, where

$$\begin{aligned}\psi(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\tilde{\psi}(\omega)\,d\omega = \int_{-\infty}^{\infty} e^{i\xi t}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(\omega-\xi)t}\tilde{\gamma}(\omega-\xi,\xi)\,d\omega\right)\tilde{\phi}(\xi)\,d\xi \\ &= \int_{-\infty}^{\infty} e^{i\xi t}\gamma(t,\xi)\tilde{\phi}(\xi)\,d\xi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\xi(t-\tau)}\gamma(t,\xi)\phi(\tau)\,d\xi\,d\tau\end{aligned} \tag{9}$$

and $\widetilde{\gamma(t,\xi)}(\omega) = \tilde{\gamma}(\omega,\xi)$ for each fixed ξ. In this case, the mapping Γ is called a pseudo-differential operator [6], while $\gamma(t,\xi)$ is called its symbol and $\tilde{\gamma}(\omega,\xi)$ its cokernel. The pair $(\Gamma, \tilde{\Gamma})$ defined by (8) and (9) may be viewed as an operational rule (7). Pseudo-differential operators arise rather naturally when Fourier transform techniques, which are readily applicable to differential equations with constant coefficients, are applied in the variable coefficient situations. A number of different settings for these operators have been proposed. (See [3], [10], [13].)

Equation (8) is a modified form of convolution in the function ring $\mathscr{K}$ and (9) is a modified form of pointwise product in the operator ring K. This is a reversal of procedures used in the original construction of the two rings and the Fourier transform between them. It is of some interest to consider this reversal of procedures more fully. The classical (backward) convolution theorem suggests the following.

DEFINITION 4. Let $\tilde{y} \in \mathscr{K}$ be nonzero, and assume that for some nonzero $\tilde{x} \in \mathscr{K}$ the ordinary convolution

$$\frac{1}{2\pi}\tilde{y} * \tilde{x}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{y}(\omega-\xi)\tilde{x}(\xi)\,d\xi = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{y}(\xi)\tilde{x}(\omega-\xi)\,d\xi \tag{10}$$

exists for a.e. $\omega \in \mathscr{R}$. The mappings $\tilde{x} \mapsto (1/(2\pi))\tilde{y} * \tilde{x} = \tilde{w}$, for all such $\tilde{x}$ in $\mathscr{K}$, is called the *convolution by* $\tilde{y}$. The corresponding mapping $x \mapsto w = y \cdot x$ in K is called *pointwise multiplication by* y.

As a ready example of (10), consider the step function $\frac{1}{2}\operatorname{sign} t$ and its Fourier transform in $\mathscr{K}$, $1/(i\omega)$. The operator pointwise product $w(t) = (\frac{1}{2}\operatorname{sign} t)\cdot x(t)$ has for its Fourier transform the function convolution

$$\tilde{w}(\omega) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\tilde{x}(\xi)\,d\xi}{\omega-\xi},$$

sometimes called the Hilbert transform [15] of $\tilde{x}$. For certain $\tilde{x}$ one should consider this is a principal value integral, since quite generally it is not an absolutely convergent Lebesgue integral. For functions (or distributions) x in K, of course, $(\frac{1}{2}\operatorname{sign} t)\cdot x(t)$ equals $\frac{1}{2}x(t)$ for $t > 0$ and equals $-\frac{1}{2}x(t)$ for $t < 0$. The distributional Fourier transform of the step function $\frac{1}{2}\operatorname{sign} t$ is $pv1/(i\omega)$.

The collection of operational rules defined by (10) includes the classical result which says that the Fourier transform of the pointwise product of two L_1-functions is the convolution (with factor $1/(2\pi)$) of the Fourier transforms of the factors, provided these latter are also in L_1. Moreover, this operational rule includes the distributional multiplier concept (when applicable here) which says that (for example) $f\phi$ is a distribution with compact support, whenever $f \in \mathscr{D}'$ and $\phi \in \mathscr{D}$, and that $\widetilde{f\phi} = (1/(2\pi))\tilde{f} * \tilde{\phi} = (1/(2\pi))\langle \tilde{f}(\xi), \tilde{\phi}(\omega-\xi)\rangle$. Indeed the latter reduces to (10) whenever $\tilde{f}$ is regular. On the other hand, for *every* ultradistribution $\tilde{f}$, the convolution $(1/(2\pi))\tilde{f} * \tilde{\phi}$ belongs to $\mathscr{K}$ for every $\tilde{\phi} \in Z$. Hence for every distribution f in $\mathscr{D}'$ and every test function ϕ in $\mathscr{D}$, the product $f\phi$ is an operator in K.

Definition 4 allows for an interpretation of the substitution of certain operators into holomorphic functions which vanish at zero. For example, since $\log(1+t)=-\sum_{n\geq 1}(-1)^n t^n/n$ ($|t|<1$), $y=\log(1+x)$ can reasonably be defined as the operator whose Fourier transform is given by

$$\tilde{y}=-2\pi\sum_{n\geq 1}\left(\frac{-1}{2\pi}\right)^n[\tilde{x}]^n/n, \quad \text{where } [\tilde{x}]^n=\overbrace{\tilde{x}*\tilde{x}*\cdots*\tilde{x}}^{n\text{ factors}},$$

whenever this last series of convolutions converges, say pointwise a.e. on $\mathcal{R}$.

A special subring (actually a field under addition and ordinary convolution) of two-sided operators of K, called exponential operators, has been studied in [23] and [25]. These consist of convolution quotients formed from the convolution ring $\mathscr{E}xp$ consisting of C^∞-functions on R which, together with all their derivatives, decay exponentially as $|t|\to\infty$. (Hence the name for such operators.) Their Fourier transforms are meromorphic in various strip-type neighborhoods of the real axis $\mathcal{R}$ of C, and every such function which is the ratio of two bounded holomorphic functions in such neighborhoods is the Fourier transform of some exponential operator. This (sub)field $\mathcal{M}_{\mathscr{E}xp}$ of exponential operators, therefore, is isomorphic to a field of meromorphic functions (in neighborhoods of $\mathcal{R}$). Certain operator transformations related to these exponential operators have been considered in [25], together with their Fourier transforms. The latter have been viewed as (continuous linear) mappings from Z into $Z_{\mathscr{E}xp}$, where $Z_{\mathscr{E}xp}$ is the ring of analytic functions consisting of the Fourier transforms of the exponentially decaying functions in $\mathscr{E}xp$. Each such pair obtained from an operator transformation and its Fourier transform yields an operational rule (7). (See [25] for some specific examples and additional properties.)

It seems appropriate to call an operator x Laplace transformable [16] if its Fourier transform $\tilde{x}$ is meromorphic in a strip-type neighborhood $N_b=\{z \mid |\operatorname{Im} z|<b\}$ of $\mathcal{R}$. In such a case, we can define the *Laplace transform* of x (as is customary) to be the function of the rotated variable p,

$$\mathscr{L}\{x\}\colon p\mapsto\tilde{x}(p/i), \quad \text{where } p=iz=i(\omega+i\rho)$$

which is meromorphic for $|\operatorname{Re} p|<b$. For such operators a *real* exponential shift (multiplier) $x(t)\mapsto e^{-\sigma t}x(t)$ (σ real with $|\sigma|<b$) can be interpreted as the inverse Fourier transform of the mapping $\tilde{x}(\omega)\mapsto\tilde{x}(\omega-i\sigma)$. The latter corresponds to the mapping $\mathscr{L}\{x\}(p)\mapsto\mathscr{L}\{x\}(p+\sigma)$ of the Laplace transforms. In particular, all exponential operators are Laplace transformable and admit to suitable real exponential shifts, in addition to all pure imaginary ones. The collection of Laplace transformable operators forms a field in the ring K containing the subfield $\mathcal{M}_{\mathscr{E}xp}$ of exponential operators.

An important (and unlimited) source of operational rules stems from considering substitution or composition in the function ring $\mathscr{K}$, that is, whenever $\tilde{x}\to\tilde{M}[\tilde{x}]$ constitutes substitution of $\tilde{x}$ into another (fixed) function or the substitution of another (fixed) function into $\tilde{x}$. We consider these special operational rules in the next section.

5. Functions of operators and operators as operational rules. One of the more interesting aspects of the operational calculus concerns the concept of "functions of operators", such as power series of the differentiation operator s, and their meanings. It is clear what meaning should be given to a polynomial function of an operator. Thus if $P(z)=a_0+a_1z+a_2z^2+\cdots+a_nz^n$ is a polynomial with complex coefficients, and if x is an operator, then $y=P[x]$ should be the operator formed by substituting x for z

and interpreting multiplication as convolution, i.e.

$$y = P[x] = a_0 + a_1 x + a_2 x * x + \cdots + a_n \overbrace{x * x * \cdots * x}^{n \text{ factors}}. \tag{11}$$

For $x = s/i$, this gives $P[s/i] = a_0 + (a_1/i)s + (a_2/i^2)s^2 + \cdots + (a_n/i^n)s^n$, where s^j (equivalently $\delta^{(j)}(t)$) is just a notation for the jth order differentiation operator. The Fourier transform of $P[s/i]$ is the polynomial $\tilde{p}(\omega) = a_0 + a_1\omega + a_2\omega^2 + \cdots + a_n\omega^n$. Hence $P[s/i]$ is the operator whose Fourier transform is $\tilde{p}(\omega)$. In the general case of (11), the Fourier transform is given by

$$\tilde{y} = \tilde{P}[\tilde{x}] = a_0 + a_1\tilde{x} + a_2\tilde{x}^2 + \cdots + a_n\tilde{x}^n, \tag{12}$$

where the right-hand side is precisely the arithmetical function obtained by substituting the function $\tilde{x}$ for z in the polynomial $P(z)$. Thus $\tilde{P}$ and P are, in a symbolic sense, the "same" function. Alternatively, the polynomial function $\tilde{p}(\omega)$ may be regarded as an element of $\mathcal{K}$, so that the mapping $\tilde{x} \mapsto \tilde{P}[\tilde{x}]$, given by (12), comes from composition in $\mathcal{K}$ with the fixed element $\tilde{p}$. The corresponding mapping $x \mapsto P[x]$, given by (11) is then viewed as induced composition in K with the fixed element p. Let us, therefore, formally distinguish the operational rules which stem from composition in $\mathcal{K}$. For this purpose, we should consider (without loss of generality since we identify functions which agree a.e.) the functions in $\mathcal{K}$ to be Borel measurable [9]. Thus if $\tilde{y}, \tilde{x} \in \mathcal{K}$, we shall say that the *composition* $\tilde{y} \circ \tilde{x}$ *belongs to* $\mathcal{K}$ (i.e. $\tilde{y} \circ \tilde{x} \in \mathcal{K}$) if for any finite-valued (Borel) measurable functions $\tilde{y}_1, \tilde{y}_2$, equivalent (equal a.e. on $\mathcal{R}$) to $\tilde{y}$ and any measurable functions $\tilde{x}_1, \tilde{x}_2$, equivalent to $\tilde{x}$, the mappings $\omega \mapsto \tilde{y}_1(\tilde{x}_1(\omega))$, $\omega \mapsto \tilde{y}_2(\tilde{x}_2(\omega))$ are defined for a.e. $\omega \in \mathcal{R}$ (they are automatically measurable and finite-valued) and are equivalent. In this way composition becomes well-defined in $\mathcal{K}$. If $\tilde{y} \circ \tilde{x} \in \mathcal{K}$, then $\tilde{x}$ must be real-valued a.e. on $\mathcal{R}$ and the measure of any set $N = \{\omega \in \mathcal{R} | \tilde{x}(\omega) \in M\}$ must be zero whenever the measure of M is zero. Indeed, any changes in the values of the function $\tilde{y}$ over a set M of measure zero, must not affect the composition except on a set N of measure zero. Thus a substituted function $\tilde{x}$ must not be too repetitive.

DEFINITION 5. Let $\tilde{y} \in \mathcal{K}$, and let $\tilde{J} = \{\tilde{x} \in \mathcal{K} | \tilde{y} \circ \tilde{x} \in \mathcal{K}\}$. (The set $\tilde{J}$ is nonempty since it contains the identity function.) The mapping $\tilde{Y}: \tilde{x} \mapsto \tilde{Y}[\tilde{x}] = \tilde{y} \circ \tilde{x} = \tilde{w}$ defined on $\tilde{J}$ is called *composition in* $\mathcal{K}$ *with the function* $\tilde{y}$. The inverse Fourier transform $Y: x \mapsto w = Y[x]$, where $\widetilde{Y[x]} = \tilde{Y}[\tilde{x}] = \tilde{y} \circ \tilde{x}$, is called *induced composition in* K *with the operator* y. The operational rule given by the pair of mappings $(Y, \tilde{Y})$ is called *composition with the function* $\tilde{y}$.

According to this definition, to each operator y there corresponds a unique operational rule $(Y, \tilde{Y})$ in which the mapping $\tilde{Y}$ is ordinary composition in $\mathcal{K}$ with the Fourier transform $\tilde{y}$ of the operator y. Thus each operator can be thought of as a special type of operational rule. More precisely, we can identify each operator y with the corresponding induced composition Y in K and its Fourier transform $\tilde{y}$ with the corresponding composition $\tilde{Y}$ in $\mathcal{K}$. In this way, Definition 3 for the Fourier transforms of mappings becomes an extension of that for operators. The induced composition mapping Y in K with the operator y has a rather interesting property (observed above for polynomials); it always maps the operator s/i to y itself, i.e. $Y[s/i] = y$, because $\widetilde{s/i} = \omega$ and so $\widetilde{Y[s/i]} = \tilde{y}(\omega)$. Hence each operator y is automatically a function $Y[s/i]$ of the differentiation operator s and each function of s is an operator; polynomials are but the simplest kinds of such functions.

The next simplest kinds are infinite power series which converge in the entire complex plane, i.e. entire functions. Composition in $\mathscr{K}$ with respect to such functions can be considered to have "holomorphic extensions" to all of $\mathscr{K}$. For example, the ordinary sine function $z \mapsto \sin z = z - z^3/(3!) + \cdots$ gives rise to the operator function $x \mapsto y = \operatorname{Sin}[x] = x - (x * x * x)/3! + \cdots$, where $\tilde{y}(\omega) = \sin(\tilde{x}(\omega))$ is a bounded function. Thus $y = \operatorname{Sin}[x]$ is always a tempered distribution, and it is unnecessary to deal directly with the infinite series. The classical result here says that $\operatorname{Sin}[x]$ is an L_1 function whenever x is [8]. On the other hand, this infinite series of convolutions will actually converge, say pointwise, for many operators x which are ordinary functions (certainly if x is continuous and has compact support and for which $\operatorname{Sin}[x](t) = (1/(2\pi)) \int_{-\infty}^{\infty} e^{i\omega t} \sin(\tilde{x}(\omega))\, d\omega$). In general, the specification of $y = \operatorname{Sin}[x]$ through its Fourier transform $\tilde{y}(\omega) = \sin(\tilde{x}(\omega))$ could be used as a method of *defining* the convergence of the infinite series of convolutions, i.e. the operational rule could be taken as a definition for a generalized type of convergence. For example, one might say that an infinite series of convolutions converges in K whenever the transformed series converges pointwise a.e. on $\mathscr{R}$. Such a definition might also be used for the convergence of any series (not necessarily convolution ones) of operators. Note that as with polynomials, it seems appropriate here to consider the sine function (and indeed every entire function) symbolically as the same function in both the original and the transformed spaces. (This is true also of any rational function, i.e. a ratio of two polynomials.)

Meromorphic functions are, perhaps, the next simplest type of functions to consider. For example, let $\tilde{f}(\omega) = \omega/(1+\omega^2)^2$, which is the classical Fourier transform of the function (operator) $f(t) = it\, e^{-|t|}/4$. Now composition in $\mathscr{K}$ with the function $\tilde{f}$ is the mapping $\tilde{F}\colon \tilde{x}(\omega) \mapsto \tilde{y}(\omega) = \tilde{f} \circ \tilde{x}(\omega) = \tilde{x}(\omega)/(1+(\tilde{x}(\omega))^2)^2$, and can be "analytically extended" to all those $\tilde{x} \in \mathscr{K}$ with $\tilde{x}(\omega) \neq \pm i$ for a.e. $\omega \in \mathscr{R}$. The inverse Fourier transform of this mapping is the mapping $F\colon x \mapsto y = F[x]$, and (for suitable x) can be identified further using ordinary power series techniques. Indeed let $\tilde{x}$ be the special meromorphic function $\tilde{g}(\omega) = 1/(2(1+\omega^2))$, which is the classical Fourier transform of $g(t) = e^{-|t|}/4$. The power series expansion about the origin of $\tilde{f}$ is given by $\sum_{n \geqq 0} (n+1)(-1)^n \omega^{2n+1}$ and is convergent for $|\omega| < 1$. But $|\tilde{g}(\omega)| < 1$, so that

$$\begin{aligned} \tilde{y}(\omega) = \tilde{F}[\tilde{g}](\omega) = \tilde{f} \circ \tilde{g}(\omega) &= \sum_{n \geqq 0} (n+1)(-1)^n (\tilde{g}(\omega))^{2n+1} \\ &= 8(1+\omega^2)^3/(1+4(1+\omega^2)^2)^2 \end{aligned}$$

holds for all (real) ω. Thus $y(t) = F[g](t) = \sum_{n \geqq 0} (n+1)(-1)^n [g(t)]^{2n+1}$ holds for all t, where $[g(t)]^n = \overbrace{g * g * \cdots * g}^{n \text{ factors}}(t)$. This last series of convolutions is uniformly convergent on compact subsets of R, and sums to

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega t}\, 8(1+\omega^2)^3}{(1+4(1+\omega^2)^2)^2}\, d\omega,$$

the classical Fourier inverse of the composition $\tilde{y} = \tilde{f} \circ \tilde{g}$. The operational rule $(F, \tilde{F})$, composition with the function $\tilde{f}$, is thus "identified" by the power series of $\tilde{f}$ about the origin. Of course, one could identify it also by the power series of $\tilde{f}$ about any other point of $\mathscr{R}$. Here one seems to need a concept of "induced analytic continuation" of operator functions.

As a nonholomorphic example, we might consider the simple step function $\omega \mapsto -i\pi \operatorname{sign} \omega$ in $\mathcal{K}$, which, through induced composition, gives rise to the operator function $x \mapsto y = Pv^{-1}[x]$, where $\tilde{y}(\omega) = -i\pi \operatorname{sign}(\tilde{x}(\omega))$. We have chosen to label this particular operator function as Pv^{-1} since the Fourier inverse of $-i\pi \operatorname{sign} \omega$ is the principal value function (distribution) $pv\, 1/t$. In this example, $\tilde{y}(\omega)$ is again a bounded function, so that y is always a tempered distribution. In particular, if $x(t) = e^{-|t|}$, then $\tilde{x}(\omega) = 2/(1+\omega^2) > 0$, and so $-i\pi \operatorname{sign}(\tilde{x}(\omega)) = -i\pi$ holds for all ω. Hence $Pv^{-1}[e^{-|t|}] = -i\pi\delta(t)$.

Of course, almost *any* complex-valued function (not necessarily a member of $\mathcal{K}$) which is defined on *any* nonempty subset of complex numbers gives rise, through induced substitution, to an operator function. The above examples furnish but a glimpse of the possibilities. Moreover, one obtains many other operator functions by substituting real-valued functions $\tilde{v} \in \mathcal{K}$ *into* the Fourier transforms $\tilde{x}$, i.e. $\tilde{x}(\omega) \mapsto \tilde{x}(\tilde{v}(\omega))$. These operator functions may be thought of as originating from "changes of variables" though the variable changes need not be one-to-one nor onto. Other classical transform pairs arise in this manner.

DEFINITION 6. Let $\tilde{v} \in \mathcal{K}$ be real-valued a.e. on $\mathcal{R}$ and satisfy the condition that the measure of the set $\{\omega \in \mathcal{R} | \tilde{v}(\omega) \in M\}$ is zero whenever the measure of M is zero. The mapping $\tilde{x} \mapsto \tilde{x} \circ \tilde{v} = \tilde{y}$, defined for all $\tilde{x} \in \mathcal{K}$, is called the *change of variable in $\mathcal{K}$ by $\tilde{v}$*. Its inverse Fourier transform $x \mapsto y$, defined for all $x \in K$, is called the *induced change of variable in K by v*. The associated operational pair of mappings (they are ring homomorphisms) is called the *change of variable by $\tilde{v}$*.

Like induced composition, the induced change of variable in K by v maps the operator s/i to v itself. If $\tilde{y}(\omega) = \tilde{x}(\tilde{v}(\omega))$ is in L_1, then $y(t) = (1/(2\pi)) \int_{-\infty}^{\infty} e^{i\omega t} \tilde{x}(\tilde{v}(\omega))\, d\omega$, and for suitable $\tilde{x}$ and $\tilde{v}$ with $\xi = \tilde{v}(\omega)$ possessing the inverse function $\tilde{u}(\xi) = \omega$,

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\tilde{u}(\xi)t} \tilde{x}(\xi) \tilde{u}'(\xi)\, d\xi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -i\tau\, e^{i(\tilde{u}(\xi)t-\tau)} x * u(\tau)\, d\tau\, d\xi.$$

In this last form, the mapping $x \mapsto y$ is sometimes called a Fourier integral operator [11].

As an elementary (non-L_1) example, consider the finite part function (distribution [30, p. 16]) $fp\, 1/(\pi t^2)$ whose Fourier transform is $|\omega|$. The induced change of variable in K by $fp\, 1/(\pi t^2)$ is, therefore, the mapping $x \mapsto y$ where $\tilde{y}(\omega) = \tilde{x}(|\omega|)$, and satisfies

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \tilde{x}(|\omega|)\, d\omega = \frac{1}{\pi} \int_{0}^{\infty} \tilde{x}(\omega) \cos \omega t\, d\omega$$

for suitable x. This last is sometimes called the Fourier cosine transform [28] of $\tilde{x}$. On the other hand, the induced composition in K with the operator $fp\, 1/(\pi\tau^2)$ (Def. 5) is the mapping $x \mapsto Fp(1^{-2}/\pi)\ [x] = w$, where $\tilde{w}(\omega) = |\tilde{x}(\omega)|$. Thus, $Fp(1^{-2}/\pi)[x](t) = (1/(2\pi)) \int_{-\infty}^{\infty} e^{i\omega t} |\tilde{x}(\omega)|\, d\omega$, for suitable x.

The special meromorphic functions $\tilde{f}$ and $\tilde{g}$ used earlier provide for an example of an operational rule based on a change of variable. We shall consider the change of variable in $\mathcal{K}$ by $\tilde{f}$. In this case, we would be interested in the composition $\tilde{h} = \tilde{g} \circ \tilde{f}$, where $\tilde{h}(\omega) = \tilde{g}(\tilde{f}(\omega)) = (1+\omega^2)^4/(2(\omega^2 + (1+\omega^2)^4))$. In order to use the power series technique here we need to subtract the limiting value $\frac{1}{2}$ of $\tilde{h}$ at $\pm\infty$ and invert the constant function $\frac{1}{2}$ separately. We obtain $\tilde{g}(\omega) = \frac{1}{2} \sum_{n \geq 0} (-1)^n \omega^{2n}$ for $|\omega| <$

1, and since $|f(\omega)|<1$, it follows that

$$\tilde{h}(\omega)=\frac{1}{2}\sum_{n\geqq 0}(-1)^n(\tilde{f}(\omega))^{2n}=\frac{1}{2}+\frac{1}{2}\sum_{n\geqq 1}(-1)^n(\tilde{f}(\omega))^{2n}$$

holds for all ω. Then $h(t)=\delta(t)/2+\frac{1}{2}\sum_{n\geqq 1}(-1)^n[f(t)]^{2n}$, where the infinite series of convolutions sums to

$$h(t)-\frac{\delta(t)}{2}=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\omega t}\{\tilde{g}(\tilde{f}(\omega))-\tfrac{1}{2}\}\,d\omega.$$

Therefore, the induced change of variable in K by $f(t)=it\,e^{-|t|}/4$ sends the function (operator) $e^{-|t|}/4$ to the operator $h(t)=\delta(t)/2+\frac{1}{2}\sum_{n\geqq 1}(-1)^n[it\,e^{-|t|}/4]^n$.

Direct "changes of variable" mappings in the original (operator) space K can also be defined (at least) for distributions in K if they map to other distributions in K. The dilation mappings U_n are examples which, however, have extensions to all of K because they turn out to be ring automorphisms. Moreover, these turn out to be *induced* changes of variables at the same time. The direct change of variable $x(t)\mapsto x(e^t)$ for suitable operators has for its Fourier transform the function mapping $\tilde{x}(\omega)\mapsto \mathcal{M}[\tilde{x}](i\omega)$, where $\mathcal{M}$ is the Mellin transformation [28]. Other examples of direct changes of variables (given for functions) appear in [29].

It has been suggested above that the convergence of an infinite series of operators might be defined as that induced from a.e. convergence of the transformed series. In view of the makeup of the transformed space $\mathcal{K}$ of functions, it does seem that this is an appropriate definition to use not only for the convergence of operator series, but also for other limit processes in K. Indeed, starting with some (any) classical a.e. limit process from and to some functions in the ring $\mathcal{K}$ one always obtains a corresponding induced limit process for some operators to operators. One might well call these "generalized convergences" in K, and attempt to characterize them intrinsically in K. There seem to be many interesting possibilities for investigations along these lines, developing an "analysis" for operators. For example, the classical Fourier transform of the function (operator) $e^{-\rho t}H(t)$ $(\rho>0)$ is the function $1/(i\omega+\rho)$, and the pointwise limit as $\rho\to 0$ of the latter exists for every $\omega\neq 0$ and yields the Fourier transform in $\mathcal{K}$ of $H(t)$. Another example is the $\lim_{\xi\to 0}\tilde{x}(\omega+\xi)-\tilde{x}(\omega)/\xi=\tilde{x}^{(1)}(\omega)$ a.e. on $\mathcal{R}$, for any absolutely continuous function $\tilde{x}\in\mathcal{K}$, as an extension of the algebraic derivative D in K. If the Fourier transform $\tilde{x}$ of an operator x is continuous at α, then with respect to a.e. convergence,

$$\lim_{n\to\infty}U_n\,e^{-i\alpha t}x(t)=\tilde{x}(\alpha)\delta(t),$$

since the Fourier transform of the left member is $\tilde{x}(\omega/n+\alpha)$. (See Corollary 6 of Part I.) If x is in L_1, then with respect to a.e. convergence,

$$x=\int_{-\infty}^{\infty}e^{-s\tau}x(\tau)\,d\tau,$$

where $e^{-s\tau}$ is the translation operator $f(t-\tau)/f(t)=\delta(t-\tau)$, since the Fourier transform of this (formal) integral is $\int_{-\infty}^{\infty}e^{-i\omega\tau}x(\tau)\,d\tau=\tilde{x}(\omega)$. Mikusiński [14] has obtained an analogous result for locally integrable functions on the half-line $t\geqq 0$.

6. An extension of the Fourier transform. Regarding $\mathcal{K}$ as the collection of all (nonpathological) functions of a real variable, we can (and shall) define transforms for arbitrary functions in analogy with the classical Fourier integral formula, where the

Fourier transform of a function is 2π times the inverse Fourier transform applied to the reflected function. To this end, let us first recall that an operator is said to be a function if it is a tempered distribution and, as such, it is regular. Thus the functions in K are all those regular distributions (i.e. locally integrable functions) whose Fourier transforms (as tempered ultradistributions) are also regular. These functions can simultaneously be considered also as elements of $\mathcal{K}$ (by simply changing the t variable to the ω variable). Indeed the functions in K can be considered to constitute precisely the intersection $K \cap \mathcal{K}$ of the operator space K and the function space $\mathcal{K}$, i.e. $x \in K \cap \mathcal{K}$ iff $t \mapsto x(t)$ is a function in K and $\omega \mapsto x(\omega)$ is the "same" function in $\mathcal{K}$. With this in view, we introduce a new mapping FT from (ostensibly) $\mathcal{K}$ onto K.

DEFINITION 7. Let f be a measurable, finite-valued function defined for a.e. real number. The operator $x \in K$ whose Fourier transform satisfies $\tilde{x} = 2\pi \check{f}(\check{f}(\omega) = f(-\omega))$ is called the FT *of* f and is denoted by FT $\{f\}$. The mapping

$$\text{FT}: f \mapsto \text{FT}\{f\} = x \quad (\text{where } \tilde{x} = 2\pi\check{f}) \tag{13}$$

from measurable, finite-valued functions onto K is called the FT.

If f and its classical Fourier transform are both L_1 functions, then $f(t) = (1/(2\pi)) \int_{-\infty}^{\infty} e^{i\xi t}\tilde{f}(\xi)\, d\xi$, and so

$$2\pi\check{f}(\omega) = 2\pi f(-\omega) = 2\pi\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\xi(-\omega)}\tilde{f}(\xi)\, d\xi\right) = \int_{-\infty}^{\infty} e^{-i\xi\omega}\tilde{f}(\xi)\, d\xi = \tilde{\tilde{f}}(\omega).$$

Hence, according to Definition 7, FT $\{f\} = \tilde{f}$. Moreover, this same conclusion applies for any function in K, since the Fourier formula $2\pi\check{f}(\omega) = \tilde{\tilde{f}}(\omega)$ also holds for tempered distributions [30, p. 189]. A direct, elementary proof of the following is included in [26].

PROPOSITION 4. *Let f be an operator which is also a function (i.e., $f \in K \cap \mathcal{K}$). Then $2\pi\check{f} = \tilde{\tilde{f}}$ holds and* FT $\{f\} = \tilde{f}$ *is the distributional Fourier transform of f.*

As an illustration, let us consider the Heaviside step function $H(t)$, which is represented by an operator in K, but is not a function in K. On the other hand by Definition 7, FT $\{H(t)\} = x$ is the operator satisfying $\tilde{x}(\omega) = 2\pi H(-\omega)$. Recalling that $\widetilde{pv(1/t)} = -i\pi \operatorname{sign} \omega$ and $\widetilde{\delta(t)} = 1$, we write $2\pi H(-\omega) = (1/i)(-i\pi \operatorname{sign}(-\omega) + i\pi)$, and conclude that FT $\{H(t)\} = (1/i)(pv(1/t) + i\pi\delta(t)) = pv(1/(it)) + \pi\delta(t)$. This last, of course, is the distributional Fourier transform of $H(t)$, as a tempered distribution. Other simple illustrations of Definition 7 are polynomials (whose distributional Fourier transforms are linear combinations of the delta function and its derivatives).

Proposition 4 implies that whenever a function f is also considered an operator, then Definition 7 for the FT of the function f is consistent with Definition 2 for the Fourier transform of the operator f. This consistency allows us to extend the Fourier transform to the vector sum $\mathcal{S} = K + \mathcal{K}$ of the vector spaces K of operators and $\mathcal{K}$ of functions. This is not a direct sum, since K and $\mathcal{K}$ have many elements in common, and the Fourier transform must be the same for the common part $K \cap \mathcal{K}$. Now $x_1 + f_1 = x_2 + f_2 \in \mathcal{S}$, with $x_j \in K$ and $f_j \in \mathcal{K}$ iff $x_1 - x_2 = f_2 - f_1 \in K \cap \mathcal{K}$, and for arbitrary $x_1 + f_1$ and $x_2 + f_2$ of $\mathcal{S}$ their sum is given by $(x_1 + f_1) + (x_2 + f_2) = (x_1 + x_2) + (f_1 + f_2)$. Because of Proposition 4, we can thus extend (unambiguously) the Fourier transform as a linear one-to-one mapping of $\mathcal{S}$ onto itself by the rule $x + f \mapsto FT\{f\} + \tilde{x}$. Here $x \in K$ and $f \in \mathcal{K}$, while $\tilde{x} \in \mathcal{K}$ and FT $\{f\} \in K$. Let us formalize these considerations with the following definition and proposition.

DEFINITION 8. The vector space sum of the spaces K of operators and $\mathcal{K}$ of functions is denoted by $\mathcal{S} = K + \mathcal{K}$. If $u = x + f (x \in K, f \in \mathcal{K})$ is an element of $\mathcal{S}$, then the *Fourier transform* of u is the element $\tilde{u} = \mathrm{FT}\{f\} + \tilde{x}$ of $\mathscr{S}$. The mapping $u \mapsto \tilde{u}$ from $\mathscr{S}$ onto $\mathscr{S}$ is called the *Fourier transform.*

PROPOSITION 5. *The Fourier transform defined in Definition* 8 *is a linear extension to $\mathscr{S}$ of the Fourier transform* (*Def.* 2) *on the space* (*ring*) *K of operators. It is one-to-one and onto $\mathscr{S}$, and is compatible with the distributional Fourier transform.*

The main advantage of extending the Fourier transform to the vector sum $\mathscr{S} = K + \mathscr{K}$ is that we can then include as members of the original space, classical functions such as $H(t)$, $t^n H(t)$, polynomials and periodic functions; they are not operators though the former two are represented by operators. Thus the extended Fourier transform should be of increased utility (encompassing the concept of Fourier series, for example [2]). However, the enlarged space is no longer a ring and the extended Fourier transform is no longer a ring isomorphism. On the other hand, the concept of an operational rule readily extends to the enlarged space. For example, the extended algebraic derivative mapping[1] $x + f(t) \mapsto D[x] - itf(t)$ has for its Fourier transform the mapping $\tilde{f}(\omega) + \tilde{x}(\omega) \mapsto \tilde{f}^{(1)}(\omega) + d\tilde{x}/d\omega$, where $\tilde{f}^{(1)}$ is the (ultra) distributional derivative of $\tilde{f}$. Here, if $f(t)$ is the polynomial $\sum a_j t^j$, then $\tilde{f}^{(1)}(\omega) = \sum (-i)^j a_j \delta^{(j+1)}(\omega)$. More generally, an operational rule on $\mathscr{S}$ is a pair of mappings $(M, \tilde{M})$, with nonempty domains $J, \tilde{J} \subseteq \mathscr{S}$, for which $u = x + f \mapsto v = M[u]$ iff $\tilde{u} = \mathrm{FT}\{f\} + \tilde{x} \mapsto \tilde{v} = \tilde{M}[\tilde{u}]$.

We shall conclude with an interesting application of the Fourier transform on the enlarged space $\mathscr{S}$. Let C_r denote the collection of continuous, right-sided functions (vanishing to the left of some number) on R. It will be considered an algebraic ring under pointwise addition and ordinary convolution; as such it is equivalent to the convolution ring used by Mikusiński to obtain his right-sided operators [20]. Using the Fourier transform of Definition 8, we can map this convolution ring C_r onto another ring C_m in which convolution corresponds to pointwise multiplication (Def. 4). In this situation, the Fourier transform actually performs as a ring isomorphism (extending that one referred to in Definition 2, since some elements of C_r are not operators). Now the quotient field of the convolution ring C_r (it has no zero divisors) is essentially the Mikusiński field of right-sided operators. Thus through the Fourier transform (further extended in the obvious way to the quotient field), the Mikusiński field becomes isomorphic to the quotient field Q of the multiplicative ring C_m. Many of the elements of Q are ordinary arithmetical fractions (functions), but some are merely formal fractions. In any case, the extended Fourier transformation yields an alternative procedure for obtaining (isomorphically) the Mikusiński field. (See [5] for the case where the right-sided functions have ordinary Laplace transforms, and also for another approach to the Mikusiński field using arithmetical functions.)

REFERENCES

[1] S. BOCHNER AND K. CHANDRASEKHARAN, *Fourier Transforms*, Annals of Mathematics Studies 19, Princeton Univ. Press, Princeton, NJ, 1948.

[2] T. K. BOEHME AND G. WYGANT, *Generalized functions on the unit circle*, Amer. Math. Monthly, 82 (1975), pp. 256–261.

[3] A. P. CALDERON AND A. ZYGMUND, *Singular integral operators and differential equations*, Amer. J. Math., 79 (1957), pp. 901–921.

[4] W. A. COPPEL, *J. B. Fourier—On the occasion of his two hundredth birthday*, Amer. Math. Monthly, 76 (1969), pp. 468–483.

[1] This formulation of the mapping is unambiguous since the equation $D[f](t) = -itf(t)$ holds for distributions.

[5] V. A. DITKIN AND A. P. PRUDNIKOV, *Integral Transforms and Operational Calculus*, Oxford, 1965.
[6] K. O. FRIEDRICKS, *Pseudo-Differential Operators*, Courant Institute of Mathematical Sciences, New York, 1968.
[7] J. M. GELFAND AND G. E. SHILOV, *Generalized Functions I, II*, Academic Press, New York, 1964, 1968.
[8] R. R. GOLDBERG, *Fourier Transforms*, Cambridge University Press, Cambridge, 1962.
[9] P. R. HALMOS, *Measure Theory*, Van Nostrand, New York, 1950.
[10] L. HÖRMANDER, *On the existence and regularity of solutions of linear pseudo-differential equations*, Enseignement math., 17 (1971), pp. 99–163.
[11] ———, *Fourier Integral Operators, I*, Acta Math., 127 (1971), pp. 79–183.
[12] C. C. HUGHES AND R. A. STRUBLE, *Neocontinuous Mikusinski operators*, Trans. Amer. Math. Soc., 185 (1973), pp. 383–400.
[13] J. J. KOHN AND L. NERINBERG, *An algebra of pseudo-differential operators*, Comm. Pure Appl. Math., 19 (1965), pp. 269–305.
[14] J. MIKUSIŃSKI, *Operational Calculus*, Pergamon, New York, 1959.
[15] M. ORTON, *Hilbert transforms, Plemelj relations, and Fourier transforms of distributions*, this Journal, 4 (1973), pp. 656–670.
[16] R. E. A. C. PALEY AND N. WIENER, *Fourier Transforms in the Complex Domain*, American Mathematical Society, Providence, RI, 1934.
[17] B. ROSS, *Fractional Calculus and its Applications*, Springer-Verlag Lecture Notes in Mathematics 457, Berlin-Heidelberg-New York, 1975.
[18] L. SCHWARTZ, *Transformation de Laplace des distributions*, Math. Seminar, Université Lund, tome supplémentaire dédié à M. Riesz, 1953, pp. 196–206.
[19] ———, *Théorie des distributions*, Hermann, Paris, 1966.
[20] R. A. STRUBLE, *On operators and distributions*, Canad. Math. Bull., 11 (1968), pp. 61–64.
[21] ———, *Operator homomorphisms*, Math. Z., 130 (1973), pp. 275–285.
[22] ———, *Representations of Fourier transforms for distributions*, Bull. Inst. Math. Acad. Sinica, 2 (1974), pp. 191–206.
[23] ———, *A two-sided operational calculus*, Studia Math., 60 (1977), pp. 239–254.
[24] ———, *The numerical-valued Fourier transform in the two-sided operational calculus*, this Journal, 8 (1977), pp. 243–257.
[25] ———, *Analytical and algebraic aspects of the operational calculus*, SIAM Rev., 19 (1977), pp. 403–436.
[26] ———, *The Fourier theorem*, J. Franklin Inst., 306 (1978), pp. 63–76.
[27] Z. SZMYDT, *On regular temperate distributions*, Studia Math., 44 (1972), pp. 309–314.
[28] E. C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals*, Oxford University Press, London, 1937.
[29] B. VAN DER POL AND H. BREMMER, *Operational Calculus Based on the Two-sided Laplace Integral*, Cambridge University Press, Cambridge, 1959.
[30] A. H. ZEMANIAN, *Distribution Theory and Transform Analysis*, McGraw-Hill, New York, 1965.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0019 $01.00/0

EXPANSIONS OF GENERALIZED COMPLETELY CONVEX FUNCTIONS*

DAN AMIR† AND ZVI ZIEGLER‡

Abstract. The concept of a generalized completely convex function is extended and a unified presentation is developed for expanding such functions by Taylor–Lidstone series. It is shown that these expansions are in fact tantamount to representation theorems for the elements of the cone of generalized completely convex functions in terms of the extreme rays.

1. Introduction. Classes of functions such that an infinity of requirements on their derivatives are imposed have been investigated by many authors. The simplest instances are the absolutely monotone and completely monotone classes. These have been studied by Bernstein, Widder and Hausdorff (see e.g. Widder [23]), and analyticity properties of functions of such classes were derived. The concept of absolute monotonicity on a finite interval has been generalized by Karlin and Ziegler [10] and by Amir and Ziegler [1] to include a class of functions obeying an infinite sequence of differential inequalities of the form

$$D_{n-1}\cdots D_0 f \geqq 0, \qquad n=0,1,\cdots,$$

where the D_i's, $i \geqq 0$, are first order differential operators of a special form (and $D_{-1}f \equiv f$). Functions of this class, under appropriate conditions on the D_i's, admit of a convergent series expansion

$$\sum_{i=0}^{\infty} a_i u_i(x)$$

where the $\{u_i\}_0^\infty$ is the Chebyshev system associated with the D_i's.

In another vein, the analyticity of f was shown to be a consequence of a weaker set of restrictions. In fact, Bernstein [2, p. 197] proved that it suffices to impose conditions of the form

$$\varepsilon_k f^{(n_k)}(x) \geqq 0, \qquad k=1,2,\cdots; \quad x \in (a,b)$$

where $\{\varepsilon_k\}_1^\infty$ is a sequence of signs, and $\{n_{k+1}/n_k\}$ is a bounded sequence.

An especially important class of functions that fits into the above category is the class of "completely convex" functions.

Widder [22] defined the "completely convex" functions on $[a, b]$ as those $f \in C^{(\infty)}[a, b]$ satisfying $(-1)^n f^{(2n)}(x) \geqq 0$ for all $a \leqq x \leqq b$ and $n = 0, 1, 2, \cdots$. He showed that the completely convex functions on $[0, 1]$ are exactly the functions which have a uniformly convergent series representation:

$$f(x)=\sum_{n=0}^{\infty}(a_n\Lambda_n(x)+b_n\Lambda_n(1-x))+c\sin \pi x \tag{1.1}$$

where $(-1)^n a_n \geqq 0$, $(-1)^n b_n \geqq 0$, $n = 1, 2, \cdots$, $c \geqq 0$ and $\Lambda_n(x)$ are the "Lidstone polynomials": $\Lambda_0(x) = x$ and $\Lambda_n(x)$ is the unique solution of $\Lambda_n^{(2n)} = \Lambda_0$, $\Lambda_n^{(2k)}(0) = \Lambda_n^{(2k)}(1) = 0$, $k = 0, 1, \cdots, n-1$. In this representation, $a_n = f^{(2n)}(0)$, $b_n = f^{(2n)}(1)$ and c is the maximal such that $f(x) - c \sin \pi x$ is completely convex.

Widder showed that every completely convex function is the restriction to $[0, 1]$ of an entire function. Extensions of this result were obtained by Boas and Polya [4], Protter [20] and Leeming and Sharma [16]. Pethe and Sharma [19] defined the

* Received by the editors September 30, 1977. This work was supported in part by the Israeli Academy of Sciences.

† Tel Aviv University, Tel-Aviv, Israel.

‡ Technion–Israel Institute of Technology, Haifa, Israel.

"completely convex*" functions on $[a, b]$ to be those completely convex functions satisfying also: $(-1)^{n+1}f^{(2n+1)}(a) \geqq 0$, $n = 0, 1, 2, \cdots$. They found an analogous representation formula:

$$f(x) = \sum_{n=0}^{\infty} (b_n\mu_n(x) - a_n\nu_n(x)) + c \cos \frac{\pi x}{2}. \tag{1.2}$$

Buckholtz and Shaw [5] showed that both results are particular cases of a more general result: Given a second order Sturm–Liouville differential operator:

$$Lf = -(pf')' + gf, \qquad p \in C^1[a, b], \quad g \in C[a, b], \quad p > 0$$

and regular homogeneous linear separable boundary conditions $B_af = B_bf = 0$ such that the fundamental solutions ($Lf_0 = 0$, $B_af_0 = 0$, $B_bf_0 = 1$ and $Lf_1 = 0$, $B_af_1 = 1$, $B_bf_1 = 0$) are nonnegative, and such that the eigenvalues λ ($Lf = \lambda f$, $B_af = B_bf = 0$, $f \neq 0$) are all positive, Buckholtz and Shaw defined the "LB-positive functions" as those infinitely differentiable functions f satisfying $L^nf(x) \geqq 0$, $B_aL^nf \geqq 0$, $B_bL^nf \geqq 0$ for all $a \leqq x \leqq b$, $n = 0, 1, 2, \cdots$. Their representation formula is:

$$f(x) = \sum_{n=0}^{\infty} (a_nf_{2n}(x) + b_nf_{2n+1}(x)) + cy_0(x) \tag{1.3}$$

where $a_n, b_n, c \geqq 0$, $Lf_n = f_{n-2}$, $B_af_n = B_bf_n = 0$ for $n \geqq 2$, and y_0 is the normalized eigenfunction corresponding to the first eigenvalue λ_0, chosen such that $y_0 \geqq 0$. In this representation $a_n = B_bL^nf$, $b_n = B_aL^nf$, and c is the largest such that $f(x) - cy_0(x)$ is LB-positive. Note that the classical case corresponds to $Lf = -f''$, with $f(0) = f(1) = 0$. A related result not contained in the Buckholtz–Shaw treatment is due to Leeming and Sharma [16]: They defined the "W_p-convex" functions as those satisfying $(-1)^nf^{(pn)}(x) \geqq 0$ and $(-1)f^{(pn+j)}(a) \geqq 0$ for all $a \leqq x \leqq b$, $j = 1, \cdots, p-2$ and $n = 0, 1, 2, \cdots$.

Shaw [21] studied the possibility of extending the LB-series expansion of functions in $C^\infty[a, b]$ to series related to an nth order linear differential operator $Lf = \sum_{i=0}^m p_i(x)f^{(m-i)}$, $p_i \in C^{m-i}[a, b]$, $p_0 > 0$ and linearly independent boundary conditions $B_if = 0$, $i = 1, \cdots, m$, such that the eigenvalue problem $Lf = \lambda f$, $B_if = 0$, $i = 1, \cdots, m$ is self-adjoint. He characterized those functions f for which the corresponding LB-series converges to f.

We present in this paper a unified approach to the theory of such series expansion and to the analysis of the concept of generalized complete convexity. In § 2 we introduce the Taylor–Lidstone series with respect to a linear operator L with an m dimensional kernel and associated linear functionals. We discuss conditions for the convergence of such series. In particular, we consider the case where L is a differential operator and the functionals involve boundary conditions.

In § 3 we generalize the concept of complete convexity, and use the associated convex cones and their extreme rays to obtain representation theorems generalizing the classical Lidstone series representations. The general framework developed in this section is utilized in § 4 to establish the corresponding representations for a general class of factorizable differential operators with certain types of boundary conditions. Using the powerful method of total positivity (see Karlin [9]) it is established that the conditions for the validity of the expansions are satisfied in such situations. Particular examples include the Rod equation with two types of boundary conditions, as well as the Sturm–Liouville operators discussed by Buckholtz and Shaw [5] and by Pethe and Sharma [19]. We close with the incorporation of the recent results of Leeming and Sharma [16] into our framework.

2. The Taylor–Lidstone expansion. Let L be a linear operator in a linear space E, defined on a linear subspace D of E, with an m-dimensional kernel N. We assume that $D \subset LD$, hence $N \subset D$. Let $B_1, \cdots, B_m$ be linear functionals on E which are total over N, i.e. such that $Lu = 0$, $B_i(u) = 0$, $i = 1, \cdots, m$ implies $u = 0$. Where L is a differential operator and $\{B_i\}$ are boundary conditions, then they are total iff there exists an associated Green's function.

The functionals $B_1, \cdots, B_m$ determine a basis $u_1^\circ, \cdots, u_m^\circ$ of N biorthogonal to $B_1, \cdots, B_m$ (i.e., with $B_j(u_i^\circ) = \delta_{ij}$). Set $Z = \bigcap_{i=1}^m B_i^{-1} 0$. The operator L restricted to $Z \cap D$ is one to one and carries $Z \cap D$ onto LD. Thus its restriction has a right inverse G, $G\colon LD \to Z \cap D$ possessing the explicit form

$$GLu = u - \sum_{i=1}^{m} B_i(u)u_i^\circ, \qquad u \in D. \tag{2.1}$$

Since $Z \cap D \subset D \subset LD$, G^nL^nu is well-defined for all $u \in D$. In particular, $u_i^k = G^k u_i^\circ$ are defined for $k = 0, 1, 2, \cdots$, $i = 1, \cdots, m$ and we call them the *Taylor–Lidstone* (TL) sequence corresponding to L, $\{B_i\}_1^m$. We observe that the u_i^k are linearly independent: if $\sum_{j=1}^n \alpha_j u_{i(j)}^{k(j)} = 0$, we may assume $k(1) = \cdots = k(r) > k(r+1) \geqq \cdots \geqq k(n)$, $i(1) < i(2) < \cdots < i(r)$ and that $\alpha_1 \neq 0$. But applying $L^{k(1)}$ we get $\sum_{j=1}^r \alpha_j \overset{\circ}{u}_{i(j)} = 0$, which implies $\alpha_1 = 0$ since the system $\overset{\circ}{u}_1, \cdots, \overset{\circ}{u}_m$ is linearly independent.

Denoting by D^n the domain of definition of L^n ($D^n = L^{n-1}D^{n-1} \cap D$) and setting $D^\infty = \bigcap_{n=1}^\infty D^n$ we get from (2.1) by induction:

$$G^nL^nu = u - \sum_{k=0}^{n-1} \sum_{i=1}^{m} B_i(L^ku)u_i^k, \quad \forall\, u \in D^n. \tag{2.2}$$

The induction step is: $L^nu = \sum_{i=1}^m B_i(L^nu)u_i^\circ + GL(L^nu)$; hence $G^nL^nu = \sum_{i=1}^m B_i(L^nu)u_i^n + G^{n+1}L^{n+1}u$.

We call $\sum_{k=0}^{n-1}\sum_{i=1}^m B_i(L^ku)u_i^k$ the *TL-expansion* of $u \in D^{n-1}$ and $\sum_{k=0}^{\infty}\sum_{i=1}^m B_i(L^ku)u_i^k$ the *TL-series* of $u \in D^\infty$. Observe that this represents a generalization of Taylor's expansion in $C[a, b]$, where $D = C^1[a, b]$, $Lf = f'$, $m = 1$, $B_1(f) = f(a)$, $\overset{\circ}{u}_1 = 1$, $Gf(x) = \int_a^x f(t)\,dt$, $u_1^k = (x-a)^k/k!$ and the expansion of f is thus $\sum_{k=0}^\infty (f^{(k)}(a)/k!)(x-a)^k$. The Lidstone series fits also as a special case of the present discussion. Here we choose $D = C^{(2)}[0, 1]$, $Lf = f''$, $B_1 f = f(0)$, $B_2 f = f(1)$, so that $\overset{\circ}{u}_1 = 1 - x$, $\overset{\circ}{u}_2 = x$. In general we can take any linear differential operator L of order m on $C[a, b]$, $Lf(x) = \sum_{i=0}^m p_i(x) f^{(n-i)}(x)$, with the linear homogeneous boundary condition $\mathscr{B} = \{B_1, \cdots, B_m\}$. Assume that the boundary value problem $Lf = 0$, $B_i(f) = 0$, $i = 1, \cdots, m$ admits of the trivial solution only, and let $u_1^\circ, \cdots, \overset{\circ}{u}_m$ be the fundamental solutions. Set $Gf(x) = \int_a^b G(x, t) f(t)\,dt$ where $G(x, t)$ is the corresponding Green's function. It is readily seen that this fits into our formal framework.

An example in an abstract setting is obtained by taking $L = \overleftarrow{S}^m$, where $\overleftarrow{S}$ is the left shift operator on the sequence space E, i.e., $(\overleftarrow{S}u)(n) = u(n+1)$, and where $B_i = e_i^*$ ($i = 1, \cdots, m$) are the coordinate functionals. In this case $u_i^\circ = e_i$ ($i = 1, \cdots, m$) are the unit vectors, and $G = \vec{S}$ where $\vec{S}$ is the right shift, viz., $(\vec{S}u)(1) = 0$, $(\vec{S}u)(n+1) = u(n)$. The corresponding TL sequence is $u_i^k = e_{mk+i}$ ($k = 0, 1, 2, \cdots$, $i = 1, \cdots, m$) so that the TL expansion is the formal expansion of u in terms of its coordinates $u \sim \sum_{k=0}^\infty \sum_{i=1}^m u(mk+i)e_{mk+i}$.

If E is any linear space, then the linear subspace spanned by the TL sequence of $L, B_1, \cdots, B_m$ is the subspace $N^\infty = \bigcup_{k=1}^\infty N(L^k)$ of generalized eigenvectors of the eigenvalue 0 of L, and $L, B_1, \cdots, B_m$ act on it as $\overleftarrow{S}^m, e_1^*, \cdots, e_m^*$ act on the sequence space. On the other hand, if u is an eigenvector corresponding to another eigenvalue $\lambda \neq 0$, then $L^n u = \lambda^n u \neq 0\ \forall n$; hence $u \notin N^\infty$.

If we have a linear topology on E, the following questions arise naturally:

(i) Find the subspace $C = \{u \in D^\infty;\ G^n L^n u \text{ converges}\}$ of elements for which the TL series converges, and its subspace $C_0 = \{u \in D^\infty;\ G^n L^n u \to 0\}$ of elements which can be expanded as a TL-series.

(ii) The analogous question when we consider the ungrouped TL-series, i.e. the series $\sum_{j=1}^\infty \alpha_j(u) v_j$ where $v_{mk+i} = u_i^k$ and $\alpha_{mk+i}(u) = B_i(L^k u)$, $i = 1, \cdots, m$.

(iii) Does the TL sequence constitute a basis in C_0, i.e. does 0 possess only the trivial representation $\sum_{j=i}^\infty 0 \cdot v_j$?

In the case of $\tilde{S}^m$, $e_1^*, \cdots, e_m^*$ in c_0 or $l_p (1 \leqq p < \infty)$, we know that the TL-expansion is just the expansion with respect to the basis consisting of unit vectors, so that the TL sequence is a basis in $C_0 = E$. The same of course applies to any space with a basis represented as a sequence space with the unit vector basis (e_n). In the case $\tilde{S}^m$, $e_1^*, \cdots, e_m^*$ is in c or m, the unit vector sequence is basic in its closed span $C_0 = c_0$.

In the Taylor case, the TL sequence $((x-a)^n/n!)$ is basic in the subspace C_0 of analytic functions in $D^\infty = C^\infty[a, b]$, both in the supremum norm and in the topology of uniform convergence of all derivatives.

In order to obtain more structure in the general case, we assume that in the given linear topology G is continuous, and that a sequence of eigenvectors of G, $(y_j)_{j=0}^\infty$, corresponding to the eigenvalues $(\mu_j)_{j=0}^\infty$ such that $|\mu_0| = |\mu_1| = \cdots = |\mu_p| > |\mu_{p+1}| \geqq |\mu_{p+2}| \geqq \cdots$ $(p \geqq 0)$, forms a basis for LD, with the coefficient functionals (φ_j). In this case, $Gu = G(\sum_{j=0}^\infty \varphi_j(u) y_j) = \sum_{j=0}^\infty \varphi_j(u) G y_j = \sum_{j=0}^\infty \mu_j \varphi_j(u) y_j$; hence $\varphi_j(Gu) = \mu_j \varphi_j(u)$, $\forall\, u \in LD$. In particular, $u_i^k = G^k u_i^\circ = \sum_{j=0}^\infty \mu_j^k \varphi_j(u_i^\circ) y_j$. We have, therefore,

$$\mu_0^{-k} u_i^k - \sum_{j=0}^{p} \left(\frac{\mu_j}{\mu_0}\right)^k \varphi_j(u_j^\circ) y_j = \sum_{j=p+1}^{\infty} \left(\frac{\mu_j}{\mu_0}\right)^k \varphi_j(u_j^\circ) y_j. \tag{2.3}$$

If we know only that $(y_j)_{j=0}^\infty$ is a basis in $Z \cap LD$ we get instead of (2.3):

$$\mu_0^{1-k} u_i^k - \sum_{j=0}^{p} \left(\frac{\mu_j}{\mu_0}\right)^{k-1} \varphi_j(u_j^1) y_j = \sum_{j=p+1}^{\infty} \left(\frac{\mu_j}{\mu_0}\right)^{k-1} \varphi_j(u_j^1) y_j. \tag{2.3'}$$

Under additional assumptions, e.g. that (y_v) is an unconditional basis, or that the $\{y_j\}$ is bounded, $\{\varphi_j\}$ is equicontinuous, and $\{\mu_j\} \in \bigcup_{p=1}^\infty l_p$, the right-hand side of (2.3) (or (2.3')) tends to 0 when $k \to \infty$. Thus, if $\mu_0 = \mu_1 = \cdots = \mu_r = -\mu_{r+1} = \cdots = -\mu_p$ we have $\mu_0^{-2k} u_i^{2k} \to u_i^+ \equiv \sum_{j=0}^p \varphi_j(u_i^\circ) y_j$, while $\mu_0^{1-2k} u_i^{2k-1} \to u_i^- \equiv \sum_{j=1}^r \varphi_j(u_i^\circ) y_j - \sum_{j=r+1}^p \varphi_j(u_i^\circ) y_j$. u_i^+ and u_i^- vanish only if $\varphi_j(u_i^\circ) = 0$ for $j = 0, \cdots, p$. If $r = p$, then $u_i^+ = u_i^-$ is either 0 or an eigenvector of μ_0.

If $L^k u \in Z$ for all k and for all $j > p$ there exists c_j such that $|\varphi_j(L^k u)| \leqq c_j |\varphi_p(L^k u)|$ for $k \geqq k_0$, then

$$\begin{aligned} |\varphi_j(u)| = |\varphi_j(G^k L^k u)| &= |\mu_j^k|\, |\varphi_j(L^k u)| \\ &\leqq c_j |\mu_j^k|\, |\varphi_p(L^k u)| \\ &= \left|\frac{\mu_j}{\mu_p}\right|^k |\varphi_p(G^k L^k u)| \\ &= c_j \left|\frac{\mu_j}{\mu_p}\right|^k |\varphi_p(u)| \to 0. \end{aligned}$$

Hence, under the former assumption on (y_j), $u = G^k L^k u = \sum_{j=0}^p \varphi_j(u) y_j$.

A case when assumptions of this type are fulfilled is when L is a differential operator on $C[a, b]$ as mentioned above. The integral operator $Gf(x) = \int_a^b G(x, t)f(t)\,dt$, is known to be compact in both the L_2-norm and the maximum norm of $C[a, b]$, so that its eigenvalue counted with multiplicities and arranged according to decreasing absolute values form a countable set with 0 as the only possible limit.

If $Lf = 0$, $B_i f = 0$, $i = 1, \cdots, n$ is a self-adjoint boundary value problem we know that (y_j) can be chosen to be an orthonormal (hence unconditional) basis of $L^2[a, b]$, and the corresponding Fourier series converges uniformly for every $f \in D \cap Z$ (Naimark [18, p. 82]).

In the nonselfadjoint case, if the boundary conditions are "regular" in the sense of Tamarkin (see Naimark [18, p. 56]) and the adjoint differential operator exists, and if the eigenvalues of G are simple poles of $G(\xi, t, \lambda)$, then there is a basis (y_j, z_j) for $L^2[a, b]$ consisting of the eigenfunctions (y_j) corresponding to the eigenvalues (μ_j) of G, and the eigenfunctions (z_j) corresponding to the eigenvalues $(\bar{\mu}_j)$ of G^* (in the ordering $|\mu_0| \geqq |\mu_1| \geqq \cdots$) such that for $f \in D \cap Z$, $f = \sum_{j=0}^{\infty} \langle f, z_j \rangle y_j$ uniformly (Naimark [18, p. 89]).

If the boundary conditions are "strongly regular" (Mihailov [17]) the eigenfunctions of G form a basis for $L^2[a, b]$ equivalent to an orthonormal basis (hence unconditional) (see Mihailov [17], Kesselman [14] or Naimark [18, p. 80]).

We next turn our attention to the nonregular separable boundary conditions. The theorem of Khromov [15] guarantees uniform convergence of the eigenfunction expansion of functions f of D^∞ satisfying the "L-analyticity condition", on $(0, a)$, $0 < a < 1$, namely

$$\left[\frac{|(d/dx)^i L^k f(x)|}{(kn+i)!}\right]^{1/(kn+i)} \leqq C, \qquad 0 \leqq k; \quad 0 \leqq i \leqq n-1, \tag{2.4}$$

for $x \in [\alpha, \beta]$, where $[\alpha, \beta] \subset (0, a)$, and c depends only on L, f and $[\alpha, \beta]$. The eigenfunction expansion converges uniformly in every interval $[0, b]$, $b < \min(R, a)$ where

$$\frac{1}{R} = \overline{\lim_{k\to\infty}} \left\{\frac{1}{k!}\left|\left(\frac{d}{dx}\right)^{k-n[k/n]} L^{[k/n]} f(x)\right|_{x=0}\right\}^{1/k}. \tag{2.5}$$

We note further that, by a theorem of Keldysh [13], the μ_j's are eventually distinct, and $\mu_j \sim 1/(jn)$ for large j, where n is the order of the equation. Hence $\{\mu_j\} \in \bigcup_{p=1}^{\infty} l_p$ and the analysis following (2.3′) applies.

These observations will assist us in establishing the convergence in the Leeming–Sharma case in § 4.

3. Generalized completely convex elements. Suppose that E is a partially ordered linear space, and that L, N, $\mathscr{B} = \{B_i\}_1^m$, $\{u_i^\circ\}$, Z, G, $\{u_i^k\}$ are as in § 2. Then the set $K = \{u \in D^\infty;\ L^k u \geqq 0,\ B_i(L^k u) \geqq 0,\ k = 0, 1, 2, \cdots;\ i = 1, \cdots, m\}$ is a proper convex cone in E which we call the cone of *completely convex elements* (with respect to $(L, \mathscr{B})$). Clearly $LK \subset K$.

When we assume that $u_i^\circ \in K$ (i.e.: $u_i^\circ \geqq 0$), $i = 1, \cdots, m$, then the u_i° are extremal in K. Indeed, if $u_i^\circ = v_1 + v_2$, $v_j \in K$ $(j = 1, 2)$, then necessarily $Lv_j = 0$ and $B_k(v_j) = 0$ for $j = 1, 2$ and $k \neq i$, so that $v_j = B_i(v_j)u_i^\circ$. The expansion $u = \sum_{i=1}^m B_i(u)u_i^\circ$, $u \in N$, shows that these are the only extreme rays of K belonging to N.

Assuming also that $G \geqq 0$, i.e. that $u \geqq 0$, $Lv = u$, $v \in Z$ implies $v \geqq 0$ (which with the assumption $u_i^\circ \geqq 0$ means that: $u \geqq 0$, $Lv = u$, $B_i(L_v) \geqq 0$, $i = 1, \cdots, m \Rightarrow v \geqq 0$), we have $GK \subset K$, hence $LK \supset LGK = K$ and $LK = K$. For $u \in K$ we have

$u - \sum_{k=0}^{n-1} \sum_{i=1}^{m} B_i(L^k u) = G^n L^n u \in K$; thus the TL expansion of u is nondecreasing and bounded by u.

If $u \in K$ is extremal, then either $Lu = 0$ and then, as observed above, $u = \alpha \overset{\circ}{u}_i$ for some $1 \leqq i \leqq m$, or $B_i(u) = 0$ for $i = 1, \cdots, m$, i.e. $u \in Z$. Moreover, Lu and Gu are extremal too. In fact, if $Lu = v_1 + v_2 \neq 0$, $v_j \in K$, then since $u \in Z$ we have $u = GLu = Gv_1 + Gv_2$, $Gv_j \in K$; hence $Gv_j = \alpha_j u$ and $v_j = LGv_j + \alpha_j Lu$ for $j = 1, 2$. If $Gu = v_1 + v_2$, $v_j \in K$ then necessarily $v_j \in Z$ and thus $u = LGu = Lv_1 + Lv_2$, $Lv_j \in K$ implying $Lv_j = \alpha_j u$ and $v_j = GLv_j = \alpha_j Gu$ for $j = 1, 2$. In particular, all the elements u_i^k ($k = 0, 1, 2, \cdots; i = 1, \cdots, m$) of the TL sequence are extremal in K.

The other extremal elements $u \in K$ must satisfy $L^k u \neq 0 \; \forall \, k$, in which case $L^k u \in Z \; \forall \, k$, i.e. u belongs to the extremal subcone $K_\infty \equiv \bigcap_{k=0}^{\infty} G^k K$ of K. Thus the extreme rays of K are those generated by the TL-sequence and the extreme rays of the extremal subcone K_∞, which we have to study in order to get extremal ray representations of the elements of K.

If $y \in K$ is an eigenvector corresponding to some positive eigenvalue μ of G, then obviously $y \in K_\infty$. These are the only eigenvectors of G in K ($\mu = 0$ is not an eigenvalue).

If we assume that the system (y_j) of eigenvectors of G satisfies the conditions specified in § 2, and that for every $j > p$ there exists a constant c_j such that

$$|\varphi_j(u)| \leqq c_j |\varphi_p(u)|, \quad \forall \, u \geqq 0 \tag{3.1}$$

then for every $u \in K_\infty$ we have $L^k u \in Z \; \forall \, k$. Hence $u = \sum_{j=0}^{p} \varphi_j(u) y_j$, and thus $K_\infty \subset \operatorname{span} \{y_0, y_1, \cdots, y_p\}$. This facilitates the task of finding ext K_∞. In particular, if $p = 0$ and $c_1, c_2, \cdots$ exist, then K_∞ is either 0 or span y_0, and we know all of ext K.

In the case of a self-adjoint linear differential operator $L, B_1, \cdots, B_m$ as described in § 2, the additional assumptions we impose in § 3 are that the fundamental solutions $u_i^\circ, \cdots, u_m^\circ$ are nonnegative and that Green's function is nonnegative. Under additional assumptions on $B_1, \cdots, B_m$, which will be described in § 4, and which are satisfied in the Sturm–Liouville case, the Green's function turns out to be a "nonnegative oscillatory kernel" (cf. Gantmacher and Krein [6], for the relevant definitions and properties of such kernels) endowing the associated eigenvalues and eigenfunctions with substantial structure. In particular, the eigenvalues (μ_j) satisfy $\mu_0 > \mu_1 > \cdots > 0$ and the eigenfunction y_0 corresponding to μ_0 is strictly positive on $(0, 1)$ and dominates the other eigenfunctions in the sense that for every j there is c_j with

$$| y_j(x)| \leqq c_j y_0(x), \qquad a \leqq x \leqq b, \tag{3.2}$$

which clearly implies (3.1). Because of the orthogonality of the eigenfunctions, y_0 is the only nonnegative eigenfunction so that in this case $K_\infty = \{\alpha y_0; \alpha \geqq 0\}$. Since $y_0(x) > 0$ on (a, b) and $u_i^\circ \geqq 0$ we have also $\mu_0^{-k} u_i^k \to u_i^+ = \alpha_i y_0 \neq 0$ for $i = 1, \cdots, m$. Choosing $x_0 \in (a, b)$ with $y_0(x_0) > 0$, we have $|\mu_0^{-k} u_i^k(x)| \leqq M$ for $i = 1, \cdots, m$ and $a \leqq x \leqq b$, so that $0 \leqq B_i(L^k u) u_i^k(x) \leqq (M/\gamma) B_i(L^k u) u_i^k(x_0)$ eventually $\forall \, u \in K$, where $0 < \gamma < \alpha_i y_0(x_0)$.

Thus, the pointwise convergence of the bounded series of nonnegative terms $\sum_{k=0}^{n-1} \sum_{i=1}^{m} B_i(L^k u) u_i^k(x_0)$ $(u \in K)$ implies the uniform convergence of the TL expansion of u. Since the same holds for $Lu = v$, we get $L(\sum_{k=0}^{n-2} \sum_{i=1}^{m} B_i(L^k u) u_i^k) = \sum_{k=0}^{n-2} \sum_{i=1}^{m} B_i(L^k v) u_i^k$ converging uniformly to $\sum_{k=0}^{\infty} \sum_{i=1}^{m} B_i(L^k v) u_i^k$. This guarantees that the convergence of the TL expansion of $u \in K$ is uniform in each derivative (following the analysis of Cartan (see e.g. Protter [20])), so that its sum w satisfies $u - w \in K_\infty$, i.e. $u - w = c y_0$.

This family of examples can be extended also to nonselfadjoint operators, as will be shown in § 4.

4. Expansions associated with *n*th order differential operators. A class of examples encompassing the classical results as well as the results of Pethe–Sharma and of Buckholtz–Shaw may be obtained by using total positivity properties of certain kernels.

We start with self-adjoint operators of order $2n$. We consider the sequences of first order differential operators

$$\tag{4.1} \begin{aligned} D_i u &= \frac{d}{dx}\frac{1}{w_i}u, \qquad i=0,1,\cdots,n-1,\\ D_i^* u &= \frac{1}{w_i}\frac{du}{dx}, \qquad i=0,1,\cdots,n-1, \end{aligned}$$

where the $\{w_i(x)\}_{i=0}^{n-1}$ are positive functions, such that w_i is of class C^{n-i} on $[0, 1]$. Define the $2n$th order differential operator

$$\tag{4.2} (Mu)(x)=(-1)^n(D_0^* \cdots D_{n-1}^* D_{n-1} \cdots D_0 u)(x).$$

This is a formally selfadjoint operator. Adjoin to it the separable boundary conditions

$$\tag{4.3} \begin{cases} D_1^* \cdots D_{n-1}^* D_{n-1} \cdots D_0 u(0)+(-1)^n c_1 u(0)=0\\ D_2^* \cdots D_{n-1}^* D_{n-1} \cdots D_0 u(0)+(-1)^{n+1} c_2 D_0 u(0)=0\\ \vdots\\ D_{n-1}^* D_{n-1} \cdots D_0 u(0)+(-1)^{2n-1} c_n D_{n-2} \cdots D_0 u(0)=0 \end{cases}$$

and

$$\begin{cases} D_1^* \cdots D_{n-1}^* D_{n-1} \cdots D_0 u(1)+(-1)^n d_1 u(1)=0\\ \vdots\\ D_{n-1}^* D_{n-1} \cdots D_0 u(1)+(-1)^{2n-1} d_n D_{n-2} \cdots D_0 u(1)=0. \end{cases}$$

We assume here that

$$\tag{4.4} 0 \leqq c_k \leqq \infty, \qquad 0 \leqq d_k \leqq \infty, \qquad 0 < c_k + d_k,$$

for all k. The case $c_k = \infty$ (or $d_k = \infty$) is taken to mean that only the corresponding term appears in that equation.

Karlin [8, Chap. 10] proved that the Green's function $G(x, t)$ of the operator M coupled with the boundary conditions (4.3) is a nonnegative oscillatory kernel (see Gantmacher and Krein [6] for the relevant definitions and properties of such kernels) endowing the associated eigenvalues and eigenfunctions with substantial structure. In fact, for the case at hand, more precise results are available, viz.;

For all k and all ordered sequences $0 < x_1 \leqq \cdots \leqq x_k < 1$, $0 < t_1 \leqq \cdots \leqq t_k < 1$, where the number of coalescences amongst the t_i's or x_i's is at most m, we have

$$\tag{4.5} G^*\begin{pmatrix} x_1, \cdots, x_k\\ t_1, \cdots, t_k \end{pmatrix} \geqq 0$$

with strict inequality if and only if the relations

$$x_i < t_{i+n}, \qquad t_i < x_{i+n}, \qquad i = 1, 2, \cdots, k-n, \tag{4.6}$$

hold.

Here

$$\operatorname{sgn} G^*\binom{x_1, \cdots, x_k}{t_1, \cdots, t_k} = \operatorname{sgn} \frac{\det \|G(x_i, t_j)\|_{i,j=1}^k}{\prod_{\mu>\nu}(x_\mu - x_\nu)\prod_{\mu>\nu}(t_\mu - t_\nu)}$$

in case all points are distinct, and is the sign of the limit of the right hand quotient as points coalesce, whenever such coalescence occurs. Relation (4.5) implies that for each k, there exists a power r_k (in our case $r_k = n+1$ for all k) such that the r_kth iterate $G^{(r_k)}(x, t)$ is ETP of order k. Thus, by a theorem of Karlin [7, Thm. 4] the operator T defined by

$$T\varphi(x) = \int_0^1 G(x, t)\varphi(t)\, dt$$

has a countable set of simple positive eigenvalues

$$\lambda_0 > \lambda_1 > \cdots,$$

decreasing to 0. Moreover, the eigenfunction $\varphi_0(x)$, corresponding to λ_0, is strictly positive on (0, 1). The boundary can be taken care of as in (Karlin [7, Thm. 5]), since the vanishing at the end points is of finite order, as $G(x, t)$ coincides with a solution of $Mu = 0$ near an end point. This ensures the fulfillment of condition (2.3). Hence the generalized "Lidstone" expansion is valid for functions satisfying $M^n u \geqq 0$, $n = 0, 1, \cdots$, and (4.3) provided that the fundamental solutions are of one sign in (0, 1). We summarize these results in the following theorem.

THEOREM 4.1. *Let M be a selfadjoint factorizable differentiable operator of the form* (4.2) *coupled with the boundary conditions* (4.3). *A function φ is said to be a generalized completely convex function with respect to M and* (4.3) *if $M^k u \geqq 0$, for all k, and if it satisfies* (4.3). *Such φ has a uniformly convergent Taylor–Lidstone expansion.*

We will later show that these results are in fact valid for certain types of nonselfadjoint operators; we first present some instances of this theorem.

We note that the boundary conditions are of the regular Sturm type (see Naimark [18, vol. 1, p. 60]) and include, as a special case, the standard second order Sturm–Liouville differential operators

$$\begin{aligned} &L(y) = -\frac{d}{dx}\left[p(x)\frac{dy(x)}{dx}\right], \qquad p(x) > 0, \\ &\alpha_1 y(0) + \beta_1 p(0) y'(0) = 0, \\ &\alpha_2 y(1) + \beta_2 p(1) y'(1) = 0. \end{aligned} \tag{4.7}$$

The generalized "Lidstone" expansion problem for such operators was discussed by Buckholtz and Shaw [5]. Another example subsumed by our discussion corresponds to the Beam (or "Rod") equation, where the operator is

$$L(y) = [r(x)y''(x)], \qquad r(x) > 0, \tag{4.8}$$

with the associated boundary conditions

$$\begin{aligned}(ry'')'(0)+c_1y(0)&=0,\\ (ry'')(0)-c_2y'(0)&=0,\\ (ry'')'(1)+d_1y(1)&=0,\\ (ry'')(1)-d_2y'(1)&=0,\end{aligned}\tag{4.9}$$

where $0\leqq c_1,\ c_2,\ d_1,\ d_2\leqq\infty$ and $0<c_k+d_k$, $k=1,2$. Some concrete examples exhibiting the explicit expansion corresponding to this case for a particular choice of r, c_i's and d_i's will be given at the end of this section (e.g. $r(x)\equiv 1$, $c_1=d_1=\infty$, $c_2=d_2=0$, produces $\lambda_0=\pi$, $\varphi_0=\sin\pi x$).

The above discussion can be generalized to nonselfadjoint operators with regular boundary conditions, following an investigation initiated by Karon [12] and elaborated by Karlin [9].

Let D_i, $i=0,1,\cdots,n-1$ be defined as in (4.1) and consider the operator $L_{n-1}=D_{n-1}\cdots D_0$, coupled with the homogeneous separable boundary conditions

$$\begin{aligned}&B_0\colon \sum_{j=0}^{n-1}\alpha_{ij}D_{j-1}\cdots D_0v(0)=0, \qquad i=1,\cdots,p,\\ &B_1\colon \sum_{j=0}^{n-1}\beta_{ij}D_{j-1}\cdots D_0v(1)=0, \qquad i=1,\cdots,n-p,\end{aligned}\tag{4.10}$$

where $D_{-1}v\equiv v$. Assume further that the coefficients appearing in the boundary conditions B_0, B_1 satisfy the following requirement

Assumption A. (i) The $p\times m$ matrix $\tilde{A}=\|(-1)^j\alpha_{ij}\|$ is SC_p (sign consistent of order p) and has rank p (a matrix J is said to be SC_p iff all $p\times p$ nonzero subdeterminants of J have the same sign).

(ii) The $(n-p)\times n$ matrix $B=\|\beta_{ij}\|$ is SC_{n-p} and has rank $n-p$.

An operator L_{n-1} with boundary conditions (4.10) satisfying Assumption A will possess a Green's function (or, equivalently, the only solution of $L_{n-1}u=0$ satisfying the conditions (4.10) is the trivial solution) iff Assumption B is satisfied (see Karlin [9, Thm. 3]), where this assumption is defined by

Assumption B. There exist increasing sets of indices $\{i_\mu\}_1^p$ and $\{j_\nu\}_1^{n-p}$ in $[0, n-1]$ such that

$$A\begin{pmatrix}1, & \cdots, & p\\ i_1, & \cdots, & i_p\end{pmatrix}\neq 0,\qquad B\begin{pmatrix}1, & \cdots, & n-p\\ j_1, & \cdots, & j_{n-p}\end{pmatrix}\neq 0$$

$$j_\mu\leqq i'_\mu,\qquad \mu=1,\cdots,p,$$

where $\{i'_\mu\}_1^p$ are the (ordered) complementary indices to $\{i_\mu\}_1^p$ in $\{0,1,\cdots,n-1\}$.

Karlin [9] proved that if the boundary conditions (4.10) satisfy Assumptions A and B, then the Green's function $G(x,t)$ of $(-1)^{n-p}L_{n-1}$ with the boundary conditions (4.10) satisfies (4.5) with strict inequality iff

$$\begin{aligned}t_\mu&<x_{\mu+n-p}, \qquad \mu=1,2,\cdots,k-(n-p)',\\ x_\mu&<t_{\mu+p}, \qquad \mu=1,2,\cdots,k-p.\end{aligned}\tag{4.11}$$

Repeating the analysis following (4.6), we may now derive

THEOREM 4.2. *Let L_{n-1} and the separable homogeneous boundary conditions* (4.10) *be given, satisfying Assumptions A and B. Assume further that the boundary conditions are regular, and that the fundamental solutions are of one sign on* $(0,1)$. *Then a* *generalized completely convex function with respect to* L_{n-1} *and* (4.10) *admits a uniformly convergent Taylor–Lidstone expansion.*

As concrete examples of the types of equations discussed here, we present the vibrating rod. We start with the rod with clamped ends. This is the special case of (4.8), (4.9) where $r(x)\equiv 1$, $c_1=c_2=d_1=d_2=\infty$, i.e. $y^{(4)}\equiv 0$, $y(0)=y'(0)=y(1)=y'(1)=0$. The eigenvalues $\{\lambda_\mu\}_0^\infty$ here are the positive solutions of

$$\cos\lambda\cosh\lambda=1$$

and the corresponding eigenfunctions are

$$y_\nu(x)=(\sinh\lambda_\nu-\sin\lambda_\nu)[\cosh\lambda_\nu x-\cos\lambda_\nu x]+(\cos\lambda_\nu-\cosh\lambda_\nu)[\sinh\lambda_\nu x-\sin\lambda_\nu x],\qquad \nu=0,1,\cdots.$$

The Green's function has the explicit form (see e.g. [11]),

$$G_1(x,t)=\begin{cases}\dfrac{t^2(1-x)^2}{6}[3x-t(2x+1)], & 0\leqq t\leqq x,\\[2mm] \dfrac{(t-1)^2x^2}{6}[(3-2x)t-x], & x\leqq t\leqq 1,\end{cases}$$

and the fundamental solutions are

$$u_1^\circ(t)=(t-1)^2(2t+1),$$
$$u_2^\circ(t)=t(t-1)^2,$$
$$u_3^\circ(t)=t^2(3-2t),$$
$$u_4^\circ(t)=t^2(1-t).$$

From the general theory, we know that y_0 dominates all y_ν, $\nu\geqq 1$, so that each $u\in C^\infty(0,1)$ satisfying

$$u^{(4n)}(t)\geqq 0,\quad \text{for } t\in(0,1),\quad u^{(4n+1)}(0)\geqq 0,\qquad u^{(4n+1)}(1)\geqq 0$$

admits of the convergent series representation

$$u(t)=\sum_{n=0}^\infty [u^{(4n)}(0)u_1^n(t)+u^{(4n+1)}(0)u_2^n(t)+u^{(4n)}(1)u_3^n(t)+u^{(4n+1)}(1)u_4^n(t)]+cy_0(t)$$

where $u_i^n=G_1^n(u_i^\circ)$, $i=1,2,3,4$; $n=0,1,\cdots$, and c is the maximum value such that $u-cy_0$ still belongs to the cone. Here $G_1(\cdot)$ is the integral operator with the kernel $G_1(x,t)$.

A similar analysis applies to the rod equation with $r(x)\equiv 1$, $c_1=d_1=\infty$, $c_2=d_2=0$, i.e. $y^{(4)}=0$ with $y(0)=y''(0)=y(1)=y''(1)=0$. The eigenvalues $\{\lambda_\nu\}_0^\infty$ here are $\lambda_\nu=(\nu+1)\pi$, $\nu=0,1,\cdots$ and the eigenfunctions are $\{\sin(\nu+1)\pi x\}_0^\infty$. The Green's function has the explicit form (see e.g. [11]),

$$G_2(x,t)=\begin{cases}\dfrac{x(1-x)(2-x)}{6}t+\dfrac{x-1}{6}t^3, & 0\leqq t\leqq x,\\[2mm] \dfrac{x(1-x)(1+x)}{6}(1-t)+\dfrac{x}{6}(t-1)^3, & x\leqq t\leqq 1,\end{cases}$$

and the fundamental solutions are

$$u_1^\circ(t)=1-t,$$
$$u_2^\circ(t)=t,$$
$$u_3^\circ(t)=\tfrac{1}{6}t(t-1)(2-t),$$
$$u_4^\circ(t)=\tfrac{1}{6}t(t^2-1).$$

Observe that here $u_3^\circ(t)\leqq 0$, $u_4^\circ(t)\leqq 0$.

From the general theory, we conclude that y_0 dominates all y_ν, $\nu\geqq 1$ (this is evident in this case directly), so that each $u\in C^\infty(0,1)$ satisfying

$$u^{(4n)}(t)\geqq 0,\quad \text{for } t\in(0,1),\qquad u^{(4n+2)}(0)\leqq 0,\qquad u^{(4n+2)}(1)\leqq 0$$

admits of the convergent series representation

$$u(t)=\sum_{n=0}^{\infty}\left[u^{(4n)}(0)u_1^n(t)+u^{(4n)}(1)u_2^n(t)+u^{(4n+2)}(0)u_3^n(t)+u^{(4n+2)}(1)u_4^n(t)\right]+c\sin\pi t$$

where $u_i^n=G_2^n(u_i^\circ)$, $i=1,2,3,4$; $n=0,1,\cdots$, and c is the maximal value such that $u(t)-c\sin\pi t$ still belongs to the cone.

We close with an example of nonregular separable boundary conditions which fits in the framework of the discussion in § 3.

The example is due to Leeming and Sharma [16] who established the validity of the expansion using different methods.

We start with a brief sketch of some of the basic properties needed for our discussion, which can be found in the paper of Leeming and Sharma. The operator is $Ly=-y^{(n)}$ with the nonregular separable boundary conditions $\mathcal{B}=\{y^{(i)}(0)=0,\ i=0,1,\cdots,n-2;\ y^{(n-1)}(1)=0\}$. The Green's function is nonnegative, and the fundamental polynomials are

$$u_i^\circ(t)=(t^i-t^{n-1})/i!,\qquad i=0,1,\cdots,n-2;\quad u_{n-1}^\circ(t)=t^{n-1}.$$

The first eigenvalue λ_0 is the smallest (in absolute value) solution of

$$\sum_{k=0}^{\infty}\frac{(-1)^k\lambda^{(nk+n-1)/n}}{(nk+n-1)!}=0.$$

It is positive, and the corresponding eigenfunction y_0 is positive in $(0,\lambda_0)$. The biorthogonal elements are $z_\nu(t)=a_\nu y_\nu(1-t)$, $\nu=0,1,\cdots$. We note also that z_0 dominates the other eigenfunctions.

Moreover, Leeming and Sharma proved that a completely convex function with respect to this set-up, which they call a completely W_p-convex function, is entire (this result in the even-order case is due to Boas and Polya [4]). Hence, we may apply now the method based on the Khromov and Keldysh results (see § 2) and conclude that each completely W_p-convex function with respect to $(L,\mathcal{B})$ admits of a convergent eigenfunction expansion.

REFERENCES

[1] D. AMIR AND Z. ZIEGLER, *On generalized absolutely monotone functions*, Israel J. Math., 7 (1969), pp. 137–146.

[2] S. N. BERNSTEIN, *Leçons sur les propriétés extrémales et la Meilleure Approximation des Fonctions Analytiques d'une Variable rèelle*, Gauthier-Villars, Paris, 1926.

[3] R. P. BOAS, *Representations of completely convex functions*, Amer. J. Math., 81 (1959), pp. 709–714.

[4] R. P. BOAS AND G. POLYA, *Influence of the signs of the derivatives of a function on its analytic character*, Duke Math. J., 9 (1942), pp. 406–424.

[5] J. D. BUCKHOLTZ AND J. K. SHAW, *Generalized completely convex functions and Sturm–Liouville operators*, this Journal, 6 (1975), pp. 812–828.

[6] V. GANTMACHER AND M. G. KREIN, *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, Mir Publishing Co., Moscow, 1950.

[7] S. KARLIN, *The existence of eigenvalues for integral operators*, Trans. Amer. Math. Soc., 113 (1964), pp. 1–17.

[8] ———, *Total Positivity*, Stanford University Press, Stanford, CA, 1968.

[9] ———, *Total positivity, interpolation by splines, and Green's functions of differential operators*, J. Approximation Theory, 4 (1971), pp. 91–112.

[10] S. KARLIN AND Z. ZIEGLER, *Generalized absolutely monotone functions*, Israel J. Math., 3 (1965), pp. 173–180.

[11] ———, *Some inequalities for generalized concave functions*, J. Approximation Theory, 13 (1975), pp. 276–293.

[12] J. KARON, *The sign-regularity properties of a class of Green's functions for ordinary differential equations*, J. Differential Equations, 6 (1969), pp. 484–502.

[13] M. V. KELDYSH, *On eigenvalues and eigenfunctions of some classes of non self-adjoint equations*, Dokl. Akad. Nauk SSSR (N.S.), 77 (1951), pp. 11–14. (In Russian.)

[14] G. M. KESSELMAN, *On the absolute convergence of expansions in eigenfunctions of certain differential operators*, Izv. Vysž. SSSR Mat., 2 (1964), pp. 82–93. (In Russian.)

[15] A. P. KHROMOV, *Eigenfunction expansions for ordinary linear differential operators with non regular separable boundary conditions*, Mat. Sb. (N.S.), 70, 112 (1966), pp. 310–329. (In Russian.)

[16] D. LEEMING AND A. SHARMA, *A generalization of the class of Completely Convex Functions*, Inequalities III, O. Shisha, ed., Academic Press, New York, 1972, pp. 177–199.

[17] V. P. MIHAILOV, *Riesz bases in* $L_2(0, 1)$, Soviet Math. Dokl., 3 (1962), pp. 851–855.

[18] M. A. NAIMARK, *Linear Differential Operators*, vol. 1, Frederick Ungar, New York, 1967.

[19] S. PETHE AND A. SHARMA, *Modified Abel expansion and a subclass of completely convex functions*, this Journal, 3 (1972), pp. 546–558.

[20] M. H. PROTTER, *A generalization of completely convex functions*, Duke Math. J., 24 (1957), pp. 205–213.

[21] J. K. SHAW, *Series expansions and linear differential operators*, this Journal, 7 (1976), pp. 311–331.

[22] D. V. WIDDER, *Completely convex functions and Lidstone series*, Trans. Amer. Math. Soc., 51 (1942), pp. 387–398.

[23] ———, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1941.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0020 $0.100/0

HEUN'S EQUATION AND THE HYPERGEOMETRIC EQUATION*

KATHRYN KUIKEN†

Abstract. The present investigation determines under which circumstances Heun's equation can be derived from the hypergeometric equation by rational substitutions.

The hypergeometric equation has precisely three regular singular points. Heun's equation has precisely four regular singular points. One can therefore pose the question as to whether one of these equations can be transformed into the other via a rational transformation. The purpose of this note is to derive the following surprising conclusion: The hypergeometric equation can be transformed into a nontrivial Heun's equation via a rational transformation if and only if this transformation is one of six quadratic polynomials. These polynomials as well as the conditions imposed on the parameters for such a transformation to exist are tabulated explicitly.

Every homogeneous linear second order differential equation with four regular singularities can be transformed into

$$\frac{d^2u}{dt^2}+\left(\frac{\gamma}{t}+\frac{\delta}{t-1}+\frac{\varepsilon}{t-d}\right)\frac{du}{dt}+\frac{\alpha\beta t-q}{t(t-1)(t-d)}u=0 \tag{1}$$

where

$$\alpha+\beta-\gamma-\delta-\varepsilon+1=0. \tag{2}$$

This is Heun's equation. It contains a large number of interesting special cases, particularly Lamé's equation [1, Vol. 3, Chap. 15, pp. 44–90, esp. pp. 57–62]. There exist several theorems which provide expansions of the solutions of (1) in terms of the solutions of the hypergeometric equation

$$z(1-z)\frac{d^2y}{dz^2}+[c-(a+b+1)z]\frac{dy}{dz}-aby=0. \tag{3}$$

(For results and literature, see [1, Vol. 1, Chap. 2, pp. 56–119, esp. p. 56, and as cited above].) Here, we shall answer the question: When can a hypergeometric equation be transformed into Heun's equation by a transformation

$$z=R(t), \tag{4}$$

where R is a rational function of t?

We shall exclude the trivial case for which $\alpha\beta=q=0$. In this case, Heun's equation has obvious and elementary solutions. We shall prove the following

THEOREM. *A hypergeometric equation* (3) *can be transformed into a nontrivial Heun's equation* (1) *by a rational transformation* (4) *if and only if the points* 0, 1, ∞, d *form a harmonic quadruplet* (*that is, if d has one of the values* $-1, \frac{1}{2}, 2$) *and R is one of the following quadratic polynomials*

$$R=t^2,\quad 1-t^2,\quad (t-1)^2,\quad 2t-t^2,\quad (2t-1)^2,\quad 4t(1-t), \tag{5}$$

where the six parameters $\alpha, \beta, \gamma, \delta, \varepsilon, q$ *of* (1) *depend in an explicitly computable form on the three parameters* a, b, c *of* (3).

* Received by the editors April 25, 1977, and in revised form December 19, 1977.
† Department of Mathematics, Polytechnic Institute of New York, Brooklyn, New York 11201.

Proof. We show first that R must be a polynomial. The transformation of (3) into (1) produces the condition

$$-ab\frac{\dot{R}^2}{R(1-R)}=\frac{\alpha\beta t-q}{t(t-1)(t-d)}\qquad\left(\dot{R}=\frac{dR}{dt}\right)\tag{6}$$

on R. Letting $R=1-S$, we see that

$$\frac{\dot{R}}{R}\frac{\dot{S}}{S}=\frac{c_0}{t}+\frac{c_1}{t-1}+\frac{c_2}{t-d}.\tag{7}$$

On the other hand, both $\dot{R}/R$ and $\dot{S}/S$ are sums of terms of the form $n(t-\lambda)^{-1}$, where n is an integer and λ is a zero or a pole of R or of S. Suppose that λ were a pole of R. Then, it would also be a pole of S, and the left-hand side of (7) would produce a term of the form $m(t-\lambda)^{-2}$, where m is an integer, which is impossible.

We now see from (7) that 0, 1, d must be the total collection of zeros of R and S. We see immediately that a k-fold zero of R implies the existence of k distinct zeros of S and vice versa. Also, we see that R and S have no common zeros. Therefore, both R and S must be of degree two and one of them must be a square of a first degree polynomial.

The polynomials of (5) with the corresponding values of d are simply obtained by observing that 0, 1, d are the total collection of zeros of R and S. The values of d are given as -1, -1, 2, 2, $\frac{1}{2}$, $\frac{1}{2}$ respectively.

The remainder of the proof of the theorem consists of elementary but tedious calculations. We give below a list of the expressions for $\alpha+\beta$, $\alpha\beta$, γ, δ, ε, q for the various cases listed in (5) as functions of a, b, c. This list comes into existence by substituting each R of (5) into the conditions (6) and

$$-\frac{\ddot{R}}{\dot{R}}+\frac{\dot{R}}{R(1-R)}[c-(a+b+1)R]=\frac{\gamma}{t}+\frac{\delta}{t-1}+\frac{\varepsilon}{t-d}\tag{8}$$

produced by the transformation of (3) into (1) by (4) and by using (2). It should be noted that in all cases the numerator of the right-hand side of (6) will cancel against one of the linear factors in the denominator to give $(\alpha\beta t-q)/[t(t-1)(t-d)]=1/(At^2+Bt+C)$.

$R=t^2$:
$$\begin{aligned}\alpha+\beta&=2a+2b\\ \alpha\beta&=4ab\\ \gamma&=-1+2c\\ \delta&=1+a+b-c\\ \varepsilon&=\delta=1+a+b-c\\ q&=0\end{aligned}$$

$R=1-t^2$:
$$\begin{aligned}\alpha+\beta&=2a+2b\\ \alpha\beta&=4ab\\ \gamma&=1-2c+2a+2b\\ \delta&=c\\ \varepsilon&=\delta=c\\ q&=0\end{aligned}$$

$R=(t-1)^2$:
$$\begin{aligned}\alpha+\beta&=2a+2b\\ \alpha\beta&=4ab\\ \gamma&=1+a+b-c\\ \delta&=2c-1\\ \varepsilon&=\gamma=1+a+b-c\\ q&=4ab\end{aligned}$$

$R=2t-t^2$:
$$\begin{aligned}\alpha+\beta&=2a+2b\\ \alpha\beta&=4ab\\ \gamma&=c\\ \delta&=1+2a+2b-2c\\ \varepsilon&=\gamma=c\\ q&=4ab\end{aligned}$$

$R=(2t-1)^2$:

$$\begin{aligned}\alpha+\beta&=2a+2b\\ \alpha\beta&=4ab\\ \gamma&=1+a+b-c\\ \delta&=\gamma=1+a+b-c\\ \varepsilon&=-1+2c\\ q&=2ab\end{aligned}$$

$R=4t(1-t)$:

$$\begin{aligned}\alpha+\beta&=2a+2b\\ \alpha\beta&=4ab\\ \gamma&=c\\ \delta&=\gamma=c\\ \varepsilon&=1+2a+2b-2c\\ q&=2ab.\end{aligned}$$

This concludes the proof of the theorem.

Note. Although Heun's equation has the easily obtained solutions

$$u_1=c,\qquad u_2=\int\exp\left(-\int A\right)\quad\text{where } A=\frac{\gamma}{t}+\frac{\delta}{t-1}+\frac{\varepsilon}{t-d}$$

in the trivial case, we observe that a hypergeometric equation (3) can be transformed into a trivial Heun's equation (1) for which $\alpha\beta=q=0$ if and only if R is one of the following thirty rational functions of order two

$$\frac{t^2}{2(t-\frac12)},\ \frac{t^2}{(t-1)^2},\ \frac{2(t-\frac12)}{t^2},\ -\frac{(t-1)^2}{2(t-\frac12)},\ -\frac{2(t-\frac12)}{(t-1)^2},\ \frac{(t-1)^2}{t^2},$$

$$\frac{(t-\frac12)^2}{t(t-1)},\ -\frac{1}{4t(t-1)},\ \frac{1}{4(t-\frac12)^2},\ \frac{t(t-1)}{(t-\frac12)^2},\ \frac{t^2}{(t-1)(t+1)},$$

$$-\frac{1}{(t-1)(t+1)},\ \frac{(t-1)(t+1)}{t^2},\ \frac{1}{t^2},\ -\frac{(t-1)^2}{4t},\ \frac{(t-1)^2}{(t+1)^2},$$

$$\frac{(t+1)^2}{4t},\ \frac{4t}{(t+1)^2},\ \frac{(t+1)^2}{(t-1)^2},\ -\frac{4t}{(t-1)^2},\ \frac{(t-2)^2}{-4(t-1)},\ -\frac{4(t-1)}{(t-2)^2},$$

$$\frac{t^2}{4(t-1)},\ \frac{t^2}{(t-2)^2},\ \frac{(t-2)^2}{t^2},\ \frac{4(t-1)}{t^2},\ \frac{(t-1)^2}{t(t-2)},\ -\frac{1}{t(t-2)},$$

$$\frac{t(t-2)}{(t-1)^2},\ \frac{1}{(t-1)^2}$$

and the points 0, 1, ∞, d form a harmonic quadruplet. Again, the connection between the values α, β, γ, δ, ε, $q=0$ of (1) and the parameters a, b, c of (3) can be calculated explicitly by use of (2), (6) and (8). One of these cases appears in the book by C. Snow [2, Eq. (9d), p. 91].

REFERENCES

[1] A. ERDÉLYI ET AL., *Higher Transcendental Functions*, (Bateman Manuscript Project), Vols. 1 and 3, McGraw-Hill, New York, 1953, 1955.

[2] CHESTER SNOW, *Hypergeometric and Legendre Functions with Applications to Integral Equations of Potential Theory*, Department of Commerce, National Bureau of Standards, Applied Mathematics Series, 19, United States Government Printing Office, Washington, D.C., 1952.

SIAM J. MATH. ANAL.
Vol. 10, No. 3, May 1979

0036-1410/79/1003-0021 $01.00/0

A ONE-PARAMETER FAMILY OF SEQUENCE TRANSFORMATIONS*

JEAN-MARC VANDEN BROECK† AND LEONARD W. SCHWARTZ†

Abstract. A one-parameter family of sequence or series transformations is presented which includes the Padé and the iterated Aitken's (Shanks) tables as special cases. The transformation table elements are rational functions with polynomial coefficients in the parameter that, when expanded for small argument, are asymptotically equivalent to the highest order partial sum from which they are formed. Numerical examples using some standard series demonstrate dramatic acceleration of convergence and the ability to move branch cuts by choosing the parameter appropriately.

1. Introduction. Summation of power series is a topic of wide interest in applied mathematics. When the general term in a power series is known, the search for efficient and powerful summation techniques falls within the realm of approximation theory. Alternatively, solutions to nonlinear differential equations arising in mechanics or elsewhere can be expressed as perturbation expansions in an independent variable or parameter. Here the general term is not known but typically 10 to 100 terms can be found using a computer. The computation of additional terms is precluded by time and storage limitation or by the accumulation of roundoff errors. In the happiest of circumstances, the high-order series coefficients are nearly proportional to the coefficients in the expansion of a standard function. Thus the remaining unknown coefficients can be replaced by those derived from the standard function. This method of "series completion", as advocated by Van Dyke (1974) and Guttmann (1975), can yield accurate approximate solutions over much of the range of interest.

Often, however, no such simple structure is apparent and a more automatic method of series summation is required. We also need a method which can analytically continue the solution into regions where the original series is divergent. Consider the formal power series expansion of a function $f(x)$

$$f(x) \sim \sum_{i=0}^{\infty} a_i x^i \tag{1.1}$$

and its associated sequence of partial sums

$$A_N = \sum_{i=0}^{N} a_i x^i \equiv [N, 0], \qquad N = 0, 1, \cdots. \tag{1.2}$$

By applying a succession of nonlinear transformations a table of approximants to $f(x)$ can be generated where the original sequence forms the first column. Thus

$$\begin{array}{lllll} [0,0] & & & & \\ [1,0] & [1,1] & & & \\ [2,0] & [2,1] & [2,2] & & \\ [3,0] & [3,1] & \cdot & \cdot & \\ [4,0] & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \quad\cdot \\ \cdot & \cdot & \cdot & & \quad\;\cdot \\ \cdot & \cdot & \cdot & & \quad\quad\cdot \end{array} \tag{1.3}$$

---

* Received by the editors January 31, 1978. This work has been supported by the Australian Research Grants Committee.

† Applied Mathematics Department, University of Adelaide, South Australia.

where, hopefully, each column is more rapidly convergent than its predecessors or, if all columns are divergent, sequences of diagonals will be convergent.

The Padé table is the most widely-studied transformation of the type (1.3). The $[N, M]$ element, now usually represented by the "mnemonic" notation $[N/M]$, is a ratio of two polynomials in x where the degree of the numerator is no greater than N and that of the denominator no greater than M. The coefficients are determined by the requirement that each rational fraction, when expanded for small x, agree with the original series to order x^{M+N}. The coefficients for a given approximant can be found as the solution to $N+M+1$ linear equations or by the computationally more efficient algorithms of Rutishauser (1954) or Wynn (1956). Padé approximants have been studied extensively and convergence of the table or certain subsequences extracted from it has been established for restricted classes of functions. Much of the recent theory is summarized in the book by Baker (1975). The theory as developed, however, is of little use when only a limited number of coefficients are known. Then (1.3) is a finite triangular array and "convergence" refers to the numerical agreement between table elements and their neighbors.

Another table of the form (1.3) can be generated by the iterated Aitken's or Shanks transformation. Each table element is formed from its three adjacent predecessors according to the formula

$$[N, M+1]=\frac{[N, M]^2-[N-1, M][N+1, M]}{2[N, M]-[N-1, M]-[N+1, M]}. \tag{1.4}$$

Virtually all the theory concerning (1.4) can be found in the papers by Lubkin (1952) and Shanks (1955). Its use is motivated by the observation, that if the original sequence elements are in geometric progression

$$[N, 0]=B+cq^N, \qquad q\neq 0, 1,$$

then one application of (1.4) to any three successive elements will yield the "correct" answer B by "filtering out" the "transient" cq^N. Thus if the columns $[N, 0], [N, 1], \cdots$ are "nearly geometric" sequences the table can be expected to converge rapidly. Note that the first derived sequence $[N, 1]$ is identical to the Padé sequence $[N/1]$. If $[N, 0]$ are the partial sums of a power series, then the elements $[N, M]$, $M=2, 3, \cdots$ are rational functions of x but with numerators and denominators of higher order than in the Padé table. The paucity of theoretical results notwithstanding, the Shanks transformation has been used successfully to accelerate the convergence of series solutions to a number of physically motivated problems. See, for example, Fenton (1972) and Schwartz (1974).

In the next section we introduce a generalized sequence transformation which includes the Padé and Shanks tables as special cases. A fundamental property of this transformation is established concerning the order of agreement of the table elements with the original power series. The following section contains several numerical applications.

2. The family of transformations. Let the table of approximants (1.3) be generated, column by column, by the formula

$$\frac{1}{[N, M+1]-[N, M]}+\frac{\alpha}{[N, M-1]-[N, M]} = \frac{1}{[N+1, M]-[N, M]}+\frac{1}{[N-1, M]-[N, M]} \tag{2.1}$$

subject to the auxiliary condition $[N, -1]=\infty$. For $\alpha=1$, (2.1) becomes Wynn's (1966) five-term identity which is an efficient method for generating the Padé table. If $\alpha=0$, on the other hand, (2.1) reduces to (1.4), the Shanks transformation. We consider now arbitrary values of α and for definitiveness will use the notation

$$[N, M]^{(\alpha)} f(x)$$

to signify a table element formed from the series expansion for f according to (2.1) with a given value of α.

A number of properties of the approximants can be readily established. In each case existence of the relevant portion of the table is assumed.

(1) If $[N, 0]$ are the partial sums of a power series in x, each element $[N, M]^{(\alpha)}$ will be a rational function of x with coefficients that are polynomials in α. This property follows immediately from the observation that (2.1) involves only the four elementary arithmetic operations.

(2) If $[N, 0]$ are the partial sums of a power series in x, the element $[N, M]^{(\alpha)}$ involves the knowledge of only the first column members $[0, 0]\,[1, 0], \cdots, [N+M, 0]$ and, subject to a "normality condition", when expanded for small x, $[N, M]^{(\alpha)}$ agrees with the original series to order x^{M+N}. This property may be established by induction. Note first that the elements in the $[N, 0]$ and $[N, 1]$ are independent of α. Hence these elements are identical to $[N/0]$ and $[N/1]$ respectively and possess the required property by definition. Assume now that elements in the $(M-1)$th and Mth columns have the Taylor series expansions

$$\begin{aligned}[N, M-1] &= \sum_{i=0}^{N+M-1} a_i x^i + \sum_{i=N+M}^{\infty} c_i^{(N,M-1)} x^i, \\ [N, M] &= \sum_{i=0}^{N+M} a_i x^i + \sum_{i=N+M+1}^{\infty} c_i^{(N,M)} x^i\end{aligned} \tag{2.2}$$

where a_i are the coefficients in the original series. The denominators of the 2nd, 3rd and 4th terms in (2.1) become

$$\begin{aligned}[N, M-1]-[N, M] &= (c_{N+M}^{(N,M-1)} - a_{N+M})x^{N+M} + O(x^{N+M+1}), \\ [N+1, M]-[N, M] &= (a_{N+M+1} - c_{N+M+1}^{(N,M)})x^{N+M+1} + O(x^{N+M+2}) \end{aligned} \tag{2.3}$$

and $$[N-1, M]-[N, M] = (c_{N+M}^{(N-1,M)} - a_{N+M})x^{N+M} + O(x^{N+M+1}).$$

Provided now that

$$c_{N+M}^{(N-1,M)} \neq a_{N+M} \quad \text{and} \quad c_{N+M}^{(N,M-1)} \neq a_{N+M}, \tag{2.4}$$

substitution of (2.3) into (2.1) yields

$$\begin{aligned}[N, M+1] &= [N, M] + (a_{N+M+1} - c_{N+M+1}^{(N,M)})x^{N+M+1} + O(x^{N+M+2}) \\ &= \sum_{i=0}^{N+M+1} a_i x^i + O(x^{n+M+2)})\end{aligned} \tag{2.5}$$

which is of the required form (2.2). The restrictions (2.4) are the normality conditions. When $\alpha=1$, they are equivalent to the normality conditions for the Padé Table, discussed by Baker (1975, p. 24) which ensure that every Padé approximant formed from a given series exists. It is sufficient for our purposes to realize that (2.4) would only be violated in exceptional circumstances since the table elements $[N-1, M]$ and $[N, M-1]$ are not functions of a_{N+M}. The result (2.5) implies that an approximant

$[N, M]^{\alpha}(f)$ is no worse an estimate of f than the highest-order partial sum from which it is formed, and that any improved convergence for large values of argument is not purchased at the expense of reduced accuracy near the origin. In the case of the Shanks Table $[N, M]^{\alpha=0}$, this important property does not appear to have been recognized previously.

(3) The series transformation $[N, M]^{(\alpha)}(f)$ is nonlinear in the sense that

$$[N, M]^{\alpha}(f+g) \neq [N, M]^{\alpha}(f)+[N, M]^{\alpha}(g)$$

for any two series f and g. However, the rules

$$[N, M]^{\alpha}(f+C)=C+[N, M]^{\alpha}(f) \tag{2.6}$$

and

$$[N, M]^{\alpha}(Kf)=K[N, M]^{\alpha}(f) \tag{2.7}$$

where K and C are arbitrary constants, are valid for any value of α. Observe that (2.1) is invariant under either transformation. On the other hand, two Padé diagonal invariance properties, namely the reciprocal formula

$$[N/N](1/f(x))=\frac{1}{[N/N](f(x))}$$

and invariance of the $[N/N]$ under the Euler transformation

$$x=\frac{Ay}{1+By}$$

are not valid for any value of α other than 1.

(4) Let $[N, 0]$ be the partial sums of a power series with real coefficients a_i. If α is also real, the approximants have the property

$$[N, M]^{\alpha}f(\bar{x})=\overline{[N, M]^{\alpha}f(x)}, \tag{2.8}$$

where the bar signifies complex conjugation. Any eventual limit must also have this property. Thus, for example, the table of approximants formed from the Maclaurin series expansion for either $\log_e(k+x)$ or $\sqrt{k+x}$, $k>0$ cannot be expected to converge on the portion of the real axis $-\infty<x<-k$. The high-order approximants will simulate this choice of branch cut by placing many poles and zeros along this line. With another branch cut, not lying along the real axis, either function is regular, albeit complex-valued, for $-\infty<x<-k$. In the next section we will present numerical evidence to illustrate that, by taking a *complex* value for α, the cut can be moved off the real axis for such functions.

It is interesting to note that the validity of the above properties is unaffected if the transformation (2.6) is made more general. A different value of α may be used to generate each column in (1.3). The general table element can now be represented as

$$[N, M]^{(\alpha_2, \alpha_3, \cdots, \alpha_M)}f(x), \tag{2.9}$$

where α_k is used to form the kth column. For example, Shanks (1955) shows that rapid convergence for certain infinite sequences can be obtained by repeated use of the Padé $[N/2]$ transformation. In the notation of (2.9) the superscript, specifying the order of procedure, would be $(1, 0, 1, \cdots, 0, 1)$ where M is necessarily an even number.

A form of (2.1) more suitable for automatic computation can be formulated as a generalization of Wynn's (1956) "ε-algorithm". A new column is inserted between each column of approximants so that the augmented table, replacing (1.3), is

$$
\begin{array}{ccccccccc}
[0,0] & & & & & & & & \\
& \varepsilon_{00} & & & & & & & \\
[1,0] & & [1,\,] & & & & & & \\
& \varepsilon_{10} & & \varepsilon_{11} & & & & & \\
[2,0] & & [2,1] & & [2,2] & & & & \\
& \varepsilon_{20} & & \varepsilon_{21} & \cdot & \cdot & & & \\
& \cdot & & \cdot & \cdot & & \cdot & & \\
[3,0] & \cdot & [3,1] & \cdot & \cdot & & \cdot & & \\
& \cdot & & \cdot & \cdot & & & \cdot & \\
\cdot & \cdot & \cdot & \cdot & \cdot & & & & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & & & & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & & & & \cdot
\end{array} \tag{2.10}
$$

Here the elements are generated by the formulas

$$\varepsilon_{N,M} = \alpha\varepsilon_{N,M-1} + \frac{1}{[N+1, M]-[N, M]}, \tag{2.11a}$$

$$[N, M+1] = [N, M] + \frac{1}{\varepsilon_{N,M} - \varepsilon_{N-1,M}}, \tag{2.11b}$$

subject to the initial condition $\varepsilon_{N,-1} = 0$. Relations (2.11) are equivalent to (2.1). This ε-algorithm formulation reduces computation time by about one-half at the expense of increased storage requirements.

3. Numerical examples. A dramatic example of rapid convergence may be obtained from the famous series for $\zeta(2)$:

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\frac{\pi^2}{6}=1.64493406\ldots. \tag{3.1}$$

This series is slowly convergent; in fact the sum of the first 61 terms in (3.1) agrees with $\pi^2/6$ to only 2 places. A standard asymptotic analysis, using the integral form of the generalized Riemann zeta function, reveals that 60 million terms would be required for 8-place accuracy. Neither the Padé ($\alpha = 1$) or Shanks ($\alpha = 0$) tables converge well; the [30, 30] approximants formed from 61 terms in the expansion are correct to only 3 and 4 places respectively. Since the $[N, M]^{\alpha=0}$ are more rapidly convergent than the $[N, M]^{\alpha=1}$, however, negative values of α are suggested. Table 1 is the array of approximants to (3.1), with $\alpha = -1$, constructed from the first 11 partial sums. Note that the [4, 5] element is already correct to 8 places. Aside from the second column, which is common to all tables, the array converges rapidly to the correct answer.

Our second and third examples are derived from the divergent expansion

$$1-1!z+2!z^2-\cdots+(-1)^n n!z^n+\cdots\sim\frac{1}{z}e^{1/z}E_1\left(\frac{1}{z}\right), \tag{3.2}$$

where E_1 is the exponential integral function as defined by Abramowitz and Stegun (1964, p. 228). The diagonal Padé approximants formed from this series can be shown to converge to the right side of (3.2) for real positive z (Baker, 1975, p. 72). Numerical tests show that the low-order approximants in the Shanks table also converge and much

TABLE 1
The table of approximants ($\alpha=-1$) formed from the first eleven partial sums of (3.1).

1.0000000					
1.250000	1.4500000				
1.361111	1.503968	1.6470970			
1.423611	1.534722	1.6459065	1.6451849		
1.463611	1.554520	1.6454554	1.6450669	1.6449342	
1.491389	1.568312	1.6452462	1.6450131	1.6449341	1.6449341
1.511797	1.578464	1.6451358	1.6449850	1.6449341	
1.527422	1.586246	1.6450720	1.6449688		
1.539768	1.592399	1.6450325			
1.549768	1.597387				
1.558032					

more rapidly. For real values of α between 0 and 1, the convergence improves more or less continuously as α decreases. As a specific example, for $z=1$, the series (3.2) should converge to $eE_1(1)=.596347362\ldots$. In Fig. 1 we plot the common logarithm of the error for the diagonal approximants with $z=1$ versus the parameter α. Note that the succession of approximants [2, 2], [4, 4], [6, 6], $\cdots$ would appear to converge to the correct answer for any α in this range. Values of α near zero appear to give about twice the accuracy of the Padé approximants. Because this calculation involves the differences of very large numbers, accuracy is ultimately limited by computer roundoff error. In general, negative values of α produced "disorderly" tables; however the choice $\alpha=-1\pm i$ produced a table which converged slightly faster than any real value tested.

The diagonal approximants $[N, N]^\alpha$, $\alpha\in[0,1]$ appear to converge to the sum indicated in (3.2) so long as $|\arg z|\neq n\pi$, $n=1,3,5,\cdots$. However, for negative z and real values of α, the approximants will be real while $E_1(-x)$, $x>0$, is necessarily complex. That is the α-real approximants choose the branch cut of the exponential integral function to lie along the negative real axis. Thus the divergent series

$$(3.3)\qquad 1+1!+2!+\cdots+n!+\cdots\sim-\frac{1}{e}E_1(-1\pm i0)\doteq.69717\pm i1.1557,$$

if summable, will have a positive or negative imaginary part according to whether the branch cut lies below or above the negative real axis. Table 2 shows the diagonal approximants found from (3.3) for 3 imaginary values of α. Each column appears to converge to $-e^{-1}E_1(-1+i0)$ indicating that the branch cut has been moved into the third quadrant. The convergence seems best for $\alpha=i$; the element $[14,14]^{\alpha=i}$, for example, is correct to within one part in 1000. In general the convergence is much poorer than for the alternating-sign series. The cut has only been moved a short distance from the real axis; proximity to the cut as well as the roundoff error limitation precludes greater accuracy. According to property (4) of the last section, replacing α by its conjugate would move the cut into the second quadrant.

The final example illustrates the amount by which a "natural" branch cut can be moved by selecting an imaginary value of α. Consider the simple series

$$(3.4)\qquad \log_e z\sim(z-1)-\frac{(z-1)^2}{2}+\frac{(z-1)^3}{3}-\cdots$$

and the approximants formed from it on the circle $z=4\,e^{i\theta}$. The approximants may be expected to converge to $\log_e 4+i\theta$ if $0\leqq\theta<\theta^*$ and to $\log_e 4+i(\theta-2\pi)$ if $\theta^*<\theta\leqq 2\pi$,

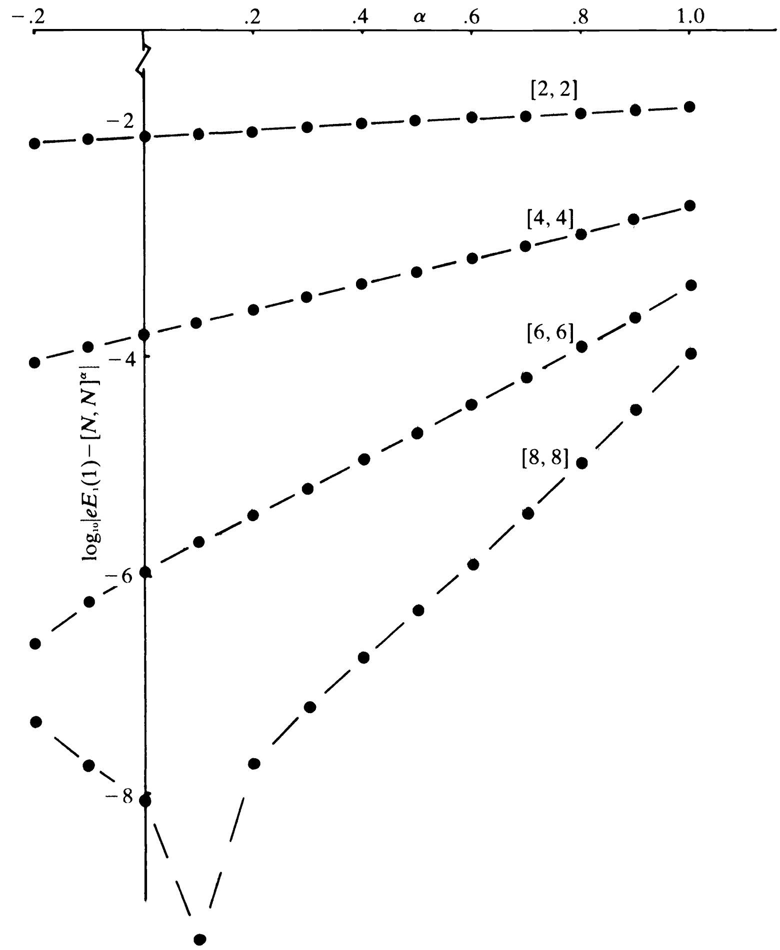

FIG. 1. *The error of diagonal approximants to* $\sum_{n=0}(-1)^n n!$ *vs.* α.

TABLE 2
Diagonal approximants to the series $\sum_{n=0} n!$.

N	$[N,N]^{\alpha=0.5i}$	$[N/N]^{\alpha=0.8i}$	$[N,N]^{\alpha=1.0i}$
1	$0+i0$	$0+i0$	$0+i0$
2	$1+6.0i$	$1+i3.75$	$1+i3.000$
3	$.8867+i.5003$	$.7835+i.6902$	$.73090+i.77631$
4	$1.0214+i1.7240$	$.9579+i1.2748$	$.9144+i1.1402$
5	$1.1418+i1.5615$	$.9388+i1.1698$	$.8500+i1.1021$
6	$1.5083+i1.1358$	$.7355+i.9746$	$.6623+i1.1067$
7	$.7384+i1.1898$	$..6979+i1.1057$	$.6913+i1.1192$
8	$.8333+i1.1845$	$.7018+i1.1127$	$.6828+i1.1455$
9	$.7361+i1.2037$	$.6772+i1.1518$	$.6938+i1.1478$
10	$.7366+i1.1741$	$.6961+i1.1509$	$.6951+i1.1537$
11	$.7318+i1.1769$	$.6995+i1.1519$	$.6944+i1.1513$
12	$.7192+i1.1588$	$.6994+i1.1553$	$.6993+i1.1552$
13	$.7193+i1.1594$	$.6995+i1.1561$	$.6976+i1.1563$
14	$.7051+i1.1533$	$.6965+i1.1558$	$.6977+i1.1565$

where θ^* specifies the angular position of the branch cut of $\log_e z$ when $|z| = 4$. In Fig. 2 we plot the imaginary components of typical approximants to the series (3.4) versus $\theta = \arg z$. For $\alpha = 0.5i$, values are shown for the imaginary parts of the [10, 10] and [20, 20] elements. Convergence is good, and the elements lie close to either the upper or lower branch values (dashed lines), except near the cut which lies at $\theta^* \approx 3.0$. For purposes of comparison, values of the Shanks element $[20, 20]^{\alpha=0}$ are presented. These points lie close to the dashed lines except near the cut, $\theta^* = \pi$. Thus the branch cut appears to have been moved about 8 degrees from the negative real axis. Similar calculations, made for other values of $|z|$, indicate that the branch cut is not a straight line, but curves away from the real axis as $|z|$ is increased. The location of the cut is determined by the value of α, but the functional dependence is unclear.

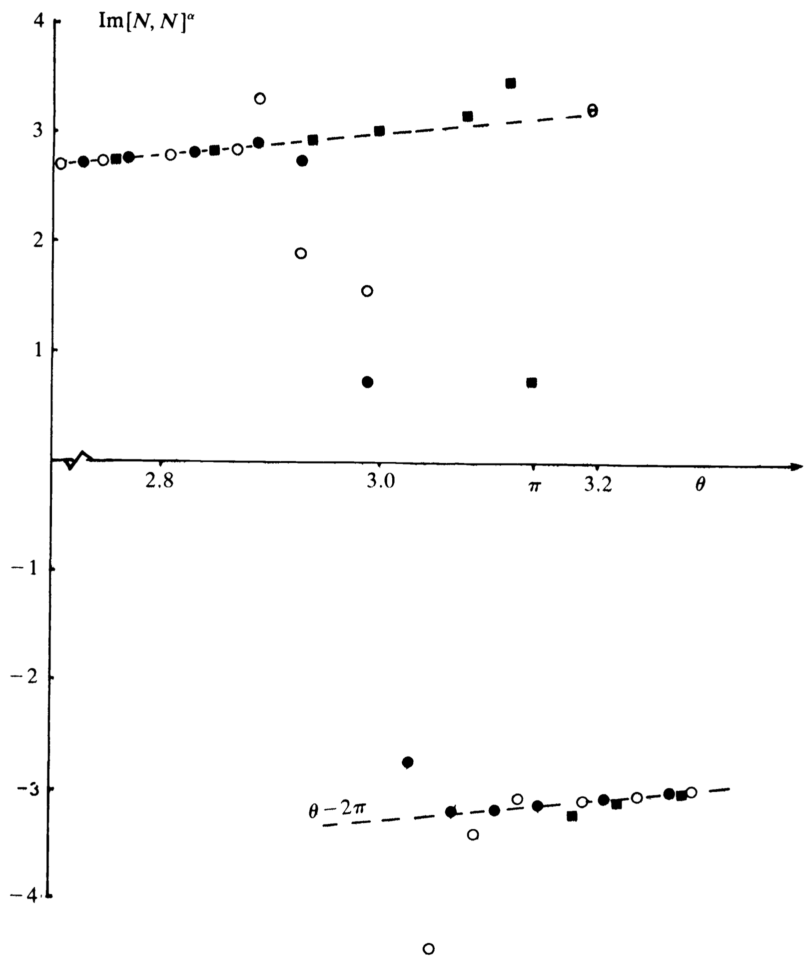

FIG. 2. *Imaginary part of* $[N, N]^\alpha$ $(\log_e 4\, e^{i\theta})$ *vs.* θ. ○ – $\alpha = 0.5i$, $N = 10$; ● – $\alpha = 0.5i$, $N = 20$; ■ – $\alpha = 0$, $N = 20$.

4. Concluding remarks. The family of transformations discussed should serve to supplement the arsenal of techniques available to the analyst who chooses to solve nonlinear problems by perturbation series expansions. To the extent that each transformation element $[N, M]^\alpha$ reproduces the series from which it is formed to appropriate order when expanded for small argument, it would appear worthwhile to investigate various choices of α in order to secure optimum convergence in a given problem. Choosing α to be complex so as to move the branch cut, is an attractive possibility where the natural cut of the Padé or Shanks tables leads to physically unacceptable discontinuities. For example, in the recent work of Vanden Broeck and Tuck (1977) on the near-stern flows past barge-like bodies, use of the Shanks table yielded converged but discontinuous free-surface profiles.

Other areas of application include the efficient approximation of known functions; our first example suggests that $[N, M]^{\alpha=-1}$ applied to sums of reciprocal powers may provide good approximations for the zeta function. Secondly, the use of iterative methods in large scale computation yields sequences of approximations to solutions, often at formidable cost. Convergence-acceleration techniques based on the simple Shanks transformation (1.4) have been applied to transonic aerodynamics calculations by Martin (1976) and others. The present work may provide a useful generalization.

REFERENCES

M. ABRAMOWITZ AND I. STEGUN (1964), *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series no. 55, Washington, DC.

G. A. BAKER, JR. (1975), *Essentials of Padé Approximants*, Academic Press, New York.

J. FENTON (1972), *A ninth-order solution for the solitary wave*, J. Fluid Mech., 53, pp. 257–271.

A. GUTTMANN (1975), *Derivation of "Mimic Functions" from regular perturbation expansions in fluid mechanics*, J. Inst. Math. Applic., 15, pp. 307–317.

S. LUBKIN (1952), *A method of summing infinite series*, J. Res. Nat. Bur. Standards, 48, pp. 228–254.

E. D. MARTIN (1976), *A technique for accelerating iterative convergence in numerical integration, with application in transonic aerodynamics*, Lecture Notes in Physics No. 47, Springer-Verlag, Berlin, pp. 123–140.

I2.8HH. RUTISHAUSER (1954), *Der Quotienten–Differenzen–Algorithmus*, Z. Angew. Math. Phys. 5, pp. 233–251.

L. W. SCHWARTZ (1974), *Computer extension and analytic continuation of Stokes' expansion for gravity waves*, J. Fluid Mech., 62, pp. 553–578.

D. SHANKS (1955), *Non-linear transformations of divergent and slowly-convergent sequences*, J. Math. and Phys., 34, pp. 1–42.

J.-M. VANDEN BROECK AND E. O. TUCK (1977), *Computation of near-bow or stern flows, using series expansion in the Froude number*, Proc. 2nd. Int. Conf. Num. Ship Hydrodynamics, Berkeley, CA.

M. D. VAN DYKE (1974), *Analysis and improvement of perturbation series*, Quart. J. Mech. Appl. Math., 27, pp. 423–450.

P. WYNN (1956), *On a device for computing the $e_m(S_n)$ transformation*, Math. Tables and other aids to Comput., 10, pp. 91–96.

——— (1966), *Upon systems of recursions which obtain among the quotients of the Padé table*, Numer. Math., 8, pp. 264–269.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0001$01.00/0

SOME ASPECTS OF OSCILLATION AND BOUNDARY PROBLEM THEORY FOR HAMILTONIAN SYSTEMS IN A B^*-ALGEBRA*

WILLIAM T. REID†

Abstract. For linear self-adjoint Hamiltonian differential systems in a B^*-algebra the topics treated include an extension of the well-known generalized polar coordinate transformation for the finite dimensional matrix case, and the derivation of two oscillation criteria for systems that may be written as self-adjoint linear differential equations of the second order. The determination of a Green's matrix for an incompatible boundary problem involving two-point boundary conditions is discussed, and, in particular, the established results are applied to reduce a certain type of vector boundary problem in a Hilbert space to known results for symmetrizable compact linear transformations in an associated Hilbert space.

1. Introduction. Within recent years there have been various extensions of oscillation theory for finite dimensional differential systems to corresponding vector and operator equations in abstract spaces—notably in Banach spaces [6], [7], [8], [11], [12], and particularly in the context of B^*-algebras, [3], [4], [5], [10], [19]. For linear self-adjoint Hamiltonian systems it is the purpose of the present paper to survey some of the difficulties that have been encountered, to establish some related criteria for oscillation, and to discuss some associated two-point boundary problems.

Section 2 is devoted to presenting some of the basic properties of Hamiltonian systems, and in § 3 there is given an extension to such systems of the generalized polar coordinate transformation initially established for finite dimensional matrix systems by Barrett [1] and the author [14]. For systems that may be written as self-adjoint linear differential equations of the second order there are given in § 4 two oscillation criteria based upon the fact that the set of such equations for which the corresponding Dirichlet functional on arbitrary compact subintervals is nonnegative forms a nonnegative cone in a certain function space. The first criterion is essentially one considered by Etgen and Pawlowski [4] and Etgen and Lewis [5], while the second criterion involves the application of certain linear transformations corresponding to those considered by the author [17] in the case of finite dimensional matrix problems. For a self-adjoint boundary problem involving a Hamiltonian differential system and two-point boundary conditions the determination of a Green's matrix is discussed in § 5, and the obtained results are applied in § 6 to reduce the consideration of certain vector boundary problems in a Hilbert space to known results on symmetrizable compact linear transformations in an allied Hilbert space.

2. Basic properties of Hamiltonian systems. Let $\mathfrak{B}$ be a complex B^*-algebra with unit E and conjugation $(\cdot)^*$, (see [18; Ch. IV, §§ 7, 8, 9] and [10; Ch. 4]), and for I an interval on the real line consider the self-adjoint Hamiltonian differential system

$$(2.1^0)\qquad \begin{aligned} L_1[U, V](t) &\equiv -V'(t)+C(t)U(t)-A^*(t)V(t)=0,\\ L_2[U, V](t) &\equiv U'(t)-A(t)U(t)-B(t)V(t)=0, \end{aligned}\qquad t\in I.$$

Throughout the subsequent discussion it will be supposed that the following hypothesis is satisfied.

* Received by the editors July 6, 1977. An elaboration of the third lecture in the series of John Barrett Memorial Lectures, The University of Tennessee, Knoxville, Tennessee, May, 1977. We regret to report the death of Professor Reid in September, 1977.

† Department of Mathematics, University of Texas at Austin, Austin, Texas 78712.

($\mathfrak{H}_0$) *A, B and C are (strongly) continuous $\mathfrak{B}$-valued functions on I, with $B(t)$ and $C(t)$ hermitian (symmetric) for $t \in I$.*

In terms of $Y: I \to \mathfrak{B} \times \mathfrak{B}$ with $Y(t) = (U(t); V(t))$, system (2.1^0) may be written as

$$L[Y](t) \equiv \mathcal{J} Y'(t) + \mathcal{A}(t) Y(t) = 0, \qquad t \in I, \tag{2.1}$$

where

$$\mathcal{J} = \begin{bmatrix} 0 & -E \\ E & 0 \end{bmatrix}, \qquad \mathcal{A}(t) = \begin{bmatrix} C(t) & -A^*(t) \\ -A(t) & -B(t) \end{bmatrix}, \qquad t \in I, \tag{2.2}$$

are elements of the matrix algebra $\mathfrak{M}_2(\mathfrak{B})$ over $\mathfrak{B}$. For $Z = (X_1, X_2) \in \mathfrak{B} \times \mathfrak{B}$, the norm $\|Z\|$ is defined to be $[\|X_1\|^2 + \|X_2\|^2]^{1/2}$, and if $M = [M_{\alpha\beta}]$, $(\alpha, \beta = 1, 2)$, belongs to $\mathfrak{M}_2(\mathfrak{B})$ then we set

$$\begin{aligned} \|M\| &= \sup \{\|MZ\| : \|Z\| \leqq 1\} \\ &= \sup \{[\|M_{11}X_1 + M_{12}X_2\|^2 + \|M_{21}X_1 + M_{22}X_2\|^2]^{1/2} : \|X_1\|^2 + \|X_2\|^2 \leqq 1\}. \end{aligned}$$

In particular, $Y(t) = (U(t); V(t))$ is called a solution of (2.1) if U and V possess (strong) derivatives which are (strongly) continuous and satisfy this equation on I. If U_0 and V_0 are elements of $\mathfrak{B}$ and $\tau \in I$, then by well-known existence theorems (see, for example, [9, § 3.4], [10; Ch. 6]) there exists a unique solution $Y(t) = (U(t); V(t))$ of (2.1) on I satisfying the initial condition $Y(\tau) = (U_0; V_0)$. Moreover, if $Y_\alpha(t) = (U_\alpha(t); V_\alpha(t))$, $(\alpha = 1, 2)$, are solutions of (2.1) then

$$\{Y_1; Y_2\}(t) \equiv Y_2^*(t) \mathcal{J} Y_1(t) \equiv V_2^*(t) U_1(t) - U_2^*(t) V_1(t)$$

is constant on I. If the value of this constant is 0, then these solutions are said to be *(mutually) conjoined* or *conjugate*; in particular, a solution Y is *self-conjoined* if $\{Y; Y\} = 0$. For $\tau \in I$ special associated self-conjoined solutions are $Y_\tau^\circ(t) = (U_\tau^\circ(t); V_\tau^\circ(t))$ and $Y_\tau^1 = (U_\tau^1(t); V_\tau^1(t))$ determined by the initial conditions

$$Y_\tau^\circ(\tau) = (0; E); \; Y_\tau^1(\tau) = (E; 0). \tag{2.3}$$

Frequently one of the following additional hypotheses will be assumed.

($\mathfrak{H}_{\mathfrak{B}}$) *$B(t) \geqq 0$ for $t \in I$; that is,* Sp $B(t) \subset [0, \infty)$, *where* "Sp $B(t)$" *denotes the spectrum of $B(t)$.*

($\mathfrak{H}_{\mathfrak{B}}^+$) *$B(t) > 0$ for $t \in I$; that is,* Sp $B(t) \subset (0, \infty)$ *for $t \in I$.*

($\mathfrak{H}_{\mathfrak{N}}$) (2.1) *is identically normal on I; that is, if I_0 is a nondegenerate subinterval of I and $Y(t) = (U(t); V(t))$ is a solution of* (2.1) *with $U(t) \equiv 0$ on I_0 then also $V(t) \equiv 0$ on I_0, and hence $Y(t) \equiv (0; 0)$ on I.*

In particular, ($\mathfrak{H}_{\mathfrak{B}}^+$) implies ($\mathfrak{H}_{\mathfrak{B}}$) and ($\mathfrak{H}_{\mathfrak{N}}$).

One may consider the linear second order differential equation

$$[R(t)U'(t) + Q(t)U(t)]' - [Q^*(t)U'(t) + P(t)U(t)] = 0, \qquad t \in I, \tag{2.4}$$

as a special instance of (2.1) whenever the coefficient functions satisfy the following condition.

($\mathfrak{H}'$) *R, P, Q are $\mathfrak{B}$-valued functions on I which are (strongly) continuous, with $R(t)$ and $P(t)$ hermitian while $R(t) > 0$ for $t \in I$.*

Indeed, with $V(t)=R(t)U'(t)+Q(t)U(t)$, under $(\mathfrak{H}')$ equation (2.4) is equivalent to a system (2.1) with

$$(2.5)\qquad A(t)=-R^{-1}(t)Q(t),\qquad B(t)=R^{-1}(t),\qquad C(t)=P(t)-Q^*(t)R^{-1}(t)Q(t),$$

and hypotheses $(\mathfrak{H}_0)$, $(\mathfrak{H}_{\mathfrak{B}}^+)$ are satisfied.

In the finite dimensional matrix case, for a system (2.1) satisfying hypotheses $(\mathfrak{H}_0)$, $(\mathfrak{H}_{\mathfrak{B}})$, $(\mathfrak{H}_{\mathfrak{N}})$ one has the following property.

$(\mathfrak{P})$ *For $\tau\in I$ there exists a $\delta=\delta_\tau>0$ such that the $U_\tau^\circ(t)$ belonging to $Y_\tau^\circ(t)$, as specified in* (2.3), *is nonsingular on* $([\tau-\delta,\tau)\cup(\tau,\tau+\delta])\cap I$.

That this property does not hold for general systems satisfying the hypotheses given above is illustrated by the following example. Let H be the Hilbert space $H=l^2$, $I=(-\infty,\infty)$, B the Banach algebra of bounded linear transformations $L[H,H]$, and, as usual, identify the transformation with its matrix in terms of the canonical unit elements $e^{(\beta)}=(\delta_{\alpha\beta})$. Then for $B(t)\equiv[(\pi/n)\delta_{nm}]=-C(t)$, and $A(t)\equiv 0$, the solution $Y_\tau^\circ(t)$ of the corresponding (2.1) has $U_\tau(t)=[\delta_{nm}\sin(\pi/n)(t-\tau)]$, $V_\tau(t)=[\delta_{nm}\cos((\pi/n)(t-\tau))]$, and for all t we have $0\in\operatorname{Sp}U_\tau(t)$ so that $U_\tau(t)$ is singular for all t. This example is essentially that of Example 1.1 of [5], and attributed therein to R. T. Lewis and S. C. Tefteller. As given in [5], however, it is presented as an example for the second order equation with $R(t)\equiv E$, $Q(t)=0$, $P(t)=-[(\pi^2/n^2)\delta_{nm}]$, and in the context of the above discussion it is to be emphasized that the above determined $[\delta_{nm}\sin((\pi/n)(t-\tau))]$ is not the matrix $U_\tau(t)$ of the solution $Y_\tau(t)=(U_\tau(t);V_\tau(t))$ of the corresponding system (2.1) with coefficients given by (2.5). Rather, for this system the corresponding $Y_\tau(t)$ has $U_\tau(t)=[\delta_{nm}(n/\pi)\sin((\pi/n)(t-\tau))]$, $V_\tau(t)=[\delta_{nm}\cos((\pi/n)(t-\tau))]$, and in this case $U_\tau(t)$ is nonsingular for all t such that $t-\tau$ is not an integer, and for $t=\tau+k$, k a positive integer, one has that the null space of $U_\tau(\tau+k)$ is one-dimensional with basis vector $e^{(k)}$. The significance of these two considerations of systems which at first sight may be thought to be equivalent will be further highlighted by the result of (2.17) below.

It is to be noted that the theory of (2.1) is no more general than the theory of such equations wherein $A(t)\equiv 0$. Indeed, if $\tau\in I$ and $Z_1(t)=Z_1(t;\tau)$, $Z_2(t)=Z_2(t;\tau)$ are solutions of the respective initial value problems

$$(2.6)\qquad \begin{aligned} &Z_1'(t)-A(t)Z_1(t)=0, && Z_1(\tau)=E,\\ &Z_2'(t)+Z_2(t)A(t)=0, && Z_2(\tau)=E, \end{aligned}$$

then $\Omega_1(t)=Z_1(t)Z_2(t)$ and $\Omega_2(t)=Z_2(t)Z_1(t)$ are solutions of the respective systems $\Omega_1'(t)+\Omega_1(t)A(t)-A(t)\Omega_1(t)=0$, $\Omega_1(\tau)=E$ and $\Omega_2'(t)=0$, $\Omega_2(\tau)=E$, so that in view of the uniqueness of solutions of such systems we have $Z_2(t)Z_1(t)\equiv E\equiv Z_1(t)Z_2(t)$. Consequently, under the substitution

$$(2.7)\qquad U(t)=Z_1(t)U_0(t),\qquad V(t)=Z_2^*(t)V_0(t)$$

system (2.1) reduces to

$$(2.1_0)\qquad L[Y_0](t)\equiv\mathcal{J}Y_0'(t)+\mathcal{A}_0(t)Y_0(t)=0,\qquad t\in I,$$

where $\mathcal{A}_0(t)=\begin{bmatrix} C_0(t) & 0\\ 0 & -B_0(t)\end{bmatrix}$, with $C_0(t)=Z_1^*(t)C(t)Z_1(t)$ and $B_0(t)=Z_2(t)B(t)Z_2^*(t)$. In particular, hypothesis $(\mathfrak{H}_0)$ for (2.1) implies $(\mathfrak{H}_0)$ for (2.1_0), and an individual hypothesis $(\mathfrak{H}_{\mathfrak{B}})$, $(\mathfrak{H}_{\mathfrak{B}}^+)$ or $(\mathfrak{H}_{\mathfrak{N}})$ holds iff the same hypothesis holds for (2.1_0).

Moreover, if $Y_\alpha(t)$, $(\alpha = 1, 2)$, are solutions of (2.1) and $Y_{0\alpha}(t)$ are the corresponding solutions of (2.1_0) satisfying (2.7) then $\{Y_1;\ Y_2\}(t) \equiv \{Y_{01};\ Y_{02}\}(t)$; in particular, Y_1 and Y_2 are conjoined solutions of (2.1) iff Y_{01} and Y_{02} are conjoined solutions of (2.1_0).

Two distinct values t_1, t_2, $(t_1 < t_2)$, on I are said to be *(mutually) conjugate* with respect to (2.1) if there exists a self-conjoined solution $Y(t) = (U(t);\ V(t))$ of this system such that $U(t) \not\equiv 0$ on $[t_1, t_2]$, while $U(t_1) = 0 = U(t_2)$. Correspondingly, (2.1) is termed *disconjugate* on a nondegenerate subinterval I_0 of I provided no two distinct values on I_0 are conjugate with respect to (2.1).

If $[a, b]$ is a nondegenerate compact subinterval of I, then a function H: $[a, b] \to \mathfrak{B}$ is said to be of class $\mathcal{D}[a, b]$ if on this subinterval H is continuous and possesses a piecewise continuous derivative which satisfies with a piecewise continuous Z: $[a, b] \to \mathfrak{B}$ the differential equation $L_2[H, Z](t) = 0$ at each point of differentiability of H; this relationship will be denoted by the symbol $\mathrm{H} \in \mathcal{D}[a, b]$: Z. Also, $\mathcal{D}_{*0}[a, b]$ and $\mathcal{D}_{0*}[a, b]$ are used to denote the subclasses of $\mathrm{H} \in \mathcal{D}[a, b]$ satisfying the respective end-conditions $\mathrm{H}(b) = 0$ and $\mathrm{H}(a) = 0$, and $\mathcal{D}_0[a, b] = \mathcal{D}_{0*}[a, b] \cap \mathcal{D}_{*0}[a, b]$.

For $\mathrm{H}_\alpha \in D[a, b]$: Z_α, $(\alpha = 1, 2)$, the "Dirichlet function" integral

$$\int_a^b \{\mathrm{Z}_2^*(s) B(s) \mathrm{Z}_1(s) + \mathrm{H}_2^*(s) C(s) \mathrm{H}_1(s)\}\, ds$$

is well-defined, and is an element $\mathbf{J}[\mathrm{H}_1, \mathrm{H}_2|a, b]$ of $\mathfrak{B}$ which on $\mathcal{D}[a, b] \times \mathcal{D}[a, b]$ is hermitian in the sense that $(\mathbf{J}[\mathrm{H}_1, \mathrm{H}_2|a, b])^* = \mathbf{J}[\mathrm{H}_2, \mathrm{H}_1|a, b]$. As is customary, the symbol $\mathbf{J}[\mathrm{H}, \mathrm{H}|a, b]$ is abbreviated to $\mathbf{J}[\mathrm{H}|a, b]$. If $\mathrm{H}_\alpha \in \mathcal{D}[a, b]$: Z_α, $(\alpha = 1, 2)$, and Z_1 is continuous and has a piecewise continuous derivative on $[a, b]$, then an integration by parts yields the relation

$$\mathbf{J}[\mathrm{H}_1, \mathrm{H}_2|a, b] = \mathrm{H}_2^* \mathrm{Z}_1|_a^b + \int_a^b \mathrm{H}_2^*(s) L_1[\mathrm{H}_1, \mathrm{Z}_1](s)\, ds.$$

If $Y(t) = (U(t);\ V(t))$ is a solution of (2.1) and $\mathrm{H} \in \mathcal{D}_0[a, b]$: Z, then $\mathbf{J}[U, \mathrm{H}|a, b] = 0$; in particular, if $Y(t) = (U(t);\ V(t))$ is a solution of (2.1) and $U(a) = 0 = U(b)$, then $\mathbf{J}[U|a, b] = 0$. Finally, if $Y(t) = (U(t);\ V(t))$ is a self-conjoined solution of (2.1) and $\mathrm{H} \in \mathcal{D}[a, b]$: Z, while there exists a Y: $[a, b] \to \mathfrak{B}$ which is continuous, with piecewise continuous derivatives, and satisfies $\mathrm{H}(t) = U(t)\mathrm{Y}(t)$ for $t \in [a, b]$, then

$$\mathbf{J}[\mathrm{H}|a, b] = \mathrm{H}^* V\mathrm{Y}|_a^b + \int_a^b [\mathrm{Z} - V\mathrm{Y}]^* B [\mathrm{Z} - V\mathrm{Y}]\, dt. \tag{2.8}$$

In particular, if $Y(t) = (U(t);\ V(t))$ is a self-conjoined solution of (2.1) with $U(t)$ nonsingular on $[a, b]$, then for $\mathrm{H} \in \mathcal{D}_0[a, b]$: Z one has $\mathrm{H}(t) = U(t)\mathrm{Y}(t)$ with $\mathrm{Y}(t) = U^{-1}(t)\mathrm{H}(t)$ and $\mathrm{Y}(a) = 0 = \mathrm{Y}(b)$, so that from the above relation we have

$$\mathbf{J}[\mathrm{H}|a, b] = \int_a^b [\mathrm{Z} - V\mathrm{Y}]^* B [\mathrm{Z} - V\mathrm{Y}]\, dt. \tag{2.9}$$

For $[a, b]$ a nondegenerate compact subinterval of I, each of the following conditions is of importance.

$\mathbf{N}[a, b]$ *There exists a self-conjoined solution* $Y(t) = (U(t);\ V(t))$ *of* (2.1) *with* $U(t)$ *nonsingular on* $[a, b]$.

$\mathbf{J}^+[a, b]$ $\mathbf{J}[\mathrm{H}|a, b]$ *is positive definite on* $\mathcal{D}_0[a, b]$.

$\Delta[a, b]$ (2.1) *is disconjugate on* $[a, b]$.

In view of the above indicated relations, as in Theorems 5.2 and 5.1 of Williams [19] one may show that for systems (2.1) satisfying hypothesis (H_0) we have the following:

(2.10) $$\mathbf{J}^+[a,b] \to \Delta[a,b];$$

(2.11) $$(\mathfrak{H}_{\mathfrak{B}}), \mathbf{N}[a,b] \to \mathbf{J}^+[a,b].$$

Also, Theorem 4.1 of [19] yields the result:

(2.12) $$(\mathfrak{H}_{\mathfrak{B}}^+), \mathbf{N}[a,b] \to U_a^\circ(t) \textit{ nonsingular on } (a,b].$$

In the finite dimensional matrix case one may show that $\Delta[a,b] \to \mathbf{N}[a,b]$ for systems (2.1) satisfying $(\mathfrak{H}_0)$ and $(\mathfrak{H}_{\mathfrak{B}})$, so that for such systems the three conditions $\mathbf{N}[a,b]$, $\mathbf{J}^+[a,b]$ and $\Delta[a,b]$ are equivalent. That this equivalence no longer persists in the general case considered here, however, is illustrated by the following example of Heimes [7]. Again, let H be the Hilbert space $H = l^2$, $I = (-\infty, \infty)$, $\mathfrak{B}$ the Banach algebra of bounded linear transformations $L[H, H]$, and consider the equation (2.4) with $R(t) \equiv E$, $Q(t) \equiv 0$, and $P(t)$ the transformation with matrix $[-k_n^2 \delta_{nm}]$, where $k_n = n\pi/(n+1)$. Then for $\tau \in I$ the corresponding solution $Y_\tau(t) = (U_\tau^\circ(t); V_\tau^\circ(t))$ of the associated system (2.1) with coefficients (2.5) is such that $U_\tau^\circ(t)$ has matrix $[\delta_{nm} k_n^{-1} \sin k_n(t-\tau)]$. If $\tau \in [0,1]$ then $U_\tau^\circ(t)$ is readily seen to be one-to-one if $t \neq \tau$ and $t \in [0,1]$, so that (2.4) is disconjugate on $[0,1]$. On the other hand, for $\tau = 0$ the transformation $U_0^\circ(1)$ is not onto, since $\Sigma_n (1/n) e^{(n)}$ is not in its range. Consequently, $U_0^\circ(1)$ is singular, and in view of (2.12) the condition $\mathbf{N}[0,1]$ does not hold for the equation, although $\Delta[0,1]$ is satisfied.

For systems (2.1) satisfying $(\mathfrak{H}_0)$, other properties established by Williams [19] are as follows.

(2.13) $$(\mathfrak{H}_{\mathfrak{B}}^+) \to (\mathfrak{P}), \qquad [19; \text{Thm. } 4.3];$$

(2.14) $$(\mathfrak{H}_{\mathfrak{B}}), U_a^\circ(t) \textit{ nonsingular on } (a,b] \to \mathbf{N}[a,b], \qquad [19; \text{Thm. } 4.4];$$

(2.15) $$(\mathfrak{H}_{\mathfrak{B}}), (\mathfrak{P}), \mathbf{N}[a,b] \to U_\tau^\circ(t) \textit{ nonsingular on } [a,\tau) \cup (\tau, b] \textit{ for } \tau \in [a,b], \ [19; \text{Thm. } 4.5];$$

(2.16) $(\mathfrak{H}_{\mathfrak{B}}), (\mathfrak{P}), \mathbf{N}[a,b] \to$ *if* τ_1, τ_2 *are distinct values on* $[a,b]$, $U_\alpha \in \mathfrak{B}$, $(\alpha = 1, 2)$, *then there is a unique solution* $Y(t) = (U(t); V(t))$ *of* (2.1) *satisfying* $U(\tau_\alpha) = U_\alpha$, $(\alpha = 1, 2)$, [19; Thm. 4.8].

An element $W \in \mathfrak{B}$ is said to be *compact* if for each bounded sequence $\{X_n\}$ in $\mathfrak{B}$ the sequence $\{WX_n\}$ contains a convergent subsequence. If $T_W: \mathfrak{B} \to \mathfrak{B}$ is defined by $T_W(X) = WX$ for each $X \in \mathfrak{B}$, then W is a compact element of $\mathfrak{B}$ iff T_W is a compact operator. The following result is of significance in connection with the discussion of the following § 4.

(2.17) *If* (2.1) *satisfies* $(\mathfrak{H}_0)$, $(\mathfrak{H}_{\mathfrak{B}}^+)$, *and* $[a,b]$ *is a nondegenerate compact subinterval of* I *on which* $C(t)$ *is compact for* $t \in [a,b]$, *then the conditions* $\mathbf{N}[a,b]$, $\mathbf{J}^+[a,b]$ *and* $\Delta[a,b]$ *are equivalent.*

Williams [19; Thm. 5.4] established this result for (2.1) under the additional requirement that $A(t)$ and $A^*(t)$ are compact for $t \in [a,b]$. This additional requirement is not needed, however, in view of the possible transformation of (2.1) into (2.1_0) by a substitution (2.7).

3. Polar coordinate transformations. Another pertinent question is the extension to differential systems (2.1) in a B^*-algebra of the generalized polar coordinate transformations established for finite dimensional matrix problems by Barrett [1] and Reid [14], [15]; see also, [16; Probs. VII.2: 6, 7, 8; VII.5: 14, 15; VII.7: 3, 4]. For A, B, C satisfying hypothesis $(\mathfrak{H}_0)$ on I, and elements Φ, Ψ of $\mathfrak{B}$, let

$$Q(t; \Phi, \Psi) = \Psi B(t)\Psi^* + \Psi A(t)\Phi^* + \Phi A^*(t)\Psi^* - \Phi C(t)\Phi^*, \tag{3.1}$$

$$M(t; \Phi, \Psi) = \Phi A(t)\Phi^* + \Psi C(t)\Phi^* + \Phi B(t)\Psi^* - \Psi A^*(t)\Psi^*. \tag{3.2}$$

In the finite n-dimensional matrix problem, if $\tau \in I$ and $Y(t) = (U(t); V(t))$ is a conjoined basis for (2.1) (i.e., the column vectors of $Y(t)$ form n linearly independent solutions of (2.1) which are mutually conjoined) satisfying $Y(\tau) = (U_0; V_0)$ we have

$$\text{(i)}\ U_0^*U_0 + V_0^*V_0 > 0; \qquad \text{(ii)}\ V_0^*U_0 - U_0^*V_0 = 0. \tag{3.3}$$

Moreover, there exist Φ_0, Ψ_0, R_0 satisfying

$$\begin{aligned} &\text{(i)}\ R_0^*R_0 = U_0^*U_0 + V_0^*V_0, \\ &\text{(ii)}\ U_0 = \Phi_0^*R_0, \qquad V_0 = \Psi_0^*R_0, \end{aligned} \tag{3.4}$$

and for any such set Φ_0, Ψ_0, R_0 we have

$$\text{(i)}\ \Phi_0\Phi_0^* + \Psi_0\Psi_0^* = E, \qquad \text{(ii)}\ \Psi_0\Phi_0^* - \Phi_0\Psi_0^* = 0, \tag{3.5}$$

and the solutions $\Phi(t)$, $\Psi(t)$, $R(t)$ of the differential systems

$$\begin{aligned} \text{(a)}\ &\Lambda_1^\circ[\Phi, \Psi](t) \equiv -\Psi'(t) - Q(t; \Phi(t), \Psi(t))\Phi(t) = 0, \\ &\Lambda_2^\circ[\Phi, \Psi](t) \equiv \Phi'(t) - Q(t; \Phi(t), \Psi(t))\Psi(t) = 0, \\ \text{(b)}\ &\Phi(\tau) = \Phi_0, \qquad \Psi(\tau) = \Psi_0, \end{aligned} \tag{3.6}$$

$$\begin{aligned} \text{(a)}\ &\Lambda^\circ[\Phi, \Psi, R](t) \equiv R'(t) - M(t; \Phi(t), \Psi(t))R(t) = 0, \\ \text{(b)}\ &R(\tau) = R_0, \end{aligned} \tag{3.7}$$

are such that

$$U(t) = \Phi^*(t)R(t), \qquad V(t) = \Psi^*(t)R(t) \quad \text{for } t \in I. \tag{3.8}$$

Conversely, if $\tau \in I$ and $\Phi(t)$, $\Psi(t)$, $R(t)$ are solutions of (3.6), (3.7), where R_0 is nonsingular and Φ_0, Ψ_0 satisfy (3.5), then $U(t)$, $V(t)$ defined by (3.8) is a conjoined basis for (2.1) with

$$R^*(t)R(t) = U^*(t)U(t) + V^*(t)V(t). \tag{3.9}$$

Now for (2.1) in a general B^*-algebra setting not all of the above statements remain true, and we shall proceed to diagnose the needed alterations to obtain a valid representation theorem.

Consider a self-conjoined solution $Y(t) = (U(t); V(t))$ of (2.1) satisfying $Y(\tau) = (U_0; V_0)$ for a given $\tau \in I$. Then (3.3ii) is a consequence of $Y(t)$ being self-conjoined, and as a partial extension of the concept of a conjugate basis we postulate that (3.3i) holds. Then the positive hermitian element $U_0^*U_0 + V_0^*V_0$ of $\mathfrak{B}$ possesses a unique positive hermitian square root $\hat{R}_0$ that commutes with all elements of $\mathfrak{B}$ that commute with $U_0^*U_0 + V_0^*V_0$ (see [10; p. 486], or [18; pp. 183, 231]; the power series expansion used in [14] to obtain the corresponding result for matrices may also be adapted to obtain the stated result) and the most general solution of (3.4i) is $R_0 = K\hat{R}_0$ where $K \in \mathfrak{B}$ is such that $K^*K = E$. Any such determined R_0 is nonsingular and conditions

(3.4ii) determine uniquely Φ_0 and Ψ_0. Moreover, with Φ_0, Ψ_0, R_0 thus determined, relations (3.3), (3.4) imply (3.5).

Now in view of the specific form of $Q(t; \Phi, \Psi)$ defined by (3.1) it follows that this function is Lipschitz in Φ, Ψ on any set $[a, b]\times\mathfrak{B}_r\times\mathfrak{B}_r$, where $\mathfrak{B}_r=\{X: X\in\mathfrak{B}, \|X\|\leqq r\}$, and consequently by well-known existence theorems (see, for example, [9; § 3.4]) there exists locally a unique solution $\Phi(t)$, $\Psi(t)$ of (3.6a, b). Now if a solution $(\Phi(t); \Psi(t))$ of (3.6a) exists on a subinterval I_0 of I containing τ, then on I_0 the functions $G(t)=\Phi(t)\Phi^*(t)+\Psi(t)\Psi^*(t)$ and $H(t)=\Psi(t)\Phi^*(t)-\Phi(t)\Psi^*(t)$ satisfy $G'(t)=Q(t)H(t)-H(t)Q(t)$, $H'(t)=G(t)Q(t)-Q(t)G(t)$, $G(\tau)=E$, $H(\tau)=0$, where for brevity we write $Q(t)$ for $Q(t; \Phi(t), \Psi(t))$, and consequently in view of the uniqueness of solution of this linear system we have $G(t)\equiv E$, $H(t)\equiv 0$ for $t\in I_0$. In particular, $0\leqq\Phi(t)\Phi^*(t)\leqq E$ and $0\leqq\Psi(t)\Psi^*(t)\leqq E$, so that $\mathrm{Sp}\,[\Phi(t)\Phi^*(t)]\subset[0, 1]$ and $\mathrm{Sp}\,[\Psi(t)\Psi^*(t)]\subset[0, 1]$ for $t\in I_0$. As the spectral radius of an hermitian element is equal to its norm, we then have $\|\Phi(t)\|^2=\|\Phi(t)\Phi^*(t)\|\leqq 1$ and $\|\Psi(t)\|^2=\|\Psi(t)\Psi^*(t)\|\leqq 1$, so that $\|\Phi(t)\|\leqq 1$ and $\|\Psi(t)\|\leqq 1$ throughout I_0. As a result of this uniform boundedness of norms of $\Phi(t)$ and $\Psi(t)$ throughout an interval of existence, the usual continuation argument for finite-dimensional systems (see, for example, [16; Thm. I.5.7]) remains applicable to conclude that the solution $\Phi(t)$, $\Psi(t)$ of (3.6a, b) is uniquely extensible to the entire interval I. For the special case of the generalizations of the sine and cosine functions, i.e., for $(U_0; V_0)=(0; E)$ or $(E; 0)$, this type of argument has been noted by Kreith and Benson [11]. Another procedure for the proof of this result is to consider an arbitrary compact subinterval $[a, b]$ of I containing τ, and on $[a, b]$ determine the solution of (3.6a, b) by iteration, starting with $\Phi_0(t)\equiv\Phi_0$, $\Psi_0(t)\equiv\Psi_0$, and consider successively

$$-\Psi'_{j+1}(t)-Q(t; \Phi_j(t), \Psi_j(t))\Phi_{j+1}(t)=0, \qquad \Phi_{j+1}(\tau)=\Phi_0,$$
$$\Phi'_{j+1}(t)-Q(t; \Phi_j(t), \Psi_j(t))\Psi_{j+1}(t)=0, \qquad \Psi_{j+1}(\tau)=\Psi_0.$$

After solutions $\Phi(t)$, $\Psi(t)$ of (3.6a, b) have been obtained, $R(t)$ is uniquely determined by (3.7a, b). Moreover, $R(t)$ is nonsingular for $t\in I$, with inverse $S(t)=R^{-1}(t)$ the solution of the corresponding linear problem $S'(t)+S(t)M(t; \Phi(t), \Psi(t))=0$, $S(\tau)=R_0^{-1}$. In particular, we have

$$\Phi(t)\Phi^*(t)+\Psi(t)\Psi^*(t)\equiv E, \qquad \Psi(t)\Phi^*(t)-\Phi(t)\Psi^*(t)\equiv 0, \qquad t\in I. \tag{3.10}$$

Now if $\Phi(t)$, $\Psi(t)$, $R(t)$ are differentiable functions satisfying systems (3.6) and (3.7), the functions $U(t)$, $V(t)$ defined by (3.8) are differentiable on I, and one may verify readily that

$$\begin{aligned} L_1[U, V]&=(\Lambda_1^\circ[\Phi, \Psi])^*R+G_1[\Phi, \Psi]R-\Psi^*\Lambda^\circ[\Phi, \Psi, R],\\ L_2[U, V]&=(\Lambda_2^\circ[\Phi, \Psi])^*R+G_2[\Phi, \Psi]R+\Phi^*\Lambda^\circ[\Phi, \Psi, R], \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} G_1[\Phi, \Psi]&=M[C\Phi^*-A^*\Psi^*]+N[A\Phi^*+B\Psi^*],\\ G_2[\Phi, \Psi]&=N[C\Phi^*-A^*\Psi^*]-M[A\Phi^*+B\Psi^*], \end{aligned} \tag{3.12}$$

with

$$M=E-\Phi^*\Phi-\Psi^*\Psi, \qquad N=\Phi^*\Psi-\Psi^*\Phi. \tag{3.13}$$

In the finite dimensional matrix case the conditions (3.10) imply that the matrix

$$\begin{bmatrix} \Phi(t) & -\Psi(t) \\ \Psi(t) & \Phi(t) \end{bmatrix}$$

is unitary, and hence also that the M and N defined by (3.13) are also equal to 0. In this case, for $\Phi(t)$, $\Psi(t)$, $R(t)$ solutions of (3.6), (3.7) we have that $Y(t)=(U(t); V(t))$ defined by (3.8) is a solution of (2.1), and in view of the relations (3.4ii) it follows that $Y(t)$ has initial value $Y(\tau)=(U_0; V_0)$. For the general B^*-algebra case the fact that $(\Phi; \Psi)$ is a solution of (3.6a) implies that the functions M and N defined by (3.13) satisfy $M'(t)\equiv 0$, $N'(t)\equiv 0$, so that these functions are indeed constant, although they need not be 0. For example, let H be the Hilbert space l^2, $\mathfrak{B}$ the B^*-algebra $L[l^2, l^2]$, and as usual identify functions on I to $\mathfrak{B}$ by their matrices. Let $A\equiv 0$, $B\equiv E$, $C(t)\equiv[C_{\alpha\beta}(t)]$ a real symmetric matrix with $C_{12}(t)\equiv 1$, and for $\tau=0$ let $U_0=[\delta_{\alpha,\beta+1}]$, $(\alpha, \beta=1, 2, \cdots)$, $V^0=0$. Then $U_0^*U_0+V_0^*V_0=E$, and (3.4) is satisfied by $R_0=E$, $\Phi_0=U_0^*=[\delta_{\beta,\alpha+1}]$, $\Psi_0=0$, in which case $M(t)\equiv E-\Phi_0^*\Phi_0-\Psi_0^*\Psi_0=[\delta_{\alpha 1}\delta_{\beta 1}]$ and $N(t)\equiv 0$. In particular, the element in the first row and first column of the matrix of $G_1[\Phi, \Psi](\tau)$ is $C_{12}(\tau)=1$, so that for $U(t)$, $V(t)$ defined by (3.8) in terms of the solutions $\Phi(t)$, $\Psi(t)$, $R(t)$ of (3.6), (3.7) we do not have $L_1[U, V](\tau)=0$.

In view of this example it is clear that in order for solutions of (2.1) to have a representation (3.8) in terms of solutions of systems (3.6), (3.7) one must augment the hypotheses beyond those which suffice in the finite dimensional matrix case. In particular, the following result is valid for systems (2.1) in the B^*-algebra context.

THEOREM 3.1. *If $\tau\in I$ and $Y(t)=(U(t), V(t))$ is a self-conjoined solution of* (2.1) *such that $Y(\tau)=(U_0; V_0)$ for which* (3.3i) *is satisfied, and there exist elements Φ_0, Ψ_0, R_0 of $\mathfrak{B}$ satisfying conditions* (3.4), (3.5) *and*

$$\Phi_0^*\Phi+\Psi_0^*\Psi=E, \qquad \Phi_0^*\Psi_0-\Psi_0^*\Phi_0=0, \tag{3.14}$$

then in terms of the solutions $(\Phi; \Psi; R)$ of (3.6), (3.7) *we have the representation* (3.8); *moreover,*

$$U^*(t)U(t)+V^*(t)V(t)>0 \quad \textit{for } t\in I. \tag{3.15}$$

Conversely, if $\Phi(t)$; $\Psi(t)$; $R(t)$ are solutions of (3.6), (3.7) *where R_0 is nonsingular and Φ_0, Ψ_0 satisfy* (3.5) *and* (3.14), *then $Y(t)=(U(t); V(t))$ defined by* (3.8) *is a self-conjoined solution of* (2.1) *and $U^*(t)U(t)+V^*(t)V(t)=R^*(t)R(t)>0$.*

4. Related criteria for equations (2.4). In this section we shall consider an equation (2.4) with coefficients that satisfy hypotheses $(\mathfrak{H}')$ on a noncompact interval $I=(c, \infty)$. If $[a, b]\subset I$, then for an equation (2.4) the class $\mathcal{D}[a, b]$ is the set of continuous functions H: $[a, b]\to\mathfrak{B}$ which have piecewise continuous derivatives; moreover, for such H we have

$$\mathbf{J}[\mathrm{H}|a, b]=\int_a^b \{\mathrm{H}^{*\prime}(t)[R(t)\mathrm{H}'(t)+Q(t)\mathrm{H}(t)]+\mathrm{H}^*(t)[Q^*(t)\mathrm{H}'(t)+P(t)\mathrm{H}(t)]\}\, dt. \tag{4.1}$$

For such equations we shall be concerned with the condition $\mathbf{J}^+[a, b]$, or the weaker condition $\mathbf{J}^{\lrcorner}[a, b]$ that $\mathbf{J}[\mathrm{H}|a, b]\geqq 0$ for arbitrary $\mathrm{H}\in\mathcal{D}_0[a, b]$.

Now let $\mathfrak{F}_{\mathfrak{B}}$ denote the class of positive linear functionals F on $\mathfrak{B}$; see, for example, Rickart [18, Ch. IV, § 5]. In particular, such an F is hermitian in the sense that $F(X)=F(X^*)$ for arbitrary $X\in\mathfrak{B}$, and $F(X)\geqq 0$, $\{F(X)>0\}$, if $X\geqq 0$, $\{X>0\}$. Also, if the coefficient functions R, P, Q satisfy $(\mathfrak{H}')$ then the related functions $R_F(t)=F(R(t))$, $P_F(t)=F(P(t))$, $Q_F(t)=F(Q(t))$ are continuous on I with $R_F(t)>0$. If $[a, b]\subset I$ and

$\eta(t)$ is a continuous complex-valued function which has piecewise continuous derivatives on $[a, b]$, and $\eta(a) = 0 = \eta(b)$, then $\mathrm{H}_\eta(t) = \eta(t)E$ is an element of $\mathscr{D}_0[a, b]$ and

$$F(\mathbf{J}[\mathrm{H}_\eta | a, b]) = \int_a^b \{\bar{\eta}'(t)[R_F(t)\eta'(t) + Q_F(t)\eta(t)] + \bar{\eta}(t)[\bar{Q}_F(t)\eta'(t) + P_F(t)\eta(t)]\}\, dt. \tag{4.2}$$

Now by the classical oscillation theory for second-order scalar equations (see, for example, [16; Chs. V, VI for real coefficients, Ch. VII for complex coefficients]) a necessary and sufficient conditions for the disconjugacy of the scalar equation

$$[R_F(t)u'(t) + Q_F(t)u(t)]' - [\bar{Q}_F(t)u'(t) + P_F(t)u(t)] = 0 \tag{4.3}$$

on a subinterval $[\alpha_0, \infty)$ of I is that for arbitrary nondegenerate compact subintervals $[a, b]$ of $[a_0, \infty)$ the integral (4.2) be nonnegative for arbitrary scalar functions η which are continuous, have piecewise continuous derivatives on $[a, b]$, and $\eta(a) = 0 = \eta(b)$. Consequently, we have the following result.

THEOREM 4.1. *If hypothesis* $(\mathfrak{H}')$ *is satisfied on* $I = (c, \infty)$, *and for arbitrary* $a_0 \in I$ *there is an* $F \in \mathfrak{F}_{\mathfrak{B}}$ *such that the scalar equation* (4.3) *is not disconjugate on* $[a_0, \infty)$, *then there is no nondegenerate subinterval* $I_0 = [a_0, \infty)$ *of* I *such that* $\mathbf{J}[\mathrm{H}|a, b] \geqq 0$ *for arbitrary* $\mathrm{H} \in \mathscr{D}_0[a, b]$. *In particular, if* $P(t) - Q^*(t)R^{-1}(t)Q(t)$ *is a compact element of* $\mathfrak{B}$ *for* $t \in I$, *then there does not exist a subinterval* $I_0 = [a_0, \infty) \subset (c, \infty)$ *on which* (2.4) *is disconjugate whenever there exists a functional* $F \in \mathfrak{F}_{\mathfrak{B}}$ *such that the scalar equation* (4.3) *is not disconjugate on arbitrary* $[a_0, \infty) \subset (c, \infty)$.

In the context of the C^*-algebra of bounded linear transformations in a Hilbert space, results of the character of the first sentence of this theorem are to be found in the papers of Etgen and Pawlowski [4], and Etgen and Lewis [5].

We shall now consider another class of related criteria for equations (2.4) satisfying hypothesis $(\mathfrak{H}')$ on an interval $I = (c, \infty)$. If $I_0 = [a_0, \infty)$, where $a_0 \in I$, and $\delta > 0$ is such that also $a_0 - \delta \in I$, then for $s \in [-\delta, \infty)$ the elements $R_s(t) = R(s+t)$, $P_s(t) = P(s+t)$, $Q_s = Q(s+t)$ of $\mathfrak{B}$ also satisfy hypothesis $(\mathfrak{H}')$ on $[a_0, \infty)$. Moreover, if $[a, b] \subset I_0$ and $\mathrm{H} \in \mathscr{D}_0[a, b]$, then $\mathrm{H}_s(t) = \mathrm{H}(t-s)$, $t \in [a+s, b+s]$ is of class $\mathscr{D}_0[a+s, b+s]$, and a simple change of variable yields the relation

$$\mathbf{J}[\mathrm{H}_s | a+s, b+s] = \int_a^b \{\mathrm{H}^{*\prime}(t)[R_s(t)\mathrm{H}'(t) + Q_s(t)\mathrm{H}(t)] + \mathrm{H}^*(t)[Q_s^*(t)\mathrm{H}'(t) + P_s(t)\mathrm{H}(t)]\}\, dt. \tag{4.4}$$

For a given $\mathrm{H} \in \mathscr{D}_0[a, b]$ the right-hand member of (4.4) defines a continuous element $\mathbf{J}_s[\mathrm{H}|a, b]$ of B on $[-\delta, \infty)$, to which one may apply continuous linear transformations corresponding to those considered for the finite-dimensional matrix case in Reid [17]. For simplicity, however, attention will be limited to the case of transformations generated by a nonnegative piecewise continuous real-valued function g with compact support on $[-\delta, \infty)$ and $\int_{-\infty}^{\infty} g(s)\, ds = 1$. A particularly important instance is the integral mean transformation determined as follows. Let $\theta_1(t) = 0$ for $|t| > 1$, $\theta_1(t) = \frac{1}{2}$ on $[-1, 1]$, and for $0 < h < \delta$ set $\theta_h(t) = h^{-1}\theta_1(h^{-1}t)$. Then $g = \theta_h(t)$ is a function of the desired type. Another important example is that of a Friedrichs mollifier function of class $\mathfrak{C}^\infty$, with support on $[-1, 1]$, and $\int_{-\infty}^{\infty} \theta_1(t)\, dt = 1$. Then whenever R, P, Q satisfy (H') on $I = (c, \infty)$, $a_0 \in I$, and $\delta > 0$ is such that also $[a_0 - \delta, \infty) \subset I$, the functions $R(t|g) = \int_{-\infty}^{\infty} g(s)R_s(t)\, ds$, $P(t|g) = \int_{-\infty}^{\infty} g(s)P_s(t)\, ds$, $Q(t|g) = \int_{-\infty}^{\infty} g(s)Q_s(t)\, ds$ are elements of $\mathfrak{B}$

for $t \in [a_0, \infty)$ which satisfy $(\mathfrak{H}')$. Moreover, if $[a, b] \subset [a_0, \infty)$ and $\mathrm{H} \in \mathcal{D}_0[a, b)$ we have

$$\int_{-\infty}^{\infty} g(s)\mathbf{J}[\mathrm{H}_s|a+s, b+s]\,ds = \int_a^b \{\mathrm{H}^{*\prime}(t)[R(t|g)\mathrm{H}'(t) + Q(t|g)\mathrm{H}(t)] + \mathrm{H}^*(t)[Q^*(t|g)\mathrm{H}'(t) + P(t|g)\mathrm{H}(t)]\}\,dt, \tag{4.5}$$

and therefore we have the following result.

THEOREM 4.2. *If hypothesis* $(\mathfrak{H}')$ *is satisfied on* $I = (c, \infty)$ *and for arbitrary* $a_0 \in I$ *and* $\delta > 0$ *such that* $[a_0 - \delta, \infty) \subset I$ *there exists a g of the above described sort with compact support in* $[-\delta, \infty)$, *and such that the functional* $\mathbf{J}_g[\mathrm{H}|a, b]$ *defined by the right-hand member of* (4.5) *fails to be nonnegative for arbitrary* $\mathrm{H} \in \mathcal{D}_0[a, b]$ *and* $[a, b] \subset [a_0, \infty)$, *then there is no nondegenerate subinterval* $[a_0, \infty)$ *of I such that* $\mathbf{J}[\mathrm{H}|a, b] \geqq 0$ *for arbitrary* $\mathrm{H} \in \mathcal{D}_0[a, b]$ *and* $[a, b] \subset [a_0, \infty)$. *In particular, if* $P(t) - Q^*(t)R^{-1}(t)Q(t)$ *is a compact element of* $\mathfrak{B}$ *for* $t \in I$, *then there does not exist a subinterval* $I = [a_0, \infty) \subset (c, \infty)$ *on which* (2.4) *is disconjugate whenever there exists a g of the above described sort such that* $\mathbf{J}_g[\mathrm{H}|a, b]$ *fails to be nonnegative for arbitrary* $\mathrm{H} \in \mathcal{D}_0[a, b]$ *and* $[a, b] \subset [a_0, \infty)$.

5. Green's functions and self-adjoint boundary problems. For a system (2.1) satisfying hypothesis $(\mathfrak{H}_0)$ let $Y_\alpha(t) = (U_\alpha(t); V_\alpha(t))$, $(\alpha = 1, 2)$, be solutions of this system which individually are self-conjoined and $\{Y_2; Y_1\} = Y_1^* \mathcal{J} Y_2 = -E$. For example, if $\tau \in I$ such a system is given by $Y_1(t) = Y_\tau^1(t)$ and $Y_2(t) = Y_\tau^\circ(t)$ as defined by (2.3). If we write in matrix form

$$\mathcal{Y}(t) = [Y_1(t) \quad Y_2(t)] = \begin{bmatrix} U_1(t) & U_2(t) \\ V_1(t) & V_2(t) \end{bmatrix}, \tag{5.1}$$

then the above specification of Y_1 and Y_2 may be written as $\mathcal{Y}^*(t)\mathcal{J}\mathcal{Y}(t) \equiv \mathcal{J}$. If $S\colon I \to \mathfrak{B} \times \mathfrak{B}$ is continuous, then for the consideration of the solvability of the nonhomogeneous equation

$$\mathcal{J}Y'(t) + \mathcal{A}(t)Y(t) = S(t), \qquad t \in I, \tag{5.2}$$

the method of variation of parameters involves seeking solutions of the form $Y(t) = \mathcal{Y}(t)T(t)$, where $T\colon I \to \mathfrak{B} \times \mathfrak{B}$. Upon substitution in (5.2) it follows that if such a solution exists, then $\mathcal{J}\mathcal{Y}(t)T'(t) = S(t)$, so that $\mathcal{J}T'(t) = \mathcal{Y}^*(t)S(t)$, $T'(t) = -\mathcal{J}\mathcal{Y}^*(t)S(t)$, and hence there exists a $T_0 \in \mathfrak{B} \times \mathfrak{B}$ such that

$$T(t) = T_0 - \frac{1}{2}\int_a^t \mathcal{J}\mathcal{Y}^*(s)S(s)\,ds + \frac{1}{2}\int_t^b \mathcal{J}\mathcal{Y}^*(s)S(s)\,ds.$$

That is, if there exists a solution $Y(t)$ of (5.2) of the form $Y(t) = \mathcal{Y}(t)T(t)$, then

$$Y(t) = \mathcal{Y}(t)T_0 - \frac{1}{2}\int_a^t \mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(s)S(s)\,ds + \frac{1}{2}\int_t^b \mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(s)S(s)\,ds. \tag{5.3}$$

In the finite dimensional matrix case the above steps are reversible, and (5.3) provides the general solution of (5.2). In the B^*-algebra context, however, additional consideration is needed. For $Y(t)$ given by (5.3) direct computation yields

$$\mathcal{J}Y'(t) + \mathcal{A}(t)Y(t) = -\mathcal{J}\mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(t)S(t), \tag{5.4}$$

and consequently for this equation to reduce to (5.2) we need $-\mathcal{J}\mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(t) \equiv \mathcal{E}$ or $\mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(t) \equiv \mathcal{J}$, where $\mathcal{E} \in \mathfrak{M}_2(\mathfrak{B})$ is the matrix representation of the identity transformation on $\mathfrak{B} \times \mathfrak{B}$. Again, in the finite dimensional matrix case this result is a ready consequence of the given relation $\mathcal{Y}^*(t)\mathcal{J}\mathcal{Y}(t) \equiv \mathcal{J}$. In general, the condition

$\mathcal{Y}^*(t)\mathcal{J}\mathcal{Y}(t) \equiv \mathcal{J}$ implies that $\mathcal{W}(t) = \mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(t)$ is such that $\mathcal{W}(t)\mathcal{J}\mathcal{Y}(t) = \mathcal{Y}(t)\mathcal{J}\mathcal{Y}^*(t)\mathcal{J}\mathcal{Y}(t) = \mathcal{Y}(t)\mathcal{J}^2 = -\mathcal{Y}(t) = \mathcal{J}^2\mathcal{Y}(t)$, so that $[\mathcal{W}(t) - \mathcal{J}]\mathcal{J}\mathcal{Y}(t) \equiv 0$. In particular, if there exists a $\tau \in I$ at which $\mathcal{Y}(\tau)$ has a right-hand inverse, then $\mathcal{W}(\tau) = \mathcal{J}$. Moreover, from (2.1) it follows that $\mathcal{W}'(t) = [\mathcal{J}\mathcal{A}(t)]\mathcal{W}(t) - \mathcal{W}(t)[\mathcal{A}(t)\mathcal{J}]$, and from the existence and uniqueness theorems for linear equations (see, for example, [10; § 6.1]) it follows that if there exists a single value $\tau \in I$ at which $\mathcal{W}(\tau) = \mathcal{J}$ then $\mathcal{W}(t) \equiv \mathcal{J}$ on I. Consequently, we have the following result.

THEOREM 5.1. *If $Y_\alpha(t) = (U_\alpha(t); V_\alpha(t))$, $(\alpha = 1, 2)$, are self-conjoined solutions of* (2.1) *satisfying $\{Y_2; Y_1\} = -E$, and there exists a $\tau \in I$ at which $\mathcal{Y}(\tau)\mathcal{J}\mathcal{Y}^*(\tau) = \mathcal{J}$, then the general solution of* (5.2) *is of the form* (5.3). *In particular, this condition holds if we have $Y_1(t) = Y^1_\tau(t)$, $Y_2(t) = Y^\circ_\tau(t)$ for some $\tau \in I$, since in this case $\mathcal{Y}(\tau) = \mathcal{E}$.*

Now associate with (5.2) two-point boundary conditions

$$M_a Y(a) + M_b Y(b) = 0, \tag{5.5}$$

where M_a and M_b are elements of $\mathfrak{M}_2(\mathfrak{B})$; that is, M_a and M_b are matrix representations of continuous linear transformations on $\mathfrak{B} \times \mathfrak{B}$ into $\mathfrak{B} \times \mathfrak{B}$. Moreover, if for a function $Y\colon [a, b] \to \mathfrak{B}$ we denote by $\hat{Y}$ the boundary element $(Y(a); Y(b))$ of $Y(t)$, then (5.5) may be written as

$$\mathcal{M}\hat{Y} = 0, \tag{5.5$'$}$$

where $\mathcal{M}$ denotes the 1×2 matrix $[M_a \ M_b]$. If $Y(t)$ is a solution of (5.2) given by (5.3), then

$$\hat{Y} = \hat{\mathcal{Y}}T_0 + \frac{1}{2}\int_a^b \mathcal{D}\hat{\mathcal{Y}}\mathcal{J}\mathcal{Y}^*(s)S(s)\,ds, \tag{5.6}$$

where $\mathcal{D}$ is the 2×2 matrix diag $\{\mathcal{E}, -\mathcal{E}\}$ of $\mathfrak{M}_2(\mathfrak{B} \times \mathfrak{B})$, and consequently condition (5.5$'$) becomes

$$0 = [\mathcal{M}\hat{\mathcal{Y}}]T_0 + \frac{1}{2}\int_a^b \mathcal{M}\mathcal{D}\hat{\mathcal{Y}}\mathcal{J}\mathcal{Y}^*(s)S(s)\,ds. \tag{5.7}$$

THEOREM 5.2. *If the matrix $\mathcal{M}\hat{\mathcal{Y}}$ of $\mathfrak{M}_2(\mathfrak{B})$ has a reciprocal $[\mathcal{M}\hat{\mathcal{Y}}]^{-1}$ satisfying $(\mathcal{M}\hat{\mathcal{Y}})([\mathcal{M}\hat{\mathcal{Y}}]^{-1}) = \mathcal{E}$, $([\mathcal{M}\hat{\mathcal{Y}}]^{-1})(\mathcal{M}\hat{\mathcal{Y}}) = \mathcal{E}$, then for arbitrary $S\colon [a, b] \to \mathfrak{B} \times \mathfrak{B}$ the equation* (5.2) *has a unique solution satisfying* (5.5), *given by*

$$Y(t) = \int_a^b \mathcal{G}(t, s)S(s)\,ds, \tag{5.8}$$

where

$$\begin{aligned} \mathcal{G}(t, s) &= -\tfrac{1}{2}\mathcal{Y}(t)[\mathcal{E} + [\mathcal{M}\hat{\mathcal{Y}}]^{-1}\mathcal{M}\mathcal{D}\hat{\mathcal{Y}}]\mathcal{J}\mathcal{Y}^*(s), \qquad a \leqq s < t \leqq b, \\ &= -\tfrac{1}{2}\mathcal{Y}(t)[-\mathcal{E} + [\mathcal{M}\hat{\mathcal{Y}}]^{-1}\mathcal{M}\mathcal{D}\hat{\mathcal{Y}}]\mathcal{J}\mathcal{Y}^*(s), \qquad a \leqq t < s \leqq b. \end{aligned} \tag{5.9}$$

As in the case of finite dimensional matrix differential equations, the $\mathcal{G}(t, s)$ of (5.9) is called the Green's matrix associated with $L[Y]$ subject to the boundary conditions (5.5). In view of the specifications of the $Y_1(t)$, $Y_2(t)$ entering into the above derivation, we also have the characteristic discontinuity along the line $t = s$,

$$\mathcal{G}(s^+, s) - \mathcal{G}(s^-, s) = -\mathcal{Y}(s)\mathcal{J}\mathcal{Y}^*(s) = -\mathcal{J}, \qquad a < s < b. \tag{5.10}$$

Moreover, as in the finite dimensional matrix differential equation case (see, for example, [16, § 8 of Ch. VII]) if one has satisfied the "self-adjointness" condition

$$M_a\mathcal{J}M_a^* - M_b\mathcal{J}M_b^* = 0, \tag{5.11}$$

then it may be established that

$$\mathscr{G}(t,s)\equiv[\mathscr{G}(s,t)]^*. \tag{5.12}$$

In particular, if $\mathscr{G}(t,s)$ is written as an element of $\mathfrak{M}_s(\mathfrak{B})$,

$$\mathscr{G}(t,s)=\begin{bmatrix} G_{11}(t,s) & G_{12}(t,s) \\ G_{21}(t,s) & G_{22}(t,s)\end{bmatrix}, \tag{5.13}$$

in view of (5.10), (5.12) the $G_{11}(t,s)$ and $G_{22}(t,s)$ may be defined on $[a,b]\times[a,b]$ as continuous elements of $\mathfrak{B}$ which are hermitian in the sense that $G_{11}(t,s)\equiv[G_{11}(s,t)]^*$, $G_{22}(t,s)\equiv[G_{22}(s,t)]^*$.

Consider, in particular, the case of boundary conditions

$$U(a)=0, \qquad U(b)=0, \tag{5.14}$$

which may be written as (5.5) with

$$M_a=\begin{bmatrix} E & 0 \\ 0 & 0\end{bmatrix}, \qquad M_b=\begin{bmatrix} 0 & 0 \\ E & 0\end{bmatrix}.$$

If it is supposed that $Y_a(t)=(U_a(t);\ V_a(t))$ defined by $U_a(a)=0$, $V_a(a)=E$ is such that $U_a(b)$ is nonsingular, then for $Y(t)=(U(t);\ V(t))$ a self-conjoined solution of (2.1) with $U(b)=0$ we have $\{Y;\ Y_a\}=Y_a^*\mathscr{J}Y\equiv -E$ iff $V(b)=U_a^{*-1}(b)$, in which case $U(a)=-E$ and $V(a)=V^*(a)$. In this instance, if $Y_1(t)=Y_a(t)$ and $Y_2(t)=Y_b(t)U_a^{*-1}(b)$, then $\mathscr{Y}(t)=[Y_1(t)\quad Y_2(t)]$ satisfies $\mathscr{Y}^*(t)\mathscr{J}\mathscr{Y}(t)\equiv\mathscr{J}$, $\mathscr{Y}(t)\mathscr{J}\mathscr{Y}^*(t)\equiv\mathscr{J}$, and

$$\mathscr{M}\hat{\mathscr{Y}}=\begin{bmatrix} U_1(a) & U_2(a) \\ U_1(b) & U_2(b)\end{bmatrix}=\begin{bmatrix} 0 & -E \\ U_a(b) & 0\end{bmatrix}$$

has inverse

$$[\mathscr{M}\hat{\mathscr{Y}}]^{-1}=\begin{bmatrix} 0 & U_a^{-1}(b) \\ -E & 0\end{bmatrix}$$

satisfying $[\mathscr{M}\hat{\mathscr{Y}}]^{-1}[\mathscr{M}\hat{\mathscr{Y}}]=\mathscr{E}=[\mathscr{M}\hat{\mathscr{Y}}][\mathscr{M}\hat{\mathscr{Y}}]^{-1}$. In this case we have $[\mathscr{G}(t,s)]^*=\mathscr{G}(s,t)$ for $(t,s)\in[a,b]\times[a,b]$, $t\neq s$, and

$$\mathscr{G}(t,s)=-\begin{bmatrix} U_2(t)U_1^*(s) & U_2(t)V_1^*(s) \\ V_2(t)U_1^*(s) & V_2(t)V_1^*(s)\end{bmatrix}, \qquad a\leqq s<t\leqq b. \tag{5.15}$$

Moreover, for arbitrary continuous $S_1\colon [a,b]\to\mathfrak{B}$ the unique solution of the boundary problem

$$\begin{aligned} &L_1[U,V](t)=S_1(t), \qquad L_2[U,V](t)=0, \\ &U(a)=0, \qquad U(b)=0, \end{aligned} \tag{5.16}$$

is given by

$$U(t)=\int_a^b G_{11}(t,s)S_1(s)\,ds, \qquad V(t)=\int_a^b G_{21}(t,s)S_1(s)\,ds, \tag{5.17}$$

where $G_{11}(t,s)=-U_2(t)U_1^*(s)$, $G_{21}(t,s)=-V_2(t)U_1^*(s)$ for $a\leqq s<t\leqq b$, and $G_{11}(t,s)=-U_1(t)U_2^*(s)$, $G_{21}(t,s)=-V_1(t)U_2^*(s)$ for $a\leqq t<s\leqq b$. In particular, $G_{11}(s^+,s)-G_{11}(s^-,s)=0$ for $s\in(a,b)$, and consequently along $t=s$ the element $G_{11}(t,s)$ may be defined in a unique manner to be continuous in (t,s) on $[a,b]\times[a,b]$ and hermitian in the sense that $G_{11}(t,s)\equiv[G_{11}(s,t)]^*$.

It is to be emphasized that the above discussion of the Green's function involves most intimately the self-adjointness of the system, and thus for second order linear differential equations with null end-conditions does not possess the generality of Heimes [7], [8, § 2], wherein no assumption of self-adjointness is made. Also, in our discussion there has been no consideration of systems wherein the coefficient functions are assumed to be merely closed linear operators, as in Heimes [8; § 3].

On the other hand, as will be discussed in the following section, the above determination of the Green's function and its hermitian character enable one to reduce the consideration of certain boundary problems in Hilbert space to known results on symmetrizable compact (completely continuous) linear transformations on an allied Hilbert space.

6. Boundary problems in Hilbert space. Now a B^*-algebra is isometrically $*$-isomorphic to the C^*-algebra of bounded linear transformations on a complex Hilbert space H, (see, for example, Rickart [18; Ch. IV]), and in this section $\mathfrak{B}$ will denote the C^*-algebra of such transformations $T: H \to H$. The coefficient functions A, B, C of (2.1) are supposed to satisfy hypothesis $(\mathfrak{H}_0)$, and in addition to the operator equation (2.1) we consider the vector equation

$$(6.1^0) \qquad \begin{aligned} L_1[u, v](t) &\equiv -v'(t) + C(t)u(t) - A^*(t)v(t) = 0, \\ L_2[u, v](t) &\equiv u'(t) - A(t)u(t) - B(t)v(t) = 0, \end{aligned}$$

where a solution of (6.1^0) is a pair of vector functions u, v on I to H which possess strong derivatives satisfying (6.1^0) on I. By the usual existence theorem (see, for example, [10; § 6.1]) for given initial values u_0, v_0 and arbitrary $\tau \in I$ there exists a unique solution $y(t) = (u(t); v(t))$ satisfying $y(\tau) = (u_0; v_0)$.

If $y_\alpha(t) = (u_\alpha(t); v_\alpha(t))$, $(\alpha = 1, 2)$, are solutions of (6.1^0), then $\{y_1; y_2\} = -(v_1(t), u_2(t))_H + (u_1(t), v_2(t))_H$ is constant on I, and if this constant is zero the vector solutions y_1 and y_2 are said to be mutually conjoined or conjugate. Also, corresponding to (2.2), the system (6.1^0) may be written as a vector equation

$$(6.1) \qquad \mathcal{J}y'(t) + \mathcal{A}(t)y(t) = 0,$$

where $y: I \to H_2 = H \times H$, and as usual the inner product $(\cdot, \cdot)_{H_2}$ for $y = (u, v)$ and $\zeta = (\xi, \eta)$ is $(y, \zeta)_{H_2} = (u, \xi)_H + (v, \eta)_H$, while $\mathcal{J}$ and $\mathcal{A}(t)$ are bounded linear transformations on H_2 with matrix representations (2.2).

For $[a, b]$ a nondegenerate compact subinterval of I, let $\mathbf{H}$ denote the class of functions $f: [a, b] \to H$ such that for arbitrary $\xi \in H$ the scalar function $(f(t), \xi)_H$ is (Lebesgue) measurable and $(f(t), f(t))_H$ is (Lebesgue) integrable on $[a, b]$. If for f and g elements of $\mathbf{H}$ we set $(f, g)_{\mathbf{H}} = \int_a^b (f(t), g(t))_H \, dt$, then $\mathbf{H}$ is a Hilbert space $\mathfrak{L}_H^2[a, b]$ with inner product $(f, g)_{\mathbf{H}}$, (see Bourbaki [2; Ch. IV, §§ 3–6; in particular, Cor. 3 on p. 209]).

Now consider the two-point vector boundary problem

$$(6.2) \qquad \begin{aligned} &L_1[u, v](t) = \lambda K(t)u(t), \qquad L_2[u, v](t) = 0, \\ &u(a) = 0, \qquad u(b) = 0, \end{aligned}$$

where $K: [a, b] \to \mathfrak{B}$ is strongly continuous and hermitian for $t \in [a, b]$. Moreover, for $Y_a(t) = (U_a(t); V_a(t))$ and $Y_b(t) = (U_b(t); V_b(t))$ solutions of the corresponding operator equation (2.1) as determined by initial conditions (2.3), suppose that $U_a(b)$ is nonsingular. As in § 5, in terms of $Y_1(t) = Y_a(t)$ and $Y_2(t) = Y_b(t)U_a^{*-1}(b)$ the corresponding Green's function has the form (5.15), and hence in view of (5.17) we have that $(u(t); v(t))$ is a solution of (6.2) iff

$$u(t)=\lambda \int_a^b G_{11}(t, s)K(s)u(s)\, ds, \tag{6.3}$$

$$v(t)=\lambda \int_a^b G_{21}(t, s)K(s)u(s)\, ds. \tag{6.4}$$

Now for $f \in \mathbf{H}$ we have that $(T_1 f)(t)=K(t)f(t)$ and $(T_2 f)(t)=\int_a^b G_{11}(t, s)f(s)\, ds$ for $t \in [a, b]$ define bounded hermitian linear transformations on $\mathbf{H}$ into $\mathbf{H}$. Moreover, $Tf = T_2 T_1 f$ is symmetrizable by T_1 in the sense that $(T_1 Tf, g)_{\mathbf{H}} = (f, T_1 Tg)_{\mathbf{H}}$ for f and g arbitrary elements of $\mathbf{H}$. In particular, if T_1 is compact, (completely continuous), then T is also compact, and if in addition T_1 is nonnegative then the theory of the integral equation (6.3) is a particular instance of the theory of symmetrizable completely continuous linear transformations as treated by Zaanen (see [20; Ch. 12], and his original papers in Indagationes Mathematicae) and Reid [13]. In particular, the spectrum of (6.3) consists of only real eigenvalues, each of finite multiplicity, and with no finite accumulation point. Moreover, there exists a set of eigenvalues λ_α and associated eigenelements u_α such that for arbitrary elements x and ξ of $\mathbf{H}$ we have $(T_1 Tx, \xi)_{\mathbf{H}} = \Sigma_\alpha \lambda_\alpha (x, T_1 u_\alpha)_{\mathbf{H}} (T_1 u_\alpha, \xi)_{\mathbf{H}}$; that is,

$$\begin{aligned}\int_a^b \int_a^b &(G_{11}(t, s)K(s)x(s), K(t)\xi(t))_H\, dt\, ds \\ &= \Sigma_\alpha \lambda_\alpha \left\{ \int_a^b (x(t), K(t)u_\alpha(t))_H\, dt \right\} \left\{ \int_a^b (K(t)u_\alpha(t), \xi(t))_H\, dt \right\}.\end{aligned}$$

REFERENCES

[1] J. H. BARRETT, *A Prüfer transformation for matrix differential equations*, Proc. Amer. Math. Soc., 8 (1957), pp. 510–518.

[2] N. BOURBAKI, *Éléments de Mathématique: Livre* VI, *Integration*, Chs. I–IV. Actualites Scientifiques et Industrielles 1175, 2nd ed., Hermann, Paris, 1965.

[3] G. J. ETGEN AND J. F. PAWLOWSKI, *Oscillation criteria for second order self-adjoint differential systems*, Pacific J. Math., 66 (1976), pp. 99–110.

[4] ———, *A comparison theorem and oscillation criteria for second order differential systems*, manuscript.

[5] G. J. ETGEN AND R. T. LEWIS, *The oscillation of ordinary differential equations in a B^*-algebra*, manuscript.

[6] T. L. HAYDEN AND H. C. HOWARD, *Oscillation of differential equations in Banach spaces*, Ann. Mat. Pura Appl., 85 (1970), pp. 383–394.

[7] K. A. HEIMES, *Two-point boundary problems in Banach spaces*, J. Differential Equations, 5 (1969), pp. 215–22.

[8] ———, *Green's functions for linear second order systems*, this Journal, to appear.

[9] E. HILLE AND R. S. PHILLIPS, *Functional Analysis and Semi-Groups*, Colloqium Publications, Vol. 31, American Mathematical Society, Providence, RI, 1957.

[10] E. HILLE, *Lectures on Ordinary Differential Equations*, Addison-Wesley, Reading, MA, 1966.

[11] K. KREITH AND D. C. BENSON, *On abstract Prüfer transformations*, Proc. Amer. Math. Soc., 26 (1970), pp. 137–140.

[12] E. S. NOUSSAIR, *Differential equations in Banach spaces*, Bull. Australian Math. Soc., 9 (1973), pp. 219–226.

[13] W. T. REID, *Symmetrizable completely continuous linear transformations in Hilbert space*, Duke Math. J., 18 (1951), pp. 41–56.

[14] ———, *A Prüfer transformation for differential systems*, Pacific J. Math., 8 (1958), pp. 575–584.

[15] ———, *Generalized polar coordinate transformations for differential systems*, Rocky Mountain J. Math., 1 (1971), pp. 383–406.

[16] ———, *Ordinary Differential Equations*, John Wiley, New York, 1971.

[17] ———, *Related oscillation criteria for higher order self-adjoint differential equations*, J. Differential Equations, to appear.

[18] C. E. RICKART, *General Theory of Banach Algebras*, Van Nostrand, New York, 1960.

[19] C. M. WILLIAMS, *Oscillation phenomena for linear differential equations in a B^*-algebra*, Ph.D. dissertation, University of Oklahoma, 1971.

[20] A. C. ZAANEN, *Linear Analysis*, Interscience, New York; North-Holland, Amsterdam, Holland, 1953.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0002$01.00/0

SINGULARITIES OF SOLUTIONS TO EXTERIOR ANALYTIC BOUNDARY VALUE PROBLEMS FOR THE HELMHOLTZ EQUATION IN THREE INDEPENDENT VARIABLES. II: THE AXISYMMETRIC BOUNDARY*

R. F. MILLAR†

Abstract. A method is developed for locating singularities of solutions to exterior boundary value problems for the axisymmetric Helmholtz equation. Through the theory of characteristics, these real singularities are related to complex singularities in the Cauchy data on the boundary. The singularities in the unknown data are found by extending into the complex domain an integral equation satisfied by the data. Results are obtained for axisymmetric Green's functions for a sphere, and for oblate and prolate spheroids. Singularities are found at the image points of the prescribed singularities. For spheroids, the results are believed to be new.

1. Introduction. The singularities of a solution to an analytic partial differential equation play a fundamental role in determining its properties. Moreover, in the computation of the solution to a boundary value problem, knowledge of the location of the singularities (if not of their precise character) is often advantageous.

Recently, a procedure was developed for locating a priori the singularities of solutions to boundary value problems for second order analytic elliptic differential equations in two independent variables [8], [9]. Subsequently, the method was extended to three-dimensional, planar boundary value problems for the Helmholtz equation [10].

In the present paper, we consider a second special type of three-dimensional problem for the Helmholtz equation, namely the class of axisymmetric problems. With respect to a cylindrical coordinate system ρ, ϕ, z, with z as the axis of symmetry, their solutions satisfy the axisymmetric Helmholtz equation

$$u_{\rho\rho}+u_{zz}+\rho^{-1}u_\rho+k^2u=0. \tag{1}$$

The regularity properties of axially-symmetric, exterior (i.e., radiative) solutions to the Helmholtz equation have been studied by Colton [2], who obtained information about the singularities of a solution from the scattered far-field pattern. In particular, he determined domains that were free of singularities. Points on the axis of symmetry necessarily were excluded from consideration; this axis is a source of difficulty in the present work also. Colton's analysis was extended to the general three-dimensional problem by Sleeman [11], who reduced the problem to a sequence of axisymmetric problems. The domains of regularity thereby obtained are axially symmetric. The general, three-dimensional, vector (i.e., electromagnetic) problem has been discussed by Weston, Bowman, and Ar [15].

As in the earlier work, we are faced here with two problems. One is the determination of the singularities in the boundary data—the solution u and its outward normal derivative $\partial u/\partial \nu$—not both of which are known. This is accomplished by using an integral equation that relates the data, and extending it analytically into the complex domain. The second problem lies in relating singularities in the data to the real singularities of the solution. Here we use the fact that singularities are borne by the complex characteristics of (1).

* Received by the editors August 3, 1976, and in revised form August 8, 1977.

† Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. This work was supported by the National Research Council of Canada under Grant A-8808. A more detailed version of this paper may be obtained from the author by request.

The remainder of the paper is organized as follows. After some preliminary discussion in § 2 of notation and symmetry properties of the boundary and the data, we introduce in § 3 the integral equation satisfied by the data. The analytic continuation of this integral equation into the complex domain of the arclength parameter is performed in § 4 and, in the following section, singularities of the data are found. In §§ 6 and 7, singularities are located for some specific cases (sphere and spheroids); in particular, image singularities in the spheroids corresponding to an exterior axisymmetric Green's function are located. Some concluding and summarizing remarks are made in § 8. A brief discussion of several functions, and the effect of a change from arclength to another parameter, will be found in the appendices (§ 9).

2. Preliminaries. We consider radiative solutions (i.e., solutions that satisfy the Sommerfeld radiation condition) to the three-dimensional Helmholtz equation in the unbounded region D exterior to a smooth, closed, analytic surface Σ that possesses axial symmetry about the z-axis. The solution u is analytic in $D \cup \Sigma$, and is axially symmetric. Thus we need only examine u in some meridian plane Π.

The plane Π cuts Σ in a simple, closed analytic curve γ, which is symmetric with respect to the z-axis. A point on γ may be specified by its arclength s. We choose a coordinate origin inside Σ, and we measure s in the positive sense from the point where the negative z-axis meets γ. A generic point in space will be specified by cylindrical polar coordinates (ρ, ϕ, z), ϕ being measured from the meridian plane Π. In this plane, a point will be specified by Cartesian coordinates (ρ, z) and, for a point on γ, we have $\rho = \rho(s)$, $z = z(s)$. We shall denote the part of γ in $\rho > 0$ by σ. If $2l$ is the length of γ, the functions $\rho(s)$ and $z(s)$ have period $2l$ and are holomorphic for real s; $\rho(s)$ is odd and $z(s)$ is even; we shall assume that $\rho(s) > 0$ for $0 < s < l$. Denote u and $\partial u/\partial \nu$ on γ by $u(s)$ and $v(s)$ respectively. Then $u(s)$ and $v(s)$ are even, holomorphic for real s, and have period $2l$.

3. Integral equation for the data. For some common boundary conditions—Dirichlet, Neumann, or linear—we are able to locate singularities in the boundary data without first determining the unknown, by using an integral equation that is satisfied by the data.

Let T denote a point in Π, exterior to γ, with rectangular coordinates (ρ_0, z_0), $\rho_0 \geqq 0$. Let P be a point in Σ with cylindrical coordinates (ρ, ϕ, z), ν be the outward unit normal vector to Σ at P, and let R be the distance from P to T. We suppose that u corresponds to a scattered field that satisfies the three-dimensional Helmholtz equation and has associated with it the time-dependence factor $e^{-i\omega t}$. Then we may represent u at T by Helmholtz's formula ([1, Chap. 1, § 4.2]):

$$4\pi u(T) = \int_\Sigma [u(P)\,\partial/\partial\nu - v(P)]\, e^{ikR}/R \, d\Sigma. \tag{2}$$

Here e^{ikR}/R is the fundamental solution to Helmholtz's equation that is outgoing for $kR \gg 1$.

Since $R > 0$, we may write the integral in (2) as a repeated integral:

$$4\pi u(T) = \int_\sigma [u(s)H(S, T) - v(s)G(S, T)]\rho(s)\, ds,$$

where S: $\rho = \rho(s)$, $z = z(s)$, is a point on σ,

$$G(S, T) \equiv 2\int_0^\pi e^{ikR}/R\, d\phi, \tag{3}$$

$$(4) \qquad H(S,T) \equiv 2\int_0^{\pi} \frac{\partial}{\partial \nu}(e^{ikR}/R)\,d\phi,$$

and we have written the surface element on Σ as $d\Sigma = \rho(s)\,d\phi\,ds$.

If neither S nor T lies on the axis of symmetry, and if T approaches S, we find that G behaves logarithmically, like a fundamental solution in two dimensions. Consequently, if T approaches a point on σ specified by arclength $t(0<t<l)$, we have

$$(5) \qquad 2\pi u(t) = \int_0^l [u(s)H(s,t) - v(s)G(s,t)]\rho(s)\,ds.$$

Here $G(s,t)$ denotes $G(S,T)$ for $S=(\rho(s), z(s))\in\sigma$, $T=(\rho(t), z(t))\in\sigma$, and similarly for $H(s,t)$. The integral is improper, but convergent, at $s=t$. Moreover, for $t=0$ or l we again obtain (5) which therefore is valid for $0\leqq t\leqq l$.

On employing an appropriate boundary condition, (5) yields an integral equation for the unknown data on $[0, l]$. We shall assume that this equation has been solved so that $u(s)$ and $v(s)$ are known on $[0, l]$, although precise knowledge of these is not needed.

Our next object is to find analytic relations in $\operatorname{Im} t>0$ and in $\operatorname{Im} t<0$ that reduce to (5) when $\operatorname{Im} t\to 0$. From the given fact that the boundary values are analytic (and therefore continuous) on σ, or from a direct examination of the continuity properties of (5), it follows by a well-known result ([12, § 4.51]) that these analytic relations are analytic continuations of each other, and of (5).

4. Analytic continuation of integral equation.

4.1. The function G. This is given by (3), with R defined for real s, t in $[0, l]$ as the nonnegative root of $R^2 \equiv \rho(s)^2+\rho(t)^2+[z(s)-z(t)]^2-2\rho(s)\rho(t)\cos\phi$. If we let $\alpha(s,t)\equiv 2\rho(s)\rho(t)$, $\beta(s,t)\equiv\rho(s)^2+\rho(t)^2+[z(s)-z(t)]^2$, $\zeta(s,t)\equiv\beta(s,t)/\alpha(s,t)$, then

$$\begin{aligned} R^2 &= \beta - \alpha\cos\phi \\ &= \alpha(\zeta - \cos\phi). \end{aligned}$$

As defined above, R is not an analytic function of s and t. To facilitate the continuation of (5), we introduce the analytic function r:

$$(6) \qquad r \equiv (\beta - \alpha\cos\phi)^{1/2},$$

where, for all real or complex s and t,

$$(7) \qquad r = |r|\,e^{i\psi},$$

with

$$(8) \qquad \psi \equiv \tfrac{1}{2}\arg(\beta - \alpha\cos\phi),$$

and

$$(9) \qquad \psi = 0 \quad \text{for} \quad 0<s<t<l.$$

Then we define a function L, for $0<s<t<l$ by

$$(10) \qquad L(s,t) \equiv 2\int_0^{\pi} e^{ikr}/r\,d\phi.$$

In a complex neighborhood of (s_0, t_0) with $0<s_0<t_0<l$, r does not vanish, and is analytic in s and t; consequently L is analytic in this neighborhood. By definition $L(s,t)=G(s,t)$, $0<s<t<l$, but we shall find that $L(s,t)\neq G(s,t)$ if $0<t<s<l$.

Equations (7) to (10) may be used to continue $L(s, t)$ analytically into the complex domain from an initial point (s_0, t_0) with $0<s_0<t_0<l$. Singularities of the *first kind* occur in L at the singular points of $\rho(s)$, $z(s)$, $\rho(t)$, or $z(t)$. Then, in continuing the integral (10), it may be necessary to deform the integration path (Γ, say) away from the real ϕ-interval $(0, \pi)$ to avoid singularities of the integrand. Singularities of the *second kind* may arise when Γ becomes pinched by singularities of the integrand; see, for example, [4, Chap. 1, § 3]. If $\alpha \to 0$ and ζ becomes unbounded, we find singularities of the *third kind*. Singularities of the first kind are easily found, and we shall not consider them further.

We discuss now a few properties of $L(s, t)$; for continuity of exposition, a fuller examination of L and other functions that arise is relegated to the appendices (§ 9).

As a function of ϕ, the singularities of r are branch points located where

$$\cos \phi = \zeta. \tag{11}$$

When $(s, t) \in (0, l)$, we have $\zeta > 1$ since $\alpha(s, t) > 0$. We define $\Phi(\equiv \Phi(\zeta))$ to be the root of (11) that is negative imaginary when $0<s<t<l$. Then the solutions to (11) are

$$\phi = \pm\Phi(\zeta) + 2n\pi, \qquad n = 0, \pm 1, \pm 2, \cdots. \tag{12}$$

The ϕ-plane is cut along the line segments that join Φ to $-\Phi$, and other corresponding pairs of solutions (12).

If s and t vary in a sufficiently small neighborhood of an initial real point (s_0, t_0), with $0<s_0<t_0<l$, the solutions to (11) move in a corresponding fashion. The contour Γ may be deformed to avoid a singularity unless a branch point approaches an endpoint of Γ. Thus we find possible singularities of the second kind if $\zeta = \pm 1$, that is, if $[\rho(s) \mp \rho(t)]^2 + [z(s) - z(t)]^2 = 0$. In each case, one endpoint of Γ is pinched by two branch points, and it may be shown that $L(s, t)$ is singular; see § 9.3.1.

We may also show that $L(s, t) \neq G(s, t)$, $0<t<s<l$, and in particular that

$$L(s, t \pm i0) = \mp \int_\Lambda e^{ikr}/r \, d\phi + G(s, t), \qquad 0<t<s<l. \tag{13}$$

Here Λ is a closed contour enclosing the cut between Φ and $-\Phi$, which is described in the anticlockwise sense. The result may be established first for t in a neighborhood of s, and then extended to $0<t<s$. Its proof requires a careful study of the loci of $\pm\Phi$, and makes use of the fact that r is an odd function of ϕ.

4.2. Analytic continuation of $\int_0^l v(s)G(s, t)\rho(s)\,ds$. Consider the function E, defined for small but nonzero $\operatorname{Im} t$ by

$$E(t) \equiv \int_0^l v(s)L(s, t)\rho(s)\,ds, \qquad 0<\operatorname{Re} t<l.$$

Since $L(s, t)$ is only logarithmically unbounded at $s = t$ (see § 9.3.1) this integral converges uniformly in t. Thus ([12, § 2.85]) E is analytic for $0<\operatorname{Re} t<l$ if $\operatorname{Im} t$ is sufficiently small but not zero.

By letting $\operatorname{Im} t \downarrow 0$ and $\uparrow 0$ in turn, we find

$$E(t \pm i0) = \int_0^l v(s)G(s, t)\rho(s)\,ds \mp \int_t^l v(s)M(s, t)\rho(s)\,ds, \qquad 0<t<l,$$

where

$$M(s, t) \equiv \int_\Lambda e^{ikr}/r \, d\phi \tag{14}$$

is analytic for s and t in complex neighborhoods of the real interval $(0, l)$; see § 9.1. Thus the continuation of $\int_0^l v(s)G(s, t)\rho(s)\,ds$ for $0<\operatorname{Re} t<l$ and for $\operatorname{Im} t$ sufficiently small is

$$\int_0^l v(s)L(s, t)\rho(s)\,ds \pm \int_t^l v(s)M(s, t)\rho(s)\,ds, \qquad \operatorname{Im} t \gtrless 0. \tag{15}$$

The path of integration from the complex point t to l is arbitrary, except that it must not sweep across any singularities of the integrand as t varies.

4.3. Analytic continuation of $\int_0^l u(s)H(s, t)\rho(s)\,ds$. We compare $H(s, t)$, given by (4), with the analytic function $N(s, t)$, where

$$\begin{aligned} N(s, t) &\equiv 2\int_\Gamma \frac{\partial}{\partial \nu}(e^{ikr}/r)\,d\phi \\ &= 2\int_\Gamma e^{ikr}(ikr-1)r^{-3}\{z'(s)[\rho(s)-\rho(t)\cos\phi]-\rho'(s)[z(s)-z(t)]\}\,d\phi \\ &= [z'(s)\,\partial/\partial\rho(s)-\rho'(s)\,\partial/\partial z(s)]L(s, t). \end{aligned} \tag{16}$$

In the previous manner, we find that $N(s, t)=H(s, t)$, $0<s<t<l$, and

$$N(s, t\pm i0) = \mp\int_\Lambda \frac{\partial}{\partial\nu}(e^{ikr}/r)\,d\phi + H(s, t), \qquad 0<t<s<l.$$

$N(s, t)$ has an integrable singularity at $s=t$; see § 9.4.1. We find that the continuation of $\int_0^l u(s)H(s, t)\rho(s)\,ds$ near the real axis is

$$\int_0^l u(s)N(s, t)\rho(s)\,ds \pm \int_t^l u(s)P(s, t)\rho(s)\,ds, \qquad \operatorname{Im} t \gtrless 0; \tag{17}$$

here

$$\begin{aligned} P(s, t) &\equiv \int_\Lambda \frac{\partial}{\partial\nu}(e^{ikr}/r)\,d\phi \\ &= [z'(s)\,\partial/\partial\rho(s)-\rho'(s)\,\partial/\partial z(s)]M(s, t). \end{aligned} \tag{18}$$

4.4. Analytic continuation of $u(t)$ and $v(t)$ near $0<t<l$. By using (15) and (17), we see that the continuation of (5) is

$$\begin{aligned} 2\pi u(t) &= \int_0^l [u(s)N(s, t)-v(s)L(s, t)]\rho(s)\,ds \\ &\quad \pm \int_t^l [u(s)P(s, t)-v(s)M(s, t)]\rho(s)\,ds, \qquad \operatorname{Im} t \gtrless 0. \end{aligned} \tag{19}$$

When $v(s)$ is prescribed, this becomes a linear Volterra integral equation for $u(t)$ in the complex domain. It is also useful in connection with the more general linear boundary condition: $v(s)=A(s)u(s)+B(s)$. For a Dirichlet boundary condition, we obtain a suitable relation by differentiating (19). Then

$$\begin{aligned} &2\pi u'(t) \pm \rho(t)[u(t)P(t, t)-v(t)M(t, t)] \\ &\quad = \int_0^l [u(s)N_t(s, t)-v(s)L_t(s, t)]\rho(s)\,ds \\ &\quad\quad \pm \int_t^l [u(s)P_t(s, t)-v(s)M_t(s, t)]\rho(s)\,ds, \qquad \operatorname{Im} t \gtrless 0. \end{aligned} \tag{20}$$

The differentiation is justifiable because $P(s, t)$ and $M(s, t)$ are suitably well-behaved near $s=t$; see §§ 9.1.1 and 9.2.1. When $s=t$ ($\neq 0$ or l), we have $\alpha=\beta=2\rho(t)^2$ and $r=2\rho(t)\sin\frac{1}{2}\phi$. The integral (14) may be evaluated by the method of residues to yield $M(t,t)=2\pi i/\rho(t)$; likewise (18) gives $P(t,t)=-\pi i z'(t)/\rho(t)^2$. Then (20) becomes

$$\begin{aligned} &2\pi[u'(t)\mp iv(t)]\mp \pi i z'(t)u(t)/\rho(t)\\ &\qquad=\int_0^l [u(s)N_t(s,t)-v(s)L_t(s,t)]\rho(s)\,ds\\ &\qquad\quad\pm\int_t^l [u(s)P_t(s,t)-v(s)M_t(s,t)]\rho(s)\,ds,\quad \operatorname{Im} t \gtrless 0. \end{aligned} \tag{21}$$

If $u(s)$ is prescribed, we obtain from (21) a linear Volterra integral equation for $v(t)$ in the complex domain.

The foregoing results are summarized in the following theorem.

THEOREM 1. *Let u and v denote the analytic boundary data for a radiative solution to the axisymmetric Helmholtz equation* (1) *in the region exterior to an analytic, axisymmetric boundary surface whose trace in a meridian plane is $\gamma:(\rho(s), z(s))$, $0\leqq s\leqq 2l$. Then the analytic continuations of u and v satisfy equations* (19) *and* (21) *in a complex neighborhood of* $0<t<l$.

5. Singularities of the boundary data. For brevity, we shall confine attention to the Neumann problem; the other linear boundary conditions introduce no essentially new difficulties. Then $v(t)$ is prescribed and analytic, and $u(t)$ is determined in some neighborhood of $0<t<l$ by (19), which we re-write as

$$2\pi u(t)\pm\int_l^t u(s)P(s,t)\rho(s)\,ds=A(t)\pm B(t),\qquad \operatorname{Im} t\gtrless 0. \tag{22}$$

where

$$A(t)\equiv\int_0^l [u(s)N(s,t)-v(s)L(s,t)]\rho(s)\,ds, \tag{23}$$

and

$$B(t)\equiv\int_l^t v(s)M(s,t)\rho(s)\,ds. \tag{24}$$

Since $u(s)$ is assumed to have been determined from (5) for $0\leqq s\leqq l$, A and B are, in principle, known.

We may continue $A(t)$ out of the neighborhood of $0<t<l$ by standard means. It is convenient to keep s real. The representation for $A(t)$ may differ from (23) in different regions of the t-plane, because poles and other singularities of N and L may produce additional terms. The function B may be analytically continued in a similar manner, except that here the path of integration is not real, and the value $B(t)$ may depend on the choice of contour. Thus A and B may be multi-valued. Consequently, $A+B$ can be continued in $\operatorname{Im} t>0$ on a Riemann surface $\mathscr{R}_+$, and $A-B$ can be extended in $\operatorname{Im} t<0$ on a Riemann surface $\mathscr{R}_-$.

Consider now the integral equation (22), in $\operatorname{Im} t>0$. The right-hand side is known and holomorphic on $\mathscr{R}_+$. Since the sets of singular points of $M(s,t)$ and $P(s,t)$ coincide, and because the integration path in the integral in (22) is identical to that in (24), $P(s,t)$ is holomorphic for s on the integration path and $t\in\mathscr{R}_+$. Consequently, the general theory of such integral equations in the complex domain (for example, [13, p. 11, § 3]) and uniform convergence of the integral at $s=l$ allows us to conclude that u is holomorphic on $\mathscr{R}_+$. And similarly, u is holomorphic on $\mathscr{R}_-$ in $\operatorname{Im} t<0$.

It is possible to show that u and v have period $2l$ and other properties mentioned in § 2. Thus the singularities of u and v are distributed with period $2l$.

The above general discussion may be specialized to determine the possible singularities of u in a given situation. We shall now apply these considerations to a few specific problems.

6. Singularities for a sphere.

6.1. Singularities of the data; Green's function. For a sphere of radius a, we have

$$\rho(s)=a \sin (s/a), \qquad z(s)=-a \cos (s/a), \tag{25}$$

and there are no singularities of the first kind. All singularities of the third kind are real, and do not correspond to singularities in the data. The singularities of the second kind, which correspond to $\zeta=\pm 1$, are given by $s=\pm t+2n\pi a$, respectively, where $n=0, \pm 1, \pm 2, \cdots$. Consequently $A(t)$ may be continued indefinitely into $\operatorname{Im} t>0$, and into $\operatorname{Im} t<0$. Moreover, $M(s, t)$ has singularities of the second kind only for $\zeta=-1$, and the path of integration in (24) is never pinched if $\operatorname{Im} t \neq 0$. Thus, if v is entire, $B(t)$ may be continued indefinitely. From (22), we conclude that $u(t)$ is holomorphic in $\operatorname{Im} t>0$ and in $\operatorname{Im} t<0$. On the real axis, $u(t)$ is known to be analytic by other considerations; or arguments like those in the last paragraph of § 3 can be used to prove analyticity. Thus $u(t)$ is entire and, since $u(t+2\pi a)=u(t)$, any singularities are at infinity in $\operatorname{Im} t>0$ and in $\operatorname{Im} t<0$.

If v is not entire, B will generally be singular at the same points as v, and u will be singular at these points also. For example, consider an exterior Green's function of the second kind, with singularity at $\rho=\rho_0(>0)$, $z=z_0$, and $\rho_0^2+z_0^2>a^2$. We express this function as $U+u$ where

$$U(\rho, z)=\int_0^{2\pi} e^{ikR}/R \, d\phi,$$

and $R^2=\rho^2+\rho_0^2+(z-z_0)^2-2\rho\rho_0 \cos \phi$. The normal derivative of the Green's function vanishes on the sphere, so

$$v(s)=2 \int_0^{\pi} e^{ikR}(ikR-1)R^{-3}\{[z(s)-z_0]\rho'(s)-[\rho(s)-\rho_0 \cos \phi]z'(s)\} \, d\phi. \tag{26}$$

This integral is quite similar to that which defines $N(s, t)$ in (16), and its singularities in the finite plane may be located in like manner. We conclude that possible singularities of v occur at those s for which $R=0$ when $\phi=0$ or $\phi=\pi$; that is, when

$$\rho(s)^2+\rho_0^2+[z(s)-z_0]^2=\pm 2\rho_0\rho(s). \tag{27}$$

We insert (25) into (27) and solve for s. The four sets of solutions are given by

$$\rho_0-iz_0=ia \exp (\pm is/a) \tag{28}$$

and

$$\rho_0+iz_0=-ia \exp (\pm is/a). \tag{29}$$

For $\phi=0$ or π (as the case may be), it is not difficult to see that the numerator of the integrand in (26) does not vanish for those s determined by (28) and (29). These, therefore, are singularities of v; u will be singular at the same points and at infinity.

6.2. Real singularities of the solution. Here we employ the theory of characteristics in the complex domain (see, for example, [3, Chap. 16]) to relate complex singularities in the data to real singularities of the solution. We make use of the property that a singularity in data that are independent is borne by the characteristics of (1) that emanate from the singular point. This result has only been proved locally for analytic equations, the data for which are singular on an analytic manifold imbedded in the initial analytic manifold; see [14] for recent work and references. However, for (1) the characteristics are parallel complex hyperplanes, and the coefficients are holomorphic except on $\rho = 0$. Consequently, the local results may be extended step-by-step at least up to $\rho = 0$. The data at each step will be related in such a manner that a singularity will be carried forward only on the appropriate characteristic. Similarly, because the data for a boundary value problem are not independent, a singularity in the boundary data may not be borne by every characteristic emanating from it; in a given case, we shall use plausibility arguments to decide whether or not a characteristic is significant in this respect.

Singularities in the solution may also arise if the initial manifold is characteristic at some of its points [5], [7]. It has been shown [9, p. 119] that these circumstances do not prevail at nonsingular points of the manifold obtained if the boundary curve is parametrized by its arclength.

The boundary is specified by (25). The characteristics through $(\rho(s), z(s))$ are

$$\rho \pm iz = \rho(s) \pm iz(s). \tag{30}$$

They meet the real ρ, z-domain in (ρ_+, z_+) and (ρ_-, z_-), where $\rho_\pm = a \sin(s_1/a) \exp(\pm s_2/a)$, $z_\pm = -a \cos(s_1/a) \exp(\pm s_2/a)$, and $s = s_1 + is_2$. These points are inverse (or images) with respect to the circle (25). When u and v are entire, there are no singularities in the data for finite s, and the interior of the sphere is singularity free, except possibly on $\rho = 0$. In fact, there is such a singularity, since a nontrivial radiative solution that is C^2 on and outside the sphere must be singular at some interior point.

In the case of the Green's function, u and v have singularities at points determined by (28) and (29). By solving (30) with these, we find that the only significant solutions are the real points $(\pm\rho_1, z_1)$, where $\rho_1 = a^2\rho_0/(\rho_0^2 + z_0^2)$, $z_1 = a^2 z_0/(\rho_0^2 + z_0^2)$. These are the image of (ρ_0, z_0) in the circle, and its reflection in the z-axis.

7. Spheroids. The treatment of the problem for a spheroid is similar to that for a sphere. However, it is inappropriate to take arclength as the parameter; the changes occasioned by introduction of a new parameter are described in § 9.5. We shall omit details, and content ourselves chiefly with a statement of results.

7.1. Oblate spheroidal boundary. A point on σ is $(\rho(\theta), z(\theta))$, where

$$\rho(\theta) = a \cos\theta, \qquad z(\theta) = b \sin\theta, \tag{31}$$

with $a > b$ and $-\pi/2 < \theta < \pi/2$. For a Neumann problem, the continuation of u is performed with the help of (A.12). In general, w (defined by (A.14)) and u will be singular where $\rho'(\theta)^2 + z'(\theta)^2 = 0$, for which

$$\theta = \pm i\eta + n\pi, \qquad n = 0, \pm 1, \pm 2, \cdots, \tag{32}$$

and

$$\eta = \tanh^{-1}(b/a). \tag{33}$$

The singularities of L, M, N and P, and their effect, may be determined as before. If w is singularity free with exception of the points (32), we find that u is holomorphic in the strip $|\mathrm{Im}\,\theta|<\eta$. There are singularities in $|\mathrm{Im}\,\theta|>\eta$ and, in particular, on $\mathrm{Im}\,\theta=\pm 2\eta$. For our purposes, the strip $|\mathrm{Im}\,\theta|<\eta$ is of most interest.

For the exterior Green's function of the second kind with singularity at (ρ_0, z_0), we find singularities in v (and in u) for values of θ that satisfy any of the following four equations:

$$e^{i\theta}=\{\rho_0-iz_0\pm[(\rho_0-iz_0)^2-c^2]^{1/2}\}/(a-b), \tag{34}$$

$$e^{i\theta}=\{\rho_0+iz_0\pm[(\rho_0+iz_0)^2-c^2]^{1/2}\}/(a+b), \tag{35}$$

$$e^{i\theta}=\{-\rho_0-iz_0\pm[(\rho_0+iz_0)^2-c^2]^{1/2}\}/(a-b), \tag{36}$$

$$e^{i\theta}=\{-\rho_0+iz_0\pm[(\rho_0-iz_0)^2-c^2]^{1/2}\}/(a+b), \tag{37}$$

with $c^2=a^2-b^2$.

The characteristics through $(\rho(\theta), z(\theta))$ are

$$\rho\pm iz=a\cos\theta\pm ib\sin\theta. \tag{38}$$

We confine attention to the strip S: $-\pi/2\leqq \mathrm{Re}\,\theta<3\pi/2$. For the upper sign in (38), $\mathrm{Im}\,\theta<0$ maps into the exterior of the ellipse (31) and $0<\mathrm{Im}\,\theta<\eta$ maps into the interior. The $\rho+iz$-plane is cut from $-c$ to c. Points in S for $\mathrm{Im}\,\theta>\eta$ map into another sheet of a Riemann surface on the cut plane. The mapping for the lower sign in (38) is obtained from this by reflecting θ in the real axis.

The relevant real singularities that correspond to (32) are the focal singularities at $\rho=\pm c$, $z=0$. Since (32) also determines points on the initial manifold that are characteristic, the focal points are branch points of the solution in the real domain; the solution will be multi-valued across the disc $0\leqq\rho\leqq c$, $z=0$. Singularities in the data for $|\mathrm{Im}\,\theta|>\eta$ will correspond to singularities in the continuation of the solution across this disc.

We write $\rho_0=a_0\cos\theta_0$, $z_0=b_0\sin\theta_0$, $a_0=c\cosh\phi_0$, $b_0=c\sinh\phi_0$, $a=c\cosh\phi$, $b=c\sinh\phi$, with $-\pi/2\leqq\theta_0\leqq\pi/2$, $\phi_0>0$, $\phi>0$. Then the significant real singularities that correspond to (34) to (37) are found at $(\pm\rho_1, z_1)$, where

$$\rho_1=a_1\cos\theta_0,\qquad z_1=b_1\sin\theta_0, \tag{39}$$

with

$$a_1=c\cosh(2\phi-\phi_0),\qquad b_1=c\sinh(2\phi-\phi_0). \tag{40}$$

The points (ρ_0, z_0) and $(\pm\rho_1, z_1)$ lie on the hyperbola

$$\rho^2\sin^2\theta_0-z^2\cos^2\theta_0=c^2\sin^2\theta_0\cos^2\theta_0. \tag{41}$$

These results seem to be new.

7.2. Prolate spheroidal boundary. Here $\rho(\theta)$ and $z(\theta)$ are given by (31), with a and b interchanged. The singularities that correspond to (32) are

$$\theta=\pm i\eta+(n+\tfrac{1}{2})\pi,\qquad n=0,\pm1,\pm2,\cdots, \tag{42}$$

where η is defined by (33). For the exterior Green's function, the solutions for $e^{i\theta}$ may be found by interchanging a and b in (34) to (37).

The real singularities that correspond to (42) are at the foci $\rho=0$, $z=\pm c$. Again, (42) determines points on the initial manifold that are characteristic. Thus we are led to expect multi-valued behavior of the solution near the foci. Since these lie on $\rho=0$ where the differential equation is singular, the actual behavior is logarithmic; see, for example, [6, p. 56]. Our present procedure does not enable us to verify this singular behavior.

For the Green's function, the additional real singularities at $(\pm\rho_1, z_1)$ may be obtained by interchanging a_1 and b_1 in (39):

(43) $$\rho_1 = b_1 \cos\theta_0, \qquad z_1 = a_1 \sin\theta_0.$$

Evidently (ρ_0, z_0) and $(\pm\rho_1, z_1)$ lie on the hyperbola conjugate to (41).

8. Concluding remarks. We have developed a method for locating the real singularities of axisymmetric solutions to exterior boundary value problems for the three-dimensional Helmholtz equation. We have shown that the image singularities in oblate and prolate spheroids of an axially symmetric, exterior Green's function of the second kind are given by (39) and (40), and (43) and (40), respectively. These results are believed to be new. Our analysis is incomplete with regard to singularities on the axis of symmetry. Their presence or absence cannot be established without further study.

No attempt has been made to extend the analysis to $n(>3)$ dimensions; certainly for $k=0$ this should not be difficult. Nor have we examined the form of the boundary data in the neighborhood of a singular point with the aid of (19) and (21). We note that these representations could be used as the basis for a stable procedure for performing the analytic continuation of the unknown data into the complex domain.

Finally, we observe that even apart from the possibility of singularities on the axis of symmetry, the existence of singularities of the third kind ($\rho(s)=0$ or $\rho(t)=0$) distinguishes an axially symmetric boundary problem from the corresponding strictly two-dimensional problem for the Helmholtz equation. For example, if the boundary is determined by $\rho=\rho(\theta)$, $z=z(\theta)$, where $\rho(\theta)=5\cos\theta-\cos 3\theta$, $z(\theta)=\sin\theta$ for $-\pi/2 \leqq \theta \leqq \pi/2$, we find singularities of the third kind at $\theta = n\pi - i\log(\sqrt{2}\pm 1)$, $n=0, \pm1, \pm2, \cdots$, in addition to solutions on the real axis of θ. These give rise to possible real singularities at the interior points $(\pm1, 0)$; there are no corresponding singularities in the two-dimensional problem.

9. Appendices.

9.1. The function *M*. Of the four functions L, M, N, and P, the most simple is M, and we shall first describe some of its properties. It is defined by (14) with r determined by (6), (7), and (8).

In general, M has singularities of the first kind at points where any of $\rho(s)$, $z(s)$, $\rho(t)$, $z(t)$ are singular; these are found in a straightforward manner. Singularities of the second kind occur when Λ is pinched. Finally, if α $(\equiv 2\rho(s)\rho(t))=0$, a singularity of the third kind occurs.

9.1.1. Singularities of the second kind. If $\zeta=-1$, the branch cut enclosed by Λ runs from $-\pi$ to π, Λ is caught between coalescing pairs of singularities at $\phi=\pm\pi$, and $M(s,t)$ is singular. We estimate the integral along segments of Λ that are pinched. For $\alpha\neq 0$, $\beta\neq 0$, we find

(A1) $$M(s,t) \sim -2i[\rho(s)\rho(t)]^{-1/2}\log\{[\rho(s)+\rho(t)]^2+[z(s)-z(t)]^2\}$$

as $\zeta\to -1$. There is no singularity as $\zeta\to 1$.

9.1.2. Singularities of the third kind. Here $\alpha\to 0$. If $\beta\to 0$ and $\zeta\to -1$, (A1) again gives the singular behavior. If $\beta\to 0$ and $\zeta(\neq -1)$ remains bounded, we find

(A2) $$M(s,t)=\alpha^{-1/2}E(s,t)+F(s,t),$$

where E and F are analytic. If $\beta\not\to 0$, then ζ is unbounded. We find

(A3) $$M(s,t)\sim -2i\beta^{-1/2}\cos(k\beta^{1/2})\log\alpha,$$

as $\alpha\to 0$, $\beta\neq 0$.

9.2. The function *P*. This function is defined by (18):

$$P(s,t)=[z'(s)\,\partial/\partial\rho(s)-\rho'(s)\,\partial/\partial z(s)]M(s,t). \tag{A4}$$

We use (A4) to determine the dominant singular behavior of P from corresponding results for M.

9.2.1. Singularities of the second kind. Here neither α nor β vanishes, and $\zeta\to-1$. From (A1), we find that

$$P(s,t)\sim-\frac{4i}{[\rho(s)\rho(t)]^{1/2}}\cdot\frac{z'(s)[\rho(s)+\rho(t)]-\rho'(s)[z(s)-z(t)]}{[\rho(s)+\rho(t)]^2+[z(s)-z(t)]^2}. \tag{A5}$$

Terms that are logarithmically singular at $\zeta=-1$ have been omitted. The second factor of the denominator in (A5) vanishes when $\zeta=-1$. In general, $P(s,t)$ has a pole, as well as a logarithmic singularity. But if $z(s)=z(t)$, the pole disappears; in particular, this occurs on $s=-t$.

9.2.2. Singularities of the third kind. If $\beta\to0$ simultaneously with α, and if $\zeta\to-1$, (A5) describes the singular behavior; if $\zeta\not\to-1$, then (A2) and (A4) give the behavior of $P(s,t)$. When $\alpha\to0$ but $\beta\not\to0$, we find that $P(s,t)$ has $\log\alpha$ behavior.

9.3. The function *L*. Here

$$L(s,t)=2\int_\Gamma e^{ikr}/r\,d\phi,$$

where Γ runs between 0 and π.

9.3.1. Singularities of the second kind. *If* $\zeta\to-1$ ($\alpha\neq0$), the situation resembles that for $M(s,t)$. We find

$$L(s,t)\sim C[\rho(s)\rho(t)]^{-1/2}\log\{[\rho(s)+\rho(t)]^2+[z(s)-z(t)]^2\}, \tag{A6}$$

as $\zeta\to-1$. Here C is a constant. If $\zeta\to1$, we obtain

$$L(s,t)\sim-[\rho(s)\rho(t)]^{-1/2}\log\{[\rho(s)-\rho(t)]^2+[z(s)-z(t)]^2\}. \tag{A7}$$

9.3.2. Singularities of the third kind. If $\beta\to0$ with α and $\zeta\to-1$ or 1, then (A6) or (A7) again is valid. If $\beta\not\to0$, ζ becomes unbounded and the branch points recede to infinity. In circumstances such that Γ may be taken to be the real interval $(0,\pi)$, we find that

$$L(s,t)\to2\pi\beta^{-1/2}\exp(ik\beta^{1/2}), \tag{A8}$$

and L is not singular. On the other hand, when Γ winds around the cut, $L(s,t)$ is a multiple of $M(s,t)$ plus an integral on $(0,\pi)$. In this case, the singular behavior of $L(s,t)$ is given by the appropriate multiple of (A3).

9.4. The function *N*. In accordance with (16), we have

$$N(s,t)=[z'(s)\,\partial/\partial\rho(s)-\rho'(s)\,\partial/\partial z(s)]L(s,t) \tag{A9}$$

and properties of N follow from the appropriate results of § 9.3.

9.4.1. Singularities of the second kind. When $\alpha\neq0$, $\beta\neq0$, and $\zeta\to-1$, (A6) and (A9) give

$$N(s,t)\sim\frac{4C}{[\rho(s)\rho(t)]^{1/2}}\cdot\frac{z'(s)[\rho(s)+\rho(t)]-\rho'(s)[z(s)-z(t)]}{[\rho(s)+\rho(t)]^2+[z(s)-z(t)]^2}. \tag{A10}$$

Here terms involving $\log(\zeta+1)$ have been omitted. In general, $N(s, t)$ has a pole at $\zeta=-1$ but, in circumstances that have already been noted after (A5), this pole dissolves and the logarithmic terms dominate. In case $\zeta\to 1$, we find that

$$N(s,t)\sim-\frac{2}{[\rho(s)\rho(t)]^{1/2}}\cdot\frac{z'(s)[\rho(s)-\rho(t)]-\rho'(s)[z(s)-z(t)]}{[\rho(s)-\rho(t)]^2+[z(s)-z(t)]^2}, \tag{A11}$$

plus terms in $\log(\zeta-1)$. A solution to $\zeta=1$ is $s=t$; here, too, the pole vanishes, and $N(s, t)$ is only logarithmically singular.

9.4.2. Singularities of the third kind. Here $\alpha\to 0$. If $\beta\to 0$ and $\zeta\to -1$, (A10) is again appropriate. Similarly, if $\zeta\to 1$, (A11) is valid. If $\beta\not\to 0$, and Γ is the real interval $(0, \pi)$, (A8) and (A9) show that $N(s, t)$ is not singular. Finally, if Γ winds around the cut, then $N(s, t)$ is a multiple of $P(s, t)$ plus the integral from 0 to π. The singularity of $N(s, t)$ is this multiple of the logarithmic terms occurring in $P(s, t)$.

9.5. Change of parameter. Here we shall note the changes that are necessary when a parameter (θ, say) different from arclength is introduced.

Since no confusion with $\rho(s)$ and $z(s)$ should arise, we assume that a point on the boundary curve γ is given by $\rho=\rho(\theta)$, $z=z(\theta)$. We suppose that θ ranges between $-\pi/2$ and $\pi/2$ as s runs from 0 to l on γ, and that $\rho(\theta)$ and $z(\theta)$ have period 2π as well as other appropriate properties. At a point on γ specified by the parameter θ, we shall denote u and $\partial u/\partial\nu$ by $u(\theta)$ and $v(\theta)$, respectively. Let $N(\theta,\chi)$ denote the function obtained from $N(s, t)$ by formally replacing $\rho(s)$, $z(s)$, $\rho'(s)$, and $z'(s)$ by $\rho(\theta)$, $z(\theta)$, $\rho'(\theta)$, and $z'(\theta)$, respectively, and $\rho(t)$, $z(t)$, $\rho'(t)$, $z'(t)$ by $\rho(\chi)$, $z(\chi)$, $\rho'(\chi)$, $z'(\chi)$, respectively; and similarly for $L(\theta,\chi)$, $M(\theta,\chi)$, and $P(\theta,\chi)$. Then the analogue of (19) is

$$\begin{aligned}2\pi u(\chi)&=\int_{-\pi/2}^{\pi/2}[u(\theta)N(\theta,\chi)-w(\theta)L(\theta,\chi)]\rho(\theta)\,d\theta\\&\quad\pm\int_{\chi}^{\pi/2}[u(\theta)P(\theta,\chi)-w(\theta)M(\theta,\chi)]\rho(\theta)\,d\theta,\qquad \operatorname{Im}\chi\gtrless 0,\end{aligned}\tag{A12}$$

and the analogue of (21) is

$$\begin{aligned}&2\pi[u'(\chi)\mp iw(\chi)]\mp\pi iz'(\chi)u(\chi)/\rho(\chi)\\&\quad=\int_{-\pi/2}^{\pi/2}[u(\theta)N_\chi(\theta,\chi)-w(\theta)L_\chi(\theta,\chi)]\rho(\theta)\,d\theta\\&\quad\quad\pm\int_{\chi}^{\pi/2}[u(\theta)P_\chi(\theta,\chi)-w(\theta)M_\chi(\theta,\chi)]\rho(\theta)\,d\theta,\qquad \operatorname{Im}\chi\gtrless 0.\end{aligned}\tag{A13}$$

Here

$$w(\theta)\equiv[\rho'(\theta)^2+z'(\theta)^2]^{1/2}v(\theta).\tag{A14}$$

In general $\rho'(\theta)^2+z'(\theta)^2\not\equiv 1$, and $v(\theta)$ may have singularities additional to those of $w(\theta)$ at points where $\rho'(\theta)^2+z'(\theta)^2=0$.

REFERENCES

[1] B. B. BAKER AND E. T. COPSON, *The Mathematical Theory of Huygens' Principle*, Clarendon Press, Oxford, 2nd ed., 1950.

[2] D. COLTON, *On the inverse scattering problem for axially symmetric solutions of the Helmholtz equation*, Quart. J. Math. Oxford, Ser. 2, 22 (1971), pp. 125–130.

[3] P. R. GARABEDIAN, *Partial Differential Equations*, John Wiley, New York, 1964.

[4] R. P. GILBERT, *Function Theoretic Methods in Partial Differential Equations*, Academic Press, New York, 1969.

[5] L. GÅRDING, T. KOTAKE AND J. LERAY, *Uniformisation et développement asymptotique de la solution du problème de Cauchy linéaire à données holomorphes; analogie avec la théorie des ondes asymptotiques et approchées* (*Problème de Cauchy, I bis et VI*), Bull. Soc. Math. France, 92 (1964), pp. 263–361.

[6] O. D. KELLOGG, *Foundations of Potential Theory*, Dover, New York, 1953.

[7] J. LERAY, *Uniformisation de la solution du problème linéaire analytique de Cauchy près de la variété qui porte les données de Cauchy* (*Problème de Cauchy I*), Bull. Soc. Math. France, 85 (1957), pp. 389–429.

[8] R. F. MILLAR, *Singularities of two-dimensional exterior solutions of the Helmholtz equation*, Proc. Cambridge Philos. Soc., 69 (1970), pp. 167–188.

[9] ———, *Singularities of solutions to linear, second order analytic elliptic equations in two independent variables I*, Applicable Anal., 1 (1971), pp. 101–121.

[10] ———, *Singularities of solutions to exterior analytic boundary value problems for the Helmholtz equation in three independent variables I*, this Journal, 7 (1976), pp. 131–156.

[11] B. D. SLEEMAN, *The three-dimensional inverse scattering problem for the Helmholtz equation*, Proc. Cambridge Philos. Soc., 73 (1973), pp. 477–488.

[12] E. C. TITCHMARSH, *The Theory of Functions*, Oxford University Press, 2nd ed., reprinted 1949.

[13] I. N. VEKUA, *New Methods for Solving Elliptic Equations*, Wiley Interscience, New York, 1967.

[14] C. WAGSCHAL, *Sur le problème de Cauchy ramifié*, J. Math. Pures Appl., 53 (1974), pp. 147–163.

[15] V. H. WESTON, J. J. BOWMAN AND E. AR, *On the electromagnetic inverse scattering problem*, Arch. Rational Mech. Anal., 31 (1968/69), pp. 199–213.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0003$01.00/0

A SINGULAR SINGULARLY-PERTURBED LINEAR BOUNDARY VALUE PROBLEM*

R. E. O'MALLEY, JR.†

Abstract. We consider the asymptotic solution of boundary value problems for the vector system

$$\dot{x} = A(t, \varepsilon)x + B(t, \varepsilon)y + C(t, \varepsilon),$$
$$\varepsilon\dot{y} = E(t, \varepsilon)x + F(t, \varepsilon)y + G(t, \varepsilon)$$

as $\varepsilon \to 0$ under the assumption that the matrix $F(t, 0)$ is singular. A full set of asymptotic solutions is constructed assuming that $F(t, 0)$ can be block-diagonalized, the reduced problem is consistent, and a new stability condition holds. Boundary value problems are then solvable if an appropriate "boundary" matrix is nonsingular for $\varepsilon \neq 0$. Such problems arise in optimal control theory, among other applications.

1. Introduction. Let us consider a linear system of the form

$$\begin{aligned} \dot{x} &= A(t, \varepsilon)x + B(t, \varepsilon)y + C(t, \varepsilon), \\ \varepsilon\dot{y} &= E(t, \varepsilon)x + F(t, \varepsilon)y + G(t, \varepsilon) \end{aligned} \tag{1}$$

for vectors x and y of dimensions n and m, respectively, for a small positive parameter ε, and for a finite t interval, say $0 \leqq t \leqq 1$. It is natural to consider (1) subject to a list of $n + m$ linearly independent boundary conditions of the form

$$\sum_{j=0}^{1} (R_j(\varepsilon)x(j) + S_j(\varepsilon)y(j)) = c(\varepsilon) \tag{2}$$

and study the asymptotic solution of (1), (2) as $\varepsilon \to 0$.

We recall that rather classical methods can be used to solve asymptotically the "regular" singularly-perturbed problem (1), (2) when $F(t, 0)$ satisfies an exponential dichotomy, i.e., its eigenvalues have either a positive or a negative real part throughout $0 \leqq t \leqq 1$ (cf., e.g., O'Malley [16], Harris [9], or Ferguson [5]). When $F(t, 0)$ is everywhere stable, for example, they show that the initial value problem for (1) has a unique solution which converges as $\varepsilon \to 0$ for $t > 0$ to the solution of the reduced system

$$\begin{aligned} \dot{X}_0 &= A(t, 0)X_0 + B(t, 0)Y_0 + C(t, 0), \\ 0 &= E(t, 0)X_0 + F(t, 0)Y_0 + G(t, 0) \end{aligned} \tag{3}$$

subject to the initial condition $X_0(0) = x(0)$. For the analogous terminal value problem, however, the solution would then be exponentially large as $\varepsilon \to 0$ for $t < 1$. More generally, boundary layers (regions of nonuniform convergence) of thickness $O(\varepsilon)$ must be expected at each endpoint for regular problems when the limiting solution within $(0, 1)$ is bounded, and this limiting solution must satisfy the reduced system (3) and n boundary conditions determined by an appropriate combination of the original conditions (2) evaluated at $\varepsilon = 0$. Because such limiting solutions involve only n boundary conditions, m linearly independent solutions of the homogeneous form of (1) are of boundary layer type, i.e., they are asymptotically negligible away from the endpoints. Considerable progress, then, has been made in determining which regular singularly-perturbed linear boundary value problems have limiting solutions as $\varepsilon \to 0$, what

* Received by the editors July 6, 1977 and in revised form May 4, 1978.
† Department of Mathematics and Committee on Applied Mathematics, University of Arizona, Tucson, Arizona 85721. This work was supported in part by the Office of Naval Research under Contract N00014-C-0326.

boundary value problems these limits satisfy within $(0, 1)$, and the nature of the endpoint boundary layers. The results are, however, more complicated than for scalar problems (cf. O'Malley [17] and O'Malley and Keller [21]). In addition to its direct utility, that information is useful in analyzing nonlinear problems (cf. Hoppensteadt [10]) and in designing numerical algorithms (cf. Flaherty and O'Malley [6]).

Here, we shall consider "singular" problems where $F(t, 0)$ is singular and of constant rank throughout $0 \leqq t \leqq 1$. (The double use of the word singular is unfortunate, but there seems no more natural alternative.) Their analysis and the behavior of their solutions are considerably more complicated than for regular problems. Specifically, we shall find that the asymptotic analysis of singular problems involves a consistency condition which did not occur for regular problems, a new stability requirement, and the occurrence of other (thicker) boundary layer regions of nonuniform convergence. These singular problems are less complicated, however, than turning point problems where $F(t, 0)$ is singular at isolated points (cf. Levinson [14], Wasow [29], and Olver [15]). Fundamental matrices for homogeneous systems (1) without turning points can be constructed as in Turrittin [25], and they could, in theory, be used to solve asymptotically nonhomogeneous problems via variation of parameters.

Our interest in such problems arose in analyzing nearly singular optimal control problems (cf. O'Malley and Jameson [20]) and in devising methods for the numerical integration of stiff differential equations (cf. Flaherty and O'Malley [6], [7]). The technique we use generalizes that developed for singular arc computations (cf. Goh [8] and Robbins [22]). Closely related methods are given in Vasil'eva [26], O'Malley and Flaherty [19] and O'Malley [18] for certain nonlinear problems.

The simplest control example is given by

$$\dot{x} = -y, \qquad x(1) = 0,$$
$$\varepsilon \dot{y} = -x, \qquad y(0) = -1.$$

Here, $-x/\varepsilon$ represents an optimal control and y, the corresponding state of a nearly singular control problem (cf. O'Malley and Jameson [20]). The solution is

$$x(t) = -\sqrt{\varepsilon}\,(e^{-t/\sqrt{\varepsilon}} - e^{-1/\sqrt{\varepsilon}}\, e^{-(1-t)/\sqrt{\varepsilon}})/(1 + e^{-2/\sqrt{\varepsilon}})$$

and $y = -\dot{x}$, so the limiting solution

$$(x(t), y(t)) \sim -(\sqrt{\varepsilon}, 1)\, e^{-t/\sqrt{\varepsilon}}$$

is asymptotocally trivial for $t > 0$ and features an $O(\sqrt{\varepsilon})$ boundary layer at $t = 0$. Note, in particular, that the corresponding control acts like an initial delta–function impulse.

2. A transformed problem and the corresponding reduced system. Under rather mild assumptions, the singular matrix $F(t, 0)$ can be block-diagonalized (cf. Sibuya [23], [24] and Chap. VII of Wasow [29]). We shall simply assume

H1 *that there exists a smooth nonsingular matrix $P(t)$ such that*

$$P^{-1}(t)F(t, 0)P(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & F_5(t, 0) & 0 \\ 0 & 0 & F_9(t, 0) \end{bmatrix} \tag{4}$$

where $-F_5$ and F_9 are stable matrices (i.e., their eigenvalues have negative real parts) *throughout $0 \leqq t \leqq 1$ of dimensions $m_2 \times m_2$ and $m_3 \times m_3$, respectively, with $m = m_1 + m_2 + m_3$, $m_1 > 0$.*

Hypotheses guaranteeing the existence of P are given in Wasow [28] and elsewhere. (We note that an analogous trichotomy was used by Hoppensteadt and Miranker [11], except that they allowed $F(t, 0)$ to have purely imaginary, but nonzero, eigenvalues.) In analogy with the parallel situation in singular optimal control (cf. Jacobson [12] or Anderson [1]), we might call problems where $m = m_1$ totally singular and those where $m > m_1$ partially singular. (If either $m_2 = 0$ or $m_3 = 0$, it may be more convenient to use a singular value decomposition $F = UDV$ where U and V are orthogonal and D is diagonal with the eigenvalues of $\sqrt{F'F}$. One then uses the transformed variable $x = U'y$ (cf. O'Malley [18]).

In general, we introduce

$$P^{-1}y = (y_1' \quad y_2' \quad y_3')' \tag{5}$$

in (1) (with the prime denoting transposition) and obtain the equivalent system

$$\begin{aligned}\dot{x} &= A(t, \varepsilon)x + B_1(t, \varepsilon)y_1 + B_2(t, \varepsilon)y_2 + B_3(t, \varepsilon)y_3 + C(t, \varepsilon),\\ \varepsilon\dot{y}_1 &= E_1(t, \varepsilon)x + \varepsilon F_1(t, \varepsilon)y_1 + \varepsilon F_2(t, \varepsilon)y_2 + \varepsilon F_3(t, \varepsilon)y_3 + G_1(t, \varepsilon),\\ \varepsilon\dot{y}_2 &= E_2(t, \varepsilon)x + \varepsilon F_4(t, \varepsilon)y_1 + F_5(t, \varepsilon)y_2 + \varepsilon F_6(t, \varepsilon)y_3 + G_2(t, \varepsilon),\\ \varepsilon\dot{y}_3 &= E_3(t, \varepsilon)x + \varepsilon F_7(t, \varepsilon)y_1 + \varepsilon F_8(t, \varepsilon)y_2 + F_9(t, \varepsilon)y_3 + G_3(t, \varepsilon),\end{aligned} \tag{6}$$

where, in blocks compatible with (4),

$$P^{-1}(FP - \varepsilon\dot{P}) = \begin{bmatrix} \varepsilon F_1 & \varepsilon F_2 & \varepsilon F_3 \\ \varepsilon F_4 & F_5 & \varepsilon F_6 \\ \varepsilon F_7 & \varepsilon F_8 & F_9 \end{bmatrix},$$

$$BP = [B_1 \quad B_2 \quad B_3], \quad P^{-1}E = [E_1' \quad E_2' \quad E_3']', \quad \text{and} \quad P^{-1}G = [G_1' \quad G_2' \quad G_3']'.$$

(Appropriate smoothness conditions on P and (1) will imply such for the coefficients in (6).) Experience with singular perturbation problems leads us to expect that the limiting solution to (6) within (0, 1) will satisfy the reduced system obtained by setting $\varepsilon = 0$ in (6), i.e.,

$$\begin{aligned}\dot{X}_0 &= A_0X_0 + B_{10}Y_{10} + B_{20}Y_{20} + B_{30}Y_{30} + C_0,\\ 0 &= E_{10}X_0 + G_{10},\\ 0 &= E_{20}X_0 + F_{50}Y_{20} + G_{20},\\ 0 &= E_{30}X_0 + F_{90}Y_{30} + G_{30},\end{aligned} \tag{7}$$

where, e.g., $A_0 = A(t, 0)$. For any solution of (7),

$$\begin{aligned}Y_{20} &= -F_{50}^{-1}(E_{20}X_0 + G_{20}),\\ Y_{30} &= -F_{90}^{-1}(E_{30}X_0 + G_{30})\end{aligned} \tag{8}$$

and there remains the system of m_1 linear equations

$$E_{10}X_0 = -G_{10} \tag{9}$$

and the nth order system of differential equations

$$\dot{X}_0 = H_0X_0 + B_{10}Y_{10} + J_0 \tag{10}$$

to determine the m_1-vector Y_{10} and the n-vector X_0. Here $H_0 = A_0 - B_{20}F_{50}^{-1}E_{20} - B_{30}F_{90}^{-1}E_{30}$ and $J_0 = C_0 - B_{20}F_{50}^{-1}G_{20} - B_{30}F_{90}^{-1}G_{30}$. For the regular problem $m_1 = 0$, neither the constraint (9) nor the Y_1 components occur, i.e., the solution of the reduced problem is obtained through the nth order system of differential equations (10).

In order to solve (7) when $n \geqq m_1 > 0$, we must be assured that (9) is consistent and that we can obtain Y_{10}. Thus, we will assume

H2 *G_{10} is in the range of E_{10} for* $0 \leqq t \leqq 1$

and

H3 $-E_{10}B_{10}$ *is stable throughout,* $0 \leqq t \leqq 1$.

(The nonsingularity of $E_{10}B_{10}$ would suffice for now, but the boundary layer structure appropriate for (6) requires H3. If $n < m_1$, (9) might overdetermine X_0 and (10) would underdetermine Y_0.) Differentiating (9), we obtain

$$E_{10}\dot{X}_0 = -\dot{E}_{10}X_0 - \dot{G}_{10},$$

so (10) and H3 imply that

$$Y_{10} = (-E_{10}B_{10})^{-1}[(E_{10}H_0 + \dot{E}_{10})X_0 + (E_{10}J_0 + \dot{G}_{10})], \tag{11}$$

and there remains a nonhomogeneous system for X_0, namely (9), (10), and (11). To determine X_0, it is convenient to introduce the projection

$$\mathscr{E} = I_n - \mathscr{Q}E_{10} \tag{12}$$

where

$$\mathscr{Q} = B_{10}(E_{10}B_{10})^{-1}. \tag{13}$$

We note that $E_{10}\mathscr{E} = 0$ while $E_{10}\mathscr{Q} = I_{m_1}$ and $\mathscr{E}B_{10} = \mathscr{E}\mathscr{Q} = 0$. Moreover, E_{10} and B_{10} have rank m_1 since $E_{10}B_{10}$ has (full) rank m_1. Thus, $\mathscr{E}$ has rank $n - m_1 \geqq 0$. Indeed, $\mathscr{Q}$ is nearly a generalized inverse of E_{10} (cf. Campbell, Meyer and Rose [2] and Campbell and Rose [3] who use such inverses more explicitly in a singular perturbations context). The use of (9), (12) implies that

$$X_0 = \mathscr{E}X_0 - \mathscr{Q}G_{10} \tag{14}$$

and (10), (11), and (14) imply the linear system

$$(\mathscr{E}X_0)^{\cdot} = \mathscr{K}(\mathscr{E}X_0) + \mathscr{L} \tag{15}$$

for $\mathscr{E}X_0$ where $\mathscr{K} = \mathscr{E}H_0 - \mathscr{Q}\dot{E}_{10}$ and $\mathscr{L} = -\mathscr{K}\mathscr{Q}G_{10} + \mathscr{E}J_0 + \dot{\mathscr{Q}}G_{10}$. Under hypotheses H1–H3, then, the solution X_0 of the reduced system (7) will be completely and uniquely determined up to later specification of a boundary value for $\mathscr{E}X_0$. It is perhaps most natural to use the condition

$$\mathscr{E}(j)X_0(j) = \mathscr{E}(j)x(j), \quad j = 0 \text{ or } 1, \tag{16}$$

presuming a boundary value (16) is supplied by (2). Other possibilities should also be considered, however.

Note that our manipulations allowed us to determine $E_{10}X_0$ from the linear equation (9) and the remaining "component" $\mathscr{E}X_0$ of X_0 from an end value problem like (15), (16). The alternative problem character of the solution for X_0 (cf. Cesari [4]), with an algebraic equation in the range of E_{10} and a differential equation in its orthogonal complement, makes it quite different from the more straightforward solution of regular problems. There, no analogue of the constraint (7) occurs, and hypothesis H1 suffices.

If H2 fails, the reduced system (7) is inconsistent, but irrelevant (cf. O'Malley [18]). A simplified example is provided by $\varepsilon\dot{y} = 1$, $y(0) = 0$ where the limiting solution for $t > 0$ is unbounded like t/ε. For regular problems, the reduced problem (3) is necessarily consistent, but inconsistency of (3) would occur in the present problem if H2 failed. The stability assumption H3 is generally needed in order for a limiting solution to exist. A simple example is provided by $\dot{x}=-y$, $\varepsilon\dot{y}=x$, $x(1)=0$, $y(0)=1$ which has the solution $(x, y)=(-\sqrt{\varepsilon}\sin(t/\sqrt{\varepsilon}), \cos(t/\sqrt{\varepsilon}))$ for which there is no limit as $\varepsilon\to 0$. We note that by changing the sign so that $\varepsilon\dot{y}=-x$, $E_{10}B_{10}<0$, and we have a limiting solution. If $E_{10}B_{10}=0$, further differentiation of (9) might allow one to determine Y_{10}, just as singular arcs of higher order are obtained in control (cf., e.g., Robbins [22]). The structure of the asymptotic solutions (i.e., the asymptotic expansions of the exact solutions) will then differ considerably from when H3 holds. An example, arising in optimal control, is

$$\begin{aligned} \dot{x}_1&=-y_1+x_2, & x_1(1)&=0,\\ \dot{x}_2&=-x_1, & x_2(1)&=0,\\ \varepsilon\dot{y}_1&=\varepsilon y_2, & y_1(0)&=1,\\ \varepsilon\dot{y}_2&=-x_2-\varepsilon y_1, & y_2(0)&=0,\end{aligned}$$

(cf. O'Malley and Jameson [20]). Here, the asymptotic solution is given by $x_1=-\dot{x}_2$ and $y_2=\dot{y}_1$ where $x_2=2\sqrt{\varepsilon}\,\mathrm{Im}\,[c\,e^{-\omega t/\sqrt[4]{\varepsilon}}]$ and $y_1=2\,\mathrm{Re}\,[c\,e^{-\omega t/\sqrt[4]{\varepsilon}}]$ for

$$c=(1-i\sqrt{(1+i\sqrt{\varepsilon})/(1-i\sqrt{\varepsilon})})^{-1} \quad\text{and}\quad \omega=e^{i\pi/4}\sqrt{1+i\sqrt{\varepsilon}}.$$

Thus, the boundary layer thickness is $O(\sqrt[4]{\varepsilon})$. Finally, if $E_{10}B_{10}$ is singular, but nonzero, progress generally might be made through preliminary algebraic manipulations (cf. Anderson [1] for an analogous control problem).

3. Construction of asymptotic solutions. A linear nonhomogeneous boundary value problem can be solved by variation of parameters once a complete set of linearly independent solutions of the corresponding homogeneous system is known. For the asymptotic solution of (1), we would need $n+m$ linearly independent asymptotic solutions. Alternatively, one could seek an outer solution of (1) (i.e., a regular perturbation of an already obtained solution of the reduced system (7)) and modify it by adding appropriate boundary layer corrections which satisfy the homogeneous version of (6) (cf. O'Malley [17] which solves corresponding scalar problems). Since we can supply $n-m_1$ boundary conditions (like (16)) for the reduced system, we can expect an outer solution under hypotheses H1–H3 to be of the form

$$(X(t,\sqrt{\varepsilon}), Y_1(t,\sqrt{\varepsilon}), Y_2(t,\sqrt{\varepsilon}), Y_3(t,\sqrt{\varepsilon}))\sim\sum_{j=0}^{\infty}(X_j(t), Y_{1j}(t), Y_{2j}(t), Y_{3j}(t))\varepsilon^{j/2} \tag{17}$$

being an asymptotic solution within $(0, 1)$ which converges to a solution $(X_0, Y_{10}, Y_{20}, Y_{30})$ of the reduced system (7) as $\varepsilon\to 0$. It would be formally determined termwise by the boundary values

$$\mathscr{E}(j)X(j,\sqrt{\varepsilon}), \qquad j=0 \text{ or } 1, \tag{18}$$

since higher order terms will satisfy a system of the form (7) with successively known nonhomogeneous terms. (Details of the process are not given, because it is completely analogous to the preceding calculation of the zeroth order terms. The expansion is in powers of $\varepsilon^{1/2}$, anticipating the boundary layer corrections obtained below.)

Since $\mathscr{E}$ has rank $n-m_1$, the outer solution is (through (18)) parameterized by $n-m_1$ vector functions of $\sqrt{\varepsilon}$, and there is need for $m+n-(n-m_1)=m+m_1$ linearly independent boundary layer solutions which are asymptotically negligible within $(0, 1)$.

For the regular problem with $m_2+m_3=m$, there would be m_3 boundary layer solutions which are decaying functions of the stereteched variable $\kappa=t/\varepsilon$ and m_2 which are decaying functions of $\rho=(1-t)/\varepsilon$ (cf., say, Harris [9]). For our singular problems, however, we shall also find m_1 boundary layer solutions depending on each of the stretched variables $\tau=t/\sqrt{\varepsilon}$ and $\sigma=(1-t)/\sqrt{\varepsilon}$. These thicker ($\sqrt{\varepsilon}\gg\varepsilon$) boundary layers (upon matching) now require all our asymptotic expansions to be power series in

$$\mu=\sqrt{\varepsilon}, \tag{19}$$

rather than ε. In order to generate these asymptotic solutions, we'll assume the coefficients in (6) to be infinitely differentiable, though finite approximations could be obtained under less smoothness.

We shall construct formal asymptotic solutions to (6) of the form

$$\begin{aligned}
x(t,\varepsilon)&=X(t,\mu)+\mu^{\alpha}z(\tau,\mu)+\mu^{\beta+1}w(\sigma,\mu)+\mu^{\gamma+2}r(\rho,\mu)+\mu^{\delta+2}h(\kappa,\mu),\\
y_1(t,\varepsilon)&=Y_1(t,\mu)+\mu^{\alpha-1}p_1(\tau,\mu)+\mu^{\beta}q_1(\sigma,\mu)+\mu^{\gamma+2}s_1(\rho,\mu)+\mu^{\delta+2}l_1(\kappa,\mu),\\
y_2(t,\varepsilon)&=Y_2(t,\mu)+\mu^{\alpha}p_2(\tau,\mu)+\mu^{\beta+1}q_2(\sigma,\mu)+\mu^{\gamma}s_2(\rho,\mu)+\mu^{\delta+2}l_2(\kappa,\mu),\\
y_3(t,\varepsilon)&=Y_3(t,\mu)+\mu^{\alpha}p_3(\tau,\mu)+\mu^{\beta+1}q_3(\sigma,\mu)+\mu^{\gamma+2}s_3(\rho,\mu)+\mu^{\delta}l_3(\kappa,\mu),
\end{aligned} \tag{20}$$

where the functions of the stretched variables

$$\tau=t/\mu,\quad \sigma=(1-t)/\mu,\quad \rho=(1-t)/\mu^2,\quad \text{and}\quad \kappa=t/\mu^2 \tag{21}$$

tend to zero as the appropriate variable tends to infinity. The outer solution (17), then, provides the asymptotic solution within (0, 1). The scalings μ^{α}, μ^{β}, μ^{γ}, and μ^{δ} for the boundary layer corrections remain free to meet various boundary conditions (2), while the remaining μ^{-1}, μ, and μ^2 factors are simply used to prevent calculation of trivial coefficients. Since the full solution (20) and the outer solution (17) both satisfy the nonhomogeneous system (6), the boundary layer corrections (with their different decay rates) must separately satisfy the corresponding homogeneous system.

Let us first seek m_1 boundary layer solutions of the homogeneous system (6) of the form

$$\left(z,\frac{1}{\mu}p_1,p_2,p_3\right)\sim\sum_{k=0}^{\infty}\left(z_k(\tau),\frac{1}{\mu}p_{1k}(\tau),p_{2k}(\tau),p_{3k}(\tau)\right)\mu^k \tag{22}$$

(cf. (20)). Thus, we'll have

$$\begin{aligned}
\frac{1}{\mu}\frac{dz}{d\tau}&=Az+\frac{1}{\mu}B_1p_1+B_2p_2+B_3p_3,\\
\frac{dp_1}{d\tau}&=E_1z+\mu F_1p_1+\mu^2F_2p_2+\mu^2F_3p_3,\\
\mu\frac{dp_2}{d\tau}&=E_2z+\mu F_4p_1+F_5p_2+\mu^2F_6p_3,\\
\mu\frac{dp_3}{d\tau}&=E_3z+\mu F_7p_1+\mu^2F_8p_2+F_9p_3.
\end{aligned} \tag{23}$$

When $\mu = 0$, this reduces to the limiting problem

$$
\begin{aligned}
&\frac{dz_0}{d\tau} = B_{10}(0)p_{10},\\
&\frac{dp_{10}}{d\tau} = E_{10}(0)z_0,\\
&0 = E_{20}(0)z_0 + F_{50}(0)p_{20},\\
&0 = E_{30}(0)z_0 + F_{90}(0)p_{30},
\end{aligned} \tag{24}
$$

so

$$
\begin{aligned}
p_{20}(\tau) &= -F_{50}^{-1}(0)E_{20}(0)z_0(\tau),\\
p_{30}(\tau) &= -F_{90}^{-1}(0)E_{30}(0)z_0(\tau),
\end{aligned} \tag{25}
$$

while

$$\frac{d^2 z_0}{d\tau^2} = B_{10}(0)E_{10}(0)z_0. \tag{26}$$

Since $\mathscr{E}(0)B_{10}(0) = 0$, $(d^2/d\tau^2)(\mathscr{E}(0)z_0) = 0$ and the only solution which decays to zero together with its first derivative as $\tau \to \infty$ is

$$\mathscr{E}(0)z_0(\tau) = 0. \tag{27}$$

Furthermore, multiplying (26) by $E_{10}(0)$ uniquely implies the decaying solution

$$E_{10}(0)z_0(\tau) = \exp\left(-\sqrt{E_{10}(0)B_{10}(0)}\,\tau\right)E_{10}(0)z_0(0). \tag{28}$$

(For the square root of an unstable matrix, we shall use the principal (unstable) square root of the diagonal entries in its Jordan form.) The definition (12) of $\mathscr{E}$ then implies that

$$z_0(\tau) = \mathscr{Q}(0)\exp\left(-\sqrt{E_{10}(0)B_{10}(0)}\,\tau\right)E_{10}(0)z_0(0). \tag{29}$$

Finally, the differential equation for p_{10} has the decaying solution

$$p_{10}(\tau) = -\left(\sqrt{E_{10}(0)B_{10}(0)}\right)^{-1}E_{10}(0)z_0(\tau). \tag{30}$$

Since $p_{10}(0)$ and $E_{10}(0)z_0(0)$ are arbitrary, we are able to provide m_1 linearly independent solutions to (24) by specifying either. Higher order coefficients in (22) satisfy nonhomogeneous forms of (24) with successively known, exponentially decaying terms. The decaying solutions (22) are thereby completely determined up to specification of either

$$p_1(0, \mu) \quad \text{or} \quad E_{10}(0)z(0, \mu). \tag{31}$$

This possibility of a choice of boundary values makes the boundary layer corrections with $O(\mu)$ boundary layers more flexible than those with $O(\varepsilon)$ layers (cf. (40)).

The classical boundary layer correction

$$(\mu^2 j, \mu^2 l_1, \mu^2 l_2, l_3) \sim \sum_{k=0}^{\infty} (\mu^2 l_k, \mu^2 l_{1k}, \mu^2 l_{2k}, l_{3k})\mu^k \tag{32}$$

must satisfy the homogeneous form of (6), so

$$
\begin{aligned}
\frac{dj}{d\kappa} &= \mu^2 Aj + \mu^2 B_1 l_1 + \mu^2 B_2 l_2 + B_3 l_3, \\
\frac{dl_1}{d\kappa} &= E_1 j + \mu^2 F_1 l_1 + \mu^2 F_2 l_2 + F_3 l_3, \\
\frac{dl_2}{d\kappa} &= E_2 j + \mu^2 F_4 l_1 + F_5 l_2 + F_6 l_3, \\
\frac{dl_3}{d\kappa} &= \mu^2 E_3 j + \mu^4 F_7 l_1 + \mu^4 F_8 l_2 + F_9 l_3.
\end{aligned}
\tag{33}
$$

The resulting limiting problem

$$
\begin{aligned}
&\frac{dj_0}{d\kappa} = B_{30}(0) l_{30}, \qquad \frac{dl_{10}}{d\kappa} = E_{10}(0) j_0 + F_{30}(0) l_{30}, \\
&\frac{dl_{20}}{d\kappa} = E_{20}(0) j_0 + F_{50}(0) l_{20} + F_{60}(0) l_{30}, \qquad \frac{dl_{30}}{d\kappa} = F_{90}(0) l_{30}
\end{aligned}
\tag{34}
$$

has the unique decaying solution

$$
\begin{aligned}
&l_{30}(\kappa) = e^{F_{90}(0)\kappa} l_{30}(0), \\
&j_0(\kappa) = B_{30}(0) F_{90}^{-1}(0) l_{30}(\kappa), \\
&l_{10}(\kappa) = (E_{10}(0) B_{30}(0) F_{90}^{-2}(0) + F_{30}(0) F_{90}^{-1}(0)) l_{30}(\kappa), \quad \text{and} \\
&l_{20}(\kappa) = -\int_\kappa^\infty e^{F_{50}(0)(\kappa - s)} (E_{20}(0) B_{30}(0) F_{90}^{-1}(0) + F_{60}(0)) l_{30}(s)\, ds
\end{aligned}
\tag{35}
$$

up to selection of $l_{30}(0)$. Higher order terms satisfy nonhomogeneous forms of (34) and are completely determined up to the m_3-dimensional initial vector

$$
l_3(0, \mu). \tag{36}
$$

This procedure, then, allows m_3 linearly independent boundary layer solutions (32) to be constructed formally, just as for the regular problem.

The terminal boundary layer solutions are found in a manner analogous to those at $t = 0$. Thus, the leading terms of the thick terminal boundary layer correction

$$
(\mu w, q_1, \mu q_2, \mu q_3) \sim \sum_{k=0}^\infty (\mu w_k, q_{1k}, \mu q_{2k}, \mu q_{3k}) \mu^k \tag{37}
$$

are

$$
\begin{aligned}
q_{20}(\sigma) &= -F_{50}^{-1}(1) E_{20}(1) w_0(\sigma), \\
q_{30}(\sigma) &= -F_{90}^{-1}(1) E_{30}(1) w_0(\sigma), \\
q_{10}(\sigma) &= \exp\left(-\sqrt{E_{10}(1) B_{10}(1)}\,\sigma\right) q_{10}(0), \\
w_0(\sigma) &= B_{10}(1) \left(\sqrt{E_{10}(1) B_{10}(1)}\right)^{-1} q_{10}(\sigma).
\end{aligned}
\tag{38}
$$

We note that (12) implies that

$$
\mathscr{E}(1) w_0(\sigma) = 0 \tag{39}
$$

so that $w_0(\sigma)=\mathcal{Q}(1)E_{10}(1)w_0(\sigma)$ and these leading terms are therefore determined up to the initial value $q_{10}(0)=(\sqrt{E_{10}(1)B_{10}(1)})^{-1}E_{10}(1)w_0(0)$. Analogous work successively provides the higher order terms in (37) up to specification of either

$$q_1(0,\mu) \quad \text{or} \quad E_{10}(1)w(0,\mu). \tag{40}$$

Finally, the leading terms of the (usual) terminal boundary layer correction

$$(\mu^2 r, \mu^2 s_1, s_2, \mu^2 s_3) \sim \sum_{k=0}^{\infty} (\mu^2 r_k, \mu^2 s_{1k}, s_{2k}, \mu^2 s_{3k})\mu^k \tag{41}$$

are given by

$$\begin{aligned} s_{20}(\rho) &= e^{-F_{50}(1)\rho} s_{20}(0), \\ r_0(\rho) &= B_{20}(1)F_{50}^{-1}(1)s_{20}(\rho), \\ s_{10}(\rho) &= (E_{10}(1)B_{20}(1)F_{50}^{-2}(1)+F_{20}(1)F_{50}^{-1}(1))s_{20}(\rho), \\ s_{30}(\rho) &= \int_\rho^\infty e^{F_{90}(1)(t-\rho)}(E_{30}(1)B_{20}(1)F_{50}^{-1}(1)+F_{80}(1))s_{20}(t)\,dt, \end{aligned} \tag{42}$$

and the solution is completely specified up to the initial m_2 vector

$$s_2(0,\mu). \tag{43}$$

We could prove the asymptotic validity of the formal series solutions (20) which we've constructed by integral equations methods (cf. Harris [9] or Vasil'eva and Butuzov [27]). Although that proof would differ somewhat from the classical (regular) ones, we regard the construction of the solutions as the most challenging aspect of this study and shall not discuss further the details of proof. Alternatively, we could base a proof on the set of linearly independent solutions to the homogeneous problem constructed (and shown to be asymptotically valid) by Turrittin [25]. The formal series solution (20) should thereby be interpreted as an asymptotic expansion of an exact solution to (6).

4. Fitting the boundary conditions. Since our construction of the outer solution (17) and of the thicker boundary layer corrections (22) and (37) distinguishes between components of $\mathscr{E}x$ and $E_{10}x$, it is natural to write

$$x = x_1 + \mathcal{Q}x_2 \tag{44}$$

for the n and m_1 dimensional vectors

$$x_1 = \mathscr{E}(t)x(t,\varepsilon) \quad \text{and} \quad x_2 = E_{10}(t)x(t,\varepsilon). \tag{45}$$

(We experienced a similar separation of components in solving the reduced problem (7) where we had a differential equation for $X_{10}=\mathscr{E}X_0$ and a linear algebraic equation for $X_{20}=E_{10}X_0$.) Instead of the x representation of (20), then, it will be more convenient to use the further decomposition

$$\begin{aligned} x_1(t,\varepsilon) &= X_1(t,\mu)+\mu^{\alpha+1}z_1(\tau,\mu)+\mu^{\beta+2}w_1(\sigma,\mu)+\mu^{\gamma+2}r_1(\rho,\mu)+\mu^{\delta+2}j_1(\kappa,\mu), \\ x_2(t,\varepsilon) &= X_2(t,\mu)+\mu^{\alpha}z_2(\tau,\mu)+\mu^{\beta+1}w_2(\sigma,\mu)+\mu^{\gamma+2}r_2(\rho,\mu)+\mu^{\delta+2}j_2(\kappa,\mu), \end{aligned} \tag{46}$$

where, e.g., $X_1(t,\mu)=\mathscr{E}(t)X(t,\mu)$ and $z_1(\tau,\mu)=(1/\mu)\mathscr{E}(\mu\tau)z(\tau,\mu)=O(1)$ by (27). We note that $E_{10}\mathscr{E}=0$ implies that we must have

$$E_{10}x_1 = 0. \tag{47}$$

Now note that the solution $(x' \quad y')'$ of the original system (1) is of the form

$$\begin{pmatrix} x \\ y \end{pmatrix} = T(t)u(t) \tag{48}$$

where

$$T(t) = \begin{pmatrix} I_n & Q & 0 & 0 & 0 \\ 0 & 0 & P_1 & P_2 & P_3 \end{pmatrix} \tag{49}$$

for

$$P = (P_1 \quad P_2 \quad P_3),$$

where each P_i is $m \times m_i$ dimensional, and u is the $n + m_1 + m$ dimensional vector

$$u(t) = (x_1' \quad x_2' \quad y_1' \quad y_2' \quad y_3')'. \tag{50}$$

Furthermore, the expansions (46) imply that all solutions (20) of the transformed problem (6) are given by

$$u(t) = U(t, \mu)k(\mu) \tag{51}$$

where the square matrix U is

$$U(t, \mu) = \begin{pmatrix} X_1(t, \mu) & \mu z_1(\tau, \mu) & \mu^2 w_1(\sigma, \mu) & \mu^2 r_1(\rho, \mu) & \mu^2 j_1(\kappa, \mu) \\ X_2(t, \mu) & z_2(\tau, \mu) & \mu w_2(\sigma, \mu) & \mu^2 r_2(\rho, \mu) & \mu^2 j_2(\kappa, \mu) \\ Y_1(t, \mu) & \dfrac{1}{\mu} p_1(\tau, \mu) & q_1(\sigma, \mu) & \mu^2 s_1(\rho, \mu) & \mu^2 l_1(\kappa, \mu) \\ Y_2(t, \mu) & p_2(\tau, \mu) & \mu q_2(\sigma, \mu) & s_2(\rho, \mu) & \mu^2 l_2(\kappa, \mu) \\ Y_3(t, \mu) & p_3(\tau, \mu) & \mu q_3(\sigma, \mu) & s_3(\rho, \mu) & l_3(\kappa, \mu) \end{pmatrix} \tag{52}$$

with fixed boundary values

$$\begin{aligned} &X_1(j, \mu) = I_n, \qquad j = 0 \text{ or } 1, \qquad z_2(0, \mu) = I_{m_1} = q_1(0, \mu), \\ &s_2(0, \mu) = I_{m_2}, \quad \text{and} \quad l_3(0, \mu) = I_{m_3}, \end{aligned} \tag{53}$$

and the $n + m + m_1$ vector $k(\mu)$ is partitioned as

$$k(\mu) = (k_1'(\mu) \quad \mu^\alpha k_2'(\mu) \quad \mu^\beta k_3'(\mu) \quad \mu^\gamma k_4'(\mu) \quad \mu^\delta k_5'(\mu))'. \tag{54}$$

We note that the boundary values (53) imply that $p_{10}(0) = -(\sqrt{E_{10}(0)B_{10}(0)})^{-1}$, and $w_{20}(0) = (\sqrt{E_{10}(1)B_{10}(1)})^{-1}$, so both $p_1(0, \mu)$ and $w_2(0, \mu)$ are nonsingular for ε sufficiently small. Also note that $U(t, \mu)$ would be an asymptotic expansion of a fundamental matrix if the corresponding $n + m + m_1$ dimensional system were homogeneous. We shall select a, β, γ, and δ so that the $k_i(0)$'s are $O(1)$ as $\varepsilon \to 0$ and nonzero if the particular k_i is not identically zero.

Putting (48) and (51) together, the boundary conditions (2) imply the $n + m$ linear equations

$$\Delta_1(\mu)k(\mu) = c(\varepsilon) \tag{55}$$

for the unknowns $k_i(\mu)$ where

$$\Delta_1(\mu) = \sum_{j=0}^{1} (R_j(\mu^2) \quad S_j(\mu^2))T(j)U(j, \mu). \tag{56}$$

An additional m_1 boundary conditions result if we impose the end condition

(57) $$\Delta_2(\mu)k(\mu) \equiv E_{10}(j)x_1(j, \mu^2) = 0, \qquad j = 0 \text{ or } 1,$$

required by consistency with (47). (Note that E_{10} has rank m_1.) Thus, the boundary value problem (1), (2) will have a unique asymptotic solution of the form

(58) $$\begin{pmatrix} x(t, \varepsilon) \\ y(t, \varepsilon) \end{pmatrix} = T(t)U(t, \mu)\Delta^{-1}(\mu)\begin{pmatrix} c(\mu^2) \\ 0 \end{pmatrix}$$

provided

H4 *the* $(n+m+m_1)\times(n+m+m_1)$ *matrix* $\Delta(\mu) = (\Delta_1'(\mu) \quad \Delta_2'(\mu))'$ *is nonsingular for* ε *sufficiently small.*

Because the boundary layer correction terms are asymptotically negligible away from one endpoint, we can considerably simplify the asymptotic calculation of $\Delta(\mu)$. Thus, up to asymptotically exponentially small terms,

$$T(0)U(0, \mu) = \begin{pmatrix} X_1(0, \mu) + \mathcal{Q}(0)X_2(0, \mu) & \mu z_1(0, \mu) + \mathcal{Q}(0) & 0 & 0 & \mu^2 j_1(0, \mu) + \mathcal{Q}(0)j_2(0, \mu) \\ \sum_{l=1}^{3} P_l(0)Y_l(0, \mu) & \frac{1}{\mu}P_1(0)p_1(0, \mu) + \sum_{l=2}^{3} P_l(0)p_l(0, \mu) & 0 & 0 & \sum_{k=1}^{2} \mu^2 P_k(0)l_k(0, \mu) + P_3(0) \end{pmatrix}$$

and likewise for $T(1)U(1, \mu)$. These imply that

(59) $$\Delta(\mu) = (\Delta_{kl}), \quad k = 1, 2; \quad l = 1, \cdots, 5$$

where

$$\Delta_{11} \sim \sum_{j=0}^{1} \left\{ R_j(X_1(j, \mu) + \mathcal{Q}(j)X_2(j, \mu)) + S_j \sum_{l=1}^{3} P_l(j)Y_l(j, \mu) \right\},$$

$$\Delta_{12} \sim R_0(\mu z_1(0, \mu) + \mathcal{Q}(0)) + \frac{1}{\mu} S_0\left(P_1(0)p_1(0, \mu) + \mu \sum_{l=2}^{3} P_l(0)p_l(0, \mu) \right),$$

$$\Delta_{13} \sim R_1(\mu^2 w_1(0, \mu) + \mu\mathcal{Q}(1)w_2(0, \mu)) + S_1\left(P_1(1)q_1(0, \mu) + \mu \sum_{l=2}^{3} P_l(1)q_l(0, \mu) \right),$$

$$\Delta_{14} \sim \mu^2 R_1(r_1(0, \mu) + \mathcal{Q}(1)r_2(0, \mu)) + S_1\left(\mu^2 P_1(1)s_1(0, \mu) + \sum_{l=2}^{3} P_l(1)s_l(0, \mu) \right),$$

$$\Delta_{15} \sim R_0(\mu^2 j_1(0, \mu) + \mathcal{Q}(0)j_2(0, \mu)) + S_0\left(\mu^2 \sum_{k=1}^{2} P_k(0)l_k(0, \mu) + P_3(0) \right),$$

$\Delta_{21} \sim E_{10}(j)X_1(j, \mu)$ with j determined in (57),

and

$$(\Delta_{22}, \Delta_{23}, \Delta_{24}, \Delta_{25}) \sim \begin{cases} \mu E_{10}(0)(z_1(0, \mu), 0, 0, \mu j_1(0, \mu)) & \text{if } j = 0, \\ \mu^2 E_{10}(1)(0, w_1(0, \mu), r_1(0, \mu), 0) & \text{if } j = 1. \end{cases}$$

Because $\Delta_{12} = O(1/\mu)(= O(1)$ only if $S_1P_1(0)p_{10}(0) = 0)$, the matrix $\Delta(\mu)$ will have an asymptotic series expansion

(60) $$\Delta(\mu) \sim \frac{1}{\mu} \sum_{j=0}^{\infty} \delta_j \mu^j.$$

It will therefore be nonsingular if the limiting matrix δ_0 is nonsingular, although $\Delta(\mu)$ can still be nonsingular for $\varepsilon \neq 0$ if $\delta_0 = 0$. If δ_l is the first nonsingular coefficient in (60),

$\Delta^{-1}(\mu)$ will be $O(\mu^{1-l})$, so the solution (58) of the given problem will generally be unbounded like $O(\mu^{-l})$. In particular, note that a bounded solution will result if $l=0$ and that the powers α, β, γ, and δ in (54) are integers. Further, the limiting solution within (0, 1) depends only on $k_1(\mu)$, so we might say which boundary conditions are appropriate for the reduced problem (7) (cf. Harris [9]). We might also consider the possibility of nonunique solutions under appropriate orthogonality assumptions if $\Delta(\mu)$ is singular.

To summarize our principal results, we have the following theorem.

THEOREM. *Under hypotheses* H1–H4, *we obtain a unique solution* (58) *of the boundary value problem* (1), (2).

5. Natural boundary value problems.

Sample problem 1. Suppose we are given a problem in the transformed form (6) with prescribed boundary values

$$x(0), \quad y_1(1), \quad y_2(1), \quad \text{and} \quad y_3(1). \tag{61}$$

Instead of actually obtaining $\Delta^{-1}(\mu)$, we can apply the boundary conditions in (51) to obtain a solution with

$$\begin{aligned} &k_1(0)=\mathscr{E}(0)x(0) \quad \text{for } X_1(0,\mu)=I_n,\\ &\alpha=0, \qquad k_2(0)=E_{10}(0)x(0)-X_{20}(0)k_1(0),\\ &\beta=0, \qquad k_3(0)=y_1(1)-Y_{10}(1)k_1(0),\\ &\gamma=0, \qquad k_4(0)=y_2(1)-Y_{20}(1)k_1(0),\\ &\delta=0, \qquad k_5(0)=y_3(0)-Y_{30}(0)k_1(0)-p_{30}(0)k_2(0). \end{aligned} \tag{62}$$

In particular, the limiting solution within (0, 1) will satisfy the reduced problem (7) and the initial condition (16) with $j=0$.

Sample problem 2. Suppose our problem is of the transformed form (6) with prescribed boundary values

$$x(1), \quad y_1(0), \quad y_2(1), \quad \text{and} \quad y_3(0). \tag{63}$$

Again, the unique solution is readily found to be of the form (51) with

$$\begin{aligned} &k_1(0)=\mathscr{E}(1)x(1) \quad \text{for } X_1(1,\mu)=I_n,\\ &\alpha=1, \qquad k_2(0)=-\sqrt{E_{10}(0)B_{10}(0)}(y_1(0)-Y_{10}(0)k_1(0)),\\ &\beta=-1, \qquad k_3(0)=\sqrt{E_{10}(1)B_{10}(1)}(E_{10}(1)x(1)-X_{20}(1)k_1(0)),\\ &\gamma=0, \qquad k_4(0)=y_2(1)-Y_{20}(1)k_1(0)-g_{20}(0)k_3(0),\\ &\delta=0, \qquad k_5(0)=y_3(0)-Y_{30}(0)k_1(0), \end{aligned} \tag{64}$$

and the limiting solution within (0, 1) satisfies the system (7) and the terminal condition (16) with $j=1$.

The general problem. As our special problems suggest, we can seek a solution (51) of the transformed system (6) plus boundary conditions. That problem will be uniquely solvable provided the reduced system (7) can be uniquely solved subject to appropriate boundary conditions. We must suppose that $E_{10}(0)x(0,\varepsilon)$ or $y_1(0,\varepsilon)$, $E_{10}(1)x(1,\varepsilon)$ or $y_1(1,\varepsilon)$, $y_2(1,\varepsilon)$, and $y_3(1,\varepsilon)$ can be obtained in order to determine uniquely the initial values (31), (36), (40), and (43) for the various boundary layer corrections of the solution (20), (46). In order to solve the reduced problem (7), separated boundary values for $\mathscr{E}(j)X(j,\varepsilon)$ need not be given (as in the sample problems). However, if the prescribed boundary values for $\mathscr{E}x$ are coupled at $t=0$ and $t=1$, existence of the solution to the resulting reduced two point problem is not a priori guaranteed. One must always be able to solve the given conditions for $y_2(1,\varepsilon)$ and $y_3(0,\varepsilon)$. Knowing $y_3(0,\varepsilon)$, for example, one would solve

$$y_3(0,\varepsilon)\sim Y_3(0,\mu)k_1(\mu)+\mu^{\alpha}p_3(0,\mu)k_2(\mu)+\mu^{\delta}k_5(\mu)$$

to obtain k_5.

Further, it is essential that at least m_1+m_3 boundary values be obtainable at $t=0$ and at least m_1+m_2 boundary values be obtainable at $t=1$ because these are the number of linearly independent boundary layer corrections decaying at those endpoints. In particular, we cannot expect to solve asymptotically initial value problems or terminal value problems for (1) unless we artificially restrict boundary values to appropriate lower dimensional manifolds (cf. Hoppensteadt [10]).

Since the solution of general problems (1), (2) relates crucially to the solution of simpler transformed problems (like our sample problems), it is convenient to solve general problems in terms of simpler "natural" ones. This generalized "shooting" method has been somewhat developed by Keller and White [13] and Ferguson [5].

REFERENCES

[1] B. D. O. ANDERSON, *Partially singular linear-quadratic control problems*, IEEE Trans. Automatic Control, AC-18 (1973), pp. 407–409.

[2] S. L. CAMPBELL, C. D. MEYER, JR. AND N. J. ROSE, *Applications of the Drazin inverse to linear systems of differential equations with singular constant coefficients*, SIAM J. Appl. Math., 31 (1976), pp. 411–425.

[3] S. L. CAMPBELL AND N. J. ROSE, *Singular perturbation of autonomous linear systems*, this Journal, to appear.

[4] L. CESARI, *Alternative methods in nonlinear analysis*, International Conference on Differential Equations, H. A. Antosiewicz, ed., Academic Press, New York, 1975, pp. 95–148.

[5] W. E. FERGUSON, JR. *A singularly perturbed linear two-point boundary value problem*, doctoral dissertation, California Inst. of Tech., Pasadena, 1975.

[6] J. E. FLAHERTY AND R. E. O'MALLEY, JR., *The numerical solution of boundary value problems for stiff differential equations*, Math. Comp., 31 (1977), pp. 66–93.

[7] ———, *Analytical and numerical methods for nonlinear singular singularly perturbed initial value problems*, to appear.

[8] B. S. GOH, *Necessary conditions for singular extremals involving multiple control variables*, SIAM J. Control, 4 (1966), pp. 716–731.

[9] W. A. HARRIS, JR., *Singularly perturbed boundary value problems revisited*, Lecture Notes in Math. 312, Springer-Verlag, Berlin, 1973, pp. 54–64.

[10] F. HOPPENSTEADT, *Properties of solutions of ordinary differential equations with a small parameter*, Comm. Pure Appl. Math., 24 (1971), pp. 807–840.

[11] F. C. HOPPENSTEADT AND W. L. MIRANKER, *Differential equations having rapidly changing solutions: Analytic methods for weakly nonlinear systems*, J. Differential Equations, 22 (1976), pp. 237–249.

[12] D. H. JACOBSON, *Totally singular quadratic minimization problems*, IEEE Trans. Automatic Control, AC-16 (1971), pp. 651–658.

[13] H. B. KELLER AND A. B. WHITE, JR., *Difference methods for boundary value problems in ordinary differential equations*, SIAM J. Numer. Anal., 12 (1975), pp. 791–802.

[14] N. LEVINSON, *Perturbations of discontinuous solutions of nonlinear systems of differential equations*, Acta Math., 82 (1951), pp. 71–106.

[15] F. W. J. OLVER, *Sufficient conditions for Ackerberg–O'Malley resonance*, this Journal, 8 (1978), pp. 328–355.

[16] R. E. O'MALLEY, JR., *Boundary value problems for linear systems of ordinary differential equations involving many small parameters*, J. Math. Mech., 18 (1969), pp. 835–856.

[17] ———, *On the asymptotic solution of boundary value problems for nonhomogeneous ordinary differential equations containing a parameter*, J. Math. Anal. Appl., 28 (1969), pp. 450–460.

[18] ———, *On singular singularly-perturbed initial value problems*, Applicable Anal., to appear.

[19] R. E. O'MALLEY, JR. AND J. E. FLAHERTY, *Singular singular-perturbation problems*, Lecture Notes in Math. 594, Springer-Verlag, Berlin, 1977, pp. 422–436.

[20] R. E. O'MALLEY, JR. AND A. JAMESON, *Singular perturbations and singular arcs I, II*, IEEE Trans. Automatic Control, AC-20 (1975), pp. 218–226; AC-22 (1977), pp. 328–337.

[21] R. E. O'MALLEY, JR. AND J. B. KELLER, *Loss of boundary conditions in the asymptotic solution of linear differential equations: II. Boundary value problems*, Comm. Pure Appl. Math., 21 (1968), pp. 263–270.

[22] H. M. ROBBINS, *A generalized Legendre–Clebsch condition for the singular cases of optimal control*, IBM J. Res. Develop., 3 (1967), pp. 361–372.

[23] Y. SIBUYA, *Sur réduction analytique d'un système d'equations différentielles ordinaires linéaires contenant un paramètre*, J. Fac. Sci. Univ. Tokyo Sect. I, 7 (1958), pp. 527–540.

[24] ———, *A block-diagonalization theorem for systems of linear ordinary differential equations and its applications*, SIAM J. Appl. Math., 14 (1966), pp. 468–475.

[25] H. L. TURRITTIN, *Asymptotic expansions of solutions of systems of ordinary linear differential equations containing a parameter*, Contrib. Theory Nonlinear Oscillations, 2 (1952), pp. 81–116.

[26] A. B. VASIL'EVA, *The development of the theory of ordinary differential equations with a small parameter multiplying a derivative in the period* 1966–1976, Russian Math. Surveys, 31 (1976), pp. 109–131.

[27] A. B. VASIL'EVA AND V. F. BUTUZOV, *Asymptotic Expansions of Solutions of Singularly Perturbed Differential Equations*, Nauka, Moscow, 1973.

[28] W. WASOW, *Asymptotic decomposition of systems of linear differential equations having poles with respect to a parameter*, Rend. Circ. Mat. Palermo, 13 (1964), pp. 329–344.

[29] ———, *Asymptotic Expansions for Ordinary Differential Equations*, John Wiley, New York, 1965.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0004$01.00/0

ON CONTINUOUS TRIANGULARIZATION OF MATRIX FUNCTIONS*

H. GINGOLD†

Abstract. This article is concerned with the continuous triangularization of matrix functions which depend continuously on several variables. By use of an algorithm analogous to the one employed for the reduction of a λ-matrix to a diagonal form, we find a continuous similarity transformation which produces the triangularization of a given matrix.

Let there be given a singular system of differential equations whose coefficient matrix depends solely on several small parameters. Then, our method may be applied to obtain the complete asymptotic expansion of a fundamental matrix solution of the singular differential system at its singular point.

1. Introduction. Let $A(\varepsilon)$ be an $n\times n$ continuous matrix function on $\bar{D}_\varepsilon$ where $\varepsilon=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_m)$ and $D_\varepsilon=\{\varepsilon\,|\,0<\varepsilon_i\leqq\varepsilon_i^0,\ i=1,2,\cdots,m\}$, ($\bar{D}_\varepsilon$ is the closure of D_ε). Consider the problem of taking $A(\varepsilon)$ into a Jordan canonical form $J(\varepsilon)$ via a nonsingular transformation $T(\varepsilon)$ such that

$$T(\varepsilon)A(\varepsilon)T^{-1}(\varepsilon)=J(\varepsilon). \tag{1.1}$$

Apparently the problem is solved by linear algebra. (See for example, [6, p. 200].) Practically (1.1) falls short of many purposes.

(i) The matrix $J(\varepsilon)$ is not necessarily a continuous function of ε on $\bar{D}_\varepsilon$.

(ii) The transforming matrices $T(\varepsilon)$ and $T^{-1}(\varepsilon)$ need not be continuous matrix functions of ε on $\bar{D}_\varepsilon$.

This statement can be easily verified by considering the example: (See [15, p. 138])

$$A(\varepsilon)=\begin{pmatrix}0&\varepsilon\\0&0\end{pmatrix},\quad \text{with } \varepsilon=\varepsilon_1.$$

For a theorem that will guarantee that a holomorphic matrix function will be taken into another given holomorphic matrix function by a nonsingular holomorphic matrix function in a neighborhood of $\varepsilon=\varepsilon_1=0$ the reader is referred to Wasow [14].

The remark above leads us to look for canonical forms for a matrix function $A(\varepsilon)$ other than the Jordan canonical form. In case that part of the eigenvalues of $A(\varepsilon)$ are separated from their complementary set of eigenvalues it was shown by Sibuya [13] and Hsieh and Sibuya [9], that $A(\varepsilon)$ may be transformed via a continuous nonsingular transformation into a blockdiagonal form. Sibuya and Hsieh also showed by a global analysis how smoothness properties of $A(\varepsilon)$ are inherited by the nonsingular transforming matrix $T(\varepsilon)$. Similar results by using a different method may be found in Coppell [5] and Gingold [7]. A normal form to which a family of matrices which depend smoothly on parameters of endomorphisms of a complex linear space can be reduced is discussed by Arnold [1].

The subject of this note is to determine conditions under which a certain matrix function $A(\varepsilon)$ which is *continuous* in a domain $\bar{D}_\varepsilon$ will be transformed into a triangular matrix via a nonsingular continuous transformation $P(\varepsilon)$, such that $P(\varepsilon)A(\varepsilon)P^{-1}(\varepsilon)$ is triangular.

Describing those conditions and a proper algorithm, one understands why triangularization of a matrix function $A(\varepsilon)$, of *one variable*, is successful when $A(\varepsilon)$ is analytic in $\varepsilon=\varepsilon_1$ in $\bar{D}_\varepsilon$. By examples in § 2, it will turn out that our results are best possible. Our algorithm will be described in §§ 3 and 4. As a corollary we will obtain Braaksma's result [3] and a weaker form of Rellich [12].

* Received by the editors June 29, 1976, and in final revised form January 13, 1978.

† Department of Mathematics, University of Utah, Salt Lake City, Utah 84112. Mailing address: Harimonim 8, Rechasim, Israel.

It is well-known that triangularization of matrix functions plays an important role in the theory of differential systems. See Perron [10], [11] and Braaksma [3]. However, Perron's method is not practical for singular linear differential systems with *unbounded* leading coefficient matrices and Braaksma's result applies only to linear differential systems whose coefficient matrix is a *meromorphic* function in one variable. Our method, for example, may produce complete information about the asymptotic behavior of a fundamental solution of the differential system $\phi(t,\varepsilon)Y'=A(\varepsilon)Y$ depending on the parameter ε for $\varepsilon\to 0$, where $\phi(t,\varepsilon)$ is a continuous mapping on $[0,a]\times\bar{D}_\varepsilon$, and $\phi(0,0)=0$.

2. A counter example. Schur's theorem tells us that for every $n\times n$ matrix A there exists a unitary matrix T such that TAT^{-1} is upper triangular, (see [2, p. 202]). If $A(\varepsilon)$ is a continuous function on $\bar{D}_\varepsilon$ the corresponding unitary transforming matrix $T(\varepsilon)$ may not be continuous on $\bar{D}_\varepsilon$. Since the eigenvalues and the eigenvectors of a matrix $A(\varepsilon)$ play the fundamental role in constructing the unitary transforming matrix $T(\varepsilon)$, we will focus our attention on them. It is well-known (see [8, vol. I, p. 267]) that if $A(\varepsilon)$ is continuous in a domain D_ε then there exist n continuous functions $\lambda_i(\varepsilon)$ such that $\lambda_i(\varepsilon)$, $i=1,2,\cdots,n$ are the eigenvalues of $A(\varepsilon)$.

PROPOSITION 2.1. *For every $N=0,1,2,\cdots$ and $N=\infty$ there exists a 2×2 matrix function $A(\varepsilon)$, $\varepsilon=\varepsilon_1$, whose entries belong to $C^N[0,\varepsilon_1^0]$ such that no continuous vector function $X(\varepsilon)$ on $[0,\varepsilon_1^0]$ can be an eigenvector of $A(\varepsilon)$ corresponding to any of its eigenvalues $\lambda_i(\varepsilon)$, $i=1,2$.*

Proof. Consider the matrix functions

$$A_N(\varepsilon)=g_N(\varepsilon)\begin{bmatrix}0 & \cos\varepsilon^{-1}\\ \sin\varepsilon^{-1} & 0\end{bmatrix}\tag{2.1}$$

where

$$g_N(\varepsilon)=\begin{cases}\varepsilon^{2N+1} & \text{if } N=1,2,\cdots,\\ \exp(-\varepsilon^{-2}) & \text{if } N=\infty.\end{cases}\tag{2.2}$$

It is easily verified that $A_N(\varepsilon)\in C^N[0,\varepsilon_1^0]$. The eigenvalues of $A_N(\varepsilon)$ are

$$\lambda_1(\varepsilon)=-\lambda_2(\varepsilon),\qquad \lambda_2(\varepsilon)=g_N(\varepsilon)\sqrt{(\cos\varepsilon^{-1})(\sin\varepsilon^{-1})}.$$

Assume that

$$X(\varepsilon)=\begin{pmatrix}x_1(\varepsilon)\\ x_2(\varepsilon)\end{pmatrix}$$

is a continuous eigenvector corresponding to $\lambda_2(\varepsilon)$. Then

$$x_2(\varepsilon)\cos\varepsilon^{-1}=x_1(\varepsilon)\sqrt{(\cos\varepsilon^{-1})(\sin\varepsilon^{-1})},\tag{2.3}$$

$$x_1(\varepsilon)\sin\varepsilon^{-1}=x_2(\varepsilon)\sqrt{(\cos\varepsilon^{-1})(\sin\varepsilon^{-1})}.\tag{2.4}$$

It is easily verified that $x_2(\varepsilon_\nu)=0$ for $\varepsilon_\nu=1/(\nu\pi)$, $\nu=1,2,\cdots$ and

$$x_1(\varepsilon_\nu)=0\quad\text{for}\quad \varepsilon_\nu=\frac{2}{(2\nu-1)\pi},\quad \nu=1,2,\cdots.$$

We let $\nu\to+\infty$ and we obtain $x_1(0)=x_2(0)=0$ which is a contradiction. A similar argument holds for $\lambda_1(\varepsilon)$.

PROPOSITION 2.2. *There exist 2×2 matrix functions of one variable $A_N(\varepsilon)\in C^N[0,\varepsilon_1^0]$, $N=1,2,\cdots$ and $N=\infty$, such that for no continuous invertible matrix function $T(\varepsilon)$ on $[0,\varepsilon_1^0]$, $T^{-1}(\varepsilon)A(\varepsilon)T(\varepsilon)$ has lower (or upper) triangular form.*

Proof. Assume that the proposition is false. Consider again $A_N(\varepsilon)$ given by (2.1) and (2.2). Let

$$T(\varepsilon)=\begin{bmatrix} t_{11}(\varepsilon) & t_{12}(\varepsilon) \\ t_{21}(\varepsilon) & t_{22}(\varepsilon) \end{bmatrix}$$

be a continuous nonsingular matrix function on $[0, \varepsilon_1^0]$. Then

$$g_N(\varepsilon)=\begin{bmatrix} 0 & \cos \varepsilon^{-1} \\ \sin \varepsilon^{-1} & 0 \end{bmatrix}\begin{bmatrix} t_{11}(\varepsilon) & t_{12}(\varepsilon) \\ t_{21}(\varepsilon) & t_{22}(\varepsilon) \end{bmatrix}=\begin{bmatrix} t_{11}(\varepsilon) & t_{12}(\varepsilon) \\ t_{21}(\varepsilon) & t_{22}(\varepsilon) \end{bmatrix}\begin{bmatrix} \lambda_1(\varepsilon) & 0 \\ d(\varepsilon) & \lambda_2(\varepsilon) \end{bmatrix} \tag{2.5}$$

where $\lambda_1(\varepsilon)$ and $\lambda_2(\varepsilon)$ are the eigenvalues of $A_N(\varepsilon)$ and $d(\varepsilon)$ is a continuous unknown function on $[0, \varepsilon_1^0]$.

We observe that

$$X(\varepsilon)=\begin{bmatrix} t_{12}(\varepsilon) \\ t_{22}(\varepsilon) \end{bmatrix}$$

must be the eigenvector corresponding to $\lambda_2(\varepsilon)$. In view of Proposition 2.1 our result follows for lower triangular matrices. Using the transpose of $A(\varepsilon)$ one easily verifies the statement about upper triangular matrices in our proposition.

We observe that

(1) the entries of $A_N(\varepsilon)$ vanish infinitely many times on $(0, \varepsilon_1^0]$,

(2) the quotient of some of two elements in $A_N(\varepsilon)$ is *not continuous* on any interval $(0, \varepsilon_1^0)$ where ε_1^0 is arbitrarily small.

This observation leads us to the next section.

3. Comparable elements and canonical forms.

DEFINITION 3.1. We say that a finite set of continuous functions on $\bar{D}_\varepsilon$ is *comparable* if the following hold.

(i) Each element of the set is either nonvanishing on D_ε or is identically zero on D_ε.

(ii) If $f_1(\varepsilon)$, $f_2(\varepsilon)$ are two elements of the set such that $f_1(\varepsilon)\cdot f_2(\varepsilon)\neq 0$ on D_ε then either there exists $\lim_{\varepsilon\to 0}[f_1(\varepsilon)/f_2(\varepsilon)]$ or there exists $\lim_{\varepsilon\to 0}[f_2(\varepsilon)/f_1(\varepsilon)]$.

The proof of the next proposition will be omitted since it is trivial.

PROPOSITION 3.1. *Let P be a given comparable set on $\bar{D}$. In $P-\{0\}$ define the relation R by $f_1(\varepsilon)Rf_2(\varepsilon)$ iff* $\lim_{\varepsilon\to 0}[f_1(\varepsilon)/f_2(\varepsilon)]\neq 0$. *Then*

(i) *this relation is an equivalence relation which induces a partition of the set $P-\{0\}$ into equivalence classes which will be denoted by $[P_i]$, such that*

$$P-\{0\}=\bigcup_{i=1}^{i=l}[P_i];$$

(ii) *in the set of equivalence classes, the following defines an order relation*

$$[P_i]<[P_j] \textit{ if } \lim_{\varepsilon\to 0}\frac{f_i(\varepsilon)}{f_j(\varepsilon)}=0$$

whenever $f_i(\varepsilon)\in[P_i]$ and $f_j(\varepsilon)\in[P_j]$, $i\neq j$;

(iii) *corresponding to the order relation in* (ii) *the set of equivalence classes possesses a minimal element which we will denote by $[P_{i_1}]$ and a maximal element which we will denote by $[P_{i_l}]$.*

The element $[P_{i_1}]$ is characterized by the fact that $[P_{i_1}]<[P_i]$ for $i=1,\cdots,l$, $i\neq i_1$. The element $[P_{i_l}]$ is characterized by the fact that $[P_i]<[P_{i_l}]$ for $i=1,\cdots,l$, $i\neq i_l$. By relabeling the indexes of $[P_i]$, $i=1,\cdots,l$ we may assume from *now on* that $i_1=1$ and $i_l=l$.

We now attempt to describe a process of triangularization and a process of diagonalization of a matrix function $A(\varepsilon)$. The processes are analogues of the algorithms described in [6, pp. 131–139] which produce canonical forms of polynomial matrices by elementary operations.

Let E denote the $n\times n$ identity matrix. Denote by E_{ij} the $n\times n$ matrix which has all of its elements zero except one element, which is one and which sits in the (i,j) place. Also denote by $\hat{S}_k$ a matrix of the form

$$\hat{S}_k=E+E_{ij}+E_{ji}-E_{ii}-E_{jj}, \tag{3.1}$$

or of the form

$$\hat{S}_k=E+b_k(\varepsilon)E_{ij}, \tag{3.2}$$

where $b_k(\varepsilon)$ is a continuous function on $\bar{D}_\varepsilon$, nonvanishing on D_ε.

Description of a triangularization process. Let $A(\varepsilon)$ be an $n\times m$ matrix function such that the elements in its first column are comparable. By Proposition 3.1 we have that $S-\{0\}$, the set of elements on the first column of $A(\varepsilon)$, (if not identically zero) possesses an element $f(\varepsilon)\neq 0$ on D_ε, $f(\varepsilon)\in[P_l]$. By multiplying $A(\varepsilon)$ on the left by matrices of the form (3.1), one brings $f(\varepsilon)\in[P_l]$ into the (1, 1) place. Since $f(\varepsilon)\in[P_l]$ every element $a_{k1}(\varepsilon)$ on the first column of $A(\varepsilon)$ may be written as $a_{k1}(\varepsilon)=f(\varepsilon)b_k(\varepsilon)$, $k=1,\cdots,n$. Thanks to the comparability of the elements of $A(\varepsilon)$ on the first column, $b_k(\varepsilon)$ are continuous functions on $\bar{D}_\varepsilon$. We continue to multiply with matrices of the form (3.2) on the left and thus we find say k_1 continuous matrices $S_1, S_2,\cdots,S_{k_1}$ on $\bar{D}_\varepsilon$ of the form (3.1) or (3.2) such that

$$S_{k_1}\cdot\cdots\cdot S_2\cdot S_1A(\varepsilon)=\begin{bmatrix} f(\varepsilon) & b_{12}(\varepsilon) & \cdot & \cdot & \cdot & b_{1m}(\varepsilon)\\ 0 & & & & & \\ \cdot & & & & & \\ \cdot & & & A_2(\varepsilon) & & \\ \cdot & & & & & \\ 0 & & & & & \end{bmatrix}. \tag{3.3}$$

Therefore the matrix function on the right of (3.3) is continuous on $\bar{D}_\varepsilon$. $A_2(\varepsilon)$ is an $(n-1)\times(m-1)$ continuous matrix function on $\bar{D}_\varepsilon$. Notice that the $b_k(\varepsilon)$ which appear in the matrices of the form (3.2) are quotients of continuous functions that sit on the first column of the matrix $A(\varepsilon)$, where the functions in the denominators may vanish at the point $\varepsilon=0$.

Assume now that for $\nu<\min\{n,m\}$ there exists k_ν matrix functions $S_1,\cdots,S_{k_\nu}$, each of them of type (3.1) or (3.2) and such that

$$S_{k_\nu}\cdot\cdots\cdot S_2\cdot S_1A(\varepsilon)=\begin{bmatrix} a_{11}(\varepsilon) & \cdots & & & & & a_{1m}(\varepsilon)\\ 0 & a_{21}(\varepsilon) & \cdots & & & & a_{2m}(\varepsilon)\\ \cdot & 0 & \ddots & & & & \\ \cdot & \cdot & & a_{ii}(\varepsilon) & \cdot & \cdot\ \cdot & a_{im}(\varepsilon)\\ \cdot & \cdot & & 0 & & & \\ \cdot & \cdot & & \vdots & & A_{i+1}(\varepsilon) & \\ 0 & 0 & & 0 & & & \end{bmatrix} \tag{3.4}$$

where the set of elements of the first column of $A_{i+1}(\varepsilon)$ is comparable on $\bar{D}_\varepsilon$, then there exist $k_{\nu+1}$ matrix functions, $S_1, \cdots, S_{k_{\nu+1}}$ each of type (3.1) or (3.2) such that $S_{k_{\nu+1}} \cdot \cdots \cdot S_2 \cdot S_1 A(\varepsilon)$ has the form (3.4) with i replaced by $i+1$.

If in each step, $A_{i+1}(\varepsilon)$ will have the set of elements of its first column comparable on $\bar{D}_\varepsilon$ one will get via a reduction process described above a canonical form of the matrix $A(\varepsilon)$ which will be either

$$\begin{bmatrix} a_{11}(\varepsilon) & \cdots & & & & & a_{1m}(\varepsilon) \\ 0 & a_{22}(\varepsilon) & & & & & \\ \vdots & 0 & \ddots & & & & \\ \vdots & \vdots & & \ddots & & & \\ 0 & 0 & \cdots & 0 & a_{nn}(\varepsilon) & \cdots & a_{nm}(\varepsilon) \end{bmatrix}, \quad (\text{if } n \leqq m) \tag{3.5}$$

or

$$\begin{bmatrix} a_{11}(\varepsilon) & \cdots & & & a_{1m}(\varepsilon) \\ 0 & & & & \vdots \\ \vdots & a_{22}(\varepsilon) & & & \vdots \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 & a_{mm}(\varepsilon) \\ \vdots & & & & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & & & 0 \end{bmatrix} \quad (\text{if } m \leqq n). \tag{3.6}$$

A similar result holds for $A(\varepsilon)$ by operating on $A(\varepsilon)$ with elementary operations from its right.

Our goal now is to describe a process by which the canonical form obtained for the matrix $A(\varepsilon)$ will be a diagonal matrix function. This will be achieved by operating on $A(\varepsilon)$ by right and left elementary operations.

Description of a diagonalization process. Let $A(\varepsilon)$ be an $n \times m$ continuous matrix function. Assume that the set of all elements of $A(\varepsilon)$ is comparable on $\bar{D}_\varepsilon$. By Proposition 3.1 we know that if the set S of elements of $A(\varepsilon)$ does not consist only of the element 0, then $S - \{0\} = \bigcup_{i=1}^{i=l} [S_i]$ and there exists a maximal element $[S_l]$.

(1) Choose an element $f(\varepsilon) \in [S_l]$ and multiply $A(\varepsilon)$ from the left and from its right by matrices of type (3.1) to make $f(\varepsilon)$ appear in the (1, 1) place in the resulting matrix function.

We multiply the new matrix thus obtained by matrices of type (3.2) from its left and from its right to obtain a new matrix function all of whose elements in the first row and on its first column (except possibly the element in the (1, 1) place) are 0.

So far we proved that there exist k_1 matrices $\hat{S}_1, \cdots, \hat{S}_{k_1}$ of the type (3.1) and (3.2) and j_1 matrices $\tilde{S}_1, \cdots, \tilde{S}_{j_1}$, of the type (3.1) and (3.2), such that

$$\left(\prod_{k=1}^{k=i_1}\hat{S}_k\right)A(\varepsilon)\left(\prod_{k=1}^{k=j_1}\tilde{S}_k\right)=\begin{bmatrix} f(\varepsilon) & 0\cdot & \cdot & \cdot\, 0 \\ 0 & & & \\ \cdot & & A_2(\varepsilon) & \\ \cdot & & & \\ \cdot & & & \\ 0 & & & \end{bmatrix}.$$

We adopt an inductive procedure. Assume that for every $i<\min\{n, m\}$ there exist ν_i matrices $\hat{S}_1, \cdots, \hat{S}_{\nu_i}$ of type (3.1) or (3.2) and ν_j matrices $\tilde{S}_1, \cdots, \tilde{S}_{\nu_j}$ of type (3.1) or (3.2) such that

$$\left[\prod_{k=1}^{k=\nu_i}\hat{S}_k\right]A(\varepsilon)\left[\prod_{k=1}^{k=\nu_j}\tilde{S}_k\right]=\begin{bmatrix} b_{11}(\varepsilon) & 0\cdots & & & 0 \\ 0 & b_{22}(\varepsilon) & 0\cdots & & 0 \\ \cdot & 0 & & & \\ \cdot & & & & \\ 0 & \cdots & b_{ii}(\varepsilon) & 0\cdots & 0 \\ & & 0 & & \\ & & \cdot & & \\ & & \cdot & A_{i+1}(\varepsilon) & \\ & & \cdot & & \\ 0 & \cdots & 0 & & \end{bmatrix},$$

where the set of elements of $A_{i+1}(\varepsilon)$ is comparable on $\bar{D}_\varepsilon$. Then, there exist (N_1+N_2) invertible matrix functions $\hat{S}_1, \cdots, \hat{S}_{N_1}, \tilde{S}_1, \cdots, \tilde{S}_{N_2}$ each of the form (3.1) or (3.2) such that

$$\left[\prod_{k=1}^{k=N_1}S_k\right]A(\varepsilon)\left[\prod_{k=1}^{k=N_2}S_k\right]=\begin{bmatrix} b_{11}(\varepsilon) & 0\cdots & & & & 0 \\ 0 & b_{22}(\varepsilon) & & & & \\ \cdot & & \ddots & & & \\ \cdot & & & & & \\ 0 & & \cdots\, 0 & b_{ll}(\varepsilon) & 0\cdots & 0 \\ 0 & & \cdots & 0 & \cdots & 0 \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & \cdot & & \cdot \\ 0\cdots & & & 0 & & 0 \end{bmatrix},$$

$l<\min\{n, m\}$, and $b_{ii}(\varepsilon)$, for $i=1, 2, \cdots, l$, are nonvanishing on D_ε.

4. Continuous eigenvectors and canonical forms.

PROPOSITION 4.1. *Let $\lambda_i(\varepsilon)$, $i=1, 2, \cdots, n$, be a continuous eigenvalue on $\bar{D}_\varepsilon$, of the continuous $n\times n$ matrix function $A(\varepsilon)$.*

Let $[A(\varepsilon)-\lambda_i(\varepsilon)E]$ satisfy the assumptions in the diagonalization process. Then, to $\lambda_i(\varepsilon)$ there corresponds a continuous eigenvector $X(\varepsilon)$ on $\bar{D}_\varepsilon$.

Proof. By the diagonalization process there exist two continuous invertible matrices on $\bar{D}_\varepsilon$, namely $L(\varepsilon)$ and $R(\varepsilon)$ such that

$$L(\varepsilon)[A(\varepsilon)-\lambda_i(\varepsilon)E]R(\varepsilon)=\begin{bmatrix} a_{11}(\varepsilon) & & & & & & \\ & \cdot & & & 0 & & \\ & & \cdot & & & & \\ & & & a_{kk}(\varepsilon) & & & \\ & & & & 0 & & \\ & 0 & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & 0 \end{bmatrix} \tag{4.1}$$

where $k<n$.

From (4.1) it follows that there exists an n-column continuous vector function on $\bar{D}_\varepsilon$, namely

$$X(\varepsilon)=\begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ x_{k+1}(\varepsilon) \\ \cdot \\ \cdot \\ \cdot \\ x_n(\varepsilon) \end{bmatrix} \tag{4.2}$$

where $x_{k+1}(\varepsilon), \cdots, x_n(\varepsilon)$ are *any* $(n-k)$ continuous functions on $\bar{D}_\varepsilon$ not all of them vanishing simultaneously such that

$$L(\varepsilon)[A(\varepsilon)-\lambda_i(\varepsilon)E]R(\varepsilon)X(\varepsilon)=0. \tag{4.3}$$

From (4.3) it follows that $R(\varepsilon)X(\varepsilon)$ is the desired continuous eigenvector corresponding to $\lambda_i(\varepsilon)$.

THEOREM 4.2. *Let $A(\varepsilon)$ be a continuous $n\times n$ matrix function on $\bar{D}_\varepsilon$. Assume also $[A_{n+1-i}(\varepsilon)-\lambda_i(\varepsilon)E_{n+1-i}]$ to satisfy the assumptions of Proposition* 4.1, *where*: E_{n+1-i}, $i=1,2,\cdots,n$, *are the* $(n+1-i)\times(n+1-i)$ *identity matrices and* $A_{n+1-i}(\varepsilon)$ *are* $(n+1-i)\times(n+1-i)$ *matrix functions to be obtained by an inductive procedure* (*to be described later*). *Then, there exists a unitary continuous function on $\bar{D}_\varepsilon$ which will be denoted by $U(\varepsilon)$ such that $U^{-1}(\varepsilon)A(\varepsilon)U(\varepsilon)$ is upper triangular.*

Proof. We follow the proof of Schur's theorem in [2, p. 202]. By Proposition 4.1 $A(\varepsilon)-\lambda_1(\varepsilon)E$ possesses a continuous eigenvector $R(\varepsilon)X_1(\varepsilon)$ on $\bar{D}_\varepsilon$, where $X_1(\varepsilon)$ is given by (4.2) and $R(\varepsilon)$ is nonsingular.

We construct any $(n-1)$ continuous vectors on $\bar{D}_\varepsilon$, $X_i(\varepsilon)$, $i=2,\cdots,n$ such that $X_i(\varepsilon)$, $i=1,2,\cdots,n$ are linearly independent on $\bar{D}_\varepsilon$. This is always possible by choosing $X_i(\varepsilon)$ to have exactly one nonvanishing coordinate. Since $X_i(\varepsilon)$, $i=1,2,\cdots,n$ are linearly independent, so are $R(\varepsilon)X_i(\varepsilon)$, $i=1,2,\cdots,n$, since $R(\varepsilon)$ is nonsingular.

We recall the Gram–Schmidt orthogonalization process from [6, Chap. IX]. Since $R(\varepsilon)X_i(\varepsilon)$, $i=1,2,\cdots,k$, $1\leqq k\leqq n$, are linearly independent on $\bar{D}_\varepsilon$ their Gramian does not vanish on $\bar{D}_\varepsilon$ and this is sufficient to produce n orthogonal, n-column vectors $Y_i(\varepsilon)$, $i=1,2,\cdots,n$, continuous on $\bar{D}_\varepsilon$ and such that

$$Y_1(\varepsilon)=R(\varepsilon)X_1(\varepsilon). \tag{4.4}$$

We construct the unitary matrix $U_1(\varepsilon)$

$$U_1(\varepsilon)=\left[\frac{Y_1(\varepsilon)}{\langle Y_1(\varepsilon), Y_1(\varepsilon)\rangle}\cdots\frac{Y_i(\varepsilon)}{\langle Y_i(\varepsilon), Y_i(\varepsilon)\rangle}\cdots\frac{Y_n(\varepsilon)}{\langle Y_n(\varepsilon), Y_n(\varepsilon)\rangle}\right], \tag{4.5}$$

where $\langle Y, Y\rangle$ denotes the Euclidean norm.

The matrix $U_1(\varepsilon)$ is easily observed to satisfy

$$U_1^{-1}(\varepsilon)A(\varepsilon)U_1(\varepsilon)=\begin{bmatrix}\lambda_1(\varepsilon) & b_{12}^1(\varepsilon) & \cdot & \cdot & \cdot\, b_{1n}^1(\varepsilon)\\ 0 & & & & \\ \cdot & & A_{n-1}(\varepsilon) & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & & & & \end{bmatrix} \tag{4.6}$$

where $A_{n-1}(\varepsilon)$ is an $(n-1)\times(n-1)$ continuous matrix function on $\bar{D}_\varepsilon$ with eigenvalues $\lambda_2(\varepsilon),\cdots,\lambda_n(\varepsilon)$.

If $[A_{n-1}(\varepsilon)-\lambda_2(\varepsilon)E_{n-1}]$ satisfies assumptions of Proposition 4.1, we can repeat the process described above, etc.:

$$U_2^{-1}(\varepsilon)A(\varepsilon)U_2(\varepsilon)=\begin{bmatrix}\lambda_1(\varepsilon) & b_{12}^2(\varepsilon) & & \cdots & b_{2n}^2(\varepsilon)\\ 0 & \lambda_2(\varepsilon) & b_{23}^2(\varepsilon) & \cdots & b_{2n}^2(\varepsilon)\\ \cdot & 0 & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & A_{n-2}(\varepsilon) & \\ & \cdot & & & \\ 0 & 0 & & & \end{bmatrix} \tag{4.7}$$

where $A_{n-2}(\varepsilon)$ is continuous on $\bar{D}_\varepsilon$ and has the eigenvalues $\lambda_3(\varepsilon),\cdots,\lambda_n(\varepsilon)$.

Assume we have found a unitary matrix $U_i(\varepsilon)$ which is continuous on $\bar{D}_\varepsilon$ and is such that

$$U_i^{-1}(\varepsilon)A(\varepsilon)U_i(\varepsilon)=\begin{bmatrix}\lambda_1(\varepsilon) & b_{12}^i(\varepsilon) & \cdots & & & b_{1n}^i(\varepsilon)\\ 0 & \lambda_2(\varepsilon) & b_{23}^i(\varepsilon) & \cdots & & b_{2n}^i(\varepsilon)\\ \cdot & 0 & & & & \\ \cdot & \vdots & & & & \\ \cdot & & & & & \\ & & & \lambda_i(\varepsilon) & \cdots & b_{in}^i(\varepsilon)\\ & & & 0 & & \\ & & & \cdot & & \\ & & & \cdot & A_{n-i}(\varepsilon) & \\ & & & \cdot & & \\ 0 & 0 & & 0 & & \end{bmatrix} \tag{4.8}$$

$i<n$ and $[A_{n-i}(\varepsilon)-\lambda_{i+1}(\varepsilon)E_{n-i}]$ satisfies the assumptions of Proposition 4.1.

Then by an induction procedure there exists a unitary matrix $U_{i+1}(\varepsilon)$ continuous on $\bar{D}_\varepsilon$ such that (4.7) is satisfied with i replaced by $i+1$, and the result follows.

It can be easily verified that in case $\varepsilon = \varepsilon_1$ and $\bar{D}_\varepsilon$ is the interval $0 < \varepsilon \leqq r < 1$ then the unitary continous matrix function on $\bar{D}_\varepsilon$,

$$U(\varepsilon) = \begin{bmatrix} \dfrac{-\sqrt{\varepsilon}}{\sqrt{1+\varepsilon}} & \dfrac{1}{\sqrt{1+\varepsilon}} \\ \dfrac{1}{\sqrt{1+\varepsilon}} & \dfrac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon}} \end{bmatrix}$$

takes the matrix

$$A(\varepsilon) = \begin{bmatrix} 0 & \varepsilon \\ 1 & 0 \end{bmatrix}$$

into an upper triangular matrix by a similarity transformation.

It is easily verified that $U(\varepsilon)$ is not differentiable at $\varepsilon = 0$. Thus, even if $A(\varepsilon)$ is analytic at $\varepsilon = 0$ the transforming matrix $U(\varepsilon)$ is not differentiable at the point $\varepsilon = 0$. We observe that by the algorithm of Proposition 4.1, the vector $X(\varepsilon)$ given by (4.2) may be taken to be analytic on $\bar{D}_\varepsilon$. This is true since the main restriction on $x_{k+1}(\varepsilon), \cdots, x_n(\varepsilon)$ is that they don't vanish simultaneously on $\bar{D}_\varepsilon$. However, the smoothness of $U(\varepsilon)$ at the point $\varepsilon = 0$ is also determined by the matrix $R(\varepsilon)$ and by the orthonormalization process.

We are ready now to point out some conditions that will guarantee the fulfillment of the assumptions that Theorem 4.2 is based on.

5. Sufficient conditions. In what follows we describe conditions that will guarantee the possibility of performing one or more steps of the algorithms proposed in §§ 3, 4.

DEFINITION 5.1. We say that the set F_0, $F_0 = \{f_1(\varepsilon), f_2(\varepsilon), \cdots,\}$ is *competent on* $\bar{D}_\varepsilon$ if:

i) $f_i(\varepsilon)$, $i = 1, 2, \cdots$ are continuous on $\bar{D}_\varepsilon$, and each $f_i(\varepsilon)$ is nonvanishing on D_ε or is identically zero;

ii) whenever $f_1(\varepsilon), f_2(\varepsilon) \in F_0$ and $c_1 f_1(\varepsilon(k)) + c_2 f_2(\varepsilon(k)) = 0$ for some constants c_1, c_2, and $\varepsilon(k)$, $k = 1, 2, \cdots$ is an infinite sequence $\varepsilon(k) \in D_\varepsilon$ with $\varepsilon(k) \to 0$ it is implied that

$$c_1 f_1(\varepsilon) + c_2 f_2(\varepsilon) = 0 \quad \text{on } \bar{D}_\varepsilon.$$

PROPOSITION 5.1. *Let F_0 be a competent set of real functions on $\bar{D}_\varepsilon$. Then F_0 is comparable on $\bar{D}_\varepsilon$.*

Proof. By Definition 5.1 it suffices to show that if $f_1(\varepsilon) \cdot f_2(\varepsilon) \neq 0$ on D_ε then there exist

$$\lim_{\varepsilon \to 0} \frac{f_1(\varepsilon)}{f_2(\varepsilon)} \quad \text{or} \quad \lim_{\varepsilon \to 0} \frac{f_2(\varepsilon)}{f_1(\varepsilon)}.$$

Denote by

$$\bar{l} = \limsup_{\varepsilon \to 0} \frac{f_1(\varepsilon)}{f_2(\varepsilon)},$$

$$\underline{l} = \liminf_{\varepsilon \to 0} \frac{f_1(\varepsilon)}{f_2(\varepsilon)}$$

where $\bar{l}, \underline{l}$ are finite or infinite.

If $\bar{l} = \underline{l}$ we are done.

If $l < \bar{l}$ then for any l, $\underline{l} < l < \bar{l}$ there exists an infinite sequence $\{\varepsilon(\nu)\}_{\nu=0}^{\infty}$, $\varepsilon(\nu) \to 0$ such that $f_1(\varepsilon(\nu))/f_2(\varepsilon(\nu)) = l$ or $f_1(\varepsilon(\nu)) - lf_2(\varepsilon(\nu)) = 0$. By definition it is implied that $f_1(\varepsilon) = f_2(\varepsilon)l$. Therefore $\bar{l} = \underline{l}$ and the result follows.

DEFINITION 5.2. Given a competent set of functions F_0 on $\bar{D}_\varepsilon$, we denote by $EX(F_0)$ the new set of elements

$$EX(F_0) = \left\{ e \,\middle|\, e = \sum_{l+r=1}^{l+r=N} c_{lr} \frac{\prod_{\nu=1}^{\nu=r} f_\nu^{i_\nu}(\varepsilon)}{\prod_{\mu=1}^{\mu=l} f_\mu^{k_\mu}(\varepsilon)}, \; f_i(\varepsilon) \in F_0 \right\} \cup \{0\}, \tag{5.1}$$

whenever e can be made continuous on $\bar{D}_\varepsilon$. We have c_{lr} as complex numbers and $l, r, N, \nu, \mu, i, j_\nu, k_\mu$ are nonnegative integers. Actually, each e, is composed of all finite sums of all finite products of functions of F_0 or of all quotients of finite products of functions of F_0 whenever they determine a continuous function on $\bar{D}_\varepsilon$.

PROPOSITION 5.2. *Let $A(\varepsilon)$ be a matrix whose elements form a competent set on $\bar{D}_\varepsilon$. Denote this set by F_A.*

Let $EX(F_A)$ be a competent set on $\bar{D}_\varepsilon$. Then the set $\{\lambda(\varepsilon)\} \cup EX(F_A)$ is competent on a closed subdomain of $\bar{D}_\varepsilon$ which contains the point 0, *where $\lambda(\varepsilon)$ is a continuous eigenvalue of $A(\varepsilon)$ on $\bar{D}_\varepsilon$.*

Proof. Given c_1, c_2, any two numbers, we have to prove that if $c_1\lambda(\varepsilon(k)) + c_2 f(\varepsilon(k)) = 0$, where $\{\varepsilon(k)\}_{k=0}^{\infty}$ is a sequence such that $\varepsilon(k) \to 0$ for $k \to \infty$, and $f(\varepsilon) \in EX(F_A)$ then $c_1\lambda(\varepsilon) + c_2 f(\varepsilon) = 0$ on some $\hat{D}_\varepsilon$. Without loss of generality assume $c_1 \neq 0$. Put

$$u(\varepsilon) = c_1\lambda(\varepsilon) + c_2 f(\varepsilon). \tag{5.2}$$

It is readily observed that

$$\lambda(\varepsilon) = c_1^{-1} u(\varepsilon) - c_2 c_1^{-1} f(\varepsilon) \tag{5.3}$$

and since

$$\det[A(\varepsilon) - \lambda(\varepsilon)E] = 0,$$

then $u(\varepsilon)$ satisfies the algebraic equation

$$u^n(\varepsilon) + p_1(\varepsilon)u^{n-1}(\varepsilon) + \cdots + p_n(\varepsilon) = 0 \tag{5.4}$$

where $p_i(\varepsilon) \in EX(F_A)$, $i = 1, 2, \cdots, n$.

From this one obtains that $u(\varepsilon(k)) = 0$ implies $p_n(\varepsilon(k)) = 0$. But $p_n(\varepsilon) \in EX(F_A)$ implies that $p_n(\varepsilon) = 0$ on $\bar{D}_\varepsilon$. This implies that the polynomial in (5.4) splits into

$$u(\varepsilon)[u^{n-1}(\varepsilon) + p_1(\varepsilon)u^{n-2}(\varepsilon) + \cdots + p_{n-1}(\varepsilon)] = 0. \tag{5.5}$$

If $u(\varepsilon)$ is not identically zero in a full neighborhood of 0 there exists a domain $\bar{D}_\varepsilon^1 \subset \bar{D}_\varepsilon$, $0 \in \bar{D}_\varepsilon^1$, such that

$$u^{n-1}(\varepsilon) + p_1(\varepsilon)u^{n-2}(\varepsilon) + \cdots + p_{n-1}(\varepsilon) = 0 \tag{5.6}$$

on $\bar{D}_\varepsilon^1$.

Since $p_{n-1}(\varepsilon) \in EX(F_A)$ one obtains again from $p_{n-1}(\varepsilon(k)) = 0$ that it implies $p_{n-1}(\varepsilon) = 0$ on $\bar{D}_\varepsilon^1$.

By an inductive procedure one obtains that $p_i(\varepsilon) = 0$ for $i = 1, 2, \cdots, n$, on a domain $\bar{D}_\varepsilon^{n-1}$ which contains the point 0. This implies $u^n(\varepsilon) = 0$ on $\bar{D}_\varepsilon^{n-1}$ and the result follows.

PROPOSITION 5.3. *With the assumptions of Proposition* 5.2 *assume also the elements of $A(\varepsilon)$ and $\lambda(\varepsilon)$ to be real on $\bar{D}_\varepsilon$. Then, the elements of $[A(\varepsilon) - \lambda(\varepsilon)E]$ are comparable on some domain $\bar{D}_\varepsilon^{n-1}$ containing the point* 0.

Proof. By Proposition 5.2 the elements of $[A(\varepsilon)-\lambda(\varepsilon)E]$ are competent on $\bar{D}_\varepsilon^{n-1}$. By Proposition 5.1 our result follows.

PROPOSITION 5.4. *Let $A(\varepsilon)$ be an $n\times n$ real analytic matrix function in the domain $\bar{D}_\varepsilon=\{\varepsilon=\varepsilon_1|0\leqq\varepsilon\leqq 1\}$. Then, there exists an $r>0$, and a unitary continuous matrix function $U(\varepsilon)$ on $\bar{D}_\varepsilon=\{\varepsilon|0\leqq\varepsilon\leqq r\}$ such that $U^{-1}(\varepsilon)A(\varepsilon)U(\varepsilon)$ has a triangular form. Moreover, $U(\varepsilon)$ is a differentiable matrix function in the variable ε, on $0<\varepsilon\leqq r$, and there exists a number $\delta>0$, such that $\varepsilon^\delta U'(\varepsilon)$ is bounded on $0<\varepsilon\leqq r$.*

Proof. Since $A(\varepsilon)$ is real analytic on $\bar{D}_\varepsilon$ its characteristic polynomial has real analytic coefficients on $\bar{D}_\varepsilon$.

It is well known (see [8, Chap. 12]) that every eigenvalue $\lambda_i(\varepsilon)$ of $A(\varepsilon)$ has an expansion of the form

$$\lambda_i(\varepsilon)=\varepsilon^{p_i/q_i}\sum_{\nu=0}^{\infty} b_{i\nu}\varepsilon^{\nu/r_i} \tag{5.7}$$

where; p_i, q_i, r_i are integers, $b_{i0}\neq 0$ and the series in (5.7) is absolutely convergent in a disk $|\varepsilon|\leqq r<1$.

Moreover, for any two functions $f_1(\varepsilon)$, $f_2(\varepsilon)$, which have the expansion (5.7) there exists either

$$\lim_{\varepsilon\to 0^+}\frac{f_1(\varepsilon)}{f_2(\varepsilon)} \quad\text{or}\quad \lim_{\varepsilon\to 0^+}\frac{f_2(\varepsilon)}{f_1(\varepsilon)}.$$

We observe that if $f(\varepsilon)$ has an expansion of the form (5.7) which is absolutely convergent then $f(\varepsilon)$ and $|f(\varepsilon)|$ are differentiable functions of ε for $0<\varepsilon\leqq r$.

Since the elements of the unitary matrix are obtained by addition, multiplication division and taking square roots of functions $f(\varepsilon)$ of the form (5.7) our result follows. In particular we proved the following proposition.

PROPOSITION 5.5. *With the assumption of Proposition 5.4 let $A(\varepsilon)$ be Hermitian; then $U^{-1}(\varepsilon)A(\varepsilon)U(\varepsilon)$ is diagonal.*

Proof. See [2, p. 202]. This proposition is a weaker form of Rellich's result [12]. By his result, the eigenvalues $\lambda_i(\varepsilon)$ of a Hermitian matrix are real analytic on $0\leqq\varepsilon\leqq r$. This in particular implies that in this case $U(\varepsilon)$ is differentiable on $0\leqq\varepsilon\leqq r$, since all elements involved in the algorithm which derives $U(\varepsilon)$ have an expansion (5.7) with $r_i=1$ and p_i/q_i a nonnegative integer.

Acknowledgment. Acknowledgment is due to Professor Wasow who introduced me to some of the references quoted.

REFERENCES

[1] V. I. ARNOLD, *On matrices depending on parameters*, Uspehi Mat. Nauk, 26 (1971), no. 2, p. 101; Russian, translation in Russian Math. Surveys, vol. 26, no. 2, pp. 29–43.
[2] R. BELLMAN, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960.
[3] B. L. J. BRAAKSMA, *Global reduction of linear differential systems involving a small singular parameter*, this Journal, 2 (1971), pp. 149–165.
[4] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
[5] W. A. COPPEL, *Dichotomies and reducibility*, J. Differential Equations, 3 (1967), pp. 500–521.
[6] F. R. GANTMACHER, *The Theory of Matrices*, Chelsea, New York, 1966.
[7] H. GINGOLD, *On the existence of a global simplifying transforming matrix*, this Journal, to appear.
[8] E. HILLE, *Analytic Function Theory*, vol. I, II, Blaisdell, New York, 4th ed., 1965.
[9] P. F. HSIEH AND Y. SIBUYA, *A global analysis of matrices of functions of several variables*, J. Math. Anal., 14 (1966), 332–340.

[10] O. PERRON, *Über eine Matrixtransformation*, Math. Z., 32 (1930), pp. 465–473.
[11] ———, *Die Stabilitatsfrage bei Differentialgleichungen*, Ibid., 32 (1930), pp. 703–728.
[12] F. RELLICH, *Störungstheorie der Spektralzerlegung, I. Mitteilung.* Math. Ann., 113 (1937), pp. 600–619.
[13] Y. SIBUYA, *Some global properties of matrices of functions of one variable*, Math. Ann., 161 (1965), pp. 67–77.
[14] W. R. WASOW, *On holomorphically similar matrices*, J. Math. Anal. Appl., 4 (1962), pp. 202–206.
[15] ———, *Asymptotic Expansions for Ordinary Differential Equations*, Wiley, New York, 1966.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0005$01.00/0

PERTURBATION AND MAXIMUM NUMBER OF INFECTIVES OF SOME *SIR* EPIDEMIC MODELS*

FRANK J. S. WANG†

Abstract. We consider a deterministic *SIR* model in which the daily contact rate is a constant $\alpha > 0$, the rate of an infected individual recovering from the disease depends only on how long it has had the disease, and the recovered individuals are permanently immune from further attack. Let $F(t)$ be the probability of an infected individual staying infected for a length of time $\geqq t$; $f(t)$ be the proportion of initial infectives who are still infective at time t and $I(0)$, $S(0)$ be the proportion of infectives and susceptibles respectively at time zero. We examine the effects of perturbation of the initial condition f on the deterministic model which is given by a system of integral equations and show that the solution of our system of equations is stable at f if and only if the L^1-norm of f is not zero. We also prove that, under appropriate conditions, there exists a positive number θ depending only on the generalized relative removal rate $\rho = (\alpha\|F\|)^{-1}$ such that when $I(0) > 0$ and $S(0) > \rho$, the epidemic occurs and at the peak of the epidemic the proportion of infectives is at least θ. A numerical example is given at the end of the paper.

1. Introduction. In an *SIR* model, the population or community under consideration is divided into the classes of susceptibles S, infectives I and removed individuals R who are isolated, dead or recovered and immune. The fractions of the total population in these classes at time t are denoted by $S(t)$, $I(t)$ and $R(t)$ respectively ([1]–[3], [6]–[10]). We assume that the population is uniform and homogeneously mixing with daily contact rate or infective rate $\alpha > 0$, that the probability of an infected individual staying infected for a length of time $\geqq t$ is $F(t)$ for some nonincreasing function $F(t)$ such that $F(0) = 1$ and $F(\infty) = 0$, and that at time $t = 0$, the population consists of only susceptibles and infectives, i.e. $S(0) + I(0) = 1$. The deterministic version of this model is given by the system of equations

$$
\begin{aligned}
&I(t) = f(t) + \int_0^t \alpha \cdot I(u)S(u)F(t-u)\,du,\\
(1)\qquad &\\
&S(t) = 1 - I(0) - \int_0^t \alpha \cdot I(u)S(u)\,du,
\end{aligned}
$$

where $f(t)$ represents the proportion of initial infectives who are still infected at time t. For any integrable function f on $[0, \infty)$, we shall use $\|f\|_{[0,T]}$ and $\|f\|$ to denote the $L^1[0, T]$ and $L^1[0, \infty)$ norm of f respectively, i.e.,

$$\|f\|_{[0,T]} = \int_0^T |f(t)|\,dt, \qquad \|f\| = \int_0^\infty |f(t)|\,dt.$$

Wang [9] proved that if $I(0) > 0$ and both $f(t)$ and $F(t)$ are integrable then the system above has a unique solution pair $(I(t), S(t))$ where $I(t)$ tends to zero as t tends to infinity and $S(t)$ tends to the unique root in $(0, \rho)$ of the equation

$$1 - S + \rho \ln (S/S(0)) = I(0) - \|f\|/\|F\|$$

where $\rho = (\alpha\|F\|)^{-1}$ is the generalized relative removal rate. This implies that if $S(0) > \rho$ and $I(0) > 0$ the epidemic occurs and the proportion $S(t)$ will drop below ρ as t tends to infinity. In case $F(t)$ is exponential, it is known that $I(t)$ first increases monotonically up to a maximum value and then decreases monotonically to zero as t tends to infinity. In

* Received by the editors July 25, 1977, and in revised form February 27, 1978.
† Department of Mathematics, University of Montana, Missoula, Montana 59812.

case the length of infection is constant, stronger initial condition [10] give the same conclusion.

When F is exponential, it can be shown by examining the solution curves in the S-I plane that the maximum value of $I(t)$ is equal to $1-\rho+\rho \ln \rho-\rho \ln S(0)$, which is greater than $1-\rho+\rho \ln \rho$ for all $\rho \leqq S(0)<1$. Since in an actual epidemic the number of original cases $I(0)$ is usually a very small fraction of the total population, the value $S(0)=1$ approximates to reality and therefore, $1-\rho+\rho \ln \rho$ approximates $\max_{0 \leqq t} I(t)$. At any rate, $\max_{0 \leqq t} I(t) \geqq 1-\rho+\rho \ln \rho$ as long as $I(0)>0$. The facts that the peak value of $I(t)$ does not go to zero with $I(0)$ and is always bounded from below by some positive number depending only on ρ (not on $I(0)>0$), illustrate very well the common observation that in many actual epidemics the peak value of the proportion of the number of infectives is quite noticeable, i.e., at the "center" of the epidemic a comparatively large proportion (compared with the size of the epidemic) of individuals will stay infected simultaneously and this peak value of $I(t)$ depends very little on how small the initial number $I(0)$ of infectives is.

In § 2, we examine the effects of perturbations of the initial conditions on the solutions of systems (1), and also on the total size of the epidemic $y=1-S(\infty)$. For each given appropriate function f_m, let $I_m(t)$ and $S_m(t)$ be the corresponding solution of (1) with forcing function $f=f_m$. We show that when $f_m \to f$ in $L^1[0, \infty)$, $\|I_m-I\| \to 0$ as $m \to \infty$ if and only if $\|I\| \neq 0$. In the last section, we extend the result concerning the maximum value of $I(t)$ to more general F. We prove that under the assumptions $\rho=(\alpha\|F\|)^{-1}<1$ and $F(t+s) \leqq F(t) \cdot F(s)$ for all $t, s \geqq 0$, the maximum number of infectives, $\max_{0 \leqq t} I(t)$, is bounded below by a positive number depending only on ρ for all $I(0)>0$. Note that the assumption $F(t+s) \leqq F(t) \cdot F(s)$ for all $t, s \geqq 0$ is equivalent to the assumption that an infected individual, who has been infected for a length of time s at time τ, has a bigger chance of being removed by the time $t+\tau$ than an infected individual who has been infected for a length of time less than s at τ. A numerical example is given at the end of § 3.

2. Perturbations of the system of integral equations. Throughout this paper we assume that the function F which represents the probability of an infected individual staying infected for a length of time t and the function f which represents the proportion of initial infectives who are still infective at time t, are both nonincreasing, integrable and satisfy $0 \leqq f(0) \leqq 1$, $F(0)=1$, $F(\infty)=f(\infty)=0$ and $\rho=(\alpha\|F\|)^{-1}<1$. We first examine the dependence of the total size of the epidemic $y=1-S(\infty)$ on the initial conditions. We write $I(t, f)$, $S(t, f)$ to denote the solution of (1) with initial forcing function f. Let $y(f)=1-S(\infty, f)$ be the size of the epidemic, i.e., the proportion of the population that finally contracts the disease. It has been shown in Wang [9] that if $\|f\|>0$ then $y(f)$ is the unique positive solution in $[1-\rho, 1)$ of the equation

$$y(f)+c(f)\, e^{-y(f)/\rho}-1=0 \tag{2}$$

where $c(f)=(1-f(0)) \cdot \exp\{f(0)/\rho-\alpha \cdot \|f\|\}$. To see that the preceding equation has a unique solution in $[1-\rho, 1)$, consider the left hand side of (2) as a function of y and denote it by $p(y)$. Then $p''>0$, $p(f(0))<0$ and $p(1)>0$ imply there exists a unique solution in $(f(0), 1)$. Since it follows from Theorem 2 of Wang [9] that $S(\infty, f) \leqq \rho$, $y=1-S(\infty, f) \geqq 1-\rho$. Thus the unique solution of (2) in $(f(0), 1)$ is indeed in $[1-\rho, 1)$.

We shall denote the set of all nonzero forcing functions f satisfying the above hypothesis by M, that is, we define $M=\{f: f \text{ is nonincreasing}, 0 \leqq f(0) \leqq 1 \text{ and } \|f\|<\infty\}$. We know that when $\|f\| \neq 0$, $S(\infty, f)<\rho$. The next lemma shows that $S(\infty, f)$ is uniformly bounded away from ρ for all f small (or large) in the sense of belonging to a set A (or B) defined in the lemma below. The proof of the last theorem in the next section uses this lemma.

LEMMA 1. *Let $y(f)$ be the unique solution in $[1-\rho, 1)$ of equation* (2), $A = \{f: 0 < f(0) \leqq (1-\rho)/2, f \in M\}$ *and* $B_\varepsilon = \{f: \|f\| \geqq \varepsilon, f \in M\}$. *Then*

$$\inf_{f \in A} y(f) > 1-\rho, \qquad \inf_{f \in B} y(f) > 1-\rho.$$

Proof. Since $c(f) \leqq (1-f(0)) \exp\{f(0)/\rho\}$, by elementary steps we conclude that

$$\sup_{f \in A} c(f) = \frac{(1+\rho)}{2} \exp\left\{\frac{1-\rho}{2\rho}\right\} < \rho \exp\left\{\frac{1}{\rho} - 1\right\},$$

$$\sup_{f \in B_\varepsilon} c(f) \leqq \rho \exp\left\{\frac{1}{\rho} - 1\right\} e^{-\alpha\varepsilon} < \rho \exp\left\{\frac{1}{\rho} - 1\right\}.$$

The lemma follows from the fact that the larger $c(f)$ the smaller the corresponding $y(f)$ and that when $c = \rho \exp\{1/\rho - 1\}$, $y = 1-\rho$.

Remark 2. Putting $\rho^* = \sup_A S(\infty, f)$ and $\rho^B = \sup_B S(\infty, f)$, the preceding lemma implies $\rho^* < \rho$ and $\rho^B < \rho$. Note that ρ^* depends only on ρ and ρ^B on ρ and ε, and that $1-\rho^*$ is the solution of (2) with $c = (1+\rho)\exp\{(1-\rho)/2\rho\}/2$ and $1-\rho^B$ is the solution of (2) with $c = \rho \exp\{1/\rho - 1 - \alpha\varepsilon\}$. Thus ρ^* is the unique solution in $(0, \rho)$ of the equation

$$X - \rho \ln X - \frac{(1+\rho)}{2} + \rho \ln\left(\frac{1+\rho}{2}\right) = 0.$$

To examine the effects of perturbations of initial conditions on system (1), we first show that $I(t, f_m) \to I(t, f)$ in $L^1[0, T]$ if $f_m \to f$ in $L^1[0, T]$ and $f_m(0) \to f(0)$.

LEMMA 3. *Suppose F is nonincreasing, integrable, $F(0) = 1$, $F(\infty) = 0$ and $0 < T < \infty$. Then $|f_m(0) - f(0)| + \|f_m - f\|_{[0,T]} \to 0$ as $m \to \infty$ implies $\|I(t, f_m) - I(t, f)\|_{[0,T]} \to 0$ as $m \to \infty$.*

Proof. For simplicity, let us write $I(t, f)$, $S(t, f)$, $I(t, f_m)$ and $S(t, f_m)$ as $I(t)$, $S(t)$, $I_m(t)$ and $S_m(t)$ respectively. Since $S_m(t) = (1 - f_m(0)) \cdot \exp\{-\alpha \|I_m\|_{[0,t]}\}$, $S(t) = (1 - f(0)) \cdot \exp\{-\alpha \|I\|_{[0,t]}\}$, it follows from

$$|e^{-x} - e^{-y}| \leqq |x - y| \quad \text{for all } x, y \geqq 0,$$

that

$$|S_m(t) - S(t)| \leqq |f_m(0) - f(0)| + \alpha \int_0^t |I_m(u) - I(u)|\, du.$$

Setting

$$G(t) = \int_0^t F(u)\, du,$$

subtracting the two equations that are satisfied by I_m and I, using the preceding inequality and changing the order of integration, we obtain, after a series of manipulations,

$$|I_m(t) - I(t)| \leqq |f_m(t) - f(t)| + \alpha \|F\| \, |f_m(0) - f(0)| + \int_0^t |I_m(u) - I(u)| (\alpha F(t-u) + \alpha^2 G(t-u))\, du.$$

The basic theorem on integral inequalities (see [4, Chap. 5]) says that the solutions $|I_m(t)-I(t)|$ of the preceding inequality are less than or equal to the solution $J_m(t)$ of the corresponding equation

$$J_m(t)=|f_m(t)-f(t)|+\alpha\|F\|\,|f_m(0)-f(0)|$$
$$+\int_0^t J_m(u)\cdot(\alpha F(t-u)+\alpha^2 G(t-u))\,du.$$

Since the hypotheses about f_m and f implies (see e.g. [5, Thm. 5.4, p. 167]) $\|J_m(t)\|\to 0$ as $m\to\infty$, this proves the lemma.

THEOREM 4. *Suppose F is nonincreasing, integrable, $F(0)=1$, $F(\infty)=0$, and α is such that $\rho=(\alpha\|F\|)^{-1}<1$. Let f_m be a sequence of nonzero functions in M which converges in $L^1[0,\infty)$ norm to a function f_0 in M. Suppose further that $f_m(0)\to f_0(0)<1-\rho$ and there exists a $g\in L^1[0,\infty)$ such that $f_m\leqq g$ for all m. Then $I(t,f_m)\to I(t,f_0)$ in $L^1[0,\infty)$ if and only if $\|f_0\|\neq 0$.*

Proof. We use the same notation as in the preceding proof. If $f_0\equiv 0$, then $S_0(t)\equiv 1$ and $\|I_0\|=0$. Now $\|f_m\|\neq 0$ for $m=1,2,\cdots$ implies

$$S_m(\infty)=S_m(0)\cdot e^{-\alpha\|I_m\|}\leqq\rho<1.$$

This shows $\|I_m\|\not\to 0$ as $m\to\infty$, since $f_m(0)\to f(0)=0$ implies $S_m(0)\to 1$.

If $\|f_0\|\neq 0$, then $S=S(\infty,f_0)=(1-f(0))\cdot\exp\{-\alpha\|I_0\|\}<\rho$; thus there exists a $T_1>0$ such that $S_0(T_1)<(S+\rho)/2=\rho'<\rho$. It then follows from Lemma 3 that

$$S_m(T_1)\leqq\rho'$$

for sufficiently large m. Without loss of generality, we assume that the preceding inequality is satisfied by all $m=0,1,\cdots$. Let $\varepsilon>0$ be given. We translate the first equation in the system (1) by T_1 to obtain a new equation

$$I_m(t+T_1)=f_m(t+T_1)+\int_0^{T_1}\alpha I_m(u)S_m(u)F(t+T_1-u)\,du$$
$$+\int_0^t \alpha I_m(T_1+u)S_m(T_1+u)F(t-u)\,du,\qquad m=0,1,2,\cdots.$$

Setting $H_m(t)=I_m(t+T_1)$ and $\alpha\rho' F(u)=L(u)$, and replacing the sum of the first two terms in the right hand side of the preceding equality by its upper bound

$$h(t)=g(t+T_1)+\alpha T_1\cdot F(t),$$

and $S(T_1+u)$ in the last integrand by ρ', we obtain

$$H_m(t)\leqq h(t)+\int_0^t H_m(u)L(t-u)\,du,\qquad m=0,1,2,\cdots.$$

Again, the solutions of these integral inequalities are less than or equal to the solution H of the corresponding equation

$$H(t)=h(t)+\int_0^t H(u)L(t-u)\,du.$$

Since $h(t)$ is integrable and $\|L\|=\alpha\rho'\|F\|=\rho'/\rho<1$, H is integrable and hence there exists a T_2 depending on T_1, ε, g, α, F but not on m, such that

$$\int_{T_2}^{\infty} H(t)\,dt<\varepsilon/2.$$

Putting $T = T_1 + T_2$, then

$$\int_T^\infty |I_m(t)|\,dt = \int_{T_2}^\infty H_m(t)\,dt \leqq \int_{T_2}^\infty H(t)\,dt < \varepsilon/2, \qquad m = 0, 1, 2, \cdots.$$

Now it follows from Lemma 3 that there exists an N depending on T and hence on ε, such that

$$\|I_m - I_0\|_{[0,T]} = \int_0^T |I_m(t) - I_0(t)|\,dt < \varepsilon/2 \quad \text{if } m > N.$$

This implies

$$\begin{aligned}\|I_m - I_0\| &\leqq \|I_m - I_0\|_{[0,T]} + \int_T^\infty |I_m(t)|\,dt + \int_T^\infty |I(t)|\,dt\\ &\leqq \varepsilon \quad \text{if } m > N,\end{aligned}$$

completing the proof.

3. The maximum number of infectives. Throughout this section we assume that F and α satisfy the hypothesis in Theorem 4 and

$$F(t+s) \leqq F(t) \cdot F(s) \quad \text{for all } t, s \geqq 0.$$

That is, we assume that $F(t+s)/F(s)$, the probability of an infected individual who has contracted the disease for a length of time s staying infected for a length of time $\geqq t+s$, is less than $F(t)$, the probability of an infected individual staying infected for a length of time $\geqq t$. (Loosely speaking, the longer the time an infected individual has the disease, the greater the chance it will be removed sooner.) Since F is not necessary exponential, we need to specify the "class-age" structure of these initial infectives. By "class-age" of an infective individual, we mean the length of time an individual has been in the infective class. Let $\varepsilon(t)$ be the class-age density function of $I(0)$, in the sense that the number of initial infected individuals having class-age between a and b is given by

$$(\text{the population size}) \cdot \int_a^b \varepsilon(t)\,dt.$$

Thus, $I(0) = \|\varepsilon(t)\|$ and

$$f_\varepsilon(t) = \int_0^\infty \varepsilon(u)\left[\frac{F(t+u)}{F(u)}\right] du, \tag{3}$$

where $F(t+u)/F(u)$ is defined to be zero if $F(u) = 0$.

Since, as was pointed out earlier, in an actual epidemic the number of original cases $I(0)$ is usually a very small fraction of the total population, we shall examine the model when $I(0)$ is small. We therefore make the assumption

$$0 < I(0) = \varepsilon = \|\varepsilon(t)\| \leqq (1-\rho)/2,$$

or equivalently,

$$1 > s(0) = 1 - I(0) \geqq (1+\rho)/2.$$

Again, when no confusion arises, we shall drop the f_ε in $I(t, f_\varepsilon)$ and $S(t, f_\varepsilon)$ and write them as $I(t)$ and $S(t)$. It follows from Remark 2 that there exists a t^* such that

$S(t^*)=(\rho+\rho^*)/2$. Since

$$S(t)=S(0)\cdot \exp\left\{-\int_0^t \alpha I(u)\,du\right\},$$

a simple calculation gives

$$\int_0^{t^*} I(u)\,du=\frac{1}{\alpha}\cdot \ln\left(\frac{2S(0)}{\rho+\rho^*}\right),$$

and

$$\int_{t^*}^{\infty} I(u)\,du=\frac{1}{\alpha}\cdot \ln\frac{\rho+\rho^*}{2s(\infty)}\geqq\frac{1}{\alpha}\cdot \ln\frac{\rho+\rho^*}{2\rho^*}. \tag{4}$$

For $t\geqq t^*$, the first equation in (1) may be written as

$$I(t^*+t)=f_\varepsilon(t^*+t)+\int_0^{t^*}\alpha I(u)S(u)F(t^*+t-u)\,du$$
$$+\int_{t^*}^{t^*+t}\alpha I(u)S(u)F(t^*+t-u)\,du.$$

Integrating this equation over t and applying the inequality $S(t^*+t)\leqq S(t^*)=(\rho+\rho^*)/2$ for $t\geqq 0$, we obtain

$$\int_{t^*}^{\infty} I(t)\,dt\leqq\int_0^{\infty} f_\varepsilon(t^*+t)\,dt+\int_0^{t^*}\alpha I(u)S(u)\left(\int_{t^*-u}^{\infty}F(t)\,dt\right)du$$
$$+\frac{(\rho+\rho^*)}{2\rho}\int_{t^*}^{\infty} I(t)\,dt.$$

The assumption $F(t+s)\leqq F(t)\cdot F(s)$ for all $t, s\geqq 0$ implies that the first two terms in the preceding sum are less than $\|F\|f_\varepsilon(t^*)$ and

$$\|F\|\cdot\int_0^{t^*}\alpha I(u)S(u)F(t^*-u)\,du=\|F\|[I(t^*)-f_\varepsilon(t^*)]$$

respectively. This together with (4) gives us the inequality

$$I(t^*)\geqq\frac{\rho-\rho^*}{2}\ln\frac{\rho+\rho^*}{2\rho^*}.$$

We summarize our results into next theorem.

THEOREM 5. *Let F be a nonincreasing integrable function on $[0,\infty)$ such that $F(0)=1, F(\infty)=0, \rho=(\alpha\cdot\|F\|)^{-1}<1$ and $F(t+s)\leqq F(t)\cdot F(s)$ for all $t, s\geqq 0$. Let ρ^* be the unique solution in $(0,\rho)$ of the equation*

$$x-\rho\ln x-\frac{1+\rho}{2}+\rho\ln\frac{1+\rho}{2}=0,$$

and let $I(t,f_\varepsilon)$, $S(t,f_\varepsilon)$ be the solution of (1) corresponding to the initial condition $f_\varepsilon(t)$ defined as in (3) where $\varepsilon(t)$ is such that $0<\varepsilon=\|\varepsilon(t)\|\leqq(1-\rho)/2$. Then

$$\max_{0\leqq t} I(t,f_\varepsilon)\geqq\frac{\rho-\rho^*}{2}\ln\frac{\rho+\rho^*}{2\rho^*}.$$

As an example, consider the case where the generalized relative removal rate $\rho = (\alpha \cdot \|F\|)^{-1} = 0.8$. Then ρ^* is the unique solution in $(0, 0.8)$ of the equation

$$X - 0.8 \ln X - 0.9843 = 0.$$

Thus $\rho^* = 0.705$ and

$$\frac{\rho - \rho^*}{2} \ln \frac{\rho + \rho^*}{2\rho^*} = 0.31\%.$$

Our theorem implies that when $I(0) > 0$ and $S(0) > 0.8$, the epidemic occurs and at the peak of the epidemic the proportion of the infective in the total population is at least 0.31%. However when $F(t)$ is exponential, a better estimate

$$1 - \rho + \rho \ln \rho = 2.14\%$$

of the maximum value of the proportion of infectives is obtained.

Acknowledgment. The author is grateful to Professor Howard Reinhardt for helpful suggestions.

REFERENCES

[1] N. T. J. BAILEY, *The Mathematical Theory of Infectious Diseases*, 2nd ed., Hafner, New York, 1975.
[2] H. W. HETHCOTE, *Qualitative analysis of communicable disease models*, Math. Biosci., 28 (1976), pp. 335–356.
[3] F. HOPPENSTEADT, *Mathematical Theories of Populations: Demographics, Genetics and Epidemics*, SIAM Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1975.
[4] V. LAKSHMIKANTHAM AND S. LEELA, *Differential and Integral Inequalities: Theory and Application*, vol. 1, Academic Press, New York, 1969.
[5] R. K. MILLER, *Nonlinear Volterra Integral Equations*, Benjamin, New York, 1971.
[6] P. WALTMAN, *Deterministic Threshold Models in the Theory of Epidemics*, Lecture Notes in Biomathematics, vol. 1, Springer-Verlag, New York, 1974.
[7] F. WANG, *Limit theorems for age and density dependent stochastic population models*, J. Math. Biology, 2 (1975), pp. 373–400.
[8] ———, *Gaussian approximation of some closed stochastic epidemic models*, J. Appl. Probability, 14 (1977), pp. 221–231.
[9] ———, *Asymptotic behavior of some deterministic epidemic models*, this Journal (1978), pp. 529–534.
[10] F. WANG AND W. DERRICK, *On deterministic epidemic model*, Bull. Inst. Math. Acad. Sinica, 6 (June, 1978), pp. 73–84.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0006$01.00/0

A NOTE ON THE ASYMPTOTIC BEHAVIOR OF SOLUTIONS OF THE KPP EQUATION*

H. J. K. MOET†

Abstract. This note is concerned with the convergence (as $t \to \infty$) to travelling waves of solutions u to the initial value problem of the KPP equation

$$u_t = u_{xx} + f(u), \qquad x \in \mathbb{R} \text{ and } t > 0.$$

A travelling wave ϕ_c is a solution of the form $u(x, t) = \phi_c(x + ct)$. Estimates for the difference between u and ϕ_c, in a moving coordinate system $\xi = x + ct$, are given in a weighted supremum norm and in weighted L^p-norm ($p \geqq 1$).

1. Introduction. This note is concerned with the asymptotic behavior as $t \to \infty$ of solutions $u(x, t)$ of the KPP equation

$$u_t = u_{xx} + f(u), \qquad x \in \mathbb{R} \text{ and } t > 0, \tag{1.1}$$

where f satisfies

$$\begin{aligned} &f: [0, 1] \to [0, \infty), \qquad f \in C^1([0, 1]);\\ &f(0) = f(1) = 0, \qquad f(u) > 0 \text{ in } (0, 1);\\ &f'(0) = \alpha > 0, \qquad f'(1) = -\beta < 0;\\ &f'(u) \leqq \alpha \text{ in } [0, 1]. \end{aligned} \tag{1.2}$$

In particular, we are interested in the convergence as $t \to \infty$ towards a travelling wave solution, i.e., a solution of the form $u(x, t) = \phi_c(x + ct)$ for some c (the velocity) with $\phi_c(-\infty) = 0$ and $\phi_c(+\infty) = 1$. It is a classical result of Kolomogoroff, Petrovsky and Piscounoff [6], whom we refer to as KPP, that with the above assumptions on f there exists a unique classical solution $u(x, t)$ of (1.1) departing at $t = 0$ from $u_0(x) \in [0, 1]$, if u_0 has at most finitely many points of discontinuity. They proved for each real $c \geqq 2\sqrt{\alpha}$ the existence of a strictly increasing travelling wave and, further, they showed that the solution starting from a step function converges to a shifted wave. The asymptotic behavior of solutions of (1.1), (1.2) has attracted an increasing amount of interest in recent years [1], [5], [8], [9] and the references therein.

Aronson and Weinberger [1] studied, among other things, the approach of solutions of (1.1), (1.2) towards the equilibrium solutions $u(x, t) \equiv 0$ and $u(x, t) \equiv 1$. For example, they proved that either $u(x, t) \equiv 0$ or $\lim_{t\to\infty} u(x, t) = 1$. Furthermore, they showed there is a number $c^* > 0$ with the property that if for some x_0 $u(x, 0) = 0$ in $(-\infty, x_0)$ then for each x and each $c > c^*$ $\lim_{t\to\infty} u(x + ct, t) = 0$.

Kametaka [5] considered the case in which $f \in C^\infty$ and the initial value u_0 belongs to $C^1(\mathbb{R})$ and is strictly increasing. Rothe [8] requires $f \in C^2$ and the initial value u_0 to be nondecreasing and continuously differentiable except for a finite number of jumps in which u_0 must be continuous from the right, moreover, 0 and 1 are the only values allowed to be assumed more than once. Both, Kametaka and Rothe, obtain uniform convergence to shifted travelling waves.

Sattinger [9] obtains his results on stability of travelling waves as a corollary of a stability theorem for nonlinear parabolic systems. Let us recall his formulation of the problem. A solution $u(x, t)$ of (1.1) which at $t = 0$ is a perturbation of a travelling wave

* Received by the editors February 10, 1977, and in revised form January 26, 1978.
† Mathematisch Instituut, Rijksuniversiteit Utrecht, Utrecht, the Netherlands.

ϕ_c, i.e., $u(x, 0)=\phi_c(x)+u_0(x)$, can be written in the form

$$u(x, t)=\phi_c(x+ct)+v(x, t; c),$$

whence, upon transformation to a moving coordinate frame $\xi = x+ct$, one obtains the following problem for v

$$\begin{aligned} v_t &= v_{\xi\xi}-cv_\xi+f(\phi_c+v)-f(\phi_c) && \xi\in\mathbb{R} \text{ and } t>0 \\ v(\xi, 0; c) &= u_0(\xi) && \xi\in\mathbb{R}. \end{aligned}$$

Let $w(\xi)=1+e^{-(c/2)\xi}$ act as a weight function on $\mathbb{R}$ and let $\mathscr{B}_{w,j}$ $(j=0, 1)$ be the Banach spaces of continuous bounded functions v on $\mathbb{R}$, corresponding to the weighted norms

$$\|v\|_{w,0}=\sup_{\xi\in\mathbb{R}} |w(\xi)v(\xi)|,$$

$$\|v\|_{w,1}=\|v\|_{w,0}+\left\|\frac{dv}{d\xi}\right\|_{w,0}.$$

Note that $\lim_{\xi\to\infty} w(\xi)v(\xi)$ must exist, since $\mathscr{B}_{w,1}$ has to be dense in $\mathscr{B}_{w,0}$. Then, assuming f is C^4, Sattinger finds for each $c>2\sqrt{\alpha}$,

$$\|v(\cdot, t; c)\|_{w,1}\leqq Ke^{-\omega t} \quad \text{as } t\to\infty,$$

for some K, ω positive (cf. [9] for details). Observe that there is no wave shift.

The main purpose of this note is to prove the following

THEOREM 1. *Every travelling wave ϕ_c, $c\geqq 2\sqrt{\alpha}$, of* (1.1), (1.2) *is asymptotically stable under perturbations u_0, having at most finitely many points of discontinuity, for which $0\leqq\phi_c(\xi)+u_0(\xi)\leqq 1$ and $e^{-(c/2)\xi}u_0(\xi)$ belongs to $L^p(\mathbb{R})$ for some $p\geqq 1$, i.e., the perturbation v satisfies, in the moving coordinate frame,*

$$|e^{-(c/2)\xi}v(\xi, t; c)|\leqq(4\pi t)^{-1/2}\, e^{-(c^2/4-\alpha)t}\left(\int_{-\infty}^{\infty} e^{-p(c/2)\xi}|u_0(\xi)|^p\, d\xi\right)^{1/p} \tag{1.3}$$

for all $(\xi, t)\in\mathbb{R}\times(0, \infty)$,

$$\left(\int_{\infty}^{\infty} e^{-q(c/2)\xi}|v(\xi, t; c)|^q\, d\xi\right)^{1/q}\leqq(4\pi t)^{-(1/2)(1/p-1/q)}\, e^{-(c^2/4-\alpha)t}\left(\int_{\infty}^{\infty} e^{-p(c/2)\xi}|u_0(\xi)|^p\, d\xi\right)^{1/p} \tag{1.4}$$

for any $q\geqq p\geqq 1$ and all $t>0$.

We observe that there are no monotonicity requirements on the initial values ϕ_c+u_0. Moreover, apart from $0\leqq\phi_c+u_0\leqq 1$ there are no restrictions on u_0 in neighborhoods of $+\infty$. The requirement $e^{-(c/2)\xi}u_0(\xi)\in L^p(\mathbb{R})$ for some $p\geqq 1$ restricts the behavior of u_0 in neighborhoods of $-\infty$ considerably. However, there is some compensation in the fact that we can treat the case $c=2\sqrt{\alpha}$. An interesting consequence of this theorem is the following. Let u be the solution of (1.1) which satisfies at $t=0$, for $\varepsilon>0$ arbitrarily small but fixed,

$$u(\xi, 0)=\begin{cases}\phi_c(\xi) & \text{if } \xi\in\{\xi\in\mathbb{R} \mid \phi_c(\xi)\leqq\varepsilon\}\\ H(\xi) & \text{elsewhere,}\end{cases}$$

i.e., $u(\xi, 0)$ is "almost" the Heaviside function. Then by our theorem the solution $u(\xi, t)$ converges uniformly to $\phi_c(\xi)$ on $(-\infty, A]$ for any $A\in\mathbb{R}$.

We conclude this section by mentioning (without proof) some results concerning the asymptotic behavior of solutions towards constant equilibrium solutions of the KPP equation. If the initial condition $u(x, 0)$ is in $L^p(\mathbb{R})$ for some $p \geqq 1$, then the solution $u(x, t)$ of (1.1) is in $L^p(\mathbb{R})$ for all $t > 0$. This implies uniform convergence on compact intervals in $\mathbb{R}$ to 1 is the best that can be expected. Sufficient conditions for uniform convergence to 1 are:

i) the intial value is bounded away from zero outside some compact interval in $\mathbb{R}$;

ii) the initial value u_0 satisfies $1 - u_0 \in L^p(\mathbb{R})$ for some $p \geqq 1$.

In the above cases a rate of convergence can be established. In (i) it is $O(e^{-(\beta-\varepsilon)t})$ as $t \to \infty$ for any $\varepsilon \in (0, \beta)$. In addition, if f' is Dini continuous at 1 [2], then the rate of convergence is $O(e^{-\beta t})$ as $t \to \infty$. In (ii) one has $\sup_{x \in \mathbb{R}} |1 - u(x, t)| = O(t^{-1/(2p)} e^{-(\beta-\varepsilon)t})$ as $t \to \infty$ for any $\varepsilon \in (0, \beta)$ and $\|1 - u(\cdot, t)\|_{L^p(\mathbb{R})} = O(t^{-(1/2)(1/p-1/q)} e^{-(\beta-\varepsilon)t})$ as $t \to \infty$ for each $q \geqq p \geqq 1$. Again, if f' is Dini continuous at 1 ε can be dropped. Similar results hold in the case of $(n > 1)$ space variables. We observe that as a consequence of (i) the solution $u(x, t)$ of (1.1), (1.2) starting from

$$u(x, 0) = \begin{cases} \phi_c(x) & \text{if } x \in \{x \in \mathbb{R} | \phi_c(x) \geqq \varepsilon\} \\ \varepsilon, & \text{elsewhere,} \end{cases}$$

for $\varepsilon > 0$ arbitrarily small but fixed and any $c \geqq 2\sqrt{\alpha}$, converges uniformly to 1 as $t \to \infty$. This proves that in order to obtain convergence towards a travelling wave a considerable restriction is necessary on the behavior of the initial perturbation in some neighborhood of $-\infty$.

2. Convergence to a traveling wave. To prove our theorem on the stability of travelling waves we need the undermentioned lemma on the asymptotic behavior of solutions of the Cauchy problem for the heat equation.

Consider

$$u_t = u_{xx}, \qquad x \in \mathbb{R} \text{ and } t > 0, \tag{2.1}$$

$$u(x, 0) = u_0(x), \qquad x \in \mathbb{R}, \tag{2.2}$$

where u_0 is a bounded continuous function having at most a finite number of discontinuity points, which belongs to $L^p(\mathbb{R})$ for some $p \geqq 1$. As is clearly seen, the function u given by

$$u(x, t) = \int_{\infty}^{\infty} E(x - y, t) u_0(y)\, dy, \tag{2.3}$$

where $E(x, t) = (4\pi t)^{-1/2} \exp(-x^2/(4t))$, solves this problem.

LEMMA 1. *Let u be the solution of the initial value problem* (2.1), (2.2) *given by* (2.3); *then u satisfies the estimates*

$$|u(x, t)| \leqq (4\pi t)^{-1/2} \|u_0\|_{L^p(\mathbb{R})} \quad \text{for all } (x, t) \in \mathbb{R} \times (0, \infty), \tag{2.4}$$

and

$$\|u(\cdot, t)\|_{L^q(\mathbb{R})} \leqq (4\pi t)^{-(1/2)(1/p-1/q)} \|u_0\|_{L^p(\mathbb{R})} \quad \text{for all } t > 0 \text{ and any } q \geqq p \geqq 1. \tag{2.5}$$

Proof. The proof of (2.4) is trivial. For (2.5) use Hölder's inequality.

Remark. Results similar to (2.4), (2.5) can be established for uniformly parabolic equations in $\mathbb{R}^n \times (0, \infty)$ having a fundamental solution which satisfies a certain estimate (see [3, Chap. 9, § 4]). This result seems to have been overlooked previously.

Proof of Theorem 1. The perturbation v satisfies

$$v_t = v_{xx} + f(\phi_c + v) - f(\phi_c), \qquad x \in \mathbb{R} \text{ and } t > 0,$$
$$v(x, 0) = u_0(x) \qquad x \in \mathbb{R}.$$

We will obtain the desired result by comparing the function v with solutions to other suitable initial value problems.

First, consider the solution p of

$$p_t = p_{xx} + f(\phi_c + p) - f(\phi_c), \qquad x \in \mathbb{R} \text{ and } t > 0,$$
$$p(x, 0) = \max\{u_0(x), 0\}, \qquad x \in \mathbb{R}.$$

Then by theorem KPP 3 (Theorem 3 of [6]) we know p is nonnegative and

$$v(x, t) \leqq p(x, t) \quad \text{for all } (x, t) \in \mathbb{R} \times (0, \infty). \tag{2.6}$$

Secondly, consider the solution q of

$$q_t = q_{xx} + f(\phi_c + q) - f(\phi_c), \qquad x \in \mathbb{R} \text{ and } t > 0,$$
$$q(x, 0) = \min\{u_0(x), 0\}, \qquad x \in \mathbb{R}.$$

Then by Theorem KPP 3 q is nonpositive and

$$q(x, t) \leqq v(x, t) \quad \text{for all } (x, t) \in R \times (0, \infty). \tag{2.7}$$

Since $f'(u) \leqq \alpha$ in $[0, 1]$ and $p(q)$ is nonnegative (nonpositive) we find

$$f(\phi_c + p) - f(\phi_c) \leqq \alpha p, \qquad (f(\phi_c + q) - f(\phi_c) \geqq \alpha q). \tag{2.8}$$

Now let r be the solution of

$$r_t = r_{xx} + \alpha r, \qquad x \in \mathbb{R} \text{ and } t > 0 \tag{2.9}$$

$$r(x, 0) = |u_0(x)|, \qquad x \in \mathbb{R}; \tag{2.10}$$

then using (2.6), (2.7), (2.8) in conjunction with theorems KPP 3, KPP 2 it is easily seen that

$$|v(x, t)| \leqq r(x, t) \quad \text{for all } (x, t) \in \mathbb{R} \times (0, \infty).$$

Transforming (2.9), (2.10) to a moving coordinate frame $\xi = x + ct$ yields

$$r_t = r_{\xi\xi} - cr_\xi + \alpha r, \qquad r(\xi, 0) = |u_0(\xi)|,$$

and after substituting $w(\xi, t) = e^{-(c/2)\xi} r(\xi, t)$

$$w_t = w_{\xi\xi} - (c^2/4 - \alpha)w, \qquad \xi \in \mathbb{R} \text{ and } t > 0,$$
$$w(\xi, 0) = e^{-(c/2)\xi} |u_0(\xi)|, \qquad \xi \in \mathbb{R},$$

from which (1.3), (1.4) easily follow.

Remark. With f given as in (1.2) KPP were able to prove the existence of a half-line of velocities $[2\sqrt{\alpha}, \infty)$ with minimal velocity $c_0 = 2\sqrt{\alpha}$. If f' does not assume its maximum value at $u = 0$, then, although a half-line of velocities exists, the minimal velocity need not be equal to $2\sqrt{\alpha}$. Hadeler and Rothe [4] show that in this case $c_0 \in [2\sqrt{\alpha}, 2\sqrt{\sigma}]$, where $\sigma = \sup\{f(u)/u \mid u \in (0, 1)\}$. It is easily seen that the proof of Theorem 1 applies to waves ϕ_c with $c \geqq 2\sqrt{\gamma}$, where $\gamma = \max\{f'(u) \mid u \in [0, 1]\}$.

Acknowledgment. The method of analysis used in this paper is based on our earlier work [7] on the asymptotic behavior of solutions of the KPP equation, done under supervision of Prof. L. A. Peletier (Technische Hogeschool Delft). The author is greatly indebted to Professor Peletier for numerous valuable and stimulating discussions with regard to this work. Finally, the author would like to thank the referee for his valuable criticism, concerning a previous version of this note, which stimulated the author to improve the contents substantially.

REFERENCES

[1] D. G. ARONSON AND H. F. WEINBERGER, *Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation*, J. A. Goldstein, ed., Partial Differential Equations and Related Topics, Lecture Notes in Math. 446, Springer-Verlag, Berlin, 1975, pp. 5–49.

[2] J. P. G. EWER AND L. A. PELETIER, *On the asymptotic behavior of solutions of semilinear parabolic equations*, SIAM J. Appl. Math., 28 (1975), pp. 43–53.

[3] A. FRIEDMAN, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, NJ, 1964.

[4] K. P. HADELER AND F. ROTHE, *Travelling fronts in nonlinear diffusion equations*, J. Math. Biol., 2 (1975), pp. 251–263.

[5] Y. KAMETAKA, *On the nonlinear diffusion equation of Kolmogorov–Petrovski–Piskounov Type*, H. Araki, ed., International Symposium on Mathematical Problems in Theoretical Physics, Lecture Notes in Physics 39, Springer-Verlag, Berlin, 1975, pp. 401–407.

[6] A. KOLMOGOROFF, I. PETROVSKY AND N. PISCOUNOFF, *Etude de l'équation de la diffusion avec croissance de la quantitité de matière et son application à un prloblème biologique*, Bull. Univ. Moscou Sér. Internationale Sec. A1, (1937), pp. 1–25.

[7] H. J. K. MOET, *Stability and asymptotic behavior of solutions of a semilinear parabolic partial differential equation*, Master's Thesis, Technische Hogeschool Delft, 1975. (In Dutch.)

[8] F. ROTHE, *Convergence to travelling fronts in semilinear parabolic equations*, Proc. Roy. Soc. Edinburgh, to appear.

[9] D. H. SATTINGER, *On the stability of waves of nonlinear parabolic systems*, Advances in Math., 22 (1976), pp. 312–355.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0007$01.00/0

LIMITS IN L_p OF CONVOLUTION TRANSFORMS WITH KERNELS $aK(at)$, $a \to 0$*

B. F. LOGAN†

Abstract. For f in L_p and K in L_1 define

$$K_a(t) = aK(at), \qquad a > 0,$$

$$f_a(x) = \int_{-\infty}^{\infty} f(t) K_a(x-t)\,dt.$$

Then it is shown that

$$\lim_{a\to 0} \|f_a\|_p = 0 \quad \text{provided } 1 < p < \infty.$$

For the case $p = 1$ the result is

$$f_a(x) = \left\{ \int_{-\infty}^{\infty} f(t)\,dt \right\} \cdot K_a(x) + \varepsilon_a(x)$$

where

$$\lim_{a\to 0} \|\varepsilon_a\|_1 = 0.$$

For functions $f(z)$ analytic in the upper half-plane and uniformly bounded in L_p norm on lines parallel to the real axis (the Hardy class H_p of the upper half-plane) it follows that

$$\lim_{y\to\infty} \left\{ \int_{-\infty}^{\infty} |f(x+iy)|^p\,dx \right\}^{1/p} = 0, \qquad (1 \leqq p < \infty).$$

A function K in L_1 may be taken as the kernel of a convolution transform on L_p; $1 \leqq p \leqq \infty$,

(1) $$\hat{f}(x) = \int_{-\infty}^{\infty} f(t) K(x-t)\,dt, \qquad f \in L_p.$$

Then $\hat{f}$, the transform of f, also belongs to L_p with

(2) $$\|\hat{f}\|_p \leqq \|K\|_1 \cdot \|f\|_p.$$

Often one is interested in the behavior of the parameterized transform

(3) $$f_a(x) = \int_{-\infty}^{\infty} f(t) K_a(x-t)\,dt$$

where

(4) $$K_a(t) = aK(at), \qquad a > 0$$

and $a \to 0$ (narrow-band filtering). In particular, for functions $f(z)$ analytic in the upper half-plane, uniformly bounded in L_p norm on lines parallel to the real axis (the Hardy class H_p of the upper half-plane) we have (Hoffman [1, p. 128])

(5) $$f(x+iy) = \int_{-\infty}^{\infty} f(t) P_y(x-t)\,dt$$

* Received by the editors April 15, 1977.
† Bell Laboratories, Murray Hill, New Jersey 07974.

where $P_y(t)$ is the Poisson kernel

$$P_y(t)=\frac{1}{\pi}\frac{y}{t^2+y^2}=\frac{1}{y}P_1\left(\frac{t}{y}\right). \tag{6}$$

It is easy to show from (5) and (6) that for f in H_p, $1\leqq p<\infty$,

$$\lim_{x\to\pm\infty} f(x+iy)=0, \quad \text{(fixed positive } y) \tag{7a}$$

$$\lim_{y\to\infty} f(x+iy)=0, \quad \text{(fixed real } x) \tag{7b}$$

$$\|f(x+iy)\|_p=\left\{\int_{-\infty}^{\infty}|f(x+iy)|^p\,dx\right\}^{1/p} \text{ is a decreasing function of } y. \tag{7c}$$

However, judging from the literature, the useful result

$$\lim_{y\to\infty}\|f(x+iy)\|_p=0 \quad \text{for } f \text{ in } H_p \quad (1\leqq p<\infty) \tag{8}$$

is apparently not generally known. Of course (7) does not imply (8). Certainly (8) is not a surprising result (obvious for $p=2$) since the Fourier transform of $P_y(t)$ tends to zero everywhere except at the origin; i.e.,

$$\int_{-\infty}^{\infty} P_y(t)\,e^{i\omega t}\,dt=e^{-y|\omega|}, \qquad y>0 \quad (-\infty<\omega<\infty).$$

However, the unsurprising result, as is often the case, is surprisingly difficult (well, say tedious) to prove, given the state of Fourier transform theory for L_p. So for the convenience of future reference we prove the more general unsurprising result:

THEOREM. *Let K belong to L_1 and f to L_p for some p satisfying $1\leqq p<\infty$. Then with f_a defined by* (3) *and* (4) *we have*

$$\lim_{a\to 0}\|f_a\|_p=0 \quad \textit{provided } 1<p<\infty. \tag{9}$$

In case $p=1$ we have

$$f_a(x)=\left\{\int_{-\infty}^{\infty} f(t)\,dt\right\}\cdot K_a(x)+\varepsilon_a(x) \tag{10}$$

where

$$\lim_{a\to 0}\|\varepsilon_a\|_1=0. \tag{11}$$

The result (8) follows by setting

$$K_a(t)=P_y(t), \qquad y=\frac{1}{a}$$

and noting that for f in H_1

$$\int_{-\infty}^{\infty} f(t)\,dt=0$$

since we can show, using the analyticity of $f(z)$ and the uniform bound on the norm of

$f(x+iy)$, that

$$F(\omega)=\int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt=0 \quad \text{for } \omega<0$$

and since f belongs to L_1, $F(\omega)$ is continuous; i.e., $F(0)=0$.

Actually (9) is valid for special K not in L_1. For example it has been shown, (Logan [2]) that (9) holds for

$$K(t)=\frac{\sin t}{\pi t}$$

and hence for finite sums

$$K(t)=\sum_1^n a_k \frac{\sin \lambda_k t}{\pi t}$$

and then for functions of the last form convolved with a function of L_1. That such kernels carry L_p into L_p $(1<p<\infty)$ is a simple consequence of the fact that the Hilbert transform of a function of L_p is also a function of L_p, provided $1<p<\infty$. It would be very desirable to find a general characterization of convolution kernels which carry L_p into L_p for all p satisfying $1<p<\infty$. One could then probably show that (9) holds for such kernels.

The reason that (9) is a bit harder to establish than (8) is the fact that the Poisson kernel belongs to $L_1 \cap L_\infty$ and hence to L_p. Our proof of (9) requires approximating K with a kernel k in $L_1 \cap L_p$. There are a lot of ways to do this. We may take

$$k(x;\lambda)=\lambda \int_{-\infty}^{\infty} s(\lambda t) K(x-t)\, dt, \qquad \lambda>0, \tag{12}$$

where

$$s(t)=\left\{\frac{\sin \pi t}{\pi t}\right\}^2$$

$$\int_{-\infty}^{\infty} s(t)\, dt=\int_{-\infty}^{\infty} |s(t)|\, dt=1. \tag{13}$$

So we have

$$\|k\|_1 \leqq \|K\|_1 \tag{14}$$

and

$$\lim_{\lambda\to\infty} \|K-k\|_1=0. \tag{15}$$

The last result is established by the standard "approximate identity" argument (see Appendix).

Since $s(t)$ belongs to L_p for every $1\leqq p\leqq\infty$ we have from (2) and (12)

$$\|k\|_p \leqq \|\lambda s(\lambda t)\|_p \cdot \|K\|_1$$

$$\|\lambda s(\lambda t)\|_p=\lambda\left\{\int_{-\infty}^{\infty} |s(\lambda t)|^p\, dt\right\}^{1/p}=\lambda^{(1-1/p)}\left\{\int_{-\infty}^{\infty} |s(t)|^p\, dt\right\}^{1/p}$$

and since $|s(t)|\leqq 1$, we have for $p\geqq 1$,

$$\int_{-\infty}^{\infty} |s(t)|^p\, dt \leqq \int_{-\infty}^{\infty} |s(t)|\, dt=1.$$

Therefore

(16) $$\|k\|_p \leqq \lambda^r \|K\|_1, \qquad r = 1 - 1/p.$$

Now we define

(17) $$k_a(t) = ak(at; \lambda)$$

and

(18) $$\phi_a(x) = \int_{-\infty}^{\infty} f(t) k_a(x-t)\, dt.$$

Then

(19) $$f_a(x) - \phi_a(x) = \int_{-\infty}^{\infty} f(t)\{K_a(x-t) - k_a(x-t)\}\, dt$$

and thus

(20) $$\|f_a - \phi_a\|_p \leqq \|K_a - k_a\|_1 \cdot \|f\|_p = \|K - k\|_1 \cdot \|f\|_p.$$

Now we want to let $\lambda \to \infty$ as $a \to 0$ in such a way that $\|\phi_a\|_p \to 0$. Then since

$$\|K - k\|_1 \to 0 \quad \text{as } \lambda \to \infty$$

we will have

$$\begin{aligned} \|f_a\|_p = \|f_a - \phi_a + \phi_a\|_p &\leqq \|f_a - \phi_a\|_p + \|\phi_a\|_p \\ &\to 0 \quad \text{as } a \to 0. \end{aligned}$$

To obtain this result we first write for any $T > 0$

(21) $$f(t) = g(t; T) + h(t; T)$$

where

$$\begin{aligned} g(t, T) &= 0, \qquad |t| \geqq T, \\ h(t; T) &= 0, \qquad |t| < T. \end{aligned}$$

That is, g is simply f truncated to the interval $(-T, T)$ and h represents the "tails" of f. We have

(22) $$\|g\|_p \leqq \|f\|_p$$

(23) $$\lim_{T \to \infty} \|f - g\|_p = 0 \qquad (1 \leqq p < \infty)$$

(24) $$\lim_{T \to \infty} \|h\|_p = 0 \qquad (1 \leqq p < \infty).$$

Now we want to use the fact that g also belongs to L_1 although its norm in L_1 may tend to ∞ as $T \to \infty$. We have from Hölder's inequality

(25) $$\int_{-\infty}^{\infty} |g(t)|\, dt = \int_{-\infty}^{\infty} |g(t)| \cdot G(t)\, dt \leqq (2T)^r \|g\|_p$$

where

$$r = 1 - \frac{1}{p} = \frac{1}{q}, \qquad G(t) = \begin{cases} 1, & |t| < T, \\ 0, & |t| \geqq T \end{cases}$$

and

$$(2T)^r = \|G\|_q.$$

Next we define

$$g_a(x; T, \lambda) = \int_{-\infty}^{\infty} k_a(t) g(x-t)\, dt \tag{26}$$

and think of k_a in L_p and g in L_1, interchanging the roles of the function g ($\sim f$) and the kernel k_a ($\sim K_a$). We have

$$\begin{aligned} \|k_a\|_p &= \left\{ \int_{-\infty}^{\infty} |ak(at)|^p\, dt \right\}^{1/p} \\ &= a^r \left\{ \int_{-\infty}^{\infty} |k(t)|^p \right\}^{1/p} = a^r \|k\|_p \end{aligned} \tag{27}$$

where $r = 1 - 1/p$. Thus we have, using (2), (16), (25), and (27),

$$\begin{aligned} \|g_a\|_p &\leqq \|k_a\|_p \cdot \|g_1\| = a^r \|k\|_p \cdot \|g\|_1 \\ &\leqq (2a\lambda T)^r \|K\|_1 \cdot \|h\|_p; \qquad r = 1 - \frac{1}{p}. \end{aligned} \tag{28}$$

Now

$$\phi_a(x) = g_a(x) + h_a(x) \tag{29}$$

where

$$h_a = h \otimes k_a$$

and since $\|k_a\|_1 = \|k\|_1 \leqq \|K\|_1$ we have

$$\begin{aligned} \|\phi_a\|_p &\leqq \|g_a\|_p + \|k_a\|_1 \cdot \|h\|_p \\ &\leqq \{(2a\lambda T)^r \cdot \|f\|_p + \|h\|_p\} \cdot \|K\|_1 \end{aligned} \tag{30}$$

where $r = 1 - 1/p$.

Now we may, for example, set

$$\lambda = a^{-\nu}, \qquad T = a^{-\nu} \qquad (0 < \nu < \tfrac{1}{2})$$

so that

$$\lim_{a \to 0} (2a\lambda T)^r = 0 \quad \text{for } p > 1 \tag{31}$$

and since

$$\lim_{T \to \infty} \|h\|_p = 0 \quad \text{for } 1 \leqq p < \infty$$

we have

$$\lim_{a \to 0} \|\phi_a\|_p = \lim_{a \to 0} \|f_a\|_p = 0 \quad \text{for } 1 < p < \infty. \tag{32}$$

Thus we have established (9). It should be noted that this result (or proof) ultimately depends on the property of the L_p norm, which offhand seems rather freakish: if $k(t)$ is a bounded function of L_1 then $ak(at)$ tends to zero in L_p for every $p > 1$ as $a \to 0$ but not for $p = 1$ (cf. (27)).

Also we note that if f belongs to L_∞ and $\lim_{t\to\pm\infty} f(t)=0$ then $\|h\|_\infty \to 0$ as $T\to\infty$ so that in this special case $\lim_{a\to 0}\|f_a\|_\infty = 0$.

The conclusion of the theorem for the case $p=1$ is easily established. For f in L_1 and

$$f_a(x)=\int_{-\infty}^{\infty} aK(at)f(x-t)\,dt, \qquad a>0$$

we define

$$F(x;a)=\frac{1}{a}f_a\left(\frac{x}{a}\right)=\int_{-\infty}^{\infty}\frac{1}{a}f\left(\frac{t}{a}\right)K(x-t)\,dt \tag{33}$$

and then write

$$F(x;a)-cK(x)=\int_{-\infty}^{\infty}\frac{1}{a}\,f\left(\frac{t}{a}\right)\{K(x-t)-K(x)\}\,dt \tag{34}$$

where

$$c=\int_{-\infty}^{\infty}\frac{1}{a}f\left(\frac{t}{a}\right)dt=\int_{-\infty}^{\infty} f(t)\,dt.$$

Then we use the "approximate identity" argument (Appendix) to conclude that

$$\lim_{a\to 0}\|F-cK\|_1=0. \tag{35}$$

Thus

$$\begin{aligned} f_a(x)&=aF(ax;a)=c\{aK(ax)\}+\varepsilon_a(x)\\ &=cK_a(x)+\varepsilon_a(x)\end{aligned} \tag{36}$$

where

$$\begin{aligned}\int_{-\infty}^{\infty}|\varepsilon_a(x)|\,dx&=\int_{-\infty}^{\infty}|aF(ax;a)-cK_a(x)|\,dx\\ &=\int_{-\infty}^{\infty}|F(x;a)-cK(x)|\,dx\end{aligned}$$

and from (35)

$$\lim_{a\to 0}\|\varepsilon_a\|_1=0. \tag{37}$$

Appendix. A convolution kernel K in L_1 satisfying

$$\int_{-\infty}^{\infty}K(t)\,dt=1 \tag{A.1}$$

is said to be an "approximate identity" (Hoffmann [1, p. 16]) for L_p in the sense that f_λ defined by

$$f_\lambda(x)=\lambda\int_{-\infty}^{\infty}K(\lambda t)f(x-t)\,dt, \qquad \lambda>0, \quad f \text{ in } L_p, \tag{A.2}$$

tends to f in L_p norm as $\lambda\to\infty$ $(1\leqq p<\infty)$.

Actually, Hoffmann imposes additional constraints on K which we do not require to obtain convergence in L_p. One should recognize that a convolution transform is to be regarded as mapping functions into functions or, here, elements of L_p into elements of L_p. The integral in (A2) does not generally make sense pointwise unless K also belongs to the complementary space L_q, $1/p+1/q=1$. Strictly speaking the convolution transform on L_p should be interpreted as the limit in L_p of convolution transforms defined by a sequence of kernels K_n in L_q (say bounded kernels of compact support) tending to K in L_1. We omit this intermediate step required for full rigor in the following argument (which is also required to rigorously establish that K carries L_p into L_p).

We assume that K belongs to L_1 and replace (A.1) by

$$\int_{-\infty}^{\infty} K(t)\,dt = c \tag{A.3}$$

since we may be interested in the case $c=0$. Then we write

$$f_\lambda(x)-cf(x)=\lambda\int_{-\infty}^{\infty} K(\lambda t)\{f(x-t)-f(x)\}\,dt=\varepsilon_\lambda(x) \tag{A.4}$$

$$|\varepsilon_\lambda(x)|\leqq\lambda\int_{-\infty}^{\infty}|K(\lambda t)|\cdot|f(x-t)-f(x)|\,dt. \tag{A.5}$$

Now ε_λ belongs to L_p and its norm is equivalently defined by

$$\|\varepsilon_\lambda\|_p=\sup_g\left|\int_{-\infty}^{\infty}\varepsilon_\lambda(x)g(x)\,dx\right|=\sup_g\int_{-\infty}^{\infty}|\varepsilon_\lambda(x)|\,|g(x)|\,dx \tag{A.6}$$

where the supremum is taken over functions g of norm 1 in the complementary space L_q, $q=p/(p-1)$. We then have from (A.6) and (A.5)

$$\|\varepsilon_\lambda\|_p\leqq\lambda\int_{-\infty}^{\infty}|K(\lambda t)|\cdot\mu_p(t)\,dt \tag{A.7}$$

where μ_p is the L_p modulus of continuity of f,

$$\mu_p(t)=\mu_p(f;t)=\left\{\int_{-\infty}^{\infty}|f(x-t)-f(x)|^p\,dx\right\}^{1/p}. \tag{A.8}$$

For f in L_p for some p satisfying $1\leqq p<\infty$, $\mu_p(t)$ is an even continuous function of t (see Achieser [3, pp. 162–163]) with

$$\mu_p(0)=0, \tag{A.9}$$

$$0\leqq\mu_p(t)\leqq2\|f\|_p. \tag{A.10}$$

Then

$$\begin{aligned}\|\varepsilon_\lambda\|_p&\leqq M(T)\int_{-T}^{T}\lambda|K(\lambda t)|\,dt+2\|f\|_p\cdot\int_{|t|>T}\lambda|K(\lambda t)|\,dt\\&\leqq M(T)\int_{-\lambda T}^{\lambda T}|K(t)|\,dt+2\|f\|_p\int_{|t|>\lambda T}|K(t)|\,dt\end{aligned} \tag{A.11}$$

where

$$M(T)=\sup_{-T<t<T}\mu_p(t). \tag{A.12}$$

Since $\mu_p(t)$ is continuous and $\mu_p(0)=0$,

$$\lim_{T\to 0} M(T)=0. \tag{A.13}$$

So if we take, for example,

$$T=\lambda^{-1/2} \tag{A.14}$$

we have

$$\lim_{\lambda\to\infty} \|\varepsilon_\lambda\|_p=0 \tag{A.15}$$

i.e.,

$$f_\lambda(x)=cf(x)+\varepsilon_\lambda(x) \tag{A.16}$$

and

$$\lim_{\lambda\to\infty} \|f_\lambda-cf\|_p=0 \qquad (1\leqq p<\infty). \tag{A.17}$$

REFERENCES

[1] KENNETH HOFFMAN, *Banach Spaces of Analytic Functions*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
[2] B. F. LOGAN, *The Convolution Kernel* (sin $at)/\pi t$ *as an Approximate Identity for* L_p *as* $a\to\infty$, Bell Telephone Laboratories Tech. Mem. Jan. 31, 1973, unpublished.
[3] N. I. ACHIESER, *Theory of Approximation*, Frederick Ungar, New York, 1956.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0008$01.00/0

INEQUALITIES AND MINIMUM NORM KERNELS FOR THE HARDY CLASS H_p*

B. F. LOGAN†

Abstract. Functions $f(z)$ analytic in the upper half-plane and uniformly bounded in L_p norm on lines parallel to the real axis are said to belong to the Hardy class H_p of the upper half plane. It is shown that

$$f(x+iy) \leqq A_p y^{-1/p} \left\{ \int_{-\infty}^{\infty} |f(t)|^p \, dt \right\}^{1/p}, \qquad f \text{ in } H_p, \quad p \geqq 1,$$

holds for

$$A_p = (4\pi)^{-1/p}$$

with equality possible for $x = 0$, $y = b > 0$, if and only if

$$f(z) = C(z+ib)^{-2/p}.$$

The best constant A_p was previously known only for $p = 2, \infty$.

The inequality is obtained by replacing the Poisson kernel $P_y(t)$ in the representation

$$f(x+iy) = \int_{\infty}^{\infty} f(t) P_y(x-t) \, dt, \qquad f \text{ in } H_p$$

by a kernel $K_y(t; p)$ of minimum L_q norm, $q = p/(p-1)$,

$$K_y(t;p) = \frac{i}{2\pi} \frac{\left(\frac{1}{2} + \frac{it}{2y}\right)^{\mu}}{t+iy} \quad \text{where } \mu = \frac{2}{p} - 1.$$

The Fourier transforms of the kernels $K_y(t; p)$, $1 < p \leqq \infty$ are given in terms of the incomplete gamma function. Also it is concluded that

$$|f(x+iy)| = o\{y^{-1/p}\}, \quad y \to \infty, \quad f \text{ in } H_p \quad (1 \leqq p < \infty).$$

Functions $f(z)$ analytic in the upper half-plane and uniformly bounded in L_p norm (for some p satisfying $1 \leqq p \leqq \infty$) on lines parallel to the real axis are said to belong to the Hardy class H_p (of the upper half-plane). This class of functions arises naturally in Fourier theory, since the Fourier transforms of these functions can be said, in a meaningful sense (cf. Logan [1]) to vanish over $(-\infty, 0)$ whether or not (i.e., for $p > 2$) their Fourier transforms are defined in the ordinary sense. Functions $g(z)$ analytic and of exponential type λ in the upper half-plane, belonging to L_p on the real axis, may be brought into H_p by the simple transformation $f(z) = e^{i\lambda z} g(z)$. H_p is just the special case $\lambda = 0$, which is, owing to the norm-multiplying property of the exponential function, a sufficiently general class to consider of functions of exponential type λ, analytic in the upper half-plane and belonging to L_p on the real axis. We should note that functions analytic in the upper half-plane and belonging to L_p on the real axis do not necessarily belong to L_p on any other line parallel to the real axis, but for functions $f(z)$ of *exponential type* λ we do have (Boas [2, 6.7.7, p. 98])

$$\left\{ \int_{-\infty}^{\infty} |f(x+iy)|^p \, dx \right\}^{1/p} = \|f_y\|_p \leqq e^{\lambda y} \|f\|_p, \qquad y > 0.$$

Hence follows the above correspondence with H_p.

* Received by the editors April 15, 1977.

† Bell Laboratories, Murray Hill, New Jersey 07974.

We will understand here and in the following that $\|f\|_p$ is the norm of f on the real axis, i.e.

$$\|f\|_p=\left\{\int_{-\infty}^{\infty}|f(t)|^p\,dt\right\}^{1/p},\qquad 1\leqq p<\infty,$$

$$\|f\|_\infty=\text{ess sup}\,|f|=\sup_g\left|\int_{-\infty}^{\infty}f(t)g(t)\,dt\right|$$

where the supremum is over functions g of L_1, $\|g\|_1=1$.

For f in H_p $(1\leqq p\leqq\infty)$ we have (Hoffman [3, p. 128])

$$f(x+iy)=\int_{-\infty}^{\infty}f(t)P_y(x-t)\,dt \tag{1}$$

where $P_y(t)$ is the Poisson kernel

$$P_y(t)=\frac{1}{\pi}\frac{y}{t^2+y^2}. \tag{2}$$

We have, applying Hölder's inequality to (1),

$$|f(x+iy)|\leqq\|f\|_p\cdot\|P_y\|_q=\frac{M_p}{y^{1/p}}\cdot\|f\|_p \tag{3}$$

where

$$\begin{aligned} M_p&=\frac{1}{\pi}\left\{\int_{-\infty}^{\infty}\frac{dt}{(1+t^2)^q}\right\}^{1/q}\\ &=\frac{1}{\pi}\{B(\tfrac{1}{2},q-\tfrac{1}{2})\}^{1/q}\qquad\text{(Beta function)}\\ &=\left(\frac{1}{4\pi}\right)^{1/p}\left\{\frac{\Gamma(2q-1)}{\Gamma^2(q)}\right\}^{1/q},\qquad q=p/p-1,\end{aligned} \tag{3a}$$

$$M_1=\frac{1}{\pi},\qquad M_2=\frac{1}{\sqrt{2\pi}},\qquad M_\infty=1.$$

Actually, the Poisson kernel does not allow equality in (3) except for the case $p=\infty$, $f(t)=\text{constant}$. Here we wish to determine the smallest number A_p such that

$$|f(x+iy)|\leqq\frac{A_p}{y^{1/p}}\|f\|_p,\qquad f\text{ in }H_p,\quad(y>0). \tag{4}$$

That $A(p,y)$ in the best inequality

$$|f(x+iy)|\leqq A(p,y)\|f\|_p$$

is of the form

$$A(p,y)=\frac{A_p}{y^{1/p}}$$

where A_p depends only on p follows from a simple change of variable argument. That is, it is sufficient to determine the best inequality for a fixed value of y, say $y=1$, and a fixed value of x, say $x=0$ (since any translate of f also belongs to H_p):

$$|f(i)|\leqq A_p\|f\|_p,\qquad f\text{ in }H_p.$$

Then if we set

$$\varphi(z)=f(bz), \qquad b>0,$$

we have φ in H_p and

$$\|\varphi\|_p=\left\{\int_{-\infty}^{\infty} | f(bt)|^p\, dt\right\}^{1/p}=b^{-1/p}\|f\|_p.$$

Then

$$|\varphi(i)|\leqq A_p\|\varphi\|_p$$

and hence

$$|f(ib)|\leqq \frac{A_p}{b^{1/p}}\|f\|_p, \qquad f \text{ in } H_p.$$

In order to determine A_p in (4) we have to use more general representations:

(5) $$f(x+iy)=\int_{-\infty}^{\infty} f(t)K_y(x-t)\,dt, \qquad f \text{ in } H_p$$

which we may derive by analytic function methods (contour integrals) or equivalently, since the Fourier transform of $f(t)$ vanishes over $(-\infty, 0)$, in the sense detailed in [1], by replacing $P_y(t)$ by a function $K_y(t)$ whose Fourier transform agrees (sense of [1]) with that of $P_y(t)$ over $(0, \infty)$, or formally

(6) $$\int_{-\infty}^{\infty} K_y(t)\, e^{-i\omega t}\, dt=e^{-\omega y}, \qquad y>0, \quad \omega>0.$$

The integral may not make sense, but formally the problem is that of extrapolating the Fourier transform (for a given fixed y) to $(-\infty, 0)$ so that $K_y(t; p)$ has minimum norm in the complementary space L_q, $q=p/(p-1)$. Of course the case $p=q=2$ is simple; the Fourier transform of $K_y(t; 2)$ must vanish for negative arguments giving

(7) $$A_2=\frac{1}{2\sqrt{\pi}} \qquad \left(\text{cf. } M_2=\frac{1}{\sqrt{2\pi}}\right).$$

The corresponding kernel is called the Cauchy kernel[1]

(8) $$K_y(t; 2)=C_y(t)=\frac{1}{2\pi}\,\frac{1}{t+iy}$$

and in fact we have, not only for H_2, but

(9) $$f(x+iy)=\int_{-\infty}^{\infty} f(t)C_y(x-t)\,dt \quad \text{for } f \text{ in } H_p \quad (y>0)$$
$$1\leqq p<\infty.$$

(We will establish a more general result later.)

Thus applying Hölder's inequality to (9) we obtain

(10) $$|f(x+iy)|\leqq \frac{C_p}{y^{1/p}}\|f\|_p, \qquad 1\leqq p<\infty,$$

[1] Sometimes $(1/(2\pi i))(1/(t-iy))$ or $(\pm i/\pi)(1/(t\pm iy))$ is called the Cauchy kernel.

where

$$C_p = \frac{1}{2\pi}\left\{\int_{-\infty}^{\infty} \frac{dt}{(1+t^2)^{q/2}}\right\}^{1/q}$$

(11)
$$= \frac{1}{2\pi}\left\{B\left(\frac{1}{2}, \frac{q-1}{2}\right)\right\}^{1/q} = \frac{1}{(4\pi)^{1/p}}\left\{\frac{\Gamma(q-1)}{\Gamma^2(q/2)}\right\}^{1/q}$$

$$C_1 = \frac{1}{2\pi}, \qquad C_2 = \frac{1}{2\sqrt{\pi}}.$$

So the Cauchy kernel gives a better inequality than the Poisson kernel for $p \leqq p_0$ for some p_0 satisfying $2 < p_0 < \infty$ ($p_0 \sim 3.2$). Now we would like to prove the following:

THEOREM. *Let f belong to H_p of the upper half-plane, where p satisfies $1 \leqq p < \infty$. Then*:

(12)
$$f(x+iy) = \int_{-\infty}^{\infty} f(t) K_y(x-t; p)\, dt, \qquad y > 0$$

where

(13)
$$K_y(t; p) = \frac{i}{2\pi} \frac{\left(\frac{1}{2} + \frac{it}{2y}\right)^{2/p-1}}{t+iy}$$

and

(14)
$$|f(x+iy)| \leqq \frac{1}{(4\pi y)^{1/p}} \|f\|_p, \qquad y > 0,$$

with equality possible in (14) *for $y = b$, $c = 0$, if and only if*

(15)
$$f(z) = \frac{A}{(z+ib)^{2/p}} \qquad (b > 0).$$

Furthermore, in (5) *we have*

(16)
$$\|K_y\|_q \geqq \frac{1}{(4\pi y)^{1/p}}, \qquad q = p/(p-1),$$

with equality holding only for $K_y(t)$ given by (13).

Actually the theorem is simple to prove, given the conclusions. In solving the extremal problem one has to guess either the extremal f or the extremal kernel and then the other is quite simple to obtain from the conditions for equality in Hölder's inequality. It seemed easier here to guess the extremal f. One seeks an f in H_p which maximizes, say $|f(ib)|$, for $\|f\|_p = 1$. A natural candidate is

(17)
$$f(z; b) = \frac{(b/\pi)^{1/p}}{(z+ib)^{2/p}}, \qquad b > 0,$$

which gives (conveniently)

(18)
$$|f(t; b)|^p = \frac{b}{\pi(t^2+b^2)} = P_b(t), \qquad \|f\|_p = 1$$

and

(19)
$$|f(ib; b)| = \left(\frac{1}{4\pi b}\right)^{1/p}.$$

Then (16) follows by applying Hölder's inequality to (5) with $y=b$, $x=0$ and f given by (17). We should remark that candidates f should be zero-free in the upper half-plane; for if $f(\gamma)=0$, $\operatorname{Im}\gamma>0$, f in H_p, then the function

$$f_0(z)=\frac{z-\bar{\gamma}}{z-\gamma}f(z)$$

belongs to H_p and $\|f_0\|_p=\|f\|_p$ with

$$|f_0(x+iy)|>|f(x+iy)| \quad \text{for } y>0.$$

As far as seeking candidates for the extremal kernel the contour integral approach is simpler than the Fourier transform approach since we do not have a direct way, except for $p=2$, to determine the norm of the kernel from its Fourier transform. Also the contour integral approach allows us to try to match the kernel with the extremal f.

So suppose f belongs to H_p and g is analytic in the upper half plane, $g(ib)=1$, and such that $g(z)/(z+ib)$ belongs to H_q, $q=p/(p-1)$. Then we have

$$\varphi(z)=\frac{f(z)g(z)}{2\pi i(z+ib)} \quad \text{in } H_1 \tag{20}$$

and hence the related function

$$h(z)=\frac{f(z)g(z)}{2\pi i(z-ib)}, \tag{21}$$

which has a simple pole at $z=ib$, belongs to L_1 on lines parallel to the real axis excluding the line $y=b$, and is analytic and uniformly in L_1 for $y\geqq c>b$; i.e.,

$$h(z+ic) \text{ belongs to } H_1 \quad \text{for } c>b. \tag{21a}$$

Now suppose $0<a<b<c$ and consider the rectangular contour formed by the lines $y=a$, $y=c$, $x=-T$, $x=+T$. We have by Cauchy's integral theorem

$$\begin{aligned} f(ib)=&\int_{-T}^{T}h(t+ia)\,dt-\int_{-T}^{T}h(t+ic)\,dt\\ &+i\int_a^b h(T+iy)\,dy-i\int_a^b h(-T+iy)\,dy. \end{aligned} \tag{22}$$

Now we have for a function f in H_p (Hoffman [3, p. 125])

$$\lim_{x\to\pm\infty}f(x+iy)=0 \quad \text{for } y>0. \tag{23}$$

Thus for φ given by (20) we have

$$\lim_{x\to\pm\infty}\varphi(x+iy)=0, \qquad y>0$$

but

$$\lim_{x\to\pm\infty}\varphi(x+iy)=\lim_{x\to\pm\infty}h(x+iy)=0, \qquad y>0.$$

Thus, taking the limit as $T\to\infty$ in (22) with $h(t+ia)$ and $h(t+ib)$ in L_1 we have

$$f(ib)=\int_{-\infty}^{\infty}h(t+ia)\,dt-\int_{-\infty}^{\infty}h(t+ic)\,dt, \qquad 0<a<b<c. \tag{24}$$

Now since $h(z+ic)$ belongs to H_1 we have

$$\int_{-\infty}^{\infty} h(t+ic)\,dt = 0. \tag{25}$$

To establish this we can use the above contour integral argument to conclude that for f in H_1

$$\int_{-\infty}^{\infty} f(x+iy)\,dx = \text{const.} \quad \text{for } y \geqq 0$$

and we have established elsewhere [4] that for f in H_p, $1 \leqq p < \infty$,

$$\lim_{y\to 0} \int_{-\infty}^{\infty} |f(x+iy)|^p\,dx = 0. \tag{26}$$

(We could also obtain (25) by appealing to the one-sided Paley–Wiener theorem for H_2 (Hoffman [3, p. 131]), since $h(t+ic)$ belongs to $H_2 \cap H_1$.)

Now we can let $a \to 0$ in (24) using the fact that (Hoffman [3, p. 128])

$$\lim_{y\to 0^+} \int_{-\infty}^{\infty} |\varphi(t+iy) - \varphi(t)|\,dt = 0$$

and hence that

$$\lim_{y\to 0^+} \int_{-\infty}^{\infty} |h(t+iy) - h(t)|\,dt = 0$$

to obtain

$$f(ib) = \int_{-\infty}^{\infty} \frac{f(t)g(t)}{2\pi i(t-ib)}\,dt, \qquad b > 0, \tag{27}$$

or replacing $f(z)$ by $f(z+a)$ (a = real), and t by $t-a$,

$$\begin{aligned} f(a+ib) &= \int_{-\infty}^{\infty} f(t)\frac{g(t-a)\,dt}{2\pi i(t-a-ib)} \\ &= \int_{-\infty}^{\infty} f(t)K_b(a-t)\,dt \end{aligned} \tag{28}$$

where

$$K_b(t) = \frac{i}{2\pi}\frac{g(-t)}{(t+ib)}, \qquad b > 0. \tag{29}$$

So here we have valid kernels for H_p. We recall the requirements on g:

$$\begin{aligned} &g(ib) = 1, \\ &G(z) = \frac{g(z)}{(z+ib)} \text{ belongs to } H_q \text{ of the upper half-plane,} \\ &q = p/(p-1) \qquad (b > 0). \end{aligned} \tag{29a}$$

(Obviously b in $G(z)$ can be replaced by any positive number.)

In particular, for $g = 1$, G belongs to H_q for every $q > 1$. The corresponding kernel is the Cauchy kernel which then is a valid kernel for H_p, $1 \leqq p < \infty$ (proving (9)).

Now let us return to f given by (15) or (17),

$$f(z)=\frac{(b/\pi)^{1/p}}{(z+ib)^{2/p}}$$

and make an appropriate choice for g in (21). In order to obtain equality throughout in

$$\begin{aligned}|f(ib)|=\left|\int_{-\infty}^{\infty} h(t)\,dt\right| &\leqq \int_{-\infty}^{\infty} |h(t)|\,dt\\ &=\int_{-\infty}^{\infty} |f(t)|\cdot\left|\frac{g(t)}{2\pi(t-ib)}\right|dt\\ &\leqq \frac{1}{2\pi}\|f\|_p\cdot\left\|\frac{g(t)}{t-ib}\right\|_q\end{aligned}$$

we need

$$|f(t)|^p=c\left|\frac{g(t)}{t-ib}\right|^q=\frac{b}{\pi(t^2+b^2)}$$

and

$$\left|\frac{f(t)g(t)}{t-ib}\right|=e^{i\theta}\frac{f(t)g(t)}{t-ib}$$

or

$$\begin{aligned}\frac{f(t)g(t)}{t-ib}&=c'|f(t)|\cdot|f(t)|^{p/q}=c'|f(t)|^p,\\ g(t)&=c'(t-ib)\frac{|f(t)|^p}{f(t)}\\ &=k\cdot\frac{(t-ib)(t+ib)^{2/p}}{t^2+b^2}=k(t+ib)^{\mu}\end{aligned}$$

where $\mu=2/p-1$; i.e.,

$$g(z)=k(z+ib)^{\mu}.$$

We require

$$g(ib)=1.$$

So we have

$$g(z)=\left(\frac{1}{2}-\frac{iz}{2b}\right)^{\mu}. \tag{30}$$

Thus we have found the extremal kernel,

$$K_y(t;p)=\frac{i}{2\pi}\frac{\left(\frac{1}{2}+\frac{it}{2y}\right)^{\mu}}{t+iy},\qquad y>0, \tag{31}$$

where $\mu=2/p-1$ and it is the only function which will satisfy

$$f(ib)=\int_{-\infty}^{\infty} f(t)K_y(-t)\,dt,\qquad f \text{ in } H_p$$

and will give equality throughout (as just argued) in

$$|f(ib)| = \left| \int_{-\infty}^{\infty} f(t) K_y(-t)\, dt \right| \leqq \int_{-\infty}^{\infty} |f(t)| \cdot |K_y(-t)|\, dy$$
$$\leqq \|f\|_p \cdot \|K_y\|_q$$

with f given by (15) ($A \neq 0$).

We have

$$(32) \qquad \begin{aligned} |K_y(t; p)| &= \frac{1}{2\pi} \frac{\left(\frac{1}{4} + \frac{t^2}{4y^2}\right)^{\mu/2}}{(t^2 + y^2)^{1/2}}, && \text{where } \mu = \frac{2}{p} - 1 \\ &= \frac{1}{(4\pi y)^{1/p}} \frac{(y/\pi)^{1/q}}{(t^2+y^2)^{1/q}}, && q = p/(p-1), \end{aligned}$$

$$(33) \qquad \left\{ \int_{-\infty}^{\infty} |K_y(t; p)|^q\, dt \right\}^{1/q} = \frac{1}{(4\pi y)^{1/p}}.$$

So to summarize, the kernel given by (13) or (31) gives the inequality (14) and then f given by (15) gives the inequality (16), with the qualification of (14) and (16) justified by the inequality argument following (29). The validity of (12) follows from (29) with $g(z)$ given by (30). Thus we have proved the Theorem.

Now regarding the Fourier transforms of the extremal kernels $K_y(t; p)$ we have

$$(34) \qquad \begin{aligned} K_y(t; 1) &= \frac{1}{4\pi y} \frac{iy - t}{iy + t} \\ &= \frac{1}{4\pi y}\left(\frac{2iy}{iy+t} - 1\right) \\ &= \frac{i}{2\pi(t+iy)} - \frac{1}{4\pi y} = C_y(t) - \frac{1}{4\pi y}. \end{aligned}$$

So for $p = 1$ the kernel just differs from the Cauchy kernel by a constant term $-1/(4\pi y)$. So we may write formally

$$(35) \qquad \begin{aligned} \int_{-\infty}^{\infty} K_y(t; 1)\, e^{-i\omega t}\, dt &= 0, && \omega < 0, \\ &= -\frac{1}{2y}\delta(\omega) + e^{-\omega y}, && \omega > 0^- \end{aligned}$$

where $\delta(\omega)$ is the Dirac "delta function".

We may write

$$(36) \qquad K_y(t; p) = \frac{(2y)^\nu}{(y+it)^\nu} \cdot \frac{1}{4\pi y} \frac{iy - t}{iy + t} = \frac{(2y)^\nu}{(y+it)^\nu}\left\{ C_y(t) - \frac{1}{4\pi y} \right\}$$

where $\nu = 2 - 2/p = 2/q$.

Note that

$$(37) \qquad K_y(t; p) = \frac{1}{y} K_1(t/y; p)$$

which can be deduced directly by a scale change in the integral (12). If in (36) we set

$$F_y(t;\nu)=\frac{(2y)^\nu}{(y+it)^\nu} \tag{38}$$

we can express $\hat{K}_y(\omega;p)$, the Fourier transform of $K_y(t;p)$, as

$$\hat{K}_y(\omega;p)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{F}_y(\lambda;\nu)\hat{C}_y(\omega-\lambda)\,d\lambda-\frac{1}{4\pi y}\hat{F}_y(\omega;\nu) \qquad (1<p\leqq\infty). \tag{39}$$

For $\nu>0$, i.e. for $1<p\leqq\infty$, and $y>0$, $F_y(t;\nu)$ has a Fourier transform given by

$$\hat{F}_y(\omega;\nu)\begin{cases}=2\pi\dfrac{(2y)^\nu}{\Gamma(\nu)}|\omega|^{\nu-1}e^{\omega y}, & \omega<0,\quad y>0,\\ =0, & \omega>0,\end{cases} \tag{40}$$

and

$$\hat{C}_y(\omega)\begin{cases}=e^{-\omega y}, & \omega>0,\quad y>0,\\ =0, & \omega<0.\end{cases} \tag{41}$$

It is easier to establish (40) in reverse; i.e., to show that

$$F_y(t;\nu)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{F}_y(\omega;\nu)\,e^{i\omega t}\,d\omega.$$

We observe (cf. (37)) that

$$\hat{K}_y(\omega;p)=\hat{K}_1(\omega y;p). \tag{42}$$

The convolution in (39) can be expressed in terms of the incomplete gamma function. We find that

$$\hat{K}_1(u;p)\begin{cases}=-\dfrac{|2u|^{\nu-1}e^{-|u|}}{\Gamma(\nu)}+e^{|u|}\gamma_\nu(2|u|), & u<0,\\ =e^{-u}, & u>0,\end{cases} \tag{43}$$

where $\nu=2-2/p=2/q$, $1<p\leqq\infty$, and

$$\gamma_\nu(x)=\int_x^\infty\frac{t^{\nu-1}e^{-t}\,dt}{\Gamma(\nu)},\qquad x>0. \tag{44}$$

In (44) we can integrate by parts to obtain

$$\hat{K}_1(u;p)=e^{|u|}\gamma_{\nu-1}(|2u|),\qquad u<0,\quad \nu=2-2/p=2/q, \tag{45}$$

which is valid for $1\leqq p\leqq\infty$ since

$$\gamma_0(x)\equiv\gamma_{-1}(x)\equiv0,\qquad x>0.$$

(A delta function of mass $-\frac{1}{2}$ arises at the origin as $p\to1$.) Graphs of $\hat{K}_1(u;p)$ are shown in Fig. 1 for $1/p=0, 0.1, 0.2, \cdots, 1.0$. Since

$$\Gamma(\nu-1)=\Gamma(1-2/p)$$

is negative for $1<p<2$ and positive for $p>2$, we have $\hat{K}_1(u;p)$ negative for $u<0$ and $1<p<2$, tending to $-\infty$ as $u\to0^-$ (cf. (43)), whereas for $p>2$, $\hat{K}_1(u;p)$ is everywhere positive and also continuous with $\hat{K}_1(u;p)=1$.

Several concluding remarks are in order. First it should be noted that the kernel $K_y(t;p)$, of minimal L_q norm, is bounded and satisfies $K_y(t;p)=O(|t|^{-\nu})$, $t\to\infty$ where $\nu=2/q$. So the kernel actually belongs to $L_{q'}$ for $q'>q/2$ which for $2\leqq q<\infty$ gives additional validity for (12), i.e., $K_y(t;p)$, $1<p\leqq 2$, is a valid kernel for $H_{p'}$, $1<p'<p/(2-p)$. On the other hand, for $p>2$, $K_y(t;p)$ belongs to $L_\infty \cap L_1$. Hence for $p>2$, $K_y(t;p)$, like the Poisson kernel, is a valid kernel for H_p, $1\leqq p\leqq\infty$. (The basic justification goes back to (29) with g given by (30).)

Secondly, the inequality due to J. Korevaar (1949) given by Boas [2, p. 102], for *entire* functions of exponential type λ $(\lambda>0)$

$$|f(x+iy)|<\{\mu_p y^{-1}\sinh p\lambda y\}^{1/p}\|f\|_p \tag{46}$$

where $\mu_p<1/\pi$ for $p>1$ and $\mu_2=1/(2\pi)$ can be replaced by

$$|f(x+iy)|\leqq\frac{e^{\lambda|y|}}{|4\pi y|^{1/p}}\|f\|_p \tag{47}$$

which for sufficiently large $|y|$ is an improvement over (46) and is asymptotically sharp as $y\to\infty$.

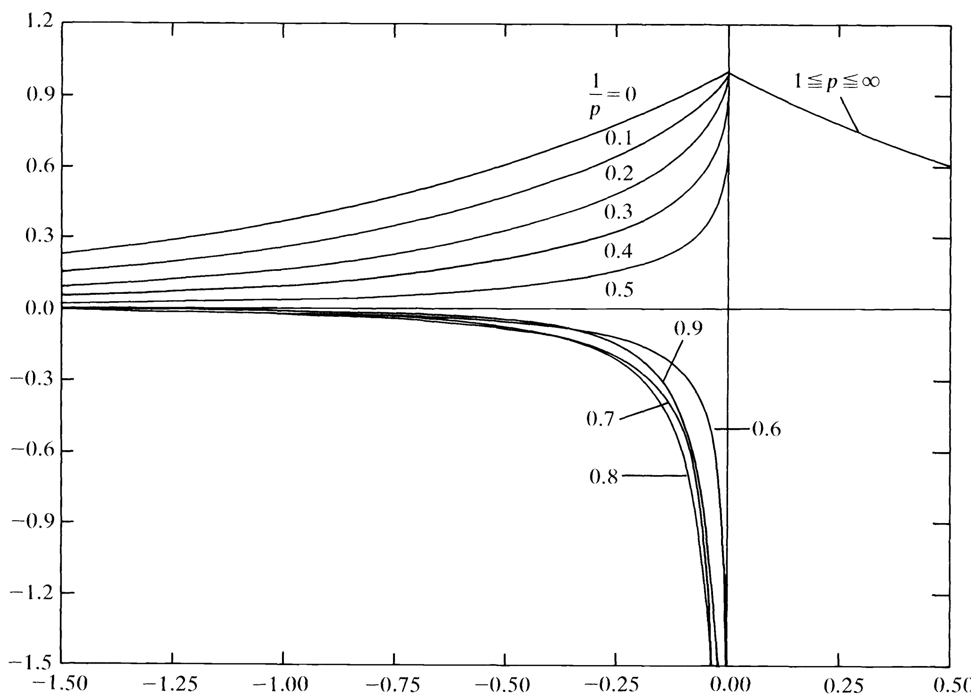

FIG. 1. *Fourier transforms of the minimum L_q-norm kernels for H_p.*

Finally we note that the result (26) and the inequality (3) imply

$$\sup_x|f(x+iy)|=o(y^{-1/p}),\qquad f \text{ in } H_p,\quad 1\leqq p<\infty. \tag{48}$$

To see this we have for $b\geqq 0$, $f(z+ib)$ in H_p and then

$$f(x+iy+ib)=\int_{-\infty}^{\infty} f(t+ib)P_y(x-t)\,dt,$$

$$\sup_x|f(x+iy+ib)|\leqq\frac{M_p}{(4\pi y)^{1/p}}\left\{\int_{-\infty}^{\infty}|f(t+ib)|^p\,dt\right\}^{1/p}.$$

Setting $b = y$ we obtain the result (48) from (26). The result (26), that the *norm* of f in H_p tends to zero as $y \to \infty$ provided $1 \leqq p < \infty$, is apparently not generally known except for the case $p = 2$ which is immediate from the Fourier integral representation (the one-sided Paley–Wiener theorem).

Thus we also have the estimate for entire functions $f(z)$ of exponential type $\lambda > 0$ belonging to L_p on the real axis ($1 \leqq p < \infty$)

$$\sup_x |f(x+iy)| = o\{|y|^{-1/p} e^{\lambda |y|}\}, \qquad |y| \to \infty. \tag{49}$$

Acknowledgment. The author gratefully acknowledges the contribution of Judith B. Seery in furnishing Fig. 1.

REFERENCES

[1] B. F. LOGAN, *Theory of analytic modulation systems*, Bell System Tech. J., 57 (1978), pp. 491–576.
[2] R. P. BOAS, JR., *Entire Functions*, Academic Press, New York, 1954.
[3] KENNETH HOFFMAN, *Banach Spaces of Analytic Functions*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
[4] B. F. LOGAN, *Limits in L_p of convolution transforms with kernels $aK(at)$, $a \to 0$*, this Journal, this issue, pp. 733–740.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0009$01.00/0

NOTE ON CONSTRUCTION OF WEIGHT FUNCTIONS*

B. F. LOGAN†

Abstract. It is well known that the zeros of successive orthogonal polynomials, associated with a nonnegative weight function, are interlaced. Conversely, any two polynomials with interlaced zeros may be regarded as orthogonal polynomials associated with a nonnegative weight function. A simple construction of such a weight function is given.

It is well known that the zeros of orthogonal polynomials are interlaced. Now suppose that $P_n(x)$ and $P_{n-1}(x)$ are given polynomials of degree n and $n-1$ having interlaced (simple) zeros. There are numerous ways of determining a nonnegative function $w(x)$ such that P_n and P_{n-1} are the orthogonal polynomials of degree n and $n-1$ associated with $w(x)$. Here we note a simple construction.

For convenience we assume that the leading coefficients of P_n and P_{n-1} are positive. Then we set

$$P_{n+1}(x)=(a_nx-b_n)P_n(x)-P_{n-1}(x) \tag{1}$$

$$\text{where } a_n>0, \quad -\infty<b_n<\infty,$$

and are otherwise arbitrary.

Let $f(x)$ be any nonnegative integrable function. Then we may take

$$w(x)=\frac{1}{P_n^2(x)}f\{P_{n+1}(x)/P_n(x)\} \tag{2}$$

or a similar form, obtained by replacing $f(x)$ by $x^{-2}f(x^{-1})$,

$$w(x)=\frac{1}{P_{n+1}^2(x)}f\{P_n(x)/P_{n+1}(x)\}. \tag{3}$$

This result, which is somewhat startling at first sight, is quite easy to obtain but, to the author's knowledge, has not been explicitly pointed out in the literature. The result may be obtained readily from identities connecting orthogonal polynomials of the first and second kinds but it is instructive to derive the result from first principles.

To obtain the result we first locate point masses at the zeros of the polynomial

$$A_{n+1}(x;\tau)=P_{n+1}(x)-\tau P_n(x) \tag{4}$$

where τ is any real number, the masses chosen so that P_n and P_{n-1} are the orthogonal polynomials for the atomic distribution. Then we will average the atomic distributions ("τ-distributions") with weight $f(\tau)\,d\tau$ to obtain the result.

Denote the zeros of A_{n+1}, P_{n+1}, and P_n by $\alpha_k=\alpha_k(\tau)$, β_k, and γ_k, respectively, ordered according to the index. We have

$$\beta_1<\gamma_1<\beta_2<\gamma_2<\cdots<\gamma_n<\beta_{n+1}. \tag{5}$$

This interlacing follows from (1) and the fact that

$$\frac{P_{n-1}(x)}{P_n(x)}=\sum_1^n\frac{\nu_k}{x-\gamma_k} \tag{6}$$

* Received by the editors February 3, 1978.
† Bell Laboratories, Murray Hill, New Jersey 07974.

where

$$\nu_k = \frac{P_{n-1}(\gamma_k)}{P'_n(\gamma_k)} > 0, \tag{7}$$

the positivity being a consequence of the interlacing of the zeros of P_n and P_{n-1} and the fact that their leading coefficients are positive (have the same sign).

Similarly we can establish from (4) and (5) that

$$\begin{aligned} &\beta_k < \alpha_k(\tau) < \gamma_k \quad \text{for } \tau > 0 \text{ and } k = 1, 2, \cdots, n, \\ &\beta_{n+1} < \alpha_{n+1}(\tau) < \infty \quad \text{for } \tau > 0, \end{aligned} \tag{8}$$

$$\begin{aligned} &\gamma_{k-1} < \alpha_k(\tau) < \beta_k \quad \text{for } \tau < 0 \text{ and } k = 2, \cdots, n+1, \\ &-\infty < \alpha_1(\tau) < \beta_1 \quad \text{for } \tau < 0, \end{aligned} \tag{9}$$

$$\begin{aligned} &\lim_{\tau \to +\infty} \alpha_k(\tau) = \gamma_k, \qquad k = 1, 2, \cdots, n \\ &\alpha_{n+1}(\tau) \to +\infty \quad \text{as } \tau \to +\infty, \end{aligned} \tag{10}$$

$$\begin{aligned} &\lim_{\tau \to -\infty} \alpha_k(\tau) = \gamma_{k-1}, \qquad k = 2, \cdots, n+1, \\ &\alpha_1(\tau) \to -\infty \quad \text{as } \tau \to -\infty. \end{aligned} \tag{11}$$

Now we seek positive masses $\lambda_k = \lambda_k(\tau)$ associated with $\alpha_k = \alpha_k(\tau)$ such that

$$\sum_1^{n+1} \lambda_k P_n(\alpha_k) Q_{n-1}(\alpha_k) = 0 \tag{12}$$

for any polynomial Q_{n-1} of degree $n-1$. Clearly the weights w_k which annihilate polynomials of degree $n-1$ at the points α_k,

$$\sum_1^{n+1} w_k Q_{n-1}(\alpha_k) = 0, \tag{13}$$

must be

$$w_k = \frac{c}{A'_{n+1}(\alpha_k)} \tag{14}$$

where c is some constant. This follows from the Lagrange interpolation formula or directly from the fact that

$$\lim_{r \to \infty} \int_{|z|=r} \frac{Q_{n-1}(z)}{A_{n+1}(z)} dz = 0.$$

So we may set

$$\lambda_k(\tau) = \frac{1}{A'_{n+1}(\alpha_k; \tau) P_n(\alpha_k)} > 0. \tag{15}$$

Here the positivity follows from

$$A'_{n+1}(x) = P'_{n+1}(x) - \tau P'_n(x), \tag{16}$$

$$\tau = \frac{P_{n+1}(\alpha_k)}{P_n(\alpha_k)}, \tag{17}$$

$$A'_{n+1}(x)=\frac{P'_{n+1}(x)P_n(\alpha_k)-P_{n+1}(\alpha_k)P'_n(x)}{P_n(\alpha_k)} \tag{18}$$

and then

$$P_n(\alpha_k)A'_{n+1}(\alpha_k)=P'_{n+1}(\alpha_k)P_n(\alpha_k)-P_{n+1}(\alpha_k)P'_n(\alpha_k). \tag{19}$$

Now use the fact that (cf. (6))

$$\frac{P_n(x)}{P_{n+1}(x)}=\sum_1^{n+1}\frac{c_k}{x-\beta_k} \quad \text{where} \quad c_k=\frac{P_n(\beta_k)}{P'_{n+1}(\beta_k)}>0 \tag{20}$$

so that

$$\begin{aligned}-\frac{d}{dx}\frac{P_n(x)}{P_{n+1}(x)}&=\frac{P'_{n+1}(x)P_n(x)-P_{n+1}(x)P'_n(x)}{P^2_{n+1}(x)}\\&=\sum_1^{n+1}\frac{c_k}{(x-\beta_k)^2}>0 \qquad (x\neq\beta_k).\end{aligned} \tag{21}$$

Hence

$$P'_{n+1}(x)P_n(x)-P_{n+1}(x)P'_n(x)=\sum_1^{n+1}c_k\frac{P^2_{n+1}(x)}{(x-\beta_k)^2}>0. \tag{22}$$

Thus the positivity of $\lambda_k(\tau)$ is established.

Now with λ_k given by (15) we have (12) holding. It will then follow from (1) that P_{n-1} is the orthogonal polynomial for the τ-distribution. We have from $A_{n+1}(\alpha_k)=0$,

$$P_{n+1}(\alpha_k)=\tau P_n(\alpha_k),$$

and then from (1)

$$P_{n-1}(\alpha_k)=(a_n\alpha_k-b_n-\tau)P_n(\alpha_k). \tag{23}$$

Hence for any polynomial Q_{n-2} of degree $n-2$,

$$\sum_1^n\lambda_kP_{n-1}(\alpha_k)Q_{n-2}(\alpha_k)=\sum_1^n\lambda_kP_n(\alpha_k)(a_n\alpha_k-b_n-\tau)Q_{n-2}(\alpha_k)=0, \tag{24}$$

since $(a_nx-b_n-\tau)Q_{n-2}(x)$ is a polynomial of degree $n-1$.

Next we observe that $\sum_1^{n+1}\lambda_k(\tau)$ is independent of τ. To see this we note that the residues of the function

$$\frac{1}{A_{n+1}(z;\tau)P_n(z)}$$

sum to zero. Thus

$$\begin{aligned}\sum_1^{n+1}\lambda_k(\tau)&=\sum_1^{n+1}\frac{1}{A'_{n+1}(\alpha_k,\tau)P_n(\alpha_k)}=-\sum_1^n\frac{1}{A_{n+1}(\gamma_k,\tau)P'_n(\gamma_k)}\\&=-\sum_1^n\frac{1}{P_{n+1}(\gamma_k)P'_n(\gamma_k)}=\sum_1^n\frac{1}{P_{n-1}(\gamma_k)P'_n(\gamma_k)}\end{aligned} \tag{25}$$

where the second line shows the independence of τ. Further, it follows from the fact that P_n and P_{n-1} are orthogonal polynomials with respect to the τ-distributions

$$\sum_1^{n+1}\lambda_k(\tau)\delta\{x-\alpha_k(\tau)\}, \qquad -\infty<\tau<\infty, \tag{26}$$

as well as the distribution based on the zeros γ_k of P_n,

$$\sum_1^n \mu_k \delta(x-\gamma_k) \quad \text{where} \quad \mu_k = \frac{1}{P_{n-1}(\gamma_k)P'_n(\gamma_k)}, \tag{27}$$

together with the fact [cf. (25)],

$$\sum_1^{n+1} \lambda_k(\tau) = \sum_1^n \mu_k, \tag{28}$$

that the distributions (26) and (27) have the same moments through order $2n-1$ (independent of τ). (Note that (27) may be obtained by letting $\tau \to \infty$ in (26), using (10), (15), (19), and (1).)

Now we average the distribution in (26) setting

$$w(x) = \sum_1^{n+1} \int_{-\infty}^{\infty} \lambda_k(\tau)\delta\{x-\alpha_k(\tau)\}f(\tau)\,d\tau. \tag{29}$$

We change variables in the integrals using

$$\tau = \frac{P_{n+1}\{\alpha_k(\tau)\}}{P_n\{\alpha_k(\tau)\}},$$

$$d\tau = \frac{P'_{n+1}\{\alpha_k\}P_n(\alpha_k) - P_{n+1}(\alpha_k)P'_n(\alpha_k)}{P_n^2(\alpha_k)}\,d\alpha_k,$$

$$\lambda_k(\tau) = \frac{1}{A'_{n+1}(\alpha_k)P_n(\alpha_k)}$$

and then from (19) we have the simple result

$$\lambda_k(\tau)\,d\tau = \frac{d\alpha_k}{P_n^2(\alpha_k)}. \tag{30}$$

Therefore, recalling (10) and (11), we have

$$\begin{aligned}\int_{-\infty}^{\infty} \lambda_k(\tau)\delta\{x-\alpha_k(\tau)\}f(\tau)\,d\tau &= \int_{\gamma_{k-1}}^{\gamma_k} \delta(x-\alpha_k)f\{P_{n+1}(\alpha_k)/P_n(\alpha_k)\}\frac{d\alpha_k}{P_n^2(\alpha_k)} \\ &= \frac{1}{P_n^2(x)}f\{P_{n+1}(x)/P_n(x)\}, \qquad \gamma_{k-1} < x < \gamma_k,\end{aligned} \tag{31}$$

which is valid for $k=1,2,\cdots,n+1$ if we set $\gamma_0=-\infty$, $\gamma_{n+1}=+\infty$. Thus we have established the result. We note from (25) that

$$\int_{-\infty}^{\infty} w(x)\,dx = \left\{\int_{-\infty}^{\infty} f(x)\,dx\right\} \sum_1^n \{P_{n-1}(\gamma_k)P'_n(\gamma_k)\}^{-1} \tag{32}$$

where γ_k are the zeros of P_n.

One can generalize the result in an obvious way by introducing additional polynomials

$$P_{m+1}(x) = (a_m x - b_m)P_m(x) - P_{m-1}(x) \qquad m \geqq n, \quad a_m > 0, \quad -\infty < b_m < \infty$$

and taking

$$w(x)=\sum_{n}^{\infty}\frac{f_k\{P_{k+1}(x)/P_k(x)\}}{P_k^2(x)},$$

$$\sum_{n}^{\infty}\int_{-\infty}^{\infty}f_k(x)\,dx<\infty,\qquad f_k(x)\geqq 0.$$

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0010$01.00/0

THE ASYMPTOTIC EXPANSION OF THE INCOMPLETE GAMMA FUNCTIONS*

N. M. TEMME†

Abstract. Earlier investigations on uniform asymptotic expansions of the incomplete gamma functions are reconsidered. The new results include estimations for the remainder and the extension of the results to complex variables.

1. Introduction. We consider the incomplete gamma functions ratios P and Q defined by

$$P(a,x)=\frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}\,dt,\qquad Q(a,x)=\frac{1}{\Gamma(a)}\int_x^\infty t^{a-1}e^{-t}\,dt. \tag{1.1}$$

We suppose first that x and a are real with

$$x\geqq 0,\qquad a>0. \tag{1.2}$$

In Temme [4] we derived asymptotic expansions of P and Q for $a\to\infty$, uniformly valid for $x\geqq 0$. In this paper we reconsider these expansions. Our new results concern the representations of the remainder in the asymptotic expansion, representations for the coefficients of the expansion for numerical applications, numerical upper bounds for the remainder in the case of real variables, and extension of the asymptotic expansions to the case of complex variables. These problems are mentioned by Olver in [2].

To describe the expansions given in [4] we introduce the following parameters

$$\lambda=x/a,\qquad \mu=\lambda-1,\qquad \eta=\{2[\mu-\ln(1+\mu)]\}^{1/2}, \tag{1.3}$$

with the convention that the square root has the sign of μ ($\mu>-1$). As a function of μ, η is monotone and infinitely differentiable on $(-1,\infty)$. Analytic properties of $\eta(\mu)$ for complex μ are considered in § 5.

The asymptotic expansions of P and Q derived in [4] follow from the representations

$$\begin{aligned}P(a,x)&=\tfrac12\operatorname{erfc}[-\eta(a/2)^{1/2}]-R_a(\eta),\\ Q(a,x)&=\tfrac12\operatorname{erfc}[\eta(a/2)^{1/2}]+R_a(\eta)\end{aligned} \tag{1.4}$$

with

$$R_a(\eta)\sim(2\pi a)^{-1/2}e^{-(1/2)a\eta^2}\sum_{k=0}^{\infty}c_k(\eta)a^{-k} \tag{1.5}$$

for $a\to\infty$, uniformly valid with respect to $\eta\in\mathbb{R}$; erfc is the complementary error function defined by

$$\operatorname{erfc}(x)=2\pi^{-1/2}\int_x^\infty e^{-t^2}\,dt. \tag{1.6}$$

The expansion (1.5) was derived by using saddle point methods. In § 2 we use a different method which yields recurrence relations for the coefficients c_k and a representation for the remainder of (1.5). In § 3 we discuss representations for c_k that can be used for numerical applications. In § 4 numerical error bounds are constructed

* Received by the editors May 31, 1977, and in revised form March 17, 1978.
† Mathematisch Centrum, Amsterdam, The Netherlands.

for the remainder of the series in (1.5) when the first n terms in the series are used. Bounds are given up to $n=10$. As a side result this section gives bounds for the remainder of the asymptotic expansion of the reciprocal gamma function $1/\Gamma(x)$ for real x. In § 5 the results are extended to complex values of a and x.

2. Recurrence relations for the coefficients and representations of the remainder. First we remark that the asymptotic expansion for $a \to \infty$, of $dR_a(\eta)/d\eta$ may be obtained by formal differentiation of (1.5). This is not proved here, but it follows from the representation of $R_a(\eta)$ in our previous paper (formula (2.10) of Temme [4]). The result is

$$\frac{dR_a(\eta)}{d\eta} \sim a(2\pi a)^{-1/2}\, e^{-(1/2)a\eta^2} \sum_{k=0}^{\infty} c_k^{(1)}(\eta) a^{-k} \tag{2.1}$$

with

$$\begin{aligned} c_0^{(1)}(\eta) &= -\eta c_0(\eta), \\ c_k^{(1)}(\eta) &= -\eta c_k(\eta) + \frac{dc_{k-1}(\eta)}{d\eta}; \qquad k \geqq 1. \end{aligned} \tag{2.2}$$

Secondly, we need the coefficients of the asymptotic expansion of the complete gamma function. Let us define

$$\Gamma^*(a) = (a/(2\pi))^{1/2}\, e^a a^{-a} \Gamma(a), \qquad a > 0. \tag{2.3}$$

Then Γ^* and $1/\Gamma^*$ have the well-known asymptotic expansions for $a \to \infty$

$$\begin{aligned} \Gamma^*(a) &\sim \sum_{k=0}^{\infty} (-1)^k \gamma_k a^{-k}, \\ 1/\Gamma^*(a) &\sim \sum_{k=0}^{\infty} \gamma_k a^{-k}. \end{aligned} \tag{2.4}$$

The first few coefficients are

$$\gamma_0 = 1, \qquad \gamma_1 = -\frac{1}{12}, \qquad \gamma_2 = \frac{1}{288}, \qquad \gamma_3 = \frac{139}{51840}.$$

Further coefficients follow from Spira [3] and Wrench [5]. Wrench gives $(-1)^k \gamma_k$ up to $k=20$ in rational form, Spira the remaining up to $k=30$. Decimal representations are also given in both references.

With these preparations we have

THEOREM 1. *Let* $\{\gamma_k\}$ *be defined by* (2.4). *Then the coefficients* c_k *of* (1.5) *satisfy the recurrence relation*

$$\begin{aligned} c_0(\eta) &= \frac{1}{\mu} - \frac{1}{\eta}, \\ \eta c_k(\eta) &= \frac{dc_{k-1}(\eta)}{d\eta} + \frac{\eta}{\mu}\gamma_k, \qquad k \geqq 1. \end{aligned} \tag{2.5}$$

Proof. By differentiating one of the formulas in (1.4) with respect to η and by using (1.1) it follows that

$$\frac{d}{d\eta} R_a(\eta) = (a/(2\pi))^{1/2}\left(1 - \frac{1}{\mu+1}\,\frac{1}{\Gamma^*(a)}\,\frac{d\mu}{d\eta}\right) e^{-(1/2)a\eta^2}. \tag{2.6}$$

From (1.3) we have

$$\frac{d\mu}{d\eta}=\frac{(\mu+1)\eta}{\mu}, \tag{2.7}$$

and substituting (2.1) and the second relation of (2.4) we obtain (2.5) by collecting equal powers of a^{-1} and using (2.2). □

As follows from [4], the coefficients c_k are holomorphic in a neighborhood of $\eta=0$. In fact the singularities of $1/\mu$ and $1/\eta$ in c_0 cancel each other. So the limiting value of c_0 for $\eta\to 0$ is well defined.

Owing to the presence of the derivative of c_{k-1} in (2.5) this formula cannot be handled easily from a numerical point of view. Further, the above mentioned cancellation of singular parts in c_0 occurs in all c_k when working with (2.5). Therefore other representations are given for these coefficients. In the next section we discuss some aspects of the Taylor expansions for small $|\eta|$-values, while for larger $|\eta|$-values a recurrence relation is constructed from which the coefficients can be computed directly. But first we give representations of the remainder in the asymptotic expansion (1.5).

From (1.4) it follows that $R_a(\infty)=R_a(-\infty)=0$. Hence, integration of (2.6) gives

$$\begin{aligned} R_a(\zeta)&=(a/(2\pi))^{1/2}\int_{-\infty}^{\zeta}\left[1-\frac{\eta}{\mu}\frac{1}{\Gamma^*(a)}\right]e^{-(1/2)a\eta^2}\,d\eta\\ &=-(a/(2\pi))^{1/2}\int_{\zeta}^{\infty}\left[1-\frac{\eta}{\mu}\frac{1}{\Gamma^*(a)}\right]e^{-(1/2)a\eta^2}\,d\eta, \end{aligned} \tag{2.8}$$

where μ as a function of η is defined implicitly in (1.3). From these representations and the recurrence relations for c_k a simple expression for the remainder follows. For this purpose we introduce the notation

$$R_a(\eta)=(2\pi a)^{-1/2}e^{-(1/2)a\eta^2}\left[\sum_{k=0}^{N-1}c_k(\eta)a^{-k}+a^{-N}G_N(\eta;a)\right], \tag{2.9}$$

$a>0$, $\eta\in\mathbb{R}$, $N=0,1,2,\cdots$. Furthermore, we need a notation for the remainder in the asymptotic expansion of $1/\Gamma^*(a)$, which is written as

$$1/\Gamma^*(a)=\sum_{k=0}^{N-1}\gamma_k a^{-k}+a^{-N}H_N(a),\qquad a>0,\qquad N=0,1,2,\cdots. \tag{2.10}$$

THEOREM 2. *Let G_N and H_N be defined by* (2.9) *and* (2.10). *Then*

$$\begin{aligned} e^{-(1/2)a\zeta^2}G_N(\zeta;a)&=a\int_{\zeta}^{\infty}\eta c_N(\eta)\,e^{-(1/2)a\eta^2}\,d\eta\\ &\quad+H_{N+1}(a)\int_{\zeta}^{\infty}\frac{\eta}{\mu}e^{-(1/2)a\eta^2}\,d\eta. \end{aligned} \tag{2.11}$$

Proof. The proof follows immediately from substitution of (2.9) and (2.10) in (2.6) (and by use of (2.5) and (2.7)). □

The second integral in (2.11) can be expressed in terms of $Q(a,x)$. The first one can be bounded if we have estimations for $|c_k(\eta)|$. From representations of c_k given in the following section it follows that $[2+\mu(\eta)]^{\kappa}|c_k(\eta)|$ is a bounded function of $\eta\in\mathbb{R}$, with

$\kappa=\frac{1}{2}$ if $k=0$, $\kappa=1$ if $k\geqq 1$. For estimating the second integral of (2.11) we define

$$C_k=\sup_{\eta\in\mathbb{R}}[2+\mu(\eta)]^{\kappa}|c_k(\eta)|,\qquad k=0,1,2,\cdots,$$

(2.12)

$$\kappa=\begin{cases}\frac{1}{2} & \text{for } k=0,\\ 1 & \text{for } k\geqq 1.\end{cases}$$

For numerical applications the following is important.

COROLLARY. *Let C_k be defined by* (2.12). *Then for $N=0,1,2,\cdots$,*

(2.13)
$$|Q(a,x)-\tfrac{1}{2}\operatorname{erfc}[\eta(a/2)^{1/2}]-e^{-(1/2)a\eta^2}(2\pi a)^{-1/2}\sum_{k=0}^{N-1}c_k(\eta)a^{-k}| \leqq Q_N(\eta;a)(2\pi a)^{-1/2}a^{-N},$$

with

(2.14)
$$Q_N(\eta;a)=(\mu+2)^{-\kappa}C_N\begin{cases}e^{-(1/2)a\eta^2}\\(2-e^{-(1/2)a\eta^2})\end{cases}+|H_{N+1}(a)|e^a a^{-a}\Gamma(a)Q(a,x),$$

where the upper term is for $\eta\geqq 0$, the lower one for $\eta\leqq 0$.

In § 4 we give more (numerical) information on C_N and H_N. With numerical values of C_N and H_N we have strict and realistic error bounds for the remainder of the uniform asymptotic expansion of $Q(a,x)$. Similar results hold for the function $P(a,x)$. For $N=0,1,2,\cdots$ we have

(2.15)
$$|P(a,x)-\tfrac{1}{2}\operatorname{erfc}[-\eta(a/2)^{1/2}]+e^{-(1/2)a\eta^2}(2\pi a)^{-1/2}\sum_{k=0}^{N-1}c_k(\eta)a^{-k}| \leqq P_N(\eta;a)(2\pi a)^{-1/2}a^{-N},$$

with

(2.16)
$$P_N(\eta;a)=(\mu+2)^{-\kappa}C_N\begin{cases}(2-e^{-(1/2)a\eta^2})\\e^{-(1/2)a\eta^2}\end{cases}+|H_{N+1}(a)|e^a a^{-a}\Gamma(a)P(a,x),$$

where the upper term is for $\eta\geqq 0$, the lower one for $\eta\leqq 0$.

Remark 1. The functions multiplying the constants C_N in (2.14) and (2.16) have quite different behavior for $\eta<0$ and $\eta>0$. This, however, is in agreement with the behavior of the functions P and Q in the same formula. In fact, the bounds P_N and Q_N give a measure for the relative accuracy for the error in the uniform expansions.

Remark 2. The asymptotic expansion (1.5) and the representation for the remainder is easily obtained by partial integration of one of the integrals in (2.8) and by using the recursions (2.5) and $H_k(a)=\gamma_k+(1/a)H_{k+1}(a)$. To demonstrate this we write in the second of (2.8) $1/\Gamma^*(a)=1+a^{-1}H_1(a)$ and we use the first line of (2.5). Then we obtain

(2.17)
$$R_a(\zeta)=(a/(2\pi))^{1/2}\int_\zeta^\infty \eta c_0(\eta)\,e^{-(1/2)a\eta^2}\,d\eta +(2\pi a)^{-1/2}H_1(a)\int_\zeta^\infty \frac{\eta}{\mu}e^{-(1/2)a\eta^2}\,d\eta.$$

Writing the first integral as $-a^{-1}\int c_0(\eta)\,d\,e^{-(1/2)a\eta^2}$, we obtain for this integral (after partial integration and using the second line of (2.5))

$$a^{-1}c_0(\zeta)\,e^{-(1/2)a\zeta^2}+a^{-1}\int \eta c_1(\eta)\,e^{-(1/2)a\eta^2}\,d\eta-\gamma_1 a^{-1}\int\frac{\eta}{\mu}e^{-(1/2)a\eta^2}\,d\eta.$$

Combining the second integral of this expression with the second integral of (2.17) and using $H_1(a)-\gamma_1=a^{-1}H_2(a)$ we arrive at (2.9) and (2.11) with $N=1$. So the process can be continued.

3. Representations of c_k. Using (2.5) with $k=1$ we obtain

$$\eta c_1(\eta)=-\frac{1}{\mu^2}\frac{d\mu}{d\eta}+\frac{1}{\eta^2}+\frac{\eta}{\mu}\gamma_1, \tag{3.1}$$

and using (2.7) we have

$$c_1(\eta)=\frac{1}{\eta^3}-\frac{1+\mu+\mu^2/12}{\mu^3}. \tag{3.2}$$

Computing higher order coefficients we notice the following structure

$$c_k(\eta)=(-1)^k\left\{\frac{Q_k(\mu)}{\mu^{2k+1}}-\frac{A_k}{\eta^{2k+1}}\right\}, \tag{3.3}$$

where Q_k is a polynomial in μ of degree $2k$ and $A_k=2^k=2^k\Gamma(k+\frac{1}{2})/\Gamma(\frac{1}{2})$.

The first few polynomials are

$$\begin{aligned} &Q_0(\mu)=1,\\ &Q_1(\mu)=1+\mu+\tfrac{1}{12}\mu^2,\\ &Q_2(\mu)=3+5\mu+\tfrac{25}{12}\mu^2+\tfrac{1}{12}\mu^3+\tfrac{1}{288}\mu^4. \end{aligned} \tag{3.4}$$

In order to preserve accuracy near $\mu=-1$ we write

$$Q_k(\mu)=(1+\mu)P_k(\mu)+(-1)^k\gamma_k\mu^{2k}. \tag{3.5}$$

P_k is a polynomial of degree $2k-2$ ($k\geqq 1$). Writing

$$P_k(\mu)=p_0^{(k)}+p_1^{(k)}\mu+\cdots+p_{2k-2}^{(k)}\mu^{2k-2} \tag{3.6}$$

we have the relation (which is easily obtained by substituting (3.6), (3.5) and (3.3) in (2.5))

$$\begin{aligned} &p_0^{(k)}=(2k-1)p_0^{(k-1)}\\ &p_j^{(k)}=(2k-1-j)[p_j^{(k-1)}+p_{j-1}^{(k-1)}], \qquad j=1,2,\cdots,2k-4,\\ &p_{2k-3}^{(k)}=2p_{2k-4}^{(k-1)}, \qquad p_{2k-2}^{(k)}=(-1)^{k-1}\gamma_{k-1}, \end{aligned} \tag{3.7}$$

with as starting polynomial $P_1(\mu)=1$, or $p_0^{(1)}=1$.

In Table 1 we give the coefficients $p_j^{(k)}$ of (3.7) for $k=1,2,\cdots,5$, $j=0,1,\cdots,2k-2$.

TABLE 1

k	$p_j^{(k)}$
1	1
2	3, 2, 1/12
3	15, 20, 25/4, 1/6, 1/288
4	105, 210, 525/4, 77/3, 49/96, 1/144, −139/51840
5	945, 2520, 9555/4, 1883/2, 12565/96, 149/72, 221/17280, −139/25920, −571/2488320

As remarked in the previous section, when computing c_k via (3.3) near $\eta=0$, one must pay attention to the cancellation of the singular parts. It may be convenient (and, when using a fixed number of word lengths on a computer, even necessary) to use representations which preserve the accuracy near $\eta = 0$.

If $|\eta|$ is small it is preferred to use expansions either in terms of η or in terms of μ. We advise expansions in η, since it gives better convergence properties. When expanding c_k in powers of μ we need (among others) the expansion of η in powers of μ. Owing to the singularity of the logarithm in (1.3), the radius of convergence of this series is 1. Other singularities for η are zeros of $\mu - \ln(1+\mu)$, but they are outside the domain $|\mu| \leqq 1$. This follows from straightforward analysis. The reader may also consult an interesting note of Diekmann [1]. The expansion of μ in powers of η has radius of convergence $2\sqrt{\pi} \simeq 3.54$. This follows from the analysis of § 5. From the recurrence relation (2.5) it is easily seen that the radius of convergence of the power series for c_k either in μ or in η is the same for all k.

We conclude this section with some information on the construction of the coefficients for the expansion of c_k in powers of η. It is convenient to start with the computation of the α_k in

$$\mu(\eta) = \alpha_1\eta + \alpha_2\eta^2 + \cdots, \tag{3.8}$$

where μ is defined implicitly in (1.3). Substitution of (3.8) in (2.7) yields the recurrence relation

$$(k+1)\ \alpha_k = \alpha_{k-1} - \sum_{j=2}^{k-1} j\alpha_j\alpha_{k-j+1}, \qquad k \geqq 2.$$

The first few are $\alpha_1 = 1$, $\alpha_2 = \frac{1}{3}$, $\alpha_3 = \frac{1}{36}$, $\alpha_4 = -\frac{1}{270}$, $\alpha_5 = \frac{1}{4320}$. With α_k we also have available the γ_k of (2.4), which are also needed in (2.5). The relation between α_k and γ_k is $\gamma_k = (-1)^k 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2k+1)\alpha_{2k+1}$, $(k = 0, 1, 2, \cdots)$.

The coefficients $c_n^{(k)}$ in the expansion $c_k(\eta) = \sum_{n=0}^{\infty} c_n^{(k)}\eta^n$ follow now from (2.5). For $k = 0$ we have

$$c_0^{(0)} = -\tfrac{1}{3}, \qquad c_k^{(0)} = (k+2)\alpha_{k+2}, \qquad k \geqq 1$$

and for general $k \geqq 1$ the recursion is

$$c_n^{(k)} = \gamma_k c_n^{(0)} + (n+2)c_{n+2}^{(k-1)}, \qquad n \geqq 0,$$

or, in terms of $c_n^{(0)}$,

$$c_n^{(k)} = \gamma_k c_n^{(0)} + \gamma_{k-1}(n+2)c_{n+2}^{(0)} + \cdots + \gamma_0(n+2)\cdots(n+2k)c_{n+2k}^{(0)}. \tag{3.9}$$

As follows from the rate of convergence of the series for c_k (with radius $2\sqrt{\pi}$) successive terms in (3.9) are decreasing in absolute value. Hence no instability problems arise when using (3.9) for the computation of $c_n^{(k)}$.

4. Bounds for the remainder in the asymptotic expansion. In Table 2 we give the numbers C_k defined in (2.12). These bounds were obtained numerically by using representations of c_k given in the foregoing section. From (3.3) and (1.3) it follows that

$$\lim_{\eta\to-\infty} c_0(\eta) = -1, \qquad \lim_{\eta\to+\infty} = [2+\mu(\eta)]^{1/2}c_0(\eta) = -2^{-1/2},$$

$$\lim_{\eta\to\pm\infty} [2+\mu(\eta)]c_k(\eta) = \pm\gamma_k, \qquad k \geqq 1.$$

TABLE 2

k	c_{2k}	c_{2k+1}
0	1	0.083
1	0.010	0.0027
2	0.0024	0.00092
3	0.0016	0.00083
4	0.0021	0.0014
5	0.0045	

Next we give details for computing the bounds H_k (defined in (2.10)) for $k = 0, 1, \cdots, 10$. It is convenient to start with details for obtaining the asymptotic expansion of $1/\Gamma(a)$. Again, a is a positive number. Our starting point is Hankel's integral

$$\frac{1}{\Gamma(a)}=\frac{1}{2\pi i}\int_{-\infty}^{(0+)} e^t t^{-a}\,dt. \tag{4.1}$$

This can be written as

$$\frac{1}{\Gamma(a)}=a^{1-a}e^a\pi^{-1}\int_{-\infty}^{\infty} e^{-(1/2)au^2}g(u)\,du$$

with

$$f(u)=\frac{ut}{1-t},\qquad -\tfrac{1}{2}u^2=t-1-\ln t,\qquad g(u)=\frac{f(u)+f(-u)}{2i}, \tag{4.2}$$

where $u\in\mathbb{R}$ and t follows the steepest descent line for (4.1) in the t-plane. More information on the relations in (4.2) is found in our previous paper [4].

The asymptotic expansion of $1/\Gamma(a)$ is obtained by expanding $g(u)$ in powers of u and integrating term by term. Let us define (for $N=0,1,2,\cdots$) the functions g_N by writing

$$g(u)=\sum_{k=0}^{N-1} a_k u^{2k}+a_N u^{2N}g_N(u),\qquad a_k=\frac{1}{(2k)!}g^{(2k)}(0);$$

all a_k are different from zero. Then the function H_N of (2.10) is given by

$$H_N(a)=(a/(2\pi))^{1/2}a_N a^N\int_{-\infty}^{\infty} e^{-(1/2)au^2}u^{2N}g_N(u)\,du.$$

It appears that g, and hence g_N, is bounded on $\mathbb{R}$. Let us define the bounds

$$G_N=\sup_{u\in\mathbb{R}}|g_N(u)|,$$

then a bound for H_N is given by

$$|H_N(a)|\leqq|\gamma_N|G_N,\qquad a>0, \tag{4.3}$$

where γ_k are the coefficients in (2.10). Table 3 gives the value of G_N for $N = 0, 1, \cdots, 11$.

TABLE 3

N	G_{2N}	G_{2N+1}
0	1	1
1	1.95	1
2	3.33	1
3	5.05	1
4	6.95	1
5	8.90	1

For $N=0, 1, 3, 5, 7, 9, 11$ the maximal function values of $|g_N(u)|$ occur at $u=0$; for $N=2, 4, 6, 8, 10$ the maxima occur in the neighborhood of $u=\pm 2\sqrt{\pi}$. These latter points are the points on the real axis marking the domain of convergence of the Taylor series of g.

With the data of Table 2 and Table 3 and relation (4.3) the bounds Q_N and P_N defined in (2.14) and (2.16) are easily computed.

5. Extension to complex variables. In this section we will show that the asymptotic expansion for P and Q given by (1.4) and (1.5) are valid for $a \to \infty$ uniformly in $|\arg a| \leqq \pi - \varepsilon_1$, $|\arg x/a| \leqq 2\pi - \varepsilon_2$ where ε_1 and ε_2 are positive numbers, $0 < \varepsilon_1 < \pi$, $0 < \varepsilon_2 < 2\pi$.

The condition on the argument of a follows from the validity of the expansions in (2.4), which are known to be uniformly valid when $|\arg a| \leqq \pi - \varepsilon_1$. As noticed in Remark 2 of § 2 the asymptotic expansion of $R_a(\eta)$ can be obtained by partial integration of one of (2.8). If we consider the second integral, one of the assumptions by partial integration will be that $\exp(-\frac{1}{2}a\eta^2)$ vanishes at infinity in a certain direction of the η-plane. If $|\arg a| < \pi$ and if it is allowed to use η-values at infinity with $\arg(a\eta^2) < \pi/2$ then the convergence of the integral is established for $|\arg a| \leqq \pi - \varepsilon_1$. From these inequalities it follows that it is sufficient to show that for large $|\eta|$ we can take $\arg \eta$ in $(-\frac{3}{4}\pi, \frac{3}{4}\pi)$. A second aspect of using the second integral of (2.8) is the possibility of joining the point ζ with ∞ such that the function $\mu(\eta)$ of the integral is holomorphic along this path and such that the point ζ can be associated unequivocally with a point in the μ-plane. In order to settle this we discuss the relation between η and the parameter μ (or λ) for complex values.

It is convenient to consider

$$\eta = [2(\lambda - 1 - \ln \lambda)]^{1/2}. \tag{5.1}$$

For $\lambda > 0$ the function η is to be interpreted as drawn in Fig. 1. This implies a choice of the square root.

We obtain a clear insight in the mapping $\lambda \to \eta(\lambda)$ and its inverse if we draw images of the half-lines l_ϕ defined by

$$l_\phi = \{\lambda \mid \lambda = \rho\, e^{i\phi}, \rho > 0\}$$

where ϕ is real, $|\phi| \leqq 2\pi$. Writing $\eta = \alpha + i\beta$ we find that the image of l_ϕ in the η-plane is governed by the equations

$$\tfrac{1}{2}(\alpha^2 - \beta^2) = \rho \cos \phi - 1 - \ln \rho,$$

$$\alpha\beta = \rho \sin \phi - \phi.$$

Taking into account the convention about the choice of the square root in (5.1) we

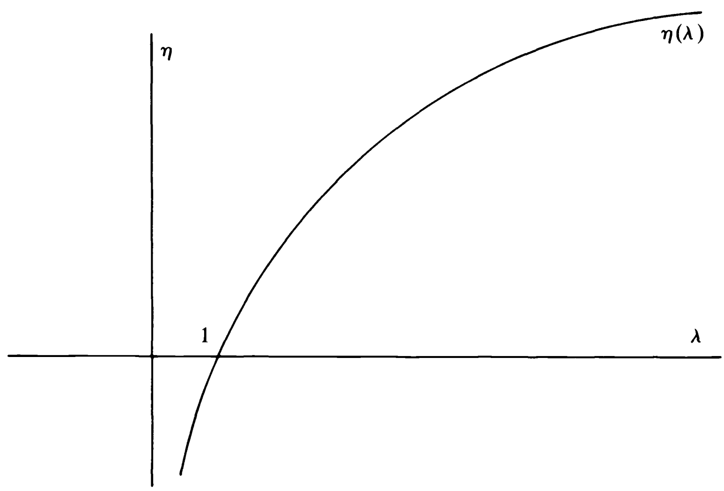

FIG. 1

obtain Fig. 2, which contains images of l_ϕ for $0 \leqq \phi \leqq 2\pi$. The complete picture for $-2\pi \leqq \phi \leqq 2\pi$ is symmetric with respect to the α-axis.

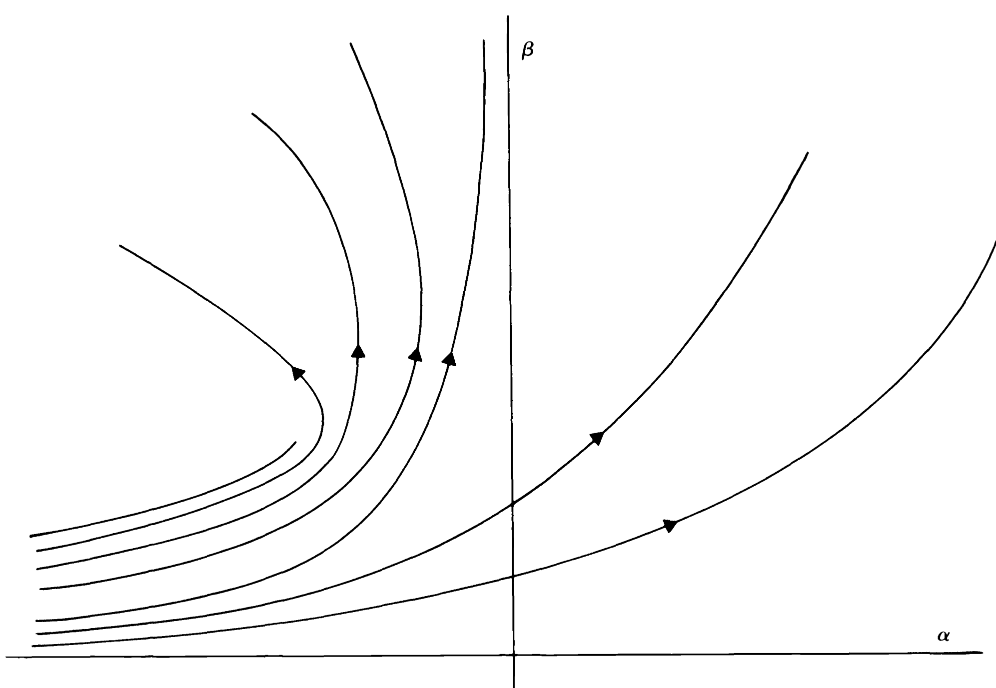

FIG. 2

The shown directions correspond to increasing values of ρ on l_ϕ. The half-lines $l_{\pm 2\pi}$ are mapped on part of the hyperbolae $\alpha\beta = \mp 2\pi$. The points $\eta^\pm = e^{\pm 3\pi i/4} 2\sqrt{\pi}$ are singular points of the mapping. Other singular points are located in other Riemann sheets of the η-plane. Convenient branch-cuts for the function $\lambda(\eta)$ are the parts of the hyperbolae $\alpha\beta = \pm 2\pi$ with $\alpha \leqq -\sqrt{2\pi}$. With the η-plane cut along these curves, lines l_ϕ with the values of ϕ outside the interval $[-2\pi, 2\pi]$ can be traced, but for our problem this is superfluous.

It is concluded that any point in the finite η-plane (not on the branch-cuts), corresponds to a point in the λ-plane with $|\arg \lambda| < 2\pi$. Consequently, if we integrate the second integral of (2.8) along a path that avoids the branch-cuts in the η-plane, the function $\mu(\eta) = \lambda(\eta) - 1$ is holomorphic. The conditions for allowing values of arg a in $(-\pi, \pi)$ are amply satisfied, since admissable directions in the η-plane can be found in the sector $-\pi < \arg \eta < \pi$.

Remark. Singular points of the mapping $\eta \to \lambda(\eta)$ can also be found by considering the derivative $d\lambda/d\eta = \lambda\eta/(\lambda - 1)$; $\lambda = 1$ gives a regular point but $\lambda = e^{2\pi i n}(n = \pm 1, \pm 2, \cdots)$ gives (due to the many-valuedness of the logarithm in (5.1)) singular points η_n satisfying $\frac{1}{2}\eta_n^2 = -2\pi i n$, $n = \pm 1, \pm 2, \cdots$.

The integration by parts procedure leads eventually to (2.9) and (2.11). From the properties of the coefficients c_k and by taking appropriate contours in (2.11) it follows that for $N = 0, 1, 2, \cdots$,

$$e^{-(1/2)a\eta^2} G_N(\eta; a) = O(1), \qquad a \to \infty$$

uniformly in $|\arg a| \leqq \pi - \varepsilon_1$, $|\arg \lambda| \leqq 2\pi - \varepsilon_2$.

Acknowledgment. Thanks are due to Mr. F. J. Burger who did nice and helpful work on writing programs for computing the coefficients and the error bounds, to the editor, Prof. F. W. J. Olver, whose remarks led to a more realistic bound for the remainder and to the referees for suggestions for a better presentation of the paper.

REFERENCES

[1] O. DIEKMANN, *Asymptotic expansion of certain numbers related to the gamma function*, Report TN 80, Mathematical Centre, Amsterdam, 1975.

[2] F. W. J. OLVER, *Unsolved problems in the asymptotic estimation of special functions*, Theory and Applications of Special Functions, R. A. Askey, ed., Academic Press, New York, 1975.

[3] R. SPIRA, *Calculating of the gamma function by Stirling's formula*, Math. Comp., 25 (1971), pp. 317–322.

[4] N. M. TEMME, *Uniform asymptotic expansions of the incomplete gamma functions and the incomplete beta functions*, Ibid., 29 (1975), pp. 1109–1114.

[5] J. W. WRENCH, JR., *Concerning two series for the gamma function*, Ibid., 22 (1968), pp. 617–626.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0011$01.00/0

REALIZABILITY THEORY OF CONTINUOUS LINEAR OPERATORS ON GROUPS*

RAMENDRA KRISHNA BOSE†

Abstract. Let G and H be a certain type of locally compact (abelian) group. $D(G)$ denotes the space of regular functions with compact support on G and $D'(G)$ is the corresponding space of distributions. Linear mappings from $D(G)$ into $D'(G)(D'(H))$ are the subject of our investigations. We have carried out some systems-theoretic investigation in a distributional setting where distributions are defined on groups. Such investigations in (Schwartz's) distributional setting have been carried out by several authors. We have chosen the distribution theory on groups as developed by F. Bruhat, K. Maurin and G. I. Kac. This choice is motivated by the existence of Bruhat's kernel theorem and the nuclearity of the space $D(G)$. Properties such as continuity, regularity, translation-invariance, causality, semipassivity and passivity (a certain positivity property) are imposed on the linear mappings and their effects are studied.

Representations of continuous linear mappings from $D(G)$ into $D'(H)$ which are regular into $C(H)$, of continuous linear and translation-invariant mappings from $D(G)$ into $D'(G)$ and of linear and scatter-semipassive mappings from $D(G)$ into $D'(G)$ are obtained. We establish one-to-one correspondence between contractions in $L_2(G)$ and linear and scatter-semipassive mappings from $D(G)$ into $D'(G)$. We also show that causality and semipassivity imply passivity in the scattering formalism and passivity implies causality in the immittance formalism. A characterization of linear, scatter-semipassive, and real mappings in terms of a positivity criterion is established.

1. Introduction. Classical realizability theory has been developed and studied by McMillan [8], and Konig and Meixner [6]. In (Schwartz's) distributional setting this theory has been studied by Zemanian [18], Wohlers and Beltrami [17], Dolezal [2] and Meidan [9]. Hackenbroch [4], and Freedman and Falb [3] have considered certain aspects of this theory where the time domain has been replaced by a locally compact abelian group. We want to give a limited exposition of the realizability theory as applied to operators generated by abstract systems defined on a locally compact (abelian) group.

Let G be a certain type of locally compact (abelian) group with $+$ as the group operation and 0 as the identity element. $D(G)$ denotes the space of regular functions with compact support on G and $D'(G)$ denotes the corresponding space of distributions. We have carried out some systems-theoretic investigation in a distributional setting. We have chosen the distribution theory on groups as developed by Bruhat [1], Maurin [7] and Kac [5]. Properties such as continuity, regularity, translation-invariance, causality, passivity and semipassivity are imposed on the linear mappings from $D(G)$ into $D'(G)$ and their effects are studied. The concept of causality on abelian and locally compact groups was first introduced by Freedman and Falb [3] and this concept of semipassivity and passivity for linear mappings on locally compact and abelian groups was first considered by Hackenbroch [4].

Representations of continuous linear mappings from $D(G)$ into $D'(H)$ which are regular into $C(H)$, of linear continuous and translation-invariant mappings from $D(G)$ into $D'(G)$ and of linear and scatter-semipassive mappings from $D(G)$ into $D'(G)$, are obtained. We establish a one-to-one correspondence between contractions in $L_2(G)$ and linear scatter-semipassive mappings from $D(G)$ into $D'(G)$. We show that causality and semipassivity imply passivity in the scattering formalism and passivity implies causality in the immittance formalism. A characterization of linear, scatter-semipassive

* Received by the editors July 29, 1977, and in revised form March 22, 1978. The material presented here is contained in the author's doctoral dissertation written at the State University of New York at Stony Brook. This work was supported by the National Science Foundation under Grant PO 33568-X00.

† Department of Mathematics, Birla Institute of Technology and Science, Pilani, Rajasthan, India.

and real mappings in terms of a positivity criterion is established. These results are generalizations of the results when $G = R^n$, which come out as special cases of our results on G, a locally compact, separable and abelian group. Some of the results are generalizations of the results of R. Meidan [9]. In the next section, we give a few pertinent definitions from the distribution theory on groups as developed by Bruhat [1], Maurin [7] and Kac [5].

2. Distribution theory on groups. Let G be a locally compact group and let $\mathcal{H}$ be a family of compact subgroups K of G such that the quotient group G/K is a Lie group. We assume that the intersection of the subgroups belonging to $\mathcal{H}$ is the subgroup $\{0\}$. This is the case if the quotient group G/G_0, where G_0 is the connected component of 0, is compact. Furthermore, if G is metrizable, there exists a decreasing sequence $\{K_n\}$ of such compact subgroups such that $\bigcap_n K_n = \{0\}$. The group is then canonically isomorphic to the projective limit of G/K_n. We assume G to be separable i.e., metrizable and countable at infinity. Maurin [7] has considered Yamabe groups (or a projective limit of Lie groups) which are locally compact and second countable.

Let K be a closed subgroup of G. We designate by I_k the canonical map of G onto the homogeneous space G/K, the left quotient group i.e., $G \ni g \to I_k g = g + K \in G/K$. If E is any space, the mapping $J_k : [f] \to [f] \circ I_k$ allows us to identify a mapping $[f]$ on G/K into E with a mapping f of G into E invariant to the right by K.

For $K \in \mathcal{H}$, the space $D(G/K)$ of indefinitely differentiable functions with compact support on the Lie group G/K is identified through J_k with a subspace $D_K(G)$ of the space $C_0(G)$ (the space of continuous functions with compact support on G). We always consider $D_K(G)$ as endowed with the topology that is carried by $D(G/K)$. This topology is finer than that induced by $C_0(G)$. For K, $H \in \mathcal{H}$ with $K \subset H$, one has $D_H(G) \subset D_K(G)$ and the topology of $D_H(G)$ is induced by that of $D_K(G)$, since $D(G/H)$ is identified as a topological subspace of $D(G/K)$.

DEFINITION 2.1. We denote by $D(G)$ the subspace of $C_0(G)$ which is $\bigcup_{K \in \mathcal{H}} D_K(G)$ equipped with the inductive limit topology. $f \in D(G)$ is called a regular function with compact support.

DEFINITION 2.2. We call a function f on G, of arbitrary support, *regular* if for every $x \in G$, there exists a neighborhood U of x and a function $\varphi \in D(G)$ such that $f(y) = \varphi(y)$ for all $y \in U$, that is, f is regular if and only if $\varphi f \in D(G)$ for all $\varphi \in D(G)$ or f is regular if and only if for all compact K of G, there exists one $\varphi \in D(G)$ such that $f(y) = \varphi(y)$ for all $y \in K$.

We designate by $E(G)$ the space of regular functions endowed with the coarsest locally convex topology such that for every $\varphi \in D(G)$, the mapping $E(G) \ni f \to \varphi f \in D(G)$ (i.e., the subspaces of $D(G)$) is continuous, that is, $E(G)$ is the projective limit of the subspaces of $D(G)$ under these mappings.

DEFINITION 2.3. We refer to the continuous linear functionals on the space $D(G)$ as *distributions*. We designate by $D'(G)$ the space of distributions on G, the dual of $D(G)$, endowed with the strong (weak) topology of the dual. The space $E'(G)$, the strong dual of $E(G)$, is identified with the space of distributions of compact support.

For the definitions of the spaces $D(G)$ and $E(G)$ when G is any locally compact group (i.e., when G/G_0 is not compact), please refer to Bruhat [1] or Maurin [7b].

3. Translation-varying linear operators. Here we study translation-varying linear mappings or operators which are continuous from $D(G)$ into $D'(H)$ where G and H are locally compact topological groups which are metrizable and countable at infinity. Meidan [9] has studied these mappings in a distributional setting where distributions are defined on R^n. Such a study of operators cannot be carried over to distribution theory as developed by Riss [11]. But such extension is possible to distribution theory as developed by Bruhat [1], Maurin [7] and Kac [5], thanks to the existence of Bruhat's kernel theorem [1].

DEFINITION 3.1. $C(H)$ is the space of continuous functions on H, endowed with the topology of compact convergence. It is a Fréchet space.

$C_0(H)$ is the space of continuous functions with compact support on H, endowed with the strict inductive limit topology. It is an LF-space.

$D(H)$ is dense in $C(H)$ because $D(H)$ is dense in $C_0(H)$ with the uniform topology and $C_0(H)$ is dense in $C(H)$. The space $C_0(H)$ is not complete under the topology of uniform convergence. Its completion is $C_{00}(H)$, the space of continuous functions vanishing at infinity.

DEFINITION 3.2. We call an operator from $D(G)$ into $D'(H)$ *regular* if its range is contained in either $C(H)$, $C_{00}(H)$, $D(H)$ or $L_2(H)$.

The following theorems show that by assuming regularity, nice characterizations of operators are obtained. Here $D'(H)$ is the weak dual of $D(H)$.

THEOREM 3.1. *Let L be a continuous linear operator from $D(G)$ into $D'(H)$. If the range of L is contained in $C(H)$, then L is continuous from $D(G)$ into $C(H)$.*

Proof. We prove this by using the closed graph theorem (Trèves [16, pp. 173]). Since the domain $D(G)$ of L is an inductive limit space, we prove the continuity of L when L is restricted to $D_j(G)$. Here the spaces $D_j(G)$ and $C(H)$ are Fréchet spaces. Let $\{\varphi_i\}$ be a sequence converging in $D_j(G)$ to φ, and such that the sequence $\{L\varphi_i\}$ converges in $C(H)$ to some ψ. We have to show that $\psi = L\varphi$ in order to prove that the graph is closed. But $\{L\varphi_i\}$ converges to ψ also in the topology of $D'(H)$ (which is weaker than that of $C(H)$). On the other hand, due to the continuity of L as an operator from $D(G)$ into $D'(H)$, $\{L\varphi_i\}$ converges in $D'(H)$ to $L\varphi$. Since $D'(H)$ is Hausdorff, the limit is unique and hence $\psi = L\varphi$.

The above theorem is also true when we replace $C(H)$ by $C_{00}(H)$ or $L_2(H)$. Raikov [10] has shown, in his "double closed graph theorem," that the closed graph theorem is applicable for operators whose range is contained in an inductive limit of Fréchet spaces. Hence we have the following theorem.

THEOREM 3.2. *Let L be a continuous linear operator from $D(G)$ into $D'(H)$. If the range of L is contained in $C_0(H)$ ($D(H)$), then L is continuous from $D(G)$ into $C_0(H)$ ($D(H)$).*

THEOREM 3.3. *Let $K_y(x)$ be a family of distributions in $D'(G)$ depending on the variable $y \in H$. If $y \to K_y(x)$ is a continuous function, then the operator L, which is defined by*

$$\psi(y) = (L\varphi)(y) = \langle K_y(x), \varphi(x)\rangle \tag{3.1}$$

where $\varphi(x) \in D(G)$, is a continuous linear mapping from $D(G)$ into $D'(H)$.

Proof. Linearity of the operator L is clear. We prove the continuity of the operator from $D(G)$ into $D'(H)$ with the help of the Banach–Steinhaus theorem (Trèves [16, p. 349]). Let $\{\varphi_i\}$ be a sequence converging to zero in $D(G)$. The sequence $\{\psi_i(y)\}$ converges to zero pointwise in H.

Let A denote the following subset of $D'(G): A = \{K_y(x): y \in B\}$ where B is a compact subset of H. For each $\varphi \in D(G)$, the function $\psi(y) = \langle K_y(x), \varphi(x)\rangle$ is, by assumption, continuous on H. Since B is a compact subset of H, $\psi(y)$ is bounded on B, for each $\varphi \in D(G)$. This means that the set A is weakly bounded in $D'(G)$. By the Banach–Steinhaus theorem, A is strongly bounded in $D'(G)$. The sequence $\{\varphi_i\}$ is bounded in $D(G)$. Hence the strong boundedness of A implies that $\sup_{i;y\in B} |\langle K_y(x), \varphi_i(x)\rangle| < \infty$. By applying Lebesgue's dominated convergence theorem, the weak convergence of $\{\psi_i\}$ with respect to regular functions of compact support follows.

COROLLARY. *Let $K_y(x)$ be a family of distributions in $D'(G)$ depending on the variable $y \in H$ such that $\psi_\varphi(y) = \langle K_y(x), \varphi(x)\rangle$ belongs to $D(H)$. Then $\varphi \to \psi_\varphi(y)$ is a continuous linear mapping of $D(G)$ into $D(H)$* (Dolezal [2]).

Proof. By Theorem 3.3, $\varphi \to \psi_\varphi(y)$ is a continuous linear mapping from $D(G)$ into $D'(H)$. By Theorem 3.2, $\varphi \to \psi_\varphi(y)$ is a continuous linear mapping from $D(G)$ into $D(H)$.

THEOREM 3.4. *Let L be a continuous linear operator from $D(G)$ into $D'(H)$ which is regular into $C(H)$ $(D(H))$. Then a family of distributions $K_y(x)$ in $D'(G)$ exists such that $(L\varphi)(y) = \langle K_y(x), \varphi(x)\rangle$.*

Proof. By Theorem 3.1 (3.2), L is continuous from $D(G)$ into $C(H)$ $(D(H))$ where the latter is equipped with its initial topology. Hence, for every fixed $y \in H$, $(L\varphi)(y)$ is a continuous functional on $D(G)$. So a distribution $K_y(x)$ exists in $D'(G)$ such that $(L\varphi)(y) = \langle K_y(x), \varphi(x)\rangle$.

The concept of regularity is also associated with the property of the extendability of the domain of definition of an operator onto spaces of distributions.

THEOREM 3.5. *Let L^t be a continuous linear operator from $C'(H)$ into $D'(G)$ for the respective weak topologies. Then its transpose L is continuous from $D(G)$ into $C(H)$. Furthermore, L^t is also continuous from $C'(H)$ into $D'(G)$ when they carry their respective strong topologies.*

Proof. The transpose of L^t, L is continuous from $D(G)$ into $C(H)$ with respect to their respective weak topologies (Trèves [16, pp. 197, 199 and 200]). But L is also continuous with respect to Mackey topologies (Robertson and Robertson [12, pp. 62, Prop. 14]). For a barrelled or metrizable locally convex Hausdorff space, the initial topology is identical to the Mackey topology (Trèves [16, p. 372, Prop. 36.3]). Consequently L is continuous from $D(G)$ into $C(H)$ for their respective initial topologies. Now, the transpose of L is L^t and is a linear continuous operator from $C'(H)$ into $D'(G)$ when these spaces carry all the topologies usually considered on the dual (Trèves [16, pp. 199 and 200]).

THEOREM 3.6. *Let L be a continuous linear operator from $D(G)$ into $D'(H)$. L^t is continuously extendable onto $C'(H)$ if and only if L is regular into $C(H)$.*

Proof. This follows from Theorem 3.1 and Theorem 3.5.

THEOREM 3.7. *Let L be a continuous linear operator from $D(G)$ into $D'(H)$ whose range is contained in $C(H)$. Let $\delta(y-z)$ denote the shifted impulse function in $C'(H)$. Then the transpose L^t defines the family of distributions $K_z(x)$ by the shifted impulse response*

$$K_z(x) = L^t\delta(y-z). \tag{3.2}$$

The family of distributions $K_z(x)$ represents the operator L by

$$(L\varphi)(z) = \langle K_z(x), \varphi(x)\rangle, \qquad \varphi(x) \in D(G). \tag{3.3}$$

Proof. By the assumption of regularity of L, L^t is continuous from $C'(H)$ into $D'(G)$. Since the function $z \to \delta(y-z)$ is continuous from H into $C'(H)$ and L^t is continuous from $C'(H)$ into $D'(G)$, it follows that the composite function $z \to L^t\delta(y-z)$ is continuous from H into $D'(G)$. By Theorem 3.3, the family $K_z(x)$ of (3.2) defines a continuous linear operator from $D(G)$ into $C(H)$ by (3.3). This operator is in fact the original operator L. Indeed, $\langle \delta(y-z), L\varphi\rangle = (L\varphi)(z)$. But, by the definition of the transpose operator, we have

$$\langle \delta(y-z), L\varphi\rangle = \langle L^t\delta(y-z), \varphi(x)\rangle.$$

Therefore $(L\varphi)(z) = \langle K_z(x), \varphi(x)\rangle$.

THEOREM 3.8. *Let L be a linear continuous operator from $D(G)$ into $D'(H)$. Its transpose L^t is a continuous linear operator from $D(H)$ into $D'(G)$. L is extendible onto $C'(G)$ $(D'(G))$ if and only if L^t is regular into $C(G)$ $(D(G))$.*

Proof. This follows from Theorem 3.6.

4. The convolutional representation of a translation-invariant continuous linear mapping of $D(G)$ into $D'(G)$ $(C(G))$. In the following G is a locally compact, abelian and second countable group.

Here we make use of Bruhat's kernel theorem and a theorem due to Maurin and Gårding [7b] to derive the convolutional representation of a translation-invariant continuous linear mapping of $D(G)$ into $D'(G)$.

DEFINITION 4.1. The element $T \in D'(G\times G)$ and bilinear separately continuous forms B on $D(G)\times D(G)$ are called kernels on G. A kernel $B(\cdot\,,\cdot)$ is left-invariant if

$$B(L_g\varphi, L_g\psi) = B(\varphi, \psi) \text{ for every } g \in G \text{ and } \varphi, \psi \in D(G),$$

where $L_g f(x) = f(-g+x)$. We denote $R_g f(x) = f(x+g)$. Since we are considering an abelian group, we have adopted the following notation.

$$\sigma_g f(x) = f(x-g) = L_g f(x) = R_{-g} f(x).$$

THEOREM 4.1 (Maurin and Gårding [7b]). *Let B be a left-invariant kernel on a Yamabe group G (not necessarily abelian). Then there exist unique distributions T, $S \in D'(G)$ such that $B(\varphi, \psi) = \langle S, \varphi^* * \psi\rangle = \langle T, \psi^* * \phi\rangle$ for every φ, $\psi \in D'(G)$, where $\psi^*(g) = \psi(-g)\Delta(-g)$ and*

$$(f*\varphi)(x) = \int f(y)\varphi(-y+x)\,dy = \int f(x+y)\varphi(-y)\,dy$$
$$= \int \Delta(-y)f(-y)\varphi(y+x)\,dy.$$

DEFINITION 4.2. A linear mapping $F: D(G) \to D'(G)$ is called *left-invariant* if $\langle FL_g\varphi, L_g\psi\rangle = \langle F\varphi, \psi\rangle$, φ, $\psi \in D(G)$; $g \in G$, i.e., if F commutes with L_g $(L_gF = FL_g)$ for every $g \in G$.

THEOREM 4.2. *F is a continuous linear and translation-invariant mapping of $D(G)$ into $D'(G) \Leftrightarrow F = T*$, $T \in D'(G)$.*

Proof. Put $B(\varphi, \psi) = \langle F\psi, \varphi\rangle$. Then B is a uniquely defined separately continuous and bilinear mapping of $D(G)\times D(G)$ into $\mathbb{C}$ (the field of complex numbers), which is also (left) translation-invariant. By Theorem 4.1 $\langle F\psi, \varphi\rangle = B(\varphi, \psi) = \langle T, \psi^* * \varphi\rangle$ where $T \in D'(G)$. Now a simple observation gives us the convolutional representation for the operator F. We have

$$(\psi^* * \varphi)(x) = \int_G \psi(y)\varphi(y+x)\,dy$$

since G is abelian ($\Delta(x) = 1$ for every $x \in G$). Therefore

$$\langle F\psi, \varphi\rangle = \langle T(x), [\psi^* * \varphi](x)\rangle = \langle T(x), \langle \psi(y), \varphi(x+y)\rangle\rangle$$
$$= \langle T * \psi, \varphi\rangle.$$

Since this holds for every $\varphi \in D(G)$, we have $F\psi = T * \psi$, $\psi \in D(G)$.

Now let $F\varphi = T * \varphi$ for all $\varphi \in D(G)$, where $T \in D'(G)$. Linearity and translation-invariance are clear. F is also continuous as a mapping from $D(G)$ into $D'(G)$. We rewrite

$$(F\varphi)(x) = (T * \varphi)(x) = \langle \sigma_x \check{T}(t), \varphi(t) \rangle = \langle K_x(t), \varphi(t) \rangle$$

where $K_x(t) = \sigma_x \check{T}(t)$. The proof follows from the Banach–Steinhaus theorem and the Lebesgue dominated convergence theorem (proof is exactly as that of Theorem 3.7).

In the following we derive the convolutional representation of a continuous linear mapping from $D(G)$ into $D'(G)$ which is regular into $C(G)$ as a special case of Theorem 3.7. We shall make use of this simple lemma.

LEMMA 4.1. *Let L be a continuous linear mapping from $D(G)$ into $D'(G)$. L is translation-invariant if and only if L^t is.*

THEOREM 4.3. *Let L be a continuous linear and translation-invariant mapping from $D(G)$ into $D'(G)$ which is regular into $C(G)$. Then L has a convolutional representation $L = K_0 *$ where K_0 is a distribution in $D'(G)$ and L is regular into $E(G)$.*

Proof. From Theorem 3.7, we have $K_z(x) = L^t \delta(y - z) = L^t \sigma_z \delta(y) = \sigma_z L^t \delta(y)$ by Lemma 4.1. Therefore $K_z(x) = \sigma_z \check{K}_0(x)$ where $\check{K}_0(x) = L^t \delta(y)$. Now

$$\begin{aligned}(L\varphi)(z) &= \langle K_z(x), \varphi(x) \rangle = \langle \sigma_z \check{K}_0(x), \varphi(x) \rangle \\ &= \langle K_0(x), \varphi(z - x) \rangle = (K_0 * \varphi)(z).\end{aligned}$$

This gives us the required convolutional representation. Since $(K_0 * \varphi)(z) \in E(G)$, L is regular into $E(G)$.

5. Immittance formalism. In this section and in the next section G is a locally compact abelian group which is also separable.

DEFINITION 5.1. A linear mapping $L: D(G) \to L_1(G)$ which satisfies the condition $\operatorname{Re} \int_G (L\varphi)(t)\bar{\varphi}(t)\, dt \geqq 0$ for all $\varphi \in D(G)$ is called *semipassive*.

Let P be a closed semigroup of positive Haar measure in the locally compact abelian group G. Let

$$E^{t_0}\varphi(t) = \begin{cases} \varphi(t) & \text{if } t \in t_0 - P, \\ 0 & \text{if } t \notin t_0 - P. \end{cases}$$

DEFINITION 5.2. A linear mapping $L: D(G) \to L_1(G)$ which satisfies the condition $\operatorname{Re} \int_{t_0 - P} (L\varphi)(t)\bar{\varphi}(t)\, dt \geqq 0$ for every $t_0 \in G$ and $\varphi \in D(G)$ is called *passive* with respect to the semigroup $P \subset G$.

DEFINITION 5.3. A linear mapping $L: D(G) \to L_1(G)$ is called *causal* with respect to P if for every $t_0 \in G$, $\varphi(t) = 0$ for all $t \in t_0 - P \Rightarrow (L\varphi)(t) = 0$ for all $t \in t_0 - P$. A similar definition holds for causal linear mappings from $L_2(G)$ into $L_2(G)$, i.e., $E^{t_0}L\varphi = E^{t_0}LE^{t_0}\varphi$ for all $t_0 \in G$ and for all $\varphi \in L_2(G)$.

THEOREM 5.1. *If L is a linear passive mapping from $D(G)$ into $L_1(G)$, then L is causal.*

Proof. To prove causality, let $\varphi(t) = 0$ for all $t \in t_0 - P$. We want to conclude that $(L\varphi)(t) = 0$ for all $t \in t_0 - P$. Let $\theta(t) = \alpha\varphi(t) + \varphi_1(t)$, where φ and φ_1 belong to $D(G)$, $\alpha \in \mathbb{C}$ and $\varphi(t) = 0$ for all $t \in t_0 - P$. By passivity, we have $\operatorname{Re} \int_{t_0 - P} (L\theta)(t)\bar{\theta}(t)\, dt \geqq 0$, i.e., $\operatorname{Re} \int_{t_0 - P} [\alpha\psi(t) + \psi_1(t)][\bar{\alpha}\bar{\varphi}(t) + \bar{\varphi}_1(t)]\, dt \geqq 0$ where $\psi(t) = (L\varphi)(t)$ and $\psi_1(t) = (L\varphi_1)(t)$. As α is arbitrary, this implies that $\int_{t_0 - P} \psi(t)\bar{\varphi}_1(t)\, dt = 0$ for all $\varphi_1 \in D(G)$ where $\varphi \in L_1(G)$. We want to prove that $\psi(t) = 0$ almost everywhere on $t_0 - P$. Let $f(t) = (E^{t_0}\psi)(t)$. We have $\int_G f(t)\bar{\varphi}_1(t)\, dt = 0$ for all $\varphi_1 \in D(G)$. Also $\int_G f(t)\bar{\varphi}_1(x - t)\, dt = 0$ for all $x \in G$ and $\varphi_1 \in D(G)$, i.e., $\int_G f(x - t)\bar{\varphi}_1(t)\, dt = 0$ for all $x \in G$ and $\varphi_1 \in D(G)$. Now by

Theorem 1.1.5 ([13]) and Proposition 2(a) of Bruhat [1] we choose a φ_1 with support contained in V, a neighborhood of 0, and satisfying the condition $\int_G \varphi_1(t)\,dt = 1$. Now

$$\int_G f(x-t)\varphi_1(t) - f(x) = \int_G [f(x-t)-f(x)]\varphi_1(t)\,dt.$$

So

$$\|f\|_1 = \|f*\varphi_1 - f\|_1 \leqq \int_G |\varphi_1(t)|\,dt \int_G |f(x-t)-f(x)|\,dx$$

$$= \int_V \|f-f_t\|_1 \varphi_1(t)\,dt < \varepsilon.$$

But, $\|f\|_1 < \varepsilon \Rightarrow f(t) = 0$ almost everywhere, i.e., $\psi(t) = 0$ almost everywhere on $t_0 - P$.

6. Scattering formalism. We now investigate the effect of certain energy constraints on linear mappings from $D(G)$ into $D'(G)$. These energy constraints are the concept of semipassivity and passivity in the framework of the scattering formalism. We restrict the range of the operator to $L_2(G)$. By Theorem 3.1, the operator is continuous from $D(G)$ into $L_2(G)$.

DEFINITION 6.1. A linear mapping L from $D(G)$ into $D'(G)$ is said to *scatter-semipassive* if

(i) range of L is contained in $L_2(G)$ and

(ii) $\|\varphi\|_2^2 - \|L\varphi\|_2^2 \geqq 0$ for every $\varphi \in D(G)$, where $\|\cdot\|$ denotes the norm in $L_2(G)$.

The linear operator L is said to be lossless if equality holds in the above definition (condition (ii)).

DEFINITION 6.2. The linear mapping L is said to be *scatter-passive* with respect to P if

(i) the range of L is contained in $L_2(G)$, and

(ii) for each $t_0 \in G$ and for every $\varphi \in D(G)$, $\int_{t_0-P} [|\varphi|^2 - |L\varphi|^2]\,dt \geqq 0$.

DEFINITION 6.3. A linear mapping $L: D(G) \to D'(G)$ is called *causal with respect to P* if for every $t_0 \in G$, and for every $\varphi \in D(G)$, the support of $L\varphi (\in D'(G))$ is contained in $(t_0 - P)'$ whenever the support of φ is contained in $(t_0 - P)'$.

THEOREM 6.1. *Let L be a linear scatter-semipassive mapping from $D(G)$ into $D'(G)$. Then L is continuous from $D(G)$ into $D'(G)$.*

Proof. It follows directly from the scatter-semipassivity and from the facts that $D(G) \subset L_2(G) \subset D'(G)$ and that the initial topologies are stronger than the induced topologies.

LEMMA 6.1. *Let L be a scatter-semipassive linear mapping from $D(G)$ into $D'(G)$. Then, it is uniquely extendable to the space $L_2(G)$. The extended operator is a contraction in $L_2(G)$.*

Proof. Due to scatter-semipassivity, L is continuous from $D(G)$ into $L_2(G)$, when the domain is endowed with the topology induced by $L_2(G)$. Since $D(G)$ is dense in $L_2(G)$, it follows that a unique continuous operator from L_2 into $L_2(G)$ exists, such that its restriction to $D(G)$ coincides with the given operator. From scatter-semipassivity, it follows that the extended operator is a contraction.

LEMMA 6.2. *Let L be a contraction in $L_2(G)$. Then L, when restricted to $D(G)$, is a scatter-semipassive linear mapping from $D(G)$ into $D'(G)$.*

Proof. As L is continuous with respect to the $L_2(G)$ topology, it must be also continuous with respect to the stronger Bruhat–Schwartz topology on the domain $D(G)$ and the weaker $D'(G)$ topology on the range.

From Lemma 6.1 and Lemma 6.2 we have the following theorem.

THEOREM 6.2. *There exists a one-to-one correspondence between linear scatter-semipassive mappings from $D(G)$ into $D'(G)$ and the contractions in $L_2(G)$.*

LEMMA 6.3. *Let L be a linear mapping from $L_2(G)$ into $L_2(G)$. If L is causal and scatter-semipassive on $L_2(G)$, then it is scatter-passive on $L_2(G)$.*

Proof. Assume that L is not scatter-passive. Then there exists a $t_0 \in G$ and a $\varphi \in L_2(G)$ such that $\int_{t_0-P}[|\varphi|^2-|L\varphi|^2]\,dt<0$. Let $\psi(x)=(E^{t_0}\varphi)(x)$. Consider the expression $\int_G[|\psi|^2-|L\psi|^2]\,dt=\int_G|\psi|^2\,dt-\int_{t_0-P}|L\psi|^2\,dt-\int_{(t_0-P)'}|L\psi|^2\,dt$ where $(t_0-P)'$ denotes the complement of t_0-P. Therefore

$$\int_G[|\psi|^2-|L\psi|^2]\,dt=\int_{t_0-P}|\varphi|^2\,dt-\int_{t_0-P}|E^{t_0}LE^{t_0}\varphi|^2\,dt$$
$$-\int_{(t_0-P)'}|L\psi|^2\,dt.$$

Since for linear operators, causality is equivalent to the condition $E^{t_0}LE^{t_0}=E^{t_0}L$, we have $\int_G[|\psi|^2-|L\psi|^2]\,dt=\int_{t_0-P}[|\varphi|^2-|L\varphi|^2]\,dt-\int_{(t_0-P)'}|L\psi|^2\,dt=J_1+J_2<0$ as $J_1<0$ by the above assumption and $J_2<0$, i.e., L is not scatter-semipassive. This is a contradiction. Hence L is scatter-passive.

Theorem 6.2 and Lemma 6.3 will be used now to prove that scatter-semipassivity and causality imply scatter-passivity in the case of linear mappings from $D(G)$ into $D'(G)$. When $G=R^n$, Meidan [9] used a theorem proved by Saeks [14] in a Hilbert resolution space to prove the above statement. However Saeks' theorem is not applicable here and we use Lemma 6.3 instead to arrive at the result.

THEOREM 6.3. *Let L be a scatter-semipassive linear mapping from $D(G)$ into $D'(G)$. If L is causal on $D(G)$, then it is also scatter-passive on $D(G)$.*

Proof. From Lemma 6.1, a linear scatter-semipassive mapping is uniquely extendable onto $L_2(G)$ along with the preservation of the scatter-semipassivity constraint on the extended mapping. We will show now that causality is also preserved for this extended mapping. Let t_0 be any element in G. L is causal on $L_2(G)$ if and only if, for every $\varphi \in L_2(G)$ with support contained in $(t_0-P)'$, the support of $L\varphi$ is also contained in $(t_0-P)'$. But this follows from the causality of L on $D(G)$ and from the continuity of L with respect to the $L_2(G)$-topology. Indeed, let φ be any function in $L_2(G)$ whose support is contained in $(t_0-P)'$. Then a sequence $\{\varphi_n\}$ exists in $D(G)$ with support $\varphi_n \subset (t_0-P)'$, which converges to φ in the $L_2(G)$-topology. By the causality of L on $D(G)$, $L\varphi_n=0$ on t_0-P for all n and by the continuity of L, $L\varphi=0$ almost everywhere on t_0-P. From Lemma 6.3, it follows that the mapping is scatter-passive.

7. A representation theorem for linear scatter-semipassive mappings from $D(G)$ into $D'(G)$. Here G is a locally compact and separable group.

THEOREM 7.1. *Let L be a linear scatter-semipassive mapping from $D(G)$ into $D'(G)$. Then, there exists a family of distributions $K_t(z)$ in $D'(G)$ where $z, t \in G$ such that*

$$(L\varphi)(t)=\langle K_t(z), \varphi(z)\rangle \quad \textit{for almost all } t \in G.$$

Proof. The continuity of L as a mapping from $D(G)$ into $L_2(G)$ follows from the linearity and scatter-semipassivity of L. Since the space $D(G)$ is nuclear and barrelled and $L_2(G)$ is a Banach space (hence quasicomplete), the mapping L is nuclear (Trèves [16, p. 511, Thm. 50.1]) and

$$(L\varphi)(t)=\sum_k \lambda_k\langle x'_k, \varphi\rangle y_k(t)$$

where the sequence $\{x'_k\}$ is bounded in $D'(G)$, the sequence $\{y_k\}$ is bounded in $L_2(G)$ and $\{\lambda_k\}$ is a complex sequence with $\sum_k |\lambda_k| < \infty$ (Trèves [16, pp. 481 and 482, Prop. 47.2, Cor. 1]). Let $P_n(t) = \sum_{k=1}^n \lambda_k y_k(t)$. Since the sequence $\{P_n\}$ converges in $L_2(G)$, there exists a subsequence $\{P_{n_i}\}$ converging absolutely almost everywhere. Let t be one such point and let $\bar{P}_{n_i} = \sum_{k=1}^{n_i} \lambda_k y_k(t) x'_k$. Then the sequence of distributions $\{\bar{P}_{n_i}\}$ converges weakly to a distribution in $D'(G)$ as $\{\langle \bar{P}_{n_i}, \varphi\rangle\}$ converges as a sequence of numbers in $\mathbb{C}$. Let us denote this distribution by the symbol K_t. Then

$$(L\varphi)(t) = \langle K_t(z), \varphi(z)\rangle$$

for almost all $t \in G$.

Incidentally this gives an alternative proof of the theorem proved by Meidan [9], where the spaces D and D' are the spaces in Schwartz's distribution theory on R^n.

The above theorem also gives a kernel representation of a linear bounded mapping from $L_2(G)$ into $L_2(G)$ (where the domain is restricted to $D(G)$).

8. Realizability condition for translation-varying mappings from $D(G)$ into $D'(G)$. The Bruhat's kernel theorem [1] asserts that there is a one-to-one correspondence between bilinear separately continuous forms B on $D(G) \times D(G)$ and the distributions f_B on $G \times G$, i.e., to each bilinear separately continuous form B on $D(G) \times D(G)$ there exists a unique distribution $f \in D'(G \times G)$ such that

$$D(G) \times D(G) \ni (\varphi, \psi) \to B(\varphi, \psi) = \langle f, \varphi \otimes \psi\rangle. \tag{8.1}$$

Clearly, every distribution $f \in D'(G \times G)$ defines a bilinear separately continuous form by the equation (8.1). The right-hand side of the equation can be used to define a composition operator as follows:

Given $f \in D'(G \times G)$ and any $\psi \in D(G)$, we define $f \cdot \psi$ as a mapping on any $\varphi \in D(G)$ by

$$\langle f \cdot \psi, \varphi\rangle = \langle f(t, x), \varphi(t)\psi(x)\rangle; \qquad x, t \in G.$$

Therefore $f \cdot \psi$ maps $D(G)$ into $\mathbb{C}$.

THEOREM 8.1. *Given $f \in D'(G \times G)$, $f \cdot$ is a continuous linear mapping of $D(G)$ into $D'(G)$.*

To get a converse, we need the following result.

LEMMA 8.1. *Let L be a continuous linear mapping of $D(G)$ into $D'(G)$. Define B from L by $B(\varphi, \psi) = \langle L\psi, \varphi\rangle$, φ, $\psi \in D(G)$. Then B is a uniquely defined separately continuous linear mapping of $D(G) \times D(G)$ into $\mathbb{C}$.*

THEOREM 8.2. *L is a continuous linear mapping of $D(G)$ into $D'(G)$ if and only if there exists an $f \in D'(G \times G)$ such that $L\psi = f \cdot \psi$ for all $D(G)$. Here f is uniquely defined by L and conversely.*

The Volterra product of kernels has been introduced by Schwartz [15] where $G = R^n$. Meidan [9] and Zemanian [18d] have considered the Volterra product in reference to scatter-semipassive kernels to obtain a scatter-semipassivity criterion in terms of the positivity of a kernel. We shall study these in a distributional setting where distributions are defined on topological groups, in a restricted form.

DEFINITION 8.1. Let f_1 and f_2 be two kernels in $D'(G \times G)$ and let L_1 and L_2 be the corresponding mappings from $D(G)$ into $D'(G)$, i.e., $L_1 = f_1 \cdot$ and $L_2 = f_2 \cdot$. Assume that the composition operator $L_2 L_1$ exists and is a linear continuous mapping from $D(G)$ into $D'(G)$ whose kernel is f. Then f is said to be the Volterra product of f_1 and f_2 and denoted by $f = f_2 \circ f_1$, $L_2 L_1 v = (f_2 \circ f_1) \cdot v$.

It may be necessary to consider the mapping L_2 in its extended form in order to make the definition meaningful. This is precisely the case for linear scatter-semipassive mappings. The following result is a consequence of the one-to-one correspondence between linear scatter-semipassive mappings and contractions in $L_2(G)$.

THEOREM 8.3. *The Volterra product turns the subset of scatter-semipassive kernels in $D'(G\times G)$ into a semigroup. This semigroup is equivalent to the semigroup of contractions in $L_2(G)$.*

DEFINITION 8.2. Let L be a linear continuous mapping from $D(G)$ into $D'(G)$. Then L^t denotes its transpose and is also a continuous linear mapping from $D(G)$ into $D'(G)$. If $f(x, y)$ denotes the kernel of L, then the kernel of L^t is $f(y, x)$.

The adjoint L^a of the operator L (when L is scatter-semipassive) is the continuous linear mapping associated with the kernel $\bar{f}(y, x)$, where the bar denotes the complex conjugate.

The following lemma establishes the existence of the Volterra product in a special case.

LEMMA 8.2. *Let L be a continuous linear mapping from $D(G)$ into $D'(G)$ such that its range is contained in $L_2(G)$ and f is the corresponding kernel. Then the Volterra product $f^t \circ f$ exists, where f^t denotes the kernel of the transpose operator L^t.*

DEFINITION 8.3. A kernel f in $D'(G\times G)$ is said to be *positive* if, for all $\varphi \in D(G)$, $\langle f\cdot \varphi, \bar{\varphi}\rangle \geqq 0$.

We give the following theorems without proof.

THEOREM 8.4. *Let L be a continuous linear mapping from $D(G)$ into $D'(G)$ which is regular into $L_2(G)$. Then, L is scatter-semipassive $\Rightarrow$ the kernel $(i-f^a\circ f)$ is positive where i denotes the kernel of the identity mapping I.*

The following is a converse to the above theorem in a restricted form.

THEOREM 8.5. *Let L be a continuous linear and real mapping from $D(G)$ into $D'(G)$ which is regular into $L_2(G)$. Then the kernel $i-f^t\circ f$ is positive $\Rightarrow L$ is scatter-semipassive.*

Acknowledgment. The author wishes to express his thanks to his adviser, Professor Armen H. Zemanian, for help and guidance. I would like to thank Professor Vaclav Dolezal for many helpful discussions. I am thankful to the referee for suggestions for the improvement of the paper.

REFERENCES

[1] F. BRUHAT, *Distribution sur un groupe localement compact et application a' l'etude des representations des groupes p-adiques*, Bull. Soc. Math. France, 89 (1961), pp. 43–75.
[2] V. DOLEZAL, *A representation of linear continuous operators on testing functions and distributions*, this Journal, 1 (1970), pp. 491–506.
[3] M. I. FREEDMAN AND P. L. FALB, *A generalized transform theory for causal operators*, SIAM J. Control, 7 (1969), pp. 452–471.
[4] W. HACKENBROCH, *Passivity and causality over locally compact abelian groups*, preprint.
[5] G. I. KAC, *Generalized functions on a locally compact group and decompositions of unitary representations*, Trudy. Moskov. Mat. Obšč., 10 (1961), pp. 3–40. (In Russian.)
[6] H. KONIG AND J. MEIXNER, *Lineare Systeme und Linear Transformationen*, Math. Nachr., 19 (1958), pp. 265–322.
[7a] K. MAURIN, *Distributionen auf Yamabe-Gruppen. Harmonische Analyse auf einer Abelschen l. k. Gruppe*, Bull. Acad. Pol. Sci. Ser. Sci. Math., 9 (1961), pp. 845–850.
[7b] ———, *Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups*, Polish Scientific Publishers, Warsaw, 1968.
[8] B. MCMILLAN, *Introduction to formal realizability theory*, Bell System Tech. J., 31 (1952), pp. 217–279 and 541–600.
[9] R. MEIDAN, *Translation varying linear operators*, SIAM J. Appl. Math., 22 (1972), pp. 419–436.

[10] D. A. RAIKOV, *Double closed-graph theorem for topological linear spaces*, Sibirsk. Mat. Z., 7, No. 2 (1966), pp. 353–372.

[11] J. RISS, *Elements de calculus differential et theorie des distributions sur les groupes abelians localement compacts*, Acta Math., 89 (1953), pp. 45–105.

[12] A. P. ROBERTSON AND W. J. ROBERTSON, *Topological Vector Spaces*, Cambridge University Press, London, 1966.

[13] W. RUDIN, *Fourier Analysis on Groups*, Wiley-Interscience, New York, 1962.

[14] R. SAEKS, *Causality in Hilbert Space*, SIAM Rev., 12 (1970), pp. 357–383.

[15] L. SCHWARTZ, *Théorie Des Noyaux*, Proceedings of the International Congress of Mathematicians (Cambridge, MA, 1950), pp. 220–230.

[16] F. TRÈVES, *Topological Vector Spaces, Distributions and Kernels*, Academic Press, New York, 1967.

[17] M. R. WOHLERS AND E. J. BELTRAMI, *Distribution theory as the basis of generalized passive network analysis*, IEEE Trans. Circuit Theory, CT-12 (1965), pp. 164–169.

[18a] A. H. ZEMANIAN, *An N-port realizability theory based on the theory of distributions*, Ibid., CT-10 (1963), pp. 265–274.

[18b] ———, *Distributional convolution and an application to network realizability theory*, Proc. Sixth Midwest Circuit Conf. (University of Wisconsin, Madison, May, 1963), pp. K1–K9.

[18c] ———, *Realizability theory of continuous linear operators*, Academic Press, New York, 1972.

[18d] ———, *Realizability conditions for time-varying and time-invariant Hilbert ports*, SIAM J. Appl. Math., 22 (1972), pp. 612–628.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0012$01.00/0

ANALYTIC SOLUTIONS OF ALGEBRAIC DIFFERENTIAL EQUATIONS*

ROGER CHALKLEY†

Abstract. For special polynomials $f_2(w)$, $f_1(w)$, $f_0(w)$ in w with analytic coefficients, the equation $f_2(w)w'^2+f_1(w)w'+f_0(w)=0$ has appeared many times in the literature. Frequently, the equation is irreducible, $\deg f_2=0$, $\deg f_1\leqq 2$, $\deg f_0\leqq 4$, and either $4f_2f_0-f_1^2$ has a multiple root or its degree is $\leqq 2$. Under these conditions, there is an algebraic transformation to simplify the equation. This paper motivates the transformation and illustrates its effectiveness in diverse situations.

1. Introduction. Let the coefficients $a_1(z), \cdots, a_9(z)$ of

$$(1)\qquad \begin{aligned} &a_1(z)w'^2+a_2(z)w^2w'+a_3(z)ww'+a_4(z)w' \\ &\qquad +a_5(z)w^4+a_6(z)w^3+a_7(z)w^2+a_8(z)w+a_9(z)=0 \end{aligned}$$

be analytic functions of a variable z on a region Ω of the complex plane such that $a_1(z)\not\equiv 0$ and set

$$\begin{aligned} F(z, w)\equiv &4a_1(z)(a_5(z)w^4+a_6(z)w^3+a_7(z)w^2+a_8(z)w+a_9(z)) \\ &-(a_2(z)w^2+a_3(z)w+a_4(z))^2. \end{aligned}$$

As in [4, § 13.2, p. 305], one can solve for w' and use a standard existence theorem to deduce the following result.

THEOREM. *Suppose z_* in Ω and w_* are complex numbers such that $a_1(z_*)\neq 0$ and $F(z_*, w_*)\neq 0$. Then, there exist distinct analytic solutions $\Phi_1(z)$ and $\Phi_2(z)$ of* (1) *on a neighborhood of z_* which satisfy $w_*=\Phi_1(z_*)$ and $w_*=\Phi_2(z_*)$. Moreover, if $\Phi(z)$ is any solution of* (1) *analytic at z_* for which $w_*=\Phi(z_*)$, then either $\Phi=\Phi_1$ or $\Phi=\Phi_2$ on some neighborhood of z_*.*

The indicated proof suggests a means to obtain information about $\Phi_1(z)$ and $\Phi_2(z)$. We proceed to motivate a condition on $F(z, w)$ as a basis for a more effective method to be given in §§ 2 and 3.

Ninety-three differential equations are listed with solution techniques or solutions in [5, pp. 355–372] and eighty-eight of these equations can be written in the form (1). The five exceptions are Nos. 373, 392, 398, 445, and 459. It is remarkable that: for eighty-seven of these eighty-eight equations, $F(z, w)$ has a representation

$$(2)\qquad F(z, w)\equiv(\varepsilon(z)w-\zeta(z))^2(\lambda(z)w^2+\mu(z)w+\nu(z))$$

with meromorphic coefficients on Ω such that $\varepsilon(z)w-\zeta(z)\not\equiv 0$. The exception is No. 372. Seventy-five of these eighty-seven equations also have

$$(3)\qquad (\mu(z))^2-4\lambda(z)\nu(z)\not\equiv 0,$$

and to each of them a transformation $w=(\alpha t^2+\beta)/(\gamma t^2+\delta)$ developed in [3, p. 463] is applicable to replace (1) by a simpler equation $R(t, t')=0$. As in [2, pp. 73–74], (3) ensures that the left member of (1) is not reducible. The nonsingular solutions can be specified explicitly for more than half of these seventy-five equations (including Nos. 400, 407, 433, 435, and 454) and, in each such case, $R(t, t')=0$ reduces to either a Riccati or Bernoulli or linear differential equation whose solutions are expressible in terms of elementary functions. For others where implicit solutions are indicated, $R(t, t')=0$ may be separable or homogeneous or etc. Typical situations are illustrated in Examples 1 through 5 of § 2.

* Received by the editors April 10, 1978.

† Department of Mathematics, University of Cincinnati, Cincinnati, Ohio 45221.

Henceforth, we suppose (1) satisfies (2) and (3). When the transformation from [3, p. 463] is applied to (1), it may not be possible to select α, β, γ, δ and the coefficients of $R(t, t')$ as meromorphic functions on Ω. However, when

$$(\mu(z_*))^2-4\lambda(z_*)\nu(z_*)\neq 0, \tag{4}$$

we shall show in § 3 that there is a neighborhood N_* of z_* relative to which the transformation can be applied so α, β, γ, δ and the ooefficients of $R(t, t')$ are analytic functions on N_*. Then, standard methods can be used to obtain local analytic solutions of $R(t, t')=0$ so that the transformation yields information about $\Phi_1(z)$ and $\Phi_2(z)$. Example 6 in § 3 illustrates this where $R(t, t')=0$ reduces to a Riccati equation without elementary solutions.

2. An algebraic transformation based on (2) and (3). The meromorphic functions defined on Ω form an ordinary differential field E_0. As in [2] and [3], we use algebraic terminology to avoid multiple-valued functions or Riemann surfaces. Let $\bar{E}_0$ be a differential field extension of E_0 which is also an algebraic closure of E_0. Then, there exist elements α, β, γ, δ in $\bar{E}_0$ which satisfy

$$\lambda w^2+\mu w+\nu\equiv(\gamma w-\alpha)(\delta w-\beta). \tag{5}$$

For example, we can select α, β, γ, δ so that

$$\lambda\alpha^2+\mu\alpha+\nu=0,\quad \beta=-\lambda\alpha-\mu,\quad \gamma=1,\quad \delta=\lambda. \tag{6}$$

We set $M=\alpha\delta-\beta\gamma$. Then, (5) and (3) give

$$M^2=(\alpha\delta+\beta\gamma)^2-4\alpha\beta\gamma\delta=\mu^2-4\lambda\nu\neq 0. \tag{7}$$

From [3, Thm. 2], the substitution $w=(\alpha t^2+\beta)/(\gamma t^2+\delta)$ relates the nonsingular solutions in $\bar{E}_0$ of (1) to the solutions in $\bar{E}_0$ of

$$4a_1Mtt'+c_4t^4+c_3t^3+c_2t^2+c_1t+c_0=0 \tag{8}$$

and $(c_3t^3+c_1t)(\gamma t^2+\delta)\neq 0$, where $c_3=M(\alpha\varepsilon-\gamma\zeta)$, $c_1=M(\beta\varepsilon-\delta\zeta)$,

$$c_4=2a_1(\alpha'\gamma-\alpha\gamma')+a_2\alpha^2+a_3\alpha\gamma+a_4\gamma^2,$$
$$c_0=2a_1(\beta'\delta-\beta\delta')+a_2\beta^2+a_3\beta\delta+a_4\delta^2,$$
$$c_2=2a_1(\alpha'\delta-\alpha\delta'+\beta'\gamma-\beta\gamma')+2a_2\alpha\beta+a_3(\alpha\delta+\beta\gamma)+2a_4\gamma\delta.$$

Singular solutions of (1), as we observed in [3, p. 461], are the solutions of both (1) and $F=0$.

Throughout, $R(t, t')$ denotes the left member of (8). To simplify computations, we may select α, β, γ, δ, ε, ζ, λ, μ, ν as any elements in $\bar{E}_0$ which satisfy (2), (3), and (5). In this section, z denotes an element of E_0 such that $z'=1$.

Example 1. We write No. 389 of [5, p. 358] as

$$w'^2-4ww'-w'+4w^2+w=0 \tag{9}$$

and find $F\equiv-(4w+1)$. This yields $R\equiv 16t(t'-t+1)$ when $\alpha=1$, $\beta=-1$, $\gamma=0$, $\delta=4$, $\varepsilon=0$, and $\zeta=-1$. The root $w_0=-\frac{1}{4}$ of $F=0$ is the only singular solution of (9). We set $t=1+u$ to replace $R=0$ by $u'-u=0$. The solutions $u_0=2Ce^z$ of $u'-u=0$ specify the nonsingular solutions

$$w_0=\frac{\alpha t_0^2+\beta}{\gamma t_0^2+\delta}=\frac{t_0^2-1}{4}=\frac{u_0^2+2u_0}{4}=C^2e^{2z}+Ce^z$$

in $\bar{E}_0$ of (9), where C is any complex number (i.e., any constant in $\bar{E}_0$).

Example 2. We write No. 397 of [5, p. 360] as

$$w'^2-2z^3w^2w'-4z^2w^3=0 \tag{10}$$

and find $F\equiv(2zw)^2w(-z^4w-4)$. This yields $R\equiv-16t(t'+2z)$ when $\alpha=0$, $\beta=4$, $\gamma=1$, $\delta=-z^4$, $\varepsilon=2z$, and $\zeta=0$. The roots $w_0=0$ and $w_0=-4/z^4$ of $F=0$ are the only singular solutions of (10). The nonzero solutions $t_0=-z^2-2C$ of $R=0$ and $\gamma t^2+\delta\neq 0$ yield the nonsingular solutions

$$w_0=\frac{\alpha t_0^2+\beta}{\gamma t_0^2+\delta}=\frac{1}{Cz^2+C^2}$$

in $\bar{E}_0$ of (10), where C is any nonzero constant.

Example 3. We write No. 456 of [5, p. 372] as

$$z(z^2-1)w'^2+2(1-z^2)ww'+zw^2-z=0 \tag{11}$$

and find $F\equiv 4(z^2-1)(w^2-z^2)$. This gives $R\equiv 4z\rho t(2z\rho t'-t^2-1)$ when $\alpha=z$, $\beta=-z$, $\gamma=1$, $\delta=1$, $\varepsilon=0$, $\zeta=2\rho$, and ρ is an element in $\bar{E}_0$ such that $\rho^2=z^2-1$. We deduce that $2\rho\rho'=2z$ and $\rho'=z/\rho$. The roots $w_0=z$ and $w_0=-z$ of $F=0$ are the only singular solutions of (11). For $i^2=-1$, the equation $iz\rho v'-v=0$ results from $R=0$ when $t=i+(1/u)$ and $u=(v-1)/(2i)$. Its nonzero solutions in $\bar{E}_0$ are $v_0=K(\rho+i)/z$, for each nonzero constant K. The nonsingular solutions in $\bar{E}_0$ of (11) are

$$w_0=\frac{z(t_0^2-1)}{(t_0^2+1)}=\frac{z(-2u_0^2+2iu_0+1)}{(2iu_0+1)}=\frac{z}{2}\left(v_0+\frac{1}{v_0}\right)=C_1\rho+C_2,$$

where $C_1=(K^2+1)/(2K)$, $C_2=i(K^2-1)/(2K)$, and $C_1^2+C_2^2=1$.

Example 4. We write No. 381 of [5, p. 357] as

$$w'^2-2zw'+w=0 \tag{12}$$

and find $F\equiv 4(w-z^2)$. This yields $R\equiv-2(2tt'-t-z)$ when $\alpha=-1$, $\beta=z^2$, $\gamma=0$, $\delta=1$, $\varepsilon=0$, and $\zeta=2$. There are no singular solutions of (12). Here, $R=0$ is homogeneous. Thus, an element w_0 in $\bar{E}_0$ is a solution of (12) if and only if there is an element t_0 in $\bar{E}_0$ such that

$$w_0=z^2-t_0^2 \quad\text{and}\quad (t_0-z)^2(2t_0+z)=C$$

for some constant C. In particular, the elements $t_0=z$ and $t_0=-z/2$ correspond to $C=0$ and specify solutions $w_0=0$ and $w_0=3z^2/4$ of (12). (The substitution $t_0=z-u_0$ yields the parametrization of [5, p. 357].)

Example 5. We write No. 443 of [5, p. 369] as

$$z^3w'^2-2z^2ww'-w'+zw^2=0 \tag{13}$$

and find $F\equiv-(4z^2w+1)$. This gives $R\equiv 8z^4(2ztt'-(t+1)(3t-1))$ when $\alpha=1$, $\beta=-1$, $\gamma=0$, $\delta=4z^2$, $\varepsilon=0$, and $\zeta=1$. There are no singular solutions of (13). Here, $R=0$ is separable. Thus, an element w_0 in $\bar{E}_0$ is a solution of (13) if and only if there is an element t_0 in $\bar{E}_0$ which satisfies

$$w_0=(t_0^2-1)/(4z^2) \quad\text{and}\quad (t_0+1)^3(3t_0-1)=Cz^6$$

for some constant C. If $C=0$, then $t_0=-1$ and $t_0=\frac{1}{3}$ specify solutions $w_0=0$ and $w_0=-2/(9z^2)$ of (13).

In general, the field of meromorphic functions defined on a region of the complex plane is the quotient field of the ring of analytic functions defined on that region. Thus, when (2) is satisfied by meromorphic coefficients on Ω, we can apply a corollary [6, p. 127] of the Gauss lemma to satisfy (2) with analytic functions ε, ζ, λ, μ, ν on Ω.

3. Analytic coefficients for (8). We suppose $a_1(z_*)\neq 0$, $F(z_*, w_*)\neq 0$, the coefficients of (2) are analytic on Ω, and (4) is satisfied. As in [1, pp. 143–144], there is an analytic function $D(z)$ on a neighborhood N_* of z_* such that

$$(D(z))^2=(\mu(z))^2-4\lambda(z)\nu(z)\neq 0, \quad \text{for each } z \text{ in } N_*. \tag{14}$$

This leads to meromorphic functions α, β, γ, δ on N_* which satisfy (5). Namely, when $\lambda\not\equiv 0$, set $\alpha=(-\mu+D)/(2\lambda)$ and use (6); or, when $\lambda\equiv 0$, set $\alpha=-\nu/\mu$ and use (6). By [6, p. 127], we can alter these definitions to obtain analytic functions α, β, γ, δ on N_* which satisfy (5). Then, the transformation of § 2 is applicable with $\Omega=N_*$ and yields analytic coefficients on N_* for (8). Moreover, (7) and (14) give $M(z)\neq 0$ for each z in N_*.

Next, we relate $\Phi_1(z)$ and $\Phi_2(z)$ to power series solutions of (8). We have $M(z_*)\neq 0$; and, $F(z_*, w_*)\neq 0$ yields $\gamma(z_*)w_*\neq\alpha(z_*)$ and $\delta(z_*)w_*\neq\beta(z_*)$. Thus, there is a nonzero complex number t_* such that

$$w_*=(\alpha(z_*)t_*^2+\beta(z_*))/(\gamma(z_*)t_*^2+\delta(z_*))$$

and $\gamma(z_*)t_*^2+\delta(z_*)\neq 0$. Let U be a neighborhood of z_* contained in N_* and let V be a neighborhood of t_* to which 0 does not belong such that $a_1(z)\neq 0$ and $\gamma(z)t^2+\delta(z)\neq 0$ for each z in U and each t in V. Then, we can solve for t' in (8) to express t' as an analytic function of z and t for z in U and t in V. As in [4, pp. 281–284], there exist power series $t=P_1(z)$ and $t=P_2(z)$ in $z-z_*$ which are solutions of (8) and converge on a neighborhood $|z-z_*|<r$ contained in U such that $t_*=P_1(z_*)$, $-t_*=P_2(z_*)$, and both $P_1(z)$ and $-P_2(z)$ map $|z-z_*|<r$ into V. This gives

$$\Phi_{\pi(s)}(z)=(\alpha(z)(P_s(z))^2+\beta(z))/(\gamma(z)(P_s(z))^2+\delta(z)),$$

where $\{\pi(1), \pi(2)\}=\{1, 2\}$, $s=1, 2$, and $|z-z_*|<r$. By substituting power series in $z-z_*$ for $\alpha(z)$, $\beta(z)$, $\gamma(z)$, and $\delta(z)$, we obtain power series in $z-z_*$ for $\Phi_1(z)$ and $\Phi_2(z)$ which converge on $|z-z_*|<r$.

Example 6. To illustrate this procedure for

$$w'^2+2w^2w'+w^4+(3zw+1)^2w(zw-1)=0, \tag{15}$$

$z_*=0$, and $w_*=1$, we find $F(z, w)\equiv(3zw+1)^2(4w)(zw-1)$ and select $\alpha(z)=0$, $\beta(z)=1$, $\gamma(z)=4$, $\delta(z)=z$, $\varepsilon(z)=-3z$, $\zeta(z)=1$, and $t_*=-\frac{1}{2}$. We solve (8) for t' to get $t'=t^2+z$. Setting $d_1=1$ and $d_n=0$ for $n\neq 1$, we verify that

$$b_{n+1}=\left(\left(\sum_{k=0}^{n} b_kb_{n-k}\right)+d_n\right)\Big/(n+1), \quad \text{for } n=0, 1, 2, \cdots, \tag{16}$$

when $P(z)=b_0+b_1z+b_2z^2+\cdots$ is a solution of $t'=t^2+z$. If $|b_0|=\frac{1}{2}$, then (16) yields $|b_n|<1$ for $n=0, 1, 2, \cdots$. Thus, for $b_0=-\frac{1}{2}$ or $b_0=\frac{1}{2}$, $P_1(z)$ and $P_2(z)$ are convergent on $|z|<1$. For $|z|<\frac{1}{6}$, we deduce

$$|4(P_s(z))^2+z|\geqq 1-2|z|-\sum_{n=2}^{\infty}4(n+1)|z|^n>\tfrac{1}{6},$$

$$\Phi_1(z)=1+0z+\tfrac{5}{4}z^2-\tfrac{7}{6}z^3+\tfrac{3}{2}z^4+\cdots,$$

and

$$\Phi_2(z)=1-2z+\tfrac{5}{4}z^2+\tfrac{5}{6}z^3-\tfrac{10}{3}z^4+\cdots.$$

Of course, one can solve (15) for w' and substitute power series directly for $\Phi_1(z)$ and $\Phi_2(z)$; but, it is difficult by such means to obtain information about convergence.

REFERENCES

[1] L. V. AHLFORS, *Complex Analysis*, 2nd ed., McGraw-Hill, New York, 1966.

[2] R. CHALKLEY, *Algebraic differential equations of the first order and the second degree*, J. Differential Equations, 19 (1975), pp. 70–79.

[3] ———, *A first-order algebraic differential equation*, Ibid., 26 (1977), pp. 458–466.

[4] E. L. INCE, *Ordinary Differential Equations*, Dover, New York, 1953.

[5] E. KAMKE, *Differentialgleichungen, Lösungsmethoden und Lösungen*, vol. 1, Chelsea, New York, 1948.

[6] S. LANG, *Algebra*, Addison-Wesley, Reading, MA, 1965.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0013$01.00/0

DECAY RATES FOR SOLUTIONS OF A CLASS OF DIFFERENTIAL-DIFFERENCE EQUATIONS*

FRED BRAUER†

Abstract. We analyze the dependence of the asymptotic behavior of solutions which tend to zero of the differential-difference equation $y'(t) = ay(t) + by(t-T)$ on the delay T. We show that if $b < 0$, the rate at which solutions tend to zero increases for small T, but decreases for larger T.

1. To study the asymptotic behavior of solutions of linear systems of ordinary differential equations with constant coefficients, it is sufficient to solve a polynomial equation whose roots—the eigenvalues of the coefficient matrix—determine the asymptotic behavior of solutions completely. For linear differential-difference equations with constant coefficients there is also a characteristic equation whose roots determine the asymptotic behavior of solutions. However, the characteristic equation is no longer a polynomial equation, and information on the location of its roots is more difficult to obtain. In addition, the characteristic equation depends on the delay as well as on the coefficients of the differential-difference equation.

The purpose of this paper is to analyze the dependence on the delay of the roots of the characteristic equation for a linear first order differential-difference equation with constant coefficients and a single delay term in the case that all roots of this characteristic equation have negative real parts. This study can be applied directly to the analysis of behavior of solutions of nonlinear first order differential-difference equations near a stable equilibrium, yielding results of interest in applications to some biological and control problems. We plan to explore some of these applications elsewhere.

2. We consider the first order linear differential-difference equation with constant coefficients

$$y'(t) = ay(t) + by(t-T). \tag{1}$$

We require $b \neq 0$ to assure an explicit dependence on the delay T. Explicit conditions on the coefficients a and b and the delay T which are equivalent to the asymptotic stability of the trivial solution of (1) have been given [1, pp. 118–120, 444–446]. We are interested in the asymptotic behavior of solutions of (1) when the trivial solution is asymptotically stable, especially in the rate of return to the equilibrium $y = 0$, and the dependence of this rate on T.

The characteristic equation corresponding to (1) is

$$(p-z)\,e^{z} + q = 0, \tag{2}$$

where

$$p = aT, \qquad q = bT. \tag{3}$$

If $\lambda < 0$ is the real part of a root of this characteristic equation, then there is a solution of (1) which decays like $e^{(\lambda/T)t}$ as $t \to \infty$. Asymptotic stability of the trivial solution of (1) is equivalent to negativity of the real part of every root of (2). We define the characteristic return time, or settling time, of a solution of (1) corresponding to a root of (2) with real part $\lambda < 0$ to be $-T/\lambda$. We define the characteristic return time for the (asymptotically stable) equilibrium $y = 0$ to be $-T/\lambda$, where λ is the largest real part of a root of (2).

* Received by the editors November 30, 1977, and in revised form June 2, 1978.
† Department of Mathematics, University of Wisconsin–Madison, Madison, Wisconsin 53706.

A necessary and sufficient condition that all roots of (2) be in the left half plane has been given.

THEOREM 1 (Hayes [2]). *The roots of* (2) *are all in the left half plane if and only if*

(4) $$p<1, \qquad p<-q<\frac{u_1}{\sin u_1},$$

where u_1 *is the root of the equation*

(5) $$u=p\tan u$$

in $0<u<\pi$, *with* $u_1=\pi/2$ *if* $p=0$.

We will always assume that the trivial solution of (1) is asymptotically stable, or equivalently, that $p=aT$ and $q=bT$ satisfy (4). The change of variable $z=\zeta-\sigma T$ transforms the characteristic equation (2) to

(6) $$[(p+\sigma T)-\zeta]\,e^{\zeta}+q\,e^{\sigma T}=0.$$

If we choose σ to be the supremum of positive real numbers for which the transformed characteristic equation (6) has all roots in the left half plane, then $\lambda=-\sigma T$, and the characteristic return time T_R for the equilibrium $y=0$ of (1) is $1/\sigma$. By Theorem 1, we must choose σ to be the supremum of positive numbers such that

(7) $$p+\sigma T=(a+\sigma)T<1,$$

(8) $$p+\sigma T=(a+\sigma)T<-q\,e^{\sigma T}=-bT\,e^{\sigma T},$$

(9) $$-q\,e^{\sigma T}=-bT\,e^{\sigma T}<\frac{u_2}{\sin u_2},$$

where u_2 is the root of

(10) $$u=(a+\sigma)T\tan u$$

in $0<u<\pi$, with $u_2=\pi/2$ if $a+\sigma=0$.

It is convenient to reformulate (8) and (9). The condition (8) is equivalent to

(11) $$(a+\sigma)T\,e^{-(a+\sigma)T}<-bT\,e^{-aT}.$$

The function

$$g(u)=\frac{u}{\sin u}\,e^{-u/\tan u}$$

is monotone increasing on $0\leqq u<\pi$, with $g(0)=1/e$, $g(\pi/2)=\pi/2$, $\lim_{u\to\pi}g(u)=\infty$. Because of (10), the condition (9) is equivalent to

(12) $$g(u_2)>-bT\,e^{-aT}.$$

It follows from (10) that u_2 decreases when σ increases. Thus an increase in σ produces a decrease in $g(u_2)$, and the maximum σ satisfying (12) is obtained by solving the corresponding equality. By an analogous argument, since an increase in σ produces an increase in the left side of both (7) and (11), the maximum σ satisfying each of these inequalities is also obtained by solving the corresponding equalities. If any of the three equalities corresponding to (7), (11), (12) has no solution for σ, then the corresponding inequality places no restriction on σ. We may summarize our results as follows:

THEOREM 2. *Let σ_1, σ_2, σ_3 be the solutions of*

$$(a+\sigma)T=1, \tag{13}$$

$$(a+\sigma)Te^{-(a+\sigma)T}=-bTe^{-aT}, \tag{14}$$

$$g(u_2)=-bTe^{-aT}, \qquad u_2=(a+\sigma)T\tan u_s, \tag{15}$$

respectively, with the understanding that $\sigma_i=\infty$ if the corresponding equation has no solution. Then the characteristic return time for the equilibrium $y=0$ of (1) *is the reciprocal of the smallest of σ_1, σ_1, σ_2.*

In order to determine the dependence of the characteristic return time on T, we must distinguish the cases $b>0$ and $b<0$.

Case I ($b>0$). If $b>0$, the only nonvacuous stability condition (4) is $a+b<0$ (which implies $a<0$), and there is no restriction on T. Since $g(u)>0$ for $0\leqq u<\pi$ and $-bTe^{-aT}<0$ for all T, there is no solution of (15). The equation (14) is equivalent to

$$a+\sigma=-be^{\sigma T} \tag{16}$$

and it is easy to see that $0\leqq\sigma_2<-a$. Since $\sigma_1=(1-aT)/T=1/T-a>-a$, the characteristic return time is $1/\sigma_2$. Implicit differentiation of (16) gives

$$\frac{d\sigma_2}{dT}=\frac{-be^{\sigma T}}{1+bTe^{\sigma T}}<0.$$

Thus σ_2 is a decreasing function of T, and the characteristic return time is an increasing function of T. We have established the following result.

THEOREM 3. *If $b>0$, $a+b<0$, the trivial solution of* (1) *is asymptotically stable for all T, and the characteristic return time for this equilibrium is a monotone increasing function of T.*

Case II ($b<0$). We begin with the assumption $a>0$; if $a\leqq 0$ the treatment is quite similar and will be outlined later. If $a>0$, $b<0$, all three of the stability conditions (4) are relevant. In particular, the delay T is restricted to the interval $0\leqq T<1/a$. The equation (13) gives $\sigma_1=(1-aT)/T$. Thus if we plot $1/\sigma_1$ against T, we see that the characteristic return time $T_R=1/\sigma$ lies above the hyperbola

$$T_R=\frac{T}{1-aT} \tag{17}$$

through the origin with a vertical asymptote at $T=1/a$.

Since the function xe^{-x} has maximum value $1/e$ for $x=1$, while $g(u)$ has minimum value $1/e$ for $u=0$, the equation (14) has no solution if $-bTe^{-aT}>1/e$, and (15) has no solution if $-bTe^{-aT}<1/e$. For $b<0$, the function $-bTe^{-aT}$ is positive and monotone increasing in T with maximum $-(b/a)(1/e)$, which is greater than $1/e$ by (4), attained at $T=1/a$. Thus there exists $T^*<1/a$ such that

$$-bT^*e^{-aT^*}=1/e, \tag{18}$$

(14) has no solution for $T>T^*$, and (15) has no solution for $0\leqq T<T^*$. For $0\leqq T\leqq T^*$, we regard σ_2 as a function of T given implicitly by (14). Differentiation of (14) with respect to T gives

$$\frac{d\sigma}{dT}=\frac{-b\sigma e^{\sigma T}}{1+bTe^{\sigma T}}=\frac{-b\sigma e^{\sigma T}}{1-(a+\sigma)T}.$$

For $T=0$, σ_1 does not exist and $\sigma_2=-(a+b)$. If $\sigma_2<\sigma_1$, then $(a+\sigma_2)T<1$, and

$d\sigma_2/dT>0$, while $d\sigma_1/dT<0$. For $T=T^*$ and $\sigma=\sigma_1$,

$$(a+\sigma)Te^{-(a+\sigma)T}=1/e=-bTe^{-aT}.$$

Thus $\sigma_1=\sigma_2$ for $T=T^*$, and $\sigma_2<\sigma_1$ for $0\leqq T<T^*$. If we plot $1/\sigma_2$ against T, we see that the characteristic return time $T_R=1/\sigma$ lies above the hyperbola (17) for $0\leqq T<T^*$ and decreases as T increases for $0\leqq T<T^*$.

For $T>T^*$, we examine (15). Since $-bTe^{-aT}$ increases as T increases and $g(u)$ is an increasing function of u, an increase in T produces an increase in u_2. Thus σ_3 decreases and the characteristic return time increases. Since $u_2>0$ for $T>T^*$, $u_2/\tan u_2<1$, and $(a+\sigma_3)T<1$. Thus $\sigma_3<\sigma_1$. If T is chosen so that the stability condition (4) is satisfied, then

$$aT<1,\qquad -bT<(u_1/\sin u_1).$$

Thus

$$g(u_1)=\frac{u_1}{\sin u_1}e^{-u_1/\tan u_1}>-bTe^{-u_1/\tan u_1}=-bTe^{-aT},$$

and $u_2<u_1$. Then

$$(a+\sigma_3)T=\frac{u_2}{\tan u_2}>\frac{u_1}{\tan u_1}=aT,$$

which implies $\sigma_3>0$. Thus for $T>T^*$, the characteristic return time is $1/\sigma_3$, an increasing function of T which remains finite so long as the equilibrium $y=0$ of (1) is asymptotically stable.

If $a\leqq 0$, $b<0$, the only relevant stability condition (4) is $-bT<u_1/\sin u_1$. The hyperbola (17) has a vertical asymptote for $T=1/a<0$ and a horizontal asymptote for $R=-1/a$ if $a<0$. If $a=0$, the hyperbola (17) is replaced by the straight line $T_R=T$. Except for these differences, the argument is identical to that for $a>0$. We have now established the following result.

THEOREM 4. *If $b<0$, the characteristic return time for the asymptotically stable equilibrium $y=0$ of* (1) *is a decreasing function of T given by* (14) *for $0\leqq T<T^*$, with T^* given by* (18), *and an increasing function of T given by* (15) *for $T>T^*$, remaining finite for all T for which the equilibrium is asymptotically stable.*

3. Consider the differential-difference equation

$$y'(t)=F[y(t),y(t-T)]. \tag{19}$$

Suppose that $F(0,0)=0$, so that $y=0$ is a solution. To study the stability of the equilibrium at $y=0$, we consider the linear approximation

$$y'(t)=F_1(0,0)y(t)+F_2(0,0)y(t-T), \tag{20}$$

where F_1 and F_2 are the partial derivatives of F with respect to the first and second variables respectively, assumed continuous. If the solution $y=0$ of (20) is asymptotocally stable, then the same is true of the solution $y=0$ of (19) [1, p. 336], and the results of § 2 yield information on the asymptotic behavior of solutions [1, pp. 355–359]. If the solution $y=0$ of (20) is unstable, it has been shown for the case $F_1(0,0)<0$, $F_2(0,0)<0$, in which instability arises if T is too large, that there is a globally asymptotically stable periodic annulus [4]. It is natural to conjecture that similar results hold in other unstable cases.

A special case of interest in various applications, either directly or in forms which can be reduced to it, is

$$y'(t)=F[y(t-T)], \tag{21}$$

where $F(0)=0$. If $F'(0)<0$, we let $r=-F'(0)>0$. Then if $rT<\pi/2$, the solution $y=0$ of (21) is asymptotically stable, while if $rT>\pi/2$ the solution $y=0$ is unstable but there is an asymptotically stable periodic solution [3]. We shall use the results of § 2 to analyze in more detail the case $rT<\pi/2$.

The linearized equation for (21) is

$$y'(t)=-ry(t-T), \tag{22}$$

where $r=-F'(0)>0$. Only one of the stability conditions (4) is relevant, namely $rT<u_1/\sin u_1$. Since $u_1=\pi/2$, this condition is $rT<\pi/2$. The return time equations become

$$\sigma T=1,$$

$$\sigma T e^{-\sigma T}=rT,$$

$$g(u_2)=rT,$$

respectively, and T^* is defined by

$$rT^*=1/e.$$

Instead of the hyperbola (17), the equation $\sigma T=1$ gives a lower bound of T for the characteristic return time. For $0\leqq rT\leqq 1/e$, the characteristic return time decreases as T increases to a minimum of $1/er$ when $rT=1/e$. For $T=0$, the characteristic return time is $1/r$. For $rT>1/e$, the characteristic return time increases monotonically as T increases. We may determine the delay T for which the characteristic return time is $1/r$, the same as for $T=0$, as follows. If $\sigma=r$, then since $u_2=\sigma T\tan u_2$, the equation $g(u_2)=rT$ becomes

$$\frac{u_2}{\sin u_2}e^{-u_2/\tan u_2}=\frac{u_2}{\tan u_2},$$

$$\log\cos u_2=-\frac{u_2}{\tan u_2}.$$

This may be solved numerically, to give $u_2=1.01125$, and then

$$rT=\frac{u_2}{\tan u_2}=0.63336.$$

Thus if $0\leqq rT<0.63336$, the characteristic return time is less than for $T=0$. It may be said that a delay in this range tends to stabilize the system, even though a larger delay tends to destroy stability.

4. All the results of this paper are for first order differential-difference equations. For systems, there are some special results on stability conditions, but these are strongly dependent on the specific form of the characteristic equation. The methods developed here can be applied to each of these specific forms to study the dependence of the characteristic return time on the delay, but it is unlikely that there is any general pattern to the results which can be obtained.

REFERENCES

[1] R. E. BELLMAN AND K. L. COOKE, *Differential-Difference Equations*, Academic Press, New York, 1963.

[2] N. D. HAYES, *Roots of the transcendental equation associated with a certain differential-difference equation*, J. London Math. Soc., 25 (1950), pp. 226–232.

[3] J. L. KAPLAN AND J. A. YORKE, *On the stability of a periodic solution of a differential-delay equation*, this Journal, 6 (1975), pp. 268–282.

[4] ———, *On the nonlinear differential delay equation* $x'(t) = -f(x(t), x(t-1))$, J. Differential Equations, 23 (1977), pp. 293–314.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0014$01.00/0

POSITIVE SOLUTIONS OF THE EQUATION $(m(t)x'(t))'+A(t)x(t)=0$*

ALLAN L. EDELSON†

Abstract. Topological methods are used to show the existence of positive solutions to the n-dimensional boundary value problem $(m(t)x')'+A(t)x=0$, $x'(0)=0=x(T)$, where $A(t)$ is an n by n symmetric matrix with continuous entries. Criteria are given for the nonexistence of solutions, and using results from the theory of positive operators, extremal properties of the solutions are proved. A comparison theorem is given, generalizing results of Reid for the case of second order equations. These results are applied to special fourth order nonselfadjoint scalar equations.

1. Introduction. This paper is concerned with the existence and general properties of positive solutions of the two point boundary value problem

$$(m(t)x'(t))'+A(t)x(t)=0, \tag{1.1}$$

$$x'(0)=x(T)=0, \tag{1.2}$$

where $x=(x_1,\cdots,x_n)$ and $A(t)=(a_{ij}(t))$ is an $n\times n$ symmetric matrix with coefficients continuous and positive on $[0,\infty)$. The real valued function $m(t)$ is also assumed continuous and positive on $[0,\infty)$. A solution to (1.1), (1.2) is positive if all coordinates are positive on $(0, T)$.

Section 2 will be devoted to a topological proof of the existence of positive solutions to (1.1), (1.2), and by utilization of results from the theory of positive operators, as in Schmitt–Smith [1], a proof that a positive solution realizes the given boundary condition in minimum time. In § 3 we will generalize comparison theorems originally proved by Reid [2] for second order equations. Finally, in § 4 we will give applications of these techniques to fourth order scalar equations.

The following notation will be used throughout: $x(t;x^0)$ is the solution to (1.1) satisfying the initial condition $x(0)=x^0$, $x'(0)=0$. I is the positive cone of R^n, i.e. $I=\{x\in R^n;\ x_i>0,\ 1\leqq i\leqq n\}$, and $\bar{I}$ is the topological closure of I. Since we are considering only nontrivial solutions to (1.1) we will normalize and assume $|x^0|=1$.

Regarding the form of the particular system under study, we note that if $\int^\infty dt/m(t)=\infty$, the transformation $\tau=\int^t ds/m(s)$ transforms (1.1) into a system of the form

$$x''+A(t)x=0. \tag{1.3}$$

Examples for which this cannot be accomplished will be given in § 4.

2. Existence and properties of positive solutions. For $x^0\in\bar{I}$, let $\tau(x^0)$ be the time at which the trajectory $x(t;x^0)$ first leaves I, if this is finite. Otherwise let $\tau(x^0)=\infty$.

LEMMA 2.1. *If $(m(x)u')'+a_{ii}(t)u=0$ is oscillatory for some i, then $\tau(x^0)<\infty$ for all $x^0\in\bar{I}$.*

Proof. Assume the condition satisfied for $i=k$, and that $x(t)\in I$ for all $t>0$. Then $(m(t)x'_k)'+\sum_{j=1}^n a_{kj}(t)x_j\geqq(m(t)x'_k)'+a_{kk}(t)x_k$. It then follows from standard comparison theorems that $x_k(t)$ has a zero. Note that a zero in $x_k(t)$ implies the trajectory leaves I, since double zeros are precluded by the positivity of the a_{ij}.

* Received by the editors June 24, 1977, and in revised form March 29, 1978.
† Department of Mathematics, University of California-Davis, Davis, California 95616.

The previous lemma provides a basic existence theorem for positive solutions to the boundary value problem (1.1), (1.2). *For the remainder of this paper we will assume the hypothesis of Lemma* 2.1 *to be satisfied.*

THEOREM 2.2. *There exists an* $x^0 \in I$ *such that* $x(\tau(x^0); x^0) = 0$.

Proof. Let B^n denote the unit ball in R^n, with boundary the unit sphere S^{n-1}. The set $\{x^0 \in \bar{I}; |x^0| = 1\}$ is homeomorphic to B^{n-1}, as is the set $\{x \in \bar{I} - I; |x| \leqq 1\}$. If P is radial contraction of vectors of length >1 onto S^{n-1}, then it follows from the continuity of solutions in their initial conditions that $x^0 \to P\,(x(\tau(x_0); x_0))$ induces a continuous function $f\colon B^{n-1} \to B^{n-1}$. Furthermore, the condition $x(\tau(x^0); x^0) = x^0$ for $x^0 \in \bar{I} - I$, implies that f restricted to S^{n-1} is the identity function. It follows from the no retraction theorem for B^n, S^{n-1} that f is onto, and hence there exists an $x^0 \in I$ such that $x(\tau(x^0); x^0) = 0$. This shows the existence of positive solutions to (1.1), (1.2).

We next show that the trajectory given by Theorem 2.2 realizes the boundary condition in minimum time. If $x(t)$ is a solution of $\mathcal{L}[x] = -(m(t)x'(t))' = A(t)x$, $x'(0) = x(\tau) = 0$, then $x(t)$ is also a solution of the integral equation.

$$x(t) = \int_0^T G(T; t, s)A(s)x(s)\,ds, \tag{2.1}$$

where $G(T; t, s)$ is a Green's function for the operator $\mathcal{L}$.

LEMMA 2.3. *The Green's function of* $\mathcal{L}[x] = -(m(t)x'(t))'$, $x'(0) = x(T) = 0$ *is positive on* $(0, T) \times (0, T)$.

Proof. In this case the Green's function is given by

$$G(T; t, s) = \begin{cases} \displaystyle\int_s^T \frac{d\xi}{m(\xi)}, & 0 \leqq t \leqq s \leqq T, \\ \displaystyle\int_t^T \frac{d\xi}{m(\xi)}, & 0 \leqq s \leqq t \leqq T, \end{cases} \tag{2.2}$$

and the lemma follows from the positivity of m.

Let $M = BC\ ([0, \infty), R^n)$ denote the Banach space of bounded continuous functions $[0, \infty) \to R^n$ with sup norm, and let $K \subset M$ be the cone consisting of all $x \in M$, $x[0, \infty) \subset I$. A bounded linear operator $A\colon M \to M$ is positive if $A(K) \subseteq K$. $r(A)$ will denote the spectral radius of A.

Following [1] we define a family of positive compact linear operators

$$(A_T x)(t) = \begin{cases} \displaystyle\int_0^T G(T; t, s)A(s)x(s)\,ds, & 0 \leqq t \leqq T, \\ 0, & t \geqq T. \end{cases} \tag{2.3}$$

The following result, a consequence of standard theorems in operator theory is proved in [1].

THEOREM 2.4. *The mapping* $T \to r(A_T)$ *is a continuous nondecreasing function of* T, *and* $r(A_T) \to 0$ *as* $T \to 0^+$.

THEOREM 2.5. *Let* T_1 *be the smallest* $T > 0$ *for which there exists a nontrivial solution to* (1.1), (1.2). *Then there exists a positive solution to* (1.1) *satisfying* $x'(0) = 0 = x(T_1)$.

Proof. Let $y(t)$ be a solution to (1.1), (1.2) with $T = T_1$, and set $y(t) = 0$ for $t > T_1$. It follows that $y = A_T(y)$ and $r(A_T) \geqq 1$. Since $r(A_T)$ is an eigenvalue of A_T with eigenfunction in K, the theorem is proved provided $r(A_T) = 1$. If $r(A_T) > 1$, then it follows from Theorem 2.4 that there exists a $T_2 < T_1$ such that $r(A_{T_2})=1$. Hence there exists a nonzero $x \in K$ such that $x = A_{T_2}x$, contradicting the definition of T_1.

3. A comparison theorem. In this section we use the methods associated with integral variational problems to prove a comparison theorem which generalizes to second order systems, a comparison theorem for focal points of second order scalar equations due to Reid. We assume all the hypotheses of § 1, and note that if $u(t)$ is the solution of the initial value problem $(m(t)u')'+a(t)u=0$, $u(0)=1$, $u'(0)=0$, with $m(t)$ and $a(t)$ positive, continuous functions, then $u(t)>0$ on $[0, t_1)$, where t_1 is the first positive zero of $u(t)$.

Let PC^n denote the class of piecewise C^n functions on $[0, T]$, and for $x \in PC^1$ define a functional

$$I[x; T] = \int_0^T [m(t)x' \cdot x' - x^*A(t)x]\, dt, \tag{3.1}$$

where x^* is the transpose of x. Let L be the left hand side of (1.1), so that $L[x] = (m(t)x')' + A(t)x$. L_i will denote the ith coordinate of L.

THEOREM 3.1. *The following conditions are equivalent, and each is necessary and sufficient for the nonexistence of solutions to* (1.1) *satisfying* $x'(0)=0=x(T_1)$, *for any* $T_1 \in (0, T]$.

(i) *For all* $x^0 \in I$, *the trajectory* $x(t; x^0)$ *satisfies* $|x(t; x^0)|>0$ *for* $0 \leqq t \leqq T$.

(ii) *There exists a nonzero function* $y(t)$ *of class* PC^2 *on* $[0, T]$ *such that*

$$y_i(0)m(0)y_i'(0) \leqq 0,$$
$$y_i(t)L_i[y](t) \leqq 0,$$
$$y_i(t)y_j(t) > 0,$$

for $1 \leqq i, j \leqq n$, *and all* $t \in [0, T]$.

(iii) *If* $x(t)$ *is a nonzero function of class* PC^1 *on* $[0, T]$ *satisfying* $x(T)=0$, *then* $I[x; T]>0$.

Proof. If (i) holds then the solution to (1.1) given by Theorem 2.2 satisfies (ii). Let $y(t)$ satisfy (ii) and define a function $h(t)$ by $x_i(t) = y_i(t)h_i(t)$. Then

$$\begin{aligned}
&m(t)(x' \cdot x') - x^*A(t)x \\
&= m(t)\left[\sum_{i=1}^n (y_i'^2h_i^2 + 2y_iy_i'h_ih_i' + y_i^2h_i'^2)\right] - \sum_{i,j=1}^n y_ih_ia_{ij}(t)y_jh_j \\
&= \left(-\sum_{i=1}^n \left(h_i^2y_iL_i[y] + h_i^2(y_im(t)y_i')' + h_i^2y_i \sum_{j=1}^n a_{ij}(t)y_j\right)\right) \\
&\qquad + \left(\sum_{i=1}^n (y_im(t)y_i'h_i^2)' - m(t)(y_iy_i')'h_i^2 - m'(t)y_iy_i'h_i^2\right) \\
&\qquad + \left(\sum_{i=1}^n m(t)y_i^2h_i'^2\right) - \sum_{i,j=1}^n y_ih_ia_{ij}(t)y_jh_j \\
&= -\sum_{i=1}^n h_i^2y_iL_i[y] + \sum_{i=1}^n (y_im(t)y_i'h_i^2)' + \sum_{i=1}^n m(t)y_i^2h_i'^2 + \sum_{i<j} a_{ij}(t)y_iy_j(h_i-h_j)^2.
\end{aligned}$$

Calculation of the terms involving $a_{ij}(t)$ uses the symmetry of A. Integrating we have $\sum_0^T (y_im(t)y_i'h_i^2)'(t)\, dt = -y_i(0)m(0)y_i'(0)h_i^2(0) \geqq 0$. The remaining terms are all positive, so that $I[x; T]>0$, and hence (ii)$\Rightarrow$(iii).

To show (iii)$\Rightarrow$(i) assume the trajectory $x(t; x^0)$ satisfies $x(\xi; x^0)=0$ for some $\xi\in(0, T)$. Define $y(t)$ by $y(t)=x(t; x^0)$ for $0\leqq t\leqq \xi$, $y(t)=0$ for $x\geqq \xi$. Then $I[y; T]=\int_0^T [m(t)y'\cdot y'-y^*A(t)y]\,dt=\int_0^\xi [m(t)x'\cdot x'-x^*A(t)x]\,dt=x^*m(t)x'|_0^\xi=0$, contradicting (iii).

It is clear, in view of Theorems 2.2 and 2.5, that (i) is necessary and sufficient for the nonexistence of solutions satisfying the given boundary condition, which completes the proof of the theorem.

Our next result, a comparison theorem which is a consequence of Theorem 3.1, generalizes Theorem A of [2]. For $n\times n$ matrices $B(t)=(b_{ij}(t))$ and $C(t)=(c_{ij}(t))$, we write $B\leqq C$ if $b_{ij}(t)\leqq c_{ij}(t)$ for all $1\leqq i, j\leqq n$, and all $t\geqq 0$. Let T_1 and T_2 denote the least $T>0$ for which there exists a nontrivial solution to

$$L_1[x]=(m(t)x')'+B(t)x=0, \qquad x'(0)=x(T)=0, \tag{3.2}$$

$$L_2[x]=(M(t)x')'+C(t)x=0, \qquad x'(0)=x(T)=0, \tag{3.3}$$

respectively.

THEOREM 3.2. *The following conditions are sufficient to guarantee that* $T_1\leqq T_2$:

(i) $$\left(\frac{M(t)}{m(t)}\right)'\geqq 0,$$

(ii) $$m(t)C(t)-M(t)B(t)\leqq 0,$$

for all $0\leqq t\leqq T_2$.

Proof. Assume that $T_1>T_2$, and let $x(t)$ be a solution to (3.2) which is positive on $[0, T_2]$. It follows that $x\cdot L_2[x]=x\cdot((M(t)x')'+C(t)x)$, and since $L_1[x]=0$, we have $x\cdot L_2[x]=m(t)(M(t)/m(t))'x\cdot x'-(1/m(t))x^*(M(t)B(t)-m(t)C(t))x\leqq 0$. The theorem then follows from Theorem 3.1(ii).

4. Applications to fourth order equations. The results of the previous sections can be applied to nonselfadjoint differential equations of the form

$$(p_2(t)x'')''+(p_1(t)x')'+q_1(t)x'+p_0(t)x=0 \tag{4.1}$$

by the transformation given in the following lemma. The proof is an elementary verification and will be omitted.

LEMMA 4.1. *Equation* 4.1 *is equivalent to a system of the form* (1.1) *under the transformation* $p_2=-m^2/a_{12}$, $p_1=-mm'(1/a_{12})'-mm''(1/a_{12})-(m/a_{12})(a_{11}+a_{22})$, $q_1=m(a_{22}/a_{12})'$, $p_0=a_{12}-[m(a_{11}/a_{12})']'+a_{11}a_{22}/a_{12}$.

Example 1. $a_{11}(t)=0$, $a_{12}(t)=a_{22}(t)=1$, $m(t)=1/t$ gives the system $((1/t)x')'+A(t)x=0$, $A(t)=\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}$, which is equivalent to the scalar equation $((-1/t^2)x'')''-[(2/t^4+1/t)x']'+x=0$.

Example 2. $a_{11}(t)=0$, $a_{12}(t)=1$, $m(t)=1/t$, $a_{22}(t)=t$, gives rise to $((1/t)x')'+A(t)x=0$, $A(t)=\begin{pmatrix}0 & 1\\ 1 & t\end{pmatrix}$, which is equivalent to the scalar equation $((-1/t^2)x'')''-[(2/t^4+1)x']'+(1/t)x'+x=0$. In both Examples 1 and 2, the equation of Lemma 2.1 is oscillatory by Leighton's oscillation criteria.

Example 3. The system $(e^tx')'+A(t)x=0$, $A(t)=\begin{pmatrix}0 & 1\\ 1 & e^{-t}\end{pmatrix}$ is equivalent to the nonselfadjoint equation $(e^{2t}x'')''+[(1+e^{2t})x']'+x'-x=0$. In this case, $\int^\infty dt/m(t)<\infty$, and the system cannot be transformed into one of the form (1.3). The methods of this paper do in fact apply since the equation of Lemma 2.1 becomes $(e^t u')'+e^{-t}u=0$, which is oscillatory on $(0,\infty)$ by the oscillation criteria of Moore [3].

REFERENCES

[1] K. SCHMITT AND H. L. SMITH, *Positive solutions and conjugate points for systems of differential equations*, J. Nonlinear Functional Analysis, to appear.

[2] WILLIAM T. REID, *A comparison theorem for selfadjoint differential equations of second order*, Ann. of Math., 65 (1957), pp. 197–202.

[3] C. A. SWANSON, *Comparison and Oscillation Theory of Linear Differential Equations*, Academic Press, New York, 1968.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0015$01.00/0

SYSTEMS OF GENERALIZED ABEL EQUATIONS*

M. LOWENGRUB† AND J. WALTON‡

Abstract. Certain mixed boundary value problems arising in the classical theory of elasticity lead to the solution of certain systems of generalized Abel integral equqtions. A method is presented where these systems are reduced to uncoupled pairs of Riemann boundary value problems. Closed form solutions are obtained. We also demonstrate how general systems of dual rèlations (given in terms of Erdélyi–Sneddon operators of fractional integration) may be reduced to these systems of Abel equations.

1. Introduction. As demonstrated in Lowengrub [3], certain mixed boundary value problems arising in the classical theory of elasticity reduce to the problem of determining functions φ_1 and φ_2 satisfying Abel type integral equations of the type,

$$\alpha(x)\int_a^x \frac{\varphi_1(t)\,dt}{(x^\rho-t^\rho)^\mu}+\beta(x)\int_x^b \frac{\varphi_2(t)\,dt}{(t^\rho-x^\rho)^\mu}=h_1(x), \qquad a<x<b, \tag{1.1}$$

$$\gamma(x)\int_x^b \frac{\varphi_1(t)\,dt}{(t^\rho-x^\rho)^\mu}+\delta(x)\int_a^x \frac{\varphi_2(t)\,dt}{(x^\rho-t^\rho)^\mu}=h_2(x), \qquad a<x<b, \tag{1.2}$$

where $0<\mu<1$, $\rho\geqq 1$ and the real valued functions $\alpha(x)$, $\beta(x)$, $\gamma(x)$ and $\delta(x)$ have derivatives satisfying Hölder conditions on (a, b). It is also assumed that $h_j(x)$, $j=1, 2$, is Hölder continuous on the interval.

In this paper we show that systems of the type (1.1) and (1.2) can be reduced to the determination of a matrix function $\Phi(z)=(\Phi_1(z), \Phi_2(z))$ analytic in the plane cut along the (a, b), satisfying certain growth conditions at ∞ and along the cut, (a, b),

$$A(x)\Phi^+(x)=e^{\mu\pi i}\overline{A(x)}\Phi(x)+F(x) \tag{1.3}$$

where $A(x)$ is a coefficient matrix with elements linear combinations of α, β, γ and δ. The system (1.3) is a coupled system. We effect a linear uncoupling of this system by introducing certain similarity transformations. The matrices associated with these transformations are explicitly computed and hence exact solutions are derived. In physical applications, such as the determination of the stress field in an inhomogeneous body containing flaws, the relevant physical quantities are expressed in terms of the matrix function $\Phi(z)$. One need not actually solve for φ_1 and φ_2 in (1.1) and (1.2). The functions $\Phi_1(z)$ and $\Phi_2(z)$ are defined (say in the case $\rho=1$) by

$$\Phi_j(z)=\frac{1}{R(z)}\int_a^b \frac{\varphi_j(t)\,dt}{(t-z)^\mu}$$

where $R(z)=[(z-a)(b-z)]^{(1-\mu)/2}$. These functions must be defined on appropriate branches. Analogous representations are introduced for $\rho\geqq 1$.

Section 2 of the paper thoroughly analyzes the case $\rho=1$ while in § 3 we choose $\rho=2$. These are the two cases of physical interest. In § 4 we consider some explicit examples: (i) $\alpha(x)=\alpha$, $\beta(x)=\beta$, $\gamma(x)=\gamma$ and $\delta(x)=\delta$ with $\mu=\frac{1}{2}$, $\rho=1$ and α, β, γ and δ constant; (ii) $\alpha(x)=\beta$, $\beta(x)=\alpha/x$, $\gamma(x)=\alpha$ and $\delta(x)=-\beta/x$ with $\mu=\frac{1}{2}$, $\rho=2$, and α, β constant. The final section demonstrates how general dual relations (given in terms of

* Received by the editors October 27, 1977, and in revised form April 20, 1978.

† Department of Mathematics, Indiana University, Bloomington, Indiana 47401. The work of this author was supported in part by the National Science Foundation under Grant NFS-MCS-75-03894A2.

‡ Texas A & M University, College Station, Texas 77840. The work of this author was supported in part by the Air Force Office of Scientific Research under Grant 77-3290.

the Erdélyi–Sneddon operators of fractional integration—see Erdélyi–Sneddon [1]) may be reduced to systems of the type (1.1) and (1.2).

It should be mentioned that scalar Abel type equations have been studied by several authors. In addition to Sakalyuk [6], the reader may wish to consult references [7], [8], [11] and [12]. These authors investigate such equations in weighted L^p-spaces.

2. The first generalized Abel system. In this section we consider the generalized Abel system of equations,

$$\alpha(x)\int_a^x \frac{\varphi_1(t)\,dt}{(x-t)^\mu}+\beta(x)\int_x^b \frac{\varphi_2(t)\,dt}{(t-x)^\mu}=f_1(x),\qquad x\in(a,b),\tag{2.1}$$

$$\gamma(x)\int_x^b \frac{\varphi_1(t)\,dt}{(t-x)^\mu}+\delta(x)\int_a^x \frac{\varphi_2(t)\,dt}{(x-t)^\mu}=f_2(x),\qquad x\in(a,b),\tag{2.2}$$

where α, β, γ, and δ satisfy conditions to be specified later. However, we do assume that f_1 and f_2 are Hölder continuous on (a, b).

As in Sakalyuk [6], we define the sectionally analytic functions

$$\Phi_j(z)=\frac{1}{R(z)}\int_a^b \frac{\varphi_j(t)}{(t-z)^\mu}\,dt,\qquad j=1,2,\tag{2.3}$$

where $R(z)=[(z-a)(b-z)]^{(1-\mu)/2}$ and the function is defined by some branch. If $\varphi_j^*(t)$ satisfies $\varphi_j(t)=\varphi_j^*(t)[(t-a)(b-t)]^{\mu+\varepsilon-1}$ where $\varepsilon>0$ and $\varphi_j^*(t)$ is Hölder on $[a,b]$, then $\Phi_j(z)$ is analytic in the plane cut along [a, b]. Moreover, the boundary limits $\Phi_j^{\pm}(x)$, where

$$\Phi_j^+(x)=\lim_{\substack{z\to x\\ \mathrm{Im}(z)>0}}\Phi_j(z),\qquad a<x<b,$$

and

$$\Phi_j^-(x)=\lim_{\substack{z\to x\\ \mathrm{Im}(z)<0}}\Phi_j(z),\qquad a<x<b,$$

are continuous. In addition,

$$\begin{aligned}\Phi_j(z)&=O(|z-a|^{(\mu-1)/2}), && z\to a,\\ \Phi_j(z)&=O(|z-b|^{(\mu-1)/2}), && z\to b,\\ \Phi_j(z)&=O(|z|^{-1}), && z\to\infty.\end{aligned}\tag{2.4}$$

A simple calculation verifies that

$$\Phi_j^+(x)=\frac{1}{R(x)}\left[e^{\mu\pi i}\int_a^x \frac{\varphi_j(t)\,dt}{(x-t)^\mu}+\int_x^b \frac{\varphi_j(t)\,dt}{(t-x)^\mu}\right],\qquad x\in(a,b),\tag{2.5}$$

and

$$\Phi_j^-(x)=-\frac{1}{R(x)}\left[\int_a^x \frac{\varphi_j(t)\,dt}{(x-t)^\mu}+e^{\mu\pi i}\int_x^b \frac{\varphi_j(t)\,dt}{(t-x)^\mu}\right].\tag{2.6}$$

It immediately follows that

$$\int_a^x \frac{\varphi_j(t)\,dt}{(x-t)^\mu}=\left[\frac{e^{\mu\pi i}\Phi_j^+(x)+\Phi_j^-(x)}{e^{2\mu\pi i}-1}\right]R(x),\qquad x\in(a,b).\tag{2.7}$$

and

$$\int_x^b \frac{\varphi_j(t)\,dt}{(t-x)^\mu} = -\left[\frac{\Phi_j^+(x)+e^{\mu\pi i}\Phi_j^-(x)}{e^{2\mu\pi i}-1}\right]R(x), \qquad j=1,2. \tag{2.8}$$

Substitution of (2.7) and (2.8) into (2.1) and (2.2) yields the boundary condition on $a<x<b$,

$$\alpha(x)\,e^{\mu\pi i}\Phi_1^+(x)-\beta(x)\Phi_2^+(x)+\alpha(x)\Phi_1^-(x)-\beta(x)\,e^{\mu\pi i}\Phi_2^-(x)=F_1(x), \tag{2.9}$$

$$-\gamma(x)\Phi_1^+(x)+\delta(x)\,e^{\mu\pi i}\Phi_2^+(x)-\gamma(x)\,e^{\mu\pi i}\Phi_1^-(x)+\delta(x)\Phi_2^-(x)=F_2(x), \tag{2.10}$$

where

$$F_j(x)=f_j(x)(e^{2\mu\pi i}-1)/R(x).$$

This substitution then reduces our problem to determining two sectionally analytic functions $\Phi_1(z)$, $\Phi_2(z)$ satisfying the growth conditions (2.4); that is, solving a coupled Riemann–Hilbert boundary value problem. Once we have determined $\Phi_j(z)$, $(j=1,2)$, the functions φ_1 and φ_2 are obtained by solving the Abel integral equations (2.7) and (2.8).

For convenience we introduce the following matrix notation; set,

$$\Phi(z)=(\Phi_1(z),\Phi_2(z))^T \quad \text{and} \quad F(x)=(F_1(x),F_2(x))^T$$

so that the boundary conditions (2.9) and (2.10) may be written in the form

$$A(x)\Phi^+(x)=-e^{\mu\pi i}\overline{A(x)}\Phi^-(x)+F(x), \qquad a<x<b, \tag{2.11}$$

where

$$A(x)=(a_{ij}(x)), \qquad i,j=1,2,$$

with

$$\begin{aligned} &a_{11}(x)=\alpha(x)\,e^{\mu\pi i}, \quad a_{12}(x)=-\beta(x),\\ &a_{21}(x)=-\gamma(x) \quad \text{and} \quad a_{22}(x)=\delta(x)\,e^{\mu\pi i}. \end{aligned} \tag{2.11a}$$

If we require the condition

$$\det A(x)=\alpha(x)\delta(x)\,e^{2\mu\pi i}-\gamma(x)\beta(x)\neq 0, \quad a<x<b, \tag{2.12}$$

then the matrix $A(x)$ is invertible and (2.11) is equivalent to the boundary condition

$$\Phi^+(x)=-e^{\mu\pi i}G(x)\Phi^-(x)+g(x), \qquad a<x<b, \tag{2.13}$$

where

$$G(x)=A^{-1}(x)\overline{A(x)} \quad \text{and} \quad g(x)=A^{-1}(x)F(x). \tag{2.14}$$

It is necessary for us to determine conditions whereby the coupled Riemann–Hilbert problem can be uncoupled. We shall effect a linear uncoupling of the system (2.9) and (2.10) by finding a nonsingular matrix $P(z)$, analytic in the complex plane (except for perhaps a finite number of poles) with a pole at infinity and such that for $a<x<b$

$$P(x)G(x)P^{-1}(x)=D(x)=(d_{ij}(x))$$

with $d_{12}(x)=d_{21}(x)=0$.

Let $\Sigma(z)=(\Sigma_1(z),\Sigma_2(z))^T$ be defined as follows:

$$\Sigma(z)=P(z)\Phi(z). \tag{2.15}$$

Note that $\Sigma(z)$ is analytic in the plane (except for perhaps a finite number of poles) cut along $a<x<b$. In addition, $\Sigma(z)$ satisfies appropriate growth conditions at infinity.

Substitution of (2.14) into the boundary condition (2.13) yields the uncoupled set of conditions.

$$\Sigma_j^+(x)=-e^{\mu\pi i}d_{jj}(x)\Sigma_j^-(x)+k_j(x), \qquad a<x<b, \tag{2.16}$$

$j=1,2$ and $k(x)=P(x)g(x)$. Thus, this procedure reduces our problem to the determination of two sectionally analytic functions $\Sigma_1(z)$ and $\Sigma_2(z)$ satisfying appropriate growth conditions at infinity and the boundary conditions (2.15). The solution to these Riemann–Hilbert problems is well known (once the index is determined) (see for example, Muskhelishvili [5]).

The main problem is to explicitly determine the matrix P. We first compute $G(x)$ from (2.14).

$$G(x)=[\det A(x)]^{-1}T(x)$$

where $T(x)=[t_{ij}(x)]$ with

$$t_{11}=t_{22}=\alpha(x)\delta(x)-\beta(x)\gamma(x), \qquad t_{12}=-2i\beta(x)\delta(x)\sin\mu\pi$$

and

$$t_{21}=-2i\alpha(x)\gamma(x)\sin\mu\pi.$$

A nonsingular matrix, $P(x)$, for which $P(x)T(x)P^{-1}(x)$ is diagonal exists if and only if $T(x)$ has two linearly independent eigenvectors. In this case, the matrix $P(x)T(x)P^{-1}(x)$ has the eigenvalues of $T(x)$ as its diagonal elements and $P^{-1}(x)$ has as its columns two independent eigenvectors of $t(x)$. If $t_{12}(x)t_{21}(x)\neq 0$ then the eigenvalues of $T(x)$ are $t_{11}(x)\pm\sqrt{t_{12}(x)t_{21}(x)}$ so that for the matrix $P^{-1}(x)$ we may take

$$P^{-1}(x)=c(x)\begin{bmatrix} t_{12} & t_{12} \\ \sqrt{t_{12}t_{21}} & -\sqrt{t_{12}t_{21}} \end{bmatrix} \tag{2.17}$$

where $c(x)$ is any nonvanishing *scalar* function.

If $t_{12}(x)t_{21}(x)=0$, then $T(x)$ has two independent eigenvectors if and only if both $t_{12}(x)$ and $t_{21}(x)$ vanish. Since $T(x)$ is nonsingular this can occur if and only if either $\delta(x)=\alpha(x)=0$ and $\gamma(x)\beta(x)\neq 0$ or $\gamma(x)=\beta(x)=0$ and $\delta(x)\alpha(x)\neq 0$. In this case, $T(x)$ is just a scalar multiple of the identity. For most applications, $t_{12}(x)t_{21}(x)\neq 0$ except perhaps at the endpoints $x=a$ and $x=b$. Such exceptional cases are easily handled. In order to simplify further work, we assume that $t_{12}(x)t_{21}(x)\neq 0$.

Let $P(x)$ be defined by (2.17) so that

$$P(x)T(x)P^{-1}(x)=\begin{pmatrix} t_{11}(x)+\sqrt{t_{12}(x)t_{21}(x)} & 0 \\ 0 & t_{11}(x)-\sqrt{t_{12}(x)t_{21}(x)} \end{pmatrix}$$

which provides the desired uncoupling of the pair of coupled Riemann–Hilbert boundary value problems. The scalar $c(x)$ is chosen so that $P(x)$ can be extended to a matrix $P(z)$ meromorphic in the plane with a pole at infinity. This enables us to explicitly determine Σ_1 and Σ_2 and hence Φ_1 and Φ_2. Inversion of (2.7), (2.8) yields our original unknown functions φ_1 and φ_2. An example of this analysis appears in § 4.

It should be emphasized that the key to the method illustrated above is the extension of the matrix $P(x)$ to a matrix $P(z)$ which is meromorphic. This naturally imposes strong conditions on the coefficients $\alpha(x)$, $\beta(x)$, $\gamma(x)$ and $\delta(x)$. However, a detailed analysis of this problem seems inappropriate since, in general, it is a simple enough matter to resolve the question for individual cases arising in applications. As illustrated later in the paper, this usually amounts to determining whether the appearance of square roots in $P(x)$ introduces branch points off of the cut (a, b).

3. The second generalized Abel system. We next consider the system

$$\alpha(x)\int_0^x \frac{\varphi_1(t)\,dt}{(x^2-t^2)^\mu}+\beta(x)\int_x^1 \frac{\varphi_2(t)\,dt}{(t^2-x^2)^\mu}=f_1(x), \qquad 0<x<1, \tag{3.1}$$

$$\gamma(x)\int_x^1 \frac{\varphi_1(t)\,dt}{(t^2-x^2)^\mu}+\delta(x)\int_0^x \frac{\varphi_2(t)\,dt}{(x^2-t^2)^\mu}=f_2(x), \qquad 0<x<1. \tag{3.2}$$

The interval $(0, 1)$ is chosen rather than (a, b) in order to simplify the analysis. It is trivial to extend to (a, b). The case $\mu=\frac{1}{2}$ in (3.1) and (3.2) has been considered by Lowengrub [3].

Analogous to the method used in § 2, we introduce the sectionally analytic functions $\Phi_1(z)$ and $\Phi_2(z)$ defined by

$$\Phi_j(z)=(z^2-1)^{\mu-1/2}\int_0^1 \frac{\varphi_j(t)\,dt}{(z^2-t^2)^\mu}, \qquad j=1, 2. \tag{3.3}$$

If $\varphi_j(t)$ satisfies $\varphi_j(t)=\varphi_j^*(t)t^\mu[t(1-t)]^{\mu+\varepsilon-1}$ where $\varepsilon>0$ and $\varphi_j^*(t)$ is Hölder continuous on $[0, 1]$, then $\Phi_j(z)$ $(j=1, 2)$ is analytic in the plane cut along $[-1, 1]$ and satisfies the following conditions,

$$\begin{aligned}&\Phi_j(z)=O(|z-1|^{\mu-1/2}) \quad \text{as } z\to 1,\\ &\Phi_j(z)=O(|z+1|^{\mu-1/2}) \quad \text{as } z\to -1, \quad \text{and}\\ &\Phi_j(z)=O(|z|^{-1}) \quad \text{as } z\to\infty.\end{aligned} \tag{3.4}$$

Moreover, the limiting values $\Phi_j^\pm(x)$ are continuous functions for $|x|<1$ except perhaps for $x=0$.

For each of the functions $(z-1)^{\mu-1/2}$, $(z+1)^{\mu-1/2}$, $(z-t)^{-\mu}$ and $(z+t)^{-\mu}$ we take as the branch cut that line lying along the positive x-axis and restrict their arguments to lie between 0 and 2π. The following limits, for $0<x<1$, are easily computed:

$$\Phi_j^+(x)=-i(1-x^2)^{\mu-1/2}\left\{\int_x^1 \frac{\varphi_j(t)\,dt}{(t^2-x^2)^\mu}+e^{\mu\pi i}\int_0^x \frac{\varphi_j(t)\,dt}{(x^2-t^2)^\mu}\right\}, \tag{3.5}$$

$$\Phi_j^-(x)=i(1-x^2)^{\mu-1/2}\left\{\int_x^1 \frac{\varphi_j(t)\,dt}{(t^2-x^2)^\mu}+e^{-\mu\pi i}\int_0^x \frac{\varphi_j(t)\,dt}{(x^2-t^2)^\mu}\right\}; \tag{3.6}$$

whereas for $-1<x<0$

$$\Phi_j^+(x)=-i(1-x^2)^{\mu-1/2}\left\{\int_{|x|}^1 \frac{\varphi_j(t)\,dt}{(t^2-x^2)^\mu}+e^{-\mu\pi i}\int_0^{|x|} \frac{\varphi_j(t)\,dt}{(x^2-t^2)^\mu}\right\}, \tag{3.7}$$

$$\Phi_j^-(x)=i(1-x^2)^{\mu-1/2}\left\{\int_{|x|}^1 \frac{\varphi_j(t)\,dt}{(t^2-x^2)^\mu}+e^{\mu\pi i}\int_0^{|x|} \frac{\varphi_j(t)\,dt}{(x^2-t^2)^\mu}\right\}. \tag{3.8}$$

It should be observed that for $-1<x<0$,

$$\Phi_j^{\pm}(x)=-\overline{\Phi_j^{\pm}(-x)}. \tag{3.9}$$

$$\Phi_j^+(x)+\Phi_j^-(x)=2\sin\mu\pi\,(1-x^2)^{\mu-1/2}\int_0^x \frac{\varphi_j(t)\,dt}{(x^2-t^2)^{\mu}}, \tag{3.10}$$

and

$$e^{-\mu\pi i}\Phi_j^+(x)+e^{\mu\pi i}\Phi_j^-(x)=-2\sin\mu\pi\,(1-x^2)^{\mu-1/2}\int_x^1 \frac{\varphi_j(t)\,dt}{(t^2-x^2)^{\mu}} \tag{3.11}$$

for $0<x<1$.

Substitution of the above into the original set of integral equations (3.1) and (3.2) reduces the problem to the following: determine two sectionally analytic functions $\Phi_1(z)$ and $\Phi_2(z)$ satisfying the conditions (3.4) and the boundary values,

$$\begin{aligned}&[\alpha(x)\Phi_1^+(x)-\beta(x)\,e^{-\mu\pi i}\Phi_2^+(x)]\\ &\quad+[\alpha(x)\Phi_1^- -\beta(x)\,e^{\mu\pi i}\Phi_2^-(x)]=F_1(x), \qquad 0<x<1,\end{aligned} \tag{3.12}$$

$$\begin{aligned}&[-\gamma(x)\,e^{-\mu\pi i}\Phi_1^+(x)+\delta(x)\Phi_2^+(x)]\\ &\quad+[\delta(x)\Phi_2^-(x)-\gamma(x)\,e^{\mu\pi i}\Phi_1^-(x)]=F_2(x), \qquad 0<x<1,\end{aligned} \tag{3.13}$$

where

$$F_j(x)=2\sin\mu\pi\,(1-x^2)^{\mu-1/2}f_j(x).$$

In matrix notation the system (3.12), (3.13) becomes

$$A(x)\Phi^+(x)=-B(x)\Phi^-(x)+F(x), \qquad 0<x<1, \tag{3.14}$$

where

$$\Phi(z)=(\Phi_1(z),\Phi_2(z))^T, \qquad F(x)=(F_1(x),F_2(x))^T,$$
$$A(x)=(a_{ij}(x))$$

with

$$a_{11}(x)=\alpha(x), \quad a_{12}(x)=-\beta(x)\,e^{-\mu\pi i}, \quad a_{21}(x)=-\gamma(x)\,e^{-\mu\pi i} \quad \text{and} \quad a_{22}(x)=\delta(x),$$

and

$$B(x)=\overline{A(x)}. \tag{3.14a}$$

In order to determine $\Phi(z)$, boundary conditions for $\Phi^{\pm}(x)$ must be extended to all of $[-1,1]$. It is clear from (3.9) how this extension is to be performed.

In particular, we obtain the system

$$\hat{A}(x)\Phi^+(x)=-\hat{B}(x)\Phi(x)+\hat{F}(x), \qquad -1<x<1, \tag{3.15}$$

where

$$\hat{A}(x)=\begin{cases}A(x), & 0<x<1,\\ \overline{-A(-x)}, & -1<x<0,\end{cases}$$
$$\hat{B}(x)=\begin{cases}B(x), & 0<x<1,\\ \overline{-B(-x)}, & -1<x<0,\end{cases} \tag{3.16a}$$

and

(3.16b)
$$\hat{F}(x)=\begin{cases} F(x), & 0<x<1, \\ \overline{F(-x)}, & -1<x<0. \end{cases}$$

As in § 1, we assume that $\hat{A}(x)$ is invertible for $-1<x<1$. The expressions in (3.16) imply that it suffices to assume that $A(x)$ is invertible for $0<x<1$, or equivalently, that

$$\alpha(x)\delta(x)\,e^{2\mu\pi i}-\gamma(x)\beta(x)\neq 0, \qquad 0<x<1.$$

This gives us the system,

(3.17)
$$\Phi^+(x)=-\hat{G}(x)\Phi^-(x)+\hat{g}(x), \qquad -1<x<1,$$

where

$$\hat{G}(x)=\hat{A}^{-1}(x)\hat{B}(x)=\hat{A}^{-1}(x)\overline{\hat{A}(x)}$$

and

$$\hat{g}(x)=\hat{A}^{-1}(x)\hat{F}(x).$$

Thus, we must now find a sectionally analytic matrix, $\Phi(z)$ satisfying appropriate conditions at 1 and -1, a growth condition $\Phi(z)=O(|z|^{-1})$ at infinity, and the boundary values (3.17). We seek a linear uncoupling of (3.17); that is, a nonsingular matrix $\hat{P}(x)$ such that $\hat{P}(x)\hat{G}(x)\hat{P}^{-1}(x)$ is a diagonal matrix for $-1<x<1$.

We first compute $\hat{G}(x)$ on $0<x<1$ and obtain,

$$\hat{G}(x)=[\det(A(x))]^{-1}T(x)$$

where $T(x)=[t_{ij}(x)]$ with $t_{11}(x)=t_{22}(x)=\delta(x)\alpha(x)-\gamma(x)\beta(x)$, $t_{12}(x)=-2i\sin\mu\pi\,\delta(x)\beta(x)$ and $t_{21}(x)=-2i\sin\mu\pi\,\alpha(x)\gamma(x)$. If we employ the arguments of § 1, we see that

(3.18)
$$\begin{aligned} P^{-1}(x)&=c(x)\begin{pmatrix} \dfrac{t_{12}(x)}{\sqrt{t_{12}(x)t_{21}(x)}} & \dfrac{t_{12}}{-\sqrt{t_{12}(x)t_{21}(x)}} \end{pmatrix} \\ &=+2i\sin\mu\pi\,c(x)\begin{pmatrix} \dfrac{-\delta(x)\beta(x)}{\sqrt{\alpha(x)\beta(x)\gamma(x)\delta(x)}} & \dfrac{-\delta(x)\beta(x)}{-\sqrt{\alpha(x)\beta(x)\gamma(x)\delta(x)}} \end{pmatrix} \end{aligned}$$

(provided $t_{12}(x)t_{21}(x)\neq 0$) produces the desired uncoupling of the boundary conditions (3.17) for $0<x<1$. In particular,

$$P(x)T(x)P^{-1}(x)=\begin{pmatrix} t_{11}(x)+\sqrt{t_{12}(x)t_{21}(x)} & 0 \\ 0 & t_{11}(x)-\sqrt{t_{12}(x)t_{21}(x)} \end{pmatrix}$$

on $0<x<1$.

Next we compute $\hat{G}(x)$ for $-1<x<0$. Since $\hat{G}(x)=\overline{\hat{G}(-x)}$ for $-1<x<0$, we find that

$$\hat{G}(x)=[\det\overline{(A(-x))}]^{-1}\overline{T(-x)}, \qquad -1<x<0.$$

Moreover, since $\alpha(x)$, $\beta(x)$, $\gamma(x)$ and $\delta(x)$ are real, the columns of $P^{-1}(x)$ are also independent eigenvectors of $T(x)$ for $0<x<1$. Hence $P(-x)T(-x)P(-x)^{-1}$ is diagonal for $-1<x<0$, and if $\hat{P}(x)$ is defined by

$$\hat{P}(x)=\begin{cases} P(x), & 0<x<1, \\ P(-x), & -1<x<0, \end{cases} \quad\text{or}\quad \begin{cases} P(x), & 0<x<1, \\ \overline{P(-x)}, & -1<x<0, \end{cases}$$

then $\hat{P}(x)\hat{G}(x)\hat{P}^{-1}(x)$ is diagonal for $-1<x<1$. If in addition $\hat{P}(x)$ can be extended to a function $\hat{P}(z)$ meromorphic in $\mathbb{C}$ with a pole at infinity, then $\Sigma(z)=\hat{P}(z)\Phi(z)$ defines a sectionally analytic function $\Sigma(z)$, satisfying properties analogous to (3.4) and the boundary condition

$$\Sigma^+(x)=-\hat{D}(x)\Sigma^-(x)+\hat{k}(x), \qquad -1<x<1, \tag{3.19}$$

where

$$\hat{D}(x)=\hat{P}(x)\hat{G}(x)\hat{P}^{-1}(x)$$

and $\hat{k}(x)=\hat{P}(x)\hat{g}(x)$. An example illustrating this analysis appears in the fourth section.

4. Examples. As a first example, we consider the set of equations (2.1), (2.2) with $\alpha(x)=\alpha$, $\beta(x)=\beta$, $\gamma(x)=\gamma$ and $\delta(x)=\delta$. Here, α, β, γ, and δ are all real constants. Since this example appears in various applications, we display the appropriate matrices and write out the solution. We assume that $\alpha\beta\gamma\delta\neq 0$.

The coefficient matrices needed in (2.13) and (2.14) are given by:

$$A^{-1}(x)=\frac{1}{\alpha\delta\, e^{2\mu\pi i}-\gamma\beta}\begin{bmatrix}\delta e^{\mu\pi i} & \beta\\ \gamma & \alpha\, e^{\mu\pi i}\end{bmatrix}, \tag{4.1}$$

$$G(x)=\frac{1}{\alpha\delta\, e^{2\mu\pi i}-\gamma\beta}\begin{bmatrix}\alpha\delta-\beta\gamma & \dfrac{-2i\beta\delta\sin\mu\pi}{\alpha\delta-\beta\gamma}\\ -2i\alpha\gamma\sin\mu\pi & \alpha\delta-\beta\gamma\end{bmatrix}, \tag{4.2}$$

$$P(x)=-2i\sin\mu\pi\begin{bmatrix}\delta\beta & \delta\beta\\ -\sqrt{\alpha\beta\gamma\delta} & \sqrt{\alpha\beta\gamma\delta}\end{bmatrix}, \tag{4.3}$$

while the diagonal matrix, $D(x)$, used in (2.15) takes the form,

$$D(x)=\frac{1}{\alpha\delta\, e^{2\mu\pi i}-\gamma\beta}\cdot\begin{bmatrix}\alpha\beta-\beta\gamma+2i\sin\mu\pi\sqrt{\alpha\beta\gamma\delta} & 0\\ 0 & \alpha\delta-\beta\gamma-2i\sin\mu\pi\sqrt{\alpha\beta\gamma\delta}\end{bmatrix}. \tag{4.4}$$

The problem is then reduced to the solution of the uncoupled Riemann–Hilbert problem: determine $\Sigma_1(z)$, and $\Sigma_2(z)$ analytic everywhere in the plane except along the cut (a, b) where Σ_1 and Σ_2 satisfy the conditions

$$\Sigma_1^+(x)=-e^{(\nu+\lambda)i}\frac{\rho}{\sigma}\Sigma_1^-(x)+k_1(x), \tag{4.5}$$

$$\Sigma_2^+(x)=-e^{(\lambda-\nu)i}\frac{\rho}{\sigma}\Sigma_2^-(x)+k_2(x), \tag{4.6}$$

$$\Sigma_1(z)=\Sigma_2(z)=O(|z|^{-1}) \quad \text{as } z\to\infty,$$

and ν, λ, ρ, σ, $k_1(x)$ and $k_2(x)$ are given by:

$$\nu=\tan^{-1}\left[\frac{2\sin\mu\pi\sqrt{\alpha\beta\gamma\delta}}{\delta\alpha-\beta\gamma}\right], \tag{4.7}$$

$$\lambda=\tan^{-1}\left[-\left(\frac{\alpha\delta+\gamma\beta}{\alpha\delta-\gamma\beta}\right)\tan\mu\pi\right], \tag{4.8}$$

$$\rho=[(\delta\alpha-\beta\gamma)^2+4\sin^2\mu\pi\,(\alpha\beta\gamma\delta)]^{1/2}, \tag{4.9}$$

(4.10) $$\sigma = [(\alpha\delta)^2 + (\gamma\beta)^2 - 2\alpha\delta\gamma\beta \cos(2\mu\pi)]^{1/2},$$

where

(4.11) $$K(x) = \begin{bmatrix} k_1(x) \\ k_2(x) \end{bmatrix} = [P(x)A^{-1}(x)]\begin{bmatrix} F_1(x) \\ F_2(x) \end{bmatrix}.$$

For example, the solution of the coupled pair of equations

(4.12) $$\int_0^x \frac{\varphi_1(t)\,dt}{(x-t)^{1/2}} + \int_x^1 \frac{\varphi_2(t)\,dt}{(t-x)^{1/2}} = f_1, \qquad x \in (0, 1),$$

and

(4.13) $$\int_x^1 \frac{\varphi_1(t)\,dt}{(t-x)^{1/2}} + \int_0^x \frac{\varphi_2(t)\,dt}{(x-t)^{1/2}} = f_2, \qquad x \in (0, 1),$$

(with f_1 and f_2 constant) requires the determination of the sectionally analytic functions $\Sigma_1(z)$, $\Sigma_2(z)$ with Σ_1 and Σ_2 vanishing at ∞ and satisfying the boundary conditions,

(4.14) $$\Sigma_1^+(x) = -\Sigma_1^-(x) + k_1, \qquad x \in (-1, 1),$$

(4.15) $$\Sigma_2^+(x) = \Sigma_2^-(x) + k_2, \qquad x \in (-1, 1),$$

where k_1 and k_2 are the complex constants

$$k_1 = (1-i)(f_1+f_2), \qquad k_2 = -(1+i)(f_1-f_2).$$

It is a simple matter to demonstrate that the solutions to the above Riemann–Hilbert problems are given by

(4.16) $$\Sigma_1(z) = \frac{1}{2\pi i}(i-1)(f_1+f_2)\left\{\frac{z}{\sqrt{(z+1)(z-1)}} - 1\right\}$$

and

(4.17) $$\Sigma_2(z) = -\frac{1}{2\pi i}(1+i)(f_1-f_2)\ln\left[\frac{z-1}{1+z}\right]$$

so that

$$\Phi_1(z) = -i[\Sigma_1(z) + \Sigma_2(z)]$$

and

$$\Phi_2(z) = -i[\Sigma_2(z) - \Sigma_1(z)].$$

The unknown functions φ_1 and φ_2 are then determined from the Abel equations (2.7) and (2.8) where $R(x) = \sqrt[4]{(1-x^2)}$. One can see that even in the simple case (4.12) and (4.13) the results are quite complicated. Fortunately, for applications, one is usually only interested in $\Phi_1(z)$ and $\Phi_2(z)$. For problems in the theory of elasticity, (see Lowengrub [3]) the stress field is expressed in terms of these two sectionally analytic functions.

Secondly, we consider the particular case of (3.1) and (3.2) with

$$\mu = \tfrac{1}{2}, \quad \alpha(x) = \beta, \quad \beta(x) = \alpha/x, \quad \gamma(x) = \alpha \quad \text{and} \quad \delta(x) = -\beta/x,$$

where α and β are constants. We show that our methods yield the same result as in

Lowengrub [3]. We consider the coupled system

$$\hat{A}(x)\Phi^+(x) = -\hat{B}(x)\Phi^-(x) + \hat{F}(x), \qquad -1 < x < 1,$$

where $A(x)$, $B(x)$, and $F(x)$ are given in (3.14a) while $\hat{A}$, $\hat{B}$ and $\hat{F}$ are defined by (3.16a, b). It follows that the condition $\det A(x) \neq 0$ implies that $\beta^2 - \alpha^2 \neq 0$. In addition, if $0 < x < 1$ we have

$$G(x) = -\frac{x}{\beta^2 - \alpha^2}\begin{bmatrix} -\left(\dfrac{\beta^2+\alpha^2}{x}\right) & \dfrac{2i\alpha\beta}{x^2} \\ -2i\alpha\beta & -\left(\dfrac{\beta^2+\alpha^2}{x}\right) \end{bmatrix}, \tag{4.18}$$

$$P^{-1}(x) = \begin{pmatrix} \dfrac{i}{x} & \dfrac{i}{x} \\ 1 & -1 \end{pmatrix}, \tag{4.19}$$

$$P(x)T(x)P^{-1}(x) = -\frac{1}{x}\begin{pmatrix} (\beta-\alpha)^2 & 0 \\ 0 & (\beta+\alpha)^2 \end{pmatrix}, \tag{4.20}$$

while if $-1 < x < 0$,

$$\overline{P(-x)}\,\overline{T(-x)}\,\overline{P^{-1}(-x)} = \frac{1}{x}\begin{pmatrix} (\beta-\alpha)^2 & 0 \\ 0 & (\beta+\alpha)^2 \end{pmatrix} \tag{4.21}$$

and $\det \overline{A(-x)} = (1/x)(\beta^2 - \alpha^2)$.

Thus, the uncoupled Riemann boundary value problem becomes: determine two sectionally analytic functions $\Sigma_1(z)$ and $\Sigma_2(z)$ that vanish at infinity and along the cut $-1 \leqq x \leqq 1$, satisfy the boundary conditions

$$\begin{aligned} \Sigma_1^+(x) &= \left(\frac{\beta-\alpha}{\beta+\alpha}\right)\Sigma_1^-(x) + k_1(x), \qquad -1 < x < 1, \\ \Sigma_2^+(x) &= \left(\frac{\beta+\alpha}{\beta-\alpha}\right)\Sigma_2^-(x) + k_2(x), \qquad -1 < x < 1. \end{aligned} \tag{4.22}$$

This is in complete agreement with Lowengrub [3], where the solution to the original pair of Abel equations is given along with an application to elasticity.

5. Application to dual relations. In what follows, use will be made of certain well-known operators of fractional integration, differentiation and Hankel transforms which were introduced by Erdélyi and Sneddon [1]. Let $S_{\eta,\alpha}$ denote the Hankel transform

$$S_{\eta,\alpha}\{f(\xi); x\} = 2^\alpha x^{-\alpha}\int_0^\infty \xi^{1-\alpha} f(\xi) J_{2\eta+\alpha}(\xi x)\, d\xi, \tag{5.1}$$

and for $\alpha > 0$ define the fractional integral operators $I_{\eta,\alpha}$ by

$$I_{\eta,\alpha}\{f(\xi); x\} = \frac{2x^{-2\alpha-2\eta}}{\Gamma(\alpha)}\int_0^x (x^2-\xi^2)^{\alpha-1}\xi^{2\eta+1} f(\xi)\, d\xi, \tag{5.2}$$

$$K_{\eta,\alpha}\{f(\xi); x\} = \frac{2x^{2\eta}}{\Gamma(\alpha)}\int_x^\infty (\xi^2-x^2)^{\alpha-1}\xi^{-2\alpha-2\eta+1} f(\xi)\, d\xi. \tag{5.3}$$

For $\alpha < 0$ the fractional derivatives $I_{\eta,\alpha}$ and $K_{\eta,\alpha}$ are defined to be the formal inverses of the operators $I_{\eta+\alpha,-\alpha}$ and $K_{\eta+\alpha,-\alpha}$ respectively. Frequent use will be made of the well-known identities relating these operators (see Sneddon [9, p. 274]).

We shall consider systems of dual relations of the form

$$(5.4)\quad \begin{aligned} &\int_0^\infty [aA(\xi)+bB(\xi)]\xi^{-2\alpha}J_\mu(\xi x)\,d\xi = f_1(x), \\ &\int_0^\infty [cA(\xi)+dB(\xi)]\xi^{-2\beta}J_\nu(\xi x)\,d\xi = f_2(x), \end{aligned} \qquad 0<x<1,$$
$$\begin{aligned} &\int_0^\infty A(\xi)J_\mu(\xi x)\,d\xi = 0, \\ &\int_0^\infty B(\xi)J_\nu(\xi x)\,d\xi = 0, \end{aligned} \qquad 1<x<\infty,$$

where a, b, c and d are constants.

Systems similar to (5.4) have been discussed by various authors. Closed forms solutions have been obtained by Lowengrub and Sneddon [4] for the cases $\alpha=\beta=-1/2$, $\mu=0$, $\nu=1$ and $\alpha=\beta=-1/2$, $\mu=1/2$, $\nu=-1/2$. In the former instance, the system (5.4) was reduced to a system of Carlemann type singular integral equations whereas in the latter the generalized Abel system (3.1), (3.2) was obtained. Westmann [10] used similar techniques to construct solutions for $\alpha=\beta$ and $\mu=\nu=2$, whereas Erdogan [2] presented a method for reducing (5.4) to an infinite system of algebraic equations.

The procedures developed for treating special cases of (5.4) have mostly been ad hoc. We shall indicate a more systematic approach for analyzing (5.4) which in certain cases reduces to the generalized Abel system (3.1), (3.2).

In operator notation the system (5.4) becomes

$$(5.5)\qquad S_{\mu/2-\alpha,2\alpha}\{[a\varphi(\xi)+b\psi(\xi)];x\}=F_1(x), \qquad 0<x<1,$$

$$(5.6)\qquad S_{\nu/2-\beta,2\beta}\{[c\varphi(\xi)+b\psi(\xi)];x\}=F_2(x), \qquad 0<x<1,$$

$$(5.7)\qquad S_{\mu/2,0}\{\varphi(\xi);x\}=g_1(x)H(1-x), \qquad x>1,$$

$$(5.8)\qquad S_{\nu/2,0}\{\psi(\xi);x\}=g_2(x)H(1-x), \qquad x>1,$$

where $\varphi(\xi)=A(\xi)/\xi$, $\psi(\xi)=B(\xi)/\xi$, $F_1(x)=2^{2\alpha}f_1(x)/x^{2\alpha}$, $F_2(x)=2^{2\beta}f_2(x)/x^{2\beta}$ and $g_1(x)$ and $g_2(x)$ are unknown functions for $0<x<1$.

Formal inversion of (5.7) and (5.8) yields

$$(5.9)\qquad \begin{aligned} \varphi(\xi) &= S_{\mu/2,0}\{g_1(x);\xi\} \\ &= S_{\mu/2,0}\cdot K_{\mu/2,\gamma}[k_1] \\ &= 2^\gamma\xi^{-\gamma}\int_0^1 x^{1-\gamma}J_{\mu+\gamma}(x\xi)k_1(x)\,dx \end{aligned}$$

and

$$(5.10)\qquad \begin{aligned} \psi(\xi) &= S_{\nu/2,0}\{g_2(x);\xi\} \\ &= S_{\nu/2,0}\cdot K_{\nu/2,\delta}[k_2] \\ &= S_{\nu/2,\delta}[k_2] \\ &= 2^\delta\xi^{-\delta}\int_0^1 x^{1-\delta}k_2(x)J_{\nu+\delta}(x\xi)\,dx \end{aligned}$$

where

$$k_1(x) = K_{\mu/2+\gamma,-\gamma}[g_1], \qquad k_2(x) = K_{\nu/2+\delta,-\delta}[g_2],$$

and γ and δ are parameters to be specified later. The manipulations involved in (5.9) and (5.10) are well known and may be found in Sneddon [9]. It should be observed that $k_1(x)$ and $k_2(x)$ vanish for $x > 1$.

Substitution of (5.9) and (5.10) into (5.5) and (5.6), followed by an application of appropriate fractional transforms, yields

$$\begin{aligned} I_{\mu/2+\alpha,\lambda}S_{\mu/2-\alpha,2\alpha}\{[a\varphi(\xi)+b\psi(\xi)]; x\} &= \hat{F}_1(x) \\ &= aS_{\mu/2-\alpha,\lambda+2\alpha}\cdot S_{\mu/2,\gamma}[k_1] + bS_{\mu/2-\alpha,\lambda+2\alpha}\cdot S_{\nu/2,\delta}[k_2] \end{aligned} \tag{5.11}$$

and

$$\begin{aligned} I_{\nu/2+\beta,\rho}\cdot S_{\nu/2-\beta,2\beta}\{[c\varphi(\xi)+d\psi(\xi)]; x\} &= \hat{F}_2(x) \\ &= cS_{\nu/2-\beta,\rho+2\beta}\cdot S_{\mu/2,\gamma}[k_1] + dS_{\nu/2-\beta,2\beta}\cdot S_{\nu/2,\delta}[k_2] \end{aligned} \tag{5.12}$$

where

$$\hat{F}_1(x) = I_{\mu/2+\alpha,\lambda}[F_1], \qquad \hat{F}_2(x) = I_{\nu/2+\beta,\rho}[F_2]$$

and λ and ρ are parameters as yet unspecified.

There are two trivial cases of (5.11) and (5.12). If $\mu = \nu$, let $\gamma = \delta$, $\lambda = -\alpha$ and $\rho = -\beta$; then these relations become

$$\begin{aligned} aS_{\mu/2-\alpha,\alpha}\cdot S_{\mu/2,\gamma}[k_1] + bS_{\mu/2-\alpha,\alpha}\cdot S_{\mu/2,\gamma}[k_2] &= \hat{F}_1(x) \\ &= aK_{\mu/2-\alpha,\alpha+\gamma}[k_1] + bK_{\mu/2-\alpha,\alpha+\gamma}[k_2] \end{aligned} \tag{5.13}$$

and

$$\begin{aligned} cS_{\mu/2-\beta,\beta}\cdot S_{\mu/2,\gamma}[k_1] + dS_{\mu/2-\beta,\beta}\cdot S_{\mu/2,\gamma}[k_2] &= \hat{F}_2(x) \\ &= cK_{\mu/2-\beta,\beta+\gamma}[k_1] + dK_{\mu/2-\beta,\beta+\gamma}[k_2]. \end{aligned} \tag{5.14}$$

If we invert relations (5.13) and (5.14), one obtains a simple algebraic system for k_1 and k_2.

A second trivial case results if $(\nu-\mu)/2+\alpha-\beta = 0$, since by choosing $\lambda+\alpha = \rho+\beta = \gamma+\alpha = \delta+\beta = 0$ we obtain

$$aS_{\mu/2-\alpha,\alpha}\cdot S_{\mu/2,-\alpha}[k_1] + bS_{\mu/2-\alpha,\alpha}\cdot S_{\nu/2,-\beta}[k_2] = \hat{F}_1(x) = ak_1 + bI_{\nu/2,\alpha-\beta}[k_2] \tag{5.15}$$

and

$$cS_{\nu/2-\beta,\beta}S_{\mu/2,-\alpha} + dS_{\nu/2-\beta,\beta}S_{\nu/2,-\beta}[k_2] = \hat{F}_2(x) = cI_{\mu/2,\beta-\alpha}[k_1] + dk_2. \tag{5.16}$$

Application of $I_{\nu/2,\alpha-\beta}$ to (5.16) yields

$$ck_1 + dI_{\nu/2,\alpha-\beta}[k_2] = I_{\nu/2,\alpha-\beta}[\hat{F}_2]. \tag{5.17}$$

It is now a simple matter to solve (5.15) and (5.17) for k_1 and k_2.

In general the system (5.11) and (5.12) cannot be solved so easily. However, it may be simplified in either of two ways. If, for λ, γ, ρ and δ, we choose $\gamma = -\alpha$ then the system becomes

$$aI_{\mu/2,(\nu-\mu)/2}[k_1] + bK_{\mu/2-\alpha,(\nu-\mu)/2+(\alpha-\beta)}[k_2] = \hat{F}_1, \tag{5.18}$$

$$cK_{\nu/2-\beta,(\mu-\nu)/2+(\beta-\alpha)}[k_1] + dI_{\nu/2,(\mu-\nu)/2}[k_2] = \hat{F}_2; \tag{5.19}$$

whereas by choosing $\gamma=(\nu-\mu)/2-\beta$, $\delta=(\mu-\nu)/2-\alpha$, $\rho=-\beta$ and $\lambda=-\alpha$ we obtain

$$aK_{\mu/2-\alpha,(\nu-\mu)/2+(\alpha-\beta)}[k_1]+bI_{\nu/2,(\mu-\nu)/2}[k_2]=\hat{F}_1, \tag{5.20}$$

$$cI_{\mu/2,(\nu-\mu)/2}[k_1]+dK_{\nu/2-\beta,(\mu-\nu)/2+(\beta-\alpha)}[k_2]=\hat{F}_2. \tag{5.21}$$

The systems (5.18)–(5.21) may be regarded as generalized Abel systems. In most applications $\alpha=\beta$ and hence we shall consider only this case for the remainder of this section. It should be observed that when $\nu-\mu$ is an even integer the operators appearing in (5.18)–(5.21) are not of fractional order. It is then a simple matter to reduce either of the systems (5.18), (5.19) or (5.20), (5.21) to a single linear ordinary differential equation. In particular, for the case considered by Westmann [10] (i.e. $\nu=\mu+2$) the system (5.18), (5.19) reduces to a simple first order linear differential equation.

If $\mu-\nu$ is not an even integer then the operators are of fractional order, with both fractional integrals and fractional derivatives appearing in both systems. However, in certain cases it is possible to reduce the systems (5.18)–(5.21) to Abel systems of the form (3.1) and (3.2). We might note that if $\mu-\nu$ is one, then the Abel system obtained will contain only operators of order $\frac{1}{2}$, whereas in general, operators of different order will occur within the same system. Provided neither of the unknown functions appears in operators of different order, a straightforward extension of the technique presented in § 3 will treat such systems.

As an example, consider the system (1) with $\nu=1$, $\mu=0$, $\alpha=\beta=\frac{1}{2}$. If we let $\lambda=\rho=-\frac{1}{2}$, $\delta=-1$ and $\gamma=0$ we obtain

$$aK_{-1/2,1/2}[k_1]+bI_{1/2,-1/2}[k_2]=\hat{F}_1(x), \tag{5.22}$$

$$cI_{0,1/2}[k_1]+dK_{0,-1/2}[k_2]=\hat{F}_2(x). \tag{5.23}$$

Since,

$$I_{1/2,-1/2}[k_2]=\int_0^x (x^2-t^2)^{-1/2}\frac{d}{dt}[tk_2(t)]\,dt,$$

$$K_{0,-1/2}[k_2]=-\int_x^1 (t^2-x^2)^{-1/2}\frac{d}{dt}[tk_2(t)]\,dt,$$

$$k_1(t)=g_1(t)$$

and

$$k_2(t)=K_{-1/2,1}[g_2]=2t^{-1}\int_t^1 g_2(u)\,du,$$

the system (5.22), (5.23) yields the generalized Abel system,

$$\frac{a}{x}\int_x^1 (t^2-x^2)^{-1/2}tg_1(t)\,dt-b\int_0^x (x^2-t^2)^{-1/2}g_2(t)\,dt=\frac{\Gamma(\frac{1}{2})}{2}\hat{F}_1(x), \qquad 0<x<1, \tag{5.24}$$

$$\frac{c}{x}\int_0^x (x^2-t^2)^{-1/2}tg_1(t)\,dt+d\int_x^1 (t^2-x^2)^{-1/2}g_2(t)\,dt=\frac{\Gamma(\frac{1}{2})}{2}\hat{F}_2(x), \qquad 0<x<1. \tag{5.25}$$

REFERENCES

[1] A. ERDÉLYI AND I. N. SNEDDON, *Fractional integration and dual integral equations*, Canad., J. Math., 14 (1962), pp. 685–693.

[2] F. ERDOGAN, *Stress distribution in bonded dissimilar materials with cracks*, J. Appl. Mech., 32 (1965), pp. 403–410.

[3] M. LOWENGRUB, *Systems of Abel type integral equations in function theoretic methods in differential equations*, R. P. Gilbert and R. J. Weinacht, ed., 1976, pp. 277–296.

[4] M. LOWENGRUB AND I. N. SNEDDON, *The stress field near a Griffith crack at the interface of two bonded dissimilar elastic half-planes*, Internat. J. Engrg. Sci., 11 (1973), pp. 1025–1034.

[5] N. I. MUSKHELISHVILI, *Some Basic Problems in the Mathematical Theory of Elasticity*, P. Noordhoff, Groningen, The Netherlands, 1965.

[6] K. D. SAKALYUK, *Abel's generalized integral equation*, Dokl. Akad. Nauk. SSSR, 131, 4 (1960), pp. 332-335.

[7] S. G. SAMKO, *The generalized Abel equation and fractional integration operators*, Differ. Uravn., 4 (1968), no. 2, pp. 157–166.

[8] ———, *Solution of generalized Abel equation by means of an equation with Cauchy kernal*, Soviet Math. Dokl., 8 (1967), no. 5, pp. 1259–1262.

[9] I. N. SNEDDON, *Mixed Boundary Value Problems in Elasticity*, North-Holland, Amsterdam, 1966.

[10] R. A. WESTMANN, *Simultaneous pairs of dual integral equations*, SIAM Rev., 7 (1965), pp. 341–348.

[11] L. v. WOLFERSDORF, *Zur Lösung der verallgemeinerten Abelschen Integralgleichnung mit konstanten Koeffizienten*, Z. Angew. Math. Mech., 49 (1969), no. 12, pp. 759–761.

[12] ———, *Uber eine Beziehung zwischen Integralen nichtganzer Ordnung*, Math. Z., 90 (1965), pp. 24–28.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0016$01.00/0

SYSTEMS OF GENERALIZED ABEL INTEGRAL EQUATIONS WITH APPLICATIONS TO SIMULTANEOUS DUAL RELATIONS*

J. R. WALTON†

Abstract. A method is presented for solving certain systems of generalized Abel integral equations by constructing an equivalent system of singular integral equations. An application is then given to a class of simultaneous dual relations of a type arising in bimedia fracture problems in elasticity. The equations discussed in this paper generalize those considered in an earlier paper of Lowengrub and Walton [SIAM J. Math. Anal., this issue, pp. 794–807].

1. Introduction. In this paper a method is presented for solving systems of generalized Abel integral equations of the type

$$\begin{aligned} &a_1(x^p)\int_0^x \frac{\alpha_1(t^p)\phi_1(t)}{(x^p-t^p)^{\mu_1}}\,dt + b_2(x^p)\int_x^1 \frac{\beta_2(t^p)\phi_2(t)}{(t^p-x^p)^{\mu_2}}\,dt = f_1(x), \\ &b_1(x^p)\int_x^1 \frac{\beta_1(t^p)\phi_1(t)}{(t^p-x^p)^{\mu_1}}\,dt + a_2(x^p)\int_0^x \frac{\alpha_2(t^p)\phi_2(t)}{(x^p-t^p)^{\mu_2}}\,dt = f_2(x). \end{aligned} \qquad 0<x<1, \tag{1}$$

Since only the cases $p = 1$ or $p = 2$ occur in applications, we shall restrict the subsequent discussion to those cases.

The equations (1) are a generalization of those analyzed in [3] for which $\mu_1 = \mu_2$ and $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$. That reference also includes a discussion of an application of such systems to problems in elasticity. In particular, a method was presented in [3] for reducing a simultaneous set of dual relations involving Hankel transforms to a simultaneous system of fractional integral and differential operators. Under certain conditions the systems obtained in that way are equivalent to one of the systems of Abel type equations for which closed form solutions were constructed in [3]. However, the conditions that must be imposed upon such simultaneous dual relations to yield Abel systems within the scope of the techniques of [3] are very restrictive.

In contrast, the method presented here is applicable to a very large class of simultaneous dual relations. Although, in general, simple closed form solutions are not obtained, it is demonstrated that the dual relations may be transformed into Abel system (1) with $p = 2$, which in turn is shown to be equivalent to a system of singular integral equations with Cauchy dominant singular part. Such systems have been studied extensively (see [2], [5]) and yield important theoretical information about the simultaneous dual relations. For example, the Noether theorems provide information on the questions of existence and uniqueness. Moreover, it can be shown that the nature of singularities exhibited by solutions of the resulting singular integral equations may be deduced from the behavior of the dominant singular part of these equations. However, this is the subject of a subsequent paper in which an application of equations similar to (1) with $p = 1$ to the study of certain bimedia fracture problems for power law viscoelastic solids is presented. In particular, in that paper it will be shown that knowledge of the singularities of solutions of the associated singular integral equations yields information about the occurrence of singular stresses in the physical problem. That generalized Abel type integral equations can play an important role in solving realistic problems for power law viscoelastic materials has already been demonstrated (see [9]).

* Received by the editors December 30, 1977, and in revised form April 26, 1978.

† Department of Mathematics, Texas A&M University, College Station, Texas 77843. This work was supported in part by the United States Air Force under AFOSR Grant 77-3290.

Very few theoretical results for simultaneous dual relations have appeared in the literature. The method described in this paper provides a means of pursuing such investigations and is useful for obtaining insight into the nature of simultaneous dual relations. We do not attempt to present a rigorous analysis of the dual relations considered here. However, it should be straightforward, albeit tedious, to do so by justifying our formal manipulations within the distributional framework employed in [7] and [8] or that developed by Braaksma and Schuitman [1], or that by A. C. McBride [4].

In §§ 2 and 3 we consider the reduction of (1) to a system of singular integral equations for $p=1$ and $p=2$ respectively. Section 4 contains the application to simultaneous dual relations.

2. First Abel system. In this section we consider the generalized Abel system (1) with $p=1$. It is assumed that

$$\begin{aligned} &\alpha_j(t)=\alpha_j^*(t)t^{\nu_j} \quad \text{and} \\ &\beta_j(t)=\beta_j^*(t)(1-t)^{\lambda_j} \qquad j=1,2, \end{aligned} \tag{2}$$

where $\alpha_j^*(t)$ and $\beta_j^*(t)$ are continuously differentiable on $[0, 1]$ and nonvanishing at the endpoints. It will prove convenient to introduce the following notation:

$$\nu_j'=\min(0,\nu_j), \qquad \lambda_j'=\min(0,\lambda_j), \tag{3}$$

$$\begin{aligned} \tilde{\alpha}_j(t)&=\alpha_j(t)t^{-\nu_j'}(1-t)^{-\lambda_j'}, \\ \tilde{\beta}_j(t)&=\beta_j(t)t^{-\nu_j'}(1-t)^{-\lambda_j'}, \end{aligned} \tag{4}$$

$$\tilde{\phi}_j(t)=\phi_j(t)t^{\nu_j'}(1-t)^{\lambda_j'}, \tag{5}$$

$$\begin{aligned} I_j(\phi_j)&=\int_0^x \frac{\alpha_j(t)\phi_j(t)}{(x-t)^{\mu_j}}\,dt, \\ K_j(\phi_j)&=\int_x^1 \frac{\beta_j(t)\phi_j(t)}{(t-x)^{\mu_j}}\,dt, \end{aligned} \tag{6}$$

$$h_j(x)=\int_0^x \frac{\tilde{\phi}_j(t)}{(x-t)^{\mu_j}}\,dt, \tag{7}$$

$$k_j(x)=\int_x^1 \frac{\tilde{\phi}_j(t)}{(t-x)^{\mu_j}}\,dt, \tag{8}$$

$$R_j(z)=[z(1-z)]^{(1-\mu_j)/2} \tag{9}$$

and

$$\Phi_j(z)=[R_j(z)]^{-1}\int_0^1 \frac{\tilde{\phi}_j(t)}{(t-z)^{\mu_j}}\,dt. \tag{10}$$

We seek solutions $\phi_j(t)$, $j=1, 2$ such that

$$\tilde{\phi}_j(t)=\frac{\phi_j^*(t)}{[t(1-t)]^{1-\mu_j-\varepsilon}} \tag{11}$$

where $\phi_j^*(t)$ is Hölder continuous on $[0, 1]$ and ε is a positive number. It follows that when a suitable branch is chosen for the multivalued function $[R_j(z)(t-z)^{\mu_j-1}]$, then

$\Phi_j(z)$ is analytic in the complex plane cut along $[0, 1]$ and satisfies the asymptotic estimates

$$\Phi_j(z)=O\left(\frac{1}{z}\right) \quad \text{as } z\to\infty \tag{12a}$$
$$\Phi_j(z)=O(z^{(\mu_j-1)/2}) \quad \text{as } z\to 0$$

and

$$\Phi_j(z)=O((1-z)^{(\mu_j-1)/2}) \quad \text{as } z\to 1. \tag{12b}$$

We remark also that $h_j(0+)=k_j(1-)=I_j(\phi_j)(0+)=K_j(\phi_j)(1-)=0$ and $h_j(1-)$, $k_j(0+)$, $I_j(\phi_j)(1-)$ and $I_j(\phi_j)(0+)$ are all finite.

Let $\Phi_j^{\pm}(x)$ denote the following limits:

$$\left.\begin{aligned}\Phi_j^+(x)&=\lim_{\substack{z\to x\\ I_m(z)>0}}\Phi_j(z),\\ \Phi_j^-(x)&=\lim_{\substack{z\to x\\ I_m(z)<0}}\Phi_j(z),\end{aligned}\right\}\quad 0<x<1.$$

It is then readily verified that

$$h_j(x)=\left[\frac{e^{\mu_j\pi i}\Phi_j^+(x)+\Phi_j^-(x)}{(e^{2\mu_j\pi i}-1)}\right]R_j(x) \quad \text{and} \tag{13}$$

$$k_j(x)=-\left[\frac{\Phi_j^+(x)+e^{\mu_j\pi i}\Phi_j^-(x)}{(e^{2\mu_j\pi i}-1)}\right]R_j(x). \tag{14}$$

From lines (2)–(5) we observe that

$$\tilde{\phi}_j(t)=\frac{\sin\mu_j\pi}{\pi}\int_0^t\frac{h_j'(y)\,dy}{(t-y)^{1-\mu_j}}, \tag{15}$$

$$=-\frac{\sin\mu_j\pi}{\pi}\int_t^1\frac{k_j'(y)\,dy}{(y-t)^{1-\mu_j}}. \tag{16}$$

Moreover, substitution of (15) into (6) followed by an integration by parts, yields

$$I_j(\phi_j)(x)=h_j(x)\tilde{\alpha}_j(x)-\frac{\sin\mu_j\pi}{\pi}\int_0^x h_j(y)K_{1,j}(x,y)\,dy \tag{17}$$

where

$$K_{1,j}(x,y)=\int_0^1\frac{\tilde{\alpha}_j'(s(x-y)+y)\,ds}{(1-s)^{\mu_j-1}s^{1-\mu_j}}.$$

It should be observed that $K_{1,j}(x, y)$ is continuous for $0<y<x<1$ and integrable on the triangle $0\leqq y\leqq x\leqq 1$.

Similarly, as a consequence of (16), we have

$$K_j(\phi_j)(x)=k_j(x)\tilde{\beta}_j(x)+\frac{\sin\mu_j\pi}{\pi}\int_x^1 k_j(y)K_{2,j}(x,y)\,dy \tag{18}$$

where

$$K_{2,j}(x,y)=\int_0^1\frac{\tilde{\beta}_j'(s(y-x)+y)\,ds}{(1-s)^{\mu_j-1}s^{1-\mu_j}},$$

and $K_{2,j}(x, y)$ is continuous for $0<x<y<1$ and integrable on $0\leqq x\leqq y\leqq 1$.

Substitution of (13), (14), (17) and (18) into (1) yields the generalized Riemann boundary equation

$$A(x)\Phi^+(x)+B(x)\Phi^-(x)+K(\Phi^+)+H(\Phi^-)=F(x) \tag{19}$$

where $\Phi=(\Phi_1, \Phi_2)^T$, $F=(f_1, f_2)^T$ and $A=(a_{ij})$, $B=(b_{ij})$, $K=(k_{ij})$ and $H=(h_{ij})$ are matrices given by

$$a_{11}=-ia_1(x)\tilde{\alpha}_1(x)R_1(x)/(2\sin\mu_1\pi),$$

$$a_{12}=ib_2(x)\tilde{\beta}_2(x)R_2(x)\,e^{-\mu_2\pi i}/(2\sin\mu_2\pi),$$

$$a_{21}=ib_1(x)\tilde{\beta}_1(x)R_1(x)\,e^{-\mu_1\pi i}/(2\sin\mu_1\pi),$$

$$a_{22}=-ia_2(x)\tilde{\alpha}_2(x)R_2(x)/(2\sin\mu_2\pi),$$

$$b_{11}=a_{11}\,e^{-\mu_1\pi i},\qquad b_{12}=a_{12}\,e^{\mu_2\pi i},\qquad b_{21}=a_{21}\,e^{\mu_1\pi i},\qquad b_{22}=a_{22}\,e^{-\mu_2\pi i},$$

$$k_{11}(\phi)=\frac{i}{2\pi}a_1(x)\int_0^x \phi(y)R_1(y)K_{1,1}(x, y)\,dy,$$

$$k_{12}(\phi)=\frac{i}{2\pi}b_2(x)\,e^{-\mu_2\pi i}\int_x^1 \phi(y)R_2(y)K_{2,2}(x, y)\,dy,$$

$$k_{21}(\phi)=\frac{i}{2\pi}b_1(x)\,e^{-\mu_1\pi i}\int_x^1 \phi(y)R_1(y)K_{2,1}(x, y)\,dy,$$

$$k_{22}(\phi)=\frac{i}{2\pi}a_2(x)\int_0^x \phi(x)R_2(y)K_{1,2}(x, y)\,dy,$$

$$h_{11}=k_{11}\,e^{-\mu_1\pi i},\quad h_{12}=k_{12}\,e^{\mu_2\pi i},\quad h_{21}=k_{21}\,e^{\mu_1\pi i}\quad\text{and}\quad h_{22}=k_{22}\,e^{-\mu_2\pi i}.$$

It is straightforward to verify that solving the Abel system (1) in the class (11) is equivalent to solving the generalized Riemann system (19), i.e., to finding sectionally analytic functions $\Phi_j(z)$, $j=1, 2$, satisfying (12) and the generalized boundary equation (19).

The generalized Riemann problem (19) may be transformed, in turn, into an equivalent system of singular integral equations with Cauchy dominant singular part by a method outlined in Gakhov [2]. In particular, define

$$\Phi_j(z)=\frac{1}{2\pi i}\int_0^1 \frac{\psi_j(t)\,dt}{(t-z)}, \tag{20}$$

and recall the Plemelj formulas [4]

$$\Phi_j^{\pm}(x)=\pm\frac{1}{2}\psi_j(x)+\frac{1}{2\pi i}\int_0^1 \frac{\psi_j(t)\,dt}{(t-x)}. \tag{21}$$

Substitution of (20) and (21) into (19) yields the system of singular integral equations

$$K^0(\psi)+k(\psi)=f, \tag{22}$$

where $\psi=(\psi_1, \psi_2)^T$ and $f=(f_1, f_2)^T$. The dominant singular part of (22), $K^0(\psi)$, is given by

$$K^0(\psi)=S(x)\psi(x)+T(x)\frac{1}{\pi i}\int_0^1 \frac{\psi(t)\,dt}{(t-x)}, \tag{23}$$

where $S=(A-B)/2$ and $T=(A+B)/2$. The operator $k(\psi)$ in (22) is easily seen to be Fredholm. The theory of systems of the type (22) is well developed. (See, for example, Muskhelishvili [5].) In particular it is known that (22) is equivalent to a system of Fredholm equations of the second kind. In this sense we may regard the above analysis as providing a solution of (1), although except for special cases, the solution is not obtainable in closed form. We remark further, that when $\mu_1=\mu_2$, the dominant singular part of (22), $K^0(\psi)$, may be substantially simplified. Important theoretical information about (1) may be obtained from (22). However, we shall postpone a consideration of this until § 4 when an application of (1) to dual relations is discussed.

3. Second Abel system. We next consider the system (1) with $p=2$. The technique employed for this case is in the same spirit as that of the previous section, with only slight modifications made necessary by the substitution of the nonunivalent function z^2 for z.

As in § 2, it is convenient to introduce certain notation. Assumptions (2) and definitions (3) and (4) are unchanged. Whereas, lines (5)–(10) are replaced by:

(24)
$$\tilde{\phi}_j(t)=\phi_j(t)t^{2\nu_j'}(1-t^2)^{\lambda_j'},$$

(25)
$$I_j(\phi_j)=\int_0^x \frac{\alpha_j(t^2)\phi_j(t)\,dt}{(x^2-t^2)^{\mu_j}},$$
$$K_j(\phi_j)=\int_x^1 \frac{\beta_j(t^2)\phi_j(t)\,dt}{(t^2-x^2)^{\mu_j}},$$

(26)
$$h_j(x)=\int_0^x \frac{\tilde{\phi}_j(t)\,dt}{(x^2-t^2)^{\mu_j}},$$

(27)
$$k_j(x)=\int_x^1 \frac{\tilde{\phi}_j(t)\,dt}{(t^2-x^2)^{\mu_j}},$$

(28)
$$R_j(z)=[z^2-1]^{1/2-\mu_j}$$

and

(29)
$$\Phi_j(z)=\frac{1}{R_j(z)}\int_0^1 \frac{\tilde{\phi}_j(t)\,dt}{(z^2-t^2)^{\mu_j}}.$$

Instead of (11), we now seek solutions $\phi_j(t)$ such that

(30)
$$\tilde{\phi}_j(t)t^{-\mu_j}=\frac{\phi_j^*(t)}{[t(1-t)]^{1-\mu_j-\varepsilon}}.$$

When a suitable branch is chosen for $(z^2-t^2)^{-\mu_j}/R_j(z)$, we see from (30), that $\Phi_j(z)$ is analytic in the complex plane cut along $[-1,1]$ and satisfies

(31)
$$\Phi_j(z)=O\left(\frac{1}{z}\right) \quad \text{as } z\to\infty \quad \text{and}$$
$$\Phi_j(z)=O((z^2-1)^{\mu_j-1/2}) \quad \text{as } z\to\pm 1.$$

Moreover, we may conclude that $h_j(0+)=k_j(1-)=I_j(\phi_j)(0+)=K_j(\phi_j)(1-)=0$ and $h_j(1-)$, $k_j(0+)$, $I_j(\phi_j)(1-)$ and $K_j(\phi_j)(0+)$ are all finite.

The limits $\Phi^{\pm}(x)$ are defined as before, only now they are computed for $-1<x<1$. It is easy to show that when $0<x<1$,

$$h_j(x)=[\Phi_j^+(x)+\Phi_j^-(x)]\frac{(1-x^2)^{1/2-\mu_j}}{2\sin\mu_j\pi} \tag{32}$$

and

$$k_j(x)=-[e^{-\mu_j\pi i}\Phi_j^+(x)+e^{\mu_j\pi i}\Phi_j^-(x)]\frac{(1-x^2)^{1/2-\mu_j}}{2\sin\mu_j\pi}, \tag{33}$$

and when $-1<x<0$

$$\Phi_j^{\pm}(x)=-\overline{\Phi_j^{\mp}(-x)}. \tag{34}$$

From (26) and (27) we obtain

$$\tilde{\phi}_j(t)=\frac{\sin\mu_j\pi}{\pi}2t\int_0^t\frac{h_j'(y)\,dy}{(t^2-y^2)^{1-\mu_j}} \tag{35}$$

$$=-\frac{\sin\mu_j\pi}{\pi}2t\int_t^1\frac{k_j'(y)\,dy}{(y^2-t^2)^{1-\mu_j}}. \tag{36}$$

Corresponding to (17) and (18), substitution of (35) and (36) into (25) yields

$$I_j(\phi_j)=h_j(x)\tilde{\alpha}_j(x^2)-\frac{\sin\mu_j\pi}{\pi}\int_0^x h_j(y)yK_{1,j}(x,y)\,dy \tag{37}$$

and

$$K_j(\phi_j)=k_j(x)\tilde{\beta}_j(x^2)+\frac{\sin\mu_j\pi}{\pi}\int_x^1 k_j(y)yK_{2,j}(x,y)\,dy \tag{38}$$

where

$$K_{1,j}(x,y)=2\int_0^1\frac{\tilde{\alpha}_j'(\sigma(x^2-y^2)+y^2)\,d\sigma}{(1-\sigma)^{\mu_j-1}\sigma^{1-\mu_j}} \tag{39}$$

and

$$K_{2,j}(x,y)=2\int_0^1\frac{\tilde{\beta}_j'(\sigma(x^2-y^2)+y^2)\,d\sigma}{(1-\sigma)^{\mu_j-1}\sigma^{1-\mu_j}}. \tag{40}$$

Substitution of (37) and (38) into (1) yields the Riemann boundary system, valid for $0<x<1$,

$$A(x)\Phi^+(x)+B(x)\Phi^-(x)+K(\Phi^+)+H(\Phi^-)=f(x) \tag{41}$$

where

$$\Phi(z)=(\Phi_1(z),\Phi_2(z))^T,$$
$$f(x)=(f_1(x),f_2(x))^T,$$
$$A(x)=(a_{ij}(x)),\quad B(x)=\overline{A(x)},\quad K=(k_{ij})\quad\text{and}\quad H=(h_{ij})$$

with

$$a_{11}(x)=\frac{a_1(x^2)\tilde{\alpha}_1(x^2)}{2\sin\mu_1\pi}(1-x^2)^{1/2-\mu_1},$$

$$a_{12}(x)=-\frac{b_2(x^2)\tilde{\beta}_2(x^2)}{2\sin\mu_2\pi}(1-x^2)^{1/2-\mu_2}\,e^{-\mu_2\pi i},$$

$$a_{21}(x)=-\frac{b_1(x^2)\tilde{\beta}_1(x^2)}{2\sin\mu_1\pi}(1-x^2)^{1/2-\mu_1}\,e^{-\mu_1\pi i},$$

$$a_{22}(x)=\frac{a_2(x^2)\tilde{\alpha}_2(x^2)}{2\sin\mu_2\pi}(1-x^2)^{1/2-\mu_2},$$

$$k_{11}(\phi)=-\frac{a_1(x^2)}{2\pi}\int_0^x \phi(y)(1-y^2)^{1/2-\mu_1}yK_{1,1}(x,y)\,dy,$$

$$k_{12}(\phi)=-\frac{b_2(x^2)}{2\pi}\,e^{-\mu_2\pi i}\int_x^1 \phi(y)(1-y^2)^{1/2-\mu_2}yK_{2,2}(x,y)\,dy,$$

$$k_{21}(\phi)=-\frac{b_1(x^2)}{2\pi}\,e^{-\mu_1\pi i}\int_x^1 \phi(y)(1-y^2)^{1/2-\mu_1}yK_{2,1}(x,y)\,dy$$

and

$$k_{22}(\phi)=-\frac{a_2(x^2)}{2\pi}\int_0^x \phi(y)(1-y^2)^{1/2-\mu_2}yK_{1,2}(x,y)\,dy.$$

The kernel of the operator h_{ij} is conjugate to that of k_{ij}.

To establish the equivalence of (1) to a system of Riemann boundary value problems it is necessary to extend the boundary equation (41) to all of $(-1, 1)$. However, from (34) and the fact that for $0<x<1$ $K(\Phi^+)+H(\Phi^-)$ is real it is clear how the extension should be effected. Specifically, if we define for $-1<x<0$

$$\begin{aligned}
a_j(x)&=a_j(-x),\\
b_j(x)&=b_j(-x),\\
k_{ij}(\phi)(x)&=k_{ij}(\phi)(-x),\\
\tilde{\alpha}_j(x)&=-\tilde{\alpha}_j(-x),\\
\beta_j(x)&=-\beta_j(-x)
\end{aligned}$$

and

$$f_j(x)=f_j(-x),$$

then

$$A(x)=-\overline{A(-x)}$$

and

$$B(x)=-\overline{B(-x)}$$

and (41) is valid for $-1<x<1$.

By an argument entirely analogous to that of § 2, the introduction of $\psi_j(t)$ through

$$\Phi_j(z)=\frac{1}{2\pi i}\int_{-1}^{1}\frac{\psi_j(t)\,dt}{(t-z)}$$

transforms (41) to an equivalent system of singular integral equations with Cauchy dominant singular part. In the next section we consider an application of (1) with $p=2$ to certain simultaneous dual relations.

4. Simultaneous dual relations. In this section we consider an application of the analysis presented in § 3 to simultaneous dual integral equations of the form

$$\begin{aligned}
&\int_0^\infty [a_1A(\xi)+b_1B(\xi)]\xi^{-2\alpha}J_\mu(\xi x)\,d\xi=F_1(x),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 0<x<1,\\
&\int_0^\infty [a_2A(\xi)+b_2B(\xi)]\xi^{-2\beta}J_\nu(\xi x)\,d\xi=F_2(x),\\
(42)\qquad&\\
&\int_0^\infty A(\xi)J_\mu(\xi x)\,d\xi=0,\\
&\qquad\qquad\qquad\qquad 1<x<\infty,\\
&\int_0^\infty B(\xi)J_\nu(\xi x)\,d\xi=0,
\end{aligned}$$

where a, b, c and d are constants. As was remarked in [3], such systems arise in bimedia fracture problems in elasticity. It was demonstrated in [3] that the system (42) may be transformed into the system

$$(43)\qquad \begin{aligned}
&a_1K_{\mu/2-\alpha,\nu-\lambda}[\eta_1]+b_1I_{\nu/2,(\mu-\nu)/2}[\eta_2]=f_1(x),\\
&a_2I_{\mu/2,(\nu-\mu)/2}[\eta_1]+b_2K_{\nu/2-\beta,\lambda-\nu}[\eta_2]=f_2(x),
\end{aligned}\qquad 0<x<1,$$

where

$$\lambda=\frac{\mu+\nu}{2}-(\alpha-\beta),$$

$$f_1(x)=2^{2\alpha}I_{\mu/2+\alpha,-\alpha}\{F_1(\xi)\xi^{-2\alpha};x\},$$

$$f_2(x)=2^{2\beta}I_{\nu/2+\beta,-\beta}\{F_2(\xi)\xi^{-2\beta};x\}$$

and η_1 and η_2 are unknown functions which vanish on $(1,\infty)$. The operators appearing in (43) are the modified Erdelyi–Kober operators introduced by Sneddon [6] and are defined as follows: if $\alpha>0$ then $I_{\eta,\alpha}$ and $K_{\eta,\alpha}$ denote the fractional integral operators

$$I_{\eta,\alpha}\{f(\xi);x\}=\frac{2}{\Gamma(\alpha)}x^{-2\eta-2\alpha}\int_0^x(x^2-\xi^2)^{\alpha-1}\xi^{2\eta+1}f(\xi)\,d\xi,$$

$$K_{\eta,\alpha}\{f(\xi);x\}=\frac{2}{\Gamma(\alpha)}x^{2\eta}\int_x^\infty(\xi^2-x^2)^{\alpha-1}\xi^{-2\eta-2\alpha+1}f(\xi)\,d\xi;$$

whereas, if $\alpha<0$, $I_{\eta,\alpha}$ and $K_{\eta,\alpha}$ are the fractional differential operators inverse to $I_{\eta+\alpha,-\alpha}$ and $K_{\eta+\alpha,-\alpha}$ respectively.

The system (43) may be regarded as a generalized Abel system. As was indicated in [3], only special cases of (43) fall within the class of Abel systems considered in that paper and for which simply closed form solutions are obtainable. In contrast, we show here that the full problem (43) may be treated by the methods of § 3. Although this approach does not provide, in general, closed form solutions of (42), it does offer a means of obtaining useful theoretical information regarding the questions of existence and uniqueness. Moreover, the system (42) is ultimately reduced to a system of Fredholm equations of the second kind.

Two observations regarding the general character of the system (43) may be made immediately. The first is that both fractional integral and differential operators appear in (43). In particular, the four operators consist of two fractional integrals and their inverses. The second observation is that only when $\alpha=\beta$ does it occur that the unknown functions, η_i, appear in operators of the same order; i.e. in (43) the two integral operators have the same order and thus also their inverses. As will become apparent later, this greatly affects the tractability of (43).

Without loss of generality we may assume $\nu>\mu$. Also, for simplicity we shall first assume that $\alpha=\beta$ and $\nu-\mu<2$, which still includes all the physically interesting cases. Later, we shall indicate the necessary adjustments in the analysis to be made when these assumptions are relaxed.

Given these restrictions, (43) becomes

$$\text{(44)}\qquad \begin{aligned} a_1K_{\mu/2-\alpha,(\nu-\mu)/2}[\eta_1]+b_1I_{\nu/2,(\mu-\nu)/2}[\eta_2]&=f_1(x),\\ a_2I_{\mu/2,(\nu-\mu)/2}[\eta_1]+b_2K_{\nu/2-\alpha,(\mu-\nu)/2}[\eta_2]&=f_2(x),\end{aligned}\qquad 0<x<1.$$

Introduction of

$$\phi_1(t)=t^{2\alpha+\mu+1}\eta_1(t),\quad \mu_1=1-(\nu-\mu)/2,\quad \alpha_1(t)=t^{\nu_1},$$
$$\nu_1=(\mu+\nu)/2-\alpha \quad\text{and}\quad \beta_1(t)=1$$

yields, in the notation of § 3,

$$\text{(45)}\qquad K_{\mu/2-\alpha,(\nu-\mu)/2}[\eta_1]=\frac{2x^{\mu-2\alpha}}{\Gamma((\nu-\mu)/2)}K_1(\phi_1)$$

and

$$\text{(46)}\qquad I_{\mu/2,(\nu-\mu)/2}[\eta_1]=\frac{2x^{-\nu}}{\Gamma\left(\frac{\nu-\mu}{2}\right)}I_1(\phi_1).$$

It should be noted that $\tilde{\alpha}_1$ and $\tilde{\beta}_1$, defined by (4), are now only power functions and that one of them is identically one. Moreover, the kernels $K_{1,1}$ and $K_{2,1}$ in (37) and (38) are easily seen to be given by

$$\text{(47)}\qquad \frac{\sin\mu_1\pi}{\pi}K_{1,1}(x,y)$$
$$=\begin{cases}0, & \nu_1\leqq 0,\\ 2\nu_1(1-\mu_1)x^{2(\nu_1-1)}\,{}_2F_1(1-\nu_1,2-\mu_1;2;x^2-y^2/x^2), & \nu_1>0,\ 0<x<y,\end{cases}$$

and

$$\text{(48)}\qquad \frac{\sin\mu_1\pi}{\pi}K_{2,1}(x,y)$$
$$=\begin{cases}-2\nu_1(1-\mu_1)y^{-2(\nu_1+1)}\,{}_2F_1(1+\nu_1,\mu_1;2;y^2-x^2/y^2), & 0<x<y,\ \nu_1<0,\\ 0, & \nu_1\geqq 0,\ 0<y<x.\end{cases}$$

Since solutions are sought for which the operators in (44) yield continuous functions, it follows that $t^{\nu}\eta_2(t)$ must vanish for $x=0$ and $x=1$. Hence we deduce that

$$(49)\qquad I_{\nu/2,(\mu-\nu)/2}[\eta_2]=\frac{x^{-\mu}}{\Gamma(1-(\nu-\mu)/2)}\int_0^x (x^2-t^2)^{(\mu-\nu)/2}\frac{d}{dt}[t^{\nu}\eta_2(t)]\,dt$$

and

$$(50)\qquad K_{\nu/2-\alpha,(\mu-\nu)/2}[\eta_2]=\frac{-x^{\nu-2\alpha}}{\Gamma(1-(\nu-\mu)/2)}\int_x^1 (t^2-x^2)^{(\mu-\nu)/2}\frac{d}{dt}[t^{2\alpha-\mu}\eta_2(t)]\,dt.$$

Define $\nu_2=(\nu+\mu)/2-\alpha$. If $\nu_2=0$, we define

$$\phi_2(t)=\frac{d}{dt}[t^{\nu}\eta_2(t)],\quad \alpha_2(t)=1,\quad \beta_2(t)=1\quad\text{and}\quad \mu_2=\frac{\nu-\mu}{2}$$

and observe that

$$(51)\qquad I_{\nu/2,(\mu-\nu)/2}[\eta_2]=\frac{x^{-\mu}}{\Gamma(1-\mu_2)}I_2(\phi_2)$$

and

$$(52)\qquad K_{\nu/2-\alpha,(\mu-\nu)/2}[\eta_2]=\frac{-x^{\nu-2\alpha}}{\Gamma(1-\mu_2)}K_2(\phi_2),$$

where $I_2(\phi_2)$ and $K_2(\phi_2)$ are given by (25).

If $\nu_2\neq 0$, we define $\alpha_2(t)=t^{\nu_2}$ and

$$\phi_2(t)=t^{-2\nu_2}\frac{d}{dt}[t^{\nu}\eta_2(t)]$$

and note that

$$t^{2\alpha-\mu}\eta_2(t)=t^{-2\nu_2}t^{\nu}\eta_2(t).$$

Line (51) is still valid but (52) must be amended. From the obvious identity

$$\frac{d}{dt}[t^{2\alpha-\mu}\eta_2(t)]=\phi_2(t)-2\nu_2t^{-2\nu_2-1}[t^{\nu}\eta_2(t)]$$

we obtain

$$(53)\qquad K_{\nu/2-\alpha,(\mu-\nu)/2}[\eta_2]=\frac{-x^{\nu-2\alpha}}{\Gamma(1-\mu_2)}K_2(\phi_2)-2\nu_2\int_x^1\frac{t^{-2\nu_2-1}[t^{\nu}\eta_2(t)]}{(t^2-x^2)^{\mu_2}}\,dt.$$

Moreover, it is straightforward to show that

$$(54)\quad t^{\nu}\eta_2(t)=\frac{-\sin\eta_2\pi}{\pi}\left[\tilde{\alpha}_2(t^2)\int_t^1\frac{2yk_2(y)\,dy}{(y^2-t^2)^{1-\mu_2}}+\int_t^1 2yk_2(y)\,dy\int_t^y\frac{\tilde{\alpha}_2'(z^2)2z\,dz}{(y^2-x^2)^{1-\mu_2}}\right].$$

We must now consider separately the two cases $\nu_2>0$ and $\nu_2<0$. For $\nu_2>0$ substitution of (54) into (53) yields

$$(55)\qquad K_{\nu/2-\alpha,(\mu-\nu)/2}[\eta_2]=\frac{-x^{\nu-2\alpha}}{\Gamma(1-\mu_2)}K_2(\phi_2)-\int_x^1 k_2(y)K_{3,2}(x,y)\,dy$$

where

$$(56)\quad \begin{aligned} K_{3,2}(x, y) = {} & \frac{2\nu_2}{\Gamma(1-\mu_2)} y^{2\mu_2-1} x^{\mu-2\alpha} + \frac{2 \sin \mu_2\pi\nu_2^2}{\pi\Gamma(2-\mu_2)} x^{\nu-2\alpha}(y^2-x^2) \\ & \cdot \int_0^1 \frac{[x^2+\tau(y^2-x^2)]^{-2}}{\tau^{\mu_2-1}(1-\tau)^{1-\mu_2}} {}_2F_1\left(1+\nu_2, 1; 2-\mu_2; \frac{\tau(y^2-x^2)}{[x^2(1-\tau)+\tau y^2]}\right) d\tau. \end{aligned}$$

It should be noted that $x^\nu K_{3,2}(x, y) \in L^1(0, 1) \times (0, 1)$ and is continuous for $0 < x \leqq y \leqq 1$.

If $\nu_2 < 0$ we obtain

$$(57)\quad K_{\nu/2-\alpha,(\mu-\nu)/2}[\eta_2] = \frac{-x^{\nu-2\alpha}}{\Gamma(1-\mu_2)} K_2(\phi_2) - \int_x^1 k_2(y) K_{4,2}(x, y)\, dy$$

with

$$(58)\quad K_{4,2}(x, y) = \frac{2\nu_2 x^{\nu-2\alpha}}{\Gamma(1-\mu_2)} y^{-\nu_2} {}_2F_1\left(1+\nu_2, \mu_2; 1; \frac{y_2-x^2}{y^2}\right).$$

Moreover, we have $x^\nu K_{4,2}(x, y) \in L^1(0, 1) \times (0, 1)$ and continuous for $0 < x \leqq y \leqq 1$. The expressions $I_2(\phi_2)$ and $K_2(\phi_2)$ appearing in (51), (52), (55), and (57) are given by (37) and (38) with

$$K_{1,2}(x, y) = \begin{cases} 0, & \nu_2 \leqq 0,\ 0 < x < y < 1, \\ \dfrac{2\nu_2(1-\mu_2)}{\sin \mu_2\pi} x^{2(\nu_2-1)} {}_2F_1\left(1-\nu_2, 2-\mu_2; 2; \dfrac{x^2-y^2}{x^2}\right), & \nu_2 > 0,\ 0 < y < x < 1, \end{cases}$$

and

$$K_{2,2}(x, y) = \begin{cases} 0, & \nu_2 \geqq 0,\ 0 < y < x < 1, \\ \dfrac{-2\nu_2(1-\mu_2)\pi}{\sin \mu_2\pi} y^{-2(\nu_2+1)} {}_2F_1\left(1+\nu_2, \mu_2; 2; \dfrac{y^2-x^2}{y^2}\right), & \nu_2 < 0,\ 0 < x < y < 1. \end{cases}$$

The subsequent observations are valid for all values of ν_2. However, for definiteness we assume $\nu_2 > 0$. Substitution of (37) and (38) into (45), (46), (51) and (55) and from there into (44) yields a generalized Riemann system of the type

$$(59)\quad A(x)\Phi^+(x) + B(x)\Phi^-(x) + K(\Phi^+, \Phi^-) = f(x), \qquad 0 < x < 1.$$

Boundary equation (59) is then extended to all of $(-1, 1)$ by the method of § 3. Alternatively, (59) may be transformed into the system of singular integral equations

$$(60)\quad S(x)\psi(x) + T(x)\frac{1}{\pi i}\int_{-1}^1 \frac{\psi(t)\, dt}{(t-x)} + \hat{K}(\psi) = f(x), \qquad -1 < x < 1,$$

by introducing

$$\Phi(z) = \frac{1}{2\pi i}\int_{-1}^1 \frac{\psi(t)\, dt}{(t-z)}.$$

Examination of the dominant singular part of (60), or equivalently the principal part of (59), yields important theoretical information about (44). For example, the number of solutions of (59) (and hence (44)) is at least as large as the number of solutions of the dominant homogeneous singular equation

(61) $$S(x)\psi(x)+T(x)\frac{1}{\pi i}\int_{-1}^{1}\frac{\psi(t)\,dt}{(t-x)}=0,$$

or its corresponding Riemann problem

(62) $$A(x)\Phi^+(x)+B(x)\Phi^-(x)=0.$$

Examination of (61) or (62) will this provide conditions necessary for the existence of a unique solution to the dual relations (42). It therefore becomes necessary to compute the index of the system (62), which from the general theory of Muskhelishvili [5], is most easily determined by actually solving (62). This can be effected by transforming (62) in the usual way [5] to a system of Fredholm equations of the second kind and then iteratively constructing the solutions. However, in certain cases the technique presented in [3] for uncoupling systems of the type (60) into two ordinary uncoupled Riemann problems will provide simple closed form solutions. To decide the applicability of the method of [3] to (62) we must examine more closely the matrices $A(x)$ and $B(x)$, the components of which are given by $A(x)=(a_{ij}(x))$, $B(x)=\overline{A(x)}$, $A(-x)=-\overline{A(x)}$ and on $0<x<1$

$$a_{11}(x)=-a_1x^{2\nu_2-\nu}[1-x^2]^{1/2-\mu_1}\frac{\Gamma(\mu_1)}{\pi}e^{-\mu_1\pi i},$$

$$a_{12}(x)=b_1x^{2\nu_2-\mu}[1-x^2]^{1/2-\mu_2}\frac{\Gamma(\mu_2)}{2\pi},$$

$$a_{21}(x)=a_2x^{2\nu_2-\mu}[1-x^2]^{1/2-\mu_1}\frac{\Gamma(\mu_1)}{\pi},$$

$$a_{22}(x)=b_2x^{2\nu_2-\mu}[1-x^2]^{1/2-\mu_2}\frac{\Gamma(\mu_2)}{2\pi}e^{-\mu_2\pi i}.$$

The first restriction to be placed on $A(x)$ is that $\det(A)(x)\neq 0$ (except perhaps for $x=0$, ± 1). Hence it is assumed that

(63) $$a_1b_2\,e^{-\pi i(\mu_1+\mu_2)}+a_2b_1\neq 0.$$

Recalling that $\mu_1+\mu_2=1$, we see that (63) is equivalent to

(64) $$\delta\equiv a_1b_2-a_2b_1\neq 0.$$

It follows that when (64) holds, the system (62) is equivalent to

(65) $$\Phi^+(x)=-\frac{1}{\delta}G(x)\Phi^-(x),$$

where $G(x)=A^{-1}(x)\overline{A(X)}=(g_{ij}(x))$, $G(-x)=\overline{G(x)}$ and on (0, 1)

$$g_{11}(x)=a_2b_1+a_1b_2\,e^{\pi i(\mu_1-\mu_2)},$$

$$g_{12}(x)=b_1b_2x^{\nu-\mu}[1-x^2]^{\mu_1-\mu_2}\frac{\Gamma(\mu_2)}{\Gamma(\mu_1)}i\sin\mu_2\pi,$$

$$g_{21}(x)=-a_1a_2x^{\mu-\nu}[1-x^2]^{\mu_2-\mu_1}\frac{\Gamma(\mu_1)}{\Gamma(\mu_2)}4i\sin\mu_1\pi,$$

$$g_{22}(x)=\overline{g_{11}(x)}.$$

Uncoupling (65) requires finding a matrix $P(z)$ analytic in the plane cut along $(-1, 1)$ and such that

$$P^{-1}(x)G(x)P(x)=D(x), \qquad -1<x<1,$$

where $D(x)$ is diagonal. As is easily shown, $P(x)$ must be of the form

$$P(x)=c(x)\begin{pmatrix} g_{12}(x) & g_{12}(x) \\ r_1-g_{11}(x) & r_2-g_{11}(x)\end{pmatrix}$$

where $c(x)$ is a scalar function and r_1, r_2 are the eigenvalues of $G(x)$ given by

$$r_1, r_2=(g_{11}+g_{22})/2\pm\sqrt{(g_{11}-g_{22})^2/4+g_{12}g_{21}}.$$

Since r_1-g_{11} and r_2-g_{11} are constants, the matrix $P(x)$ has an analytic extension to the cut plane if and only if $x^{\mu-\nu}[1-x^2]^{\mu_2-\mu_1}$ has no branch point at infinity. Recalling that $\mu_1-\mu_2=1-(\nu-\mu)$ and $0<\nu-\mu<2$, we conclude that (65) can be uncoupled if and only if $\nu-\mu=1$. This condition is satisfied by the systems arising in applications to bimedia crack problems in elasticity [3]. When the restriction $0<\nu-\mu<2$ is withdrawn, it becomes apparent that (65) uncouples whenever $\nu-\mu$ is an odd integer. When $\nu-\mu$ is an even integer, the original system (43) was shown in [3] to reduce to a single linear ordinary differential equation. Hence, it is apparent that (43) is greatly simplified when $\nu-\mu$ is an integer. Moreover, a simple closed form solution is obtained whenever $\nu_1=\nu_2=0$. It now is a simple matter to compute the indices of the two uncoupled Riemann problems and obtain conditions for the solvability of (44). In particular, the indices provide necessary conditions for uniqueness to hold for the dual relations (42).

It remains to consider how removing the restrictions $0<\nu-\mu<2$ and $\alpha=\beta$ affects the analysis of (44). Maintaining $\alpha=\beta$ but allowing $\nu-\mu$ to be any positive number does not affect the general character of (44) and requires only minor alterations in the analysis presented above. Since the calculations involved are rather tedius we shall dispense with a detailed analysis of this case. However, if $\alpha\neq\beta$ the behavior of (43) is substantially different from that of (44). To illustrate the difference we shall consider (43) with $\alpha>\beta$, and for simplicity we assume $0<\nu-\mu<2$. Note that in this case $\nu-\lambda>0$, and we may define n to be the least positive integer greater than $\nu-\lambda$, i.e., $n-1<\nu-\lambda\leqq n$. To avoid a case argument we shall assume $n-1<\nu-\lambda<n$ and $n\geqq 2$.

The following identity is easily verified:

$$K_{\mu/2-\alpha,\nu-\lambda}(\eta_1)=\frac{4x^{\mu-2\alpha}}{\Gamma\left(\frac{\nu-\mu}{2}\right)\Gamma(\alpha-\beta)}\int_x^1 (t^2-x^2)^{\alpha-\beta-1}t\,dt \cdot\int_t^1 (y^2-t^2)^{(\nu-\mu)/2-1}y^{2\beta-\nu+1}\eta_1(y)\,dy.$$

Therefore, if we define $\nu_1=(\mu+\nu)/2-\beta$, $\alpha_1(t)=t^{\nu_1}$, $\beta_1(t)=1$, $\mu_1=1-(\nu-\mu)/2$ and $\phi_1(t)=\eta_1(t)t^{2\beta-\nu+1}$ we obtain

$$I_{\mu/2,(\nu-\mu)/2}(\eta_1)=\frac{2x^{-\nu}}{\Gamma(1-\mu_1)}I_1(\phi_1) \tag{66}$$

and

$$K_{\mu/2-\alpha,\nu-\lambda}(\eta_1)=\frac{4x^{\mu-2\alpha}}{\Gamma(1-\mu_1)\Gamma(\alpha-\beta)}\int_x^1 (t^2-x^2)^{\alpha-\beta-1}tK_1(\phi_1)\,dt \tag{67}$$

where $I_1(\phi_1)$ and $K_1(\phi_1)$ are defined as in (25).

Moreover, it is straightforward to show that

$$K_{\mu/2-\beta,\lambda-\nu}(\eta_2)=\frac{2(-1)^n}{\Gamma(\lambda-\nu+n)}x^{\nu-2\beta}\int_x^1 (t^2-x^2)^{\lambda-\nu+n-1}tD_t^n[t^{2\alpha-\mu}\eta_2(t)]\,dt \tag{68}$$

and

$$I_{\nu/2,(\mu-\nu)/2}(\eta_2)=\frac{2x^{-\mu-2}}{\Gamma\left(\frac{\mu-\nu}{2}+1\right)}\int_0^x (x^2-t^2)^{(\mu-\nu)/2}t^{\nu-\mu+1}D_t[t^{\mu+2}\beta_2(t)]\,dt, \tag{69}$$

where $D_t=[(1/2t)(d/dt)]$. Now let $\phi_2(t)=tD_t^n[t^{2\alpha-\mu}\eta_2(t)]$, $\alpha_2(t)=1$, $\beta_2(t)=1$ and $\mu_2=1-n+\nu-\lambda$, and let $K_2(\phi_2)$ and $I_2(\phi_2)$ be as in (25). We then have from (68)

$$K_{\nu/2-\beta,\lambda-\nu}(\eta_2)=\frac{2(-1)^n x^{\nu-2\beta}}{\Gamma(1-\mu_2)}K_2(\phi_2) \tag{70}$$

and after some manipulation (69) becomes

$$I_{\nu/2,(\mu-\nu)/2}(\eta_2)=\int_0^x I_2(\phi_2)K_{2,1}(x,y)\,dy, \tag{71}$$

where

$$\begin{aligned}K_{2,1}(x,y)=&\frac{4yx^{\nu-2\alpha-4}(x^2-y^2)^{\alpha-\beta-1}}{\Gamma(1-\mu_2)\Gamma(\alpha-\beta)}\\ &\cdot\left\{\frac{(\mu+1-\alpha)}{(\alpha-\beta)}x^2(x^2-y^2){}_2F_1\left(\alpha-\frac{(\mu+\nu)}{2},1-\frac{(\nu-\mu)}{2};(\alpha-\beta)+1;\frac{(x^2-y^2)}{x^2}\right)\right.\\ &\left.+{}_2F_1\left(\alpha-\frac{(\mu+\nu)}{2}-1,1-\frac{(\nu-\mu)}{2};(\alpha-\beta);\frac{(x^2-y^2)}{x^2}\right)\right\}.\end{aligned}$$

Once again a generalized Riemann boundary value problem equivalent to (42) is obtained by substituting (32), (33), (37), and (38) into (43). However, in this case the resulting system has a singular (or degenerate) principal part. In particular, the system has the form

$$A(x)\Phi^+(x)+B(x)\Phi^-(x)+k(\Phi^+,\Phi^-)=F(x)$$

with $A(x)=(a_{ij}(x))$, $B(x)=\overline{A(x)}$, $a_{11}(x)=a_{12}(x)=0$,

$$a_{21}(x)=a_2x^{-\nu}\tilde{\alpha}_1(x^2)(1-x^2)^{1/2-\mu_1}/(\Gamma(1-\mu_1)\sin\mu_1\pi)$$

and

$$a_{22}(x)=b_2(-1)^{n+1}x^{\nu-2\beta}e^{-\mu_2\pi i}(1-x^2)^{1/2-\nu_2}/(\Gamma(1-\mu_2)\sin\mu_2\pi).$$

It is now apparent that $A(x)$ and $B(x)$ are not invertible and the methods of this section do not effect a simplification of (43), or hence of (42). Hence, the behavior of (42) is substantially different for $\alpha=\beta$ and $\alpha\neq\beta$. Evidently, this is due to the fact that when $\alpha\neq\beta$, the function η_1 in (43) appears in operators of different order, as does η_2. For $\alpha=\beta$ this does not occur. That this is important is easily illustrated by considering a single generalized Abel equation of the type

$$a(x)\int_0^x\frac{\phi(t)\,dt}{(x-t)^{\mu_1}}+b(x)\int_x^1\frac{\phi(t)\,dt}{(t-x)^{\mu_2}}=f(x),\qquad 0<x<1. \tag{72}$$

Applying the methods of this paper to (72) shows that (72) is equivalent to the singular integral equation

$$A(x)\Psi(x)+B(x)\frac{1}{i\pi}\int_0^1 \frac{\psi(t)\,dt}{(t-x)}+k(\psi)=F(x).$$

If $\mu_1=\mu_2=\mu$, then

$$A(x)=i\tan\frac{\pi}{2}\mu(a(x)+b(x)),$$

$$B(x)=(a(x)-b(x))$$

and

$$k(\psi)\equiv 0.$$

However, if $\mu_1\neq\mu_2$, say $\mu_1>\mu_2$, then

$$A(x)=i\tan\left(\frac{\pi}{2}\mu_1\right)a(x),$$

$$B(x)=a(x)$$

and

$$k(\cdot)\not\equiv 0.$$

The integrals in (72) correspond to fractional integral operators of order $1-\mu_1$ and $1-\mu_2$ respectively. Thus, if the orders of the operators in (72) are the same the essential structure of the equation is governed by the functions $a(x)+b(x)$ and $a(x)-b(x)$; whereas when the orders are different, the fundamental properties of (72) are determined only by the coefficient of the operator of lowest order. This is analogous to the characteristic behavior of a differential equation being determined by the coefficients of the terms of highest order.

REFERENCES

[1] B. L. J. BRAAKSMA AND A. SCHUITMAN, *Some classes of Watson transforms and related integral equations for generalized functions*, this Journal, 7 (1976), pp. 771–798.

[2] F. D. GAKHOV, *Boundary Value Problems*, Pergamon Press, Oxford, 1966.

[3] M. LOWENGRUB AND J. R. WALTON, *Systems of generalized Abel equations*, this Journal, to appear.

[4] A. C. MCBRIDE, *A theory of fractional integration for generalized functions*, this Journal, 6 (1975), pp. 583–599.

[5] N. I. MUSKHELISHVILI, *Singular Integral Equations*, P. Noordhoff, Groningen, The Netherlands, 1953.

[6] I. N. SNEDDON, *Mixed Boundary Value Problems in Potential Theory*, North-Holland, Amsterdam, 1966.

[7] J. R. WALTON, *A distributional approach to dual integral equations of Titchmarsh type*, this Journal, 6 (1975), pp. 628–643.

[8] ———, *The question of uniqueness for dual integral equations of Titchmarsh type*, Proc. Roy. Soc. Edinburgh Sect. A, 76A (1977), pp. 267–282.

[9] ———, A. NACHMAN AND R. A. SCHAPERY, *The sliding of a rigid indentor over a power law viscoelastic halfspace*, Quart. J. Mech. and Appl. Math., 31 (1978), no. 3, pp. 295–321.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0017$01.00/0

A METHOD OF GENERATING INTEGRAL RELATIONS BY THE SIMULTANEOUS SEPARABILITY OF GENERALIZED SCHRÖDINGER EQUATIONS*

DIETER SCHMIDT† AND GERHARD WOLF†

Abstract. One of the most important methods in the theory of special functions of mathematical physics is that of generating integral relations for these functions by the simultaneous separability of the 3-dimensional wave equation in different orthogonal coordinate systems. In the present paper it will be shown that a consequent application of this principle of simultaneous separability to more general partial differential equations and higher dimensions yields various types of integral relations for the solutions of a wide class of ordinary differential equations which especially contains all second-order equations of Fuchsian type.

Introduction. Let D be a domain (nonvoid open connected set) in the k-dimensional complex vector space $\mathbb{C}^k$ with $\mathbb{N} \ni k \geqq 2$ and let $p = (p_\kappa)\colon D \to \mathbb{C}^k$ and $q\colon D \to \mathbb{C}$ be analytic functions. In this paper we consider the generalized Schrödinger equation

$$Aw := \Delta w + p(x)^t \cdot \operatorname{grad} w + q(x)w = 0, \tag{1}$$

where Δ denotes the Laplace operator; grad, the gradient and $p(x)^t$ the transpose of $p(x)$.

In § 1 we introduce several orthogonal curvilinear coordinate systems, namely ellipsoidal, sphero-conal and special forms of spherical and rectangular coordinates. We give the representations of the operator A in terms of these coordinates, which directly imply sufficient conditions for separability. The most interesting result of § 1 is that the "special" Schrödinger operator A with coefficients

$$\begin{aligned} p_\kappa(x) &= \alpha x_\kappa + \beta_\kappa / x_\kappa, \qquad \kappa = 1, \cdots, k, \\ q(x) &= \gamma \cdot \sum_{\kappa=1}^{k} x_\kappa^2 + \delta + \sum_{\kappa=1}^{k} (\varepsilon_\kappa / x_\kappa^2), \end{aligned} \tag{2}$$

where α, β_κ, γ, δ, ε_κ are complex parameters, separates simultaneously in all four coordinate systems specified above and that its separation yields a wide class of ordinary differential equations especially containing all second-order equations of Fuchsian type and some of their confluent forms.

Since the results of § 1 can be readily verified, we have merely stated the facts and omitted all the proofs. The proofs are essentially the same as in the 3-dimensional case and can be carried out by direct computation or, more elegantly, by the use of Lie theory ([12], [18], [5], [10]). It also can be shown that the sufficient conditions for separability and simultaneous separability stated in § 1 are necessary, too.

In the first part of § 2 we establish a general principle to obtain $(k-1)$-linear integral relations for the solutions of k ordinary linear differential equations occurring with the separation of a k-dimensional partial differential operator and which are thus linked by $k-1$ separation parameters. Such theorems, in a more or less abstract formulation, are well known in multiparameter eigenvalue theory ([1], [2], [16]). We have restricted ourselves to a special formulation which enables us to meet the various situations of § 1. Furthermore, we have restricted our formulation of the integral relations only to proper integrals, since, in this paper, we particularly want to point out the more formal aspects of the method. The corresponding relations with improper integrals can be obtained in the same way.

---

* Received by the editors September 27, 1977, and in revised form March 14, 1978.

† Fachbereich Mathematik, Universität Essen–Gesamthochschule, 4300 Essen 1, West Germany.

In the second part of § 2 the most important applications of the above stated principle to different situations of § 1 are discussed. Of course, one can always apply the above stated principle whenever our general Schrödinger operator separates in one coordinate system; however, then there is the problem of finding suitable solutions of the partial differential equation which can serve as kernels. In the case of our special Schrödinger operator this problem can be solved due to its simultaneous separability. Separation in one coordinate system yields product solutions in terms of these variables, which then may serve as nontrivial kernels for integral relations obtained by separation in another coordinate system. This method yields various types of integral relations for the solutions of the special ordinary differential equations (14), (22.1), (22.2), (29.1), (29.2), and (34). Only the most interesting cases, especially those which lead to new types of integral relations, are discussed here.

Explicit examples and applications of our integral relations, especially with regard to special functions of mathematical physics, will be treated in a later paper.

The present paper was stimulated by a series of papers of Leitner and Meixner [7], [8], [9] as well as by the papers of Erdélyi [4] and Sleeman [15].

In [7], [8], [9] Leitner and Meixner made an approach to a unifying concept for generating integral relations for the special functions of mathematical physics by studying the simultaneous separability of the 3-dimensional Schrödinger equation (1) with $p=0$. This concept was carried on in the thesis of Turner [17], which was initiated by Leitner. Their investigations were restricted to those pairs of coordinate systems which share a common coordinate to be separated out. Hence, their integral relations were linear.

In earlier papers Lambe and Ward [6] and Erdélyi [4] obtained linear integral relations and equations for Heun polynomials and Heun functions by the simultaneous separability of the 3-dimensional special Schrödinger equation (1) where $q=0$ and p is given in (2) with $\alpha=0$ in terms of sphero-conal and spherical coordinates, which share the common coordinate r. Later on, Sleeman [15] obtained quadratic integral relations and equations for the solutions of the Heun equation by the simultaneous separability of the same Schrödinger equation in terms of ellipsoidal and spherical coordinates.

1. Separability of Schrödinger equations in k-dimensional orthogonal coordinate systems.

1.1. General orthogonal coordinates. Let G be a domain in the k-dimensional complex vector space $\mathbb{C}^k$ with $\mathbb{N}\ni k\geqq 2$ and

$$\mathbb{C}^k\supset G\ni z=(z_\kappa)\mapsto x=\phi(z)=(\phi_\kappa(z))\in\mathbb{C}^k$$

be an analytic transformation. We call ϕ "orthogonal" if

$$\sum_{\kappa=1}^{k}\frac{\partial\phi_\kappa}{\partial z_\rho}\cdot\frac{\partial\phi_\kappa}{\partial z_\sigma}=\delta_{\rho\sigma}\cdot g_\rho,\qquad \rho,\sigma\in\{1,\cdots,k\},\tag{3}$$

where $\delta_{\rho\sigma}$ denotes the Kronecker symbol and the g_ρ are analytic functions satisfying

$$g_\rho(z)\neq 0,\qquad z\in G;\quad \rho=1,\cdots,k.\tag{4}$$

If $w:D\to\mathbb{C}$ is an analytic function with domain $D\subset\mathbb{C}^k$ and $\tilde{w}:=w\circ\phi$, our Schrödinger operator A in terms of the new variable $z=(z_1,\cdots,z_k)$ becomes

$$\begin{gathered}Aw=\tilde{A}\tilde{w}:=\sum_{\kappa=1}^{k}\frac{1}{g_\kappa}[\cdots]+(q\circ\phi)\cdot\tilde{w},\\ [\cdots]=\frac{\partial^2\tilde{w}}{\partial z_\kappa^2}+\frac{1}{2}\Bigg(\sum_{\substack{\rho=1\\ \rho\neq\kappa}}^{k}\frac{1}{g_\rho}\frac{\partial g_\rho}{\partial z_\kappa}-\frac{1}{g_\kappa}\frac{\partial g_\kappa}{\partial z_\kappa}+\varphi_\kappa\Bigg)\frac{\partial\tilde{w}}{\partial z_\kappa},\end{gathered}\tag{5}$$

where

$$\varphi_\kappa = 2 \cdot \sum_{\rho=1}^{k} \frac{\partial \phi_\rho}{\partial z_\kappa} \cdot (p_\rho \circ \phi), \qquad \kappa = 1, \cdots, k. \tag{5.1}$$

On the other hand, the p_ρ are expressible in terms of the φ_κ. The orthogonality relation (3) directly yields

$$p_\kappa \circ \phi = \frac{1}{2} \sum_{\rho=1}^{k} \frac{1}{g_\rho} \cdot \frac{\partial \phi_\kappa}{\partial z_\rho} \cdot \varphi_\rho, \qquad \kappa = 1, \cdots, k. \tag{5.2}$$

1.2. Ellipsoidal coordinates. Let $a = (a_1, \cdots, a_k) \in \mathbb{C}^k$ be a fixed vector with $a_\kappa \neq a_\rho$ $(\kappa \neq \rho)$. Ellipsoidal coordinates $\xi = (\xi_1, \cdots, \xi_k)$, which are related to rectangular coordinates $x = (x_1, \cdots, x_k)$ by

$$\sum_{\kappa=1}^{k} \frac{x_\kappa^2}{\xi_\rho - a_\kappa} = 1, \qquad \rho = 1, \cdots, k, \tag{6}$$

can be introduced by

$$\mathbb{C}^k \supset G \ni \xi \mapsto x = \phi(\xi) = (\phi_\kappa(\xi)) \in \mathbb{C}^k$$

where $G \subset (\mathbb{C} \backslash \{a_1, \cdots, a_k\})^k$ is a domain and the ϕ_κ are analytic functions with

$$\phi_\kappa(\xi)^2 = (\xi_\kappa - a_\kappa) \cdot \prod_{\substack{\rho=1 \\ \rho \neq \kappa}}^{k} \left(\frac{\xi_\rho - a_\kappa}{a_\rho - a_\kappa} \right), \qquad \kappa = 1, \cdots, k. \tag{7}$$

At each point $\xi \in G$ with $\xi_\rho \neq \xi_\sigma$ $(\rho \neq \sigma)$, ϕ satisfies the orthogonality relations (3) with

$$g_\kappa(\xi) = \tfrac{1}{4} f(\xi_\kappa; a)^{-1} \cdot \prod_{\substack{\rho=1 \\ \rho \neq \kappa}}^{k} (\xi_\kappa - \xi_\rho), \tag{8}$$

where

$$f(t; \xi) := \prod_{\rho=1}^{k} (t - \xi_\rho). \tag{9}$$

We now introduce the determinant

$$P_k(\xi) := \det \begin{bmatrix} \xi_1^{k-1} & \cdots & \xi_k^{k-1} \\ \vdots & & \vdots \\ \xi_1 & \cdots & \xi_k \\ 1 & \cdots & 1 \end{bmatrix} = \prod_{1 \leq \sigma < \rho \leq k} (\xi_\sigma - \xi_\rho) \tag{10}$$

and assume for the following

$$G \subset \{\xi : P_k(\xi) \neq 0\}.$$

Furthermore, let $\overset{\kappa}{\xi}$ denote the $(k-1)$-dimensional vector $(\xi_1, \cdots, \xi_{\kappa-1}, \xi_{\kappa+1}, \cdots, \xi_k)$. Then by (5) the representation of our Schrödinger operator A in ellipsoidal coordinates becomes

$$\begin{gathered} \tilde{A} = \frac{1}{P_k(\xi)} \sum_{\kappa=1}^{k} (-1)^{\kappa-1} P_{k-1}(\overset{\kappa}{\xi}) \cdot \tilde{A}_\kappa, \\ \tilde{A}_\kappa = 4 f(\xi_\kappa; a) \cdot \left[\frac{\partial^2}{\partial \xi_\kappa^2} + \frac{1}{2} \left(\sum_{\rho=1}^{k} \frac{1}{\xi_\kappa - a_\rho} + \varphi_\kappa \right) \frac{\partial}{\partial \xi_\kappa} \right] + \psi_\kappa, \end{gathered} \tag{11}$$

where

$$\varphi_\kappa = \sum_{\rho=1}^{k} \frac{\phi_\rho \cdot (p_\rho \circ \phi)}{\xi_\kappa - a_\rho} \tag{11.1}$$

and ψ_κ are any functions with

$$q \circ \phi = \frac{1}{P_k(\xi)} \cdot \sum_{\kappa=1}^{k} (-1)^{\kappa-1} P_{k-1}(\overset{\kappa}{\xi}) \cdot \psi_\kappa. \tag{11.2}$$

If φ_κ and ψ_κ depend only on the variable ξ_κ, then $\tilde{A}$ may be written in the form

$$\tilde{A} = \frac{1}{P_k(\xi)} \cdot \det \begin{bmatrix} \tilde{A}_1 & \cdots & \tilde{A}_k \\ \xi_1^{k-2} & & \xi_k^{k-2} \\ \vdots & & \vdots \\ \xi_1 & \cdots & \xi_k \\ 1 & \cdots & 1 \end{bmatrix} \tag{12}$$

where $\tilde{A}_\kappa$ is an ordinary differential operator with respect to ξ_κ, which just means that $P_k(\xi) \cdot \tilde{A}$ is separable ([13], [12]). We say: "A separates in ellipsoidal coordinates".

Especially, if p and q have the form (2), we obtain from (11.1)

$$\varphi_\kappa = \alpha + \sum_{\rho=1}^{k} \frac{\beta_\rho}{\xi_\kappa - a_\rho}. \tag{13.1}$$

If we choose

$$\psi_\kappa = \gamma \xi_\kappa^k + \left(\delta - \gamma \cdot \sum_{\rho=1}^{k} a_\rho\right) \cdot \xi_\kappa^{k-1} + \sum_{\rho=1}^{k} \frac{\varepsilon_\rho}{\xi_\kappa - a_\rho} \cdot \prod_{\substack{\sigma=1 \\ \sigma \neq \rho}}^{k} (a_\rho - a_\sigma), \tag{13.2}$$

also (11.2) is satisfied. Therefore, our special Schrödinger operator separates in ellipsoidal coordinates. Now, using well-known facts on separated solutions of separable operators, we can establish the following

PROPOSITION 1. *For $\kappa = 1, \cdots, k$ let $v_\kappa\colon G_\kappa \to \mathbb{C}$ be analytic with domain $G_\kappa \subset \mathbb{C} \backslash \{a_1, \cdots, a_k\}$, such that $G \subset \times_{\kappa=1}^{k} G_\kappa$. Furthermore, let $w\colon D \to \mathbb{C}$ be analytic with domain $D \subset \mathbb{C}^k$, such that $\phi(G) \subset D$. Finally, let $w \neq 0$ and*

$$(w \circ \phi)(\xi) = \prod_{\kappa=1}^{k} v_\kappa(\xi_\kappa).$$

Then w is a solution of our special Schrödinger equation

$$Aw = 0,$$

iff there exist separation constants $(\lambda_0, \cdots, \lambda_{k-2}) \in \mathbb{C}^{k-1}$, such that the v_κ, $(\kappa = 1, \cdots, k)$, are solutions of the ordinary differential equation

$$\begin{aligned} &\prod_{\rho=1}^{k} (z - a_\rho) \cdot \left[v'' + \frac{1}{2}\left(\alpha + \sum_{\rho=1}^{k} \frac{1+\beta_\rho}{z - a_\rho}\right) v' \right] \\ &\quad + \frac{1}{4}\left(\sum_{\rho=1}^{k} \frac{\varepsilon_\rho}{z - a_\rho} \prod_{\substack{\sigma=1 \\ \sigma \neq \rho}}^{k} (a_\rho - a_\sigma) + \gamma z^k + \delta' z^{k-1} + \sum_{\rho=0}^{k-2} \lambda_\rho z^\rho \right) v = 0 \end{aligned} \tag{14}$$

where $\delta' = \delta - \gamma \cdot \sum_{\rho=1}^{k} a_\rho$.

The differential equation (14) has the only singular points $a_1, \cdots, a_k$ and ∞. The finite points a_κ are regular singular points with characteristic exponents ν_κ^1, ν_κ^2 determined by

$$\nu_\kappa^1+\nu_\kappa^2=\tfrac{1}{2}(1-\beta_\kappa), \qquad \nu_\kappa^1\cdot\nu_\kappa^2=\tfrac{1}{4}\varepsilon_\kappa, \qquad \kappa=1,\cdots,k. \tag{15.1}$$

If $\alpha=\gamma=\delta=0$, the point ∞ is also a regular singular point and the differential equation is the most general second-order equation of Fuchsian type with $k+1$ singularities [14, p. 136]. The characteristic exponents ν_∞^1, ν_∞^2 of the point ∞ are then determined by

$$\nu_\infty^1+\nu_\infty^2=\frac{1}{2}\cdot\sum_{\kappa=1}^{k}(1+\beta_\kappa)-1, \qquad \nu_\infty^1\nu_\infty^2=\frac{1}{4}\lambda_{k-2}. \tag{15.2}$$

The remaining $k-2$ separation constants $\lambda_0,\cdots,\lambda_{k-3}$ are the so-called "accessory parameters".

We would mention that in the case $k=2$ and $\alpha=\varepsilon_\kappa=0$ equation (14) is a confluent form of the Heun equation [6]. Thus, special cases are the Mathieu equation as well as the spheroidal wave equation. In the case $k=3$ and $\alpha=\gamma=\delta=\varepsilon_\kappa=0$, (14) is the Heun equation ([4], [15]).

1.3. Sphero-conal coordinates. As in the case of ellipsoidal coordinates let $a=(a_1,\cdots,a_k)\in\mathbb{C}^k$ be a vector with $a_\kappa\neq a_\rho$ $(\kappa\neq\rho)$. Sphero-conal coordinates $\zeta=(\zeta_1,\cdots,\zeta_k)$, which are related to rectangular coordinates $x=(x_1,\cdots,x_k)$ by

$$\zeta_1=\sum_{\kappa=1}^{k}x_\kappa^2, \qquad \sum_{\kappa=1}^{k}\frac{x_\kappa^2}{\zeta_\rho-a_\kappa}=0, \qquad \rho=2,\cdots,k, \tag{16}$$

can be introduced by

$$\mathbb{C}^k\supset G\ni\zeta\mapsto x=\phi(\zeta)=(\phi_\kappa(\zeta))\in\mathbb{C}^k,$$

where $G\subset(\mathbb{C}\backslash\{0\})\times(\mathbb{C}\backslash\{a_1,\cdots,a_k\})^{k-1}$ is a domain and the ϕ_κ are analytic functions with

$$\phi_\kappa(\zeta)^2=\zeta_1\cdot\prod_{\rho=1}^{\kappa-1}\left(\frac{\zeta_{\rho+1}-a_\kappa}{a_\rho-a_\kappa}\right)\cdot\prod_{\rho=\kappa+1}^{k}\left(\frac{\zeta_\rho-a_\kappa}{a_\rho-a_\kappa}\right), \qquad \kappa=1,\cdots,k. \tag{17}$$

Using the notations of (9) we find that ϕ satisfies the orthogonality relations (3) with

$$\begin{aligned} g_1(\zeta)&=\frac{1}{4}\frac{1}{\zeta_1},\\ g_\kappa(\zeta)&=-\frac{1}{4}f(\zeta_\kappa;a)^{-1}\cdot\zeta_1\cdot\prod_{\substack{\rho=2\\ \rho\neq\kappa}}^{k}(\zeta_\kappa-\zeta_\rho), \qquad \kappa=2,\cdots,k, \end{aligned} \tag{18}$$

at each point ζ of a region $G\subset\{\zeta: P_{k-1}(\overset{1}{\check{\zeta}})\neq 0\}$.

Now by (5) the representation of our Schrödinger operator A in terms of sphero-conal coordinates becomes

$$\begin{aligned} \tilde{A}&=\tilde{A}_1+(\zeta_1\cdot P_{k-1}(\overset{1}{\check{\zeta}}))^{-1}\cdot\sum_{\kappa=2}^{k}(-1)^{\kappa-1}P_{k-2}(\overset{1,\kappa}{\check{\zeta}})\tilde{A}_\kappa,\\ \tilde{A}_1&=4\zeta_1\frac{\partial^2}{\partial\zeta_1^2}+2(k+\varphi_1)\frac{\partial}{\partial\zeta_1}+\psi_1,\\ \tilde{A}_\kappa&=4f(\zeta_\kappa;a)\left[\frac{\partial^2}{\partial\zeta_\kappa^2}+\frac{1}{2}\left(\sum_{\rho=1}^{k}\frac{1}{\zeta_\kappa-a_\rho}+\varphi_\kappa\right)\frac{\partial}{\partial\zeta_\kappa}\right]+\psi_\kappa \qquad \kappa=2,\cdots,k, \end{aligned} \tag{19}$$

where

$$\varphi_1=\sum_{\rho=1}^{k}\phi_\rho\cdot(p_\rho\circ\phi),\qquad \varphi_\kappa=\sum_{\rho=1}^{k}\frac{\phi_\rho\cdot(p_\rho\circ\phi)}{\zeta_\kappa-a_\rho},\qquad \kappa=2,\cdots,k,\tag{19.1}$$

and ψ_κ are any functions with

$$q\circ\phi=\psi_1+\Big(\zeta_1\cdot P_{k-1}\big(\overset{\substack{1\\ \vee}}{\zeta}\big)\Big)^{-1}\sum_{\kappa=2}^{k}(-1)^{\kappa-1}P_{k-2}\big(\overset{\substack{1,\kappa\\ \vee}}{\zeta}\big)\cdot\psi_\kappa.\tag{19.2}$$

If φ_κ and ψ_κ depend only on the variable ζ_κ, then $\tilde{A}$ may be written in the form

$$\tilde{A}=\frac{1}{P_{k-1}\big(\overset{\substack{1\\ \vee}}{\zeta}\big)}\det\begin{bmatrix}\tilde{A}_1 & \tilde{A}_2 & \cdots & \tilde{A}_k\\ 1/\zeta_1 & \zeta_2^{k-2} & & \zeta_k^{k-2}\\ 0 & \vdots & & \vdots\\ \vdots & \zeta_2 & & \zeta_k\\ 0 & 1 & & 1\end{bmatrix}\tag{20}$$

where $\tilde{A}_\kappa$ is an ordinary differential operator with respect to ζ_κ. Therefore, $P_{k-1}(\overset{1}{\zeta})\cdot\tilde{A}$ is separable and we say: "A separates in sphero-conal coordinates".

In the case of our special Schrödinger operator with coefficients p and q of the form (2) we obtain from (19.1)

$$\varphi_1=\alpha\zeta_1+\beta,\qquad \varphi_\kappa=\sum_{\rho=1}^{k}\frac{\beta_\rho}{\zeta_k-a_\rho},\qquad \kappa=2,\cdots,k,\tag{21.1}$$

where $\beta=\sum_{\rho=1}^{k}\beta_\rho$. If we choose

$$\psi_1=\gamma\zeta_1+\delta,\qquad \psi_\kappa=\sum_{\rho=1}^{k}\frac{\varepsilon_\rho}{\zeta_\kappa-a_\rho}\prod_{\substack{\sigma=1\\ \sigma\neq\rho}}^{k}(a_\rho-a_\sigma),\qquad \kappa=2,\cdots,k,\tag{21.2}$$

(19.2) also is satisfied. Thus, our special Schrödinger operator also separates in sphero-conal coordinates. Hence, we can establish the following

PROPOSITION 2. *Let $v_\kappa\colon G_\kappa\to\mathbb{C}$ $(\kappa=1,\cdots,k)$ be analytic with domains $G_1\subset\mathbb{C}\backslash\{0\}$ and $G_\kappa\subset\mathbb{C}\backslash\{a_1,\cdots,a_k\}$ $(\kappa=2,\cdots,k)$, such that $G\subset\times_{\kappa=1}^{k}G_\kappa$. Further, let $w\colon D\to\mathbb{C}$ be analytic with domain $D\subset\mathbb{C}^k$, such that $\phi(G)\subset D$. Finally, let $w\neq 0$ and*

$$(w\circ\phi)(\zeta)=\prod_{\kappa=1}^{k}v_\kappa(\zeta_\kappa).$$

Then w is a solution of our special Schrödinger equation

$$Aw=0,$$

iff there exist separation constants $(\lambda_0,\cdots,\lambda_{\kappa-2})\in\mathbb{C}^{k-1}$, such that v_1 is a solution of

$$zv''+\frac{1}{2}(\alpha z+(k+\beta))v'+\frac{1}{4}\Big(\gamma z+\delta+\frac{\lambda_{k-2}}{z}\Big)v=0\tag{22.1}$$

where $\beta=\sum_{\kappa=1}^{k}\beta_\kappa$, and the v_κ $(\kappa=2,\cdots,k)$ are solutions of

$$\begin{aligned}&\prod_{\rho=1}^{k}(z-a_\rho)\Big[v''+\frac{1}{2}\Big(\sum_{\rho=1}^{k}\frac{1+\beta_\rho}{z-a_\rho}\Big)v'\Big]\\ &\quad+\frac{1}{4}\Big(\sum_{\rho=1}^{k}\frac{\varepsilon_\rho}{z-a_\rho}\prod_{\substack{\sigma=1\\ \sigma\neq\rho}}^{k}(a_\rho-a_\sigma)+\sum_{\rho=0}^{k-2}\lambda_\rho z^\rho\Big)v=0.\end{aligned}\tag{22.2}$$

The differential equation (22.1) has the only singularities 0 and ∞. The point 0 is a regular singular point with characteristic exponents μ_1^1, μ_1^2 determined by

$$\mu_1^1+\mu_1^2=1-\frac{1}{2}\sum_{\kappa=1}^{k}(1+\beta_\kappa),\qquad \mu_1^1\mu_1^2=\frac{1}{4}\lambda_{k-2}. \tag{23}$$

If $\alpha=\gamma=\delta=0$, the point ∞ is also a regular singular point and (22.1) is an Euler equation. Equation (22.1) can always be integrated by confluent hypergeometric functions.

The differential equation (22.2) is the general second-order equation of Fuchsian type already obtained in (14) in the case $\alpha=\gamma=\delta=0$.

1.4. Spherical and rectangular coordinates. Since we want the ordinary differential equations obtained by separation to be in an appropriate "normal form", we use in this paper an algebraic form of spherical coordinates $\eta=(\eta_1,\cdots,\eta_k)$, which are related to rectangular coordinates $x=(x_1,\cdots,x_k)$ by

$$\eta_1=\sum_{\rho=1}^{k}x_\rho^2,\qquad \frac{x_{\kappa-1}^2}{\eta_\kappa}+\frac{\sum_{\rho=\kappa}^{k}x_\rho^2}{\eta_\kappa-1}=0,\qquad \kappa=2,\cdots,k. \tag{24}$$

These can be introduced by

$$\mathbb{C}^k\supset G=\mathop{\times}_{\kappa=1}^{k}G_\kappa\ni\eta\mapsto x=\phi(\eta)=(\phi_\kappa(\eta))\in\mathbb{C}^k,$$

where $G_1\subset\mathbb{C}\backslash\{0\}$, $G_\kappa\subset\mathbb{C}\backslash\{0,1\}$ $(\kappa=2,\cdots,k)$ are domains and the ϕ_κ are analytic functions with

$$\begin{aligned}&\phi_1(\eta)^2=\eta_1\cdot\eta_2,\\ &\phi_\kappa(\eta)^2=\eta_1\cdot\eta_{\kappa+1}\cdot\prod_{\rho=2}^{\kappa}(1-\eta_\rho),\qquad 2\leqq\kappa\leqq k-1,\\ &\phi_k(\eta)^2=\eta_1\cdot\prod_{\rho=2}^{k}(1-\eta_\rho).\end{aligned} \tag{25}$$

From (5) we find, that our general Schrödinger operator A in terms of spherical coordinates has the form

$$\begin{aligned}&\tilde{A}=\tilde{A}_1+\frac{1}{\eta_1}\sum_{\kappa=2}^{k}(-1)^{\kappa-1}\cdot\left(\prod_{\rho=2}^{\kappa-1}(\eta_\rho-1)\right)^{-1}\cdot\tilde{A}_\kappa,\\ &\tilde{A}_1=4\eta_1\frac{\partial^2}{\partial\eta_1^2}+2(k+\varphi_1)\frac{\partial}{\partial\eta_1}+\psi_1,\\ &\tilde{A}_\kappa=4\eta_\kappa(\eta_\kappa-1)\left[\frac{\partial^2}{\partial\eta_\kappa^2}+\frac{1}{2}\left(\frac{1}{\eta_\kappa}+\frac{k+1-\kappa}{\eta_\kappa-1}+\varphi_\kappa\right)\frac{\partial}{\partial\eta_\kappa}\right]+\psi_\kappa,\qquad \kappa=2,\cdots,k,\end{aligned} \tag{26}$$

where

$$\begin{aligned}&\varphi_1=\sum_{\rho=1}^{k}\phi_\rho\cdot(p_\rho\circ\phi),\\ &\varphi_\kappa=\frac{1}{\eta_\kappa}\cdot\phi_{\kappa-1}\cdot(p_{\kappa-1}\circ\phi)+\frac{1}{\eta_\kappa-1}\cdot\sum_{\rho=\kappa}^{k}\phi_\rho\cdot(p_\rho\circ\phi),\qquad \kappa=2,\cdots,k,\end{aligned} \tag{26.1}$$

and ψ_κ are any functions with

$$q \circ \phi = \psi_1 + \frac{1}{\eta_1} \cdot \sum_{\kappa=2}^{k} (-1)^{\kappa-1} \Big(\prod_{\rho=2}^{\kappa-1} (\eta_\rho - 1) \Big)^{-1} \cdot \psi_\kappa. \tag{26.2}$$

If φ_κ and ψ_κ depend only on the variable η_κ, then $\tilde{A}$ is separable and may be written in the form

$$\tilde{A} = \det \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 & \cdots & \tilde{A}_{k-1} & \tilde{A}_k \\ \frac{1}{\eta_1} & 1 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\eta_2 - 1} & 1 & & & \\ \vdots & \vdots & & & 1 & 0 \\ 0 & 0 & & \cdots & \frac{1}{\eta_{k-1} - 1} & 1 \end{bmatrix} \tag{27}$$

where $\tilde{A}_k$ is an ordinary differential operator with respect to η_κ. We say: "A separates in spherical coordinates".

In the case of our special Schrödinger operator with coefficients p and q of the form (2) we obtain from (26.1)

$$\varphi_1 = \alpha\eta_1 + \beta, \qquad \varphi_\kappa = \frac{\beta_{\kappa-1}}{\eta_\kappa} + \frac{\sum_{\rho=\kappa}^{k} \beta_\rho}{\eta_\kappa - 1}, \qquad \kappa = 2, \cdots, k, \tag{28.1}$$

where $\beta = \sum_{\rho=1}^{k} \beta_\rho$. If we choose

$$\psi_1 = \gamma\eta_1 + \delta, \qquad \psi_\kappa = -\frac{\varepsilon_{\kappa-1}}{\eta_\kappa} \quad (2 \leqq \kappa \leqq k-2), \qquad \psi_k = -\frac{\varepsilon_{k-1}}{\eta_k} + \frac{\varepsilon_k}{\eta_k - 1} \tag{28.2}$$

(26.2) also is satisfied. Thus, our special Schrödinger operator separates in spherical coordinates and we can establish

PROPOSITION 3. *Let $v_\kappa: G_\kappa \to \mathbb{C}$ $(\kappa = 1, \cdots, k)$ and $w: D \to \mathbb{C}$ be analytic with domain $D \subset \mathbb{C}^k$ such that $\phi(G) \subset D$. Further, let $w \neq 0$ and*

$$(w \circ \phi)(\eta) = \prod_{\kappa=1}^{k} v_\kappa(\eta_\kappa).$$

Then w is a solution of our special Schrödinger equation

$$Aw = 0,$$

iff there exist separation constants $(\lambda_0, \cdots, \lambda_{k-2}) \in \mathbb{C}^{k-1}$ such that v_1 is a solutions of

$$zv'' + \frac{1}{2}(\alpha z + (k + \beta))v' + \frac{1}{4}\Big(\gamma z + \delta + \frac{\lambda_{k-2}}{z}\Big)v = 0 \tag{29.1}$$

with $\beta = \sum_{\rho=1}^{k} \beta_\rho$ and the v_κ $(\kappa = 2, \cdots, k)$ are solutions of

$$\begin{aligned} &z(z-1)\Big[v'' + \frac{1}{2}\Big(\frac{1+\beta_{\kappa-1}}{z} + \frac{\sum_{\rho=\kappa}^{k}(1+\beta_\rho)}{z-1}\Big)v'\Big] \\ &\quad + \frac{1}{4}\Big(\frac{-\varepsilon_{\kappa-1}}{z} + \frac{\lambda_{k-1-\kappa}}{z-1} + \lambda_{k-\kappa}\Big)v = 0 \end{aligned} \tag{29.2}$$

where $\lambda_{-1} = \varepsilon_k$.

The differential equation (29.1) is identical with equation (22.1) and thus can always be integrated by confluent hypergeometric functions.

The differential equations (29.2) are of Fuchsian type with the 3 singular points 0, 1, and ∞ and thus can always be integrated by hypergeometric functions. By use of the Riemann P-notation, equations (29.2) may be symbolized by

$$P\begin{Bmatrix} 0 & 1 & \infty & \\ \nu^1_{\kappa-1} & \mu^1_\kappa & -\mu^1_{\kappa-1} & z \\ \nu^2_{\kappa-1} & \mu^2_\kappa & -\mu^2_{\kappa-1} & \end{Bmatrix}, \qquad \kappa = 2, \cdots, k, \tag{30}$$

where the ν^1_κ, ν^2_κ are determined by (15.1) and the μ^1_κ, μ^2_κ by

$$\mu^1_\kappa + \mu^2_\kappa = 1 - \frac{1}{2}\sum_{\rho=\kappa}^{k}(1+\beta_\rho), \qquad \mu^1_\kappa \cdot \mu^2_\kappa = \tfrac{1}{4}\lambda_{k-1-\kappa}, \qquad \kappa = 1, \cdots, k-1, \tag{30.1}$$

and

$$\mu^\rho_k = \nu^\rho_k, \qquad \rho = 1, 2. \tag{30.2}$$

Finally, we give a simple transformation of rectangular coordinates, such that the corresponding ordinary differential equations obtained by separation are also in an appropriate form.

Let $\theta = (\theta_1, \cdots, \theta_k) \in \times_{\kappa=1}^{k} G_\kappa$ with domains $G_\kappa \subset \mathbb{C}\backslash\{0\}$ be related to rectangular coordinates $x = (x_1, \cdots, x_k) = \phi(\theta)$ by

$$x^2_\kappa = \theta_\kappa, \qquad \kappa = 1, \cdots, k. \tag{31}$$

Our general Schrödinger operator in terms of the variable θ then becomes

$$\begin{aligned} \tilde{A} &= \sum_{\kappa=1}^{k} \tilde{A}_\kappa, \\ \tilde{A}_\kappa &= 4\theta_\kappa \frac{\partial^2}{\partial\theta^2_\kappa} + 2(1+\phi_\kappa \cdot (p_\kappa \circ \phi))\frac{\partial}{\partial\theta_\kappa} + \psi_\kappa, \end{aligned} \tag{32}$$

where the ψ_κ are any functions with

$$q \circ \phi = \sum_{\kappa=1}^{k} \psi_\kappa. \tag{32.1}$$

If φ_κ and ψ_κ depend only on θ_κ, then $\tilde{A}$ is separable and may be written in the form

$$\tilde{A} = \det\begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 & \cdots & \tilde{A}_{k-1} & \tilde{A}_k \\ -1 & 1 & 0 & & 0 & 0 \\ 0 & -1 & 1 & & \vdots & \vdots \\ \vdots & \vdots & & & 1 & 0 \\ 0 & 0 & & \cdots & -1 & 1 \end{bmatrix} \tag{33}$$

where $\tilde{A}_\kappa$ is an ordinary differential operator with respect to θ_κ.

In the case of our special Schrödinger operator we have

$$\phi_\kappa \cdot (p_\kappa \circ \phi) = \alpha\theta_\kappa + \beta_\kappa, \qquad \kappa = 1, \cdots, k, \tag{33.1}$$

and with

$$\psi_\kappa := \gamma\theta_\kappa + \delta_\kappa + \varepsilon_\kappa/\theta_\kappa, \qquad \kappa = 1, \cdots, k, \tag{33.2}$$

where $\sum_{\kappa=1}^{k} \delta_\kappa = \delta$, condition (32.1) is satisfied. Thus, our special Schrödinger operator separates in the coordinates $\theta = (\theta_1, \cdots, \theta_k)$ and we can establish

PROPOSITION 4. *Let $v_\kappa: G_\kappa \to \mathbb{C}$ $(\kappa = 1, \cdots, k)$ and $w: D \to \mathbb{C}$ be analytic with domain $D \subset \mathbb{C}^k$ such that $\phi(G) \subset D$. Further, let $w \neq 0$ and*

$$(w \circ \phi)(\theta) = \prod_{\kappa=1}^{k} v_\kappa(\theta_\kappa).$$

Then w is a solution of our special Schrödinger equation

$$Aw = 0,$$

iff there exist separation constants $(\lambda_1, \cdots, \lambda_k) \in \mathbb{C}^k$ with $\sum_{\kappa=1}^{k} \lambda_\kappa = \delta$ such that for $\kappa = 1, \cdots, k$ v_κ is a solution of

$$zv'' + \frac{1}{2}(\alpha z + (1 + \beta_\kappa))v' + \frac{1}{4}\left(\gamma z + \lambda_\kappa + \frac{\varepsilon_\kappa}{z}\right)v = 0. \tag{34}$$

The differential equations (34) are of the same type as equation (22.1) and thus can always be integrated by confluent hypergeometric functions. Obviously, the indices ν_κ^1, ν_κ^2 of the regular singular point 0 of the κth equation (34) are determined by (15.1).

2. Integral relations.

2.1. A general principle of generating integral relations. Let $G_\kappa \subset \mathbb{C}$ $(\kappa = 1, \cdots, k)$ be domains and

$$r_\kappa, p_\kappa, q_\kappa, c_\kappa^\rho: G_\kappa \to \mathbb{C}, \qquad \kappa = 1, \cdots, k; \quad \rho = 0, \cdots, k-2, \tag{35}$$

be analytic functions. We then define second-order ordinary differential operators A_κ (with respect to $z_\kappa \in G_\kappa$) by

$$A_\kappa v_\kappa := r_\kappa v_\kappa'' + p_\kappa v_\kappa' + q_\kappa v_\kappa, \qquad \kappa = 1, \cdots, k, \tag{36}$$

and with these the second-order partial differential operator A (with respect to $z = (z_1, \cdots, z_k) \in G := \times_{\kappa=1}^{k} G_\kappa$) by

$$A := \det \begin{bmatrix} A_1 & \cdots & A_k \\ c_1^{k-2} & & c_k^{k-2} \\ \vdots & & \vdots \\ c_1^0 & \cdots & c_k^0 \end{bmatrix} =: \sum_{\kappa=1}^{k} (-1)^{\kappa-1} d_\kappa(\overset{\kappa}{\check{z}}) A_\kappa. \tag{37}$$

Since A is separable, we can establish the following theorem.

THEOREM 2.1. *Let*

(i) *$w: G = \times_{\kappa=1}^{k} G_\kappa \to \mathbb{C}$ be an analytic solution of*

$$Aw = 0;$$

(ii) *for $\kappa = 1, \cdots, k-1$, $v_\kappa: G_\kappa \to \mathbb{C}$ be an analytic solution of*

$$A_\kappa v_\kappa + \left(\sum_{\rho=0}^{k-2} \lambda_\rho c_\kappa^\rho\right) v_\kappa = 0,$$

where $(\lambda_0, \cdots, \lambda_{k-2}) \in \mathbb{C}^{k-1}$;

(iii) *for $\kappa = 1, \cdots, k-1$, $\omega_\kappa: G_\kappa \to \mathbb{C}$ be an analytic function with*

$$(\omega_\kappa r_\kappa)' = p_\kappa \cdot \omega_\kappa;$$

(iv) *for* $\kappa = 1, \cdots, k-1$, $\mathfrak{C}_\kappa$ *be a path in* G_κ, *such that*

$$\left[\omega_\kappa \cdot r_\kappa \cdot \left(\frac{\partial w}{\partial z_\kappa} \cdot v_\kappa - w \cdot v'_\kappa\right)\right]_{\mathfrak{C}_\kappa} = 0.$$

Then $v_k\colon G_k \to \mathbb{C}$, *defined by*

$$v_k(z_k) := \int_{\mathfrak{C}_1} \cdots \int_{\mathfrak{C}_{k-1}} d_k(\overset{k}{z}) w(z) \cdot \prod_{\kappa=1}^{k-1} (\omega_\kappa(z_\kappa) v_\kappa(z_\kappa))\, dz_1 \cdots dz_{k-1},$$

is an analytic solution of

$$A_k v_k + \left(\sum_{\rho=0}^{k-2} \lambda_\rho c_k^\rho\right) v_k = 0.$$

Proof. We consider

$$u \mapsto L_\kappa u := A_\kappa u + \left(\sum_{\sigma=0}^{k-2} \lambda_\sigma c_\kappa^\sigma\right) u.$$

Condition (iii) implies that $\omega_\kappa \cdot L_\kappa$ is formally self-adjoint

$$w \cdot \omega_\kappa \cdot L_\kappa u - u \cdot \omega_\kappa \cdot L_\kappa w = \frac{\partial}{\partial z_\kappa}\left[\omega_\kappa \cdot r_\kappa \cdot \left(w \cdot \frac{\partial u}{\partial z_\kappa} - u \cdot \frac{\partial w}{\partial z_\kappa}\right)\right].$$

Since $L_\kappa v_\kappa = 0$ by (ii), we get

$$\omega_\kappa \cdot v_\kappa \cdot L_\kappa w = \frac{\partial}{\partial z_\kappa}\left[\omega_\kappa \cdot r_\kappa \cdot \left(v_\kappa \cdot \frac{\partial w}{\partial z_\kappa} - v'_\kappa \cdot w\right)\right]$$

and therefore by (iv)

$$\int_{\mathfrak{C}_\kappa} \omega_\kappa(z_\kappa) \cdot v_\kappa(z_\kappa) \cdot (L_\kappa w)(z)\, dz_\kappa = 0$$

identically with respect to $\overset{\kappa}{z}$. On multiplying this by

$$d_\kappa \cdot \prod_{\substack{\rho=1 \\ \rho \neq \kappa}}^{k-1} (\omega_\rho \cdot v_\rho),$$

integrating $(k-2)$-times and changing the order of integration, we find

$$(*) \qquad \int_{\mathfrak{C}_1} \cdots \int_{\mathfrak{C}_{k-1}} d_\kappa(\overset{\kappa}{z}) \cdot \prod_{\rho=1}^{k-1} (\omega_\rho(z_\rho) v_\rho(z_\rho)) \cdot (L_\kappa w)(z)\, dz_1 \cdots dz_{k-1} = 0$$

for $\kappa = 1, \cdots, k-1$. On the other hand, (37) and (i) imply

$$\sum_{\kappa=1}^{k} (-1)^{\kappa-1} d_\kappa(\overset{\kappa}{z})(L_\kappa w)(z) = (Aw)(z) = 0.$$

On multiplying this by $\prod_{\rho=1}^{k-1} (\omega_\rho v_\rho)$, integrating $(k-1)$ times and using $(*)$ for $\kappa = 1, \cdots, k-1$, we see that $(*)$ also holds for $\kappa = k$. Hence, by definition of v_k and by use of the fact that differentiation may be carried out under the integral sign we finally obtain $L_k v_k = 0$.

We have formulated Theorem 2.1 only for proper integrals. If we deal with improper integrals—one knows that these play the more important role in applications—we have to take care that all repeated improper integrals, which occur in the definition of v_k and also in (*), converge locally uniformly with respect to the remaining variables and are independent of the order of integration.

A special case of Theorem 2.1 should be pointed out, since it yields a reduction principle.

THEOREM 2.2. *Let* $k \geqq 3$ *and let the assumptions* (ii) *and* (iii) *of Theorem* 2.1 *be given. Further, let*

$$c_1^\rho = 0, \qquad \rho = 0, \cdots, k-3.$$

Then we can choose

$$d_\kappa(\overset{\kappa}{\check{z}}) =: c_1^{k-2}(z_1) \cdot \tilde{d}_\kappa(\overset{1,\kappa}{\check{z}}), \qquad \kappa = 2, \cdots, k,$$
$$\tilde{A}_\kappa := A_\kappa + \lambda_{k-2} c_\kappa^{k-2},$$

and

$$\tilde{A} := \det \begin{bmatrix} \tilde{A}_2 & \cdots & \tilde{A}_k \\ c_2^{k-3} & & c_k^{k-3} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ c_2^0 & \cdots & c_k^0 \end{bmatrix} = \sum_{\kappa=2}^{k} (-1)^{\kappa-2} \tilde{d}_\kappa(\overset{1,\kappa}{\check{z}}) \cdot \tilde{A}_\kappa.$$

Now, let

(i′) $\tilde{w}: \tilde{G} := \times_{\kappa=2}^{k} G_\kappa \to \mathbb{C}$ *be an analytic solution of*

$$\tilde{A}\tilde{w} = 0;$$

(iv′) *for* $\kappa = 2, \cdots, k-1$, $\mathfrak{C}_\kappa$ *be a path in* G_κ, *such that*

$$\left[\omega_\kappa \cdot r_\kappa \cdot \left(\frac{\partial \tilde{w}}{\partial z_\kappa} \cdot v_\kappa - \tilde{w} \cdot v_\kappa'\right)\right]_{\mathfrak{C}_\kappa} = 0$$

and $\mathfrak{C}_1$ *be any path in* G_1.

Then $w: G \to \mathbb{C}$, *defined by*

$$w(z) := v_1(z_1) \cdot \tilde{w}(\overset{1}{\check{z}}),$$

satisfies (i) *and* (iv) *of Theorem* 2.1 *and* $v_k: G_k \to \mathbb{C}$, *defined in Theorem* 2.1, *becomes*

$$v_k(z_k) = \gamma \cdot \int_{\mathfrak{C}_2} \cdots \int_{\mathfrak{C}_{k-1}} \tilde{d}_k(\overset{1,\kappa}{\check{z}}) \cdot \tilde{w}(\overset{1}{\check{z}}) \cdot \prod_{\kappa=2}^{k-1} (\omega_\kappa(z_\kappa) v_\kappa(z_\kappa))\, dz_2 \cdots dz_{k-1},$$

where

$$\gamma = \int_{\mathfrak{C}_1} c_1^{k-2}(z_1)\omega_1(z_1)v_1(z_1)^2\, dz_1.$$

We notice that in the case of $c_1^{k-2} \cdot \omega_1 \cdot v_1 \neq 0$ one can always find a path $\mathfrak{C}_1$ in G_1 such that $\gamma \neq 0$.

2.2. Integral relations for special functions. Let $a_1, \cdots, a_k$ be different points in $\mathbb{C}$ and $\nu_\kappa^1, \nu_\kappa^2$ ($\kappa = 1, \cdots, k$), α, γ, δ and λ_ρ ($\rho = 0, \cdots, k-2$) be complex parameters. Let $\delta' = \delta - \gamma \cdot \sum_{\rho=1}^{k} a_\rho$.

Our aim is to get integral relations for the solutions of the differential equation

$$\text{(38)} \quad \begin{aligned} &\prod_{\rho=1}^{k} (z-a_\rho) \cdot \left[v'' + \left(\frac{\alpha}{2} + \sum_{\rho=1}^{k} \frac{1-\nu_\rho^1-\nu_\rho^2}{z-a_\rho}\right)v'\right] \\ &\quad + \left(\sum_{\rho=1}^{k} \frac{\nu_\rho^1\nu_\rho^2}{z-a_\rho} \cdot \prod_{\substack{\sigma=1 \\ \sigma\neq\rho}}^{k} (a_\rho - a_\sigma) + \frac{\gamma}{4}z^k + \frac{\delta'}{4}z^{k-1} + \sum_{\rho=0}^{k-2} \lambda_\rho z^\rho\right)v = 0, \end{aligned}$$

by applying the methods of § 2.1 to the situations of §§ 1.2 and 1.3 where (38) occurs in connection with the separation of the special Schrödinger operator in ellipsoidal and sphero-conal coordinates.

As kernels for the integral relations we shall use the product solutions of the special Schrödinger equation in terms of spherical and rectangular coordinates to be found by Propositions 3 and 4.

2.2.1. Kernels in terms of spherical coordinates. Let $\overset{1}{\mu}_\kappa, \overset{2}{\mu}_\kappa$ $(\kappa = 1, \cdots, k-1)$ be complex parameters with

$$(39) \qquad 1-\overset{1}{\mu}_\kappa-\overset{2}{\mu}_\kappa=\sum_{\rho=\kappa}^{k}(1-\overset{1}{\nu}_\rho-\overset{2}{\nu}_\rho), \qquad \kappa=1,\cdots,k-1.$$

The kernels in terms of spherical coordinates to be found by Proposition 3 are then of the form

$$(40) \qquad V(\eta)=R(\eta_1)\cdot K(\overset{1}{\check{\eta}}), \qquad K(\overset{1}{\check{\eta}})=\prod_{\kappa=2}^{k}K_\kappa(\eta_\kappa),$$

where R is a solution of

$$(40.1) \qquad zv''+\left(\frac{\alpha}{2}z+(1-\overset{1}{\mu}_1-\overset{2}{\mu}_1)\right)v'+\left(\frac{\gamma}{4}z+\frac{\delta}{4}+\frac{\overset{1}{\mu}_1\overset{2}{\mu}_1}{z}\right)v=0,$$

and

$$(40.2) \qquad K_\kappa\in P\begin{Bmatrix} o & 1 & \infty & \\ \overset{1}{\nu}_{\kappa-1} & \overset{1}{\mu}_\kappa & -\overset{1}{\mu}_{\kappa-1} & z \\ \overset{2}{\nu}_{\kappa-1} & \overset{2}{\mu}_\kappa & -\overset{2}{\mu}_{\kappa-1} & \end{Bmatrix}, \qquad \kappa=2,\cdots,k,$$

with $\overset{\rho}{\mu}_k=\overset{\rho}{\nu}_k$ $(\rho=1,2)$.

The solutions of (40.1) may be written in terms of confluent hypergeometric functions. We have to distinguish 3 cases. Let

$$(41) \qquad \tilde{\alpha}:=(\alpha^2-4\gamma)^{1/2}, \qquad \tilde{\delta}:=\delta-\alpha\cdot\sum_{\rho=1}^{k}(1-\overset{1}{\nu}_\rho-\overset{2}{\nu}_\rho).$$

Then one easily finds [3, vol. II], that the solutions of (40.1) are given by the following.

Case 1. $\tilde{\alpha}\neq 0$.

$$(40.1.1) \quad R(z)=\exp\left(\frac{1}{4}(\tilde{\alpha}-\alpha)z\right)\cdot z^{\overset{1}{\mu}_1}\cdot {}_1\mathfrak{F}_1\left(\frac{\tilde{\delta}}{2\tilde{\alpha}}+\frac{1}{2}(1+\overset{1}{\mu}_1-\overset{2}{\mu}_1);1+\overset{1}{\mu}_1-\overset{2}{\mu}_1;-\frac{\tilde{\alpha}}{2}z\right),$$

Case 2. $\tilde{\alpha}=0$, $\tilde{\delta}=:-\tau^2\neq 0$.

$$(40.1.2) \quad \begin{aligned} R(z)&=\exp\left(-\tfrac{1}{4}\alpha z\right)\cdot z^{\overset{1}{\mu}_1}\cdot\exp\left(\tau z^{1/2}\right)\\ &\quad\cdot {}_1\mathfrak{F}_1\left(\tfrac{1}{2}+\overset{1}{\mu}_1-\overset{2}{\mu}_1;1+2\overset{1}{\mu}_1-2\overset{2}{\mu}_1;-2\tau z^{1/2}\right),\end{aligned}$$

Case 3. $\tilde{\alpha}=\tilde{\delta}=0$.

$$(40.1.3) \qquad R(z)=\exp\left(-\tfrac{1}{4}\alpha z\right)\cdot\left(c_1z^{\overset{1}{\mu}_1}+c_2z^{\overset{2}{\mu}_1}\right), \qquad c_1,c_2\in\mathbb{C},$$

where ${}_1\mathfrak{F}_1(a;c;z)$ denotes any solution of the confluent hypergeometric equation

$$zu''+(c-z)u'-au=0.$$

Obviously, the K_κ ($\kappa=2,\cdots,k$) are given by

$$\begin{aligned}K_\kappa(z)&=z^{\nu^1_{\kappa-1}}\cdot(1-z)^{\mu^1_\kappa}\\&\cdot{}_2\mathfrak{F}_1(\nu^1_{\kappa-1}+\mu^1_\kappa-\mu^1_{\kappa-1},\,\nu^1_{\kappa-1}+\mu^1_\kappa-\mu^2_{\kappa-1};\,1+\nu^1_{\kappa-1}-\nu^2_{\kappa-1};\,z)\end{aligned}\tag{40.2.1}$$

with $\mu^\rho_k=\nu^\rho_k$ ($\rho=1,2$), where ${}_2\mathfrak{F}_1(a,b;c;z)$ denotes any solution of the hypergeometric differential equation

$$z(1-z)u''+(c-(a+b+1)z)u'-abu=0.$$

2.2.2. Kernels in terms of rectangular coordinates. Let τ_κ ($\kappa=1,\cdots,k$) be complex parameters with

$$\sum_{\kappa=1}^{k}\tau_\kappa^2+\tilde{\delta}=0\tag{42}$$

where $\tilde{\delta}$ is given by (41). The kernels in terms of rectangular coordinates to be found by Proposition 4 are then of the form

$$W(\theta)=\prod_{\kappa=1}^{k}W_\kappa(\theta_\kappa),\tag{43}$$

where W_κ is a solution of

$$zv''+\left(\frac{\alpha}{2}z+(1-\nu^1_\kappa-\nu^2_\kappa)\right)v'+\left(\frac{\gamma}{4}z+\frac{\alpha(1-\nu^1_\kappa-\nu^2_\kappa)-\tau^2_\kappa}{4}+\frac{\nu^1_\kappa\nu^2_\kappa}{z}\right)v=0.\tag{43.1}$$

In the same way as in § 2.2.1 for (40.1) one finds that the W_κ are given by the following.

Case 1. $\tilde{\alpha}\neq0$.

$$\begin{aligned}W_\kappa(z)&=\exp\left(\frac{1}{4}(\tilde{\alpha}-\alpha)z\right)\cdot z^{\nu^1_\kappa}\\&\cdot{}_1\mathfrak{F}_1\left(-\frac{\tau^2_\kappa}{\tilde{\alpha}}+\frac{1}{2}(1+\nu^1_\kappa-\nu^2_\kappa);\,1+\nu^1_\kappa-\nu^2_\kappa;\,-\frac{\tilde{\alpha}}{2}z\right).\end{aligned}\tag{43.1.1}$$

Case 2. $\tilde{\alpha}=0$.

$$\begin{aligned}W_\kappa(z)&=\exp\left(-\frac{\alpha}{4}z\right)\cdot z^{\nu^1_\kappa}\\&\cdot\exp(\tau_\kappa z^{1/2})\cdot{}_1\mathfrak{F}_1(\tfrac{1}{2}+\nu^1_\kappa-\nu^2_\kappa;\,1+2\nu^1_\kappa-2\nu^2_\kappa;\,-2\tau_\kappa z^{1/2}),\qquad\tau_\kappa\neq0,\\W_\kappa(z)&=\exp\left(-\frac{\alpha}{4}z\right)\cdot(c_1z^{\nu^1_\kappa}+c_2z^{\nu^2_\kappa}),\qquad c_1,c_2\in\mathbb{C},\qquad\tau_\kappa=0,\end{aligned}\tag{43.1.2}$$

where $\tilde{\alpha}$ is given by (41) and ${}_1\mathfrak{F}_1(a;c;z)$ denotes any solution of the confluent hypergeometric equation.

2.2.3. Two types of integral relations. Let $G_\kappa\subset\mathbb{C}\setminus\{a_1,\cdots,a_k\}$ ($\kappa=1,\cdots,k$) be domains and

$$\omega(z)=\exp\left(\frac{\alpha}{2}z\right)\prod_{\rho=1}^{k}(z-a_\rho)^{-\nu^1_\rho-\nu^2_\rho}.\tag{44}$$

Application of Theorem 2.1 in connection with § 1.2 then yields the next theorem.

THEOREM 2.3. *Let* $k \geqq 2$.

(i) *Let* $w: \times_{\kappa=1}^{k} G_\kappa \to \mathbb{C}$ *denote either the function* V *in* (40) *or the function* W *in* (43) *in terms of ellipsoidal coordinates.*

(ii) *Let* $v_\kappa: G_\kappa \to \mathbb{C}$ $(\kappa = 1, \cdots, k-1)$ *be solutions of* (38).

(iii) *Let for* $\kappa = 1, \cdots, k-1$, $\mathfrak{C}_\kappa$ *be a path in* G_κ *such that*

$$\left[\omega(z_\kappa) \cdot \prod_{\rho=1}^{k} (z_\kappa - a_\rho) \cdot \left(\frac{\partial w}{\partial z_\kappa}(z) v_\kappa(z_\kappa) - v'_\kappa(z_\kappa) w(z)\right)\right]_{\mathfrak{C}_\kappa} = 0$$

identically with respect to $\overset{\kappa}{\check{z}}$.

Then $v_k: G_k \to \mathbb{C}$ *defined by*

$$v_k(z_k) = \int_{\mathfrak{C}_1} \cdots \int_{\mathfrak{C}_{k-1}} P_{k-1}(\overset{k}{\check{z}}) \cdot w(z) \cdot \prod_{\kappa=1}^{k-1} (\omega(z_\kappa) v_\kappa(z_\kappa))\, dz_1 \cdots dz_{k-1}$$

is an analytic solution of (38).

Special cases of integral relations of this type are for $k = 2$ the well-known (linear) integral relations for Mathieu and spheroidal wave functions ([11], [3, vol. III]), and for $k = 3$ the (quadratic) integral relations for Heun functions found by Sleeman [15]. It should be noted that there is a mistake in [15]: the operator in (4.2) of [15] and therefore the following kernels have to depend also on δ and ε.

Application of Theorems 2.1 and 2.2 in connection with § 1.3 yields the following theorem.

THEOREM 2.4. *Let* $k \geqq 3$ *and* $\alpha = \gamma = \delta = 0$.

(i) *Let* $\tilde{w}: \times_{\kappa=2}^{k} G_\kappa \to \mathbb{C}$ *denote the function* K *in* (40) *with* $\mu_1^1 \cdot \mu_1^2 = \lambda_{k-2}$ *in terms of sphero-conal coordinates.*

(ii) *Let* $v_\kappa: G_\kappa \to \mathbb{C}$ $(\kappa = 2, \cdots, k-1)$ *be solutions of* (38).

(iii) *Let for* $\kappa = 2, \cdots, k-1$, $\mathfrak{C}_\kappa$ *be a path in* G_κ *such that*

$$\left[\omega(z_\kappa) \cdot \prod_{\rho=1}^{k} (z_\kappa - a_\rho) \cdot \left(\frac{\partial \tilde{w}}{\partial z_\kappa}(\overset{1}{\check{z}}) v_\kappa(z_\kappa) - v'_\kappa(z_\kappa) \tilde{w}(\overset{1}{\check{z}})\right)\right]_{\mathfrak{C}_\kappa} = 0$$

identically with respect to $\overset{1,\kappa}{\check{z}}$.

Then $v_k: G_k \leftarrow \mathbb{C}$ *defined by*

$$v_k(z_k) = \int_{\mathfrak{C}_2} \cdots \int_{\mathfrak{C}_{k-1}} P_{k-2}(\overset{1,k}{\check{z}}) \cdot \tilde{w}(\overset{1}{\check{z}}) \cdot \prod_{\kappa=2}^{k-2} (\omega(z_\kappa) v_\kappa(z_\kappa))\, dz_2 \cdots dz_{k-1}$$

is an analytic solution of (38).

Special cases of integral relations of this type for $k = 3$ are the (linear) integral relations for Heun functions found by Erdélyi [4] and Lambe and Ward [6].

REFERENCES

[1] F. M. ARSCOTT, *Two-parameter eigenvalue problems in differential equations*, Proc. London Math. Soc., 14 (1964), pp. 459–470.

[2] ———, *Transform theorems for two-parameter eigenvalue problems in Hilbert Space*, Proc. Conference on the Theory of Ordinary and Partial Differential Equations, 1974, Lect. Notes in Math., vol. 415, Springer, New York, 1975, pp. 302–307.

[3] A. ERDÉLYI, W. MAGNUS, F. OBERHETTINGER AND F. G. TRICOMI, *Higher Transcendental Functions*, vols. I, II, III, McGraw-Hill, New York, 1953–1955.

[4] A. ERDÉLYI, *Integral equations for Heun functions*, Quart. J. Math. Oxford Ser,, 13 (1942), pp. 107–112.

[5] E. G. KALNINS, W. MILLER JR. AND P. WINTERNITZ, *The group $O(4)$, separation of variables and the hydrogen atom*, SIAM J. Appl. Math., 30 (1976), pp. 630–664.

[6] C. G. LAMBE AND D. R. WARD, *Some differential equations and associated integral equations*, Quart. J. Math. Oxford Ser., 5 (1934), pp. 81–87.

[7] A. LEITNER UND J. MEIXNER, *Simultane Separierbarkeit von verallgemeinerten Schwingungsgleichungen*, Arch. Math., 10 (1959), pp. 387–391.

[8] ———, *Zur Theorie der hypergeometrischen Funktionen*, Arch. Math. (Basel), 10 (1959), pp. 452–459.

[9] ———, *Eine Verallgemeinerung der Sphäroidfunktionen*, Ibid., 11 (1960), pp. 29–39.

[10] A. A. MAKAROV, J. A. SMORODINSKY, KH. VALIEV AND P. WINTERNITZ, *A systematic search for nonrelativistic systems with dynamical symmetries*, Nuovo Cimento A, L11 (1967), No. 4, pp. 1061–1084.

[11] J. MEIXNER UND F. W. SCHÄFKE, *Mathieusche Funktionen und Sphäroidfunktionen*, Springer, Berlin, 1954.

[12] W. MILLER, JR., *Symmetry and Separation of Variables*, Addison-Wesley, Reading, MA, 1977.

[13] H. D. NIESSEN, *Algebraische Untersuchungen über separierbare Operatoren*, Math. Z., 94 (1966), pp. 328–348.

[14] F. W. SCHÄFKE UND D. SCHMIDT, *Gewöhnliche Differentialgleichungen*, Springer, Berlin, 1973.

[15] B. D. SLEEMAN, *Non-linear integral equations for Heun functions*, Proc. Edinburgh Math. Soc., 16 (1969), pp. 281–289.

[16] ———, *Multiparameter eigenvalue problems and k-linear operators*, Proc. Conference on the Theory of Ordinary and Partial Differential Equations, Lecture Notes in Math., vol. 280, Springer, New York, 1972, pp. 347–353.

[17] W. W. TURNER, *A method of generating integral representations*, Ph.D. thesis, Dept. of Math., Michigan State University, East Lansing, 1963.

[18] P. WINTERNITZ, YA. A. SMORODINSKII, M. UHLIR AND I. FRIS, *Symmetry groups in classical and quantum mechanics*, Soviet J. Nuclear Phys., 4 (1967), pp. 444–450.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0018$01.00/0

ON A FREQUENCY DOMAIN CONDITION USED IN THE THEORY OF VOLTERRA EQUATIONS*

GUSTAF GRIPENBERG†

Abstract. This paper studies a local frequency domain condition used by Londen and Staffans in the existence theory for abstract Volterra equations with subdifferential nonlinearities.

1. Introduction and statement of results. This note considers the relationship between various conditions on the kernel k in the nonlinear Volterra equation

$$u(t)+\int_0^t k(t-s)gu(s)\,ds \quad \ni f(t), \quad t\in R_+=[0,\infty) \tag{1.1}$$

where k, g and f are given and u is the unknown function taking its values in a real Hilbert space H. The nonlinear operator g is assumed to be the subdifferential of a convex, lower semicontinuous function $\varphi: H\to(-\infty,\infty]$, $\varphi\not\equiv+\infty$. A standard assumption on the real-valued kernel k is that it should satisfy

$$k \text{ is locally absolutely continuous on } R_+ \text{ and } k(0)>0. \tag{1.2}$$

Barbu [1] studied the existence of solutions of (1.1) under the condition that the kernel is of positive type, i.e.

$$\int_0^T v(t)\int_0^t k(t-s)v(s)\,ds\,dt \geqq 0 \tag{1.3}$$

for all $v\in C(R_+;R)$ and all $T>0$. One could clearly obtain the same result if (1.3) is replaced by the assumption that the kernel is of T-positive type for some $T>0$, where we use

DEFINITION 1. A function a is said to be of *T-positive type*, $T>0$ fixed, if $a\in L^1(0,T;R)$ and $\int_0^T v(t)\int_0^t a(t-s)v(s)\,ds\,dt\geqq 0$ for all $v\in C([0,T];R)$.

Londen [2] studied the equation (1.1) when k satisfies (1.2) and

$$k' \text{ is of bounded variation on } [0,T] \text{ for some } T>0. \tag{1.4}$$

Later Londen and Staffans weakened (1.4) to the condition that there exists $T_0>0$ such that

$$\sup_{0\leqq\tau\leqq T_0}\ \sup_{\omega\in R_+} \omega\int_0^\tau \sin(\omega t)k'(t)\,dt<\infty. \tag{1.5}$$

In order to investigate the implications of this assumption we need the following

DEFINITION 2. A function a is said to be of *T-F-positive type* if $a\in L^1(0,T;R)$ and $\int_0^T \cos(n\pi T^{-1}t)a(t)\,dt\geqq 0$ for all integers n.

We will establish the following relationships between the definitions presented here:

PROPOSITION 1. *If $T>0$ and the function a is of T-F-positive type, then it is of T-positive type but the converse does not necessarily hold. If the function a is of T-positive type, then it is of T_1-positive type for all T_1, $0<T_1\leqq T$ but this statement does not necessarily hold for T-F-positivity.*

Now we can state our main result.

* Received by the editors March 15, 1978, and in revised form July 24, 1978.
† Institute of Mathematics, Helsinki University of Technology, Espoo, Finland.

THEOREM. *Assume that $T>0$ and that*

(1.6) $$k \text{ is absolutely continuous on } [0, T]$$

and

(1.7) $$\sup_{\omega\in R_+} \omega \int_0^T \sin(\omega t)k'(t)\,dt<\infty.$$

Then $k=k_1+k_2$ where

(1.8) *k_1 is absolutely continuous on $[0, T]$ and k_1' is of bounded variation on this interval*

and

(1.9) $$k_2 \text{ is of } T\text{-}F\text{-positive type.}$$

Concerning the converse of this statement we have

PROPOSITION 2. *For every $T>0$ there exists a function of T-F-positive type that satisfies* (1.6) *but not* (1.5) *for any T_0, $0<T_0\leqq T$.*

It is not too difficult to see that if k satisfies (1.2) and if for some $T>0$ we have $k=k_1+k_2$ on $[0, T]$ where k_1 satisfies (1.8) and k_2 is of T-positive type, then there exists a solution of (1.1) (under appropriate conditions on the function f) provided that

$$\sup\{c\,|\,k_1(t)-c \text{ is of } \tau\text{-positive type for some } \tau\in(0, T]\}>-k_1(0).$$

To prove this statement one proceeds in the same way as in [2], but one has to use [2, line (3.26)] (for the function k_1) and Definition 1 (for the function k_2-c) to obtain the crucial [2, line (3.34)]. Note that this is a more general result than the one established in [3] since an easy calculation using the definition shows that if a is of T-F-positive type then $a(t)-T^{-1}\int_0^T a(s)\,ds$ is also of T-F-positive type.

2. Proof of Proposition 1. Assume that $v\in C([0, T]; R)$ and define

(2.1) $$v(t)=0,\quad t\in(-T, 0),\qquad v(t+2T)=v(t),\quad t\in R.$$

When we also define for $a\in L^1(0, t; R)$

(2.2) $$a(t)=a(-t),\quad t\in(-T, 0),\qquad a(t+2T)=a(t),\quad t\in R$$

then it follows from Fubini's theorem that

$$\int_0^T v(t)\int_0^t a(t-s)v(s)\,ds\,dt=2^{-1}\int_{-T}^T v(t)\int_{-T}^T a(t-s)v(s)\,ds\,dt$$

and hence an application of Plancherel's theorem implies that

(2.3) $$\int_0^T v(t)\int_0^t a(t-s)v(s)\,ds\,dt=2^{-1}(2T)^2\sum_{n=-\infty}^{\infty}\hat{a}(n)|\hat{v}(n)|^2$$

where we have utilized the definition

(2.4) $$\hat{u}(n)=(2T)^{-1}\int_{-T}^T e^{-int}u(t)\,dt.$$

Combining (2.2), (2.3) and (2.4) we see that T-F-positivity implies T-positivity.

It is obvious that if a function a is of T-positive type then it is also of T_1-positive type for all T_1, $0<T_1\leqq T$, because if $v\in C([0, T_1]; R)$ then we put $v(t)=0$ on $(T_1, T]$ and use a standard argument.

Let $T>0$ be arbitrary. The function $a(t)=\cos(\pi T^{-1}t)$ is obviously of T-F-positive type. A calculation shows however that

$$\int_0^{T/2} \cos(2\pi(T/2)^{-1}t)\cos(\pi T^{-1}t)\,dt = -T(15\pi)^{-1}<0$$

and hence $a(t)=\cos(\pi T^{-1}t)$ is not of $T/2$-F-positive type. This result combined with the previous one shows that T-positivity does not imply T-F-positivity and the proof is completed.

3. Proof of the Theorem. Let $T>0$ be arbitrary and define

$$b(t)=c_1+Tt-t^2/2, \qquad t\in[0,T] \tag{3.1}$$

where c_1 is some real constant. A calculation shows that if n is an integer, $n\neq 0$, then

$$\int_0^T \cos(n\pi T^{-1}t)b(t)\,dt = -(n\pi)^{-2}T^3. \tag{3.2}$$

Define

$$k_1(t)=\alpha T^{-1}b(t)+k(T), \qquad t\in[0,T] \tag{3.3}$$

where $\alpha=\sup_{\omega\in R_+}\omega\int_0^T \sin(\omega t)k'(t)\,dt$. Clearly k_1, thus defined, satisfies (1.8).

An integration by parts in (1.7) shows that we also have

$$\omega^2\int_0^T \cos(\omega t)(k(t)-k(T))\,dt \geqq -\alpha, \qquad \omega\in R_+. \tag{3.4}$$

Let $k_2(t)=k(t)-k_1(t)$ and let n be a strictly positive integer. Then we have by (3.2)–(3.4) (putting $\omega=n\pi T^{-1}$)

$$\begin{aligned}&\int_0^T \cos(n\pi T^{-1}t)k_2(t)\,dt\\ &\quad=\int_0^T \cos(n\pi T^{-1}t)(k(t)-k(T))\,dt-\alpha T^{-1}\int_0^T \cos(n\pi T^{-1}t)b(t)\,dt\geqq 0.\end{aligned} \tag{3.5}$$

Finally we choose the constant c_1 to be sufficiently negative so that

$$\int_0^T k_2(t)\,dt \geqq 0. \tag{3.6}$$

Now (3.5) and (3.6) imply that (1.9) holds and the proof is completed.

4. Proof of Proposition 2. Let the sequence $\{q_i\}_{i=1}^{\infty}$ be defined by

$$q_1=1, \quad q_i=2^{8^i}, \quad i\geqq 2. \tag{4.1}$$

Define the sequence $\{k_n\}_{n=0}^{\infty}$ by

$$k_n=\begin{cases} r^{-7/4} & \text{if } n=2^i r \text{ for some } r\in N,\ q_i\leqq r<q_{i+1},\\ 0 & \text{for all other } n.\end{cases} \tag{4.2}$$

It follows directly from (4.2) that

$$\sum_{n=0}^{\infty}|k_n|<\infty, \tag{4.3}$$

and since (4.1) implies that $2^{2i}\sum_{n=q_i}^{\infty} n^{-3/2} \leqq 2^{-i}$, $i \geqq 2$ we also have

$$\sum_{n=0}^{\infty} |nk_n|^2 < \infty. \tag{4.4}$$

Let $T > 0$ be arbitrary and define the function k by

$$k(t) = \sum_{n=0}^{\infty} k_n \cos(n\pi T^{-1}t), \qquad t \in [0, T]. \tag{4.5}$$

In view of (4.3) the sum converges uniformly and we note that (4.4) implies that k is absolutely continuous on $[0, T]$. It is obvious that k is of T-F-positive type. To establish the conclusion of the corollary it is consequently sufficient to show that

$$\sup_{\omega \in R_+} \omega \int_0^{2^{-m}T} \sin(\omega t) k'(t)\, dt = +\infty, \qquad m = 0, 1, \cdots. \tag{4.6}$$

Fix an arbitrary nonnegative integer m and define

$$h(t) = \sum_{n=2^{m+1}q_{m+1}}^{\infty} k_n \cos(n\pi T^{-1}t), \qquad t \in [0, T]. \tag{4.7}$$

Since $k'(t) - h'(t)$ is of bounded variation on $[0, 2^{-m}T]$ it follows that

$$\sup_{\omega \in R_+} \left| \omega \int_0^{T_1} \sin(\omega t)(k'(t) - h'(t))\, dt \right| < \infty, \tag{4.8}$$

where we have defined $T_1 = 2^{-m}T$. Performing an integration by parts we get

$$\omega \int_0^{T_1} \sin(\omega t) h'(t)\, dt = -\omega^2 \int_0^{T_1} \cos(\omega t)(h(t) - h(T_1))\, dt \tag{4.9}$$

and due to the uniform convergence on the right side in (4.9) when we insert (4.7), we can integrate termwise with the result that

$$\begin{aligned} \omega \int_0^{T_1} & \sin(\omega t) h'(t)\, dt \\ &= \sum_{n=2^{m+1}q_{m+1}}^{\infty} k_n((n\pi)^2(\omega T)^{-2}((n\pi)^2(\omega T)^{-2} - 1)^{-1}\omega \\ &\qquad \cdot \sin(\omega T_1)\cos(n\pi T^{-1}T_1) + (1 - (n\pi)^2(\omega T)^{-2})^{-1} n\pi T^{-1} \\ &\qquad \cdot \cos(\omega T_1)\sin(n\pi T^{-1}T_1)). \end{aligned} \tag{4.10}$$

Next we define the sequence $\{h_j\}_{j=0}^{\infty}$ by

$$h_j = \begin{cases} r^{-7/4} & \text{if } j = 2^i r \text{ for some } r \in N,\ q_{m+i} \leqq r < q_{m+i+1},\ i \geqq 1, \\ 0 & \text{for all other } j. \end{cases} \tag{4.11}$$

We note that when $n \geqq 2^{m+1}q_{m+1}$ then $k_n \cos(n\pi T^{-1}T_1) = k_n$ and $k_n \sin(n\pi T^{-1}T_1) = 0$ since by definition $k_n = 0$ if $nT^{-1}T_1 \not\equiv 0 \bmod 2$. Consequently we have by (4.10) and (4.11) that

$$\omega \int_0^{T_1} \sin(\omega t) h'(t)\, dt = \pi T_1^{-1} \sum_{j=0}^{\infty} h_j (j/\beta)^2((j/\beta)^2 - 1)^{-1}\beta \sin(\pi\beta) \tag{4.12}$$

where $\beta = \omega T_1 \pi^{-1}$. Let $i \geqq 1$ be arbitrary and choose ω so that

$$\beta = 2^{i+1} q_{m+i+1} - 3/2. \tag{4.13}$$

Then we have

$$\begin{aligned}\sum_{j=2^{i+1}q_{m+i+1}}^{\infty} & h_j (j/\beta)^2((j/\beta)^2-1)^{-1}\beta \sin(\pi\beta) \\ & \geqq 2^{i+1}(1-(\tfrac{3}{2})(2^{i+1}q_{m+i+1})^{-1})q_{m+i+1} \sum_{r=q_{m+i+1}}^{\infty} r^{-7/4},\end{aligned} \tag{4.14}$$

since $(j/\beta)^2((j/\beta)^2-1)^{-1} \geqq 1$ when $j \geqq 2^{i+1}q_{m+i+1}$ and since $\sin(\pi\beta)=1$ by (4.13). It is easy to see that

$$\begin{aligned} q_{m+i+1} \sum_{r=q_{m+i+1}}^{\infty} r^{-7/4} & \geqq q_{m+i+1} \int_{q_{m+i+1}}^{\infty} t^{-7/4}\, dt \\ & = (\tfrac{4}{3}) q_{m+i+1}^{1/4}. \end{aligned} \tag{4.15}$$

On the other hand a simple calculation shows that (4.11) and (4.13) imply

$$\begin{aligned}\sum_{j=0}^{2^{i+1}q_{m+i+1}-1} & h_j (j/\beta)^2((j/\beta)^2-1)^{-1}\beta \sin(\pi\beta) \\ & \geqq -(\tfrac{4}{3})(2^{i+1}q_{m+i+1})^{-1}(1-(\tfrac{3}{2})(2^{i+1}q_{m+i+1})^{-1})^{-1} \sum_{j=0}^{2^{i+1}q_{m+i+1}-1} j^2 h_j,\end{aligned} \tag{4.16}$$

since $h_j((j/\beta)^2-1)^{-1} \geqq -(\tfrac{4}{3})h_j$ when $j < 2^{i+1}q_{m+i+1}$. But we also have

$$\begin{aligned}\sum_{j=0}^{2^{i+1}q_{m+i+1}-1} j^2 h_j & \leqq 2^{2i} \sum_{r=q_{m+1}}^{q_{m+i+1}-1} r^{2-7/4} \\ & \leqq 2^{2i} \int_{q_{m+1}}^{q_{m+i+1}} t^{1/4}\, dt \leqq (\tfrac{4}{5}) 2^{2i} q_{m+i+1}^{5/4}.\end{aligned} \tag{4.17}$$

Letting finally $i \to \infty$ we see from (4.12) and (4.14)–(4.17) that

$$\sup_{\omega \in R_+} \omega \int_0^{T_1} \sin(\omega t) h'(t)\, dt = +\infty.$$

Combining this result with (4.8) and the fact that m was arbitrary we obtain (4.6) and the proof is completed.

REFERENCES

[1] V. BARBU, *Nonlinear Volterra equations in a Hilbert space*, this Journal, 6 (1975), pp. 728–741.

[2] S-O. LONDEN, *On an integral equation in a Hilbert space*, this Journal, 8 (1977), pp. 950–970.

[3] S-O. LONDEN AND O. J. STAFFANS, *A note on Volterra equations in a Hilbert space*, Proc. Amer. Math. Soc., 70 (1978), pp. 57–62.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0019$01.00/0

SOLUTION OF H-EQUATIONS BY ITERATION*

C. T. KELLEY†

Abstract. A generalization of the Chandrasekhar H-equation is solved by iteration. Such equations are of interest in heat transfer.

1. Introduction and statement of results.

Consider the integral equation

$$f(x,\omega)=1+\omega f(x,\omega)\int_0^\infty \frac{t}{x+t}\varphi(t)f(t,\omega)\,dt. \tag{1.1}$$

In (1.1), ω is a complex parameter, f is the function to be found, and φ is a measurable function on $(0,\infty)$.

Under certain assumptions on φ it is known that (1.1) has a solution $H(x,\omega)$ that is analytic in ω for $|\omega|<1$. This solution is of importance in many applications. We call H the physical solution to (1.1). The question considered in this paper is the following: When can H be found by solving (1.1) by iteration?

The question has been answered if $\varphi\geqq 0$, $\varphi(x)=0$ for $0\leqq x\leqq 1$, and $\int_1^\infty t^2\varphi(t)\,dt<\infty$. In this case H is continuous for $|\omega|\leqq 1$, $x\geqq 1$. These assumptions are natural in neutron transport theory. Bowden and Zwiefel [1] have shown that if $\int_1^\infty \varphi(t)\,dt=\frac{1}{2}$ then H may be found for $|\omega|<1$ by an iterative method. In a more general setting, Mullikin and the author [7] showed that H may be computed for $|\omega|\leqq 1$ in this way.

In many cases of physical importance, however, φ may become negative. The methods of Bowden and Zweifel [1] and Rall [9], then give the best known results. If, however, φ is not integrable on $(0,\infty)$, these methods fail as the integral in (1.1) need not even be defined in the Lebesgue sense.

This paper is motivated by work of Crosbie and Sawheny [5], [6]. They consider (1.1) for $\varphi(x)=\frac{1}{2}J_1(x)$, where J_1 is the Bessel function of order one. They obtained numerical results that indicated that H may be found by an iterative method for quite general φ.

We assume φ satisfies the following conditions:

(A1) $\varphi\in L_p(0,\infty)$ for some P, $1\leqq P<\infty$.

(A2) $\displaystyle\int_0^\infty \left|\frac{\varphi(t)}{t}\right|dt<\infty.$

(A3) Let $k(x)=\lim_{N\to\infty}\int_0^N c^{-|x|t}t\varphi(t)\,dt$; then $k\geqq 0$, and $k\in L_1(-\infty,\infty)$.

(A4) $\int_{-\infty}^\infty k(x)\,dx=1$.

(A5) $\lim_{N\to\infty}\int_0^N \varphi(t)\,dt=\frac{1}{2}$.

(A6) If f is a measurable, nonnegative decreasing function on $(0,\infty)$ and

$$0<\sup_{0<x\leqq 1}|xf(x)|,<\infty,\quad\text{then}\quad \lim_{N\to\infty}\int_0^N \varphi(t)f(t)\,dt>0.$$

The reader should note that (A4) is merely a normalization condition on φ. As will become clear, (A5) implies (A4). The purpose of the normalization is to make H analytic in the disc $|\omega|<1$. Our assumptions imply that the integral in (1.1) is defined for

* Received by the editors December 30, 1977, and in revised form April 3, 1978. This work was sponsored by the United States Army under Contract No. DAAG29-75-C-0024 and the National Science Foundation under Grant No. MCS75-17385 A01.

† Mathematics Research Center, University of Wisconsin–Madison, Madison, Wisconsin. Now at: Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650.

nonnegative decreasing f by

$$\int_0^\infty \frac{t}{x+t}\varphi(t)f(t)\,dt = \lim_{N\to\infty}\int_0^N \frac{t}{x+t}\varphi(t)f(t)\,dt. \tag{1.2}$$

For $\varepsilon \geqq 0$; let $\mathscr{C}(\varepsilon)$ be the space of complex-valued functions continuous on $[\varepsilon, \infty)$ and having finite limits at infinity. $\mathscr{C}(\varepsilon)$ is a Banach space under the sup norm. Moreover Dini's theorem holds for $\mathscr{C}(\varepsilon)$ in the sense that if a sequence of real valued functions f_n in $\mathscr{C}(\varepsilon)$ converge monotonically upward to f in $\mathscr{C}(\varepsilon)$ pointwise, and $\lim_{x\to\infty} f_n(x) = \lim_{x\to\infty} f(x)$ for all n, then f_n converges to f in the topology of $\mathscr{C}(\varepsilon)$.

We will show in § 2 that H is an analytic $\mathscr{C}(0)$ valued function of ω for $|\omega| < 1$ and is a continuous $\mathscr{C}(\varepsilon)$ valued function of ω for $|\omega| \leqq 1$.

Under assumptions (A1)–(A6) we prove the following iteration results.

THEOREM 2.1. *For* $|\omega| \leqq 1$, $x \geqq 0$ *define*

$$H_0(x, \omega) = 1,$$

$$H_{n+1}(x, \omega) = 1 + \omega H_n(x, \omega)\int_0^\infty \frac{t}{x+t} H_n(t, \omega)\varphi(t)\,dt, \qquad n \geqq 0.$$

Then for $0 < \varepsilon < 1$,

(i) H_n *converges to* H *in* $\mathscr{C}(0)$ *uniformly in* ω *for* $|\omega| \leqq 1 - \varepsilon$.

(ii) H_n *converges to* H *in* $\mathscr{C}(\varepsilon)$ *uniformly in* ω *for* $|\omega| \leqq 1$.

THEOREM 2.2. *Let* $0 < \varepsilon < 1$, $|\omega| \leqq 1$, *define for* $x \geqq 0$,

$$K_0(x, \omega) = 1$$

$$K_{n+1}(x, \omega) = \left[1 - \omega \int_0^\infty \frac{t}{x+t}\varphi(t)K_n(t, \varphi)\,dt\right]^{-1}.$$

Then

(i) K_n *converges to* H *in* $\mathscr{C}(0)$ *uniformly in* ω *for* $|\omega| \leqq 1 - \varepsilon$;

(ii) K_n *converges to* H *in* $\mathscr{C}(\varepsilon)$ *uniformly in* ω *for* $|\omega| \leqq 1$.

Implicit in the statements of these theorems is the fact that all integrals exist in the sense of (1.2). For convenience we write, for g measurable,

$$\lim_{N\to\infty}\int_0^N g(t)dt = \int_* g(t)\,dt \tag{1.3}$$

when the limit on the left exists.

2. Proofs of Theorems 2.1 and 2.2. Equation (1.1) is intimately connected to the theory of Wiener–Hopf equations [8]. This connection plays a vital role in the proofs of our results, and we give the relevant details here. For $f \in L_1(-\infty, \infty)$, $\hat{f}$ will denote the Fourier transform of f.

Let $k(x)$ be as in (A3). Assumptions (A3) and (A4) imply that for $|\omega| < 1$, the equation,

$$\gamma(x, \omega) - \omega\int_0^\infty k(x-y)\gamma(y, \omega)\,dy = \omega k(x), \qquad x > 0, \tag{2.1}$$

has a unique solution $\gamma(x, \omega) \in L_1(0, \infty)$. For $0 \leqq \omega < 1$, $\gamma \geqq 0$, and γ is an analytic $L_1(0, \infty)$-valued function of ω for $|\omega| < 1$. We write

$$\gamma(x, \omega) = \sum_{n=1}^\infty \omega^n \gamma_n(x). \tag{2.2}$$

In (2.2), $\gamma_n(x)\geqq 0$ for $x>0$, $\gamma_n\in L_1(0,\infty)$. Then the unique solution to (1.1) which is analytic in ω for $|\omega|<1$ and $\mathscr{C}[0,\infty]$ valued is

$$H(x,\omega)=1+\hat{\gamma}(ix,\omega). \tag{2.3}$$

For $|\omega|<1$, we have the factorization, valid for real λ,

$$(1-\omega\hat{k}(\lambda))(1+\hat{\gamma}(\lambda,\omega))(1+\hat{\gamma}(-\lambda,\omega))=1. \tag{2.4}$$

From (2.3) and (2.4) we conclude

$$\frac{\partial}{\partial x}H(x,\omega)<0 \quad \text{for } x>0, \qquad 0\leqq\omega<1, \tag{2.5}$$

and

$$\frac{\partial}{\partial \omega}H(x,\omega)>0 \quad \text{for } x\geqq 0, \qquad 0\leqq\omega<1. \tag{2.6}$$

$$H(x,\omega)=1+\sum_{n=1}^{\infty}\omega^n\hat{\gamma}_n(ix), \qquad |\omega|<1. \tag{2.7}$$

$$\hat{\gamma}_n(ix)>0 \quad \text{for } n\geqq 1, \qquad x\geqq 0. \tag{2.8}$$

$$\hat{\gamma}_n(i\,\cdot\,)\in\mathscr{C}[0,\infty] \quad \text{for } n\geqq 1. \tag{2.9}$$

Let Σ denote the class of nonnegative decreasing continuous functions f on $(0,\infty)$ which satisfy

$$\sup_{0<x\leqq 1}|xf(x)|<\infty. \tag{2.10}$$

For $f\in\Sigma$ define

$$Lf(x)=\int_* \frac{t}{x+t}\varphi(t)f(t)\,dt, \tag{2.11}$$

$$Mf(x)=\int_0^\infty \frac{x}{x+t}\varphi(t)f(t)\,dt. \tag{2.12}$$

The integral defining M is a Lebesgue integral for each fixed $x>0$ by assumption (A1). Assumptions (A1) and (A6) imply that Mf and Lf are in $\mathscr{C}(0)$ for every $f\in\Sigma$. We note that for $0\leqq\omega<1$, H is in Σ. We may rewrite (1.1) as

$$H(\omega)=1+\omega L(H(\omega))H(\omega). \tag{2.13}$$

We now prove the main lemma.

LEMMA 2.1. *Let $\varepsilon>0$; then*

$$\lim_{\omega\to 1^-}H(\omega)=H(1)$$

exists in $\mathscr{C}(\varepsilon)$. Moreover $H(1)\in\Sigma$, and for $x>0$, we have

$$H(1)=1+L(H(1))H(1). \tag{2.14}$$

Proof. By (2.4) we have that

$$H(0,\omega)=(1-\omega)^{-1/2}=1+\omega\int_*\varphi(t)H(t,\omega)\,dt\,H(0,\omega).$$

Hence

$$H^{-1}(\omega)=(1-\omega)^{1/2}+\omega M(H(\omega)). \tag{2.15}$$

Consider the function F given for $x\geqq 0$ by,

$$F(x)=\int_*\frac{t}{x+t}\varphi(t)\,dt. \tag{2.16}$$

Assumptions (A2) and (A6) imply that $F\in\mathscr{C}(0)$ and $F\geqq 0$. For $\varepsilon>0, x\geqq 0$, we have

$$F(x)-F(x+\varepsilon)=\int_0^\infty\frac{\varepsilon}{(x+t)(x+\varepsilon+t)}\varphi(t)\,dt>0.$$

Therefore F is decreasing. By (2.13), we have

$$H^{-1}(x,\omega)\geqq\omega xF(x)\geqq\omega xF(0). \tag{2.17}$$

Hence, for $x\geqq\varepsilon$, $0\leqq\omega<1$, $H(x,\omega)\leqq(1/(\omega\varepsilon))F^{-1}(0)$. Hence, for $x>0$, $\lim_{\omega\to 1^-}H(x,\omega)=H(x,1)$ exists by (2.6). Moreover $H(x,1)$ is decreasing in x and $\sup_{0<x\leqq 1}|xH(x,1)|\leqq F^{-1}(0)<\infty$. Hence $MH(1)\in\mathscr{C}(0)$ by assumption (A1). By the dominated convergence theorem and assumption (A1) we have, for $x>0$,

$$MH(1)(x)=\lim_{\omega\to 1^-}MH(\omega)(x)=H^{-1}(x,\omega). \tag{2.18}$$

Hence $H(1)\in\Sigma$ and $H(1)$ satisfies (2.14). Dini's theorem then implies that the limit in (2.18) exists in $\mathscr{C}(\varepsilon)$ for $\varepsilon>0$. This completes the proof.

Now for $x>0$, $|H(x,\omega)|\leqq\sum_{n=1}^\infty|\omega|^n\hat{\gamma}_n(ix)+1=H(x,|\omega|)\leqq H(x,1)$. Hence for $\varepsilon\geqq 0, H$ is an analytic $\mathscr{C}(\varepsilon)$-valued function of ω for $|\omega|<1$. For $\varepsilon>0, H$ is a continuous $\mathscr{C}(\varepsilon)$ valued function of ω for $|\omega|\leqq 1$.

We now prove the main results.

Proof of Theorem 2.1. Let $0<\varepsilon<1$. Observe that if $g(x)=1+\int_0^\infty e^{-xt}G(t)\,dt$, and $G\in L_1$ is nonnegative, then $1+Lg(x)$ has the same form as

$$Lg(x)=\int_0^\infty e^{-xt}\left[k(t)+\int_0^\infty k(x+t)G(s)\,ds\right]dt. \tag{2.19}$$

The definition of H_n implies, therefore that for $m\geqq 1$ there are functions $P_{n,m}\in L_1(0,\infty)$, $P_{n,m}\geqq 0$, such that for $x\geqq 0$, $|\omega|\leqq 1$,

$$H_n(x,\omega)=\sum_{m=0}^{2^n-1}\omega^m\int_0^\infty e^{-xt}P_{n,m}(x)\,dx. \tag{2.20}$$

It is easy to see that

$$P_{n,m}(x)=\gamma_m(x),\qquad 1\leqq m\leqq n. \tag{2.21}$$

We claim that $P_{n,m}(x)\leqq\gamma_m(x)$ for all m and n. We prove this by induction on n. For $n=0$ the result is clear. Assume (2.21) for $n=N$; from the definition of H_n and (2.19) we have

$$
\begin{aligned}
P_{N+1,m}(t) &= \sum_{\substack{j+l=m-1\\ j,l\geqq 1}} \int_0^t P_{N,l}(t-s)\int_0^\infty k(s+r)P_{N,j}(r)\,dr\,ds \\
&\quad + \int_0^t P_{N,m-1}(t-s)k(s)\,ds + \int_0^\infty k(t+s)P_{N,m-1}(s)\,ds \\
&\leqq \sum_{\substack{j+l=m-1\\ j,l\geqq 1}} \int_0^t \gamma_l(t-s)\int_0^\infty k(s+r)\gamma_j(r)\,dr\,dx \\
&\quad + \int_0^t \gamma_{m-1}(t-s)k(s)\,ds + \int_0^\infty k(t+s)\gamma_{m-1}(s)\,ds \\
&= \gamma_m(x).
\end{aligned} \tag{2.22}
$$

The last equality is a consequence of the fact that

$$\hat{\gamma}_m(ix) = \sum_{\substack{j+l=m-1\\ j,l\geqq 1}} \hat{\gamma}_j(ix)L(\hat{\gamma}_l(i\,\cdot\,))(x) + \hat{\gamma}_{m-1}(ix) + L(\hat{\gamma}(i\,\cdot\,))(x), \tag{2.23}$$

which is an easy consequence of (1.1).

A similar argument shows that

$$P_{n,m} \leqq P_{n+1,m}. \tag{2.24}$$

Hence we have, for $n,\ k\geqq 0,\ x\geqq 0,\ |\omega|\leqq 1$,

$$|H_{n+k}(x,\omega)-H_n(x,\omega)| \leqq H(x,|\omega|)-H_n(x,|\omega|) \leqq \sum_{m=n+1}^{\infty} \hat{\gamma}_m(ix)|\omega|^m. \tag{2.25}$$

For $x\geqq \varepsilon$, $|\omega|\leqq 1$ we have

$$\sum_{m=n+1}^{\infty} \hat{\gamma}_m(ix)|\omega|^m \leqq H(x,1) - \sum_{m=1}^{n} \hat{\gamma}_m(ix). \tag{2.26}$$

The right side of (2.26) goes to zero as n becomes large uniformly for $x\geqq\varepsilon$, $|\omega|\leqq 1$. This proves (ii).

For $x\geqq 0$, $|\omega|\leqq 1-\varepsilon$ we have

$$\sum_{m=n+1}^{\infty} \hat{\gamma}(ix)|\omega|^m \leqq H(x,1-\varepsilon) - \sum_{m=1}^{n} \hat{\gamma}(ix)|1-\varepsilon|^m. \tag{2.27}$$

The right side of (2.27) converges to zero as n becomes large uniformly for $x\geqq 0$, $|\omega|\leqq 1-\varepsilon$. This proves (i) and completes the proof.

Proof of Theorem 2.2. Note that $H_0(x,\omega)=K_0(x,\omega)\leqq H(x,\omega)$. An induction argument similar to that used in the proof of Theorem 2.1 will show that, for $x>0$, $0\leqq\omega\leqq 1$,

$$K_n(x,\omega)\leqq H(x,\omega), \tag{2.28}$$

and

$$K_n(x,\omega)[1-\omega L(H_{n-1}(\cdot\,,\omega)(x))]^{-1} \geqq H_n(x,\omega). \tag{2.29}$$

As in the proof of Theorem 2.1 we also have for $|\omega|\leqq 1$, $x>0$, $n\geqq 0$, $m\geqq 0$,

$$|K_{n+m}(x,\omega)-K_n(x,\omega)| \leqq K_{n+m}(x,|\omega|)-K_n(x,|\omega|). \tag{2.30}$$

This completes the proof.

Systems of nonlinear equations similar to (1.1) are of interest in the kinetic theory of gases [2], [4]. It is possible that the methods of this paper will generalize to that setting.

REFERENCES

[1] R. L. BOWDEN AND P. F. ZWEIFEL, *A Banach Space Analysis of the Chandrasekhar H-equation*, Astrophysical J., 210 (1976), pp. 178–183.

[2] E. E. BURNISTON AND C. E. SIEWERT, *Half-space analysis basic to the time dependent BGK model in the kinetic theory of gases*, J. Math. Phys., 18 (1977), pp. 376–380.

[3] S. CHANDRASEKHAR, *Radiative Transfer*, Dover, New York, 1960.

[4] T. S. CHANG, J. T. KRIESE AND C. E. SIEWERT, *Elementary solutions of coupled model equations in the kinetic theory of gases*, Internat. J. Engrg. Sci., 12 (1974), pp. 441–470.

[5] A. L. CROSBIE AND T. R. SAWHENY, *Application of Ambarzumian's method to radiant interchange in a rectangular cavity*, J. Heat Transfer, C96 (1974), pp. 191–196.

[6] ———, *Radiant interchange in a nonisothermal rectangular cavity*, AIAA J., 13 (1975), pp. 425–431.

[7] C. T. KELLEY AND T. W. MULLIKIN, *Solution by interation of H-equations in multigroup neutron transport*, J. Math. Phys., 19 (1978), pp. 500–501.

[8] M. G. KRIEN, *Integral equations on a half-line with kernel depending upon the difference of the arguments*, Amer. Math. Soc. Transl., (2), 22 (1962), pp. 163–288.

[9] L. B. RALL, *Quadratic equations in Banach spaces*, Rend. Circ. Mat. Palermo, 10 (1961), pp. 314–332.

SIAM J. MATH. ANAL.
Vol. 10, No. 4, July 1979

0036-1410/79/1004-0020$01.00/0

CALCULATION OF EXTREMUM PROBLEMS FOR UNIVALENT FUNCTIONS*

ECKHARD GRASSMANN† AND JON ROKNE†

Abstract. Let S be the usual class of univalent functions in $\{|z|<1\}$ normalized by $f(z)=z+\sum_{i=2}^{\infty} a_i z^i$ and V_n the coefficient region of S. It is well known that f corresponds to a boundary point of V_n if and only if f satisfies a quadratic equation of the form $Q(w)\,dw^2=R(z)\,dz^2$ called Schiffers equation that maps $\{|z|<1\}$ onto a slit domain. We treat the following problems numerically for V_4:

1. Given Q find R and f.
2. Find the function that maximizes Re $e^{i\varphi}a_4$ with the constraint that a_2 and a_3 are some given complex numbers in V_3. In this case Schiffers equation is a sufficient condition for f to be extremal.

The critical trajectories of $Q(w)\,dw^2$ and $R(z)\,dz^2$ are in each case displayed graphically for some particular examples.

1. Introduction. Let S be the usual class of univalent functions in the unit disc that are normalized by $f(z)=z+\sum_{i=2}^{\infty} a_i z^i$. It is easy to prove existence of solutions of extremum problems in this class, but extremely hard to actually solve them. In fact it seems to be a lucky coincidence if such a problem can be solved explicitly. Today, modern computers give us so much more calculating power which we feel has not been exploited sufficiently for this type of problem. The purpose of this paper is to show that a computer can be used successfully to find explicit solutions for problems involving the coefficients up to order four. Even though our methods are not restricted to four, this seems to be our limit for practical reasons (financial and otherwise).

As in [5] and [6] our main aim is to give heuristic insight. Accordingly, we chose, in case of doubt, an approach that promised most successful solutions rather than one which would work in every case.

Our main line of attack are the theories of [10]. As it is done there, we denote by V_n the nth coefficient region of S. To each boundary point of V_n there corresponds a unique function $w=f(z)$ and this function satisfies an equation of the form

$$Q(w)\,dw^2=R(z)\,dz^2. \tag{1}$$

Here Q and R are rational functions. They have a pole of order $n+1$ at zero. At ∞ Q has at least a triple zero so the quadratic differential $Q(w)\,dw^2$ has at most a simple pole there and the numerator of Q has degree $n-2$. Furthermore, f maps $\{|z|=<1\}$ onto the complement of finitely many slits satisfying $Q(w)\,dw^2 \geqq 0$ containing ∞. Conversely, each function f that satisfies these conditions corresponds to a boundary point of V_n. We call these functions D_n-functions.

We treat two problems. Problem 1 is to find R if Q is given. This problem has according to [10, pp. 111, 112] a solution. It is unique if there is only one slit that does not split. Problem 2 is to maximize Re $e^{i\varphi}a_4$ with the constraint that a_2 and a_3 are some given complex numbers, and φ some given real number. According to the general coefficient theorem an equation of the form (1) is then necessary and sufficient (see [7]).

2. Problem 1. As mentioned in the introduction, there is to each quadratic function $q_5+q_4w+q_3w^2$ at least one D_4-function $f(z)=w$, and a rational function R such that the differential equation

$$Q(w)\,dw^2=\frac{q^5+q_4w+q_2w^2}{w^5}\,dw^2=R(z)\,dz^2$$

* Received by the editors January 1, 1977, and in final revised form January 16, 1978.
† Department of Computer Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

is satisfied. Our first task is to find sufficiently many necessary conditions for R. We therefore assume we have a rational function $R(z)$ and a mapping function $f(z)$ such that the above equation is satisfied. Throughout, we will also assume that there is only one boundary slit and that it does not split, i.e. both zeros of Q lie in the image domain,[1] and therefore R has two zeros inside the unit disc. Then the solution is unique. Because $R(z)\,dz^2 = -R(z)z^2\,d\varphi \geqq 0$ on $\{|z|=1\}$ the symmetry principle implies that then R is also zero at $1/\bar{\alpha}_i$. Since the slit is analytic there is a unique α_0 on $\{|z|=1\}$ that corresponds to the tip of the slit, so R has at this point a double zero (at least). These are all the zeros of R. We get:

$$R(z) = \alpha \frac{(z-\alpha_1)(z-\alpha_2)(z-1/\overline{\alpha_1})(z-1/\overline{\alpha_2})(z-\alpha_0)^2}{z^5}.$$

Since $z^2R(z)$ is real on $\{|z|=1\}$ we get, according to the symmetry principle,

$$z^2R(z) = \overline{\frac{1}{\bar{z}^2} R(1/\bar{z})},$$

Evaluating the right hand side we get:

$$\overline{\frac{1}{\bar{z}^2} R\left(\frac{1}{\bar{z}}\right)} = \bar{\alpha}\left[\overline{\frac{\prod_{i=0}^2 (1/\bar{z}-\alpha_i)(1/\bar{z}-1/\bar{\alpha}_i)}{1/\bar{z}^3}}\right]$$

$$= \bar{\alpha}\frac{\prod_{i=0}^2 (1-\bar{\alpha}_i z)(1-z/\alpha_i)}{z^3}$$

$$= \bar{\alpha}\frac{\prod \bar{\alpha}_i}{\prod \alpha_i}\frac{\prod (1/\bar{\alpha}_i - z)(\alpha_i - z)}{z^3}$$

and comparing with the left hand side we get that:

$$\alpha \prod_{i=0}^{2} \alpha_i = \bar{\alpha} \prod_{i=0}^{2} \bar{\alpha}_i \tag{2}$$

is real. If f is to be in S the terms of order 5 of the Laurent series of the poles of R and Q also have to agree. This gives

$$q_5 = \alpha \prod_{i=0}^{2} \frac{\alpha_i}{\bar{\alpha}_i}. \tag{3}$$

For Problem 2 we will have more conditions of this type.

Next we will deal with two crucial conditions. First of all the critical points of Q and R have to match if the differential equation is to have a univalent solution. We denote by β_1, β_2 the zeros of $Q(w)$, so $Q(w)$ can be written

$$Q(w) = \beta \frac{(w-\beta_1)(w-\beta_2)}{w^5}.$$

We obtain as consistency condition:

$$\int_{\beta_1}^{\beta_2} \sqrt{Q(w)}\,dw = \int_{\alpha_1}^{\alpha_2} \sqrt{R(z)}\,dz. \tag{4}$$

[1] This assumption can be removed by a slightly more complicated program. It holds if f is a support point of S (see [11]). This is the most interesting case.

Here the integration is to be performed over some path in the z-plane and the image-path in the w-plane, or equivalently over paths that are homotopic to those. For practical purposes we chose straight lines. Equation (4) holds for some evaluation of the square roots which leaves us with an ambiguity. This ambiguity is the cause of a lot of troubles, and necessitates the use of plots to check the solutions, as we will see later.

The last condition is easy to miss: There is a point a_∞ on $\{|z|=1\}$ that corresponds to ∞ in the w-plane. Since $\int \sqrt{|Q(w)|}\,|dw|$ is the same along either side of the boundary slit the same has to be true for the corresponding arcs in the z-plane which means that

$$\int_{\alpha_0}^{\alpha_\infty} \sqrt{|R(e^{i\varphi})|}\,d\varphi = \int_{\alpha_\infty}^{\alpha_0} \sqrt{|R(e^{i\varphi})|}\,d\varphi \tag{5}$$

where the integrals are followed in the counterclockwise sense. This equation determines a_∞. From this we get the consistency condition

$$\int_{\alpha_1}^{\alpha_\infty} \sqrt{R(z)}\,dz = \int_{\beta_1}^{\infty} \sqrt{Q(w)}\,dw. \tag{6}$$

We have now the right number of conditions: The unknowns are α_∞, α_0, α_1, α_2, α. The conditions are $|\alpha_0|=1$, $|\alpha_\infty|=1$, the complex equations (3), (4) and (6) and the conditions (2) and (5), i.e. we have as many equations as we have unknowns. In order to simplify the numerical calculations we reduce the ten real equations to five real equations using algebraic and analytic techniques in the following manner. The following constants were computed once for all the iterations

$$p_1 = \int_{\beta_1}^{\beta_2} \sqrt{Q(w)}\,dw, \qquad p_2 = \int_{\beta_1}^{\infty} \sqrt{Q(w)}\,dw, \qquad p_3 = \prod_{i=o}^{2} \beta_i.$$

Then for each iteration the following was computed with α_0 and α_∞ normalized to 1:

$$\alpha = p_3 \left(\prod_{i=0}^{2} \frac{b_i}{\bar{b}_i} \right)^{-1}, \qquad f_1 + if_2 = \int_{\alpha_1}^{\alpha_2} \sqrt{R(z)}\,dz - p_1.$$

α_∞ via a Newton iteration on equation (5) (see § 5 #1 for details)

$$f_3 + if_4 = \int_{\alpha_1}^{\alpha_\infty} \sqrt{R(z)}\,dz - p_2, \qquad f_5 = \operatorname{Im}\left(\alpha \prod_{i=0}^{2} \alpha_i \right)$$

and the five real equations are then

$$f_i = 0, \qquad i = 1, 2, \cdots, 5.$$

3. Univalence of the solution. The question is of course whether the equations of the last section are sufficient for univalence and the answer is unfortunately in general no.

To see this, observe Fig. 1. $Q(w)$ is in this case the left-hand side of an equation (1) corresponding to the Koebe function $k(z) = z/(1-z)^2$ and the extremum problem $\max \operatorname{Re} a_4$. Schiffers variation gives in this case $Q(w) = (10w^2 + 6w + 1)/w^5$.[2] The correct solution is shown in Fig. 2, a wrong solution in Figs. 3 and 4. The lines shown are the trajectories of $Q(w)\,dw^2 \geqq 0$ respectively $R(z)\,dz^2 \geqq 0$ that end at the zeros and the boundary slit. f should map the trajectories of the z-plane onto those of the w-plane. It is quite obvious that this cannot be the case in the wrong solution since they are linked in a wrong way. The tip, which is found by numerical integration of equation (1), starts at

[2] In this case, equation (1) is not uniquely determined by k.

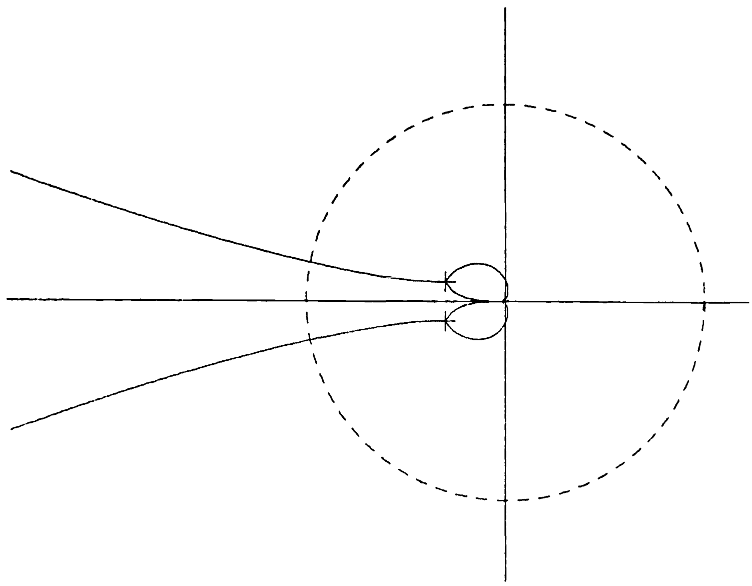

FIG. 1

the wrong place in Fig. 3. This is another indication that something went wrong. This will be explained in a later section.

This gives us a clue how to decide whether we have a right solution or not. It is known that the trajectories starting at the critical points β_i of Q, and ∞ cut the w-plane into three half-plane domains (so called since $\zeta = \int_{\cdot}\sqrt{Q(w)}\, dw$ maps those domains 1–1 onto a half plane), and finitely many strip domains (so called because ζ maps those domains onto a strip), there not being any circular, ring or ergodic flow domains (see [10, pp. 68, 69]). The trajectories bounding the half plane domains all have limiting endpoints at zero, and there are three limiting directions there. Those belonging to the same half-plane domain enclose an angle of $2\pi/3$. f does not rotate these trajectories since $f'(0) = 1$ and it is easy to determine which corresponds to which.

We are now ready to prove univalence in particular cases assuming that (2)–(6) are satisfied. (The general case seems unnecessarily complicated and does not give much

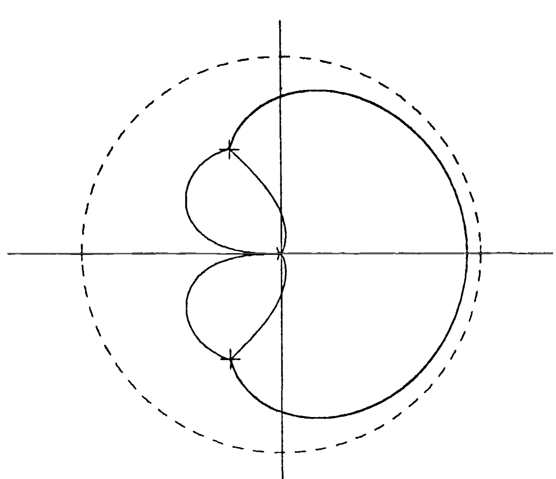

FIG. 2

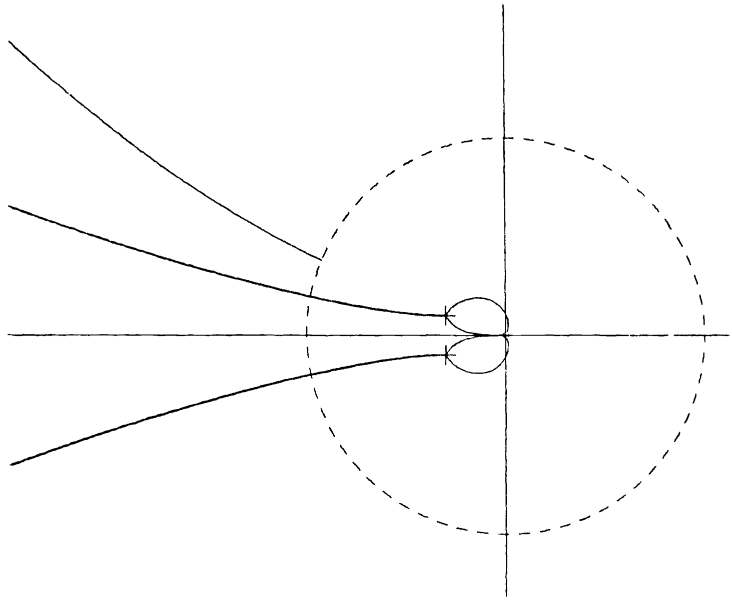

FIG. 3

insight.) Let us take Figs. 5 and 6 and explain the various arcs first: The tip of the slit has been found by integrating the differential equation (1) numerically along the line segment $\overline{\alpha_1\alpha_0}$. This corresponds function-theoretically to an analytic continuation of the function elements $w = f(z)$ given by (1) along the segment $\alpha_1\alpha_0$. At the end we get $\beta_0 = f(\alpha_0)$ which is, if there is a unique function f, the tip of the boundary slit. From there the trajectory $Q(w)\,dw^2 \geqq 0$ was followed in the direction away from the region to give the complementary slit. For the following argument it should be considered an illustration *only* since it uses the existence of f. On the other hand there must be a trajectory of $Q(w)\,dw^2 \geqq 0$ coming from ∞ and ending at zero and the only open place is near our trajectory, so we can assume that our picture is qualitatively correct. The other solid lines are the trajectories starting at the β_i, respectively the α_i, with the exception of the trajectory from α_0 to the origin. There are two strip domains, S_1 and S_2 and one half-plane domain H_1 adjacent to α_1, and strip domains R_1 and R_2 and a half-plane

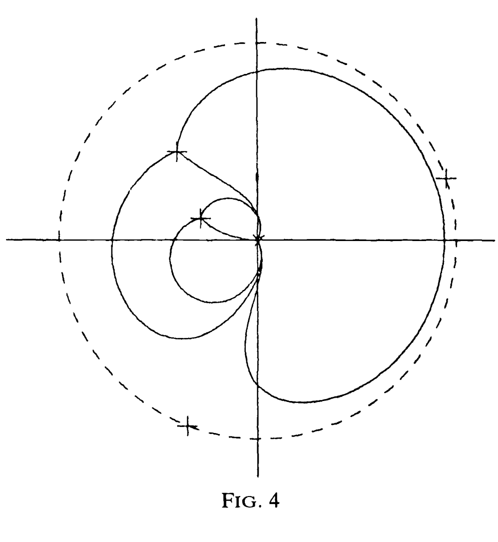

FIG. 4

domain K_1 adjacent to β_1. The functions $\zeta=\int_{\alpha_1}\sqrt{R(z)}\,dz$ respectively $\eta=\int_{\beta_1}\sqrt{Q(w)}\,dw$ are univalent in these respective domains. Therefore the equation $\zeta=\eta$ gives a univalent function in the half-plane domain $H_1\to K_1$. For the strip domains we have to show that the associated strips in the ζ, respectively η-plane, have the same width. First, we consider the strip R_1 whose other side contains ∞ and the boundary slit respectively S_1 whose other side contains the unit circle. The width of the strip in the ζ-plane is $|\mathrm{Im}\int_{\alpha_1}^{\alpha_\infty}\sqrt{R(z)}\,dz|$ taken along the straight line while the width of the strip in the η-plane is $|\mathrm{Im}\int_{\beta_1}^{\infty}\sqrt{Q(w)}\,dw|$ and the two are equal according to (6). Call this number h_1. If we consider now the line segment $\alpha_1\alpha_2$ we observe that we first cross a bit of the half-plane domain H_1 then S_1 until we meet the trajectory coming from α_0 (not in picture but it must end at zero for geometric reasons, see [10, pp. 61, 62]). Then we cross S_1 again then S_2 and then a piece of a half-plane domain adjacent to α_2, H_2. So the total increment of Im ζ is $\pm(2h_1+h_2)$ where h_2 is the width of the second strip domain. The same reasoning gives us that the increment of Im η along the straight line joining β_1 and β_2 is $\pm(2h_1+h_2')$ where h_2' is the width of the second strip in the η-plane. Equation (4) gives us now at once $h_1=h'$ and the equation $\zeta=\eta$ gives a univalent function in S_1 and S_2.

From the definition of these mappings it is obvious that they "match" along the trajectories separating H_1 from S_1, respectively S_2, but we also have to prove that this holds along the trajectory from α_0 to 0 which separates S_1 from itself. For this we observe that equation (6) insures that α_∞ corresponds to ∞. Since the unit circle is a trajectory the two circular segments from α_∞ correspond to the two sides of the trajectory coming from ∞. Equation (5) insures now that α_0 corresponds to one point in the w-plane only, no mattter whether we approach it from below or above. The same now holds for the whole trajectory, joining α_0 with zero. We have shown that the equation $\zeta=\eta$ provides a univalent function in $cl(S_1\cup S_2\cup H_1)$. It is now easy to see that equation (4) implies that α_1 corresponds in the mapping to β_1, and that the image of the straight line segment $\alpha_1\alpha_2$ is homotopic to the straight line segment $\beta_1\beta_2$. A repetition of the previous argument shows that $\zeta=\eta$ also gives a univalent map in the remaining two half-plane domains H_2 and H_3. Since these again match at their boundaries we have a univalent function defined in the unit circle, thus completing the proof.

Remark. The trajectories that are boundary to R_1 leave the picture. In our earlier attempts they were "lost" and it was just a guess that they would come back, as shown in Fig. 5. The univalence of f can actually be shown without them but this is somewhat more complicated, since it is then also not certain that what appears to be the boundary slit is really the boundary slit because it might a priori be in a bounded component of the complement of these trajectories, due to the ambiguity of the square roots for the integration of (1). We could have avoided some of these difficulties by following the trajectory from ∞, and using equation (5) to determine its length. This would have involved a transformation $u=1/w$, since the computer does not handle ∞, and a converter for the length of the trajectory; i.e. it would have been slightly more complicated. We are grateful to the referee for this suggestion.

4. Problem 2. In this case we need the equations for Problem 1 and further the function f must have the right coefficients a_2 and a_3. We could of course not expect a solution and most certainly not a "correct" one if (a_2, a_3) were not in V_3. To evaluate for a given Q and R, a_2 and a_3, we set

$$Q(w)=\frac{q_5}{w^5}+\frac{q_4}{w^4}+\frac{q_3}{w^3}, \qquad R(z)=\frac{r_5}{z^5}+\frac{r_4}{z^4}+\frac{r_3}{z^3}+\cdots,$$

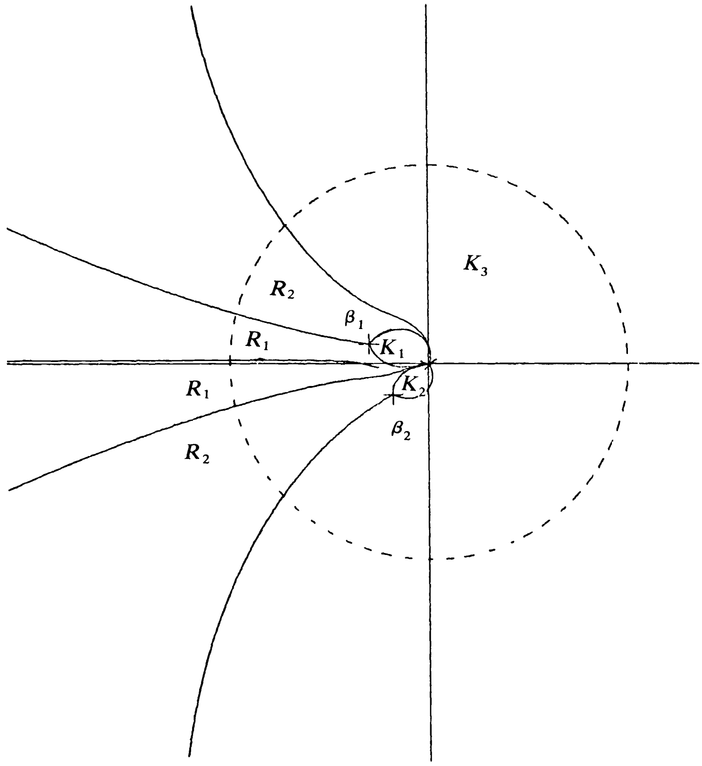
K_3
R_2
β_1
R_1
K_1
R_1
K_2
β_2
R_2

FIG. 5

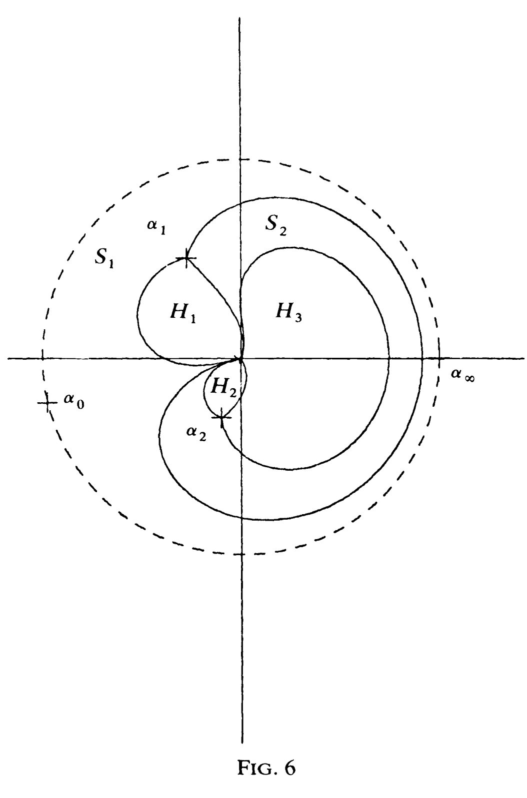
α_1
S_2
S_1
H_1
H_3
H_2
α_∞
α_0
α_2

FIG. 6

$$f(z)=z+a_2z^2+a_3z^3+\cdots,$$
$$[f'(z)]^{-2}=[1+2a_2z+3a_3z^2+\cdot]^{-2}=1-4a_2z-6a_3z^2+12a_2^2z^2+\cdots.$$

So

$$R(z)f'(z)^{-2}=\left[\frac{r_5}{z^5}+\frac{r_4}{z^4}+\frac{r_3}{z^3}\right][1-4a_2z+z^2(12a_2^2-6a_3)+\cdots]$$
$$=\frac{r_5}{z^5}+\frac{r_4-4a_2r_5}{z^4}+\frac{r_3-4a_2r_4+r_5(12a_2^2-6a_3)}{z^3}+\cdots.$$

On the other hand we get for $Q(w)$:

$$Q(w)=q_5[z(1+a_2z+a_3z^2+\cdots)]^{-5}+q_4[z(1+a_2z+\cdots)]^{-4}+q_3[z+\cdots]^{-3}$$
$$=\frac{q_5}{z^5}+\frac{q_4-5a_2q_5}{z^4}+\frac{q_3-4a_2q_4+q_5[15a_2^2-5a_3]}{z^3}+\cdots.$$

And (1) therefore yields the equations:

$$q_5=r_5 \quad (\text{meaning } f\in S),$$
$$q_4-5a_2q_5=r_4-4a_2r_5, \tag{7}$$
$$q_3-4a_2q_4+q_5(15a_2^2-5a_3)=r_3-4a_2r_4+r_5(12a_2^2-6a_3). \tag{8}$$

Of course the r_i have also to be expressed in terms of the α_i. For this see § 5. If these equations are solved together with the equations of Problem 1 ((2), (3), (4), (5) and (6)) and if f is univalent which is to be shown as in the case at the end of the last paragraph, then we have the solution of the problem: Maximize Re $(-q_5a_4)$ subject to the condition that a_2 and a_3 are the given values. (See [4] and [7].) Since we can multiply both sides of equation (1) by a positive constant without affecting any conditions we can assume that $|q_5|=1$ or $-q_5=e^{i\varphi}$ where φ is some real number.

As in Problem 1 we reduced the number of real equations from 14 to 7 using algebraic techniques. A plot for the solution continuum for this problem is displayed in Figs. 7 and 8. The input values were $a_1=1.92-0.44_i$, $a_2=2.66-1.29_i$ and $\varphi=0.0$.

5. The computational procedure. Programs were written that accomplished the following: 1) solved a set of nonlinear equations using the sequential secant method; 2) evaluated the equations (2)–(6) for Problem 1 and (2)–(8) for Problem 2; 3) plotted the two continua; 4) located the tip of the slit from ∞; 5) plotted the slit. The procedures differ slightly between the two problems and we will identify the differences as they arise.

1. *The sequential secant method.* The main program calls a subroutine to evaluate equations (2)–(6) or (2)–(8). It does not recognize that it deals with a set of complex equations since the values of the function are returned in real form. We are therefore simply solving a set of real equations, 5 in the case of Problem 1 and 7 in the case of Problem 2. We used the sequential secant method of [1] and obtained reasonable results. A problem in the case of nonlinear equations of this kind is of course to obtain starting values for the iterations.

Since the Koebe function has a known solution, we therefore started by solving problems close to it using the solution for the Koebe function as input guesses. We could then use the solutions for these problems as guesses for problems further away from the Koebe function. This procedure was repeated until we had a satisfactory number of solutions. An algorithmic implementation of a technique also including predicting

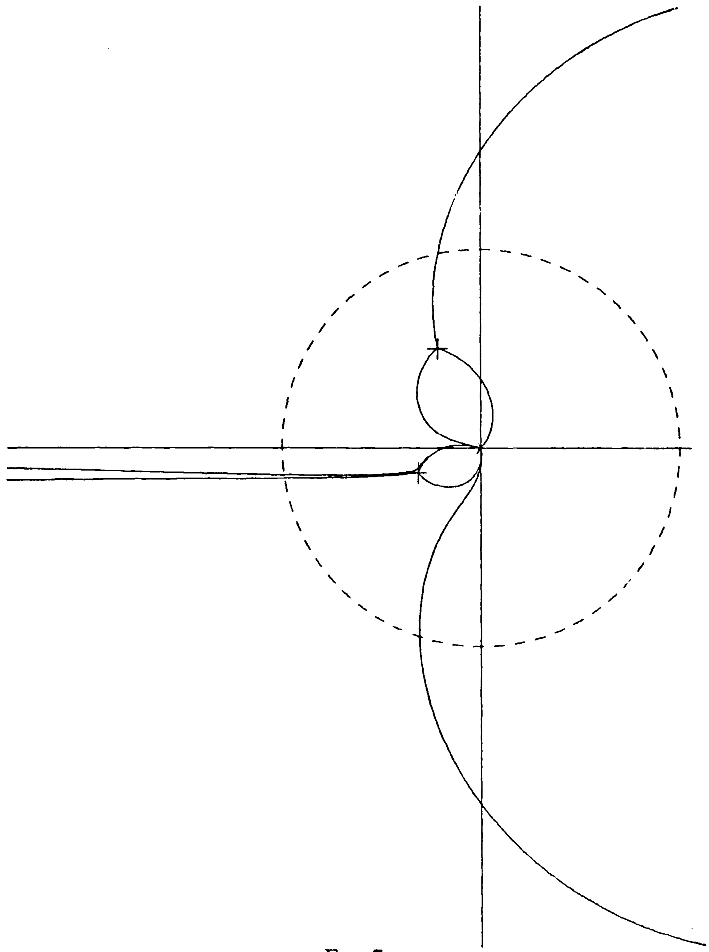

FIG. 7

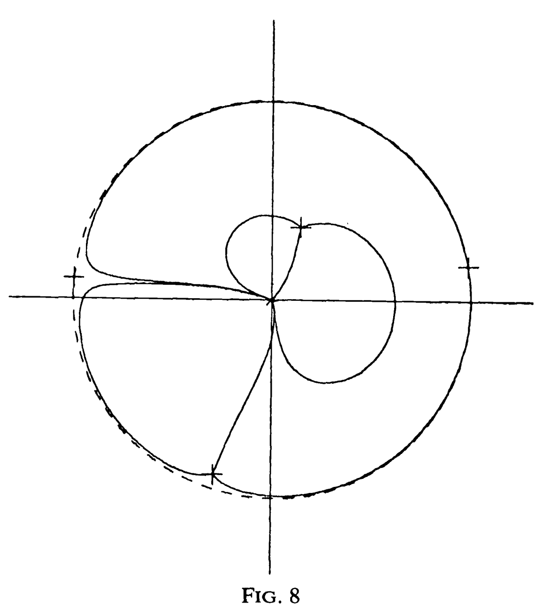

FIG. 8

initial guesses was done in [3] and it was suggested to us by Rheinboldt [9]. In view of the complexity of the program to evaluate (2)–(6) and (2)–(8), we decided not to implement this suggestion, but we did implement a simple form of this idea. The program would take a solution and perturb the inputs by fixed amounts. It would then attempt to find a solution to the perturbed program, using the solution to the last problem as input data. Since our main interest was to find some solutions we felt that this approach was satisfactory.

The sequential secant method was successful if either the norm of the solution vector was small or the last solution vector was close. We used a tolerance of 10^{-6} in both cases.

2. *Evaluating the equations.* A subroutine was written for each problem that evaluated the equations (2)–(6), respectively (2)–(8). This subroutine again called routines to perform a number of tasks.

The first task was to evaluate $\int_{\alpha_1}^{\alpha_2} R(z)\,dz$. This was accomplished using Simpson's rule over 2^n subdivisions ($n=6$, 7 or 8). In some cases, however, the path of integration went almost through the origin which is a pole of order 5. We therefore encountered severe numerical difficulties and we had to integrate via $-0.5+0.5i$ or $-0.5i$ depending on the case. The flags discussed in [6] were of course used to keep track of the branch of the square root.

Secondly, an iteration was performed to find α_∞. We know that α_∞ satisfies the equation

$$\int_{\alpha_0}^{\alpha_\infty} \sqrt{|R(z)|}\,d\varphi + \int_{\alpha_0}^{\alpha_\infty} \sqrt{|R(z)|}\,d\varphi = 0. \tag{9}$$

This equation is analytic in $\arg \alpha_\infty$; hence it may be solved iteratively using Newton's method. Letting (9) be written as $f(\arg \alpha_\infty)=0$ we get

$$\frac{df(\arg \alpha_\infty)}{d \arg \alpha_\infty} = 2\sqrt{|R(z)|}$$

and the iteration

$$\alpha_\infty^{i+1} = \alpha_\infty^{(i)} - \frac{f(\alpha_\infty^{(i)})}{2\sqrt{|R(z)|}}, \qquad i=0, 1, 2, 3, \cdots.$$

This iteration scheme converged rapidly in all cases when the initial guess was $\alpha_\infty^{(0)} = -\alpha_0$. The integrals $\int_{\alpha_0}^{\alpha_\infty} \sqrt{|R(z)|}\,d\varphi$ and $\int_{\alpha_0}^{\alpha_\infty} \sqrt{|R(z)|}\,d\varphi$ were evaluated using the same subroutine taking care to go in the correct directions around the unit circle. No flags were needed in this routine since we were dealing with real integrals. Again, Simpson's rule was used. The integral $\int_{\alpha_1}^{\alpha_\infty} \sqrt{R(z)}\,dz$ was computed using the same routine as for $\int_{\alpha_1}^{\alpha_2} \sqrt{R(z)}\,dz$.

A problem arose in evaluating $\int_{\beta_1}^{\infty} \sqrt{Q(w)}\,dw$ due to the improper integral. This was solved by standard techniques. In fact we define $S(u)=\sqrt{\beta(1-u\beta_1)(1-u\beta_2)}$ and $u=1/w$. Then

$$\begin{aligned}\int_{\beta_1}^{\infty} \sqrt{Q(w)}\,dw &= \int_{1/\beta_1}^{0} -\sqrt{Q(1/u)}\frac{1}{u^2}\,du\\ &= \int_0^{1/\beta_1} \frac{S(u)}{\sqrt{u}}\,du\\ &= \int_0^{1/\beta_1} \frac{S(u)-S(0)}{\sqrt{u}}\,du + \frac{3}{2}\frac{1}{\beta_1}S(0),\end{aligned}$$

the last transformation being done to avoid the singular endpoint at 0. In the case of Problem 1, this integral was evaluated once in the main program since the value of the integral did not change as the iteration proceeded. In the case of Problem 2, this integral had to be evaluated for each iteration. For Problem 2, the subroutine evaluating the function also had to compute the numbers $\beta_1, \beta_2, \alpha_0, \alpha_1, \alpha_2$ and α from the input values β, a_2 and a_3 at the beginning of the routine. Let the real unknowns be $x_1, x_2, \cdots, x_7$. These are assigned input approximations. Then using equations (8) and (9) we may write the following sequential assignments:

$$\alpha^1 \leftarrow x_1 + ix_2$$ [3]

$$\alpha_2 \leftarrow x_3 + ix_4$$

$$\alpha_0 \leftarrow e^{ix_7}$$

$$\rho \leftarrow \prod_{i=0}^{2} \frac{\alpha_i}{\overline{\alpha_i}}$$

$$b_1 \leftarrow x_5 + ix_6$$

$$r_5 \leftarrow \frac{\alpha_1\alpha_2}{\overline{\alpha_1\alpha_2}}$$

$$r_4 \leftarrow -\left(\left(\frac{\alpha_2+\alpha_2}{\overline{\alpha_1\alpha_2}}\right) + (\alpha_1 * \alpha_2)\left(\frac{1}{\overline{\alpha_1}} + \frac{1}{\overline{\alpha_2}}\right)\right)$$

$$r_3 \leftarrow \alpha_1\alpha_2 + \frac{1}{\overline{\alpha_1\alpha_2}} + (\alpha_1 + \alpha_2)\left(\frac{1}{\overline{\alpha_1}} + \frac{1}{\overline{\alpha_2}}\right)$$

$$r_3 \leftarrow r_5 - r_4\left(\alpha_0 + \frac{1}{\overline{\alpha_0}}\right) + \frac{\alpha_0}{\overline{\alpha_0}} r_3$$

$$r_4 \leftarrow -r_5\left(\alpha_0 + \frac{1}{\overline{\alpha_0}}\right) + \frac{r_4\alpha_0}{\overline{\alpha_0}}$$

$$r_5 \leftarrow r_5 \frac{\alpha_0}{\overline{\alpha_0}}$$

$$\beta_2 \leftarrow \frac{-\alpha_1}{(1 + 5a_2\alpha_1 + (\alpha_1/\rho)(r4 - 4r_5a_2))}$$

$$\alpha \leftarrow \frac{\beta\beta_1\beta_2}{\rho}$$

$$r_3 \leftarrow r_3\alpha$$

$$r_4 \leftarrow r_4\alpha$$

$$r_5 \leftarrow r_5\alpha$$

$$q_3 \leftarrow \beta$$

$$q_4 \leftarrow -\beta(\beta_1 + \beta_2)$$

$$q_5 \leftarrow \beta\beta_1\beta_2$$

$$v \leftarrow q_3 - r_3 + 4a_2(r_4 - q_4) + a_2^2(15q_5 - 12r_5) + a_3(6r_5 - 5q_5)$$

[3] The notation $\leftarrow$ means "assigned to" as used in computer languages.

where v is now the error in equation (8) which in the solution process provides us with the value of one of the equations. The above rearrangement of the equations allows us to solve systems in 7 unknowns rather than 9 which a simple application of (7) and (8) would give us.

In Problem 1, we evaluated the coefficients a_2, a_3 and a_4 of equations (7) and (8). In order to do this we wrote a routine that computed the coefficients of a polynomial from its roots. We then applied this routine to the roots of $R(z)$ and $Q(w)$ that the program had calculated obtaining the coefficients r_i, $i = 2, \cdots, 5$ and q_i, $i = 3, 4, 5$. By rearranging (7) and (8) as well as the next equation obtained from (1) we could compute a_2, a_3 and a_4. In Problem 2 the coefficients a_2, a_3 and a_4 were evaluated in the same manner for checking purposes.

3. *Plotting the continua.* In both Problem 1 and Problem 2 the programs were now identical from this point on. We first plotted the trajectories of $Q(w)\,dw^2 \geqq 0$ and $R(z)\,dz_2 \geqq 0$ respectively. At each zero of Q and R we may go in three directions. We have, for instance, for $Q(w)$:

$$0 = \arg\,[Q(w)\,dw^2] = \arg \beta \left[\frac{(w-\beta_1)(w-\beta_2)}{w^5}\,dw^2\right].$$

For $w \to \beta_1$ we get

$$0 = \arg \beta \frac{w-\beta_1}{w^5}\,dw^3$$

and therefore

$$\arg dw = \frac{1}{3}\left[\arg \beta \frac{(w-\beta_2)}{w^5} + 2n\pi\right]$$

which gives three directions $2\pi/3$ radians apart. We stepped a small step in one of the directions, then a few steps using simple steps. From then on a predictor-corrector method was used (see Milne [8, p. 65] for method, and our paper [6]). There are problems, of course, as to where the trajectories would terminate, but these were solved using appropriate programming. In general an additional complication arose out of the fact that the plots in the w-plane went off the paper and back in again. This was solved using appropriate counters and flags. In some cases, however, the trajectories went so far out that for practical purposes we had to terminate before it returned towards the origin. One such return trajectory is missing in a couple of plots. In Figs. 7 and 8 for example, we let our computer run for 20 minutes but the trajectory close to the slit did not return. The trajectory could of course be obtained by starting from the origin. We omitted that final programming effort since we felt it would not add substantially to the displayed results.

4. *The tip of the slit.* In order to find the tip of the slit the equation

$$\sqrt{R(z)}\,dz = \sqrt{Q(w)}\,dw$$

was integrated from the initial point β_i, $i = 1, 2$, to b_0, the tip of the slit (see also § 3). The initial directions for this integration are chosen from

$$dw = \frac{1}{3}\left(\arg\left|\left(\frac{R(z)}{z-\beta_i}\right)\left(\frac{w-\alpha_i}{Q(z)}\right)\right| + 2\pi\right) + dz.$$

Since this gives six possible endpoints we choose the two that coincide.

If one were to choose a wrong initial direction for the integration one would follow a curve in the w-plane that encloses the same angle with the trajectories of $Q(w)\,dw^2$ as the line segment $\overline{\alpha_i\alpha_0}$ makes with the trajectories of $R(z)\,dz^2$ but would start in a wrong half-plane (or slit)-domain. It is then very unlikely that one would reach a close point starting from the other zero of R, respectively Q. It is interesting to note that in some cases of "wrong solutions" the computer accepted "the tip of the slit" in bounded components (see Fig. 4).

Having found the endpoint of the slit, the slit was plotted using the same technique as for the continua.

A total of 15 routines of various kinds was needed for Problem 1 and 16 for Problem 2.

5. *Numerical examples.* In order that one should get an appreciation of the kinds of calculations one has to do we give a numerical example both of Problem 1 and of Problem 2.

In Problem 1 we input two flags +1, and +1. These are used to control the sign of $\int_{\alpha_1}^{\alpha_2}\sqrt{R(z)}\,dz$ and $\int_{\alpha_1}^{\alpha_\infty}\sqrt{R(z)}\,dz$ since these may change from one set of solutions to the next.

We then input

$$\beta_1=-0.3785+i0.1785$$

$$\beta_2=0.3-i0.1$$

$$\beta=\pi.$$

Since we are using the sequential secant method we need $n+1$ guesses for n unknowns. The unknowns are d_1, d_2 and α and the input guesses are given in the following table.

TABLE 1

Re (α_1)	Im (α_1)	Re (α_2)	Im (α_2)	α
−0.10568	0.5097029	−0.35697298	−0.90227015	3.1949666
−0.1056845	0.5097029	−0.35697298	−0.90227	3.1949666
−0.1056845	0.509702	−0.35697298	−0.90227015	3.1949666
−0.1056845	0.5097029	−0.35697	−0.90227015	3.1949666
−0.1056845	0.5097029	−0.35697298	−0.90227015	3.1949666
−0.1056845	0.5097029	−0.3567298	−0.90227	3.19496

After the two iterations we computed

$$\alpha_1=-0.10565422+i0.50969774$$

$$\alpha_2=-0.35661963-0.90139553$$

$$\alpha=3.1950142$$

with an error of 0.98900644E-7. The coefficients are

$$\alpha_2=1.9324569-i0.40410985$$

$$a_3=2.7450348-1.0865342.$$

This example was plotted in Figs. 9, 10. Both in this example and in the two following examples the unit circles give the scale for the numerical values.

In the case of Problem 2 the solutions were somewhat harder to obtain due to the sensitivity of the solutions to the input values.

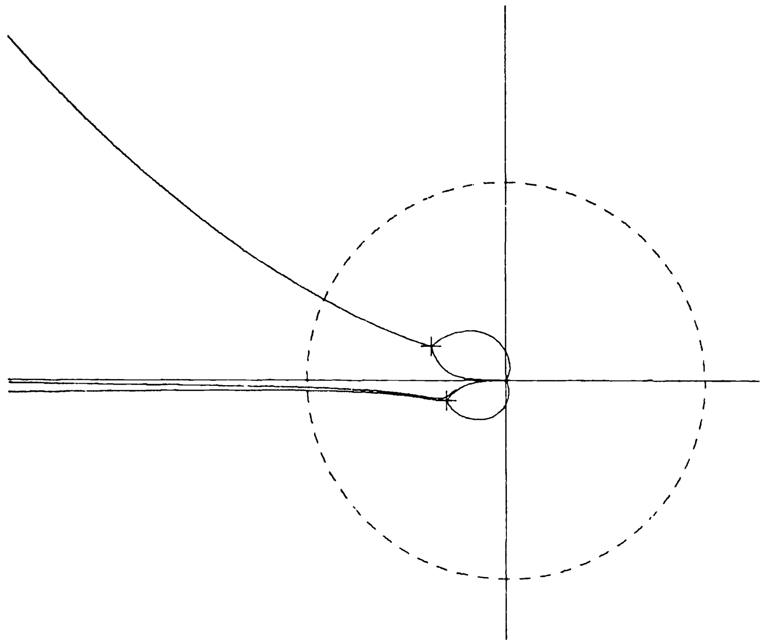

FIG. 9

The solution graphs of Figs. 7 and 8 were obtained starting with the inputs

$$a_2 = 1.92 - 0.44i$$

$$a_3 = 2.66 - 1.29i$$

$$= 0.0.$$

Since there are seven real equations we need eight sets of input approximations. These are given in Table 2.

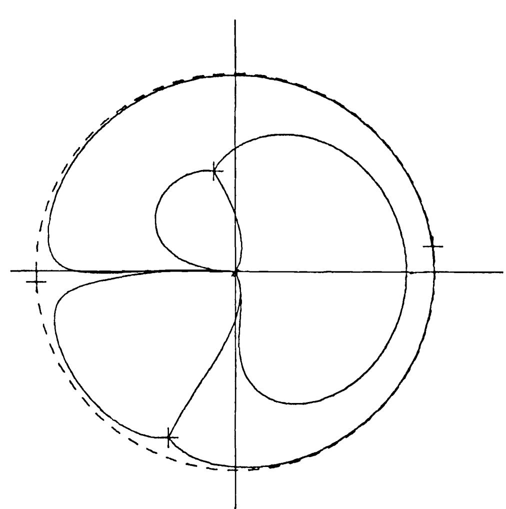

FIG. 10

TABLE 2

Re (α_1)	Im (α_1)	Re (α_2)	Im (α_2)	Re (β_1)	Im (β_1)	β_0
0.14077	0.39930598	−0.18048628	−0.70184641	−0.28441	0.50898933	3.1328576
0.14077439	0.3993	−0.18048628	−0.70184641	−0.28441	0.50898933	3.1328576
0.14077439	0.39930598	−0.1804	−0.70184641	−0.28441	0.50898933	3.1328576
0.14077439	0.39930598	−0.18048628	−0.7018	−0.28441	0.50898933	3.1328576
0.14077439	0.39930598	−0.18048628	−0.70184641	−0.2844	0.50898933	3.1328576
0.14077439	0.39930598	−0.18048628	−0.70184641	−0.28441	0.50898	3.1328576
0.10477439	0.39930598	−0.18048628	−0.70184641	−0.28441	0.50898933	3.1328
0.10477439	0.39930598	−0.18048628	−0.70184641	−0.28441	0.50898933	3.1328576

The sum of the absolute values of the errors in these equations was of the order of 3×10^{-2} for all sets of inputs. Even so, the solution process needed 25 iterations in order to find a close enough solution. This suggests that the constant for linear convergence for the sequential secant method is very small for this problem. The output in this case was

$$\alpha_1 = 0.14137042 + 0.36632225i$$

$$\alpha_2 = -0.29189849 - 0.87970653i$$

$$\beta_1 = -0.21578116 + 0.50209839i$$

$$\beta_0 = 3.0355212$$

from which one could calculate

$$\beta_2 = -0.31494625 - 0.12220681.i$$

With these values we recalculated the inputs a_2 and a_3 as well as the coefficient a_4. It was of course encouraging to observe the following output

$$a_2 = 1.9200000 - 0.44000000i$$

$$a_3 = 2.6600000 - 1.2900000i$$

$$a_4 = 3.1378254 - 2.3594080i.$$

A small perturbation in the input data to

$$a_2 = 1.92 - 0.44i$$

$$a_3 = 2.66 - 1.3i$$

$$= 0.0$$

resulted in the outputs

$$\alpha_1 = 0.21334869 + 0.30924486i$$

$$\alpha_2 = -0.37487270 - 0.8742856i$$

$$\beta_1 = -0.03777511 + 0.63257063i$$

$$\beta_0 = 2.9923485$$

as well as a calculated

$$\beta_2 = -0.30700235 - 0.11339281i.$$

This is of course not surprising since this function is fairly close to the Koebe-function.

6. *Practical problems.* In the previous paragraphs we described the numerical method used. The method is not foolproof even for Problem 1, as can be seen from the following numerical calculations.

We start with

$$\beta_1 = -0.3 + 0.2i$$

$$\beta_2 = -0.3 - 0.09i$$

$$\beta_0 = -\pi$$

and increment Re (β_1) by -0.05 each time we compute a solution. We get the following sequence of solutions:

TABLE 3

Re (β_1)	α_1	α_2	$\theta(\alpha_0 = e^{i\theta})$
-0.3	$-0.08143+0.4243i$	$-0.3829-0.7793i$	3.1122
-0.35	$-0.0695+0.4775i$	$-0.4826-1.1501i$	3.1662
-0.4	$0.0438+0.5261i$	$-0.4444-1.3245i$	3.2100
-0.45	$-0.0145+0.5655i$	$-0.3758-1.4572i$	3.2416
-0.50	$-0.0163+0.5972i$	$-0.2963-1.5642i$	3.2654
-0.55	$0.07437+0.6030i$	$0.3410-1.5808i$	2.9946
-0.60	$0.07356+0.6192i$	$0.3555-1.4095i$	2.9824
-0.65	$0.0689+0.6347i$	$0.3498-1.229i$	2.9655
-0.70	$0.0682+0.6493i$	$0.2738-0.8622i$	2.952

Since the problem only admits solutions for which $|\alpha_i| \leqq 1$, we clearly found several "bad" solutions, i.e. either wrong solutions or solutions that solve the equations but not the problem. We note that α_2, the cause of the problem, took a trip outside the unit circle but returned later. The question then can be posed: Is the last solution in Table 3 a good solution, i.e. a solution to the original problem? The answer can only be given by producing a plot for this case. In Figs. 13 and 14 we see the plot for that case. We note that the graph of the trajectories looks turned only but topologically alright. One is tempted to mend this problem by turning the solution appropriately but with closer inspection we note that Im $\int_{\alpha_1}^{\alpha_\infty} \sqrt{R(z)}\, dz$ is equal to the sum of the widths of S_1 and S_2 in Fig. 14 while $\int_{\beta_1}^{\infty} \sqrt{Q(w)}\, dw$ is only the width of R_2. It turns out that the width of S_2 is the difference of the two. Note here that the boundary slit is in a bounded component.

We tried to rescue such "wrong" solutions in a few cases by replacing α_i by $1/\bar{\alpha}_i$ and some similar tricks but never got close enough to the correct solution to insure convergence. A program that can handle split trajectories might actually help here.

Almost every set of about 10 consecutive solutions with α_i's about 0.05 apart displayed this peculiar behavior.

We are left with the distinct impression that the problem in this form is not tractable using continuation methods. Human intervention is needed at every solution point.

The above description of the method and the program does not do proper justice to the number of difficulties that arose in the programming of the problem. We believe that these difficulties are inherent to this type of approach and we maintain that such problems should only be solved if one has a large amount of time at one's disposal both for programming, for debugging and for general problem-shooting.

It is surprising how fast the trajectory from ∞ moves into the zeros and respectively how fast the α's move towards the boundary. It seems as if the zero "likes" the trajectory from ∞. It is not at all clear to us why this is the case. Figures 9–12 illustrate this behavior.

Problem 2 is another matter again. Not only does it contain all the difficulties of Problem 1, but also the following: Some algebraic functions, most notably the Koebe function, do not determine $Q(w)$ uniquely. The β_i have then a freedom for a given a_i and this implies that the Jacobian of the system of equations is singular.

If we are very close to such a function, the system is very nearly singular and the iterative methods of solution perform very poorly. We also tried quasi Newton methods with no more success than with the sequential secant method.

The next difficulty is that starting at any place and perturbing one is likely to hit such a function sooner or later. The β_i then have a jump discontinuity and the old values for the β_i are too far off to ensure convergence of the sequential secant method.

To get some solutions we started out with a solution of Problem 1 that looked far away from the Koebe function and perturbed the a_i. We believe that the problem is solved in principle if we can produce solutions with a_2 and a_3 being not too complicated numbers. In this context it should be pointed out that in the third coefficient region the part of the boundary plot corresponding to functions whose complementary slit does not split is quite small (see pictures in [10]). We should therefore in the case of V_4 not expect that the a_2 and a_3 are one digit numbers. We admit that for a_2 and a_3 given at random in V_3 a solution would still require a substantial amount of work and computer time.

In Figures 15–18 we have shown two more representative solutions to Problem 1. Figures 19–24 again show a sequence of three solutions to Problem 1 showing the rapidity by which the α's move towards the boundary.

6. Concluding remarks. Since this approach to these problems is so hard, in fact much harder than we expected, the following approaches gain interest:

(a) One could replace the integrals by a numerical integration of (1) along the line segments $\overline{\beta_1\beta_2}$ and $\overline{\beta_1\infty}$ and evaluate R at the endpoint as criterion for closeness. This would remove some of the difficulties, but the ambiguity of the square roots would stay and one would probably lose accuracy, a typical differential equation solver having lower accuracy than equation solvers.

(b) One could find the boundary-slit by integrating $Q(z)\,dz^2 \geqq 0$ starting at ∞, and then use the techniques of [2] to find the mapping function. An obvious difficulty here is to implement the conditions on the coefficients without using too much computer time. This approach would on the other hand have the advantage that higher coefficients could be handled without much more difficulty.

(c) Another way would be to use the Biberbach–Eilenberg class instead of S. This promising approach has been followed by J. Hummel.

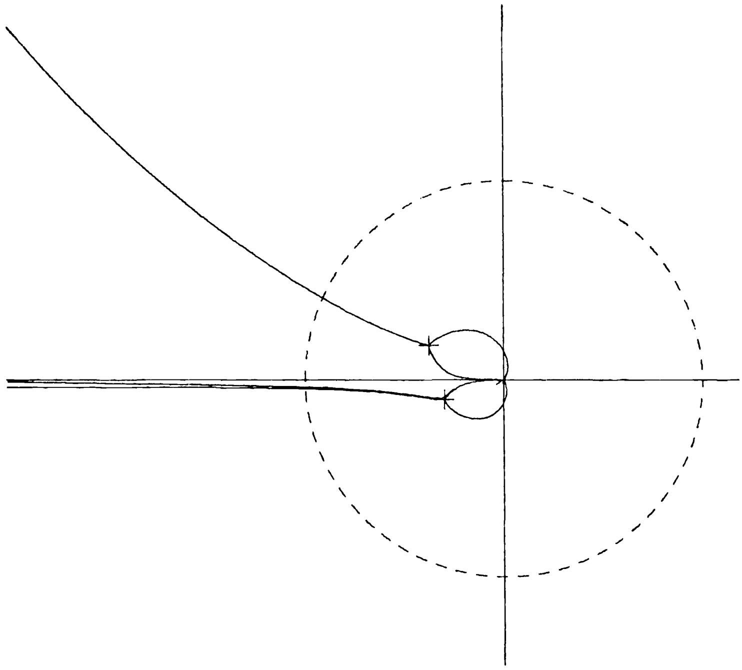

FIG. 11

(d) Since higher coefficients seem to be too hard for practical reasons, one could use the theory of elliptic integrals instead of Simpson's rule. This would reduce the computer time and thus give a broader scope for V_4. It could not be used for higher coefficients (suggestion by the referee).

Still, in spite of all the difficulties, we feel very strongly that such programs should be undertaken. If for nothing else, it is valuable for the same reason that it is valuable for

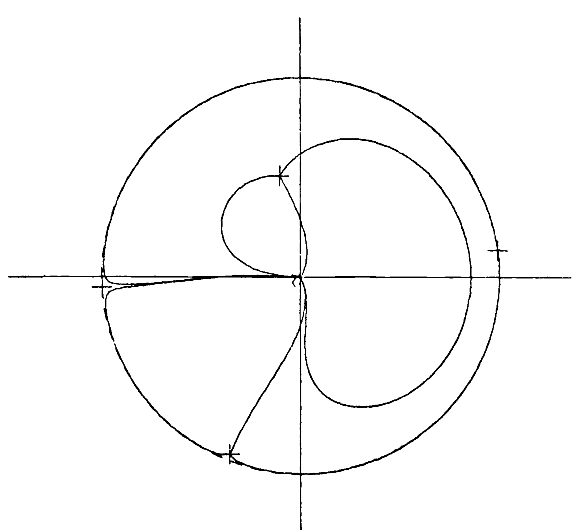

FIG. 12

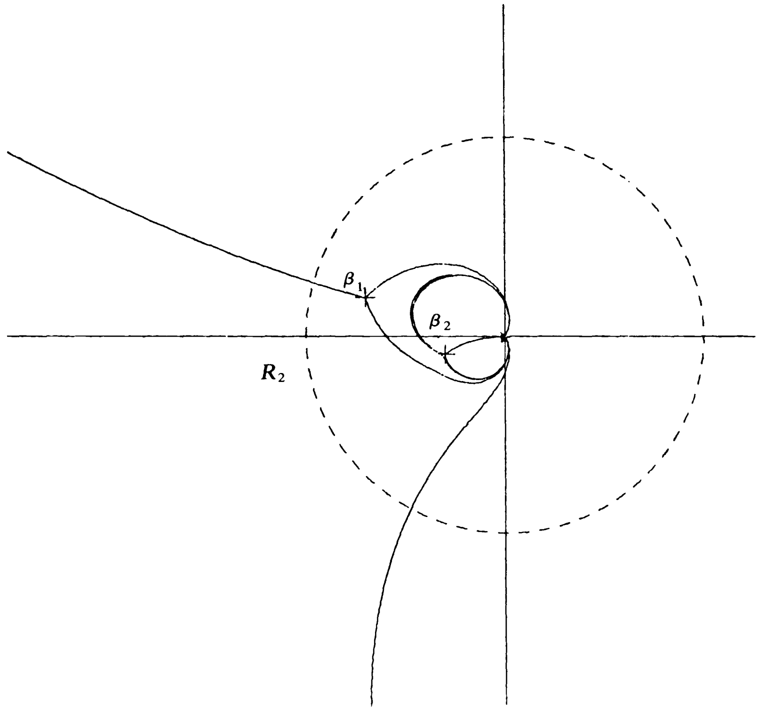

FIG. 13

a student to calculate examples of an abstract theory, namely to see what is going on.

It is also needless to say that for mathematicians who share our taste it is sometimes just as satisfying to solve a problem as to prove a theorem as there are a lot of problems remaining in the direction of [5], [6] and the present paper, which range from mere term projects to problems of almost unlimited difficulty.

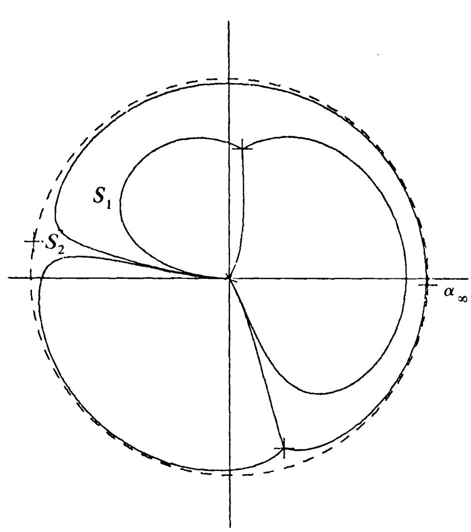

FIG. 14

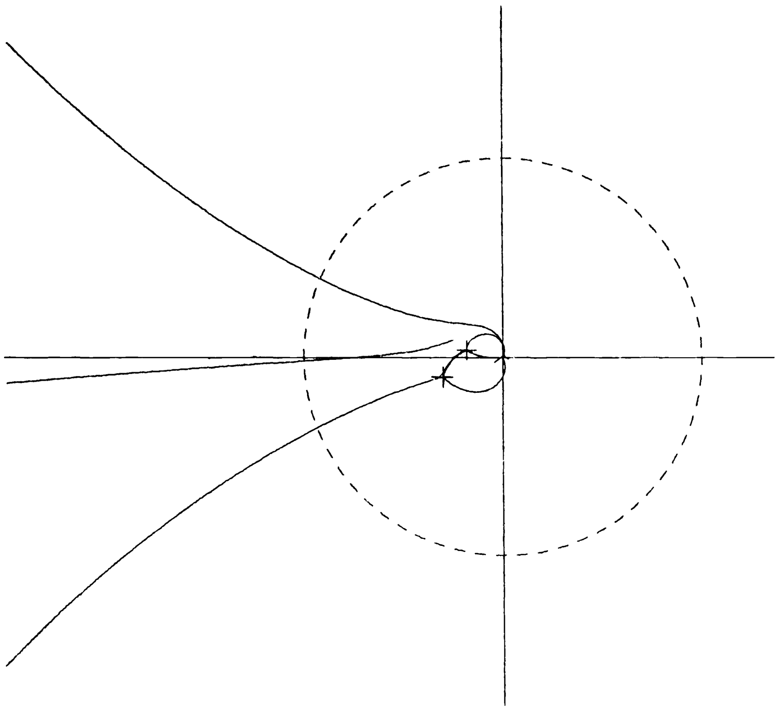

FIG. 15

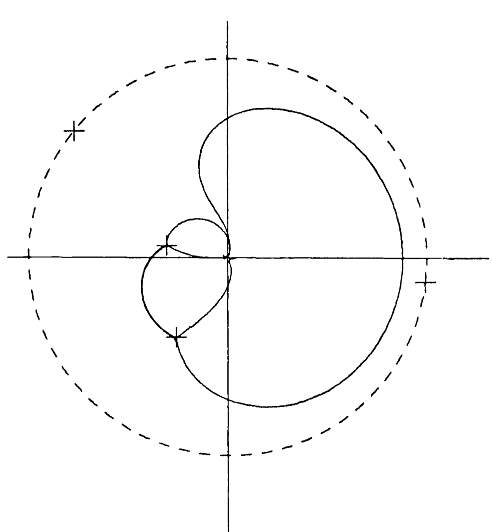

FIG. 16

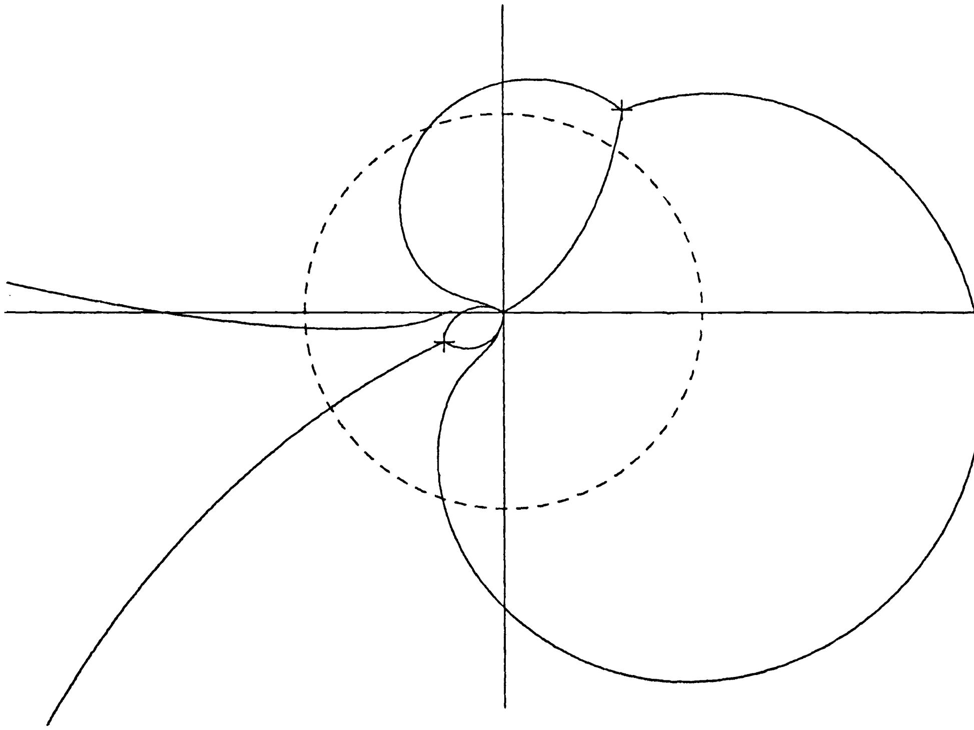

FIG. 17

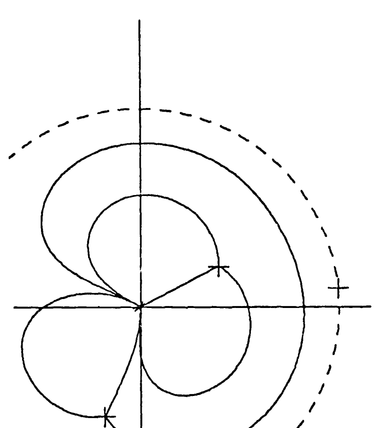

FIG. 18

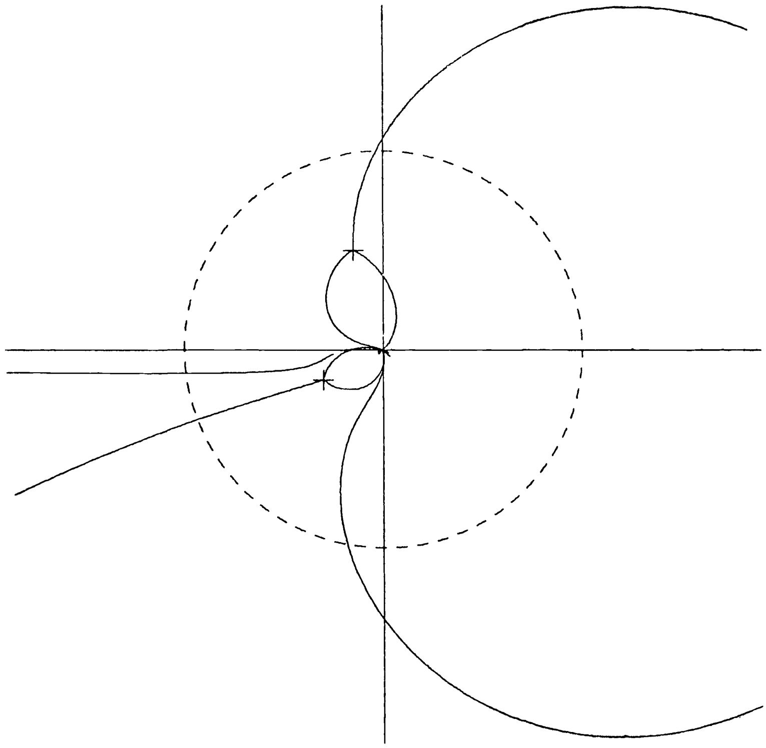

FIG. 19

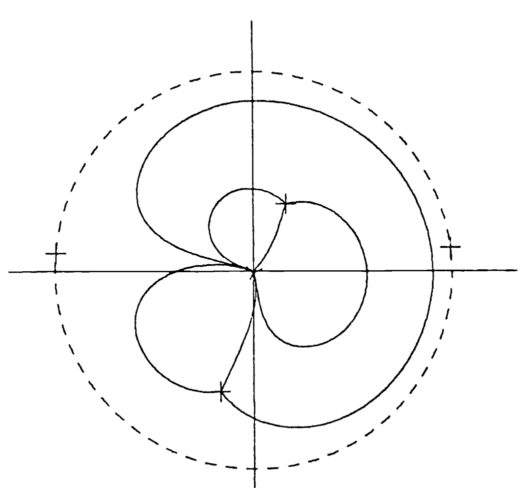

FIG. 20

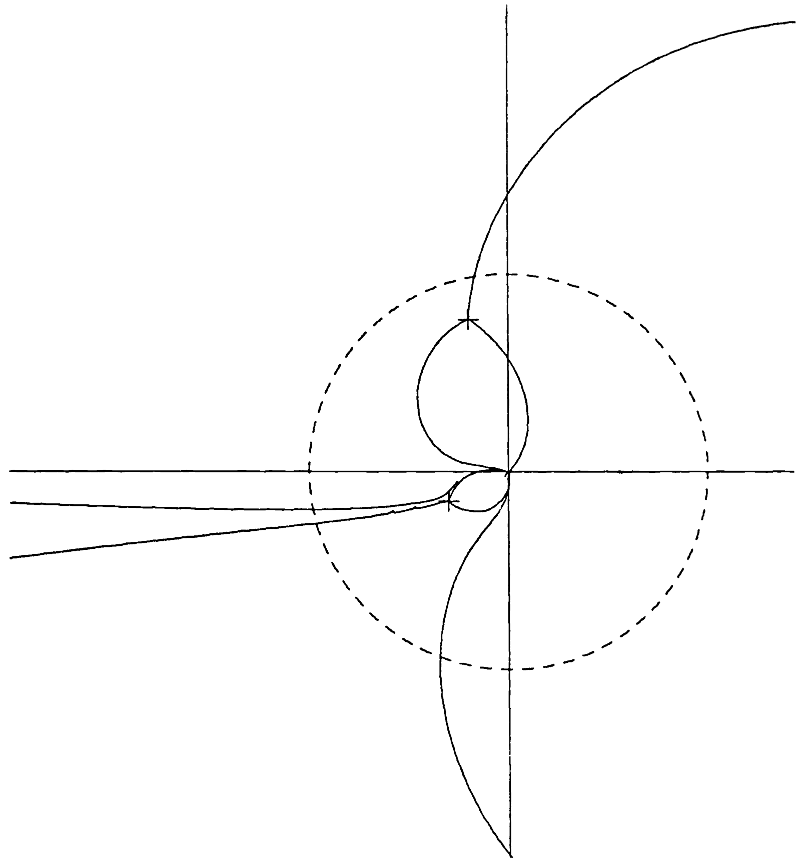

FIG. 21

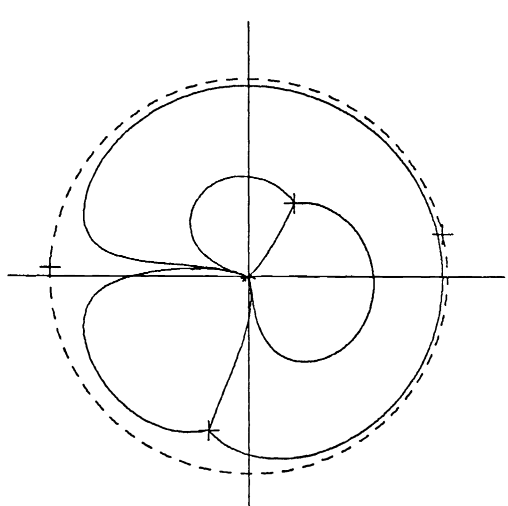

FIG. 22

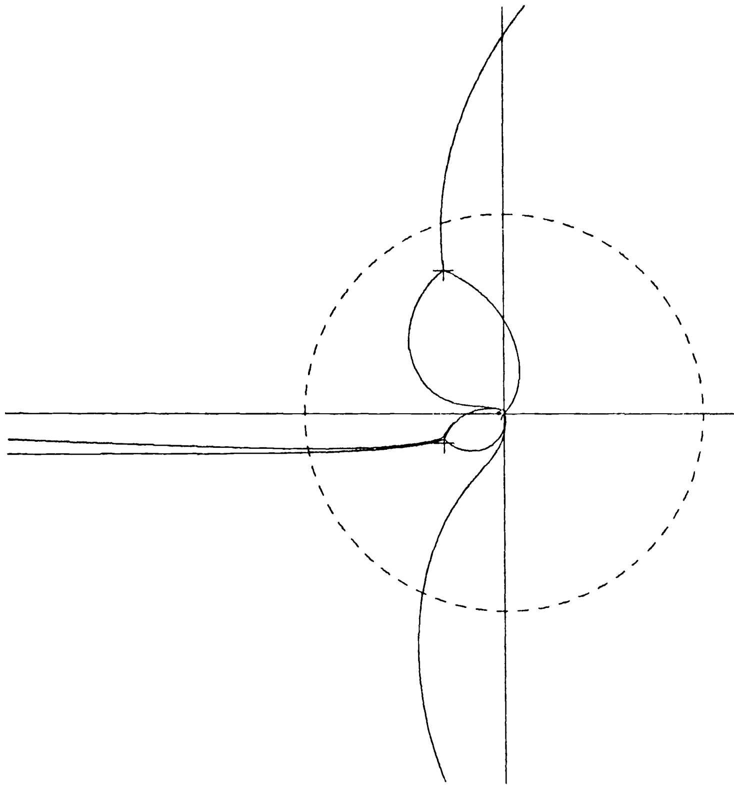

FIG. 23

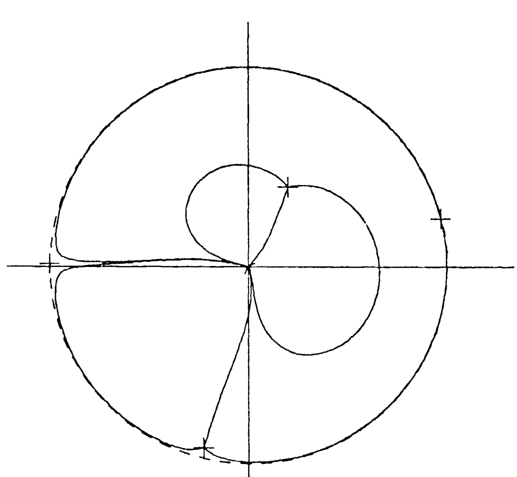

FIG. 24

REFERENCES

[1] L. BITTNER, *Eine Verallgemeinerung des Sekantenverfahrens*, Wiss. Z. Techn. Hochsch. Dresden, 9 (1959), pp. 325–329.

[2] D. GAIER, *Konstruktive Methoden der Konformen Abbildung*, Springer-Verlag, Berlin, 1964.

[3] W. B. CRAGG AND G. W. STEWART, *A Stable Variant of the Secant Method for Solving Nonlinear Equations*, Dept. of Comp. Sciences, Carnegie-Mellon University Publications, Pittsburgh, PA.

[4] E. G. GRASSMANN, *The existence of quadratic differentials in simply connected regions of the complex plane*, Canad. J. Math., 25 (1973), pp. 83–91.

[5] E. G. GRASSMANN AND J. G. ROKNE, *An Explicit Calculation of Some Sets of Minimal Capacity*, this Journal, 6 (1975), pp. 242–249.

[6] ———, *Calculation of Some Extremal Conformal Mappings*, this Journal, to appear.

[7] J. A. JENKINS, *Univalent Functions and Conformal Mapping*, Springer-Verlag, 2nd ed., New York, 1965.

[8] W. E. MILNE, *Numerical Solution of Differential Equations*, John Wiley, New York and London, 1953.

[9] W. C. RHEINBOLDT, *An adaptive continuation process for solving systems of nonlinear equations*, Computer Science Technical Report Series, TR393, University of Maryland, College Park, MD.

[10] A. C. SCHAEFFER AND D. C. SPENCER, *Coefficient Regions for Schlicht Functions*, American Mathematical Society Publications, Providence, RI, 1950.

[11] G. SCHOBER, *Topics in the Theory of Univalent Functions*, Springer-Verlag, New York, 1975.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0001 $01.00/0

ON THE EXISTENCE OF POSITIVE SOLUTIONS FOR A CLASS OF SEMILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS*

K. J. BROWN† AND H. BUDIN†

Abstract. We consider semilinear elliptic boundary value problems of the form

$$Lu = \lambda f(x, u)$$

with Dirichlet boundary conditions. By using variational methods, we show how changes in the sign of $f(x, \cdot)$ lead to multiple positive solutions of the equation for sufficiently large λ. In addition more detailed results are obtained for autonomous ordinary differential equations by using simple quadrature arguments.

1. Introduction. In this paper we discuss the semilinear elliptic boundary value problem

$$\text{(1.1)}\qquad \begin{aligned} &Lu = \lambda f(x, u) \quad \text{for } x \in \Omega;\\ &u(x) = 0 \quad \text{for } x \in \partial\Omega. \end{aligned}$$

We assume that $\Omega \subset \mathbf{R}^n$ is a smooth bounded open set and that

$$(Lu)(x) = -\sum_{i,j=1}^{n} \partial_i(a_{ij}(x)\,\partial_j u(x)) + c(x)u(x)$$

where $-L$ is uniformly elliptic on Ω and the coefficients $a_{ij} \in C^{1+\alpha}(\bar\Omega)$, $c \in C^{\alpha}(\bar\Omega)$ and c is nonnegative.

We show how changes in the sign of $f(x, \cdot)$ lead to multiple positive solutions of (1.1). We assume that $f: \bar\Omega \times \mathbf{R}^+ \to \mathbf{R}$ satisfies the following:

(i) $f \in C^{\alpha}(\bar\Omega \times \mathbf{R}^+)$.

(ii) $f(x, 0) > 0$ for $x \in \bar\Omega$.

(iii) There exists $a_1, a_2, \cdots, a_n \in \mathbf{R}$ such that $0 < a_1 < a_2 < \cdots < a_n$ and $f(x, a_i) \leqq 0$ for $x \in \bar\Omega$ and $i = 1, 2, \cdots, n$.

(iv) If $F(x, u) = \int_0^u f(x, s)\,ds$, there exists $b_1, \cdots, b_{n-1} \in \mathbf{R}$ such that $a_1 < b_1 < a_2 < \cdots < b_{n-1} < a_n$ and $F(x, b_i) > F(x, u)$ for $x \in \bar\Omega$ and $0 \leqq u < b_i$, $i = 1, 2, \cdots, n-1$.

Roughly speaking the above hypotheses imply that, for fixed x, the graph of $f(x, \cdot)$ has n positive humps and n negative humps, each positive hump having greater area than the previous negative hump.

In § 2 we prove our main result

THEOREM 1.2. *If f satisfies (i)–(iv), then, for sufficiently large $\lambda > 0$, (1.1) has at least n nonnegative solutions $\phi_1, \phi_2, \cdots, \phi_n$ such that $\phi_1 \leqq \phi_2 \leqq \cdots \leqq \phi_n$. Moreover, for such λ, if $f \in C^{1+\alpha}(\bar\Omega \times \mathbf{R}^+)$ and*

$$\begin{aligned} &Lv = \lambda f_u(x, \phi_k(x))v \quad \textit{for } x \in \Omega,\\ &v(x) = 0 \quad \textit{for } x \in \partial\Omega \end{aligned}$$

has only the trivial solution for $k = i$ and $k = i + 1$, then (1.1) has another solution lying between ϕ_i and ϕ_{i+1}.

Finally, in § 3, we consider the simple special case where (1.1) is an autonomous ordinary differential equation and f satisfies conditions analogous to (i)–(iv). In this case

* Received by the editors July 8, 1977, and in revised form January 9, 1978.

† Department of Mathematics, Heriot-Watt University, Riccarton, Currie, Edinburgh, Great Britain.

we can obtain results which are somewhat more detailed than Theorem 1.2 by using a generalization of a quadrature technique due to Laetsch [3].

2. Nonautonomous partial differential equations. In this section we prove Theorem 1.2 by using variational and upper and lower solution techniques. We assume that f satisfies (i)–(iv) and that L and Ω are as in § 1.

First we modify the function f. Let

$$f_1(x,u)=\begin{cases} f(x,0) & \text{if } u<0,\\ f(x,u) & \text{if } 0\leqq u\leqq a_1,\\ f(x,a_1) & \text{if } u>a_1.\end{cases}$$

Suppose $\lambda>0$ and (λ, u) is a solution of

$$\begin{aligned} &Lu=\lambda f_1(x,u) \quad \text{for } x\in\Omega,\\ &u(x)=0 \quad \text{for } x\in\partial\Omega. \end{aligned} \tag{2.1}$$

If $A=\{x\in\Omega\colon u(x)>a_1\}$, then $Lu(x)\leqq 0$ for $x\in A$ and $u(x)=a_1$ if $x\in\partial A$. Hence by the maximum principle $u(x)=a_1$ for $x\in A$. Thus $u(x)\leqq a_1$ for $x\in\Omega$. Hence all solutions of (2.1) are also solutions of (1.1).

We now introduce a functional whose critical points are solutions of (2.1). Let E denote the Sobolev space $W_0^{1,2}(\Omega)$. It is well known that, since $-L$ is uniformly elliptic, the inner product

$$(u,v)_E=\sum_{i,j=1}^{n}\int_\Omega (a_{ij}\,\partial_i u\,\partial_j v+cuv)\,dx \quad \text{for } u,v\in E$$

is equivalent to the usual inner product in $W_0^{1,2}(\Omega)$. We denote the corresponding norm by $\|\cdot\|_E$. Consider the functional $I_1(\lambda,\cdot)\colon E\to\mathbf{R}$ such that

$$\begin{aligned} I_1(\lambda,u)&=\tfrac12\|u\|_E^2-\lambda\int_\Omega F_1(x,u(x))\,dx\\ &=\tfrac12\|u\|_E^2-\lambda J_1(u), \end{aligned}$$

where $F_1(x,u)=\int_0^u f_1(x,s)\,ds$.

LEMMA 2.2. *For all $\lambda>0$ there exists $u_1(\lambda)\in E$ such that $I_1(\lambda,\cdot)$ attains its global minimum at $u_1(\lambda)$. Moreover $u_1(\lambda)$ is a classical solution of* (1.1) *and satisfies* $0\leqq u_1(\lambda)\leqq a_1$.

Proof. We first show that $u\to I_1(\lambda,u)$ is weakly lower semicontinuous. Let $\{u_k\}$ be a sequence in E converging weakly to u. It is well known that $\|u\|_E\leqq \liminf_{k\to\infty}\|u_k\|$. By the Sobolev embedding theorems $\{u_k\}$ converges strongly to u in $L^p(\Omega)$ if $p<2n/(n-2)$ and $n\geqq 3$. By the definition of F_1 there exist constants K_1 and K_2 such that $|F_1(x,u)|\leqq K_1+K_2|u|$ for all $(x,u)\in\bar\Omega\times\mathbf{R}$. Hence $\{F_1(\cdot\,,u_k(\cdot))\}$ converges to $F_1(\cdot\,,u(\cdot))$ in $L^p(\Omega)$ and so in $L^1(\Omega)$ i.e. $\lim_{k\to\infty}J_1(u_k)=J_1(u)$. Therefore $I_1(\lambda,u)\leqq\liminf_{k\to\infty}I_1(\lambda,u_k)$ and so I_1 is weakly lower semicontinuous. Similar, simpler, arguments are possible for $n=1,2$.

It follows from a standard theorem in the calculus of variations, see e.g. Vainberg [5], that $I_1(\lambda,\cdot)$ attains its minimum on any weakly closed set. In particular, if $R>0$, $I_1(\lambda,\cdot)$ attains its minimum on B_R where B_R denotes the closed ball center 0 radius R in

E. Moreover

$$\begin{aligned} I_1(\lambda, u) &\geqq \tfrac{1}{2}\|u\|_E^2 - \lambda \int_\Omega |F_1(x, u(x))|\, dx \\ &\geqq \tfrac{1}{2}\|u\|_E^2 - \lambda \int_\Omega (K_1 + K_2|u(x)|)\, dx \\ &\geqq \tfrac{1}{2}\|u\|_E^2 - \lambda (K_3 + K_4\|u\|_E) \end{aligned}$$

for all $u \in E$, where K_3 and K_4 are constants. Hence there exists $R(\lambda) > 0$ such that $I_1(\lambda, u) \geqq 1$ if $\|u\|_E \geqq R(\lambda)$. Thus, as $I_1(\lambda, 0) = 0$, $I_1(\lambda, \cdot)$ attains its global minimum at an interior point $u_1(\lambda)$ of $B_{R(\lambda)}$. It is easy to show that $I_1(\lambda, \cdot)$ is Fréchet differentiable with Fréchet derivative at u given by

$$I_1'(\lambda, u)v = (u, v)_E - \lambda \int_\Omega f_1(x, u(x))v(x)\, dx \tag{2.3}$$

for all $v \in E$. Since $u_1(\lambda)$ is a global minimum for $I_1(\lambda, \cdot)$, $I_1'(\lambda, u_1(\lambda)) = 0$ and so by (2.3)

$$\sum_{i,j=1}^n \int_\Omega a_{ij}(x)\, \partial_i u_1(\lambda)(x)\, \partial_j v(x)\, dx - \lambda \int_\Omega f_1(x, u_1(\lambda)(x))v(x)\, dx = 0$$

for all $v \in E$, i.e. $u_1(\lambda)$ is a weak solution of (2.1). Because of the uniform boundedness of f_1, it can be proved, using bootstrapping arguments, that $u_1(\lambda)$ is also a classical solution of (2.1) and so $0 \leqq u_1(\lambda) \leqq a_1$. This completes the proof.

Remark. It is easy to see that (1.1) has at least one positive solution between 0 and a_1 by observing that $u = 0$ and $u = a_1$ are lower and upper solutions respectively of (1.1). The more complicated variational approach above, however, enables us to investigate the existence of further solutions.

Now let

$$f_2(x, u) = \begin{cases} f(x, 0) & \text{if } u < 0, \\ f(x, u) & \text{if } 0 \leqq u \leqq a_2, \\ f(x, a_2) & \text{if } u > a_2 \end{cases}$$

and let $I_2(\lambda, \cdot)\colon E \to \mathbf{R}$ such that

$$\begin{aligned} I_2(\lambda, u) &= \tfrac{1}{2}\|u\|_E^2 - \lambda \int_\Omega F_2(x, u(x))\, dx \\ &= \tfrac{1}{2}\|u\|_E^2 - \lambda J_2(u) \end{aligned}$$

where $F_2(x, u) = \int_0^u f_2(x, s)\, ds$.

Arguing in exactly the same way as above we can prove

LEMMA 2.4. *For all $\lambda > 0$ there exists $u_2(\lambda) \in E$ such that $I_2(\lambda, \cdot)$ attains its global minimum at $u_2(\lambda)$. Moreover $u_2(\lambda)$ is a classical solution of* (1.1) *and satisfies* $0 \leqq u_2(\lambda) \leqq a_2$.

We now show that if λ is sufficiently large, then $u_1(\lambda) \neq u_2(\lambda)$ provided that $\partial\Omega$ is sufficiently smooth. We shall assume that $\partial\Omega$ is of class $C^{2+\alpha}$, that $x \to \beta(x)$ is of class $C^{1+\alpha}$ on $\partial\Omega$ where $\beta(x)$ denotes the outward unit normal to $\partial\Omega$ at x and that there exists $\varepsilon > 0$ such that for $x \in \partial\Omega$ there exists a ball B of radius ε such that $x \in \partial B$ and $B \subseteq \bar{\Omega}$. In this case there exists $c > 0$ such that, if $\delta > 0$ and the diameter of $\Omega \geqq 2\delta$, then $m(\Omega_\delta) \leqq c\delta$ where $\Omega_\delta = \{x \in \Omega : \text{dist}\,(x, \partial\Omega) < \delta\}$ and m denotes Lebesgue measure on $\mathbf{R}^n$ (see Nussbaum [4]).

LEMMA 2.5. *If Ω satisfies the above assumption and λ is sufficiently large, then* $\sup\{u_2(\lambda)(x): x \in \Omega\} > a_1$ *and so* $u_2(\lambda) \neq u_1(\lambda)$.

Proof. We shall show that there exists $w \in E$ such that $I_2(w, \lambda) < I_2(u, \lambda)$ for all $u \in E$ satisfying $0 \leqq u \leqq a_1$. Let $\alpha = \inf\{F(x, b_1) - F(x, u): x \in \bar{\Omega} \text{ and } 0 \leqq u \leqq a_1\}$. Then $\alpha > 0$ as f satisfies condition (iv). If $u \in E$ satisfies $0 \leqq u \leqq a_1$, then

$$\begin{aligned} J_2(u) &= \int_\Omega F_2(x, u(x))\, dx \\ &= \int_\Omega F(x, u(x))\, dx \qquad (2.6) \\ &\leqq \int_\Omega F(x, b_1)\, dx - \alpha m(\Omega). \end{aligned}$$

Let $\delta > 0$ and let $\phi_\delta: \mathbf{R}^n \to \mathbf{R}$ be a nonnegative C^∞ function such that $\phi_\delta(x) = 0$ if $\|x\| \geqq \delta$ and $\int_{\mathbf{R}^n} \phi_\delta(x)\, dx = 1$. Let $h: \mathbf{R}^n \to \mathbf{R}$ such that $h(x) = b_1$ if $x \in \Omega - \Omega_{2\delta}$ and $h(x) = 0$ otherwise. Let $w_\delta = \phi_\delta * h$ i.e. $w_\delta(x) = \int_{\mathbf{R}^n} \phi_\delta(x-y) h(y)\, dy$. w_δ is usually called a mollification of h. It follows from the standard theory of mollifiers (see e.g. Adams [1]) that w_δ is a C^∞ function with support contained in $\Omega - \Omega_\delta$ and so $w_\delta \in E$. Moreover $0 \leqq w_\delta(x) \leqq b_1$ for all $x \in \mathbf{R}^n$ and $w_\delta(x) = b_1$ for $x \in \Omega - \Omega_{3\delta}$. Hence

$$\begin{aligned} J_2(w_\delta) &= \int_\Omega F_2(x, w_\delta(x))\, dx \\ &= \int_{\Omega - \Omega_{3\delta}} F(x, b_1)\, dx + \int_{\Omega_{3\delta}} F(x, w_\delta(x))\, dx \\ &= \int_\Omega F(x, b_1)\, dx - \int_{\Omega_{3\delta}} F(x, b_1)\, dx + \int_{\Omega_{3\delta}} F(x, w_\delta(x))\, dx \qquad (2.7) \\ &\geqq \int_\Omega F(x, b_1)\, dx - 2Km(\Omega_{3\delta}) \\ &\geqq \int_\Omega F(x, b_1)\, dx - 6Kc\delta \end{aligned}$$

where $K = \sup\{|F(x, u)|: x \in \Omega \text{ and } 0 \leqq u \leqq b_1\}$.

By (2.6) and (2.7) we can choose and fix δ sufficiently small so that there exists $\eta > 0$ such that $w = w_\delta$ satisfies

$$J_2(w) > J_2(u) + \eta \qquad (2.8)$$

for all $u \in E$ satisfying $0 \leqq u \leqq a_1$. Therefore for all such u

$$\begin{aligned} I_2(\lambda, w) - I_2(\lambda, u) &= \tfrac{1}{2}\|w\|_E^2 - \tfrac{1}{2}\|u\|_E^2 - \lambda(J_2(w) - J_2(u)) \\ &\leqq \tfrac{1}{2}\|w\|_E^2 - \lambda\eta \\ &< 0 \quad \text{for } \lambda \text{ sufficiently large.} \end{aligned}$$

Hence for such λ the global minimum of I_2 cannot be attained at any $u \in E$ such that $0 \leqq u \leqq a_1$. Therefore $\sup\{u_2(\lambda): x \in \Omega\} > a_1$ and so $u_2(\lambda) \neq u_1(\lambda)$. This completes the proof of the theorem.

By letting

$$f_k(x, u) = \begin{cases} f(x, 0) & \text{if } u < 0, \\ f(x, u) & \text{if } 0 \leqq u \leqq a_k, \\ f(x, a_k) & \text{if } u > a_k \end{cases}$$

for $k = 1, 2, \cdots, n$ and arguing as in Lemmas 2.4 and 2.5 we obtain

LEMMA 2.9. *If Ω is as in Lemma 2.5 and λ is sufficiently large, then there exist n distinct nonnegative classical solutions $u_1(\lambda), u_2(\lambda), \cdots, u_n(\lambda)$ of (1.1) such that*

$$a_{i-1} < \sup \{u_i(\lambda)(x) : x \in \Omega\} \leqq a_i$$

for $i = 1, \cdots, n$ and where $a_0 = 0$.

We complete the proof of Theorem 1.2 by using arguments based on upper and lower solutions. The definitions and results we use can be found in Amann [2].

First we produce solutions to (1.1) which are the limits of iterations starting from upper solutions.

LEMMA 2.10. *Suppose Ω and λ are as in Lemma 2.9. Then there exist n distinct nonnegative classical solutions $\phi_1(\lambda), \phi_2(\lambda), \cdots, \phi_n(\lambda)$ of (1.1) such that*

(a) *$\phi_i(\lambda)$ is the maximal nonnegative solution of (1.1) satisfying $u \leqq a_i$ i.e. if u is a solution of (1.1) with $u \leqq a_i$ then $u \leqq \phi_i(\lambda)$;*

(b) *$a_{i-1} < \sup \{\phi_i(\lambda)(x) : x \in \Omega\} \leqq a_i$.*

Proof. It is easy to see that $u = a_i$ is an upper solution of (1.1) and is greater than or equal to the (lower) solution $u_i(\lambda)$ for all $x \in \Omega$. Hence by Amann [2] the sequence defined by the following iteration scheme

$$w_0 = a_i; \qquad Lw_{n+1} = \lambda f(x, w_n) \quad \text{for } x \in \Omega;$$
$$w_{n+1}(x) = 0 \quad \text{for } x \in \partial\Omega,$$

for $n = 0, 1, 2, \cdots$ satisfies

$$a_i \geqq w_1 \geqq w_2 \geqq \cdots \geqq w_n \geqq u_i(\lambda).$$

Moreover $\lim_{n\to\infty} w_n(x) = \phi_i(\lambda)(x)$ for $x \in \Omega$ where $\phi_i(\lambda)$ is the maximal solution of (1.1) satisfying $u \leqq a_i$. Hence $\phi_i(\lambda) \geqq u_i(\lambda)$ and this completes the proof.

The next lemma produces lower solutions close to the solutions $\phi_i(\lambda)$.

LEMMA 2.11. *Suppose f, Ω and λ are as above and in addition $f \in C^{1+\alpha}(\bar{\Omega} \times \mathbf{R}^+)$. If*

$$(2.12) \qquad \begin{aligned} &Lv = \lambda f_u(x, \phi_i(\lambda)(x))v \quad \text{for } x \in \Omega, \\ &v(x) = 0 \quad \text{for } x \in \partial\Omega \end{aligned}$$

has only the trivial solution, then there exists $\varepsilon > 0$ such that

(1) *$\phi_i(\lambda)$ is the only solution of (1.1) in the ball $B_\varepsilon(\phi_i(\lambda))$ radius ε and center $\phi_i(\lambda)$ in $C^{2+\alpha}(\bar{\Omega})$;*

(2) *there exists a lower solution $\psi_i(\lambda) \neq \phi_i(\lambda)$ of (1.1) in $B_\varepsilon(\phi_i(\lambda))$;*

(3) *the sequence defined by*

$$u_0(x) = \psi_i(\lambda); \qquad Lu_{n+1}(x) = \lambda f(x, u_n(x)) \quad \text{for } x \in \Omega;$$
$$u_{n+1}(x) = 0 \quad \text{for } x \in \partial\Omega,$$

for $n = 0, 1, 2, \cdots$ satisfies

$$(2.13) \qquad \psi_i(\lambda) \leqq u_1 \leqq u_2 \leqq \cdots \leqq u_n \leqq a_i$$

and $\lim_{n\to\infty} u_n(x) = \phi_i(\lambda)(x)$ for $x \in \Omega$.

Proof. Let $D=\{u\in C^{2+\alpha}(\bar{\Omega}): u(x)=0 \text{ for } x\in\partial\Omega\}$. Then D is a Banach space with respect to the usual $C^{2+\alpha}(\bar{\Omega})$ norm. Since (2.12) has only the trivial solution, it follows that $G: D\to C^{\alpha}(\bar{\Omega})$ where $(Gu)(x)=Lu(x)-\lambda f_u(x,\phi_i(\lambda)(x))u(x)$ is a linear homeomorphism. If $H: D\to C^{\alpha}(\bar{\Omega})$ where $(Hu)(x)=Lu(x)-\lambda f(x,u(x))$, the Fréchet derivative of H at $\phi_i(\lambda)$ is G and so (1) follows from the inverse function theorem.

Define $K: D\times\mathbf{R}\to C^{\alpha}(\bar{\Omega})$ by $[K(u,\mu)](x)=Lu(x)-\lambda f(x,u)+\mu$. Then the Fréchet derivative of $u\to K(u,\mu)$ at $(u,\mu)=(\phi_i(\lambda),0)$ is G and so by the implicit function theorem there exists $\delta>0$ and a continuous function $\theta: (-\delta,\delta)\to D$ such that $\theta(0)=\phi_i(\lambda)$ and $H(\theta(\mu),\mu)=0$ for $\mu\in(-\delta,\delta)$. If $\mu>0$, then

$$L\theta(\mu)=\lambda f(x,\theta(\mu))-\mu<\lambda f(x,\theta(\mu))$$

i.e. $\theta(\mu)$ is a lower solution of (1.1).

Since a_i and $\theta(\mu)$ are upper and lower solutions respectively of (1.1) with $\theta(\mu)\leqq a_i$ for small μ, the iteration $\{u_n\}$ starting from $\theta(\mu)$ satisfies (2.13) and $\lim_{n\to\infty} u_n(x)=U(x)$ where U is a solution of (1.1). We now show that, provided $\mu>0$ is sufficiently small, $U=\phi_i(\lambda)$. Since $U\leqq a_i$ and $\phi_i(\lambda)$ is maximal, $U\leqq\phi_i(\lambda)$. Hence, if $\|\theta(\mu)-\phi_i(\lambda)\|=r$ in $C(\bar{\Omega})$, then $\|\phi_i(\lambda)-U\|\leqq r$ in $C(\bar{\Omega})$. Therefore there exists a constant K_1 independent of r such that $\|f(\cdot\,,\phi_i(\lambda))-f(\cdot\,,U)\|\leqq K_1 r$ in $L_p(\Omega)$. Since

$$(2.14)\qquad \begin{aligned} &L(\phi_i(\lambda)-U)=\lambda[f(x,\phi_i(\lambda))-f(x,U)] \quad \text{for } x\in\Omega,\\ &\phi_i(\lambda)(x)-U(x)=0 \quad \text{for } x\in\partial\Omega\end{aligned}$$

by the Agmon–Douglis–Nirenberg L_p estimates there exists a constant K_2 such that $\|\phi_i(\lambda)-U\|\leqq K_2 r$ in the Sobolev space $W_{2,p}(\Omega)$. Hence, if we choose $p>n$, it follows by an embedding theorem that there exists a constant K_3 such that $\|\phi_i(\lambda)-U\|\leqq K_3 r$ in $C^{\alpha}(\bar{\Omega})$. It follows that there exists a constant K_4 such that $\|f(\cdot\,,\phi_i(\lambda))-f(\cdot\,,U)\|<K_4 r$ in $C^{\alpha}(\bar{\Omega})$. Finally, by (2.14) and the Schauder estimates there exists a constant K_5 such that $\|\phi_i(\lambda)-U\|<K_5 r$ in $C^{2+\alpha}(\bar{\Omega})$. Choose and fix $\mu=\mu_0$ so small that $K_5 r<\varepsilon$ and let $\psi_i(\lambda)=\theta(\mu_0)$. Then, if U is the limit of the iteration starting from $\psi_i(\lambda)$, $U\in B_\varepsilon(\phi_i(\lambda))$ and U is a solution of (1.1) and so, by (1), $U=\phi_i(\lambda)$. This completes the proof of the lemma.

It is now easy to complete the proof of Theorem 1.2. Fix i, $1\leqq i\leqq n-1$. Since a_{i+1} and $\psi_i(\lambda)$ are upper and lower solutions respectively of (1.1), the sequences defined by the following iteration schemes

$$\begin{aligned} u_0=\psi_i(\lambda);\qquad & Lu_{n+1}=\lambda f(x,u_n) \quad \text{for } x\in\Omega,\\ & u_{n+1}(x)=0 \quad \text{for } x\in\partial\Omega,\\ v_0=a_{i+1};\qquad & Lv_{n+1}=\lambda f(x,v_n) \quad \text{for } x\in\Omega,\\ & v_{n+1}(x)=0 \quad \text{for } x\in\partial\Omega\end{aligned}$$

for $n=0,1,2,\cdots$ satisfy

$$\psi_i(\lambda)\leqq u_1\leqq\cdots\leqq u_n\leqq v_n\leqq v_1\leqq a_{i+1}.$$

Moreover $\lim_{n\to\infty} u_n(x)=\phi_i(\lambda)$ and $\lim_{n\to\infty} v_n(x)=\phi_{i+1}(\lambda)$. Hence Amann [2, Thm. 1] asserts the existence of another solution of (1.1) lying between $\phi_i(\lambda)$ and $\phi_{i+1}(\lambda)$.

3. Autonomous ordinary differential equations. We now study

$$(3.1)\qquad \begin{aligned} &-u''(t)=\lambda f(u(t)) \quad \text{for } t\in[0,1],\\ &u(0)=0=u(1),\end{aligned}$$

where $\lambda > 0$ and $f: \mathbf{R}^+ \to \mathbf{R}$ satisfies the following:

(a) f has continuous derivative.

(b) $f(0) > 0$.

(c) There exists $a_1, a_2, \cdots, a_n \in \mathbf{R}$ *such that* $0 < a_1 < a_2 < \cdots < a_n$ and $f(a_i) \leqq 0$ for $i = 1, 2, \cdots, n$.

(d) If $F(u) = \int_0^u f(s)\,ds$, there exist $b_1, \cdots, b_{n-1} \in \mathbf{R}$ with $a_1 < b_1 < a_2 < \cdots < b_{n-1} < a_n$ such that $f(b_i) > 0$ and $F(b_i) > F(u)$ for $0 \leqq u < b_i$, $i = 1, 2, \cdots, n-1$.

If we extend f so that $f(u) > 0$ for all $u < 0$, then all solutions of (3.1) are positive on $(0, 1)$. Moreover, because (3.1) is autonomous, any positive solution attains its maximum at $t = \frac{1}{2}$ and is symmetric with respect to $t = \frac{1}{2}$. Hence u is a solution of (3.1) if and only if u is a solution of

$$-u''(t) = \lambda f(u(t)) \quad \text{for } t \in (0, \tfrac{1}{2}); \qquad u(0) = 0 = u'(\tfrac{1}{2}). \tag{3.2}$$

Also, if u satisfies (3.2), $u(\frac{1}{2}) = \sup\{u(t): t \in [0, 1]\} = \|u\|$.

It is easy to see that if u satisfies (3.2) then

$$\tfrac{1}{2}[u'(t)]^2 + \lambda F(u(t)) = \lambda F(\|u\|) \quad \text{for } t \in [0, \tfrac{1}{2}]. \tag{3.3}$$

Hence $\|u\|$ is such that $F(\|u\|) > F(u)$ for $0 \leqq u < \|u\|$. We shall prove a converse of this result which will enable us to discuss the multiplicity of solutions of (3.1). Let $S = \{u > 0: f(u) > 0 \text{ and } F(u) > F(s) \text{ for all } s,\ 0 \leqq s < u\}$.

THEOREM 3.4. *If $\rho \in S$, there exists a unique $\lambda > 0$ such that there is a solution (λ, u) of* (3.1) *satisfying $\|u\| = \rho$. Moreover $\rho \to \lambda(\rho)$ is a continuous function on S.*

Proof. As the proof is very similar to that in Laetsch [3, Thm. 2.1] we shall merely sketch it briefly.

Suppose that (λ, u) is a solution of (3.1) with $\|u\| = \rho$. Then by (3.3)

$$[u'(t)]^2 = 2\lambda(F(\rho) - F(u(t))) \quad \text{for } t \in [0, \tfrac{1}{2}]$$

and so

$$t(2\lambda)^{1/2} = \int_0^{u(t)} (F(\rho) - F(s))^{-1/2}\,ds \quad \text{for } t \in [0, \tfrac{1}{2}]. \tag{3.5}$$

Putting $t = \frac{1}{2}$ we obtain

$$\lambda^{1/2} = 2^{1/2} \int_0^{\rho} (F(\rho) - F(s))^{-1/2}\,ds. \tag{3.6}$$

Hence λ (if it exists) is uniquely determined by ρ.

If, given $\rho \in S$, we define $\lambda(\rho)$ by (3.6) and $u(t)$ by (3.5), it is straightforward to verify that u is twice differentiable, u satisfies (3.2) and $u(\frac{1}{2}) = \rho$. The continuity of $\lambda(\cdot)$ is implied by (3.6) and this completes the proof.

If $r = \inf\{u > 0: f(u) = 0\}$, clearly $(0, r) \subseteq S$. It is shown in Laetsch [3, Thms. 2.6 and 2.9] that $\lim_{\rho \to 0} \lambda(\rho) = 0$ and $\lim_{\rho \to r^-} \lambda(\rho) = \infty$. We now derive some further asymptotic properties of $\lambda(\cdot)$. Let $\beta_i = \inf\{u > b_i: f(u) = 0\}$ and $\alpha_i = \inf\{u: (u, \beta_i) \subseteq S\}$. Then $a_i \leqq \alpha_i < b_i < \beta_i \leqq a_{i+1}$ and $(\alpha_i, \beta_i) \subseteq S$ for $i = 1, 2, \cdots, n-1$. Moreover we have

THEOREM 3.7.

$$\text{(i)} \quad \lim_{\rho \to \alpha_i^+} \lambda(\rho) = \infty; \qquad \text{(ii)} \quad \lim_{\rho \to \beta_i^-} \lambda(\rho) = \infty.$$

Proof. (i) Suppose firstly that $f(\alpha_i)>0$. Since S is open, $\alpha_i \notin S$ and so there exists k, $0<k<\alpha_i$ such that $F(\alpha_i)=F(k)$. Clearly k must be a local maximum for F and so $f(k)=0$. Hence, if $M=\max\{|f'(u)|:0\leqq u\leqq b_i\}$, then $|f(u)|\leqq M|u-k|$ for $0\leqq u\leqq b_i$. Let $N=\max\{|f(u)|:0\leqq u\leqq b_i\}$. Then, if $\alpha_i<\rho<b_i$ and $0\leqq u<\rho$,

$$F(\rho)-F(u)=F(\rho)-F(\alpha_i)+F(k)-F(u)$$

$$=(\rho-\alpha_i)f(\xi)+(k-u)f(\eta) \quad \text{where } \xi\in(\alpha_i,\rho)$$

and $\eta\in(k,u)$

$$\leqq N(\rho-\alpha_i)+M(k-u)^2.$$

Hence

$$[\lambda(\rho)]^{1/2}=2^{1/2}\int_0^\rho (F(\rho)-F(u))^{-1/2}\,du$$

$$\geqq 2^{1/2}\int_0^{\alpha_i}(N|\rho-\alpha_i|+M|k-u|^2)^{-1/2}\,du$$

$$=\int_0^{\alpha_i} H_\rho(u)\,du.$$

As $\rho\to\alpha_i^+$, H_ρ is a nondecreasing sequence of measurable functions. Hence by the monotone convergence theorem,

$$\lim_{\rho\to\alpha_i^+}[\lambda(\rho)]^{1/2}=\lim_{\rho\to\alpha_i^+}\int_0^{\alpha_i}H_\rho(u)\,du=\lim_{\rho\to\alpha_i^+}\int_0^{\alpha_i}2^{1/2}M^{-1/2}|k-u|^{-1}\,du=\infty.$$

Suppose now that $f(\alpha_i)=0$ i.e. $F'(\alpha_i)=0$. It is easy to show that $\int_0^{\alpha_i}(F(\alpha_i)-F(u))^{-1/2}\,du=\infty$. Now

$$\lim_{\rho\to\alpha_i^+}[\lambda(\rho)]^{1/2}\geqq\lim_{\rho\to\alpha_i^+}\int_0^{\alpha_i}(F(\rho)-F(u))^{-1/2}\,du$$

$$=\int_0^{\alpha_i}(F(\alpha_i)-F(u))^{-1/2}\,du, \qquad \text{by the monotone convergence theorem}$$

$$=\infty.$$

(ii) Let $K_1=\max\{|f(u)|:0\leqq u\leqq\beta_i\}$ and $K_2=\max\{|f'(u)|:0\leqq u\leqq\beta_i\}$. Since $f(\beta_i)=0$,

$$|f(u)|\leqq K_2|u-\beta_i| \quad \text{if } 0\leqq u<\beta_i.$$

Hence, if $0\leqq u\leqq\rho<\beta_i$,

$$F(\rho)-F(u)=F(\rho)-F(\beta_i)+F(\beta_i)-F(u)$$

$$=(\beta_i-\rho)f(\xi)+(\beta_i-u)f(\eta)$$

$$\text{where } \xi\in(\rho,\beta_i) \text{ and } \eta\in(u,\beta_i)$$

$$\leqq K_1(\beta_i-\rho)+K_2(\beta_i-u)^2.$$

Hence, if $0<\rho<\beta_i$,

$$[\lambda(\rho)]^{1/2}\geqq 2^{1/2}\int_0^\rho (K_1(\beta_i-\rho)+K_2(\beta_i-u)^2)^{-1/2}\,du$$

$$=\int_0^{\beta_i}G_\rho(u)\,du$$

where $G_\rho(u)=2^{1/2}(K_1|\beta_i-\rho|+K_2|\beta_i-u|^2)^{-1/2}\chi_{[0,\rho]}$ and $\chi_{[0,\rho]}$ denotes the characteristic function of $[0,\rho]$. As $\{G_\rho\}$ is a nondecreasing sequence of measurable functions, by the monotone convergence theorem

$$\lim_{\rho\to\beta_i^-}[\lambda(\rho)]^{1/2}\geqq\lim_{\rho\to\beta_i^-}\int_0^{\beta_i}G_\rho(u)\,du$$
$$=\int_0^{\beta_i}2^{1/2}K_2^{-1/2}|\beta_i-u|^{-1}\,du=\infty.$$

We can now prove the following result on the multiplicity of solutions of (3.1).

THEOREM 3.8 (a). *For all $\lambda>0$ there exists a solution (λ,u) of* (3.1) *such that $\|u\|<r$.*

(b) *If $\lambda>\inf\{\lambda(\rho):\rho\in(\alpha_i,\beta_i)\}$, there exist at least two solutions (λ,u) of* (3.1) *such that $\alpha_i<\|u\|<\beta_i$ for $i=1,2,\cdots,n-1$.*

(c) *If (λ,u) is any solution of* (3.1) *such that $\alpha_i<\|u\|<\beta_i$, then $\lambda>4\alpha_ik^{-1}$ where $k=\sup\{|f(u)|:0\leqq u\leqq\beta_i\}$.*

Proof. (a) follows from the continuity of $\rho\to\lambda(\rho)$ on $(0,r)$ and the facts that $\lim_{\rho\to0}\lambda(\rho)=0$ and $\lim_{\rho\to r^-}\lambda(\rho)=\infty$.

(b) follows from the continuity of $\rho\to\lambda(\rho)$ on (α_i,β_i) and the facts that $\lim_{\rho\to\alpha_i^+}\lambda(\rho)=\lim_{\rho\to\beta_i^-}\lambda(\rho)=\infty$.

(c) (λ,u) satisfies (3.1) if and only if u satisfies the integral equation

$$u(x)=\lambda\int_0^1G(x,y)f(u(y))\,dy$$

where G is the Green's function corresponding to $-u''$ and zero boundary conditions, i.e.

$$G(x,y)=\begin{cases}x(1-y)&\text{for }0\leqq x\leqq y\leqq1,\\(1-x)y&\text{for }0\leqq y\leqq x\leqq1.\end{cases}$$

Hence

$$|u(x)|\leqq\lambda\int_0^1G(x,y)|f(u(y))|\,dy$$
$$\leqq\lambda\sup_{x,y\in[0,1]}G(x,y)\sup_{y\in[0,1]}|f(u(y))|.$$

Hence, if $\alpha_i<\|u\|<\beta_i$, $\alpha_i<\frac14\lambda\sup\{|f(u)|:0\leqq u\leqq\beta_i\}$ and so $\lambda>4\alpha_ik^{-1}$.

REFERENCES

[1] R. A. ADAMS, *Sobolev Spaces*, Academic Press, New York, 1975.
[2] H. AMANN, *Existence of multiple solutions for nonlinear boundary value problems*, Indiana Univ. Math. J., 21 (1972), pp. 925–935.
[3] T. W. LAETSCH, *The number of solutions of a nonlinear two point boundary value problem*, Ibid., 20 (1970), pp. 1–13.
[4] R. NUSSBAUM, *Positive solutions of nonlinear elliptic boundary value problems*, J. Math. Anal. Appl., 51 (1975), pp. 461–482.
[5] M. M. VAINBERG, *Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations*, John Wiley, New York, 1973.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0002 $1.00/0

SPECIAL FUNCTIONS, STIELTJES TRANSFORMS AND INFINITE DIVISIBILITY*

MOURAD E. H. ISMAIL† AND DOUGLAS H. KELKER‡

Abstract. We establish the complete monotonicity of several quotients of Whittaker (Tricomi) functions and of parabolic cylinder functions. These results are used to show that the F distribution of any positive degrees of freedom (including fractional) is infinitely divisible and self-decomposable. We also prove the infinite divisibility of several related distributions, including the square of a gamma variable. We also prove that $x^{(\nu-\mu)/2}I_\mu(\sqrt{x})/I_\nu(\sqrt{x})$ is a completely monotonic function of x when $\mu>\nu>-1$. This result and the complete monotonicity of $x^{(\nu-\mu)/2}K_\nu(\sqrt{x})/K_\mu(\sqrt{x})$, $\mu>\nu>-1$, are used to introduce two new continuous infinitely divisible probability distributions. The limiting cases contain the reciprocal of a gamma distribution and a distribution whose probability density function is a "generalized" theta function. The first distribution is used as a mixing distribution to introduce a new, two parameter, symmetric, infinitely divisible probability distribution on the real line, which contains the Student t distribution as a limiting case. We also establish the complete monotonicity of $K_\nu(b\sqrt{x})/K_\nu(a\sqrt{x})$ and $I_\nu(a\sqrt{x})/I_\nu(b\sqrt{x})$ for $b>a>0$ and $\nu>-1$. We also obtain some results on the zeros of combinations of modified Bessel functions.

1. Introduction and main results. A probability distribution μ on the half line $(0, \infty)$ is infinitely divisible if for every n, $n=1,2,3,\cdots$, there exists a probability distribution μ_n, on $(0,\infty)$, such that

$$\int_0^\infty e^{-xt}\,d\mu=\left\{\int_0^\infty e^{-xt}\,d\mu_n\right\}^n, \qquad n=1,2,\cdots.$$

A function f, defined and having continuous derivatives of all orders for $x\in(0,\infty)$, is called completely monotonic if $(-1)^n f^{(n)}(x)\geqq 0$, $x\in(0,\infty)$. The classes of completely monotonic functions and infinitely divisible distributions are related by, see Feller [12, p. 425],

THEOREM 1.1. *The function $w(x)$ is the Laplace transform of an infinitely divisible distribution if and only if $w(x)=e^{-h(x)}$, where $h(0^+)=1$ and $h'(x)$ is completely monotonic.*

Because of Bernstein's theorem the usual technique of proving the complete monotonicity of a function is by showing the positivity of its inverse Laplace transform. This is cumbersome sometimes and the Stieltjes transform is often easier to work with, Ismail [19], [20]. The Stieltjes transform is, at least formally, a two-fold Laplace transform

$$(\mathscr{G}\,d\mu)(z)=\int_0^\infty \frac{d\mu(t)}{z+t}=\int_0^\infty e^{-zx}\int_0^\infty e^{-xt}\,d\mu(t)\,dx.$$

A Stieltjes transform of a positive measure $d\mu$ is the Laplace transform of an infinitely divisible density if $\mathscr{G}(d\mu)(0)=1$, since $\int_0^\infty e^{-xt}\,d\mu(t)$ is obviously completely monotonic and all completely monotonic densities are infinitely divisible, as shown in Goldie [13].

The representation and inversion theorems for the Stieltjes transform are

THEOREM 1.2 (*The Representation Theorem*). *If*

(i) *$F(z)$ is analytic for $|\arg z|<\pi/\alpha$ for some α, $0<\alpha<1$,*

(ii) *$F(z)=o(1)$ as $z\to\infty$ and $F(z)=o(|z|^{-1})$ as $z\to 0$, uniformly in every sector $|\arg z|\leqq\pi/\alpha'$, $\alpha'>\alpha$,*

* Received by the editors May 6, 1977, and in revised form March 20, 1978.

† Department of Applied Mathematics, McMaster University, Hamilton, Ontario, Canada 68S 4K1.
‡ Department of Mathematics, University of Alberta, Edmonton, Canada T6G 2G1.

then

$$F(x)=\frac{1}{\pi}\int_{0^+}^{\infty}\frac{dt}{x+t}\frac{1}{2\pi i}\int_C \frac{zF(te^z)e^{z/2}}{\pi^2+z^2}\,dz, \qquad x\in(0,\infty), \tag{1.1}$$

where C is a rectifiable closed curve going around $[-i\pi, i\pi]$ in the positive direction and lying in the strip $|\mathrm{Im}\, z|<\pi/\alpha$.

THEOREM 1.3 (*The Inversion Theorem*). *If $F(z)=\int_0^\infty (z+t)^{-1}\,d\mu(t)$ then*

$$\mu(t_2)-\mu(t_1)=\lim_{\eta\to 0^+}(2\pi i)^{-1}\int_{t_1}^{t_2}\{F(-t-i\eta)-F(-t+i\eta)\}\,dt,$$

where $\mu(t)$ is normalized by $\mu(0)=\mu(0^+)=0$ and $\mu(t)=\frac{1}{2}\{\mu(t^+)+\mu(t^-)\}$ for $t>0$.

For a proof of Theorem 1.2 see Hirschman and Widder [18, pp. 210 and 235]. A proof of Theorem 1.3 can be found in Stone [31].

Our first set of results is on the infinite divisibility of the quotient of two gamma variables and related distributions. In particular this establishes the infinite divisibility of the F distribution. In doing so we arrived at new entries for the Stieltjes transform. Our results in this direction are the following.

THEOREM 1.4. *We have the following integral representations, for $|\arg z|<\pi$*

$$z^{-1/2}D_{-\nu-1}(\sqrt{z})/D_{-\nu}(\sqrt{z})=\{\sqrt{2\pi}\,\Gamma(1+\nu)\}^{-1}\int_0^\infty \frac{|D_{-\nu}(i\sqrt{t})|^{-2}}{(z+t)\sqrt{t}}\,dt, \qquad \nu>0, \tag{1.2}$$

$$\begin{aligned}&z^{-1}-z^{-3/2}D_{-\nu-1}(z^{-1/2})/D_{-\nu}(z^{-1/2})\\ &\qquad=\{\sqrt{2\pi}\,\Gamma(1+\nu)\}^{-1}\int_0^\infty \frac{|D_{-\nu}(it^{-1/2})|^{-2}}{(z+t)t^{3/2}}\,dt, \qquad \nu>0,\end{aligned} \tag{1.3}$$

$$\frac{\psi(a+1,c+1,z)}{\psi(a,c,z)}=\int_0^\infty \frac{t^{-c}e^{-t}|\psi(a,c,te^{i\pi})|^{-2}}{(z+t)\Gamma(a+1)\Gamma(a-c+1)}\,dt, \qquad a>0,\quad c<1, \tag{1.4}$$

and

$$\begin{aligned}\frac{1}{z}\frac{\psi(a,c-1,z)}{\psi(a,c,z)}&=\{\Gamma(a)\Gamma(a-c+2)\}^{-1}\int_0^\infty \frac{t^{-c}e^{-t}}{z+t}|\psi(a,c,te^{i\pi})|^{-2}\,dt,\\ &\qquad a>0,\quad 1<c<a+1.\end{aligned} \tag{1.5}$$

THEOREM 1.5. *The distribution of the quotient of two gamma random variables is self-decomposable, hence is infinitely divisible.*

THEOREM 1.6. *The noncentral chi-square distribution is infinitely divisible for all degrees of freedom (including fractional degrees of freedom).*

THEOREM 1.7. *The square of a gamma variable is infinitely divisible.*

Theorems 1.4–1.7 will be proved in § 3. Theorem 1.5 was stated as an open problem in Steutel [30]. Our second set of results is to introduce some new infinitely divisible distributions and point out a few more entries for the Stieltjes transform. These results are as follows.

THEOREM 1.8. *Let $-1<\nu<\mu$. The function $x^{(\nu-\mu)/2}K_\nu(\sqrt{x})/K_\mu(\sqrt{x})$ is a completely monotonic function of x with*

$$\left\{\frac{x^{\nu/2}K_\nu(\sqrt{x})}{2^{\nu-1}\Gamma(\nu)}\right\}\Big/\left\{\frac{x^{\mu/2}K_\mu(\sqrt{x})}{2^{\mu-1}\Gamma(\mu)}\right\}=\int_0^\infty e^{-xt}\rho_1(t,\nu,\mu)\,dt, \tag{1.6}$$

where $\rho_1(t, \nu, \mu)$ is a probability density function on $(0, \infty)$ of an infinitely divisible distribution.

THEOREM 1.9. *Let $-1<\nu<\mu$. The function $x^{(\nu-\mu)/2}I_\mu(\sqrt{x})/I_\nu(\sqrt{x})$ is a completely monotonic function of x with*

$$\left\{\frac{x^{\nu/2}}{2^\nu\Gamma(\nu+1)I_\nu(\sqrt{x})}\right\}\Big/\left\{\frac{x^{\mu/2}}{2^\mu\Gamma(\mu+1)I_\mu(\sqrt{x})}\right\}=\int_0^\infty e^{-xt}\rho_2(t,\nu,\mu)\,dt, \tag{1.7}$$

where $\rho_2(t, \nu, \mu)$ is a probability density function, on $(0, \infty)$, of an infinitely divisible distribution.

The limiting case $\mu\to\infty$ of Theorem 1.9 is of interest, as we shall see in § 6, as is the following

THEOREM 1.10. *For $\nu>0$, we have*

$$x^{\nu/2}2^{-\nu}\{I_\nu(\sqrt{x})\Gamma(\nu+1)\}^{-1}=\int_0^\infty e^{-xt}\rho_2(t,\nu,\infty)\,dt, \tag{1.8}$$

where $\rho_2(t, \nu, \infty)$, $t\in(0,\infty)$, is an infinitely divisible probability density.

The above distributions are very natural companions to the distributions W_1 and W_2 of Hartman [16].

$$\left\{\frac{I_\mu(t)}{I_0(t)}\right\}\Big/\left\{\frac{I_\mu(\tau)}{I_0(\tau)}\right\}=\int_0^\infty \exp(-r\mu^2)W_1(dr,t,\tau),\qquad 0<t<\tau\leqq\infty. \tag{1.9}$$

$$\left\{\frac{K_\mu(\tau)}{K_0(\tau)}\right\}\Big/\left\{\frac{K_\mu(t)}{K_0(t)}\right\}=\int_0^\infty \exp(-r\mu^2)W_2(dr,t,\tau),\qquad 0<t<\tau\leqq\infty. \tag{1.10}$$

In all the above formulas $I_\nu(x)$ and $K_\nu(x)$ are the modified Bessel functions of the first and third kind, respectively. Hartman [16] proved that the distributions W_1 and W_2, of (1.9) and (1.10), are infinitely divisible; see also Hartman and Watson [17]. It might be of interest to note that as $\mu\to\infty$ the distribution $\rho_1(t, \nu, \mu)$ tends to the reciprocal of a gamma distribution which is the mixing distribution for the Student t distribution with $2\nu+2$ degrees of freedom. The information on the Student t and the variance mixtures can be found in [19] and [22]. The infinite divisibility of the Student t distribution was proved by Grosswald [15] and by Ismail [19] by proving the infinite divisibility of the reciprocal of a gamma distribution. Therefore the distribution $\rho_1(t, \nu, \mu)$ is the mixing distribution of another infinitely divisible distribution $\rho_3(t, \nu, \mu)$ depending on the two parameters (degrees of freedom) μ, ν, and as $\mu\to\infty$ this new distribution $\rho_3(t, \nu, \mu)$ will converge to the Student t distribution. The probability density functions ρ_1, ρ_2 and ρ_3 will be given explicitly in § 4, see (4.3), (4.14) and (4.8), respectively. As it will turn out, the representation (1.7) or (1.8) can be interpreted as a generalized Mittag-Leffler expansion for Bessel functions, while (1.8) is essentially a Fourier–Bessel expansion. These expansions will be derived in § 4. As an immediate consequence of Theorems 1.8 and 1.9 and the convolution property of the Laplace transform we see that the probability density functions $\rho_1(t, \nu, \mu)$ and $\rho_2(t, \nu, \mu)$ have the reproducing property

$$\rho_j(t,\alpha,\beta)*\rho_j(t,\beta,\gamma)=\rho_j(t,\alpha,\gamma),\qquad \gamma>\beta>\alpha>0,\qquad j=1,2, \tag{1.11}$$

where "*" is the usual convolution

$$f(t)*g(t)=\int_0^t f(\theta)g(t-\theta)\,d\theta,\qquad t>0, \tag{1.12}$$

associated with the Laplace transform.

In § 2 we show that the infinitely divisible distributions of Theorems 1.5, 1.7, 1.8, 1.9 and 1.10 are members of an important class of infinitely divisible distributions, the self-decomposable distributions or the class L. Section 3 contains proofs of Theorems 1.4–1.7, while § 4 contains proofs of Theorems 1.8–1.10 and some related results. In § 5 we study the infinite divisibility of certain functions of normal variables. Section 6 includes additional results on Bessel functions. We shall prove the following

THEOREM 1.11. *The functions* $(b/a)^{\nu}K_{\nu}(b\sqrt{x})/K_{\nu}(a\sqrt{x})$ *and* $(b/a)^{\nu}I_{\nu}(a\sqrt{x})/I_{\nu}(b\sqrt{x})$ *are self-decomposable Laplace transforms, hence are infinitely divisible and completely monotonic functions of* x *for* $b>a>0$.

THEOREM 1.12. *Assume* $\nu>0$. *If* $f(z)$ *and* $g(z)$ *are analytic functions in the half plane* $\operatorname{Re} z \geqq 0$ *with no common zeros and* $\operatorname{Re}\{g(z)/f(z)\} \geqq 0$ *for* $\operatorname{Re} z \geqq 0$, *then the functions*

$$\phi(z)=f(z)K_{\nu-1}(z)+g(z)K_{\nu}(z) \tag{1.13}$$

and

$$\zeta(z)=g(z)I_{\nu-1}(z)+f(z)I_{\nu}(z) \tag{1.14}$$

will have no zeros in the half planes $\operatorname{Re} z \geqq 0$ *and* $\operatorname{Re} z>0$ *respectively.*

COROLLARY 1.13. *Assume* $\nu>0$. *Then*

$$\operatorname{Re}\{K_{\nu-1}(z)/K_{\nu}(z)\}>0 \quad \textit{for } \operatorname{Re} z \geqq 0$$

and

$$\operatorname{Re}\{I_{\nu-1}(z)/I_{\nu}(z)\}>0 \quad \textit{for } \operatorname{Re} z>0.$$

Corollary 1.13 follows immediately from the proof of Theorem 1.12. The complete monotonicity of the quotients appearing in Theorem 1.11 is certainly new. The only other known results on monotonicity of quotients of Bessel functions can be found in Hartman and Watson [17], Ismail [20] and Lorch [26]. Recently Berndt and Glasser [2] evaluated certain integrals involving quotients of Bessel functions. Erdélyi and Kermack [8] proved that under the assumptions of Theorem 1.12, the function

$$\eta(z)=: f(z)K'_{\nu}(z)-g(z)K_{\nu}(z), \tag{1.15}$$

will have no zeros with nonnegative real part. Our result on the zeros of $\phi(z)$ of (1.13) is stronger than Erdélyi and Kermack's as will be explained in § 6. The problem of showing that the zeros of combinations of Bessel functions, like $\phi(z)$, $\zeta(z)$, and $\eta(z)$, lie off the closed right half plane was encountered in solving boundary value problems, see for example Carslaw and Jaeger [3] and [4]. Furthermore, integral representations for quotients of Bessel functions of order zero were also used in solving heat conduction problems, [4]. Section 6 also includes a method for summing Bessel series. We conclude the paper by mentioning a few related open problems in the last section, § 7.

2. The class L. Let $S_n=X_1+\cdots+X_n$ where the X_i's are independent random variables. Let $S_n^*=(S_n-b_n)/a_n$ where a_n and b_n are constants with $a_n\to\infty$, $a_{n+1}/a_n\to 1$. A distribution is in class L if it is the limit distribution of such a sequence S_n^*.

A random variable with characteristic function $\phi(t)$ belongs to the class L if and only if for each a, $0<a<1$, the ratio $\phi(t)/\phi(at)$ is a characteristic function. (For distributions on $[0,\infty)$ the term characteristic function may be replaced by Laplace transform.)

LEMMA 2.1. *A nonnegative random variable is a member of the class* L *if its Laplace transform is of the form* $e^{-h(x)}$, *with* $h'(x)=\int_0^\infty 1/(x+t)\,d\mu(t)$, $d\mu(t)\geqq 0$.

Proof. $e^{-h(x)}/e^{-h(ax)}=e^{-(h(x)-h(ax))}$, $0<a<1$.

By Theorem 1.1 it is sufficient to show that $(d/dx)(h(x)-h(ax))$ is completely monotonic.

$$\frac{d}{dx}(h(x)-h(ax))=\int_0^\infty \frac{1}{x+t}\,d\mu(t)-\int_0^\infty \frac{1}{x+t/a}\,d\mu(t).$$

Since $0<a<1$, the difference is clearly completely monotonic.

The quotient of two gamma variables, the square of a gamma variable, and the infinitely divisible distributions of Theorems 1.8, 1.9 and 1.10 have Laplace transforms of the form of Lemma 2.1 and hence are members of class L.

Consider a probability density of the form

$$\int_0^\infty (2\pi u)^{-1/2}\exp(-x^2/2u)\,dG(u),$$

where G is a distribution on $[0,\infty)$. This is called a variance mixture of the normal distribution. The Student t distribution is one of these mixtures. In the proof of the infinite divisibility of the t distribution (Grosswald [15], Ismail [19]), the mixing distribution G was shown to have a Laplace transform of the form of Lemma 2.1. In particular, G is the distribution of the reciprocal of a gamma variable, so this distribution is a member of the class L. But the t distribution itself is also in class L as is shown in the following lemma.

LEMMA 2.2. *A variance mixture of the normal distribution is in class L if the mixing distribution is in class L.*

Proof. Let G be the mixing distribution. Since G is in class L, there is a distribution G_a such that

$$\int_0^\infty e^{-tu}\,dG(u)\Big/\int_0^\infty e^{-a^2tu}\,dG(u)=\int_0^\infty e^{-tu}\,dG_a(u),\qquad a>0,\quad 0<a^2<1,\quad t>0.$$

Replacing t by $t^2/2$ we have

$$\frac{\int_0^\infty \exp(-t^2u/2)\,dG(u)}{\int_0^\infty \exp(-(at)^2u/2)\,dG(u)}=\int_0^\infty \exp(-t^2u/2)dG_a(u),\qquad -\infty<t<\infty,\quad 0<a<1.$$

The numerator on the left is the characteristic function of the normal mixture, and the term on the right is the characteristic function of a normal mixture. Hence the variance mixture of the normal distribution with the self-decomposable mixing distribution G is a member of the class L.

3. The proofs of Theorems 1.4 to 1.7. Representations (1.4) and (1.3) of Theorem 1.4 will be derived in the proofs of Theorems 1.5 and 1.7, respectively. Representation (1.5) is obtained from (1.4) by the use of the relation $\psi(a,c;z)=z^{1-c}\psi(a-c+1,2-c;z)$. Representation (1.2) is a special case of (1.4) obtained with the use of the relation

$$D_{-\nu}(z)=2^{-\nu/2}\exp\left\{-\frac{1}{4}z^2\right\}\psi\left(\frac{\nu}{2},\frac{1}{2};\frac{z^2}{2}\right).\tag{3.1}$$

This relation is found in Erdélyi et al. [9, p. 267]. Representations (1.2)–(1.5) of Theorem 1.4 can also be expressed in terms of the Whittaker function $W_{K,\mu}$.

3.1. Proof of Theorem 1.5. A gamma variable has the density

$$(\Gamma(\alpha))^{-1}\beta^{-\alpha}x^{\alpha-1}\exp(-x/\beta),\qquad x>0,\quad \beta>0,\quad \alpha>0.$$

The chi-square distribution with k degrees of freedom is a gamma with $\alpha = k/2$, $\beta = 2$. If X and Y are independent gamma variables with parameters (α, β), (α_0, β_0), respectively, then $Z = X/Y$ has the density

$$\Gamma(\alpha+\alpha_0)[\Gamma(\alpha)\Gamma(\alpha_0)]^{-1}(\beta_0/\beta)^{\alpha} z^{\alpha-1}\left(1+\frac{\beta_0}{\beta}z\right)^{-(\alpha+\alpha_0)}, \qquad z>0.$$

The density of an F variable with n and m degrees of freedom is this density with $\alpha = n/2$, $\alpha_0 = m/2$, $\beta_0 = n$, $\beta = m$.

Without loss of generality we will let $\beta = \beta_0$. The Laplace transform of the density of Z is $[\Gamma(\alpha_0)]^{-1}\Gamma(\alpha+\alpha_0)\psi(\alpha, 1-\alpha_0; t)$, where ψ is the Tricomi ψ function as defined in Erdélyi et al. [9]. To prove that the density is infinitely divisible, it is sufficient to show that $-(d/dz)\log\psi$ is completely monotonic on $(0, \infty)$.

Let $a = \alpha$, $c = 1-\alpha_0$, and

$$G(z) = \frac{-\psi'(a, c; z)}{\psi(a, c; z)} = \frac{a\psi(a+1, c+1; z)}{\psi(a, c; z)}, \tag{3.2}$$

by the differential relation of the ψ function.

The distribution of the quotient of two gamma variables is infinitely divisible for all $\alpha > 0$, $\alpha_0 > 0$, if $G(z)$ is completely monotonic on $(0, \infty)$ for all $a > 0$ and $c < 1$. We will show more. We will show that $G(z)$ is a Stieltjes transform of a nonnegative function; i.e.,

$$G(z) = \int_0^\infty \frac{1}{z+t} g(t)\, dt, \quad \text{with } g(t) \geqq 0. \tag{3.3}$$

Conditions (i) and (ii) of the representation theorem, Theorem 1.2, are verifiable by reference to the asymptotic expansion of $\psi(a, c; z)$ for $|\arg z| < \frac{3}{2}\pi$, given by equation (1) in [9, p. 278], and the fact that $\psi(a, c; z)$ has no zeros in the region $|\arg z| \leqq \pi$ [32, p. 672]. So G is a Stieltjes transform:

$$G(z) = \int_0^\infty (z+t)^{-1}\, d\theta(t).$$

Choose α, $\frac{2}{3} < \alpha < 1$ so that $\psi(a, c; z)$ has no zeros in $|\arg z| < \pi/\alpha$. Evaluating the contour integral of (1.1) as the sum of the residues at $z = \pm i\pi$ yields the relation

$$\frac{d\theta(t)}{dt} \equiv g(t) = \lim_{\eta\to 0^+} \frac{G(-t-i\eta)-G(-t+i\eta)}{2\pi i}, \qquad t>0.$$

We will use the notation $-t-i\eta = e^{-i\pi}(t+i\eta)$, $-t+i\eta = e^{i\pi}(t-i\eta)$.

$$g(t) = -\lim_{\eta\to 0^+}\frac{1}{2\pi i} \cdot \left[\frac{\psi'(a, c, e^{-i\pi}(t+i\eta))\psi(a, c, e^{i\pi}(t-i\eta)) - \psi(a, c, e^{-i\pi}(t+i\eta))\psi'(a, c, e^{i\pi}(t-i\eta))}{\|\psi(a, c, e^{i\pi}(t-i\eta))\|^2}\right]. \tag{3.4}$$

To show that $g(t) \geqq 0$ we need to show that

$$\lim_{\eta\to 0^+}\frac{1}{2\pi i}(-\text{numerator}) \geqq 0.$$

For the present, assume that c is not an integer. Then ψ has the representation

$$\lim_{\eta\to 0^+}\psi(a, c, e^{\pm i\pi}(t\mp i\eta)) = \frac{\Gamma(1-c)}{\Gamma(a-c+1)}y_1(-t) - e^{\mp i\pi c}(-1)^{1-c}\frac{\Gamma(c-1)}{\Gamma(a)}y_2(-t)$$

$-a < \min(0, 1-c)$, where $y_1(a, c; z)$ and $y_2(z, c; z)$ are two solutions of the confluent hypergeometric equation with the Wronskian $(1-c)z^{-c}e^z$. These relations are found on pages 252, 253 and 263 of Erdélyi et al. [9].

Write the limit as $K_1 y_1 - e^{\mp i\pi c} K_2 y_2$. Then the numerator of G can be written, (evaluated at $-t$), as

$$(K_1 y_1' - e^{-i\pi c} K_2 y_2')(K_1 y_1 - e^{i\pi c} K_2 y_2) - (K_1 y_1 - e^{-i\pi c} K_2 y_2)(K_1 y_1' - e^{i\pi c} K_2 y_2')$$
$$= K_1 K_2 [y_1 y_2' 2i \sin(\pi c) - y_1' y_2 2i \sin(\pi c)].$$

Since $y_1 y_2' - y_1' y_2 = (1-c)(-t)^{-c} e^{-t}$ and $\sin(\pi c) = \pi(\Gamma(c-1)\Gamma(2-c))^{-1}$, it follows that

$$-\frac{1}{2\pi i}(\text{numerator of } g) = \frac{t^{-c} e^{-t}}{\Gamma(a)\Gamma(a-c+1)} > 0 \tag{3.5}$$

for $t > 0$, $a > 0$, $c < 1$, c not an integer.

We have been assuming that c is not an integer. But all of the functions used are continuous with respect to c and so the final expression is also valid for c a negative integer or zero. Thus $G(z) = \int_0^\infty 1/(z+t) g(t)\, dt$ with $g(t) \geqq 0$ and the quotient of two gamma variables is infinitely divisible.

Remark. 3.2. By substituting (3.5) into equation (3.4) and the result into (3.3) we get representation (1.3) of Theorem 1.4.

3.3. Proof of Theorem 1.6. We are fairly sure that this result is known to some people, but we have not seen it stated anywhere.

For integer degrees of freedom, k, and noncentrality parameter θ, starting with the normal density it is straightforward to show that the Laplace transform of the noncentral chi-square distribution is

$$(1-2t)^{-k/2} \exp(-t\theta/(1-2t)) = (1-2t)^{-k/2} \exp\left(+\frac{\theta}{2} - \theta/(2-4t)\right).$$

This is the product of the Laplace transform of a gamma distribution and the Laplace transform of a compound Poisson distribution, both of which are infinitely divisible for all real $k > 0$.

For $2\nu + 2$ degrees of freedom the density of the noncentral chi-square distribution can be written as

$$\tfrac{1}{2} \exp(-(\theta + x)/2)(x/\theta)^{\nu/2} I_\nu((\theta x)^{1/2}).$$

Without noting that this is the density of a noncentral chi-square, Feller [12, Chap. 13, § 3] discusses this density as an infinitely divisible Bessel function density.

3.4. Proof of Theorem 1.7. The square of a gamma variable has density

$$(2\Gamma(2\nu))^{-1} \beta^{-2\nu} x^{\nu-1} \exp(-\sqrt{x}/\beta), \qquad \nu > 0.$$

Without loss of generality, let $\beta = 1$. The Laplace transform of the density is found in Erdélyi et al. [11, p. 147] to be

$$\frac{(2t)^{-\nu}}{2} \exp((8t)^{-1}) D_{-2\nu}[(2t)^{-1/2}]$$

where D is a parabolic cylinder function. In terms of the ψ function, Erdélyi et al. [9, p. 267], the Laplace transform is $\frac{1}{2}(4t)^{-\nu} \psi(\nu, \frac{1}{2}, (4t)^{-1})$.

We need to verify conditions (i) and (ii) of the Stieltjes representation theorem for the function

$$G_1(z)=\frac{\nu}{z}-\frac{\nu}{4z^2}\frac{\psi(\nu+1,\frac{3}{2},(4z)^{-1})}{\psi(\nu,\frac{1}{2},(4z)^{-1})}. \tag{3.6}$$

From the comments about ψ in Theorem 1.5, condition (i) is clearly satisfied.

With the use of the asymptotic formulas on pages 262 and 278 of Erdélyi et al. [9], it is straightforward to verify condition (ii) and hence G_1 is a Stieltjes transform. As in Theorem 1.5 we need to show that $\lim_{\eta\to 0^+}(G_1(-t-i\eta)-G_1(-t+i\eta))/(2\pi i)\geqq 0$, $t>0$.

In the limit the ν/z terms cancel. The remainder of the proof is almost identical to the corresponding part of the proof of Theorem 1.5. The terms $e^{i\pi}(t+i\eta)^{-1}$ and $e^{-i\pi}(t-i\eta)^{-1}$ are substituted for $e^{-i\pi}(t+i\eta)$ and $e^{i\pi}(t-i\eta)$ in all the formulas. The factor $(4z)^{-2}$ gives a term multiplied by $(4z)^{-2}$ and two terms that go to zero as $\eta\to 0^+$. Proceeding as in Theorem 1.5, the limit as $\eta\to 0^+$ of (numerator of G_1)$/2\pi i$ is

$$(4t^{3/2}\Gamma(\nu)\Gamma(\nu+\tfrac{1}{2}))^{-1}\exp(-(4t)^{-1}). \tag{3.7}$$

This quantity is positive for $t>0$, and hence the square of a gamma variable is infinitely divisible.

Remark 3.5. Representation (1.3) of Theorem 1.4 is obtained by combining relations (3.2), (3.3), (3.4), (3.6), and (3.7) in the appropriate manner and replacing each ψ function by the corresponding parabolic cylinder function D. It can also be proved directly by showing that the left-hand side of (1.3) is a Stieltjes transform then inverting the transform. The only additional information required is the behavior of $D_{-\nu}(z)$ for large $|z|$ and this is available in Olver [28].

4. Some infinitely divisible distributions. The present section contains proofs of Theorems 1.8–1.10 and the explicit formulas for the densities $\rho_1(t,\nu,\mu)$ and $\rho_2(t,\nu,\mu)$.

For the proof of Theorem 1.8 and later theorems we need the integral representation (Grosswald [15] and Ismail [19])

$$z^{-1/2}K_{\nu-1}(\sqrt{z})/K_\nu(\sqrt{z})=2\pi^{-2}\int_0^\infty\frac{t^{-1}}{z+t}\{J_\nu^2(\sqrt{t})+Y_\nu^2(\sqrt{t})\}^{-1}\,dt, \tag{4.1}$$
$$\nu>0,\quad |\arg z|<\pi.$$

4.1. Proof of Theorem 1.8. The function $z^{(\nu-\mu)/2}K_\nu(\sqrt{z})/K_\mu(\sqrt{z})$ is analytic in the cut plane $|\arg z|<\pi$ because $K_\mu(z)$ has no zeros with $|\arg z|\leqq\pi/2$, Erdélyi et al. [10, p. 62]. We now apply Theorem 1.1. Let

$$h(x)=\ln\{x^{(\mu-\nu)/2}K_\mu(\sqrt{x})/K_\nu(\sqrt{x})\}.$$

Using the differential recurrence relation for K_ν (Erdélyi et al. [10]), we get, for $x>0$,

$$2h'(x)=x^{-1/2}\{K_{\nu-1}(\sqrt{x})/K_\nu(\sqrt{x})\}-x^{-1/2}\{K_{\mu-1}(\sqrt{x})/K_\mu(\sqrt{x})\}.$$

Hence from the above integral representation we obtain

$$h'(x)=\pi^{-2}\int_0^\infty\frac{t^{-1}}{x+t}[\{J_\nu^2(\sqrt{t})+Y_\nu^2(\sqrt{t})\}^{-1}-\{J_\mu^2(\sqrt{t})+Y_\mu^2(\sqrt{t})\}^{-1}]\,dt. \tag{4.2}$$

Nicholson's formula (Watson [33, p. 479])

$$J_\nu^2(z)+Y_\nu^2(z)=8\pi^{-2}\int_0^\infty K_0(2z\sinh t)\cosh(2\nu t)\,dt,\qquad \operatorname{Re} z>0,$$

shows that $J_\nu^2(t)+Y_\alpha^2(t)$ is an increasing function of α. Thus the integrand in (4.2) is a positive multiple, independent of x, of $(x+t)^{-1}$, hence $h'(x)$ is completely monotonic. It remains to show that $h(0^+)=0$. From (12) and (13) of p. 5, (37) of p. 9 in Erdélyi et al. [10] and the well known relationship

$$\Gamma(z)\Gamma(1-z)=\pi \csc(\pi z),$$

we see that, for $\alpha>0$, $(z^\alpha K_\alpha(z)/(\Gamma(\alpha)2^{\alpha-1}))\to 1$ as $z\to 0$. This shows that $h(0^+)=0$ and the proof is complete.

Remark 4.2. The explicit form of $\rho_1(t,\nu,\mu)$ is

$$\rho_1(t,\nu,\mu)=\frac{\Gamma(\mu)2^{\mu-\nu}}{\Gamma(\nu)\pi}\int_0^\infty e^{-tu}\frac{u^{(\nu-\mu)/2}\{J_\mu(\sqrt{u})Y_\nu(\sqrt{u})-J_\nu(\sqrt{u})Y_\mu(\sqrt{u})\}\,du}{J_\mu^2(\sqrt{u})+Y_\mu^2(\sqrt{u})} \tag{4.3}$$

which follows from (2.1) in Ismail [20] and the observation that the Stieltjes transform is a two-fold Laplace transform.

Remark 4.3. It is shown in Ismail [20] that as $\mu\to\infty$, $K_\mu(x)\sim 2^{\mu-1}x^{-\mu}\Gamma(\mu)$. This and (1.6) imply

$$\frac{x^{\nu/2}K_\nu(\sqrt{x})}{2^{\nu-1}\Gamma(\nu)}=\int_0^\infty e^{-xt}\rho_1(t,\nu,\infty)\,dt,$$

so that [11. (39), p. 283], $\rho_1(t,\nu,\infty)$ is the probability density function of the reciprocal of a gamma distribution since

$$\rho_1(t,\nu,\infty)=\frac{4^{-\nu}t^{-\nu-1}}{\Gamma(\nu)}e^{-(1/4)/t},\qquad t\in(0,\infty). \tag{4.4}$$

The probability density function for the Student t distribution with k degrees of freedom can be written as

$$\frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\,\Gamma(k/2)}(1+x^2/k)^{-(k+1)/2}=\frac{(k/2)^{k/2}}{\Gamma(k/2)}\int_0^\infty \frac{e^{-x^2/(2u)}}{\sqrt{2\pi u}}u^{-(k+2)/2}\,e^{-k/2u}\,du, \tag{4.5}$$

because a variance mixture of normal distributions has the form $\int_0^\infty (2\pi u)^{-1/2}\,e^{-x^2/2u}dG(u)$, where G is the mixing distribution; see [21] and [22]. This suggests that the Student t admits a generalization, generalized Student t distribution, which is a variance mixture of the normal distribution and a mixing distribution with probability density function $(2k)^{-1}\rho_1(t/(2k),k/2,\mu)$ or $(4\nu)^{-1}\rho_1(t/(4\nu),\nu,\mu)$. Define

$$\rho_3(x,\nu,\mu)=\int_0^\infty (2\pi t)^{-1/2}\,e^{-x^2/2t}(4\nu)^{-1}\rho_1(\tfrac{1}{4}t/\nu,\nu,\mu)\,dt;$$

hence

$$\rho_3(x,\nu,\mu)=(2\pi)^{-1/2}(4\nu)^{-1/2}(4\nu)^{-1}\int_0^\infty\int_0^\infty e^{-tu/4\nu}\,e^{-x^2/2t}t^{-1/2}g(u,\nu,\mu)\,dt\,du, \tag{4.6}$$

with

$$g(u,\nu,\mu)=\frac{\Gamma(\mu)}{\Gamma(\nu)}\frac{2^{\mu-\nu}}{\pi}u^{(\nu-\mu)/2}\frac{\{J_\mu(\sqrt{u})Y_\nu(\sqrt{u})-J_\nu(\sqrt{u})Y_\mu(\sqrt{u})\}}{\{J_\mu^2(\sqrt{u})+Y_\mu^2(\sqrt{u})\}}. \tag{4.7}$$

The inner integral in (4.6) is a Laplace transform and can be evaluated using (27) in

Erdélyi et al. [11, p. 146]. Therefore

$$\rho_3(x,\nu,\mu)=\frac{\Gamma(\mu)}{4\pi\Gamma(\nu+1)}\int_0^\infty e^{-|x|\sqrt{\mu/2\nu}}\left(\frac{\sqrt{u}}{2}\right)^{\nu-\mu}\left(\frac{2\nu}{u}\right)^{1/2} \tag{4.8}$$
$$\cdot\frac{\{J_\mu(\sqrt{u})Y_\nu(\sqrt{u})-J_\nu(\sqrt{u})Y_\mu(\sqrt{u})\}\,du}{\{J_\mu^2(\sqrt{u})+Y_\mu^2(\sqrt{u})\}}.$$

These considerations lead to

THEOREM 4.4. *The function $\rho_3(x,\nu,\mu)$ of (4.8) is the probability density function for an infinitely divisible probability distribution. As $\mu\to\infty$, this distribution tends to the Student t distribution with 2ν degrees of freedom.*

All that remains to be proved is the evaluation of the limiting distribution $\rho_3(x,\nu,\infty)$. The asymptotic behavior of $J_\mu(x)$, $Y_\mu(x)$ as $\mu\to\infty$ and $x>0$ can be determined from their explicit forms, (2) and (4), [10, p. 4],

$$J_\mu(z)=\sum_{m=0}^{\infty}\frac{(z/2)^{\nu+2m}(-1)^m}{\Gamma(m+\nu+1)m!},$$

$$Y_\mu(z)=[\sin(\nu\pi)]^{-1}[J_\nu(z)\cos(\nu\pi)-J_{-\nu}(z)].$$

Indeed $J_\mu(x)\sim(x/2)^\mu/\Gamma(\mu+1)$ and $Y_\mu(x)\sim-(2/x)^\mu\Gamma(\mu)/\pi$, and the limiting density function is

$$\rho_3(x,\nu,\infty)=\tfrac{1}{4}[\Gamma(\nu+1)]^{-1}\int_0^\infty e^{-|x|\sqrt{u/2\nu}}\left(\frac{\sqrt{u}}{2}\right)^{\nu}\left(\frac{2\nu}{u}\right)^{1/2}J_\nu(\sqrt{u})\,du$$
$$=\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{2}\right)^{\nu/2}\int_0^\infty e^{-|x|\theta}\theta^\nu J_\nu(\theta\sqrt{2\nu})\,d\theta$$
$$=\frac{\Gamma(\nu+\frac{1}{2})}{\sqrt{2\nu\pi}\,\Gamma(\nu)}\left(1+\frac{x^2}{2\nu}\right)^{-\nu-1/2},$$

and a comparison with (4.5) identifies $\rho_3(x,\nu,\infty)$ as the probability density function for the Student t distribution.

LEMMA 4.5. *The Bessel function of the first kind $J_\nu(z)$ has infinitely many positive zeros $\{j_{\nu,n}\}_{n=1}^\infty$; all are simple zeros and for every fixed n, $j_{\nu,n}$ is a continuous increasing function of ν, $\nu>-1$. If $\nu+1\geqq\mu>\nu>-1$, then*

$$0<j_{\nu,1}<j_{\mu,1}<j_{\nu,2}<\cdots.$$

Proof. This lemma follows from the fact that $j_{\nu,n}$ increases with ν ($\nu>-1$), Watson [33, p. 508] and that

$$j_{\mu,k}<j_{\nu+1,k}<j_{\nu,k+1},$$

Watson [33, p. 479].

For the proof of Theorem 1.9 we will need Lemma 4.5 and the following representation (Mittag-Leffler expansion)

$$I_{\nu+1}(z)/I_\nu(z)=2z\sum_{n=1}^\infty(z^2+j_{\nu,n}^2)^{-1},\qquad \nu>-1.$$

This formula can be found, for example, in Erdélyi et al. [10, (3) p. 61] upon using the following relation between $I_\nu(z)$ and $J_\nu(iz)$, see [10, p. 5],

$$I_\nu(z)=e^{-i\nu\pi/2}J_\nu(z\,e^{i\pi/2})=\sum_{k=0}^\infty\left(\frac{1}{2}z\right)^{2k+\nu}\Big/[k!\Gamma(\nu+k+1)]. \tag{4.9}$$

4.6. Proof of Theorem 1.9. Let $h(x)$ be the natural logarithm of the left hand side in (1.7). Clearly $h(x)$ is well defined since $z^{-\alpha}I_\alpha(z)$ has no real zeros and is analytic in z, for $\alpha>-1$. Clearly, by the Mittag-Leffler expansion and the differential recurrence relation for I_ν (Erdélyi et al. [10, p. 79]) we get

$$\begin{aligned}2h'(x)&=\frac{1}{\sqrt{x}}\frac{I_{\nu+1}(\sqrt{x})}{I_\nu(\sqrt{x})}-\frac{1}{\sqrt{x}}\frac{I_{\mu+1}(\sqrt{x})}{I_\mu(\sqrt{x})}\\ &=2\sum_{n=1}^{\infty}\frac{j_{\mu,n}^2-j_{\nu,n}^2}{(x+j_{\nu,n}^2)(x+j_{\mu,n}^2)}.\end{aligned}$$

Therefore $h'(x)$ is completely monotonic because $j_{\mu,n}>j_{\nu,n}$ for $\mu>\nu$. What is left is to check that $h(0)=0$, which is certainly the case, in view of (4.9). This completes the proof of Theorem 1.9.

Our next result is a generalization of the Mittag-Leffler expansion.

THEOREM 4.7. *We have*

$$z^{(\nu-\mu)/2}I_\mu(\sqrt{z})/I_\nu(\sqrt{z})=-2\sum_{n=1}^{\infty}\frac{j_{\nu,n}^{\nu+1-\mu}}{z+j_{\nu,n}^2}J_\mu(j_{\nu,n})/J'_\nu(j_{\nu,n}), \tag{4.10}$$

for $\mu>\nu>-1$, $|\arg z|<\pi$.

Proof. Let us first assume that

$$F(z)=:z^{(\nu-\mu)/2}I_\mu(\sqrt{z})/I_\nu(\sqrt{z})=\int_0^\infty\frac{d\theta(t)}{z+t}. \tag{4.11}$$

The relation betweem $I_\nu(z)$ and $J_\nu(z)$ is provided by (4.9), namely

$$I_\nu(z\,e^{\pm i\pi/2})=e^{\pm i\pi\nu/2}J_\nu(z). \tag{4.12}$$

From the inversion theorem, Theorem 1.3, it is not too difficult to see that when t lies strictly between the squares of two successive positive zeros of $J_\nu(z)$, or in $(0, j_{\nu,1}^2)$, the function $\theta(t)$ will be absolutely continuous with

$$2\pi i\frac{d\theta(t)}{dt}=\lim_{\eta\to0^+}\{F(-t-i\eta)-F(-t+i\eta)\}=F(r\,e^{-i\pi})-F(t\,e^{i\pi})=0,$$

by (4.12). This shows that $\theta(t)$ is a step function with jumps at the points $\{j_{\nu,k}^2\}_{k=1}^\infty$, the squares of the positive zeros of $J_\nu(z)$. Next we consider the case $j_{\nu,k-1}<t_1<j_{\nu,k}<t_2<j_{\nu,k+1}$, where $j_{\nu,0}$ is interpreted as zero. Rewrite $F(z)$ as

$$F(z)=z^{(\nu-\mu)/2}\frac{I_\mu(\sqrt{z})}{(z+j_{\nu,k}^2)}\{(z+j_{\nu,k}^2)/I_\nu(\sqrt{z})\}. \tag{4.13}$$

Observe that if $s_1<0<s_2$,

$$(2\pi i)^{-1}\lim_{\eta\to0^+}\int_{s_1}^{s_2}\left\{\frac{1}{-t-i\eta}-\frac{1}{-t+i\eta}\right\}dt=\lim_{\eta\to0^+}\int_{s_1}^{s_2}\frac{\eta\pi^{-1}\,dt}{t^2+\eta^2}=1,$$

and that

$$\lim_{z \to j_{\nu,k}^2 e^{\pm i\pi}} z^{(\nu-\mu)/2} I_\mu(\sqrt{z})(z + j_{\nu,k}^2)/I_\nu(\sqrt{z})$$

$$= \lim_{z \to j_{\nu,k}^2 e^{\pm i\pi}} 2z^{(\nu-\mu+1)/2} I_\mu(\sqrt{z})/I'_\nu(\sqrt{z}) = -2j_{\nu,k}^{\nu-\mu+1} J_\mu(j_{\nu,k})/J'_\nu(j_{\nu,k}).$$

Formally, we have $\int (z+t)^{-1}\, d\theta(t)$ equal to the right side of (4.10). Thus if F is a Stieltjes transform then (4.10) holds. However, the function $z^{-\nu/2} I_\nu(\sqrt{z})$ $\nu > -1$ is analytic and has no zeros with $|\arg z| < \pi$ because all the zeros of $z^{-\nu} I_\nu(z)$ lie on the imaginary axis. $F(z) - \int_0^\infty (z+t)^{-1}\, d\theta(t)$ is obviously analytic in $|\arg z| < 3\pi/2$ and the behavior near $z = 0$ and $z = \infty$ can be determined from (4.9), (5) and (6) on p. 86 of Erdélyi et al. [10]. Thus the representation theorem, Theorem 1.2, is applicable to $F(z) - \int_0^\infty (z+t)^{-1}\, d\theta(t)$ and (4.10) holds.

As an immediate corollary of Theorem 4.7 we obtain the explicit form of the probability density function $\rho_2(t, \nu, \mu)$ of Theorem 1.6.

COROLLARY 4.8. *The probability density function* $\rho_2(t, \nu, \mu)$ *is given by*

$$\rho_2(t, \nu, \mu) = -2^{\mu-\nu+1} \frac{\Gamma(\mu+1)}{\Gamma(\nu+1)} \sum_{n=1}^\infty j_{\nu,n}^{\nu-\mu+1} \frac{J_\mu(j_{\nu,n})}{J'_\nu(j_{\nu,n})} e^{-j_{\nu,n}^2 t}, \qquad t \in (0, \infty). \tag{4.14}$$

In the special case $\mu = \nu + 1$ we get, using (54) in Erdélyi et al. [10, p. 11],

$$\rho_2(t, \nu, \nu+1) = 4(\nu+1) \sum_{n=1}^\infty e^{-j_{\nu,n}^2 t}, \tag{4.15}$$

and when $\nu = \frac{1}{2}$, (4.15) reduces to

$$\rho_2(t, \tfrac{1}{2}, \tfrac{3}{2}) = 6 \sum_{n=1}^\infty e^{-n^2\pi^2 t}, \tag{4.16}$$

since $J_{1/2}(z) = \sqrt{(2/(\pi z))} \sin z$.

The function $\rho_2(t, \frac{1}{2}, \frac{3}{2})$ is a theta function so $\rho_2(t, \nu, \nu+1)$ may be considered as a natural generalized theta function.

Remark 4.9. When $\nu + 1 \geqq \mu > \nu > -1$, the numbers $J_\mu(j_{\nu,n})/J'_\nu(j_{\nu,n})$, $n = 1, 2, \cdots$ are negative because of the simplicity of the zeros $\{j_{\nu,n}\}_{n=1}^\infty$ and the interlacing property of Lemma 4.5. See Fig. 1. In the case $\mu > \nu + 1$ some of the numbers $J_\mu(j_{\nu,n})/J'_\nu(j_{\nu,n})$ might change sign but, due to the reproducing property (1.11), $\rho_2(t, \nu, \mu)$ will be nonnegative for all $\mu > \nu > -1$.

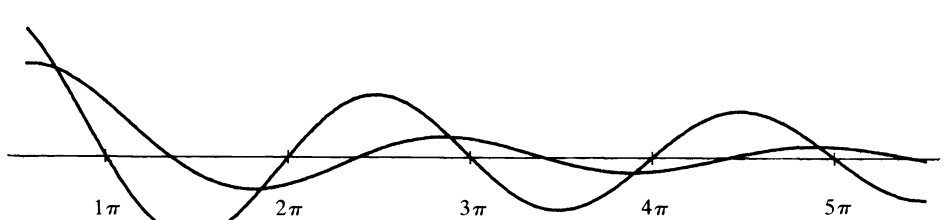

FIG. 1. *The graph of* $J_{1/2}$ *and* $J_{3/2}$ *illustrating the negativity of* $J_\mu(j_{\nu,n})/J'_\nu(j_{\nu,n})$.

We conclude this section by finding the explicit form of the limiting distribution $\rho_2(t, \nu, \infty)$ and the corresponding Mittag-Leffler expansion. The proof is almost identical with our proof of Theorem 4.7 and will be omitted.

THEOREM 4.10. *We have*

$$z^{\nu/2}/I_\nu(\sqrt{z}) = -2\sum_{n=1}^{\infty}\frac{j_{\nu,n}^{\nu+1}}{z+j_{\nu,n}^2}\{J'_\nu(j_{\nu,n})\}^{-1} \tag{4.17}$$

and

$$\rho_2(t,\nu,\infty) = \frac{-2^{1-\nu}}{\Gamma(\nu+1)}\sum_{n=1}^{\infty} j_{\nu,n}^{\nu+1}\{J'_\nu(j_{\nu,n})\}^{-1}e^{-j_{\nu,n}^2 t}, \qquad t\in(0,\infty). \tag{4.18}$$

Remark 4.11. Special cases of the densities $\rho_2(t,\nu,\mu)$ and $\rho_2(t,\nu,\infty)$ appear in Feller [12]; see formulas (5.9) and (5.11) in Chapter 10.

5. The infinite divisibility of other functions of normal variables. Goldie [13] has shown that nonnegative random variables with completely monotonic densities are infinitely divisible. Steutel [29] has shown that mixtures of the Laplace density of the form $\int_0^\infty u/2\exp(-|x|u)\,dG(u)$ are infinitely divisible; i.e., densities symmetric about zero and completely monotonic on $(0,\infty)$ are infinitely divisible. These results can be used to easily show the infinite divisibility of certain densities. For example, let Z be $N(0,1)$ and let $Y=Z^{2k}$ for k a positive integer. Y has the density

$$(\sqrt{2\pi}\,k)^{-1}y^{(1/(2k))-1}\exp(-\tfrac{1}{2}y^{1/k}), \qquad y>0.$$

This density is a product of completely monotonic functions if $k\geqq 1$. Let $W=Z^{2k+1}$. The density of W is

$$(\sqrt{2\pi}(2k+1))^{-1}|w|^{-2k/(2k+1)}\exp(-\tfrac{1}{2}|w|^{2/(2k+1)}), \qquad -\infty<w<\infty.$$

For $k\geqq 1$ this density is a product of functions which are completely monotonic on $(0,\infty)$, and hence the density is a mixture of Laplace densities. Thus if Z is $N(0,1)$, then Z^k is an infinitely divisible variable for any integer $k\geqq 1$.

Let X have the gamma density $(\Gamma(\alpha)\beta^\alpha)^{-1}x^{\alpha-1}\exp(-x/\beta)$, $x>0$. Let $Y=X^\nu$. The density of Y is

$$(\nu\Gamma(\alpha)\beta^\alpha)^{-1}y^{\alpha/\nu-1}\exp\left(-\frac{1}{\beta}y^{1/\nu}\right), \qquad y>0.$$

This is completely monotonic and hence infinitely divisible if $\nu\geqq\max[\alpha,1]$. Thus if X is a chi-square variable with k degrees of freedom, $k\geqq 2$, X^ν is infinitely divisible for any real ν such that $\nu\geqq k/2$.

The t density with k degrees of freedom is

$$\Gamma\left(\frac{k+1}{2}\right)\left(\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)\right)^{-1}\left(1+\frac{t^2}{k}\right)^{-(k+1)/2}, \qquad -\infty<t<\infty.$$

Again, the density of t^m is easily shown to be completely monotonic on $(0,\infty)$ for m a positive integer, $m\geqq 2$. Thus t^m is an infinitely divisible variable for $m=2,3,4,\cdots$.

For the F distribution, the density of f^ν is easily seen to be completely monotonic and hence infinitely divisible if ν is greater than or equal to one half of the degrees of freedom of the numerator.

The reciprocals of chi-square, Cauchy, and F variables are infinitely divisible. We now show that the reciprocal of a normal variable is not infinitely divisible, and neither, in general, is the reciprocal of a t variable.

Let Z have a $N(0,1)$ distribution. Assume that Z^{-1} is infinitely divisible. Then $Y=\sqrt{2u}\,Z^{-1}$ is infinitely divisible for all $u>0$. The density of Y is

$$\pi^{-1/2}y^{-2}\sqrt{u}\exp(-uy^{-2})$$

and the characteristic function is found in Oberhettinger [27, p. 13] to be

$$\phi(t)=\frac{\sqrt{\pi}}{2}\sum_{0}^{\infty}\frac{(-1)^n(\sqrt{u}\,t)^n}{n!\Gamma((n+1)/2)}.$$

Infinitely divisible characteristic functions have no zeros so under our assumptions this characteristic function is strictly positive for all real t. In particular, for fixed $t>0$,

$$\sqrt{u}\exp(-u)\phi(t)>0 \quad \text{for all } u>0.$$

Hence

$$\frac{\sqrt{\pi}}{2}\int_0^{\infty}\sum_{0}^{\infty}\frac{u^{n/2+1/2}(-t)^n e^{-u}}{n!\Gamma((n+1)/2)}\,du>0.$$

The sum converges absolutely, so the integral and summation sign may be interchanged. Integrating term by term, we get that the sum is the Taylor series expansion of $\sqrt{\pi}\,4^{-1}(1-t)\exp(-t)$ which is negative for $0<t<1$. This contradicts the assumption of the infinite divisibility of Z^{-1}.

The t distribution with 3 degrees of freedom has the density

$$2(\sqrt{3}\,\pi)^{-1}\left(1+\frac{t^2}{3}\right)^{-2}.$$

The density of $Y=1/t$ is

$$2(\sqrt{3}\,\pi)^{-1}y^2\left(1+\frac{y^2}{3}\right)^{-2}$$

with characteristic function

$$1.5(1-\sqrt{3}|u|)\exp(-\sqrt{3}|u|)$$

which is negative at $u=1$. Hence t^{-1} is not an infinitely divisible variable.

6. Further results. We start the current section by a proof of Theorem 1.11.

6.1. Proof of Theorem 1.11. For positive x, the function $K_\nu(\sqrt{x})$ does not vanish and is infinitely differentiable. Consequently, the function

$$f(x)=\ln\{K_\nu(a\sqrt{x})/K_\nu(b\sqrt{x})\}, \qquad x>0, \tag{6.1}$$

is a well-defined infinitely differentiable function. An appeal to (4.1) and the differential recurrence relation for K_ν yield

$$\begin{aligned}2f'(x)&=\frac{b}{\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})}-\frac{a}{\sqrt{x}}\frac{K_{\nu-1}(a\sqrt{x})}{K_\nu(a\sqrt{x})}\\&=2\pi^{-2}\int_0^{\infty}\{(x+t/b^2)^{-1}-(x+t/a^2)^{-1}\}t^{-1}\{J_\nu^2(\sqrt{t})+Y_\nu^2(\sqrt{t})\}^{-1}\,dt.\end{aligned} \tag{6.2}$$

From this identity it is easy to see by direct differentiation that $f'(x)$ is completely monotonic since $a\leqq b$. Therefore

$$\left(\frac{b}{a}\right)^{\nu}K_\nu(b\sqrt{x})/K_\nu(a\sqrt{x})=e^{-[f(x)-\nu\log(b/a)]}, \tag{6.3}$$

hence is infinitely divisible by Theorem 1.1 since $f(0+)=\nu\log(b/a)$, and the complete monotonicity of $K_\nu(b\sqrt{x})/K_\nu(a\sqrt{x})$ follows from (6.3), the complete monotonicity of

$f'(x)$ and the formula for the nth derivative of a composite function, the Faa Di Bruno formula. The complete monotonicity of $I_\nu(a\sqrt{x})/I_\nu(b\sqrt{x})$ can be proved along the same lines. This completes the proof of Theorem 1.11.

6.2. Proof of Theorem 1.12. The integral representation (4.1) implies

$$\frac{K_{\nu-1}(z)}{K_\nu(z)}=2\pi^{-2}\int_0^\infty \frac{zt^{-1}}{z^2+t}\{J_\nu^2(\sqrt{t})+Y_\nu^2(\sqrt{t})\}^{-1}\,dt, \qquad \nu\geqq 0, \quad |\arg z|<\pi/2. \tag{6.4}$$

The relationship (6.4) and the observation

$$\operatorname{Re}\left\{\frac{z}{z^2+t}\right\}=\operatorname{Re}\left\{\frac{zz^{-2}+tz}{|z^2+t|^2}\right\}=\frac{\{|z|^2+t\}}{|z^2+t|}\operatorname{Re} z$$

show that $\operatorname{Re} K_{\nu-1}(z)/K_\nu(z)>0$ if $\operatorname{Re} z>0$. Rewrite (1.13) when $f(z)\neq 0$, as

$$\phi(z)=f(z)K_\nu(z)\left\{\frac{K_{\nu-1}(z)}{K_\nu(z)}+\frac{g(z)}{f(z)}\right\}.$$

Thus $\phi(z)\neq 0$ when $\operatorname{Re} z>0$ and $f(z)\neq 0$. If $f(z)=0$, $\phi(z)\neq 0$ since $f(z)$ and $g(z)$ have no common zeros. It remains to consider the case $\operatorname{Re} z=0$. Let $z=e^{\pm i\pi/2}t$, $t>0$. The connection between the modified Bessel functions of the third kind $K_\nu(z)$ and the Hankel functions $H_\nu^{(1)}(z)$ and $H_\nu^{(2)}(z)$ is, Erdélyi et al. [10, (15), p. 5],

$$K_\nu(te^{i\pi/2})=-\frac{i\pi}{2}e^{-i\nu\pi/2}H_\nu^{(2)}(t) \quad\text{and}\quad K_\nu(te^{-i\pi/2})=\frac{i\pi}{2}e^{i\nu\pi/2}H_\nu^{(1)}(t),$$

where, Erdélyi [10, (5) and (6), p. 4],

$$J_\nu(z)+iY_\nu(z)=H_\nu^{(1)}(z) \quad\text{and}\quad J_\nu(z)-iY_\nu(z)=H_\nu^{(2)}(z).$$

Hence

$$\operatorname{Re}\frac{K_{\nu-1}(te^{\pm i\pi/2})}{K_\nu(te^{\pm i\pi/2})}=\operatorname{Re}\left\{e^{\pm i\pi/2}\frac{J_{\nu-1}(t)\mp iY_{\nu-1}(t)}{J_\nu(t)\mp iY_\nu(t)}\right\}=\frac{J_\nu(t)Y_{\nu-1}(t)-J_{\nu-1}(t)Y_\nu(t)}{J_\nu^2(t)+Y_\nu^2(t)}$$
$$=\frac{2}{\pi t}\{J_\nu^2(t)+Y_\nu^2(t)\}^{-1},$$

by (35) [10, p. 80]. Thus $\operatorname{Re}\{K_{\nu-1}(z)/K_\nu(z)\}>0$ even when z is purely imaginary. Using the same argument used when $\operatorname{Re} z>0$, we see that $\phi(z)$ has no zeros in the half plane $\operatorname{Re} z\geqq 0$. This argument can be repeated to prove that $\zeta(z)$ has no zeros with positive real parts.

Remark 6.3. Our proof depended only on the fact that

$$\operatorname{Re}\{K_{\nu-1}(z)/K_\nu(z)\}>0 \quad \text{for } \operatorname{Re} z\geqq 0, \tag{6.5}$$

while Erdélyi and Kermack's result depended on

$$\operatorname{Re}\{-K_\nu'(z)/K_\nu(z)\}>0 \quad \text{for } \operatorname{Re} z\geqq 0. \tag{6.6}$$

Clearly (6.6) is stronger than (6.5) since

$$-\frac{K_\nu'(z)}{K_\nu(z)}=\frac{\nu}{z}+\frac{K_{\nu-1}(z)}{K_\nu(z)}$$

holds.

We conclude the present section by illustrating a method for summing Bessel series. Recall that if $F(z)$ is a meromorphic function with a finite number of poles then, under mild additional conditions, the series $\sum_{n=1}^{\infty} F(n)$ and $\sum_{n=1}^{\infty}(-1)^n F(n)$ can be summed by integrating $F(z)z\cot \pi z$ and $F(z)z \operatorname{cosec} z$ along a rectangular contour. The sequence $\pi, 2\pi, 3\pi, \cdots$ is the sequence of positive zeros of $J_{1/2}(z)$, since $J_{1/2}(z)$ is $\sqrt{2/(\pi z)}\sin z$. In potential and heat conduction problems, see Greenwood [14], the Bessel series $\sum_{n=1}^{\infty} F(j_{\nu,n})$ and $\sum_{n=1}^{\infty} F(j_{\nu,n})/J_{\nu-1}(j_{\nu,n})$ often arise. Observe that $J_{\nu-1}(j_{\nu,n})$ is $J'_\nu(j_{\nu,n})$; hence the sign of $J_{\nu-1}(j_{\nu,n})$ is $(-1)^n$. Therefore the aforementioned Bessel series can be summed by integrating $F(z)J_{\nu-1}(z)/J_\nu(z)$ or $z^\nu F(z)/J_\nu(z)$ around rectangles. Indeed one can prove the following general theorem by using the Stieltjes transform.

THEOREM 6.4. *If $F(z)$ is a single-valued entire function with the asymptotic behavior $F(z)=O(z^{-\nu/2-1/2}e^{\sqrt{z}})$, as $|z|\to\infty$, uniformly in every sector $|\arg z|\leqq \pi-\varepsilon$, $0<\varepsilon<\pi$, then*

$$\frac{z^{\nu/2}F(z)}{I_\nu(\sqrt{z})}=-2\sum_{n=1}^{\infty}\frac{j_{\nu,n}^{\nu+1}F(-j_{\nu,n}^2)}{J'_\nu(j_{\nu,n})(z+j_{\nu,n}^2)},\qquad \nu>-1. \tag{6.7}$$

We omit the proof of Theorem 6.4 because it is straightforward. Note that Theorem 6.4 can be extended to meromorphic functions $F(z)$ with a finite number of poles. This can be done by replacing $F(z)$ by $F(z)-G(z)$ where the rational function $G(z)$ is chosen such that $F-G$ is an entire single valued function, then express F as the sum of G and $F-G$ and apply Theorem 6.4 to $F-G$. One can easily state and prove an analogous theorem by replacing $z^{\nu/2}/I_\nu(\sqrt{z})$ in (6.7) by $z^{-1/2}I_{\nu+1}(\sqrt{z})/I_\nu(\sqrt{z})$.

7. Open problems. As we saw in § 5, for the F distribution, the density of f^ν is infinitely divisible when ν is at least one half of the degrees of freedom of the numerator. We conjecture that the following holds:

Problem 7.1. If X is a gamma or an F variable, then X^ν is infinitely divisible for all $\nu\geqq 1$.

In § 3 we saw how the infinite divisibility of the F distribution led us to study the complete monotonicity of $\psi(a+1, c+1, x)/\psi(a, c, x)$. This suggests

Problem 7.2. Prove the complete monotonicity of $\psi(a, c, x)/\psi(\alpha, \gamma, x)$ for $a-\alpha=c-\gamma>0$, $a-c+1>0$ and $\alpha-\gamma+1>0$.

Motivated by the existing proofs of the complete monotonicity of $x^{(\nu-\mu)/2}K_\nu(\sqrt{x})/K_\mu(\sqrt{x})$ one might try solving Problem 7.2 by either calculating the inverse Stieltjes transform of $\psi(a, c, x)/\psi(\alpha, \gamma, x)$ or show that the negative of the logarithmic derivative of that function is completely monotonic. The direct computation of the Stieltjes transform will establish the result if one can prove the positivity of the double integral

$$\int_0^\infty\int_0^{\pi/2} e^{-r^2}r^{2\mu+2}(\cos\theta)^{2\mu+1}(\sin\theta)^{2\mu}\{J_\nu(rt\sin\theta)Y_{\nu-1}(rt\cos\theta) - Y_\nu(rt\sin\theta)J_{\nu-1}(rt\cos\theta)\}r\,dr\,d\theta,$$

for all positive t, μ, ν. For fixed θ, the function

$$J_\nu(x\sin\theta)Y_{\nu-1}(x\cos\theta)-Y_\nu(x\sin\theta)J_{\nu-1}(x\cos\theta),$$

as a function of x, has infinitely many positive zeros. The second approach is to mimic the proof of Theorem 1.11 by using (1.4) and an analogue of Nicholson's formula for the ψ function. Unfortunately this analogue of Nicholson's formula is not known. Recently

Durand [6] and Durand et al. [7] derived a Nicholson's formula for the Gegenbauer functions. The next step is to study the Jacobi functions of the first and second kind and hopefully derive the corresponding Nicholson type formula for them. This will yield a similar formula for the confluent hypergeometric function as a limiting case.

Problem 7.3. Derive a formula of Nicholson type for the Jacobi functions or their limiting case, the confluent hypergeometric functions.

We might add that the addition theorem for the Jacobi polynomials might be helpful. This addition theorem was discovered by Koornwinder [23]–[25]. An account of the addition formulas is available in Askey [1].

The methods used in the present paper and in Ismail [19], [20] utilized the explicit form of the functions under consideration. These functions, however, are solutions of certain differential equations. It is very likely that a differential equations approach can be used to handle these problems. This suggests

Problem 7.4. Find a differential equation technique that guarantees the complete monotonicity of quotients of (nonoscillatory) solutions belonging to different eigenvalues of an eigenvalue problem.

In conclusion we would like to mention an open problem on the monotonicity of the quotient of the density $\rho_2(t, \nu, \mu)$ by its average on $(0, t)$.

Problem 7.5. The function $\rho_2(t, \nu, \mu)/\{t^{-1} \int_0^t \rho_2(s, \nu, \mu)\, ds\}$, $\mu > \nu \geqq 0$ is a decreasing function of t for $t > 0$.

The special case $\mu = 1$, $\nu = 0$ of Problem 7.5 was conjectured by Clements and Edelstein [5] and arose in their investigation of certain diffusion processes.

REFERENCES

[1] R. A. ASKEY, *Orthogonal Polynomials and Special Functions*, vol. 21, Society for Industrial and Applied Mathematics, Philadelphia, 1975.

[2] B. BERNDT AND M. L. GLASSER, *A new class of Bessel function integrals, Aequationes Math.*, 16 (1977), pp. 183–186.

[3] H. S. CARSLAW AND J. C. JAEGER, *Some problems in the mathematical theory of heat conduction*, Philos. Magazine, 26 (1938), pp. 473–495.

[4] ———, *Some two dimensional problems in conduction of heat with circular symmetry*, Proc. London Math. Soc., 46 (1940), pp. 361–388.

[5] J. C. CLEMENTS AND M. EDELSTEIN, *A new class of diffusion problems*, Appl. Math. Notes, 1 (1975), pp. 82–89.

[6] L. DURAND, *Nicholson-type integrals for products of Gegenbauer functions and related topics*, Theory and Applications of Special Functions, R. Askey, ed., Academic Press, New York, 1975, pp. 353–374.

[7] L. DURAND, P. M. FISHBANE AND L. M. SIMMONS, *Expansion formulas and addition theorems for Gegenbauer functions*, J. Math. Phys., 17 (1976), pp. 1933–1948.

[8] A. ERDÉLYI AND W. O. KERMACK, *Note on the equation* $f(z)K_n'(z) - g(z)K_n(z) = 0$, Proc. Cambridge Philos. Soc., 41 (1945), pp. 74–75.

[9] A. ERDÉLYI, W. MAGNUS, F. OBERHETTINGER AND F. G. TRICOMI, *Higher Transcendental Functions*, vol. 1, McGraw-Hill, New York, 1953.

[10] ———, *Higher Transcendental Functions*, vol. 2, McGraw-Hill, New York, 1953.

[11] ———, *Tables of Integral Transforms*, vol. 1, McGraw-Hill, New York, 1954.

[12] W. FELLER, *An Introduction to Probability Theory and its Applications*, vol. 2, John Wiley, New York, 1966.

[13] C. M. GOLDIE, *A class of infinitely divisible random variables*, Proc. Cambridge Philos. Soc., 63 (1967), pp. 1141–1143.

[14] J. A. GREENWOOD, *A simple method of summing Bessel series*, Ibid., 64 (1968), pp. 705–710.

[15] E. GROSSWALD, *The student t distribution of any degrees of freedom is infinitely divisible*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 36 (1976), pp. 103–109.

[16] P. HARTMAN, *Completely monotone families of solutions of nth order linear differential equations and infinitely divisible distributions*, Ann. Scoula Norm. Sup. Pisa, (4) 3 (1976), pp. 267–287.

[17] P. HARTMAN AND G. S. WATSON, "*Normal*" *distribution functions on spheres and modified Bessel functions*, Ann. Probability, 2 (1974), pp. 593–607.

[18] I. I. HIRSCHMAN AND D. V. WIDDER, *The Convolution Transform*, Princeton University Press, Prunceton, 1955.

[19] M. E. H. ISMAIL, *Bessel functions and the infinite divisibility of the student t distribution*, Ann. Probability, 5 (1977), pp. 582–585.

[20] ———, *Integral representations and complete monotonicity of various quotients of Bessel functions*, Canad. J. Math., 29 (1977), pp. 1198–1207.

[21] M. E. H. ISMAIL AND D. H. KELKER, *The Bessel polynomials and the Student t distribution*, this Journal, 7 (1976), pp. 82–91.

[22] D. H. KELKER, *Infinite divisibility and variance mixtures of the normal distribution*, Ann. Math. Statist., 42 (1971), pp. 802–808.

[23] T. KOORNWINDER, *The addition formula for Jacobi polynomials*, Summary of results, Indag. Math., 34 (1972), pp. 188–191.

[24] ———, *The addition formula for Jacobi polynomials and spherical harmonics*, SIAM J. Appl. Math., 25 (1973), pp. 236–246.

[25] ———, *Jacobi polynomials, II, An analytic proof of the product formula*, this Journal, 5 (1974), pp. 125–137.

[26] L. LORCH, *Inequalities for some Whittaker functions*, Arch. Math. (Brno), 3 (1967), pp. 1–9.

[27] F. OBERHETTINGER, *Tabellen zur Fourier Transformation*, Springer-Verlag, Berlin, 1957.

[28] F. W. J. OLVER, *Asymptotics and Special Functions*, Academic Press, New York, 1974.

[29] F. W. STEUTEL, *A class of infinitely divisible mixtures*, Ann. Math. Statist., 39 (1968), pp. 1153–1157.

[30] ———, *Some recent results in infinite divisibility*, Stochastic Processes Appl., 1 (1973), pp. 125–143.

[31] M. H. STONE, *Linear Transformations in Hilbert Spaces and their Applications to Analysis*, American Mathematical Society Colloquium Publications, American Mathematical Society, Providence, RI, 1932.

[32] F. G. TRICOMI, *Über die Abzählung der nullstellen der konfluenten hypergeometrischen Funktionen*, Math. Z., 52 (1950), pp. 669–675.

[33] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, Cambridge, 1944.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0003 $01.00/0

INEQUALITIES FOR ULTRASPHERICAL AND LAGUERRE POLYNOMIALS*

Dedicated to Professor Walter T. Scott

J. BUSTOZ† AND N. SAVAGE†

Abstract. R. Askey and G. Gasper have proved that

$$\sum_{k=0}^{n} \frac{(\lambda)_k}{k!} \frac{(\lambda)_{n-k}}{(n-k)!} \frac{\sin (k+1)\theta}{(k+1)\sin\theta} > 0, \qquad 0<\theta<\pi,$$

for $0 \leqq \lambda \leqq 2$ and Askey proved it for $\lambda = 3$. Here we prove this inequality for $2<\lambda<3$ by considering inequalities of the form $P_n^\alpha(x)P_{n+1}^\beta(x) - P_{n+1}^\alpha(x)P_n^\beta(x) > 0$, $0<x<1$, for the ultraspherical polynomials P_n^α. We also prove an inequality similar to the above for Laguerre polynomials.

1. Introduction. R. Askey and G. Gasper have proved [2] that the inequality,

$$\sum_{k=0}^{n} \frac{(\lambda)_k}{k!} \frac{(\lambda)_{n-k}}{(n-k)!} \frac{\sin (k+1)\theta}{(k+1)\sin\theta} > 0, \qquad 0<\theta<\pi, \tag{1.1}$$

holds for $0 \leqq \lambda \leqq 2$, and Askey also proved it for $\lambda = 3$ [1]. Here we will prove that (1.1) also holds for $2<\lambda<3$ and we will prove that (1.1) is a special case of a more general type of inequality valid for ultraspherical and Laguerre polynomials.

Denote the sum in (1.1) by $A_n(\theta)/\sin\theta$. Then

$$\sum_{k=0}^{\infty} A_k(\theta) z^{k+1} = \frac{(1-z)^{-\lambda}}{2i(\lambda-1)}\{(1-e^{i\theta}z)^{-\lambda+1} - (1-e^{-i\theta}z)^{-\lambda+1}\}.$$

If we define $B_k(\theta)$ by

$$\frac{(1-r)^{-\lambda}}{2i(\lambda-1)}(1-e^{i\theta}r)^{-\lambda+1} = \sum_{k=0}^{\infty} B_k(\theta) r^k,$$

then $\sum_{k=0}^{\infty} A_k(\theta) z^{k+1} = \sum_{k=0}^{\infty} [B_k(\theta) + \bar{B}_k(\theta)] z^k$ so that $A_k = 2 \operatorname{Re} B_{k+1}(\theta)$. Now for a suitable contour C,

$$2i(\lambda-1)B_n(\theta) = \frac{1}{2\pi i}\int_c \frac{(1-z)^{-\lambda}(1-e^{i\theta}z)^{-\lambda+1}}{z^{n+1}}\, dz.$$

If we set $w = z\, e^{i\theta/2}$ then

$$\begin{aligned} 2i(\lambda-1)B_n(\theta) &= \frac{e^{in\theta/2}}{2\pi i}\int_{c'} \frac{(1-2\cos(\theta/2)w + w^2)^{-\lambda}}{w^{n+1}}\, dw \\ &\quad - \frac{e^{i(n+1)\theta/2}}{2\pi i}\int_{c'} \frac{(1-2\cos(\theta/2)w + w^2)^{-\lambda}}{w^{n}}\, dw \\ &= P_n^\lambda(\cos(\theta/2))\, e^{in(\theta/2)} - P_{n-1}^\lambda(\cos(\theta/2))\, e^{i(n+1)\theta/2} \end{aligned}$$

where $P_n^\lambda(x)$ denotes the ultraspherical polynomial of degree n and we have used the generating relation

$$(1-2xz+z^2)^{-\lambda} = \sum_{k=0}^{\infty} P_k^\lambda(x) z^k.$$

* Received by the editors October 6, 1977, and in revised form March 7, 1978.

† Department of Mathematics, Arizona State University, Tempe, Arizona 85281.

Since $A_n(\theta)=2\,\mathrm{Re}\,B_{n+1}(\theta)$ we find

$$\frac{A_n(\theta)}{\sin\theta}=\frac{\sin[(n+1)(\theta/2)]P^\lambda_{n+1}(\cos(\theta/2))-\sin[(n+2)(\theta/2)]P^\lambda_n(\cos(\theta/2))}{(\lambda-1)\sin\theta}.$$

Then since $P^1_n(\cos\theta)=\sin[(n+1)\theta]/(\sin\theta)$ we have

$$\sum_{k=0}^{n}\frac{(\lambda)_k}{k!}\frac{(\lambda)_{n-k}}{(n-k)!}\frac{\sin[(k+1)\theta]}{(k+1)\sin\theta}$$

$$=\frac{P^1_n(\cos(\theta/2))P^\lambda_{n+1}(\cos(\theta/2))-P^1_{n+1}(\cos(\theta/2))P^\lambda_n(\cos(\theta/2))}{2(\lambda-1)\cos(\theta/2)}. \tag{1.2}$$

Inequality (1.1) and the identity (1.2) led us to look for inequalities of the form

$$\frac{P^\alpha_n(x)P^\beta_{n+1}(x)-P^\alpha_{n+1}(x)P^\beta_n(x)}{\beta-\alpha}>0,\qquad 0<x<1, \tag{1.3}$$

for ultraspherical polynomials, and

$$\frac{L^\alpha_n(x)L^\beta_{n+1}(x)-L^\alpha_{n+1}(x)L^\beta_n(x)}{\beta-\alpha}>0,\qquad 0<x<\infty, \tag{1.4}$$

for Laguerre polynomials.

In § 2 we will prove that (1.3) holds on the lines $\beta=\alpha+1$, $\alpha>0$ and $\beta=\alpha+2$, $\alpha>1/2$ (Theorems 2.1 and 2.3). Then we fill in the region between the lines $\alpha=\beta$ and $\beta=\alpha+2$ (Lemma 2.2 and Theorem 2.4). In § 3 we prove similar results for (1.4).

2. Ultraspherical polynomials. In this section we will need the following identities (see [6]). We will frequently suppress the independent variable and write P^λ_n for $P^\lambda_n(x)$.

$$(n+1)P^\lambda_{n+1}=2(n+\lambda)xP^\lambda_n-(n+2\lambda-1)P^\lambda_{n-1},\qquad n\geqq 1,\quad\text{where}$$
$$P^\lambda_0=1,\qquad P^\lambda_1=2\lambda x. \tag{2.1}$$

$$P^\lambda_n(1)=\binom{n+2\lambda-1}{n}. \tag{2.2}$$

$$(1-x^2)\frac{d}{dx}P^\lambda_n=(n+2\lambda-1)P^\lambda_{n-1}-nxP^\lambda_n,\qquad n\geqq 1. \tag{2.3}$$

$$\frac{d}{dx}P^\lambda_n=2\lambda P^{\lambda+1}_{n-1},\qquad n\geqq 1. \tag{2.4}$$

$$\lim_{\lambda\to 0}\frac{P^\lambda_n(x)}{\lambda}=\frac{2}{n}T_n(x),\qquad n\geqq 1,\quad\text{where}\quad T_n(\cos\theta)=\cos n\theta. \tag{2.5}$$

$$(1-x^2)^{\lambda-1/2}P^\lambda_n=\frac{(-2)^n}{n!}\frac{\Gamma(n+\lambda)\Gamma(n+2\lambda)}{\Gamma(\lambda)\Gamma(2n+2\lambda)}[(1-x^2)^{n+\lambda-1/2}]^{(n)}. \tag{2.6}$$

$$2\lambda(1-x^2)P^{\lambda+1}_n=(n+2\lambda)P^\lambda_n-(n+1)xP^\lambda_{n+1},\qquad n\geqq 1. \tag{2.7}$$

Define $\Delta_n(x;\alpha,\beta)=P^\alpha_nP^\beta_{n+1}-P^\alpha_{n+1}P^\beta_n$. Throughout this section we will make extensive use of a derivative relation for Δ_n. Computing $(d/dx)\Delta_n=\Delta'_n$ and using (2.3) and (2.1) a computation gives

$$(1-x^2)\,\Delta_n' = (2\alpha-1)x\,\Delta_n + 2(\beta-\alpha)P_n^\beta[P_n^\alpha - xP_{n+1}^\alpha]. \tag{2.8}$$

Equation (2.8) can be rewritten as

$$[(1-x^2)^{\alpha-1/2}\,\Delta_n]' = 2(\beta-\alpha)(1-x^2)^{\alpha-3/2}P_n^\beta[P_n^\alpha - xP_{n+1}^\alpha]. \tag{2.9}$$

An equivalent form of (2.9) which will be useful is obtained as follows. From (2.3) we have

$$\begin{aligned}(1-x^2)(P_{n+1}^\alpha)' &= (n+2\alpha)P_n^\alpha - (n+1)xP_{n+1}^\alpha \\ &= (n+2\alpha)[P_n^\alpha - xP_{n+1}^\alpha] + (2\alpha-1)xP_{n+1}^\alpha.\end{aligned} \tag{2.10}$$

Thus $(1-x^2)(P_{n+1}^\alpha)' - (2\alpha-1)xP_{n+1}^\alpha = (n+2\alpha)[P_n^\alpha - xP_{n+1}^\alpha]$, and this last can be written as

$$[(1-x^2)^{\alpha-1/2}P_{n+1}^\alpha]' = (n+2\alpha)(1-x^2)^{\alpha-3/2}[P_n^\alpha - xP_{n+1}^\alpha]. \tag{2.11}$$

Hence (2.9) can be rewritten as

$$[(1-x^2)^{\alpha-1/2}\,\Delta_n]' = \frac{2(\beta-\alpha)}{n+2\alpha}P_n^\beta[(1-x^2)^{\alpha-1/2}P_{n+1}^\alpha]'. \tag{2.12}$$

If we define F_n^α by

$$F_n^\alpha = (1-x^2)^{\alpha-1/2}P_n^\alpha, \tag{2.13}$$

and we define $\delta_n(x;\alpha,\beta)$ by

$$\delta_n = F_n^\alpha F_{n+1}^\beta - F_{n+1}^\alpha F_n^\beta, \tag{2.14}$$

then we note that F_n^α has the same zeros in $(-1, 1)$ as P_n^α and also $\operatorname{Sgn} P_n^\alpha = \operatorname{Sgn} F_n^\alpha$. Further, since

$$\delta_n = (1-x^2)^{\alpha+\beta-1}\,\Delta_n \tag{2.15}$$

we also have that $\operatorname{Sgn}\delta_n = \operatorname{Sgn}\Delta_n$ in $-1<x<1$. Equation (2.12) can then be expressed as

$$[(1-x^2)^{-\beta+1/2}\delta_n]' = \frac{2(\beta-\alpha)}{n+2\alpha}(1-x^2)^{-\beta+1/2}F_n^\beta(F_{n+1}^\alpha)'. \tag{2.16}$$

We get from (2.6) that

$$(F_{n+1}^\alpha)' = \frac{-(n+2)}{2(\alpha-1)}(n+2\alpha)F_{n+2}^{\alpha-1}. \tag{2.17}$$

Thus (2.16) can be written as

$$[(1-x^2)^{-\beta+1/2}\,\delta_n]' = -\frac{(\beta-\alpha)(n+2)}{(\alpha-1)}(1-x^2)^{-\beta+1/2}F_n^\beta F_{n+2}^{\alpha-1}. \tag{2.18}$$

We will need some identities involving the functions F_n^α. From (2.1) we get

$$(n+1)F_{n+1}^\alpha = 2(n+\alpha)xF_n^\alpha - (n+2\alpha-1)F_{n-1}^\alpha. \tag{2.19}$$

From (2.11) we get

$$(1-x^2)(F_{n+1}^\alpha)' = (n+2\alpha)[F_n^\alpha - xF_{n+1}^\alpha]. \tag{2.20}$$

Computing $(1-x^2)(F_n^\alpha)'$ directly and using (2.4) we get

$$(1-x^2)(F_n^\alpha)' = 2\alpha F_{n-1}^{\alpha+1} - (2\alpha-1)xF_n^\alpha. \tag{2.21}$$

Finally from (2.7) we get

$$2\alpha F_n^{\alpha+1} = (n+2\alpha)F_n^\alpha - (n+1)xF_{n+1}^\alpha. \tag{2.22}$$

We will say that the zeros of two polynomials $A(x)$ and $B(x)$ *interlace* if between any two consecutive zeros of $A(x)$ there is precisely one zero of $B(x)$ and vice versa.

LEMMA 2.1. *If $\Delta_n(x;\alpha,\beta)\neq 0$ for $0<x<1$ then the zeros of P_n^α and P_n^β interlace.*

Proof. Let $x_1<x_2$ be consecutive zeros of P_n^α. Suppose $P_n^\beta\neq 0$ in $[x_1, x_2]$. Then

$$\Delta_n(x_1;\alpha,\beta) = -P_{n+1}^\alpha(x_1)P_n^\beta(x_1),$$
$$\Delta_n(x_2;\alpha,\beta) = -P_{n+1}^\alpha(x_2)P_n^\beta(x_2).$$

Since the zeros of P_n^α and P_{n+1}^α interlace, $P_{n+1}^\alpha(x_1)$ and $P_{n+1}^\alpha(x_2)$ have opposite signs. Hence $\Delta_n(x_1;\alpha,\beta)$ and $\Delta_n(x_2;\alpha,\beta)$ have opposite signs.

We will need to know the value of $\Delta_n(x;\alpha,\beta)$ when $x=0$ and $x=1$. Since $P_n^\lambda(0)=0$ when n is odd, it follows that $\Delta_n(0;\alpha,\beta)=0$. Using (2.2) we can compute

$$\Delta_n(1;\alpha,\beta) = \frac{2(\beta-\alpha)\Gamma(n+2\alpha)\Gamma(n+2\beta)}{\Gamma(n+1)\Gamma(n+2)\Gamma(2\alpha)\Gamma(2\beta)}.$$

Thus $\Delta_n(1;\alpha,\beta)>0$ for $\beta>\alpha>0$, and $\Delta_n(1;\alpha,\beta)<0$ for $-1/2<\alpha<0$, $\beta>0$. The values of $\Delta_n(x;\alpha,\beta)$ at $x=0$, $x=1$ will be important because we will be proving positivity of $\Delta_n(x;\alpha,\beta)$ in $0<x<1$ by proving that $(1-x^2)^{\alpha-1/2}\Delta_n$ has positive extrema in $0<x<1$.

THEOREM 2.1. *If $\alpha>0$ then $\Delta_n(x;\alpha,\alpha+1)>0$ for $0<x<1$. If $-\frac{1}{2}<\alpha<0$ then $\Delta_n(x;\alpha,\alpha+1)<0$ for $0<x<1$.*

Proof. Set $\beta=\alpha+1$ in (2.9) and get

$$[(1-x^2)^{\alpha-1/2}\Delta_n]' = 2(1-x^2)^{\alpha-3/2}P_n^{\alpha+1}[P_n^\alpha - xP_{n+1}^\alpha]. \tag{2.23}$$

We will show that the relative extrema of $(1-x^2)^{\alpha-1/2}\Delta_n$ are all positive if $\alpha>0$ and all negative if $-\frac{1}{2}<\alpha<0$. The points of relative extrema occur when $P_n^{\alpha+1}=0$ or when $P_n^\alpha = xP_{n+1}^\alpha$. This gives two cases. From (2.7) we have

$$2\alpha(1-x^2)P_{n+1}^{\alpha+1} = (n+2\alpha+1)P_{n+1}^\alpha - (n+2)xP_{n+2}^\alpha \tag{2.24}$$

and from (2.1)

$$(n+2)P_{n+2}^\alpha = 2(n+\alpha+1)xP_{n+1}^\alpha - (n+2\alpha)P_n^\alpha. \tag{2.25}$$

Substitution of (2.25) in (2.24) gives

$$2\alpha(1-x^2)P_{n+1}^{\alpha+1} = [(n+2\alpha+1)-2(n+\alpha+1)x^2]P_{n+1}^\alpha + (n+2\alpha)xP_n^\alpha. \tag{2.26}$$

We proceed with consideration of the two cases.

Case 1. In this case $P_n^{\alpha+1}(x)=0$, and we have $\Delta_n = P_n^\alpha P_{n+1}^{\alpha+1}$. Since $P_n^{\alpha+1}=0$ we get from (2.7) that

$$(n+2\alpha)P_n^\alpha = (n+1)xP_{n+1}^\alpha. \tag{2.27}$$

From (2.26) we get, using (2.27),

$$\begin{aligned}2\alpha(1-x^2)P_n^\alpha P_{n+1}^{\alpha+1} &= [(n+2\alpha+1)-2(n+\alpha+1)x^2]P_{n+1}^\alpha P_n^\alpha + (n+2\alpha)x(P_n^\alpha)^2 \\ &= \frac{(n+2\alpha)}{(n+1)x}[(n+2\alpha+1)-2(n+\alpha+1)x^2](P_n^\alpha)^2 \\ &\qquad + (n+2\alpha)x(P_n^\alpha)^2.\end{aligned} \tag{2.28}$$

Thus

$$\begin{aligned}&2\alpha x(1-x^2)P_n^\alpha P_{n+1}^{\alpha+1} \\ &= \frac{\{[(n+2\alpha+1)-2(n+\alpha+1)x^2](n+2\alpha)+(n+2\alpha)(n+1)x^2\}(P_n^\alpha)^2}{n+1} \\ &= \frac{(n+2\alpha)}{n+1}(n+2\alpha+1)(1-x^2)(P_n^\alpha)^2.\end{aligned} \tag{2.29}$$

Equation (2.29) gives

$$\Delta_n = \frac{(n+2\alpha)(n+2\alpha+1)}{2\alpha(n+1)x}(P_n^\alpha)^2. \tag{2.30}$$

From (2.30) we see that at a point where $P_n^{\alpha+1}=0$ we have $\Delta_n>0$ provided $\alpha>0$ and $0<x<1$, while if $\alpha<0$ and $0<x<1$ then $\Delta_n<0$.

Case 2. In this case we have $P_n^\alpha = xP_{n+1}^\alpha$ and Δ_n becomes $\Delta_n = P_{n+1}^\alpha[xP_{n+1}^{\alpha+1} - P_n^{\alpha+1}]$. From (2.26) we obtain

$$2\alpha x(1-x^2)P_{n+1}^{\alpha+1} = [(n+2\alpha+1)-2(n+\alpha+1)x^2]P_n^\alpha + (n+2\alpha)x^2P_n^\alpha, \tag{2.31}$$

and (2.7) becomes

$$2\alpha(1-x^2)P_n^{\alpha+1} = (2\alpha-1)P_n^\alpha. \tag{2.32}$$

Hence we have

$$\begin{aligned}&2\alpha(1-x^2)[xP_{n+1}^{\alpha+1} - P_n^{\alpha+1}] \\ &= [(n+2\alpha+1)-2(n+\alpha+1)x^2+(n+2\alpha)x^2-(2\alpha-1)]P_n^\alpha \\ &= (n+2)(1-x^2)P_n^\alpha.\end{aligned} \tag{2.33}$$

Consequently when $P_n^\alpha = xP_{n+1}^\alpha$ we have

$$\Delta_n = \frac{(n+2)x(P_{n+1}^\alpha)^2}{2\alpha}. \tag{2.34}$$

Thus $\Delta_n>0$ if $\alpha>0$ and $0<x<1$. If $-\frac{1}{2}<\alpha<0$ and $0<x<1$ then $\Delta_n<0$. This completes the proof.

Theorem 2.1 has a limiting case as $\alpha\to 0$. Since $P_n^\alpha \equiv 0$ for $n \geqq 1$, when $\alpha=0$ we need to look at $\Delta_n(x;\alpha,\alpha+1)/\alpha$. Using (2.25) we find

$$\lim_{\alpha\to 0}\frac{\Delta_n(x;\alpha,\alpha+1)}{\alpha} = \frac{2}{n}T_nP_{n+1}^1 - \frac{2}{n+1}T_{n+1}P_n^1.$$

Denote the limit above by Δ_n^*. From (2.23) we find

$$[(1-x^2)^{-1/2}\Delta_n^*]' = 4(1-x^2)^{-3/2}P_n^1\left[\frac{1}{n}T_n - \frac{x}{n+1}T_{n+1}\right]. \tag{2.35}$$

Proceeding as in the proof of Lemma 2.2, we find that if $P_n^1=0$ then $\Delta_n^*=2(T_n)^2/(nx)$, and if $T_n/n=xT_{n+1}/(n+1)$ then $\Delta_n^*=(2(n+2)/(n+1)^2)x(T_{n+1})^2$. Thus $\Delta_n^*>0$ for $0<x<1$.

We can include this limiting case in the statement of Theorem 2.1 if we restate it as

THEOREM 2.2. *If* $\alpha>-\frac{1}{2}$ *and* $0<x<1$ *then* $\Delta_n(x;\alpha,\alpha+1)/\alpha>0$.

As a consequence of Theorem 2.2 and $\Delta_n(x;\alpha,\beta)=-\Delta_n(x;\beta,\alpha)$ we have

COROLLARY 2.2. *If* $\alpha>\frac{1}{2}$ *and* $0<x<1$ *then* $\Delta_n(x;\alpha,\alpha-1)/(\alpha-1)<0$.

THEOREM 2.3. $\Delta_n(x;\alpha,\alpha+2)>0$ *for* $\alpha\geqq\frac{1}{2}$, $0<x<1$.

Proof. If we set $\beta=\alpha+2$ *in* (2.9) *we get*

$$[(1-x^2)^{\alpha-1/2}\Delta_n]'=4(1-x^2)^{\alpha-3/2}P_n^{\alpha+2}[P_n^\alpha-xP_{n+1}^\alpha]. \tag{2.35a}$$

From (2.26) we get

$$2(\alpha+1)(1-x^2)P_{n+1}^{\alpha+2}=[(n+2\alpha+3)-2(n+\alpha+2)x^2]P_{n+1}^{\alpha+1}+(n+2\alpha+2)xP_n^{\alpha+1}. \tag{2.36}$$

From (2.7) we get

$$2(\alpha+1)(1-x^2)P_n^{\alpha+2}=(n+2\alpha+2)P_n^{\alpha+1}-(n+1)xP_{n+1}^{\alpha+1}. \tag{2.37}$$

Now consider the two cases under which the right side of (2.35a) vanishes in (0, 1).

Case 1. $P_n^{\alpha+2}=0$.

In this case $\Delta_n=P_n^\alpha P_{n+1}^{\alpha+2}$. From (2.37) we have

$$(n+2\alpha+2)P_n^{\alpha+1}=(n+1)xP_{n+1}^{\alpha+1}. \tag{2.38}$$

Substituting for $P_n^{\alpha+1}$ from (2.7) and for $P_{n+1}^{\alpha+1}$ from (2.26) after multiplying both sides by $2\alpha(1-x^2)$, we get

$$\begin{aligned}(n+2\alpha+2)&[(n+2\alpha)P_n^\alpha-(n+1)xP_{n+1}^\alpha]\\&=(n+1)x\{[(n+2\alpha+1)-2(n+\alpha+1)x^2]P_{n+1}^\alpha\\&\qquad+(n+2\alpha)xP_n^\alpha\}.\end{aligned} \tag{2.39}$$

Equation (2.39) reduces to

$$\begin{aligned}[(n+2\alpha+2)&(n+2\alpha)-(n+1)(n+2\alpha)x^2]P_n^\alpha\\&=\{(n+1)x[(n+2\alpha+1)-2(n+\alpha+1)x^2]\\&\qquad+(n+2\alpha+2)(n+1)x\}P_{n+1}^\alpha.\end{aligned} \tag{2.40}$$

If we substitute (2.38) into (2.36) we get

$$\begin{aligned}2(\alpha+1)&(1-x^2)P_{n+1}^{\alpha+2}\\&=\frac{(n+2\alpha+2)}{(n+1)x}[(n+2\alpha+3)-2(n+\alpha+2)x^2]P_n^{\alpha+1}\\&\qquad+(n+2\alpha+2)xP_n^{\alpha+1},\end{aligned} \tag{2.41}$$

or, equivalently,

$$\frac{2(\alpha+1)(n+1)x}{(n+2\alpha+2)(n+2\alpha+3)}P_{n+1}^{\alpha+2}=P_n^{\alpha+1}. \tag{2.42}$$

From (2.7), substituting for P_{n+1}^α from (2.40), we get

$$\frac{2\alpha(1-x^2)}{n+2\alpha}P_n^{\alpha+1}=\frac{(n+2\alpha+1)(1-x^2)P_n^\alpha}{2n(1-x^2)+2\alpha(2-x^2)+2(\frac{3}{2}-x^2)}. \tag{2.43}$$

Hence,

$$(2.44)\quad \Delta_n = P_n^{\alpha}P_{n+1}^{\alpha+2} = \frac{(n+2\alpha)(n+2\alpha+1)(n+2\alpha+2)(n+2\alpha+3)(P_n^{\alpha})^2}{4\alpha(\alpha+1)(n+1)x[2n(1-x^2)+2\alpha(2-x^2)+2(\frac{3}{2}-x^2)]}$$

which is positive for $\alpha>0$ and $0<x<1$. Now we turn to the other critical points of (2.35).

Case 2. $P_n^{\alpha} = xP_{n+1}^{\alpha}$.

In this case $\Delta_n = P_{n+1}^{\alpha}[xP_{n+1}^{\alpha+2} - P_n^{\alpha+2}]$. From (2.26) using $P_n^{\alpha} = xP_{n+1}^{\alpha}$ we get

$$(2.45)\qquad P_{n+1}^{\alpha+1} = \frac{[(n+2\alpha+1)-(n+2)x^2]}{2\alpha x(1-x^2)}P_n^{\alpha}$$

and from (2.7) we get

$$(2.46)\qquad P_n^{\alpha+1} = \frac{(2\alpha-1)P_n^{\alpha}}{2\alpha(1-x^2)}.$$

Substituting (2.45) and (2.46) into (2.36) we get

$$(2.47)\qquad \begin{aligned} &4\alpha(\alpha+1)(1-x^2)^2xP_{n+1}^{\alpha+2} \\ &\quad = \{[(n+2\alpha+3)-2(n+\alpha+2)x^2][(n+2\alpha+1)-(n+2)x^2] \\ &\qquad\qquad +(2\alpha-1)(n+2\alpha+2)x^2\}P_n^{\alpha}. \end{aligned}$$

Now from (2.37) substituting (2.45) and (2.46) we get

$$(2.48)\qquad \begin{aligned} 4\alpha(\alpha+1)(1-x^2)^2P_n^{\alpha+2} &= \{(2\alpha-1)(n+2\alpha+2) \\ &\quad -(n+1)[(n+2\alpha+1)-(n+2)x^2]\}P_n^{\alpha}. \end{aligned}$$

Subtracting (2.48) from (2.47) we get, after simplifying,

$$(2.49)\qquad \begin{aligned} &4\alpha(\alpha+1)(1-x^2)[xP_{n+1}^{\alpha+2} - P_n^{\alpha+2}] \\ &\quad = \{2(n+\alpha+2)[(n+2\alpha+1)-(n+2)x^2] \\ &\qquad\qquad -(2\alpha-1)(n+2\alpha+2)\}P_n^{\alpha}. \end{aligned}$$

Now $\Delta_n = x^{-1}P_n^{\alpha}[xP_{n+1}^{\alpha+2} - P_n^{\alpha+2}]$ so from (2.49) we have

$$(2.50)\quad \Delta_n = \frac{2(n+\alpha+2)[(n+2\alpha+1)-(n+2)x^2]-(2\alpha-1)(n+2\alpha+2)}{4\alpha(\alpha+1)x(1-x^2)}(P_n^{\alpha})^2.$$

Set $R_n(x) = 2(n+\alpha+2)[(n+2\alpha+1)-(n+2)x^2]-(2\alpha-1)(n+2\alpha+2)$. Note that $R_n(1) = (2\alpha-1)(n+2) \geqq 0$ for $\alpha \geqq \frac{1}{2}$ and $R_n(0) = 2n^2+(4\alpha+7)n+8\alpha+6>0$ for $\alpha \geqq \frac{1}{2}$. Thus $R_n(x)>0$ for $\alpha \geqq \frac{1}{2}$, $0<x<1$ and so is Δ_n. Thus the function $(1-x^2)^{\alpha-1/2}\Delta_n(x;\alpha,\alpha+2)$ has positive extrema for $\alpha \geqq \frac{1}{2}$, $0<x<1$; and the proof is complete.

COROLLARY 2.3. *If* $\alpha \geqq \frac{5}{2}$ *then* $\Delta_n(x;\alpha,\alpha-2)<0$ *for* $0<x<1$.

LEMMA 2.2. *If* $\frac{1}{2}<\alpha<\beta<\gamma$ *and* $\Delta_n(x;\alpha,\gamma)>0$ *for* $0<x<1$ *then* $\Delta_n(x;\alpha,\beta)>0$ *for* $0<x<1$.

Proof. From Lemma 2.1 and from the fact that the positive zeros of P_n^{λ} are monotone decreasing functions of λ, it follows that the zeros of P_n^{α} and P_n^{γ} are interlaced as are the zeros of P_n^{α} and P_n^{β}. We consider

$$[(1-x^2)^{-\beta+1/2}\,\delta_n]' = \frac{2(\beta-\alpha)}{n+2\alpha}(1-x^2)^{-\beta+1/2}F_n^{\beta}(F_{n+1}^{\alpha})'.$$

Case 1. Suppose $F_n^{\beta}=0$. Then $\delta_n = F_n^{\alpha}F_{n+1}^{\beta}$. Let $w_1>w_2>\cdots$ be the zeros of F_n^{β}

in $(0, 1)$. Then $\operatorname{Sgn} F^{\beta}_{n+1}(w_j)=(-1)^j$. The interlacing of zeros for F^{α}_n and F^{β}_n give $\operatorname{Sgn} F^{\alpha}_n(w_j)=(-1)^j$. Hence $\delta_n(w_j; \alpha, \beta)>0$.

Case 2. Suppose $(F^{\alpha}_{n+1})'=0$. Let $w_1>w_2>\cdots$ be the zeros of $(F^{\alpha}_{n+1})'$ in $(0, 1)$. From (2.20) we find

$$\delta_n(w_j; \alpha, \beta)=\frac{n+2}{2}(1-w_j^2)F^{\alpha}_{n+1}(w_j)\frac{F^{\beta-1}_{n+2}(w_j)}{\beta-1}.$$

We have $\operatorname{Sgn} F^{\alpha}_{n+1}(w_j)=(-1)^{j+1}$. From (2.17) we see that the zeros of $(F^{\alpha}_{n+1})'$ are the zeros of $F^{\alpha-1}_{n+2}$. Since $\Delta_n(x; \alpha, \gamma)>0$ we have that $\operatorname{Sgn}(F^{\gamma-1}_{n+2}(w_j)/(\gamma-1))=(-1)^{j+1}$. Let $z_1>z_2>\cdots$ be the zeros of $F^{\gamma-1}_{n+2}$ in $(0, 1)$ and $y_1>y_2>\cdots$ the zeros of $F^{\beta-1}_{n+2}$ in $(0, 1)$. Then we have $w_{k+1}<z_k<y_k<w_k$. Also, $F^{\beta-1}_{n+2}(x)/(\beta-1)>0$ for x in a left neighborhood of 1. Hence $\operatorname{Sgn}(F^{\beta-1}_{n+2}(w_j)/(\beta-1))=(-1)^{j+1}$ and $\delta_n(w_j; \alpha, \beta)>0$. This completes the proof.

THEOREM 2.4. $\Delta_n(x; \alpha, \beta)/(\beta-\alpha)>0$ *for* $0<x<1$ *if* (α, β) *lies in either of the regions*

(i) $\frac{1}{2}<\alpha<\beta\leqq\alpha+2$

(ii) $\frac{1}{2}<\beta<\alpha\leqq\beta+2$.

Proof. The case for region (i) follows immediately from Theorem 2.3 and Lemma 2.2 with $\gamma=\alpha+2$. Then (ii) follows from (i) by noting that $\Delta_n(x; \alpha, \beta)=-\Delta_n(x; \beta, \alpha)$.

Theorem 2.4 gives the missing values $2<\lambda<3$ for inequality (1.1). We need only note that the line segment $\alpha=1, 2<\beta<3$, is contained in the region (i). Thus we have

COROLLARY 2.4. *If* $2<\lambda<3$ *then*

$$\sum_{k=0}^{n}\frac{(\lambda)_k}{k!}\frac{(\lambda)_{n-k}}{(n-k)!}\frac{\sin(k+1)\theta}{(k+1)\sin\theta}>0, \qquad 0<\theta<\pi.$$

It would be interesting to know the exact region in the (α, β)-plane for which $\Delta_n(x; \alpha, \beta)\neq 0$, $0<x<1$. We note that our results do not for example include inequality (1.1) for $0<\lambda\leqq\frac{1}{2}$.

3. Generalized Laguerre polynomials. In this section we prove theorems similar to those of § 2 but for generalized Laguerre polynomials. These are defined by

$$(3.1)\qquad (n+1)L^{\alpha}_{n+1}=(2n+\alpha+1-x)L^{\alpha}_n-(n+\alpha)L^{\alpha}_{n-1}, \qquad n\geqq 1,$$

where $L^{\alpha}_0=1$, $L^{\alpha}_1=1+\alpha-x$, and also

$$(3.2)\qquad L^{\alpha}_n(0)=\binom{n+\alpha}{n}.$$

We will need the following identities.

$$(3.3)\qquad x\frac{d}{dx}L^{\alpha}_n=nL^{\alpha}_n-(n+\alpha)L^{\alpha}_{n-1}, \qquad \alpha>-1, \quad n\geqq 1,$$

$$(3.4)\qquad \frac{d}{dx}L^{\alpha}_n=-L^{\alpha+1}_{n-1},$$

$$(3.5)\qquad L^{\alpha}_{n+1}=L^{\alpha+1}_{n+1}-L^{\alpha+1}_n.$$

Set $D_n(x; \alpha, \beta)=L^{\alpha}_nL^{\beta}_{n+1}-L^{\alpha}_{n+1}L^{\beta}_n$. Lemma 2.1 holds for Laguerre polynomials so we have

LEMMA 3.1. *If* $D_n(x; \alpha, \beta)\neq 0$ *then the zeros of* L^{α}_n *and* L^{β}_n *interlace.*

There is a derivative formula for D_n analogous to (2.9) for Δ_n. It is

$$(3.6)\qquad [e^{-x}x^{\beta}D_n]'=(\beta-\alpha)e^{-x}x^{\beta-1}L^{\alpha}_nL^{\beta-1}_{n+1}.$$

THEOREM 3.1. *$D_n(x;\alpha,\alpha+1)>0$ for $x>0$ if $\alpha>-1$, and also*

$$D_n(x;\alpha,\alpha-1)<0 \quad \text{for } x>0 \text{ if } \alpha>0.$$

Proof. For contrast we give a different type of argument than was used in Theorem 2.1 although a proof along the lines of Theorem 2.1 can be constructed.

Setting $\beta=\alpha+1$ in (3.6) we find

$$[e^{-x}x^{\alpha+1}D_n]'=e^{-x}x^{\alpha}L_n^{\alpha}L_{n+1}^{\alpha}.$$

We observe that the function $e^{-x}x^{\alpha+1}D_n$ vanishes at $x=0$ and at $x=+\infty$. Thus it will be sufficient to show that $D_n>0$ at the relative extrema of $e^{-x}x^{\alpha+1}D_n$. These occur either where $L_n^{\alpha}=0$ or $L_{n+1}^{\alpha}=0$. Let $x_1<x_2<\cdots<x_n$ be the zeros of L_n^{α}. If $L_n^{\alpha}=0$ then $D_n=-L_{n+1}^{\alpha}L_n^{\alpha+1}$. The zeros of L_n^{α} and L_{n+1}^{α} interlace so Sgn $L_{n+1}^{\alpha}(x_j)=(-1)^j$. From the identity $L_n^{\alpha+1}=L_n^{\alpha}-(L_n^{\alpha})'$ which follows from (3.4) and (3.5) we get $L_n^{\alpha+1}(x_j)=-(L_n^{\alpha})'(x_j)$. Since Sgn $(L_n^{\alpha})'(x_j)=(-1)^j$ we have Sgn $L_n^{\alpha+1}(x_j)=(-1)^{j+1}$ and hence $D_n(x_j;\alpha,\alpha+1)>0$. Next let $z_1<z_2<\cdots<z_{n+1}$ be the zeros of L_{n+1}^{α}. Then $D_n(z_j;\alpha,\alpha+1)=L_n^{\alpha}(z_j)L_{n+1}^{\alpha+1}(z_j)$. Now Sgn $L_n^{\alpha}(z_j)=(-1)^{j+1}$ and the identity $L_{n+1}^{\alpha+1}=L_{n+1}^{\alpha}-(L_{n+1}^{\alpha})'$ implies that Sgn $L_{n+1}^{\alpha+1}(z_j)=(-1)^{j+1}$. Hence $D_n(z_j;\alpha,\alpha+1)>0$. We conclude that $D_n(x;\alpha,\alpha+1)>0$ for $\alpha>-1$, $x>0$. To prove the second part, note that $D_n(x;\alpha,\beta)=-D_n(x;\beta,\alpha)$, and the proof goes as in the proof of Corollary 2.2.

THEOREM 3.2. *$D_n(x;\alpha,\alpha+2)>0$ for $x>0$ if $\alpha>0$, and also*

$$D_n(x;\alpha,\alpha-2)>0 \quad \text{for } x>0 \text{ if } \alpha>2.$$

Proof. First we observe that since $D_n(x;\alpha,\beta)=-D_n(x;\beta,\alpha)$ (3.6) can be written as

$$[e^{-x}x^{\alpha}D_n(x;\alpha,\beta)]'=(\beta-\alpha)\,e^{-x}x^{\alpha-1}L_n^{\beta}L_{n+1}^{\alpha-1}.$$

Setting $\beta=\alpha+2$ we have

$$[e^{-x}x^{\alpha}D_n]'=2e^{-x}x^{\alpha-1}L_n^{\alpha+2}L_{n+1}^{\alpha-1},$$

and using (3.5), which implies $L_{n+1}^{\alpha-1}=L_{n+1}^{\alpha}-L_n^{\alpha}$, we get

$$[e^{-x}x^{\alpha}D_n]'=2e^{-x}x^{\alpha-1}L_n^{\alpha+2}[L_{n+1}^{\alpha}-L_n^{\alpha}].$$

The relative extrema of $e^{-x}x^{\alpha}D_n$ occur when either $L_n^{\alpha+2}=0$ or $L_{n+1}^{\alpha}=L_n^{\alpha}$. First we establish some identities which will be useful. From (3.3) and (3.4) we get

$$xL_n^{\alpha+1}=(n+\alpha+1)L_n^{\alpha}-(n+1)L_{n+1}^{\alpha}. \tag{3.7}$$

From (3.1) and (3.7) we get

$$xL_{n+1}^{\alpha+1}=(n+\alpha+1)L_n^{\alpha}-(n+1-x)L_{n+1}^{\alpha}. \tag{3.8}$$

Case 1. If $L_n^{\alpha+2}=0$ then $D_n=L_n^{\alpha}L_{n+1}^{\alpha+2}$ and (3.7) with α replaced by $\alpha+1$ gives

$$(n+\alpha+2)L_n^{\alpha+1}=(n+1)L_{n+1}^{\alpha+1}. \tag{3.9}$$

Multiplying (3.9) by x and substituting for $L_n^{\alpha+1}$ from (3.7) and for $L_{n+1}^{\alpha+1}$ from (3.8) gives

$$(n+\alpha+2)[(n+\alpha+1)L_n^{\alpha}-(n+1)L_{n+1}^{\alpha}]=(n+1)[(n+\alpha+1)L_n^{\alpha}-(n+1-x)L_{n+1}^{\alpha}].$$

This reduces to

$$(n+\alpha+1)(\alpha+1)L_n^{\alpha}=(n+1)(x+\alpha+1)L_{n+1}^{\alpha}. \tag{3.10}$$

Replacing α by $\alpha+1$ in (3.8) and then substituting for $L_{n+1}^{\alpha+1}$ from (3.9) gives

$$(n+1)L_{n+1}^{\alpha+2}=(n+\alpha+2)L_n^{\alpha+1}. \tag{3.11}$$

Substituting in (3.7) for L_{n+1}^{α} from (3.10) gives

$$(x+\alpha+1)L_n^{\alpha+1}=(n+\alpha+1)L_n^{\alpha}. \tag{3.12}$$

Hence $D_n=((n+\alpha+2)(n+\alpha+1)/(n+1)(x+\alpha+1))[L_n^{\alpha}]^2$. Thus $D_n>0$ for $x>0$ and $\alpha>-1$ at the points where $L_n^{\alpha+2}=0$.

Case 2. If $L_{n+1}^{\alpha}=L_n^{\alpha}$ then $D_n=L_{n+1}^{\alpha}[L_{n+1}^{\alpha+2}-L_n^{\alpha+2}]$ and from (3.8) we get

$$xL_{n+1}^{\alpha+1}=(\alpha+x)L_n^{\alpha}. \tag{3.13}$$

From (3.7) we get

$$xL_n^{\alpha+1}=\alpha L_n^{\alpha}. \tag{3.14}$$

If we replace α by $\alpha+1$ in (3.8) and use (3.13) and (3.14) we find from (3.8)

$$x^2L_{n+1}^{\alpha+2}=[\alpha(n+\alpha+2)-(\alpha+x)(n+1-x)]L_n^{\alpha} \tag{3.15}$$

Similarly from (3.7) we obtain

$$x^2L_n^{\alpha+2}=[\alpha(n+\alpha+2)-(\alpha+x)(n+1)]L_n^{\alpha}. \tag{3.16}$$

Hence

$$x[L_{n+1}^{\alpha+2}-L_n^{\alpha+2}]=(x+\alpha)L_n^{\alpha}. \tag{3.17}$$

Thus $D_n=((x+\alpha)/x)(L_n^{\alpha})^2$. Hence $D_n>0$ for $x>0$ and $\alpha>0$ at the points where $L_{n+1}^{\alpha}=L_n^{\alpha}$, and we conclude the proof of the first part of the lemma. The second part follows as in the proof of Corollary 2.3.

The proofs of the next lemma and theorem are similar to the proofs of Lemma 2.2 and Theorem 2.4 and consequently are omitted. An important difference is that the zeros of L_n^{α} are monotone increasing functions of α.

LEMMA 3.2. *If* $0<\alpha<\beta<\gamma$ *and* $D_n(x;\alpha,\gamma)>0$ *for* $x>0$ *then* $D_n(x;\alpha,\beta)$ *for* $x>0$.

THEOREM 3.3. $D_n(x;\alpha,\beta)/(\beta-\alpha)>0$ *for* $x>0$ *in the regions*:

(i) $0<\alpha<\beta\leqq\alpha+2$,

(ii) $0<\beta<\alpha\leqq\beta+2$.

4. Remarks. Differential identities similar to (2.9) were used by O. Szász [5] and by G. Gasper [3] in proving inequalities of Turan type. We observe that the inequalities of §§ 2 and 3 can be written in the determinant forms

$$\frac{\Delta_n(x;\alpha,\beta)}{\beta-\alpha}=\frac{1}{\beta-\alpha}\begin{vmatrix}P_n^{\alpha} & P_{n+1}^{\alpha}\\ P_n^{\beta} & P_{n+1}^{\beta}\end{vmatrix}>0, \tag{4.1}$$

$$\frac{D_n(x;\alpha,\beta)}{\beta-\alpha}=\frac{1}{\beta-\alpha}\begin{vmatrix}L_n^{\alpha} & L_{n+1}^{\alpha}\\ L_n^{\beta} & L_{n+1}^{\beta}\end{vmatrix}>0. \tag{4.2}$$

This suggests the possibility of generalizing inequalities (4.1) and (4.2) to $n\times n$ determinants as was done by S. Karlin and G. Szegö [4] for determinants of Turan type.

REFERENCES

[1] R. ASKEY, *Some absolutely monotonic functions*, Studia Sci. Math., 9 (1974), pp. 51–56.

[2] R. ASKEY AND G. GASPER, *Positive Jacobi polynomial sums II*, Amer. J. Math., 98 (1976), pp. 709–738.

[3] G. GASPER, *On the extension of Turan's inequality to Jacobi polynomials*, Duke Math. J., 38 (1971), pp. 415–428.

[4] S. KARLIN AND G. SZEGÖ, *On certain determinants whose elements are orthogonal polynomials*, J. Analyse Math., 8 (1961), pp. 1–157.

[5] O. SZÁSZ, *Identities and inequalities concerning orthogonal polynomials and Bessel functions*, J. Analyse Math., 1 (1951), pp. 116–134.

[6] G. SZEGÖ, *Orthogonal Polynomials*, Colloquium Publications, vol. 23, 3rd ed., American Mathematical Society, Providence, RI, 1967.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0004 $01.00/0

FRANKL–MORAWETZ PROBLEM IN $\mathbb{R}^3$*

A. K. AZIZ† AND M. SCHNEIDER‡

Abstract. Using a variation of the a, b, c method, in this paper we obtain sufficient conditions for the uniqueness of the solution of the boundary value problem $L[u]=K(x_3)[u_{x_1x_1}+u_{x_2x_2}]+u_{x_3x_3}+ru=f$ in G and $u|_{\phi^{(0)}\cup\phi^{(2)}}=0$, where $K(x_3)$ is a function of one variable satisfying sgn $K(x_3)=$ sgn x_3, and r is a function of three variables. Let G be a bounded simply connected region in $\mathbb{R}^3$ such that $L[u]$ is defined on G, whose boundary ∂G consists of the following surfaces:

(i) a surface $\phi^{(0)}(x)=0$ lying in $x_3>0$ and which intersects the plane $x_3=0$ in the circle $x_1^2+x_2^2=1$;
(ii) the characteristic surface $\phi^{(1)}(x)=-[x_1^2+x_2^2]^{1/2}+\int_{x_3}^0[-K(t)]^{1/2}\,dt=0$;
(iii) a surface $\phi^{(2)}(x)=0$ which intersects $x_3=0$ in the circle $x_1^2+x_2^2=1$.

Special cases of the present problem have been dealt with in the literature. However, it appears the proofs of some of these uniqueness results contain crucial gaps. For more details we refer to the Introduction and references contained therein.

1. Introduction. Consider the equation

$$\tilde{L}[u]:=k(x_3)(u_{x_1x_1}+u_{x_2x_2})+u_{x_3x_3}+r(x)u=f(x) \tag{1.1}$$

in a bounded simply-connected region G of $\mathbb{R}^3$, where the function $k(x_3)\gtreqless 0$ for $x_3\gtreqless 0$ and the region G is bounded by the surfaces of the form: A piecewise smooth surface $\phi^{(0)}(x)=0$ lying in $x_3>0$ which intersects the plane $x_3=0$ in the circle $x_1^2+x_2^2=1$. For $x_3>0$ by the characteristic surface of (1.1)

$$\phi^{(1)}(x)=-(x_1^2+x_2^2)^{1/2}+\int_{x_3}^0(-k(t))^{1/2}\,dt=0 \tag{1.2}$$

and a piecewise smooth surface $\phi^3(x)=0$ which intersects the plane $x_3=0$ in $x_1^2+x_2^2=1$ and satisfies the condition

$$k(x_3)([\phi_{x_1}^{(3)}]^2+[\phi_{x_2}^{(3)}]^2)+[\phi_{x_3}^{(3)}]^2\geqq 0. \tag{1.3}$$

The condition (1.3) implies that the surface $\phi^{(3)}(x)=0$ lies inside the characteristic triangle $\phi^{(1)}(x)=0$ and the characteristic surface

$$\phi^{(2)}(x)=(x_1^2+x_2^2)^{1/2}-1+\int_{x_3}^0(-k(t))^{1/2}\,dt=0. \tag{1.4}$$

In this paper using a variation of the a, b, c method we obtain sufficient conditions for the uniqueness of the solution of the boundary value problem

$$\tilde{L}[u]=f \quad \text{in } G, \tag{1.5}$$

$$u|_{\phi^{(0)}\cup\phi^{(3)}}=0. \tag{1.6}$$

We note that the surface $\phi^{(3)}=0$ may coincide with the characteristic surface $\phi^{(2)}=0$. In fact G. D. Karatoprakliev [3], J. S. Papadakis [6] and M. H. Protter [7] give uniqueness theorems for the case $\phi^{(3)}\equiv\phi^{(2)}$. In [3] a uniqueness result is given for the case $\phi^3=\phi^2$, for the boundary value problem

$$\tilde{L}[u]=f \quad \text{in } G, \tag{1.7}$$

---

* Received by the editors August 11, 1977 and in final revised form July 25, 1978.

† Department of Mathematics, University of Maryland, and Institute for Physical Science and Technology, College Park, Maryland 20742. The work of this author was supported in part by the U.S. Energy Research and Development Administration under Contract ERDA E(40-1)3443.

‡ Technische Universität Berlin, Berlin, West Germany.

$$u|_{\phi^{(0)}\cup\phi^{(1)}\cup\phi^{(2)}}=0. \tag{1.8}$$

In general, it is evident that in this case the problem is overdetermined. In [6] the uniqueness of the solution for the boundary value problem (1.5), (1.6) is considered for the case $\phi^2=\phi^3$ and $k(x_3)=\pm 1$ for $x_3 \gtreqless 0$. The proof of the uniqueness in [6] contains a gap on page 168. By private communication with the author we know that he presented a corrected version of this uniqueness result at the 1975 spring meeting of the American Mathematical Society (Missouri) not only for the case $k(x_3)=\pm 1$ but also for the case $k(x_3)=x_3$ and the author mentions that the method of proof may be extended to the case $k(x_3)=x_3^n$, where n is an odd integer.

Similarly the proof in [7] contains a crucial gap on page 445, where it is stated that the surface integral over S_4 is positive-semidefinite. On the contrary, this surface integral is negative definite. The second author in [8] gives a uniqueness proof for the problem considered here, with the assumption that the solution lies in a restricted function space.

Our proof of the uniqueness, which is a generalization of the techniques used in [1], is based on the a, b, c method. Since this method is also useful for proving the existence of weak and semistrong solutions by providing a priori estimates, we briefly indicate the technique of how to use the method in $\mathbb{R}^3$ for a general linear equation of second order (see Remark 2.1 below).

2. A priori estimates for the Frank–Morawetz problem. We consider the differential operator (1.1) in the form

$$L[u]=(A^{ik}u_{x_i})_{x_k}+Ru, \qquad i,k=1,2,3, \tag{2.1}$$

where

$$\begin{gathered} A^{ik}(x)=0 \quad \text{for } i\neq k,\\ A^{ii}(x)\in C^1(\bar G_+)\cap C^1(\bar G_-), \qquad R(x)\in C^1(\bar G), \quad \text{and}\\ x=(x_1,x_2,x_3), \quad G_+=G\cap\{x_3>0\}, \quad G_-=G\cap\{x_3<0\}. \end{gathered} \tag{2.2}$$

The repeated indices as a subscript or a superscript denote a summation over i, k = 1, 2, 3.

Let $\alpha^0(x)$, $\alpha^i(x)$, $i=1,2,3$, be real valued arbitrary functions with

$$\alpha^0(x)\in C^2(\bar G_+)\cap C^2(\bar G_-), \qquad \alpha^i(x)\in C^1(\bar G_+)\cap C^2(\bar G_-); \tag{2.3}$$

then we get, using $u_{x_k}u_{x_ix_j}=(u_{x_k}u_{x_i})_{x_j}-u_{x_i}u_{x_kx_j}$,

$$\alpha^1u_{x_1}(A^{ii}u_{x_i})_{x_i}=[\alpha^1A^{ii}u_{x_1}u_{x_i}]_{x_i}-(\alpha^1)_{x_i}A^{ii}u_{x_1}u_{x_i}-\tfrac12[\alpha^1A^{ii}u_{x_i}^2]_{x_1}+\tfrac12(\alpha^1A^{ii})_{x_1}u_{x_i}^2 \tag{2.4}$$

as well as corresponding terms for

$$\alpha^2u_{x_2}(A^{ii}u_{x_i})_{x_i}, \qquad \alpha^3u_{x_2}(A^{ii}u_{x_i})_{x_i}$$

and

$$\alpha^0u(A^{ii}u_{x_i})_{x_i}=[\alpha^0uA^{ii}u_{x_i}]_{x_i}-\tfrac12[(\alpha^0)_{x_i}A^{ii}u^2]_{x_i}+\tfrac12(\alpha^0_{x_i}A^{ii})_{x_i}u^2-\alpha^0A^{ii}u_{x_i}u_{x_i}. \tag{2.5}$$

By (2.4) and (2.5) we get for the differential operator (2.1) the identity

$$2(\alpha^0u+\alpha^iu_{x_i})L[u]=P^i_{x_i}+(-a_{\infty})u^2+\sum_{i,k=1}^{3}(-a_{ik})u_{x_i}u_{x_k} \tag{2.6}$$

where

$$P^1 = 2(\alpha^0 u + \alpha^i u_{x_i})A^{11}u_{x_1} + \alpha^1(ru^2 - A^{ii}u_{x_i}^2) - u^2 A^{11}\alpha^0_{x_1},$$
$$(2.7)\qquad P^2 = 2(\alpha^0 u + \alpha^i u_{x_i})A^{22}u_{x_2} + \alpha^2(ru^2 - A^{ii}u_{x_i}^2) - u^2 A^{22}\alpha^0_{x_2},$$
$$P^3 = 2(\alpha^0 u + \alpha^i u_{x_i})A^{33}u_{x_3} + \alpha^3(ru^2 - A^{ii}u_{x_i}^2) - u^2 A^{33}\alpha^0_{x_3};$$

$$a_\infty = -2\alpha^0 R + (R\alpha^i)_{x_i} - (A^{ii}\alpha^0_{x_i})_{x_i},$$
$$a_{11} = -(A^{11}\alpha^i)_{x_i} + 2A^{11}(\alpha^1_{x_i} + \alpha^0),$$
$$a_{22} = -(A^{22}\alpha^i)_{x_i} + 2A^{22}(\alpha^2_{x_2} + \alpha^0),$$
$$(2.8)\qquad a_{33} = -(A^{33}\alpha^i)_{x_i} + 2A^{33}(\alpha^3_{x_3} + \alpha^0),$$
$$a_{12} = a_{21} = \alpha^2_{x_1}A^{11} + \alpha^1_{x_2}A^{22},$$
$$a_{13} = a_{31} = \alpha^3_{x_1}A^{11} + \alpha^1_{x_3}A^{33},$$
$$a_{23} = a_{32} = \alpha^3_{x_2}A^{22} + \alpha^2_{x_3}A^{33}.$$

Let $n = (n_1, n_2, n_3)$ denote the outer normal to ∂G. From (2.6) by use of Green's theorem, we obtain

$$(2.9)\qquad \iiint\limits_{G_+\cup G_-} 2(\alpha^0 u + \alpha^i u_{x_i})L[u]\, dx_1\, dx_2\, dx_3 + \iiint\limits_{G_+\cup G_-} \left(a_\infty u^2 + \sum_{i,k=1}^{3} a_{ik}u_{x_i}u_{x_k}\right) dx_1\, dx_2\, dx_3 = \iint\limits_{\partial G_+\cup\partial G_-} P^i n_i\, ds.$$

We shall show that for a quasi-regular solution u of the Frankl–Morawetz problem (1.5), $\tilde{L}[u] = 0$ in G with $u|_{\phi^{(0)}\cup\phi^{(3)}} = 0$, the functions α^0 and α^i, $i = 1, 2, 3$, can be determined so that

$$(2.10)\qquad 0 \leqq \iint\limits_{\partial G_+\cup\partial G_-} P^i n_i\, ds = \iiint\limits_{G_+\cup G_-} \left(a_\infty u^2 + \sum_{i,k=1}^{3} a_{ik}u_{x_i}u_{x_k}\right) dx_1\, dx_2\, dx_3 \leqq 0;$$

thus it will follow from (2.10) that $u \equiv 0$ in G. Hereby the functions α^0, α^i have the further property that the boundary integrals over the "parabolic" surfaces $G \cap \{x_3 = 0\pm\}$ vanish by cancellation.

DEFINITION 2.1. If the coefficients of (2.1) satisfy conditions (2.2) and the functions α^0, α^i satisfy conditions (2.3), then we call a function $u(x)$ a *quasi-regular solution* of (2.1) if the following hold [4, p. 234]:

(i) $u(x)$ satisfies $\tilde{L}[u] = f(\in C^0(\bar{G}))$.

(ii) The integrals

$$\iint\limits_{G\cap\{x_3=0\pm\}} P^i n_i\, ds$$

in (2.9) exist.

(iii) If $G_\pm(\varepsilon)$ are regions with boundaries $\partial G_\pm(\varepsilon)$ lying entirely in G_+ and G_-, then surface integrals along the surfaces $\partial G_\pm(\varepsilon)$ which result from the application of Green's theorem to the integrals

$$\iiint_{G_\pm(\varepsilon)} uL[u]\,d\tau, \qquad \iiint_{G_\pm(\varepsilon)} u_{x_j}L[u]\,d\tau, \qquad j=1,2,3,$$

have a limit when $\partial G_\pm(\varepsilon)$ approaches the boundary of G_+ and G_-.
To motivate our method of proof we first give a uniqueness theorem for a special Frankl–Morawetz problem, namely, we consider the case where in (1.1), $k(x_3)=\operatorname{sign} x_3|x_3|^m$, $m>0$.

THEOREM 2.1. *The equation*

$$\tilde{L}[u]=k(x_3)(u_{x_1x_1}+u_{x_2x_2})+u_{x_3x_3}+r(x)u=f(x)$$

where

$$k(x_3)=\operatorname{sign} x_3|x_3|^m, \quad m>0; \quad r(x)\in C^1(\bar{G}), \quad f(x)\in C^0(\bar{G}),$$

has at most one quasi-regular solution in G satisfying the boundary condition $u|_{\phi^{(0)}\cup\phi^{(3)}}=0$ *if the function* $r(x)$, *the surfaces* $\phi^{(0)}$ *and* $\phi^{(3)}$, *satisfy the following conditions*:

$$-\frac{4}{m+2}r-\boldsymbol{\alpha}\cdot\operatorname{grad} r\geqq 0 \quad \text{in } G, \tag{2.11}$$

$$\boldsymbol{\alpha}\cdot\operatorname{grad}\phi^{(0)}|_{\phi^{(0)}}=\alpha^i n_i|_{\phi^{(0)}}>0, \tag{2.12}$$

$$\boldsymbol{\alpha}\cdot\operatorname{grad}\phi^{(3)}|_{\phi^{(3)}}=\alpha^i n_i|_{\phi^{(3)}}\geqq 0, \tag{2.13}$$

where

$$\boldsymbol{\alpha}=(\alpha^1,\alpha^2,\alpha^3)=\left(x_1,x_2,\frac{2}{m+2}x_3\right).$$

Proof. Suppose there exist two solutions u_1 and u_2; let $u=u_1-u_2$. By (2.9)

$$\iint_{\partial G_+\cup\partial G_-} P^i n_i\,dS=\iiint_{G_+\cup G_-}\left(a_\infty u^2+\sum_{i,k=1}^{3} a_{ik}u_{x_i}u_{x_k}\right)dx_1\,dx_2\,dx_3 \tag{2.14}$$

where the functions P^1, a_∞, a_{ik} correspond to (2.7) and (2.8).

(i) By choice of

$$\alpha^1=x_1, \quad \alpha^2=x_2, \quad \alpha^3=\frac{2}{m+2}x_3, \quad \alpha^0=\frac{m+1}{m+2}, \tag{2.15}$$

we have

$$\sum_{i,k=1}^{3} a_{ik}u_{x_i}u_{x_k}\equiv 0, \qquad a_\infty=\frac{4}{m+2}r+\boldsymbol{\alpha}\cdot\operatorname{grad} r;$$

thus the right side of (2.14) is nonpositive under the condition (2.11).

(ii) For the boundary integrals over the surfaces $\phi^{(0)}$ and $\phi^{(3)}$ we get from (2.7)

$$I_{03}:=\iint_{\phi^{(0)}\cup\phi^{(3)}}\{2(\alpha^0u+\alpha^iu_{x_i})A^{kk}u_{x_k}+\alpha^k(ru^2-A^{ii}u_{x_i}^2)-u^2A^{kk}\alpha^0_{x_k}\}n_k\,dS.$$

With the choice (2.15), $u_{x_i}=(\partial u/\partial n)n_i$ on the boundary, $u|_{\phi^{(0)}\cup\phi^{(3)}}=0$, and we have

$$I_{03}=\iint_{\phi^{(0)}\cup\phi^{(3)}}\left(\frac{\partial u}{\partial n}\right)^2\{\alpha^i n_i A^{kk}n_k n_k\}\,dS.$$

For $x_3>0$ (in the elliptic part of the region) we have $A^{kk}n_kn_k>0$; on the boundary $\phi^{(3)}$ we have $A^{kk}n_kn_k \geqq 0$ by condition (1.3); thus $I_1>0$ with (2.12) and (2.13).

(iii) It remains to show that

$$I_1 := \iint_{\phi^{(1)}} P^i n_i \, dS \geqq 0$$

where

$$\phi^{(1)} = -(x_1^2+x_2^2)^{1/2} + \int_{x_3}^{0} (-k(t))^{1/2}\, dt = -r + \frac{2}{m+2}(-x_3)^{(m+2)/2} = 0.$$

By a simple calculation with $x_1 = r\cos\phi$, $x_2 = r\sin\phi$ we have

$$\text{(2.16)}\quad I_1 = 2\left(\frac{m+2}{2}\right)^{m/(m+2)} \int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} \left\{\frac{m+1}{m+2} r^{m/(m+2)+1} u u_r + r^{m/(m+2)+2} u_r^2\right\} dr\, d\phi.$$

Now by integration by parts we have

$$\text{(2.17)}\quad \frac{m+2}{m+1} \int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{m/(m+2)+1} u u_r \, dr\, d\phi = -\int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{m/(m+2)} u^2 \, dr\, d\phi.$$

Thus

$$\begin{aligned} \left| -\int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{m/(m+2)} u^2 \, dr\, d\phi \right| &= \frac{m+2}{m+1} \int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{m/(2(m+2))} u r^{(3m+4)/(2(m+2))} u_r \, dr\, d\phi \\ &\leqq \frac{m+2}{m+1} \left(\int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{m/(m+2)} u^2 \, dr\, d\phi \right)^{1/2} \\ &\quad \cdot \left(\int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{(3m+4)/(m+2)} u_r^2 \, dr\, d\phi \right)^{1/2}. \end{aligned}$$

We now have

$$\text{(2.18)}\quad \int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{m/(m+2)} u^2 \, dr\, d\phi \leqq \left(\frac{m+2}{m+1}\right)^2 \int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} r^{(3m+4)/(m+2)} u_r^2 \, dr\, d\phi.$$

Using (2.17) in (2.18) we have

$$\int_{\phi=0}^{2\pi} \int_{r=0}^{R(\phi)} \left\{\frac{m+1}{m+2} r^{m/(m+2)+1} u u_r + r^{m/(m+2)+2} u_r^2\right\} dr\, d\phi \geqq 0.$$

The last inequality implies $I_1 \geqq 0$.

(iv) With the choice (2.15) the boundary integrals

$$\iint_{G\cap\{x_3=0\pm\}} P^i n_i \, dS$$

vanish by cancellation.

Under the hypothesis of Theorem 2.1 we obtain from (i)–(iv) that

$$\text{(2.19)}\quad u|_{\phi^{(0)}} = \left.\frac{\partial u}{\partial n}\right|_{\phi^{(0)}} = 0.$$

By elliptic theory (see [5, p. 60]) we get $u \equiv 0$ in G_+, but this implies $u \equiv 0$ in G_-, since in this case $u|_{x_3=0} = (\partial u/\partial n)|_{x_3=0} = 0$ and $u|_{\phi^{(3)}} = 0$.

The proof of Theorem 2.1 indicates the way for proving an analogous uniqueness theorem for Frankl–Morawetz problem in the general case when the function $k(x_3) \gtreqless 0$ whenever $x_3 \gtreqless 0$.

The central idea is to choose the functions α^0, α^i, $i = 1, 2, 3$, in such a manner that the quadratic form $\sum_{i,k=1}^3 a_{ik} u_{x_i} u_{x_k} \equiv 0$ in G and the boundary integral $\iint_{\partial G_+ \cup \partial G_-} P^i n_i \, dS$ in (2.10) is nonnegative.

We consider

$$L[u] = k(x_3)(u_{x_1x_1} + u_{x_2x_2}) + u_{x_3x_3} + r(x)u = f(x) \tag{2.20}$$

where it is assumed that the functions $k(x_3)$, $r(x)$, $f(x)$, $\alpha^0(x)$ and $\boldsymbol{\alpha} = (\alpha^1, \alpha^2, \alpha^3)$ satisfy the following hypotheses.

H_1. (a) $k(x_3) \in C^0(\bar{G}) \cap C^3(\bar{G}_+) \cap C^3(\bar{G}_-)$, $\quad r(x) \in C^1(\bar{G})$, $\quad f(x) \in C^0(\bar{G})$,

(b)
$$\alpha^1 = x_1, \qquad \alpha^2 = x_2,$$
$$\alpha^3 = -|k(x_3)|^{-1/2} \int_{x_3}^0 |k(t)|^{1/2} \, dt \quad \text{and}$$
$$2\alpha^0 = 1 - \operatorname{sign} x_3 \frac{k'}{2} |k(x_3)|^{-3/2} \int_{x_3}^0 |k(t)|^{1/2} \, dt.$$

H_2. (a) $\alpha^0_+ - \alpha^0_-$, $\quad \alpha^0_{x_3+} - \alpha^0_{x_3-} = 0$, $\quad \alpha^3_+ - \alpha^3_- = 0 \quad$ in $G \cap \{x_3 = 0\pm\}$,

(b)
$$-a_\infty = 2\alpha^0 r - (r\alpha^i)_{x_i} + k(x_3)(\alpha^0_{x_1x_1} + \alpha^0_{x_2x_2}) + \alpha^0_{x_3x_3} \geqq 0$$

in G_+ and G_-, and a_∞ is integrable over G_+ and G_-,

(c)
$$\boldsymbol{\alpha} \cdot \operatorname{grad} \phi^{(0)}|_{\phi^{(0)}} = \alpha^i n_i|_{\phi^{(0)}} > 0,$$
$$\boldsymbol{\alpha} \cdot \operatorname{grad} \phi^{(3)}|_{\phi^{(3)}} = \alpha^i n_i|_{\phi^{(3)}} \geqq 0.$$

H_3. (a) $\{(-k)^{1/2} r^2 g_r - 2g^2\}|_{\phi^{(1)}} \geqq 0, \qquad \{g_r(r)\}|_{\phi^{(1)}} \geqq 0$

and g^2/g_r^2 is integrable on $\phi^{(1)}$, where

$$r = \int_{x_3}^0 (-k(t))^{1/2} \, dt, \quad g(r) = \alpha^0 (-k)^{1/2} r + h(r), \quad h_r(r) = (-k)^{1/2} r \alpha^0_r(r).$$

(b)
$$\lim_{r \to 0} g(r) = \lim_{r \to 0} h(r) = 0 \quad \text{on } \phi^{(1)}.$$

THEOREM 2.2. *If the hypotheses* H_1–H_3 *hold, then there exists at most one quasi-regular solution of* (2.20) *in G satisfying the boundary condition* $u|_{\phi^{(0)} \cup \phi^{(3)}} = 0$.

Proof. Suppose u_1 and u_2 are two solutions of (2.20) satisfying the boundary conditions $u_i|_{\phi^{(0)} \cup \phi^{(3)}} = 0$, $i = 1, 2$; let $u = u_1 - u_2$. From (2.9) we have

$$\iint_{\partial G_+ \cup \partial G_-} P^i n_i \, dS = \iiint_{G_+ \cup G_-} \left(a_\infty u^2 + \sum_{i,k=1}^3 a_{ik} u_{x_i} u_{x_k} \right) dx_1 \, dx_2 \, dx_3. \tag{2.21}$$

(i) Using α^0, α^i, $i = 1, 2, 3$, as given in H_1(b), we have

$$\sum_{i,k=1}^3 a_{ik} u_{x_i} u_{x_k} \equiv 0,$$
$$\alpha_\infty = -2\alpha^0 r + (r\alpha^i)_{x_i} - k(x_3)(\alpha^0_{x_1x_1} + \alpha^0_{x_2x_2}) - \alpha^0_{x_3x_3};$$

thus the right side in (2.21) is nonpositive under the hypothesis H_2(b).

(ii) From (2.7) and H_2(c), using the same argument as in Theorem 2.1 we obtain

$$I_{03}=\iint_{\phi^{(0)}\cup\phi^{(3)}} P^i n_i\,dS \geqq 0.$$

(iii) It only remains to show that

$$I_1=\iint_{\phi^{(1)}} P^i n_i\,dS \geqq 0$$

where

$$\phi^{(1)}=-r+\int_{x_3}^{0}(-k(t))^{1/2}\,dt=0.$$

To this end we consider

$$I_1=\int_{\phi=0}^{2\pi}\int_{r=0}^{R}[2\alpha^0(-k)^{1/2}ruu_r+2(-k)^{1/2}r^2u_r^2(-k)^{1/2}r\alpha_r^0]\,dr\,d\phi. \tag{2.22}$$

From H_3(a, b) and integration by parts we have

$$\int_{\phi=0}^{2\pi}\int_{r=0}^{R}2uu_rh(r)\,dr\,d\phi=-\int_{\phi=0}^{2\pi}\int_{r=0}^{R}u^2h_r(r)\,dr\,d\phi=-\int_{\phi=0}^{2\pi}\int_{r=0}^{R}u^2(-k)^{1/2}r\alpha_r^0\,dr\,d\phi.$$

Thus

$$I_1=\int_{\phi=0}^{2\pi}\int_{r=0}^{R}\{2uu_r[\alpha^0(-k)^{1/2}r+h(r)]+2(-k)^{1/2}r^2u_r^2\}\,dr\,d\phi. \tag{2.23}$$

Similarly, we get

$$\begin{aligned}\int_{\phi=0}^{2\pi}\int_{r=0}^{R}2uu_r[\alpha^0(-k)^{1/2}r+h(r)]\,dr\,d\phi&=\int_{\phi=0}^{2\pi}\int_{r=0}^{R}2uu_rg(r)\,dr\,d\phi\\&=\int_{\phi=0}^{2\pi}\int_{r=0}^{R}u^2g_r(r)\,dr\,d\phi.\end{aligned} \tag{2.24}$$

Using the above relation we obtain

$$I_1=\int_{\phi=0}^{2\pi}\int_{r=0}^{R}\{-u^2g_r(r)+2(-k)^{1/2}r^2u_r^2\}\,dr\,d\phi. \tag{2.25}$$

Since by H_3(a), $g_r(r)\geqq 0$ on $\phi^{(1)}$, integration by parts and Schwarz inequality yield

$$\begin{aligned}\left|-\int_{\phi=0}^{2\pi}\int_{r=0}^{R}u^2g_r(r)\,dr\,d\phi\right|&=\left|\int_{\phi=0}^{2\pi}\int_{r=0}^{R}ug_r(r)^{1/2}2u_rg_r(r)^{-1/2}g(r)\,dr\,d\phi\right|\\&\leqq\left(\int_{\phi=0}^{2\pi}\int_{r=0}^{R}u^2g_r\,dr\,d\phi\right)^{1/2}\left(\int_{\phi=0}^{2\pi}\int_{r=0}^{R}4u_r^2\frac{g^2}{g_r}\,dr\,d\phi\right)^{1/2}.\end{aligned}$$

Thus

$$\int_{\phi=0}^{2\pi}\int_{r=0}^{R}u^2g_r(r)\,dr\,d\phi\leqq\int_{\phi=0}^{2\pi}\int_{r=0}^{R}4u_r^2\frac{g^2}{g_r}\,dr\,d\phi.$$

Noting that

$$\int_{\phi=0}^{2\pi}\int_{r=0}^{R} 2uu_r g(r)\,dr\,d\phi = -\int_{\phi=0}^{2\pi}\int_{r=0}^{R} u^2 g_r(r)\,dr\,d\phi,$$

we conclude from (2.25) that

$$\int_{\phi=0}^{2\pi}\int_{r=0}^{R}\left\{4u_r^2\frac{g^2}{gr}+2uu_r g(r)\right\}dr\,d\phi \geqq 0. \tag{2.26}$$

From (2.25) and (2.26) we obtain $I_1=\iint_{\phi^{(1)}} P^i n_i\,dS \geqq 0$ provided

$$2(-k)^{1/2}r^2 \geqq 4\frac{g^2}{g_r} \quad \text{on } \phi^{(1)},$$

which is precisely our first condition in H_3(a).

(iv) With the choice of the functions α^0, α^i, in H_1(b) and the hypothesis H_2(a), the boundary integrals $\iint_{G\cap\{x_3=0\pm\}} P^i n_i\,dS$ vanish by addition.

Under the hypotheses of Theorem 2.2, we get from (2.21) that

$$u|_{\phi^{(0)}} = \left.\frac{\partial u}{\partial n}\right|_{\phi^{(0)}} = 0,$$

from which it follows as in the proof of Theorem 2.1 that $u\equiv 0$ in G.

Remark 2.1. If $k(x_3)=\operatorname{sign} x_3|x_3|^m$, $m>0$, then we have a special case of Theorem 2.1. To see that the conditions of Theorem 2.1 are satisfied, we observe that in this special case $h_r(r)=0$, and since $\lim_{r\to 0} h(r)=0$ we have $h(r)=0$,

$$g(r)=\frac{m+1}{m+2}\left(\frac{m+2}{2}\right)^{m/(m+2)} r^{m/(m+2)+1} \quad \text{and} \quad (-k)_r^{1/2} g_r - 2g^2 = 0.$$

Remark 2.2. The method used for the uniqueness theorem is also useful for proving the existence of weak and semistrong solutions by providing a priori estimates. To show how to use this method for a general linear equation of second order in $\mathbb{R}^3$ we consider the differential operator

$$L[u]=(A^{ik}u_{x_i})_{x_k}+B^i u_{x_i}+Ru, \qquad i,k=1,2,3, \tag{2.27}$$

where $A^{ik}(x)=0$ for $i\neq k$. The equation (2.27) may be written as a Pfaffian form of third degree by the use of ε-tensor

$$\varepsilon_{irl}=\begin{cases}1 & \text{for an even permutation } i, r, l \text{ of } 1, 2, 3,\\ 0 & \text{for two same indices,}\\ -1 & \text{for an odd permutation } i, r, l \text{ of } 1, 2, 3.\end{cases} \tag{2.28}$$

Corresponding to (2.27) we introduce the forms

$$\begin{aligned} d_n u &= \tfrac{1}{2}\varepsilon_{irl}A^{ik}u_{x_k}[dx^r, dx^l]\\ &= A^{11}u_{x_1}[dx^2, dx^3]-A^{22}u_{x_2}[dx^1, dx^3]+A^{33}u_{x_3}[dx^1, dx^2],\\ \theta &= \tfrac{1}{2}\varepsilon_{irl}B^i[dx^r, dx^l],\\ \tilde{R} &= R-\sum_{i=1}^{3}\frac{\partial B_i}{\partial x_i}.\end{aligned} \tag{2.29}$$

For the geometrical interpretation of $d_n u$ we refer the reader to [2]. The differential operator (2.27) now may be expressed as a Pfaffian form of third degree given by

$$L[u][dx^1, dx^2, dx^3] = [d, d_n u] + [d, u\theta] + Ru[dx^1, dx^2, dx^3]. \tag{2.30}$$

Let $\alpha^0(x)$, $\alpha^i(x)$, $i = 1, 2, 3$, be real valued arbitrary functions and

$$\omega_2 = \alpha^1[dx^2, dx^3] - \alpha^2[dx^1, dx^3] + \alpha^3[dx^1, dx^2];$$

then we consider the second order Pfaffian form

$$\Omega = 2(\alpha^0 i + \alpha^i u_{x_i})\, d_n u + (Ru^2 - A^{ii} u_{x_i}^2)\omega_2 - u^2 d_n \alpha^0 + \alpha^0 u^2 \theta. \tag{2.31}$$

A formal calculation gives

$$[d, \Omega] = 2(\alpha^0 u + \alpha^i u_{x_i}) L[u][dx^1, dx^2, dx^3] + \left(a_{\infty} u^2 + \sum_{i,k=1}^{3} a_{ik} u_{x_i} u_{x_k} \right) \cdot [dx^1, dx^2, dx^3] \tag{2.32}$$

where $a_{ik} = a_{ki}$ and these functions depend on A^{ik}, α^0 and α^i, $i = 1, 2, 3$, in a way similar to (2.7), (2.8). By the use of the Green theorem we have

$$\iint_{\partial G} \Omega = \iiint_{G} [d, \Omega] \tag{2.33}$$

from which we obtain a priori estimates by special choice of the functions α^0, $\alpha^i(x)$.

REFERENCES

[1] A. K. AZIZ AND M. SCHNEIDER, *Uniqueness of Frankl-Morawetz problem in* $\mathbb{R}^2$, Monatsh. Math., 85 (1977), pp. 265–276.

[2] W. HAACK AND W. WENDLAND, *Über Partielle und Pfaffsche Differentialgleichungen*, Birkhäuser, Basel and Stuttgart, 1969.

[3] G. D. KARATOPRAKLIEV, *On some equations of a mixed type in multidimensional regions*, C.R. Acad. Bulgare Sci., 24 (1971), pp. 1597–1600.

[4] S. G. MIKHLIN, *Linear Equations of Mathematical Physics*, Holt Reinhart and Winston, New York, 1967.

[5] C. MIRANDA, *Partial Differential Equations of Elliptic Type*, Springer-Verlag, Berlin, 1970.

[6] J. PAPADAKIS, *A boundary value problem for equations of mixed type*, Bull. Soc. Math. Grèce, 14 (1973), pp. 157–171.

[7] M. H. PROTTER, *New boundary value problems for the wave equation and equations of mixed type*, J. Rational Mech. Anal., 3 (1954), pp. 435–446.

[8] M. SCHNEIDER, *Über Randwertprobleme für Gleichungen vom gemischten Typ. I, II*, Ann. Polon. Math., XXXIV (1977), pp. 1–12; XXXIV (1977), pp. 43–45.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0005 $01.00/0

A BIFURCATION PROBLEM FOR A NONLINEAR HYPERBOLIC PARTIAL DIFFERENTIAL EQUATION*

G. F. WEBB†

Abstract. The stability of equilibrium solutions of the damped nonlinear wave equation $w_{tt}+2\alpha w_t-w_{xx}-\lambda f(w)=0$, $\alpha>0$, $\lambda\geqq 0$, is investigated using Lyapunov stability techniques. Under appropriate conditions on f it is shown that for $\lambda_n<\lambda\leqq\lambda_{n+1}$, $\lambda_n=n^2/f'(0)$, $n=0,1,2,\cdots$, there are exactly $2n+1$ equilibrium solutions and all solutions exist globally and approach exactly one of them as t approaches ∞.

1. Introduction. The objective of this paper is to study the asymptotic behavior of solutions to the damped nonlinear wave equation

$$
\begin{aligned}
&w_{tt}(x,t)+2\alpha w_t(x,t)-w_{xx}(x,t)-\lambda f(w(x,t))=0, \qquad 0<x<\pi, \qquad t\geqq 0,\\
&w(0,t)=w(\pi,t), \qquad t\geqq 0, \qquad\qquad (1.1)\\
&w(x,0)=\phi(x), \qquad w_t(x,0)=\psi(x), \qquad 0<x<\pi,
\end{aligned}
$$

where α is a positive constant and λ is a nonnegative parameter. The conditions we require on f allow the cases that $f(w)=\sin w$ (sine-Gordon equation) and $f(w)=aw-bw^3$; $a, b>0$ (Klein–Gordon equation). Out results extend and refine results obtained by different methods due to R. W. Dickey in [5], [6] and A. J. Callegari and E. L. Reiss in [1]. The approach we take is to reformulate (1.1) in a Banach space as an abstract first order ordinary differential equation whose solutions define a dynamical system. We then apply Lyapunov stability techniques to investigate the large time behavior of the solutions as a function of the parameter λ. Our treatment is analogous to that of N. Chafee and E. F. Infante in [2] who considered a similar problem for a parabolic equation. We will follow the development of D. Henry [7] who extended and abstracted the ideas of Chafee and Infante in the parabolic case.

Our results can be summarized in the following way: Let $\lambda_n=n^2/f'(0)$, $n=1,2,\cdots$. If $0\leqq\lambda\leqq\lambda_1$, then all solutions of (1.1) exist globally and approach the zero stationary solution $\phi_0\equiv 0$ as $t\to+\infty$. If $\lambda_n<\lambda\leqq\lambda_{n+1}$, then there are $2n$ additional stationary solutions $\phi_{k,\pm}$, $k=1,\cdots,n$, and every solution of (1.1) exists globally and approaches exactly one of them as $t\to+\infty$. In this case $\phi_{1,\pm}$ are locally asymptotically stable and ϕ_0, $\phi_{k,\pm}$, $k=2,\cdots,n$, are all unstable.

In § 2 we present the abstract theory which we shall require. In § 3 we formulate (1.1) abstractly and apply the results of § 2. For the sake of completeness and clarity we will provide careful details for all proofs except those which may be found elsewhere.

2. The abstract problem. Let X be a Banach space with norm $\|\cdot\|$ and let $A: X\to X$. We make the following assumption on A:

(2.1) A is a closed linear operator in X and there exists a closed linear operator B in X such that $B^2=-A$ and B^{-1} exists with domain all of X and is compact from X to X.

Let $[D(B)]$ be the Banach space which is the domain of B with the norm $\|\phi\|_B=\|B\phi\|$, $\phi\in D(B)$. Let Y denote the Banach space which is $[D(B)]\times X$ with the norm

* Received by the editors February 10, 1978.

† Department of Mathematics, Vanderbilt University, Nashville, Tennessee 37235. This work was supported in part by the National Science Foundation under Grant NSF 75-06332 A01.

$\|(\phi, \psi)\|_Y = (\|\phi\|_B + \|\psi\|)^{1/2}$, $\phi \in D(B)$, $\psi \in X$. Let $\alpha \in \mathbb{R}$ and let $\mathscr{A}: Y \to Y$ by

$$\mathscr{A}(\phi, \psi) = (\psi, A\phi - 2\alpha\psi), \qquad D(\mathscr{A}) = D(A) \times D(B). \tag{2.2}$$

We make the following assumption on $\mathscr{A}$:

(2.3) $\mathscr{A}$ is the infinitesimal generator of a strongly continuous group $T(t)$, $t \in \mathbb{R}$ in Y satisfying $|T(t)| \leqq Me^{-\gamma t}$, $t \in \mathbb{R}$, where M and γ are real constants.

We make the following assumptions on the (nonlinear) operator F from $[D(B)]$ to X:

(2.4) F is continuously (Fréchet) differentiable from $[D(B)]$ to X in the sense that (i) for each $\hat{\phi} \in D(B)$, $F(\hat{\phi} + \phi) = F(\hat{\phi}) + L\phi + g(\phi)$ for all $\phi \in D(B)$, where $L: [D(B)] \to X$ is bounded, linear, and everywhere defined, and $g: [D(B)] \to X$ satisfies $\|g(\phi_1) - g(\phi_2)\| \leqq c_1(r)\|\phi_1 - \phi_2\|_B$ for all $\phi_1, \phi_2 \in D(B)$ such that $\|\phi_1\|_B$, $\|\phi_2\|_B \leqq r$ for some continuous function $c_1: \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $c_1(0) = 0$; and (ii) $L \overset{\text{def}}{=} F'(\hat{\phi})$ is continuous in $\hat{\phi}$ from $[D(B)]$ to the Banach algebra of bounded linear operators from $[D(B)]$ to X;

(2.5) $F(D(B)) \subset D(B)$ and there is a continuous increasing function $c_2: \mathbb{R}^+ \to \mathbb{R}^+$ such that $\|F(\phi)\|_B \leqq c_2(\|\phi\|_B)$ for all $\phi \in D(B)$.

We now define the (nonlinear) operator $\mathscr{F}$ from Y to Y by

$$\mathscr{F}((\phi, \psi)) = (0, F(\phi)), \qquad D(\mathscr{F}) = Y. \tag{2.6}$$

The following fundamental theorem of I. Segal will provide the local existence and the uniqueness of solutions of (1.1). Its proof may be found in [11, Thm. 1, Corollary 1.5, and Lemma 3.1].

THEOREM 2.1. *Let* (2.1), (2.3), *and* (2.4) *hold. The following are true*:

(i) *for each* $(\phi, \psi) \in Y$ *the maximal interval* $[0, t_0)$, $t_0 = t_0(\phi, \psi)$, *of the necessarily unique continuous function* $u(\cdot\,; \phi, \psi)$ *from such an interval to* Y *such that*

$$u(t; \phi, \psi) = T(t)(\phi, \psi) + \int_0^t T(t-s)\mathscr{F}(u(s; \phi, \psi))\, ds \tag{2.7}$$

has positive length, and either $t_0 = +\infty$ *or* $\|u(\cdot\,; \phi, \psi)\|_Y$ *is not bounded on* $[0, t_0)$;

(ii) $u(t; \phi, \psi)$ *is a continuous function of* t *and* (ϕ, ψ) *in the sense that if* $(\phi, \psi) \in Y$, $0 \leqq t < t_0(\phi, \psi)$, *and* $\varepsilon > 0$, *then there exists* $\delta > 0$ *such that if* $\|(\phi, \psi) - (\hat{\phi}, \hat{\psi})\|_Y < \delta$ *then* $t < t_0(\hat{\phi}, \hat{\psi})$ *and* $\|u(t; \phi, \psi) - u(t; \hat{\phi}, \hat{\psi})\|_Y < \varepsilon$; *and*

(iii) *if* $(\phi, \psi) \in D(\mathscr{A})$ *then* $u(t; \phi, \psi) \in D(\mathscr{A})$ *for* $0 \leqq t < t_0(\phi, \psi)$, *and on* $[0, t_0(\phi, \psi))$, $u(\cdot\,; \phi, \psi)$ *is continuously differentiable and satisfies*

$$(d/dt)u(t; \phi, \psi) = \mathscr{A}u(t; \phi, \psi) + \mathscr{F}(u(t; \phi, \psi)), \qquad 0 \leqq t < t_0(\phi, \psi). \tag{2.8}$$

Now we define π_1, π_2 as the projections of Y onto $[D(B)]$ and X, respectively, given by $\pi_1(\phi, \psi) = \phi$ and $\pi_2(\phi, \psi) = \psi$. Then (2.8) implies that for $(\phi, \psi) \in D(\mathscr{A})$, $0 \leqq t < t_0(\phi, \psi)$,

$$(d/dt)\pi_1 u(t; \phi, \psi) = \pi_2 u(t; \phi, \psi), \tag{2.9}$$

$$(d^2/dt^2)\pi_1 u(t; \phi, \psi) + 2\alpha(d/dt)\pi_1 u(t; \phi, \psi) = A\pi_1 u(t; \phi, \psi) + F(\pi_1 u(t; \phi, \psi)). \tag{2.10}$$

The following theorem will be of essential importance when we apply Lyapunov stability methods:

THEOREM 2.2. *Let* (2.1), (2.3), (2.4), *and* (2.5) *hold and let* $\gamma > 0$ *in* (2.3). *Let* $(\phi, \psi) \in Y$ *and let the solution* $u(\cdot\,; \phi, \psi)$ *of* (2.7) *be defined and bounded on* $[0, \infty)$. *Then*

$\{u(t;\phi,\psi): t \geqq 0\}$ is a pre-compact set in Y.

Proof. Since $\{u(t;\phi,\psi): t\geqq 0\}$ is bounded in Y, we have that $\{B\pi_1 u(t;\phi,\psi): t\geqq 0\}$ is bounded in X. By (2.5) $\{BF(\pi_1 u(t;\phi,\psi)): t\geqq 0\}$ is bounded in X. Since B^{-1} is compact in X, we must have that $\{F(\pi_1 u(t;\phi,\psi)): t\geqq 0\}$ is pre-compact in X. Thus $\{\mathcal{F}(u(t;\phi,\psi)): t\geqq 0\}$ is pre-compact in Y, which together with the strong continuity of $T(t)$, $t\in\mathbb{R}$, implies that for every $\tau\geqq 0$, $K_\tau \stackrel{\text{def}}{=} \{\int_0^\tau T(s)\mathcal{F}(u(t-s;\phi,\psi))\,ds: t\geqq \tau\}$ is pre-compact in Y. Now let N be a positive constant such that $\|\mathcal{F}(u(t;\phi,\psi))\|_Y\leqq N$ for $t\geqq 0$ and define for every $\tau\geqq 0$, $L_\tau \stackrel{\text{def}}{=} \{(\hat\phi,\hat\psi)\in Y: \|(\hat\phi,\hat\psi)\|_Y \leqq Me^{-\gamma\tau}(\|(\phi,\psi)\|_Y+N/\gamma)\}$. Then $u(t;\phi,\psi)\in L_\tau+K_\tau$ for $t\geqq\tau$, since

$$\left\|T(t)(\phi,\psi)+\int_\tau^t T(s)\mathcal{F}(u(t-s;\phi,\psi))\,ds\right\|_Y \leqq Me^{-\gamma\tau}(\|(\phi,\psi)\|_Y+N/\gamma).$$

To establish the conclusion it suffices to show that if $\{t_n\}$ increases to $+\infty$, then $\{u(t_n;\phi,\psi)\}$ has a convergent subsequence in Y. To achieve this we will use the concept of the α-measure of noncompactness due to K. Kuratowski [9]: if E is a bounded subset of Y then $\alpha[E]$ is the infimum of $\varepsilon>0$ such that E can be covered by a finite number of sets of diameter no larger than ε. We require the following facts: if E_1, E_2 are bounded subsets of Y, then (i) $E_1\subset E_2$ implies $\alpha[E_1]\leqq\alpha[E_2]$; (ii) $\alpha[E_1]=0$ iff E_1 is pre-compact; (iii) $\alpha[E_1\cup E_2]=\max\{\alpha[E_1],\alpha[E_2]\}$; and (iv) $\alpha[E_1+E_2]\leqq\alpha[E_1]+\alpha[E_2]$ ((i), (ii), (iii) follow directly from the definition of $\alpha[\cdot]$ and (iv) is proved in [4]). Let $\varepsilon>0$. There exists k such that $\alpha[L_{t_k}+K_{t_k}]<\varepsilon$. Then, $\alpha[\{u(t_n;\phi,\psi): n=1,2,\cdots\}]=\alpha[\{u(t_n;\phi,\psi): n=k,\quad k+1,\cdots\}]\leqq\alpha[L_{t_k}+K_{t_k}]<\varepsilon$. Hence, $\alpha[\{u(t_n;\phi,\psi): n=1,2,\cdots\}]=0$ and the proof is complete.

Our next two theorems, which are similar to Theorems 5.1.1 and 5.1.3 in [7], use the fact the property of being the infinitesimal generator of a strongly continuous group is preserved under bounded perturbations [8, Thm. 2.1, p. 495].

THEOREM 2.3. *Let* (2.1), (2.3), *and* (2.4) *hold. Let $\hat\phi$ be an equilibrium point of* (2.10), *that is, $\hat\phi\in D(A)$ and $A\hat\phi+F(\hat\phi)=0$. Let $F(\hat\phi+\phi)=F(\hat\phi)+F'(\hat\phi)\phi+g(\phi)$ for all $\phi\in D(B)$ as in* (2.4). *Let $\mathcal{L}: Y\to Y$ be defined by $\mathcal{L}(\phi,\psi)=(0,F'(\hat\phi)\phi)$ for all $(\phi,\psi)\in Y$ and let the strongly continuous group $\hat T(t)$, $t\in\mathbb{R}$ in Y whose infinitesimal generator is $\mathcal{A}+\mathcal{L}$ satisfy*

$$|\hat T(t)|\leqq \hat M e^{-\hat\gamma t}\quad \text{for } t\geqq 0 \text{ and constants } \hat M\geqq 1,\ \hat\gamma>0. \tag{2.11}$$

Then $(\hat\phi,0)$ is locally asymptotically stable in the sense that

(2.12) *there exists $\varepsilon>0$, $N\geqq 1$, and $\delta>0$ such that if $\|(\phi,\psi)-(\hat\phi,0)\|_Y\leqq\varepsilon$ then the solution $u(t;\phi,\psi)$ of* (2.7) *exists on $[0,\infty)$ and satisfies $\|u(t;\phi,\psi)-(\hat\phi,0)\|_Y\leqq Ne^{-\delta t}\|(\phi,\psi)-(\hat\phi,0)\|_Y$ for all $t\geqq 0$.*

Proof. Since $g(0)=0$, there exists $\varepsilon_1>0$ such that $\|g(\phi)\|\leqq(\hat\gamma/(2\hat M))\|\phi\|_B$ if $\|\phi\|_B<\varepsilon_1$. Let $\varepsilon=\varepsilon_1/(2\hat M)$, let $\|(\phi,\psi)-(\hat\phi,0)\|_Y\leqq\varepsilon$, and let $t_1\leqq\infty$ be the largest extended real number such that $\|u(t;\phi,\psi)-(\hat\phi,0)\|_Y\leqq\varepsilon_1$ for $0\leqq t<t_1$, where $u(t;\phi,\psi)$ is defined on its maximal interval of existence. We will use the fact that

$$\begin{aligned} w(t) &\stackrel{\text{def}}{=} \hat T(t)(\phi,\psi)+\int_0^t \hat T(t-s)h(s)\,ds \\ &= T(t)(\phi,\psi)+\int_0^t T(t-s)(\mathcal{L}w(s)+h(s))\,ds \end{aligned} \tag{2.13}$$

for all $(\phi,\psi)\in Y$, $t\in\mathbb{R}$, and $h\in C(\mathbb{R};Y)$.

(Condition (2.13) holds for $(\phi, \psi) \in D(\mathscr{A}) = D(\mathscr{A}+\mathscr{L})$ and $h \in C^1(\mathbb{R}; Y)$ by Theorem 1.19 of [8, p. 486] and hence for all $(\phi, \psi) \in Y$, $h \in C(\mathbb{R}; Y)$ by the strong continuity of $T(t)$, $t \in \mathbb{R}$ and $\hat{T}(t)$, $t \in \mathbb{R}$). Then, for $0 \leqq t < t_1$ we have

$$\begin{aligned} u(t; \phi, \psi) &= T(t)(\phi, \psi) + \int_0^t T(t-s)(0, F(\hat{\phi}) \\ &\qquad + F'(\hat{\phi})(\pi_1 u(s; \phi, \psi) - \hat{\phi}) + g(\pi_1 u(s; \phi, \psi) - \hat{\phi}))\, ds \\ &= \hat{T}(t)(\phi, \psi) + \int_0^t \hat{T}(t-s)(0, F(\hat{\phi}) \\ &\qquad - F'(\hat{\phi})(\hat{\phi}) + g(\pi_1 u(s; \phi, \psi) - \hat{\phi}))\, ds. \end{aligned}$$

Since $\pi_1 u(t; \hat{\phi}, 0) \equiv \hat{\phi}$ and $g(0) = 0$, we have for $0 \leqq t < t_1$,

$$\begin{aligned} \|u(t; \phi, \psi) - (\hat{\phi}, 0)\|_Y &\leqq \hat{M} e^{-\hat{\gamma}t} \|(\phi, \psi) - (\hat{\phi}, 0)\|_Y + \int_0^t \hat{M} e^{-\hat{\gamma}(t-s)} \|g(\pi_1 u(s; \phi, \psi) - \hat{\phi})\|\, ds \\ &\leqq \hat{M} e^{-\hat{\gamma}t} \|(\phi, \psi) - (\hat{\phi}, 0)\|_Y + \frac{\hat{\gamma}}{2} \int_0^t e^{-\hat{\gamma}(t-s)} \|\pi_1 u(s; \phi, \psi) - \hat{\phi}\|_B\, ds. \end{aligned}$$

Thus, for $0 \leqq t < t_1$,

$$e^{\hat{\gamma}t} \|u(t; \phi, \psi) - (\hat{\phi}, 0)\|_Y \leqq \hat{M} \|(\phi, \psi) - (\hat{\phi}, 0)\|_Y + \frac{\hat{\gamma}}{2} \int_0^t e^{\hat{\gamma}s} \|u(s; \phi, \psi) - (\hat{\phi}, 0)\|_Y\, ds$$

and Gronwall's lemma implies

$$\|u(t; \phi, \psi) - (\hat{\phi}, 0)\|_Y \leqq \hat{M} \|(\phi, \psi) - (\hat{\phi}, 0)\|_Y e^{-(\hat{\gamma}/2)t} \leqq \varepsilon_1 e^{-(\hat{\gamma}/2)t}.$$

But this means that $t_1 = \infty$ and also (2.12) holds with $N = \hat{M}$ and $\delta = \hat{\gamma}/2$.

THEOREM 2.4. *Let (2.1), (2.3), and (2.4) hold. Let $\hat{\phi}$ be an equilibrium point of (2.10), that is, $\hat{\phi} \in D(A)$ and $A\hat{\phi} + F(\hat{\phi}) = 0$. Let $F(\hat{\phi} + \phi) = F(\hat{\phi}) + F'(\hat{\phi})\phi + g(\phi)$ for all $\phi \in D(B)$ as in (2.4), let $\mathscr{L}: Y \to Y$ be defined by $\mathscr{L}(\phi, \psi) = (0, F'(\hat{\phi})\phi)$ for all $(\phi, \psi) \in Y$, and let $\hat{T}(t)$, $t \in \mathbb{R}$, be the strongly continuous group in Y whose infinitesimal generator is $\mathscr{A} + \mathscr{L}$. Let $Y = Y_1 \oplus Y_2$ with $Y_1 \neq 0$ and let σ_1, σ_2 be the projections of Y onto Y_1, Y_2, respectively. For each $t \in \mathbb{R}$ let $\hat{T}(t) Y_1 \subset Y_1$ and $\hat{T}(t) Y_2 \subset Y_2$ and let $K \geqq 1$ and $\omega > 0$ be constants such that*

$$\|\hat{T}(t)(\phi, \psi)\|_Y \leqq K e^{3\omega t} \|(\phi, \psi)\|_Y \quad \text{for } t \leqq 0,\ (\phi, \psi) \in Y_1, \tag{2.14}$$

$$\|\hat{T}(t)(\phi, \psi)\|_Y \leqq K e^{\omega t} \|(\phi, \psi)\|_Y \quad \text{for } t \geqq 0,\ (\phi, \psi) \in Y_2. \tag{2.15}$$

Then $(\hat{\phi}, 0)$ is unstable in the sense that

(2.16) *there exists $\varepsilon > 0$ and a sequence $\{(\phi_n, \psi_n)\}$ in Y such that $\{\|(\phi_n, \psi_n) - (\hat{\phi}, 0)\|_Y\} \to 0$ and $\sup_{t \geqq 0} \{\|u(t; \phi_n, \psi_n) - (\hat{\phi}, 0)\|_Y\} \geqq \varepsilon$ (where the supremum is taken over the maximal interval of existence of $u(t; \phi_n, \psi_n)$).*

Proof. Choose $r > 0$ such that $K(|\sigma_1| + |\sigma_2|) c_1(r)/\omega < \frac{1}{2}$ and $K|\sigma_2| c_1(r) r/\omega < r/(4K)$ where c_1 is as in (2.4). Let $y \in Y_1$ such that $K|\sigma_1| c_1(r) r/\omega < \|y\|_Y/2 < r/(4K)$. We claim that for each $\tau \geqq 0$ there exists a unique function $v(\cdot): (-\infty, \tau] \to Y$ such that

$$\begin{aligned} v(t) &= \hat{T}(t-\tau)y - \int_t^\tau \hat{T}(t-s)\sigma_1(0, g(\pi_1 v(s)))\, ds \\ &\quad + \int_{-\infty}^t \hat{T}(t-s)\sigma_2(0, g(\pi_1 v(s)))\, ds, \qquad t \leqq \tau, \end{aligned} \tag{2.17}$$

$$\|v(t)\|_Y \leqq r\,e^{2\omega(t-\tau)}, \qquad t \leqq \tau. \tag{2.18}$$

To establish this claim define the complete metric space $\{Q, |||\cdot|||\}$ by

$$Q = \{v \in C(-\infty, \tau;\, Y) : \|v(t)\|_Y \leqq r\,e^{2\omega(t-\tau)},\ t \leqq \tau\}$$

$$|||v||| = \sup_{t \leqq \tau} e^{-2\omega(t-\tau)}\|v(t)\|_Y.$$

Define $H: Q \to Q$ with $(Hv)(t)$ as the right-hand side of (2.17). Then H maps Q into Q, since for $v \in Q$, $t \leqq \tau$,

$$\begin{aligned}\|(Hv)(t)\|_Y &\leqq K\,e^{3\omega(t-\tau)}\|y\|_Y + \int_t^\tau K\,e^{3\omega(t-s)}|\sigma_1|c_1(r)\,r\,e^{2\omega(s-\tau)}\,ds\\ &\quad + \int_{-\infty}^t K\,e^{\omega(t-s)}|\sigma_2|c_1(r)r\,e^{2\omega(s-\tau)\,ds}\\ &\leqq K(\|y\|_Y + (|\sigma_1| + |\sigma_2|)c_1(r)r/\omega)\,e^{2\omega(t-\tau)}.\end{aligned}$$

Also, H is a strict contraction in Q, since for v, $\hat{v} \in Q$, $t \leqq \tau$,

$$\begin{aligned}\|(Hv)(t) - (H\hat{v})(t)\|_Y &\leqq \int_t^\tau K\,e^{3\omega(t-s)}|\sigma_1|c_1(r)\|v(s) - \hat{v}(s)\|_Y\,ds\\ &\quad + \int_{-\infty}^t K\,e^{\omega(t-s)}|\sigma_2|c_1(r)\|v(s) - \hat{v}(s)\|_Y\,ds\\ &\leqq Kc_1(r)(|\sigma_1| + |\sigma_2|)2\,e^{2\omega(t-\tau)}|||v - \hat{v}|||/\omega.\end{aligned}$$

By the contraction mapping theorem, H has a unique fixed point $v(t) = v(t;\tau) \in Q$, which is also the unique solution of (2.17).

Now define $y_n = v(0; n) + (\hat{\phi}, 0)$, $n = 1, 2, \cdots$. We claim that

$$v(t; n) + (\hat{\phi}, 0) = u(t; y_n), \qquad n = 1, 2, \cdots, \quad 0 \leqq t \leqq n. \tag{2.19}$$

As in (2.13)

$$\begin{aligned}v(t; n) &= T(t-n)y + \int_n^t T(t-s)[(0, F'(\hat{\phi})\pi_1\sigma_1 v(s; n)) + \sigma_1(0, g(\pi_1 v(s; n)))]\,ds\\ &\quad + \int_{-\infty}^t T(t-s)[(0, F'(\hat{\phi})\pi_1\sigma_2 v(s; n)) + \sigma_2(0, g(\pi_1 v(s; n)))]\,ds\\ &= T(t)\Big(\hat{T}(-n)y + \int_n^0 T(-s)[(0, F'(\hat{\phi})\pi_1\sigma_1 v(s; n)) + \sigma_1(0, g(\pi_1 v(s; n)))]\,ds\\ &\quad + \int_{-\infty}^0 T(-s)[(0, F'(\hat{\phi})\pi_1\sigma_2 v(s; n)) + \sigma_2(0, g(\pi_1 v(s; n)))]\,ds\Big)\\ &\quad + \int_0^t T(t-s)(0, F'(\hat{\phi})\pi_1 v(s; n) + g(\pi_1 v(s; n)))\,ds\\ &= T(t)v(0; n) + \int_0^t T(t-s)(0, F(\hat{\phi} + \pi_1 v(s; n)) - F(\hat{\phi}))\,ds\\ &= T(t)y_n - (\hat{\phi}, 0) + \int_0^t T(t-s)(0, F(\hat{\phi} + \pi_1 v(s; n)))\,ds.\end{aligned}$$

By the uniqueness of solutions to (2.7), (2.19) must hold.

Finally, $\{\|y_n-(\hat{\phi},0)\|_Y\}\to 0$ by (2.18), and $\sup_{0\leqq t\leqq n}\|u(t;y_n)-(\hat{\phi},0)\|_Y\geqq \|y\|_Y/2$, since $\|u(n;y_n)-(\hat{\phi},0)\|_Y=\|v(n;n)\|_Y\geqq\|y\|_Y-\int_{-\infty}^{n}Ke^{\omega(n-s)}|\sigma_2|c_1(r)re^{2\omega(s-n)}\,ds=\|y\|_Y-K|\sigma_2|c_1(r)r/\omega$.

3. The bifurcation problem. We now apply the abstract theory developed in § 2 to (1.1). Let $X=L^2(0;\pi;\mathbb{R})$ with inner-product $\langle\cdot,\cdot\rangle$. Let $A\colon X\to X$ be defined by $A\phi=\phi''$, $D(A)=\{\phi\in X\colon \phi,\phi'$ are absolutely continuous, $\phi''\in X$, $\phi(0)=\phi(\pi)=0\}$. We observe that $A\phi=\sum_{n=1}^{\infty}-n^2(\phi,\chi_n)\chi_n$, $\phi\in D(A)$, where $\chi_n(x)=(2/\pi)^{1/2}\sin nx$, $n=1,2,\cdots$, is a complete set of orthonormal vectors in X. Let $B\colon X\to X$ be defined by $B\phi=\sum_{n=1}^{\infty}n(\phi,\chi_n)\chi_n$, $\phi\in D(B)=\{\phi\in X:\sum_{n=1}^{\infty}n^2\langle\phi,\chi_n\rangle^2<\infty\}\supset\{\phi\in X:\phi$ is absolutely continuous, $\phi'\in X$, $\phi(0)=\phi(\pi)=0\}$. We observe that $B^2=-A$, $B^{-1}\phi=\sum_{n=1}^{\infty}(1/n)(\phi,\chi_n)\chi_n$, $\phi\in D(B^{-1})=X$, and B^{-1} is compact from X to X. Thus, (2.1) is satisfied.

If $\mathscr{A}$ is defined as in (2.2), then (2.3) is satisfied, where $T(t)$, $t\in\mathbb{R}$ is given by

$$T(t)(\phi,\psi)=\left(\sum_{n=1}^{\infty}[\langle\phi,\chi_n\rangle U_n(t)+\langle\psi,\chi_n\rangle V_n(t)]\chi_n,\right.\qquad(3.1)$$
$$\left.\sum_{n=1}^{\infty}[\langle\phi,\chi_n\rangle U'_n(t)+\langle\psi,\chi_n\rangle V'_n(t)]\chi_n\right),\qquad(\phi,\psi)\in Y,\quad t\in\mathbb{R},$$

$$(3.2)\ U_n(t)=\begin{cases}e^{-\alpha t}(\cos h(\sqrt{\alpha^2-n^2}\,t)+(\alpha/\sqrt{\alpha^2-n^2})\sin h(\sqrt{\alpha^2-n^2}\,t)), & n^2<\alpha^2,\\ e^{-\alpha t}(1+\alpha t), & n^2=\alpha^2,\\ e^{-\alpha t}(\cos(\sqrt{n^2-\alpha^2}\,t)+(\alpha/\sqrt{n^2-\alpha^2})\sin(\sqrt{n^2-\alpha^2}\,t)), & n^2>\alpha^2,\end{cases}$$

$$(3.3)\qquad V_n(t)=\begin{cases}(e^{-\alpha t}/\sqrt{\alpha^2-n^2})\sin h(\sqrt{\alpha^2-n^2}\,t), & n^2<\alpha^2,\\ t\,e^{-\alpha t}, & n^2=\alpha^2,\\ (e^{-\alpha t}/\sqrt{n^2-\alpha^2})\sin(\sqrt{n^2-\alpha^2}\,t), & n^2>\alpha^2,\end{cases}$$

[12, p. 112–115]. From (3.1) we see that there exist constants $M\geqq 1$ and $\gamma\in\mathbb{R}$ as in (2.3). Furthermore, if $\alpha>0$, then we may take $\gamma>0$.

The conditions we require on f in (1.1) are exactly the same as those in [2]. Let $f\colon\mathbb{R}\to\mathbb{R}$ such that:

(3.4) f is twice continuously differentiable on $\mathbb{R}$;

(3.5) $f(0)=0$ and $f'(0)>0$;

(3.6) $\limsup_{|x|\to+\infty} f(x)/x\leqq 0$;

(3.7) $\operatorname{sgn} f''(x)=-\operatorname{sgn} x$ for all $x\in\mathbb{R}$.

For a given $\lambda\geqq 0$ define $F\colon D(B)\to X$ by

$$(F(\phi))(x)=\lambda f(\phi(x)),\qquad \phi\in D(B),\quad 0\leqq x\leqq\pi.\qquad(3.8)$$

Our next goal will be to verify that F satisfies (2.4) and (2.5). We first observe that $\sup_{0\leqq x\leqq\pi}|\phi(x)|\leqq\sqrt{\pi}\|B\phi\|$ and $\|\phi\|\leqq\|B\phi\|$ for all $\phi\in D(B)$.

To establish (i) of (2.4) let $\hat{\phi}\in D(B)$, let $(L\phi)(x)\overset{\text{def}}{=}\lambda f'(\hat{\phi}(x))\phi(x)$ for $0<x<\pi$, and let $c(r)=\sup_{|x|\leqq\sqrt{\pi}(\|B\hat{\phi}\|+r)}|f''(x)|$. Let $g(\phi)=F(\hat{\phi}+\phi)-F(\hat{\phi})-L\phi$ for $\phi\in D(B)$, let

$r>0$, and let $\|\phi_1\|_B, \|\phi_2\|_B \leqq r$. Then

$$\begin{aligned}
\|g(\phi_1)-g(\phi_2)\|^2 &= \int_0^\pi \lambda^2 |f(\hat{\phi}(x)+\phi_1(x)) \\
&\qquad - f(\hat{\phi}(x)+\phi_2(x)) - f'(\hat{\phi}(x))(\phi_1(x)-\phi_2(x))|^2\,dx \\
&= \lambda^2 \int_0^\pi \left[\int_{\hat{\phi}(x)+\phi_2(x)}^{\hat{\phi}(x)+\phi_1(x)} \left(\int_{\hat{\phi}(x)}^{s} f''(\tau)\,d\tau\right) ds\right]^2 dx \\
&\leqq \lambda^2 c(r)^2 \int_0^\pi [(\phi_1(x)^2-\phi_2(x)^2)/2]^2\,dx \\
&\leqq (\lambda^2 c(r)^2/4)\pi \|B(\phi_1+\phi_2)\|^2 \|B(\phi_1-\phi_2)\|^2 \\
&\leqq c_1(r)^2 \|\phi_1-\phi_2\|_B^2
\end{aligned}$$

where $c_1(r)=\lambda c(r)\sqrt{\pi}\, r$. A similar argument shows that $\|L\phi\| \leqq \lambda\ (\sup_{|x|\leqq\sqrt{\pi}\|B\hat{\phi}\|} |f'(x)|)\cdot \|\phi\|_B$ for all $\phi \in D(B)$ and $L \overset{\text{def}}{=} F'(\hat{\phi})$ is continuous in $\hat{\phi}$ as in (ii) of (2.4). Next let $\phi \in D(B)$ and define $\hat{c}(r)=\lambda \sup_{|x|\leqq\sqrt{\pi r}} |f'(x)|$. Then $F(\phi)$ is absolutely continuous, $(F(\phi))(0)=(F(\phi))(\pi)=0$, and $(F(\phi))' \in X$, since

$$\begin{aligned}
\int_0^\pi |(F(\phi))'(x)|^2\,dx &= \int_0^\pi |\lambda f'(\phi(x))\phi'(x)|^2\,dx \\
&\leqq \hat{c}(\|B\phi\|)^2 \|B\phi\|^2.
\end{aligned}$$

Thus, we have shown that (2.5) holds with $c_2(r)=\hat{c}(r)r$.

We may now apply Theorems 2.1, 2.2, 2.3, and 2.4. Our next goal will be to prove

THEOREM 3.1. *Let $\alpha \geqq 0$. For each $(\phi, \psi) \in Y$ the solution $u(t; \phi, \psi)$ of (2.7) exists and is bounded on $[0, \infty)$.*

Proof. We define the Lyapunov functional $V\colon Y \to \mathbb{R}$, $D(V)=Y$, by

$$V(\phi,\psi) = (\tfrac{1}{2})(\|\phi\|_B^2 + \|\psi\|^2) - \lambda \int_0^\pi J(\phi(x))\,dx \tag{3.9}$$

where $J(x)=\int_0^x f(s)\,ds$. It follows from (3.6) that there is a constant $k>0$ such that $J(x) \leqq x^2/(4\lambda)+k$ for all $x \in \mathbb{R}$. Then, for $(\phi,\psi) \in Y$ we have

$$\begin{aligned}
V(\phi,\psi) &\geqq (\tfrac{1}{2})(\|\phi\|_B^2+\|\psi\|^2) - \lambda \int_0^\pi (\phi(x)^2/(4\lambda)+k)\,dx \\
&\geqq (\tfrac{1}{4})\|\phi\|_B^2 + (\tfrac{1}{2})\|\psi\|^2 - \lambda\pi k.
\end{aligned}$$

Thus, for $(\phi,\psi) \in Y$, $0 \leqq t < t_0(\phi,\psi)$, we have

$$\|u(t;\phi,\psi)\|_Y^2 \leqq 4(V(u(t;\phi,\psi)) + \lambda\pi k). \tag{3.10}$$

If $(\phi,\psi) \in D(\mathscr{A})$ then the symmetry of B, (2.9), and (2.10) imply that for $0 \leqq t < t_0(\phi,\psi)$

$$\begin{aligned}
(d/dt)V(u(t;\phi,\psi)) &= \langle B(d/dt)\pi_1 u(t;\phi,\psi), B\pi_1 u(t;\phi,\psi)\rangle \\
&\quad + \langle (d/dt)\pi_2 u(t;\phi,\psi), \pi_2 u(t;\phi,\psi)\rangle \\
&\quad - \lambda\langle f(\pi_1 u(t;\phi,\psi)), (d/dt)\pi_1 u(t;\phi,\psi)\rangle \\
&= \langle -A\pi_1 u(t;\phi,\psi) + (d^2/dt^2)\pi_1 u(t;\phi,\psi) \\
&\quad - \lambda f(\pi_1 u(t;\phi,\psi)), (d/dt)\pi_1 u(t;\phi,\psi)\rangle \\
&= -2\alpha\|(d/dt)\pi_1 u(t;\phi,\psi)\|^2 \\
&= -2\alpha\|\pi_2 u(t;\phi,\psi)\|^2.
\end{aligned} \tag{3.11}$$

Then, (3.11) implies that for all $(\phi, \psi) \in D(\mathcal{A})$, $0 \leqq t < t_0(\phi, \psi)$,

$$V(u(t; \phi, \psi)) = V(\phi, \psi) - 2\alpha \int_0^t \|\pi_2 u(s; \phi, \psi)\|^2 \, ds. \tag{3.12}$$

Since $D(\mathcal{A})$ is dense in Y, the continuity of V and of $u(t; \phi, \psi)$ in the sense of Theorem 2.1 imply that (3.12) must hold for all $(\phi, \psi) \in Y$, $0 \leqq t < t_0(\phi, \psi)$. Since $\alpha \geqq 0$, (3.10) and part (i) of Theorem 2.1 yield the desired conclusion.

THEOREM 3.2. *Let $\alpha \geqq 0$. Define $S(t)y = u(t; y)$ for $t \geqq 0$, $y \in Y$. Then, $S(t)$, $t \geqq 0$, is a dynamical system in Y in the sense that:*

(3.13) *for each $t \geqq 0$, $S(t)$ is a continuous mapping from Y to Y;*

(3.14) *for each $y \in Y$, $S(\cdot)y$ is a continuous function from $[0, \infty)$ to Y;*

(3.15) *$S(0) = I$, the identity mapping on Y;*

(3.16) *$S(t)S(s)y = S(t+s)y$ for all $s, t \geqq 0$, $y \in Y$.*

Proof. The conclusions are true by virtue of Theorem 2.1 and Theorem 3.1. We observe that (3.16) holds because of the uniqueness of solutions to (2.7).

The following theorem is proved in [7, Thm. 4.3.3]:

THEOREM 3.3. *Let $S(t)$, $t \geqq 0$ be a dynamical system in the complete metric space C. Let $\hat{y} \in C$ and let the orbit $\gamma(\hat{y}) \stackrel{\text{def}}{=} \{S(t)\hat{y} : t \geqq 0\}$ of $\hat{y}$ be a pre-compact set in C. Then the ω-limit set $\omega(\hat{y}) \stackrel{\text{def}}{=} \{y \in C$: there exists $t_n \to \infty$ such that $S(t_n)\hat{y} \to y\}$ of $\hat{y}$ is nonempty, compact, connected, and* dist $(S(t)\hat{y}, \omega(\hat{y})) \to 0$ *as $t \to \infty$.*

THEOREM 3.4. *Let $\alpha > 0$. Let $E \stackrel{\text{def}}{=} \{(\phi, \psi) \in Y : \phi \in D(A),\ A\phi + F(\phi) = 0$, and $\psi = 0\}$ be the set of equilibrium points of* (2.10). *Then,* dist $(u(t; \phi, \psi), E) \to 0$ *as $t \to \infty$ for all $(\phi, \psi) \in Y$.*

Proof. Since $\alpha > 0$, we may take $\gamma > 0$ in (2.3). By virtue of Theorems 2.2, 3.1, 3.2, and 3.3 the solutions $u(t; \phi, \psi)$ of (2.7) define a dynamical system in which orbits are pre-compact and ω-limit sets are nonempty, compact, connected, and dist $(u(t; \phi, \psi), \omega(\phi, \psi)) \to 0$ as $t \to \infty$. Thus, it suffices to show that if $(\hat{\phi}, \hat{\psi}) \in \omega(\phi, \psi)$ for a given $(\phi, \psi) \in Y$, then $(\hat{\phi}, \hat{\psi}) \in E$. Since (3.10) and (3.12) hold for all $(\phi, \psi) \in Y$, we have that $V(u(t; \phi, \psi))$ is bounded below and nonincreasing for $t \geqq 0$. Let $q = \lim_{t\to\infty} V(u(t; \phi, \psi))$ and let $(\hat{\phi}, \hat{\psi}) \in \omega(\phi, \psi)$, so that $(\hat{\phi}, \hat{\psi}) = \lim_{n\to\infty} u(t_n; \phi, \psi)$ for some sequence $t_n \to \infty$. Since V is continuous, $V((\hat{\phi}, \hat{\psi})) = q$. But if $(\hat{\phi}, \hat{\psi}) \in \omega(\phi, \psi)$, then so is $u(t; \hat{\phi}, \hat{\psi})$ for all $t \geqq 0$, so that $V(u(t; \hat{\phi}, \hat{\psi})) = q$ for all $t \geqq 0$. From (3.12) we must have that $\pi_2 u(t; \hat{\phi}, \hat{\psi}) = \hat{\psi} = 0$ for all $t \geqq 0$. Now we observe from (3.1) that $(d/dt)\pi_1 T(t)(\phi, \psi) = \pi_2 T(t)(\phi, \psi)$ for all $(\phi, \psi) \in Y$, $t \in \mathbb{R}$. Then we see from (2.7) that $(d/dt)\pi_1 u(t; \phi, \psi) = \pi_2 u(t; \phi, \psi)$ for all $(\phi, \psi) \in Y$ and $t \geqq 0$. Thus, we must have $(d/dt)\pi_1 u(t; \hat{\phi}, \hat{\psi}) = 0$ for all $t \geqq 0$, so that $u(t; \hat{\phi}, \hat{\psi}) = (\hat{\phi}, \hat{\psi})$ for all $t \geqq 0$. From (2.7) we see that $T(t)(\hat{\phi}, \hat{\psi}) = (\hat{\phi}, \hat{\psi}) - \int_0^t T(s)\mathcal{F}(\hat{\phi}, \hat{\psi}) \, ds$ is differentiable with respect to t, so that $(\hat{\phi}, \hat{\psi}) \in D(\mathcal{A})$ and $\mathcal{A}(\hat{\phi}, \hat{\psi}) = -\mathcal{F}(\hat{\phi}, \hat{\psi})$, or equivalently, $(\hat{\phi}, \hat{\psi}) \in E$.

The proof of the following theorem is given in [2, Theorem 5.5].

THEOREM 3.5. *Let $\lambda_n = n^2/f'(0)$, $n = 0, 1, \cdots$, and let $\lambda_n < \lambda \leqq \lambda_{n+1}$ for some $n = 0, 1, \cdots$. There exist exactly $2n+1$ functions $\phi \in X$ satisfying*

$$\begin{aligned} &A\phi + F(\phi) = 0, \qquad \phi \in D(A), \quad \textit{or equivalently}, \\ &\phi''(x) + \lambda f(\phi(x)) = 0, \qquad 0 < x < \pi, \quad \phi(0) = \phi(\pi) = 0. \end{aligned} \tag{3.17}$$

If we designate these functions as ϕ_0 and $\phi_{k,\pm}$, $k = 1, \cdots, n$, then $\phi_0 \equiv 0$ and for

$k=1,\cdots,n$, $\phi'_{k,+}(0)>0$, $\phi'_{k,-}(0)<0$, $\phi_{k,\pm}$ *vanishes exactly* $(k-1)$ *times in* $(0,\pi)$, *and* $\operatorname{sgn} f(\phi_{k,\pm}(x))=\operatorname{sgn}\phi_{k,+}(x)$ *for all* $x\in(0,\pi)$.

In [2] it is proved that as λ increases through λ_n the pair of equilibrium solutions $\phi_{n,\pm}$ bifurcates from ϕ_0 in the sense that $\sup_{0\leqq x\leqq\pi}|\phi'_{n,\pm}(x)|\to 0$ as $\lambda\to\lambda_n$ from above. As our last theorem we prove

THEOREM 3.6. *Let* $\alpha>0$. *Let* λ_n, $n=1,2,\cdots$, *and let* ϕ_0, $\phi_{k,\pm}$, $k=1,\cdots,n$, *be as in Theorem* 3.5. *The following are true*:

(3.18) *if* $0\leqq\lambda\leqq\lambda_1$ *and* $(\phi,\psi)\in Y$, *then the solution* $u(t;\phi,\psi)$ *of* (2.7) *converges in* Y *to* $(0,0)$ *as* $t\to\infty$;

(3.19) *if* $\lambda_n<\lambda\leqq\lambda_{n+1}$ *for some* $n=1,2,\cdots$ *and* $(\phi,\psi)\in Y$, *then the solution* $u(t;\phi,\psi)$ *of* (2.7) *converges in* Y *to exactly one of* $(\phi_0,0)$, $(\phi_{k,\pm},0)$, $k=1,\cdots,n$ *as* $t\to\infty$;

(3.20) *if* $\lambda_n<\lambda\leqq\lambda_{n+1}$ *for some* $n=1,2,\cdots$, *then* $(\phi_{1,\pm},0)$ *are locally asymptotically stable in the sense of Theorem* 2.3 *and* $(\phi_0,0)$, $(\phi_{k,\pm},0)$, $k=2,\cdots,n$, *are unstable in the sense of Theorem* 2.4.

Our proof of Theorem 3.6 is based on the ideas of [7, § 5.3]. We first prove the following lemma:

LEMMA 3.7. *Let* $a\in C[0,r;\mathbb{R}]$ *with* $r>0$ *and let* $g,h\in C^2[0,r;\mathbb{R}]$ *such that* $g(0)=h(0)$, $g'(0)=h'(0)>0$, *and* $h''+ah=0$ *on* $[0,r]$. *The following hold*:

(3.21) *if* $h>0$ *on* $(0,r)$ *and* $g''+ag\geqq h''+ah$ *on* $(0,r)$, *then* $g\geqq h$ *on* $(0,r]$;

(3.22) *if* $h>0$ *on* $(0,r)$ *and* $g''+ag<(\leqq)h''+ah$ *on* $(0,r)$, *then* $g<(\leqq)h$ *on* $(0,r]$;

(3.23) *if* $g>0$ *on* $(0,r)$ *and* $g''+ag<(\leqq)h''+ah$ *on* $(0,r)$, *then* $g<(\leqq)h$ *on* $(0,r]$.

Proof. To prove (3.21) observe first that $\lim_{x\to 0}g(x)/h(x)=\lim_{x\to 0}g'(x)/h'(x)=1$. Since $(g/h)'=(hg'-gh')/h^2$, it suffices to show that $hg'-gh'\geqq 0$ on $(0,r)$. But $(hg'-gh')(0)=0$ and $(hg'-gh')'=hg''-gh''\geqq hh''+ah^2-agh-gh''=(h-g)(h''+ah)=0$ on $(0,r)$. The proof of (3.22) is similar to the proof of (3.21) except that one shows $hg'-gh'<(\leqq)0$ in $(0,r)$. To prove (3.23) observe that since $h(0)=0$ and $h'(0)>0$, there exists $r_1\in(0,r)$ such that $h>0$ on $(0,r_1)$. Assume $h(r_1)=0$. Then $g<(\leqq)h$ on $(0,r_1]$ by (3.22). So $0<g(r_1)<(\leqq)h(r_1)=0$, yielding a contradiction. Thus, $h(r_1)>0$ and in fact $h>0$ on $(0,r)$. Now (3.23) follows immediately from (3.22).

Proof of Theorem 3.6. Parts (3.18) and (3.19) follow immediately from Theorems 3.4 and 3.5, so we have only to prove (3.20). Let $\lambda_n<\lambda\leqq\lambda_{n+1}$, $n=1,2,\cdots$, and let $\hat\phi$ denote one of the equilibrium solutions ϕ_0, $\phi_{k,\pm}$, $k=1,\cdots,n$. Let $\mathscr{L}(\phi,\psi)=(0,F'(\hat\phi)\phi)$ for all $(\phi,\psi)\in Y$ and let $\hat T(t)$, $t\in\mathbb{R}$, be the strongly continuous group generated by $\mathscr{A}+\mathscr{L}$. We shall give an explicit formula for $\hat T(t)$, $t\in\mathbb{R}$, by using the method of separation of variables to solve

$$
\begin{aligned}
&w_{tt}+2\alpha w_t-w_{xx}-\lambda f'(\hat\phi)w=0, \qquad 0<x<\pi,\quad t\in\mathbb{R},\\
&w(x,0)=\phi(x),\quad w_t(x,0)=\psi(x),\quad w(0,t)=w(\pi,t)=0.
\end{aligned}
\tag{3.24}
$$

Let $\mu_1<\mu_2<\cdots$ be the eigenvalues and $\hat\chi_1,\hat\chi_2,\cdots$ the corresponding orthonormal sequence of eigenfunctions corresponding to

$$
\begin{aligned}
&\chi''(x)+\lambda f'(\hat\phi(x))\chi(x)=-\mu\chi(x), \qquad 0<x<\pi,\\
&\chi(0)=\chi(\pi)=0
\end{aligned}
\tag{3.25}
$$

[3, Thm. 4.2, p. 199 and Thm. 2.1, p. 2.12]. Then, for all $(\phi, \psi) \in Y$, $t \in \mathbb{R}$,

$$\hat{T}(t)(\phi, \psi) = \Big(\sum_{n=1}^{\infty} [\langle \phi, \hat{\chi}_n \rangle \hat{U}_n(t) + \langle \psi, \hat{\chi}_n \rangle \hat{V}_n(t)] \hat{\chi}_n, \tag{3.26}$$
$$\sum_{n=1}^{\infty} [\langle \phi, \hat{\chi}_n \rangle \hat{U}'_n(t) + \langle \psi, \hat{\chi}_n \rangle \hat{V}'_n(t)] \hat{\chi}_n \Big)$$

where $\hat{U}_n(t)$ is as in (3.2) and $\hat{V}_n(t)$ is as in (3.3), but with n^2 replaced by μ_n.

From the representation (3.26) it is clear that $(\hat{\phi}, 0)$ will be locally asymptotically stable in the sense of (2.12) provided $\mu_1 > 0$ (that is, $\hat{M} \geqq 1$ and $\hat{\gamma} > 0$ can be chosen as in (2.11)). Further, we claim that $(\hat{\phi}, 0)$ will be unstable in the sense of (2.16) provided $\mu_1 < 0$. To establish this claim let $Z_n = \{(\phi, \psi) \in Y : \phi = a\hat{\chi}_n,\ \psi = b\hat{\chi}_n$ for some $a, b \in \mathbb{R}\}$, $n = 1, 2, \cdots$, and observe that $Y = \sum_{n=1}^{\infty} Z_n \oplus$ and $\hat{T}(t) Z_n \subset Z_n$ for $t \in \mathbb{R}$, $n = 1, 2, \cdots$. If $\mu_n < 0$ let $Z_{n,1} = \{(a\hat{\chi}_n,\ a(-\alpha + \sqrt{\alpha^2 - \mu_n})\hat{\chi}_n) \in Z_n : a \in \mathbb{R}\}$, $Z_{n,2} = \{(a\hat{\chi}_n,\ a(-\alpha - \sqrt{\alpha^2 - \mu_n})\hat{\chi}_n) \in Z_n : a \in \mathbb{R}\}$, and observe that $Z_n = Z_{n,1} \oplus Z_{n,2}$. Also, $T(t) Z_{n,1} \subset Z_{n,1}$ for $t \in \mathbb{R}$, since from (3.24), $T(t)(a\hat{\chi}_n,\ a(-\alpha + \sqrt{\alpha^2 - \mu_n})\hat{\chi}_n) = (a\, e^{(-\alpha + \sqrt{\alpha^2 - \mu_n})t} \hat{\chi}_n,\ a(-\alpha + \sqrt{\alpha^2 - \mu_n})\, e^{(-\alpha + \sqrt{\alpha^2 - \mu_n})t} \hat{\chi}_n)$, and similarly $T(t) Z_{n,2} \subset Z_{n,2}$. If $\mu_N < 0$ and $\mu_{N+1} \geqq 0$, let $Y_1 = \sum_{n=1}^{N} Z_{n,1} \oplus$ and $Y_2 = (\sum_{n=1}^{N} Z_{n,2} \oplus) \oplus (\sum_{n=N+1}^{\infty} Z_n \oplus)$. Then Y_1 and Y_2 satisfy the hypothesis of Theorem 2.4 and furthermore (2.14) and (2.15) are satisfied for appropriately chosen $K \geqq 1$ and $\omega > 0$.

Now take $\hat{\phi} = \phi_{1,+}$ and we will show that $\mu_1 > 0$. From Theorem 3.5, $\hat{\phi} > 0$ on $(0, \pi)$, $f(\hat{\phi}) > 0$ on $(0, \pi)$, and $\hat{\phi}'(0) > 0$. Let χ be the unique solution of

$$\chi'' + \lambda f'(\hat{\phi})\chi = 0, \qquad 0 < x < \pi, \tag{3.27}$$
$$\chi(0) = 0, \qquad \chi'(0) = f'(0).$$

We claim that $\chi > 0$ on $(0, \pi]$. To establish this claim define $\rho(x) = f(\hat{\phi}(x))/\hat{\phi}'(0)$. Then $\rho > 0$ on $(0, \pi)$, $\rho(0) = 0$, and $\rho'(0) = f'(0) > 0$. Using (3.17) and (3.7) we have that $\rho'' + \lambda f'(\hat{\phi})\rho = f''(\hat{\phi})\hat{\phi}'^2/\phi'(0) < 0$ on $(0, \pi)$. By (3.23) $\rho < \chi$ on $(0, \pi]$, which implies $\chi > 0$ on $(0, \pi]$. Set $\hat{\chi}(x) = \hat{\chi}_1(x) f'(0)/\hat{\chi}'_1(0)$ ($\hat{\chi}'_1(0) \neq 0$, since $\hat{\chi}_1(0) = \hat{\chi}'_1(0) = 0$ implies $\hat{\chi}_1 \equiv 0$). Then $\hat{\chi}'(0) > 0$ and since $\hat{\chi}_1$ has no zeros in $(0, \pi)$ [3, Thm. 2.1, p. 212], $\hat{\chi} > 0$ on $(0, \pi)$. Now assume $\mu_1 \leqq 0$, so that $\mu_1 \hat{\chi} \leqq 0$. Then $0 = \chi'' + \lambda f'(\hat{\phi})\chi = \hat{\chi}'' + (\mu_1 + \lambda f'(\hat{\phi}))\hat{\chi} \leqq \hat{\chi}'' + \lambda f'(\hat{\phi})\hat{\chi}$ on $(0, \pi)$ and by (3.21) we must have $\hat{\chi} \geqq \chi > 0$ on $(0, \pi]$. But $\hat{\chi}(\pi) = 0$ and so the assumption $\mu_1 \leqq 0$ must be false.

Next, take $\hat{\phi} = \phi_0$. From (3.25) we see that $\mu_1 = 1 - \lambda f'(0) < 0$, since $\lambda > \lambda_1 = 1/f'(0)$. Lastly, take $\hat{\phi} = \phi_{k,+}$, $k = 2, \cdots, n$. From Theorem 3.5 and the fact that $\phi'' = -\lambda f(\phi)$ on $(0, \pi)$ we see that there exists $x_1 \in (0, \pi)$ such that $\hat{\phi}(x_1) < 0$ and $\hat{\phi}'(x_1) = 0$. Let χ be as in (3.27) for this $\hat{\phi}$. We claim that $\chi(x_1) < 0$. To see this claim observe that from (3.17) $\hat{\phi}'$ satisfies $(\hat{\phi}')'' + \lambda f'(\hat{\phi})\hat{\phi}' = 0$ on $[0, \pi]$. Then the Wronskian $\hat{\phi}'\chi' - \hat{\phi}''\chi$ is constant on $[0, \pi]$, since $\hat{\phi}'$ and χ are linearly independent solutions ($\hat{\phi}'(0) = 1$, $\chi(0) = 0$) of the same homogeneous equation [3, Thm. 6.1, p. 83]. Thus, $\hat{\phi}'(x_1)\chi'(x_1) - \hat{\phi}''(x_1)\chi(x_1) = -\hat{\phi}''(x_1)\chi(x_1) = \hat{\phi}'(0) f'(0) > 0$, since $\phi'(0) > 0$ by Theorem 3.5 and $f'(0) > 0$ by (3.5). By Theorem 3.5, $\hat{\phi}''(x_1) = -\lambda f(\hat{\phi}(x_1)) > 0$, which means $\chi(x_1) < 0$ as claimed. As before set $\hat{\chi}(x) = \hat{\chi}_1(x) f'(0)/\hat{\chi}'_1(0)$. Assume that $\mu_1 \geqq 0$. Then, $0 = \chi'' + \lambda f'(\hat{\phi})\chi = \hat{\chi}'' + (\mu_1 + \lambda f'(\hat{\phi}))\hat{\chi} \geqq \hat{\chi}'' + \lambda f'(\hat{\phi})\hat{\chi}$ and by (3.23) we must have $\hat{\chi} \leqq \chi$ on $(0, \pi]$. But $\hat{\chi} > 0$ on $(0, \pi)$ and $\chi(x_1) < 0$, so that the assumption $\mu_1 \geqq 0$ is false. The cases $\phi_{k,-}$, $k = 1, \cdots, n$ are handled in a similar fashion.

REFERENCES

[1] A. J. CALLEGARI AND E. L. REISS, *Nonlinear stability problems for the sine-Gordon equation*, J. Math. and Physics, 14 (1973), pp. 267–276.

[2] N. CHAFEE AND E. F. INFANTE, *A bifurcation problem for a nonlinear parabolic equation*, Applicable Anal., 4 (1974), pp. 17–37.

[3] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

[4] G. DARBO, *Punti uniti in transformazioni a condominio non compatto*, Rend. Sem. Mat. Univ. Padova, 24 (1955), pp 84–92.

[5] R. W. DICKEY, *Stability theory for the damped sine-Gordon equation*, SIAM J. Appl. Math., 30 (1976), pp. 248–262.

[6] ———, *Asymptotic behavior of solutions of hyperbolic semilinear partial differential equations*, J. Math. Anal. Appl., 55 (1976), pp. 380–393.

[7] D. HENRY, *Geometric theory of semilinear parabolic equations*, monograph, to appear.

[8] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.

[9] K. KURATOWSKI, *Topology*, *Vol.* I, Academic Press, New York, 1966.

[10] M. REED, *Abstract Non-Linear Wave Equations*, Springer-Verlag, New York, 1976.

[11] I. SEGAL, *Non-linear semi-groups*, Ann. Math., 78 (1963), pp. 339–364.

[12] H. F. WEINBERGER, *Partial Differential Equations*, Ginn-Blaisdell, Waltham, MA, 1965.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0006 $1.00/0

SOME OPERATIONAL FORMULAS INVOLVING THE OPERATORS xD, $x\Delta$ AND FRACTIONAL DERIVATIVES*

RICHARD TREMBLAY†

Abstract. We consider the operator $\{D^{1-\delta}\prod_{j=1}^{m}(xD+\alpha_j)^r x^\delta\}^n$ ($\delta=0$ or 1; m, n, r integers, and α_j arbitrary) which contains as a special case the operator $(D(xD)^r)^n$ previously studied by Carlitz. We also consider the analogous operator involving the finite difference operator Δ. Some general operational formulas are established from which interesting relationships may be deduced. Further generalizations of operational formulas for the fractional derivative are also given. In particular a law of exponents is deduced for the general fractional operator $D_z^\alpha(z^\alpha D_z^\alpha)^r$ where α may be nonintegral.

1. Introduction. The fractional derivative of the function $F(z)$ is a generalization of the familiar derivative $D^nF(z)=d^nF(z)/(dz)^n$ where the order "n" is replaced by arbitrary (integral, rational, irrational or complex) "α", and denoted by $D_z^\alpha F(z)$. The literature contains many examples of the use of fractional derivatives in ordinary differential [9] and integral equations [8]. There exists an extensive fractional calculus for D_z^α and the most important representations which have been proposed are reviewed in [13]. This reference contains a list of selected formulas and theorems on fractional differentiation such as the Leibniz rule, the law of exponents, the generalized Taylor's series, etc.

This present paper has two parts. In the first (§§ 2 and 3) we treat the following operational formulas

$$\left\{D^{1-\delta}\prod_{j=1}^{m}(xD+\alpha_j)^r x^\delta\right\}^n = D^{(1-\delta)\theta n}x^{\delta(1-\theta)n}\prod_{j=1}^{m}(xD+\alpha_j+1-\theta n)_n^r x^{\delta\theta n}D^{(1-\delta)(1-\theta)n} \tag{1.1}$$

where δ, $\theta=0$ or 1, and the equivalent of (1.1) for the difference operator Δ defined by $\Delta f(x)=f(x+1)-f(x)$

$$\prod_{i=0}^{n-1}\left[\Delta^{1-\delta}\prod_{j=1}^{m}\{[x+\delta i-(1-\delta)(i+1)]\Delta+\alpha_j\}^r(x+i)^\delta\right] = \Delta^{(1-\delta)\theta n}(x)_n^{\delta(1-\theta)}\prod_{j=1}^{m}(\{x+n(\delta-\theta)\}\Delta+\alpha_j+1-\theta n)_n^r(x)_n^{\delta\theta}\Delta^{(1-\delta)(1-\theta)n}, \tag{1.2}$$

δ, $\theta=0$ or 1 and $(x)_i=x(x+1)\cdots(x+i-1)$ for $i\geqq 1$, $(x)_0=1$. (Each of the above relations contains four identities; we use the parameters δ and θ for brevity).

In these identities r, m and n are nonnegative integers and the α_j are arbitrary constants.

Note that the operators

$$((x+q)\Delta+\alpha)\Delta \quad \text{and} \quad ((x+s)\Delta+\beta)\Delta$$

where α, β, q and s are constants, do not commute; it is understood that the products in

* Received by the editors August 18, 1977, and in final revised form May 19, 1978.

† Department of Mathematics, Royal Military College of Canada, Kingston, Ontario, K7L 2W3. This work was supported by the Defense Research Board under Grant DRB 3610-607.

(1.2) mean

$$\prod_{i=0}^{n-1} \mu_i = \mu_0 \cdot \mu_1 \cdot \cdots \cdot \mu_{n-1} \tag{1.3}$$

and must be read from right to left.

The proofs of these results are given in § 2.

The operators involved in (1.1) contain the differential operator studied previously by Carlitz [4]

$$(D(xD)^r)^n = \sum_{s=0}^{rn} A^{(r)}(n, s)x^s D^{s+n} \tag{1.4}$$

where

$$A^{(r)}(n, s) = \frac{1}{s!} \sum_{j=0}^{s} (-1)^{s-j} \binom{s}{j} (j+1)^r \cdots (j+n)^r. \tag{1.5}$$

The expression (1.5) has been obtained by application of a general result developed in [3, p. 746] and $A^{(r)}(n, s)$ contains as a special case the Stirling numbers of the second kind.

The particular operator $(DxD)^n$ has been introduced by Lardner [11] and generalized by many authors, for example Osipov [14] and more recently Al-Salam and Ismail [2].

Many other operational formulas can be deduced from (1.1) and (1.2). For example, we easily prove

$$(D^m(x^m D^m)^r)^n = D^{mn}(x^{mn} D^{mn})^r \tag{1.6}$$

using the well-known Boole's formula

$$x^n D^n = (xD - n + 1)_n. \tag{1.7}$$

Other formulas are given in § 3 which involve the numbers $A^{(r)}(n, s)$.

Now, the question is: Can we generalize these results to fractional order? The second part of this paper (§ 4) is devoted to this question. The equivalent of (1.1) for the fractional derivative D_z^α is given in Theorem 4.2. Moreover, the formula (1.6) suggests the fractional operator $D_z^\alpha(z^\alpha D_z^\alpha)^r$ and we prove that (1.6) remains valid if m is an arbitrary number. More precisely, we show that it is a consequence of a law of exponents for the operator $D_z^\alpha(z^\alpha D_z^\alpha)^r$ given in Theorem 4.3. This theorem is a generalization of the law of exponents $D_z^\alpha D_z^\beta = D_z^{\alpha+\beta}$ which appears in Theorem 4.1. Note that the extension of (1.2) to the finite difference of fractional order Δ^α remains to be done. To establish such a result, we need the equivalent of Theorem 4.1 for $\Delta^\alpha \Delta^\beta = \Delta^{\alpha+\beta}$ which, to the best of the author's knowledge, does not exist. Little attention seems to have been given to the operator Δ^α. The author knows only of Diaz and Osler's paper [7] which treats this subject.

2. Proofs of (1.1) and (1.2). To establish the proofs of these operational formulas, we can, as usual for such formulas, proceed by mathematical induction on the integers m, n and r, without particular difficulty. We can also readily prove these relationships by operating with each side respectively of (1.1) and (1.2) on x^s and $(x)_s$. If there is equality for $s = 0, 1, 2, \cdots$, according to theorems previously stated by Carlitz [6], the operators involved on each side are equivalent. The important elements of the inductive proof of (1.1) and (1.2) are the following principles.

A product of operators of the form D and $(xD + \alpha_j)$ (or Δ and $[(x+s)\Delta + \alpha_j]$) may be rearranged at will, provided that if qD's (or $q\Delta$'s) are moved from the left to the right of a particular operator $(xD+\alpha_j)$ (or $[(x+s)\Delta+\alpha_j]$), then this operator becomes $(xD+\alpha_j+q)$ (or $[(x+s+q)\Delta+\alpha_j+q]$). For instance, we have

$$D^q(xD+\alpha_j)=(xD+\alpha_j+q)D^q \tag{2.1}$$

and similarly

$$\Delta^q[(x+s)\Delta+\alpha_j]=[(x+s+q)\Delta+\alpha_j+q]\Delta^q. \tag{2.2}$$

Similarly, if qD's (or $q\Delta$'s) are moved from the right to the left, we have the same result as if $-qD$'s (or $-q\Delta$'s) are moved from the left to the right.

The same ideas apply in reverse in a product of operators x and $(xD+\alpha_j)$ (or $(x+s)$ and $[(x+s)\Delta+\alpha_j]$). For instance, we have

$$(xD+\alpha_j)x^q=x^q(xD+\alpha_j+q) \tag{2.3}$$

and

$$[(x+s)\Delta+\alpha_j](x+s)_q=(x+s)_q[(x+s+q)\Delta+\alpha_j+q]. \tag{2.4}$$

Finally, $(xD+\alpha)$ and $(xD+\beta)$ commute. Moreover, $[(x+s)\Delta+\alpha]$ and $[(x+q)\Delta+\beta]$ do not commute. The reason is that when Δ moves from the left to the right of $(x\Delta+\alpha_j)$, $x\Delta$ changes to $(x+1)\Delta$. We must pay special attention to the products involved in formula (1.2), whose sense is given by (1.3).

Note that these principles readily prove Boole's law (1.7) and its equivalent in terms of Δ

$$(x)_n\Delta^n=(x\Delta-n+1)_n, \tag{2.5}$$

also the following

$$D^nx^n=(xD+1)_n \tag{2.6}$$

and

$$\Delta^n(x-n)_n=(x\Delta+1)_n \tag{2.7}$$

which will be of use in the next section.

3. Other operational formulas. We next deduce some formulas which are more or less direct consequences of the main results (1.1) and (1.2). In particular, we obtain some formulas involving the coefficient $A^{(r)}(n,s)$ defined by (1.5).

First, with the help of (1.7), we can easily deduce from (1.1) the following relationship

$$(D^{1-\delta}(xD)^rx^\delta)^n=D^{(1-\delta)n}(x^nD^n)^rx^{\delta n} \tag{3.1}$$

where $\delta=0$ or 1. Formula (3.1) leads us readily to the following expression

$$(D(xD)^r)^n\cdot(D(xD)^s)^m=\begin{cases}x^{-n}(x(Dx)^{r-s})^n(D(xD)^s)^{m+n} & \text{if } r\geqq s,\\ (D(xD)^r)^{m+n}(x(Dx)^{s-r})^mx^{-m} & \text{if } r\leqq s,\end{cases} \tag{3.2}$$

and more generally

$$\begin{aligned}&(D^{r_1}(x^{r_1}D^{r_1})^{n-1})^{m_1}\cdot(D^{r_2}(x^{r_2}D^{r_2})^{n-2})^{m_2}\cdots\\ &\qquad\cdots(D^{r_{n-1}}x^{r_{n-1}}D^{r_{n-1}})^{m_{n-1}}D^{r_nm_n}\\ &=D^{r_1m_1}x^{r_1m_1}D^{r_1m_1+r_2m_2}x^{r_1m_1+r_2m_2}\cdots\\ &\qquad\cdots D^{r_1m_1+\cdots+r_{n-1}m_{n-1}}x^{r_1m_1+\cdots+r_{n-1}m_{n-1}}D^{r_1m_1+\cdots+r_nm_n}.\end{aligned} \tag{3.3}$$

The special case $r_i = m_i = 1$ $(i = 1, 2, \cdots, n)$ of (3.3) can be found in [6, p. 253]. Similarly, with the help of (2.5), we obtain from (1.2) the analogue of (3.1) as

$$(3.4)\quad \prod_{i=0}^{n-1} [\Delta^{1-\delta}\{[x+\delta i-(1-\delta)(i+1)]\Delta\}^r (x+i)^\delta] = \Delta^{(1-\delta)n}[(x+n\delta-n)_n\Delta^n]^r(x)_n^\delta$$

where $\delta = 0$ or 1.

Note that (3.4) with $\delta = 0$ can be written in the more elegant form

$$(3.5)\qquad [(\Delta x)^r\nabla]^n = [\Delta^n x^{(n)}]^r\nabla^n$$

where ∇ is defined by $\nabla f(x) = f(x) - f(x-1)$ and $x^{(0)} = 1$, $x^{(n)} = x(x-1)\cdots(x-n+1)$ for $n \geqq 1$.

The above formulas are obtained from (1.1) and (1.2) when we put $\alpha_j = 0$. The replacement $\alpha_j = -jn$ leads us easily to

$$(3.6)\qquad \begin{aligned} &\{x^\delta(xD-n+1)^r(xD-2n+1)^r\cdots(xD-mn+1)^rD^{1-\delta}\}^n \\ &\quad = D^{(1-\delta)\theta n}x^{(\delta+\theta-\theta\delta)n}(x^{mn}D^{mn})^r x^{(\delta-1)\theta n}D^{(1-\delta)(1-\theta)n} \end{aligned}$$

where δ and θ are 0 or 1, which contains as a special case the formulas (12) and (13) of Klamkin and Newman [10]. These formulas can be useful in the solution of certain differential equations (see [10]). Note that the special cases $\delta = 1$, $\theta = 0$ and $\delta = \theta = 1$ in (3.6) are the same.

The formula (3.1) can be used to prove other types of operational formulas. For example, we now give a direct proof of the "Leibniz formula", proved in [1], for the operator $\theta = x(1+xD) = xDx$,

$$(3.7)\qquad (xDx)^n xuv = x\sum_{s=0}^{n}\binom{n}{s}[(xDx)^{n-s}u]\cdot[(xDx)^s v].$$

Indeed, by use of (3.1) with $r = \delta = 1$ and (2.6), formula (3.7) becomes

$$(3.8)\qquad (xD+1)_n xuv = x\sum_{s=0}^{n}\binom{n}{s}[(xD+1)_{n-s}u]\cdot[(xD+1)_s v].$$

Making the change of variable $x = 1/y$ and using (1.7), we see that $yD_y = -xD_x$ and (3.8) reduces to the classical Leibniz rule

$$(3.9)\qquad D^n uv = \sum_{s=0}^{n}\binom{n}{s}D^{n-s}u\cdot D^s v.$$

Now, with the help of (3.1) with $\delta = 0$, it is easy to obtain from (3.7) the following

$$(3.10)\quad (xDx)^n x(D^n u)\cdot(D^n v) = x^{n+1}\sum_{s=0}^{n}\binom{n}{s}[(DxD)^{n-s}D^s u]\cdot[(DxD)^s D^{n-s}v],$$

which reduces to equation (5.1) of [5, p. 384] if $u = 1/x$.

The analogue of (3.7) for the difference operator Δ is

$$(3.11)\qquad \begin{aligned} \Delta^n(x-n+1)_n uv/x = x\sum_{s=0}^{n}\binom{n}{s}&\{\Delta^{n-s}(x-n+s+1)_{n-s}u(x+s)/x\} \\ &\cdot\{\Delta^s(x-s+1)_s v(x)/x\} \end{aligned}$$

which can be established by induction.

The same elementary change of variable can also serve to obtain other relationships. For example, starting with the equation (6.6) of [4, p. 387]

$$(xD+1)_n^r = \sum_{s=0}^{rn} A^{(r)}(n,s)x^s D^s, \tag{3.12}$$

we can easily deduce

$$(x^n D^n)^r = \sum_{s=0}^{rn} (-1)^{rn-s} A^{(r)}(n,s)(xD+1)_s. \tag{3.13}$$

Special cases $r=1$ of (3.12) and (3.13) are given in [4] and [5]. Formulas (3.12) and (3.13) imply the summation formula

$$A^{(r)}(n,k) = \sum_{s=k}^{rn} A^{(r)}(n,s)\binom{n+s}{n+k}\frac{s!}{k!}(-1)^{rn-s}. \tag{3.14}$$

Now, with the help of (3.14) and [4, p. 387]

$$(x\Delta+1)_n^r = \sum_{s=0}^{rn} A^{(r)}(n,s)(x)_s \Delta^s, \tag{3.15}$$

we can deduce readily that the analogue of (3.13) is

$$((x)_n \Delta^n)^r = \sum_{s=0}^{rn} (-1)^{rn-s} A^{(r)}(n,s)(x\Delta+1)_s. \tag{3.16}$$

Other formulas can be obtained with the help of the fundamental expansions (3.12), (3.13), (3.15) and (3.16). For instance, using the fact that

$$\Delta^n (x-n+1)_n \nabla^n = (\Delta x \nabla)^n \tag{3.17}$$

and

$$\nabla^n (x)_n \Delta^n = (\nabla x \Delta)^n, \tag{3.18}$$

we can obtain from (3.15) and (3.16)

$$\nabla^n (x\Delta+1)_n^r = \sum_{s=0}^{rn} A^{(r)}(n,s)\nabla^{n-s}(\nabla x \Delta)^s \tag{3.19}$$

and

$$((x)_n \Delta^n)^r \nabla^n = \sum_{s=0}^{rn} A^{(r)}(n,s)(-1)^{rn-s}(\Delta x \nabla)^s \nabla^{n-s} \tag{3.20}$$

where ∇ is defined above.

4. Some generalizations involving the fractional derivative. As we mentioned above, the most important representations for the fractional derivative D_z^α as well as a list of selected formulas and theorems on fractional differentiation can be found in [13]. Among these results, the generalizations of the Leibniz rule (3.9), such as

$$D_z^\alpha u(z)v(z) = \sum_{s=-\infty}^{+\infty} \binom{\alpha}{\gamma+s} D_z^{\alpha-\gamma-s} u(z) \cdot D_z^{\gamma+s} v(z), \tag{4.1}$$

play a prominent part in this theory. More complex extensions of the Leibniz rule have been given by Osler [15] (see the bibliography of [13]).

In this section, we generalize certain formulas contained in the previous sections.

Firstly, we prove the validity of (2.1) and (2.3) when q is arbitrary, i.e. the fundamental relationships

$$D_z^\beta(zD+\alpha)f(z)=(zD+\alpha+\beta)D_z^\beta f(z) \tag{4.2}$$

and

$$(zD+\alpha)z^\beta f(z)=z^\beta(zD+\alpha+\beta)f(z) \tag{4.3}$$

for all values of α and β. Formula (4.3) is obvious. For the proof of (4.2), using (4.1) with $\gamma=0$, we have

$$D_z^\beta(zD+\alpha)f(z)=zD_z^\beta[Df(z)]+\beta D_z^{\beta-1}[Df(z)]+\alpha D_z^\beta f(z). \tag{4.4}$$

Here, we must pay special attention to (4.4) because it is not always true that $D_z^\alpha\cdot D_z^\beta=D_z^{\alpha+\beta}$. For instance [13, p. 262]

$$D_z^{-1}Df(z)=\int_0^z f'(z)\,dz=f(z)-f(0)\neq D_z^0f(z)=f(z).$$

For (4.4), we must appeal to the following theorem on the law of exponents given in [13] (see also [12]).

THEOREM 4.1. *Let $f(z)$ be analytic on the simply connected open set R containing the point $z=0$. Assume $f(0)\neq 0$, and that $p\neq -1,-2,\cdots$. If $p-\alpha\neq -1,-2,\cdots$, then for $z\in R-\{0\}$ we have*

$$D_z^\beta D_z^\alpha z^pf(z)=D_z^{\beta+\alpha}z^pf(z).$$

If $p-\alpha=-N$, $N=1,2,\cdots$, then for $z\in R-\{0\}$,

$$D_z^\beta D_z^\alpha z^pf(z)=D_z^{\beta+\alpha}z^pf(z)-\sum_{n=0}^{N-1}\frac{f^{(n)}(0)\Gamma(1+p+n)}{n!\Gamma(1+p-\alpha-\beta+n)}z^{p-\alpha-\beta+n}.$$

We assume, as always, that all functions on which operators perform satisfy the appropriate existence and differentiability conditions. Now,

$$D_z^\beta[Df(z)]=D_z^{\beta+1}f(z)-\frac{f(0)}{\Gamma(-\beta)}z^{-\beta-1}$$

and

$$\begin{aligned}D_z^\beta(zD+\alpha)f(z)&=z\left(D_z^{\beta+1}f(z)-\frac{f(0)}{\Gamma(-\beta)}z^{-\beta-1}\right)\\&\quad+\beta\left(D_z^\beta f(z)-\frac{f(0)}{\Gamma(-\beta+1)}z^{-\beta}\right)+\alpha D_z^\beta f(z)\\&=(zD+\alpha+\beta)D_z^\beta f(z)\end{aligned}$$

and this completes the proof of (4.2). Similarly, if D_z^β is moved from the right to the left of $(zD+\alpha)$, we have the same result as if $D_z^{-\beta}$ is moved from the left to the right, which can be proved directly with Theorem 4.1 for all values of β.

The following theorem generalizes the result (1.1) to the fractional derivative.

THEOREM 4.2. *Let $f(z)$ be analytic on the simply connected open set R containing the point $z=0$. Assume $f(0)\neq 0$, δ, $\theta=0$ or 1, m, n are positive integers, $t\in\{0,1,2,\cdots\}$, $l\in\{2,3,\cdots,n-1\}$ and $i\in\{2,3,\cdots,l\}$ with $n>2$. If $l\beta\neq lt+i-1$ when $\delta=0$ and*

$n>2$, then for $z\in R-\{0\}$ we have

$$\left\{D_z^{(1-\delta)\beta}\prod_{j=1}^{m}(zD+\alpha_j)^r z^{\delta\beta}\right\}^n f(z)=D_z^{(1-\delta)\theta\beta n}z^{\delta(1-\theta)\beta n}\cdot\prod_{i=0}^{n-1}\prod_{j=1}^{m}[zD+\alpha_j-i\beta+(1-\theta)\beta n]^r\cdot z^{\delta\theta\beta n}D_z^{(1-\delta)(1-\theta)\beta n}f(z). \tag{4.5}$$

If $l\beta=lt+i-1$ when $\delta=0$ and $n>2$, then for $z\in R-\{0\}$,

$$\left\{D_z^{\beta}\prod_{j=1}^{m}(zD+\alpha_j)^r\right\}^n f(z)=D_z^{\theta\beta n}\prod_{i=0}^{n-1}\prod_{j=1}^{m}[zD+\alpha_j-i\beta+(1-\theta)\beta n]^r\cdot D_z^{(1-\theta)\beta n}\left\{f(z)-\sum_{s=0}^{[n/l]\cdot l\beta-1}\frac{f^{(s)}(0)}{s!}z^s\right\} \tag{4.6}$$

where $[n/l]$ is the integer part of n/l.

Proof. By iteration of (4.2), we easily obtain

$$D_z^{\beta}\prod_{j=1}^{m}(zD+\alpha_j)^r f(z)=\prod_{j=1}^{m}(zD+\alpha_j+\beta)^r D_z^{\beta}f(z), \tag{4.7}$$

if $f(z)$ is analytic at $z=0$. Now $D_z^{\beta}f(z)$ is a function in the form $z^{-\beta}H(z)$, if $\beta\neq 1,2,\cdots$, where $H(z)$ is analytic at $z=0$ [13, Thm. 18.1]. Consequently, the function involved in (4.7) is also of this form and we can iterate (4.7) once more. We obtain

$$\left\{D_z^{\beta}\prod_{j=1}^{m}(zD+\alpha_j)^r\right\}^2 f(z)=D_z^{2\theta\beta}\prod_{i=0}^{1}\prod_{j=1}^{m}[zD+\alpha_j-i\beta+2(1-\theta)\beta]^r D_z^{2(1-\theta)\beta}f(z) \tag{4.8}$$

valid for all values of β. This proves (4.5) for $\delta=0$ and $n=1,2$. In (4.8), the function is now in the form $z^{-2\beta}G(z)$, where $G(z)$ is analytic at $z=0$. Also, for $n=3,4,\cdots$, successive application of (4.1), by virtue of Theorem 4.1, forces us to reject in (4.5) the values $-2\beta-1=-1,-2,-3,\cdots,-3\beta-1=-1,-2,\cdots$, etc., except the integers $\beta=0,1,2,\cdots$. In other words, for $n(>2)$ iterations, (4.5) with $\delta=0$ is valid if $l\beta\neq lt+i-1$, where $l=2,\cdots,n-1$, $t=0,1,\cdots$, and $i=2,\cdots,l$. In the special case that $l\beta=lt+i-1$, if we operate termwise on the Maclaurin series of $f(z)$, the presence of the factor $1/\Gamma(1+s-[n/l]\cdot(lt+i-1))$ in the series resulting from $[n/l]$ iterations causes the first $[n/l]\cdot(lt+i-1)$ terms to vanish and we obtain (4.6). By analytic continuation, the result is valid on $R-\{0\}$. The case $\delta=1$ of (4.5) can be easily proven from (4.3) and the proof gives rise to no particular difficulties; therefore this proof is omitted.

For the last example, we shall prove the following result

$$\{D_z^{(1-\delta)}(z^{\alpha}D_z^{\alpha})^r z^{\delta\alpha}\}^n f(z)=D_z^{(1-\delta)\alpha n}(z^{\alpha n}D_z^{\alpha n})^r z^{\delta\alpha n}f(z) \tag{4.9}$$

where $\delta=0$ or 1, with only restrictions $i\alpha\neq -1,-2,\cdots$ for $i=1,2,\cdots,n$ when $\delta=1$ (obtained from Theorem 4.3). We suppose that $f(z)$ is analytic at $z=0$ in (4.9). Relationship (4.9) reduces to (3.1) if $\alpha=1$.

In fact, this formula can be obtained with the help of the following relationship

$$\{D_z^{(1-\delta)\alpha}(z^{\alpha}D_z^{\alpha})^r z^{\delta\alpha}\}\cdot\{D_z^{(1-\delta)\beta}(z^{\beta}D_z^{\beta})^r z^{\delta\beta}\}=D_z^{(1-\delta)(\alpha+\beta)}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^{\delta(\alpha+\beta)} \tag{4.10}$$

where $\delta=0$ or 1. Again, we must pay special attention to the law of exponents (4.10), which is not valid for all values of parameters α and β. Indeed, if $r=1$ and $\delta=0$, we have with $\beta=-\alpha=1$

$$(D_z^{-1}z^{-1}D_z^{-1})\cdot(DzD)f(z)$$
$$=\int_0^z \frac{1}{v}\int_0^v [uf''(u)+f'(u)]\,du\,dv$$
$$=\int_0^z f'(v)\,dv=f(z)-f(0)\neq D_z^0z^0D_z^0f(z)=f(z).$$

The general situation is examined in the following theorem. The study [12] of the analytic properties of fractional derivatives plays a decisive part in this theorem.

THEOREM 4.3. *Let $f(z)$ be analytic on the simply connected open set R containing the point $z=0$. Also, assume $f(0)\neq 0$, $\delta=0$ or 1, $M, N\in\{1,2,3,\cdots\}$ and r is a nonnegative integer. If $\gamma\neq -N$, $\gamma-\beta\neq -M$ when $\delta=0$, and $\gamma+\beta\neq -N$, $\gamma+\alpha+\beta\neq -M$ when $\delta=1$, then for $z\in R-\{0\}$ we have*

$$\{D_z^{(1-\delta)\alpha}(z^\alpha D_z^\alpha)^r z^{\delta\alpha}\}\cdot\{D_z^{(1-\delta)\beta}(z^\beta D_z^\beta)^r z^{\delta\beta}\}\cdot z^\gamma f(z) = \{D_z^{(1-\delta)(\alpha+\beta)}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^{\delta(\alpha+\beta)}\}\cdot z^\gamma f(z); \tag{4.11}$$

if $\gamma\neq -N$, $\gamma-\beta=-M$ when $\delta=0$, then for $z\in R-\{0\}$,

$$\{D_z^\alpha(z^\alpha D_z^\alpha)^r\}\cdot\{D_z^\beta(z^\beta D_z^\beta)^r\}\cdot z^\gamma f(z)=\{D_z^{\alpha+\beta}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r\}\cdot z^\gamma f(z) - \sum_{s=0}^{M-1}\left(\frac{\Gamma(1+\gamma+s)}{\Gamma(1+\gamma-\alpha-\beta+s)}\right)^{r+1}\frac{f^{(s)}(0)}{s!}z^{s+\gamma-\alpha-\beta}. \tag{4.12}$$

Remarks. (i) If $r=0$ and $\delta=0$, Theorem 4.3 reduces to Theorem 4.1.

(ii) If $\alpha=0,1,2,\cdots$ in (4.12), $1/\Gamma(1+\gamma-\alpha-\beta+s)$ is zero and the finite sum vanishes. Relationship (4.11) is clearly true when α and β are natural numbers.

Proof. We shall demonstrate below (4.11) and (4.12) by operating with the operators

$$D_z^{(1-\delta)\beta}(z^\beta D_z^\beta)^r z^{\delta\beta},\qquad D_z^{(1-\delta)\alpha}(z^\alpha D_z^\alpha)^r z^{\delta\alpha}\quad\text{and}$$
$$D_z^{(1-\delta)(\alpha+\beta)}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^{\delta(\alpha+\beta)}$$

(respectively for $\delta=0$ and 1), termwise on the Maclaurin series of $f(z)$. This procedure only proves the results for z inside the circle of convergence of the Maclaurin series. However, we recall Theorem 3.1 in [12] (which describes the analycity of $D_z^\alpha z^p f(z)$ with reference to the three variables z, α and p) and this theorem affirms that all terms in (4.11) and (4.12) are analytic functions of z for $z\in R-\{0\}$. Thus the results are true for z in the full domain $z\in R-\{0\}$ by analytic continuation.

Recalling the fact that [13, p. 243]

$$D_z^\alpha z^p=\frac{\Gamma(1+p)}{\Gamma(1+p-\alpha)}z^{p-\alpha} \tag{4.13}$$

where the only restriction is that $p\neq -1,-2,-3,\cdots$, and expanding $f(z)$ in a Maclaurin series, and operating termwise with the operator $D_z^\beta(z^\beta D_z^\beta)^r$ we get

$$D_z^\beta(z^\beta D_z^\beta)^r z^\gamma f(z)=\begin{cases}\displaystyle\sum_{s=0}^{\infty}\frac{f^{(s)}(0)}{s!}\left(\frac{\Gamma(1+\gamma+s)}{\Gamma(1+\gamma-\beta+s)}\right)^{r+1}z^{\gamma-\beta+s} & \text{if } \gamma-\beta \neq -1,-2,-3,\cdots,\\ \displaystyle\sum_{s=M}^{\infty}\frac{f^{(s)}(0)}{s!}\left(\frac{\Gamma(1+\gamma+s)}{\Gamma(1+\gamma-\beta+s)}\right)^{r+1}z^{\gamma-\beta+s} & \text{if } \gamma-\beta=-M,\ M=1,2,3,\cdots.\end{cases}$$

Now, operating with $D_z^\alpha(z^\alpha D_z^\alpha)^r$ on the above series, we have

$$\{D_z^\alpha(z^\alpha D_z^\alpha)^r\}\cdot\{D_z^\beta(z^\beta D_z^\beta)^r\}z^\gamma f(z)$$

$$(4.14)\qquad =\begin{cases}\displaystyle\sum_{s=0}^{\infty}\frac{f^{(s)}(0)}{s!}\left(\frac{\Gamma(1+\gamma+s)}{\Gamma(1+\gamma-\alpha-\beta+s)}\right)^{r+1} z^{\gamma-\alpha-\beta+s} & \text{if } \gamma-\beta \neq -1,-2,-3,\cdots, \\ \displaystyle\sum_{s=M}^{\infty}\frac{f^{(s)}(0)}{s!}\left(\frac{\Gamma(1+\gamma+s)}{\Gamma(1+\gamma-\alpha-\beta+s)}\right)^{r+1} z^{\gamma-\alpha-\beta+s} & \text{if } \gamma-\beta=-M, \\ & M=1,2,3,\cdots.\end{cases}$$

We note at once that the first series of (4.14) is $D_z^{\alpha+\beta}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^\gamma f(z)$, and thus formula (4.11) with $\delta=0$ is proved. The second series above is just the first series with the first M terms subtracted away, and thus the validity of (4.12) is clear.

The case $\delta=1$ of (4.11) can also easily be proved. In the same way, expanding $f(z)$ in a Maclaurin series and operating termwise with $(z^\beta D_z^\beta)^r z^\beta$ we get

$$(z^\beta D_z^\beta)^r z^\beta \cdot z^\gamma f(z)=\sum_{s=0}^{\infty}\frac{f^{(s)}(0)}{s!}\left(\frac{\Gamma(1+\gamma+\beta+s)}{\Gamma(1+\gamma+s)}\right)^r z^{\gamma+\beta+s}$$

with the simple restriction $\gamma+\beta\neq -N$, $N=1,2,3,\cdots$. Operating next with $(z^\alpha D_z^\alpha)^r z^\alpha$ on the above series, we have

$$\{(z^\alpha D_z^\alpha)^r z^\alpha\}\cdot\{(z^\beta D_z^\beta)^r z^\beta\}z^\gamma f(z)=\sum_{s=0}^{\infty}\frac{f^{(s)}(0)}{s!}\left(\frac{\Gamma(1+\gamma+\alpha+\beta+s)}{\Gamma(1+\gamma+s)}\right)^r z^{\gamma+\alpha+\beta+s}$$

if $\gamma+\beta+\alpha\neq -1,-2,-3,\cdots$. This is simply $(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^{\alpha+\beta}\cdot z^\gamma f(z)$ and then the relationship (4.11) with $\delta=1$ is proved. This completes the proof of Theorem 4.3.

Note that we can prove Theorem 4.3 in another way. Let ${}_zO_\beta^\alpha$ denote a fractional operator defined by

$$(4.15)\qquad {}_zO_\beta^\alpha\{\cdot\}=\frac{\Gamma(\beta)}{\Gamma(\alpha)}z^{1-\beta}D_z^{\alpha-\beta}z^{\alpha-1}\{\cdot\},\qquad \beta\neq 0,-1,-2,\cdots.$$

With the help of (4.13), we can now obtain ${}_zO_\beta^\alpha f(z)$, where $f(z)$ is analytic at $z=0$, by operating on the power series for $f(z)$ term by term. We get

$$(4.16)\qquad {}_zO_\beta^\alpha f(z)=\sum_{s=0}^{\infty}\frac{f^{(s)}(0)}{s!}\frac{(\alpha)_s}{(\beta)_s}z^s.$$

This series has the same circle of convergence as the power series for $f(z)$ about $z=0$ and this operation has only the restriction $\beta\neq 0,-1,-2,\cdots$.Note that this restriction vanishes if we divide both sides of (4.16) by $\Gamma(\beta)$. The result of operating with ${}_zO_\beta^\alpha/\Gamma(\beta)$ on the first $M+1$ terms of the series in (4.16) is zero if $\beta=-M$. Clearly, two operators ${}_zO_b^a$ and ${}_zO_\beta^\alpha$ commute and for the proof of Theorem 4.3, we need only the following properties:

$$(4.17)\qquad {}_zO_\beta^\alpha\, {}_zO_\gamma^\beta f(z)={}_zO_\gamma^\alpha f(z)\qquad \beta,\gamma\neq 0,-1,-2,\cdots$$

and

$$(4.18)\qquad {}_zO_\beta^\alpha z^\gamma f(z)=\frac{\Gamma(\beta)\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)}z^\gamma\, {}_zO_{\beta+\gamma}^{\alpha+\gamma}f(z)\qquad \beta,\alpha+\gamma\neq 0,-1,-2,\cdots$$

which can be easily deduced from the definition (4.15) and (4.16).

Remark. If $\beta+\gamma=-M, M=0,1,2,\cdots$, then

$$(4.19)\quad {}_zO_\beta^\alpha z^\gamma f(z)=\sum_{s=M+1}^{\infty}\frac{f^{(s)}(0)}{s!}\frac{\Gamma(\alpha+\gamma+s)\Gamma(\beta)}{\Gamma(\beta+\gamma+s)\Gamma(\alpha)}z^{\gamma+s},\qquad \beta\neq 0,-1,-2,\cdots$$

since $1/\Gamma(\beta+\gamma+s)=0$ for $s=0,1,2,\cdots,M$. Consequently, if $\beta+\gamma=-M$, $M=0,1,2,\cdots$, we have

$$(4.20)\qquad {}_zO_\theta^\beta\,{}_zO_\beta^\alpha z^\gamma f(z)={}_zO_\theta^\alpha z^\gamma\left\{f(z)-\sum_{s=0}^{M}\frac{f^{(s)}(0)}{s!}z^s\right\}.$$

Now, from (4.15), we have

$$(4.21)\qquad z^\alpha D_z^\alpha=\frac{{}_zO_{1-\alpha}^1}{\Gamma(1-\alpha)}$$

and

$$(4.22)\qquad D_z^\alpha z^\alpha=\Gamma(1+\alpha)\,{}_zO_1^{1+\alpha}.$$

Thus, we have

$$D_z^\alpha(z^\alpha D_z^\alpha)^r\cdot D_z^\beta(z^\beta D_z^\beta)^r z^\gamma f(z)=z^{-\alpha}\left(\frac{{}_zO_{1-\alpha}^1}{\Gamma(1-\alpha)}\right)^{r+1}\cdot z^{-\beta}\left(\frac{{}_zO_{1-\beta}^1}{\Gamma(1-\beta)}\right)^{r+1}z^\gamma f(z)$$

and using the shift rule (4.18) $r+1$ times (and (4.20) if $\gamma-\beta=-M, M=1,2,\cdots$) and other properties mentioned above, we obtain

$$\{D_z^\alpha(z^\alpha D_z^\alpha)^r\}\cdot\{D_z^\beta(z^\beta D_z^\beta)^r\}z^\gamma f(z)$$

$$=z^{-\alpha-\beta}\left(\frac{1}{\Gamma(1-\alpha-\beta)}\right)^{r+1}({}_zO_{1-\beta-\alpha}^{1-\beta}\,{}_zO_{1-\beta}^1)^{r+1}z^\gamma f(z)$$

$$=\begin{cases} z^{-\alpha-\beta}\left(\dfrac{1}{\Gamma(1-\alpha-\beta)}\,{}_zO_{1-\beta-\alpha}^1\right)^{r+1}z^\gamma f(z) & \text{if } \gamma-\beta\neq-1,-2,-3,\cdots,\\ z^{-\alpha-\beta}\left(\dfrac{1}{\Gamma(1-\alpha-\beta)}\,{}_zO_{1-\alpha-\beta}^1\right)^{r+1}z^\gamma\left\{f(z)-\displaystyle\sum_{s=0}^{M-1}\frac{f^{(s)}(0)}{s!}z^s\right\} & \\ & \text{if } \gamma-\beta=-M, M=1,2,3,\cdots.\end{cases}$$

$$=\begin{cases} D_z^{\alpha+\beta}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^\gamma f(z) & \text{if } \gamma-\beta\neq-1,-2,-3,\cdots,\\ D_z^{\alpha+\beta}(z^{\alpha+\beta}D_z^{\alpha+\beta})^r z^\gamma\left\{f(z)-\displaystyle\sum_{s=0}^{M-1}\frac{f^{(s)}(0)}{s!}z^s\right\} & \text{if } \gamma-\beta=-M, M=1,2,3,\cdots.\end{cases}$$

We prove (4.11) with $\delta=1$ in a similar way using (4.22) and the shift rule (4.18).

Note that the operator ${}_zO_\beta^\alpha$ has many other properties and, with various representations of the fractional derivatives reviewed in [13], we can find the corresponding representations for the operator ${}_zO_\beta^\alpha$ with the help of definition (4.15). In particular, with the Pochhammer representation of the fractional derivative as given in [13, p. 256, Eq. (13.6)], we obtain

$$
\begin{aligned}
{}_zO^{\alpha}_{\beta}f(z) &= \frac{-\Gamma(\beta)\, e^{-i\pi\beta} z^{1-\beta}}{4\Gamma(\alpha)\Gamma(\beta-\alpha)\sin(\alpha\pi)\sin[(\beta-\alpha)\pi]} \\
&\quad \cdot \int^{(z+,0+,z-,0-)} f(\xi)\xi^{\alpha-1}(z-\xi)^{\beta-\alpha-1}\, d\xi
\end{aligned} \tag{4.23}
$$

where the only restriction is $\beta \neq 0, -1, -2, \cdots$. (We can prove that the singularities $\alpha = 0, \pm 1, \pm 2, \cdots$ and $\beta - \alpha = 0, \pm 1, \pm 2, \cdots$ can be removed). Representation (4.23) is valid for $z \in R - \{0\}$, R being an open, simply connected set in the complex plane containing the origin, where $f(z)$ is an analytic function. A publication discussing the various properties of this operator is in process of preparation (for more details concerning the operator ${}_zO^{\alpha}_{\beta}$ and its applications, see [16]).

Acknowledgments. The author wishes to thank the referees for several remarks which have improved the paper, in particular the proofs of the main results of § 2. Thanks are due to Dr. A. J. Barrett of R.M.C. for his advice in the preparation of this paper.

REFERENCES

[1] W. A. AL-SALAM, *Operational representations for the Laguerre and other polynomials*, Duke Math. J., 31 (1964), pp. 127–142.

[2] W. A. AL-SALAM AND M. E. H. ISMAIL, *Some operational formulas*, J. Math. Anal. Appl., 51 (1975), pp. 208–218.

[3] L. CARLITZ, *On arrays of numbers*, Amer. J. Math., 54 (1932), pp. 739–752.

[4] ———, *Some operational formulas*, Math. Nachr., 45 (1970), pp. 379–389.

[5] ———, *Some reduction formulas for generalized hypergeometric functions*, this Journal, 1 (1970), pp. 243–245.

[6] ———, *A theorem on differential operators*, Amer. Math. Monthly, 83 (1976), pp. 351–354.

[7] J. B. DIAZ AND T. J. OSLER, *Differences of fractional order*, Math. Comput., 28 (1974), pp. 183–202.

[8] A. ERDÉLYI, *An integral equation involving Legendre functions*, J. Soc. Indust. Appl. Math., 12 (1964), pp. 15–30.

[9] T. P. HIGGINS, *The use of fractional integral operators for solving nonhomogeneous differential equations.* Document D1-82-0677, Boeing Scientific Research Laboratories, Seattle, Washington, 1967.

[10] M. S. KLAMKIN AND D. J. NEWMAN, *On the reducibility of some linear differential operators*, Amer. Math. Monthly, 66 (1959), pp. 293–295.

[11] T. S. LARDNER, *Relations between ${}_0F_3$ and Bessel functions*, SIAM Rev. 11 (1969), pp. 69–72.

[12] J. L. LAVOIE, T. J. OSLER AND R. TREMBLAY, *Fundamental properties of fractional derivatives via Pochhammer integrals*, Lecture Notes in Mathematics on Fractional Calculus and Its Applications, No. 457, Springer-Verlag, Berlin-Heidelberg-New York, 1975, pp. 323–356.

[13] ———, *Fractional derivatives and special functions*, SIAM Rev. 18 (1976), pp. 240–268.

[14] S. OSIPOV, *On the expansion of a polynomial of the operator B_α*, U.S.S.R. Comput. Math. and Math. Phys., 3 (1963), pp. 250–256.

[15] T. J. OSLER, *A further extension of the Leibniz rule to fractional derivatives and its relation to Parseval's formula*, this Journal, 3 (1972), pp. 1–16.

[16] R. TREMBLAY, *Une contribution à la théorie de la dérivée fractionnaire*, thèse de doctorat, Université Laval, Québec, Canada, 1974.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0007 $01.00/0

COUNTABLY INFINITE NONLINEAR TIME-VARYING ACTIVE ELECTRICAL NETWORKS*

A. H. ZEMANIAN†

Abstract. This work presents existence and uniqueness theorems for the currents and voltages in a countably infinite RLC electrical network for which the total power dissipated in all the resistors or stored in all the capacitors and inductors is allowed to be infinite. This relaxation of the finite-power condition prevents the use of a number of Hilbert-space techniques and requires instead a more graph-theoretical approach. The latter has previously been used to analyze linear time-invariant networks. The main contribution of the present work is that it encompasses, under certain conditions, time-varying active networks with nonlinear resistors, inductors, and capacitors.

1. Introduction. The theory of infinite electrical networks is a comparatively recent development in the networks literature, the seminal work in this area being Flanders' 1971 paper [4] on locally finite, purely resistive, electrical networks. The initial works were devoted to linear networks, but lately attention has shifted to nonlinear ones. Some powerful results in the latter direction have been obtained by Dolezal [2]. The present work is another contribution to the theory of nonlinear infinite networks. It is distinguished from the works of Dolezal and Flanders in that it does not require that the total power within the network be finite. Finite total power allows the current flows obtained by Dolezal and Flanders to be approximated by current flows in finite subnetworks and to be analyzed by certain Hilbert-space techniques. However, the total power within an infinite electrical network need not be finite, and the current flows corresponding to infinite power are not encompassed by those Hilbert-space techniques.

An alternative approach, which we call "limb analysis", is to use graph-theoretic methods to partition the network into a sequence of finite subnetworks, which can be analyzed recursively [7]. The only requirements, other than restrictions on the element values, are that Ohm's law (i.e., the voltage-current relationships imposed by the branch parameters, be they resistive or reactive) and Kirchhoff's node and loop laws be satisfied. Consequently, current flows that generate infinite power become allowable. Moreover, the branch currents determined by this procedure in the finite subnetworks are not merely approximations to the corresponding branch currents in the infinite network but are instead precisely equal to them.

We should point out that in a countably infinite network, Kirchhoff's node law is required to hold only at the nodes of finite degree, not at those of infinite degree. An attempt to apply that law to nodes of infinite degree leads to a contradiction [9]. Thus, Kirchhoff's node law is an assertion concerning a finite number of branches, as is his loop law.

For a simple example of limb analysis, consider the infinite linear resistive network of Fig. 1. For simplicity, we allow no voltage or current sources, even though they can be easily incorporated into the analysis. (In the following, we allude to certain subnetworks called "limbs" and "orbs" and to certain branches called "joints" and "chords". They are defined in the next section, but what they are in this example should be clear.) The upper horizontal branches induce one "limb", the lower horizontal branches induce a second "limb", the vertical branch is chosen as the one and only "joint", and the diagonal branches are the "chords". We assume that a current j flows in from infinity

* Received by the editors January 31, 1978.

† State University of New York at Stony Brook, Stony Brook, New York 11794. This research was supported by the National Science Foundation under Grant MCS-75-05268-A03.

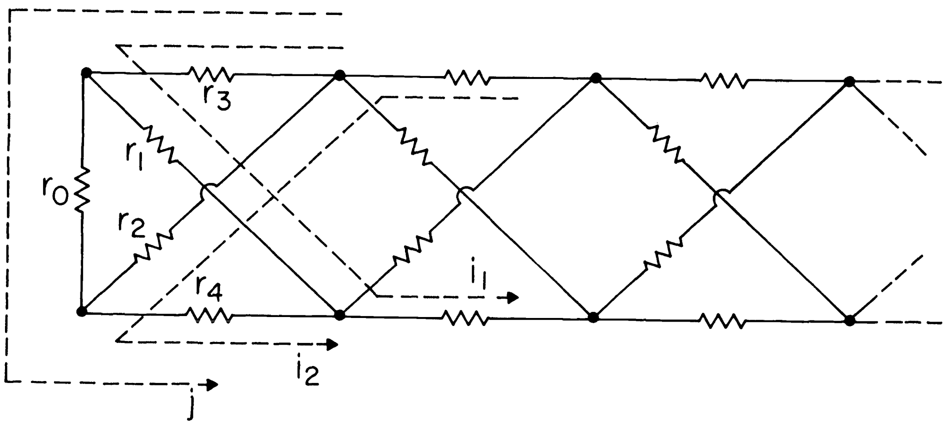

FIG. 1

along the upper limb, down through the joint, and out to infinity along the lower limb. This path is called the "joint orb". The value of j may be chosen arbitrarily. We also assume that a current i_1 (and i_2) flows along the upper limb, diagonally down through the branch with resistance r_1 (respectively, r_2) and then out toward infinity along the lower limb. These paths are called "chord orbs". The currents i_1 and i_2 are treated as unknowns. The branch currents resulting from the superposition of j, i_1, and i_2 satisfy Kirchhoff's node law at every node. We now write Kirchhoff's loop law for the r_0, r_1, r_4 loop and the r_0, r_3, r_2 loop.

$$\begin{aligned} r_1i_1-r_4i_2&=(r_0+r_4)j,\\ -r_3i_1+r_2i_2&=(r_0+r_3)j. \end{aligned} \tag{1.1}$$

Upon choosing j, we can solve these equations for i_1 and i_2 so long as the determinant $r_1r_2-r_3r_4$ is not zero. A similar procedure can then be applied to the next four branches to the right in Fig. 1 to solve for the currents in the next two diagonal branches. Continuing in this fashion, we can compute all the chord currents and therefore all the branch currents.

If however $r_1r_2=r_3r_4$, then the bridge consisting of the corresponding four branches and the vertical branch is balanced and (1.1) cannot be solved. As a matter of fact, if an infinite sequence of such bridges out along Fig. 1 are balanced, an infinite sequence of pairs of equations like (1.1) cannot be solved, whatever be the choices of the limbs and joints. This balancing implies that the network of Fig. 1 can carry no nonzero currents.

For general infinite networks, the question arises as to what conditions can be imposed on the resistance values to ensure that nonzero currents can flow. For Fig. 1, the answer is that all but a finite number of bridges are to be unbalanced. For general networks, just how the requirement of not too much balancing is to be prescribed is problematic. However, a sufficient condition is suggested by Fig. 1: If the limb branches are close enough to short circuits and the chords are close enough to open circuits, then no balancing occurs anywhere, and limb analysis will determine the possible current flows. This is the basic idea we shall exploit.

Other contributions of this work are as follows. Limb analysis is extended to the case where nonlinear time-varying active resistors, inductors, and capacitors are allowed in the network. A first result along these lines was given in [8] for nonlinear purely resistive networks. It says in effect that, if there exists a choice of limbs and joints for which the slopes of the resistance functions are small enough in the limb branches and large enough in the chords, then there exists a unique current flow for each choice of the joint currents. Sections 3 to 6 extend this result to the case where series capacitors occur in the limb branches and parallel inductors occur in the chords. Section 7 indicates how mutual coupling can be incorporated into the analysis. An existence (but not uniqueness) result for the current flows is obtained in §§ 8 to 10 by weakening the slope restrictions on the resistors, inductors, and capacitors but adding a boundedness condition on them. A stronger existence theorem is obtained in § 11 for purely resistive nonlinear networks by totally removing the slope restrictions and imposing only boundedness and continuity requirements.

2. The chainlike structure of a countably infinite network. R. Halin [5] proved that every locally finite network has a certain "chainlike" topology. This result, which we extended to countably infinite networks in [7], is fundamental to limb analysis, and so we describe it in this section.

Let N be a countably infinite network and M a subnetwork of N. A node of M is said to be N-infinite (or N-finite) if its degree as a node of N is infinite (or, respectively, finite). Thus, a node of M may be both M-finite and N-infinite.

A countably infinite network N is called *chainlike* if the following holds true: N can be partitioned into a sequence of finite subnetworks N_p:

$$N = \bigcup_{p=1}^{\infty} N_p$$

where each branch of N belongs to one and only one N_p. Moreover,

$$N_p \cap N_{p+1} = V_{p+1} \cup W_{p+1}, \qquad p = 1, 2, 3, \cdots,$$

where the following conditions are satisfied.

1. V_{p+1} consists of m_{p+1} N-finite nodes (but no branches), where $m_{p+1} < \infty$, and W_{p+1} consists of n_{p+1} N-infinite nodes (but no branches), where $n_{p+1} < \infty$.

2. The sequence $\{m_{p+1}\}_{p=1}^{\infty}$ is monotonic increasing but not necessarily strictly so, whereas the sequence $\{n_{p+1}\}_{p=1}^{\infty}$ need not be monotonic. Some or all of the m_{p+1} or n_{p+1} may equal zero. (m_{p+1} either tends to ∞ or remains constant for all p sufficiently large.)

3. For every p, V_{p+1} shares no nodes in common with $\bigcup_{s=1}^{p-1} N_s$; also, if $|p-m|>1$, then $N_p \cap N_m$ is a finite set (possibly void) of N-infinite nodes.

4. In each N_{p+1}, there are m_{p+1} node-distinct finite paths from the nodes in V_{p+1} to m_{p+1} of the m_{p+2} nodes in V_{p+2}.

This ends the definition of the adjective "chainlike". (A finite network or a network consisting of a countable infinity of finite components is encompassed by this definition when $m_{p+1} = 0$ for all p.)

LEMMA 1. *Every countably infinite network is chainlike.*

This was established in [7; § 4] for connected networks; the extension to disconnected networks is immediate.

The chainlike structure is illustrated in Fig. 2. Each finite subnetwork is contained within the dotted lines labeled by the N_p. The union of the paths indicated in item 4 above is a collection of one-sided paths, which we call *spines*; these are indicated by the horizontal solid lines of Fig. 2. (A one-ended path is an infinite connected graph all nodes of which have degree 2 except for one of them, which has degree 1.) The infinite nodes of N will also be called *spines*; we represent the latter in Fig. 2 by the horizontal dot-dash lines and think of them as one-ended paths of short circuits. Thus, a spine is

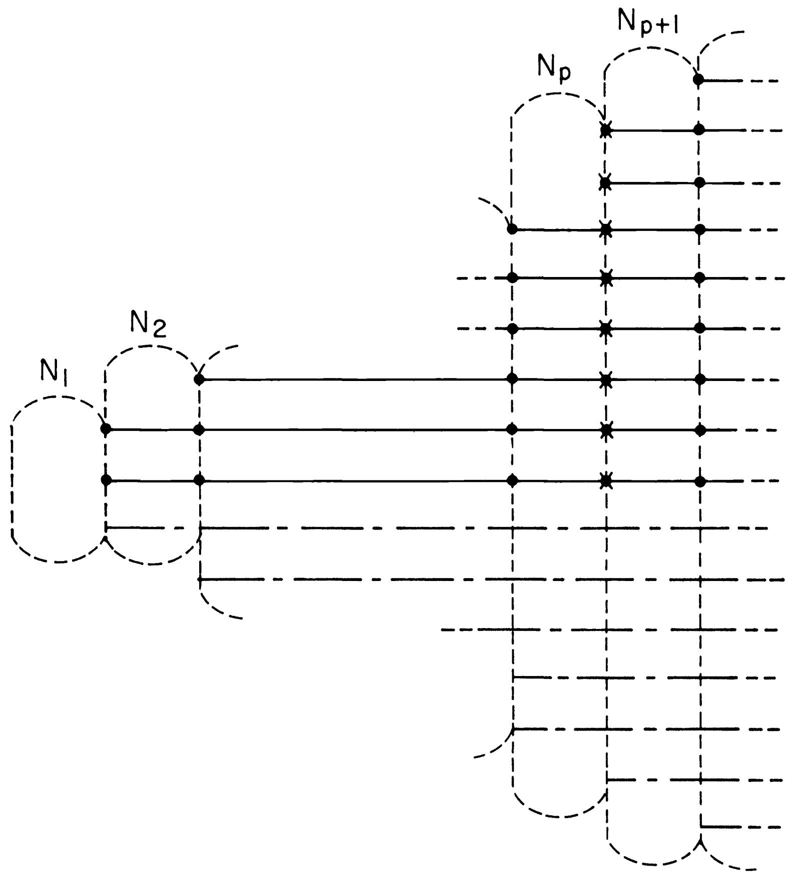

FIG. 2

either (i) a one-ended path, which starts at some node of, say, V_m and then passes through exactly one node of each set $V_{m+1}, V_{m+2}, \cdots$, and possibly other nodes as well, or (ii) an infinite node, which meets an infinity of the N_p but not necessarily all the N_p after the first N_m that it meets. The nodes of V_{p+1} are represented by the small crosses in Fig. 2. The set of spines is maximal in the sense that it contains all the N-infinite nodes and that no other one-ended path exists in N that does not meet any spines. It follows from our definition that every finite component of N will be contained in a single N_p and will not meet any spine.

As was shown in [7], branches in N can be added to the set of spine branches to obtain a spanning forest F in N each component of which either is a spanning tree in a finite component of N or alternately contains one and only one spine and does not meet any other spine. Each component of F is called a *limb*, and every branch of F is called a *limb branch.*

Next, we add branches to F to obtain a forest T that is a spanning tree in each component of N as follows. Add to $F \cap N_1$ as many branches in N_1 as possible to obtain a forest that is a spanning tree in each component of N_1. Assuming that a forest T_p, which is a spanning tree in each component of $N_1 \cup \cdots \cup N_p$, exists and contains $F \cap (N_1 \cup \cdots \cup N_p)$, we add to $T_p \cup (F \cap N_{p+1})$ as many branches in N_{p+1} as possible to obtain a forest T_{p+1} that is a spanning tree in each component of $N_1 \cup \cdots \cup N_{p+1}$. T_{p+1}

will contain $F \cap (N_1 \cup \cdots \cup N_{p+1})$. Continuing in this fashion, we obtain T. All the branches in T that are not in F will be called *joints*. Finally, all branches in N that are not in T are called *chords*. We shall refer to the set of all limbs and the set of all joints chosen in this way as a *full set of limbs* and a *full set of joints*. The partition $N = \bigcup_{p=1}^{\infty} N_p$, the spines, the limbs, the spanning forests F and T, the joints, and the chords can usually be chosen in many different ways for a given countably infinite network.

We assume henceforth that every branch has an orientation with respect to which the currents and voltages in the branch are measured. A branch voltage will mean a voltage drop in the direction of the branch's orientation.

We shall need a few more facts, which were established in [7]. Assume choices of F and T have been made. Given any chord d whose two nodes lie in different limbs, say, L_1 and L_2, there exists a unique endless path (i.e., an infinite connected subgraph whose nodes all have degree 2) that contains d and lies entirely within $L_1 \cup d \cup L_2$. Alternatively, if d's two nodes lie in the same limb, say, L_1, then there is a unique loop (i.e., a finite connected subgraph every node of which has degree 2) that is contained in $d \cup L_1$. We shall call that endless path or alternatively that loop the *d-orb* or a *chord orb*. If d is in N_p, then the d-orb is contained in $\bigcup_{s=p}^{\infty} N_s$, when the d-orb is an endless path, and in N_p, when the d-orb is a loop. In the same way, any joint generates a unique joint orb, which must be an endless path since a joint's nodes must lie in different limbs. It is also true that no more than a finite number of joint orbs or chord orbs can pass through any branch.

Furthermore, for any given chord d in N_p, there exists a unique loop that is contained in $T \cup d$. We call this loop the *d-tree loop* and *chord-tree loop*. When d's two nodes lie in different limbs, the d-tree loop is contained in $\bigcup_{s=1}^{p} N_s$; when d's two nodes lie in the same limb, the d-tree loop is contained in N_p and is in fact identical to the d-orb. We assign orientations to the d-orb and the d-tree loop that coincide on d with d's orientation.

Assume that the currents are zero in all the joints and chords except for one of them, say, the chord or joint d, which carries a current of $i \neq 0$. It was shown in [7; § 5] that, if Kirchhoff's node law is satisfied, all the limb branches have zero currents as well except for those in the d-orb. The latter limb branches carry a current of i in the direction of the d-orb's orientation. If b is a limb branch in the d-orb, we say that d *induces* $\pm i$ in b, the plus (minus) sign being chosen if the orientations of b and the d-orb agree (or, respectively, disagree). Similarly, we say that d *induces* in itself the current i. The following result comes from [7].

LEMMA 2. *Let N be a countably infinite network and choose F and T as above. An arbitrary assignment of joint and chord currents uniquely determines under Kirchhoff's node law the current in any limb branch b as the finite sum of all the currents induced in b by the joints and chords. In other words, if a current flow in N satisfies Kirchhoff's node law, then the current in any branch is equal to the finite sum of currents induced in that branch by the chords and joints.*

3. Assumptions on the network's elements. R^n denotes n-dimensional real Euclidean space, and $t \in R^1$ is the time variable. T denotes a positive number in R^1. In this work all functions and variables are real-valued. $L_2(\alpha, \beta)$ is the customary Hilbert space of (equivalence classes of) quadratically integrable functions on the interval (α, β) with respect to Lebesgue measure. Here, we allow $-\infty \leqq \alpha < \beta \leqq \infty$. The norm for $L_2(\alpha, \beta)$ is

$$\|v\| = \left[\int_{\alpha}^{\beta} |v(t)|^2 \, dt \right]^{1/2}.$$

Every branch of a network is assumed to have an orientation and every voltage $v(t)$ and current $i(t)$ in a given branch are measured with respect to the orientation of that branch. The work "voltage" is understood to mean voltage drop.

The assumptions we impose throughout most of this paper upon our network elements and upon the structure of the network are as follows.

I. *Sources.* Every voltage source v^s or current source i^s is a mapping of R^1 into R^1. We always assume that $v^s, i^s \in L_2(0, T)$. The range values $v^s(t)$ and $i^s(t)$ are the instantaneous voltage and current of those sources. These quantities are assumed to be known.

II. *Resistors.* We assume that the voltage $v(t)$ and current $i(t)$ in each resistor are related by $v(t) = r(i(t), t)$, where the resistance function $r(\cdot, \cdot)$ is a mapping of $R^1 \times (0, T)$ into R^1. It is not required that $r(\cdot, \cdot)$ be a nonnegative function. We assume furthermore that for any $a, b \in R^1$

$$|r(a, t)| \leqq Q_r |a| \tag{3.1}$$

and

$$|r(a, t) - r(b, t)| \leqq K_r |a - b|, \tag{3.2}$$

where Q_r and K_r are constants not depending upon a, b, or t.

III. *Conductors.* The voltage and current in every conductor are assumed to be related by $i(t) = g(v(t), t)$, where $g(\cdot, \cdot)$ maps $R^1 \times (0, T)$ into R^1 and need not be nonnegative. In addition, we assume that for any $a, b \in R^1$

$$|g(a, t)| \leqq Q_g |a| \tag{3.3}$$

and

$$|g(a, t) - g(b, t)| \leqq K_g |a - b|. \tag{3.4}$$

The constants Q_g and K_g do not depend upon a, b, or t.

IV. *Capacitors.* The voltage $v(t)$ and charge $q(t)$ on each capacitor are assumed to be related by

$$v(t) = \gamma(q(t), t), \tag{3.5}$$

where $\gamma(\cdot, \cdot)$ maps $R^1 \times (0, T)$ into R^1; $\gamma(\cdot, \cdot)$ need not be a nonnegative function. Furthermore, it is assumed that each capacitor has zero initial charge at $t = 0$ (i.e., $q(0) = 0$) and that the following inequalities hold for any $a, b \in R^1$ and any $t, \tau \in (0, T)$.

$$|\gamma(a, t)| \leqq A_\gamma |a|, \tag{3.6}$$

$$|\gamma(a, t) - \gamma(b, t)| \leqq P_\gamma |a - b|, \tag{3.7}$$

$$|\gamma(a, t) - \gamma(a, \tau)| \leqq L_\gamma |a| \, |t - \tau|. \tag{3.8}$$

A_γ, L_γ, and P_γ are constants not depending upon the time or charge variables.

V. *Inductors.* The current $i(t)$ and the flux linkages $\phi(t)$ in each inductor are assumed to be related by

$$i(t) = \lambda(\phi(t), t) \tag{3.9}$$

where $\lambda(\cdot, \cdot)$ maps $R^1 \times (0, T)$ into R^1; $\lambda(\cdot, \cdot)$ is not required to be nonnegative. We also assume that $\phi(0) = 0$ and that the following hold for any $a, b \in R^1$ and any

$t, \tau \in (0, T)$.

(3.10) $$|\lambda(a, t)| \leqq A_\lambda |a|,$$

(3.11) $$|\lambda(a, t) - \lambda(b, t)| \leqq P_\lambda |a - b|,$$

(3.12) $$|\lambda(a, t) - \lambda(a, \tau)| \leqq L_\lambda |a|\, |t - \tau|.$$

Here again, A_λ, P_λ, and L_λ are constants that don't depend upon the time or flux-linkage variables.

VI. *Structure of the network.* We assume that a full set $\mathscr{L}$ of limbs and a full set $\mathscr{J}$ of joints can be so chosen that every limb branch is either a voltage source, a resistor, a capacitor, or a series connection of any two or all three of these elements and that every chord is either a current source, a conductor, an inductor, or a parallel connection of any two or all three of these elements. We finally assume that there is no mutual coupling.

4. Some consequences of the assumptions. Conditions II–V allow every resistor, conductor, capacitor, and inductor to be nonlinear, time-varying, and active.

With regard to the resistors, (3.1) and (3.2) imply respectively that, for $i, i' \in L_2(0, T)$, $v(t) = r(i(t), t)$, and $v'(t) = r(i'(t), t)$, we have

(4.1) $$\|v\| \leqq Q_r \|i\|$$

and

(4.2) $$\|v - v'\| \leqq K_r \|i - i'\|,$$

where $\|\cdot\|$ is the norm for $L_2(0, T)$. Thus, $i \mapsto r(i(\cdot), \cdot)$ is a continuous mapping of $L_2(0, T)$ into $L_2(0, T)$. Also, (3.1) implies that $r(0, t) = 0$. Conversely, when $r(0, t) = 0$, (3.1) with Q_r replaced by K_r follows from (3.2). In general, however, we allow $Q_r < K_r$.

Similar assertions hold for conductors. In particular, for $v, v' \in L_2(0, T)$, $i(t) = g(v(t), t)$, and $i'(t) = g(v'(t), t)$, we have

(4.3) $$\|i\| \leqq Q_g \|v\|$$

and

(4.4) $$\|i - i'\| \leqq K_g \|v - v'\|.$$

Now, consider the capacitors. Since $q(0) = 0$ by assumption, the charge $q(t)$ is related to the current $i(t)$ by $q(t) = \int_0^t i(\omega)\, d\omega$. In view of (3.5), (3.6), and Schwarz's inequality, we may write

$$\|v\|^2 = \int_0^T |\gamma(q(t), t)|^2\, dt \leqq A_\gamma^2 \int_0^T \left| \int_0^t i(\omega)\, d\omega \right|^2 dt$$
$$\leqq A_\gamma^2 \int_0^T t \int_0^t |i(\omega)|^2\, d\omega\, dt \leqq \frac{1}{2}(A_\gamma \|i\| T)^2.$$

Thus,

(4.5) $$\|v\| \leqq Q_\gamma \|i\|$$

where $Q_\gamma = A_\gamma T/\sqrt{2}$. This implies that

$$i \mapsto v = \gamma\left(\int_0 i(\omega)\, d\omega, \cdot \right)$$

is a mapping of $L_2(0, T)$ into $L_2(0, T)$.

Virtually the same manipulation based upon (3.7) instead of (3.6) yields

$$\|v-v'\| \leqq K_\gamma \|i-i'\| \tag{4.6}$$

where $K_\gamma = P_\gamma T/\sqrt{2}$ and $v' = \gamma(\int_0^{\cdot} i'(\omega)\, d\omega, \cdot)$.

For any inductor the flux linkages $\phi(t)$ are related to the voltage $v(t)$ by $\phi(t) = \int_0^t v(\omega)\, d\omega$. We may use (3.7) through (3.9) with the same arguments as those for a capacitor to obtain the following inequalities.

$$\|i\| \leqq Q_\lambda \|v\|, \tag{4.7}$$

$$\|i-i'\| \leqq K_\lambda \|v-v'\|. \tag{4.8}$$

Here, $Q_\lambda = A_\lambda T/\sqrt{2}$ and $K_\lambda = P_\lambda T/\sqrt{2}$.

Now, consider a branch consisting of a series connection of a resistor and a capacitor, where either one but not both of these elements may be zero. Then, the branch voltage $v(t)$ is related to the branch current $i(t)$ by

$$v(t) = z(i, t) = r(i(t), t) + \gamma\left(\int_0^t i(\omega)\, d\omega, t\right). \tag{4.9}$$

By (4.1) and (4.5),

$$\|v\| \leqq (Q_r + Q_\gamma)\|i\|. \tag{4.10}$$

Similarly, if also $v'(t) = z(i', t)$, then by (4.2) and (4.6)

$$\|v-v'\| \leqq (K_r + K_\gamma)\|i-i'\|. \tag{4.11}$$

Thus, $i \mapsto v$ is a Lipschitz mapping of $L_2(0, T)$ into $L_2(0, T)$.

For a parallel connection of a conductor and an inductor, where either but not both of these elements may be zero, we have

$$i(t) = y(v, t) = g(v(t), t) + \lambda\left(\int_0^t v(\omega)\, d\omega, t\right). \tag{4.12}$$

The inequalities (4.3), (4.7), (4.4), (4.8) with $i'(t) = y(v', t)$ lead to

$$\|i\| \leqq (Q_g + Q_\lambda)\|v\|, \tag{4.13}$$

$$\|i-i'\| \leqq (K_g + K_\lambda)\|v-v'\| \tag{4.14}$$

so that here too $v \mapsto i$ is a Lipschitz mapping of $L_2(0, T)$ into $L_2(0, T)$.

5. The determining equations. Assume that a partition $N = \bigcup_{p=1}^\infty N_p$, a full set of limbs, and a full set of joints have been chosen in accordance with § 2 and condition V. Let the indices of the limb branches of N_p be $\nu = 1, \cdots, n$ and the indices of the chords of N_p be $\mu = n+1, \cdots, n+m$. Also, let k denote the indices of the joints in N_p. We can use Lemma 2, which is a consequence of Kirchhoff's node law, to write an equation for the current i_ν in each limb branch b_ν of the subnetwork N_p.

$$i_\nu(t) - \sum_\mu^{Y,\nu} (\pm) y_\mu(v_\mu, t) = h_\nu(t) + \sum_k^{J,\nu} (\pm) j_k(t) + \sum_\mu^{I,\nu} (\pm) i_\mu^s(t). \tag{5.1}$$

Here, the summation $\sum^{Y,\nu}$ is over all chords b_μ in N_p that induce nonzero currents in b_ν and have nonzero admittances y_μ as indicated in (4.12). Thus, chords that are purely current sources are not included. As was mentioned in § 2, the plus (minus) sign is used with y_μ if the orientations of b_ν and the b_μ-orb agree (disagree). $v_\mu(t)$ is the voltage in b_μ. $h_\nu(t)$ is the sum of all the currents induced in b_ν by all the joints and chords in

$\bigcup_{s=1}^{p-1} N_s$. $h_\nu(t)=0$ if b_ν is not a spine branch. $j_k(t)$ is the current in the kth joint of N_p. This is a free parameter except that it must be chosen in conformity with the kth joint's parameters. The summation $\sum^{J,\nu}$ is over all joints in N_p that induce nonzero currents in b_ν, and the plus or minus sign is chosen if the orientations of b_ν and the kth joint's orb agree or disagree respectively. $i_\mu^s(t)$ is the value of the current source in the μth chord of N_p. $\sum^{I,\nu}$ is a summation over all chords in N_p that induce nonzero currents in b_ν and have nonzero current sources. The plus or minus sign is chosen with i_μ^s in the same way as it was for y_μ.

We can also write an equation for the voltage v_μ in the chord b_μ of N_p by applying Kirchhoff's loop law to the b_μ-tree loop:

$$v_\mu(t)-\sum_{\nu}^{Z,\mu}(\pm)z_\nu(i_\nu,t)=e_\mu(t)+\sum_{k}^{J,\mu}(\pm)w_k(t)+\sum_{\nu}^{V,\mu}(\pm)v_\nu^s(t). \tag{5.2}$$

$\sum^{Z,\mu}$ is a summation over all limb branches b_ν in N_p that lie in the b_μ-tree loop and have nonzero impedances z_ν as indicated in (4.11). The plus (minus) sign is used with z_ν if the orientations of b_ν and the b_μ-tree loop disagree (agree). $i_\nu(t)$ is the current in b_ν. $e_\mu(t)$ is the sum of all the voltages of the branches in that portion of the b_μ-tree loop that lies in $\bigcup_{s=1}^{p-1} N_s$, those voltages being measured with respect to a tracing of the b_μ-tree loop that disagrees with b_μ's orientation. $w_k(t)$ is the voltage in the kth joint of N_p; it may be chosen arbitrarily so long as it conforms with the kth joint's parameters. The summation $\sum^{J,\mu}$ is over all joints in N_p lying in the b_μ-tree loop; the plus (minus) sign is used with w_k if the orientations of the kth joint and of the b_μ-tree loop disagree (agree). $v_\nu^s(t)$ is the value of the voltage source in the νth limb branch of N_p. $\sum^{V,\mu}$ is a summation over all limb branches b_ν in N_p that lie in the b_μ-tree loop and have nonzero voltage sources. The plus or minus sign is chosen with v_ν^s in the same way as it was for z_ν.

We shall now write the simultaneous equations given by (5.1) and (5.2) in matrix form. First, let $x(t)$ be the unknown $(n+m)\times 1$ vector of limb-branch currents and chord voltages in N_p:

$$x(t)=[i_1(t),\cdots,i_n(t),v_{n+1}(t),\cdots,v_{n+m}(t)]^{\mathrm{T}}.$$

The superscript T denotes the matrix transpose. Let

$$d(t)=[d_1(t),\cdots,d_{n+m}(t)]^{\mathrm{T}}$$

where, for $\nu=1,\cdots,n$, $d_\nu(t)$ is the right-hand side of (5.1) and, for $\mu=n+1,\cdots,n+m$, $d_\mu(t)$ is the right-hand side of (5.2). If all the joint currents and voltages have been chosen as members of $L_2[0,T]$, if all the voltages and currents in $\bigcup_{s=1}^{p-1} N_s$ have already been determined by prior computations as members of $L_2[0,T]$, and if all voltage and current sources are given as members of $L_2[0,T]$, then the right-hand sides in (5.1) and (5.2) are determined as members of $L_2[0,T]$; that is, $d(t)$ is a known vector. Finally, set

$$w(\cdot,t)=\begin{bmatrix} 0 & a(\cdot,t) \\ b(\cdot,t) & 0 \end{bmatrix} \tag{5.3}$$

where $a(\cdot,t)$ is an $n\times m$ matrix and $b(\cdot,t)$ is an $m\times n$ matrix whose elements involve y_μ and z_ν respectively. More specifically, the νth row of $w(\cdot,t)$ has the term $\pm y_\mu(\cdot,t)$ (or zero) in the μth column if that term appears (or, respectively, does not appear) in the summation $\sum_\mu^{Y,\nu}$. Similarly, the μth row of $w(\cdot,t)$ has the term $\pm z_\nu(\cdot,t)$ (or zero) in the νth column if that term appears (or, respectively, does not appear) in the summation

$\sum_{\nu}^{Z,\mu}$. With these definitions, the matrix notation for (5.1) and (5.2) is

$$[I-w(\cdot\,,t)]x(t)=d(t), \tag{5.4}$$

where I is the $(n+m)\times(n+m)$ identity matrix.

6. Existence and uniqueness. We now let H be the direct sum of $n+m$ replications of $L_2[0, T]$. That is,

$$H=\{\alpha=[\alpha_1,\cdots,\alpha_{n+m}]^{\mathrm{T}}\colon \alpha_i\in L_2[0,T]\}$$

and the norm of H is given by

$$\|\alpha\|^2=\sum_{i=1}^{n+m}\|\alpha_i\|^2.$$

Thus, the vectors x and d of the preceding section are members of H. Let

$$x'(t)=[i_1'(t),\cdots,i_n'(t),v_{n+1}'(t),\cdots,v_{n+m}'(t)]^{\mathrm{T}}$$

be another member of H. Then, with the understanding that t is the independent variable of the functions whose norms are being taken, we may write

$$\begin{aligned}\|w(x,t)-w(x',t)\|^2&=\sum_{\nu=1}^{n}\left\|\sum_{\mu}^{Y,\nu}(\pm)[y_\mu(v_\mu,t)-y_\mu(v_\mu',t)]\right\|^2\\&\quad+\sum_{\mu=n+1}^{n+m}\left\|\sum_{\nu}^{Z,\mu}(\pm)[z_\nu(i_\nu,t)-z_\nu(i',t)]\right\|^2\\&\leqq\sum_{\nu=1}^{n}\left[\sum_{\mu}^{Y,\nu}\|y_\mu(v_\mu,t)-y_\mu(v_\mu',t)\|\right]^2\\&\quad+\sum_{\mu=n+1}^{n+m}\left[\sum_{\nu}^{Z,\mu}\|z_\nu(i_\nu,t)-z_\nu(i_\nu',t)\|\right]^2.\end{aligned}$$

Since $(\sum_{i=1}^{k}\beta_i)^2\leqq k\sum_{i=1}^{k}\beta_i^2$, the right-hand side is bounded by

$$\sum_{\mu=n+1}^{n+m}M_\mu\|y_\mu(v_\mu,t)-y_\mu(v_\mu',t)\|^2+\sum_{\nu=1}^{n}M_\nu\|z_\nu(i_\nu,t)-z_\nu(i_\nu',t)\|^2$$

where M_μ (or M_ν) is the number of terms in all the rows of $w(\cdot\,,t)$ in which y_μ (or, respectively, z_ν) appears. In view of (4.14) and (4.11), the last expression is bounded by

$$\sum_{\mu=n+1}^{n+m}M_\mu(K_{g_\mu}+K_{\lambda_\mu})^2\|v_\mu-v_\mu'\|^2+\sum_{\nu=1}^{n}M_\nu(K_{r_\nu}+K_{\gamma_\nu})^2\|i_\nu-i_\nu'\|^2.$$

Here, $K_{g_\mu}+K_{\lambda_\mu}$ and $K_{r_\nu}+K_{\gamma_\nu}$ are the constants corresponding to y_μ and z_μ in accordance with (4.14) and (4.11) respectively. If in addition to our prior assumptions we also assume that

$$(K_{g_\mu}+K_{\lambda_\mu})^2<M_\mu^{-1},\qquad \mu=n+1,\cdots,n+m, \tag{6.1}$$

and

$$(K_{r_\nu}+K_{\gamma_\nu})^2<M_\nu^{-1},\qquad \nu=1,\cdots,n, \tag{6.2}$$

then we can conclude that

$$\|w(x,t)-w(x',t)\|\leqq\delta\|x-x'\| \tag{6.3}$$

where $\delta<1$.

Set $(Wx)(t) = w(x, t)$. So, (5.4) becomes

$$(I - W)x = d. \tag{6.4}$$

The inequality (6.3) asserts that W is a contractive mapping of H into H. Consequently, $I - W$ is a bijection of H into H. (Indeed, for any fixed $d \in H$, set $F_d \cdot = d + W \cdot$. Then, F_d is also contractive and therefore has a unique fixed point $x_0 \in H$, which implies our assertion.)

In view of (6.4) we can conclude that once the voltage-current pairs for the joints in N have been chosen and the currents and voltages in the limb branches and chords of $\bigcup_{s=1}^{p-1} N_s$ have been determined, then the currents and voltages in the limb branches and chords of N_p are also uniquely determined by (5.4). Upon applying this argument recursively to $N_1, N_2, \cdots$, we can conclude with the following theorem. [Note that the argument in this section employs all the assumptions of § 3 except for (3.8) and (3.12).]

THEOREM 1. *Let N be a countably infinite network satisfying the conditions* I–VI *of* § 3 *except for* (3.8) *and* (3.12). (Those two inequalities may or may not be satisfied.) *Assume in addition that, for the chosen partition $N = \bigcup_{p=1}^{\infty} N_p$ in the chainlike structure and the choices of $\mathscr{L}$ and $\mathscr{J}$ indicated in condition* VI, *the constants $K_{g_\mu} + K_{\lambda_\mu}$ corresponding to the chord admittances in N_p satisfy* (6.1) *and the constants $K_{r_\nu} + K_{\gamma_\nu}$ corresponding to the limb-branch impedances in N_p satisfy* (6.2) *for every p.* (By definition, $K_{\lambda_\mu} = P_{\lambda_\mu} T/\sqrt{2}$ and $K_{\gamma_\nu} = P_{\gamma_\nu} T/\sqrt{2}$.) *Then, any assignment of voltage-current pairs in $L_2[0, T] \times L_2[0, T]$ to the joints (in conformity with the joint's parameters) uniquely determines under Ohm's law and Kirchhoff's node and loop laws the voltage-current pairs from $L_2[0, T] \times L_2[0, T]$ for all the branches of N. These voltage-current pairs can be computed by solving recursively equations of the form* (6.4) *for the finite subnetworks $N_1, N_2, \cdots$.*

Let us make a few more remarks before leaving this section: It is the restriction (6.1) coupled with (4.13) and (4.14) that requires in effect that the chords be sufficiently close to open circuits, and it is the restriction (6.2) coupled with (4.10) and (4.11) that requires the limb branches to be sufficiently close to short circuits.

Another way to determine M_μ is as follows: Let P_μ be the orb for the μth chord in N_p. For any limb branch b lying in $N_p \cap P_\mu$, count the number of chords in N_p whose orbs pass through b and whose admittances are not zero. Add together all such numbers obtained by letting b traverse all the limb branches that lie in $N_p \cap P_\mu$. This gives M_μ.

To get M_ν, consider the νth limb branch b_ν in N_p, and let d be any chord in N_p whose d-tree loop contains b_ν. Count the number of limb branches in N_p that lie in the d-tree loop and have nonzero impedances. Add together all such numbers obtained by letting d traverse all the chords in N_p whose chord-tree loops contain b_ν. This gives M_ν.

7. Mutual coupling. Under certain circumstances we can allow dependent current sources $u(\sigma, t)$ as parallel-connected elements in the chords of N_p and dependent voltage sources $f(\sigma, t)$ as series-connected elements in the limb branches of N_p so long as u and f are in $L_2[0, T]$ whenever σ is in $L_2[0, T]$. When $\sigma(t)$ is a voltage or a current in $\bigcup_{s=1}^{p-1} N_s$, u and f appear as known quantities in the right-hand sides of (5.1) and (5.2), and the analysis proceeds as before. On the other hand, we cannot allow $\sigma(t)$ to be a voltage or current in $\bigcup_{s=p+1}^{\infty} N_s$, for then our recursive method would fail.

When mutual coupling occurs within a particular N_p, we require that the corresponding dependent sources have the form $u(v, t)$ and $f(i, t)$, where $v(t)$ is a chord voltage in N_p and $i(t)$ is a limb-branch current in N_p. Such sources appear as unknown terms in the left-hand sides of (5.1) and (5.2), but W maintains the form indicated in (5.3). If these terms correspond to transconductors, transresistors, mutual capacitors, or mutual inductors and if sufficiently strong conditions like those of § 3 and (6.1) and (6.2) as well are imposed upon them, then (6.3) can be satisfied again. In this way Theorem 1 can be extended to encompass such mutual coupling.

8. Some results from functional analysis. Our next object is to get another existence theorem by weakening part of the hypothesis of Theorem 1 but adding other assumptions. We will lose uniqueness in the process. We start with some results from functional analysis.

H continues to be the Hilbert space defined in § 6. Every $f=[f_1, \cdots, f_{n+m}]^{\mathrm{T}} \in H$ can be extended as the zero vector outside of $[0, T]$. This we do. Let supp h denote the support of any function h, and let $f(\cdot+s)$ be the translation of f through the distance $s \in R^1$. Then, $\operatorname{supp}[f(\cdot+s)-f(\cdot)] \subset[-|s|, T+|s|]$. Moreover, $f(\cdot+s)-f(\cdot)$ is a member of the direct sum of $n+m$ replicates of $L_2(-\infty, \infty)$, whose norm in the latter space is also denoted by $\|f(\cdot+s)-f(\cdot)\|$.

LEMMA 3. *A subset K of H is conditionally compact if and only if the following hold.*

(i) *K is bounded under the norm of H.*

(ii) *As $s \rightarrow 0$,*

$$\|f(\cdot+s)-f(\cdot)\|^2=\sum_{i=1}^{n+m} \int_{-\infty}^{\infty}|f_i(t+s)-f_i(t)|^2\, dt \rightarrow 0$$

uniformly for all $f \in K$.

Lemma 3 is the extension of the M. Riesz–Tamarkin theorem to the direct sum H of the $n+m$ replicates of $L_2[0, T]$. See [1, p. 32] or [3, pp. 298–299].

LEMMA 4. *Let D be a closed sphere in H and let B and C be continuous functions from D into H satisfying the following*:

(a) *$Bx+Cx \in D$ for all $x \in D$.*

(b) *$C(D)$ is conditionally compact.*

(c) *There exists a real number $\eta<1$ such that $\|Bx-Bx'\| \leqq \eta\|x-x'\|$ for all $x, x' \in D$.*

Then, there exists at least one $x_0 \in D$ such that $Bx_0+Cx_0=x_0$.

A reference for Lemma 4 is [6, p. 126].

LEMMA 5. *Let R and S be continuous functions from H into H, let $W=R+S$, let d be any fixed member of H, and let the following be satisfied.*

(i) *There is a real number $\delta<1$ such that $\|Wx\| \leqq \delta\|x\|$ for all $x \in H$.*

(ii) *For any closed sphere X in H, $S(X)$ is conditionally compact.*

(iii) *There is a real number $\eta<1$ such that $\|Rx-Rx'\| \leqq \eta\|x-x'\|$ for all $x, x' \in H$.*

Then, there exists at least one $x_0 \in H$ such that $x_0-Wx_0=d$.

Proof. Let B be the function $d+R\,\cdot: x \mapsto d+Rx$, let $C=S$, and let

$$D=\left\{x \in H:\|x\| \leqq \frac{\|d\|}{1-\delta}\right\}.$$

B and C are continuous functions on D. Set $z=d/(1-\delta)$. Then, for all $x \in D$,

$$\begin{aligned} \|Bx+Cx\| &=\|d+Wx\| \leqq\|d\|+\|Wx\| \leqq\|d\|+\delta\|x\| \\ & \leqq(1-\delta)\|z\|+\delta\|z\|=\|z\| . \end{aligned}$$

Thus, condition (a) of Lemma 4 is satisfied.

Since $S=C$, condition (b) of Lemma 4 is also satisfied. Finally, for any $x, x' \in D$,

$$\|Bx-Bx'\|=\|Rx-Rx'\| \leqq \eta\|x-x'\| .$$

So, (c) is satisfied too. Our conclusion now follows from Lemma 4.

9. Another estimate. In addition to the conditions of § 3, let us now assume that for all capacitors and inductors the defining equations (3.5) and (3.9) and the inequalities (3.6) to (3.8) and (3.10) to (3.12) hold for all t and τ in a neighborhood Ω of the compact interval $[0, T]$. Also, let $\varepsilon > 0$ be so small that, whenever $t \in [0, T]$ and $|s| \leqq \varepsilon$, we have $s \in \Omega$ and $t+s \in \Omega$. For these restrictions on t and s and for any inductor, we obtain from (3.11) and (3.12)

$$\begin{aligned}|\lambda(b, t+s)-\lambda(a, t)| &\leqq |\lambda(b, t+s)-\lambda(a, t+s)|+|\lambda(a, t+s)-\lambda(a, t)|\\ &\leqq P_\lambda|b-a|+L_\lambda|as|.\end{aligned}$$

Since $|\alpha+\beta|^2 \leqq 2\alpha^2+2\beta^2$ for $\alpha, \beta \in R^1$,

$$|\lambda(b, t+s)-\lambda(a, t)|^2 \leqq 2(P_\lambda|b-a|)^2+2(L_\lambda|as|)^2. \tag{9.1}$$

Next, let $\rho(t)=1$ for $0<t<T$ and $\rho(t)=0$ otherwise. Any function defined on a subset of R^1 is understood to be extended outside its domain as the zero function. With this convention, every function in $L_2[0, T]$ is zero outside $[0, T]$. Also, since $\lambda(0, t)=0$, it follows that, for any $v \in L_2[0, T]$,

$$\operatorname{supp} \lambda\Big(\rho(\cdot+s) \int_0^{\cdot+s} v(\omega)\, d\omega, \cdot+s\Big) \subset [-s, T-s].$$

In the following manipulations, we let $\int = \int_{-\infty}^{\infty}$ and we understand that it is the independent variable of the functions in $L_2(-\infty, \infty)$ whose norms are being taken. Then,

$$\begin{aligned}&\Big\|\lambda\Big(\rho(t+s) \int_0^{t+s} v(\omega)\, d\omega, t+s\Big)-\lambda\Big(\rho(t) \int_0^t v(\omega)\, d\omega, t\Big)\Big\|^2\\ &\quad=\int \Big|\lambda\Big(\rho(t+s) \int_0^{t+s} v(\omega)\, d\omega, t+s\Big)-\lambda\Big(\rho(t) \int_0^t v(\omega)\, d\omega, t\Big)\Big|^2 dt.\end{aligned} \tag{9.2}$$

By (9.1) this is bounded by

$$\begin{aligned}&2P_\lambda^2 \int \Big|\rho(t+s) \int_0^{t+s} v(\omega)\, d\omega-\rho(t) \int_0^t v(\omega)\, d\omega\Big|^2 dt\\ &\quad+2L_\lambda^2|s|^2 \int \Big|\rho(t) \int_0^t v(\omega)\, d\omega\Big|^2 dt.\end{aligned} \tag{9.3}$$

By Schwarz's inequality, the second term is bounded by

$$2L_\lambda^2|s|^2 \int_0^T \Big|\int_0^t v(\omega)\, d\omega\Big|^2 dt \leqq 2L_\lambda^2|s|^2 \int_0^T t \int_0^t |v(\omega)|^2\, d\omega\, dt.$$

But, supp $v \subset [0, T]$, and so the right-hand side is dominated by $(L_\lambda|s| T\|v\|)^2$.

Furthermore, with $|s| \leqq \varepsilon$ as above and $\varepsilon < T$, the first term in (9.3) is for $s>0$ equal to

$$2P_\lambda^2 \int_{-s}^T \Big|\int_t^{t+s} v(\omega)\, d\omega\Big|^2 dt$$

and for $s<0$ equal to

$$2P_\lambda^2 \int_0^{T-s} \Big|\int_{t+s}^t v(\omega)\, d\omega\Big|^2 dt.$$

By Schwarz's inequality the squared magnitude inside both of these integrals is no larger than $|s|\,\|v\|^2$, and so the first term in (9.3) is bounded by $2P_\lambda^2|s|(T+|s|)\|v\|^2$.

Altogether then, (9.2) is bounded by

$$2P_\lambda^2|s|(T+|s|)\|v\|^2+(L_\lambda|s|T\|v\|)^2, \tag{9.4}$$

which tends to zero as $s\to 0$. A similar result holds for our capacitors.

10. Another existence theorem. Instead of conditions (6.1) and (6.2), we will now require that, for $\nu=1,\cdots,n$ and $\mu=n+1,\cdots,n+m$,

$$M_\mu(Q_{g_\mu}+Q_{\lambda_\mu})^2<1, \tag{10.1}$$

$$M_\nu(Q_{r_\nu}+Q_{\gamma_\nu})^2<1, \tag{10.2}$$

$$M_{g_\mu}K_{g_\mu}^2<1, \tag{10.3}$$

$$M_{r_\nu}K_{r_\nu}^2<1. \tag{10.4}$$

Here, M_{g_μ} (or M_{r_ν}) is the value that M_μ (respectively, M_ν) would take if all chord inductors (limb-branch capacitors) were replaced by open (respectively, short) circuits. That is, if R is the matrix operator obtained by setting all inductor and capacitor terms in W equal to zero, then M_{g_μ} and M_{r_ν} are determined from R in the same way as M_μ and M_ν are determined from W. Alternatively, we may follow the procedures given in the last two paragraphs of § 6 so long as open circuits and short circuits are not counted as branches. Thus, $M_{g_\mu}\leqq M_\mu$ and $M_{r_\nu}\leqq M_\nu$.

It can be seen from (3.1) through (3.4) that the conditions imposed on the network elements by (10.1) through (10.4) are no stronger and in general weaker than those imposed by (6.1) and (6.2).

THEOREM 2. *Let N be a countably infinite electrical network whose elements satisfy conditions* I–VI *of* § 3 *with the added proviso that the equations and inequalities of conditions* IV *and* V *hold for all t and τ in a neighborhood Ω of the compact interval* $[0, T]$. *Assume that, for the chosen partition $N=\bigcup_{p=1}^\infty N_p$ for the chainlike structure and the choices of $\mathcal{L}$ and $\mathcal{J}$ indicated in condition* VI, *the constants Q_{g_μ}, K_{g_μ}, Q_{r_ν}, K_{r_ν}, $Q_{\lambda_\mu}=A_{\lambda_\mu}T/\sqrt{2}$, $Q_{\gamma_\nu}=A_{\gamma_\nu}T/\sqrt{2}$ corresponding to the electrical elements in N_p satisfy* (10.1)–(10.4). *Then, any assignment of voltage-current pairs from $L_2[0,T]\times L_2[0,T]$ to all the joints* (*in conformity with the joints' parameters*) *and any specification of the initial currents and voltages in the inductors and capacitors respectively determine at least one collection of voltage-current pairs from $L_2[0,T]\times L_2[0,T]$ for all the branches of N that satisfies Ohm's law and Kirchhoff's node and loop laws.*

Proof. We will show that the hypothesis of Lemma 5 is satisfied by the operator W defined in § 6. As before, it is understood that t is the independent variable of the functions whose norms are being taken. For any

$$x=[i_1,\cdots,i_n,v_{n+1},\cdots,v_{n+m}]^{\mathrm{T}}\in H,$$

we have

$$\begin{aligned}\|Wx\|^2&=\sum_{\nu=1}^{n}\left\|\sum_{\mu}^{Y,\nu}(\pm)y_\mu(v_\mu,t)\right\|^2+\sum_{\mu=n+1}^{n+m}\left\|\sum_{\nu}^{Z,\mu}(\pm)z_\nu(i_\nu,t)\right\|^2\\ &\leqq\sum_{\nu=1}^{n}\left[\sum_{\mu}^{Y,\nu}\|y_\mu(v_\mu,t)\|\right]^2+\sum_{\mu=n+1}^{n+m}\left[\sum_{\nu}^{Z,\mu}\|z_\nu(i_\nu,t)\|\right]^2.\end{aligned}$$

Since $(\sum_{i=1}^k\beta_i)^2\leqq k\sum_{i=1}^k\beta_i^2$, the right-hand side is bounded by

$$\sum_{\mu=n+1}^{n+m}M_\mu\|y_\mu(v_\mu,t)\|^2+\sum_{\nu=1}^{n}M_\nu\|z_\nu(i_\nu,t)\|^2.$$

By (4.13) and (4.10), this is dominated by

$$\sum_{\mu=n+1}^{n+m} M_\mu(Q_{g_\mu}+Q_{\lambda_\mu})^2\|v_\mu\|^2+\sum_{\nu=1}^{n} M_\nu(Q_{r_\nu}+Q_{\gamma_\nu})^2\|i_\nu\|^2.$$

By virtue of (10.1) and (10.2) we can conclude that there exist a $\delta<1$ such that $\|Wx\|\leqq\delta\|x\|$. Thus, condition (i) of Lemma 5 is satisfied.

Now, let $W=R+S$, where R (or S) is the matrix operator consisting of the conductor and resistor (respectively, inductor and capacitor) terms in W. Let M_{λ_μ} (or M_{γ_ν}) be the value that M_μ (respectively, M_ν) would take if all chord conductors (limb-branch resistors) were replaced by open (respectively, short) circuits. Consider S. By the same manipulations as in the preceding paragraph but based upon (4.7) and (4.5), we obtain

$$\|Sx\|^2\leqq\sum_{\mu=n+1}^{n+m} M_{\lambda_\mu}Q_{\lambda_\mu}^2\|v_\mu\|^2+\sum_{\nu=1}^{n} M_{\gamma_\nu}Q_{\gamma_\nu}^2\|i_\nu\|^2,$$

which implies that $S(X)$ is a bounded set in H for every closed sphere X in H. That is, $S(X)$ satisfies condition (i) of Lemma 3.

Next, let t and s be restricted as in § 9. Then, in accordance with that section, we may write for any $x\in H$

$$\begin{aligned}&\|(Sx)(\cdot+s)-(Sx)(\cdot)\|^2\\&=\sum_{\nu=1}^{n}\left\|\sum_{\mu}^{\lambda,\nu}(\pm)\left[\lambda_\mu\left(\rho(t+s)\int_0^{t+s}v_\mu(\omega)\,d\omega,t+s\right)-\lambda_\mu\left(\rho(t)\int_0^t v_\mu(\omega)\,d\omega,t\right)\right]\right\|^2\\&\quad+\sum_{\mu=n+1}^{n+m}\left\|\sum^{\gamma,\mu}(\pm)\left[\gamma_\nu\left(\rho(t+s)\int_0^{t+s}i_\nu(\omega)\,d\omega,t+s\right)-\gamma_\nu\left(\rho(t)\int_0^t i_\nu(\omega)\,d\omega,t\right)\right]\right\|^2\end{aligned}$$

where $\sum^{\lambda,\nu}$ (or $\sum^{\gamma,\mu}$) is the summation that $\sum^{Y,\nu}$ (respectively, $\sum^{Z,\mu}$) would become if all conductor (respectively, resistor) terms were set equal to zero. By virtue of § 9, especially (9.4), the last expression is bounded by

$$\begin{aligned}&\sum_{\nu=1}^{n}\left\{\sum_{\mu}^{\lambda,\nu}[2P_{\lambda_\mu}^2|s|(T+|s|)\|v_\mu\|^2+(L_{\lambda_\mu}|s|T\|v_\mu\|)^2]^{1/2}\right\}^2\\&\qquad+\sum_{\mu=n+1}^{n+m}\left\{\sum^{\gamma,\mu}[2P_{\gamma_\nu}|s|(T+|s|)\|i_\nu\|^2+(L_{\gamma_\nu}|s|T\|i_\nu\|)^2]^{1/2}\right\}^2.\end{aligned}$$

As $s\to0$, the last expression tends to zero uniformly for all x in any given closed sphere X in H. Thus, $S(X)$ satisfies condition (ii) of Lemma 3.

By Lemma 3, therefore, $S(X)$ is conditionally compact in H for every closed sphere X in H. In other words, condition (ii) of Lemma 5 is satisfied. Furthermore, by just the same manipulations as those of the first paragraph of this proof in conjunction with (4.8) and (4.6), we have for every $x, x'\in H$

$$\|Sx-Sx'\|^2\leqq\sum_{\mu=n+1}^{n+m} M_{\lambda_\mu}K_{\lambda_\mu}^2\|v_\mu-v'_\mu\|^2+\sum_{\nu=1}^{n} M_{\gamma_\nu}K_{\gamma_\nu}^2\|i_\nu-i'_\nu\|^2,$$

which shows that S is continuous on H.

In the same way the inequalities (4.2) and (4.4) yield

$$\|Rx-Rx'\|^2\leqq\sum_{\mu=n+1}^{n+m} M_{g_\mu}^2K_{g_\mu}^2\|v_\mu-v'_\mu\|^2+\sum_{\nu=1}^{n} M_{r_\nu}K_{r_\nu}^2\|i_\nu-i'_\nu\|^2.$$

By virtue of (10.3) and (10.4), there is an $\eta < 1$ such that

$$\|Rx - Rx'\| \leqq \eta \|x - x'\|.$$

Thus, R satisfies condition (iii) of Lemma 5 and is therefore continuous on H.

We can now invoke Lemma 5 to conclude the proof of Theorem 2.

11. A stronger existence theorem for a purely resistive time-invariant network. When N consists only of time-invariant continuous conductors and resistors, we can strengthen Theorem 2 by totally dropping the Lipschitz conditions (3.2) and (3.4) and using only boundedness conditions like (3.1) and (3.3). We assume now that each conductor $g: v \mapsto i$ and each resistor $r: i \mapsto v$ is a (linear or nonlinear) continuous mapping of R^1 into R^1 and that

$$|g(v)| \leqq Q_g |v|, \tag{11.1}$$

$$|r(i)| \leqq Q_r |i|, \tag{11.2}$$

where Q_g and Q_r are constants. In general, there will be fixed voltage sources in series with the resistors and fixed current sources in parallel with the conductors, but no other sources will be allowed. We also assume that there are no inductors, capacitors, or mutual couplings. We assume still further that, among the possible choices of the partition $N = \bigcup_{p=1}^{\infty} N_p$, the full set $\mathscr{L}$ of limbs, and the full set $\mathscr{J}$ of joints, there is one for which every limb branch is a series connection of a resistor and a voltage source, either but not both of which may be zero, and every chord is a parallel connection of a conductor and a current source, either but not both of which may be zero. Finally, we let $\nu = 1, \cdots, n$ be the indices of the limb branches and $\mu = n+1, \cdots, n+m$ be the indices of the chords in any given N_p, and we assume that

$$M_{r_\nu} Q_{r_\nu}^2 < 1, \qquad M_{g_\mu} Q_{g_\mu}^2 < 1 \tag{11.3}$$

for all ν and μ, where M_{r_ν} and M_{g_μ} are defined as in § 10.

For any N_p, we can write determining equations just like (5.1) and (5.2) except that the arguments in t disappear and the $y_\mu(v_\mu(t), t)$ and $z_\nu(i_\nu(t), t)$ are replaced by $g_\mu(v_\mu)$ and $r_\nu(i_\nu)$. In the same way as before, this defines the (in general, nonlinear) matrix operator W as a mapping of real $(n+m)$-dimensional Euclidean space R^{n+m} into R^{n+m}. The manipulation indicated in the first paragraph of the proof of Theorem 2 then shows that, for every $x = [i_1, \cdots, i_n, v_{n+1}, \cdots, v_{n+m}]^{\mathrm{T}} \in R^{n+m}$, we have

$$\|Wx\|^2 \leqq \sum_{\mu=n+1}^{n+m} M_{g_\mu} Q_{g_\mu}^2 v_\mu^2 + \sum_{\nu=1}^{n} M_{r_\nu} Q_{r_\nu}^2 i_\nu^2,$$

which by virtue of (11.3) yields $\|Wx\| \leqq \delta \|x\|$ for some fixed $\delta < 1$.

Now, let $d \in R^{n+m}$ be given and set

$$D = \left\{ x \in R^{n+m} : \|x\| \leqq \frac{\|d\|}{1-\delta} \right\}.$$

Also, let $z = d/(1-\delta)$. Then, for every $x \in D$,

$$\|d + Wx\| \leqq \|d\| + \delta \|x\| \leqq (1-\delta)\|z\| + \delta \|z\| = \|z\|.$$

Thus, $d + W \cdot$ maps D into D. Since all resistors and conductors are continuous on R^1, $d + W \cdot$ is continuous on D. So, we may invoke Brouwer's fixed point theorem [3, p. 468] to conclude that that there exists at least one $x_0 \in D$ such that $x_0 - Wx_0 = d$. Upon applying this argument inductively to $N_1, N_2, N_3, \cdots$, we obtain

THEOREM 3. *Let N be a countably infinite electrical network that satisfies the conditions states in the first paragraph of this section. Then, any assignment of voltage-current pairs from $R^1 \times R^1$ to all the joints determines (but not necessarily uniquely) a collection of voltage-current pairs from $R^1 \times R^1$ for all the branches of N that satisfies Ohm's law and Kirchhoff's node and loop laws.*

We finally mention that both Theorems 2 and 3 can also be extended to networks with mutual couplings as indicated in § 7.

REFERENCES

[1] M. S. BERGER, *Nonlinearity and Functional Analysis*, Academic Press, New York, 1977.
[2] V. DOLEZAL, *Nonlinear Networks*, Elsevier, New York, 1977.
[3] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators, Part I*, Interscience, New York, 1966.
[4] H. FLANDERS, *Infinite networks: I—Resistive networks*, IEEE Trans. Circuit Theory, CT-18 (1971), pp. 326–331.
[5] R. HALIN, *Some path problems in graph theory*, Abh. Math. Sem. Hamburg, 44 (1975), pp. 175–186.
[6] R. H. MARTIN, JR., *Nonlinear Operators and Differential Equations in Banach Spaces*, Wiley-Interscience, New York, 1976.
[7] A. H. ZEMANIAN, *The limb analysis of countably infinite electrical networks*, J. Combinatorial Theory Ser. B, 24 (1978), pp. 76–93.
[8] ———, *Nonlinear, resistive, countably infinite, electrical networks*, Applicable Anal., to appear.
[9] ———, *Infinite electrical networks*, Proc. IEEE, 64 (1976), pp. 6–17.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0008 $01.00/0

A NONHOMOGENEOUS INTEGRODIFFERENTIAL EQUATION IN HILBERT SPACE*

RALPH W. CARR† AND KENNETH B. HANNSGEN‡

Abstract. Let $\mathbf{y}(t, \mathbf{x}, \mathbf{f})$ denote the solution of $\mathbf{y}'(t)+\int_0^t [d+a(t-s)]\mathbf{L}\mathbf{y}(s)\,ds=\mathbf{f}(t)$, $t \geqq 0$, $\mathbf{y}(0)=\mathbf{x}$, where $d \geqq 0$ and $\mathbf{L}$ is a selfadjoint densely defined operator on a Hilbert space $\mathcal{H}$ with $\mathbf{L} \geqq \Lambda > 0$. Let $\mathbf{U}(t)\mathbf{x}=\mathbf{y}(t, \mathbf{x}, \mathbf{0})$. By analyzing a related scalar equation with parameter, we find sufficient conditions on the kernel $a(t)$ for $\|\mathbf{U}(t)\| \to 0 (t \to \infty)$ and $\int_0^\infty \|\mathbf{U}(t)\|\,dt < \infty$. These results and a resolvent formula can be combined to reveal the behavior of $\mathbf{y}(t, \mathbf{x}, \mathbf{f})$ as $t \to \infty$.

1. Introduction. Let $\mathbf{L}$ be a selfadjoint (possibly unbounded) linear operator on a Hilbert space $\mathcal{H}$, with spectral decomposition

$$\mathbf{L}\mathbf{x}=\int_{-\infty}^{\infty} \lambda \, d\mathbf{E}_\lambda \, \mathbf{x}$$

for $\mathbf{x}$ in $\mathcal{D}$, the domain of $\mathbf{L}$. We assume that the spectrum of $\mathbf{L}$ is contained in an interval $[\Lambda, \infty)$ with $\Lambda > 0$, so that L is a positive operator. We study the asymptotic behavior $(t \to \infty)$ of solutions of the initial value problem

$$\mathbf{y}'(t)+\int_0^t [d+a(t-s)]\mathbf{L}\mathbf{y}(s)\,ds=\mathbf{f}(t) \qquad (t \geqq 0), \tag{1.1}$$

$$\mathbf{y}(0)=\mathbf{y}_0 \tag{1.2}$$

$(' = d/dt)$, where $\mathbf{y}_0$ and $\mathbf{f}(t)$ belong to $\mathcal{H}$, $d \geqq 0$, and the real-valued kernel a satisfies

$$\begin{array}{l} a \in C(R^+) \cap L^1(0, 1).\ a \text{ is nonnegative, nonincreasing, and} \\ \text{convex on } R^+,\ 0 < a(0+) \leqq \infty, \text{ and } a(\infty)=0. \end{array} \tag{1.3}$$

(In this paper, $R^+=(0, \infty)$, $\bar{R}^+=[0, \infty)$.) See [9] for a discussion (with references and an example) of applications of (1.1) to viscoelasticity theory. Here we remark only that conditions like (1.3) are natural in such applications. Our most precise results on asymptotic behavior will require additional hypotheses lacking evident physical interpretations.

The resolvent kernel of (1.1) is defined by the formula

$$\mathbf{U}(t)=\int_{-\infty}^{\infty} u(t, \lambda)\, d\mathbf{E}_\lambda, \tag{1.4}$$

where $u(t, \lambda)$ is the solution of the scalar problem

$$u'(t)+\lambda \int_0^t [d+a(t-s)]u(s)\,ds=0, \qquad u(0)=1 \tag{1.5}$$

* Received by the editors March 21, 1978. This work was supported by the United States Army under Contract No. DAAG29-75-C-0024 and the National Science Foundation under Grant MPS 74-06403 A01.
† Department of Mathematics, St. Cloud State University, St. Cloud, Minnesota, 56301.
‡ Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. Support by the Mathematics Research Center for a short visit to Madison is gratefully acknowledged.

with parameter λ $(\Lambda \leqq \lambda < \infty, 0 \leqq t < \infty)$. Under certain additional conditions on $a(t)$, we shall show in Theorem 2.2 that

$$\sup_{\Lambda \leqq \lambda < \infty} |u(t, \lambda)| \to 0 \qquad (t \to \infty), \tag{1.6}$$

$$\int_0^\infty \sup_{\Lambda \leqq \lambda < \infty} |u(t, \lambda)|\, dt < \infty. \tag{1.7}$$

It is clear that (1.6) and (1.7) imply, respectively,

$$\|\mathbf{U}(t)\| \to 0 \quad (t \to \infty) \quad \text{and} \quad \int_0^\infty \|\mathbf{U}(t)\|\, dt < \infty. \tag{1.8}$$

In view of the resolvent formula

$$\mathbf{y}(t) = \mathbf{U}(t)\mathbf{y}_0 + \int_0^t \mathbf{U}(t-s)\mathbf{f}(s)\, ds \tag{1.9}$$

for the solution of (1.1), (1.2) (see Theorem 2.1), (1.8) shows, for instance, that $\mathbf{y}(t)$ has a limit in $\mathcal{H}(t \to \infty)$ if $\mathbf{f}(t)$ does and $\|\mathbf{y}\| \in L^1(R^+)$ if $\|f\| \in L^1(R^+)$.

Our results extend those of [10] with respect to the conditions on $a(t)$ as $t \to 0, t \to \infty$. In particular, our results imply that (1.6) and (1.7) hold if a satisfies (1.3) and $-a'$ is convex. Thus, for example, our results include the class of kernels $a(t) = t^{-\beta} (0 < \beta < 1)$.

2. Statement of results. A solution of (1.1) is a continuously differentiable function $\mathbf{y}$ from $\bar{R}^+$ to $\mathcal{H}$ such that $\mathbf{Ly}(t)$ is continuous in t on $\bar{R}^+$ and (1.1) holds. Hille and Phillips [11, pp. 58–89] give the general theory of Bochner integration, which we shall use in studying (1.1) and (1.9). See [19] for the functional calculus of selfadjoint operators.

Our first result, to be proved in § 3, summarizes some earlier work and establishes the resolvent formula.

THEOREM 2.1. (i) *Let* (1.3) *hold. Then the operator* $\mathbf{U}(t)$ *defined by* (1.4) *and* (1.5) *is bounded on* $\mathcal{H}$ *with* $\|\mathbf{U}(t)\| \leqq 1 (t \in \bar{R}^+)$. $\mathbf{U}(t)$ *commutes with* $\mathbf{L}$ *on* $\mathcal{D}$ *and is strongly continuous on* $\bar{R}^+$.

(ii) *If* $\mathbf{y}_0 \in \mathcal{D}$, *if* $\mathbf{f}$: $\bar{R}^+ \to \mathcal{H}$ *is continuous with* $\mathbf{f}(t) \in \mathcal{D}$ *for all t, and if* $\mathbf{Lf}$ *is Bochner integrable on each finite subinterval of* R^+, *then* (1.9) *gives the unique solution of* (1.1), (1.2).

Remark. If $\mathbf{y}_0$, $\mathbf{f}$ are in $\mathcal{H}$ but not necessarily in $\mathcal{D}$, then (as shown in [8] for constant $\mathbf{f}$) (1.9) gives the unique weak solution of an integrated form of (1.1), (1.2).

In proving (1.7), we shall need the technical hypothesis

(2.1) $a(t) = b(t) + c(t)$, *where b and c satisfy* (1.3) *except that either* $b(0+) = 0$ *or* $c(0+) = 0$ *is permitted. Moreover,*

$$\text{(i)} \quad \int_1^\infty t^{-1} b(t)\, dt < \infty \textit{ and}$$

$$\text{(ii)} \quad -c' \quad \textit{is convex on } R^+.$$

The Fourier transform of a will be denoted

$$\hat{a}(\tau) \equiv \varphi(\tau) - i\tau\theta(\tau) \equiv \int_0^\infty e^{-i\tau t} a(t)\, dt. \tag{2.2}$$

Under hypothesis (1.3), $\hat{a}(\tau)$ is continuous, and $\varphi(\tau)$ and $\theta(\tau)$ are nonnegative for $\tau>0$ [4].

The frequency conditions

$$(2.3)\qquad \begin{aligned} &\text{(i)}\quad \varphi(\tau)>0 \qquad (\tau>0),\\ &\text{(ii)}\quad \limsup_{\tau\to\infty}\frac{\theta(\tau)}{\varphi(\tau)}<\infty \end{aligned}$$

are crucial for (1.6) and (1.7); we indicate briefly their role. From [4] we know that if (1.3) holds, (2.3(i)) fails to hold if and only if $a(t)$ is piecewise linear with changes of slope only at integer multiples of a single positive number; $u(t,\lambda)$ is then asymptotic $(t\to\infty)$ to a nonconstant periodic function, so neither (1.6) nor (1.7) holds. If, on the other hand, (1.3) and (2.3i) hold, a result of Shea and Wainger [20] shows that

$$\int_0^\infty |u(t,\lambda)|\,dt<\infty \qquad (\lambda>0).$$

It is then easy to show from (1.5) that

$$(2.4)\qquad \hat{u}(\tau,\lambda)=\frac{1}{\lambda[\varphi(\tau)+i\tau(\lambda^{-1}-\theta(\tau)-d\tau^{-2})]} \qquad (\tau>0).$$

In proving Theorem 2.2, we shall show that if λ is sufficiently large, $d/\tau^2+\theta(\tau)=\lambda^{-1}$ for exactly one positive number $\tau=\omega(\lambda)$ with $\omega(\lambda)$ continuous and $\omega(\lambda)\uparrow\infty$ as $\lambda\uparrow\infty$. From (2.4), it follows that

$$\int_0^\infty |u(t,\lambda)|\,dt\geqq|\hat{u}(\omega(\lambda))|\geqq\theta(\omega(\lambda))/\varphi(\omega(\lambda)).$$

This shows the necessity of (2.3(ii)) for (1.7).

THEOREM 2.2. *Assume that* (1.3) *holds. Then*

(i) (1.6) *holds if* (2.3) *holds,*

(ii) (1.7) *holds if* (2.1) *and* (2.3) *holds, and*

(iii) *if* (1.7) *holds, then* (2.3) *holds.*

If (1.3) and (2.3(i)) hold and $a(0+)<\infty$, then (2.3(ii)) holds if and only if $a(t)$ is strongly positive (that is, $(1+\tau^2)\varphi(\tau)$ is bounded away from zero; see [10]). In § 7 below we shall give an example (with $a(0+)=\infty$) where $a(t)$ is strongly positive but (2.3(ii)) does not hold. In the same section, we shall prove the following positive result.

COROLLARY 2.1. *If* (2.1) *holds and either* (i) $c\equiv 0$, $b(0+)<\infty$, *and b is strongly positive, or* (ii)

$$\limsup_{x\to 0+}\frac{\int_0^x b(t)\,dt}{\int_0^x c(t)\,dt}<\infty,$$

then (1.6) *and* (1.7) *hold.*

Thus, in particular, (1.8) holds if $a(t)$ satisfies (1.3) and $-a'(t)$ is convex.

Integration of (1.5) (see also [11], [4]) shows that

$$u(t,\lambda)+\lambda\int_0^t [d+a(t-s)]\int_0^s u(r,\lambda)\,dr\,ds=1,$$

so that when (1.3) and (2.3(i)) hold,

$$\int_0^\infty u(t,\lambda)\,dt=1\bigg/\left(\lambda\int_0^\infty [d+a(t)]\,dt\right)$$

(interpreted as zero if $d+a(t)\notin L^1(R^+)$.) Thus in Theorem 2.2 we have

$$\int_0^\infty \mathbf{U}(t)\,dt=\mathbf{L}^{-1}\Big/\left(\int_0^\infty [d+a(t)]\,dt\right).$$

Detailed statements about the asymptotic nature of $u(t,\lambda)$ as $t\to\infty$ (λ fixed) are given for certain special cases by Levin and Nohel [13], [14], by the second author [5], and by Wong and Wong [24]. For example if $d+a(t)=t^{-\beta}(0<\beta<1)$, Corollary 2.1 applies and [5, Cor. 3.3] shows that

$$u(t,\lambda)\sim\frac{C}{\lambda}t^{\beta-2}\qquad (t\to\infty).$$

On the other hand (see § 4), there is a $C'>0$ such that

$$\int_0^\infty |u(t,\lambda)|\,dt\geq C'\lambda^{-1/(2-\beta)}\qquad (\lambda>0). \tag{2.5}$$

Thus the asymptotic behavior of $u(t,\lambda)$ as $\lambda\to\infty$ is not completely clear. See [14], [9] for further discussion.

Another useful example, where (1.5) can be solved explicitly, is $d+a(t)=e^{-t}$. Then [10] (1.5) reduces to an ordinary differential equation, and

$$\begin{aligned} u(t,\lambda)&=e^{-t/2}(\cos\mu t+\tfrac{1}{2}\mu^{-1}\sin\mu t)\qquad (\lambda\neq\tfrac14),\\ u(t,\tfrac14)&=e^{-t/2}(1+\tfrac12 t),\end{aligned} \tag{2.6}$$

where $\mu=\frac12(4\lambda-1)^{1/2}$ (λ and μ may be real or complex). For this example, we remark that

(i) $t\to\mathbf{U}(t)$ is not continuous in the norm topology if $\mathbf{L}$ is unbounded, since $\|\mathbf{U}(t)-\mathbf{U}(s)\|\geqq|u(t,\lambda)-u(s,\lambda)|$ for λ in the spectrum of $\mathbf{L}$.

(ii) $\int_0^\infty|u(t,\lambda)|\,dt\geqq 1$ ($\lambda>0$). This is proved in [10].

(iii) $u(\cdot,\lambda e^{i\varphi})\notin L^p(R^+)$ if $p\geqq 1$, $e^{i\varphi}\neq 1$, and $\lambda>0$ is sufficiently large.

Dafermos [2] and Slemrod [21] study equations similar to (1.1) as linear models in viscoelasticity and fluid mechanics respectively. [2] concerns equations of a much more general form than (1.1), but in overlapping cases Dafermos' results on asymptotic behavior require $d>0$, $a(0+)<\infty$, and $\|\mathbf{f}'\|\in L^1(R^+)$. The corresponding conclusion in [2] is that $\|\mathbf{y}'(t)\|+\|\mathbf{L}^{1/2}y'(t)\|\to 0$ as $t\to\infty$ if $\mathbf{y}_0$ is in the domain of $\mathbf{L}^{1/2}$. The hypotheses in [21] are quite close to ours, and sufficient conditions are given for solutions to tend to zero ($t\to\infty$) in a fading memory space. Neither [2] nor [21] contains an analogue of (1.8); consequently, the results cover a more restricted class of forcing terms and do not give conditions ensuring that $\|\mathbf{y}\|\in L^1(R^+)$.

Our results and methods are closer to those of Friedman and Shinbrot [3], who obtain L^p estimates ($1\leqq p<\infty$) for the resolvent (fundamental solution) $\mathbf{S}(t)$ of

$$\mathbf{y}(t)+\int_0^t h(t-s)\mathbf{L}\mathbf{y}(s)=\mathbf{F}(t) \tag{2.7}$$

in Banach space. Formal differentiation of (2.7) yields (1.1) if $h'(t)=d+a(t)$, $h(0)=0$, $\mathbf{F}'=\mathbf{f}$. For their L^p estimates, Friedman and Shinbrot require at least $h(0)>0$, $h'\in L^1(R^+)$.

Miller and Wheeler [17] use procedures similar to those of [3] to study the equation

$$\mathbf{y}'(t)=-\mathbf{L}\mathbf{y}(t)-\int_0^t a(t-s)(\mathbf{L}+l\mathbf{I})\mathbf{y}(s)\,ds+\mathbf{f}(t) \tag{2.8}$$

in Hilbert space. Here $\mathbf{L}$ is self-adjoint and bounded below and has a compact resolvent. Miller and Wheeler give conditions under which the resolvent for (2.8) may be decomposed into an exponential polynomial with finite-dimensional projections as coefficients and a remainder ("residual resolvent") $\mathbf{R}(t)$ with $\|\mathbf{R}(t)\| \in L^p(R^+)$.

The proofs of these results in [3] and [17] use the operational calculus based on contour integrals and estimates such as

$$\int_0^\infty |r(t,\lambda)|^p\,dt \leqq C|\lambda|^{-\delta} \tag{2.9}$$

($|\arg\lambda| \leqq (\pi/2)-\varepsilon$), where $\varepsilon, \delta > 0$ and $r(t,\lambda)$ is the solution of a certain scalar equation (analogous to (1.5)) with complex parameter λ. Remarks (ii) and (iii) following (2.6) above show that estimates like (2.9) need not hold for our function $u(t,\lambda)$.

Questions of existence and uniqueness are easily settled under the hypotheses of our theorems, as will be seen in § 3. In Theorem 2.1, only the sufficient conditions for the resolvent formula are new.

For a broader treatment of existence, uniqueness, and continuous dependence for equations like (1.1) in Banach space, see Miller [16]; further discussion of the resolvent formula (1.9) will also be found in [16].

Finally, we remark that nonlinear versions of (1.1) are under active study by many authors. See, for example, [1], [15], [25], [26], [27].

3. Proof of Theorem 2.1. (i) The proof of Theorem 2 of [6], with $a(t)$ replaced by $d+a(t)$ (and the last equalities corrected to read $2V(0)=u^2(0)=1$), shows that $|u(t,\lambda)| \leqq 1$ ($t \in \bar{R}^+$, $\lambda \in R^+$), so $\|\mathbf{U}(t)\| \leqq 1$ and $\mathbf{U}(t)\mathbf{L} = \mathbf{L}\mathbf{U}(t)$ on $\mathcal{D}$. Since

$$\|\mathbf{U}(t)\mathbf{x} - \mathbf{U}(s)\mathbf{x}\|^2 = \int_\Lambda |u(t,\lambda) - u(s,\lambda)|^2\,d(\mathbf{E}_\lambda \mathbf{x}, \mathbf{x}),$$

the continuity of $u(t,\lambda)$ in t and the dominated convergence theorem imply that $\mathbf{U}(t)$ is strongly continuous.

The computations for (ii) are formally the same as those for Theorem 2 of [9], where a and $\mathbf{Lf}$ are continuous on $\bar{R}^+$. To simplify formulas we take $d=0$ since this does not change the following argument. It is obvious that the function $\mathbf{y}(t)$ of (1.9) satisfies (1.2). Let $T(t)$ be the triangle $\{0 \leqq r < s \leqq t\}(t>0)$, and let $h(t) = \int_0^t a(s)\,ds$. Since $\mathbf{L}\mathbf{U}(r)\mathbf{y}_0 = \mathbf{U}(r)\mathbf{L}\mathbf{y}_0$ is continuous ($r \in \bar{R}^+$), $a(s-r)\mathbf{L}\mathbf{U}(r)\mathbf{y}_0$ is in $L^1(T(t))$ and

$$\begin{aligned} \mathbf{y}_0 - \int_0^t \int_0^s a(s-r)\mathbf{L}\mathbf{U}(r)\mathbf{y}_0\,dr\,ds \\ &= \mathbf{y}_0 - \int_0^t h(t-r)\mathbf{L}\mathbf{U}(r)\mathbf{y}_0\,dr \\ &= \mathbf{y}_0 - \int_0^t h(t-r)\left[\lim_{M\to\infty}\int_{\lambda_0}^M \lambda u(r,\lambda)\,d\mathbf{E}_\lambda \mathbf{y}_0\right] dr \\ &= \mathbf{y}_0 - \int_{\lambda_0}^\infty \left[\lambda \int_0^t h(t-r)u(r,\lambda)\,dr\right] d\mathbf{E}_\lambda\,\mathbf{y}_0. \end{aligned} \tag{3.1}$$

The expression in brackets here is just $1-u(t,\lambda)$, as one sees by integrating (1.5); thus the left-hand side of (3.1) is equal to $\mathbf{U}(t)\mathbf{y}_0$ and differentiation establishes that

$$\frac{d}{dt}[\mathbf{U}(t)\mathbf{y}_0] = -\int_0^t a(t-s)\mathbf{L}\mathbf{U}(s)\mathbf{y}_0\,ds. \tag{3.2}$$

We observe next that the strong continuity and uniform boundedness of $\mathbf{U}$ ensure that the function

$$a(t-s)\mathbf{LU}(s-r)\mathbf{f}(r) = a(t-s)\mathbf{U}(s-r)\mathbf{Lf}(r)$$

is strongly measurable on $T(t)$. In view of our hypotheses and [11, Thm. 3.5.4], the following lemma establishes this. (Compare [16, Lemma 2.1].)

LEMMA 3.1. *If* $\mathbf{g}$: $R^+ \to \mathcal{H}$ *belongs to* $B^1(0, t)$, *then the function* $\mathbf{G}(s, r) = \mathbf{U}(s-r)\mathbf{g}(r)$ *is strongly measurable on* $T(t)$.

Proof. To simplify notation, take $t = 1$. For each positive integer n, let $\mathbf{U}_{n,j} = \mathbf{U}(j/n)$, $E_{n,j} = ((j-1)/n, j/n](1 \leqq j \leqq n)$. Let $\mathbf{g}_n$ be a sequence of countably-valued functions

$$\mathbf{g}_n(r) = \sum_{k=1}^{\infty} \chi_{n,k}(r)\mathbf{g}_{n,k}$$

($\chi_{n,k}$ = the characteristic function of a measurable set $\Omega_{n,k}$) such that $\mathbf{g}_n(r) \to \mathbf{g}(r)(n \to \infty)$ except on a set Z of measure zero. For $(s, r) \in T(1)$, let $j(s, r, n)$ be the integer such that $s - r \in E_{n,j(s,r,n)}$, and let

$$\mathbf{G}_n(s, r) = \mathbf{U}_{n,j(s,r,n)}\mathbf{g}_n(r).$$

Then $\mathbf{G}_n(s, r)$ is measurable and countably valued since $T(1)$ is the union of the measurable sets

$$\{s - r \in E_{n,m}\} \cap \{r \in \Omega_{n,k}\}$$

$(1 \leqq m \leqq n, 1 \leqq k < \infty)$, on each of which $\mathbf{G}_n$ is constant. For fixed $(s, r) \in T(1)$, $r \notin Z$,

$$\begin{aligned}\|\mathbf{G}_n(s, r) - \mathbf{G}(s, r)\| &\leqq \|\mathbf{U}_{n,j(s,r,n)}[\mathbf{g}_n(r) - \mathbf{g}(r)]\| \\ &\quad + \|[\mathbf{U}_{n,j(s,r,n)} - \mathbf{U}(s-r)]\mathbf{g}(r)\|.\end{aligned}$$

As $n \to \infty$, the first term tends to zero, since $\|U\| \leqq 1$ and $\mathbf{g}_n(r) \to \mathbf{g}(r)$; by strong continuity, the second term tends to zero as well. Thus $\mathbf{G}(s, r)$ is the limit almost everywhere of countably-valued measurable functions, and the lemma is proved.

Continuing the proof of Theorem 2.1, we note that

$$\int_0^t \int_0^s a(t-s)\|\mathbf{U}(s-r)\mathbf{Lf}(r)\|\, dr\, ds \leqq \int_0^t a(s)\, ds \int_0^t \|\mathbf{Lf}(r)\|\, dr < \infty,$$

so $a(t-s)\mathbf{U}(s-r)\mathbf{Lf}(r) \in B^1(T(t))$. Then, using Fubini's theorem, a change of variable, and the fact that $\mathbf{L}$ is closed, we may compute

$$\begin{aligned}\int_0^t a(t-s)\mathbf{L}\int_0^s \mathbf{U}(s-r)\mathbf{f}(r)\, dr\, ds &= \int_0^t \int_0^s a(t-s)\mathbf{LU}(s-r)\mathbf{f}(r)\, dr\, ds \\ &= \int_0^t \int_0^{t-r} [a(t-r-s)\mathbf{LU}(s)\mathbf{f}(r)]\, ds\, dr \qquad (3.3)\\ &= -\int_0^t \frac{d}{dt}[\mathbf{U}(t-r)\mathbf{f}(r)]\, dr,\end{aligned}$$

where the last step uses (3.2) and $\mathbf{f}(r)$ in place of $\mathbf{y}_0$. It is clear from these equalities that the integrand in the last expression is locally Bochner integrable in (t, r); using Fubini's theorem, we see that this expression (and hence the left-hand side (3.3)) is equal to

$$\mathbf{f}(t) - \frac{d}{dt}\int_0^t \mathbf{U}(t-r)\mathbf{f}(r)\, dr.$$

In view of (3.2) this establishes (1.1).

For uniqueness, we pass to the weak, integrated version of (1.1), (1.2) and project on $\mathbf{E}_\Lambda\mathscr{H}$; see [9] or [7] for details.

4. Proof of Theorem 2.2. Reduction to two estimates. We assume without further mention that $d+a(t)$ has been rescaled, if necessary, so that $\Lambda=1$. The functions a', b', and c' are redefined where necessary so as to be continuous from the left on $\mathbb{R}^+$. We let $A(t)=\int_0^t a(r)\,dr$.

The proof relies on detailed information about $\hat{a}$ (see (2.2)). See [4], [20] for earlier versions of these ideas.

LEMMA 4.1. *Suppose* (1.3) *holds. Then φ and θ are continuously differentiable on* $\mathbb{R}^+$ *with*

$$\frac{1}{2\sqrt{2}}A\left(\frac{1}{\tau}\right)\leqq|\hat{a}(\tau)|\leqq 4A\left(\frac{1}{\tau}\right)\qquad(\tau>0),\tag{4.1}$$

$$|\hat{a}'(\tau)|\leqq 40\int_0^{1/\tau}ra(r)\,dr\qquad(\tau>0),\tag{4.2}$$

$$\frac{1}{5}\int_0^{1/\tau}ra(r)\,dr\leqq\theta(\tau)\leqq 12\int_0^{1/\tau}ra(r)\,dr\qquad(\tau>0),\tag{4.3}$$

$$-\theta'(\tau)\geqq\frac{\tau}{5}\int_0^{1/\tau}r^3a(r)\,dr\qquad(\tau>0).\tag{4.4}$$

Our proof is adapted from [10, Lemma 2.2]. We exploit the fact that $da'(t)$ is a positive measure on R^+ and adapt the convention, consistent with our choice of $a'(t)=a'(t-)$ that when $0\leqq x\leqq y$ and $f\in L^1(da'(t))$,

$$\int_x^y f(t)\,da'(t)=\int_{[x,y)}f(t)\,da'(t).$$

Convexity of $a(t)$ implies $a(t/2)-a(t)\geqq -ta'(t)/2$, and hence

$$2\int_0^{t/2}a(r)\,dr\geqq ta\left(\frac{t}{2}\right)\geqq ta(t)-\frac{t^2}{2}a'(t)\geqq 0\qquad(t\geqq 0).\tag{4.5}$$

In particular (4.5) shows that $ta(t)+(t^2/2)|a'(t)|=o(1)$ $(t\to 0+)$. We also have $ta'(t)=o(1)$ $(t\to\infty)$, as a consequence of (1.3) and

$$\int_T^\infty r\,da'(r)=a(T)-Ta'(T)<\infty\qquad(T>0).\tag{4.6}$$

Two integrations by parts in (2.2) yield the formula

$$\hat{a}(\tau)=\tau^{-2}\int_0^\infty(1-i\tau r-e^{-i\tau r})\,da'(r)\qquad(\tau>0),\tag{4.7}$$

where (4.5) and (4.6) assure vanishing of the boundary terms and absolute convergence of the integral.

Following [20], we let $J(u)=iu(1-e^{iu})-2(1+iu-e^{iu})$; then

$$|J(u)|\leqq\tfrac{1}{4}u^3\qquad(0\leqq u\leqq 1),\quad\text{and}\quad|J(u)|\leqq 2(u+2)\qquad(u\geqq 0).\tag{4.8}$$

(4.8), combined with Fubini's theorem, justifies differentiation of (4.7) and gives us

$$\hat{a}'(\tau)=\tau^{-3}\int_0^\infty J(-\tau r)\,da'(r) \qquad (\tau>0). \tag{4.9}$$

The inequalities (4.1) and (4.2) now follow as in [20].

From (4.7) we have

$$\theta(\tau)=\tau^{-2}\int_0^\infty rK(\tau r)\,da'(r) \qquad (\tau>0), \tag{4.10}$$

with

$$K(u)=1-\frac{\sin u}{u} \qquad (u>0).$$

Note that

$$K(u)\geqq\begin{cases}\dfrac{u^2}{6}-\dfrac{u^2}{120}\geqq\dfrac{u^2}{30} & (0\leqq u\leqq 1)\\[2ex] 1-\max\left\{\sin 1,\dfrac{2}{\pi}\right\}\geqq\dfrac{1}{10} & (u\geqq 1).\end{cases}$$

Therefore,

$$\theta(\tau)\geqq\frac{1}{30}\int_0^{1/\tau} r^3\,da'(r)+\frac{1}{10\tau^2}\int_{1/\tau}^\infty r\,da'(r), \qquad (\tau>0). \tag{4.11}$$

The relations

$$\begin{aligned}\int_{1/\tau}^\infty r\,da'(r)&=-\frac{1}{\tau}a'\left(\frac{1}{\tau}\right)+a\left(\frac{1}{\tau}\right),\\ \int_0^{1/\tau} r^3\,da'(r)&=\frac{1}{\tau^3}a'\left(\frac{1}{\tau}\right)-\frac{3}{\tau^2}a\left(\frac{1}{\tau}\right)+6\int_0^{1/\tau} ra(r)\,dr,\end{aligned} \tag{4.12}$$

along with (4.11), give us the first inequality in (4.3). The second inequality follows from the estimates

$$|K(u)|\leqq 2u^2 \quad (0\leqq u\leqq 1) \quad\text{and}\quad |K(u)|\leqq 2 \qquad (u>0),$$

which, along with (4.12), yield

$$\theta(\tau)\leqq 2\int_0^{1/\tau} r^3\,da'(r)+2\tau^{-2}\int_{1/\tau}^\infty r\,da'(r)\leqq 12\int_0^{1/\tau} ra(r)\,dr.$$

To prove (4.4) we differentiate (4.10), which yields

$$-\tau^4\theta'(\tau)=\int_0^\infty H(\tau r)\,da'(r), \qquad (\tau>0), \tag{4.13}$$

where

$$H(u)=3(u-\sin u)-u(1-\cos u).$$

Then

$$(4.14)\qquad \begin{aligned} &H^{(j)}(0)=0 \qquad (j=0,1,2,3,4),\\ &H^{(5)}(0)=2,\\ &-4\leqq -3\sin u-u\cos u=H^{(6)}(u)\leqq 0 \qquad (0\leqq u\leqq 1),\end{aligned}$$

so

$$(4.15)\qquad H(u)\geqq \frac{2u^5}{5!}-\frac{4u^6}{6!}=\frac{u^5}{180}(3-u)\geqq\frac{u^5}{100}\qquad (0\leqq u\leqq 1),$$

and

$$(4.16)\qquad H'(u)\geqq \frac{2u^4}{4!}-\frac{4u^5}{5!}=\frac{u^4}{60}(5-2u)\geqq\frac{u^4}{20}\qquad (0\leqq u\leqq 1).$$

If we can show that $H''(u)\geqq 0$ $(1\leqq u\leqq 4)$, it will then follow from (4.15) and (4.16) that

$$(4.17)\qquad H(u)\geqq\frac{1}{100}+\frac{1}{20}(u-1)=\frac{1}{100}(5u-4)\qquad (1\leqq u\leqq 4).$$

But $H^{(3)}(u)=u\sin u\geqq 0$ $(0\leqq u\leqq \pi)$; by (4.14), we conclude that $H''(u)\geqq 0$ $(0\leqq u\leqq\pi)$. Then since $H^{(3)}(u)\leqq 0$ $(\pi\leqq u\leqq 4\pi/3)$, $H^{(2)}(u)\geqq H^{(2)}(4\pi/3)>0$ $(\pi\leqq u\leqq 4)$, so (4.17) follows.

It is easy to see that

$$(4.18)\qquad H(u)\geqq u-3\qquad (u\geqq 4),$$

and thus (4.13), along with (4.15), (4.17), and (4.18), gives

$$(4.19)\qquad -\tau^4\theta'(\tau)\geqq\frac{\tau^5}{100}\int_0^{1/\tau} r^5\,da'(r)+\frac{1}{100}\int_{1/\tau}^{4/\tau}(5\tau r-4)\,da'(r)+\int_{4/\tau}^{\infty}(\tau r-3)\,da'(r),$$

and

$$(4.20)\qquad \begin{aligned} &\frac{\tau^5}{100}\int_0^{1/\tau} r^5\,da'(r)=\tfrac{1}{100}a'\left(\frac{1}{\tau}\right)-\frac{\tau}{20}a\left(\frac{1}{\tau}\right)+\frac{\tau^5}{5}\int_0^{1/\tau} r^3a(r)\,dr,\\ &\frac{1}{100}\int_{1/\tau}^{4/\tau}(5\tau r-4)\,da'(r)=\frac{4}{25}a'\left(\frac{4}{\tau}\right)-\frac{1}{100}a'\left(\frac{1}{\tau}\right)-\frac{\tau}{20}a\left(\frac{4}{\tau}\right)+\frac{\tau}{20}a\left(\frac{1}{\tau}\right),\\ &\int_{4/\tau}^{\infty}(\tau r-3)\,da'(r)=-a'\left(\frac{4}{\tau}\right)+\tau a\left(\frac{4}{\tau}\right).\end{aligned}$$

Combining (4.19) and (4.20) we obtain (4.4). This completes the proof of Lemma 4.1.

Using (4.3) and (4.4) we see that the equation

$$(4.21)\qquad \theta(\omega)+\frac{d}{\omega^2}=\frac{1}{\lambda}$$

defines a strictly increasing, continuously differentiable function $\omega=\tilde{\omega}(\lambda)$ on a subinterval (λ_0,∞) of $\mathbb{R}^+$ with $\tilde{\omega}(\lambda)\to\infty$ as $\lambda\to\infty$. This provides the missing step in the proof of Theorem 2.2(iii) outlined just above the statement of the theorem.

Fix $t_1>0$ such that $a(t_1)>0$ and let $\rho=6t_1^{-1}$. Set $\omega(\lambda)=\max\{\rho,\tilde{\omega}(\lambda)\}$ if $\tilde{\omega}(\lambda)$ is defined, $\omega(\lambda)=\rho$ otherwise.

(4.3) and (4.21) yield

$$\frac{1}{\lambda} \geqq \theta(\omega) \geqq \frac{1}{5} \int_0^{1/\omega} ra(r)\, dr, \tag{4.22}$$

and

$$\int_0^{1/\omega} ra(r)\, dr \geqq \frac{1}{2\omega^2} a\left(\frac{1}{\omega}\right) \geqq \frac{a(t_1)}{2\omega^2}. \tag{4.23}$$

In particular, (4.22) and (4.23) show that

$$\omega^2(\lambda) \geqq \frac{a(t_1)}{10} \lambda. \tag{4.24}$$

In [10] it was also shown that when $a(0+) < \infty$, $\omega^2(\lambda)$ is bounded above by some constant times λ. Such an estimate is not available to us when $a(0+) = \infty$, and this causes the principal new difficulties in the proof of Theorem 2.2.

When $a(t) = t^{-\beta}$ $(0 < \beta < 1, t > 0)$, direct computations show that $\omega(\lambda) = K\lambda^{1/(2-\beta)}$ $(K = K(\beta) > 0)$. Now the inequality $\int_0^\infty |u(t, \lambda)|\, dt \geqq |\hat{u}(\omega(\lambda))| \geqq \theta(\omega(\lambda))/\varphi(\omega(\lambda))$ can be used to derive the estimate (2.5).

From (4.3), (4.21), and (4.23) we obtain

$$\frac{1}{\lambda} \leqq 12 \int_0^{1/\omega} ra(r)\, dr + \frac{d}{\omega^2} \leqq \left(12 + \frac{2d}{a(t_1)}\right) \int_0^{1/\omega} ra(r)\, dr \tag{4.25}$$

whenever

$$\lambda \geqq \lambda_0 = \max\left\{\left(\theta(\rho) + \frac{d}{\rho^2}\right)^{-1}, 1\right\}.$$

On the other hand when $1 \leqq \lambda \leqq \lambda_0$, $\omega(\lambda) = \rho$ and we have

$$\frac{1}{\lambda} \leqq 1 \leqq \lambda_0 \left(12 + \frac{2d}{a(t_1)}\right) \int_0^{1/\rho} ra(r)\, dr. \tag{4.26}$$

Then, combining (4.22), (4.25) and (4.26) we find that

$$\frac{1}{5} \int_0^{1/\omega} ra(r)\, dr \leqq \frac{1}{\lambda} \leqq C \int_0^{1/\omega} ra(r)\, dr \qquad (\lambda \geqq 1), \tag{4.27}$$

where $C = [12 + 2d/(a(t_1))]\lambda_0$. Define

$$\begin{aligned} &D(\tau) \equiv D(\tau, \infty) = \hat{a}(\tau) - i\, d\tau^{-1} \\ &D(\tau, \lambda) = D(\tau) + i\tau\lambda^{-1} = \varphi(\tau) + i\tau(\lambda^{-1} - \theta(\tau) - d\tau^{-2}) \qquad (1 \leqq \lambda \leqq \infty, \tau > 0). \end{aligned} \tag{4.28}$$

If (2.3(i)) holds then $|D(\tau, \lambda)| \geqq \varphi(\tau) > 0$ $(\tau > 0, 1 \leqq \lambda \leqq \infty)$ and [4] gives the representation

$$\pi u(t, \lambda) = \frac{1}{\lambda} \int_0^\infty \mathrm{Re}\left\{\frac{e^{i\tau t}}{D(\tau, \lambda)}\right\} dt \qquad (0 < t < \infty, 0 < \lambda < \infty). \tag{4.29}$$

(The integral is improper at $\tau = \infty$; by (1.3) and (4.1) $[D(\tau, \lambda)]^{-1}$ is continuous at $\tau = 0$ and for every $\tau > 0$.) Moreover the result of Shea and Wainger [20] shows that

$$\int_0^\infty |u(t, \lambda)|\, dt < \infty \qquad (\lambda > 0), \tag{4.30}$$

and in [6] it was shown that (1.3) implies

$$\sup_{t\geqq 0}\sup_{1\leqq\lambda\leqq\infty}|u(t,\lambda)|=1. \tag{4.31}$$

For the remainder of the proof, unless noted otherwise, we assume that (1.3), (2.1), and (2.3) hold. Return to (4.29) and integrate by parts. There results the formula

$$\pi u(t,\lambda)=\operatorname{Re}\left\{\frac{1}{it\lambda}\int_0^\infty e^{i\tau t}\frac{D_\tau(\tau,\lambda)}{[D(\tau,\lambda)]^2}\,d\tau\right\}\qquad (t>0,\lambda>0). \tag{4.32}$$

Relations (4.1) and (4.2) show that the boundary terms vanish and that the integral converges absolutely when $d\neq 0$. Absolute convergence of the integral when $d=0$ is assured by an estimate of Shea and Wainger [20, pp. 322–323], namely

$$\int_0^\rho \frac{\int_0^{1/\tau} ra(r)\,dr}{(\int_0^{1/\tau} a(r)\,dr)^2}\,d\tau<\infty. \tag{4.33}$$

Note that

$$\frac{1}{D(\tau,\lambda)}-\frac{1}{D(\tau)}=\frac{-i\tau\lambda^{-1}}{D(\tau,\lambda)D(\tau)}.$$

Then

$$\begin{aligned}\frac{D_\tau(\tau,\lambda)}{[D(\tau,\lambda)]^2}&=\frac{D_\tau(\tau,\lambda)}{[D(\tau)]^2}\left[1-\frac{i\tau\lambda^{-1}}{D(\tau,\lambda)}\right]^2\\&=\frac{[D'(\tau)+i\lambda^{-1}]}{[D(\tau)]^2}\left[1-\frac{2i\tau\lambda^{-1}}{D(\tau)}\right]-\frac{\tau^2D_\tau(\tau,\lambda)}{\lambda^2[D(\tau)]^2D(\tau,\lambda)}\left[\frac{2}{D(\tau)}+\frac{1}{D(\tau,\lambda)}\right].\end{aligned} \tag{4.34}$$

Define

$$\begin{aligned}u_1(t)&=\frac{1}{t}\int_0^\rho e^{i\tau t}\frac{D'(\tau)}{[D(\tau)]^2}\,d\tau &&(t>0),\\u_2(t)&=\frac{1}{t}\int_0^\rho \frac{e^{i\tau t}}{[D(\tau)]^2}\left[1-\frac{2D'(\tau)}{D(\tau)}\right]d\tau &&(t>0),\\u_3(t)&=\frac{1}{t}\int_0^\rho e^{i\tau t}\frac{2\tau}{[D(\tau)]^3}\,d\tau &&(t>0),\\u_4(t,\lambda)&=\frac{-1}{\lambda^3 t}\int_0^\rho e^{i\tau t}\frac{\tau^2D_\tau(\tau,\lambda)}{[D(\tau)]^3D(\tau,\lambda)}\left[\frac{2}{D(\tau)}+\frac{1}{D(\tau,\lambda)}\right]d\tau &&(t>0),\\u_5(t,\lambda)&=\frac{1}{t\lambda}\int_\rho^\infty e^{i\tau t}\frac{D_\tau(\tau,\lambda)}{[D(\tau,\lambda)]^2}\,d\tau &&(t>0).\end{aligned} \tag{4.35}$$

Referring to (4.32) and (4.34) we see that

$$u(t,\lambda)=\operatorname{Im}\{\lambda^{-1}u_1(t)+i\lambda^{-2}u_2(t)+\lambda^{-3}u_3(t)+u_4(t,\lambda)+u_5(t,\lambda)\}. \tag{4.36}$$

In § 6 we will show that if (1.3), (2.1) and (2.3) hold then

$$|u_j(t,\lambda)|\leqq Mq(t)\qquad (t>0,1\leqq\lambda<\infty,j=4,5), \tag{4.37}$$

where, now and henceforth, M (or M_j) denotes a positive constant independent of λ whose value may change from line to line, and

$$q(t)=t^{-2}\int_0^t b(r)\,dr+t^{-2}+t^{-1}b(t)-b'(t), \qquad (t>0).$$

The assumption (2.1(i)) combined with an integration by parts shows that

$$\int_1^T t^{-2}\int_0^t b(r)\,dr\,dt=-T^{-1}\int_0^T b(t)\,dt+\int_0^1 b(t)\,dt+\int_1^T \frac{b(t)}{t}\,dt$$

from which it follows easily that $q(t)\in L^1(1,\infty)$.

In view of (4.30), (4.36), and (4.37) we find that $u_j(t)\in L^1(1,\infty)$ $(j=1,2,3)$, and this, along with (4.31), implies that

$$\int_0^\infty \sup_{1\leqq\lambda<\infty}|u(t,\lambda)|\,dt<\infty.$$

This is the assertion of Theorem 2.2(ii).

Similar estimates which hold when (1.3) and (2.3) hold, but without the assumptions of (2.1), can then be used to show that

$$(4.38)\quad |u_1(t)|+|u_2(t)|+|u_3(t)|+|u_4(t,\lambda)|+|u_5(t,\lambda)|\leqq\frac{M}{t} \qquad (1\leqq\lambda<\infty,\ 0<t),$$

from which Theorem 2.2(i) follows. (Theorem 2.2(iii) has already been proved above.)

5. Two more lemmas—some important estimates. We would like to integrate by parts in the formulas for u_4 and u_5 in order to bring out another factor of t^{-1}. Hypothesis (2.1(i)) appears to be too weak to permit this, so we separate $\hat{b}'$ into a "small" part and a differentiable part.

We again let $J(u)=iu(1-e^{iu})-2(1-iu-e^{iu})$ and define

$$(5.1)\quad \begin{aligned} \beta^0(t,\tau)&=\tau^{-3}\int_0^t J(-\tau r)\,db'(r) \qquad (0<t<\infty,\ 0<\tau<\infty),\\ \beta^\infty(t,\tau)&=\tau^{-3}\int_t^\infty J(-\tau r)\,db'(r) \qquad (0<t<\infty,\ 0<\tau<\infty).\end{aligned}$$

Observe that $\hat{b}(\tau)$ and $\hat{b}'(\tau)$ are given by expressions analogous to (4.7) and (4.9) respectively so that, in particular

$$(5.2)\qquad \hat{b}'(\tau)=\beta^0(t,\tau)+\beta^\infty(t,\tau) \qquad (0<t<\infty,\ 0<\tau<\infty).$$

LEMMA 5.1. *If* (1.3) *and* (2.1) *hold then* $\hat{c}(\tau)$ *is twice continuously differentiable,* $\partial\beta^0/\partial\tau(t,\tau)$ *exists and is continuous* $(t>0,\ \tau>0)$, *and*

$$(5.3)\qquad |\hat{c}''(\tau)|\leqq 6000\int_0^{1/\tau} r^2 c(r)\,dr \qquad (\tau>0),$$

$$(5.4)\qquad |\beta^\infty(t,\tau)|\leqq 40\tau^{-2}(b(t)-tb'(t)) \qquad (\tau>0,\ t>0),$$

$$(5.5)\qquad \left|\frac{\partial\beta^0}{\partial\tau}(t,\tau)\right|\leqq 500\tau^{-2}\int_0^t b(r)\,dr \qquad (\tau>0,\ t>0),$$

$$(5.6)\qquad \begin{aligned}&\text{(i)}\quad |\beta^0(t,\tau)|\leqq 40\int_0^{1/\tau} rb(r)\,dr \qquad (\tau>0,\, t>0),\\ &\text{(ii)}\quad |\hat{c}'(\tau)|\leqq 40\int_0^{1/\tau} rc(r)\,dr \qquad (\tau>0).\end{aligned}$$

Proof. Three integrations by parts show that

$$(5.7)\qquad \hat{c}(\tau)=-i\tau^{-3}\int_0^\infty \left(1-i\tau r+\frac{(i\tau r)^2}{2}-e^{-i\tau r}\right) dc''(r),$$

where we use

$$rc(r)+r^2|c'(r)|+r^3c''(r-)\to 0 \qquad (r\to 0+)$$

and

$$c(r)+r|c'(r)|+r^2c''(r-)\to 0 \qquad (r\to\infty),$$

which are consequences of (2.1) and which assure that the various boundary terms vanish. Note that $dc''(r)$ is a negative measure; three differentiations of (5.7) yield

$$\hat{c}''(\tau)=i\tau^{-5}\int_0^\infty K(-\tau r)\,dc''(r) \qquad (\tau>0),$$

where

$$K(u)=-12\left(1+iu+\frac{(iu)^2}{2}-e^{iu}\right)+6iu(1+iu-e^{iu})+u^2(1-e^{iu}).$$

The remainder of the proof of (5.3) now follows as in [10, Lemma 6.1(ii)].

To obtain (5.4) we first consider the case where $\tau t\geqq 1$. (4.8) and (5.1) give us

$$\begin{aligned}|\beta^\infty(t,\tau)|&\leqq 2\tau^{-3}\int_t^\infty (\tau r+2)\,db'(r)\leqq 2\tau^{-2}(b(t)-3tb'(t))\\ &\leqq 40\tau^{-2}(b(t)-tb'(t)).\end{aligned}$$

On the other hand if $\tau t<1$ then (4.8) yields

$$\begin{aligned}|\beta^\infty(t,\tau)|&\leqq 6\int_t^{1/\tau} r^3\,db'(r)+2\tau^{-3}\int_{1/\tau}^\infty (\tau r+2)\,db'(r)\\ &=6\left[\tau^{-3}b'\left(\frac{1}{\tau}\right)-t^3b'(t)-3\tau^{-2}b\left(\frac{1}{\tau}\right)+3t^2b(t)+6\int_t^{1/\tau} rb(r)\,dr\right]\\ &\quad+2\tau^{-2}b\left(\frac{1}{\tau}\right)-6\tau^{-3}b'\left(\frac{1}{\tau}\right)\\ &\leqq -6t^3b'(t)+18t^2b(t)+36b(t)\int_t^{1/\tau} r\,dr\\ &\leqq 40\tau^{-2}(b(t)-tb'(t)),\end{aligned}$$

thus giving us (5.4) in both cases.

Differentiation of (5.1) yields

$$\frac{\partial\beta^0}{\partial\tau}(t,\tau)=\tau^{-4}\int_0^t K(-\tau r)\,db'(r) \qquad (\tau>0,\, t>0),$$

where $K(u)=6(1+iu-e^{iu})-4iu(1-e^{iu})+u^2e^{iu}$, and we have

$$|K(u)|\leqq u^4 \qquad (0\leqq u\leqq 1), \quad \text{and} \quad |K(u)|\leqq 20(1+u^2) \qquad (u>0).$$

If $\tau t>1$,

$$\begin{aligned}\left|\frac{\partial\beta^0}{\partial\tau}(t,\tau)\right| &\leqq 40\int_0^{1/\tau} r^4\,db'(r)+20\tau^{-4}\int_{1/\tau}^{t}(\tau^2r^2+1)\,db'(r)\\ &=40\left[\tau^{-4}b'\left(\frac{1}{\tau}\right)-4\tau^{-3}b\left(\frac{1}{\tau}\right)+12\int_0^{1/\tau} r^2b(r)\,dr\right]\\ &\quad+20\tau^{-2}\left[t^2b'(t)-\tau^{-2}b'\left(\frac{1}{\tau}\right)-2tb(t)+2\tau^{-1}b\left(\frac{1}{\tau}\right)+2\int_{1/\tau}^{t} b(r)\,dr\right]\\ &\quad+20\tau^{-4}\left[b'(t)-b'\left(\frac{1}{\tau}\right)\right]\\ &\leqq 40\tau^{-2}\int_{1/\tau}^{t} b(r)\,dr+480\int_0^{1/\tau} r^2b(r)\,dr\\ &\leqq 500\tau^{-2}\int_0^t b(r)\,dr.\end{aligned}$$

If $\tau t<1$ then

$$\left|\frac{\partial\beta^0}{\partial\tau}(t,\tau)\right|\leqq\int_0^t r^4\,db'(r)\leqq 12\int_0^t r^2b(r)\,dr\leqq\frac{12}{\tau^2}\int_0^t b(r)\,dr,$$

so we have established (5.5) in both cases. (5.6) is obtained in the same way as (4.2). This completes the proof of Lemma 5.1.

In order to obtain estimates on the size of u_5 we will need lower bounds on $D(\tau,\lambda)$. We use $\omega=\omega(\lambda)$ as defined in § 4.

LEMMA 5.2. *If* (1.3) *holds, then*

$$|D(\tau,\lambda)|\geqq M\tau\int_0^{1/\tau} ra(r)\,dr \qquad \left(\frac{\rho}{2}\leqq\tau\leqq\frac{\omega}{2}\right), \tag{5.8}$$

$$|D(\tau,\lambda)|\geqq M\frac{|\tau-\omega|}{\lambda} \qquad \left(\tau\geqq\frac{\omega}{2}\right). \tag{5.9}$$

Proof. The estimates (4.3), (4.4), and (4.27) will be exploited throughout without explicit mention.

For $\rho/2\leqq\tau<\omega=\rho$,

$$\operatorname{Re} D(\tau,\lambda)=\varphi(\tau)\geqq M\geqq M_1\tau\int_0^{1/\tau} ra(r)\,dr\geqq M_2\frac{|\tau-\omega|}{\lambda}$$

which establishes (5.8) and (5.9) in this trivial case.

In all other cases we start with

$$-\theta'(\tau) \geqq \frac{\tau}{5}\int_0^{1/\tau} r^3 a(r)\,dr \geqq \frac{1}{80\tau^3} a\left(\frac{1}{\tau}\right), \tag{5.10}$$

and use both parts of this estimate to obtain

$$\begin{aligned}
|\operatorname{Im} D(\tau,\lambda)| &= \tau\left|\frac{1}{\lambda}-\frac{d}{\tau^2}-\theta(\tau)\right| \\
&\geqq \tau\left|\frac{1}{\lambda}-\frac{d}{\omega^2}-\theta(\tau)\right| \\
&\geqq \tau|\theta(\omega)-\theta(\tau)| \\
&= \tau\left|\int_\omega^\tau \theta'(s)\,ds\right| \\
&\geqq \frac{\tau}{10}\left|\int_\omega^\tau s\int_0^{1/s} r^3 a(r)\,dr\,ds\right| + \frac{\tau}{160}\left|\int_\omega^\tau s^{-3} a\left(\frac{1}{s}\right) ds\right| \\
&\geqq \frac{\tau}{10}\left|\int_0^{1/\tau} r^3 a(r)\int_\omega^\tau s\,ds\,dr\right| + \frac{\tau}{160}\left|\int_\omega^\tau s^{-3} a\left(\frac{1}{s}\right) ds\right| \\
&= \frac{\tau|\tau-\omega|(\tau+\omega)}{20}\int_0^{1/\tau} r^3 a(r)\,dr + \frac{\tau}{160}\left|\int_{1/\tau}^{1/\omega} r a(r)\,dr\right|.
\end{aligned} \tag{5.11}$$

In the last step of this computation we used Fubini's theorem on the first integral and a change of variable ($r = 1/s$) on the second. Then

$$\begin{aligned}
\frac{\tau|\tau-\omega|(\tau+\omega)}{20}\int_0^{1/\tau} r^3 a(r)\,dr &\geqq \frac{\tau^2|\tau-\omega|}{20}\left[\int_0^{1/(4\tau)} + \int_{1/(4\tau)}^{1/(2\tau)}\right] r^3 a(r)\,dr \\
&\geqq \frac{|\tau-\omega|}{6000}\tau^{-2}\left(a\left(\frac{1}{4\tau}\right) + a\left(\frac{1}{2\tau}\right)\right).
\end{aligned} \tag{5.12}$$

Define $f(t) = ta(t) - \frac{1}{2}t^2 a'(t)$ $(t>0)$. Observe that $f(t)$ is nonnegative, left continuous (by our convention $a'(t) = a'(t^-)$), and satisfies

$$ta(t) \leqq f(t) \leqq \int_0^t a(r)\,dr \qquad (t>0). \tag{5.13}$$

Here the first inequality is immediate from the definition and the second is obtained by integrating by parts twice in the inequality $\frac{1}{2}\int_0^t r^2\,da'(r) \geqq 0$. From (4.5) we also have

$$f(t) \leqq ta\left(\frac{t}{2}\right) \qquad (t>0), \tag{5.14}$$

so that $f(0+) = 0$ and f is bounded on $[0, 1/\tau]$ for every $\tau > 0$. Let $S(\tau) = \sup_{0<x<1/\tau}\{f(x)\}$, and for each $\tau > 0$ choose $\delta = \delta(\tau) \in (0, 1]$ so that $f(\delta/\tau) \geqq \frac{1}{2}S(\tau)$.

The proof now splits into three cases.

Case 1. If $1 \geqq \delta(\tau) \geqq \frac{1}{2}$ then (5.14) and (5.13) imply

$$\begin{aligned}\frac{|\tau-\omega|}{6000\tau^2}a\left(\frac{1}{4\tau}\right) &\geqq \frac{|\tau-\omega|}{6000\tau}\left[\frac{\delta}{\tau}a\left(\frac{\delta}{2\tau}\right)\right]\\ &\geqq \frac{|\tau-\omega|}{6000\tau}f\left(\frac{\delta}{\tau}\right)\\ &\geqq \frac{|\tau-\omega|}{12000\tau}S(\tau) \qquad (5.15)\\ &\geqq \frac{|\tau-\omega|}{12000}\left(\frac{1}{\tau}\cdot \sup_{0<x<1/\tau}\{xa(x)\}\right)\\ &\geqq \frac{|\tau-\omega|}{12000}\int_0^{1/\tau} ra(r)\,dr.\end{aligned}$$

If $\rho/2 \leqq \tau \leqq \omega/2$ then $|\omega-\tau| \geqq \frac{1}{2}\omega \geqq \tau$ so (5.11), (5.12), and (5.15) combine to give (5.8).

On the other hand, if $\tau \geqq \frac{1}{2}\omega$ then (5.11), (5.12), (5.15) and (4.27) yield

$$\begin{aligned}|\mathrm{Im}\, D(\tau,\lambda)| &\geqq \frac{|\tau-\omega|}{12000}\int_0^{1/\tau} ra(r)\,dr + \frac{\tau}{160}\left|\int_{1/\tau}^{1/\omega} ra(r)\,dr\right|\\ &\geqq \frac{|\tau-\omega|}{12000}\int_0^{1/\omega} ra(r)\,dr \geqq M\frac{|\tau-\omega|}{\lambda},\end{aligned}$$

which is (5.9).

Case 2. If $0<\delta(\tau)<\frac{1}{2}$ and $f(1/\tau) \geqq \frac{1}{2}f(\delta/\tau)$ then, again using (5.14), we obtain

$$\begin{aligned}\frac{|\tau-\omega|}{6000\tau^2}a\left(\frac{1}{2\tau}\right) &\geqq \frac{|\tau-\omega|}{6000\tau}f\left(\frac{1}{\tau}\right)\\ &\geqq \frac{|\tau-\omega|}{12000\tau}f\left(\frac{\delta}{\tau}\right)\\ &\geqq \frac{|\tau-\omega|}{24000}S(\tau) \qquad (5.16)\\ &\geqq \frac{|\tau-\omega|}{24000}\left(\frac{1}{\tau}\sup_{0<x<1/\tau}\{xa(x)\}\right)\\ &\geqq \frac{|\tau-\omega|}{24000}\int_0^{1/\tau} ra(r)\,dr.\end{aligned}$$

Combining (5.11), (5.12), and (5.16) we complete the computations as in Case 1 for both (5.8) and (5.9).

Case 3. If $0<\delta(\tau)<\frac{1}{2}$ and $f(1/\tau) \leqq \frac{1}{2}f(\delta/\tau)$ then we use (4.7) along with the estimate $1-\cos x \geqq x^2/4$ $(0<x<1)$, and then apply (5.13) to obtain

$$
\begin{aligned}
2 \operatorname{Re} D(\tau, \lambda) = 2\varphi(\tau) &= 2\tau^{-2} \int_0^\infty (1-\cos \tau r)\, da'(r) \\
&\geqq \frac{1}{2} \int_0^{1/\tau} r^2\, da'(r) \\
&= \int_0^{1/\tau} a(r)\, dr - f\left(\frac{1}{\tau}\right) \\
&= \tau \int_0^{1/\tau} ra(r)\, dr + \int_0^{1/\tau} (1-\tau r) a(r)\, dr - f\left(\frac{1}{\tau}\right) \qquad (5.17)\\
&\geqq \tau \int_0^{1/\tau} ra(r)\, dr + (1-\delta) \int_0^{\delta/\tau} a(r)\, dr - f\left(\frac{1}{\tau}\right) \\
&\geqq \tau \int_0^{1/\tau} ra(r)\, dr + \frac{1}{2} f\left(\frac{\delta}{\tau}\right) - f\left(\frac{1}{\tau}\right) \\
&\geqq \tau \int_0^{1/\tau} ra(r)\, dr.
\end{aligned}
$$

This is (5.8).

When $\tau \geqq \omega/2$, (5.11), (5.17), and (4.27) combine to yield

$$
\begin{aligned}
\sqrt{2}\, |D(\tau, \lambda)| &\geqq \operatorname{Re} D(\tau, \lambda) + |\operatorname{Im} D(\tau, \lambda)| \\
&\geqq \frac{\tau}{2} \int_0^{1/\tau} ra(r)\, dr + \frac{\tau}{160} \left| \int_{1/\tau}^{1/\omega} ra(r)\, dr \right| \qquad (5.18)\\
&\geqq \frac{|\tau-\omega|}{160} \int_0^{1/\omega} ra(r)\, dr \\
&\geqq M \frac{|\tau-\omega|}{\lambda},
\end{aligned}
$$

which is (5.9). This completes the proof of Lemma 5.2.

6. Proof of estimate (4.37). Define $\Delta(t, \tau) = \beta^0(t, \tau) + \hat{c}'(\tau) + id/\tau^2$ $(\tau>0, t>0)$, and let

$$\mu_{41}(t, \lambda) = \frac{1}{\lambda^3 t} \int_0^\rho e^{i\tau t} \frac{\tau^2(\Delta(t, \tau) + i\lambda^{-1})}{[D(\tau)]^3 D(\tau, \lambda)} \left[\frac{2}{D(\tau)} + \frac{1}{D(\tau, \lambda)}\right] d\tau \qquad (t>0),$$

$$\mu_{42}(t, \lambda) = \frac{1}{\lambda^3 t} \int_0^\rho e^{i\tau t} \frac{\tau^2 \beta^\infty(t, \tau)}{[D(\tau)]^3 D(\tau, \lambda)} \left[\frac{2}{D(\tau)} + \frac{1}{D(\tau, \lambda)}\right] d\tau \qquad (t>0),$$

$$\mu_{51}(t, \lambda) = \frac{1}{\lambda t} \int_\rho^\infty e^{i\tau t} \frac{(\Delta(t, \tau) + i\lambda^{-1})}{[D(\tau, \lambda)]^2} \qquad (t>0),$$

$$\mu_{52}(t, \lambda) = \frac{1}{\lambda t} \int_\rho^\infty e^{i\tau t} \frac{\beta^\infty(t, \tau)}{[D(\tau, \lambda)]^2}\, d\tau \qquad (t>0).$$

Thus $u_j(t,\lambda)=\mu_{j1}(t,\lambda)+\mu_{j2}(t,\lambda)$ $(t>0, 1\leqq\lambda<\infty, j=4,5)$. We now integrate by parts on μ_{41} and μ_{51} in order to bring out another factor of t^{-1}. This gives us

$$\begin{aligned}-i\lambda^3t^2\mu_{41}(t,\lambda)=&-\frac{e^{i\rho t}\rho^2(\Delta(t,\rho)+i\lambda^{-1})}{D^3(\rho)D(\rho,\lambda)}\left[\frac{2}{D(\rho)}+\frac{1}{D(\rho,\lambda)}\right]\\ &+\int_0^\rho e^{i\tau t}\Big[\frac{2\tau(\Delta(t,\tau)+i\lambda^{-1})+\tau^2\Delta_\tau(t,\tau)}{D^3(\tau)D(\tau,\lambda)}\left(\frac{2}{D(\tau)}+\frac{1}{D(\tau,\lambda)}\right)\\ &-\frac{\tau^2(\Delta(t,\tau)+i\lambda^{-1})}{D^3(\tau)D(\tau,\lambda)}\left(\frac{8D'(\tau)}{D^2(\tau)}+\frac{5D'(\tau)+2i\lambda^{-1}}{D(\tau)D(\tau,\lambda)}+\frac{2D_\tau(\tau,\lambda)}{D^2(\tau,\lambda)}\right)\Big]\,d\tau\end{aligned}\tag{6.1}$$

and

$$\begin{aligned}-i\lambda t^2\mu_{51}(t,\lambda)=&\frac{e^{i\rho t}(\Delta(t,\rho)+i\lambda^{-1})}{D^2(\rho,\lambda)}\\ &+\int_\rho^\infty e^{i\tau t}\left[\frac{\Delta_\tau(t,\tau)}{D^2(\tau,\lambda)}-\frac{2(\Delta(t,\tau)+i\lambda^{-1})D_\tau(\tau,\lambda)}{D^3(\tau,\lambda)}\right]d\tau.\end{aligned}\tag{6.2}$$

In (6.1) the estimates of Lemmas 4.1 and 5.1 assure absolute convergence of the integral and vanishing boundary terms at $\tau=0$. In (6.2) absolute convergence and vanishing boundary terms at $\tau=\infty$ is a consequence of Lemmas 5.1 and 5.2 along with (2.3). Inequalities (4.1) and (2.3(i)) imply that

$$\inf_{0<\tau<\rho}\ \inf_{1\leqq\lambda\leqq\infty}|D(\tau,\lambda)|\equiv\gamma>0.$$

From (4.2) and Lemma 5.1 we find the estimates

$$\begin{aligned}&\max\{\tau|D_\tau(\tau,\lambda)|,\tau|\Delta(t,\tau)+i\lambda^{-1}|\}\leqq 40\tau\int_0^{1/\tau}ra(r)\,dr+d\tau^{-1}+\tau\lambda^{-1}, &&(t,\tau>0)\\ &\tau^2|\Delta_\tau(t,\tau)|\leqq 500\int_0^t b(r)\,dr+6000\tau^2\int_0^{1/\tau}r^2c(r)\,dr+2\,d\tau^{-1} &&(t,\tau>0)\end{aligned}\tag{6.3}$$

and from (4.28) and (4.3) we have $|D(\tau)|\geqq|\operatorname{Im}D(\tau)|\geqq d\tau^{-1}$. When $d=0$ we have a lower bound on $|D(\tau)|$ from (4.1). Thus, referring to (6.1), we obtain

$$\begin{aligned}|\lambda^3t^2\mu_{41}(t,\lambda)|&\leqq M+M\int_0^\rho\Big\{\frac{\tau\int_0^{1/\tau}ra(r)\,dr+\tau+d\tau^{-1}+\int_0^t b(r)\,dr+\tau^2\int_0^{1/\tau}r^2c(r)\,dr}{\max\{d\tau^{-1},\int_0^{1/\tau}a(r)\,dr\}\cdot\gamma^4}\\ &\qquad+\left(\frac{\tau\int_0^{1/\tau}ra(r)\,dr+\tau+d\tau^{-1}}{\max\{d\tau^{-1},\int_0^{1/\tau}a(r)\,dr\}}\right)^2\gamma^{-4}\Big\}\,d\tau\\ &\leqq M\left(\int_0^t b(r)\,dr+1\right);\end{aligned}$$

in other words

$$|\mu_{41}(t,\lambda)|\leqq Mq(t).\tag{6.4}$$

From (5.4) it is clear that

$$|\lambda^3 t\mu_{42}(t,\lambda)| \leqq M \int_0^\rho \frac{[b(t)-tb'(t)]}{\gamma^5}\,d\tau \leqq \frac{M\rho}{\gamma^5}(b(t)-tb'(t)) \leqq \frac{M\rho}{\gamma^5} tq(t).$$

Taken together with (6.4), this shows that

$$|u_4(t,\lambda)| \leqq Mq(t) \tag{6.5}$$

which is the case $j=4$ of (4.37).

In order to obtain a similar estimate on $u_5(t,\lambda)$ we partition the interval $[\rho/2,\infty)$ into four sets

$$\begin{aligned}\left[\frac{\rho}{2},\infty\right) &= \left[\frac{\rho}{2},\frac{\omega}{2}\right) \cup \left[\frac{\omega}{2},\omega-\frac{\rho}{2}\right) \cup \left[\omega-\frac{\rho}{2},\omega+\frac{\rho}{2}\right) \cup \left[\omega+\frac{\rho}{2},\infty\right)\\ &= E_1 \cup E_2 \cup E_3 \cup E_4.\end{aligned}$$

We use the estimates of Lemma 5.2 on $E_1 \cup E_2 \cup E_4$ and (2.3) on E_3 for lower bounds on $|D(\tau,\lambda)|$. Lemmas 4.1 and 5.1 will again give upper bounds on the numerators.

We know from (6.3) that

$$\begin{aligned}|\Delta_\tau(t,\tau)| &\leqq M\left(\tau^{-2}\int_0^t b(r)\,dr + \int_0^{1/\tau} r^2 c(r)\,dr + d\tau^{-3}\right)\\ &\leqq M_1 t^2 q(t)\tau^{-2} \qquad \left(\tau > \frac{\rho}{2}\right),\end{aligned} \tag{6.6}$$

and

$$\max\{|\Delta(t,\tau)+i\lambda^{-1}|, |D_\tau(\tau,\lambda)|\} \leqq M\int_0^{1/\tau} ra(r)\,dr \qquad \left(\tau \geqq \frac{\rho}{2}\right). \tag{6.7}$$

Using the estimate $\int_{1/\omega}^{2/\omega} ra(r)\,dr \leqq \frac{3}{2}\omega^{-2}a(\omega^{-1}) \leqq 3\int_0^{1/\omega} ra(r)\,dr$, along with (4.27) and (6.7), we obtain

$$\max\{|\Delta(t,\tau)+i\lambda^{-1}|, |D_\tau(\tau,\lambda)|\} \leqq M\int_0^{2/\omega} ra(r)\,dr \leqq 20M\lambda^{-1} \qquad \left(\tau \geqq \frac{\omega}{2}\right). \tag{6.8}$$

Returning to (6.2), we observe that

$$|t^2\mu_{51}(t,\lambda)| \leqq M + \int_\rho^\infty \left\{\frac{|\Delta_\tau(t,\tau)|}{\lambda|D(\tau,\lambda)|^2} + \frac{2|\Delta(t,\tau)+i\lambda^{-1}|\,|D_\tau(\tau,\lambda)|}{\lambda|D(\tau,\lambda)|^3}\right\} d\tau. \tag{6.9}$$

From (6.6), (6.7), (5.8), and (4.27) it follows that

$$\begin{aligned}&\int_{E_1}\left\{\frac{|\Delta_\tau(t,\tau)|}{\lambda|D(t,\lambda)|^2} + \frac{2|\Delta(t,\tau)+i\lambda^{-1}|\,|D_\tau(\tau,\lambda)|}{\lambda|D(\tau,\lambda)|^3}\right\} d\tau\\ &\leqq Mt^2q(t)\left[\frac{1}{\lambda\int_0^{2/\omega} ra(r)\,dr}\cdot\int_{\rho/2}^{\omega/2}\frac{d\tau}{\tau^4\int_0^{1/\tau} ra(r)\,dr} + \frac{1}{\lambda\int_0^{2/\omega} ra(r)\,dr}\int_{\rho/2}^{\omega/2}\frac{d\tau}{\tau^3}\right]\\ &\leqq M_1 t^2 q(t)\int_{\rho/2}^\infty\left[\frac{1}{\tau^2}+\frac{1}{\tau^3}\right]d\tau \leqq M_2 t^2 q(t).\end{aligned} \tag{6.10}$$

From (6.6), (6.8), (5.9), (4.24), and (4.27) we find that

$$\begin{aligned}
&\int_{E_2\cup E_4}\left\{\frac{|\Delta_\tau(t,\tau)|}{\lambda|D(\tau,\lambda)|^2}+\frac{2|\Delta(t,\tau)+i\lambda^{-1}|\cdot|D_\tau(\tau,\lambda)|}{\lambda|D(\tau,\lambda)|^3}\right\}d\tau\\
&\leqq Mt^2q(t)\left[\int_{\omega/2}^{\omega-\rho/2}+\int_{\omega+\rho/2}^{\infty}\frac{1}{(\tau-\omega)^2}+\frac{1}{(\tau-\omega)^3}\,d\tau\right] \qquad (6.11)\\
&\leqq M_1t^2q(t).
\end{aligned}$$

Then (6.6), (6.8), (4.3), (4.27) and (2.3) yield

$$\begin{aligned}
&\int_{E_3}\left\{\frac{|\Delta_\tau(t,\tau)|}{\lambda|D(\tau,\lambda)|^2}+\frac{2|\Delta(t,\tau)+i\lambda^{-1}|\cdot|D_\tau(\tau,\lambda)|}{\lambda|D(\tau,\lambda)|^3}\right\}d\tau\\
&\leqq Mt^2q(t)\int_{\omega-\rho/2}^{\omega+\rho/2}\left\{\frac{\theta(\omega)\theta(\tau)}{\varphi^2(\tau)}+\frac{\theta^3(\omega)}{\varphi^3(\tau)}\right\}d\tau \qquad (6.12)\\
&\leqq M_1t^2q(t).
\end{aligned}$$

Now (6.9) through (6.12) imply that

$$|\mu_{51}(t,\lambda)|\leqq Mq(t). \qquad (6.13)$$

It remains to establish an estimate on

$$|t\mu_{52}(t,\lambda)|\leqq\int_\rho^\infty\frac{|\beta^\infty(t,\tau)|}{\lambda|D(\tau,\lambda)|^2}\,d\tau. \qquad (6.14)$$

From (5.4) and (5.8) we argue as in (6.10) to obtain

$$\int_{E_1}\frac{|\beta^\infty(t,\tau)|}{\lambda|D(\tau,\lambda)|^2}\,d\tau\leqq\frac{Mtq(t)}{\lambda\int_0^{2/\omega}ra(r)\,dr}\int_{\rho/2}^{\omega/2}\frac{d\tau}{\tau^2}\leqq M_1tq(t). \qquad (6.15)$$

As in (6.11), (5.4) and (5.9) yield

$$\int_{E_2\cup E_4}\frac{|\beta^\infty(t,\tau)|}{\lambda|D(\tau,\lambda)|^2}\,d\tau\leqq Mtq(t)\left[\int_{\omega/2}^{\omega-\rho/2}+\int_{\omega+\rho/2}^{\infty}\right]\frac{d\tau}{(\tau-\omega)^2}\leqq M_1tq(t). \qquad (6.16)$$

Finally, (5.4) and (2.3) give us

$$\int_{E_3}\frac{|\beta^\infty(t,\tau)|}{\lambda|D(\tau,\lambda)|^2}\,d\tau\leqq Mtq(t)\int_{\omega-\rho/2}^{\omega+\rho/2}\frac{\theta(\tau)\theta(\omega)}{\varphi^2(\tau)}\,d\tau\leqq M_1tq(t). \qquad (6.17)$$

Combining (6.13) through (6.17) we have now established that

$$|u_5(t,\lambda)|\leqq Mq(t).$$

This completes the case $j=5$ of (4.37).

In order to prove (4.38) one need only apply the estimates of Lemmas 4.1 and 5.2 along with (4.27) and (4.33) to the functions as defined in (4.35) in the same manner as we have done in this section, noting that in this case the decomposition $a(t)=b(t)+c(t)$ from (2.1) is never used.

7. Proof of Corollary 2.1 and an example. In view of Theorem 2.2 it suffices to show that either hypothesis (i) or hypothesis (ii) of the corollary implies (2.3).

If $c \equiv 0$ and $a(0+) = b(0+) < \infty$ and $a(t)$ is strongly positive then there exist constants $\eta > 0$ and $\mu > 0$ such that

$$\varphi(\tau) \geqq \frac{\eta}{1+\tau^2} \geqq \mu \frac{a(0)}{2\tau^2} \geqq \mu \int_0^{1/\tau} ta(t)\,dt \geqq \frac{\mu}{12}\theta(\tau) > 0 \qquad (\tau \geqq \rho),$$

and

$$\varphi(\tau) \geqq \frac{\eta}{1+\rho^2} > 0 \qquad (0 < \tau \leqq \rho).$$

Here we have used (4.3). This establishes (2.3) when (i) holds.

Assume now that hypothesis (ii) holds. If $a(0+) < \infty$ then [18, Cor. 2.1 and 2.2] imply that c is strongly positive, and hypothesis (iii) assures that, for sufficiently small $x_0 > 0$, there exists $\beta > 0$ such that

$$\frac{\int_0^x a(t)\,dt}{\int_0^x c(t)\,dt} \leqq \beta \qquad (0 < x \leqq x_0),$$

and that there exists $\nu > 0$ such that

$$\operatorname{Re} \hat{c}(\tau) \geqq \nu\tau^{-2} \qquad (\tau \geqq \rho).$$

Therefore

$$\begin{aligned}\varphi(\tau) &\geqq \operatorname{Re} \hat{c}(\tau) \\ &\geqq \nu\tau^{-2} \\ &\geqq \frac{\nu}{c(0)\tau} \int_0^{1/\tau} c(t)\,dt \\ &\geqq \frac{\nu}{\beta c(0)\tau} \int_0^{1/\tau} a(t)\,dt \\ &\geqq \frac{\nu}{4\beta c(0)} \frac{|\hat{a}(\tau)|}{\tau} \\ &\geqq M\theta(\tau) > 0 \qquad (\tau \geqq \max\{\rho, x_0^{-1}\}).\end{aligned}$$

Here we have used (4.1). The condition (2.3(i)) is satisfied, because a is strongly positive. This establishes (2.3) when (ii) holds and $a(0+) < \infty$.

Assume now that (ii) holds and $a(0+) = \infty$. Then $c(0+) = \infty$, and we define

$$c_1(t) = \begin{cases} \frac{1}{2}(t-t_1)^2 c''(t_1) + (t-t_1)c'(t_1) + c(t_1) & (0 < t \leqq t_1) \\ c(t) & (t > t_1) \end{cases}$$

and $c_2(t) = c(t) - c_1(t)$, $(t > 0)$. Then c_1 and c_2 both satisfy (H), $-c_2'(t)$ is convex $c_2 \in L^1(R^+)$, $c_1(0+) < \infty$, and hence $c_2(0+) = \infty$.

By a result of O. J. Staffans [23, Thm. 2(iii)]

$$\alpha = \inf_{\tau > 0} \left\{ \frac{\operatorname{Re} \hat{c}_2(\tau)}{|\hat{c}_2(\tau)|^2} \right\} > 0.$$

Furthermore $c_1(0+) < \infty$ and hypothesis (ii) imply that for some $x_0 > 0$, $\beta > 0$,

$$\frac{\int_0^x a(t)\,dt}{\int_0^x c_2(t)\,dt} < \beta < \infty \qquad (0 < x \leqq x_0).$$

Thus

$$\begin{aligned}\varphi(\tau) &\geqq \operatorname{Re} \hat{c}_2(\tau)\\ &\geqq \alpha|\hat{c}_2(\tau)|^2\\ &\geqq \frac{\alpha}{8}\left(\int_0^{1/\tau} c_2(t)\,dt\right)^2\\ &\geqq \frac{\alpha}{8\beta^2}\left(\int_0^{1/\tau} a(t)\,dt\right)^2\\ &\geqq \frac{\alpha}{8\beta^2}\,\frac{a(t_1)}{\tau}\int_0^{1/\tau} a(t)\,dt\\ &\geqq \frac{\alpha a(t_1)}{8\beta^2}\,\frac{|\hat{a}(\tau)|}{\tau}\\ &\geqq M\theta(\tau)>0 \qquad (\tau \geqq \max\{\rho, x_0^{-1}\}).\end{aligned}$$

Again (2.3(i)) is trivial, and we have established (2.3) when $a(0+)=\infty$. This completes the proof of Corollary 2.1.

We conclude with an example where (2.3(i)) holds and the kernel is even strongly positive but (2.3(ii)) is not satisfied.

Let $b_k(t)=(1-2^{2^k}t)\chi_{[0,2^{-2^k}]}(t)$ $(t \geqq 0,\ k=0,1,2,3,\cdots)$ where χ_E denotes the characteristic function of the set E. Define

$$a(t)=\sum_{k=0}^{\infty} b_k(t) \qquad (t>0).$$

This sum is finite for each $t>0$ and one easily checks that $a(t)$ satisfies (1.3) with $a(0+)=\infty$. A direct computation shows that $a \in L^1(0,\infty)$, and

$$\varphi(\tau)=\sum_{k=0}^{\infty} 2^{2^k}\frac{(1-\cos 2^{-2^k}\tau)}{\tau^2} \qquad (\tau>0). \tag{7.1}$$

We first show that $a(t)$ is strongly positive. Note that

$$\tfrac{1}{2}u^2 \geqq 1-\cos u \geqq \tfrac{1}{4}u^2 \qquad (0 \leqq u \leqq 1), \tag{7.2}$$

and let $m=[\log_2 \log_2 \tau]$ for $\tau \geqq 2$, where $[\,\cdot\,]$ denotes the greatest integer function. Thus $2^{2^m} \leqq \tau \leqq 2^{2^{m+1}}$. (7.1) and (7.2) imply that

$$\varphi(\tau) \geqq \frac{1}{4}\sum_{k=m+1}^{\infty} 2^{-2^k} \geqq \frac{1}{4}2^{-2^{m+1}}=\frac{1}{4}\left[\frac{1}{2^{2^m}}\right]^2 \geqq \frac{1}{4\tau^2} \qquad (\tau \geqq 2). \tag{7.3}$$

For $0<\tau \leqq 2$ we use (4.7) and (7.2) to obtain

$$\varphi(\tau)=\tau^{-2}\int_0^{\infty}(1-\cos \tau t)\,da'(t) \geqq \frac{1}{4}\int_0^{1/\tau} t^2\,da'(t) \geqq K>0, \tag{7.4}$$

where $K=\frac{1}{4}\int_0^{1/2} t^2\,da'(t)$ is a fixed positive constant.

(7.3) and (7.4) show that $a(t)$ is strongly positive, even though $da'(t)$ is a purely singular measure. (Compare [18, § 4].)

We next show that $a(t)$ does not satisfy (2.3(ii)). Let $\tau_n = 2^{2^n}(2\pi)$ $(n=0,1,2,3,\cdots)$.

Then, referring to (7.1), (7.2) and (7.3), and using the fact that $\operatorname{Re}\hat{b}_k(\tau_n) = 0$ $(k = 0, 1, 2, \cdots, n)$, we find that

$$\varphi(\tau_n) \leqq \frac{1}{2} \sum_{k=n+1}^{\infty} \frac{2^{2^k}(2^{-2^k}\tau_n)^2}{\tau_n^2} = \frac{1}{2} \sum_{k=n+1}^{\infty} 2^{-2^k} \leqq 2^{-2^{n+1}} = \frac{(2\pi)^2}{\tau_n^2}. \tag{7.5}$$

From (4.3) we have

$$\theta(\tau_n) \geqq \frac{1}{5} \int_0^{1/\tau} ta(t)\,dt \geqq \frac{1}{5} \sum_{k=1}^{n} \int_0^{1/\tau_n} t(1-\tau_n t)\,dt = \frac{n}{30\tau_n^2}. \tag{7.6}$$

Comparison of (7.5) and (7.6) shows that

$$\frac{\theta(\tau_n)}{\varphi(\tau_n)} \geqq \frac{n}{120\pi^2} \to \infty \qquad (n \to \infty, \tau_n \to \infty);$$

thus (2.3(ii)) fails to hold. We note that $da'(t)$ being a purely singular measure was not the critical aspect of this counterexample. One could use $\alpha(t) = a(t) + e^{-t}$, where $a(t)$ is as defined above, and observe the same phenomenon.

Acknowledgment. This paper is based, in part, on the first author's Ph.D. thesis, written at the University of Wisconsin under the supervision of Professor John A. Nohel. His help in the preparation of this paper is gratefully acknowledged.

REFERENCES

[1] M. G. CRANDALL, S.-O. LONDEN AND J. A. NOHEL, *An abstract nonlinear Volterra integrodifferential equation*, MRC Tech. Sum. Rep. #1684, Univ. of Wisconsin Mathematics Research Center, Madison, Wisconsin, 1976.

[2] C. M. DAFERMOS, *An abstract Volterra equation with applications to linear viscoelasticity*, J. Differential Equations, 7 (1970), pp. 554–569.

[3] A. FRIEDMAN AND M. SHINBROT, *Volterra integral equations in Banach space*, Trans. Amer. Math. Soc., 126 (1967), pp. 131–179.

[4] K. B. HANNSGEN, *Indirect Abelian theorems and a linear Volterra equation*, Ibid., 142 (1969), pp. 539–555.

[5] ———, *A Volterra equation with completely monotonic convolution kernel*, J. Math. Anal. Appl., 31 (1970), pp. 459–471.

[6] ———, *A Volterra equation with parameter*, this Journal, 4 (1973), pp. 22–30.

[7] ———, *A Volterra equation in Hilbert space*, this Journal, 5 (1974), pp. 412–416.

[8] ———, *A linear Volterra equation in Hilbert space*, this Journal, 5 (1974), pp. 927–940.

[9] ———, *The resolvent kernel of an integrodifferential equation in Hilbert space*, this Journal, 7 (1976), pp. 481–490.

[10] ———, *Uniform L^1 behavior for an integrodifferential equation with parameter*, this Journal, 8 (1977), pp. 626–639.

[11] E. HILLE AND R. S. PHILLIPS, *Functional Analysis and Semi-Groups*, American Mathematical Society, Providence, RI, 1957.

[12] J. J. LEVIN, *The asymptotic behavior of the solution of a Volterra equation*, Proc. Amer. Math. Soc., 14 (1963), pp. 534–541.

[13] J. J. LEVIN AND J. A. NOHEL, *On a system of integrodifferential equations occurring in reactor dynamics*, J. Math. Mech., 9 (1960), pp. 347–368.

[14] ———, *On a system of integrodifferential equations occurring in reactor dynamics*, II, Arch. Rat. Mech. Anal., 11 (1962), pp. 210–243.

[15] S.-O. LONDEN, *An existence result on a Volterra equation in Banach space*, Trans. Amer. Math. Soc., to appear.

[16] R. K. MILLER, *Volterra integral equations in a Banach space*, Funkcial. Ekvac., 18 (1975), pp. 163–194.

[17] R. K. MILLER AND R. L. WHEELER, *Asymptotic behavior for a linear Volterra integral equation in Hilbert space*, J. Differential Equations, 23 (1977), pp. 270–284.

[18] J. A. NOHEL AND D. A. SHEA, *Frequency domain methods for Volterra equations*, Advances in Math., 22 (1976), pp. 278–304.

[19] F. RIESZ AND B. SZ.-NAGY, *Functional Analysis*, Frederick Ungar, New York, 1955.
[20] D. F. SHEA AND S. WAINGER, *Variants of the Wiener–Lévy theorem, with applications to stability problems for some Volterra integral equations*, Amer. J. Math., 97 (1975), pp. 312–343.
[21] M. SLEMROD, *A hereditary partial differential equation with applications in the theory of simple fluids*, Arch. Rat. Mech. Anal., 62 (1976), pp. 303–321.
[22] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1946.
[23] O. J. STAFFANS, *An inequality for positive definite Volterra kernels*, Proc. Amer. Math. Soc., 58 (1976), pp. 205–210.
[24] J. S. W. WONG AND R. WONG, *Asymptotic solutions of linear Volterra integral equations with singular kernels*, Trans. Amer. Math. Soc., 189 (1974), pp. 185–200.
[25] R. C. MACCAMY, *An integro-differential equation with application in heat flow*, Quart. Appl. Math., 35 (1977), pp. 1–19.
[26] ———, *A model for one-dimensional, nonlinear viscoelasticity*, Ibid., 35 (1977), pp. 21–33.
[27] C. M. DAFERMOS AND J. A. NOHEL, *Energy methods for nonlinear hyperbolic Volterra integrodifferential equations*, to appear.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0009 $01.00/0

STABILITY CONDITIONS FOR SECOND ORDER ORDINARY DIFFERENTIAL EQUATIONS WITH PERIODIC COEFFIENTS*

EARL R. BARNES†

Abstract. We give several stability tests for canonical second order ordinary differential equations with periodic coefficients. The tests assume a knowledge of the integrals of coefficients in the differential equations and are the best possible when no further information is available.

1. Introduction. Consider the canonical linear second order ordinary differential equation

$$\begin{aligned} \dot{y}_1 &= -\beta(t)y_1 - \gamma(t)y_2, \\ \dot{y}_2 &= \alpha(t)y_1 + \beta(t)y_2 \end{aligned} \tag{1.1}$$

where $\alpha(t)$, $\beta(t)$, $\gamma(t)$ are periodic of period $T>0$, and Lebesgue integrable on $[0, T]$. Hill's equation

$$y''+p(t)y=0, \qquad p(t+T)=p(t), \tag{1.2}$$

which arises in numerous applications, is included as a special case by taking $y_1=y'$, $y_2=y$. A classical result due to Lyapunov states that all solutions of (1.2) are bounded on $(-\infty, \infty)$ if

$$p(t)\begin{cases} \geqq 0 \\ \not\equiv 0. \end{cases} \quad \text{and} \quad T\int_0^T p(t)\,dt<4, \tag{1.3}$$

This result was generalized by Krein in [1]. He showed that all solutions of (1.2) are bounded if for some integer $n \geqq 1$,

$$p(t)\geqq\frac{n^2\pi^2}{T^2} \quad \text{and} \quad T\int_0^T p(t)\,dt<n^2\pi^2+2\pi n(n+1)\tan\frac{\pi}{2(n+1)}. \tag{1.4}$$

In this paper we derive similar results for equation (1.1). For example, we show that all solutions of (1.1) are bounded on $(-\infty, \infty)$ if for some integer $n \geqq 0$,

$$\alpha(t)\geqq\frac{n\pi}{T}, \quad \gamma(t)\geqq\frac{n\pi}{T}, \quad \left(\alpha(t)-\frac{n\pi}{T}\right)\left(\gamma(t)-\frac{n\pi}{T}\right)-\beta^2(t)\geqq 0, \tag{1.5a}$$

and

$$\begin{aligned} \int_0^T \{\alpha(t)+|\beta(t)|+\gamma(t)\}\,dt &< 2n\pi+2(n+1)\log\left[\frac{1+\cos\dfrac{n\pi}{4(n+1)}}{1+\sin\dfrac{n\pi}{4(n+1)}}\right] \\ &\quad +4(n+1)\left[\frac{\cos\dfrac{n\pi}{4(n+1)}-\sin\dfrac{n\pi}{4(n+1)}}{1+\sin\dfrac{n\pi}{4(n+1)}+\cos\dfrac{n\pi}{4(n+1)}}\right]. \end{aligned} \tag{1.5b}$$

* Received by the editors December 22, 1977.

† Mathematical Sciences Department, IBM T. J. Watson Research Center, Yorktown Heights, New York 10598. This work was supported in part by a Vinton Hayes Fellowship from the Department of Electrical Engineering, Massachusetts Institute of Technology.

The tests (1.3) and (1.4) are derived in [2] by variational methods. These methods also apply to the more general equation (1.1). However, the variational problems that arise are, in this case, considerably more difficult to solve. We give the solutions in § 4. In § 5 we list the stability tests that are implied by these solutions and work out an example. In §§ 2 and 3 we review some preliminaries from [2, Chap. 8] that are necessary for the variational formulation of stability conditions.

2. Preliminaries. For convenience we write (1.1) in vector notation as

$$J\dot{y} = H(t)y \tag{2.1}$$

where J and $H(t)$ are the 2×2 matrices

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad H(t) = \begin{pmatrix} \alpha(t) & \beta(t) \\ \beta(t) & \gamma(t) \end{pmatrix}. \tag{2.2}$$

$H(t)$ is called a Hamiltonian matrix and the differential equation (1.1) is called a canonical Hamiltonian system. The totality of Hamiltonians, under ordinary matrix algebra, forms a vector space which we denote by $\mathscr{L}^3$ and term Hamiltonian space. We define a norm on $\mathscr{L}^3$ by

$$\|H(t)\| = \int_0^T \{|\alpha(t)| + |\beta(t)| + |\gamma(t)|\}\, dt. \tag{2.3}$$

A 3-dimensional model of $\mathscr{L}^3$ is constructed in [2, Chap. 8]. It consists of open connected sets $\mathcal{O}_n$, $\mathscr{H}_n$, $n = 0, \pm 1, \pm 2, \cdots$, and closed connected sets π_n^{*-}, π_n^{*+}, π_n^{**}, $n = 0, \pm 1, \pm 2, \cdots$, and may be obtained by rotating Fig. 2.1 about a line through the points π_n^{**}, $n = 0, \pm 1, \pm 2, \cdots$. We denote this model by R^3 and write

$$R^3 = \left(\bigcup_{n=-\infty}^{\infty} \mathcal{O}_n\right) \cup \left(\bigcup_{n=-\infty}^{\infty} \mathscr{H}_n\right) \cup \left(\bigcup_{n=-\infty}^{\infty} \pi_n^{*-}\right) \cup \left(\bigcup_{n=-\infty}^{\infty} \pi_n^{*+}\right) \cup \left(\bigcup_{n=-\infty}^{\infty} \pi_n^{**}\right).$$

Each Hamiltonian $H(t)$ is associated with a point in exactly one of the sets in R^3. This association defines a many-to-one continuous mapping. For simplicity we shall say that a given $H(t)$ belongs to the set with which it is associated under this mapping.

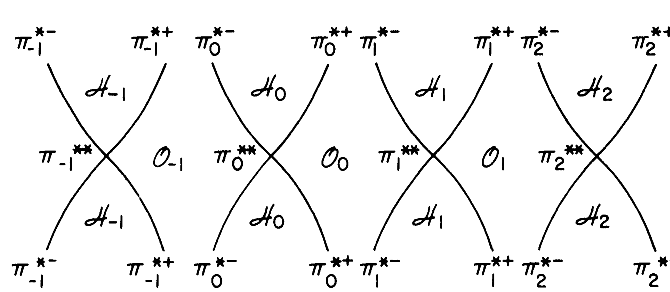

FIG. 2.1. R^3.

The sets $\mathcal{O}_n$ and $\mathscr{H}_n$ represent zones of stability and instability for systems of the form (1.1).

If $H(t) \in \mathcal{O}_n$ for some n, then all solutions of (1.1) are bounded on the entire line $(-\infty, \infty)$.

If $H(t) \in \mathscr{H}_n$ for some n, then all solutions of (1.1) grow unboundedly as $t \to +\infty$ or as $t \to -\infty$.

If $H(t) \in \pi_n^{*-} \cup \pi_n^{*+}$ for some n, then (1.1) has exactly one periodic solution $(y(t+T)=y(t))$ of period T if n is even, and exactly one antiperiodic solution $(y(t+T)=-y(t))$ of period T if n is odd.

If $H(t) \in \pi_n^{**}$, then all solutions of (1.1) are periodic of period T if n is even, and all solutions are antiperiodic of period T if n is odd.

R^3 is a faithful representation of $\mathscr{L}^3$ in the sense that the preimage of each $\mathscr{O}_n$ in $\mathscr{L}^3$ is open, connected and simply connected and its boundary is the preimage of the sets $\pi_n^{*+} \cup \pi_n^{**}$ and $\pi_{n+1}^{*-} \cup \pi_{n+1}^{**}$. These preimages are connected and simply connected, as are the sets $\pi_n^{*+} \cup \pi_n^{**}$ and $\pi_{n+1}^{*-} \cup \pi_{n+1}^{**}$. The preimage of each $\mathscr{H}_n$ is connected, but not simply connected, and its boundary is formed by the preimages of the sets $\pi_n^{*-} \cup \pi_n^{**}$, $\pi_n^{*+} \cup \pi_n^{**}$. The fact that R^3 is a faithful representation of $\mathscr{L}^3$ will be of particular importance in § 3 where we shall state, without proofs, certain results which are obviously true in R^3 but require proofs in $\mathscr{L}^3$. The proofs for $\mathscr{L}^3$ are given in [2].

DEFINITION. We define the rotation ϕ_y of a solution $y(t)$ of (1.1) to be the angle through which $y(t)$ turns as t goes from 0 to T.

A brief calculation shows that

$$\phi_y = \int_0^T \frac{H(t)y(t) \cdot y(t)}{y(t) \cdot y(t)}\, dt. \tag{2.4}$$

The following characterization of the sets $\mathscr{O}_n$, $\mathscr{H}_n$, π_n^{*-}, π_n^{*+}, π_n^{**} in terms of rotations is given in [2, p. 654].

THEOREM 2.1. *$H(t) \in \mathscr{O}_n$ if and only if the rotation of any solution $y(t) \not\equiv 0$ of* (1.1) *satisfies $n\pi < \phi_y < (n+1)\pi$.*

$H(t) \in \mathscr{H}_n$ if and only if there exist two linearly independent solutions y^1 and y^2 of (1.1) *with rotations $\phi_{y^1} = \phi_{y^2} = n\pi$ and there are also solutions with rotations both less than and greater than $n\pi$. In this case we have $(n-1)\pi < \phi_y < (n+1)\pi$ for any solution $y(t) \not\equiv 0$.*

$H(t) \in \pi_n^{*-}$ if and only if $(n-1)\pi < \phi_y \leqq n\pi$ for any solution $y(t) \not\equiv 0$ and there is exactly one linearly independent solution $y(t)$ such that $\phi_y = n\pi$.

$H(t) \in \pi_n^{*+}$ if and only if for any solution $y(t) \not\equiv 0$ we have $n\pi \leqq \phi_y < (n+1)\pi$ and there is exactly one linearly independent solution y such *that $\phi_y = n\pi$.*

$H(t) \in \pi_n^{**}$ if and only if $\phi_y = n\pi$ for any solution $y(t) \not\equiv 0$ of (1.1).

This theorem remains valid if the interval $[0, T]$ is replaced by any interval $[t_0, t_0+T]$ in the definition of the rotation. Thus if $H(t) \in \pi_n^{*\pm}$, $n \neq 0$, a given solution of (2.1) may always be assumed to satisfy $y_1(0) = 0$.

Sufficient conditions for a constant Hamiltonian to belong to each of the sets π_n^{**}, $\mathscr{O}_n$, $n = 0, \pm 1, \pm 2, \cdots$, are given in [2, p. 658]. According to these results, the constant Hamiltonian

$$H_{c_n} = \begin{pmatrix} c_n & 0 \\ 0 & c_n \end{pmatrix} \tag{2.5}$$

belongs to $\mathscr{O}_n$ if

$$\frac{n\pi}{T} < c_n < \frac{(n+1)\pi}{T}, \tag{2.6}$$

and to π_n^{**} if $c_n = n\pi/T$, $n = 0, \pm 1, \pm 2, \cdots$. This follows immediately from (2.4) and Theorem 2.1.

3. Stability tests. In this section we give several sufficient conditions for all solutions of (1.1) to be bounded. When one of these conditions holds, it also identifies the stability zone $\mathcal{O}_n$ in which $H(t)$ lies. The conditions are based on results which are proved in [2, Chap. 8] and stated here without proofs. However, if one is willing to accept Fig. 2.1 as a faithful representation of $\mathcal{L}^3$, no proofs are required.

Let

$$\pi_n^+ = \pi_n^{*+} \cup \pi_n^{**}, \qquad \pi_{n+1}^- = \pi_{n+1}^{*-} \cup \pi_{n+1}^{**}.$$

These sets are simply connected and closed. Their union is the boundary of the stability zone $\mathcal{O}_n$. Let

$$\rho_n^+ = \inf_{H(t)\in\pi_n^+} \|H(t) - H_{c_n}\|, \tag{3.1}$$

$$\rho_{n+1}^- = \inf_{H(t)\in\pi_{n+1}^-} \|H(t) - H_{c_n}\|, \tag{3.2}$$

where c_n satisfies (2.6). Strictly speaking, we should write $\rho_n^+(c_n)$ and $\rho_{n+1}^-(c_n)$ to indicate the dependence of ρ_n^+ and ρ_{n+1}^- on c_n. But we shall not do this. In § 5, c_n will be taken to be an endpoint, or the midpoint, of the interval (2.6).

THEOREM 3.1. *Let $H(t)$ be a Hamiltonian satisfying*

$$\|H(t) - H_{c_n}\| < \min\{\rho_n^+, \rho_{n+1}^-\}. \tag{3.3}$$

Then $H(t) \in \mathcal{O}_n$ and consequently all solutions of (1.1) *are bounded.*

The proof of this theorem is trivial if one accepts Fig. 2.1 as an accurate description of $\mathcal{L}^3$. Inequality (3.3) says that $H(t)$ lies in an open sphere centered at $H_{c_n} \in \mathcal{O}_n$ and contained in the closure of $\mathcal{O}_n$. It follows that $H(t) \in \mathcal{O}_n$.

In the next section we show how to compute the numbers ρ_n^+, ρ_{n+1}^-, $n = 0, \pm 1, \pm 2, \cdots$. Inequality (3.3) then becomes a practical test for stability. In the remainder of this section we explain how tests for stability can be made in terms of just one of the numbers ρ_n^+, ρ_{n+1}^-. Tests (1.3), (1.4), (1.5) are of this type. Their derivations require a well-known comparison theorem for linear second order ordinary differential equations.

Recall that a square matrix H is said to be nonnegative definite if the inner product $Hy \cdot y$ is $\geqq 0$ for each vector y for which Hy is defined. We indicate that H is nonnegative definite by writing $H \geqq 0$. If in this definition the inequality $Hy \cdot y \geqq 0$ holds with strict inequality for $y \neq 0$ we say that H is positive definite and write $H > 0$. Similarly, $H_1 \geqq H_2$ means that $H_1 - H_2 \geqq 0$, etc. The matrix

$$H = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$$

is nonnegative definite if and only if

$$\alpha \geqq 0, \quad \gamma \geqq 0, \quad \text{and} \quad \alpha\gamma - \beta^2 \geqq 0. \tag{3.4}$$

THEOREM 3.2 (Comparison Theorem). *Let $y^1(t)$ and $y^2(t)$ be solutions of* (2.1) *with Hamiltonians $H_1(t)$ and $H_2(t)$ respectively, and let $y^1(0) = y^2(0) \neq 0$. Let* $\arg(y^i(t))$ *be a continuous branch of the argument of $y^i(t)$ with* $\arg(y^1(0)) = \arg(y^2(0))$. *If $H_1(t) \geqq H_2(t)$, then $y^1(t)$ rotates "ahead" of $y^2(t)$ in the sense that* $\arg(y^1(t)) \geqq \arg(y^2(t))$. *If there is a set of positive measure in the interval* $(0, t_0)$ *on which $H_1(t) > H_2(t)$, then* $\arg(y^1(t)) >$ $\arg(y^2(t))$ *for $t \geqq t_0$.*

The following theorem is now a consequence of Theorem 2.1.

THEOREM 3.3. *Let c_n be a constant satisfying* (2.6) *so that $H_{c_n} \in \mathcal{O}_n$. If the Hamiltonian $H(t)$ in* (2.1) *satisfies $H(t) \geqq H_{c_n}$, that is, if*

$$\alpha(t) \geqq c_n, \quad \gamma(t) \geqq c_n, \quad \text{and} \quad (\alpha(t) - c_n)(\gamma(t) - c_n) - \beta^2(t) \geqq 0, \tag{3.5}$$

then $H(t)$ *lies in one of the sets* $\mathcal{O}_n$, π_{n+1}^{*-}, π_{n+1}^{**}, π_{n+1}^{*+}, $\mathcal{H}_{n+1}, \cdots$, *to the right of* $\pi_n^+ = \pi_n^{*+} \cup \pi_n^{**}$ *in Fig.* 2.1.

Similarly, if

$$(3.6) \qquad \alpha(t) \leqq c_n, \quad \gamma(t) \leqq c_n, \quad \text{and} \quad (\alpha(t)-c_n)(\gamma(t)-c_n)-\beta^2(t) \geqq 0,$$

then $H(t)$ *lies in one of the sets* $\mathcal{O}_n$, π_n^{**}, $\mathcal{H}_n$, $\pi_n^{*-}, \cdots$, *to the left of* $\pi_{n+1}^- = \pi_{n+1}^{*-} \cup \pi_{n+1}^{**}$ *in Fig.* 2.1.

THEOREM 3.4. *Let* $H(t)$ *be a Hamiltonian satisfying*

$$(3.7) \qquad \|H(t)-H_{c_n}\| < \rho_{n+1}^-$$

where H_{c_n} *and* ρ_{n+1}^- *are defined by* (2.5) *and* (3.2) *respectively. Then* $H(t)$ *lies in one of the sets* $\mathcal{O}_n$, π_n^{*+}, π_n^{**}, $\mathcal{H}_n, \cdots$, *to the left of* $\pi_{n+1}^- = \pi_{n+1}^{*-} \cup \pi_{n+1}^{**}$ *in Fig.* 2.1.

Similarly, if

$$(3.8) \qquad \|H(t)-H_{c_n}\| < \rho_n^+$$

then $H(t)$ *lies in one of the sets* $\mathcal{O}_n$, π_{n+1}^{*-}, π_{n+1}^{**}, π_{n+1}^{*+}, $\mathcal{H}_{n+1}, \cdots$, *to the right of* $\pi_n^+ = \pi_n^{*+} \cup \pi_n^{**}$ *in Fig.* 2.1.

THEOREM 3.5. *If the Hamiltonian* $H(t)$ *satisfies* (3.5) *and* (3.7) *for some integer n, then* $H(t) \in \mathcal{O}_n$.

Similarly, if $H(t)$ *satisfies* (3.6) *and* (3.8) *for some integer n, then* $H(t) \in \mathcal{O}_n$.

COROLLARY. *If* $H(t)$ *satisfies* (3.5) *and* (3.7) *for some integer n with* $c_n = n\pi/T$, *then* $H(t) \in \mathcal{O}_n \cup \pi_n^{**}$ *and consequently all solutions of* (2.1) *are bounded.*

Similarly, if $H(t)$ *satisfies* (3.6) *and* (3.8) *for some integer n with* $c_n = ((n+1)\pi)/T$, *then* $H(t) \in \mathcal{O}_n \cup \pi_{n+1}^{**}$ *and consequently all solutions of* (2.1) *are bounded.*

In the next section we shall see than when $c_n = n\pi/T$, $n \geqq 0$,

$$(3.9) \quad \rho_{n+1}^- = 2(n+1)\log\left[\frac{1+\cos\dfrac{n\pi}{4(n+1)}}{1+\sin\dfrac{n\pi}{4(n+1)}}\right] + 4(n+1)\left[\frac{\cos\dfrac{n\pi}{4(n+1)}-\sin\dfrac{n\pi}{4(n+1)}}{1+\cos\dfrac{n\pi}{4(n+1)}+\sin\dfrac{n\pi}{4(n+1)}}\right].$$

Test (1.5) is therefore a direct application of this corollary.

4. An optimal control problem. We turn now to the problem of computing the numbers ρ_n^+, ρ_{n+1}^-. We begin with the case $n \geqq 0$ and compute ρ_{n+1}^-. For simplicity we drop the subscript n on c_n.

From (3.2) and (2.3) we see that

$$(4.1) \qquad \rho_{n+1}^- = \inf \int_0^T \{|\alpha(t)-c|+|\beta(t)|+|\gamma(t)-c|\}\,dt$$

where the infimum is taken subject to $H(t) \in \pi_{n+1}^-$. The requirement $H(t) \in \pi_{n+1}^-$ implies that the differential equation (2.1) has a periodic solution of period T if $n+1$ is even, and an antiperiodic solution of period T if $n+1$ is old. That is, the differential equations

$$(4.2) \qquad \begin{aligned} \dot{y}_1 &= -\beta(t)y_1 - \gamma(t)y_2, \\ \dot{y}_2 &= \alpha(t)y_1 + \beta(t)y_2 \end{aligned}$$

with boundary conditions

$$y_1(0) = y_1(T) = 0, \qquad y_2(0) = (-1)^{n+1} y_2(T) \neq 0 \tag{4.3}$$

have a solution. Problem (4.2), (4.3) also has a solution if $H(t) \in \pi_{n+1}^{+} = \pi_{n+1}^{*+} \cup \pi_{n+1}^{**}$. It is also clear from Fig. 2.1 that the infimum (4.1) does not change if π_{n+1}^{-} is replaced by the larger set $\pi_{n+1}^{-} \cup \pi_{n+1}^{+}$. Thus computing ρ_{n+1}^{-} is equivalent to solving the optimal control problem (4.1), (4.2), (4.3). It turns out that the infimum (4.1) is not attained at any $H(t) \in \pi_{n+1}^{-} \cup \pi_{n+1}^{+}$. We determine the infimum by the following device.

We choose a large positive constant l and impose the constraints

$$|\alpha(t)| \leqq l, \quad |\beta(t)| \leqq l, \quad |\gamma(t)| \leqq l \tag{4.4}$$

on the controls $\alpha(t)$, $\beta(t)$, $\gamma(t)$. Let $0 < \varepsilon < m$ be fixed constants. We replace the boundary conditions in (4.3) by the conditions

$$y_1(0) = y_1(T) = 0, \quad y_2(0) \in [\varepsilon, m], \quad (-1)^{n+1} y_2(T) \in [\varepsilon, m]. \tag{4.5}$$

We then consider the problem of computing (4.1) subject to the differential equation (4.2) with boundary conditions (4.5), and subject to the constraints (4.4). We must also impose the constraint

$$\phi_y = (n+1)\pi \tag{4.6}$$

on y. For l sufficiently large, this problem can be shown to have a solution by standard existence theorems in optimal control theory. See for example [3, p. 259]. As we shall see, the optimal trajectory y corresponding to this solution also satisfies the boundary conditions (4.3). We can therefore obtain ρ_{n+1}^{-} as the limiting value of (4.1) subject to (4.2), (4.4), (4.5), and (4.6), as $l \to \infty$.

For a fixed l, sufficiently large, let

$$H(t) = \begin{pmatrix} \alpha(t) & \beta(t) \\ \beta(t) & \gamma(t) \end{pmatrix} \tag{4.7}$$

denote a solution of (4.1), (4.2), (4.4), (4.5), (4.6), and let $y(t)$ denote the corresponding optimal trajectory. Then according to Pontryagin's maximum principle there exists a nonzero function $\psi(t) = (\psi_1(t), \psi_2(t))$ satisfying

$$\begin{gathered} \dot{\psi}_1 = \beta(t)\psi_1 - \alpha(t)\psi_2, \\ \dot{\psi}_2 = \gamma(t)\psi_1 - \beta(t)\psi_2, \\ \psi_2(0) = \psi_2(T) = 0, \end{gathered} \tag{4.8}$$

and a constant $\psi_0 \leqq 0$, such that

$$\begin{aligned} &\max_{\alpha,\beta,\gamma} \{-\psi_1(t)(\beta y_1(t) + \gamma y_2(t)) + \psi_2(t)(\alpha y_1(t) + \beta y_2(t)) + \psi_0[|\alpha - c| + |\beta| + |\gamma - c|]\} \\ &\quad = -\psi_1(t)(\beta(t) y_1(t) + \gamma(t) y_2(t)) + \psi_2(t)(\alpha(t) y_1(t) - \beta(t) y_2(t)) \\ &\qquad + \psi_0[|\alpha(t) - c| + |\beta(t)| + |\gamma(t) - c|] \end{aligned} \tag{4.9}$$

for almost all t in $[0, T]$ where the maximum is taken subject to

$$|\alpha| \leqq l, \quad |\beta| \leqq l, \quad |\gamma| \leqq l.$$

All nonzero solutions of (4.8) have the form

$$\psi_1(t) = \mp \kappa y_2(t), \qquad \psi_2(t) = \pm \kappa y_1(t),$$

or, in vector notation,

$$\psi(t) = \mp\kappa J y(t) \tag{4.10}$$

for some constant $\kappa > 0$. If we substitute this into (4.9) we see that $H(t)$ must provide a pointwise maximum to the expression

$$\pm\kappa y \cdot Hy + \psi_0[|\alpha - c| + |\beta| + |\gamma - c|] \tag{4.11a}$$

over α, β, γ subject to (4.4). From this it follows that $\psi_0 \neq 0$, for the condition $\psi_0 = 0$ implies that at least one entry in $H(t)$ has absolute value l for each $t \in [0, T]$. It therefore follows from (2.4) that $|\phi_y| \to \infty$ as $l \to \infty$. This contradicts (4.6). We therefore have $\psi_0 < 0$, and by properly scaling $\psi(t)$, we shall take $\psi_0 = -1$. Having done this, we can compute $\alpha(t)$, $\beta(t)$, $\gamma(t)$ in terms of y, κ, and l by maximizing (4.11a) subject to (4.4).

To simplify matters we eliminate κ from the expressions for $\alpha(t)$, $\beta(t)$, $\gamma(t)$ by defining

$$z_1(t) = \sqrt{\kappa}\, y_1(t), \qquad z_2(t) = \sqrt{\kappa}\, y_2(t).$$

Expression (4.11a) can now be written as

$$\pm\alpha z_1^2(t) \pm 2\beta z_1(t) z_2(t) \pm \gamma z_2^2(t) - [|\alpha - c| + |\beta| + |\gamma - c|]. \tag{4.11b}$$

Since this expression must be a maximum with respect to α, β, γ, subject to (4.4), we must have

$$\alpha(t) = \begin{cases} \pm l & \text{if } z_1^2(t) > 1, \\ c & \text{if } z_1^2(t) < 1, \end{cases} \tag{4.12a}$$

$$\beta(t) = \begin{cases} \mp l & \text{if } 2z_1(t)z_2(t) < -1, \\ 0 & \text{if } |2z_1(t)z_2(t)| < 1, \\ \pm l & \text{if } 2z_1(t)z_2(t) > 1, \end{cases} \tag{4.12b}$$

$$\gamma(t) = \begin{cases} \pm l & \text{if } z_2^2(t) > 1, \\ c & \text{if } z_2^2(t) < 1. \end{cases} \tag{4.12c}$$

On the surfaces $z_1^2 = 1$, $z_2^2 = 1$, $|2z_1z_2| = 1$, condition (4.9) does not uniquely determine α, β, and γ.

Clearly $z_1(t)$ and $z_2(t)$ satisfy

$$\dot z_1 = -\beta(t) z_1 - \gamma(t) z_2,$$
$$\dot z_2 = \alpha(t) z_1 + \beta(t) z_2,$$

or, in vector notation

$$J\dot z = H(t) z \tag{4.13}$$

and

$$z_1(0) = z_1(T) = 0, \quad z_2(0) > 0, \quad (-1)^{n+1} z_2(T) > 0.$$

We emphasize that the signs in (4.10), (4.11), and (4.12) must be chosen in a consistent manner. Thus if the $-$ sign is used in (4.10) we must set $\pm 1 = 1$ and $\mp 1 = -1$ in (4.11) and (4.12). Since we are computing ρ_{n+1}^-, the sign in (4.10) must be chosen so that $\phi_z = (n+1)\pi \geqq \pi$ since $H(t) \in \pi_{n+1}^- \cup \pi_{n+1}^+$ and $n + 1 \geqq 1$. We shall see shortly that the choice of the $-$ sign causes this to happen.

The procedure for determining $H(t)$ is now clear. The surfaces

$$z_1^2 = 1, \quad z_2^2 = 1, \quad |2z_1z_2| = 1 \tag{4.14}$$

divide the second quadrant of the z_1, z_2 plane into seven regions which we have labeled R_i, $i=0,\cdots,6$, in Fig. 4.1 below. In the interior of each of these regions α, β, and γ are constant and are determined explicitly by (4.12). Thus on the interior of each R_i we have, by (4.13),

$$\frac{d}{dt}z(t)\cdot Hz(t)=2\dot{z}(t)\cdot Hz(t)=-2JHz(t)\cdot Hz(t)=0.$$

This last equality follows from the fact that J rotates vectors through 90°. It follows that solutions of (4.12), (4.13) travel along curves of the form

$$\alpha z_1^2+2\beta z_1z_2+\gamma z_2^2=E \tag{4.15}$$

in phase space. E is constant on the interior of each R_i. We have plotted these curves in Figs. 4.1 and 4.2 for z in the second quadrant with $l=10$ and $c=5$. For z in the remaining quadrants, the curves (4.15) can clearly be obtained as reflections of the second quadrant. Figure 4.1 is obtained when the $-$ sign is chosen in (4.10) and Fig. 4.2 is obtained when the $+$ sign is chosen. In the case we are considering, we clearly cannot have $z_1^2(t)+z_2^2(t)\leqq 1$ for $0\leqq t\leqq T$, for this would imply that $\phi_y<(n+1)\pi$, contradicting (4.6). This shows that $z(t)$ must enter one of the regions R_1, R_2, R_3, R_4, R_5, R_6. And since $\phi_z=(n+1)\pi\geqq\pi$, $z(t)$ must describe an orbit in Fig. 4.1. Thus the $-$ sign must be chosen in (4.10).

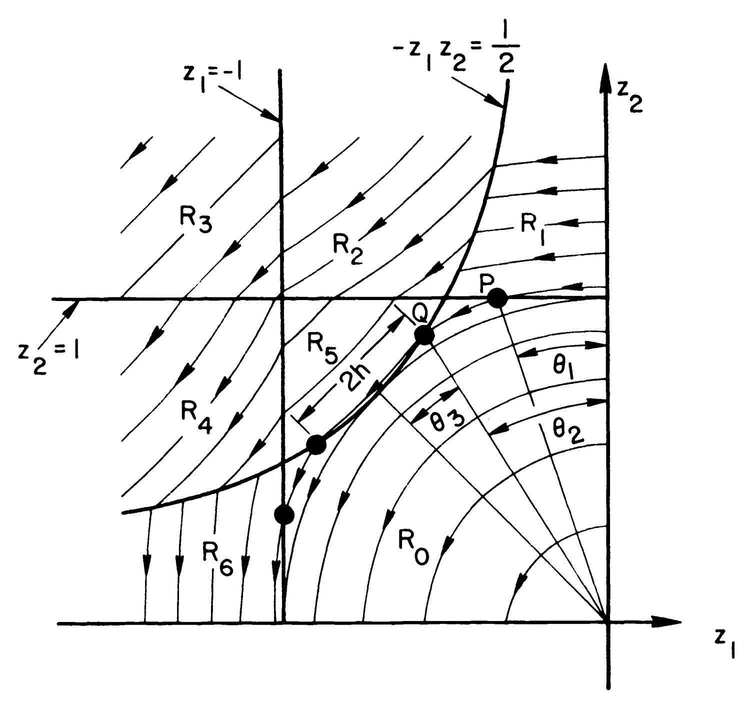

FIG. 4.1. $P=(\xi,1)$, $Q=(\eta_1,\eta_2)$.

If one plots the entire phase portrait of the system (4.12), (4.13), he observes that all orbits close and that the phase portrait is symmetric with respect to the z_1 and z_2 axes. Since the optimal trajectory $z(t)$ satisfies $\phi_z=(n+1)\pi$, it also satisfies $z_1(0)=z_1(T)=0$, $z_2(0)=(-1)^{n+1}z_2(T)$. The optimal trajectory, $y(t)=(1/\sqrt{\kappa})z(t)$ therefore satisfies the boundary conditions (4.3). This shows that the solution of (4.1), (4.2), (4.4), (4.5), (4.6) is also a solution of (4.1), (4.2), (4.3), (4.4), (4.6). By letting $l \to \infty$ we obtain the value ρ_{n+1}^- defined by (3.2).

Since the optimal trajectory must satisfy $\phi_z = (n+1)\pi$ and must start and end on the z_2 axis, it must cross the z_1 axis $n+1$ times. It must therefore pass through $2(n+1)$ quadrants, counting repetitions. By the symmetry of the phase portrait, the optimal $z(t)$ must spend time

$$\frac{T}{2(n+1)}$$

in each quadrant. But in regions R_1, R_2, R_3, R_4, R_5, R_6, at least one of the quantities $|\alpha(t)|$, $|\beta(t)|$, $|\gamma(t)|$ has value l and so the time that $z(t)$ spends in these regions approaches zero as $l \to \infty$. It follows that the time that $z(t)$ spends in one crossing of R_0 approaches $T/(2(n+1))$ as $l \to \infty$. We can therefore find the limiting position of $z(t)$ in R_0, as $l \to \infty$, by finding the trajectory which spends exactly time $T/(2(n+1))$ in R_0.

We have seen that $z_1^2(t) + z_2^2(t) > 1$ for $z(t) \in R_0$. It follows that $z(t)$ enters R_0 from R_1 at a point $P = (\xi, 1)$ where $-\frac{1}{2} < \xi < 0$. $z(t)$ then travels along the circle $z_1^2 + z_2^2 = \xi^2 + 1$ reaching R_5 at a point $Q = (\eta_1, \eta_2)$. See Fig. 4.1. As $l \to \infty$, ξ and η approach limiting values which we again denote by ξ and η, respectively.

Let $\tau_i(l)$ denote the time at which $z(t)$ enters R_i for the first time and let $t_i(l)$ denote the time required for $z(t)$ to pass through R_i. Then $z_1(\tau_0(l)) = \xi$, $z_2(\tau_0(l)) = 1$, $z_1(\tau_5(l)) = \eta_1$, and $z_2(\tau_5(l)) = \eta_2$. Since $z(t)$ passes through one quadrant in time $T/(2(n+1))$, clearly

$$\frac{1}{2} t_0(l) = \tau_5(l) - \tau_0(l) \to \frac{T}{4(n+1)}$$

as $l \to \infty$.

$z(t)$ leaves R_5 at the point $Q' = (-\eta_2, -\eta_1)$ and reaches R_6 at the point $P' = (-1, -\xi)$. It follows from (4.12) that

$$\begin{aligned} &\alpha(t) = c, \quad \gamma(t) = l, \quad \beta(t) = 0 \quad \text{for } z(t) \in R_1, \\ &\alpha(t) = c, \quad \gamma(t) = c, \quad \beta(t) = 0 \quad \text{for } z(t) \in R_0, \\ &\alpha(t) = c, \quad \gamma(t) = c, \quad \beta(t) = -l \quad \text{for } z(t) \in R_5, \\ &\alpha(t) = l, \quad \gamma(t) = c, \quad \beta(t) = 0 \quad \text{for } z(t) \in R_6. \end{aligned} \tag{4.16}$$

We therefore have

$$\begin{aligned} \int_0^{T/(2(n+1))} \{|\alpha(t) - c| + |\beta(t)| + |\gamma(t) - c|\}\, dt &= 2 \int_0^{\tau_0(l)} |l - c|\, dt + \int_{\tau_5(l)}^{\tau_5(l)+t_5(l)} l\, dt \\ &= 2|l - c|\tau_0(l) + l t_5(l), \end{aligned}$$

and since each of the $2(n+1)$ quadrants that $z(t)$ passes through is just a copy of Fig. 4.1 we have

$$\int_0^T \{|\alpha(t) - c| + |\beta(t)| + |\gamma(t) - c|\}\, dt = 2(n+1)\{2|l - c|\tau_0(l) + l t_5(l)\}. \tag{4.17}$$

All that remains is to compute the limiting value of this expression as $l \to \infty$. This will give the value of ρ_{n+1}^-, $n \geqq 0$.

On the interval $(0, \tau_0(l))$ we have $z(t) \in R_1$ and it follows from (4.16) and (4.13) that

$$\frac{dz_2}{dz_1} = -\frac{cz_1}{lz_2}.$$

Thus as $l \to \infty$, the portion of the trajectory $z(t)$ lying in R_1 approaches the horizontal line segment $\xi \leqq z_1 \leqq 0$, $z_2 = 1$, uniformly in t. From now on we shall not distinguish ξ and η from their limiting values as $l \to \infty$. We have

$$\xi = \int_0^{\tau_0(l)} \dot{z}_1(t)\, dt = -l\tau_0(l) \cdot \frac{1}{\tau_0(l)} \int_0^{\tau_0(l)} z_2(t)\, dt,$$

and since $z_2(t) \to 1$ uniformly in t as $l \to \infty$, we have

$$|\xi| = \lim_{l \to \infty} l\tau_0(l) = \lim_{l \to \infty} |l - c|\tau_0(l). \tag{4.18}$$

For $z(t) \in R_5$ we have, by (4.16) and (4.13)

$$\dot{z}_1 = lz_1 - cz_2$$

so that

$$z_1(t) = z_1(\tau_5(l))\, e^{l(t-\tau_5(l))} - c \int_{\tau_5(l)}^{t} e^{l(t-s)} z_2(s)\, ds.$$

if we take $t = \tau_5(l) + t_5(l)$ we obtain

$$-\eta_2 = \eta_1\, e^{lt_5(l)} - c \int_{\tau_5(l)}^{\tau_5(l)+t_5(l)} e^{l(\tau_5(l)+t_5(l)-s)} z_2(s)\, ds. \tag{4.19}$$

For $z(s) \in R_5$ we clearly have

$$l(\tau_5(l) + t_5(l) - s) \leqq lt_5(l) = \int_{\tau_5(l)}^{\tau_5(l)+t_5(l)} |\beta(t)|\, dt < 2\rho_{n+1}^{-}$$

for l sufficiently large. Thus

$$|e^{l(\tau_5(l)+t_5(l)-s)} z_2(s)| \leqq e^{2\rho_{n+1}},$$

and since $t_5(l) \to 0$ as $l \to \infty$, the integral in (4.19) approaches 0 as $l \to \infty$. It follows that

$$\log\left(-\frac{\eta_2}{\eta_1}\right) = \lim_{l \to \infty} lt_5(l). \tag{4.20}$$

The limiting value of (4.17) can now be expressed in terms of ξ and η through (4.18) and (4.20).

Let θ_1 and θ_2 denote the angles which the respective line segments connecting the origin and the points $(\xi, 1)$, (η_1, η_2) make with the positive z_2 axis. Let $\theta_3 = \pi/4 - \theta_2$. We then have

$$\tan \theta_1 = |\xi|$$

and

$$\cot \theta_2 = -\frac{\eta_2}{\eta_1},$$

and it follows from (4.17), (4.18), (4.20) that

$$\rho_{n+1}^{-} = 2(n+1)\{2 \tan \theta_1 + \log \cot \theta_2\}. \tag{4.21}$$

For $z(t) \in R_0$ let

$$\theta(t) = \mathrm{Tan}^{-1} \frac{z_2(t)}{z_1(t)}.$$

It then follows from (4.16) and (4.13) that

$$\dot{\theta}(t) = c.$$

And since $z(t)$ goes from $(\xi, 1)$ to (η_1, η_2) in time $T/(4(n+1))$ we have

$$\theta_2 - \theta_1 = \frac{cT}{4(n+1)}.$$

Let $2h$ denote the distance from (η_1, η_2) to $(-\eta_2, -\eta_1)$. It is then clear from Fig. 4.1 that

$$\sin\theta_3 = \frac{h}{\sqrt{\xi^2+1}} \quad \text{and} \quad \sin\theta_1 = \frac{|\xi|}{\sqrt{\xi^2+1}}.$$

Moreover,

$$2h = [(\eta_1+\eta_2)^2 + (\eta_2+\eta_1)^2]^{1/2} = [2(\eta_1^2 + 2\eta_1\eta_1 + \eta_2^2)]^{1/2} = [2(\xi^2+1-1)]^{1/2} = \sqrt{2}|\xi|$$

so that

$$\sin\theta_3 = \frac{1}{\sqrt{2}}\sin\theta_1.$$

On the other hand

$$\begin{aligned}
\sin\theta_3 &= \sin\left(\frac{\pi}{4} - \theta_2\right) = \sin\left(\frac{\pi}{4} - \frac{cT}{4(n+1)} - \theta_1\right) \\
&= \sin\left(\frac{\pi}{4} - \frac{cT}{4(n+1)}\right)\cos\theta_1 - \cos\left(\frac{\pi}{4} - \frac{cT}{4(n+1)}\right)\sin\theta_1 \\
&= \frac{1}{\sqrt{2}}\left(\cos\frac{cT}{4(n+1)} - \sin\frac{cT}{4(n+1)}\right)\cos\theta_1 \\
&\quad - \frac{1}{\sqrt{2}}\left(\cos\frac{cT}{4(n+1)} + \sin\frac{cT}{4(n+1)}\right)\sin\theta_1.
\end{aligned}$$

It follows that

$$\tan\theta_1 = \frac{\sin\theta_1}{\cos\theta_1} = \frac{\cos\dfrac{cT}{4(n+1)} - \sin\dfrac{cT}{4(n+1)}}{1 + \cos\dfrac{cT}{4(n+1)} + \sin\dfrac{cT}{4(n+1)}}. \tag{4.22}$$

Similarly,

$$\sin\theta_3 = \sin\left(\frac{\pi}{4} - \theta_2\right) = \frac{1}{\sqrt{2}}\sin\theta_1 = \frac{1}{\sqrt{2}}\sin\left(\theta_2 - \frac{cT}{4(n+1)}\right)$$

which implies that

$$\cot\theta_2 = \frac{1 + \cos\dfrac{cT}{4(n+1)}}{1 + \sin\dfrac{cT}{4(n+1)}}. \tag{4.23}$$

Substituting (4.22) and (4.23) into (4.21) we obtain the formula

$$\rho_{n+1}^{-}=2(n+1)\log\left[\frac{1+\cos\dfrac{cT}{4(n+1)}}{1+\sin\dfrac{cT}{4(n+1)}}\right]+4(n+1)\left[\frac{\cos\dfrac{cT}{4(n+1)}-\sin\dfrac{cT}{4(n+1)}}{1+\cos\dfrac{cT}{4(n+1)}+\sin\dfrac{cT}{4(n+1)}}\right]. \tag{4.24}$$

By taking $c=n\pi/T$ we obtain (3.9).

Since (4.1) is a decreasing function of l, (4.24) gives a lower bound on the value of the optimal control problem (4.1), (4.2), (4.4), (4.5), (4.6) for arbitrary values of the parameters c, l, ε, m satisfying

$$0<c<\frac{(n+1)\pi}{T},\qquad 0<\varepsilon<m,\quad l \text{ sufficiently large.}$$

For the case $c_n=n\pi/T$ the following lemma gives the simpler, but slightly smaller, lower bound π.

LEMMA 4.1. *If $c_n=n\pi/T$ and $n\geqq 0$, then $\rho_{n+1}^{-}>\pi$ and $\lim_{n\to\infty}\rho_{n+1}^{-}=\pi$.*

Proof. Observe that

$$\sqrt{2}\sin\frac{\pi}{4(n+1)}=\sqrt{2}\sin\left(\frac{\pi}{4}-\frac{n\pi}{4(n+1)}\right)=\cos\frac{n\pi}{4(n+1)}-\sin\frac{n\pi}{4(n+1)}.$$

This implies that

$$\lim_{n\to\infty}4(n+1)\left[\cos\frac{n\pi}{4(n+1)}-\sin\frac{n\pi}{4(n+1)}\right]=\sqrt{2}\pi. \tag{4.25}$$

We therefore have

$$\lim_{n\to\infty}4(n+1)\,\frac{\cos\dfrac{n\pi}{4(n+1)}-\sin\dfrac{n\pi}{4(n+1)}}{1+\cos\dfrac{n\pi}{4(n+1)}+\sin\dfrac{n\pi}{4(n+1)}}=\frac{\sqrt{2}\pi}{1+\sqrt{2}}. \tag{4.26}$$

Now consider the first term in (3.9). We have

$$\begin{aligned}\log&\left[\frac{1+\cos\dfrac{n\pi}{4(n+1)}}{1+\sin\dfrac{n\pi}{4(n+1)}}\right]\\&=\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}\cos^k\frac{n\pi}{4(n+1)}-\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}\sin^k\frac{n\pi}{4(n+1)}\\&=\left[\cos\frac{n\pi}{4(n+1)}-\sin\frac{n\pi}{4(n+1)}\right]\left\{1+\sum_{k=1}^{\infty}\frac{(-1)^k}{k+1}\left[\sum_{j=0}^{k}\cos^{k-j}\frac{n\pi}{4(n+1)}\sin^j\frac{n\pi}{4(n+1)}\right]\right\}.\end{aligned}$$

It therefore follows from (4.25) that

$$\lim_{n\to\infty}2(n+1)\log\frac{1+\cos\dfrac{n\pi}{4(n+1)}}{1+\sin\dfrac{n\pi}{4(n+1)}}=\frac{\pi}{\sqrt{2}}\left\{1+\sum_{k=1}^{\infty}(-1)^k\left(\frac{1}{\sqrt{2}}\right)^k\right\}=\frac{\pi}{1+\sqrt{2}}. \tag{4.27}$$

The fact that $\lim_{n\to\infty} \rho_{n+1}^- = \pi$ now follows from (4.26) and (4.27). It is easy to verify that ρ_{n+1}^- is a decreasing function of n for $n \geqq 0$.

Consider now the problem of computing ρ_n^+, $n \geqq 1$. This problem has an obvious formulation as an optimal control problem similar to (4.1). In this case the optimal trajectory $z(t)$ passes through one quadrant in time $T/(2n)$. Moreover, the time that $z(t)$ spends in the interior of each R_i, $i > 0$, approaches 0 as $l \to \infty$. On the other hand, the sum of the angles through which $z(t)$ rotates in one crossing of R_0 is given by $ct_0(l)$. If $t_0(l)$ were to approach $T/(2n)$ as $l \to \infty$, we would have, for l sufficiently large

$$ct_0(l) > \pi/2$$

by (2.6). But this is impossible since R_0 lies in one quadrant of phase space. It follows that $z(t)$ must remain fixed for some time $\Delta T > 0$ at a point in the boundary of R_0. This means that the + sign must be chosen in (4.10), since the system (4.12), (4.13) has no rest points (except the origin) when the − sign is chosen. This is readily seen from the phase portrait Fig. 4.1. The phase portrait of the system (4.12), (4.13) corresponding to the + sign in (4.10) is shown in Fig. 4.2. The point $(-1/\sqrt{2}, 1/\sqrt{2})$ is the only rest point in the boundary of R_0. We must therefore have $z(0) = (0, 1)$.

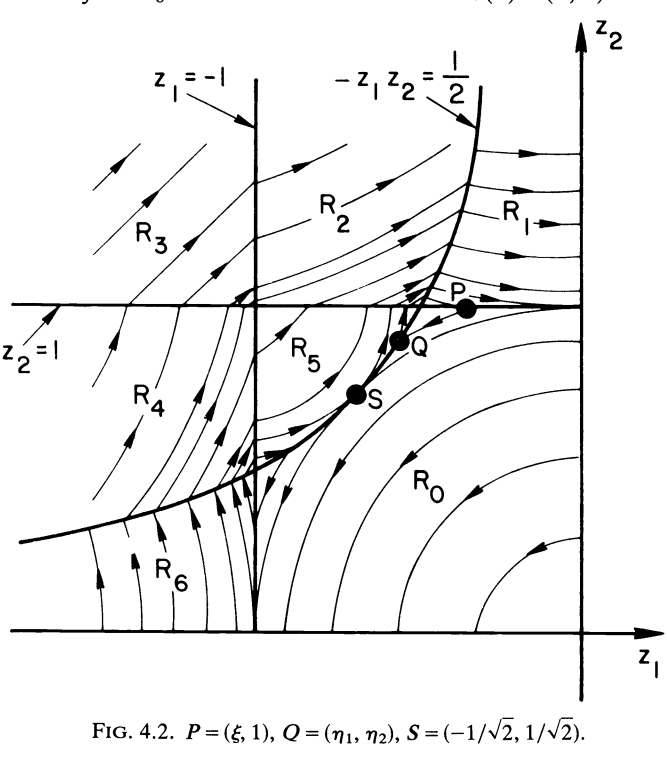

FIG. 4.2. $P = (\xi, 1)$, $Q = (\eta_1, \eta_2)$, $S = (-1/\sqrt{2}, 1/\sqrt{2})$.

When the + sign is used in (4.10) the − sign must be used in (4.11b). At $(z_1(t), z_2(t)) = (-1/\sqrt{2}, 1/\sqrt{2})$, (4.11b) is maximized by taking $\alpha = c$, $\gamma = c$ and $\beta \geqq 0$. In particular we can take $\beta = c$. By substituting these values in (4.13) one verifies that $(-1/\sqrt{2}, 1/\sqrt{2})$ is a rest point. Similarly, $(-1/\sqrt{2}, -1/\sqrt{2})$, $(1/\sqrt{2}, -1/\sqrt{2})$, and $(1/\sqrt{2}, 1/\sqrt{2})$ are rest points in the remaining quadrants.

Since $z(t)$ makes one crossing of R_0 in time $T/(2n) - \Delta T$ we must have

$$\frac{\pi}{2} = c\left(\frac{T}{2n} - \Delta T\right)$$

so that

$$\Delta T = \frac{T}{2n} - \frac{\pi}{2c}.$$

Clearly

$$\int_0^{T/(2n)} \{|\alpha(t)-c| + |\beta(t)| + |\gamma(t)-c|\}\, dt = c\, \Delta T = \frac{cT}{2n} - \frac{\pi}{2}.$$

It follows that

(4.28) $$\rho_n^+ = 2nc\, \Delta T = cT - n\pi, \qquad n \geqq 1.$$

The case $n = 0$ is special. In this case we must have $\phi_y = 0$. In particular $y(t)$ does not rotate around the origin. Thus in the control theoretic formulation of the problem of determining ρ_0^+ it is not possible to assert that $y_1(0) = y_1(T) = 0$. All we know is that $y(0) = y(T)$. However, it is clear that the optimal trajectory satisfies

(4.29) $$y_1(0) + \lambda y_2(0) = 0, \qquad y_1(T) + \lambda y_2(T) = 0$$

for some constant $\lambda \neq 0$. That is, the optimal trajectory starts and ends on a line segment of the form

$$y_1 + \lambda y_2 = 0$$

and forms a closed orbit.

If in the control problem associated with ρ_0^+ one imposes the boundary conditions (4.29) he finds that the adjoint variables $\psi_1(t)$, $\psi_2(t)$ satisfy the transversality conditions

$$\lambda\psi_1(0) - \psi_2(0) = 0, \qquad \lambda\psi_1(T) - \psi_2(T) = 0$$

and the differential equation (4.8). It follows that $\psi(t)$ is related to $y(t)$ by an equation of form (4.10). Thus the theory proceeds as in the case $n > 0$. The phase portrait in the z space must contain a closed orbit which does not encircle the origin. The phase portrait must therefore be given by Fig. 4.2, which means that the + sign must be chosen in (4.10).

The optimal trajectory $z(t)$ must enter R_0 at a point $P = (\xi, 1)$, $-\frac{1}{2} < \xi \leqq 0$. It crosses R_0 and enters R_5 at a point $Q = (\eta_1, \eta_2)$. All of this is accomplished in time $t_0(l)$. Clearly $ct_0(l) \leqq \pi/4$. If $cT \geqq \pi/4$, then $\xi = 0$ and $(\eta_1, \eta_2) = (-1/\sqrt{2},\ 1/\sqrt{2})$ is a rest point for $z(t)$. $z(t)$ remains at this point for a time ΔT which approaches $T - \pi/(4c)$ as $l \to \infty$. $z(t)$ leaves the rest point and crosses R_5, first entering R_2, and then R_1, at points which approach $(-\frac{1}{2}, 1)$ as $l \to \infty$. Finally, $z(t)$ returns to the starting point $(\xi, 1)$ after time T has elapsed. Clearly

$$\int_0^T \{|\alpha(t)-c| + |\beta(t)| + |\gamma(t)-c|\}\, dt = c\, \Delta T + l t_5(l) + |l-c|(t_2(l) + t_1(l)).$$

The limiting value of this expression as $l \to \infty$ is easily found to be

(4.30) $$\rho_0^+ = cT - \frac{\pi}{4} + \log \sqrt{2} + \tfrac{1}{2}.$$

If $cT < \pi/4$, then $\xi < 0$ and $\eta_2 > 1/\sqrt{2}$, and it is easily shown that

(4.31) $$\rho_0^+ = \log \frac{1}{\eta_2} + \frac{1}{2} - |\xi| = \log \sqrt{2}(\sqrt{1+|\xi|^2} - |\xi|) + \frac{1}{2} - |\xi|$$

where

$$|\xi|=\frac{\cos(cT)-\sin(cT)}{1+\cos(cT)+\sin(cT)}.$$

We have now determined ρ_n^+ and ρ_{n+1}^- for $n=0,1,2,\cdots$. We leave it to the reader to determine ρ_n^+ and ρ_{n+1}^- for $n=-1,-2,\cdots$. The results are

$$\rho_n^+=2|n|\log\frac{1+\cos\dfrac{cT}{4n}}{1+\sin\dfrac{cT}{4n}}+4|n|\left[\frac{\cos\dfrac{cT}{4n}-\sin\dfrac{cT}{4n}}{1+\cos\dfrac{cT}{4n}+\sin\dfrac{cT}{4n}}\right], \tag{4.32}$$

$n=-1,-2,\cdots$, and

$$\rho_{n+1}^-=|c|T-|n+1|\pi,\qquad n=-2,-3,\cdots, \tag{4.33}$$

where $c=c_n$ satisfies (2.6).

Finally, ρ_0^- can be obtained from the formula for ρ_0^+ by replacing c with $|c|$. Thus

$$\rho_0^-=\begin{cases}|c|T-\pi/4+\log\sqrt{2}+\frac{1}{2} & \text{if } |c|T\geqq\pi/4,\\ \log\sqrt{2}(\sqrt{1+|\xi|^2}-|\xi|)+\frac{1}{2}-|\xi| & \text{if } |c|T<\pi/4,\end{cases} \tag{4.34}$$

where

$$|\xi|=\frac{\cos|c|T-\sin|c|T}{1+\cos|c|T+\sin|c|T}.$$

5. Practical stability tests. In this section we list several stability tests that can be deduced from results obtained in § 4.

THEOREM 5.1. *All solutions of* (1.1) *are bounded if* $\alpha(t)$, $\beta(t)$, *and* $\gamma(t)$ *satisfy* (1.5) *for some* $n\geqq 0$.

Proof. This follows from the corollary to Theorem 3.5 and formula (3.9).

THEOREM 5.2. *All solutions of* (1.1) *are bounded if for some integer n, unrestricted in sign,*

$$\alpha(t)\geqq\frac{n\pi}{T},\quad \gamma(t)\geqq\frac{n\pi}{T},\quad \left(\alpha(t)-\frac{n\pi}{T}\right)\left(\gamma(t)-\frac{n\pi}{T}\right)-\beta^2(t)\geqq 0,$$

and

$$\int_0^T\{|\alpha(t)|+|\beta(t)|+|\gamma(t)|\}\,dt<(2|n|+1)\pi.$$

Proof. For $n\geqq 0$ the result follows from Theorem 5.1 and Lemma 4.1. For $n<0$ the result follows from the corollary to Theorem 3.5, together with the fact that $\rho_{n+1}^-=\pi$ for $c_n=n\pi/T$ and $n<-1$. See (4.33). Also, $\rho_0^->\pi$ for $c=-\pi/T$.

THEOREM 5.3. *All solutions of* (1.1) *are bounded if for some integer* $n<0$,

$$\begin{gathered}\alpha(t)\leqq\frac{(n+1)\pi}{T},\qquad \gamma(t)\leqq\frac{(n+1)\pi}{T},\\ \left(\alpha(t)-\frac{(n+1)\pi}{T}\right)\left(\gamma(t)-\frac{(n+1)\pi}{T}\right)-\beta^2(t)\geqq 0,\end{gathered} \tag{5.1}$$

and

$$\int_0^T \{|\alpha(t)|+|\beta(t)|+|\gamma(t)|\}\,dt$$

$$<2|n+1|\pi+2|n|\log\left[\frac{1+\cos\dfrac{(n+1)\pi}{4n}}{1+\sin\dfrac{(n+1)\pi}{4n}}\right]+4|n|\left[\frac{\cos\dfrac{(n+1)\pi}{4n}-\sin\dfrac{(n+1)\pi}{4n}}{1+\cos\dfrac{(n+1)\pi}{4n}+\sin\dfrac{(n+1)\pi}{4n}}\right].$$

Proof. This theorem is a consequence of the corollary to Theorem 3.5 and formula (4.32).

THEOREM 5.4. *All solutions of* (1.1) *are bounded if for some integer n, unrestricted in sign*, (5.1) *holds and*

$$\int_0^T \left\{\left|\alpha(t)-\frac{(n+1)\pi}{T}\right|+|\beta(t)|+\left|\gamma(t)-\frac{(n+1)\pi}{T}\right|\right\}dt<\pi. \tag{5.2}$$

Proof. For $n<0$, ρ_n^+ is given by (4.32). It follows as in the proof of Lemma 4.1 that $\rho_n^+>\pi$ for $c=(n+1)\pi/T$ and $n<0$. Moreover,

$$\lim_{n\to-\infty}\rho_n^+=\pi.$$

Thus for $n<0$ the conclusion of the theorem follows from Theorem 5.3.

For $n>0$, (4.28) shows that $\rho_n^+=\pi$ if $c=(n+1)\pi/T$. Also, (4.30) shows that $\rho_0^+>\pi$ for $c=\pi/T$. Thus for $n\geqq 0$ the theorem follows from the corollary to Theorem 3.4.

THEOREM 5.5. *All solutions of* (1.1) *are bounded if for some integer n we have*

$$\int_0^T \left\{\left|\alpha(t)-\frac{(2n+1)\pi}{2T}\right|+|\beta(t)|+\left|\gamma(t)-\frac{(2n+1)\pi}{2T}\right|\right\}dt<\frac{\pi}{2}.$$

Proof. If one takes c_n to be the midpoint of the interval (2.6), then it is easily shown that $\rho_{n+1}^-\geqq\pi/2$ and $\rho_n^+\geqq\pi/2$ for each n. The conclusion of the theorem therefore follows from Theorem 3.1.

Example. At $\lambda=0$ all solutions of the differential equations

$$\begin{aligned}\dot y_1&=-\lambda(\sin(2\pi t)\cos(2\pi t))y_1-\lambda(\sin^2(2\pi t))y_2,\\ \dot y_2&=\lambda(\cos^2(2\pi t))y_1+\lambda(\sin(2\pi t)\cos(2\pi t))y_2\end{aligned} \tag{5.3}$$

are bounded. We wish to determine how large we can make λ before we encounter unbounded solutions. This example has been chosen so that the answer can be determined explicitly. We wish to compare this with the estimates given by Theorem 5.1.

Let

$$\alpha(t)=\lambda\cos^2(2\pi t),\quad \gamma(t)=\lambda\sin^2(2\pi t),\quad \beta(t)=\lambda\sin(2\pi t)\cos(2\pi t)$$

where $\lambda>0$. These functions are periodic of period $\frac{1}{2}$ and

$$\int_0^{1/2}\{\alpha(t)+|\beta(t)|+\gamma(t)\}\,dt=\frac{\lambda}{2}+\frac{\lambda}{2\pi}.$$

It therefore follows from Theorem 5.1, with $n=0$, that the Hamiltonian

$$H(t)=\begin{pmatrix}\alpha(t) & \beta(t)\\ \beta(t) & \gamma(t)\end{pmatrix}$$

lies in $\mathcal{O}_0$ if

$$\frac{\lambda}{2}+\frac{\lambda}{2\pi}<\rho_1^-=3.386294.$$

Thus all solutions of (5.3) are bounded if

$$0\leqq\lambda<5.13727.$$

We shall now show that, in fact, all solutions are bounded if

$$0\leqq\lambda<2\pi=6.283185$$

and that 2π is the largest we can make λ without encountering unbounded solutions. First observe that, if $\lambda=2\pi$

$$y_1(t)=\cos(2\pi t),\qquad y_2(t)=\sin(2\pi t) \tag{5.4}$$

is an antiperiodic solution of (5.3) satisfying $\phi_y=\pi$. Then observe that if $y(t)$ is any other solution of (5.3), with $\lambda=2\pi$, we have, by (2.4),

$$\begin{aligned}\phi_y&=2\pi\int_0^{1/2}\frac{y_1^2\cos^2(2\pi t)+2y_1y_2\sin(2\pi t)\cos(2\pi t)+y_2^2\sin^2(2\pi t)}{y_1^2+y_2^2}\,dt\\ &=2\pi\int_0^{1/2}\frac{(y_1\cos(2\pi t)+y_2\sin(2\pi t))^2}{y_1^2+y_2^2}\,dt\\ &\leqq2\pi\int_0^{1/2}\frac{(y_1^2+y_2^2)(\cos^2(2\pi t)+\sin^2(2\pi t))}{y_1^2+y_2^2}\,dt=\pi,\end{aligned}$$

and equality holds if and only if $y(t)$ is a multiple of the solution (5.4). It follows from Theorem 2.1 that $H(t)\in\pi_1^{*-}$ for $\lambda=2\pi$. We must therefore have $H(t)\in\mathcal{O}_0$ for $0<\lambda<2\pi$.

REFERENCES

[1] M. G. KREIN, *On certain problems on the maximum and minimum characteristic values and on Liapunov zones of stability*, Prikl. Mat. Meh., 15 (1951). (In Russian.) English translation, Amer. Math. Soc. Transl., Ser. 2, 1 (1955), pp. 163–187.

[2] V. A. YAKUBOVICH AND V. M. STARZHINSKII, *Linear Differential Equations with Periodic Coefficients*, John Wiley, New York, 1975.

[3] E. B. LEE AND L. MARKUS, *Foundations of Optimal Control Theory*, John Wiley, New York, 1967.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0010 $01.00/0

ON THE A-ACCEPTABILITY OF PADÉ APPROXIMATIONS*

ARIEH ISERLES†

Abstract. The problem of A-acceptability of the Padé approximations is explored. The conjecture of Ehle, according to which these approximations are not A-acceptable for $n \geqq m+3$ is verified in the case $n-m \not\equiv 2$ (mod. 4).

1. Introduction. The Padé approximations to the exponential function have been an object of extensive research in the recent years. These approximations are a natural outcome when one solves ordinary or partial linear initial value problems.

The Padé approximation $R_{n,m}$ is a rational function,

$$R_{n,m}(z) := P_{n,m}(z)/Q_{n,m}(z),$$

$$P_{n,m}(z) := \sum_{k=0}^{m} \frac{(n+m-k)!\,m!}{n!\,k!\,(m-k)!} z^k,$$

$$Q_{n,m}(z) := \sum_{k=0}^{n} \frac{(n+m-k)!}{k!\,(n-k)!}(-z)^k.$$

One verifies readily that

$$R_{n,m}(z) - e^z = \mathcal{O}(z^{n+m+1}),$$

We call a rational function A-acceptable if it is analytic in the complex left half-plane and there its modulus does not exceed one. Clearly, when the solution of the stable scalar linear differential equation $\dot{x} = \lambda x$, $x(0)=1$, $\operatorname{Re} \lambda < 0$, is approximated by an A-acceptable rational function, the approximation itself is stable.

Birkhoff and Varga [1] prove that $R_{n,n}$ is A-acceptable for every $n \geqq 1$. Ehle [2] shows that $R_{n+1,n}$ and $R_{n+2,n}$ are A-acceptable for every $n \geqq 0$. Furthermore, Ehle conjectures that these, namely the diagonal and the first two subdiagonals of the Padé tableau, are the only A-acceptable functions of this type. Actually, Ehle himself [2] and Nørsett [3] show that $R_{n+3,n}$, $R_{n+4,n}$ and $R_{n+3,0}$ are not A-acceptable for any $n \geqq 0$.

The purpose of this paper is to verify the conjecture of Ehle in certain general cases. We prove that if $n-m \not\equiv 2$ (mod. 4) and $n \geqq m+3 \geqq 3$ then $R_{n,m}$ is not A-acceptable and the conjecture is valid.

2. The case $n-m \not\equiv 2$ (mod. 4). According to the maximum modulus theorem $R_{n,m}$ is A-acceptable if and only if the three following conditions are simultaneously valid:

A. $R_{n,m}$ is analytical in the left half plane.

B. $|R_{n,m}(it)| \leqq 1$ for every real t.

C. $\lim_{|z|\to\infty} |R_{n,m}(z)| \leqq 1$.

Obviously, the condition C implies that $n \geqq m$. The condition A is equivalent to the absence of zeros of $Q_{n,m}$ in the whole left half plane.

We now explore the condition B in greater detail:

We define

$$\omega_{n,m}(t) := |Q_{n,m}(it)|^2 - |P_{n,m}(it)|^2.$$

As a matter of fact, the condition B is equivalent to the nonnegativity of $\omega_{n,m}$ for every

* Received by the editors September 27, 1977.

† King's College, University of Cambridge, Cambridge CB2 1ST, England.

real t. Furthermore, $\omega_{n,m}(t)=\omega_{n,m}(-t)$ and $\omega_{n,m}$ is an even polynomial. Thus

$$\omega_{n,m}(t)=\sum_{k=0}^{n} A_k^{(n,m)}t^{2k}.$$

We know that $R_{n,m}(it)=\exp(it)+\mathcal{O}(|t|^{n+m+1})$. Therefore

$$|R_{n,m}(it)|^2=1+\mathcal{O}(|t^{n+m+1}|),$$

implying $\omega_{n,m}(t)=\mathcal{O}(|t^{n+m+1}|)$. Hence, $A_k^{(n,m)}=0$ for every $0\leqq k\leqq[(n+m)/2]$. We now determine the expression for $A_k^{(n,m)}$, $[(n+m+2)/2]\leqq k\leqq n$.

LEMMA 1.

$$A_{n-l}^{(n,m)}=(-1)^l\frac{m!(m+l)!(n-m-1-l)!}{l!n!(n-l)!(n-m-1-2l)!},\qquad 0\leqq l\leqq\left[\frac{n-m-1}{2}\right].$$

Proof. $m\leqq n$ implies $m\leqq[(n+m)/2]$ and thus $|P_{n,m}(it)|^2$, as well as the lower-order terms of $|Q_{n,m}(it)|^2$, does not contribute to $\omega_{n,m}$. Therefore, $\omega_{n,m}$ equals to the high-order terms of $|Q_{n,m}(it)|^2$.

But

$$Q_{n,m}(it)=(-1)^n\left\{\frac{m!}{0!n!}(it)^n-\frac{(m+1)!}{1!(n-1)!}(it)^{n-1}+\frac{(m+2)!}{2!(n-2)!}(it)^{n-2}-\cdots\right\}$$

and we see by introspection that

$$A_{n-l}^{(n,m)}=\left[\frac{(m+l)!}{l!(n-l)!}\right]^2+2\sum_{i=1}^{l}(-1)^i\frac{(m+l+i)!(m+l-i)!}{(n-l+i)!(n-l-i)!(l+i)!(l-i)!}.$$

Hence, in order to prove the argument of the lemma, it suffices to show that

$$\sum_{i=1}^{l}(-1)^i\frac{(m+l+i)!(m+l-i)!}{(n-l+i)!(n-l-i)!(l+i)!(l-i)!}$$
$$=\frac{1}{2}\frac{(m+l)!}{l!(n-l)!}\left[\frac{(m+l)!}{l!(n-l)!}+(-1)^l\frac{m!(n-m-1-l)!}{n!(n-m-1-2l)!}\right]$$

for every $0\leqq l\leqq[(n-m-1)/2]$.

We define

$$Z(m,n,l):=\sum_{i=0}^{l}(-1)^i\frac{(m+l+i)!(m+l-i)!}{(l+i)!(l-i)!(n-l+i)!(n-l-i)!}.\tag{1}$$

We obtain

$$Z(m,n,l)=\sum_{i=0}^{l}(-1)^i\frac{(m+l+i-1)!(m+l-i-1)!}{(l+i)!(l-i)!(n-l+i)!(n-l-i)!}(m^2+2ml+(l+i)(l-i))$$
$$=(m^2+2ml)\sum_{i=0}^{l}(-1)^i\frac{((m-1)+l+i)!((m-1)+l-i)!}{(l+i)!(l-i)!(n-l+i)!(n-l-i)!}$$
$$+\sum_{i=0}^{l-1}(-1)^i\frac{(m+(l-1)+i)!(m+(l-1)-i)!}{((l-1)+i)!((l-1)-i)!((n-1)-(l-1)+i)!((n-1)-(l-1)-i)!}.$$

Therefore, Z obeys the partial difference equation

$$Z(m,n,l)=(m^2+2ml)Z(m-1,n,l)+Z(m,n-1,l-1).\tag{2}$$

We know that $n\geqq l$. Thus, in order to set the initial conditions for (2), we need to compute $Z(\tilde{m},\tilde{n},0)$ and $Z(0,\tilde{n},\tilde{l})$ for every $\tilde{n}$, $\tilde{m}\leqq\tilde{n}$ and $\tilde{l}\leqq[\tilde{n}/2]$.

Obviously, direct substitution into (1) gives $Z(\tilde{m}, \tilde{n}, 0)=(\tilde{m}!/\tilde{n}!)^2$, in full agreement with the argument of the lemma.

On the other hand

$$Z(0, \tilde{n}, \tilde{l})=\sum_{i=0}^{\tilde{l}}(-1)^i \frac{1}{(\tilde{n}-\tilde{l}+i)!(\tilde{n}-l-i)!}=\frac{1}{(2\tilde{n}-2\tilde{l})!}\sum_{i=0}^{\tilde{l}}(-1)^i\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-\tilde{l}+i}$$

$$=\frac{(-1)^{\tilde{l}}}{(2\tilde{n}-2\tilde{l})!}\sum_{i=0}^{\tilde{l}}(-1)^i\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-i}.$$

Thus, in order to prove the argument of the lemma in this case, it suffices to verify that

$$\sum_{i=0}^{\tilde{l}}(-1)^i\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-i}=\frac{1}{2}\frac{(2\tilde{n}-2\tilde{l})!}{(\tilde{n}-\tilde{l})!}\left[\frac{(-1)^{\tilde{l}}}{(\tilde{n}-\tilde{l})!}+\frac{(\tilde{n}-1-\tilde{l})!}{\tilde{n}!(\tilde{n}-1-2\tilde{l})!}\right]$$

$$=(-1)^{\tilde{l}}\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-\tilde{l}}+\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}}.$$

Indeed,

$$\begin{aligned}
&\sum_{i=0}^{\tilde{l}}(-1)^i\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-i}\\
&=\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}}+(-1)^{\tilde{l}}\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-\tilde{l}}+\sum_{i=1}^{\tilde{l}-1}(-1)^i\left[\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-1}+\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-i-1}\right]\\
&=\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}}+(-1)^{\tilde{l}}\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-\tilde{l}}+\sum_{i=1}^{\tilde{l}-1}(-1)^i\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-i}-\sum_{i=2}^{\tilde{l}}(-1)^i\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-i}\\
&=\left\{\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}}-\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-1}\right\}+(-1)^{\tilde{l}}\left\{\binom{2\tilde{n}-2\tilde{l}}{\tilde{n}-\tilde{l}}-\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-\tilde{l}}\right\}\\
&=\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}}+(-1)^{\tilde{l}}\binom{2\tilde{n}-2\tilde{l}-1}{\tilde{n}-\tilde{l}},
\end{aligned}$$

and the case $Z(0, \tilde{n}, \tilde{l})$ is verified.

Now we can apply the induction argument: let's suppose that we proved the lemma for $Z(m-1, n, l)$ and for $Z(m, n-1, l-1)$. Substituting in (2), we get

$$\begin{aligned}
Z(m, n, l)=\frac{1}{2}\Bigg\{&(m^2+2ml)\frac{(m-1+l)!}{l!(n-l)!}\left[\frac{(m-1+l)!}{l!(n-l)!}+(-1)^l\frac{(m-1)!(n-m-l)!}{n!(n-m-2l)!}\right]\\
&+\frac{(m+l-1)!}{(l-1)!(n-l)!}\left[\frac{(m+l-1)!}{(l-1)!(n-l)!}-(-1)^l\frac{m!(n-m-1-l)!}{(n-1)!(n-m-2l)!}\right]\Bigg\}\\
=\frac{1}{2}\Bigg\{&\left[\frac{(m+l-1)!}{l!(n-l)!}\right]^2(m^2+2ml+l^2)\\
&+(-1)^l\frac{(m-1+l)!(m-1)!(n-m-1-l)!}{l!(n-l)!n!(n-m-2l)!}(m^2n+mnl-m^3-3m^2l-2ml^2)\Bigg\}\\
=\frac{1}{2}\Bigg\{&\left[\frac{(m+l-1)!(m+l)}{l!(n-l)!}\right]^2\\
&+(-1)^l\frac{(m-1+l)!(m-1)!(n-m-1-l)!}{l!(n-l)!n!(n-m-2l)!}m(m+l)(n-m-2l)\Bigg\}\\
=\frac{1}{2}&\frac{(m+l)!}{l!(n-l)!}\left[\frac{(m+l)!}{l!(n-l)!}+(-1)^l\frac{m!(n-m-1-l)!}{n!(n-m-1-2l)!}\right]
\end{aligned}$$

Q.E.D.

LEMMA 2. *When* $n-m\equiv 0$ (mod. 4) *or* $n-m\equiv 3$ (mod. 4), $n\geqq m+1$, *the Padé approximation* $R_{n,m}$ *is not A-acceptable*

Proof. If $n-m=4k$, $k\geqq 1$, then, according to Lemma 1,

$$\omega_{n,m}(t)=-2k\frac{m!}{(m+4k)!(m+2k)(m+2k+1)}t^{2m+4k+2}+\mathcal{O}(t^{2m+4k+4})$$

and if $n-m=4k+3$, $k\geqq 0$, then

$$\omega_{n,m}(t)=-\frac{m!}{(m+4k+3)!(m+2k+2)}t^{2m+4k+4}+\mathcal{O}(t^{2m+4k+6}).$$

In both cases, for $t\neq 0$ close enough to zero, we get $\omega_{n,m}(t)<0$, violating the condition B.

Thus, the Padé approximations $R_{n,m}$, for $n\geqq m+1$, $n-m\equiv 0$ or 3 (mod. 4), are not A-acceptable. Q.E.D.

LEMMA 3. *The Padé approximation* $R_{n,m}$, *for* $n\geqq m+5$, $n-m\equiv 1$ (mod. 4) *is not A-acceptable.*

Proof. If $n=m+4k+1$, $k\geqq 1$, we set

$$\hat{s}:=\left[6\frac{(m+2k-1)(m+2k+3)}{(2k-1)(2k+2)}\right]^{1/2}. \tag{3}$$

We shall show that $\omega_{n,m}(\hat{s})<0$:

$$\begin{aligned}\omega_{n,m}(\hat{s})&=\sum_{l=m+2k+1}^{n}A_l^{(n,m)}\hat{s}^{2l}\\&=\hat{s}^{2m+4k+2}\sum_{l=0}^{2k}A_{m+2k+1+l}^{(n,m)}\hat{s}^{2l}\\&=\hat{s}^{2m+4k+2}\Big\{A_{m+2k+1}^{(n,m)}+A_{m+2k+2}^{(n,m)}\hat{s}^2+A_{m+2k+3}^{(n,m)}\hat{s}^4\\&\qquad+\sum_{l=1}^{k-1}\hat{s}^{4l+2}\big(A_{m+2k+2l+2}^{(n,m)}+A_{m+2k+2l+3}^{(n,m)}\hat{s}^2\big)\Big\}.\end{aligned}$$

Thus it suffices to show that

$$A_{m+2k+1}^{(n,m)}+A_{m+2k+2}^{(n,m)}\hat{s}^2+A_{m+2k+3}^{(n,m)}\hat{s}^4<0, \tag{i}$$

$$A_{m+2k+2l+2}^{(n,m)}+A_{m+2k+2l+3}^{(n,m)}\hat{s}^2<0\quad\text{for every } 1\leqq l\leqq k-1. \tag{ii}$$

But, according to Lemma 1,

$$\begin{aligned}&A_{m+2k+1}^{(n,m)}+A_{m+2k+2}^{(n,m)}\hat{s}^2+A_{m+2k+4}^{(n,m)}\hat{s}^4\\&\quad=\frac{m!}{(m+4k+1)!}\Big\{\frac{1}{m+2k+1}-\frac{1}{2}\frac{(2k+1)!}{(2k-1)!}\frac{(m+2k-1)!}{(m+2k+2)!}\hat{s}^2\\&\qquad\qquad+\frac{1}{24}\frac{(2k+2)!}{(2k-2)!}\frac{(m+2k-2)!}{(m+2k+3)!}\hat{s}^4\Big\},\end{aligned}$$

and, substituting $\hat{s}$ as defined in (3), we get

$$A^{(n,m)}_{m+2k+1}+A^{(n,m)}_{m+2k+2}\hat{s}^2+A^{(n,m)}_{m+2k+4}\hat{s}$$

$$=\frac{m!}{(m+4k+1)!}\left\{\frac{1}{m+2k+1}-\frac{3}{2}\frac{(m+2k-1)(m+2k+3)!(2k+1)!(m+2k-1)!}{(2k-1)(2k+2)(2k-1)!(m+2k+2)!}\right\}$$

$$<\frac{m!}{(m+4k+1)!}\left\{\frac{1}{m+2k+1}-\frac{3}{2}\frac{1}{m+2k+1}\right\}<0,$$

and (i) is valid.

Now we verify (ii):

$$A^{(n,m)}_{m+2k+2p+2}+A^{(n,m)}_{m+2k+2p+3}\hat{s}^2$$

$$=c^2\{(2k+2p+1)(2k-2p-1)\hat{s}^2 \\ -2(m+2k-2p-1)(m+2k+2p+3)(4p+3)(p+1)\}$$

where

$$c^2=2\frac{m!(m+2k-2p-2)!(2k+2p+1)!}{(m+4k+1)!(2k-2p-1)!(m+2k+2p+3)!(4p+4)!}.$$

Thus, substituting the value of $\hat{s}$, it suffices to show that

$$3((m+2k+1)^2-4)((k+1)(2k-1)-3p-2p^2) \\ \leqq 2((m+2k+1)^2-4(p+1)^2)(4p^2+7p+3)(2k-1)(k+1)$$

for every $1\leqq p\leqq k-1$.

We get a quadratic inequality in m:

$$(m^2+2(2k+1)m+4k^2+4k-3)A_p\geqq B_p \tag{4}$$

where

$$A_p:=2(4p+3)(p+1)(2k-1)(k+1)-3(k+p+1)(2k-2p-1),$$

$$B_p:=8p(p+1)(p+2)(4p+3)(2k-1)(k+1).$$

A sufficient condition for (4) for every $m\geqq 0$ is $A_p>0$, $B_p\geqq((2k+1)^2-4)A_p$. But

$$A_p=(k+1)(2k-1)(8p^2+14p+3)+9p+6p^2>0$$

and $B_p\geqq((2k+1)^2-4)A_p$ is equivalent to the inequality

$$(2k-1)(k+1)[4k^2+4k-3-2p^2-4p]+(4k^2+4k-3)(9p+6p^2)\geqq 0,$$

which is obvious. Thus, (4) holds for every $m\geqq 0$, implying the validity of condition (ii) for every $1\leqq p\leqq k-1$.

Hence, indeed $\omega_{n,m}(\hat{s})<0$ and the Padé approximation $R_{n,m}$, where $n\geqq m+5$, $n-m\equiv 1 \pmod 4$ is not A-acceptable. Q.E.D.

So far, we verified the conjecture of Ehle for the case $n\geqq m+3$, $n-m\not\equiv 2 \pmod 4$. The objective of the next section is to explore the conjecture in the case $n-m\equiv 2$ (mod. 4).

The case $n-m\equiv 2$ (mod. 4). In the cases $n-m\not\equiv 2$ (mod. 4), $n\geqq m+3$ we proved the nonacceptability by showing that the condition B is violated. This is not necessarily true when $n-m\equiv 2$ (mod. 4). Trivial examples are $R_{6,0}$ and $R_{10,0}$. Nevertheless, these two approximations violate the condition A.

We have to point out that, generally speaking, the condition B is more essential for the possible disproving of the A acceptability. This is a consequence of theorem of Saff and Varga [4], which states that for every integer $m \geqq 0$ there exists an integer $\tau_m \geqq 0$ such that the polynomials $P_{m,n}$, $n \geqq \tau_m$, have all their zeros in the left half-plane. The implication of the condition A for $R_{n,m}$ is obvious. Thus, going far enough along any diagonal, we arrive at approximations which satisfy the condition A.

Computer tests show that in the cases $n = m+6$ and $n = m+10$ integers K_6 and K_{10} exist, so that the condition B is violated for $R_{m+6,m}$, $m \geqq K_6$ (resp. $R_{m+10,m}$, $m \geqq K_{10}$) and the K's are smaller than the respective τ's of Saff and Varga. Thus these two diagonals of the Padé tableau satisfy the Ehle conjecture.

These tests provide some indication for further research, aimed at the complete characterization of A-acceptable Padé approximations.

REFERENCES

[1] G. BIRKHOFF AND R. S. VARGA, *Discretization errors for well-set Cauchy problems*, I. J. Math. and Phys., 44 (1965), pp. 1–23.

[2] B. L. EHLE, *A-stable methods and Padé approximations to the exponential*, this Journal, 4 (1973), pp. 671–680.

[3] S. P. NØRSETT, *C-polynomials for rational approximation to the exponential function*, Numer. Math., 25 (1975), pp. 39–56.

[4] E. B. SAFF AND R. S. VARGA, *On the zeros and poles of Padé approximants to e^z*, Ibid., 25 (1975), pp. 1–14.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0011 $01.00/0

A SET OF ORTHOGONAL POLYNOMIALS THAT GENERALIZE THE RACAH COEFFICIENTS OR 6−j SYMBOLS*

RICHARD ASKEY† AND JAMES WILSON‡

Abstract. A very general set of orthogonal polynomials with five free parameters is given explicitly, the orthogonality relation is proved and the three term recurrence relation is found.

1. Introduction. A hypergeometric series has the form $\sum a_n$ with a_{n+1}/a_n a rational function of n. A basic hypergeometric series has a_{n+1}/a_n a rational function of q^n for a fixed q. The standard notation will be used. It is

$$ {}_rF_s\left(\begin{matrix} a_1, \cdots, a_r \\ b_1, \cdots, b_s \end{matrix}; \ x\right) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_r)_n}{(b_1)_n \cdots (b_s)_n} \frac{x^n}{n!}, \tag{1.1} $$

where

$$ (a)_n = \begin{cases} a(a+1)\cdots(a+n-1), & n=1,2,\cdots, \\ 1, & n=0, \end{cases} \tag{1.2} $$

for hypergeometric series and

$$ {}_{r+1}\varphi_r\left(\begin{matrix} a_1, \cdots, a_{r+1} \\ b_1, \cdots, b_r \end{matrix}; \ q, x\right) = \sum_{n=0}^{\infty} \frac{(a_1;q)_n \cdots (a_{r+1};q)_n x^n}{(b_1;q)_n \cdots (b_r;q)_n (q;q)_n} \tag{1.3} $$

with

$$ (a;q)_n = \begin{cases} (1-a)\cdots(1-aq^{n-1}), & n=1,2,\cdots, \\ 1, & n=0, \\ \dfrac{1}{(1-aq^{-n})\cdots(1-aq^{-1})}, & n=-1,-2,\cdots, \end{cases} \tag{1.4} $$

for basic hypergeometric series.

For readers who are unacquainted with basic hypergeometric series, observe that

$$ \lim_{q\to 1} \frac{(q^\alpha;q)_n}{(q^\beta;q)_n} = \frac{(\alpha)_n}{(\beta)_n}. $$

There are reasons for using $(a;q)_n$ in (1.3) rather than $(q^\alpha;q)_n$ which go beyond a desire for a notation that is easy to set in type. There are times when we want "a" to be negative, and we can only make q^α negative by taking α complex. It is possible to do this but unnecessary. Also there are times when we want "a" to be independent of q. Again it is possible to take $\alpha = (\log a)/(\log q)$ so that q^α is independent of q, but it is unnecessary if we use $(a;q)_n$ rather than $(q^\alpha;q)_n$.

In [12] it was pointed out that

$$ p_n(\lambda(x)) = {}_4F_3\left(\begin{matrix} -n, n+\alpha+\beta+1, -x, x+\gamma+\delta+1 \\ \alpha+1, \beta+\delta+1, \gamma+1 \end{matrix}; \ 1\right) \tag{1.5} $$

$$ \lambda(x) = x(x+\gamma+\delta+1) \tag{1.6} $$

* Received by the editors March 3, 1978, and in revised form May 5, 1978.

† Department of Mathematics, University of Wisconsin, Madison, Wisconsin 53706. The work of this author was sponsored by the United States Army under Contract DAAG29-75-C-0024.

‡ Department of Mathematics, Purdue University, West Lafayette, Indiana 47907.

is a polynomial of degree n in $\lambda(x)$ which is orthogonal on $x=0, 1, \cdots, N$ when $\alpha+1$, $\beta+\delta+1$ or $\gamma+1$ is $-N$. This orthogonality relation is equivalent to Racah's orthogonality for functions that are usually called Racah coefficients or $6-j$ symbols. We will call these polynomials Racah polynomials. Racah was unaware of the existence of these polynomials, but he was the first to find an orthogonality relation equivalent to the orthogonality relation for (1.5) which is given in [12]. Since almost all the classical orthogonal polynomials are named after a rediscoverer rather than the original discoverer, we felt it was appropriate to err in the opposite way and name the polynomials after the first person to treat them, even if he was unaware of the orthogonal polynomials buried in his results. These polynomials contain as limiting cases the classical polynomials of Jacobi, Laguerre and Hermite and their discrete analogues which go under the names of Hahn, Meixner, Krawtchouk and Charlier polynomials. All of these polynomials can be given as hypergeometric series. Since basic hypergeometric extensions of the classical polynomials have been found [8], [2] it is natural to look for a basic hypergeometric extension of (1.5). The right polynomials to consider are balanced ${}_4\varphi_3$'s

$$p_n(\mu(x); a, b, c, d; q) = p_n(\mu(x)) = {}_4\varphi_3\left(\begin{matrix} q^{-n}, q^{n+1}ab, q^{-x}, q^{x+1}cd \\ aq, bdq, cq \end{matrix}; q, q\right) \tag{1.7}$$

where

$$\mu(x) = q^{-x} + q^{x+1}cd. \tag{1.8}$$

Since

$$p_n(\mu(x)) = 1 + \sum_{k=1}^{n} \frac{(q^{-n}; q)_k (q^{n+1}ab; q)_k q^k}{(aq; q)_k (bdq; q)_k (cq; q)_k (q; q)_k} \prod_{j=0}^{k-1} [1 + q^{2j+1}cd - q^j(\mu(x))]$$

it is clear that $p_n(\mu(x))$ is a polynomial of degree n in the variable $\mu(x)$.

The adjective "balanced" refers to a condition put on the parameters. For basic hypergeometric series it means that the product of the numerator parameters times q is the product of the denominator parameters. In this case $q^{-n+n+1}abq^{-x+x+1}cdq = abcdq^3$.

2. Orthogonality. Assume that one of aq, cq or bdq is q^{-N}. Then the orthogonality relation is

$$\sum_{x=0}^{N} p_n(\mu(x); a, b, c, d; q) p_m(\mu(x); a, b, c, d; q) w(x) = 0, \qquad m \neq n, \quad 0 \leqq m, n \leqq N, \tag{2.1}$$

where

$$w(x) = \frac{(cdq; q)_x (1 - cdq^{2x+1})(aq; q)_x (bdq; q)_x (cq; q)_x (abq)^{-x}}{(q; q)_x (1 - cdq)(cda^{-1}q; q)_x (b^{-1}cq; q)_x (dq; q)_x}.$$

Observe that

$$\begin{aligned}(q^{-x}d^{-1}; q)_m (q^{x+1}c; q)_m &= \prod_{j=0}^{m-1} (1 - d^{-1}q^{-x+j})(1 - cq^{x+j+1}) \\ &= \prod_{j=0}^{m-1} (1 + cq^{2j+1}d^{-1} - d^{-1}q^j\mu(x))\end{aligned}$$

is a polynomial of degree m in $\mu(x)$. To prove (2.1) for $m\neq n$ it will suffice to show that

$$(2.2)\quad I=\sum_{x=0}^{N} p_n(\mu(x);a,b,c,d;q)(q^{-x}d^{-1};q)_m(q^{x+1}c;q)_m w(x)=0;\qquad m<n.$$

The advantage of (2.2) over (2.1) is that the polynomial of degree m can be attached to the weight function. Using the definition of $p_n(\mu(x))$ in (2.2) gives

$$\begin{aligned}
I&=\sum_{x=0}^{N}\sum_{k=0}^{n}\frac{(q^{-n};q)_k(q^{n+1}ab;q)_k(q^{-x};q)_k(q^{x+1}cd;q)_kq^k}{(aq;q)_k(bdq;q)_k(cq;q)_k(q;q)_k}\\
&\quad\cdot\frac{(q^{-x}d^{-1};q)_m(q^{x+1}c;q)_m(cdq;q)_x(1-cdq^{2x+1})(aq;q)_x(bdq;q)_x(cq;q)_x}{(q;q)_x(1-cdq)(a^{-1}cdq;q)_x(b^{-1}cq;q)_x(dq;q)_x(abq)^x}\\
&=\sum_{k=0}^{n}\sum_{x=k}^{N}\frac{(q^{-n};q)_k(q^{n+1}ab;q)_k(aq^{k+1};q)_{x-k}(bdq^{k+1};q)_{x-k}}{(q;q)_k(a^{-1}cdq;q)_x(b^{-1}cq;q)_x(q;q)_{x-k}}\\
&\quad\cdot\frac{(cq^{k+1};q)_{x-k+m}(cdq;q)_{x+k}(1-cdq^{2x+1})(-1)^{k+m}q^{k-x(k+m+1)+\binom{k}{2}+\binom{m}{2}}}{(dq;q)_{x-m}(1-cdq)(ab)^xd^m}\\
&=\frac{q^{\binom{m}{2}}(-1)^m}{d^m}\sum_{k=0}^{n}\frac{(q^{-n};q)_k(q^{n+1}ab;q)_kq^{k(k+1)/2}}{(q;q)_k(-1)^kq^{k(k+m+1)}(ab)^k}\cdot\sum_{x=0}^{N-k}\frac{(aq^{k+1};q)_x(bdq^{k+1};q)_x}{(a^{-1}cdq;q)_{x+k}}\\
&\quad\cdot\frac{(cq^{k+1};q)_{x+m}(cdq;q)_{x+2k}(1-cdq^{2x+2k+1})}{(b^{-1}cq;q)_{x+k}(q;q)_x(dq;q)_{x+k-m}(1-cdq)}\left(\frac{1}{abq^{k+m+1}}\right)^x\\
&=\frac{q^{\binom{m}{2}}(-1)^m}{d^m}\sum_{k=0}^{n}\frac{(q^{-n};q)_k(q^{n+1}ab;q)_k(cdq;q)_{2k}(cq^{k+1};q)_m(1-cdq^{2k+1})(-1)^kq^{-mk-k}}{(q;q)_k(a^{-1}cdq;q)_k(b^{-1}cq;q)_k(dq;q)_{k-m}(1-cdq)q^{\binom{k}{2}}(ab)^k}\\
&\quad\cdot\sum_{x=0}^{N-k}\frac{(cdq^{2k+1};q)_x(1-cdq^{2k+2x+1})(aq^{k+1};q)_x(bdq^{k+1};q)_x(cq^{k+m+1};q)_x}{(q;q)_x(1-cdq^{2k+1})(a^{-1}cdq^{k+1};q)_x(b^{-1}cq^{k+1};q)_x(dq^{k-m+1};q)_x}\\
&\quad\cdot\left(\frac{q^{-m-k-1}}{ab}\right)^x.
\end{aligned}$$

The sum on x can be evaluated since it is a very well poised ${}_6\varphi_5$. The required sum is

$$\begin{aligned}
(2.3)\quad {}_6\varphi_5&\left(\begin{matrix}a^2,\ aq,\ -aq,\ b,\ c,\ q^{-N}\\ a,\ -a,\ a^2b^{-1}q,\ a^2c^{-1}q,\ a^2q^{N+1}\end{matrix};q,\frac{a^2q^{N+1}}{bc}\right)\\
&=\frac{(a^2q;q)_N(a^2qb^{-1}c^{-1};q)_N}{(a^2qb^{-1};q)_N(a^2qc^{-1};q)_N}\\
&=\frac{(a^2q;q)_\infty(a^2qb^{-1}c^{-1};q)_\infty(a^2q^{N+1}b^{-1};q)_\infty(a^2q^{N+1}c^{-1};q)_\infty}{(a^2q^{N+1};q)_\infty(a^2q^{N+1}b^{-1}c^{-1};q)_\infty(a^2qb^{-1};q)_\infty(a^2qc^{-1};q)_\infty}.
\end{aligned}$$

A basic hypergeometric series

$${}_{p+1}\varphi_p\left(\begin{matrix}a_1,\cdots,a_{p+1}\\ b_1,\cdots,b_p\end{matrix};q,x\right)$$

is called well poised if

$$a_1q=a_2b_1=\cdots=a_{p+1}b_p.$$

It is very well poised if $a_2=qb_1$, $a_3=-a_2$, or equivalently, $a_2=qa_1^{1/2}$, $a_3=-qa_1^{1/2}$. The effect of the very in very well poised is to introduce $(1-a_1q^{2k})/(1-a_1)$ as a factor in the term involving x^k. Unfortunately this definition is not sufficiently precise to uniquely determine a well poised or a very well poised series, since nothing is said about the value of x. The usual value of x which leads to series that can be summed is

$$x=\left(\frac{qb_1\cdots b_p}{a_1a_2\cdots a_{p+1}}\right)^{1/2}.$$

That is its value in (2.3). However there are cases when a slightly different choice of x is the right one.

Up until now no assumptions on q have been made other than the implicit assumption that there are no zeros in the denominator. To make the calculations that follow a little easier we will assume $|q|<1$ and define $(a;q)_\infty$ by

$$(a;q)_\infty=\prod_{n=0}^{\infty}(1-aq^n). \tag{2.4}$$

Since we are only dealing with polynomials it is easy to remove this restriction on q. A proof of (2.3) using orthogonal polynomials is given in [1]. One needs to set $bc=aq$ in the formula in Theorem 12. A proof is also given in [10, (3.3, 1.4)]. However in the appendix in [10] this formula is given with some misprints.

Using (2.3) in I gives

$$I=\frac{q^{\binom{m}{2}}(-1)^m(cq;q)_m}{d^m}\sum_{k=0}^{n}\frac{(q^{-n};q)_k(q^{n+1}ab;q)_k(cdq;q)_{2k}(1-cdq^{2k+1})}{(q;q)_k(a^{-1}cdq;q)_k(b^{-1}cq;q)_k(1-cdq)}$$
$$\cdot\frac{(cq^{m+1};q)_k(cdq^{2k+2};q)_\infty(a^{-1}b^{-1}c;q)_\infty(a^{-1}dq^{-m};q)_\infty\cdot(b^{-1}q^{-m};q)_\infty(-1)^kq^{-(k^2/2)-(k/2)-mk}(ab)^{-k}}{(dq;q)_{k-m}(cq;q)_k(a^{-1}cdq^{k+1};q)_\infty(b^{-1}cq^{k+1};q)_\infty\cdot(dq^{k-m+1};q)_\infty(a^{-1}b^{-1}q^{-k-m-1};q)_\infty}$$
$$=\frac{q^{\binom{m}{2}}(-1)^m(a^{-1}b^{-1}c;q)_\infty(a^{-1}dq^{-m};q)_\infty(b^{-1}q^{-m};q)_\infty(cdq^2;q)_\infty(cq;q)_m}{d^m(a^{-1}cdq;q)_\infty(b^{-1}cq;q)_\infty(dq;q)_\infty(a^{-1}b^{-1}q^{-m-1};q)_\infty}$$
$$\cdot\sum_{k=0}^{n}\frac{(q^{-n};q)_k(q^{n+1}ab;q)_k(cq^{m+1};q)_kq^k}{(q;q)_k(abq^{m+2};q)_k(cq;q)_k}.$$

This sum is a balanced ${}_3\varphi_2$, and so can be summed using

$${}_3\varphi_2\left(\begin{matrix}q^{-n},aq^n,b\\ c,d\end{matrix};q,q\right)=\frac{(c/b;q)_n(d/b;q)_nb^n}{(c;q)_n(d;q)_n} \tag{2.5}$$

when $abq=cd$. The final result is

$$I=\frac{A(abc^{-1}q;q)_n(q^{-m};q)_n(cq^{m+1})^n}{(abq^{m+2};q)_n(cq;q)_n}$$

where A is the coefficient of the sum above. So $I=0$ for $m=0,1,\cdots,n-1$.

The value of the sum in (2.1) when $m=n$ can be found from this sum. However, it is easier to obtain it from results in the next section.

3. Recurrence relation. If $p_n(x)$ are orthogonal with respect to a positive measure then

$$xp_n(x)=A_np_{n+1}(x)+B_np_n(x)+C_np_{n-1}(x);\qquad p_{-1}(x)\equiv0. \tag{3.1}$$

If the measure has infinitely many points of support then (3.1) holds for $n=0,1,\cdots$.

When there are only finitely many point masses, say $N+1$, then (3.1) holds for $n=0, 1, \cdots, N-1$, and when $n=N$ the zeros of $p_{N+1}(x)$ determine the location of the point masses. For a proof of this old fact see [3]. It is implicit in some of Chebychev's work on continued fractions. We have shown that $\{p_n(\mu(x))\}$ is orthogonal, so (3.1) becomes

$$\mu(x)p_n(\mu(x)) = A_n p_{n+1}(\mu(x)) + B_n p_n(\mu(x)) + C_n p_{n-1}(\mu(x)). \tag{3.2}$$

When $x=0$, $p_n(\mu(0))=1$, so (3.2) can be written as

$$\begin{aligned}[\mu(x)-\mu(0)]p_n(\mu(x)) &= A_n[p_{n+1}(\mu(x))-p_n(\mu(x))] \\ &\quad - C_n[p_n(\mu(x))-p_{n-1}(\mu(x))].\end{aligned} \tag{3.3}$$

A_n is determined by equating the highest powers of $\mu(x)$. It is

$$A_n = \frac{(1-abq^{n+1})(1-aq^{n+1})(1-bdq^{n+1})(1-cq^{n+1})}{(1-abq^{2n+1})(1-abq^{2n+2})} \tag{3.4}$$

since

$$p_n(\mu(x)) = \frac{(q^{-n};q)_n(q^{n+1}ab;q)_n q^n(-1)^n q^{\binom{n}{2}}}{(aq;q)_n(bdq;q)_n(cq;q)_n(q;q)_n}[\mu(x)]^n + \text{lower terms.} \tag{3.5}$$

The easiest way to find C_n is to first simplify (3.3). A routine calculation gives

$$\begin{aligned} &p_{n+1}(\mu(x)) - p_n(\mu(x)) \\ &= \frac{-q^{-n}(1-q^{2n+2}ab)(1-q^{-x})(1-q^{x+1}cd)}{(1-aq)(1-bdq)(1-cq)} \\ &\quad \cdot {}_4\varphi_3\left(\begin{matrix} q^{-n}, q^{n+2}ab, q^{-x+1}, q^{x+2}cd \\ aq^2, bdq^2, cq^2 \end{matrix}; q, q\right). \end{aligned} \tag{3.6}$$

So (3.3) can be rewritten as

$$\begin{aligned} &-{}_4\varphi_3\left(\begin{matrix} q^{-n}, q^{n+1}ab, q^{-x}, q^{x+1}cd \\ aq, bdq, cq \end{matrix}; q, q\right) \\ &= \frac{-A_n q^{-n}(1-q^{2n+2}ab)}{(1-aq)(1-bdq)(1-cq)} {}_4\varphi_3\left(\begin{matrix} q^{-n}, q^{n+2}ab, q^{-x+1}, q^{x+2}cd \\ aq^2, bdq^2, cq^2 \end{matrix}; q, q\right) \\ &\quad + \frac{C_n q^{-n+1}(1-q^{2n}ab)}{(1-aq)(1-bdq)(1-cq)} {}_4\varphi_3\left(\begin{matrix} q^{-n+1}, q^{n+1}ab, q^{-x+1}, q^{x+2} \\ aq^2, bdq^2, cq^2 \end{matrix}; q, q\right). \end{aligned} \tag{3.7}$$

Now there are a couple of ways to proceed. If $x=1$ then all the ${}_4\varphi_3$'s can be evaluated, but the reduction is more complicated than it has to be. Another way is to set $q^{-x}=aq$ and use (2.5) on all the series. This calculation gives

$$C_n = \frac{cdq(1-q^n)(1-bq^n)(1-abc^{-1}q^n)(1-ad^{-1}q^n)}{(1-abq^{2n+1})(1-abq^{2n})}. \tag{3.8}$$

Formula (3.2) is an analogue for balanced ${}_4\varphi_3$'s of one of the contiguous relations of Gauss. Set

$$\varphi = {}_4\varphi_3\left(\begin{matrix} a, b, c, d \\ e, f, g \end{matrix}; q, q\right)$$

where $abcdq = efg$ and one of a, b, c or d is q^{-n}. Also set

$$\varphi(a+, b-) = {}_4\varphi_3\left(\begin{matrix} aq, bq^{-1}, c, d \\ e, f, g \end{matrix}; q, q\right).$$

Then (3.2) becomes

$$\begin{aligned} &b(1-b)(a-e)(a-f)(a-g)(aq-b)\varphi(a-, b+) \\ (3.9) \qquad &+\begin{bmatrix} ab(1-c)(1-d)(aq-b)(a-bq)(b-a) \\ -b(1-b)(a-e)(a-f)(a-g)(aq-b) \\ +a(1-a)(e-b)(f-b)(g-b)(a-bq) \end{bmatrix}\varphi \\ &-a(1-a)(e-b)(f-b)(g-b)(a-bq)\varphi(a+, b-) = 0, \end{aligned}$$

or

$$\begin{aligned} &b(1-b)(a-e)(a-f)(a-g)(aq-b)[\varphi(a-, b+)-\varphi] \\ (3.10) \qquad &+a(1-a)(e-b)(f-b)(g-b)(a-bq)[\varphi-\varphi(a+, b-)] \\ &+ab(1-c)(1-d)(aq-b)(a-bq)(b-a)\varphi = 0. \end{aligned}$$

To find the sum of (2.1) when $m = n$, call it h_n. Then

$$A_n h_{n+1} = \sum_x \mu(x) p_n(\mu(x)) p_{n+1}(\mu(x)) w(x)$$

and

$$\begin{aligned} C_n h_{n-1} &= \sum_x \mu(x) p_n(\mu(x)) p_{n-1}(\mu(x)) w(x) \\ &= A_{n-1} h_n \end{aligned}$$

so

$$(3.11) \qquad h_n = \frac{C_n}{A_{n-1}} h_{n-1} = \cdots = \frac{C_n \cdots C_1}{A_{n-1} \cdots A_0} h_0.$$

The ${}_6\varphi_5$ sum (2.3) gives

$$(3.12) \qquad h_0 = \frac{(cdq^2; q)_\infty \left(\frac{c}{ab}; q\right)_\infty \left(\frac{d}{a}; q\right)_\infty \left(\frac{1}{b}; q\right)_\infty}{\left(\frac{cdq}{a}; q\right)_\infty \left(\frac{cq}{b}; q\right)_\infty (dq; q)_\infty \left(\frac{1}{abq}; q\right)_\infty}$$

and so

$$h_n = \frac{(q; q)_n (1-abq)(bq; q)_n (ad^{-1}q; q)_n (abc^{-1}q; q)_n (cdq)^n}{(abq; q)_n (1-abq^{2n+1})(aq; q)_n (bdq; q)_n (cq; q)_n} h_0.$$

Once this formula has been found some of the mystery of § 2 can be removed. It is natural to ask where the weight function came from. Observe that

$${}_4\varphi_3\left(\begin{matrix} q^{-n}, q^{n+1}ab, q^{-x}, q^{x+1}cd \\ aq, bdq, cq \end{matrix}; q, q\right)$$

is symmetric in n and x when (a, b) is changed into (c, d). This symmetry carries over to $w(x)$ and h_n, that is $w(n)$ is just h_0/h_n with (a, b) interchanged with (c, d). The reason for this is that a matrix that is orthogonal by rows is also orthogonal by columns. The usefulness of this remark was mentioned by Karlin and McGregor [9] in connection with the Hahn and dual Hahn polynomials. Also see Eagleson [6]. In fact this is how we found the weight function. However we could not give a proof by this method without first proving the recurrence relation (3.3) directly. This would be very tedious, so it is preferable to prove the orthogonality directly.

To show from the recurrence relation that the masses must be located at $x = 0, 1, \cdots, N$, observe first that $A_N = 0$, since one of aq, bdq and cq was assumed to be q^{-N}. For definiteness take $bdq = q^{-N}$. The other cases are handled in a similar fashion. Formula (3.3) will hold when $n = N$ if we can show that

$$(1-q^{-x})(1-q^{x+1}cd)p_N(\mu(x)) = C_N[p_N(\mu(x)) - p_{N-1}(\mu(x))].$$

Both sides vanish when $x = 0$, since $p_n(\mu(0)) = 1$. Use (3.6) on $p_N(\mu(x)) - p_{N-1}(\mu(x))$ and the value of C_N given in (3.8), to see that this is equivalent to

$$\begin{aligned}&{}_3\varphi_2\left(\begin{matrix} q^{N+1}ab, q^{-x}, q^{x+1}cd \\ aq, cq \end{matrix}; q, q\right)\\ &= \frac{cdq(1-q^N)(1-bq^N)(1-abc^{-1}q^N)(1-ad^{-1}q^N)}{(1-abq^{2N})(1-abq^{2N+1})}\\ &\quad\cdot \frac{(-q^{1-N})(1-abq^{2N})}{(1-aq)(1-bdq)(1-cq)}\, {}_3\varphi_2\left(\begin{matrix} q^{N+1}ab, q^{1-x}, q^{x+2}cd \\ aq^2, cq^2 \end{matrix}; q, q\right),\end{aligned} \tag{3.13}$$

when $x = 1, 2, \cdots, N$. The series is terminated by q^{1-x}, so it is correct to replace the factors $(q^{-N}; q)_k/(q^{-N}; q)_k$ by 1, since they do not vanish. Since $x = 1, 2, \cdots, N$ the series in (3.13) can be summed using (2.5). Again a simple calculation shows that (3.13) holds for $x = 1, 2, \cdots, N$. Thus the recurrence relation (3.2) holds when $n = N$ and $x = 0, 1, \cdots, N$. Therefore the point masses must be located at $x = 0, 1, \cdots, N$.

4. Summary and miscellaneous results. For ease of reference we state the two main results again:

$$\sum_{x=0}^{N} p_n(\mu(x); a, b, c, d; q)p_m(\mu(x); a, b, c, d; q)w(x) = \delta_{m,n}h_n, \quad aq, bdq \quad \text{or} \quad cq = q^{-N}, \tag{4.1}$$

where

$$p_n(\mu(x); a, b, c, d; q) = {}_4\varphi_3\left(\begin{matrix} q^{-n}, q^{n+1}ab, q^{-x}, q^{x+1}cd \\ aq, bdq, cq \end{matrix}; q, q\right), \tag{4.2}$$

$$\mu(x) = q^{-x} + q^{x+1}cd, \tag{4.3}$$

$$w(x) = \frac{(cdq; q)_x(1-cdq^{2x+1})(aq; q)_x(bdq; q)_x(cq; q)_x}{(q; q)_x(1-cdq)(a^{-1}cdq; q)_x(b^{-1}cq; q)_x(dq; q)_x(abq)^x}, \tag{4.4}$$

$$\begin{aligned}h_n &= \frac{(q; q)_n(1-abq)(bq; q)_n(ad^{-1}q; q)_n(abc^{-1}q; q)_n(cdq)^n}{(abq; q)_n(1-abq^{2n+1})(aq; q)_n(bdq; q)_n(cq; q)_n}\\ &\quad\cdot \frac{(cdq^2; q)_\infty(a^{-1}b^{-1}c; q)_\infty(a^{-1}d; q)_\infty(b^{-1}; q)_\infty}{(a^{-1}cdq; q)_\infty(b^{-1}cq; q)_\infty(dq; q)_\infty(a^{-1}b^{-1}q^{-1}; q)_\infty}.\end{aligned} \tag{4.5}$$

These infinite products look like they must have $|q| < 1$ before they make sense. However, since one of aq, bdq or cq is q^{-N} these products all reduce to finite products.

For example, when $aq = q^{-N}$, then

$$\frac{(cdq^2;q)_\infty(a^{-1}b^{-1}c;q)_\infty(a^{-1}d;q)_\infty(b^{-1};q)_\infty}{(a^{-1}cdq;q)_\infty(b^{-1}cq;q)_\infty(dq;q)_\infty(a^{-1}b^{-1}q^{-1};q)_\infty}$$
$$=\frac{(cdq^2;q)_\infty(b^{-1}cq^{N+1};q)_\infty(dq^{N+1};q)_\infty(b^{-1};q)_\infty}{(cdq^{2+N};q)_\infty(b^{-1}cq;q)_\infty(dq;q)_\infty(b^{-1}q^N;q)_\infty}$$
$$=\frac{(cdq^2;q)_N(b^{-1};q)_N}{(b^{-1}cq;q)_N(dq;q)_N}.$$

$$\begin{aligned}&-(1-q^{-x})(1-q^{x+1}cd)p_n(\mu(x);a,b,c,d;q)\\ &\quad= A_n p_{n+1}(\mu(x);a,b,c,d;q)-(A_n+C_n)p_n(\mu(x);a,b,c,d;q)\\ &\qquad+C_n p_{n-1}(\mu(x);a,b,c,d;q),\end{aligned} \tag{4.6}$$

where $p_{-1}(\mu(x);a,b,c,d;q)\equiv 0$ and

$$A_n=\frac{(1-abq^{n+1})(1-aq^{n+1})(1-bdq^{n+1})(1-cq^{n+1})}{(1-abq^{2n+1})(1-abq^{2n+2})}, \tag{4.7}$$

$$C_n=\frac{q(1-q^n)(1-bq^n)(c-abq^n)(d-aq^n)}{(1-abq^{2n})(1-abq^{2n+1})}. \tag{4.8}$$

When aq, bdq or cq is q^{-N} then (4.6) holds for $n=0,1,\cdots,N-1$ and all x if the basic hypergeometric series that define $p_n(\mu(x))$ are assumed to terminate so that $p_n(\mu(x))$ is a polynomial of degree n, and holds when $n=N$ when $x=0,1,\cdots,N$. When none of aq, bdq or cq is equal to q^{-N} then (4.6) holds for all x for $n=0,1,\cdots$.

For the polynomials $p_n(\mu(x))$ to be orthogonal with respect to a positive measure it is necessary and sufficient that

$$A_{n-1}C_n>0, \tag{4.9}$$

See [7, Chap. II, Thm. 1.5]. If (4.9) holds for $n=1,2,\cdots$, then the measure has infinitely many points of support; when it holds for $n=1,2,\cdots,N$ then the measure can be taken to have support on $N+1$ points. In this paper we have only considered some cases when the measure is purely discrete and is supported on a finite set of points. In a later paper we will treat the general cases where the measure has both an absolutely continuous part and a discrete part.

There are many special cases of the orthogonality relation (4.1) which are interesting. The polynomials with $d=0$ and $cq=q^{-N}$ were discovered by Hahn, and their weight function was found a couple of years ago by Andrews and Askey. Delsarte [4], Dunkl [5] and Stanton [11] have considered special cases of these polynomials. The polynomials are called dual Hahn polynomials when $b=0$. The orthogonality when $aq=q^{-N}$ was also found by Andrews and Askey.

Another interesting special case is Stanton's q-analogue of the Krawtchouk polynomials. These are

$$K_n(q^{-x};c,q^{-N-1};q)={}_3\varphi_2\left(\begin{matrix}q^{-n},q^{-x},-fq^{k+1}\\ 0,q^{-N}\end{matrix};q,q\right).$$

To obtain these from the q-Racah polynomials (1.7) first set $c=d=0$ and $aq=q^{-N}$, then set $b=-q^{N+1}f^{-1}$ so $ab=-f^{-1}$. The weight function is then

$$w(x)=\frac{(q^{-N};q)_x}{(q;q)_x}\left(-\frac{f}{q}\right)^x.$$

When $q \to 1$ this converges to $((-N)_x/x!)f^x(-1)^x = \binom{N}{x} f^x$, which is the weight function for the Krawtchouk polynomials, when $f = p/(1-p)$.

A word of caution about characterization theorems needs to be said. There are many theorems that say "the classical polynomials are the only polynomials to have a given property". Such theorems are often misleading. For example, Eagleson [6] showed that the Charlier, Krawtchouk and Meixner polynomials are the only polynomials that are self dual. He is able to prove this theorem and yet miss the polynomials $p_n(\mu(x); a, b, a, b)$, which are clearly symmetric in n and x because his definition of self dual or symmetrizable is too restricted. A characterization theorem that leads to new orthogonal polynomials is usually interesting, one that says the classical polynomials are the only polynomials with a given property is usually much less interesting and if it keeps people from looking for new polynomials it is harmful.

REFERENCES

[1] G. ANDREWS AND R. ASKEY, *Enumeration of partitions: The role of Eulerian series and q-orthogonal polynomials*, Higher Combinatorics, M. Aigner, D. Reidel, eds., Dordrecht, Boston, 1977, pp. 3–26.

[2] ———, *Some basic hypergeometric analogues of the classical orthogonal polynomials and applications*, to appear.

[3] C. DE BOOR AND G. H. GOLUB, *The numerically stable reconstruction of a Jacobi matrix from spectral data*, Linear Algebra and Appl., 21 (1978), pp. 245–260.

[4] P. DELSARTE, *Bilinear forms over a finite field, with applications to coding theory*, MBLE Report R327, April 1976.

[5] C. DUNKL, *An addition theorem for some q-Hahn polynomials*, Monatsh. Math., 85 (1978), p. 537.

[6] G. K. EAGLESON, *A duality relation for discrete orthogonal polynomials*, Studia Scient. Math. Hung., 3 (1968), pp. 127–136.

[7] G. FREUD, *Orthogonal Polynomials*, Pergamon Press, Oxford, 1971.

[8] W. HAHN, *Uber Orthogonalpolynome, die q-Differenzengleichungen genügen*, Math. Nach., 2 (1949), pp. 4–34.

[9] S. KARLIN AND J. L. MCGREGOR, *The Hahn polynomials, formulas and an application*, Scripta Math., 26 (1961), pp. 33–46.

[10] L. J. SLATER, *Generalized Hypergeometric Functions*, Cambridge University Press, Cambridge, 1966.

[11] D. STANTON, *Some basic hypergeometric polynomials arising from finite classical groups*, Ph.D. thesis, Univ. Wisconsin, Madison, 1977.

[12] J. WILSON, *Hypergeometric series recurrence relations and some new orthogonal functions*, Ph.D. thesis, Univ. Wisconsin, Madison, 1978.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0012 $01.00/0

ON AN ABSTRACT INTEGRAL EQUATION*

GUSTAF GRIPENBERG†

Abstract. The existence of solutions of the nonlinear Volterra equation

$$u(t)+\int_0^t k(t-s)gu(s)\,ds \ni f(t)$$

is studied in a real Hilbert space. The nonlinear operator g is assumed to be the subdifferential of a convex function. The results obtained extend earlier ones by Barbu (SIAM J. Math. Anal., 1975), Londen (SIAM J. Math. Anal., 1977) and Londen and Staffans (Proc. Amer. Math. Soc., 1978).

1. Introduction and statement of results. The nonlinear Volterra equation

$$u(t)+\int_0^t k(t-s)gu(s)\,ds \ni f(t), \qquad t\in R_+=[0,\infty), \tag{1.1}$$

is studied in a real Hilbert space H. Here k, g and f are given and u is the unknown function. The kernel k is real-valued and f maps R_+ into H. The nonlinear mapping g is assumed to be the subdifferential of a convex, lower semicontinuous function $\varphi: H\to(-\infty,\infty]$ and is hence maximal monotone. We are interested in existence and uniqueness of solutions of (1.1). This problem has been studied by Barbu [1], Londen [6], Londen and Staffans [7] in the case $g=\partial\varphi$, and by the author [4], as well as independently by Crandall and Nohel [3] (using a different approach), for more general operators g (maximal monotone on H or m-accretive on a real Banach space X). In these latter papers the nonlinear operator can also depend explicitly on t.

In this paper we obtain the following existence and uniqueness result for (1.1) which is not covered by earlier work.

THEOREM. *Assume that*

$$k\in W^{1,1}_{\text{loc}}(R_+;R), \tag{1.2}$$

$$k(0)>0, \tag{1.3}$$

(1.4) *there exists* $T_0>0$ *and* $c_0>0$ *such that if* $0<t<T_0$; *then* $\operatorname{var}(k';[t,T_0])\leqq c_0\log t^{-1}$,

(1.5) $g=\partial\varphi$ *where* φ *is a proper, convex, lower semicontinuous function*: $H\to(\infty,\infty]$,

$$f\in W^{1,2}_{\text{loc}}(R_+;H), \tag{1.6}$$

$$f(0)\in D(\varphi). \tag{1.7}$$

Then there exists a unique function $u: R_+\to H$ *such that*

$$u\in W^{1,2}_{\text{loc}}(R_+;H), \tag{1.8}$$

(1.9) *there exists a function* $w\in L^2_{\text{loc}}(R_+;H)$ *such that*

(1.10) $u(t)\in D(g), \qquad w(t)\in gu(t)$ *a.e.* $t\in R_+$

and

$$u(t)+\int_0^t k(t-s)w(s)\,ds=f(t), \qquad t\in R_+. \tag{1.11}$$

---

* Received by the editors October 14, 1977, and in revised form March 24, 1978.

† Institute of Mathematics, Helsinki University of Technology, SF-02150 Espoo 15, Finland.

Here $W^{1,2}_{\mathrm{loc}}(R_+; H)$ denotes the set $\{u \mid u \in L^2_{\mathrm{loc}}(R_+; H),\ u' \in L^2_{\mathrm{loc}}(R_+; H)\}$ where $u' = du/dt$ is the distributional derivative.

The novelty of this theorem lies in the fact that it allows $\lim_{t\to 0+} (k(t) - k(0))t^{-1} = +\infty$, a case excluded in [1], [3], [4], [6], [7], [9] (for further details see [5]).

Note that the assumption (1.4) is crucial for the proof. If $H = R^n$, $n \geqq 1$ then the uniqueness assertion of Theorem 1 is still of some interest since g is not assumed to be strictly monotone.

The difficult step in the proof of the existence consists in showing that certain approximate solutions converge. To accomplish this we use a technique of inverting the kernel that has earlier been used in [3], [4], [9].

Finally we remark that one could also use the fixed-point technique developed in [3], [4] to prove this theorem, but the main parts of the proof would still be essentially the same.

2. Proof of the theorem. We study the approximating equation

$$u_\lambda(t) + \int_0^t k(t-s) g_\lambda u_\lambda(s)\, ds = f(t), \qquad t \in R_+, \tag{2.1}$$

where $\lambda > 0$ and g_λ is the Yosida approximation of g (i.e., $g_\lambda = \lambda^{-1}(I - J_\lambda)$, $J_\lambda = (I + \lambda g)^{-1}$). It is easy to see that under the assumptions of the theorem the equation (2.1) has a unique solution u_λ which satisfies

$$\sup_{\lambda > 0} \|u'_\lambda\|_{L^2(0,T;H)} < \infty, \tag{2.2}$$

$$\sup_{\lambda > 0} \|g_\lambda u_\lambda\|_{L^2(0,T;H)} < \infty \tag{2.3}$$

for any finite T (see e.g. [6]). If one can show that the functions u_λ converge uniformly on some interval $[0, T]$ to a function u then it is rather easy to deduce that this u is a local solution of (1.1). But before we can establish any convergence properties we must establish a certain local result for resolvents of Volterra equations, (for other properties of these resolvents see e.g. [8]).

LEMMA 2.1. *Assume that $T_0 > 0$ and that a and b are real functions on $(0, T_0]$ such that*

$$a \text{ is positive, nonincreasing on } (0, T_0] \text{ and } a \in L^2(0, T_0; R) \tag{2.4}$$

and

$$\operatorname{var}(b; [t, T_0]) \leqq a(t) \quad \text{on } (0, T_0]. \tag{2.5}$$

Then there exists T_1, $0 < T_1 \leqq T_0$ and a constant c_1 such that the function r defined by

$$r(t) + \int_0^t b(t-s) r(s)\, ds = a(t) \quad \text{on } (0, T_0] \tag{2.6}$$

satisfies

$$\operatorname{var}(r; [t, T_1]) \leqq c_1 a(t) \quad \text{on } (0, T_1]. \tag{2.7}$$

Proof. Let the sequence $\{b_m\}_{m=1}^\infty$ be such that $b_m \in C^1(0, T_0; R)$, $m \geqq 1$,

$$b_m \to b \quad \text{in } L^1(0, T_0; R) \text{ as } m \to \infty \tag{2.8}$$

and

$$\int_t^{T_0} |b'_m(s)|\, ds \leqq a(t) \quad \text{on } (0, T_0], \quad m \geqq 1. \tag{2.9}$$

This is possible by (2.4) and (2.5). Define the functions r_m, $m \geqq 1$, by

$$r_m(t) + \int_0^t b_m(t-s) r_m(s)\, ds = b_m(t) \quad \text{on } [0, T_0]. \tag{2.10}$$

Now it is easy to see that $r_m \in C^1(0, T_0; R)$ and it follows from (2.8) and (2.10) that $r_m \to r$ in $L^1(0, T_0; R)$ as $m \to \infty$. Hence it is sufficient to establish (2.7) with r replaced by r_m for all m. Choose T_1 such that

$$\int_0^{T_1} |b_m(s)|\, ds \leqq \tfrac{1}{2}, \qquad m \geqq 1. \tag{2.11}$$

Fix m. It follows from (2.4) and (2.9) that there exists a constant c_2 (independent of m) such that

$$|b_m(t)| \leqq c_2 a(t) \quad \text{on } (0, T_1]. \tag{2.12}$$

Now (2.4), (2.9), (2.10) and (2.12) yield that for some constant c_3 we have $\|b_m\|_{L^2(0,T_1;R)} \leqq c_3$ and $\|r_m\|_{L^2(0,T_1;R)} \leqq c_3$. Consequently we get from (2.10) by Schwarz inequality that

$$|r_m(t)| \leqq c_3^2 + |b_m(t)| \quad \text{on } (0, T_1]. \tag{2.13}$$

Let t_0, $0 < t_0 \leqq T_1$, be arbitrary and define the continuously differentiable functions d and e on $[t_0, T_1]$ by

$$d(t) = \int_0^{t_0} b_m(t-s) r_m(s)\, ds, \qquad e(t) = \int_{t_0}^t b_m(t-s) r_m(s)\, ds. \tag{2.14}$$

By use of (2.9), (2.12), (2.13), Fubini's theorem and Schwarz inequality, it is not difficult to see that there exists a constant c_4 such that

$$\int_{t_0}^{T_1} |d'(s)|\, ds \leqq c_4. \tag{2.15}$$

It is also easy to deduce that

$$\int_{t_0}^{T_1} |e'(s)|\, ds \leqq \left(|r_m(t_0)| + \int_{t_0}^{T_1} |r'_m(s)|\, ds \right) \int_0^{T_1} |b_m(s)|\, ds. \tag{2.16}$$

Since it follows from (2.10) and (2.14) that

$$\int_{t_0}^{T_1} |r'_m(s)|\, ds \leqq \int_{t_0}^{T_1} (|b'_m(s)| + |d'(s)| + |e'(s)|)\, ds$$

we conclude from (2.9), (2.11)–(2.13), (2.15) and (2.16) that there exists a constant c_5 such that

$$\int_{t_0}^{T_1} |r'_m(s)|\, ds \leqq c_5(1 + a(t_0)).$$

As t_0 was arbitrary we deduce from this inequality that there exists a constant c_1 such that (2.7) holds with r replaced by r_m. This suffices to complete the proof of Lemma 2.1 as noted above.

We are going to apply Lemma 2.1 to the case when $a(t)=c_0 \log t^{-1}$ and $b(t)=k'(t)$. The next lemma is crucial for the rest of the proof as it gives the needed convergence properties of the functions u_λ.

LEMMA 2.2. *Let the assumptions of the theorem hold. Then there exists a number T_2, $0<T_2\leqq T_1$ which depends only on the behavior of k on $(0, T_1)$ such that the functions u_λ converge uniformly on $[0, T_2]$ toward a function u.*

Proof. Without loss of generality we may assume that $k(0)=1$ (otherwise replace the nonlinear operator g by $k(0)g$). By (1.2) and (1.6) we can differentiate (2.1) and obtain

$$u'_\lambda(t)+g_\lambda u_\lambda(t)+\int_0^t k'(t-s)g_\lambda u_\lambda(s)\,ds=f'(t) \quad \text{a.e. } t\in R_+. \tag{2.17}$$

Using the resolvent property of r (defined in (2.6) with $b=k'$) we have from (2.17) that

$$u'_\lambda(t)+g_\lambda u_\lambda(t)-\int_0^t r(t-s)u'_\lambda(s)\,ds=f'(t)-\int_0^t r(t-s)f'(s)\,ds \quad \text{a.e. } t\in(0, T_1). \tag{2.18}$$

Let t_n be a sequence of positive real numbers such that $t_n\to 0$ when $n\to\infty$. Define

$$r_n(t)=\begin{cases} r(t_n), & 0<t\leqq t_n,\\ r(t), & t_n\leqq t\leqq T_1.\end{cases} \tag{2.19}$$

Then if $0<t<T_1$

$$\int_0^t r(t-s)u'_\lambda(s)\,ds=\int_0^t (r(t-s)-r_n(t-s))u'_\lambda(s)\,ds+\int_0^t r_n(t-s)u'_\lambda(s)\,ds \tag{2.20}$$

and

$$\int_0^t r_n(t-s)u'_\lambda(s)\,ds=r(t_n)u_\lambda(t)-r_n(t)u_\lambda(0)+\int_0^t u_\lambda(t-s)\,dr_n(s). \tag{2.21}$$

Subtract (2.18) with $\lambda=\mu$ from (2.18), form the scalar product of the resulting equation and $v_{\lambda,\mu}(t)\overset{\text{def}}{=}u_\lambda(t)-u_\mu(t)$; finally integrate over $(0, t)$. The result is by (2.20) and (2.21). ($|\cdot|$ is the norm and $(\cdot,\cdot)$ the scalar product in H.)

$$\begin{aligned}\tfrac{1}{2}|v_{\lambda,\mu}(t)|^2=&-\int_0^t (g_\lambda u_\lambda(s)-g_\mu u_\mu(s), v_{\lambda,\mu}(s))\,ds\\ &+\int_0^t\left(\int_0^s (r(s-p)-r_n(s-p))v'_{\lambda,\mu}(p)\,dp, v_{\lambda,\mu}(s)\right)ds\\ &+\int_0^t\left(r(t_n)v_{\lambda,\mu}(s)+\int_0^s v_{\lambda,\mu}(s-p)\,dr_n(p), v_{\lambda,\mu}(s)\right)ds,\\ &\qquad t\in[0, T_1].\end{aligned} \tag{2.22}$$

Using (2.3) the definition of the Yosida approximation g_λ and the monotonicity of g we see that there exists a constant c_6 such that the first term on the right side in the equation in (2.22) is $\leqq\frac{1}{2}c_6^2(\lambda+\mu)$. If we apply a quadratic integral inequality (see [2, Lemma A5]), we obtain from (2.22)

$$\begin{aligned}|v_{\lambda,\mu}(t)|\leqq\ & c_6(\lambda+\mu)^{1/2}+\|r-r_n\|_{L^1(0,T_1;R)}\|v'_{\lambda,\mu}\|_{L^1(0,T_1;H)}\\ &+\int_0^t\left(|r(t_n)|\,|v_{\lambda,\mu}(s)|+\int_0^s |v_{\lambda,\mu}(s-p)|\,|dr_n(p)|\right)ds, \qquad t\in[0, T_1].\end{aligned} \tag{2.23}$$

It follows from (2.2) that $\|v'_{\lambda,\mu}\|_{L^1(0,T;H)}$ remains bounded and so (2.19) and (2.23) combined with Gronwall's lemma imply that there exists a constant c_7 such that

$$\max_{0\leqq s\leqq t}|v_{\lambda,\mu}(s)|\leqq c_7\left[(\lambda+\mu)^{1/2}+\int_0^{t_n}|r(s)-r(t_n)|\,ds\right] \tag{2.24}$$
$$\cdot\exp\left[t(|r(t_n)|+\int_{t_n}^{T_1}|dr(s)|)\right],\qquad t\in[0,T_1].$$

By (1.4) and Lemma 2.1 we get for certain positive constants c_8 and c_9 (independent of n)

$$\exp\left[t(|r(t_n)|+\int_{t_n}^{T_1}|dr(s)|)\right]\leqq c_8 t_n^{-c_9 t},\qquad t\in[0,T_1]. \tag{2.25}$$

From (1.5) and Lemma 2.1 we conclude that

$$\int_0^{t_n}|r(s)-r(t_n)|\,ds\leqq c_0c_1(t_n-t_n\log t_n). \tag{2.26}$$

If we choose $T_2>0$ such that $c_9T_2\leqq\frac{1}{2}$, then (2.24)–(2.26) imply that

$$\limsup_{\lambda,\mu\to 0}\max_{0\leqq s\leqq T_2}|v_{\lambda,\mu}(s)|\leqq c_0c_1c_7c_8t_n^{1/2}(1-\log t_n), \tag{2.27}$$

and letting $n\to\infty$ in (2.27) we obtain the conclusion of Lemma 2.2.

The rest of the proof now follows easily (for details see [6] or [7]) and using a translation-iteration argument we can continue the solution to R_+. The uniqueness of the function u is established in the same way as the convergence of the approximate solutions u_λ.

REFERENCES

[1] V. Barbu, *Nonlinear Volterra equations in a Hilbert space*, this Journal, 6 (1975), pp. 728–741.

[2] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland/American Elsevier, Amsterdam, 1973.

[3] M. G. Crandall and J. A. Nohel, *An abstract functional differential equation and a related nonlinear Volterra equation*, Israel J. Math., 29 (1978), pp. 313–328.

[4] G. Gripenberg, *On a nonlinear Volterra integral equation in a Banach space*, J. Math. Anal. Appl., to appear.

[5] ———, *On a frequency domain condition used in the theory of Volterra equations*, this Journal, 10 (1979), pp. 839–843.

[6] S-O. Londen, *On an integral equation in a Hilbert space*, this Journal, 8 (1977), pp. 950–970.

[7] S-O. Londen and O. J. Staffans, *A note on Volterra equations in a Hilbert space*, Proc. Amer. Math. Soc., 70 (1978), pp. 57–62.

[8] J. J. Levin, *Resolvents and bounds for linear and nonlinear Volterra equations*, Trans. Amer. Math. Soc., 288 (1977), pp. 207–222.

[9] R. C. MacCamy, *Stability theorems for a class of functional differential equations*, SIAM J. Appl. Math., 30 (1976), pp. 557–576.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0013 $01.00/0

SOME MULTIPLE POWER SERIES WITH ZERO-ONE COEFFICIENTS*

L. CARLITZ†

Abstract. The paper is concerned with sums of the type

$$S_{n,j}=\sum x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n} \qquad (n>1),$$

where the summation is over either

$$(*) \qquad ja_i \leqq a_1+a_2+\cdots+a_n \qquad (1\leqq j\leqq n;\ 1\leqq i\leqq n)$$

or

$$(**) \qquad a_1+a_2+\cdots+a_n=ja_i+(n-j)b_i \qquad (0\leqq j\leqq n;\ 1\leqq i\leqq n);$$

the a_i and b_i are nonnegative integers. It is proved, for example, that for the first type with $j=2$, the sum is a rational function with denominator equal to $\prod_{1\leqq i<k\leqq n}(1-x_ix_k)$.

Several combinatorial applications are obtained by specializing the x_i. For example it is proved that the number of nonnegative solutions of the system

$$a_1+a_2+\cdots+a_n=N, \qquad (n-1)a_i\leqq N \quad (1\leqq i\leqq n)$$

is equal to the binomial coefficient

$$\binom{k+n-s-1}{n-1} \qquad (N=k(n-1)+s,\ 0\leqq s<n-1).$$

The final section of the paper is concerned with multiple Dirichlet series

$$\Phi_{n,j}=\sum m_1^{-s_1}m_2^{-s_2}\cdots m_n^{-s_n},$$

where the smmation is over all positive integers m_i such that

$$m_i^j \mid m_1m_2\cdots m_n \qquad (1\leqq j\leqq n;\ 1\leqq i\leqq n).$$

The $\Phi_{n,j}$ are expressed as products involving series satisfying $(*)$; in particular

$$\Phi_{n,n-1}=\frac{\zeta(\alpha)\zeta(\sigma-s_1)\cdots\zeta(\sigma-s_n)}{\zeta((n-1)\sigma)},$$

where $\sigma=s_1+\cdots+s_n$ and $\zeta(s)$ is the Riemann zeta-function.

1. Introduction and summary. Let $n\geqq 2$ and put

$$(1.1) \qquad S_n=S_n(x_1,x_2,\cdots,x_n)=\sum x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n},$$

where the summation is over all nonnegative $a_1,a_2,\cdots,a_n$ that satisfy

$$(1.2) \qquad \begin{aligned} a_1&\leqq a_2+a_3+\cdots+a_n\\ a_2&\leqq a_1+a_3+\cdots+a_n\\ &\ \vdots \qquad\qquad \vdots\\ a_n&\leqq a_1+a_2+\cdots+a_{n-1}. \end{aligned}$$

For $n=2$ the conditions (1.2) reduce to $a_1\leqq a_2$, $a_2\leqq a_1$, so that $a_1=a_2$. It follows that

$$(1.3) \qquad S_2=\frac{1}{1-x_1x_2}.$$

* Received by the editors April 22, 1977, and in revised form September 6, 1978.
† Department of Mathematics, Duke University, Durham, North Carolina 27706.

For $n=3$ we shall show that

$$(1.4)\qquad S_3=\frac{1+x_1x_2x_3}{(1-x_2x_3)(1-x_1x_3)(1-x_1x_2)},$$

while, for $n=4$,

$$(1.5)\qquad S_4=\frac{1+\sigma_3-2\sigma_4-2\sigma_1\sigma_4+\sigma_3\sigma_4+\sigma_4^2}{(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)(1-x_2x_3)(1-x_2x_4)(1-x_3x_4)},$$

where σ_i denotes the ith elementary symmetric function of x_1, x_2, x_3, x_4.

For arbitrary $n \geqq 2$, we prove that

$$(1.6)\qquad S_n=\frac{P_n(x_1, x_2, \cdots, x_n)}{\prod_{1\leqq i<j\leqq n}(1-x_ix_j)},$$

where $P_n \equiv P_n(x_1, x_2, \cdots, x_n)$ is a symmetric polynomial in $x_1, x_2, \cdots, x_n$ with integral coefficients.

Generalizing (1.1), we put, for $1 \leqq j \leqq n$,

$$(1.7)\qquad S_{n,j} \equiv S_{n,j}(x_1, x_2, \cdots, x_n)=\sum x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n},$$

where now the summation is over all nonnegative $a_1, a_2, \cdots, a_n$ such that

$$(1.8)\qquad ja_i \leqq a_1+a_2+\cdots+a_n \qquad (i=1, 2, \cdots, n).$$

Clearly

$$(1.9)\qquad S_{n,2}(x_1, x_2, \cdots, x_n)=S_n(x_1, x_2, \cdots, x_n).$$

For $j=1$, the conditions (1.8) are automatically satisfied; hence

$$(1.10)\qquad S_{n,1}=\frac{1}{(1-x_1)(1-x_2)\cdots(1-x_n)}.$$

For $j=n$, (1.8) implies

$$n(a_1+\cdots+a_n) \leqq n(a_1+\cdots+a_n).$$

Thus the inequalities become equalities, so that $a_1=a_2=\cdots=a_n$. Hence

$$(1.11)\qquad S_{n,n}=\frac{1}{1-x_1x_2\cdots x_n}.$$

We shall show that

$$(1.12)\qquad S_{n,n-1}=\frac{1+p+\cdots+p^{n-2}}{(1-p/x_1)(1-p/x_2)\cdots(1-p/x_n)},$$

where $p=x_1x_2\cdots x_n$.

In view of these special results it seems plausible that $S_{n,j}$ is a rational function of $x_1, x_2, \cdots, x_n$:

$$(1.13)\qquad S_{n,j}=N_{n,j}/D_{n,j} \qquad (1 \leqq j \leqq n),$$

where $N_{n,j}$ is a symmetric polynomial in $x_1, x_2, \cdots, x_n$ and

$$(1.14)\qquad D_{n,j}=\prod(1-x_1x_2\cdots x_j),$$

the product extending over all products of j of the x_i.

A variant of (1.8) that is somewhat more symmetrical is the following:

(1.15) $$a_1+a_2+\cdots+a_n=ja_i+(n-j)b_i \qquad (i=1,2,\cdots,n),$$

where both the a_i and b_i are nonnegative integers. Note that (1.15) implies

$$n\sum_{i=1}^{n} a_i=j\sum_{i=1}^{n} a_i+(n-j)\sum_{i=1}^{n} b_i,$$

so that

$$a_1+a_2+\cdots+a_n=b_1+b_2+\cdots+b_n.$$

We now put

(1.16) $$\bar{S}_{n,j}\equiv\bar{S}_{n,j}(x_1,x_2,\cdots,x_n)=\sum x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n} \qquad (0\leqq j\leqq n),$$

where the summation is over all nonnegative a_i, b_i satisfying (1.15). Note that

$$\bar{S}_{n,n-1}=S_{n,n-1},\ \bar{S}_{n,n}=S_{n,n}.$$

We shall show that $\bar{S}_{n,j}$ satisfies the relation

(1.17) $$\bar{S}_{n,j}(x_1^j,\cdots,x_n^j)=\bar{S}_{n,n-j}(px_1^{-(n-j)},\cdots,px_n^{-(n-j)}) \qquad (0\leqq j\leqq n),$$

where $p=x_1x_2\cdots x_n$. Indeed if we let

(1.18) $$F_{n,j}(x_1,\cdots,x_n;y_1,\cdots,y_n)=\sum x_1^{a_1}\cdots x_n^{a_n}y_1^{b_1}\cdots y_n^{b_n},$$

where the summation is over all nonnegative a_i, b_i satisfying (1.15), then

(1.19) $$F_{n,j}(x_1,\cdots,x_n;y_1,\cdots,y_n)=F_{n,n-j}(y_1,\cdots,y_n;x_1,\cdots,x_n) \qquad (0\leqq j\leqq n).$$

It is clear from the definitions that the various power series defined above have coefficients equal to zero or one. If in any of the series we take $x_1=\cdots=x_n=x$, we get a number of simple combinatorial results. For example (1.4) reduces to

$$S_3(x)=\frac{1+x^3}{(1-x^2)^3}.$$

It follows that the number of solutions in nonnegative a_1, a_2, a_3 of the system

$$\begin{aligned}&a_1+a_2+a_3=N,\\ &2a_i\leqq N, \qquad (i=1,2,3)\end{aligned}$$

is equal to

(1.20) $$\begin{cases}\binom{m+2}{2} & (N=2m),\\[2mm] \binom{m+1}{2} & (N=2m+1).\end{cases}$$

More generally (1.12) reduces to

$$S_{n,n-1}=\frac{1+x^n+\cdots+x^{n(n-2)}}{(1-x^{n-1})^n}.$$

It follows that the number of nonnegative solutions of the system

$$\begin{aligned}&a_1+a_2+\cdots+a_n=N,\\ &(n-1)a_i\leqq N \qquad (i=1,2,\cdots,n)\end{aligned}$$

is equal to

$$\binom{k+n-s-1}{n-1} \qquad (N=k(n-1)+s,\ 0\leqq s<n-1). \tag{1.21}$$

The special result (1.5) implies the following: Let $f_4(m)$ denote the number of solutions in nonnegative a_i of

$$\begin{aligned} &a_1+a_2+a_3+a_4=m, \\ &2a_i\leqq m \qquad\qquad (1\leqq i\leqq 4). \end{aligned}$$

Then

$$\begin{aligned} &f_4(2m+1)=4\binom{m+2}{3}, \\ &f_4(2m)=\tfrac{1}{3}(m+1)(2m^2+4m+3). \end{aligned}$$

In the last section of the paper we define the multiple Dirichlet series

$$\Phi_{n,j}(s_1, s_2, \cdots, s_n)=\sum m_1^{-s_1}m_2^{-s_2}\cdots m_n^{-s_n} \qquad (1\leqq j\leqq n),$$

where the summation is over all positive integers $m_1, m_2, \cdots, m_n$ satisfying

$$m_i^j \mid m_1m_2\cdots m_n \qquad (1\leqq i\leqq n).$$

For example

$$\begin{aligned} &\Phi_{n,1}(s_1, s_2, \cdots, s_n)=\zeta(s_1)\zeta(s_2)\cdots\zeta(s_n), \\ &\Phi_{n,n}(s_1, s_2, \cdots, s_n)=\zeta(s_1+s_2+\cdots+s_n), \end{aligned}$$

where $\zeta(s)$ is the Riemann zeta function.

We shall show that

$$\Phi_{n,j}(s_1, s_2, \cdots, s_n)=\prod_p S_{n,j}(p^{-s_1}, p^{-s_2}, \cdots, p^{-s_n}),$$

where $S_{n,j}$ is defined by (1.7). In particular, we have

$$\Phi_{n,n-1}(s_1, s_2, \cdots, s_n)=\frac{\zeta(\sigma)\zeta(\sigma-s_1)\cdots\zeta(\sigma-s_2)}{\zeta((n-1)\sigma)}, \tag{1.22}$$

where $\sigma=s_1+s_2+\cdots+s_n$.

As an application of (1.22), it is proved that $\delta_n(m)$, the number of solutions of

$$\begin{aligned} &m_1m_2\cdots m_n=m, \\ &m_i^{n-1}\mid m \qquad\qquad (1\leqq i\leqq n), \end{aligned}$$

satisfies

$$\delta_n(m)=\sum_{a^nb^{n-1}c^{n(n-1)}=m}\tau_n(b)\mu(c),$$

where μ is the Möbius function and

$$\zeta^n(s)=\sum_{m=1}^{\infty}\tau_n(m)m^{-s},$$

so that $\tau_n(m)$ is the number of solutions of $m_1m_2\cdots m_n=m$.

2. Proof of (1.4) and (1.5). We first prove (1.4). The conditions

$$\begin{aligned} &C\colon\ a_1 \leqq a_2+a_3, \quad a_2 \leqq a_1+a_3, \quad a_3 \leqq a_1+a_2, \\ &C_1\colon\ a_1 > a_2+a_3, \quad a_2 \geqq 0, \quad a_3 \geqq 0, \\ &C_2\colon\ a_2 > a_1+a_3, \quad a_1 \geqq 0, \quad a_3 \geqq 0, \\ &C_3\colon\ a_3 > a_1+a_2, \quad a_1 \geqq 0, \quad a_2 \geqq 0 \end{aligned}$$

are mutually exclusive. Moreover their union is

$$a_1 \geqq 0, \quad a_2 \geqq 0, \quad a_3 \geqq 0.$$

It follows that

$$\frac{1}{(1-x_1)(1-x_2)(1-x_3)} = S_3 + \Big(\sum_{C_1}\sum_{C_2}\sum_{C_3}\Big)\, x_1^{a_1} x_2^{a_2} x_3^{a_3}. \tag{2.1}$$

Now

$$\begin{aligned} \sum_{C_1} &= \sum_{a_2,a_3=0}^{\infty} x_2^{a_2} x_3^{a_3} \sum_{a_1=a_2+a_3+1}^{\infty} x_1^{a_1} \\ &= \sum_{a_2,a_3=0}^{\infty} x_2^{a_2} x_3^{a_3} \frac{x_1^{a_2+a_3+1}}{1-x_1} \\ &= \frac{x_1}{1-x_1} \frac{1}{(1-x_1x_2)(1-x_1x_3)} \end{aligned} \tag{2.2}$$

and similarly for $\sum_{C_2}$ and $\sum_{C_3}$. Substituting in (2.1), we get (1.4).

To prove (1.5) we consider the conditions

$$\begin{aligned} &C\colon\ a_1 \leqq a_2+a_3+a_4, \quad a_2 \leqq a_1+a_3+a_4, \quad a_3 \leqq a_1+a_2+a_4, \quad a_4 \leqq a_1+a_2+a_3, \\ &C_1\colon\ a_1 > a_2+a_3+a_4, \quad a_2 \geqq 0, \quad a_3 \geqq 0, \quad a_4 \geqq 0, \\ &C_2\colon\ a_2 > a_1+a_3+a_4, \quad a_1 \geqq 0, \quad a_3 \geqq 0, \quad a_4 \geqq 0, \\ &C_3\colon\ a_3 > a_1+a_2+a_4, \quad a_1 \geqq 0, \quad a_2 \geqq 0, \quad a_4 \geqq 0, \\ &C_4\colon\ a_4 > a_1+a_2+a_3, \quad a_1 \geqq 0, \quad a_2 \geqq 0, \quad a_3 \geqq 0. \end{aligned}$$

The conditions C, C_1, C_2, C_3, C_4 are mutually exclusive and their union is

$$a_1 \geqq 0, \quad a_2 \geqq 0, \quad a_3 \geqq 0, \quad a_4 \geqq 0.$$

It follows that

$$\frac{1}{(1-x_1)(1-x_2)(1-x_3)(1-x_4)} = S_4 - \sum_{i=1}^{4} \sum_{C_i}, \tag{2.3}$$

where

$$\begin{aligned} \sum_{C_1} &= \sum_{a_2,a_3,a_4=0}^{\infty} x_2^{a_2} x_3^{a_3} x_4^{a_4} \sum_{a_1=a_2+a_3+a_4+1}^{\infty} x_1^{a_1} \\ &= \sum_{a_2,a_3,a_4=0}^{\infty} x_2^{a_2} x_3^{a_3} x_4^{a_4} \frac{x_1^{a_1+a_2+a_3+1}}{1-x_1} \\ &= \frac{x_1}{1-x_1} \frac{1}{(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)} \end{aligned}$$

and similarly for $\sum_{C_2}, \sum_{C_3}, \sum_{C_4}$. Thus (2.3) becomes

$$
\begin{aligned}
S_4 = {} & \frac{1}{(1-x_1)(1-x_2)(1-x_3)(1-x_4)} - \frac{x_1}{1-x_1} \frac{1}{(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)} \\
& - \frac{x_2}{1-x_2} \frac{1}{(1-x_1x_2)(1-x_2x_3)(1-x_2x_4)} \\
& - \frac{x_3}{1-x_3} \frac{1}{(1-x_1x_3)(1-x_2x_3)(1-x_3x_4)} \\
& - \frac{x_4}{1-x_4} \frac{1}{(1-x_1x_4)(1-x_2x_4)(1-x_3x_4)}.
\end{aligned} \tag{2.4}
$$

It is clear from (2.4) that S_4 is rational and that the denominator is at most

$$(1-x_1)(1-x_2)(1-x_3)(1-x_4)(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)(1-x_3x_3)(1-x_2x_4)(1-x_2x_4).$$

We now show that the factors

$$1-x_1, \quad 1-x_2, \quad 1-x_3, \quad 1-x_4$$

drop out when S_4 is written in reduced form. Consider

$$
\begin{aligned}
& \frac{1}{(1-x_1)(1-x_2)(1-x_3)(1-x_4)} - \frac{x_1}{1-x_1} \frac{1}{(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)} \\
& \frac{1}{1-x_1} \left\{ \frac{1}{(1-x_2)(1-x_3)(1-x_4)} - \frac{x_1}{(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)} \right\}.
\end{aligned}
$$

The quantity in braces $\{\cdots\}$ vanishes when $x_1 = 1$. Hence the factor $1-x_1$ drops out. Similarly for the remaining three factors

$$1-x_2, \quad 1-x_3, \quad 1-x_4.$$

Thus the denominator is at most

$$(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)(1-x_2x_3)(1-x_2x_4)(1-x_3x_4). \tag{2.5}$$

Not all the factors in (2.5) can drop out since S_4 is clearly not a polynomial. Thus at least one factor, say $1-x_1x_2$, remains. But since S_4 is symmetric in x_1, x_2, x_3, x_4 it follows that the remaining ones also survive.

We may accordingly write

$$S_4 = \frac{N_4}{(1-x_1x_2)(1-x_1x_3)(1-x_1x_4)(1-x_2x_3)(1-x_2x_4)(1-x_3x_4)}, \tag{2.6}$$

where N_4 is symmetric in x_1, x_2, x_3, x_4. Also it is evident from (2.4) that N_4 is of degree 8 and that it reduces to 1 when $x_1 = x_2 = x_3 = x_4 = 0$.

Since

$$S_4(x_1, x_2, x_3, 0) = S_3(x_1, x_2, x_3),$$

comparison of (2.6) with (1.4) gives

$$N_4 = 1 + \sigma_3 + \sigma_4(a_0 + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 + a_4\sigma_4), \tag{2.7}$$

where $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ are the elementary symmetric functions of x_1, x_2, x_3, x_4 while a_0, a_1, a_2, a_3, a_4 are independent of the x_i.

In order to evaluate coefficients $a_0, \cdots, a_4$ we examine the coefficients of the monomials

$$x_1x_2x_3x_4,\quad x_1^2x_2x_3x_4,\quad x_1^2x_2^2x_3x_4,\quad x_1^2x_2^2x_3^2x_4,\quad x_1^2x_2^2x_3^2x_4$$

in the right hand side of (2.6.). This leads to the following relations:

$$3+a_0=1,$$
$$3+a_1=1,$$
$$3+a_0+a_2=1,$$
$$6+3a_1+a_3=1,$$
$$6+3a_0+a_4=1.$$

Hence $a_0=-2$, $a_1=-2$, $a_2=0$, $a_3=1$, $a_4=1$, so that

$$N_4=1+\sigma_3-2\sigma_4-2\sigma_1\sigma_4+\sigma_3\sigma_4+\sigma_4^2. \tag{2.8}$$

This completes the proof of (1.5).

For $x_1=x_2=x_3=x_4=x$, (1.5) reduces to

$$\begin{aligned} S_4(x)\equiv S_4(x,x,x,x)&=\frac{1+2x^2+4x^3+x^4}{(1-x^2)^4}\\ &=\frac{1-x+3x^2+x^3}{(1-x)(1-x^2)^3}. \end{aligned} \tag{2.9}$$

3. Rationality of S_n. We now prove the following

THEOREM 3.1. *For arbitrary $n\geqq 2$,*

$$S_n(x_1,x_2,\cdots,x_n)=\frac{P_n(x_1,x_2,\cdots,x_n)}{\prod_{1\leqq j<k\leqq n}(1-x_jx_k)}, \tag{3.1}$$

where $P_n\equiv P_n(x_1,x_2,\cdots,x_n)$ is a symmetric polynomial in $x_1,x_2,\cdots,x_n$ with integral coefficients and of degree $n(n-2)$.

Proof. The proof is similar to the proof of (1.5). Let C, C_i denote the conditions

$$C:\ 2a_j\leqq a_1+a_2+\cdots+a_n \qquad (j=1,2,\cdots,n),$$
$$C_i:\ 2a_i>a_1+a_2+\cdots+a_n,\quad a_j\geqq 0 \qquad (i,j=1,2,\cdots,n).$$

The conditions $C, C_1, C_2,\cdots,C_n$ are mutually exclusive and their union is

$$a_1\geqq 0,\quad a_2\geqq 0,\cdots,a_n\geqq 0.$$

It then follows that

$$S_n=\frac{1}{(1-x_1)(1-x_2)\cdots(1-x_n)}-\sum_{i=1}^{n}\sum_{C_i}x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}. \tag{3.2}$$

We have

$$\begin{aligned} \sum_{C_1}x_1^{a_1}x_2^{a_2}\cdots x_n^{a_3}&=\sum_{a_2,\cdots,a_n=0}^{\infty}x_2^{a_2}\cdots x_n^{a_n}\sum_{a_1=a_2+\cdots+a_n+1}^{\infty}x_1^{a_1}\\ &=\sum_{a_2,\cdots,a_n=0}^{\infty}x_2^{a_2}\cdots x_n^{a_n}\frac{x_1^{a_2+\cdots a_n+1}}{1-x_1}\\ &=\frac{x_1}{1-x_1}\frac{1}{(1-x_1x_2)\cdots(1-x_1x_n)} \end{aligned}$$

and similarly for the remaining $\sum c_i$. Substituting in (3.2), we get

$$S_n = \frac{1}{(1-x_1)(1-x_2)\cdots(1-x_n)} - \sum \frac{x_1}{1-x_1}\frac{1}{(1-x_1x_2)\cdots(1-x_1x_n)} \tag{3.3}$$

where the sum on the right denotes a symmetric sum.

From (3.3) it is evident that S_n is a rational function of the x_i and that the denominator is at most

$$\prod_{1\leqq i\leqq n}(1-x_i)\cdot\prod_{1\leqq j<k\leqq n}(1-x_jx_k).$$

To show that the linear factors drop out, consider

$$\frac{1}{(1-x_1)(1-x_2)\cdots(1-x_n)} - \frac{x_1}{1-x_1}\frac{1}{(1-x_1x_2)\cdots(1-x_1x_3)}$$
$$= \frac{1}{1-x_1}\left\{\frac{1}{(1-x_2)\cdots(1-x_n)} - \frac{1}{(1-x_1x_2)\cdots(1-x_1x_n)}\right\}.$$

For $x_1 = 1$ the quantity in braces vanishes and therefore the factor $1-x_1$ does indeed drop out of the denominator. Similarly for the remaining linear factors. As for the factors $1-x_jx_k$, since S_n is not a polynomial, at least one, say $1-x_1x_2$, must remain. But then by symmetry the remaining ones must also survive.

Finally it follows from (3.3) that the degree of the numerator $P_n(x_1, x_2, \cdots, x_n)$ is equal to $n(n-1)-n = n(n-2)$, as stated.

4. The sum $S_{n,j}$. Let $1\leqq j\leqq n$ and put

$$S_{n,j} = S_{n,j}(x_1, x_2, \cdots, x_n) = \sum x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}, \tag{4.1}$$

where the summation is over all nonnegative a_i that satisfy

$$ja_i \leqq a_1+a_2+\cdots+a_n \qquad (1\leqq i\leqq n). \tag{4.2}$$

For $j=1$, (4.2) is automatically satisfied so that

$$S_{n,1} = \frac{1}{(1-x_1)(1-x_2)\cdots(1-x_n)}. \tag{4.3}$$

For $n=2$ it is evident that

$$S_{n,2}(x_1, x_2, \cdots, x_n) = S_n(x_1, x_2, \cdots, x_n), \tag{4.4}$$

where $S_n(x_1, x_2, \cdots, x_n)$ is defined by (1.1) and (1.2).

For $j=n$, the conditions (4.2) imply

$$n(a_1+a_2+\cdots+a_n)\leqq n(a_1+a_2+\cdots+a_n),$$

so that the inequalities become equalities. It follows that $a_1=a_2=\cdots=a_n$ and therefore

$$S_{n,n}(x_1, x_2, \cdots, x_n) = \frac{1}{1-x_1x_2\cdots x_n}. \tag{4.5}$$

We shall now show that

$$S_{n,n-1}(x_1, x_2, \cdots, x_n) = \frac{1+p+\cdots+p^{n-2}}{(1-p/x_1)(1-p/x_2)\cdots(1-p/x_n)}. \tag{4.6}$$

Put

$$b_i = s-(n-1)a_i \qquad (1 \leqq i \leqq n), \tag{4.7}$$

where

$$s = a_1 + a_2 + \cdots + a_n.$$

Then

$$\sum_{i=1}^{n} b_i = ns-(n-1)\sum_{i=1}^{n} a_i,$$

so that

$$b_1 + b_2 + \cdots + b_n = s. \tag{4.8}$$

Also, by (4.7),

$$b_1 \equiv b_2 \equiv \cdots \equiv b_n \equiv s \pmod{n-1}. \tag{4.9}$$

Conversely, if $b_1, b_2, \cdots, b_n$ are nonnegative integers satisfying

$$b_1 \equiv b_2 \equiv \cdots \equiv b_n \pmod{n-1} \tag{4.10}$$

and we define s by (4.8), we can solve (4.7) for the a_i.

It is convenient to define

$$S_{n,n-1}^{(k)} = \sum x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \qquad (0 \leqq k < n-1), \tag{4.11}$$

where the summation is over all nonnegative a_i satisfying

$$(n-1)a_i \leqq a_1 + a_2 + \cdots + a_n \qquad (1 \leqq i \leqq n)$$

and

$$a_1 + a_2 + \cdots + a_n \equiv k \pmod{n-1}. \tag{4.12}$$

Then by (4.10),

$$b_1 \equiv b_2 \equiv \cdots \equiv b_n \equiv k \pmod{n-1}.$$

Put

$$b_i = (n-1)c_i + k \qquad (i \leqq i \leqq n), \tag{4.13}$$

where the c_i are nonnegative integers. Moreover the c_i are otherwise unrestricted.

Hence (4.11) becomes

$$\begin{aligned} S_{n,n-1}^{(k)} &= \sum x_1^{(s-b_1)/(n-1)} \cdots x_n^{(s-b_n)/(n-1)} \\ &= \sum_{c_1,\cdots,c_n=0}^{\infty} x_1^{c_2+\cdots+c_n+k} \cdots x_n^{c_1+\cdots+c_{n-1}+k} \\ &= (x_1 x_2 \cdots x_n)^k \sum_{c_1,\cdots,c_n=0}^{\infty} (x_2 \cdots x_n)^{c_1} (x_1 x_3 \cdots x_n)^{c_2} \cdots (x_1 \cdots x_{n-1})^{c_n} \\ &= \frac{(x_1 x_2 \cdots x_n)^k}{(1-x_2 \cdots x_n)(1-x_1 x_3 \cdots x_n) \cdots (1-x_1 \cdots x_{n-1})}. \end{aligned}$$

If we put $p = x_1 x_2 \cdots x_n$, this becomes

$$S_{n,n-1}^{(k)} = \frac{p^k}{(1-p/x_1)(1-p/x_2)\cdots(1-p/x_n)} \qquad (1 \leqq k < n-1). \tag{4.14}$$

Summing over k, we get

$$S_{n,n-1} = \frac{1+p+\cdots+p^{n-2}}{(1-p/x_1)(1-p/x_2)\cdots(1-p/x_n)}. \tag{4.15}$$

This completes the proof of (4.6).

A refinement of $S_{n,n-1}$ may be mentioned. Let $i_1, i_2, \cdots, i_n$ be nonnegative integers and put

$$S^*_{i_1,\cdots,i_n} = \sum a_1^{i_1} a_2^{i_2} \cdots a_n^{i_n} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \tag{4.16}$$

where the summation is over all nonnegative a_i satisfying

$$(n-1)a_i \leqq a_1 + a_2 + \cdots + a_n \qquad (1 \leqq i \leqq n). \tag{4.17}$$

Then

$$\sum_{i_1,\cdots,i_n=0}^{\infty} S^*_{i_1,\cdots,i_n} \frac{z_1^{i_1}\cdots z_n^{i_n}}{i_1!\cdots i_n!} = \sum_{a_1,\cdots,a_n} (x_1 e^{z_1})^{a_1} \cdots (x_n e^{z_n})^{a_n},$$

where again the summation is over all nonnegative a_i satisfying (4.17). Hence, by (4.6), we get

$$\begin{aligned} \sum_{i_1,\cdots,i_n=0}^{\infty} S^*_{i_1,\cdots,i_n} \frac{z_1^{i_1}\cdots z_n^{i_n}}{i_1!\cdots i_n!} &= S_{n,n-1}(x_1 e^{z_1}, \cdots, x_n e^{z_n}) \\ &= \frac{1+p_z+\cdots+p_z^{n-2}}{(1-p_z/x_1 e^{z_1})\cdots(1-p_z/x_n e^{z_n})}, \end{aligned} \tag{4.18}$$

where

$$p_z = x_1 x_2 \cdots x_n e^{z_1+z_2+\cdots+z_n} = p e^{z_1+z_2+\cdots+z_n}.$$

5. The sums $\bar{S}_{n,j}$ and $F_{n,j}$. Let $0 \leqq j \leqq n$. We define

$$\bar{S}_{n,j} \equiv \bar{S}_{n,j}(x_1, x_2, \cdots, x_n) = \sum x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \tag{5.1}$$

and

$$F_{n,j} \equiv F_{n,j}(x_1, \cdots, x_n; y_1, \cdots, y_n) = \sum x_1^{a_1} \cdots x_n^{a_n} y_1^{b_1} \cdots y_n^{b_n}, \tag{5.2}$$

where, in each case, the summation is over all nonnegative a_i and b_i that satisfy

$$ja_i + (n-j)b_i = a_1 + \cdots + a_n \qquad (1 \leqq i \leqq n). \tag{5.3}$$

Summing over i, we find that (5.3) gives

$$j \sum_{i=1}^n a_i + (n-j) \sum_{i=1}^n b_i = n \sum_{i=1}^n a_i.$$

Hence

$$a_1 + \cdots + a_n = b_1 + \cdots + b_n, \tag{5.4}$$

so that the conditions (5.3) are symmetrical in the a_i and the b_i. It follows that

$$F_{n,j}(x_1, \cdots, x_n; y_1, \cdots, y_n) = F_{n,n-j}(y_1, \cdots, y_n; x_1, \cdots, x_n). \tag{5.5}$$

In the next place, by (5.1) and (5.3), we have

$$\begin{aligned}\bar{S}_{n,j}(x_1^j, \cdots, x_n^j) &= \sum x_1^{s-(n-j)b_1} \cdots x_n^{s-(n-j)b_1} \\ &= \sum (x_1^{1-n+j}x_2 \cdots x_n)^{b_1} \cdots (x_1 \cdots x_{n-1}x_n^{1-n+j})^{b_n} \\ &= \sum (px_1^{-n+j})^{b_1}(px_2^{-n+j})^{b_2} \cdots (px_n^{-n+j})^{b_n},\end{aligned}$$

where $p = x_1x_2 \cdots x_n$. Therefore

$$\bar{S}_{n,j}(x_1^j, \cdots, x_n^j) = \bar{S}_{n,n-j}(px_1^{-(n-j)}, \cdots, px_n^{-(n-j)}). \tag{5.6}$$

It is immediate from (5.1) and (5.2) that

$$\bar{S}_{n,j}(x_1, \cdots, x_n) = F_{n,j}(x_1, \cdots, x_n; 1, \cdots, 1), \tag{5.7}$$

so that $\bar{S}_{n,j}$ is expressible in terms of F_{nj}. We can also express $F_{n,j}$ in terms of $\bar{S}_{n,j}$. Indeed it follows from (5.3) that

$$\begin{aligned}&F_{n,j}(x_1, \cdots, x_n; y_1^{n-j}, \cdots, y_n^{n-j}) \\ &\quad= \sum x_1^{a_1} \cdots x_n^{a_n} y_1^{s-ja_1} \cdots y_n^{s-jz_n} \qquad (s = a_1 + \cdots + a_n) \\ &\quad= \sum (x_1y_1^{-j}q)^{a_1} \cdots (x_ny_n^{-j}q)^{a_n} \qquad (q = y_1y_2 \cdots y_n).\end{aligned}$$

Thus

$$F_{n,j}(x_1, \cdots, x_n; y_1^{n-j}, \cdots, y_n^{n-j}) = \bar{S}_{n,j}(x_1y_1^{-j}q, \cdots, x_ny_n^{-j}q), \tag{5.8}$$

where $q = y_1y_2 \cdots y_n$.

We have also

$$\begin{aligned}&F_{n,j}(x_1^j, \cdots, x_n^j; y_1, \cdots, y_n) \\ &\quad= \sum x_1^{s-(n-j)b_1} \cdots x_n^{s-(n-j)b_n} y_1^{b_1} \cdots y_n^{b_n} \qquad (s = b_1, \cdots, b_n) \\ &\quad= \sum (x_1^{-(n-j)}y_1p)^{b_1} \cdots (x_n^{-(n-j)}y_np)^{b_n} \qquad (p = x_1x_2 \cdots x_n).\end{aligned}$$

It follows that

$$F_{n,j}(x_1^j, \cdots, x_n^j; y_1, \cdots, y_n) = \bar{S}_{n,n-j}(x_1^{-(n-j)}y_1p, \cdots, x_n^{-(n-j)}y_np), \tag{5.9}$$

where $p = x_1x_2 \cdots x_n$.

Comparison of (5.9) with (5.8) again gives (5.6).

We shall now show that when j and n are relatively prime, $\bar{S}_{n,j}$ and $F_{n,j}$ can be "evaluated". It follows from (5.3) that when $(j, n) = 1$ then

$$\begin{aligned}a_1 \equiv a_2 \equiv \cdots \equiv a_n &\quad (\mathrm{mod}\; n-j) \\ b_1 \equiv b_2 \equiv \cdots \equiv b_n &\quad (\mathrm{mod}\; j).\end{aligned}$$

Moreover, by (5.3),

$$na_i \equiv s \ (\mathrm{mod}\; n-j), \quad nb_i \equiv s \ (\mathrm{mod}\; j) \qquad (1 \leqq i \leqq n),$$

where $s = a_1 + \cdots + a_n = b_1 + \cdots + b_n$. Hence if we define n' by means of

$$nn' \equiv 1 \ (\mathrm{mod}\; j(n-j)),$$

we get

$$a_i \equiv n's \ (\mathrm{mod}\; n-j), \quad b_i \equiv n's \ (\mathrm{mod}\; j) \qquad (1 \leqq i \leqq n).$$

Thus we may put, for $1 \leqq i \leqq n$,

$$(5.10)\qquad \begin{aligned} a_i &= (n-j)c_i + u \quad && (0 \leqq u < n-j), \\ b_i &= jd_i + v && (0 \leqq v < j), \end{aligned}$$

where c_i and d_i are nonnegative. Substituting from (5.10) in (5.3) and (5.4), we get

$$(5.11)\qquad \left.\begin{aligned} c_1 + c_2 + \cdots + c_n + u - v &= j(c_i + d_i) \\ d_1 + d_2 + \cdots + d_n + v - u &= (n-j)(c_i + d_i) \end{aligned}\right. \quad (1 \leqq i \leqq n).$$

It follows immediately from (5.11) that $c_i + d_i$ is independent of i; we put

$$(5.12)\qquad t = c_i + d_i \qquad (1 \leqq i \leqq n),$$

so that (5.11) becomes

$$(5.13)\qquad \begin{aligned} c_1 + c_2 + \cdots + c_n + u - v &= jt, \\ d_1 + d_2 + \cdots + d_n + v - u &= (n-j)t. \end{aligned}$$

Note that adding together corresponding sides of the two equations in (5.13) results in a tautology.

Suppose t, u, v fixed. Choose $c_1, c_2, \cdots, c_n$ so that $0 \leqq c_i \leqq t$, $1 \leqq i \leqq n$, and the first of (5.13) is satisfied. Then the d_i obtained from (5.12) will satisfy the second of (5.13).

Returning to (5.2) it follows from (5.10) that

$$(5.14)\qquad F_{n,j} = \sum p^u q^v x_1^{(n-j)c_1} \cdots x_n^{(n-j)c_n} y_1^{jd_1} \cdots y_n^{jd_n},$$

where

$$p = x_1 x_2 \cdots x_n, \qquad q = y_1 y_2 \cdots y_n;$$

the summation in (5.14) is over t, u, v, c_i, d_i satisfying (5.12) and (5.13). It will be convenient to define

$$(5.15)\qquad \Phi_{t,u,v} = \sum x_1^{(n-j)c_1} \cdots x_n^{(n-j)c_i} y_1^{jd_1} \cdots y_n^{jd_n},$$

where t, u, v are fixed and the summation on the right is over c_i, d_i satisfying (5.12) and (5.13). Thus, by (5.14),

$$(5.16)\qquad F_{n,j} = \sum_{t=0}^{\infty} \sum_{u=0}^{n-j-1} \sum_{v=0}^{j-1} \Phi_{t,u,v}.$$

To evaluate $\Phi_{t,u,v}$ consider the product

$$(5.17)\qquad \prod_{i=1}^{n} (1 + x_i z + x_i^2 z^2 + \cdots + x_i^t z^t) = \sum_{k=0}^{nt} \sigma_k^{(t)} z^k,$$

where $\sigma_k^{(t)} = \sigma_k^{(t)}(x_1, \cdots, x_n)$ is a symmetric polynomial of weight k. It follows that

$$(5.18)\qquad \Phi_{t,u,v} = q^{njt} \sigma_{jt-u+v}^{(k)} (x_1^{n-j} y_1^{-j}, \cdots, x_n^{n-j} y_n^{-j}).$$

Therefore, by (5.14) and (5.16),

$$(5.19)\quad F_{n,j} = \sum_{t=0}^{\infty} \sum_{u=0}^{n-j-1} \sum_{v=0}^{j-1} p^u q^{njt+v} \sigma_{jt-u+v}^{(t)} (x_1^{n-j} y_1^{-j}, \cdots, x_n^{n-j} y_n^{-1}) \qquad (0 < j < n).$$

We may accordingly state

THEOREM 5.1. *The series*

$$F_{n,j} \equiv F_{n,j}(x_1, \cdots, x_n; y_1, \cdots, y_n) \qquad ((n,j) = 1)$$

defined by (5.2) *and* (5.3) *satisfies* (5.19), *where*

$$\sigma_k^{(t)} = \sigma_k^{(t)}(x_1, x_2, \cdots, x_n) \qquad (0 \leqq k \leqq nt)$$

is a symmetric polynomial in the x_i *of weight* k.

In particular, for $j = 1$, (5.19) reduces to

$$F_{n,1} = \sum_{t=0}^{\infty} \sum_{u=0}^{n-2} p^u q^{nt} \sigma_{t-u}^{(t)}(x_1^{n-1}, \cdots, x_n^{n-1}; y_1, \cdots, y_n)$$

$$= \sum_{u=0}^{n-2} p^u \sum_{t=0}^{\infty} q^{nt} \sigma_t^{(t+u)}(x_1^{n-1} y_1^{-1}, \cdots, x_n^{n-1} y_n^{-1}).$$

It is clear from (5.17) that

$$\sigma_k^{(t)} = \sigma_k^{(t')} \qquad (k \leqq t < t').$$

Hence

$$\sum_{t=0}^{\infty} q^{nt} \sigma_t^{(t+u)}(x_1^{n-1} y_1, \cdots, x_n^{n-1} y_n)$$

$$= \prod_{i=1}^{n} (1 + (x_i y_i)^{n-1} + (x_i y_i)^{2n-2} + \cdots)$$

$$= \prod_{i=1}^{n} (1 - (x_i y_i)^{n-1})^1.$$

It follows that

$$F_{n,1}(x_1, \cdots, x_n; y_1, \cdots, y_n) = \frac{1 + p + \cdots + p^{n-2}}{\prod_{i=1}^{n} (1 - (x_i y_i)^{n-1})}. \tag{5.20}$$

For $y_1 = \cdots = y_n = 1$, this reduces to

$$\bar{S}_{n,1}(x_1, \cdots, x_n) = \frac{1 + p + \cdots + p^{n-2}}{\prod_{i=1}^{n} (1 - x_i^{n-1})};$$

hence by (5.6)

$$S_{n,n-1}(p x_1^{-(n-1)}, \cdots, p x_n^{-(n-1)}) = \frac{1 + p + \cdots + p^{n-2}}{\prod_{i=1}^{n} (1 - x_i^{n-1})}, \tag{5.21}$$

since $\bar{S}_{n,n-1} = S_{n,n-1}$. Now if we put

$$\bar{x}_i = p x_i^{-(n-1)} \qquad (i = 1, 2, \cdots, n), \qquad \bar{p} = \bar{x}_1 \bar{x}_2 \cdots \bar{x}_n,$$

then

$$\bar{p} = p^n p^{-(n-1)} = p, \qquad x_i^{-(n-1)} = \bar{p}/\bar{x}_i.$$

Thus (5.21) is equivalent to (4.15).

This furnishes a partial check on (5.19). Unfortunately the method does not enable us to simplify (5.19) for $1 < j < n-1$.

6. Some applications. In (4.15) take $x_1 = \cdots = x_n = x$ and we get

$$S_{n,n-1}(x) = \frac{1 + x^n + x^{2n} + \cdots + x^{n(n-2)}}{(1 - x^{n-1})^n}, \tag{6.1}$$

where

$$S_{n,n-1}(x)=S_{n,n-1}(x, x, \cdots, x).$$

The right hand side of (6.1) is equal to

$$\sum_{j=0}^{n-2} x^{nj} \sum_{k=0}^{\infty} \binom{n+k-1}{k} x^{(n-1)k} = \sum_{m=0}^{\infty} x^m \sum_{nj+(n-1)k=m} \binom{n+k-1}{n-1}. \tag{6.2}$$

The equation

$$nj+(n-1)k=m \qquad (0 \leqq j < n-1;\ k \geqq 0) \tag{6.3}$$

has at most one solution (j, k). For assume a second solution (j', k'). Then

$$n(j-j') = -(n-1)(k-k'),$$

which implies $n-1 \mid j-j'$ and therefore $j=j'$.

Given the solution (j, k) of (6.3) we may write

$$m=(n-1)(j+k)+j.$$

Hence (6.3) is solvable if and only if

$$m=(n-1)r+s \qquad (0 \leqq s < n-1;\ s \geqq r). \tag{6.4}$$

It follows that the right hand side of (6.2) is equal to

$$\sum_{m=0}^{\infty} \binom{n+r-s-1}{n-1} x^m, \tag{6.5}$$

where r, s are uniquely determined by (6.4).

As for the left hand side of (6.1), we have

$$S_{n,n-1}(x)=\sum x^{a_1+a_2+\cdots+a_n}$$

where the summation is over nonnegative x_i such that

$$(n-1)a_i \leqq a_1+a_2+\cdots+a_n \qquad (1 \leqq i \leqq n).$$

It follows that the coefficient of x^m is equal to the number of solutions of the system

$$\begin{aligned} &a_1+a_2+\cdots+a_n=m, \\ &(n-1)a_i \leqq m \qquad\qquad (1 \leqq i \leqq n). \end{aligned} \tag{6.6}$$

If we denote the number of solutions of (6.6) by $N_n(m)$ we have therefore

$$N_n(m)=\binom{n+r-s-1}{n-1}, \tag{6.7}$$

where r and s are determined by (6.4).

On the other hand the number of solutions of the system

$$a_1+a_2+\cdots+a_n=m, \qquad a_i \leqq t, \tag{6.8}$$

is evidently equal to the coefficient of x^m in

$$\begin{aligned} (1+x+\cdots+x^t)^n = \left(\frac{1-x^{t+1}}{1-x}\right)^n &= \sum_{j=0}^{n} (-1)^j \binom{n}{j} x^{j(t+1)} \sum_{k=0}^{\infty} \binom{n+k-1}{n-1} x^k \\ &= \sum_{m=0}^{\infty} x^m \sum_{j(t+1)+k=m} (-1)^j \binom{n}{j} \binom{n+k-1}{n-1}. \end{aligned}$$

For $t=[m/(n-1)]$, (6.7) reduces to (6.6). Thus we obtain the binomial identity

$$\sum_{j(t+1)\leqq m}(-1)^j\binom{n}{j}\binom{n+m-j(t+1)-1}{n-1}=\binom{n+r-s-1}{n-1}, \tag{6.9}$$

where $t=[m/(n-1)]$ and r, s, if they exist, are determined by

$$m=(n-1)r+s \qquad (0\leqq s<n-1;\ s\leqq r).$$

We remark that, by finite differences,

$$\sum_{j=0}^{\infty}(-1)^j\binom{n}{j}\binom{n+m-j(t+1)-1}{n-1}=0.$$

Moreover, in the sum on the left of (6.9), $j=n$ provided

$$n\left(\left[\frac{m}{n-1}\right]+1\right)\leqq m,$$

so that

$$\frac{m}{n-1}<\frac{m}{n-1}+1\leqq\frac{m}{n},$$

which is impossible. Similarly $j=n-1$ is ruled out.

Put

$$m=t(n-1)+s, \qquad 0\leqq s<n-1.$$

Then the maximum admissible value of j is determined by

$$(t+1)j\leqq t(n-1)+s,$$

that is

$$j\leqq\frac{t(n-1)+s}{t+1}=n-1-\frac{n-1-s}{t+1}.$$

Hence for large n, there may be approximately $n/(t+1)$ excluded terms in the left number of (6.9).

It would be of interest to give a direct proof of (6.9). The formula is apparently in neither [1] nor [3].

7. Other applications. In (4.18) take

$$x_1=\cdots=x_n=x, \qquad z_1=\cdots=z_n=z.$$

The extreme right member of (4.18) reduces to

$$\begin{aligned}&\frac{1+x^n e^{nz}+\cdots+x^{n(n-2)}e^{n(n-2)z}}{(1-x^{n-1})^n}\\ &=\sum_{j=0}^{n-2}x^{jn}e^{jnz}\sum_{k=0}^{\infty}\binom{n+k-1}{n-1}x^{k(n-1)}\\ &=\sum_{m=0}^{\infty}x^m\sum_{jn+k(n-1)=m}\binom{n+k-1}{n-1}e^{jnz}.\end{aligned}$$

Exactly as above, the inner sum on the extreme right is equal to

$$\binom{n+r-s-1}{n-1}e^{snz}, \tag{7.1}$$

where r, s are uniquely determined by

$$m=(n-1)r+s \qquad (0 \leqq s<n-1;\, s \leqq r). \tag{7.2}$$

If this cannot be satisfied the sum in question vanishes.

The sum (7.1) is equal to

$$\binom{n+r-s-1}{n-1} \sum_{f=0}^{\infty} (sn)^f \frac{z^f}{f!}. \tag{7.3}$$

Hence if we put

$$N_n(m,f)=\sum \frac{f!}{i_1! \cdots i_n!} a_1^{i_1} \cdots a_n^{i_n}, \tag{7.4}$$

where the summation is over all nonnegative $a_1, \cdots, a_n$, $i_1, \cdots, i_n$ satisfying

$$\begin{aligned} &a_1+a_2+\cdots+a_n=m \\ &(n-1)a_j \leqq m \qquad\qquad (1 \leqq j \leqq n) \\ &i_1+i_2+\cdots+i_n=f, \end{aligned} \tag{7.5}$$

it follows from (7.3) that

$$N_n(m,f)=\binom{n+r-s-1}{n-1}(sn)^f, \tag{7.6}$$

where r, s are determined by (7.2)

Comparison of (7.6) with (6.7) gives

$$N_n(m,f)=(sn)^f N_n(m). \tag{7.7}$$

In the next place, it is not difficult to show that

$$N_n(m,f,t)=\sum \frac{f!}{i_1! \cdots i_n!} a_1^{i_1} \cdots a_n^{i_n},$$

where the summation is over all nonnegative $a_1, \cdots, a_n$; $i_1, \cdots, i_n$ satisfying

$$\begin{aligned} &a_1+a_2+\cdots+a_n=m, \\ &a_j \leqq t \qquad\qquad (1 \leqq h \leqq n) \\ &i_1+i_2+\cdots+i_n=f, \end{aligned}$$

is equal to the coefficient of x^m in

$$(sn)^f \left(\frac{1-x^{t+1}}{1-x}\right)^n.$$

Hence, taking $t=[m/(u-1)]$, we get

$$N_n(m,f)=(sn)^f \sum_{j(t+1) \leqq m} (-1)^j \binom{n}{j} \binom{n+m-j(t+1)-1}{n-1}. \tag{7.8}$$

Thus comparison of (7.8) with (7.6) again leads to the identity (6.9).

Another simple application is implied by (2.9). We state this equation in the form

$$\sum x^{a_1+a_2+a_3+a_4}=\frac{1+2x^2+4x^3+x^4}{(1-x^2)^4}, \tag{7.9}$$

where the summation on the left is over all nonnegative a_1, a_2, a_3, a_4 such that

$$2a_i \leqq a_1+a_2+a_3+a_4 \qquad (i \leqq i \leqq 4).$$

Let $f_4(m)$ denote the number of nonnegative solutions of the system

$$\begin{aligned} &a_1+a_2+a_3+a_4 = m \\ &2a_i \leqq m \qquad\qquad (1 \leqq i \leqq 4), \end{aligned}$$

so that the left hand side of (7.9) is equal to

$$\sum_{m=0}^{\infty} f_4(m)x^m.$$

The right hand side of (7.9) is equal to

$$(1+2x^2+4x^3+x^4)\sum_{m=0}^{\infty}\binom{m+3}{3}x^{2m}.$$

It follows that

$$\begin{aligned} &f_4(2m+1) = 4\binom{m+2}{3}, \\ (7.10)\qquad &f_4(2m) = \binom{m+3}{3}+2\binom{m+2}{3}+\binom{m+1}{3} = \tfrac{1}{3}(m+1)(2m^2+4m+3). \end{aligned}$$

8. Consider the multiple Dirichlet series

(8.1) $$\Phi_{n,j}(s_1, s_2, \cdots, s_n) = \sum m_1^{-s_1}m_2^{-s_2}\cdots m_n^{s_n} \qquad (1 \leqq j \leqq n).$$

where the summation is over all positive integers $m_1, m_2, \cdots, m_n$ satisfying

(8.2) $$m_i^j \mid m_1m_2\cdots m_n \qquad (1 \leqq i \leqq n).$$

Corresponding to $F_{n,j}(x_1, \cdots, x_n; y_1, \cdots, y_n)$ we may define

$$\Phi_{n,j}(s_1, \cdots, s_n; s_1', \cdots, s_n') = \sum m_1^{-s_1}\cdots m_n^{-s_n}m_1'^{-s_1}\cdots m_n'^{-s_n} \qquad (0 \leqq j \leqq n),$$

where the summation is over all positive integers $m_1, \cdots, m_n; m_1', \cdots, m_n'$ satisfying

$$m_1m_2\cdots m_n = m_i^j m_i'^{n-j} \qquad (1 \leqq i \leqq n).$$

However we shall not take the space to discuss properties of $\Phi_{n,j}(s_1, \cdots, s_n; s_1', \ldots, s_n')$.

For $j = 1$, the conditions (8.2) are satisfied and therefore

(8.3) $$\Phi_{n,1}(s_1, s_2, \cdots, s_n) = \zeta(s_1)\zeta(s_2)\cdots\zeta(s_n),$$

where $\zeta(s)$ is the Riemann zeta function. For $j = n$, (8.2) implies

$$m_1^n m_2^n \cdots m_n^n \mid (m_1m_2\cdots m_n)^n$$

and so each vertical bar in (8.2) becomes equality. It follows that $m_1 = m_2 = \cdots = m_n$. Hence

(8.4) $$\Phi_{n,n}(s_1, s_2, \cdots, s_n) = \zeta(s_1+s_2+\cdots+s_n),$$

For $1 \leqq j \leqq n$, it is easily seen that

$$\Phi_{n,j}(s_1, s_2, \cdots, s_n) = \prod_p \Phi_{n,j,p}(s_1, s_2, \cdots, s_n),$$

where the product extends over all primes p, and

$$\Phi_{n,j,p}(s_1, s_2, \cdots, s_n) = \sum p^{-a_1 s_1 - a_2 s_2 - \cdots - a_n s_n}$$

where the summation is over all nonnegative $a_1, a_2, \cdots, a_n$ such that

$$ja_i \leqq a_1 + a_2 + \cdots + a_n \qquad (1 \leqq i \leqq n).$$

It follows at once that

$$\Phi_{n,j,p}(s_1, s_2, \cdots, s_n) = s_{n,j}(p^{-s_1}, p^{-s_2}, \cdots, p^{-s_n}), \tag{8.5}$$

where $S_{n,j}$ is defined by (4.1). Therefore

$$\Phi_{n,j}(s_1, s_2, \cdots, s_n) = \prod_p S_{n,j}(p^{-s_1}, p^{-s_2}, \cdots, p^{-s_n}). \tag{8.6}$$

In particular, for $j = n-1$, we have by (4.15),

$$\begin{aligned} &S_{n,n-1}(p^{-s_1}, p^{-s_2}, \cdots, p^{-s_n}) \\ &= \frac{1 + p^{-\sigma} + \cdots + p^{-(n-2)\sigma}}{(1 - p^{-\sigma+s_1})(1 - p^{-\sigma+s_2}) \cdots (1 - p^{-\sigma+n})} \\ &\quad \cdot \frac{1 - p^{-(n-1)s}}{(1 - p^{-\sigma})(1 - p^{-\sigma+s_1})(1 - p^{-\sigma+s_2}) \cdots (1 - p^{-\sigma+s_n})}, \end{aligned}$$

where

$$\sigma = s_1 + s_2 + \cdots + s_n.$$

It therefore follows from (8.6) that

$$\Phi_{n,n-1}(s_1, s_2, \cdots, s_n) = \frac{\zeta(\sigma)\zeta(\sigma - s_1)\zeta(\sigma - s_2) \cdots \zeta(\sigma - s_n)}{\zeta((n-1)\sigma)}. \tag{8.7}$$

We now take $s_1 = \cdots = s_n = s$. Then (8.7) reduces to

$$\Phi_{n,n-1}(s) \equiv \Phi_{n,n-1}(s_1, s_2, \cdots, s_n) = \frac{\zeta(ns)\zeta^n((n-1)s)}{\zeta(n(n-1)s)}, \tag{8.8}$$

while (8.1) yields

$$\Phi_{n,n-1}(s) = \sum_{m=1}^{\infty} \frac{\delta_n(m)}{m^s}, \tag{8.9}$$

where $\delta_n(m)$ is defined as the number of solutions of

$$\begin{aligned} &m_1 m_2 \cdots m_n = m \\ &m_i^{n-1} \mid m \qquad\qquad (1 \leqq i \leqq n). \end{aligned} \tag{8.10}$$

Put

$$\zeta^n(s) = \sum_{m=1}^{\infty} \frac{\tau_n(m)}{m^s},$$

so that $\tau_n(m)$ is the number of solutions $m_1, m_2, \cdots, m_n$ of $m_1 m_2 \cdots m_n = m$. Then

$$\begin{aligned} \frac{\zeta(ns)\zeta^n((n-1)s)}{\zeta(n(n-1)s)} &= \sum_{a=1}^{\infty} \frac{1}{a^{ns}} \sum_{b=1}^{\infty} \frac{\tau_n(b)}{b^{(n-1)s}} \sum_{c=1}^{\infty} \frac{\mu(c)}{c^{n(n-1)s}} \\ &= \sum_{m=1}^{\infty} \frac{1}{m^s} \sum_{a^n b^{n-1} c^{n(n-1)} = m} \tau_n(b)\mu(c). \end{aligned}$$

Therefore, by (8.8) and (8.9), we have

$$\delta_n(m) = \sum_{a^n b^{n-1} c^{n(n-1)} = m} \tau_n(b)\mu(c). \tag{8.11}$$

For a discussion of arithmetic functions see for example [2, Chap. 4].

REFERENCES

[1] H. W. GOULD, *Combinatorial Identities*, revised edition, published by the author, Morgantown, WV, 1972.

[2] I. NIVEN AND H. S. ZUCKERMAN, *An Introduction to the Theory of Numbers*, third edition, John Wiley, New York, 1972.

[3] J. RIORDAN, *Combinatorial Identities*, John Wiley, New York, 1968.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0014 $01.00/0

ANALYTIC FUNCTIONS RELATED TO THE DISTRIBUTIONS OF EXPONENTIAL GROWTH*

RICHARD D. CARMICHAEL†

Abstract. We study the relationship between certain classes of analytic functions in tubes, the distributions of exponential growth $\mathscr{K}'_p$, $p \geqq 1$, and the Fourier transform spaces K'_p, $p \geqq 1$, of such distributions. Representations of the analytic functions are obtained in terms of the Fourier–Laplace transform of distributions in $\mathscr{K}'_p$, and when the analytic functions are considered as elements in K'_p we obtain representations of them in terms of the Fourier transform in K'_p of certain elements in $\mathscr{K}'_p$. In every case the distributions which yield these representations are analyzed. Further, we obtain strong boundedness properties of the analytic functions when considered as elements of K'_p, and certain of our analytic functions are shown to have distributional boundary values in the strong (and weak) topology of K'_p. Our results are motivated by analysis of V. S. Vladimirov who has considered similar problems to those that we study but for spaces of analytic functions that are properly contained in the spaces that we define in this paper. The spectral functions of Vladimirov which are associated with his analytic functions are distributions in $\mathscr{S}'$ which are defined by continuous functions of power increase in $\mathbb{R}^n$ while the distributions which correspond to spectral functions in this paper are distributions in $\mathscr{K}'_p$ which are defined by continuous functions of exponential growth in $\mathbb{R}^n$, a more general situation than that of Vladimirov.

1. Introduction. V. S. Vladimirov [17, Chap. 5] has obtained considerable information concerning the relation of analytic functions defined in tubes and the tempered distributions $\mathscr{S}'$. He has characterized the analytic functions in tubes which obtain distributional boundary values in $\mathscr{S}'$ and has studied the relation between certain classes of analytic functions and the associated spectral functions [17, p. 230] of elements of the class.

In this paper we shall be concerned primarily with the distributions of exponential growth $\mathscr{K}'_p$, $p \geqq 1$, and their Fourier transforms K'_p, $p \geqq 1$, which are larger classes of distributions than $\mathscr{S}'$. Sebastião e Silva [13], [14] was the first to define distributions of this type. The spaces $\mathscr{K}'_1 \equiv \Lambda_\infty$ and $K'_1 \equiv \mathscr{k}'$ were defined by Sebastião e Silva in 1-dimension and were later extended to n-dimensions and studied by Hasumi [8], Yoshinaga [18], and Zieleźny [19]. Recently Sampson and Zieleźny [10] have defined and studied the distributions $\mathscr{K}'_p$, $p > 1$, and their Fourier transforms K'_p, $p > 1$, in n-dimensions.

The principal motivation of this paper is the work of Vladimirov first contained in [16] and later contained in [17, § 26.4]. We define two different types of spaces of analytic functions in tubes both of which are more general than the $H_p(A; C)$ spaces of [17, p. 238] by considering functions whose growth is known for Im (z) in some sense bounded away from the origin in $\mathbb{R}^n$, a situation motivated for this author by the researches of Beltrami and Wohlers [1], Lauwerier [9], and Swartz [15] and also motivated by properties of the Fourier–Laplace transform of certain elements in the spaces of distributions of exponential growth $\mathscr{K}'_p$, $p \geqq 1$. We study the relationship between our new analytic functions and the distributions $\mathscr{K}'_{p,}$, $p \geqq 1$, and their Fourier transforms K'_p, $p \geqq 1$, both of which properly contain $\mathscr{S}'$. We show that elements in our analytic function spaces can be represented as the Fourier–Laplace transform of certain distributions in $\mathscr{K}'_p$ and analyze these distributions, as Vladimirov has done with respect to his $H_p(A; C)$ spaces and the associated spectral functions. We further present representations of the analytic functions as elements in K'_p in terms of Fourier transforms in K'_p of certain elements in $\mathscr{K}'_p$ and present boundedness information for

* Received by the editors September 15, 1977, and in revised form March 27, 1978.
† Department of Mathematics, Iowa State University, Ames, Iowa 50011. Permanent address: Department of Mathematics, Wake Forest University, Winston-Salem, North Carolina 27109.

the analytic functions considered as elements in K'_p. Further, for one of our new type of spaces of analytic functions we show that the elements in these spaces obtain distributional boundary values in the strong (and weak) topology of K'_p and present a general condition under which analytic functions in tubes have distributional boundary values in K'_p in the strong (and weak) topology of K'_p.

2. Notation and definitions. In this section we state the basic notation to be used in this paper. Let $t \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. We define $\langle t, y\rangle = t_1y_1 + \cdots + t_ny_n$ and similarly define $\langle t, z\rangle$, $t \in \mathbb{R}^n$, $z \in \mathbb{C}^n$. Let α denote an n-tuple of nonnegative integers. D_t^α denotes the differential operator $D_t^\alpha = D_1^{\alpha_1}D_2^{\alpha_2}\cdots D_n^{\alpha_n}$ where $D_j = -1/(2\pi i)(\partial/\partial t_j)$, $j = 1, \cdots, n$. Similarly we define D_z^α. If $k = (k_1, k_2, \cdots, k_n)$ is an n-tuple of integers, we define $t^k = t_1^{k_1}\cdots t_n^{k_n}$ with similar definition for z^k. When the components of k are nonnegative integers we also define $|k| = k_1 + k_2 + \cdots + k_n$ and $k! = k_1!k_2!\cdots k_n!$. Throughout this paper $N(0, m)$ will denote the closed ball in $\mathbb{R}^n$ of radius $m > 0$ having center at the origin.

A set $C \subset \mathbb{R}^n$ is a cone (with vertex at zero) if $y \in C$ implies $\lambda y \in C$ for all positive scalars λ. The intersection of C with the unit sphere $|y| = 1$ in $\mathbb{R}^n$ is called the projection of C and is denoted pr (C). Let C' be a cone such that pr $(\bar{C}') \subset$ pr (C); C' will be called a compact subcone of C. $O(C)$ will denote the convex envelope of C. T^C will denote $T^C = \mathbb{R}^n + iC$, a subset of $\mathbb{C}^n$. T^C will be called a tubular cone if the cone C is open and will be called a tubular radial domain if C is both open and connected. The function

$$u_C(t) = \sup_{y \in \mathrm{pr}(C)} (-\langle t, y\rangle)$$

is the indicatrix of the cone C. Throughout the paper the sets C^* and C_* will denote $C^* = \{t \in \mathbb{R}^n : u_C(t) \leqq 0\}$ and $C_* = \mathbb{R}^n \backslash C^*$. It follows that $C^* = \{t : \langle t, y\rangle \geqq 0, y \in C\}$ and $C_* = \{t : \langle t, y\rangle < 0, y \in C\}$. For any cone C, C^* will be called the dual cone of C. Note that both C^* and C_* are cones in $\mathbb{R}^n$. The number

$$\rho_C = \sup_{t \in C_*} \frac{u_{O(C)}(t)}{u_C(t)}$$

characterizes the nonconvexity of the cone C [17, p. 220], and we have $\rho_C \geqq 1$ always [17, p. 220]. We shall be considering open cones having a finite number of components in this paper; so throughout this paper $1 \leqq \rho_C < \infty$ [17, p. 220, Lemma 3] for all cones C.

Let C be a cone. We have [17, § 25.1]

$$(2.1) \qquad -\langle t, y\rangle \leqq |y| u_{O(C)}(t), \qquad u_{O(C)}(t) \leqq \rho_C u_C(t), \qquad t \in C_*, \quad y \in O(C).$$

Let C be an open connected cone and let C'_* be a compact subcone of C_*. There exists a number $\xi > 0$ depending on C'_* such that [17, eq. (28), p. 241]

$$(2.2) \qquad \xi|t| \leqq u_C(t) \leqq |t|, \qquad t \in C'_*.$$

Let $\phi(t) \in L^1(\mathbb{R}^n)$. We define the Fourier transform of $\phi(t)$ by

$$\hat{\phi}(x) = \mathscr{F}[\phi(t); x] = \int_{\mathbb{R}^n} \phi(t)\, e^{2\pi i\langle x, t\rangle}\, dt$$

while the inverse Fourier transform of $\phi(t)$ is

$$\mathscr{F}^{-1}[\phi(t); x] = \int_{\mathbb{R}^n} \phi(t)\, e^{-2\pi i\langle x, t\rangle}\, dt.$$

Throughout this paper all needed definitions and terminology concerning distributions, such as support of a distribution, will be that of L. Schwartz [11]. We shall denote the support of a function f and of a distribution V by supp (f) and supp (V). All terminology from the theory of topological vector spaces and their dual spaces used in this paper such as bounded set in a topological vector space and strongly bounded set in a dual space, can be found in Edwards [4] and Friedman [5, Chap. 1].

The notion of distributional boundary value of an analytic function in the weak or strong topology of a distribution (generalized function) space, with the function being analytic in a tubular radial domain, will be exactly the same in this paper as discussed in [2, p. 767, last paragraph].

3. The test spaces $\mathcal{K}_p$ and K_p, $p \geqq 1$. $\mathcal{K}_p$, $p \geqq 1$, is the space of all functions $\phi \in C^\infty(\mathbb{R}^n)$ for which the products $(\exp(k|t|^p) D_t^\alpha(\phi(t)))$, $|\alpha| \leqq k$, are bounded over $\mathbb{R}^n$ for $k = 0, 1, 2, \cdots$. We define the system of norms

$$ {}_p\|\phi\|_k = \sup_{\substack{t \in \mathbb{R}^n \\ |\alpha| \leqq k}} e^{k|t|^p} |D_t^\alpha(\phi(t))|, \qquad k = 1, 2, \cdots, \tag{3.1} $$

in $\mathcal{K}_p$. These norms have increasing strength; that is, ${}_p\|\phi\|_k \leqq {}_p\|\phi\|_{k+1}$, $k = 1, 2, \cdots$. If we put $M_k(t) = \exp(k|t|^p)$, $k = 1, 2, \cdots$, we recognize that $\mathcal{K}_p$ is an example of the $K\{M_k\}$ spaces of Gel'fand and Shilov [6, p. 86]. We define the topology of $\mathcal{K}_p$ by the norms (3.1), and $\mathcal{K}_p$ becomes a complete countably normed space [6, p. 88] and a locally convex topological vector space. We further note that $\{M_k(t)\} = \{\exp(k|t|^p)\}$ satisfies the properties (M) and (N) of [6, p. 111]. Thus the system of norms

$$ {}_p\|\phi\|'_k = \sup_{|\alpha| \leqq k} \int_{\mathbb{R}^n} e^{k|t|^p} |D_t^\alpha(\phi(t))| \, dt, \qquad k = 1, 2, \cdots, \tag{3.2} $$

is equivalent to the system (3.1) in $\mathcal{K}_p$ [6, pp. 111–112]. Any $C^\infty(\mathbb{R}^n)$ function which is $\mathcal{O}(\exp(k|t|^p))$ for some constant k is a multiplier in $\mathcal{K}_p$. The convergence of a sequence $\{\phi_v\}$ in $\mathcal{K}_p$ is determined by the topology of $\mathcal{K}_p$. Thus a sequence $\{\phi_v\}$ is said to converge to zero in $\mathcal{K}_p$ as $v \to v_0$ if $\phi_v \in \mathcal{K}_p$ for each v and $\lim_{v \to v_0} {}_p\|\phi_v\|_k = 0$ (or equivalently $\lim_{v \to v_0} {}_p\|\phi_v\|'_k = 0$) for each $k = 1, 2, \cdots$. It is easy to prove that $\mathcal{K}_p \subseteq \mathcal{K}_r$, $p \geqq r \geqq 1$; and if a sequence converges to zero in $\mathcal{K}_p$, the same sequence converges to zero in $\mathcal{K}_r$.

K_1 is the space of all $C^\infty(\mathbb{R}^n)$ functions which can be extended to $\mathbb{C}^n$ to be an entire analytic function such that

$$ {}_1\|\psi\|_k^* = \sup_{z \in V_k} (1+|z|)^k |\psi(z)| < \infty, \qquad k = 1, 2, \cdots, \tag{3.3} $$

where $V_k = \{z \in \mathbb{C}^n : |\mathrm{Im}(z_j)| \leqq k, j = 1, \cdots, n\}$. The topology defined by the semi-norms (3.3) makes K_1 into a locally convex topological vector space. Any $C^\infty(\mathbb{R}^n)$ function which can be extended to be an entire function of polynomial growth in any of the sets V_k is a multiplier in K_1. A sequence $\{\psi_v\}$ converges to zero in K_1 as $v \to v_0$ if $\psi_v \in K_1$ for each v and $\lim_{v \to v_0} {}_1\|\psi_v\|_k^* = 0$, $k = 1, 2, \cdots$.

K_p, $p > 1$, is the space of all $C^\infty(\mathbb{R}^n)$ functions which can be extended to be entire analytic functions such that

$$ {}_p\|\psi\|_k^* = \sup_{\substack{z \in \mathbb{C}^n \\ z = x+iy}} (1+|z|)^k \exp[-(|y|^q/k)] |\psi(z)| < \infty, \qquad k = 1, 2, \cdots, \tag{3.4} $$

where $1/p + 1/q = 1$. K_p becomes a locally convex topological vector space under the topology defined by the semi-norms (3.4). A sequence $\{\psi_v\}$ converges to zero in K_p as $v \to v_0$ if $\psi_v \in K_p$ for each v and $\lim_{v \to v_0} {}_p\|\psi_v\|_k^* = 0$, $k = 1, 2, \cdots$.

THEOREM 3.1. *The Fourier transform is a topological isomorphism of* $\mathcal{K}_p$ *onto* K_p, $p \geqq 1$. *The inverse Fourier transform is a topological isomorphism of* K_p *onto* $\mathcal{K}_p$, $p \geqq 1$. $\psi \in K_p$ *satisfies* $\psi(x) = \hat{\phi}(x)$, $\phi \in \mathcal{K}_p$, *if and only if* $\phi(t) = \mathcal{F}^{-1}[\psi(x); t]$.

Proof. For $p = 1$ the results have been obtained by Hasumi [8, pp. 97–99] and Zieleźny [19, p. 113] and for $p > 1$ by Sampson and Zieleźny [10].

The following lemma will be useful.

LEMMA 3.1. *The space* $\mathcal{K}_p$, $p \geqq 1$, *is a Fréchet nuclear space, a perfect space, and a Montel space.*

Proof. See [10], [6, Chap. 2], [7, pp. 178–182], and [5, Chaps. 1 and 2].

4. The distribution spaces $\mathcal{K}'_p$ and K'_p, $p \geqq 1$. The distribution spaces $\mathcal{K}'_p$ and K'_p, $p \geqq 1$, are the spaces of continuous linear functions on $\mathcal{K}_p$ and K_p, respectively.

Sampson and Zieleźny [10] have obtained the following characterization of $\mathcal{K}'_p$, $p \geqq 1$, in which $\mathcal{D}'$ refers to the distributions of L. Schwartz [11].

THEOREM 4.1 [10]. *A distribution* $V \in \mathcal{D}'$ *is in* $\mathcal{K}'_p$, $p \geqq 1$, *if and only if there exist an integer* $k \geqq 0$, *an* n*-tuple* α *of nonnegative integers, and a bounded continuous function* $f(t)$ *on* $\mathbb{R}^n$ *such that*

$$V_t = D_t^\alpha(e^{k|t|^p} f(t)). \tag{4.1}$$

Recall from § 3 that $\mathcal{K}_p \subseteq \mathcal{K}_r$, $p \geqq r \geqq 1$, and convergence of a sequence in $\mathcal{K}_p$ implies convergence of that same sequence in $\mathcal{K}_r$. It follows that $\mathcal{K}'_r \subseteq \mathcal{K}'_p$, $p \geqq r \geqq 1$, and the injection is continuous in the strong (and weak) topology. In particular $\mathcal{K}'_1 \subseteq \mathcal{K}'_p$ for all $p \geqq 1$. The reader is asked to remember this important fact.

Sampson and Zieleźny [10] have defined the Fourier transform of elements $V \in \mathcal{K}'_p$, $p \geqq 1$. Each $V \in \mathcal{K}'_p$ has a Fourier transform $U = \hat{V} = \mathcal{F}[V]$ which is an element in K'_p defined by the Parseval formula

$$\langle U, \psi \rangle = \langle V, \check{\phi} \rangle, \quad \phi \in \mathcal{K}_p, \quad \psi = \hat{\phi} \in K_p, \quad \check{\phi}(t) = \phi(-t). \tag{4.2}$$

Similarly the inverse Fourier transform of an element $U \in K'_p$, $p \geqq 1$, is the element $V = \mathcal{F}^{-1}[U] \in \mathcal{K}'_p$ defined by the relation

$$\langle V, \phi \rangle = \langle U, \check{\psi} \rangle, \quad \psi \in K_p, \quad \phi(t) = \mathcal{F}^{-1}[\psi(x); t] \in \mathcal{K}_p, \quad \check{\psi}(x) = \psi(-x). \tag{4.3}$$

Because of Theorem 3.1 we have the following basic result.

THEOREM 4.2 [10]. *The Fourier transform defined by* (4.2) *is a topological isomorphism of* $\mathcal{K}'_p$ *onto* K'_p, $p \geqq 1$. *The inverse Fourier transform defined by* (4.3) *is a topological isomorphism of* K'_p *onto* $\mathcal{K}'_p$, $p \geqq 1$. $U \in K'_p$ *satisfies* $U = \mathcal{F}[V]$, $V \in \mathcal{K}'_p$, *if and only if* $V = \mathcal{F}^{-1}[U]$.

5. Technical results. The results obtained in this section will be used to establish the main results of this paper which are contained in §§ 7–9. From § 2 recall the definitions of the dual cone $C^* = \{t: u_C(t) \leqq 0\}$ of a cone C, the cone $C_* = \mathbb{R}^n \backslash C^*$, the convex envelope $O(C)$ of C, and the convexity function ρ_C.

The first two lemmas in this section concern inequalities involving functions which are defined on compact subcones of the cone C_*.

LEMMA 5.1. *Let* γ *be an* n*-tuple of nonnegative integers. Let* $n \geqq 1$ *be an integer,* $R > 0$, *and* $q > 1$. *For the open connected cone* C *we have*

$$(1 + |t|)^{n+1+|\gamma|} \leqq M \exp[2\pi R(u_C(t))^q], \qquad t \in C'_* \subset C_*, \tag{5.1}$$

where C'_* *is an arbitrary compact subcone of* C_* *and the constant* M *depends on* C'_*.

Proof. Given $C'_* \subset C_*$ there exists a number $\xi > 0$ depending on C'_* such that (2.2) holds for $t \in C'_*$. For any $R > 0$, (2.2) yields

$$0 < \exp[2\pi R(\xi|t|)^q] \leqq \exp[2\pi R(u_C(t))^q], \qquad t \in C'_*. \tag{5.2}$$

But $((1+|t|)^{-n-1-|\gamma|}\exp[2\pi R(\xi|t|)^q]) \to \infty$ as $|t| \to \infty$ for $t \in \mathbb{R}^n$. This fact and (5.2) combine to show

$$((1+|t|)^{-n-1-|\gamma|}\exp[2\pi R(u_C(t))^q]) \to \infty \quad \text{as } |t| \to \infty, t \in C'_*. \tag{5.3}$$

Further, if we restrict our attention to any closed ball $N(0, m)$ of the origin in $\mathbb{R}^n$ of radius $m > 0$, it is evident that we can choose a constant $Q_m > 1$ depending on $m > 0$ such that

$$Q_m(1+|t|)^{-n-1-|\gamma|}\exp[2\pi R(\xi|t|)^q] \geqq 1, \qquad t \in N(0, m). \tag{5.4}$$

By (2.2) and (5.4) we now have

$$Q_m(1+|t|)^{-n-1-|\gamma|}\exp[2\pi R(u_C(t))^q] \geqq 1, \qquad t \in N(0, m) \cap C'_*. \tag{5.5}$$

The conclusion (5.1) can now be obtained from (5.3) and (5.5).

LEMMA 5.2. *Let C be an open connected cone and $y \in O(C)$ be arbitrary but fixed. Let p and q be real numbers related by $1/p + 1/q = 1$, $p > 1$, and let $B > 0$. Let C'_* be an arbitrary compact subcone of $C_* = \mathbb{R}^n \backslash C^*$ and let $\xi > 0$ be the number of* (2.2). *For every $\eta \in (0, 1)$ we have*

$$\begin{aligned} &\sup_{t \in C'_*} \exp[2\pi|y|(\rho_C)u_C(t) - 2\pi B(1-2q\eta)(u_C(t))^q] \\ &\qquad \leqq \sup_{\delta > 0} \exp[2\pi|y|(\rho_C)\delta - 2\pi B(\xi^q - 2q\eta)\delta^q]. \end{aligned} \tag{5.6}$$

Further if $\eta \in (0, 1)$ is now fixed such that $(\xi^q - 2q\eta) > 0$ for the fixed $\xi > 0$ and $q > 1$, the inequality (5.6) *is continued as*

$$\leqq \exp\left[\left(\frac{2\pi}{p}\right)\left(\frac{1}{qB(\xi^q - 2q\eta)}\right)^{p/q} \rho_C^p |y|^p\right]. \tag{5.7}$$

Proof. Using (2.2) we have for all $t \in C'_* \subset C_*$ that

$$-2\pi B(1-2q\eta)(u_C(t))^q \leqq -2\pi B(\xi^q - 2q\eta)|t|^q \tag{5.8}$$

and

$$2\pi|y|(\rho_C)u_C(t) \leqq 2\pi|y|(\rho_C)|t|, \qquad t \in C'_* \subset C_*. \tag{5.9}$$

Combining (5.8) and (5.9) yields

$$\begin{aligned} &\exp[2\pi|y|(\rho_C)u_C(t) - 2\pi B(1-2q\eta)(u_C(t))^q] \\ &\qquad \leqq \exp[2\pi|y|(\rho_C)|t| - 2\pi B(\xi^q - 2q\eta)|t|^q], \qquad t \in C'_* \subset C_*. \end{aligned} \tag{5.10}$$

Again using (2.2) we have $|t| \geqq u_C(t) > 0$ for $t \in C'_* \subset C_* = \mathbb{R}^n \backslash C^* = \{t\colon u_C(t) > 0\}$. By this and (5.10) the inequality (5.6) follows.

Recall that $\eta \in (0, 1)$ is arbitrary in (5.6); that is, (5.6) holds for every $\eta \in (0, 1)$. Now fix $\eta \in (0, 1)$ such that $(\xi^q - 2q\eta) > 0$ for the fixed $\xi > 0$ depending on C'_* and $q > 1$. Consider the function

$$f(\delta) = \exp[2\pi|y|(\rho_C)\delta - 2\pi B(\xi^q - 2q\eta)\delta^q], \qquad \delta > 0. \tag{5.11}$$

By considerations from the calculus and recalling that $1/p + 1/q = 1$, $p > 1$, $q > 1$, we see that $f(\delta)$ attains its maximum at

$$\delta=\left(\frac{|y|\rho_C}{qB(\xi^q-2q\eta)}\right)^{p/q}>0.$$

Thus

$$\begin{aligned}&\sup_{\delta>0}\exp[2\pi|y|(\rho_C)\delta-2\pi B(\xi^q-2q\eta)\delta^q]\\ (5.12)\qquad &\le\exp\left[2\pi|y|\rho_C\left(\frac{|y|\rho_C}{qB(\xi^q-2q\eta)}\right)^{p/q}-2\pi B(\xi^q-2q\eta)\left(\frac{|y|\rho_C}{qB(\xi^q-2q\eta)}\right)^{p}\right].\end{aligned}$$

Since $1/p+1/q=1$, $p>1$, $q>1$, then some simple, but tedious, arithmetic yields

$$\begin{aligned}&2\pi|y|\rho_C\left(\frac{|y|\rho_C}{qB(\xi^q-2q\eta)}\right)^{p/q}-2\pi B(\xi^q-2q\eta)\left(\frac{|y|\rho_C}{qB(\xi^q-2q\eta)}\right)^{p}\\ &\quad=\frac{2\pi}{p}\left(\frac{1}{qB(\xi^q-2q\eta)}\right)^{p/q}\rho_C^p|y|^p.\end{aligned}$$

From this last equality and (5.12) we obtain (5.7). The proof is complete.

The following two lemmas show various properties of certain integrals over arbitrary compact subcones of $C_*=\mathbb{R}^n\backslash C^*$.

LEMMA 5.3. *Let C be an open connected cone and let C' be an arbitrary open compact subcone of $O(C)$. Let C'_* be an arbitrary compact subcone of C_*. Let $B>0$ and $q>1$. Let $g(t)$ be a continuous function of $t\in\mathbb{R}^n$ which satisfies*

$$(5.13)\qquad |g(t)|\le K(C'_*,\eta)\exp[-2\pi B(1-q\eta)(u_C(t))^q],\qquad t\in C'_*\subset C_*,$$

for every $\eta\in(0,1)$, where $K(C'_*,\eta)$ is a constant depending on C'_* and on η. Let $z_0\in T^{C'}=\mathbb{R}^n+iC'$ be arbitrary but fixed and let $z\in N'(z_0,r)\subset T^{C'}$, where $N'(z_0,r)$ is an open neighborhood of z_0 with radius $r>0$ whose closure is in $T^{C'}$. Then for any n-tuple γ of nonnegative integers, the integral

$$h_\gamma(z)=\int_{C'_*}t^\gamma g(t)\,e^{2\pi i\langle z,t\rangle}\,dt$$

converges absolutely and uniformly for $z\in N'(z_0,r)$.

Proof. From (2.1) we have

$$(5.14)\qquad -\langle y,t\rangle\le|y|(\rho_C)u_C(t),\qquad t\in C_*,\quad y\in O(C).$$

For $z=x+iy\in N'(z_0,r)$ we can choose a real number $T>0$ such that $|y|\le T$, $y=\operatorname{Im}(z)$. Combining this fact with (5.13) and (5.14) we have for all $z=x+iy\in N'(z_0,r)$ that

$$\begin{aligned}|h_\gamma(z)|&\le K(C'_*,\eta)\int_{C'_*}|t^\gamma|\,e^{-2\pi\langle y,t\rangle}\exp[-2\pi B(1-q\eta)(u_C(t))^q]\,dt\\ (5.15)\qquad &\le K(C'_*,\eta)\int_{C'_*}\exp[2\pi T(\rho_C)u_C(t)-2\pi B(1-q\eta)(u_C(t))^q]\frac{(1+|t|)^{n+1+|\gamma|}}{(1+|t|)^{n+1}}\,dt\end{aligned}$$

where n is the dimension. Replace $|y|$ by T in (5.9); then (5.8) and (5.9) yield (5.10) with $|y|$ replaced by T. Thus using (5.1) with $R=Bq\eta$ and using (5.10) with T replacing $|y|$, we continue (5.15) and get

$$
\begin{aligned}
|h_\gamma(z)| &\leqq MK(C'_*, \eta) \int_{C'_*} \exp[2\pi T(\rho_C) u_C(t) - 2\pi B(1-q\eta)(u_C(t))^q] \\
&\qquad \cdot \exp[2\pi B q\eta (u_C(t))^q](1+|t|)^{-n-1}\, dt \qquad (5.16)\\
&\leqq MK(C'_*, \eta) \int_{C'_*} \exp[2\pi T(\rho_C)|t| - 2\pi B(\xi^q - 2q\eta)|t|^q](1+|t|)^{-n-1}\, dt
\end{aligned}
$$

where $\xi = \xi(C'_*) > 0$ is the number of (2.2), the constant M depending on C'_* is obtained from (5.1); and (5.16) holds for every $\eta \in (0, 1)$. We now fix $\eta \in (0, 1)$ such that $(\xi^q - 2q\eta) > 0$. With this choice of η, the proof of Lemma 5.2 combined with (5.16) now yields

$$
\begin{aligned}
|h_\gamma(z)| &\leqq MK(C'_*, \eta) \exp\left[\left(\frac{2\pi}{p}\right)\left(\frac{1}{qB(\xi^q - 2q\eta)}\right)^{p/q} \rho_C^p T^p\right] \int_{C'_*} (1+|t|)^{-n-1}\, dt \qquad (5.17)\\
&\leqq K'(C'_*, \eta) \exp\left[\left(\frac{2\pi}{p}\right)\left(\frac{1}{qB(\xi^q - 2q\eta)}\right)^{p/q} \rho_C^p T^p\right]
\end{aligned}
$$

for all $z \in N'(z_0, r)$, where $K'(C'_*, \eta)$ is a constant depending on the fixed C'_* and on the now fixed $\eta \in (0, 1)$ such that $(\xi^q - 2q\eta) > 0$. Everything on the right of the last inequality of (5.17) is independent of $z \in N'(z_0, r)$. Thus the integral defining $h_\gamma(z)$ converges absolutely and uniformly for $z \in N'(z_0, r)$. The proof is complete.

LEMMA 5.4. *Let C be an open connected cone and let C'_* be an arbitrary compact subcone of C_*. Let $g(t)$ be a continuous function of $t \in \mathbb{R}^n$ which satisfies* (5.13) *for $t \in C'_*$ with $B > 0$, $q > 1$, and $\eta \in (0, 1)$ being arbitrary. Let $\xi = \xi(C'_*) > 0$ be the number in* (2.2) *and let*

$$
A = \frac{1}{p}\left(\frac{1}{Bq}\right)^{p/q}, \qquad \frac{1}{p} + \frac{1}{q} = 1. \qquad (5.18)
$$

Then for any n-tuple γ of nonnegative integers

$$
\left| \int_{C'_*} t^\gamma g(t)\, e^{2\pi i \langle z, t\rangle}\, dt \right| \leqq K'(C'_*, \eta) \exp\left[2\pi A \left(\frac{1}{\xi^q - 2q\eta}\right)^{p/q} \rho_C^p |y|^p\right] \qquad (5.19)
$$

for $z = x + iy \in T^{O(C)} = \mathbb{R}^n + iO(C)$, where $K'(C'_*, \eta)$ is a constant depending on C'_* and on $\eta \in (0, 1)$ which is now fixed such that $(\xi^q - 2q\eta) > 0$.

Proof. Condition (5.14) holds for all $t \in C_*$ and $y \in O(C)$. Thus the analysis of (5.15), (5.16), and (5.17) with T replaced by $|y|$ proves (5.19).

In the following two lemmas we prove $(e^{-2\pi\langle y, t\rangle} g(t)) \in L^r$, $1 \leqq r < \infty$, under certain assumptions on the function $g(t)$ for y in a cone.

LEMMA 5.5. *Let C be an open connected cone and let C' be an arbitrary compact subcone of $O(C)$. Let $m = m(C') > 0$ be fixed depending on C'. Let $g(t)$ be a continuous function of $t \in \mathbb{R}^n$ such that*

$$
|g(t)| \leqq K(C', m) \exp[2\pi(\langle \Omega, t\rangle + A|\Omega|^p)], \qquad t \in \mathbb{R}^n, \qquad (5.20)
$$

where $A > 0$, $p > 1$, $K(C', m)$ is a constant depending on C' and on m, and (5.20) *holds independently of $\Omega \in (C' \setminus (C' \cap N(0, m)))$ for arbitrary $C' \subset O(C)$. Further, for any compact subcone $C'_* \subset C_*$ let $g(t)$ satisfy*

$$
|g(t)| \leqq M(C'_*, \eta) \exp[-2\pi B(1-q\eta)(u_C(t))^q], \qquad t \in C'_*, \qquad (5.21)
$$

for every $\eta \in (0, 1)$ *where*

$$B = \frac{1}{q}\left(\frac{1}{Ap}\right)^{q/p}, \qquad \frac{1}{p} + \frac{1}{q} = 1, \tag{5.22}$$

and $M(C'_*, \eta)$ *is a constant depending on* C'_* *and on* η. *Then*

$$e^{-2\pi\langle y,t\rangle} g(t) \in L^r, \qquad 1 \leqq r < \infty, \tag{5.23}$$

for $y \in (C' \backslash (C' \cap N(0, m)))$ *arbitrary but fixed.*

Proof. Let r be arbitrary, $1 \leqq r < \infty$. Let y be arbitrary but fixed in $(C' \backslash (C' \cap N(0, m)))$. Then $y \in C'$ and $|y| > m > 0$. Choose a real number ζ such that $0 < (m/|y|) < \zeta < 1$. Now put $\Omega = \zeta y$. Since $y \in C'$ and C' is a cone then $\Omega = \zeta y \in C'$. Further, $|\Omega| = \zeta |y| > m$; thus $\Omega = \zeta y \in (C' \backslash (C' \cap N(0, m)))$. Since (5.20) holds independently of $\Omega \in (C' \backslash (C' \cap N(0, m)))$ we choose $\Omega = \zeta y$ and obtain from (5.20) that

$$\begin{aligned} |e^{-2\pi\langle y,t\rangle} g(t)|^r &\leqq K^r(C', m)\, e^{-2\pi r\langle y,t\rangle} \exp\left[2\pi r(\zeta\langle y, t\rangle + A\zeta^p |y|^p)\right] \\ &\leqq K^r(C', m) \exp\left[2\pi r A\zeta^p |y|^p\right] \exp\left[-2\pi r(1-\zeta)\langle y, t\rangle\right] \end{aligned} \tag{5.24}$$

for all $t \in \mathbb{R}^n$.

For the arbitrary compact subcone C' of $O(C)$ we apply [17, Lemma 2, p. 223] and obtain a number $\delta > 0$ and an open cone $(C^*)'$ both depending on C' such that $C^* \subset (C^*)'$ and

$$\langle y, t\rangle \geqq \delta |y|\,|t|, \qquad y \in C', \quad t \in (C^*)'. \tag{5.25}$$

Since $(1-\zeta) > 0$ in (5.24) then using (5.24) and (5.25) we have

$$\begin{aligned} &\int_{(C^*)'} |e^{-2\pi\langle y,t\rangle} g(t)|^r \, dt \\ &\qquad \leqq K^r(C', m) \exp\left[2\pi r A\zeta^p |y|^p\right] \int_{(C^*)'} \exp\left[-2\pi r\delta(1-\zeta)|y|\,|t|\right] dt. \end{aligned} \tag{5.26}$$

From (5.26) and Schwartz [12, p. 39, Theorem 32] we obtain

$$\begin{aligned} &\int_{(C^*)'} |e^{-2\pi\langle y,t\rangle} g(t)|^r \, dt \\ &\qquad \leqq K^r(C', m) Z_n \exp\left[2\pi r A\zeta^p |y|^p\right] \int_0^\infty s^{n-1} \exp\left[-2\pi r\delta(1-\zeta)|y| s\right] ds \\ &\qquad = K^r(C', m) Z_n \exp\left[2\pi r A\zeta^p |y|^p\right](n-1)!(2\pi r\delta(1-\zeta)|y|)^{-n} \end{aligned} \tag{5.27}$$

where Z_n is the area of the unit sphere in $\mathbb{R}^n$ and we have integrated by parts $(n-1)$ times on the last integral in (5.27).

Let us now put $C'_* = \mathbb{R}^n \backslash (C^*)'$. Since $C^* \subset (C^*)'$ and $(C^*)'$ is an open cone then C'_* is a compact subcone of C_* and (5.21) holds for this C'_* by hypothesis. By analysis exactly like that in Lemmas 5.3 and 5.4 we have

$$\int_{C'_*} |e^{-2\pi\langle y,t\rangle} g(t)|^r \, dt \leqq M'(C'_*, \eta) \exp\left[2\pi r A\left(\frac{1}{\xi^q - 2q\eta}\right)^{p/q} \rho^p_C |y|^p\right] \tag{5.28}$$

where $M'(C'_*, \eta)$ is a constant depending on C'_* and on the $\eta \in (0, 1)$ which is now fixed such that $(\xi^q - 2q\eta) > 0$ for $\xi = \xi(C'_*) > 0$ being the number in (2.2) corresponding to our present compact subcone $C'_* \subset C_*$.

The open cone $(C^*)'$ for which (5.27) holds is fixed depending on the compact subcone $C' \subset O(C)$. Then the compact subcone $C'_* \subset C_*$ for which (5.28) holds was defined by $C'_* = \mathbb{R}^n \backslash (C^*)'$. Thus $(C^*)' \cup C'_* = \mathbb{R}^n$ and $(C^*)' \cap C'_* = \varnothing$. Using these facts together with (5.27) and (5.28) we conclude that $(e^{-2\pi\langle y,t\rangle} g(t)) \in L^r$, $1 \leqq r < \infty$, for y arbitrary but fixed in $(C' \backslash (C' \cap N(0, m)))$. The proof is complete.

LEMMA 5.6. *Let C, $C' \subset O(C)$, and $m = m(C') > 0$ be as in Lemma 5.5. Let $g(t)$ be a continuous function of $t \in \mathbb{R}^n$ which satisfies* (5.20) *for $p \geqq 1$ and $A \geqq 0$ and which satisfies* supp $(g) \subseteq C^*$. *Then* (5.23) *holds for $y \in (C' \backslash (C' \cap N(0, m)))$.*

Proof. The open cone $(C^*)'$ for which (5.25) holds contains C^*. Thus by the proof of Lemma 5.5 leading to the inequality (5.27) and the fact that supp $(g) \subseteq C^*$, the desired conclusion (5.23) is obtained.

In several of our proofs below it is important to know that the estimate (5.20) in fact implies the estimate (5.21). We prove this now.

LEMMA 5.7. *Let C, $C' \subset O(C)$, and $m = m(C') > 0$ be as in Lemma 5.5. Let $g(t)$ be a continuous function of $t \in \mathbb{R}^n$ which satisfies* (5.20) *for $A > 0$ and $p > 1$; and suppose* (5.20) *holds independently of $\Omega \in (C' \backslash (C' \cap N(0, m)))$ for arbitrary $C' \subset O(C)$. Then for any compact subcone $C'_* \subset C_*$, $g(t)$ satisfies* (5.21) *where $\eta \in (0, 1)$ is arbitrary, $1/p + 1/q = 1$, and B is given by* (5.22).

Proof. Throughout this proof C'_* is an arbitrary but fixed compact subcone of C_*. By [17, Lemma, p. 241], for arbitrary $\eta \in (0, 1)$ there exists a compact subcone $C' = C'(C'_*, \eta)$ of $C \subseteq O(C)$, depending on C'_* and on η, such that for any $t \in C'_*$ there exists a point $y_t^0 \in \text{pr}\,(C')$ where

$$-\langle t, y_t^0 \rangle \geqq (1-\eta) u_C(t). \tag{5.29}$$

Throughout the rest of this proof C' will denote the compact subcone of $C \subseteq O(C)$ just obtained corresponding to the arbitrary $\eta \in (0, 1)$ and t represents an arbitrary point in $C'_* \subset C_*$. Let us put

$$y_t = y_t^0 \left(\frac{u_C(t)}{Ap} \right)^{q/p}. \tag{5.30}$$

Since C' is a cone, $y_t^0 \in \text{pr}\,(C')$, and $u_C(t) > 0$ for $t \in C'_* \subset C_* = \{t : u_C(t) > 0\}$, then $y_t \in C' \subset C$ for any $t \in C'_*$.

We now choose a real number $R > 0$ such that

$$R > (Ap(m^{p/q}))/\xi \tag{5.31}$$

with $m = m(C') > 0$ being the fixed number in the hypothesis corresponding to the now fixed C' depending on C'_* and on $\eta \in (0, 1)$ and with $\xi = \xi(C'_*) > 0$ being the fixed number in (2.2) depending on C'_*. Consider $t \in C'_*$ such that $|t| > R > 0$. For such t the point $y_t \in C'$ of (5.30) satisfies

$$|y_t| = \left(\frac{u_C(t)}{Ap} \right)^{q/p} \geqq \left(\frac{\xi |t|}{Ap} \right)^{q/p} > \left(\frac{\xi R}{Ap} \right)^{q/p} > m \tag{5.32}$$

where we have used (2.2), (5.31), and the fact that $y_t^0 \in \text{pr}\,(C')$. Thus for $t \in C'_*$ such that $|t| > R > 0$, $y_t \in (C' \backslash (C' \cap N(0, m)))$. Applying the hypothesis (5.20) where we let $\Omega = y_t$, we have

$$|g(t)| \leqq K(C', m) \exp\left[2\pi(\langle y_t, t \rangle + A|y_t|^p)\right], \qquad t \in C'_*, \quad |t| > R. \tag{5.33}$$

By (5.29) and (5.30) we have for all $t \in C'_*$ that

$$\langle y_t, t\rangle = \left(\frac{u_C(t)}{Ap}\right)^{q/p} \langle y_t^0, t\rangle \leqq -(1-\eta)u_C(t)\left(\frac{u_C(t)}{Ap}\right)^{q/p}. \tag{5.34}$$

Also for all $t \in C'_*$ we have by (5.30) that

$$|y_t|^p = \left(\frac{u_C(t)}{Ap}\right)^q \tag{5.35}$$

since $y_t^0 \in \mathrm{pr}\,(C')$. Using (5.34) and (5.35) in (5.33) we get

$$|g(t)| \leqq K(C', m) \exp\left[2\pi A\left(\frac{u_C(t)}{Ap}\right)^q - 2\pi(1-\eta)u_C(t)\left(\frac{u_C(t)}{Ap}\right)^{q/p}\right], \quad t \in C'_*, \quad |t| > R. \tag{5.36}$$

Using the fact that $1/p + 1/q = 1$, $p > 1$, the number B in (5.22), and some simple, but tedious, arithmetic we obtain

$$2\pi A\left(\frac{u_C(t)}{Ap}\right)^q - 2\pi(1-\eta)u_C(t)\left(\frac{u_C(t)}{Ap}\right)^{q/p} = -2\pi B(1-q\eta)(u_C(t))^q. \tag{5.37}$$

Putting (5.37) in (5.36) we have

$$|g(t)| \leqq K(C', m) \exp\left[-2\pi B(1-q\eta)(u_C(t))^q\right], \qquad t \in C'_*, \quad |t| > R. \tag{5.38}$$

We now obtain a growth like (5.38) for $t \in C'_*$ such that $|t| \leqq R$ for the fixed $R > 0$ of (5.31). For the point $y_t^0 \in \mathrm{pr}\,(C')$ of (5.29) corresponding to $t \in C'_* \subset C_*$ put

$$y'_t = Qy_t^0 \tag{5.39}$$

for fixed $Q > m > 0$. Then $y'_t \in (C' \backslash (C' \cap N(0, m)))$ for each $t \in C'_*$. Again recalling that (5.20) holds for all $t \in \mathbb{R}^N$ independently of $\Omega \in (C' \backslash (C' \cap N(0, m)))$, we now choose $\Omega = y'_t$ in (5.20) and obtain from (5.20), (5.29), and the fact that $y_t^0 \in \mathrm{pr}\,(C')$ that

$$|g(t)| \leqq K(C', m) \exp\left[2\pi AQ^p\right] \exp\left[-2\pi Q(1-\eta)u_C(t)\right], \qquad t \in C'_*. \tag{5.40}$$

Now $\eta \in (0, 1)$ in (5.29) and recall that $u_C(t) > 0$ for $t \in C'_* \subset C_*$. Thus

$$\exp\left[-2\pi Q(1-\eta)u_C(t)\right] \leqq 1, \qquad t \in C'_*. \tag{5.41}$$

Using (5.40), (5.41), and (2.2), we have the following inequalities for all $t \in C'_*$ such that $|t| \leqq R$ and for B being the number in (5.22):

$$\begin{aligned} |g(t)| &\leqq K(C', m) \exp\left[2\pi AQ^p\right] \exp\left[2\pi B(1-q\eta)(u_C(t))^q\right] \\ &\quad \cdot \exp\left[-2\pi B(1-q\eta)(u_C(t))^q\right] \\ &\leqq K(C', m) \exp\left[2\pi AQ^p\right] \exp\left[2\pi BR^q\right] \exp\left[-2\pi B(1-q\eta)(u_C(t))^q\right]. \end{aligned} \tag{5.42}$$

Now (5.38) holds for all $t \in C'_*$ such that $|t| > R$ and (5.42) holds for all $t \in C'_*$ such that $|t| \leqq R$. Recall that the compact subcone $C' \subset C \subseteq O(C)$ constructed by using [17, Lemma, p. 241] in the sentence culminating in (5.29) depends on both C'_* and on $\eta \in (0, 1)$ and it is this C' that we have used to obtain both (5.38) and (5.42). Since $m = m(C')$ depends on C', it too depends on C'_* and on $\eta \in (0, 1)$. Thus combining (5.38) and (5.42) we can choose a constant $M(C'_*, \eta)$ depending on $C'_* \subset C_*$ and on $\eta \in (0, 1)$ such that (5.21) holds for all $t \in C'_* \subset C_*$. This completes the proof of Lemma 5.7.

The proofs and conclusions of Lemmas 5.5–5.7 hold equally well if the $m>0$ is arbitrary and independent of the $C'\subset O(C)$. We need this fact in § 9.

The next two lemmas concern elements of $\mathcal{K}'_p$, $p\geqq 1$.

LEMMA 5.8. *Let* $V\in\mathcal{K}'_p$, $p\geqq 1$. *Then* $\{\exp(-2\pi\langle y,t\rangle)V_t\colon y\in\mathbb{R}^n, |y|\leqq Q\}$ *is a strongly bounded set in* $\mathcal{K}'_p$ *where* $Q>0$ *is arbitrary but fixed.*

Proof. We first note that $V\in\mathcal{K}'_p$ implies $(e^{-2\pi\langle y,t\rangle}V_t)\in\mathcal{K}'_p$ for any $y\in\mathbb{R}^n$ because $e^{-2\pi\langle y,t\rangle}$ is a multiplier in $\mathcal{K}_p$, $p\geqq 1$, as a function of $t\in\mathbb{R}^n$. Applying Theorem 4.1 we have $V_t=D_t^\alpha(\exp(k|t|^p)f(t))$ for some n-tuple α of nonnegative integers, some integer $k\geqq 0$, and some bounded continuous function $f(t)$ on $\mathbb{R}^n$. Let Φ be an arbitrary bounded set in $\mathcal{K}_p$ and let $\phi\in\Phi$. We have by the generalized Leibnitz rule that

$$\begin{aligned}\langle e^{-2\pi\langle y,t\rangle}V_t,\phi(t)\rangle&=\langle D_t^\alpha(e^{k|t|^p}f(t)),e^{-2\pi\langle y,t\rangle}\phi(t)\rangle\\ &=(-1)^{|\alpha|}\int_{\mathbb{R}^n}e^{k|t|^p}f(t)\sum_{\beta+\gamma=\alpha}\frac{\alpha!}{\beta!\gamma!}\left(\frac{1}{i}\right)^{|\beta|}y^\beta e^{-2\pi\langle y,t\rangle}D_t^\gamma(\phi(t))\,dt\\ &=(-1)^{|\alpha|}\sum_{\beta+\gamma=\alpha}\frac{\alpha!}{\beta!\gamma!}\left(\frac{1}{i}\right)^{|\beta|}y^\beta I_y(\gamma)\end{aligned}\tag{5.43}$$

where

$$I_y(\gamma)=\int_{\mathbb{R}^n}e^{k|t|^p}f(t)\,e^{-2\pi\langle y,t\rangle}D_t^\gamma(\phi(t))\,dt.\tag{5.44}$$

We first prove our result for $p=1$. For $y\in\mathbb{R}^n$ such that $|y|\leqq Q$, $Q>0$ being arbitrary but fixed, we have

$$\begin{aligned}|I_y(\gamma)|&\leqq M\int_{\mathbb{R}^n}e^{k|t|}e^{2\pi|y||t|}|D_t^\gamma(\phi(t))|\,dt\\ &\leqq M\int_{\mathbb{R}^n}e^{-k|t|}e^{(2k+2\pi Q)|t|}|D_t^\gamma(\phi(t))|\,dt\end{aligned}\tag{5.45}$$

where $M>0$ is a bound on the bounded continuous function $f(t)$. Choose $r\geqq\max(|\alpha|,2k+2\pi Q)$ and recall that γ satisfies $|\gamma|\leqq|\alpha|$ since $\gamma+\beta=\alpha$ in (5.43). Continuing (5.45) we obtain

$$|I_y(\gamma)|\leqq M\int_{\mathbb{R}^n}e^{-k|t|}e^{r|t|}|D_t^\gamma(\phi(t))|\,dt\leqq M\,{}_1\|\phi\|_r\int_{\mathbb{R}^n}e^{-k|t|}\,dt\tag{5.46}$$

since $\phi\in\Phi\subseteq\mathcal{K}_1$. But Φ being a bounded set in $\mathcal{K}_1$ implies the existence of a constant W_r such that ${}_1\|\phi\|_r\leqq W_r$ for all $\phi\in\Phi$. Thus from (5.46) we have for each γ, $\beta+\gamma=\alpha$, that

$$|I_y(\gamma)|\leqq MW_r\int_{\mathbb{R}^n}e^{-k|t|}\,dt=M'_r\tag{5.47}$$

for all $\phi\in\Phi$, and the constant $M'_r>0$ in (5.47) is independent of $\phi\in\Phi$. Thus for $p=1$ and $y\in\mathbb{R}^n$ such that $|y|\leqq Q$, we have from (5.43) and (5.47) that

$$|\langle e^{-2\pi\langle y,t\rangle}V_t,\phi(t)\rangle|\leqq M'_r\sum_{\beta+\gamma=\alpha}\frac{\alpha!}{\beta!\gamma!}Q^{|\beta|},\qquad \phi\in\Phi,$$

with the bound being independent of $\phi\in\Phi$. Hence $\{\langle e^{-2\pi\langle y,t\rangle}V_t,\phi(t)\rangle\colon\phi\in\Phi, y\in\mathbb{R}^n, |y|\leqq Q\}$ is a bounded set in the complex plane. Since Φ is an arbitrary bounded set in $\mathcal{K}_1$, this proves that $\{e^{-2\pi\langle y,t\rangle}V_t\colon y\in\mathbb{R}^n, |y|\leqq Q\}$ is a strongly bounded set in $\mathcal{K}'_1$ as desired.

We now prove the lemma for the case that $p>1$. Let $1/p+1/q=1$ and let $m=(q^{-1}(kp)^{-(q/p)})$ for the fixed integer $k\geqq 0$ in the representation of $V\in\mathcal{K}'_p$. Let $y\in\mathbb{R}^n$ such that $|y|\leqq Q$ and let $\phi\in\Phi$. From (5.44) with $p>1$ we have

$$\begin{aligned}|I_y(\gamma)|&\leqq M\int_{\mathbb{R}^n}(e^{-k|t|^p}e^{-2\pi\langle y,t\rangle})(e^{(2k+1)|t|^p}|D_t^\gamma(\phi(t))|)\,e^{-|t|^p}\,dt\\ &\leqq M e^{m|2\pi y|^q}\int_{\mathbb{R}^n}(e^{(2k+1)|t|^p}|D_t^\gamma(\phi(t))|)\,e^{-|t|^p}\,dt\end{aligned}\tag{5.48}$$

since

$$e^{-k|t|^p}e^{-2\pi\langle y,t\rangle}\leqq e^{m|2\pi y|^q},\qquad \frac{1}{p}+\frac{1}{q}=1,\qquad m=(q^{-1}(kp)^{-(q/p)}),$$

by a technique in Sampson and Zieleźny [10]. Choose $r\geqq\max(|\alpha|,2k+1)$ and recall that $|\gamma|\leqq|\alpha|$ in (5.43) since $\beta+\gamma=\alpha$ there. Continuing (5.48) we get

$$\begin{aligned}|I_y(\gamma)|&\leqq M e^{m(2\pi Q)^q}\int_{\mathbb{R}^n}(e^{r|t|^p}|D_t^\gamma(\phi(t))|)\,e^{-|t|^p}\,dt\\ &\leqq M e^{m(2\pi Q)^q}{}_p\|\phi\|_r\int_{\mathbb{R}^n}e^{-|t|^p}\,dt.\end{aligned}\tag{5.49}$$

Applying the fact that $\phi\in\Phi$, a bounded set in $\mathcal{K}_p$, we have the existence of a constant W_r such that ${}_p\|\phi\|_r\leqq W_r$ for all $\phi\in\Phi$. Using this fact, (5.49), and (5.43) we continue now exactly as in the case $p=1$ and obtain that $\{e^{-2\pi\langle y,t\rangle}V_t\colon y\in\mathbb{R}^n,|y|\leqq Q\}$ is a strongly bounded set in $\mathcal{K}'_p$ for $p>1$ where $Q>0$ is arbitrary but fixed. The proof is complete.

LEMMA 5.9. *Let $V\in\mathcal{K}'_p$, $p\geqq 1$. Then*

$$\lim_{\substack{y\to 0\\ y\in\mathbb{R}^n}} e^{-2\pi\langle y,t\rangle}V_t=V_t\tag{5.50}$$

in the strong (and weak) topology of $\mathcal{K}'_p$.

Proof. Let ϕ be an arbitrary element of $\mathcal{K}_p$. Noting the first two sentences of the proof of Lemma 5.8 and using the generalized Leibnitz rule, we have by a calculation as in (5.43) that

$$\langle V,(e^{-2\pi\langle y,t\rangle}-1)\phi(t)\rangle=(-1)^{|\alpha|}\sum_{\beta+\gamma=\alpha}\frac{\alpha!}{\beta!\,\gamma!}I_y(\beta,\gamma)\tag{5.51}$$

where

$$I_y(\beta,\gamma)=\int_{\mathbb{R}^n}e^{k|t|^p}f(t)\Big(\Big(\frac{1}{i}\Big)^{|\beta|}y^\beta e^{-2\pi\langle y,t\rangle}-D^\beta(1)\Big)D_t^\gamma(\phi(t))\,dt.\tag{5.52}$$

Since

$$\Big(\frac{1}{i}\Big)^{|\beta|}y^\beta e^{-2\pi\langle y,t\rangle}-D^\beta(1)=\begin{cases}e^{-2\pi\langle y,t\rangle}-1, & \beta=(0,0,\cdots,0),\\ \Big(\dfrac{1}{i}\Big)^{|\beta|}y^\beta e^{-2\pi\langle y,t\rangle}, & \beta\neq(0,0,\cdots,0),\end{cases}$$

then no matter what the n-tuples of nonnegative integers β and γ are, $\beta+\gamma=\alpha$, we have

$$\lim_{\substack{y\to 0\\ y\in\mathbb{R}^n}}e^{k|t|^p}f(t)\Big(\Big(\frac{1}{i}\Big)^{|\beta|}y^\beta e^{-2\pi\langle y,t\rangle}-D^\beta(1)\Big)D_t^\gamma(\phi(t))=0\tag{5.53}$$

pointwise for almost all $t \in \mathbb{R}^n$. Since we desire a convergence result as $y \to 0$, $y \in \mathbb{R}^n$, in this lemma, it suffices to consider $y \in \mathbb{R}^n$ such that $|y| \leqq Q$ for some fixed $Q > 0$. For such y we have for each β and γ, $\beta + \gamma = \alpha$, and all $t \in \mathbb{R}^n$ that

$$\tag{5.54} \begin{aligned} &\left| e^{k|t|^p} f(t) \left(\left(\frac{1}{i}\right)^{|\beta|} y^\beta e^{-2\pi\langle y,t\rangle} - D^\beta(1) \right) D_t^\gamma(\phi(t)) \right| \\ &\leqq M e^{k|t|^p} |D_t^\gamma(\phi(t))| (Q^{|\beta|} e^{2\pi Q|t|} + 1) \end{aligned}$$

where $M > 0$ is the bound on $f(t)$; and the right side of (5.54) is an L^1 function on $\mathbb{R}^n$, since $\phi \in \mathcal{K}_p$, which is independent of $y \in \mathbb{R}^n$ such that $|y| \leqq Q$. Because of this and (5.53) it follows from the Lebesgue dominated convergence theorem that the integral $I_y(\beta, \gamma)$ of (5.52) satisfies $I_y(\beta, \gamma) \to 0$ as $y \to 0$, $y \in \mathbb{R}^n$, for each β and γ, $\beta + \gamma = \alpha$. By this and (5.51) we get the convergence (5.50) in the weak topology of $\mathcal{K}_p'$ since ϕ is an arbitrary element of $\mathcal{K}_p$. But by Lemma 3.1, $\mathcal{K}_p$ is a Montel space. Hence by Edwards [4, Corollary 8.4.9, p. 510] we have the convergence (5.50) in the strong topology of $\mathcal{K}_p'$ as well. The proof is complete.

6. The analytic functions. In this section we define the analytic functions with which we are concerned in this paper. In each of the definitions which follow C is an open cone in $\mathbb{R}^n$ and C' is an arbitrary compact subcone of C. Also $p \geqq 1$ and $A \geqq 0$ are real numbers.

We first recall the functions $H_p(A; C)$ of Vladimirov [17, p. 238]. A function $f(z)$ belongs to the class $H_p(A; C)$ if it is analytic in the tubular cone $T^C = \mathbb{R}^n + iC \subset \mathbb{C}^n$ and satisfies

$$\tag{6.1} |f(z)| \leqq K(C')(1+|z|)^N (1+|y|^{-M}) e^{2\pi A|y|^p}, \qquad z = x+iy \in T^{C'} = \mathbb{R}^n + iC',$$

where C' is an arbitrary compact subcone of C, $K(C')$ is a constant depending on C', and N and M are nonnegative real numbers which do not depend on C'. Further we define

$$\tag{6.2} H_p(A+\varepsilon; C) = \bigcap_{A'>A} H_p(A'; C), \qquad H_0(C) = H_1(0; C).$$

The 2π in the exponential term in (6.1) simply reflects the way we have defined the Fourier transform in this paper and putting it there does not alter the $H_p(A; C)$ functions of Vladimirov [17, p. 238].

Let $m > 0$. Throughout the rest of this paper $T(C'; m)$ will denote the set $T(C'; m) = \mathbb{R}^n + i(C' \backslash (C' \cap N(0, m)))$ where $N(0, m)$ is a closed ball in $\mathbb{R}^n$ of radius $m > 0$ and with center at the origin.

We shall say that a function $f(z)$ belongs to the class $G_p(A; C)$ if for each compact subcone C' of C there exists a fixed $m = m(C') > 0$ depending on C' such that $f(z)$ is analytic in $T(C'; m)$ and satisfies

$$\tag{6.3} |f(z)| \leqq K(C', m)(1+|z|)^N e^{2\pi A|y|^p}, \qquad z = x+iy \in T(C'; m),$$

where $K(C', m)$ is a constant depending on C' and on m and N is a nonnegative real number which does not depend on C' or on m.

We shall say that a function $f(z)$ belongs to the class $F_p(A; C)$ if, for each compact subcone C' of C, $f(z)$ is analytic in $T^{C'} = \mathbb{R}^n + iC'$ and satisfies

$$\tag{6.4} |f(z)| \leqq K(C', m)(1+|z|)^N e^{2\pi A|y|^p}, \qquad z = x+iy \in T(C'; m),$$

where $m > 0$ is arbitrary and independent of any quantity or object, $K(C', m)$ is a constant depending on C' and on m, and N is a nonnegative real number which does not depend on C' or on m.

The three spaces of analytic functions in tubular cones which we have defined in this section are related as follows:

$$H_p(A; C) \subset F_p(A; C) \subset G_p(A; C), \qquad p \geqq 1, \quad A \geqq 0.$$

In the remainder of this paper we shall analyze these spaces of analytic functions.

In general we have been motivated in our research to consider analytic functions in tubular cones which satisfy growth conditions like (6.3) and (6.4), in which $\operatorname{Im}(z)$ is in some sense bounded away from the origin in the cone, by the researches in 1 dimension of Beltrami and Wohlers [1], Lauwerier [9], and Swartz [15]. In the present paper the spaces $F_p(A; C)$ and $G_p(A; C)$ are particularly motivated by properties of the Fourier–Laplace transform of elements in $\mathcal{K}'_p$, $p \geqq 1$, as we shall see in §§ 7–9, and by the $H_p(A; C)$ functions of Vladimirov. The representation of analytic functions in tubular cones by the Fourier–Laplace transform of generalized functions is of fundamental importance in problems of quantum field theory.

7. The spaces $G_p(A; C)$. Throughout this section C will be an open connected cone. For such cones, elements of $G_p(A; C)$ satisfy properties as given in the following theorem.

THEOREM 7.1. *For the open connected cone C let $f(z) \in G_p(A; C)$. For any compact subcone $C' \subset C$ let $m = m(C') > 0$ denote the fixed real number depending on C' from the definition of $G_p(A; C)$. There exists a unique element $V = D_t^\alpha(g(t)) \in \mathcal{K}'_1$, where α is an n-tuple of nonnegative integers and $g(t)$ is a continuous function of $t \in \mathbb{R}^n$, such that for $p \geqq 1$ and $A \geqq 0$*

$$f(z) = z^\alpha \mathcal{F}[e^{-2\pi\langle y,t\rangle} g(t); x], \qquad z = x + iy \in T(C'; m), \tag{7.1}$$

with the Fourier transform being in the L^2 sense. If $p \geqq 1$ and $A = 0$ or if $p > 1$ and $A > 0$, (7.1) becomes

$$f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle, \qquad z \in T(C'; m). \tag{7.2}$$

For $p \geqq 1$ and $A \geqq 0$ we have

$$f(z) = \mathcal{F}[e^{-2\pi\langle y,t\rangle} V_t], \qquad z = x + iy \in T(C'; m), \tag{7.3}$$

with the equality (7.3) holding in K'_r for each $r \geqq 1$ and

(7.4) $\{f(z): y = \operatorname{Im}(z) \in (C' \setminus (C' \cap N(0, m))), |y| \leqq Q_m\}$ *is a strongly bounded set in K'_r for each $r \geqq 1$ where $Q_m > m > 0$.*

Further, for $p \geqq 1$ and $A \geqq 0$, $g(t)$ satisfies

$$|g(t)| \leqq K(C', m) \exp[2\pi(A|y|^p + |y|\,|t|)], \qquad t \in \mathbb{R}^n, \tag{7.5}$$

where $C' \subset C$ is arbitrary, the inequality (7.5) is independent of $y \in (C' \setminus (C' \cap N(0, m)))$, and the constant $K(C', m)$ depends on $C' \subset C$ and on m. If $p \geqq 1$ and $A = 0$ or if $p = 1$ and $A \geqq 0$, then $\operatorname{supp}(g) = \operatorname{supp}(V) \subseteq \{t: u_C(t) \leqq A\}$; *while if $p > 1$ and $A > 0$, then for any compact subcone $C'_* \subset C_* = \mathbb{R}^n \setminus C^*$, $g(t)$ satisfies*

$$|g(t)| \leqq M(C'_*, \eta) \exp[-2\pi B(1 - q\eta)(u_C(t))^q], \qquad t \in C'_* \subset C_*, \tag{7.6}$$

where $\eta \in (0, 1)$ *is arbitrary*, $1/p + 1/q = 1$,

$$B = \frac{1}{q}\left(\frac{1}{Ap}\right)^{q/p}, \tag{7.7}$$

and $M(C'_*, \eta)$ *is a constant depending on* $C'_* \subset C_*$ *and on* η.

Proof. Throughout this proof C' represents an arbitrary compact subcone of C and $m = m(C') > 0$ is the dependent number from the definition of $G_p(A; C)$ corresponding to $f(z)$. Let $p \geqq 1$ and $A \geqq 0$ for the moment. Let $\varepsilon > 0$ be fixed. For $f(z) \in G_p(A; C)$ we choose an n-tuple $\alpha = (\alpha_1, \cdots, \alpha_n)$ of nonnegative integers such that

$$|z^{-\alpha} f(z)| \leqq K'(C', m)(1 + |z|)^{-n-\varepsilon} e^{2\pi A|y|^p}, \qquad z = x + iy \in T(C'; m), \tag{7.8}$$

for some constant $K'(C', m)$ where n is the dimension of $\mathbb{C}^n$. We emphasize that α is independent of C' and of m since the choice of α depends only on N in (6.3) which is independent of $C' \subset C$ and of m. We put

$$g(t) = \int_{\mathbb{R}^n} z^{-\alpha} f(z)\, e^{-2\pi i\langle z, t\rangle}\, dx, \qquad z = x + iy \in T(C'; m), \tag{7.9}$$

which is a continuous function of $t \in \mathbb{R}^n$. By an application of the Cauchy–Poincaré theorem [17, p. 198] as in the proof of [3, Thm. 1, p. 846] we have that $g(t)$ is independent of $y = \operatorname{Im}(z) \in (C' \backslash (C' \cap N(0, m)))$ because of (7.8). From (7.8) and (7.9) we get for all $t \in \mathbb{R}^n$ that

$$\begin{aligned} |g(t)| &\leqq K'(C', m)\, e^{2\pi A|y|^p} e^{2\pi\langle y, t\rangle} \int_{\mathbb{R}^n} (1 + |x|)^{-n-\varepsilon}\, dx \\ &\leqq K(C', m) \exp\left[2\pi(\langle y, t\rangle + A|y|^p)\right] \end{aligned} \tag{7.10}$$

where $K(C', m)$ is a constant, and (7.10) holds independently of $y \in (C' \backslash (C' \cap N(0, m)))$ since $g(t)$ is independent of such $y = \operatorname{Im}(z)$. The desired conclusion (7.5) follows immediately from (7.10) for $p \geqq 1$ and $A \geqq 0$.

Now let $A > 0$ and $p > 1$. Since $g(t)$ is a continuous function of $t \in \mathbb{R}^n$ and satisfies (7.10) for $t \in \mathbb{R}^n$ where C' is an arbitrary compact subcone of $C \subseteq O(C)$ and (7.10) holds independently of $y \in (C' \backslash (C' \cap N(0, m)))$ then by the proof of Lemma 5.7, $g(t)$ satisfies (7.6) for B being given by (7.7).

We now obtain (7.1). Because of (7.8) we have $(z^{-\alpha} f(z)) \in L^1 \cap L^2$ as a function of $x = \operatorname{Re}(z)$ for arbitrary but fixed $y \in (C' \backslash (C' \cap N(0, m)))$. Thus we can write (7.9) as

$$e^{-2\pi\langle y, t\rangle} g(t) = \mathscr{F}^{-1}[z^{-\alpha} f(z); t], \qquad z = x + iy \in T(C'; m), \tag{7.11}$$

with this inverse Fourier transform being in the L^1 or L^2 sense. Using the Plancherel theory for Fourier transforms we have $(e^{-2\pi\langle y, t\rangle} g(t)) \in L^2$ and

$$z^{-\alpha} f(z) = \mathscr{F}[e^{-2\pi\langle y, t\rangle} g(t); x], \qquad z = x + iy \in T(C'; m), \tag{7.12}$$

where this Fourier transform is in the L^2 sense. Equation (7.1) follows immediately from (7.12).

Because of the growth (7.5), the fact that $g(t)$ is a continuous function of $t \in \mathbb{R}^n$, and Theorem 4.1 for $p = 1$ we have $g(t) \in \mathscr{K}'_1$. Hence

$$V = D_t^\alpha(g(t)) \in \mathscr{K}'_1. \tag{7.13}$$

Since $\mathscr{K}'_1 \subseteq \mathscr{K}'_r$, $r \geqq 1$, we also conclude that $V \in \mathscr{K}'_r$ for each $r \geqq 1$. (It is in fact easy to prove directly that any continuous function $g(t)$ *in* $\mathbb{R}^n$ which satisfies a growth condition like (7.10) or (7.5) is a continuous linear functional on $\mathscr{K}_r$ for any $r \geqq 1$; hence $g(t) \in \mathscr{K}'_r$

from which it follows that $V = D_t^\alpha(g(t)) \in \mathcal{K}_r'$, $r \geqq 1$. This comment applies in all the proofs of our theorems in §§ 7, 8, and 9 and will not be repeated.)

Now let $p \geqq 1$ and $A = 0$ or let $p = 1$ and $A \geqq 0$. Using exactly the same proof as in [3, last paragraph on p. 846 through the top of p. 847], which works in our present setting, the inequality (7.8) and the definition (7.9) of $g(t)$ yield supp $(g) \subseteq \{t: u_C(t) \leqq A\}$ for either of the cases $p \geqq 1$ and $A = 0$ or $p = 1$ and $A \geqq 0$. In particular when $A = 0$ we have supp $(g) \subseteq C^*$. Since $\{t: u_C(t) \leqq A\}$ is a regular set [11, pp. 98–99] then supp $(V) =$ supp (g), and we have obtained another of our conclusions.

We now prove (7.2). Let $p \geqq 1$ and $A = 0$ or $p > 1$ and $A > 0$. The function $g(t)$ satisfies (7.10) for $t \in \mathbb{R}^n$ when $p \geqq 1$ and $A \geqq 0$. Further $g(t)$ satisfies (7.6) when $p > 1$ and $A > 0$. Thus when $p > 1$ and $A > 0$, $g(t)$ satisfies (5.20) and (5.21), and we conclude by the proof of Lemma 5.5 that $(e^{-2\pi\langle y,t\rangle} g(t)) \in L^r$, $1 \leqq r < \infty$, for $y \in (C' \backslash (C' \cap N(0, m)))$. Also recall that if $p \geqq 1$ and $A = 0$ we have proved supp $(g) \subseteq C^*$. So if $p \geqq 1$ and $A = 0$, $g(t)$ satisfies the hypothesis of Lemma 5.6, and in this case also we have $(e^{-2\pi\langle y,t\rangle} g(t)) \in L^r$, $1 \leqq r < \infty$, for $y \in (C' \backslash (C' \cap N(0, m)))$. So for either of the cases $p \geqq 1$ and $A = 0$ or $p > 1$ and $A > 0$ the Fourier transform in (7.12) (i.e. in (7.1)) can be considered to be the L^1 transform as well as the L^2 transform. We now form $\langle V, e^{2\pi i\langle z,t\rangle}\rangle$. As will be seen in the proof of Theorem 7.2, this is a well-defined analytic function of $z \in T(C'; m)$ because of our concluded results on $g(t)$ for either of the cases $p \geqq 1$ and $A = 0$ or $p > 1$ and $A > 0$. Using distributional differentiation we have

$$\begin{aligned}\langle V, e^{2\pi i\langle z,t\rangle}\rangle &= (-1)^{|\alpha|}\langle g(t), D_t^\alpha(e^{2\pi i\langle z,t\rangle})\rangle \\ &= z^\alpha \int_{\mathbb{R}^n} g(t)\, e^{2\pi i\langle z,t\rangle}\, dt = z^\alpha \mathcal{F}[e^{-2\pi\langle y,t\rangle} g(t); x]\end{aligned} \tag{7.14}$$

for $z \in T(C'; m)$, and the Fourier transform is in either the L^1 or L^2 sense because $(e^{-2\pi\langle y,t\rangle} g(t)) \in L^r$, $1 \leqq r < \infty$, as we have seen. Thus for $p \geqq 1$ and $A = 0$ or $p > 1$ and $A > 0$, (7.1) and (7.14) combine to prove (7.2).

It remains to prove (7.3) and (7.4) for $p \geqq 1$ and $A \geqq 0$. Recall that we have concluded $V \in \mathcal{K}_r'$ for each $r \geqq 1$. For any such r let $\psi(x) \in K_r$. Let $\phi(t) \in \mathcal{K}_r$ be that element such that $\psi(x) = \mathcal{F}[\phi(t); x]$ (Theorem 3.1). Then $\psi(x) = \mathcal{F}[\check{\phi}(t); -x]$, $\check{\phi}(t) = \phi(-t)$, and $\check{\phi}(t) = \mathcal{F}^{-1}[\psi(x); -t]$ so that

$$e^{-2\pi\langle y,t\rangle} \check{\phi}(t) = \int_{\mathbb{R}^n} \psi(x)\, e^{2\pi i\langle z,t\rangle}\, dx \tag{7.15}$$

for y arbitrary but fixed in $(C' \backslash (C' \cap N(0, m)))$. We have $(e^{-2\pi\langle y,t\rangle} \phi(t)) \in \mathcal{K}_r$ for any $\phi \in \mathcal{K}_r$; hence $V \in \mathcal{K}_r'$ implies $(e^{-2\pi\langle y,t\rangle} V_t) \in \mathcal{K}_r'$. Further, $f(x+iy) \in K_r'$ as a function of $x \in \mathbb{R}^n$ for $y \in (C' \backslash (C' \cap N(0, m)))$ because of the growth condition (6.3). So for $y = \operatorname{Im}(z) \in (C' \backslash (C' \cap N(0, m)))$ we use (7.1) and obtain

$$\langle f(z), \psi(x)\rangle = \int_{\mathbb{R}^n} z^\alpha \psi(x) \mathcal{F}[e^{-2\pi\langle y,t\rangle} g(t); x]\, dx \tag{7.16}$$

with the Fourier transform being in the L^2 sense. Recalling our analysis immediately preceding (7.12) we remember that $(e^{-2\pi\langle y,t\rangle} g(t)) \in L^2$, $y \in (C' \backslash (C' \cap N(0, m)))$, for $p \geqq 1$ and $A \geqq 0$; and since $\psi \in K_r$ then $(z^\alpha \psi(x)) \in L^1 \cap L^2$ as a function of $x = \operatorname{Re}(z)$. Thus a familiar property of the L^2 Fourier transform yields

$$\int_{\mathbb{R}^n} z^\alpha \psi(x) \mathcal{F}[e^{-2\pi\langle y,t\rangle} g(t); x]\, dx = \int_{\mathbb{R}^n} e^{-2\pi\langle y,t\rangle} g(t) \mathcal{F}[z^\alpha \psi(x); t]\, dt \tag{7.17}$$

and the Fourier transform on the right of the equality in (7.17) is an L^1 transform as well as an L^2 transform since $(z^\alpha\psi(x)) \in L^1 \cap L^2$. Combining (7.16) and (7.17) we have

$$
\begin{aligned}
\langle f(z), \psi(x)\rangle &= \int_{\mathbb{R}^n} e^{-2\pi\langle y,t\rangle} g(t) \int_{\mathbb{R}^n} z^\alpha \psi(x)\, e^{2\pi i\langle x,t\rangle}\, dx\, dt \\
&= \int_{\mathbb{R}^n} g(t) \int_{\mathbb{R}^n} z^\alpha \psi(x)\, e^{2\pi i\langle z,t\rangle}\, dx\, dt \qquad (7.18)\\
&= (-1)^{|\alpha|} \int_{\mathbb{R}^n} g(t)\Big(D_t^\alpha \int_{\mathbb{R}^n} \psi(x)\, e^{2\pi i\langle z,t\rangle}\, dx\Big)\, dt
\end{aligned}
$$

where the differentiation under the integral sign is valid. Now using (7.15), the fact that $(e^{-2\pi\langle y,t\rangle} V_t) \in \mathscr{K}'_r$, and (4.2), we continue (7.18) and obtain

$$
\begin{aligned}
\langle f(z), \psi(x)\rangle &= (-1)^{|\alpha|} \int_{\mathbb{R}^n} g(t) D_t^\alpha (e^{-2\pi\langle y,t\rangle} \check{\phi}(t))\, dt \\
&= \langle V, e^{-2\pi\langle y,t\rangle} \check{\phi}(t)\rangle \qquad (7.19)\\
&= \langle e^{-2\pi\langle y,t\rangle} V, \check{\phi}(t)\rangle \\
&= \langle \mathscr{F}[e^{-2\pi\langle y,t\rangle} V_t], \psi(x)\rangle.
\end{aligned}
$$

Here ψ is an arbitrary element of K_r and our calculations (7.15)–(7.19) hold for $y = \operatorname{Im}(z) \in (C'\backslash(C' \cap N(0, m)))$. Thus (7.19) proves (7.3) for each $r \geqq 1$.

To prove (7.4) we note that by the proof of Lemma 5.8, $\{e^{-2\pi\langle y,t\rangle} V_t\colon y \in (C'\backslash(C' \cap N(0, m))), |y| \leqq Q_m\}$, $Q_m > m > 0$, is a strongly bounded set in $\mathscr{K}'_r$ for each $r \geq 1$ since $V \in \mathscr{K}'_r$ for such r. But the Fourier transform is a strongly continuous mapping from $\mathscr{K}'_r$ onto K'_r. Using (7.3) we thus obtain

$$
\begin{aligned}
&\{f(z)\colon y = \operatorname{Im}(z) \in (C'\backslash(C' \cap N(0, m))), |y| \leqq Q_m\} \\
&\quad = \{\mathscr{F}[e^{-2\pi\langle y,t\rangle} V_t]\colon y \in (C'\backslash(C' \cap N(0, m))), |y| \leqq Q_m\}
\end{aligned}
$$

is a strongly bounded set in K'_r; this proves (7.4). The proof of Theorem 7.1 is complete.

We emphasize again that the constructed $V = D_t^\alpha(g(t)) \in \mathscr{K}'_1$ of Theorem 7.1 is independent of the compact subcones $C' \subset C$ because the choice of the n-tuple α for which (7.8) holds depends only on the number N of (6.3) which is independent of $C' \subset C$ and of $m = m(C') > 0$.

We now obtain converse results to Theorem 7.1. Notice that in its most general sense, inequality (7.5) states that the function $g(t)$ grows no faster than a constant times $\exp[k|t|]$ for $t \in \mathbb{R}^n$, where $k > 0$ is a constant. This is the reason for the assumed growth (7.21) below. Our first converse result will be for $p > 1$ and $A > 0$. Recall that $O(C)$ denotes the convex envelope of a cone C and that ρ_C characterizes the convexity of C.

THEOREM 7.2. *Let $p > 1$ and $A > 0$. Let C be an open connected cone in $\mathbb{R}^n$. Let V be a finite sum*

$$V = \sum_\alpha D_t^\alpha(g_\alpha(t)) \qquad (7.20)$$

where the $g_\alpha(t)$ are continuous functions of $t \in \mathbb{R}^n$ such that for each n-tuple α of nonnegative integers $g_\alpha(t)$ satisfies

$$|g_\alpha(t)| \leqq K_\alpha\, e^{k_\alpha|t|}, \qquad t \in \mathbb{R}^n, \qquad (7.21)$$

for some constant K_α and some $k_\alpha \geqq 0$ depending on α. Further assume that each $g_\alpha(t)$ satisfies

$$|g_\alpha(t)| \leqq M_\alpha(C'_*, \eta) \exp[-2\pi B(1 - q\eta)(u_C(t))^q], \qquad t \in C'_* \subset C_*, \qquad (7.22)$$

where C'_* is an arbitrary compact subcone of $C_* = \mathbb{R}^n \backslash C^*$, $\eta \in (0, 1)$ is arbitrary, $1/p + 1/q = 1$, $B > 0$ is given by (7.7), and $M_\alpha(C'_*, \eta)$ is a constant depending on α, C'_*, and η. Then $V \in \mathcal{K}'_1$. Further the function $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ and any derivative of it belong to $G_p(AE\rho_C^p; O(C))$ for some constant E, and (7.3) and (7.4) hold for $z \in T(C'; m)$, $C' \subset O(C)$, $m = m(C') > 0$.

Proof. Each $g_\alpha(t) \in \mathcal{K}'_1$ because of (7.21), the continuity of $g_\alpha(t)$, and Theorem 4.1. Thus $V \in \mathcal{K}'_1 \subseteq \mathcal{K}'_r$, $r \geqq 1$, as desired.

We now consider

$$f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle = \sum_\alpha z^\alpha \int_{\mathbb{R}^n} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt. \tag{7.23}$$

(Here we have formally differentiated distributionally. By our succeeding analysis this calculation is valid for certain z and we specify these z below.) We shall prove the existence and analyticity of $f(z)$ for certain z. To do so it suffices to consider the function

$$h_\alpha(z) = \int_{\mathbb{R}^n} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt \tag{7.24}$$

for each α in the sum in (7.23). Let C' be an arbitrary compact subcone of $O(C)$. By [17, Lemma 2, p. 223] there exists a real number $\delta = \delta(C') > 0$ and an open cone $(C^*)'$, both depending on C', such that $C^* \subset (C^*)'$ and

$$\langle y, t\rangle \geqq \delta |y|\,|t|, \qquad y \in C' \subset O(C), \quad t \in (C^*)'. \tag{7.25}$$

For this cone $(C^*)'$, we put $C'_* = \mathbb{R}^n \backslash (C^*)'$. Then C'_* is a cone, $C'_* \cup (C^*)' = \mathbb{R}^n$, and $C'_* \cap (C^*)' = \varnothing$. Further, we have that C'_* is a compact subcone of $C_* = \mathbb{R}^n \backslash C^*$ since $C^* \subset (C^*)'$. We also now choose the real number $m_\alpha = m_\alpha(C') > 0$ depending on α and on C' to be

$$m_\alpha = (k_\alpha/(2\pi\delta)) + 1 \tag{7.26}$$

where $k_\alpha \geqq 0$ is from (7.21). If $y \in C'$ and $|y| > m_\alpha$ then by (7.26) we have

$$k_\alpha - 2\pi\delta|y| < -2\pi\delta < 0. \tag{7.27}$$

We are now ready to consider $h_\alpha(z)$ for each α. Because of the properties of the cones $(C^*)'$ and C'_* obtained above we can rewrite $h_\alpha(z)$ in (7.24) as

$$h_\alpha(z) = \int_{(C^*)'} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt + \int_{C'_*} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt = I_1^\alpha(z) + I_2^\alpha(z). \tag{7.28}$$

For the chosen $m_\alpha > 0$ in (7.26) let z_0 be an arbitrary but fixed point in $T(C'; m_\alpha) = \mathbb{R}^n + i(C' \backslash (C' \cap N(0, m_\alpha)))$. (For purposes of the proof of Theorem 7.2, we assume without loss of generality that the arbitrary compact subcone $C' \subset O(C)$ is in fact open; any compact subcone of $O(C)$ is contained in some open compact subcone of $O(C)$.) Choose an open neighborhood $N'(z_0, r)$ of z_0 with radius $r > 0$ whose closure is contained in $T(C'; m_\alpha)$. Let $z \in N'(z_0, r)$ and let γ be an arbitrary n-tuple of nonnegative integers. Corresponding to $I_1^\alpha(z)$ in (7.28) and the m_α in (7.26), we apply

$$\left| \int_{(C^*)'} t^\gamma g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt \right| \leqq K_\alpha \int_{(C^*)'} |t^\gamma|\, e^{k_\alpha|t|}\, e^{-2\pi\langle y,t\rangle}\, dt$$

$$\leqq K_\alpha \int_{(C^*)'} |t^\gamma| \exp\left[(k_\alpha - 2\pi\delta|y|)|t|\right] dt \tag{7.29}$$

$$\leqq K_\alpha \int_{(C^*)'} |t^\gamma|\, e^{-2\pi\delta|t|}\, dt$$

$$\leqq K_\alpha Z_n \int_0^\infty s^{|\gamma|+n-1}\, e^{-2\pi\delta s}\, ds = K_\alpha Z_n (|\gamma|+n-1)!(2\pi\delta)^{-|\gamma|-n}.$$

Here we have used [12, Thm. 32, p. 39] and integration by parts $(|\gamma|+n-1)$ times in the last two steps in (7.29), where Z_n is the area of the unit sphere in $\mathbb{R}^n$. Inequality (7.29) proves that the integral on the left there converges absolutely and uniformly for $z \in N'(z_0, r)$ and for any γ. Thus the differentiation D_z^γ can be taken under the integral sign in $I_1^\alpha(z)$ for $z \in N'(z_0, r)$ with the resulting integral converging absolutely and uniformly. We conclude that the integral $I_1^\alpha(z)$ in (7.28) and any derivative of it exists for $z \in N'(z_0, r)$ and for any n-tuple γ of nonnegative integers which proves that $I_1^\alpha(z)$ is analytic at z_0.

Recall that $C'_* = \mathbb{R}^n \backslash (C^*)'$ constructed above is a compact subcone of C_*. Using the assumption (7.22) and Lemma 5.3 we conclude that the integral defining $I_2^\alpha(z)$ in (7.28) and any derivative $D_z^\gamma(I_2^\alpha(z))$ of it converges absolutely and uniformly for $z \in N'(z_0, r)$. Thus $D_z^\gamma(I_2^\alpha(z))$ exists for any n-tuple γ of nonnegative integers with $z \in N'(z_0, r)$, and $I_2^\alpha(z)$ is analytic at z_0. We can now conclude that $h_\alpha(z)$ given in (7.24) is analytic at z_0 since both $I_1^\alpha(z)$ and $I_2^\alpha(z)$ of (7.28) are. Since z_0 is an arbitrary point in $T(C'; m_\alpha)$ then $h_\alpha(z)$ is analytic in $T(C'; m_\alpha)$. For each of the finite number of n-tuples α in the sum (7.20) there is an m_α given by (7.26). We now put

$$m = \max_\alpha \{m_\alpha\}. \tag{7.30}$$

It follows that for each α the corresponding integral $h_\alpha(z)$ in (7.24) is analytic in $T(C'; m)$ for the m chosen in (7.30). Hence the sum on the right of (7.23) is analytic in $T(C'; m)$. For $z \in T(C'; m)$ the use of the distributional derivative in (7.23) is justified, and we conclude that $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ is analytic in $T(C'; m)$, $C' \subset O(C)$, for the fixed $m > 0$ chosen in (7.30). It follows that any derivative of $f(z)$ is also analytic in $T(C'; m)$.

For any compact subcone $C' \subset O(C)$ and the corresponding $m > 0$ chosen in (7.30), we now show that $f(z)$ satisfies the desired growth. By (7.22) and Lemma 5.4 with $\gamma = (0, \cdots, 0)$ in (5.19), we have for each α in (7.20) and the corresponding integral $I_2^\alpha(z)$ in (7.28) that

$$|I_2^\alpha(z)| \leqq M'_\alpha(C'_*, \eta) \exp\left[2\pi A E \rho_C^p |y|^p\right], \qquad z = x + iy \in T^{C'}, \quad C' \subset O(C), \tag{7.31}$$

where $M'_\alpha(C'_*, \eta)$ is a constant and

$$E = (\xi^q - 2q\eta)^{-(p/q)}. \tag{7.32}$$

Here $\xi = \xi(C'_*) > 0$ is fixed corresponding to the fixed C'_*, $1/p + 1/q = 1$, and $\eta \in (0, 1)$ is fixed from Lemma 5.4 such that $(\xi^q - 2q\eta) > 0$.

For any of the α in (7.20) we now consider the integral $I_1^\alpha(z)$ of (7.28) for $z \in T(C'; m)$ where $m > 0$ is given in (7.30). Since $m \geqq m_\alpha$ for each m_α chosen in (7.26) then (7.27) holds if $|y| > m$, $y \in C'$. Thus by the analysis of (7.29) with $\gamma = (0, \cdots, 0)$

there we have

$$|I_1^\alpha(z)| \leqq K_\alpha Z_n (n-1)!(2\pi\delta)^{-n} = Q_\alpha(C') < \infty, \qquad z \in T(C'; m), \tag{7.33}$$

where the constant $Q_\alpha(C')$ depends on α and on the compact subcone $C' \subset O(C)$ since $\delta = \delta(C')$ depends on C'. Combining (7.28), (7.31), and (7.33) we get for each α in (7.20) and for $z = x + iy \in T(C'; m)$ that

$$\begin{aligned} |h_\alpha(z)| \leqq |I_1^\alpha(z)| + |I_2^\alpha(z)| &\leqq Q_\alpha(C') + M'_\alpha(C'_*, \eta) \exp[2\pi A E \rho_C^p |y|^p] \\ &\leqq (Q_\alpha(C') + M'_\alpha(C'_*, \eta)) \exp[2\pi A E \rho_C^p |y|^p] \end{aligned} \tag{7.34}$$

for E being given in (7.32). Recall that the choice of $(C^*)'$ in (7.25) depended on C' so that $C'_* = \mathbb{R}^n \backslash (C^*)'$ also depends on C'. Further $\eta \in (0, 1)$ has been fixed in (7.31) such that $(\xi^q - 2q\eta) > 0$. Combining (7.23), (7.24), and (7.34) we see that $f(z)$ satisfies the growth (6.3) for $z \in T(C'; m)$, where $m > 0$ was chosen in (7.30), with A in the exponential term in (6.3) being replaced by $(AE\rho_C^p)$. We conclude that $f(z) \in G_p(AE\rho_C^p; O(C))$ as desired.

Again let C' be an arbitrary compact subcone of $O(C)$ and let $m = m(C') > 0$ be chosen as in (7.30). Let γ be any n-tuple of nonnegative integers. Using (7.23), (7.24), (7.28), and the generalized Leibnitz rule, we have

$$D_z^\gamma(f(z)) = \sum_\alpha \sum_{\beta+\mu=\gamma} \frac{\gamma!}{\beta!\mu!} D_z^\beta(z^\alpha)[D_z^\mu(I_1^\alpha(z)) + D_z^\mu(I_2^\alpha(z))], \qquad z \in T(C'; m), \tag{7.35}$$

where β and μ are n-tuples of nonnegative integers. By our analysis above and our definition of the differential operator D_z^μ in § 2 we have

$$D_z^\mu(I_1^\alpha(z)) = (-1)^{|\mu|} \int_{(C^*)'} t^\mu g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt, \qquad z \in T(C'; m), \tag{7.36}$$

and by the proof of Lemma 5.3

$$D_z^\mu(I_2^\alpha(z)) = (-1)^{|\mu|} \int_{C'_*} t^\mu g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt, \qquad z \in T(C'; m), \tag{7.37}$$

for each α in (7.20). By (7.27) and the analysis of (7.29), $D_z^\mu(I_1^\alpha(z))$, $z \in T(C'; m)$, given by (7.36) is bounded by a constant depending only on α, μ, and $C' \subset O(C)$ and not depending on $z \in T(C'; m)$. Further, by the proof of Lemma 5.4, $D_z^\mu(I_2^\alpha(z))$, $z \in T(C'; m)$, given by (7.37) satisfies the growth (5.19) for any μ. It is important to note that the constant term on the right of (5.19) depends only on C'_* and on the now fixed $\eta \in (0, 1)$ such that $(\xi^q - 2q\eta) > 0$ and not on z; and in our present situation $C'_* = \mathbb{R}^n \backslash (C^*)'$ in $I_2^\alpha(z)$ depends on $C' \subset O(C)$ since $(C^*)'$ does. Using these facts along with (7.35), (7.36), and (7.37), we see that $D_z^\gamma(f(z))$, $z \in T(C'; m)$, satisfies the growth (6.3) with the exponential term being $\exp[2\pi AE\rho_C^p|y|^p]$ for any n-tuple γ of nonnegative integers where E is given by (7.32). Thus for any γ, $D_z^\gamma(f(z)) \in G_p(AE\rho_C^p; O(C))$ as we desired to show since we had previously concluded that any derivative of $f(z)$ is analytic in $T(C'; m)$, C' being an arbitrary compact subcone of $O(C)$, $m = m(C') > 0$ being chosen as in (7.30).

It remains to prove the desired conclusions (7.3) and (7.4) in the present theorem. Let $\psi \in K_r$, $r \geqq 1$, and recall that $V \in \mathcal{K}'_r$ for each $r \geqq 1$. For an arbitrary compact subcone $C' \subset O(C)$ and the corresponding $m = m(C') > 0$ we use (7.23) and a change of order of integration, which is valid here, to obtain

$$\begin{aligned}\langle f(z),\psi(x)\rangle &= \sum_\alpha \int_{\mathbb{R}^n} z^\alpha \psi(x) \int_{\mathbb{R}^n} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt\, dx \\ &= \sum_\alpha \int_{\mathbb{R}^n} g_\alpha(t) \int_{\mathbb{R}^n} z^\alpha \psi(x)\, e^{2\pi i\langle z,t\rangle}\, dx\, dt \qquad (7.38)\\ &= \sum_\alpha (-1)^{|\alpha|} \int_{\mathbb{R}^n} g_\alpha(t) \left(D_t^\alpha \int_{\mathbb{R}^n} \psi(x)\, e^{2\pi i\langle z,t\rangle}\, dx\right) dt\end{aligned}$$

for $z = x + iy \in T(C'; m)$. Now constructing the function $\check{\phi}(t) = \mathscr{F}^{-1}[\psi(x); -t] \in \mathscr{K}_r$ such that (7.15) holds as in the proof of Theorem 7.1 and arguing as in (7.19), we see that the conclusion (7.3), namely

$$f(z) = \mathscr{F}[e^{-2\pi\langle y,t\rangle} V_t], \qquad z = x + iy \in T(C'; m), \quad C' \subset O(C), \qquad (7.39)$$

with this equality holding in K_r' for any $r \geqq 1$, follows from (7.38) for the V in (7.20). Now that we have (7.39) the conclusion (7.4) in the present theorem follows from (7.39) exactly as (7.4) followed from (7.3) in the proof of Theorem 7.1. This completes the proof of Theorem 7.2.

By combining Theorems 7.1 and 7.2 we obtain the following interesting corollary.

COROLLARY 7.1. *Let C be an open connected cone and let $p > 1$ and $A > 0$. If $f(z) \in G_p(A; C)$ then $f(z)$ and any derivative of it can be extended to be an element of $G_p(AE\rho_C^p; O(C))$ for some constant $E > 0$.*

Proof. By the proof of Theorem 7.1 there exists $V = D_t^\alpha(g(t)) \in \mathscr{K}_1'$ for which (7.2) holds, where $g(t)$ is continuous over $\mathbb{R}^n$ and satisfies (7.5) and (7.6). The proof of Theorem 7.2 now yields $D_z^\gamma(\langle V, e^{2\pi i\langle z,t\rangle}\rangle) \in G_p(AE\rho_C^p; O(C))$ for any n-tuple γ of nonnegative integers. Hence for any compact subcone $C' \subset O(C)$ there exists an $m = m(C') > 0$ such that $\langle V, e^{2\pi i\langle z,t\rangle}\rangle$ is analytic in $T(C'; m)$. Thus using (7.2) to extend $f(z)$ to sets $T(C'; m)$, $C' \subset O(C)$, $m = m(C')$, we have that $D_z^\gamma(f(z)) \in G_p(AE\rho_C^p; O(C))$ for any n-tuple γ of nonnegative integers. The proof is complete.

Theorem 7.2 is a converse of Theorem 7.1 for $p > 1$ and $A > 0$. We now obtain a converse of Theorem 7.1 corresponding to the conclusions of Theorem 7.1 obtained for when $p \geqq 1$ and $A = 0$.

THEOREM 7.3. *Let C be an open connected cone. Let V be a finite sum, $V = \sum_\alpha D_t^\alpha(g_\alpha(t))$, where each $g_\alpha(t)$ is a continuous function in $\mathbb{R}^n$ which satisfies* (7.21) *and has support in $C^* = \{t: u_C(t) \leqq 0\}$. Then $V \in \mathscr{K}_1'$, $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ and any derivative of it belong to $G_p(0; O(C))$, $p \geqq 1$, and* (7.3) *and* (7.4) *hold for $z \in T(C'; m)$, $C' \subset O(C)$, $m = m(C') > 0$.*

Proof. $V \in \mathscr{K}_1' \subseteq \mathscr{K}_r'$, $r \geqq 1$, as in the proof of Theorem 7.2. Since $\operatorname{supp}(g_\alpha) \subseteq C^*$ for each α, (7.23) becomes

$$f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle = \sum_\alpha z^\alpha \int_{C^*} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt. \qquad (7.40)$$

Let C' be an arbitrary compact subcone of $O(C)$. Then the open cone $(C^*)'$ obtained from [17, Lemma 2, p. 223] for which (7.25) holds contains $C^* = \{t: u_C(t) \leqq 0\}$. Thus using (7.25), (7.26), and (7.27), we argue exactly as in (7.29) to conclude that each of the integrals on the right of (7.40) is analytic in $T(C'; m_\alpha)$ for the $m_\alpha = m_\alpha(C')$ chosen in (7.26). Then defining m as in (7.30) we have from (7.40) that the sum on the right is analytic in $T(C'; m)$. Further for these z the distributional derivative calculation in (7.40) is valid. We conclude that $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ is analytic in $T(C'; m)$. Now using analysis as in (7.29) with $\gamma = (0, \cdots, 0)$ there, we obtain a growth like (7.33) for each of the integrals on the right of (7.40) for $z \in T(C'; m)$. From this and (7.40) we have that $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ satisfies the growth of the space $G_p(0; O(C)), p \geq 1$, and $f(z) \in G_p(0; O(C))$ as desired. Further, any derivative of the present $f(z)$ also belongs to $G_p(0; O(C))$ by a similar argument as in the proof of Theorem 7.2. The results (7.3) and (7.4) holding for $z \in T(C'; m)$, $C' \subset O(C)$, $m = m(C') > 0$, follow now by exactly the same arguments as in the proof of Theorem 7.2. This completes the proof of Theorem 7.3.

Theorem 7.3 is a generalization of [2, Thm. 2]. We note here that the boundary value conclusion of [2, Thm. 2] is false; however, we can conclude results like (7.2) and (7.3) in [2, Thm. 2]. The boundary value conclusion in [2, Thm. 2] is false because the constructed m in the proof is fixed and not arbitrary; $f(z)$ in [2, (19), p. 775] is defined for $z \in T(C'; m)$ for fixed $m > 0$ and we can not let $\operatorname{Im}(z) \to 0$ as indicated there. As we shall see in succeeding sections in this paper, the m in $T(C'; m)$ must be arbitrary in order to obtain boundary value results.

The following corollary results from Theorems 7.1 and 7.3 just as Corollary 7.1 followed from Theorems 7.1 and 7.2. The proof will be left to the reader.

COROLLARY 7.2. *Let C be an open connected cone. If $f(z) \in G_p(0; C)$, $p \geq 1$, then $f(z)$ and any derivative of it can be extended to be an element of $G_p(0; O(C))$.*

8. Distributional boundary values of the spaces $F_1(A; C)$. In this section and the next we show that elements of the spaces $F_p(A; C)$ obtain distributional boundary values in K'_p, $p \geq 1$. In the present section we study the spaces $F_1(A; C)$, $A \geq 0$, which generalize the functions G^b_C defined in [2, p. 772]. (The growth (6.4) of elements in $F_1(A; C)$, $A \geq 0$, is known for $\operatorname{Im}(z)$ being arbitrarily bounded away from the origin in compact subcones and the constant $K(C', m)$ in (6.4) depends on how $\operatorname{Im}(z)$ is bounded away from the origin. This is more general than the growth [2, (12), p. 772]. Note also that the arbitrary $\sigma > 0$ in [2, (12), p. 772] is actually unnecessary for the results obtained for G^b_C in [2].) We generalize the results [2, Thms. 1 and 3, Corollary 1] for the spaces $F_1(A; C)$, which are more appropriate spaces in which to obtain results of this type, and obtain new information.

Our first theorem in this section generalizes and considerably strengthens [2, Thm. 1]; we show by the following result that the weak convergence in [2, Thm. 1] can be replaced by strong convergence in $\mathcal{K}'_r$, $r \geq 1$, and that more information concerning the analytic functions is obtainable.

THEOREM 8.1. *Let C be an open connected cone. Let C' be an arbitrary compact subcone of C. Let $f(z) \in F_1(A; C)$, $A \geq 0$. There exists a unique element $V = D^\alpha_t(g(t)) \in \mathcal{K}'_1$ with* supp $(V) \subseteq \{t: u_C(t) \leq A\}$ *such that*

$$f(z) = z^\alpha \mathcal{F}[e^{-2\pi\langle y,t\rangle} g(t); x], \qquad z = x + iy \in T^{C'}, \tag{8.1}$$

with this Fourier transform being in the L^2 sense,

$$f(z) = \mathcal{F}[e^{-2\pi\langle y,t\rangle} V_t], \qquad z = x + iy \in T^{C'}, \tag{8.2}$$

with the equality (8.2) *holding as an equality in K'_r for each $r \geq 1$,*

(8.3) *$\{f(z): y = \operatorname{Im}(z) \in C', |y| \leq Q\}$ is a strongly bounded set in K'_r for each $r \geq 1$ where $Q > 0$ is arbitrary but fixed,*

and

(8.4) *$f(z) \to \mathcal{F}[V] \in K'_r$ in the strong (and weak) topology of K'_r for each $r \geq 1$ as $y = \operatorname{Im}(z) \to 0$, $y \in C' \subset C$, with this boundary value being obtained independently of how $y \to 0$ in $C' \subset C$.*

The function $g(t)$ is a continuous function of $t \in \mathbb{R}^n$ with $\operatorname{supp}(g) \subseteq \{t: u_C(t) \leqq A\}$ *which satisfies*

$$|g(t)| \leqq M(C', m) \exp[2\pi(\langle y, t\rangle + A|y|)], \qquad t \in \mathbb{R}^n, \tag{8.5}$$

independently of $y \in (C' \backslash (C' \cap N(0, m)))$ for each $C' \subset C$ with $m > 0$ being arbitrary where $M(C', m)$ is a constant depending on C' and on m.

Proof. Let C' be an arbitrary compact subcone of C and let $m > 0$ be arbitrary. $f(z) \in F_1(A; C)$ satisfies the growth (6.4) with $p = 1$. As in the proof of Theorem 7.1 we choose an n-tuple α of nonnegative integers such that

$$|z^{-\alpha} f(z)| \leqq K'(C', m)(1 + |z|)^{-n-\varepsilon} e^{2\pi A|y|}, \qquad z = x + iy \in T(C'; m), \tag{8.6}$$

for some constant $K'(C', m)$ with n being the dimension and $\varepsilon > 0$ fixed. Here α depends only on N in (6.4) and hence is independent of $C' \subset C$ and of the arbitrary $m > 0$. We define

$$g(t) = \int_{\mathbb{R}^n} z^{-\alpha} f(z)\, e^{-2\pi i\langle z, t\rangle}\, dx, \qquad t \in \mathbb{R}^n, \quad z \in T(C'; m), \tag{8.7}$$

and $V = D_t^\alpha(g(t))$. By (8.6), $(z^{-\alpha} f(z)) \in L^1 \cap L^2$ as a function of $x = \operatorname{Re}(z)$ for $y \in (C' \backslash (C' \cap N(0, m)))$. Thus from (8.7) we have

$$e^{-2\pi\langle y, t\rangle} g(t) = \mathscr{F}^{-1}[z^{-\alpha} f(z); t], \qquad z = x + iy \in T(C'; m), \tag{8.8}$$

with this inverse Fourier transform being in either the L^1 or L^2 sense, and we have $(e^{-2\pi\langle y, t\rangle} g(t)) \in L^2$ for $y \in (C' \backslash (C' \cap N(0, m)))$ from Plancherel theory. We can now proceed by exactly the same details as in the proof of Theorem 7.1 to obtain that $g(t)$ is a continuous function of $t \in \mathbb{R}^n$, $g(t)$ is independent of $y \in (C' \backslash (C' \cap N(0, m)))$, $C' \subset C$, $m > 0$ arbitrary, $g(t)$ satisfies (8.5), $V \in \mathscr{K}_1'$, and $\operatorname{supp}(V) = \operatorname{supp}(g) \subseteq \{t: u_C(t) \leqq A\}$. Notice also that $V = D_t^\alpha(g(t))$ is independent of $C' \subset C$ and of $m > 0$ because both α and $g(t)$ are. Further, from (8.8) and Plancherel theory we have

$$z^{-\alpha} f(z) = \mathscr{F}[e^{-2\pi\langle y, t\rangle} g(t); x], \qquad z = x + iy \in T(C'; m), \tag{8.9}$$

with this inverse Fourier transform being in either the L^1 or L^2 sense, and we have $(e^{-2\pi\langle y, t\rangle} g(t)) \in L^2$ for $y \in (C' \backslash (C' \cap N(0, m)))$ from Plancherel theory. We can now proceed by exactly the same details as in the proof of Theorem 7.1 to obtain that $g(t)$ is a continuous function of $t \in \mathbb{R}^n$, $g(t)$ is independent of $y \in (C' \backslash (C' \cap N(0, m)))$, $C' \subset C$, $m > 0$ arbitrary, $g(t)$ satisfies (8.5), $V \in \mathscr{K}_1'$, and $\operatorname{supp}(V) = \operatorname{supp}(g) \subseteq \{t: u_C(t) \leqq A\}$. Notice also that $V = D_t^\alpha(g(t))$ is independent of $C' \subset C$ and of $m > 0$ because both α and $g(t)$ are. Further, from (8.8) and Plancherel theory we have

Statement (8.4) remains to be proved. As in previous theorems we have $V \in \mathscr{K}_1' \subseteq \mathscr{K}_r'$, $r \geqq 1$. Using Lemma 5.9, the fact that the Fourier transform defined by (4.2) is a strongly continuous mapping from $\mathscr{K}_r'$ onto K_r', and (8.2) we obtain (8.4). The boundary value $\mathscr{F}[V] \in K_r'$ is obtained independently of how $y \to 0$, $y \in C' \subset C$, since V is independent of such y. The proof of Theorem 8.1 is complete.

If $A = 0$ in Theorem 8.1 we can conclude a further result as seen in the following theorem which generalizes [2, Thm. 3].

THEOREM 8.2. *Let C and $C' \subset C$ be as in Theorem* 8.1. *Let $f(z) \in F_1(0; C)$. There exists a unique element $V \in \mathscr{K}_1'$ with* $\operatorname{supp}(V) \subseteq C^* = \{t: u_C(t) \leqq 0\}$ *such that* (8.1), (8.2), (8.3), *and* (8.4) *hold and such that*

$$f(z) = \langle V, e^{2\pi i\langle z, t\rangle}\rangle, \qquad z \in T^{C'}, \quad C' \subset C. \tag{8.10}$$

Proof. The unique element $V = D_t^\alpha(g(t)) \in \mathscr{K}_1'$ with supp $(V) \subseteq C^*$ such that (8.1), (8.2), (8.3), and (8.4) hold is obtained from Theorem 8.1 with $A = 0$ there, where $g(t)$ is defined in (8.7). By the proof of Theorem 8.1, supp $(V) =$ supp $(g) \subseteq C^*$ and $g(t)$ satisfies (8.5) independently of $y = \operatorname{Im}(z) \in (C' \setminus (C' \cap N(0, m)))$ for $C' \subset C$, $m > 0$ arbitrary, and $A = 0$. Thus by Lemma 5.6 and the fact that $m > 0$ is arbitrary, $(e^{-2\pi\langle y,t\rangle} g(t)) \in L^r(\mathbb{R}^n)$, $1 \leqq r < \infty$, for $y \in C' \subset C$; hence the Fourier transform in (8.1) can be interpreted in both the L^1 and L^2 sense here. With this fact, a computation as in (7.14), and the conclusion (8.1) in our present theorem, we obtain (8.10) as desired. (With the properties obtained for $g(t)$ in this theorem, we shall prove in Theorem 8.4 that the form $\langle V, e^{2\pi i\langle z,t\rangle}\rangle$ is a well-defined element of $F_1(0; O(C))$. Thus there is no problem in forming $\langle V, e^{2\pi i\langle z,t\rangle}\rangle$, $z \in T^{C'}$, here.) The proof is complete.

The following interesting theorem is of independent interest since it gives a general condition under which analytic functions in tubes obtain boundary values in K_r', $r \geqq 1$. Let $Q > 0$ be fixed. Let $m > 0$ be arbitrary such that $0 < m < Q$. Let C be an open connected cone and let C' be an arbitrary compact subcone of C. Put

$$T(Q; C'; m) = T(C'; m) \cap \{z \in T^{C'} : |\operatorname{Im}(z)| < Q\}.$$

THEOREM 8.3. *Let C, $C' \subset C$, Q, m, and $T(Q; C'; m)$ be as in the preceding paragraph. For each $C' \subset C$ let $f(z)$ be analytic in $T^{C'} \cap \{z : |\operatorname{Im}(z)| < Q\}$ and satisfy*

$$|f(z)| \leqq K(C', m)(1 + |z|)^N, \qquad z \in T(Q; C'; m), \tag{8.11}$$

where $m > 0$ is arbitrary, $K(C', m)$ is a constant depending on $C' \subset C$ and on m, and N is a nonnegative real number which is independent of $C' \subset C$ and of the $m > 0$. There exists a unique element $V \in \mathscr{K}_1'$ such that $f(z) \to \mathscr{F}[V] \in K_r'$ in the strong topology of K_r' for each $r \geqq 1$ as $y = \operatorname{Im}(z) \to 0$, $y \in C' \subset C$, with the boundary value being obtained independently of how $y \to 0$ in $C' \subset C$.

The proof of this theorem is obtained by using essentially the proof of Theorem 8.1 and will be omitted.

We now obtain a converse to Theorem 8.2 which generalizes [2, Corollary 1]. Recall in Theorem 8.2 that the constructed $V \in \mathscr{K}_1'$ has the form $V = D_t^\alpha(g(t))$ where $g(t)$ is continuous in $\mathbb{R}^n$, has supp $(g) \subseteq C^*$, and satisfies (8.5) for $A = 0$ independently of $y \in (C' \setminus (C' \cap N(0, m)))$ for each $C' \subset C$ with $m > 0$ arbitrary. This accounts for our assumption on V in the following theorem.

THEOREM 8.4. *Let C be an open connected cone. Let V be a finite sum $V = \sum_\alpha D_t^\alpha(g_\alpha(t))$, where each $g_\alpha(t)$ is a continuous function on $\mathbb{R}^n$ with support in $C^* = \{t : u_C(t) \leqq 0\}$. For each compact subcone $C' \subset O(C)$ let each $g_\alpha(t)$ satisfy*

$$|g_\alpha(t)| \leqq M_\alpha(C', m)\, e^{2\pi\langle \Omega, t\rangle}, \qquad t \in \mathbb{R}^n, \tag{8.12}$$

independently of $\Omega \in (C' \setminus (C' \cap N(0, m)))$ for $m > 0$ arbitrary where $M_\alpha(C', m)$ is a constant depending on $C' \subset O(C)$ and on $m > 0$ for each α. Then $V \in \mathscr{K}_1'$ with supp $(V) \subseteq C^*$. *Further the function $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ and any derivative of it belong to $F_1(0; O(C))$ and* (8.2), (8.3), *and* (8.4) *hold for $C' \subset O(C)$.*

Proof. The conclusions $V \in \mathscr{K}_1' \subseteq \mathscr{K}_r'$, $r \geqq 1$, and supp $(V) \subseteq C^*$ follow from the assumptions on the $g_\alpha(t)$ by similar arguments as we have used before. Now as in (7.40) we formally write

$$f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle = \sum_\alpha z^\alpha \int_{C^*} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt \tag{8.13}$$

where each integral is taken over C^* since supp $(g_\alpha) \subseteq C^*$ for each α. Let C' be an arbitrary compact subcone of $O(C)$, and as in the proof of Theorem 7.2 we can assume

C' is open. Let z_0 be an arbitrary but fixed point in $T^{C'} = \mathbb{R}^n + iC'$. Let $N'(z_0, r)$ be an open neighborhood of z_0 with radius $r > 0$ whose closure is in $T^{C'}$. Let $z \in N'(z_0, r)$. There exists a fixed $Q > 0$ such that $|y| = |\mathrm{Im}\,(z)| > Q$ for all $z \in N'(z_0, r)$. Recalling that (8.12) holds for $m > 0$ arbitrary and independently of $\Omega \in (C' \backslash (C' \cap N(0, m)))$ we now choose $m = Q/2$ and $\Omega = y/2$ for $z = x + iy \in N'(z_0, r)$. Then $|\Omega| = |y|/2 > Q/2 = m$, and $\Omega = y/2 \in C'$ since C' is a cone and $y = \mathrm{Im}\,(z) \in C'$. Thus $\Omega = y/2 \in (C' \backslash (C' \cap N(0, Q/2)))$ for any $z = x + iy \in N'(z_0, r)$ and (8.12) holds for this $m = Q/2$ and $\Omega = y/2$. Let γ be an arbitrary n-tuple of nonnegative integers. With $z \in N'(z_0, r)$ and the choice $\Omega = \mathrm{Im}\,(z)/2$ as above we apply (8.12) and (7.25) (i.e. [17, Lemma 2, p. 223]) to obtain a real number $\delta = \delta(C') > 0$ such that

$$
\begin{aligned}
\left|\int_{C^*} t^\gamma g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt\right| &\leqq M_\alpha(C', Q/2) \int_{C^*} |t^\gamma|\, e^{2\pi\langle y/2,t\rangle}\, e^{-2\pi\langle y,t\rangle}\, dt \\
&\leqq M_\alpha(C', Q/2) \int_{C^*} |t^\gamma|\, e^{-\pi\langle y,t\rangle}\, dt \\
&\leqq M_\alpha(C', Q/2) \int_{C^*} |t^\gamma|\, e^{-\pi\delta Q|t|}\, dt \qquad (8.14)\\
&\leqq M_\alpha(C', Q/2) Z_n \int_0^\infty s^{|\gamma|+n-1}\, e^{-\pi\delta Q s}\, ds \\
&= M_\alpha(C', Q/2) Z_n (|\gamma| + n - 1)!\, (\pi\delta Q)^{-|\gamma|-n}
\end{aligned}
$$

where Z_n is the area of the unit sphere in $\mathbb{R}^n$. (Here we have used [12, p. 39, Thm. 32] and integrated by parts $(|\gamma| + n - 1)$ times as in (7.29). We have also used the fact that for $z = x + iy \in N'(z_0, r)$, $|y| > Q$ for our above chosen Q. Thus in (7.25) we have $\langle y, t\rangle \geqq \delta|y|\,|t| > \delta Q|t|$ for any $z \in N'(z_0, r)$ and $t \in C^*$ since $C^* \subset (C^*)'$ in (7.25).) Inequality (8.14) proves that we can differentiate under the integral sign with respect to z for any of the integrals on the right of (8.13) and for any differential operator D_z^γ, $z \in N'(z_0, r)$; and the resulting integral converges absolutely and uniformly for $z \in N'(z_0, r)$. Thus each of the integrals on the right of (8.13) is analytic at $z_0 \in T^{C'}$. Since z_0 was arbitrary in $T^{C'}$ we conclude that the sum on the right of (8.13) is analytic in $T^{C'}$, the calculation in (8.13) is valid for $z \in T^{C'}$, and $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ is analytic for $z \in T^{C'}$ where C' is an arbitrary compact subcone of $O(C)$.

We now show that $f(z)$ satisfies the growth (6.4) for $A = 0$. Let $m > 0$ be arbitrary and let $C' \subset O(C)$ be arbitrary. Let $z = x + iy \in T(C'; m)$. Choose $\Omega = y/2$, $y = \mathrm{Im}\,(z)$, in (8.12). Applying (8.12) with $\Omega = y/2 \in (C' \backslash (C' \cap N(0, m/2)))$ and using (7.25) we have for each α and $z = x + iy \in T(C'; m)$ that

$$\left|\int_{C^*} g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt\right| \leqq M_\alpha(C', m/2) Z_n (n-1)!(\pi\delta m)^{-n} \qquad (8.15)$$

by analysis as in (8.14) where Z_n is the area of the unit sphere in $\mathbb{R}^n$. From the sentence containing (7.25) we have $\delta = \delta(C') > 0$ depends on $C' \subset O(C)$ in (8.15). It follows from (8.15) and (8.13) that $f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle$ satisfies (6.4) with $A = 0$ for $z \in T(C'; m)$, $C' \subset O(C)$, $m > 0$ arbitrary. Thus $f(z) \in F_1(0; O(C))$ as desired.

Let γ be any n-tuple of nonnegative integers. Let $C' \subset O(C)$ be arbitrary and $m > 0$ be arbitrary. From (8.13) and the generalized Leibniz rule we have

$$D_z^\gamma(f(z)) = \sum_\alpha \sum_{\beta+\mu=\gamma} \frac{\gamma!}{\beta!\,\mu!} D_z^\beta(z^\alpha)(-1)^{|\mu|} \int_{C^*} t^\mu g_\alpha(t)\, e^{2\pi i\langle z,t\rangle}\, dt, \qquad z \in T^{C'}, \qquad (8.16)$$

where the differentiation under the integral has been seen to be valid for any $z \in T^{C'}$. (Recall (8.14).) Using (8.16) and analysis as in (8.15) it follows that (6.4) holds for $A=0$ with $z \in T(C'; m)$, $m>0$ arbitrary. Thus $D_z^\gamma(f(z)) \in F_1(0; O(C))$ for any n-tuple γ of nonnegative integers as desired.

Now using (8.13) and analysis exactly like that of (7.38) and (7.39) we obtain the conclusion (8.2) in the present theorem. Now that we have this equality in K'_r for each $r \geqq 1$, the conclusions (8.3) and (8.4) in this theorem follow exactly as (8.3) and (8.4) followed from (8.2) in the proof of Theorem 8.1. The proof is complete.

Of course we could have stated Theorems 8.2 and 8.4 for $F_p(0; C)$ and $F_p(0; O(C))$, respectively, $p \geqq 1$, since for $A=0$ the exponential term in (6.4) is 1 for any $p \geqq 1$.

We formulate the following corollary of Theorems 8.2 and 8.4.

COROLLARY 8.1. *Let C be an open connected cone. If $f(z) \in F_1(0; C)$ then $f(z)$ and any derivative of it can be extended to be an element of $F_1(0; O(C))$.*

9. Distributional boundary values of the spaces $F_p(A; C)$, $p>1$, $A>0$. In this section we obtain results like those of § 8 for the spaces $F_p(A; C)$, $p>1$, $A>0$. In general the ideas, techniques, and much of the analysis needed to prove the results in this section have been developed in §§ 7 and 8. Thus we shall simply state the results of this section and invite the interested reader to supply the proofs. Our first result corresponds to Theorems 8.1 and 8.2.

THEOREM 9.1. *Let C be an open connected cone and let C' be an arbitrary compact subcone of C. Let $f(z) \in F_p(A; C)$, $p>1$, $A>0$. There exists a unique element $V = D_t^\alpha(g(t)) \in \mathscr{K}'_1$, where α is a fixed n-tuple of nonnegative integers and $g(t)$ is a continuous function of $t \in \mathbb{R}^n$, such that*

$$f(z) = z^\alpha \mathscr{F}[e^{-2\pi\langle y,t\rangle} g(t); x], \qquad z = x+iy \in T^{C'}, \tag{9.1}$$

with this Fourier transform being in the L^2 sense,

$$f(z) = \langle V, e^{2\pi i\langle z,t\rangle}\rangle, \qquad z \in T^{C'}, \tag{9.2}$$

$$f(z) = \mathscr{F}[e^{-2\pi\langle y,t\rangle} V_t], \qquad z = x+iy \in T^{C'} \tag{9.3}$$

with equality (9.3) *holding in K'_r for each $r \geqq 1$,*

(9.4) *$\{f(z) \colon y = \operatorname{Im}(z) \in C', |y| \leqq Q\}$ is a strongly bounded set in K'_r for each $r \geqq 1$ where $Q>0$,*

and

(9.5) *$f(z) \to \mathscr{F}[V] \in K'_r$ in the strong (and weak) topology of K'_r for each $r \geqq 1$ as $y = \operatorname{Im}(z) \to 0$, $y \in C' \subset C$, with this boundary value being obtained independently of how $y \to 0$ in $C' \subset C$.*

Further, $g(t)$ satisfies

$$|g(t)| \leqq M(C', m) \exp[2\pi(\langle y, t\rangle + A|y|^p)], \qquad t \in \mathbb{R}^n, \tag{9.6}$$

independently of $y \in (C' \backslash (C' \cap N(0, m)))$, $m>0$ being arbitrary, where $M(C', m)$ is a constant depending on C' and on m; and for any compact subcone $C'_* \subset C_* = \mathbb{R}^n \backslash C^*$ we have

$$|g(t)| \leqq M(C'_*, \eta) \exp[-2\pi B(1-q\eta)(u_C(t))^q], \qquad t \in C'_* \subset C_*, \tag{9.7}$$

where $\eta \in (0, 1)$ is arbitrary, $1/p + 1/q = 1$, B is given by (7.7), *and $M(C'_*, \eta)$ is a constant depending on C'_* and on η.*

It is interesting to note that when $p>1$ and $A>0$ in $F_p(A; C)$, the knowledge (9.7) on $g(t)$ and hence on V takes the place of the knowledge $\operatorname{supp}(g)=\operatorname{supp}(V)\subseteq C^*$ obtained for $F_p(0; C)$, $p\geqq 1$, and the knowledge $\operatorname{supp}(g)=\operatorname{supp}(V)\subseteq\{t: u_C(t)\leqq A\}$ obtained for $F_1(A; C)$, $A>0$, in the theorems of § 8. The same comment holds with respect to the analysis of § 7 also.

We now state a converse to Theorem 9.1.

THEOREM 9.2. *Let C be an open connected cone. Let C' be an arbitrary compact subcone of $O(C)$ and let $m>0$ be arbitrary. Let $p>1$ and $A>0$. Let V be a finite sum, $V=\sum_\alpha D_t^\alpha(g_\alpha(t))$, where the $g_\alpha(t)$ are continuous functions of $t\in\mathbb{R}^n$ such that for each n-tuple α of nonnegative integers and each $C'\subset O(C)$*

$$|g_\alpha(t)|\leqq M_\alpha(C', m)\exp[2\pi(\langle\Omega, t\rangle+A|\Omega|^p)],\qquad t\in\mathbb{R}^n, \tag{9.8}$$

with this inequality holding independently of $\Omega\in(C'\backslash(C'\cap N(0, m)))$, $m>0$ arbitrary, and $M_\alpha(C', m)$ is a constant depending on α, C', and m. Further, assume that each $g_\alpha(t)$ satisfies

$$|g_\alpha(t)|\leqq M_\alpha(C'_*, \eta)\exp[-2\pi B(1-q\eta)(u_C(t))^q],\qquad t\in C'_*\subset C_*, \tag{9.9}$$

where C'_* is an arbitrary compact subcone of $C_*=\mathbb{R}^n\backslash C^*$, $\eta\in(0,1)$ is arbitrary, $1/p+1/q=1$, B is given by (7.7), and $M_\alpha(C'_*, \eta)$ is a constant depending on α, C'_*, and η. Then $V\in\mathcal{K}'_1$. Further the function $f(z)=\langle V, e^{2\pi i\langle z,t\rangle}\rangle$ and any derivative of it belongs to $F_p(AE\rho_C^p; O(C))$ for some constant $E>0$ and (9.3), (9.4), and (9.5) hold for $C'\subset O(C)$.

Note that the analytic functions in Theorems 9.1 and 9.2 satisfy the hypothesis of Theorem 8.3 and hence are guaranteed to have strong K'_r, $r\geqq 1$, boundary values by Theorem 8.3.

COROLLARY 9.1. *Let C be an open connected cone. If $f(z)\in F_p(A; C)$, $p>1$, $A>0$, then $f(z)$ and any derivative of it can be extended to be an element of $F_p(AE\rho_C^p; O(C))$ for some constant $E>0$.*

Acknowledgment. The author expresses his sincere gratitude to the referees for their excellent comments and suggestions concerning this paper. The author also thanks the Department of Mathematics of Iowa State University for giving him the opportunity of visiting Iowa State during 1978–1979.

REFERENCES

[1] E. J. BELTRAMI AND M. R. WOHLERS, *Distributions and the Boundary Values of Analytic Functions*, Academic Press, New York, 1966.

[2] R. D. CARMICHAEL, *Distributions of exponential growth and their Fourier transforms*, Duke Math. J., 40 (1973), pp. 765–783.

[3] ———, *Distributional boundary values of functions analytic in tubular radial domains*, Indiana Univ. Math. J., 20 (1971), pp. 843–853.

[4] R. E. EDWARDS, *Functional Analysis*, Holt, Rinehart and Winston, New York, 1965.

[5] A. FRIEDMAN, *Generalized Functions and Partial Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1963.

[6] I. M. GEL'FAND AND G. E. SHILOV, *Generalized Functions*, vol. 2, Academic Press, New York, 1968.

[7] ———, *Generalized Functions*, vol. 3, Academic Press, New York, 1967.

[8] M. HASUMI, *Note on the n-dimensional tempered ultra-distributions*, Tôhoku Math. J., 13 (1961), pp. 94–104.

[9] H. A. LAUWERIER, *The Hilbert problem for generalized functions*, Arch. Rational Mech. Anal., 13 (1963), pp. 157–166.

[10] G. SAMPSON AND Z. ZIELEŹNY, *Hypoelliptic convolution equations in $\mathcal{K}'_p$, $p>1$*, preprint (1975).

[11] L. SCHWARTZ, *Théorie des Distributions*, Hermann, Paris, 1966.

[12] ———, *Mathematics for the Physical Sciences*, Addison-Wesley, Reading, MA, 1966.

[13] J. SEBASTIÃO E SILVA, *Le calcul opérationnel au point de vue des distributions*, Portugal. Math., 14 (1955), pp. 105–132.
[14] ———, *Les fonctions analytiques comme ultra-distributions dans le calcul opérationnel*, Math. Ann., 136 (1958), pp. 58–96.
[15] C. SWARTZ, *Distributional boundary values*, Indiana Univ. Math. J., 21 (1972), pp. 689–692.
[16] V. S. VLADIMIROV, *On functions holomorphic in tubular cones*, Izv. Akad. Nauk SSSR Ser. Mat., 27 (1963), pp. 75–100.
[17] ———, *Methods of the Theory of Functions of Many Complex Variables*, M.I.T. Press, Cambridge, MA, 1966.
[18] K. YOSHINAGA, *On spaces of distributions of exponential growth*, Bull. Kyushu Inst. Tech. Math. Natur. Sci., 6 (1960), pp. 1–16.
[19] Z. ZIELEŹNY, *On the space of convolution operators in* $\mathcal{K}'_1$, Studia Math., 31 (1968), pp. 111–124.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0015 $01.00/0

ON A HILL'S EQUATION WITH TWO GAPS IN ITS SPECTRUM*

WALLACE GOLDBERG† AND HARRY HOCHSTADT‡

Abstract. It is known that when the potential in a Hill's equation is such that there are two gaps in the spectrum of the equation then the potential satisfies a certain nonlinear fourth order differential equation. This equation cannot be solved by any other standard methods. However, a method is presented whereby, one can show that for suitable choices of the parameters in the equation, one can find solutions which are periodic and furthermore, are also elliptic functions. This generalizes a comparable result when there is only one gap in the spectrum.

1. Introduction.

We consider the Hill's equation

$$y''+[\lambda-q(z)]y=0, \qquad q(z+\pi)=q(z). \tag{1}$$

(Relevant background information can be found in references [2], [6], [9].) The spectrum of (1) consists of an infinity of intervals

$$(\lambda_0,\lambda_1'),(\lambda_2',\lambda_1),(\lambda_2,\lambda_3'),(\lambda_4',\lambda_3),\cdots. \tag{2}$$

These intervals are called stability intervals while the intervals

$$(-\infty,\lambda_0),(\lambda_1',\lambda_2'),(\lambda_1,\lambda_2),(\lambda_3',\lambda_4')\cdots \tag{3}$$

are called instability intervals. All but the first interval in (3) is finite and may shrink to a point under special conditions. Erdélyi [3] discovered situations where all but a finite number of the finite instability intervals vanish when $q(z)$ is a suitable elliptic function. Lax [8] showed that a function $q(z)$ which satisfies the nth order Korteweg–de Vries equation requires (1) to have no more than n nonvanishing finite instability intervals.

If all finite instability intervals vanish, then $q(z)$ in (1) is necessarily a constant. Proofs of this fact may be found in [1], [6].

When precisely one of the finite instability intervals does not vanish, then, as was shown by Hochstadt [6], $q(z)$ satisfies the nonlinear differential equation

$$q''=3q^2+Aq+B \tag{4}$$

where A and B are suitable constants. If precisely n finite instability intervals fail to vanish, then $q(z)$ satisfies a differential equation of the form

$$q^{(2n)}=H(q,q',\cdots,q^{(2n-2)}), \tag{5}$$

where H is a polynomial of maximal degree $n+2$. This result was established by Goldberg [4] who also showed [5] that (5) is equivalent to the nth order Korteweg–de Vries equation. Therefore, (5) is both necessary and sufficient for n finite instability intervals to fail to vanish.

In particular, for the case $n=2$, Goldberg [4] showed that (5) reduces to

$$q^{(4)}=10qq''+A(q''-3q^2)+5(q')^2-10q^3+Bq+C, \tag{6}$$

where A, B and C are suitable constants.

---

* Received by the editors June 13, 1978.

† Queens College of CUNY, Flushing, New York 11367.

‡ Polytechnic Institute of New York, Department of Mathematics, Brooklyn, New York 11201. The work of this author supported in part by the National Science Foundation under Grant MCS 76-06637A01.

When one instability interval fails to vanish, Hochstadt [7] showed that a $q(z)$ which satisfies (4) must be of the form $q(\pm z+\tau)$ where $q(z)$ is an even function. From a different approach, McKean [10] proves the existence of 2^n potential functions corresponding to (5).

Magnus[1] has conjectured that (6) has solutions which can be found by the following "Ansatz." Let

$$q=2\beta s^2+\gamma \tag{7}$$

and seek solutions of (6) that also satisfy

$$(s')^2=s^4+as^3+bs^2+cs+d \tag{8}$$

where β, γ, a, b, c and d are constants to be chosen appropriately. We are then led to the following theorem.

THEOREM. *For a suitable choice of constants A, B and C,* (6) *will have two solutions $q_1(z)$, $q_2(z)$ which have period π. These two potential functions, when inserted in* (1), *lead to Hill's equations with at most two nonvanishing instability intervals. Only one of these potentials, say $q_1(z)$, will also satisfy* (4).

The second potential, say $q_2(z)$, will satisfy (6) *but not* (4) *and will lead to a Hill's equation with precisely two nonvanishing instability intervals. All other such potentials found via* (7) *and* (8) *are of the form $q_2(\pm z+\tau)$, where $q_2(z)$ is an even function and τ an arbitrary translation.*

For two nonvanishing finite instability intervals, our approach produces two periodic potential functions which, as follows from (8), *are elliptic functions.*

2. Preliminary results.

LEMMA 1. *Equation* (6)

$$q^{(4)}=10qq''+A(q''-3q^2)+5(q')^2-10q^3+Bq+C,$$

where A, B and C are real constants, has real periodic nontrivial solutions with real periods if A, B, and C are chosen appropriately.

Proof. By replacing q by $q-A/10$, (6) takes on the somewhat simpler form

$$q^{(4)}=10qq''+5(q')^2-10q^3+Bq+C. \tag{9}$$

Next, we show that it is possible to find solutions of (9) such that

$$q=2\beta s^2+\gamma, \tag{7}$$

where

$$(s')^2=s^4+as^3+bs^2+cs+d \tag{8}$$

for some real constants β, γ, a, b, c and d. From (7) and (8)—(see the Appendix for details)—we obtain the following identities.

$$\begin{aligned} q^{(4)}=&\,240\beta s^6+336\beta as^5+(105\beta a^2+240\beta b)s^4\\ &+(180\beta c+130\beta ab)s^3+(84\beta ac+32\beta b^2+144\beta d)s^2\\ &+(30\beta bc+60\beta ad)s+(3\beta c^2+16\beta bd), \end{aligned} \tag{10}$$

[1] We are indebted to Wilhelm Magnus for suggesting this approach and for many helpful discussions.

$$
\begin{aligned}
10qq'' &= 240\beta^2 s^6 + 200\beta^2 a s^5 + (160\beta^2 b + 120\gamma\beta)s^4 \\
&\quad + (120\beta^2 c + 100\beta a\gamma)s^3 + (80\beta^2 d + 80\beta b\gamma)s^2 \\
&\quad + (60\beta\gamma c)s + (40\beta\gamma d), \\
5(q')^2 &= 80\beta^2 s^6 + 80\beta^2 a s^5 + 80\beta^2 b s^4 + 80\beta^2 c s^3 + 80\beta^2\, d s^2, \\
-10q^3 &= -80\beta^3 s^6 - 120\beta^2\gamma s^4 - 60\beta\gamma^2 s^2 - 10\gamma^3
\end{aligned}
$$

and

$$Bq + C = 2\beta B s^2 + \gamma B + C.$$

The right hand side of (9) becomes

$$
\begin{aligned}
&(320\beta^2 - 80\beta^3)s^6 + (280\beta^2 a)s^5 + (240\beta^2 b + 120\gamma\beta - 120\beta^2\gamma)s^4 \\
&+ (200\beta^2 c + 100\beta a\gamma)s^3 + (2\beta B + 160\beta^2 d + 80\beta b\gamma - 60\beta\gamma^2)s^2 \qquad (11) \\
&+ (60\beta\gamma c)s + (C + 40\beta\gamma d - 10\gamma^3 + \gamma B).
\end{aligned}
$$

Comparing coefficients of s^6 in (10) and (11) we get

$$240\beta = 320\beta^2 - 80\beta^3 \qquad (12.1)$$

which has roots $\beta = 0, 1, 3$. For nontrivial solutions of (6) it is necessary that $\beta \neq 0$. Therefore

$$\beta = 1 \quad \text{or} \quad \beta = 3.$$

Comparing coefficients of s^5 in (10) and (11) we get

$$336\beta a = 280\beta^2 a. \qquad (12.2)$$

Since $\beta \neq 0$, we must have

$$a = 0. \qquad (12.3)$$

A comparison of the coefficients of s^4 now gives us

$$240\beta b = 240\beta^2 b + 120\gamma\beta(1-\beta). \qquad (12.4)$$

For $\beta = 1$ (12.4) is an identity and γ remains arbitrary, while for $\beta = 3$ we get

$$\gamma = 2b. \qquad (12.5)$$

The coefficients of s^3 in (10) and (11) now imply that

$$200\beta^2 c = 180\beta c. \qquad (12.6)$$

Since $\beta \neq 0$, we must also have

$$c = 0. \qquad (12.7)$$

The coefficients of s^2 and 1 give us

$$32\beta b^2 + 144\beta d = 2\beta B + 160\beta^2 d + 80\beta b\gamma - 60\beta\gamma^2 \qquad (12.8)$$

and

$$16\beta b d = C + 40\beta\gamma d - 10\gamma^3 + \gamma B, \qquad (12.9)$$

respectively, while the coefficients of s lead to an identity. To summarize, for $\beta = 1$ the results in (12) are

$$\begin{aligned} &a = c = 0, \quad \gamma \text{ arbitrary}, \\ &B = 16b^2 - 8d - 40b\gamma + 30\gamma^2 \\ &C = 4[4bd - \gamma(5\gamma^2 - 10b\gamma + 4b^2 + 8d)] \end{aligned} \tag{13}$$

while for $\beta = 3$ (12) gives us

$$\begin{aligned} &a = c = 0, \quad \gamma = 2b, \\ &B = 56(b^2 - 3d) \\ &C = 16(9bd - 2b^3). \end{aligned} \tag{14}$$

From the above, we conclude that (9) is equivalent to

$$(s')^2 = s^4 + bs^2 + d \tag{15}$$

under the change of variable (8). Nontrivial solutions can arise only from the choices $\beta = 1$ or $\beta = 3$.

If we let $s' = V$, then $s'' = V(dV/ds)$ and a differentiation of (15) with respect to z yields

$$\frac{dV}{ds} = \frac{s(2s^2 + b)}{V}. \tag{16}$$

For $b \geqq 0$, a simple phase-plane analysis shows that solutions of (16) are never closed curves. For large s and V the solutions behave asymptotically as

$$V \approx \pm s^2.$$

For $b < 0$, we can replace V by $(-b/2)V$ and s by $(-b/2)^{1/2}s$ so that, without loss of generality, we can consider

$$\frac{dV}{ds} = \frac{2s(s^2 - 1)}{V}, \tag{17}$$

which has solutions of the form

$$V^2 = s^4 - 2s^2 + K. \tag{18}$$

Equation (17) has three critical points in the (s, V)-plane, namely $(0, 0)$ and $(\pm 1, 0)$. The origin is a center and $(\pm 1, 0)$ are saddle points. The solutions of (17) which pass through $(\pm 1, 0)$ are the separatrices

$$V = \pm(s^2 - 1).$$

Figure 1 illustrates what happens in the phase plane. All real periodic solutions of (17) must correspond to closed loops in the interior of the region bounded by the separatrices.

LEMMA 2. *Equation* (4)

$$q'' = 3q^2 + Aq + B,$$

where A and B are real constants, is equivalent to

$$(s')^2 = s^4 + as^3 + bs^2 + cs + d, \tag{8}$$

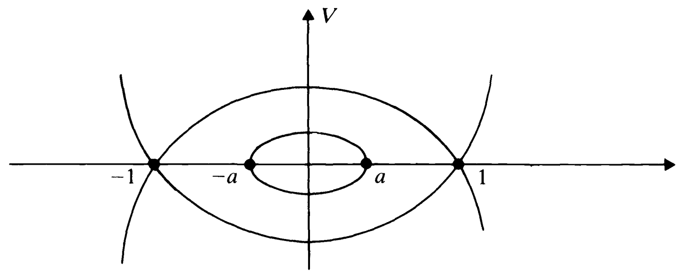

FIG. 1

under the change of variable

$$q = 2\beta s^2 + \gamma \tag{7}$$

only when $\beta = 0, 1$. *When* $\beta = 1$ (8) *takes on the form*

$$(s')^2 = s^4 + bs^2 + d. \tag{15}$$

Proof. If we replace q by $q - A/6$, (4) takes on the simpler form

$$q'' = 3q^2 + B. \tag{19}$$

Using (7) and (8) in (19)—(see the Appendix for details)—we get the following identities.

$$q'' = 12\beta s^4 + 10\beta a s^3 + 8\beta b s^2 + 6\beta c s + 4\beta d \tag{20}$$

and

$$3q^2 + B = 12\beta^2 s^4 + 12\gamma\beta s^2 + 3\gamma^2 + B. \tag{21}$$

Comparing coefficients of s^4 in (20) and (21), we get

$$12\beta = 12\beta^2 \tag{22.1}$$

which has roots $\beta = 0, 1$. When $\beta = 1$, the coefficients of s^3 and s in (20) and (21) imply that

$$a = c = 0 \tag{22.2}$$

while the coefficients of s^2 and 1 give us

$$b = \tfrac{3}{2}\gamma \tag{23.3}$$

and

$$4d = 3\gamma^2 + B. \tag{22.4}$$

For nontrivial periodic solutions of (4) to exist, we must have $\beta = 1$ and (22.3) requires that $\gamma < 0$.

Therefore, if $q(z)$ is a nontrivial real periodic solution of (4) then $q(z)$ also satisfies (6) and can only correspond to the choice $\beta = 1$.

LEMMA 3. *If* (17) *has real periodic solutions with a given real period, then there exist two such solutions that are also even functions of* z. *If the period is denoted by* $2T$, *then in the interval* $(0, T)$ *such a solution is a monotonic function of* z.

Proof. Suppose such periodic solutions correspond to the closed loop in Fig. 1. The solution of (15) corresponding to that loop is given by

$$\left(\frac{ds}{dz}\right)^2 = V^2 = s^4 - 2s^2 + 2a^2 - a^4. \tag{23}$$

This curve is symmetric with respect to both s and V axes and passes through $(a, 0)$ and $(-a, 0)$.

One solution of (23) which is even is given by

$$z=\int_{-a}^{s(z)} \frac{d\tau}{(\tau^4-2\tau^2+2a^2-a^4)^{1/2}}. \tag{24}$$

Clearly

$$s(T)=a \tag{25}$$

and (24) defines s for $0 \leqq z \leqq T$. Now s is extended to $(-T, 0)$ as an even function and beyond $(-T, T)$ as a periodic function. Clearly $s(z+T)$ will also be even and periodic. The two choices of β give rise to four nontrivial even periodic solutions of (6).

From the preceding discussion it is clear that periodic solutions of (6) can be found by solving (15). Then

$$\frac{ds}{dz}=(s^4-2s^2+2a^2-a^4)^{1/2} \tag{26}$$

where $2a^2-a^4$ is a suitable constant of integration. For $0 \leqq a < 1$, periodic solutions are found and their period is given by

$$2T=4\int_0^a \frac{d\tau}{(\tau^4-2\tau^2+2a^2-a^4)^{1/2}}. \tag{27}$$

For $a=1$ we get the separatrices and for $a=0$ we get the constant solution $s=0$.

LEMMA 4. *The period* $2T$, *as a function of* a $(0 \leqq a < 1)$, *is monotonically increasing.*

Proof. Equation (27) is rewritten as

$$2T=4\int_0^a \frac{d\tau}{(a^2-\tau^2)^{1/2}(2-a^2-\tau^2)^{1/2}}.$$

We now substitute $\tau = a \sin\theta$ into the above to obtain

$$2T=4\int_0^{\pi/2} \frac{d\theta}{(2-a^2-a^2\sin^2\theta)^{1/2}}. \tag{28}$$

The Maclaurin series of the integrand of (28) as a function of a^2 is

$$\frac{1}{\sqrt{2}}+\sum_{n=1}^{\infty} \frac{(2n-1)(2n-3)\cdots(5)(3)}{2^{n-1/2}}\left(\frac{1+\sin^2\theta}{2}\right)^n a^{2n}. \tag{29}$$

Every coefficient of a^2 is positive, so as a increases so does the sum of the series. Hence, T is a monotonically increasing function of a.

We note that, had we not simplified (16) into (17), the leading term in (29) would have been $(-2b)^{-1/2}$. By choosing b suitably, it is now easy to show that the minimum period is less than π, the period of $q(z)$ in (1).

In the Ansatz (7), (8) it was assumed that q and $(s')^2$ are polynomials in s of degrees 2 and 4, respectively. One might well ask whether polynomials of other degree could be used. The following lemma will address itself to this point.

LEMMA 5. *If we assume that* (6) *can be solved by setting*

$$q=\sum_{j=0}^{n} a_j s^j \tag{30}$$

where

$$(s')^2 = \sum_{j=0}^{k} b_j s^j, \tag{31}$$

then necessarily $k = n+2$. This procedure can be expected to work in general only if $n \leqq 3$.

Proof. A direct calculation (of the type outlined in the appendix) shows that

$$\text{degree } q^{(4)} = n + 2k - 4, \tag{32}$$

$$\text{degree of righthand side of (6)} = \max(3n, 2n + k - 2). \tag{33}$$

Since the degrees of the left and right sides of (6) must be equal we see that necessarily $k = n+2$.

In (30) we have $n+1$ and in (31) $n+3$ unknown coefficients. Thus we have a total of $2n+4$ unknowns. Equation (6) will reduce to a polynomial of degree $3n$, which has to vanish identically. This leads to $3n+1$ equations. In order to solve for $2n+4$ unknowns we now require that

$$3n + 1 \leqq 2n + 4$$

so that

$$n \leqq 3.$$

It might still be possible to use this procedure for $n > 3$, but then the $3n+1$ equations could no longer be independent.

N.B. The seven equations in the proof of Lemma 1 do in fact reduce to six.

3. Proof of the theorem. When precisely two finite instability intervals fail to vanish, (6) must be satisfied. Lemma 1 and Lemma 4 show the existence of two nontrivial periodic solutions of (6) with period π. Lemma 2 tells us that only one of these two solutions satisfies (6) and fails to satisfy (4). Lemma 3 now gives us an even function $q_2(z)$ which is periodic with period π, satisfies (6) and doesn't satisfy (4) such that

$$q = 6s^2 + \gamma$$

where $(s')^2$ is a polynomial in s.

Appendix. In this section we indicate how (10), (11), (20), (21), (32) and (33) were calculated.

If $q = 2\beta s^2 + \gamma$ where $(s')^2 = P(s)$, a polynomial in s, then

$$q' = 4\beta s\sqrt{P},$$

$$q'' = 2\beta s\frac{dP}{ds} + 4\beta P,$$

$$q''' = \left[6\beta\frac{dP}{ds} + 2\beta s\frac{d^2P}{ds^2}\right]\sqrt{P}$$

and

$$q^{(4)} = 3\beta\left(\frac{dP}{ds}\right)^2 + \beta s\frac{dP}{ds}\frac{d^2P}{ds^2} + 8\beta P\frac{d^2P}{ds^2} + 2\beta sP\frac{d^3P}{ds^3}.$$

P is differentiated a sufficient number of times with respect to s and is then substituted into the above terms.

REFERENCES

[1] G. Borg, *Eine Umkehrung der Sturm–Liouvilleschen Eigenwertaufgabe. Bestimmung der Differentialgleichung durch die Eigenwerte*, Acta Math., 78 (1946), pp. 1–96.

[2] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

[3] A. Erdélyi, *Higher Transcendental Functions*, vol. III, McGraw-Hill, New York, 1955.

[4] W. Goldberg, *On the determination of a Hill's equation from its spectrum*, J. Math. Anal. Appl. 51, 3 (1975), pp. 705–723.

[5] ———, *Necessary and sufficient conditions for determining a Hill's equation from its spectrum*, J. Math. Anal. Appl., 55, 3 (1976), pp. 549–554.

[6] H. Hochstadt, *On the determination of a Hill's equation from its spectrum*, Arch. Rational Mech. Anal., 19 (1965), pp. 353–362.

[7] ———, *An inverse problem for a Hill's equation*, J. Differential Equations, 20, 1 (1976), pp. 53–60.

[8] P. D. Lax, *Periodic solutions of the kdv equations*, Lect. in Appl. Math. AMS 15 (1974), pp. 85–96.

[9] W. Magnus and S. Winkler, *Hill's Equation*, Wiley-Interscience, New York, 1966.

[10] H. P. McKean and P. von Moerbeke, *The spectrum of Hill's equation*, Inven. Math., 30 (1975), pp. 217–274.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0016 $01.00/0

ALGEBRAIC METHOD FOR SOLVING LINEAR PARTIAL DIFFERENTIAL EQUATIONS WITH VARIABLE COEFFICIENTS. II: FUNDAMENTAL SOLUTIONS*

SERGE VASILACH†

Abstract. This paper is devoted to the construction of fundamental solutions of linear partial differential equations with variable coefficients, by means of the algebraic method which is given in previous papers by S. Vasilach (SIAM J. Math. Anal., 1975 and 1979).

Introduction. For an introduction to the algebraic method for solving linear differential and partial differential equations with variable coefficients see our previous papers [1], [2]. This method is based on the composition product of tensor products of kernel distributions (on the general definition of the composition products of this kind, see our papers [1], [2], [3], [4]) and is characterised by the fact that all operations for determining fundamental solutions are algebraic operations.

Fundamental solutions in $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$.

1. Preliminaries. In [2, § 5] we have shown that the fundamental solution of the equation

$$(1)\qquad \frac{\partial^{m+n}\{E\}}{\partial x^m \partial y^n}+\sum_{j=0}^{m-1}\sum_{k=0}^{n-1} a_{jk}(x,y)\frac{\partial^{j+k}\{E\}}{\partial x^j \partial y^k}=\delta(x-\alpha)\otimes\delta(y-\beta),$$

in which (a_{jk}), $(j,k)\in[0,m-1]\times[0,n-1]$, are operators of multiplication in $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$, is given by

$$(2)\qquad \{E\}=\{\delta(x-\alpha)\otimes\delta(y-\beta)+\sum_{\nu\in\mathbb{N}}(-1)^\nu H^{(\nu)}(x,y,\alpha,\beta)\}\circ\{Y(x-\alpha)^m\circ Y(y-\beta)^n\}$$

where $Y(x-\alpha)^m$ (resp. $Y(y-\beta)^n$) is the mth (resp. nth) composition power of Heaviside's kernel $Y(x-\alpha)$ (resp. $Y(y-\beta)$), and $H^{(\nu)}$ is the νth composition power of the function

$$(3)\qquad H(x,y,\alpha,\beta)=\sum_{j=0}^{m-1}\sum_{k=0}^{n-1}(-1)^{j+k}\frac{\partial^{j+k}}{\partial\alpha^j\partial\beta^k}\Big(\frac{(x-\alpha)^{m-1}}{(m-1)!}\otimes\frac{(y-\beta)^{n-1}}{(n-1)!}a_{jk}(\alpha,\beta)\Big).$$

Likewise, in [2, § 5, No. 3.2], we have given as example, the fundamental solution of the hyperbolic equation

$$(4)\qquad \frac{\partial^2\{E\}}{\partial x\partial y}+a(x,y)\{E\}=\delta(x-\alpha)\otimes\delta(y-\beta).$$

In the present paper, we will determine the fundamental solutions of other linear partial differential equations with variable coefficients by applying the same algebraic method.

2. Fundamental solutions of first order linear partial differential equations. In [2, No. 1.3], we have seen that $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$ is a right (resp. left) composition module over the composition algebra $\mathscr{D}_{(-\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(+\Gamma_{\alpha\beta})}$ (resp. $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}_{(-\Gamma_{\alpha\beta})}$).

* Received by the editors March 28, 1977, and in revised form July 24, 1978.

† Départment de Mathématiques, Université Laval, Québec 10, Canada. This article was written while the author was a visitor at University Paris VI (France) in 1976, in the French-Canadian cultural exchange program. Present address: Résidence Les Charmes, 34 Place Champ de Ville, 27400 Louviers, France.

Consider now the equation

$$a(x, y)\frac{\partial\{T\}}{\partial x}+b(x, y)\frac{\partial\{T\}}{\partial y}+c(x, y)\{T\}=S(x, y) \tag{5'}$$

where $a(x, y)$, $b(x, y)$, $c(x, y)$ are operators of multiplication and $S(x, y)$ is a given element of $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$. The fundamental solution $\{E\}$ of (5′) is given by the equation

$$a(x, y)\frac{\partial\{E\}}{\partial x}+b(x, y)\frac{\partial\{E\}}{\partial y}+c(x, y)\{E\}=\delta(x-\alpha)\otimes\delta(y-\beta). \tag{5}$$

For $a(x, y)\neq 0$ in $(+\Gamma_{xy})$, equation (5) is equivalent to the equation

$$\frac{\partial\{E_1\}}{\partial x}+\frac{b(x, y)}{a(x, y)}\frac{\partial\{E_1\}}{\partial y}+\frac{c(x, y)}{a(x, y)}\{E_1\}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{a(x, y)}. \tag{6}$$

The composition of (6) to the left with $Y(x-\alpha)\otimes\delta(y-\beta)$ yields

$$[\delta(x-\alpha)\otimes\delta(y-\beta)+\{H_1(x, y, \alpha, \beta)\}]\circ\{E_1\}=\frac{Y(x-\alpha)\otimes\delta(y-\beta)}{a(\alpha, \beta)} \tag{7}$$

where, by virtue of [2, formula (3.42)], one has

$$\{H_1(x, y, \alpha, \beta)\}=Y(x-\alpha)\frac{b(\alpha, y)}{a(\alpha, y)}\delta'_y(y-\beta)+Y(x-\alpha)\frac{c(\alpha, y)}{a(\alpha, y)}\delta(y-\beta) \tag{8}$$

and from (6) and (7) we deduce

$$\{E_1\}=\left[\frac{1}{a(\alpha, \beta)}\sum_{\nu=0}^{\infty}(-1)^{\nu}\{H_1\}^{\nu}\right]\circ(Y(x-\alpha)\otimes\delta(y-\beta)) \tag{9}$$

in which $\{H_1\}^{\nu}$ is the νth composition power of $\{H_1\}$.

Thus, for $\nu=2$ we obtain

$$\begin{aligned}\{H_1\}^{(2)}=&\left\{\int_{\alpha}^{x}\frac{b(\xi, y)}{a(\xi, y)}\mathrm{d}\xi\frac{\partial}{\partial y}\left\{\frac{b(\alpha, y)}{a(\alpha, y)}\delta'_y(y-\beta)\right\}\right\}\\ &+\left\{\int_{\alpha}^{x}\frac{b(\xi, y)}{a(\xi, y)}\mathrm{d}\xi\frac{\partial}{\partial y}\left\{\frac{c(\alpha, y)}{a(\alpha, y)}\delta(y-\beta)\right\}\right\}\\ &+\left\{\int_{\alpha}^{x}\frac{c(\xi, y)}{a(\xi, y)}\mathrm{d}\xi\frac{b(\alpha, y)}{a(\alpha, y)}\delta'_y(y-\beta)\right\}\\ &+\left\{\int_{\alpha}^{x}\frac{c(\xi, y)}{a(\xi, y)}\mathrm{d}\xi\frac{c(\alpha, y)}{a(\alpha, y)}\delta(y-\beta)\right\}.\end{aligned} \tag{10}$$

For $b(x, y)\neq 0$ in $(+\Gamma_{xy})$, equation (5) is equivalent to the equation

$$\frac{\partial\{E_2\}}{\partial y}+\frac{a(x, y)}{b(x, y)}\frac{\partial\{E_2\}}{\partial x}+\frac{c(x, y)}{b(x, y)}\{E_2\}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{b(x, y)}. \tag{11}$$

In a similar way one can show that the solution $\{E_2\}$ of (11) can be expressed as

$$\{E_2\}=\left[\frac{1}{b(\alpha, \beta)}\sum_{\nu=0}^{\infty}(-1)^{\nu}\{H_2\}^{\nu}\right]\circ\delta(x-\alpha)\otimes Y(y-\beta) \tag{12}$$

in which $\{H_2\}^\nu$ is the νth composition power of the kernel

(13) $$\{H_2\}=Y(y-\beta)\left[\frac{a(x,\beta)}{b(x,\beta)}\delta'_x(x-\alpha)+\frac{c(x,\beta)}{b(x,\beta)}\delta(x-\alpha)\right].$$

In brief, the formal solution of (5) is given by

(14) $$\{E\}=\begin{cases}\left[\dfrac{1}{a(\alpha,\beta)}\sum\limits_{\nu=0}^{\infty}(-1)^\nu\{H_1\}^\nu\right]\circ(Y(x-\alpha)\otimes\delta(y-\beta))=\{E_1\}\\ \qquad\qquad\text{with } a(x,y)\neq 0 \text{ in } (+\Gamma_{xy}),\\ \left[\dfrac{1}{b(\alpha,\beta)}\sum\limits_{\nu=0}^{\infty}(-1)^\nu\{H_2\}^\nu\right]\circ(\delta(x-\alpha)\otimes Y(y-\beta))=\{E_2\}\\ \qquad\qquad\text{for } b(x,y)\neq 0 \text{ in } (+\Gamma_{xy}).\end{cases}$$

Keeping in mind the expressions (8) and (13) of $\{H_1\}$ and $\{H_2\}$ respectively, we find that the fundamental solution $\{E\}$ of the equation (5) given by (14) is an element of $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$.

Then, the solution $\{T\}$ of (5′) may be written in the form (cf. [2, § 1.2 formula (12)]):

$$\{T\}=\{E\}\circ\{S\}=\langle E(x,y,\xi,\eta),S(\xi,\eta,\alpha,\beta)\rangle_{\xi\eta}$$

and has meaning if $\{S\}$ belongs to the composition algebra $\mathscr{D}_{(-\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(+\Gamma_{\alpha\beta})}$. Under these conditions $\{T\}$ is an object of $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha,\beta})}$, considered as a right composition module over the composition algebra $\mathscr{D}_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$.

In particular, for a, b, c real or complex constant functions we obtain

$$\{H_1\}=Y(x-\alpha)\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)$$

and

$$\{E_1\}=\frac{1}{a}\sum_{\nu=0}^{\infty}(-1)^\nu\{H_1\}^\nu\circ(Y(x-\alpha)\otimes\delta(y-\beta)).$$

But

$$\{H_1\}^\nu=Y(x-\alpha)^\nu_{(\alpha,x)}\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)^\nu_{(\beta,y)},$$

where (α, x) (resp. (β, y)) means that the composition powers are taken with respect to the pair of variables (α, x) (resp. (β, y)).

Then we have

$$\{H_1\}^\nu\circ(Y(x-\alpha)\otimes\delta(y-\beta))=Y(x-\alpha)^{\nu+1}_{(\alpha,x)}\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)^\nu_{(\beta,y)}$$

whence

$$\begin{aligned}\{E_1\}&=\frac{1}{a}\sum_{\nu=0}^{\infty}(-1)^\nu Y(x-\alpha)^{\nu+1}_{(\alpha,x)}\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)^\nu_{(\beta,y)}\\&=\frac{1}{a}\sum_{\nu=0}^{\infty}(-1)^\nu\left\{\frac{(x-\alpha)^\nu}{\nu!}\right\}\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)^\nu_{(\beta,y)}.\end{aligned}$$

Now, the series

$$\sum_{\nu=0}^{\infty}(-1)^\nu\frac{(x-\alpha)^\nu}{\nu!}\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)^\nu_{(\beta,y)}$$

may be considered as a Taylor's expansion series of the composition exponential kernel

$$\exp\left[-(x-\alpha)\otimes\left(\frac{b}{a}\delta'_y(y-\beta)+\frac{c}{a}\delta(y-\beta)\right)\right] \tag{15}$$
$$=\exp\left[-\frac{c}{a}(x-\alpha)\right]\exp\left[-(x-\alpha)\otimes\frac{b}{a}\delta'_y(y-\beta)\right].$$

Therefore we have

$$\{E_1\}=\frac{1}{a}Y(x-\alpha)\exp\left(-\frac{c}{a}(x-\alpha)\right)\exp\left[-(x-\alpha)\otimes\frac{b}{a}\delta'_y(y-\beta)\right]. \tag{16}$$

Consider, on the other hand, the Dirac kernel $\delta(y-\beta-(b/a)(x-\alpha))$ as a distribution with respect to (β, y) and as an indefinitely differentiable function with respect to (α, x). Then, the expansion in Taylor's series of this kernel gives us

$$\delta\left(y-\beta-\frac{b}{a}(x-\alpha)\right)=\sum_{\nu=0}^{\infty}(-1)^\nu\frac{(x-\alpha)^\nu}{\nu!}\left(\frac{b}{a}\right)^\nu\otimes\delta_y^{(\nu)}(y-\beta)$$
$$=\exp\left[-(x-\alpha)\otimes\frac{b}{a}\delta'_y(y-\beta)\right].$$

Therefore the solution $\{E_1\}$ of the equation

$$\frac{\partial\{E_1\}}{\partial x}+\frac{b}{a}\frac{\partial\{E_1\}}{\partial y}+\frac{c}{a}\{E_1\}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{a} \tag{17}$$

in which a, b, c are real or complex constants is given by

$$\{E_1\}=\frac{1}{a}Y(x-\alpha)\exp\left[-\frac{c}{a}(x-\alpha)\right]\delta\left(y-\beta-\frac{b}{a}(x-\alpha)\right). \tag{18}$$

In a similar way one can show that the solution $\{E_2\}$ of the equation

$$\frac{a\,\partial\{E_2\}}{b\,\partial x}+\frac{\partial\{E_2\}}{\partial y}+\frac{c}{b}\{E_2\}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{b} \tag{19}$$

where a, b, c are real or complex constants, is of the form

$$\{E_2\}=\frac{1}{b}Y(y-\beta)\exp\left[-\frac{c}{b}(y-\beta)\right]\delta\left(x-\alpha-\frac{a}{b}(y-\beta)\right), \tag{20}$$

where $\delta(x-\alpha-(a/b)(y-\beta))$ is a distribution in (α, x) and an indefinitely differentiable function with respect to the pair of variables (β, y) with:

$$\delta\left(x-\alpha-\frac{a}{b}(y-\beta)\right)=\sum_{\nu=0}^{\infty}(-1)^\nu\left(\frac{a}{b}\right)^\nu\frac{(y-\beta)^\nu}{\nu!}\otimes\delta_x^{(\nu)}(x-\alpha). \tag{21}$$

Consequently, the fundamental solution $\{E\}$ of the equation (5) is given, for a, b, c real or complex constants, by

$$\{E\}=\begin{cases}\dfrac{1}{a}Y(x-\alpha)\exp\left[-\dfrac{c}{a}(x-\alpha)\right]\delta(y-\beta-\dfrac{b}{a}(x-\alpha))=\{E_1\}\\ \qquad\qquad\text{with }\delta(y-\beta-(b/a)(x-\alpha))\in\mathscr{D}'_{(+\Gamma_y)(-\Gamma_\beta)}\hat{\otimes}\mathscr{E}_{\alpha x};\\ \dfrac{1}{b}\exp\left[\dfrac{a}{b}(y-\beta)\right]\delta(x-\alpha-\dfrac{a}{b}(y-\beta))=\{E_2\}\\ \qquad\qquad\text{with }\delta(x-\alpha-(a/b))\in\mathscr{D}'_{(+\Gamma_x)(-\Gamma_\alpha)}\hat{\otimes}\mathscr{E}_{\beta y}.\end{cases} \tag{22}$$

Of course, the solution of the equation (5′) is given by

$$\{T\}=\{E\}\circ\{S\}. \tag{23}$$

Remark. For $\alpha=0$, $\beta=0$, $\{E\}$ coincides with the solution obtained by means of the algebraic operational calculus of distributions with support in R_+^2, which constitutes a commutative convolution algebra without zero divisors (cf. [5, Chap. IV, § 4, No. 4, formula (21)]).

3. Fundamental solutions of linear partial differential equations of order $\geqq 2$. In this section it is required to find the fundamental solution of the equation

$$a(x,y)\frac{\partial^p\{E\}}{\partial x^p}+b(x,y)\frac{\partial^q\{E\}}{\partial x^q}=\delta(x-\alpha)\otimes\delta(y-\beta) \tag{24}$$

in which $a(x,y)$ and $b(x,y)$ are operators of multiplication in $\mathscr{D}'_{(+\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(-\Gamma_{\alpha\beta})}$.

For $a(x,y)\neq 0$ in $(+\Gamma_{xy})$, (24) is equivalent to the equation

$$\frac{\partial^p\{E_1\}}{\partial x^p}+\frac{b(x,y)}{a(x,y)}\frac{\partial^q\{E_1\}}{\partial y^q}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{a(x,y)}. \tag{25}$$

According to our algebraic method, the composition to the left of (25) with the kernel $Y(x-\alpha)^p\otimes\delta(y-\beta)$ gives, for the solution $\{E_1\}$, the formal series

$$\{E_1\}=\left[\frac{1}{a(\alpha,\beta)}\sum_{\nu=0}^{\infty}(-1)^\nu\{H_1\}^\nu\right]\circ(Y(x-\alpha)^p\otimes\delta(y-\beta)) \tag{26}$$

where $\{H_1^\nu\}$ is the νth composition power of the kernel

$$\{H_1\}=Y(x-\alpha)^p\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y^{(q)}(y-\beta). \tag{27}$$

Then we have

$$\begin{aligned}
&\{H_1\}^0=\delta(x-\alpha)\otimes\delta(y-\beta),\\
&\{H_1\}^1=Y(x-\alpha)^p\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y^{(q)}(y-\beta),\\
&\{H_1\}^2=\{Y(x-\alpha)^p\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y^{(q)}(y-\beta)\}^2\\
&\qquad=\left\{\int_\alpha^x\frac{(x-\xi)^{p-1}}{(p-1)!}\frac{(\xi-\alpha)^{p-1}}{(p-1)!}\frac{b(\xi,y)}{a(\xi,y)}d\xi\frac{\partial^q}{\partial y^q}\left[\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y^{(q)}(y-\beta)\right]\right\},\\
&\qquad\cdots\quad\cdots\quad\cdots\quad\cdots\\
&\{H_1\}^\nu=\left\{\int_\alpha^x\frac{(x-\xi)^{p-1}}{(p-1)!}\frac{b(\xi,y)}{a(\xi,y)}d\xi\int_\alpha^\xi\frac{(\xi-\xi_1)^{p-1}}{(p-1)!}d\xi_1\cdots\right.\\
&\qquad\cdots\int_\alpha^{\xi_{\nu-3}}\frac{(\xi_{\nu-3}-\xi_{\nu-2})^{p-1}}{(p-1)!}\frac{(\xi_{\nu-2}-\alpha)^{p-1}}{(p-1)!}d\xi_{\nu-2}\\
&\qquad\cdot\frac{\partial^q}{\partial y^q}\left[\frac{b(\xi_1,y)}{a(\xi_1,y)}\frac{\partial^q}{\partial y^q}\left(\frac{b(\xi_2,y)}{a(\xi_2,y)}\right)\right]\cdots\\
&\qquad\cdot\frac{\partial^q}{\partial y^q}\left\{\frac{b(\xi_{\nu-2},y)}{a(\xi_{\nu-2},y)}\frac{\partial^q}{\partial y^q}\left(\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y^{(q)}(y-\beta)\right)\right\}\Bigg\}\cdots
\end{aligned} \tag{28}$$

Likewise, for $b(x, y) \neq 0$ in $(+\Gamma_{xy})$, the fundamental solution $\{E_2\}$ of the equation

$$\frac{\partial^q\{E_2\}}{\partial y^q}+\frac{a(x, y)}{b(x, y)}\frac{\partial^p\{E_2\}}{\partial x^p}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{b(x, y)} \tag{29}$$

is given by

$$\{E_2\}=\left[\frac{1}{b(\alpha, \beta)}\sum_{\nu=0}^{\infty}(-1)^\nu\{H_2\}^\nu\right]\circ(\delta(x-\alpha)\otimes Y(-\beta)^q) \tag{30}$$

in which the kernel $\{H_2\}$ has the expression

$$\{H_2\}=\delta_x^{(p)}(x-\alpha)\frac{a(x, \beta)}{b(x, \beta)}Y(y-\beta)^q. \tag{31}$$

Assume now that a, b, c are real or complex constant functions. Then we obtain

$$\{H_1\}=Y(x-\alpha)^p\otimes\frac{b}{a}\delta_y^{(q)}(y-\beta) \tag{32}$$

whence

$$\{H_1\}^\nu=Y(x-\alpha)^{\nu p}\otimes\left(\frac{b}{a}\right)^\nu\delta_y^{(\nu q)}(y-\beta)$$

and

$$\{E_1\}=\left[\sum_{\nu=0}^{\infty}(-1)^\nu Y(x-\alpha)^{\nu p}\otimes\left(\frac{b}{a}\right)^\nu\delta_y^{(\nu q)}(y-\beta)\right]\circ\frac{1}{a}(Y(x-\alpha)^p\otimes\delta(y-\beta)). \tag{33}$$

In particular, for $p=q=2$, the fundamental solution $\{E\}$ of the equation

$$a\frac{\partial^2\{E\}}{\partial x^2}+b\frac{\partial^2\{E\}}{\partial y^2}=\delta(x-\alpha)\otimes\delta(y-\beta) \tag{34}$$

is given by

$$\{E_1\}=\frac{1}{a}\left[\sum_{\nu=0}^{\infty}(-1)^\nu Y(x-\alpha)^{2\nu}\otimes\left(\frac{b}{a}\right)^\nu\delta_y^{(2\nu)}(y-\beta)\right]\circ(Y(x-\alpha)^2\otimes\delta(y-\beta)). \tag{35}$$

Now, if we set $\{E_1\}=\frac{1}{2}\{E_1\}2$, and if we write

$$\{E_1\}=\left[\frac{1}{a}\sum_{\nu=0}^{\infty}Y(x-\alpha)^{2\nu+1}\otimes\left(i\sqrt{\frac{b}{a}}\right)^{2\nu}\left[\delta_y^{(2\nu)}(y-\beta)\right]\right.\circ(Y(x-\alpha)\otimes\delta(y-\beta))$$

where $i=\sqrt{-1}$; then by adding and subtracting in the right handside of (35) the term

$$\left[\frac{1}{2a}\sum_{\nu=0}^{\infty}Y(x-\alpha)^{2\nu}\otimes\left(i\sqrt{\frac{b}{a}}\right)^{2\nu-1}\delta_y^{(2\nu-1)}(y-\beta)\right]\circ(Y(x-\alpha)\otimes\delta(y-\beta))$$

we obtain

$$\begin{aligned}\{E_1\} = \frac{1}{2a}\Big[& Y(x-\alpha)\otimes\delta(y-\beta)+\{(x-\alpha)\}\otimes i\sqrt{\frac{b}{a}}\,\delta'_y\,(y-\beta)\\ & +\left\{\frac{(x-\alpha)^2}{2!}\right\}\otimes\left(i\sqrt{\frac{b}{a}}\right)^2\delta''_y(y-\beta)+\cdots\\ & \cdots+\left\{\frac{(x-\alpha)^\nu}{\nu!}\right\}\otimes\left(i\sqrt{\frac{b}{a}}\right)^\nu\delta_y^{(\nu)}\,(y-\beta)+\cdots\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta)\\ +\frac{1}{2a}\Big[& Y(x-\alpha)\otimes\delta(y-\beta)-\{(x-\alpha)\}\otimes i\sqrt{\frac{b}{a}}\,\delta'_y\,(y-\beta)\\ & +\left\{\frac{(x-\alpha)^2}{2!}\right\}\otimes\left(i\sqrt{\frac{b}{a}}\right)^2\delta''_y(y-\beta)-\cdots\\ & \cdots+(-1)^\nu\left\{\frac{(x-\alpha)^\nu}{\nu!}\right\}\otimes\left(i\sqrt{\frac{b}{a}}\right)^\nu\delta_y^{(\nu)}\,(y-\beta)+\cdots\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta).\end{aligned} \tag{36}$$

But, we have

$$\sum_{\nu=0}^{\infty}\frac{(x-\alpha)^\nu}{\nu!}\otimes\left(i\sqrt{\frac{b}{a}}\right)^\nu\delta_y^{(\nu)}\,(y-\beta)=\delta\left(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\right) \tag{37}$$

and

$$\sum_{\nu=0}^{\infty}(-1)^\nu\frac{(x-\alpha)^\nu}{\nu!}\otimes\left(i\sqrt{\frac{b}{a}}\right)^\nu\delta_y^{(\nu)}\,(y-\beta)=\delta\left(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\right). \tag{38}$$

Therefore

$$\begin{aligned}\{E_1\}=\frac{1}{2a}Y(x-\alpha)\Big[&\delta\left(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\right)\\ &+\delta\left(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\right)\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta)\end{aligned} \tag{39}$$

where $\delta(y-\beta\pm i\sqrt{(b/a)}(x-\alpha))$ should be considered as distributions in (β, y) and as indefinitely differentiable functions in (α, x).

Likewise, by writing the solution of (34) in the form $\{E_2\}$ given by

$$\frac{\partial^2\{E_2\}}{\partial y^2}+\frac{a\,\partial^2\{E_2\}}{b\,\partial x^2}=\frac{\delta(x-\alpha)\otimes\delta(y-\beta)}{b} \tag{40}$$

we find

$$\begin{aligned}(41)\quad \{E_2\}=\frac{1}{2b}Y(y-\beta)\Big[&\delta\left(x-\alpha+i\sqrt{\frac{a}{b}}(y-\beta)\right)\\ &+\delta\left(x-\alpha-i\sqrt{\frac{a}{b}}(y-\beta)\right)\Big]\circ\delta(x-\alpha)\otimes Y(y-\beta).\end{aligned} \tag{41}$$

with $\delta(x-\alpha\pm i\sqrt{(b/a)}(y-\beta))$ considered as distributions in (α, x) and as indefinitely differentiable functions in (β, y).

Let us show that $\{E\}$ given by (39) (resp. (41)) satisfies the equation (34). For $\{E_1\}$ we obtain

$$\frac{\partial\{E_1\}}{\partial x}=\frac{1}{a}Y(x-\alpha)\otimes\delta(y-\beta)$$
$$+\frac{1}{2a}i\sqrt{\frac{b}{a}}\,Y(x-\alpha)\Big[\delta'\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)$$
$$-\delta'\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta)$$

and

$$\frac{\partial^2\{E_1\}}{\partial x^2}=\frac{1}{a}\delta(x-\alpha)\otimes\delta(y-\beta)$$
$$+\frac{1}{2a}Y(x-\alpha)\Big(i\sqrt{\frac{b}{a}}\Big)^2\Big[\delta''\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)$$
$$+\delta''\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta)$$
$$=\frac{1}{a}\delta(x-\alpha)\otimes\delta(y-\beta)+\frac{1}{2a}Y(x-\alpha)\Big[-i\sqrt{\frac{b}{a}}\frac{\partial}{\partial\alpha}\delta'\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)$$
$$+i\sqrt{\frac{b}{a}}\frac{\partial}{\partial\alpha}\delta'\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta)$$
$$=\frac{1}{a}\delta(x-\alpha)\otimes\delta(y-\beta)+\frac{i}{2a}\sqrt{\frac{b}{a}}\,Y(x-\alpha)\Big[\delta'\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)$$
$$-\delta'\Big(v-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\Big].$$

Furthermore, we have

$$\frac{\partial^2\{E_1\}}{\partial y^2}=\frac{1}{2a}Y(x-\alpha)\Big[\delta''\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)$$
$$+\delta''\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\Big]\circ Y(x-\alpha)\otimes\delta(y-\beta)$$
$$=\frac{1}{2a}\Big\{\int_\alpha^x\Big[\delta''\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\xi)\Big)+\delta''\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\xi)\Big)\Big]\,d\xi\Big\}.$$

But

$$\delta''\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\xi)\Big)=i\sqrt{\frac{a}{b}}\frac{\partial}{\partial\xi}\delta'\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\xi)\Big)$$

and

$$\delta''\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\xi)\Big)=-i\sqrt{\frac{a}{b}}\frac{\partial}{\partial\xi}\delta'\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\xi)\Big).$$

Therefore

$$\frac{\partial^2\{E_1\}}{\partial y^2}=-\frac{i}{2a}\sqrt{\frac{a}{b}}\delta\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)+\frac{i}{2a}\sqrt{\frac{a}{b}}\delta\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\circ Y(x-\alpha)$$

whence

$$a\frac{\partial^2\{E_1\}}{\partial x^2}+b\frac{\partial^2\{E_1\}}{\partial y^2}=\delta(x-\alpha)\otimes\delta(y-\beta).$$

In a similar way or by reason of symmetry one shows that $\{E_2\}$ given by (41) satisfies also the equation (34).

Let us return to $\{E_1\}$ given by (38). If we set $\lambda = i\sqrt{b/a}$ we obtain

$$\begin{aligned}\{E_1\}&=\frac{1}{2a}Y(x-\alpha)\Big[\delta(y-\beta+\lambda(x-\alpha))+\delta(y-\beta-\lambda(x-\alpha))]\circ Y(x-\alpha)\otimes\delta(y-\beta)\\ &=\frac{1}{2a}\Big\{\int_\alpha^x\Big[\delta(y-\beta+\lambda(x-\xi))+\delta(y-\beta-\lambda(x-\xi))\Big]\,\mathrm{d}\xi\Big\}\\ &=\frac{1}{2a}\cdot\frac{1}{\lambda}\Big[Y(y-\beta+\lambda(x-\alpha))-Y(y-\beta-\lambda(x-\alpha))\Big]\end{aligned}$$

where $Y(y-\beta+\lambda(x-\alpha))$ (resp. $Y(y-\beta-\lambda(x-\alpha))$ is defined by the Taylor series

$$Y(y-\beta+\lambda(x-\alpha))=\sum_{\nu=0}^{\infty}\lambda^\nu\Big\{\frac{(x-\alpha)^\nu}{\nu!}\Big\}\frac{\partial^\nu}{\partial y^\nu}Y(y-\beta)$$

$$\Big(\text{resp.}\quad\sum_{\nu=0}^{\infty}(-1)^\nu\lambda^\nu\Big\{\frac{(x-\alpha)^\nu}{\nu!}\Big\}\frac{\partial^\nu}{\partial y^\nu}Y(y-\beta)\Big).$$

Therefore

$$\{E_1\}=\frac{1}{2}\frac{i}{\sqrt{ab}}\Big[Y\Big(y-\beta-i\sqrt{\frac{b}{a}}(x-\alpha)\Big)-Y\Big(y-\beta+i\sqrt{\frac{b}{a}}(x-\alpha)\Big)\Big].\tag{42}$$

Likewise, for $\{E_2\}$ given by (41), we obtain

$$\{E_2\}=\frac{i}{2\sqrt{ab}}\Big[Y\Big(x-\alpha-i\sqrt{\frac{a}{b}}(y-\beta)\Big)-Y\Big(x-\alpha+i\sqrt{\frac{a}{b}}(y-\beta)\Big)\Big].\tag{43}$$

In particular, for $a=b=1$, the elliptic equation

$$\frac{\partial^2\{E\}}{\partial x^2}+\frac{\partial^2\{E\}}{\partial y^2}=\delta(x-\alpha)\otimes\delta(y-\beta)$$

has the fundamental solution

$$\{E\}=\begin{cases}\tfrac{1}{2}Y(x-\alpha)[\delta(y-\beta+i(x-\alpha))\\ \qquad +\delta(y-\beta-i(x-\alpha))]\circ Y(x-\alpha)\otimes\delta(y-\beta)=\{E_1\}\\ \qquad \text{with } \delta(y-\beta\pm i(x-\alpha)) \text{ considered as distributions in}\\ \qquad (\beta, y) \text{ and indefinitely differentiable functions in } (\alpha, x);\\ \tfrac{1}{2}Y(y-\beta)[\delta(x-\alpha+i(y-\beta)\\ \qquad +\delta(x-\alpha-i(y-\beta))]\circ\delta(x-\alpha)\otimes Y(y-\beta)=\{E_2\}\\ \qquad \text{with } \delta(x-\alpha\pm i(y-\beta)) \text{ considered as distributions in}\\ \qquad (\alpha, x) \text{ and indefinitely differentiable functions in } (\beta, y).\end{cases}\tag{44}$$

Keeping in mind (42) and (43), we find that (44) is equivalent to

$$
(45)\qquad \{E\}=\begin{cases}(i/2)[Y(y-\beta-i(x-\alpha))-Y(y-\beta+i(x-\alpha))] \\ \qquad \text{with } Y(y-\beta\pm i(x-\alpha)) \text{ distributions in } (\beta, y) \text{ and} \\ \qquad \text{indefinitely differentiable functions in } (\alpha, x); \\ (i/2)[Y((x-\alpha)-i(y-\beta))-Y(x-\alpha+i(y-\beta))] \\ \qquad \text{with } Y(x-\alpha\pm i(y-\beta)) \text{ distributions in } (\alpha, x) \text{ and} \\ \qquad \text{indefinitely differentiable functions in } (\beta, y).\end{cases}
$$

4. Fundamental solution of the hyperbolic equation. For $a\equiv 1, b=-\lambda^2, \lambda>0$, the fundamental solution of the hyperbolic equation

$$
\frac{\partial^2\{E\}}{\partial x^2}-\lambda^2\frac{\partial^2\{E\}}{\partial y^2}=\delta(x-\alpha)\otimes\delta(y-\beta)
$$

is given by

$$
(46)\ \{E\}=\begin{cases}\tfrac{1}{2}Y(x-\alpha)[\delta(y-\beta+\lambda(x-\alpha)) \\ \qquad +\delta(y-\beta-\lambda(x-\alpha))]\circ Y(x-\alpha)\otimes\delta(y-\beta)=\{E_1\} \\ \qquad \text{with } \delta(y-\beta\pm\lambda(x-\alpha)) \text{ considered as distributions in} \\ \qquad (\beta, y) \text{ and indefinitely differentiable functions in } (\alpha, x); \\ \dfrac{-1}{2\lambda^2}Y(y-\beta)\Big[\delta(x-\alpha+\dfrac{1}{\lambda}(y-\beta)) \\ \qquad +\delta(x-\alpha-\dfrac{1}{\lambda}(y-\beta))\Big]\circ\delta(x-\alpha)\otimes Y(y-\beta)=\{E_2\} \\ \qquad \text{with } \delta(x-\alpha\pm(1/\lambda)(y-\beta)) \text{ distributions in } (\alpha, x) \text{ and} \\ \qquad \text{indefinitely differentiable functions in } (\beta, y);\end{cases}
$$

which is equivalent to

$$
(47)\qquad \{E\}=\begin{cases}\dfrac{1}{2\lambda}\Big\{Y(y-\beta+\lambda(x-\alpha))-Y(y-\beta-\lambda(x-\alpha))\Big\}=\{E_1\} \\ \qquad \text{with } Y(y-\beta\pm\lambda(x-\alpha)) \text{ distributions in } (\beta, y) \text{ and} \\ \qquad \text{indefinitely differentiable functions in } (\alpha, x); \\ \dfrac{1}{2\lambda}\Big\{Y(x-\alpha-\dfrac{1}{\lambda}(y-\beta))-Y(x-\alpha+\dfrac{1}{\lambda}(y-\beta))\Big\}=\{E_2\} \\ \qquad \text{with } Y(x-\beta\pm(1/\lambda)(y-\beta)) \text{ distributions in } (\alpha, x) \text{ and} \\ \qquad \text{indefinitely differentiable functions in } (\beta, y).\end{cases}
$$

5. Fundamental solution of parabolic equations. For $p=2, q\simeq 1, a(x, y)\neq 0$ in $(+\Gamma_{xy})$, the equation

$$
(48)\qquad a(x, y)\frac{\partial^2\{E\}}{\partial x^2}+b(x, y)\frac{\partial\{E\}}{\partial y}=\delta(x-\alpha)\otimes\delta(v-\beta)
$$

has the formal solution

$$\{E_1\}=\left[\frac{1}{a(\alpha,\beta)}\sum_{\nu=0}^{\infty}(-1)^{\nu}\{H_1\}^{(\nu)}\right]\circ(Y(x-\alpha)^2\otimes\delta(y-\beta))$$

where, according to (27)

$$\{H_1\}=\left\{Y(x-\alpha)^2\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y'(y-\beta)\right\}$$

and

$$\{H_1\}^{(\nu)}=\left\{Y(x-\alpha)^2\frac{b(\alpha,y)}{a(\alpha,y)}\delta_y'(y-\beta)\right\}^{(\nu)}.$$

Likewise, for $b(x,y)\neq 0$ in $(+\Gamma_{xy})$ we obtain the formal solution

$$\{E_2\}=\left[\sum_{\nu=0}^{\infty}\frac{1}{b(\alpha,\beta)}(-1)^{\nu}\{H_2\}^{\nu}\right]\circ(\delta(x-\alpha)\otimes Y(y-\beta)) \tag{49}$$

in which, according to (31),

$$\{H_2\}^{\nu}=\left\{Y(y-\beta)\frac{a(x,\alpha)}{b(x,\beta)}\delta_x''(x-\alpha)\right\}^{(\nu)}.$$

In particular, for $a=1, b=-1$, we obtain for the parabolic equation the fundamental formal solution

(50) $\{E\}=$

$$\{E_1\}=\sum_{\nu=0}^{\infty}\left\{\frac{(x-\alpha)^{2\nu+1}}{(2\nu+1)!}\right\}\otimes\delta_y^{(\nu)}(y-\beta)=\text{formal series}$$

belonging to the ring $[[(\mathscr{E}_{\alpha x}\otimes\mathscr{E}'_{\beta y})^{\mathbb{N}}]]$ of formal series whose terms belong to the composition algebra $\mathscr{E}_{\alpha x}\otimes\mathscr{E}'_{\beta y}$;

$$\{E_2\}=-\sum_{\nu=0}^{\infty}\delta_x^{(2\nu)}(x-\alpha)\otimes\frac{\{(y-\beta)^{\nu}\}}{\nu!},$$ as an element of $[[(\mathscr{E}'_{\alpha x}\otimes\mathscr{E}_{\beta y})^{\mathbb{N}}]]$;

on the composition ring of the form $[[(\mathscr{D}_{(-\Gamma_{xy})}\hat{\otimes}\mathscr{D}'_{(+\Gamma_{\alpha\beta})})^{\mathbb{N}}]]$ (cf. [2, § 4.2]).

Remark. For $\alpha=\beta=0$, the solution (50) coincides with the solution of the parabolic equation

$$\frac{\partial^2\{E\}}{\partial x^2}-\frac{\partial\{E\}}{\partial y}=\delta(x)\otimes\delta(y)$$

obtained by means of our algebraic operational calculus of distributions (cf. [8, formula (6)]).

Our next paper will be devoted to the boundary value problems and to the problems of convergence of the formal solutions obtained by our algebraic method.

Acknowledgment. The author wishes to thank the referee for numerous helpful suggestions.

REFERENCES

[1] S. VASILACH, *Algebraic method for solving linear differential equations whose coefficients are functions of one variable*, this Journal, 6 (1975), pp. 295–311.

[2] ———, *Algebraic method for solving linear partial differential equations with variable coefficients. Part I: Basic theory*, this Journal, 10 (1979) pp. 586–602.

[3] ———, *Produit de composition dans des produits tensoriels*, C.R. Acad. Sci. Paris Sér. A–B, 284 (1967), pp. 100–103, 179–182.

[4] ———, *Product of Composition in Tensor Products of bimodules*, Atti. Acad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. 52 (1972), pp. 277–286.

[5] ———, *Calcul opérationnel algébrique des distributions à support dans R^n_+, $n \geqq 1$*, Rev. Math. Pures Appl., 2 (1957), pp. 181–238; I, 4 (1959), pp. 185–219; II, 5 (1960), pp. 493–531; III, 6 (1961), pp. 69–100; IV, V, 8 (1963), pp. 19–66.

[6] ———, *Sur une classe d'espaces de distributions*, Boll. Un. Mat. Ital., 5 (1970), pp. 745–760.

[7] L. SCHWARTZ, *Théorie des Distributions*, Hermann, Paris, 1966.

[8] S. VASILACH, Solution of the equation $\lambda^2 \frac{\partial^2\{E\}}{\partial x^2} - \mu^2 \frac{\partial\{E\}}{cy} = \delta(x) \otimes \delta(y)$ by means of the algebraic operational calculus of distributions with support in R^2_+, Publ. Inst. Math. (Beograd), 13 (1972), no. 27, pp. 127–141.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-017 $01.0/0

STABILITY OF LINEAR VOLTERRA INTEGRODIFFERENTIAL EQUATIONS OF ORDER n*

R. K. MILLER† AND A. N. MICHEL‡

Abstract. A result is proved which relates integrability of the set of resolvent functions for an nth order Volterra integrodifferential equation with integrable kernels and a certain natural Laplace transform condition.

The authors wish to provide a proof of a result announced in [4]. This result generalizes the work in § V of [2] concerning the integrability of the set of resolvent functions for an nth order Volterra integrodifferential equation with L^1-kernels. Related work on the problem where only the integral resolvent is considered can be found in the interesting work of Callier and Desoer [1]. Shea and Wainger [5] have provided an interesting generalization of the work in [2] for the first order case. Jordan and Wheeler [3] have improved the results of Shea and Wainger.

THEOREM. *Let A_j be constant matrices for $j=0,1,\cdots,n$, let $B_j\in L^1(0,\infty)$ be integrable matrix valued functions for $j=0,1,\cdots,n+1$, and define*

$$a(t)=B_{n+1}(t)+\sum_{j=0}^{n}\left(\frac{A_jt^{n-j}}{(n-j)!}+\int_0^t\int_0^{t_1}\int_0^{t_2}\cdots\int_0^{t_j}B_{n-j}(t_{j+1})\,dt_{j+1}\cdots dt_1\right).$$

Assume that $A_0+B_0(t)\not\equiv 0$. Let $r_0(t)$ be the resolvent of $a(t)$, that is, r_0 is the unique solution of

$$r_0(t)=-a(t)+\int_0^t a(t-\tau)r_0(\tau)\,d\tau,\qquad t\geqq 0, \tag{1}$$

and for $j=0,1,\cdots,n$ define

$$r_{n+1-j}(t)=\frac{d^j}{dt^j}\left\{\frac{t^n}{n!}I-\int_0^t\int_0^{t_1}\cdots\int_0^{t_n}r_0(t_{n+1})\,dt_{n+1}\cdots dt_1\right\}, \tag{2}$$

or equivalently in terms of Laplace transforms

$$r_0^*(s)=-(I-a^*(s))^{-1}a^*(s),\qquad r_j^*(s)=s^{-j}(I-a^*(s))^{-1}.$$

Then the following are equivalent:

(a) $\det s^{n+1}(I-a^*(s))\neq 0$ *for all s with* $\operatorname{Re} s\geqq 0$.

(b) $r_j\in L^1(0,\infty)$ *for $j=0,1,\cdots,n+1$.*

Remark 1. If $n=-1$ so that $a\equiv B_0\in L^1(0,\infty)$, then this theorem reduces to the following result of Paley and Wiener [6, p. 60]: $r_0\in L^1(0,\infty)$ if and only if $\det(I-B_0^*(s))\neq 0$ when $\operatorname{Re} s\geqq 0$.

Remark 2. If $B_{n+1}\equiv 0$, then the equivalence of (a) and (b) was proved in [2].

Remark 3. It was shown in [2] that (a) implies that $r_0\in L^1(0,\infty)$.

Remark 4. This theorem corrects an error in the original result—Theorem 1 of [4].

Proof. Assume that (a) is true. By Remark 3, $r_0\in L^1(0,\infty)$. It remains to show that $r_j\in L^1(0,\infty)$ for $1\leqq j\leqq n+1$ in order to see that (b) is true. Let $b_{n+1}(t)$ be a C^∞-smooth

* Received by the editors December 19, 1977, and in revised form June 1, 1978.

† Mathematics Department, Iowa State University, Ames, Iowa 50011. The work of this author was supported by the National Science Foundation under Grant MCS77-03863.

‡ Electrical Engineering Department, Iowa State University, Ames, Iowa 50011. The work of this author was supported by the National Science Foundation under Grant ENG75-14093.

function with compact support in $[0, \infty)$ such that

$$\int_0^\infty |b_{n+1}(t) - B_{n+1}(t)|\, dt < \varepsilon.$$

Choose ε so small that if

$$a_0(t) = b_{n+1}(t) + \{a(t) - B_{n+1}(t)\},$$

then

$$\det s^{n+1}(I - a_0^*(s)) \neq 0 \quad \text{for } \operatorname{Re} s \geqq 0. \tag{3}$$

This is possible because (a) is true.

Define $\Delta B = B_{n+1} - b_{n+1}$ and let ρ_0 be the resolvent of a_0, that is $\rho_0^*(s) = -(I - a_0^*(s))^{-1} a_0^*(s)$. Let $\phi \in C(R^+)$ be a given, fixed bounded function. Define

$$f(t) = \int_0^t \phi(\tau)\, d\tau,$$

and let $x(t)$ be the solution of the equation

$$x(t) = f(t) + \int_0^t a(t-\tau) x(\tau)\, d\tau.$$

If $*$ denotes the convolution integral, then this can be expressed as $f = I * \phi$ and $x = f + a * x$. Equivalently x solves

$$x = f + a_0 * x + \Delta B * x,$$

or

$$x - a_0 * x = f + \Delta B * x.$$

Since ρ_0 is the resolvent of a_0, then

$$x = f + \Delta B * x - \rho_0 * f - \rho_0 * \Delta B * x$$

or

$$x = (I - \rho_0 * I) * \phi + (\Delta B - \rho_0 * \Delta B) * x. \tag{4}$$

By Remark 2 and by (3) it follows that $\rho_0 \in L^1(0, \infty)$ and $\rho_1 = I - \rho_0 * I \in L^1(0, \infty)$.

Since $\Delta B - \rho_0 * \Delta B \in L^1(0, \infty)$ and since

$$\begin{aligned} \det(I - \Delta B^*(s) + \rho_0^*(s)\, \Delta B^*(s)) &= \det\{(I - a_0^*(s))^{-1}(I - a_0^*(s) + \Delta B^*(s))\} \\ &= \det\{(I - a_0^*(s))^{-1}(I - a^*(s))\} \neq 0 \end{aligned}$$

for $\operatorname{Re} s \geqq 0$, then by the Paley–Wiener theorem (Remark 1) the resolvent H of $\Delta B - \rho_0 * \Delta B$ is in $L^1(0, \infty)$. The solution of (4) can be written as

$$\begin{aligned} x &= (I - \rho_0 * I) * \phi - H * (I - \rho_0 * I) * \phi \\ &= (\rho_1 - H * \rho_1) * \phi = H_1 * \phi \end{aligned}$$

where $H_1 = \rho_1 - H * \rho_1 \in L^1(0, \infty)$. On the other hand the solution x of $x = f + a * x$, $f = I * \phi$ is

$$x = f - r_0 * f = (I - r_0 * I) * \phi,$$

and $r_1 = I - r_0 * I$. Thus r_1 and H_1 are the same function. Since $H_1 \in L^1$ then $r_1 \in L^1(0, \infty)$.

A similar argument can be used for the functions $r_2, r_3, \cdots, r_{n+1}$ successively. Let $f = I * I * \phi$ and let x solve

$$x = f + a * x = f + a_0 * x + \Delta B * x$$

where a_0 and ΔB are as defined above. If ρ_0 is the resolvent of a_0, then as before

$$\begin{aligned} x &= f - \rho_0 * f + (\Delta B - \rho_0 * \Delta B) * x \\ &= \rho_2 * \phi + (\Delta B - \rho_0 * \Delta B) * x \end{aligned}$$

and there exist functions H, $\rho_2 = I * \rho_1$ and $H_2 = \rho_2 - H * \rho_2$ all in $L^2(0, \infty)$ such that

$$\begin{aligned} x &= \rho_2 * \phi - H * \rho_2 * \phi \\ &= (\rho_2 - H * \rho_2) * \phi = H_2 * \phi. \end{aligned}$$

Thus $r_2 = H_2 \in L^1(0, \infty)$. Continue in this manner to see that (b) is true.

To prove that (b) implies (a) assume that (a) is false. Then there is a point s_0 with $\operatorname{Re} s_0 \geqq 0$ such that $\det s_0^{n+1}(I - a^*(s_0)) = 0$. This means that

$$r_{n+1}^*(s) = s^{-n-1}(I - a^*(s))^{-1}$$

has a singularity at $s = s_0$. Thus r_{n+1} cannot be in $L^1(0, \infty)$ and (b) is false. Q.E.D.

REFERENCES

[1] F. M. CALLIER AND C. A. DESOER, *Necessary and sufficient conditions for stability for n-input, n-output convolution feedback systems with a finite number of unstable poles*, IEEE Trans. Automatic Control, AC-18 (1973), pp. 295–298.

[2] S. I. GROSSMAN AND R. K. MILLER, *Nonlinear Volterra integrodifferential systems with L^1-kernels*, J. Differential Equations, 13 (1973), pp. 551–566.

[3] G. S. JORDON AND R. L. WHEELER, *A generalization of the Wiener–Lévy theorem applicable to some Volterra equations*, Proc. Amer. Math. Soc., 57 (1976), pp. 109–114.

[4] R. K. MILLER AND A. N. MICHEL, *Stability of multivariable feedback systems containing elements which are open loop unstable*, Proceedings 13th Annual Allerton Conference on Circuit and System Theory, Univ. of Illinois at Urban-Champaign, Urbana, 1975.

[5] D. F. SHEA AND S. WAINGER, *Variants of the Wiener–Lévy theorem, with applications to stability problems for some Volterra integral equations*, Amer. J. Math., 97 (1975), pp. 312–343.

[6] R. E. A. C. PALEY AND N. WIENER, *Fourier Transforms in the Complex Domain*, American Mathematical Society Colloquium Publications, Providence, RI, 1934.

SIAM J. MATH. ANAL.
Vol. 10, No. 5, September 1979

0036-1410/79/1005-0018 $01.00/0

THE UNSETTLED PROBLEM OF M. G. KREIN ON NONNEGATIVE POLYNOMIALS IN T-SYSTEMS*

R. K. S. RATHORE† AND P. N. AGRAWAL†

Abstract. The unsettled problem of M. G. Krein was as follows: If $\{u_i\}_{i=0}^{2m}$ $(m = 1, 2, \cdots)$ is a T-system on $[a, b]$ and $T = \{t_1, t_2, \cdots, t_k\} \subset [a, b]$ is a set of distinct points containing only one of the end points a, b and $k \leqq m$, then does there always exist a nonnegative linear combination of u_i's vanishing precisely on T in $[a, b]$? In this note we settle this problem by furnishing a counterexample for which such a linear combination does not exist.

In this note we furnish a counterexample for the unsettled problem of M. G. Krein on the existence of nonnegative polynomials in T-systems $\{u_i\}_{i=0}^{2m}$ $(m = 1, 2, \cdots)$ on an interval $[a, b]$, vanishing precisely on a given set $T = \{t_1, t_2, \cdots, t_k\}$ containing only one of the end points a or b.

The problem arose in the following manner. M. G. Krein [2] (see Karlin and Studden [1, Thm. 5.1, p. 28]) proved the following result.

THEOREM 1. *Let $\{u_i\}_{i=0}^{n}$ be a T-system on $[a, b]$. Let $T = \{t_1, t_2, \cdots, t_k\} \subset [a, b]$. Define*

$$w(t_i) = \begin{cases} 2, & t_i \in (a, b), \\ 1, & t_i = a \text{ or } b. \end{cases}$$

Then, (a) *if $\sum_{i=1}^{k} w(t_i) \leqq n$, there exists a nontrivial, nonnegative polynomial $u(t) = \sum_{i=0}^{n} \beta_i u_i(t)$ vanishing precisely on T and at no other point of $[a, b]$. The only exception is that if $n = 2m$ and exactly one of the end points a or b is in T then $u(t)$ may vanish at the other end point as well.*

(b) *If any one of the following further conditions holds then without exception the polynomial may be constructed to vanish precisely at $t_1, t_2, \cdots, t_k$:*

(i) *$\{u_i\}_{i=0}^{n-1}$ (i.e. the set of functions $u_0, u_1, u_2, \cdots, u_{n-1}$ excluding u_n) is a T-system.*

(ii) *$\{u_i\}_{i=0}^{n}$ is a T-system on an interval $[a', b']$ containing $[a, b]$ where $a' < a < b < b'$.*

(iii) *$\{u_i\}_{i=0}^{n}$ is an ET-system of order 2 on $[a, b]$.*

Regarding the exceptional case in the above theorem Karlin and Studden [1, p. 30] remark that "it is worthwhile to notice the slightly weaker conclusion in the case where only one of the end points a or b is contained in T (and $n = 2m$). Whether or not this exceptional case can be eliminated from the theorem has not been settled."

In the sequel we show that this exceptional case cannot be eliminated from the theorem. For this purpose we employ an interesting example of a T-system due to Zielke [3].

PROPOSITION 1. *The set $\{p_i\}_{i=0}^{n}$ of functions defined by*

$$p_0(t) = 1,$$

$$p_1(t) = (1-t)t,$$

$$p_i(t) = (1-t)t^{i-2}(t^2-1), \qquad i = 2, \cdots, n,$$

is a T-system on $[-1, 1]$ if $n = 2m$ $(m = 1, 2, \cdots)$.

To render the paper self contained we first give an elementary new proof of the above proposition.

* Received by the editors December 9, 1977.

† Department of Mathematics, Indian Institute of Technology, Kanpur, India.

Proof. Let $P(t)=\sum_{i=0}^{n}\alpha_i p_i(t)$ be a nontrivial polynomial in $\{p_i(t)\}_{i=0}^{n}$ where n is even. To prove the proposition we have to show that $P(t)$ can have at most n distinct zeros on $[-1, 1]$.

We notice first that if $\alpha_n=0$, then $P(t)$ is a polynomial of degree at most n and hence it can have at most n zeros on $[-1, 1]$. Thus without loss of generality we can assume that $\alpha_n>0$. The rest of the proof is divided in six parts:

(i) If $\alpha_0>0$, since $\sum_{i=1}^{n}\alpha_i p_i(t)$ is zero at $t=1$ and tends to $-\infty$ as $t\to+\infty$, it follows that $P(t)$ has at least one zero on $(1,\infty)$. Hence $P(t)$ (being a polynomial of degree $n+1$) can have at most n zeros on $[-1, 1]$.

(ii) If $\alpha_0<0$ and $\alpha_1\geqq 0$, $\sum_{i=1}^{n}\alpha_i p_i(t)$ is $-2\alpha_1\leqq 0$ at $t=-1$ and tends to $+\infty$ as $t\to-\infty$. Hence $P(t)$ must have one zero on $(-\infty,-1)$ and therefore at most n zeros on $[-1, 1]$.

(iii) If $\alpha_0, \alpha_1<0$ and $P(t)$ has $n+1$ distinct zeros on $[-1, 1]$, by Rolle's theorem $P'(t)$ has n distinct zeros on $[-1, 1]$. However, $P'(1)=-\alpha_1>0$ and $P'(t)\to-\infty$ as $t\to\infty$. Hence $P'(t)$ has a zero on $(1,\infty)$, which is a contradiction, since $P'(t)$ is a polynomial of degree n.

(iv) If $\alpha_0=0$ and $\alpha_1>0$ then $P(t)/(1-t)$ is $-\alpha_1<0$ at $t=-1$ and tends to $+\infty$ as $t\to-\infty$ and therefore has one zero on $(-\infty,-1)$. Hence also $P(t)$ has one zero on $(-\infty,-1)$ and consequently at most n zeros on $[-1, 1]$.

(v) If $\alpha_0=0$ and $\alpha_1<0$ then $P(t)/(1-t)$ is $\alpha_1<0$ at $t=1$ and tends to $+\infty$ as $t\to+\infty$. Hence $P(t)/(1-t)$ and therefore also $P(t)$ has one zero on $(1,\infty)$. Thus $P(t)$ has at most n zeros on $[-1, 1]$.

(vi) Finally if $\alpha_0=\alpha_1=0$, $P(t)$ has a double zero at $t=1$. Hence $P(t)$ can have at most n distinct zeros on $[-1, 1]$.

This completes the proof of the proposition.

Our counterexample to the possibility of the elimination of the exceptional case in Krein's theorem is the following.

PROPOSITION 2. *Let $T=\{t_i\}_{i=1}^{k}$ be any subset of $[-1, 1]$ such that*

$$-1<t_1<t_2<\cdots<t_k=1.$$

Then there exists no polynomial $u(t)=\sum_{i=0}^{n}\beta_i p_i(t)$ $(n=2m;\ m=1,2,\cdots)$ which is nonnegative and which vanishes only on T in $[-1, 1]$.

Proof. Assume on the contrary that there exists a polynomial $u(t)=\sum_{i=0}^{n}\beta_i p_i(t)\geqq 0$ vanishing precisely on T in $[-1, 1]$. Since $t_k=1$, we have $u(1)=0$. Since $p_i(1)=0$, $i=1,2,\cdots,n$, and $p_0(1)=1\neq 0$, we have $\beta_0=0$. Thus

$$u(t)=\sum_{i=1}^{n}\beta_i p_i(t).$$

Since $p_i(-1)=0$, $i=2,3,\cdots,n$, $p_1(-1)=-2\neq 0$ and $u(-1)\neq 0$ as $-1\notin T$, it follows that $\beta_1\neq 0$.

Now

$$\begin{aligned} u(t)&=(1-t)\left\{\beta_1 t+(t^2-1)\sum_{i=2}^{n}\beta_i t^{i-2}\right\}\\ &=(1-t)Q(t), \end{aligned}$$

say. Then $Q(-1)=-\beta_1$ and $Q(1)=\beta_1$. Hence since β_1 is nonzero, $Q(t)$ changes sign on $(-1, 1)$. But $(1-t)$ is positive on $(-1, 1)$. Therefore $(1-t)Q(t)=u(t)$ changes sign on $(-1, 1)$. This contradicts the assumed nonnegativity of $u(t)$ on $[-1, 1]$ and completes the proof.

In view of the above proposition and the theorem of Krein we obtain the following three additional properties of the T system $\{p_i(t)\}_{i=0}^n$ (n even):

(i) $\{p_i(t)\}_{i=0}^n$ is not a T-system on any interval $[a, b]$ properly containing the interval $[-1, 1]$.

(ii) $\{p_i(t)\}_{i=0}^n$ is not an ET-system of order 2 on $[-1, 1]$.

(iii) The subspace spanned by $\{p_i(t)\}_{i=0}^n$ contains no even dimensional subspace generated by a T-system.

The property (iii), in particular, entails the result of Zielke [3] that the subspace spanned by $\{p_i(t)\}_{i=0}^n$ has no Markov basis.

REFERENCES

[1] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, John Wiley and Sons, New York, 1966.

[2] M. G. Krein, *The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals and their further development*, Amer. Math. Soc. Transl., Ser. 2, 12 (1959), pp. 1–122.

[3] R. Zielke, *Tchebyshev systems that can not be transformed into Markov systems*, Manuscripta Math., 17 (1975), pp. 67–71.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0001 $01.00/0

ASYMPTOTIC EXPANSION OF MULTIPLE FOURIER TRANSFORMS*

P. N. SHIVAKUMAR† AND R. WONG†

Abstract. Asymptotic expansion of multi-dimensional Fourier transforms is derived. An explicit expression for the remainder term is also given, from which an error bound can readily be obtained.

1. Introduction. Let $f(x)$ be a complex-valued locally integrable function on R^n, and let $T_\alpha[f]$ denote the Fourier transform of f, i.e.,

$$T_\alpha[f]=\frac{1}{(2\pi)^{n/2}}\int_{R^n} f(x)\, e^{i\alpha\cdot x}\, dx, \tag{1.1}$$

where $x=(x_1,\cdots,x_n)$, $\alpha=(\alpha_1,\cdots,\alpha_n)$ and $\alpha\cdot x=\alpha_1x_1+\cdots+\alpha_nx_n$. In (1.1) we only assume that the Cauchy limit

$$\lim_{L\to\infty}\int_{|x|<L} f(x)\, e^{i\alpha\cdot x}\, dx$$

exists for every $\alpha\in R^n$, where $|x|=\sqrt{x_1^2+\cdots+x_n^2}$. Thus the function f need not belong to $L^1(R^n)$.

The problem of finding the asymptotic behavior of $T_\alpha[f]$, when $n=3$ and as $|\alpha|\to\infty$, was first considered by Duffin [4], and subsequently treated in more detail, when $n=2$, by Duffin and Shaffer [5]. In both papers [4] and [5], the function $f(x)$ is assumed to be expressible in the form

$$f(x)=\sum_{p,q} c_{pq}x^pr^q+\varphi(x). \tag{1.2}$$

Here $\varphi\in C^\infty(R^n)$, q is a real number, $p=(p_1,\cdots,p_n)$ is a multi-index of nonnegative integers, $x^p=x_1^{p_1}\cdots x_n^{p_n}$, $r=\sqrt{x_1^2+\cdots+x_n^2}$, and the sum is finite.

In Duffin's analysis use was made of so-called *mollifier functions*. Due to a degree of arbitrariness of these functions, his derivation will not lead to the construction of error bounds for the asymptotic approximations of [4] and [5]. The main purpose of this paper is to provide an explicit expression for the remainder term in the asymptotic expansion of $T_\alpha[f]$ from which error bounds can readily be obtained. Indeed, we show that if f has continuous partial derivatives of order up to and including $2m$ in R^n (except possibly at the origin) and satisfies (1.2), then under certain additional assumptions

$$\begin{aligned}T_\alpha[f]&=\sum{}' c_{pq}L(q)\,\partial_\alpha^p(|\alpha|^{-q-n})\\&\quad+\sum{}'' c_{pq}L^*(q)\,\partial_\alpha^p(|\alpha|^{2l}\log|\alpha|)+\delta_m(\alpha),\end{aligned} \tag{1.3}$$

where $L(q)$ and $L^*(q)$ are explicitly given constants and

$$\delta_m(\alpha)=\frac{(-1)^m}{|\alpha|^{2m}}T_\alpha[\Delta^m\varphi]. \tag{1.4}$$

In (1.3), $\sum'$ excludes those q's for which $q+n$ is a negative even integer, while $\sum''$

* Received by the editors October 11, 1977, and in final revised form May 11, 1978. This research was supported in part by the National Research Council of Canada under Grants NRC-A-8783 and NRC-A-7359.

† Department of Mathematics and Astronomy, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.

includes only those q's for which $q+n=-2l$, a negative even integer. In (1.4), Δ denotes the Laplacian.

The secondary purpose of this paper is to present a true extension of the asymptotic expansion of the one-dimensional Fourier transform to the multi-dimensional case. In several directions, our results are more general than the corresponding ones given in [4] and [5]. The approach presented here is based on a concept recently used by Olver in obtaining error bounds for stationary phase approximation [8].

Results like those of our paper, which deal with the asymptotic expansions of multi-dimensional Fourier transforms, have been given by Jones [6, Thm. 9.7, p. 328]. However, Jones considers Fourier transforms of generalized functions and gives no error estimate for the approximation. For other results concerning asymptotic evaluation of multiple integrals of Fourier type, which involve only one large parameter, see Jones and Kline [7], Chako [3] and Bleistein and Handelsman [1].

2. Assumptions. We shall use rather standard notation. For a multi-index $p=(p_1,\cdots,p_n)$, we let $D^p=D_1^{p_1}\cdots D_n^{p_n}$ be a differentiation of order $|p|=p_1+\cdots+p_n$, where $D_j=\partial/\partial x_j$. For any open set $\Omega\subset R^n$ and for each nonnegative m, the set $C^m(\Omega)$ consists of all complex functions f in Ω whose derivatives D^pf exists for each multi-index p with $|p|\leqq m$, and are continuous functions in Ω. The Laplacian is denoted by $\Delta=\partial^2/\partial x_1^2+\cdots+\partial^2/\partial x_n^2$. In polar coordinates $x=(r,\theta_1,\cdots,\theta_{n-1})=(r,\theta)$,

$$\Delta=\frac{\partial^2}{\partial r^2}+\frac{n-1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}L(\theta,D_\theta),\tag{2.1}$$

where $L(\theta,D_\theta)$ is a second-order linear differential operator with the independent variables $\theta_1,\cdots,\theta_{n-1}$. To simplify some of the formalism, we also use the symbol

$$\partial_x^p=(i)^{-|p|}D^p=\left(\frac{1}{i}\frac{\partial}{\partial x_1}\right)^{p_1}\cdots\left(\frac{1}{i}\frac{\partial}{\partial x_n}\right)^{p_n}.\tag{2.2}$$

The following assumptions are similar to those adopted in [8] for the one-dimensional case.

(A_1) *$f\in C^{2m}(R^n/\{0\})$, m being a nonnegative integer.*

(A_2) *$f(x)$ is expressible in the form* (1.2) *with $q+|p|>-n$ for all multi-index p and all real numbers q under the summation sign. Let $Q=\max\{q+|p|\}$ with the maximum also taken over all p and q in* (1.2). *We require*

$$2m-1\leqq Q+n<2m+1.\tag{2.3}$$

As $r\to 0^+$, we assume

$$(\Delta^j\varphi)(x)=O(r^{Q-2j+1}),\qquad j=0,1,\cdots,m,\tag{2.4}$$

if $Q+n\neq 2m-1$, and

$$(\Delta^j\varphi)(x)=O(r^{Q-2j+2}),\qquad j=0,1,\cdots,m,\tag{2.5}$$

if $Q+n=2m-1$.

(A_3) *For some $\rho>0$, the integrals*

$$\int_{|x|\geqq\rho}(\Delta^jf)(x)\,e^{i\alpha\cdot x}\,dx,\qquad j=0,1,\cdots,m,$$

converge uniformly for all sufficiently large $|\alpha|$.

Remarks. (i) if $f\in C^\infty(R^n/\{0\})$ then there always exists a nonnegative integer m satisfying (2.3). If f is an odd number of times continuously differentiable, a result similar to Theorem 1 (below) can be obtained. The statement of the result is, however, more complicated, and we shall not give it explicitly.

(ii) In (A_2), if $Q+n=2m-1$ and if $\Delta^j\varphi$ satisfies (2.4) instead of (2.5) then we must stop the series at the term preceding the last and let φ denote the new remainder term. For $j=0, 1, \cdots, m-1$, $\varphi(x)$ satisfies

$$(\Delta^j\varphi)(x)=O(r^{Q-2j}).$$

This condition will lead to an asymptotic expansion of $T_\alpha[f]$ with one less term than that to be obtained under the condition (2.5).

(iii) The assumptions (A_1), (A_2) and (A_3) ensure that the original Fourier transform (1.1) converges uniformly for all sufficiently large values of $\alpha \in R^n$. Moreover, they imply that the Fourier transform

$$T_\alpha[\Delta^m\varphi]=\frac{1}{(2\pi)^{n/2}}\int_{R^n}(\Delta^m\varphi)(x)\,e^{i\alpha\cdot x}\,dx \tag{2.6}$$

also exists uniformly for all sufficiently large values of $\alpha \in R^n$. We shall prove this assertion only for the case $Q+n\neq 2m-1$ in (2.3). The case when $Q+n=2m-1$ can be handled similarly. First we note that by (2.4)

$$(\Delta^m\varphi)(x)=O(r^{Q-2m+1})\quad \text{as } r\to 0^+.$$

Since $Q+n>2m-1$, near $x=0$ the improper integral in (2.6) exists absolutely and uniformly for all values of α. Next we observe that from (1.2) and (2.1), it follows that

$$(\Delta^m\varphi)(x)=(\Delta^m f)(x)-\sum \psi_{pq}(\theta)r^{|p|+q-2m},$$

where the ψ_{pq}'s are linear combinations of products of the sine and cosine functions of $\theta=(\theta_1, \cdots, \theta_{n-1})$. Since $Q+n<2m+1$, the powers of r in the last sum are all less than $1-n$. Hence, by condition (A_3) with $j=m$, the integral in (2.6) also exists at infinity, uniformly for all large values of $\alpha \in R^n$. This establishes our assertion.

(iv) If $f(x)$ is a function of r only, say $f(x)=g(r)$, then it is well known that $T_\alpha[f]$ depends only on $|\alpha|$. More specifically, we have

$$T_\alpha[f]=|\alpha|^{(2-n)/2}\int_0^\infty r^{n/2}g(r)J_{(n-2)/2}(|\alpha|r)\,dr, \tag{2.7}$$

where $J_\nu(r)$ denotes the Bessel function of the first kind; see [2, Thm. 40]. The integral on the right side of (2.7) is known as the Hankel transform of $g(r)$. Asymptotic expansions of Hankel transforms are available, complete with error bounds; see [10].

3. Abel limits. For each $\varepsilon>0$ the Abel mean $A_\varepsilon(f)$ is defined to be the integral

$$A_\varepsilon(f)=\int_{R^n} f(x)\,e^{-\varepsilon|x|}\,dx. \tag{3.1}$$

It is clear that if $f\in L^1(R^n)$ then $\lim_{\varepsilon\to 0}A_\varepsilon(f)=\int_{R^n}f(x)\,dx$. However, these Abel means can be well-defined even when $f\notin L^1(R^n)$. In this section, we wish to explicitly evaluate the Abel limit.

$$\lim_{\varepsilon\to 0} T_\alpha[e^{-\varepsilon r}x^p r^q], \qquad |p|+q+n>0. \tag{3.2}$$

LEMMA 1. *For $|p|+q+n>0$ but $q+n\neq 0, -2, -4, \cdots,$*

$$\lim_{\varepsilon\to 0} T_\alpha[e^{-\varepsilon r}x^p r^q]=L(q)\,\partial_\alpha^p(|\alpha|^{-q-n}), \tag{3.3}$$

where

$$L(q)=2^{q+n/2}\Gamma\left(\frac{q+n}{2}\right)\Big/\Gamma\left(-\frac{q}{2}\right). \tag{3.4}$$

Proof. If $|p|=0$ and $q+n>0$ then it is known that

$$\lim T_\alpha[e^{-\varepsilon r}r^q]=L(q)\cdot|\alpha|^{-q-n}; \tag{3.5}$$

see [6, p. 230]. Examining the derivation of this result in detail reveals that the desired identity (3.3) can be obtained by simply differentiating (3.5) with respect to α.

If $q+n<0$ then we must proceed in a different manner. First we make the change of variables $x=r\xi$ so that

$$\begin{aligned} T_\alpha[e^{-\varepsilon r}x^p r^q]&=\frac{1}{(2\pi)^{n/2}}\int_{|\xi|=1}\xi^p\int_0^\infty r^{|p|+q+n-1}e^{-(\varepsilon-i\beta)r}\,dr\,dS_\xi \\ &=\frac{\Gamma(|p|+q+n)}{(2\pi)^{n/2}}\int_{|\xi|=1}\xi^p(\varepsilon-i\beta)^{-q-n-|p|}\,dS_\xi \end{aligned} \tag{3.6}$$

where $\beta=\sum_{k=1}^n\alpha_k\xi_k$ and dS_ξ is the surface element on the unit sphere $|\xi|=1$. Next we observe that as long as $q+n$ is not a negative odd integer, (3.6) can be rewritten as

$$T_\alpha[e^{-\varepsilon r}x^p r^q]=\frac{\Gamma(|p|+q+n)}{(2\pi)^{n/2}(q+n)_{|p|}}\partial_\alpha^p\int_{|\xi|=1}(\varepsilon-i\beta)^{-q-n}\,dS_\xi, \tag{3.7}$$

where $(\lambda)_s=\lambda(\lambda+1)\cdots(\lambda+s-1)$. To evaluate the last integral, we subject it to an orthogonal transformation

$$\xi_k'=\sum_{j=1}^n b_{kj}\xi_j \tag{3.8}$$

with determinant $+1$ and $b_{1j}=\alpha_j/|\alpha|$, $j=1,\cdots,n$; see [2, p. 70]. Then we have

$$\sum_{k=1}^n\xi_k'^2=\sum_{k=1}^n\xi_k^2=1,\qquad \beta=\sum_{k=1}^n\alpha_k\xi_k=|\alpha|\xi_1', \tag{3.9}$$

and the right-hand side of (3.7) becomes

$$\frac{\Gamma(|p|+q+n)}{(2\pi)^{n/2}(q+n)_{|p|}}\partial_\alpha^p\int_{|\xi'|=1}(\varepsilon-i|\alpha|\xi_1')^{-q-n}\,dS_{\xi'}.$$

Taking the limit in $\varepsilon\to0$ yields

$$\begin{aligned} \lim_{\varepsilon\to0}T_\alpha[e^{-\varepsilon r}x^p r^q]&=\frac{\Gamma(|p|+q+n)}{(2\pi)^{n/2}(q+n)_{|p|}}\partial_\alpha^p(|\alpha|^{-q-n}) \\ &\quad\cdot(-i)^{-q-n}\int_{|\xi'|=1}(\xi_1')^{-q-n}\,dS_{\xi'}. \end{aligned} \tag{3.10}$$

The integral in (3.10) can be evaluated by using polar coordinates $\xi'=(1,\theta)=(1,\theta_1,\cdots,\theta_{n-1})$ with

$$\xi_1'=\sin\theta_{n-1}\sin\theta_{n-2}\cdots\sin\theta_2\sin\theta_1$$

and

$$dS_{\xi'}=\sin^{n-2}\theta_{n-1}\sin^{n-3}\theta_{n-2}\cdots\sin\theta_2\,d\theta_1\cdots d\theta_{n-1}.$$

If ξ' varies on the unit sphere, then $\theta_2, \cdots, \theta_{n-1}$ vary each in the interval $(0, \pi)$, whereas θ_1 varies in the interval $(0, 2\pi)$. Thus

$$\int_{|\xi'|=1} (\xi_1')^{-q-n}\, dS_{\xi'} = [1+(-1)^{-q-n}] \prod_{j=0}^{n-2} \int_0^{\pi} \sin^{-q-n+j}\theta\, d\theta.$$

Furthermore, it follows from the identity

$$\int_0^{\pi} \sin^t\theta\, d\theta = \frac{\sqrt{\pi}\Gamma[(t+1)/2]}{\Gamma[1+(t/2)]} \qquad (t \geqq 0)$$

that

$$(3.11) \quad (-i)^{-q-n} \int_{|\xi'|=1} (\xi_1')^{-q-n}\, dS_{\xi'} = \pi^{n/2} 2^{q+n} \Gamma\left(\frac{q+n}{2}\right) \Big/ \left(\Gamma(q+n)\Gamma\left(-\frac{q}{2}\right)\right).$$

Upon inserting (3.11) into (3.10), we obtain from (3.4)

$$(3.12) \qquad \lim_{\varepsilon \to 0} T_\alpha[e^{-\varepsilon r} x^p r^q] = L(q)\, \partial_\alpha^p(|\alpha|^{-q-n}),$$

provided that $q+n$ is not a negative odd integer.

To show that (3.12) also holds when $q+n$ is a negative odd integer, we put $q+n=-m$ and observe that

$$(3.13) \qquad \partial_\alpha^p\left\{(\varepsilon - i\beta)^m\left[\log(\varepsilon - i\beta) - \sum_{k=1}^m \frac{1}{k}\right]\right\} = \frac{\Gamma(|p|-m)m!\xi^p}{(\varepsilon - i\beta)^{|p|-m}}.$$

In place of (3.7), we now rewrite (3.6) in the form

$$T_\alpha[e^{-\varepsilon r} x^p r^q] = \frac{(-1)^{m+1}}{(2\pi)^{n/2} m!} \partial_\alpha^p \int_{|\xi|=1} (\varepsilon - i\beta)^m\left[\log(\varepsilon - i\beta) - \sum_{k=1}^m \frac{1}{k}\right] dS_\xi.$$

The orthogonal transformation (3.8)–(3.9) then gives

$$(3.14) \qquad \lim_{\varepsilon \to 0} T_\alpha[e^{-\varepsilon r} x^p r^q] = -\frac{i^m}{(2\pi)^{n/2} m!} \partial_\alpha^p(|\alpha|^m) \int_{|\xi'|=1} \xi_1'^m \log \xi_1'\, dS_{\xi'}.$$

(Note that $\int_{|\xi'|=1} \xi_1'^m\, dS_{\xi'} = 0$ if m is an odd integer.) The integral on the right-hand side of (3.14) can be evaluated by using polar coordinates, as in (3.11). The result is

$$(3.15) \qquad \int_{|\xi'|=1} \xi_1'^m \log \xi_1'\, dS_{\xi'} = i^{2m+1} \pi^{(n+1)/2} \frac{\Gamma[(m+1)/2]}{\Gamma[(m+n)/2]}.$$

Inserting (3.15) into (3.14), we again obtain (3.12) with $q+n=-m$ and m being an odd integer. This completes the proof of Lemma 1.

LEMMA 2. *For* $|p|+q+n>0$ *and* $q+n=-2l$, $l=0, 1, 2, \cdots$,

$$(3.16) \qquad \lim_{\varepsilon \to 0} T_\alpha[e^{-\varepsilon r} x^p r^q] = L^*(q)\, \partial_\alpha^p(|\alpha|^{2l} \log |\alpha|),$$

where $L^*(q) = -1/\Gamma\left(\frac{n}{2}\right)$ *if* $l=0$ *and*

$$L^*(q) = (-1)^{l+1}\left\{2^l l!\Gamma\left(\frac{n}{2}\right)[n(n+2)\cdots(n+2l-2)]\right\}^{-1}$$

if $l = 1, 2, \cdots$.

Proof. Since the argument here is similar to that for Lemma 1, we only give a sketch of the proof. In place of (3.13), we use the identity

$$\partial_\alpha^p[(\varepsilon-i\beta)^{2l}\log(\varepsilon-i\beta)]=-\frac{\Gamma(|p|-2l)(2l)!\xi^p}{(\varepsilon-i\beta)^{|p|-2l}}. \tag{3.17}$$

Equation (3.14) then becomes

$$\lim_{\varepsilon\to 0} T_\alpha[e^{-\varepsilon r}x^p r^q]=\frac{(-1)^{l+1}}{(2\pi)^{n/2}(2l)!}\,\partial_\alpha^p(|\alpha|^{2l}\log|\alpha|)\int_{|\xi'|=1}\xi_1'^{2l}\,dS_{\xi'}. \tag{3.18}$$

(Note that $\partial_\alpha^p(|\alpha|^{2l})=0$.) The last integral can be evaluated as in (3.11), and we obtain (3.16) as desired.

We close this section with one further result concerning the Abel limits, which we shall need later in our discussion.

LEMMA 3. *If f is locally integrable on R^n and*

$$\lim_{\rho\to 0}\int_{|x|<\rho} f(x)\,dx=L \tag{3.19}$$

exists then the Abel mean $A_\varepsilon(f)$ converges to L, i.e.,

$$\lim_{\varepsilon\to 0}\int_{R^n} f(x)\,e^{-\varepsilon|x|}\,dx=L. \tag{3.20}$$

Proof. This result is well-known when $n=1$; see [9, p. 26]. The general case follows from this special one, when the integrals in (3.19) and (3.20) are expressed as iterated integrals in polar coordinates.

4. Main theorem. With the preliminary lemmas established in the preceding section, we are now ready to state and prove our main result concerning the asymptotic expansion of $T_\alpha[f]$.

THEOREM 1. *Assume that conditions* (A_1), (A_2) *and* (A_3) *hold. Then*

$$T_\alpha[f]=\sum{}' c_{pq}L(q)\,\partial_\alpha^p(|\alpha|^{-q-n})+\sum{}'' c_{pq}L^*(q)\,\partial_\alpha^p(|\alpha|^{2l}\log|\alpha|)+\delta_m(\alpha), \tag{4.1}$$

where, in $\sum'$, we exclude those q's for which $q+n$ is a negative even integer, while, in $\sum''$, we include only those q's for which $q+n=-2l$, a negative even integer. The error term is given by

$$\delta_m(\alpha)=\frac{(-1)^m}{|\alpha|^{2m}}T_\alpha[\Delta^m\varphi]. \tag{4.2}$$

Proof. For any $\varepsilon>0$, we have from (1.2)

$$T_\alpha[e^{-\varepsilon r}f]=\sum c_{pq}T_\alpha[e^{-\varepsilon r}x^p r^q]+T_\alpha[e^{-\varepsilon r}\varphi]. \tag{4.3}$$

Since $T_\alpha[f]$ exists, by letting $\varepsilon\to 0$ in (4.3) and applying Lemmas 1, 2 and 3 we obtain

$$T_\alpha[f]=\sum{}' c_{pq}L(q)\,\partial_\alpha^p(|\alpha|^{-q-n})+\sum{}'' c_{pq}L^*(q)\,\partial_\alpha^p(|\alpha|^{2l}\log|\alpha|)+\delta_m(\alpha), \tag{4.4}$$

where

$$\delta_m(\alpha)=\lim_{\varepsilon\to 0} T_\alpha[e^{-\varepsilon r}\,\varphi]. \tag{4.5}$$

We must show that

$$\lim_{\varepsilon\to 0} T_\alpha[e^{-\varepsilon r}\,\varphi]=\frac{(-1)^m}{|\alpha|^{2m}}T_\alpha[\Delta^m\varphi]. \tag{4.6}$$

Put $u(x)=e^{-\varepsilon r+i\alpha\cdot x}$ and recall that as $r\to 0^+$, $(\Delta^j\varphi)(x)=O(r^{Q-2j+1})$ if $Q+n>2m-1\geqq 1$ and $(\Delta^j\varphi)(x)=O(r^{Q-2j+2})$ if $Q+n=2m-1\geqq 1$. By Green's theorem

$$\int_{|x|<\rho}(u\Delta\varphi-\varphi\Delta u)\,dx=\int_{|x|=\rho}\left(u\frac{\partial\varphi}{\partial r}-\varphi\frac{\partial u}{\partial r}\right)d\sigma \tag{4.7}$$

where $d\sigma$ is the element of surface area on $|x|=\rho$. A simple calculation shows

$$\Delta u=\left[\varepsilon^2-\frac{2i\varepsilon}{r}\alpha\cdot x-|\alpha|^2-\frac{(n-1)}{r}\varepsilon\right]u.$$

We first substitute this into (4.7), next observe that

$$\int_{|x|<\rho}e^{i\alpha\cdot x}[\Delta\varphi+|\alpha|^2\varphi]\,dx=\int_{|x|=\rho}e^{i\alpha\cdot x}\frac{\partial\varphi}{\partial r}\,d\sigma-\int_{|x|=\rho}\varphi\frac{\partial}{\partial r}(e^{i\alpha\cdot x})\,d\sigma$$

again by Green's theorem, and finally let $\rho\to\infty$. The result is

$$\begin{aligned}T_\alpha[e^{-\varepsilon r}(\Delta\varphi+|\alpha|^2\varphi)]&=\varepsilon^2T_\alpha[e^{-\varepsilon r}\varphi]-2i\varepsilon T_\alpha\left[\left(\frac{\alpha\cdot x}{r}\right)e^{-\varepsilon r}\varphi\right]\\&\quad-(n-1)\varepsilon T_\alpha\left[\frac{1}{r}e^{-\varepsilon r}\varphi\right].\end{aligned} \tag{4.8}$$

By using Lemmas 1, 2 and 3, it can be shown that the limits, as $\varepsilon\to 0$, of $T_\alpha[e^{-\varepsilon r}\varphi]$, $T_\alpha[((\alpha\cdot\chi)/r)\,e^{-\varepsilon r}\varphi]$ and $T_\alpha[(1/r)\,e^{-\varepsilon r}\varphi]$ all exist. Hence the right-hand side of (4.8) tends to zero as $\varepsilon\to 0$, and

$$\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\varphi]=\frac{(-1)}{|\alpha|^2}\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\,\Delta\varphi]. \tag{4.9}$$

This procedure can be repeated m times and finally leads to

$$\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\,\varphi]=\frac{(-1)^m}{|\alpha|^{2m}}\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\,\Delta^m\varphi]. \tag{4.10}$$

Since $T_\alpha[\Delta^m\varphi]$ exists by Remark (iii) in § 2, we have

$$\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\varphi]=\frac{(-1)^m}{|\alpha|^{2m}}T_\alpha[\Delta^m\varphi]$$

by Lemma 3. This demonstrates (4.6), and completes the proof of Theorem 1.

5. Asymptotic properties and error bounds. If $g\in L^1(R^n)$ then by the Riemann–Lebesgue lemma

$$T_\alpha[g]\to 0\quad\text{as }|\alpha|\to\infty; \tag{5.1}$$

see, e.g., [2, p. 57]. However, it can easily be shown that (5.1) holds as long as $T_\alpha[g]$ exists uniformly for all sufficiently large values of $\alpha\in R^n$. Taking $g=\Delta^m\varphi$, we have by Remark (iii) of § 2

$$\delta_m(\alpha)=o(|\alpha|^{-2m})\quad\text{as }|\alpha|\to\infty. \tag{5.2}$$

This confirms the asymptotic nature of (4.1).

To obtain a bound for the error term $\delta_m(\alpha)$, we replace assumption (A_3) by

(A_3') *For some* $\rho>0$ *and for each* $j=0,1,\cdots,m-1$, *the integral*

$$\int_{|x|\geqq\rho}(\Delta^jf)(x)\,e^{i\alpha\cdot x}\,dx$$

converges uniformly for all large $|\alpha|$. *Furthermore,* $\int_{|x|\geqq\rho}|(\Delta^m f)(x)|\,dx$ *is finite*;

and require, instead of (2.3),

$$2m-1\leqq Q+n<2m. \tag{5.3}$$

Under these conditions, $\Delta^m\varphi$ is absolutely integrable and

$$|\delta_m(\alpha)|\leqq\frac{1}{(2\pi)^{n/2}|\alpha|^{2m}}\int_{R^n}|(\Delta^m\varphi)(x)|\,dx. \tag{5.4}$$

We now give a generalization of the asymptotic expansion of the one-dimensional Fourier transform; see [8, Theorem 2]. We write

$$f(x)\sim\sum c_{pq}x^p r^q \quad \text{as } r\to 0+ \tag{5.5}$$

to mean that if we truncate the series after a finite number of terms and if we let Q be the maximum of $|p|+q$ taken over all p and q under the finite sum, then the remainder $\varphi(x)$ satisfies either $\varphi(x)=O(r^{Q+2})$ or $\varphi(x)=O(r^{Q+1})$, as $r\to 0$, depending on whether or not $Q+n$ is an odd integer. (In (5.5), we of course assume that $q+|p|>-n$ for all p and all q). The following result is an immediate consequence of Theorem 1.

THEOREM 2. *If* (i) *f is a C^∞ function on R^n except possibly at the origin*; (ii) *f possesses an asymptotic expansion of the form* (5.5); (iii) *the asymptotic expansion of any derivative of f* is obtained by differentiating (5.5); *and* (iv) *each of the integrals* $\int(\Delta^j f)(x)\,e^{i\alpha\cdot x}\,dx$, $j=0,1,\cdots$, *converges at $x=\infty$ uniformly for all large values of α; then as $|\alpha|\to\infty$, the asymptotic expansion of $T_\alpha[f]$ is obtained by substituting* (5.5) *in* (1.1) *and integrating formally term by term in the sense of* (3.3) *and* (3.15).

Example 1. Consider the double Fourier transform

$$I(\alpha_1,\alpha_2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{i(\alpha_1x+\alpha_2y)}}{r^{3/2}[1+(x+y)^2]}\,dx\,dy \tag{5.6}$$

where $r=\sqrt{x^2+y^2}$. In order to apply the results in § 4, we write

$$\frac{1}{r^{3/2}[1+(x+y)^2]}=r^{-3/2}+\varphi(x,y).$$

Note that $\varphi\notin C^1(R^2)$. Since the function on the left is absolutely integrable and $T_\alpha[r^{-3/2}]$ exists uniformly for all large values of $|\alpha|$, so does $T_\alpha[\varphi]$. In the notation of § 2, we have

$$Q=-\tfrac{3}{2},\qquad n=2,\qquad m=0.$$

From (4.1) it follows that

$$I(\alpha_1,\alpha_2)=\frac{\Gamma^2(\frac14)}{2\pi|\alpha|^{1/2}}+\delta(\alpha_1,\alpha_2), \tag{5.7}$$

where $\delta(\alpha_1,\alpha_2)=T_\alpha[\varphi]$. Since $\varphi\notin L^1(R^2)$, the error estimate (5.4) does not hold for $m=0$. However, by Lemma 3, we have

$$\delta(\alpha_1,\alpha_2)=\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\varphi]=-\frac{1}{|\alpha|^2}\lim_{\varepsilon\to 0}T_\alpha[e^{-\varepsilon r}\,\Delta\varphi]. \tag{5.8}$$

The last equality follows from (4.9). A simple estimation gives

$$|\Delta\varphi|\leqq\frac{52}{r^{3/2}[1+(x+y)^2]},$$

which implies that $\Delta\varphi \in L^1(R^2)$ and $\lim_{\varepsilon\to 0} T_\alpha[e^{-\varepsilon r}\,\Delta\varphi] = T_\alpha[\Delta\varphi]$. Therefore

$$|\delta(\alpha_1, \alpha_2)| \leqq \frac{52}{|\alpha|^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{r^{3/2}[1+(x+y)^2]}\,dx\,dy,$$

in view of (5.8). The integral on the right can be explicitly evaluated. The final result is

$$(5.9)\qquad |\delta(\alpha_1, \alpha_2)| \leqq \frac{52\sqrt{\pi}\,\Gamma^2(\frac{1}{4})}{2^{1/4}|\alpha|^2}.$$

It should be pointed out that $T_\alpha[\Delta^2\varphi]$ does not exist and hence the technique employed in (5.8) can not be repeated to obtain a bound of a lower asymptotic order of magnitude.

Example 2. Consider the triple Fourier integral

$$(5.10)\qquad I(\alpha_1, \alpha_2, \alpha_3) = \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{x\,y\,z}{(1+r)^6}\,e^{i(\alpha_1 x+\alpha_2 y+\alpha_3 z)}\,dx\,dy\,dz,$$

where $r = \sqrt{x^2+y^2+z^2}$. Note that the function $xyz/(1+r)^6$ is not absolutely integrable on R^3. Put

$$\frac{x\,y\,z}{(1+r)^6} = xyz(1-6r+21\,r^2)+\varphi(x, y, z).$$

Then, in the notation of § 2, we have

$$Q = 5, \qquad n = 3, \qquad m = 4.$$

From (4.1) it follows that

$$(5.11)\qquad I(\alpha_1, \alpha_2, \alpha_3) = 12\sqrt{\frac{2}{\pi}}\,\partial_\alpha^p(|\alpha|^{-4})+\delta(\alpha_1, \alpha_2, \alpha_3),$$

where the multi-index p is given by $p = (1, 1, 1)$ and the remainder satisfies

$$(5.12)\qquad \delta(\alpha_1, \alpha_2, \alpha_3) = \frac{1}{|\alpha|^8}T_\alpha[\Delta^4\varphi].$$

To estimate $T_\alpha[\Delta^4\varphi]$, we write $\varphi(x, y, z) = xyzF(r)$ and observe that

$$\Delta^4\varphi(x, y, z) = xyz\left(\frac{d^2}{dr^2}+\frac{8}{r}\frac{d}{dr}\right)^4 F(r).$$

An elementary estimation gives

$$|\Delta^4\varphi| \leqq \frac{C}{r^2(1+r)^2}$$

with $C = 304283520$. Thus by (5.4)

$$(5.13)\qquad |\delta(\alpha_1, \alpha_2, \alpha_3)| \leqq \sqrt{\frac{\pi}{2}}\cdot\frac{C}{|\alpha|^8}.$$

REFERENCES

[1] N. BLEISTEIN AND R. A. HANDELSMAN, *Multidimensional stationary phase an alternative derivation*, this Journal, 6 (1975), pp. 480–487.

[2] S. BOCHNER AND K. CHANDRASEKHARAN, *Fourier Transforms*, Ann. of Math. Studies, No. 19, Princeton University Press, Princeton, NJ, 1949.

[3] N. CHAKO, *Asymptotic expansions of double and multiple integrals arising in diffraction theory*, J. Inst. Math. Appl., 1 (1965), pp. 372–422.

[4] R. J. DUFFIN, *Discrete potential theory*, Duke Math. J., 20 (1953), pp. 233–251.

[5] R. J. DUFFIN AND D. H. SHAFFER, *Asymptotic expansion of double Fourier transforms*, Ibid., 27 (1960), pp. 581–596.

[6] D. S. JONES, *Generalized Functions*, McGraw-Hill, London, 1966.

[7] D. S. JONES AND M. KLINE, *Asymptotic expansion of multiple integrals and the method of stationary phase*, J. Mathematical Phys., 37 (1958), pp. 1-28.

[8] F. W. J. OLVER, *Error bounds for stationary phase approximations*, this Journal, 5 (1974), pp. 19-29.

[9] E. C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals*, Clarendon Press, Oxford, 1962.

[10] R. WONG, *Error bounds for asymptotic expansions of Hankel transforms*, this Journal, 7 (1976), pp. 799–808.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0002 $01.00/0

HOMEOMORPHISMS OF COMPACT, CONVEX SETS AND THE JACOBIAN MATRIX*

ANDREU MAS-COLELL†

Abstract. Let $K \subset R^n$ be a compact, convex polyhedron and f: $K \to R^n$ a C^1 function. The problem of existence of a global inverse for f is studied. It is shown (Theorem 1) that f has an inverse, if, for every $x \in K$, the Jacobian of f at x, $Jf(x)$, is such that for every linear space spanned by a face of K containing x the determinant of the linear map from L to L formed by projecting $Jf(x)$ on L has positive sign. Theorem 2 is a similar result for K with smooth boundary. The theorems generalize the well-known Gale–Nikaido theorems, which originated in some problems of mathematical economics.

1. Introduction. Let $K \subset R^n$ be a compact, convex set. Without loss of generality we assume that K has a nonempty interior. Let F: $K \to R^n$ be a C^1 function. The derivative map of F at x is denoted $DF(x)$. Given a coordinate system the Jacobian matrix of F at $x \in K$ is denoted $JF(x)$. We want to find sets of local conditions, i.e., conditions on $JF(x)$ only, implying that F is one to one and so, a homeomorphism.

It is well known that the nonsingularity everywhere of $JF(x)$ will not do; see Fig. 1.

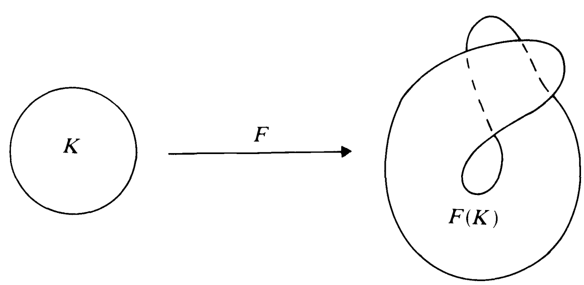

FIG. 1

A set of sufficient conditions is provided by the theorems of Gale and Nikaido ([2], [6, Chap. VII]), which were stimulated by some problems in mathematical economics:

(i) Let K be a rectangle. If for every $x \in K$, $JF(x)$ is a P matrix (i.e., every principal minor of $J(x)$ has positive sign), then F is a homeomorphism.

(ii) If for every $x \in K$, $JF(x)$ is positive quasidefinite (i.e., $v'JF(x)v > 0$ for all $x \in R^n$, $v \neq 0$), then F is a homeomorphism.

It will be shown here that the result can be obtained under substantially weaker hypotheses. In particular, for points $x \in \operatorname{Int} K$ our analogue of (i) will impose sign restrictions only on the principal minor of order n.

More specifically, consider (i) above. The set K is a rectangle, i.e., it is of the form $K = \{x \in R^n : s^i \leqq x^i \leqq r^i\}$. For every nonempty subspace $L \subset R^n$ let Π_L: $R^n \to L$ denote the perpendicular projection map. The condition that $JF(x)$ be a P matrix is equivalent to the requirement that for every coordinate subspace $L \subset R^n$, the linear map $\Pi_L \cdot DF(x)$: $L \to L$ preserves orientation, i.e., has a positive determinant. We will show that, with K a general polyhedron, F is a homeomorphism if for every $x \in K$ and every subspace $L \subset R^n$ spanned by a face of K which includes x, the linear map $\Pi_L \cdot DF(x)$: $L \to L$ preserves orientation, i.e., has positive determinant (the subspace

* Received by the editors March 13, 1978, and in final revised form September 18, 1978. This paper was written in April 1977 while the author was visiting Bonn University, made possible by the financial support of the Sonderforschungsbereich 21.

† Department of Economics and Mathematics, University of California, Berkeley, California 94720.

spanned by a convex set is the translation to the origin of the minimal affine space containing the set). So, if K is a rectangle, $JF(x)$ needs to be a P matrix only at the vertices of K and for $x \in \operatorname{Int} K$ the only requirement is that $JF(x)$ have a positive determinant.

Observe also that, in contrast with (i), our conditions are coordinate free, in the sense that their formulation does not rely on a previous choosing of coordinates. This will be emphasized in the statement of the theorem. Consider now (ii) and suppose that the boundary of K, denoted ∂K, is smooth (a C^1 hypersurface, say). For $x \in \partial K$, T_x is the tangent plane of ∂K at x (see Fig. 2). We will show that F is a homeomorphism if: (a) $JF(x)$ has a positive determinant for every $x \in K$, and (b) for every $x \in \partial K$, $JF(x)$ is positive quasidefinite on T_x, i.e., $v'JF(x)v > 0$ whenever $v \neq 0$ and $v \in T_x$.

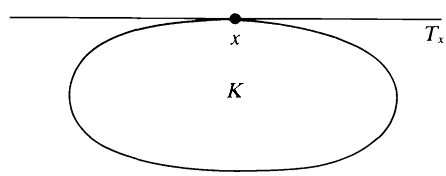

FIG. 2

The mathematical tool for the proofs is fixed-point index theory (see Milnor [5], Guillemin–Pollack [4]), in particular, the powerful Poincaré–Hopf theorem. That index theory could be the key to the sort of generalization of the Gale–Nikaido theorem given here was surmised by H. Scarf [8] in view of the Eaves and Scarf analysis in [1] of the index theory associated with the linear (and nonlinear) complementarity problem (see also Saigal and Simon [7]).

It is worth emphasizing that our results are not of a purely differential topological nature; they hold for domains K which are *convex* sets. It should be clear from the inspection of Fig. 3 how counterexamples can be constructed for nonconvex K and maps F satisfying (a) and (b) of the paragraph previous to the last.

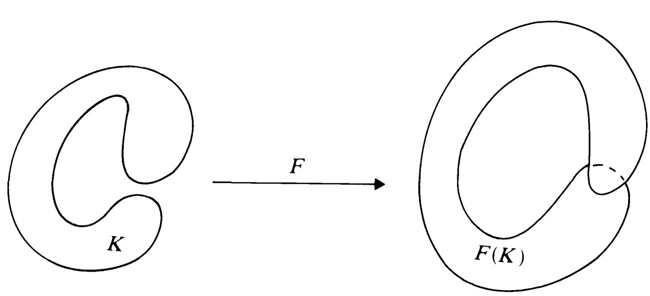

FIG. 3

2. Statement of theorems. Terminology and notation are as in the Introduction.

THEOREM 1. *Let $K \subset R^n$ be a compact, convex polyhedron of full dimension and $F: K \to R^n$ a C^1 function. If for every $x \in K$ and subspace $L \subset R^n$ spanned by a face of K which includes x, the map $\Pi_L \cdot DF(x)$: $L \to L$ has a positive determinant, then F is one-to-one and so, a homeomorphism.*

THEOREM 2. *Let $K \subset R^n$ be a compact, convex set of full dimension with a C^1 boundary ∂K and $F: K \to R^n$ a C^1 function. If for every $x \in K$, $DF(x)$ has a positive determinant and if for all $x \in \partial K$, $DF(x)$ is positive quasidefinite on T_x (i.e., $v'DF(x)v > 0$ for $v \in T_x$, $v \neq 0$), then F is one-to-one and so, a homeomorphism.*

3. Demonstration.

1. It may be useful if we first sketch the main idea of the proof, which is very simple. We first extend F to the whole of R^n in a certain simple manner which preserves differentiability except in a set of measure zero and has the property that whenever differentiable the extended function has a positive Jacobian determinant. Now take any point of R^n, say, $0 \in R^n$. It turns out that for our purposes we can assume that $F^{-1}(0)$ lies entirely in the region of differentiability. This means that the sum of the indexes of F at points $x \in F^{-1}(0)$ equals the sum of the signs of the Jacobian determinant, i.e., the sum is $\geqq 1$. But, after verifying that the extended F satisfies appropriate boundary conditions, we appeal to a topological index theorem to conclude that the sum must be $\leqq 1$. Hence $F^{-1}(0)$ is a singleton set.

2. We let $K \subset R^n$ be a general compact, convex polyhedron of full dimension and prove Theorem 1. We shall see at the end that Theorem 2 is essentially a corollary of Theorem 1.

We note first that it suffices to prove that $F|\operatorname{Int} K$ is one-to-one. Indeed, we can always extend F to a K' containing K in its interior and sufficiently similar to K for all hypotheses on $DF(x)$ to be still satisfied.

3. For every $x \in R^n$ let $z(x) \in K$ be the foot of x, i.e., $z(x)$ is the (unique) element of K minimizing $\|x-z\|$ for $z \in K$. Of course, $z(x)=x$ for $x \in K$; see Fig. 4.

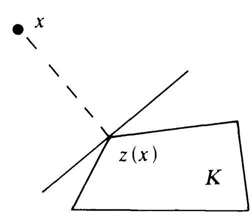

FIG. 4

We now extend $F: K \to R^n$ to the whole of R^n by letting a function $\hat{F}: R^n \to R^n$ be defined by $\hat{F}(x)=F(z(x))+x-z(x)$; see Fig. 5. For any $y \in F(K)$ define $\hat{F}_y: R^n \to R^n$ by $\hat{F}_y(x)=\hat{F}(x)-y$.

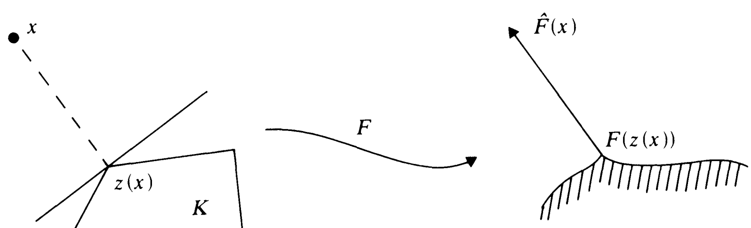

FIG. 5

4. Let S_r, B_r be, respectively, the sphere and ball of radius r. We claim that for any $y \in F(K)$ and any r sufficiently large, $\hat{F}_y|S_r$ has degree one, i.e., it is homotopic, with respect to $R^n \backslash \{0\}$, to the identity in S_r. Indeed, it suffices to verify that for r sufficiently large and any $y \in F(K)$, if $x \in S_r$, then $x'\hat{F}_y(x)>0$. Take $r> \max_{z \in K, y \in F(K)} \|F(z)-z-y\|=s$. Then

$$\begin{aligned} x'\hat{F}_y(x) &= \|x\|^2 - x'(z(x)+y-F(z(x))) \\ &\geqq \|x\|^2 - \|x\| \, \|z(x)+y-F(z(x))\| \geqq r^2 - rs > 0. \end{aligned}$$

5. The region $A=\{x\in R^n: \hat{F}$ is not C^1 at $x\}$ contains no open set. This is clear since $z(x)$ is C^1 everywhere except at $x\in K$ with $x-z(x)$ perpendicular at $z(x)$ to more than one face of K and those x are contained in a finite number of hyperplanes. Since $\hat{F}$ is Lipschitzian, $\hat{F}(A)$ contains no open set.

6. We now state the basic lemma. The proof is postponed to 8.

LEMMA 1. *Let K be a polyhedron and F satisfy the hypothesis of Theorem 1. Then if $x\notin A$, $|D\hat{F}(x)|$ is positive.*

Of course, $|D\hat{F}(x)|$ denotes the determinant of the linear map $D\hat{F}(x)$.

7. Let $r>0$ be a fixed number with $K\subset B_r$ and $\hat{F}_y|S_r$ of degree one for any $y\in F(K)$. By the Poincaré–Hopf theorem (see Milnor [5]), if $\hat{F}_y^{-1}(0)\cap A=\varnothing$, then $\sum_{x\in\hat{F}_y^{-1}(0)\cap B_r}\operatorname{sign}|D\hat{F}_y(x)|=1$, which, by the lemma, means that $\hat{F}_y^{-1}(0)\cap B_r$ is a singleton set.

Suppose now that $F|\operatorname{Int}K$ were not one-to-one, i.e., there are $x_1, x_2\in\operatorname{Int}K$ with $x_1\neq x_2$ and $F(x_1)=F(x_2)$. By the implicit function theorem there are disjoint open sets $V_1, V_2\subset\operatorname{Int}K$ with $x_1\in V_1$, $x_2\in V_2$ and $F(V_1)\cap F(V_2)\neq\varnothing$ open. Since $\hat{F}(A)$ contains no open set, there is $y\in F(V_1)\cap F(V_2)$ such that $y\notin\hat{F}(A)$. But then $\hat{F}_y^{-1}(0)\cap A=\varnothing$ and $F^{-1}(y)\subset\hat{F}_y^{-1}(0)\cap B_r$ is not a singleton set. This contradiction establishes that $F|\operatorname{Int}K$ must be one-to-one and concludes the proof of Theorem 1.

8. We now prove Lemma 1.

Let $x\notin A$. Then $x-z(x)$ is perpendicular to a single face of K, which, of course, includes $z(x)$. Let L be the subspace spanned by this face and $L^\perp$ the subspace orthogonal to L. We then have that for small $v\in L$, $z(x+v)=z(x)+v$ and so, $\hat{F}(x+v)=F(z(x)+v)+x-z(x)$; hence, $D\hat{F}(x)v=DF(z(x))v$. For $v\in L^\perp$, $z(x+v)=z(x)$ and so, $\hat{F}(x+v)=F(z(x))+x+v-z(x)$; hence $D\hat{F}(x)v=v$. Therefore, if we choose an orthogonal coordinate system whose k first coordinates generate L, $J\hat{F}(x)$, the matrix of $D\hat{F}(x)$ with respect to this coordinate system, takes the form

$$J\hat{F}(x)=\begin{bmatrix} J_kF(z(x)) & \begin{matrix}0\\ I\end{matrix}\end{bmatrix},\qquad \text{where } J_kF(z(x)) \text{ are}$$

the first k columns of $JF(x)$. So $|D\hat{F}(x)|=|J\hat{F}(x)|=|J_{kk}F(z(x))|$, where $J_{kk}F(z(x))$ are the first k rows of $J_kF(z(x))$. But $J_{kk}F(z(x))$ is the matrix of $\Pi_L\cdot DF(z(x))\colon L\to L$ which by hypothesis is positive.

9. We now prove Theorem 2.

LEMMA 2. *Under the hypothesis of Theorem 2, if $x\in\partial K$ and $L\subset T_x$ is a subspace, then $\Pi_L\cdot DF(x)\colon L\to L$ has a positive determinant.*

If the lemma holds, the proof is concluded since we can approximate K by a polyhedron K' and if K' is sufficiently close to K, Lemma 2 implies that the hypotheses of Theorem 1 are satisfied.

Lemma 2 is a well-known fact. Choose an orthogonal coordinate system such that the first k coordinates generate L and the nth is perpendicular to T_x and let $JF(x)$ be the matrix of $DF(x)$ in this coordinate system. Then $J_{n-1,n-1}F(x)$, the matrix formed by deleting the nth row and column is positive quasidefinite. This is the assumption of the theorem. But any principle minor of a positive quasidefinite matrix is positive (see, for example, Nikaido [6, p. 374]); this applies to $J_{kk}F(x)$, the matrix of $\Pi_L\cdot DF(x)\colon L\to L$, and yields the lemma. Q.E.D.

Acknowledgment. This paper was written in April 1977 while I was visiting the Universität Bonn for the academic year. The stay was made possible by the financial support of the Sonderforschungsbereich 21 which is gratefully acknowledged. The problem treated in the paper was initially brought to my attention by H. Scarf. Thanks are also due to D. Gale. R. Saigal and two referees saved me from a serious mishap. Working independently from me and from each other the solution to the problem has been arrived at by at least two other sets of researchers: C. Garcia and W. Zangwill [3] on the one hand and G. Chichilnisky, M. Hirsch, and H. Scarf on the other. The proofs are, in every case, different.

REFERENCES

[1] C. EAVES AND H. SCARF, *The solution of systems of piecewise linear equations*, Math. Operations Res., 1, 1 (1976) pp. 1–27.

[2] D. GALE AND H. NIKAIDO, *The Jacobian matrix and global univalence of mappings*, Math. Ann., 159, 2 (1965) pp. 81–93.

[3] C. GARCIA AND W. ZANGWILL, *On univalence and P matrices*, Report 7737, Center for Mathematical Studies in Business and Economics, University of Chicago, 1977; Linear Algebra and Appl., to appear.

[4] V. GUILLEMIN AND A. POLLACK, *Differential Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[5] J. MILNOR, *Topology from the Differential Viewpoint*, The University Press of Virginia, Charlottesville, VA, 1965.

[6] H. NIKAIDO, *Convex Structures and Economic Theory*, Academic Press, New York, 1968.

[7] R. SAIGAL AND C. SIMON, *Generic properties of the complementarity problem*, Math. Prog., 4 (1973) pp. 324–325.

[8] H. SCARF, Private communication, 1975.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0003 $01.00/0

APPLICATIONS OF A CONVOLUTION THEOREM TO JACOBI POLYNOMIALS*

JOHN L. LEWIS†

Abstract. In this paper we use a convolution theorem to study certain polynomials in the unit disk whose coefficients are constant multiples of Jacobi polynomials. We show these polynomials have no zeros in the unit disk. This result extends a theorem of Ruscheweyh on Gegenbauer polynomials, and is related to results of Askey and Gasper on Jacobi polynomials.

1. Introduction. Let $\Delta=\{z:|z|<1\}$, and $T=\{z:|z|=1\}$. Given σ, $-\infty<\sigma\leqq 1$, let $S(\sigma)$ be the class of normalized starlike functions of order σ in Δ. An analytic function f in Δ is in $S(\sigma)$ if and only if $f(0)=0$, $f'(0)=1$, and

$$\operatorname{Re}[zf'(z)/f(z)]\geqq\sigma,\qquad z\in\Delta. \tag{1.1}$$

Using the Poisson integral formula and integrating (1.1), it follows that $f\in S(\sigma)$ if and only if

$$\log(f(z)/z)=-2(1-\sigma)\int_T\log(1-e^{-i\theta}z)\,d\mu(e^{i\theta}),\qquad z\in\Delta, \tag{1.2}$$

for some probability measure μ on T. Observe from (1.2) that

$$|f(z)/z|\geqq 2^{-2(1-\sigma)},\qquad z\in\Delta. \tag{1.3}$$

The function

$$z/(1-z)^{2(1-\sigma)}=\sum_{n=1}^{\infty}[(2-2\sigma)_{n-1}/(n-1)!]z^n, \tag{1.4}$$

where $(a)_0=1$, $(a)_n=a\cdot(a+1)\cdots(a+n-1)$, $n\geqq 1$, is in $S(\sigma)$, as follows easily from (1.1). Let $C(\sigma)$ denote the class of analytic functions g in Δ satisfying $g(0)=0$, $g'(0)=1$, and the condition that

$$\operatorname{Re}[zg'(z)/h(e^{i\theta}z)]\geqq 0,\qquad z\in\Delta, \tag{1.5}$$

for some $h\in S(\sigma)$ and θ real.

For fixed σ, $-\infty<\sigma\leqq 1$, and f, g, analytic in Δ with $f(0)=g(0)=0$, define the convolution of f and g [denoted $(f*g)_\sigma$] as follows: if $f(z)=\sum_{n=1}^{\infty}a_nz^n$, $g(z)=\sum_{n=1}^{\infty}b_nz^n$, then

$$(f*g)_\sigma(z)=\sum_{n=1}^{\infty}[(n-1)!a_nb_n/(2-2\sigma)_{n-1}]z^n,$$

when $z\in\Delta$. In [11], Ruscheweyh, and in [8], the author, independently obtained the following generalization of a theorem of Suffridge [14].

THEOREM A. *Let $f\in S(\sigma)$ and $g\in C(\sigma)$. Then $(f*g)_\sigma\in C(\sigma)$.*

Let $P_n^{(\alpha,\beta)}$ denote the Jacobi polynomials defined by

$$P_n^{(\alpha,\beta)}(x)=[(1+\alpha)_n/n!]F[-n,n+\alpha+\beta+1,1+\alpha,(1-x)/2].$$

Here F is the hypergeometric function. Clearly, $P_n^{(\alpha,\beta)}(1)=(1+\alpha)_n/n!$. In this paper we use Theorem A to prove

* Received by the editors March 29, 1978, and in revised form August 14, 1978.

† Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506.

THEOREM 1. *Let* $0 \leqq \lambda \leqq \alpha+\beta$ *and* $\alpha \geqq \beta > -\infty$. *Then*

$$\sum_{k=0}^{n} \frac{(1+\lambda)_{n-k}}{(n-k)!} \frac{(1+\lambda)_k}{k!} \frac{P_k^{(\alpha,\beta)}(x)}{P_k^{(\alpha,\beta)}(1)} z^k \neq 0, \qquad n=1,2,\cdots, \tag{1.6}$$

for $-1 \leqq x \leqq 1$ *and* $z \in \Delta$.

If we let $z \to -1$ in (1.6), it follows that

$$\sum_{k=0}^{n} (-1)^k \frac{(1+\lambda)_{n-k}}{(n-k)!} \frac{(1+\lambda)_k}{k!} \frac{P_k^{(\alpha,\beta)}(x)}{P_k^{(\alpha,\beta)}(1)} \geqq 0, \qquad n=1,2,\cdots, \tag{1.7}$$

when $-1 \leqq x \leqq 1$. We note that Askey and Gasper in [2], and Gasper in [5], showed that (1.7) holds whenever either (a) $0 \leqq \lambda \leqq \alpha+\beta$, $\alpha \geqq -\frac{1}{2}$, or (b) $-1 < \lambda \leqq \alpha+\beta$, $|\beta| \leqq \alpha$. Also Bustoz and Savage [3] proved that (1.7) is valid for $1 < \lambda < 2$ and $\alpha = \beta = \frac{1}{2}$. Askey [1] verified (1.7) for $\lambda = 2$, $\alpha = \beta = \frac{1}{2}$, and pointed out that this inequality is false when $\lambda > 2$, $\alpha = \beta = \frac{1}{2}$. Thus it appears likely that the restrictions on λ in Theorem 1 can be relaxed. However, the assumption $\alpha \geqq \beta$ in Theorem 1 cannot be replaced by $\alpha < \beta$. Indeed, if $n = 1$, $x = -1$, then the polynomials in (1.6) have a zero at

$$z = -P_1^{(\alpha,\beta)}(1)/[P_1^{(\alpha,\beta)}(-1)] = (1+\alpha)/(1+\beta),$$

and $|z| < 1$ when $|\alpha| < \beta$.

Next we note that Ruscheweyh [12] proved Theorem 1 when $\alpha = \beta \geqq 0$ and $\lambda = 0$. We sketch his proof. In this case the function in (1.6) can be written in the form: $z^{-1}(f * zg')_{(1/2-\alpha)}(z)$, where

$$f(z) = z(1+z^2-2xz)^{-(\alpha+1/2)}, \qquad z \in \Delta, \tag{1.8}$$

$g(0) = 0$, and

$$g'(z) = (1-z^{n+1})/(1-z), \qquad z \in \Delta.$$

It is easily checked that $f \in S(\frac{1}{2}-\alpha)$ and $g \in C(\frac{1}{2}-\alpha)$. Hence by Theorem A,

$$z(f*g)'_{(1/2-\alpha)}(z) = (f * zg')_{(1/2-\alpha)}(z) = \varphi(e^{i\theta}z)P(z), \qquad z \in \Delta,$$

for some $\varphi \in S(\frac{1}{2}-\alpha)$, θ real, and P analytic in Δ with $\operatorname{Re} P > 0$. From (1.3) we conclude that (1.6) holds when $\alpha = \beta \geqq 0$ and $\lambda = 0$.

Our proof of Theorem 1 is similar. However, the functions involved in the convolution are more complicated when $\alpha > \beta$ and $0 < \lambda \leqq \alpha+\beta$. In § 2 we study, for fixed α, β, as in Theorem 1, the following generating function for Jacobi polynomials (see [10, § 132]):

$$\begin{aligned} G(z,x) &= z(1-z)^{-(1+\alpha+\beta)} F[\tfrac{1}{2}(1+\alpha+\beta), \tfrac{1}{2}(2+\alpha+\beta), 1+\alpha, 2(x-1)z/(1-z)^2] \\ &= \sum_{n=0}^{\infty} \frac{(1+\alpha+\beta)_n}{n!} \frac{P_n^{(\alpha,\beta)}(x)}{P_n^{(\alpha,\beta)}(1)} z^{n+1}. \end{aligned} \tag{1.9}$$

Observe that $G(\cdot, x) = f$ in (1.8) when $\alpha = \beta$. Using Lemma 1 we show in Lemma 2 that $G(\cdot, x)$ is in $S(\frac{1}{2}(1-\alpha-\beta))$ when $|\beta| \leqq \alpha$ and $-1 \leqq x \leqq 1$. In § 3 we consider the polynomials $q_n(\cdot, \lambda)$, $n = 1, 2, \cdots$, defined by

$$q_n(z,\lambda) = \sum_{k=0}^{n} \frac{(1+\lambda)_{n-k}}{(n-k)!} \frac{(1+\lambda)_k}{k!} z^k. \tag{1.10}$$

In Lemma 3 we prove that $n!q_n(\cdot, \lambda)/(1+\lambda)_n$, is the derivative of a function in $C(\frac{1}{2}(1-\alpha-\beta))$ when $0 \leqq \lambda \leqq \alpha+\beta$ and $n = 1, 2, \cdots$. Since (1.6) can be written in the form:

$$z^{-1}(G(\cdot, x) * zq_n(\cdot, \lambda))_{(1-\alpha-\beta)/2}(z) \neq 0, \tag{1.11}$$

when $-1 \leqq x \leqq 1$ and $z \in \Delta$, it then follows from Theorem A, as in the previous special case, that Theorem 1 is true.

2. Several lemmas. We first prove

LEMMA 1. *Let α and β be as in Theorem 1. Then the function*

$$z \to zF(1+\alpha+\beta, 1+\beta, 1+\alpha, z), \qquad z \in \Delta,$$

is in $S(\frac{1}{2}(1-\alpha-\beta))$.

Proof. Lemma 1 is trivial if $\alpha = \beta$, since the above function is then of the form (1.4). Hence we assume $\beta < \alpha$. Let $a = 1+\alpha+\beta$, $b = 1+\beta$, $c = 1+\alpha$. We write F for $F(a, b, c, z)$. Recall that F satisfies in Δ the differential equation:

$$z(1-z)F'' + [c-(a+b+1)z]F' - abF = 0. \tag{2.1}$$

Let ω be the meromorphic function in Δ defined by

$$zF'/F = a\omega/(1-\omega). \tag{2.2}$$

To prove Lemma 1 it clearly suffices to show $\operatorname{Re}[zF'/F] \geqq -a/2$, in Δ. Since the function, $z \to az(1-z)^{-1}$, maps Δ onto $\{\zeta\colon \operatorname{Re}\zeta > -a/2\}$, this inequality is equivalent to $|\omega| < 1$. To show $|\omega| < 1$, we employ a useful technique (see [6], [9], [13]). First note that $|\omega| < 1$ in $\{z\colon |z| < s\}$, provided $s > 0$ is small enough, since $\omega(0) = 0$. Second, let ρ be the supremum of the numbers $s > 0$ for which $|\omega| < 1$ in $\{z\colon |z| < s\}$. If $\rho < 1$, there exists z_0 with $|z_0| = \rho$ for which $|\omega(z_0)| = 1$. From the maximum modulus theorem, we have $|z|^{-1}|\omega(z)| \leqq \rho^{-1}$ in $\{z\colon |z| \leqq \rho\}$. Hence if $z = re^{i\theta}$,

$$\frac{\partial}{\partial\theta}\log(|z|^{-1}|\omega(z)|)|_{z=z_0} = -\operatorname{Im}(z\omega'/\omega)(z_0) = 0,$$

$$r\frac{\partial}{\partial r}\log(|z|^{-1}|\omega(z)|)|_{z=z_0} = \operatorname{Re}(z\omega'/\omega)(z_0) - 1 \geqq 0.$$

We conclude that

$$z_0\omega'(z_0) = t\omega(z_0), \qquad t \geqq 1. \tag{2.3}$$

Multiplying (2.2) by F and differentiating, we get

$$z(zF')' = zF'a\omega(1-\omega)^{-1} + aFz\omega'(1-\omega)^{-2}, \tag{2.4}$$

for $|z| < \rho$. From (2.1),

$$\begin{aligned} z(zF')' &= zF' + z^2F'' \\ &= -[c-(a+b+1)z]zF'/(1-z) + abzF/(1-z) + zF', \end{aligned} \tag{2.5}$$

when $|z| < \rho$. Equating (2.4) to (2.5) and combining terms in zF' and F, we obtain

$$\begin{aligned} &[abz/(1-z) - az\omega'/(1-\omega)^2]F \\ &\quad = zF'[((a+1)\omega - 1)/(1-\omega) + (c-(a+b+1)z)/(1-z)]. \end{aligned}$$

Multiplying this equality by $(1-z)(1-\omega)^2/F$ and using (2.2), it follows that

$$[abz(1-\omega)^2-az\omega'(1-z)]=a\omega[(1-z)((a+1)\omega-1)+(c-(a+b+1)z)(1-\omega)].$$

We let $z \to z_0$ in the above equality, and use (2.3) to replace $z_0\omega'$ by $t\omega$. Rearranging terms in the resulting equality, we get

$$Az_0=B,$$

where

$$\begin{aligned} A &= ab(1-\omega)^2+at\omega+a\omega((a+1)\omega-1)+a\omega(a+b+1)(1-\omega) \\ &= a[b+(a-b+t)\omega], \end{aligned}$$

and

$$B=a\omega[t+(a+1)\omega-1+c(1-\omega)]=a\omega[t+c-1+(a+1-c)\omega].$$

We now substitute the expressions for a, b, and c in terms of α and β, in the above equalities. Since $|\omega|=1$, $t\geqq 1$, $\alpha>\beta$, and $\alpha+\beta\geqq 0$, it follows that

$$|A|=|B|=(1+\alpha+\beta)|\alpha+t+\omega(1+\beta)|>0.$$

Hence, $|z_0|=\rho=1$. We have reached a contradiction to our assumption that $\rho<1$. This concludes the proof of Lemma 1.

Next we observe for α, β, as in Theorem 1, that the function

$$w \to F(\tfrac{1}{2}(1+\alpha+\beta), \tfrac{1}{2}(2+\alpha+\beta), 1+\alpha, w)$$

is analytic in $\{w\colon w \notin [1,\infty)\}$. Also, this function and its derivative have continuous one-sided limits as $w \to r>1$ from either above or below $[1,\infty)$ (see [16, chap. 14]). Since $z \to 2(x-1)z/(1-z)^2$ maps Δ univalently onto $\{w\colon w \notin [(1-x)/2,\infty)\}$, it follows that $G(\cdot, x)$ in (1.9) is analytic in Δ for $-1\leqq x\leqq 1$. If $x=\cos\varphi$, $0\leqq\varphi\leqq\pi$, $z=e^{i\theta}$, and $z\neq 1$, then

$$2(x-1)z/(1-z)^2=\sin^2(\varphi/2)/\sin^2(\theta/2). \tag{2.6}$$

Hence, $2(x-1)z/(1-z)^2=1$ in $\Delta\cup T$ only for $z=e^{\pm i\varphi}$. We also note for later use that

$$2(x-1)(\operatorname{Im} z)\operatorname{Im}[z/(1-z)^2]\leqq 0, \tag{2.7}$$

when $z\in\Delta$ and $-1\leqq x\leqq 1$. From (2.6) and the first observation, we see for $-1\leqq x\leqq 1$ that $G(\cdot, x)$ in (1.9) and its derivative extend continuously to $\Delta\cup T-\{1, e^{\pm i\varphi}\}$, when $x=\cos\varphi$, $0\leqq\varphi\leqq\pi$.

We use these facts to prove

LEMMA 2. *Let α and β be as in Theorem* 1. *Then $G(\cdot, x)$ is in $S(\frac{1}{2}(1-\alpha-\beta))$ for $-1\leqq x\leqq 1$.*

Proof. Lemma 2 is trivial if $x=1$, since

$$G(z,1)=z(1-z)^{-(1+\alpha+\beta)}, \qquad z\in\Delta.$$

Also if $x=-1$, then from (1.9) and the fact that

$$P_n^{(\alpha,\beta)}(-1)=(-1)^n(1+\beta)_n/n!, \qquad n=0,1,\cdots,$$

we find

$$G(z,-1)=zF(1+\alpha+\beta, 1+\beta, 1+\alpha, -z), \qquad z\in\Delta.$$

From Lemma 1 we conclude that $G(\cdot,-1)$ is in $S(\frac{1}{2}(1-\alpha-\beta))$.

For fixed x, $-1<x<1$, define σ in Δ by

$$2(x-1)z/(1-z)^2=-4\sigma(z)/(1-\sigma(z))^2, \qquad z\in\Delta. \tag{2.8}$$

From the remarks preceding Lemma 2, we see that σ is analytic in Δ, extends continuously to $\bar{\Delta}$, and $|\sigma|\leqq 1$. If $z\in\Delta$ and $|\sigma(z)|\leqq\frac{1}{2}$, then from (2.8)

$$|\sigma(z)/z|=\tfrac{1}{2}|(x-1)(1-\sigma(z))^2/(1-z)^2|\geqq|x-1|/32,$$

while if $|\sigma|\geqq\frac{1}{2}$, then

$$|\sigma(z)/z|\geqq\tfrac{1}{2}\geqq|x-1|/32.$$

Using these inequalities, (2.8), and (1.3), we get

$$\begin{aligned}|z^{-1}G(z,x)|&=|[(1-\sigma(z))/(1-z)]^{(1+\alpha+\beta)}|\,|G(\sigma(z),-1)/\sigma(z)|\\&=[2|x-1|^{-1}|\sigma(z)|/|z|]^{(1+\alpha+\beta)/2}|G(\sigma(z),-1)|/|\sigma(z)|\\&\geqq 16^{-(1+\alpha+\beta)/2}2^{-(1+\alpha+\beta)}\\&=8^{-(1+\alpha+\beta)}.\end{aligned}$$

If $x=\cos\varphi$, $0<\varphi<\pi$, then from this inequality and the remark preceding Lemma 2, we deduce that $zG'(\cdot,x)/G(\cdot,x)$ is analytic in Δ and extends continuously to $\Delta\cup T-\{e^{\pm i\varphi},1\}$.

If $\zeta=-4z/(1-z)^2$, let

$$H(z)=\frac{(dF/d\zeta)[\frac{1}{2}(1+\alpha+\beta),\frac{1}{2}(2+\alpha+\beta),1+\alpha,-4z/(1-z)^2]}{F[\frac{1}{2}(1+\alpha+\beta),\frac{1}{2}(2+\alpha+\beta),1+\alpha,-4z/(1-z)^2]},$$

when $z\in\Delta$. Taking the logarithmic derivative of $G(\cdot,-1)$, we get

$$\begin{aligned}\operatorname{Re}[zG'(z,-1)/G(z,-1)]&=\operatorname{Re}[1+(1+\alpha+\beta)z/(1-z)]\\&\quad-4\operatorname{Re}[z(1+z)(1-z)^{-3}H(z)],\end{aligned}$$

for $z\in\Delta\cup T-\{\pm1\}$. If $z\in T-\{\pm1\}$, note that

$$\begin{aligned}&\operatorname{Re}[1+(1+\alpha+\beta)z/(1-z)]=(1-\alpha-\beta)/2,\\&\operatorname{Re}[z(1+z)/(1-z)^3]=0,\\&(\operatorname{Im} z)\operatorname{Im}[z(1+z)/(1-z)^3]<0.\end{aligned}\tag{2.9}$$

These inequalities and the fact that $G(\cdot,-1)$ is in $S(\frac{1}{2}(1-\alpha-\beta))$ imply

$$(\operatorname{Im} z)(\operatorname{Im} H(z))\leqq 0, \qquad z\in T-\{\pm1\}. \tag{2.10}$$

Again, if $x=\cos\varphi$, $0<\varphi<\pi$, is fixed, then

$$\begin{aligned}&\operatorname{Re}[zG'(z,x)/G(z,x)]\\&\quad=\operatorname{Re}[1+(1+\alpha+\beta)z/(1-z)]+2(x-1)\operatorname{Re}[z(1+z)(1-z)^{-3}H(\sigma(z))],\end{aligned}\tag{2.11}$$

where σ is as in (2.8). Let $\tau=\{e^{i\theta}: -\varphi<\theta<\varphi\}$. If $z\in\tau-\{1\}$, then from (2.6) we see that $2(x-1)z/(1-z)^2>1$, and thereupon from (2.8) that $\sigma(z)\in T-\{\pm1\}$. Also from (2.7), we deduce

$$(\operatorname{Im} z)(\operatorname{Im}\sigma(z))>0, \qquad z\in\tau-\{1\}.$$

Using (2.11), (2.9), (2.10), and the above inequality, we obtain

$$\begin{aligned}
&\operatorname{Re}[zG'(z,x)/G(z,x)] \\
&\quad=\tfrac{1}{2}(1-\alpha-\beta)+2(1-x)\operatorname{Im}[z(1+z)(1-z)^{-3}]\operatorname{Im}H(\sigma(z)) \qquad (2.12)\\
&\quad\geqq\tfrac{1}{2}(1-\alpha-\beta),
\end{aligned}$$

when $z\in\tau-\{1\}$. Suppose now that $z\in T-\tau$ and $z\neq e^{\pm i\varphi}$. Then $0<2(x-1)z/(1-z)^2<1$, and so, $-1<\sigma(z)<1$. Hence, $\operatorname{Im}H(\sigma(z))=0$, and the second function on the right hand side of (2.11) is zero at z. Thus (2.12) also holds for $z\in T-\tau$ and $z\neq e^{\pm i\varphi}$.

It remains to examine the behavior of $zG'(\cdot,x)/G(\cdot,x)$, near $e^{\pm i\varphi}$ and 1. To do this let $a=\frac{1}{2}(1+a+\beta)$, $b=\frac{1}{2}(2+\alpha+\beta)$, $c=1+\alpha$, and suppose that

$$a+b-c=\tfrac{1}{2}+\beta\neq\text{integer}. \qquad (2.13)$$

Under this additional assumption on β we have

$$\begin{aligned}
F(a,b,c,w)&=A_1F(a,b,a+b-c+1,1-w) \qquad (2.14)\\
&\quad+A_2(1-w)^{c-(a+b)}F(c-a,c-b,c-a-b+1,1-w),
\end{aligned}$$

when $|\arg(1-w)|<\pi$, and

$$\begin{aligned}
F(a,b,c,w)&=B_1(-w)^{-a}F(a,1-c+a,1-b+a,w^{-1}) \qquad (2.15)\\
&\quad+B_2(-w)^{-b}F(b,1-c+b,1-a+b,w^{-1}),
\end{aligned}$$

when $|\arg(-w)|<\pi$. Here arg denotes the principal argument and

$$\begin{aligned}
A_1&=\Gamma(c)\Gamma(c-a-b)/[\Gamma(c-a)\Gamma(c-b)],\\
A_2&=\Gamma(c)\Gamma(a+b-c)/[\Gamma(a)\Gamma(b)], \qquad (2.16)\\
B_1&=\Gamma(c)\Gamma(b-a)/[\Gamma(b)\Gamma(c-a)],\\
B_2&=\Gamma(c)\Gamma(a-b)/[\Gamma(a)\Gamma(c-b)],
\end{aligned}$$

(see [4, § 2.10] for these formulas).

If $c-(a+b)>0$, then from (2.14)

$$F'(a,b,c,w)=-A_1ab/(a+b-c+1)+A_2(a+b-c)(1-w)^{c-a-b-1}+o(1),$$

as $w\to1$, in such a way that $|\arg(1-w)|<\pi$. Moreover,

$$F(a,b,c,w)^{-1}=[A_1+A_2(1-w)^{c-(a+b)}]^{-1}+O(|1-w|),$$

as $w\to1$, with $|\arg(1-w)|<\pi$. Therefore,

$$\begin{aligned}
F'(a,b,c,w)/F(a,b,c,w)&=-ab/(a+b-c+1) \qquad (2.17)\\
&\quad+A_2(a+b-c)(1-w)^{c-a-b-1}[A_1+A_2(1-w)^{c-(a+b)}]^{-1}+o(1),
\end{aligned}$$

as $w\to1$, $|\arg(1-w)|<\pi$. Put $w=2(x-1)z/(1-z)^2$, where $x=\cos\varphi$, $0<\varphi<\pi$. Observe from (2.6) that $w\to1$ as $z\to e^{i\varphi}$. If $c-a-b>1$, then from (2.17), we clearly have

$$\operatorname{Im}[F'(a,b,c,w)/F(a,b,c,w)]=\operatorname{Im}H(\sigma(z))\to0, \qquad (2.18)$$

as $z \to e^{i\varphi}$. If $0 < c-a-b < 1$, observe from (2.16) that $A_2/A_1 < 0$. Let N be the largest positive integer such that $N(c-a-b) < 1$. Then from (2.17) we have

$$\operatorname{Im} H(\sigma(z)) = (c-a-b)\operatorname{Im}\left\{(1-w)^{-1}\sum_{k=1}^{N}[(|A_2|/|A_1|)(1-w)^{c-a-b}]^k\right\} + o(1), \tag{2.19}$$

as $w \to 1$, $|\arg(1-w)| < \pi$. Note that the function

$$\zeta \to (1-\zeta)^{-\gamma}, \qquad 1^{-\gamma} = 1,$$

maps $\{\zeta : \operatorname{Im}\zeta \leqq 0\}$ into itself for $0 < \gamma < 1$. Since by (2.7),

$$(\operatorname{Im} z)(\operatorname{Im} w) \leqq 0, \qquad z \in \Delta,$$

it follows from this observation and (2.19) that

$$\limsup_{z\to e^{i\varphi}} \operatorname{Im} H(\sigma(z)) \leqq 0, \tag{2.20}$$

when $0 < c-a-b < 1$.

We now observe from (2.9) that

$$\operatorname{Re}[z(1+z)(1-z)^{-3}] = O(|z-e^{i\varphi}|) = O(|1-w|),$$

as $z \to e^{i\varphi}$. Hence, from (2.17),

$$\lim_{z\to e^{i\varphi}} \operatorname{Re}[z(1+z)/(1-z)^3]|H(\sigma(z))| = 0.$$

This equality, (2.9), (2.11), (2.18), and (2.20) imply for $c-a-b > 0$,

$$\begin{aligned} &\liminf_{z\to e^{i\varphi}} \operatorname{Re}[zG'(z,x)/G(z,x)] \\ &\geqq \tfrac{1}{2}(1-\alpha-\beta) + 2(1-x)\liminf_{z\to e^{i\varphi}}\{\operatorname{Im}[z(1+z)(1-z)^{-3}]\operatorname{Im} H(\sigma(z))\} \qquad (2.21) \\ &\geqq \tfrac{1}{2}(1-\alpha-\beta). \end{aligned}$$

If $c-a-b < 0$, we rewrite (2.14) in the form:

$$F(a,b,c,w) = (1-w)^{c-a-b}\psi(w),$$

where

$$\begin{aligned} \psi(w) &= A_1(1-w)^{a+b-c}F(a,b,a+b-c+1,1-w) \\ &\quad + A_2F(c-a,c-b,c-a-b+1,1-w) \\ &= A_1(1-w)^{a+b-c} + A_2 + O(|1-w|), \end{aligned}$$

as $w \to 1$, $|\arg(1-w)| < \pi$. Hence

$$\psi'(w) = A_1(c-a-b)(1-w)^{a+b-c-1} - A_2(c-a)(c-b)/(c-a-b+1) + o(1),$$

as $w \to 1$, $|\arg(1-w)| < \pi$. Taking the logarithmic derivative of $F(a,b,c,\cdot)$, we get

$$\begin{aligned} &F'(a,b,c,w)/F(a,b,c,w) \\ &= (a+b-c)/(1-w) + [A_1(c-a-b)(1-w)^{a+b-c-1} \qquad (2.22) \\ &\quad - A_2(c-a)(c-b)/(c-a-b+1)][A_1(1-w)^{a+b-c} + A_2]^{-1} + o(1) \\ &= (a+b-c)/(1-w) - (c-a)(c-b)/(c-a-b+1) + J(w) + o(1), \end{aligned}$$

as $w \to 1$, $|\arg(1-w)| < \pi$, where

$$J(w) = A_1(c-a-b)(1-w)^{a+b-c-1}[A_1(1-w)^{a+b-c} + A_2]^{-1}.$$

As in the proof of (2.18) and (2.20), it follows that

$$\limsup_{z \to e^{i\varphi}} \operatorname{Im} J(w) \leqq 0. \tag{2.23}$$

Now,

$$\begin{aligned} &2(x-1)\operatorname{Re}[z(1+z)(1-z)^{-3}(1-w)^{-1}] \\ &\quad = 2(x-1)\operatorname{Re}[z(1+z)(1-z)^{-1}(1+z^2-2xz)^{-1}] \\ &\quad = \operatorname{Re}[-2(1-z)^{-1} + (1-e^{-i\varphi}z)^{-1} + (1-e^{i\varphi}z)^{-1}] \\ &\quad \geqq \tfrac{1}{2} + \operatorname{Re}[-2(1-z)^{-1} + (1-e^{i\varphi}z)^{-1}] \\ &\quad \geqq \tfrac{1}{2} - \tfrac{1}{2} + o(1) \\ &\quad = o(1), \end{aligned} \tag{2.24}$$

as $z \to e^{i\varphi}$. Here we have used the fact that $\operatorname{Re}(1-\zeta)^{-1} = \frac{1}{2}$ for ζ in T. From (2.9), (2.11), and (2.22)–(2.24), it follows, as previously, that (2.21) holds when $c-a-b<0$. Since,

$$\operatorname{Re}[zG'(z,x)/G(z,x)] = \operatorname{Re}[\bar{z}G'(\bar{z},x)/G(\bar{z},x)], \qquad z \in \Delta,$$

we conclude from (2.21) that

$$\liminf_{z \to e^{-i\varphi}} \operatorname{Re}[zG'(z,x)/G(z,x)] \geqq \tfrac{1}{2}(1-\alpha-\beta). \tag{2.25}$$

Finally, we examine the behavior of $zG'(\cdot,x)/G(\cdot,x)$ as $z \to 1$. From (2.15) and the fact that $b-a=\frac{1}{2}$, we have

$$\begin{aligned} F(a,b,c,w) &\equiv (-w)^{-a}[B_1F(a,1-c+a,1-b+a,w^{-1}) \\ &\qquad + B_2(-w)^{-1/2}F(b,1-c+b,1-a+b,w^{-1})] \\ &= (-w)^{-a}[B_1 + B_2(-w)^{-1/2} + O(|w|^{-1})], \end{aligned}$$

as $w \to \infty$, $|\arg(-w)| < \pi$. Taking the logarithmic derivative of $F(a,b,c,\cdot)$, we get

$$\begin{aligned} F'(a,b,c,w)/F(a,b,c,w) &= -a/w + [\tfrac{1}{2}B_2(-w)^{-3/2} + O(|w|^{-2})][B_1 + O(|w|^{-1/2})]^{-1} \\ &= -a/w + (B_2/2B_1)(-w)^{-3/2} + O(|w|^{-2}) \end{aligned}$$

as $w \to \infty$, $|\arg(-w)| \leqq \pi$. Using this equality with $w = 2(x-1)z/(1-z)^2$, the fact that $a = \frac{1}{2}(1+\alpha+\beta)$, and (2.11), we obtain

$$\begin{aligned} &zG'(z,x)/G(z,x) \\ &\quad = 1 + (1+\alpha+\beta)z/(1-z) - (1+\alpha+\beta)(1+z)/[2(1-z)] \\ &\qquad - (B_2/2B_1)[2(1-x)z]^{-1/2}(1+z) + o(1) \\ &\quad = \tfrac{1}{2}(1-\alpha-\beta) - (B_2/B_1)[2(1-x)]^{-1/2} + o(1) \end{aligned}$$

as $z \to 1$. Since by (2.16), $B_2/B_1 \leqq 0$, it follows that

$$\lim_{z \to 1} \operatorname{Re}[zG'(z,x)/G(z,x)] \geqq \tfrac{1}{2}(1-\alpha-\beta).$$

From this inequality, (2.12), (2.21), (2.25), and the minimum principle for harmonic functions, we conclude

$$\operatorname{Re}[zG'(z,x)/G(z,x)]>\tfrac{1}{2}(1-\alpha-\beta), \qquad z\in\Delta.$$

This proves Lemma 2 when $\frac{1}{2}+\beta\neq$ integer. To prove Lemma 2 when $\frac{1}{2}+\beta$ is an integer, it suffices to observe for fixed $z\in\Delta$ and $-1\leqq x\leqq 1$ that $zG'(z,x)/G(z,x)$ is continuous as a function of α and β.

3. Proof of Theorem 1. We now consider the polynomials $q_n(\cdot,\lambda)$ defined in (1.10). We prove

LEMMA 3. *If* $\lambda\geqq 0$, *then* $n!q_n(\cdot,\lambda)/(1+\lambda)_n$ *is the derivative of a function in* $C(\frac{1}{2}(1-\lambda))$ *for* $n=1,2,\cdots$.

Proof. First observe that $q_n(z,\lambda)$ is the nth coefficient in the Maclaurin series expansion of the function:

$$w\to(1-zw)^{-(1+\lambda)}(1-w)^{-(1+\lambda)}, \qquad w\in\Delta.$$

Hence by Cauchy's theorem,

$$q_n(z,\lambda)=(1/(2\pi i))\int_{|w|=\rho}(1-zw)^{-(1+\lambda)}(1-w)^{-(1+\lambda)}\,dw/w^{n+1},$$

when $z\in\Delta\cup T$ and $\rho<1$. Let $z=e^{i\theta}$ and $w=e^{-i\theta/2}\zeta$. Changing variables in the above equality, we get

$$\begin{aligned} q_n(e^{i\theta},\lambda)&=(1/(2\pi i))\,e^{in\theta/2}\int_{|\zeta|=\rho}\left(1-2\cos\frac{\theta}{2}\zeta+\zeta^2\right)^{-(1+\lambda)}d\zeta/\zeta^{n+1}\\ &=e^{in\theta/2}Q_n^{1+\lambda}[\cos(\theta/2)], \end{aligned} \tag{3.1}$$

where

$$Q_n^{1+\lambda}=[(2+2\lambda)_n/(\tfrac{3}{2}+\lambda)_n]P_n^{(\lambda+1/2,\lambda+1/2)},$$

is the Gegenbauer polynomial of degree n corresponding to $1+\lambda$. It is well known that $Q_n^{1+\lambda}$ has n distinct zeros in $(-1,1)$. Hence $q_n(\cdot,\lambda)$ has its zeros on T. Moreover, if $x_1=\cos(\theta_1/2)$ and $x_2=\cos(\theta_2/2)$, $\theta_1<\theta_2$, are two zeros of $Q_n^{1+\lambda}$, then an argument similar to [15, Thm. 6.3.1] shows that

$$\theta_2-\theta_1>2\pi/(n+1+\lambda), \tag{3.2}$$

for $\lambda\geqq 0$. Let $\arg q_n(\cdot,\lambda)=\operatorname{Im}\log q_n(\cdot,\lambda)$ in Δ, where $\operatorname{Im}\log q_n(0,\lambda)=0$. Let $E=\{\theta: q_n(e^{i\theta},\lambda)=0\}$ and put

$$\arg q_n(e^{i\theta},\lambda)=\lim_{r\to 1}\arg q_n(re^{i\theta},\lambda), \qquad \theta\notin E.$$

Note that $\arg q_n(\cdot,\lambda)$ is continuous on $T-\{e^{i\theta}:\theta\in E\}$, and

$$\lim_{\varepsilon\to 0}[\arg q_n(e^{i\theta+\varepsilon},\lambda)-\arg q_n(e^{i(\theta-\varepsilon)},\lambda)]=-\pi, \tag{3.3}$$

when $\theta\in E$.

We claim for $\varphi_1,\varphi_2\notin E$ and $\varphi_1<\varphi_2$ that

$$\arg q_n(e^{i\varphi_2},\lambda)-\arg q_n(e^{i\varphi_1},\lambda)>-\pi-\tfrac{1}{2}(1+\lambda)(\varphi_2-\varphi_1). \tag{3.4}$$

To prove this claim choose φ_2' so that $\varphi_2-\varphi_2'$ is divisible by 2π, and $\varphi_1<\varphi_2'\leqq\varphi_1+2\pi$. Suppose that $q_n(\cdot,\lambda)$ has m zeros on the arc: $\{e^{i\varphi}:\varphi_1\leqq\varphi\leqq\varphi_2'\}$. Then from the periodicity of $\arg q_n(\cdot, \lambda)$, and (3.1)–(3.3), we get

$$\begin{aligned}\arg q_n(e^{i\varphi_2}, \lambda) - \arg q_n(e^{i\varphi_1}, \lambda)\\ &= n(\varphi_2' - \varphi_1)/2 - m\pi\\ &= \tfrac{1}{2}(n+1+\lambda)(\varphi_2' - \varphi_1) - m\pi - \tfrac{1}{2}(1+\lambda)(\varphi_2' - \varphi_1)\\ &> -\pi - \tfrac{1}{2}(1+\lambda)(\varphi_2 - \varphi_1).\end{aligned}$$

Lemma 3 follows from (3.4) and an argument similar to either Suffridge [14, Thm. 2] or Kaplan [7]. We prefer to argue as in Kaplan. Let

$$p(e^{i\varphi}) = \sup_{\substack{-\infty<\theta\leqq\varphi\\ \theta\notin E}} [\arg q_n(e^{i\theta}, \lambda) + \tfrac{1}{2}(1+\lambda)\theta] - \tfrac{1}{2}(1+\lambda)\varphi - \pi/2.$$

Clearly,

$$p(e^{i\varphi}) \geqq \arg q_n(e^{i\varphi}, \lambda) - \pi/2, \qquad \varphi \notin E,$$

Also, from (3.4) we deduce

$$p(e^{i\varphi}) \leqq \arg q_n(e^{i\varphi}, \lambda) + \pi/2, \qquad \varphi \notin E.$$

Hence,

$$|p(e^{i\varphi}) - \arg q_n(e^{i\varphi}, \lambda)| \leqq \pi/2, \qquad \varphi \notin E. \tag{3.5}$$

It is easily seen that p is periodic of period 2π, and

$$p(e^{i\varphi_2}) - p(e^{i\varphi_1}) \geqq -\tfrac{1}{2}(1+\lambda)(\varphi_2 - \varphi_1), \tag{3.6}$$

when $\varphi_1 < \varphi_2$.

We now define f analytic in Δ by the Poisson integral,

$$\log(f(z)/z) = \frac{i}{2\pi}\int_0^{2\pi} p(e^{i\varphi})\frac{1+e^{-i\varphi}z}{1-e^{-i\varphi}z}\,d\varphi, \qquad z \in \Delta.$$

Observe for $z = re^{i\psi}$ that

$$\arg(f(z)/z) = (1/(2\pi))(1-r^2)\int_0^{2\pi} p(e^{i(\varphi+\psi)})(1+r^2-2r\cos\varphi)^{-1}\,d\varphi. \tag{3.7}$$

(3.6) and (3.7) imply that

$$\arg[e^{-i\psi_2}f(re^{i\psi_2})] - \arg[e^{-i\psi_1}f(re^{i\psi_1})] \geqq -\tfrac{1}{2}(1+\lambda)(\psi_2 - \psi_1),$$

when $\psi_1 < \psi_2$ and $0 < r < 1$. Hence, $f/f'(0) \in S(\tfrac{1}{2}(1-\lambda))$. From (3.5), (3.7), and a corresponding Poisson integral for $\arg q_n(\cdot, \lambda)$, it follows that

$$|\arg(f(z)/z) - \arg q_n(z, \lambda)| \leqq \pi/2.$$

Thus,

$$zq_n(z, \lambda) = f(z)P(z), \qquad z \in \Delta,$$

where P is analytic in Δ with positive real part. If $|f'(0)|^{-1}f'(0) = e^{i\theta}$, let

$$h(z) = f(e^{-i\theta}z)/(|f'(0)|), \qquad z \in \Delta,$$

and define g analytic in Δ by $g(0) = 0$,

$$g'(z) = n!\,q_n(z, \lambda)/(1+\lambda)_n, \qquad z \in \Delta.$$

Then $h \in S(\frac{1}{2}(1-\lambda))$ and g, h, satisfy (1.5), so $g \in C(\frac{1}{2}(1-\lambda))$. Hence Lemma 3 is true.

We remark that Lemma 3 fails when $-1 < \lambda < 0$, essentially because inequality (3.2) is reversed in this case. To prove Theorem 1 first observe that (1.6) can be written in the form (1.11). Second, observe that $C(\frac{1}{2}(1-\lambda))$ is contained in $C(\frac{1}{2}(1-\alpha-\beta))$ for $0 \leqq \lambda \leqq \alpha+\beta$. Theorem 1 follows from these observations, Lemmas 2–3, and Theorem A, as in § 1. We omit the details. The proof of Theorem 1 is now complete.

REFERENCES

[1] R. ASKEY, *Some absolutely monotonic functions*, Studia Sci. Math., 9 (1974), pp. 51–56.
[2] R. ASKEY AND G. GASPER, *Positive Jacobi polynomial sums II*, Amer. J. Math., 3, 98 (1976), pp. 709–737.
[3] J. BUSTOZ AND N. SAVAGE, *Inequalities for ultraspherical and Laguerre polynomials*, this Journal, to appear.
[4] A. ERDÉLYI, *Higher Transcendental Functions*, I, McGraw-Hill, New York, 1953.
[5] G. GASPER, *Positive sums of the classical orthogonal polynomials*, this Journal, 8 (1977), 423–447.
[6] I. S. JACK, *Functions starlike and convex of order α*, J. London Math. Soc., (2) (3) (1971), pp. 469–474.
[7] W. KAPLAN, *Close to convex schlicht functions*, Michigan Math. J., 1 (1953), pp. 169–185.
[8] J. LEWIS, *Convolutions of starlike functions*, Indiana Univ. Math. J., 27 (1978), pp. 671–688.
[9] T. MACGREGOR, *A subordination for convex functions of order α*, J. London Math. Soc., 9 (1975), pp. 530–536.
[10] E. RAINVILLE, *Special Functions*, Macmillan Company, New York, 1965.
[11] S. RUSCHEWEYH, *Linear operators between classes of prestarlike functions*, to appear.
[12] ———, *On the Kakeya–Eneström theorem and Gegenbauer polynomial sums*, to appear.
[13] ———, *New criteria for univalent functions*, Proc. Amer. Math. Soc., 49 (1975), pp. 109–115.
[14] T. J. SUFFRIDGE, *Starlike Functions as Limits of Polynomials*, Advances in Complex Function Theory, Springer Lecture Notes, 505, Springer-Verlag, New York, pp. 164–202.
[15] G. SZEGÖ, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, 23, New York, 1975.
[16] E. WHITTAKER AND G. WATSON, *A Course of Modern Analysis*, Cambridge University Press, London, 1969.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0004 $01.00/0

A SENSITIVITY ANALYSIS OF SYSTEMS CONSISTING OF LINEAR BLOCKS*

VACLAV DOLEZAL†

Abstract. The paper presents a simple yet quite general method for analyzing the sensitivity of a performance characteristic of a system consisting of linear blocks. The analysis is based on properties of a gradient of an abstract function. Applications in network theory and in the theory of feedback systems are discussed.

Introduction. In this paper our main objective is the sensitivity analysis of systems consisting of finitely many linear blocks. The basic idea concerning sensitivity is standard, but the present setting is much more general. In the traditional sensitivity analysis it is assumed [1], [2] that some elements of the system under consideration depend on a real parameter K. Here, the "parameter K" has a broader meaning; it can be, for example, an operator describing some particular block, a vector, etc.

To describe our approach in more detail, consider a system S, whose specified performance characteristic F depends on certain quantities $X_1^0, X_2^0, \cdots, X_n^0$. Assume that these quantities are in some Banach space $\mathscr{L}$, and that they are allowed to vary within certain limits. Thus, the performance characteristic F appears as a function of an n-vector $X = (X_1, X_2, \cdots, X_n)$, which lies in a given vicinity U_0 of the nominal n-vector $X^0 = (X_1^0, X_2^0, \cdots, X_n^0)$. Suppose that we can find linear operators $f_1, f_2, \cdots, f_n$ such that

$$F(X) - F(X^0) = \sum_{i=1}^{n} f_i(X_i - X_i^0) + \Psi(X - X^0) \tag{i}$$

for every X in U_0, where the size of the quantity $\Psi(X - X^0)$ is much smaller than the numbers $\|X_i - X_i^0\|$, $i = 1, 2, \cdots, n$, i.e., $\Psi(X - X^0)$ can be neglected in the first approximation. Then we can adopt the operator f_i or its norm $\|f_i\|$ as a measure of the sensitivity of F to variations $X_i - X_i^0$ of the nominal value X_i^0.

Note that we did not specify the nature of the quantity $F(X)$; it can be an operator, element of some fixed Banach space $\mathscr{H}$, etc.

As it is suggested by equation (i) the operators f_i, $i = 1, 2, \cdots, n$ are the entries of the (strong) gradient of F at X^0. This fact turns out to be particularly useful, if the quantities X_i, $i = 1, 2, \cdots, n$ have the meaning of operators describing certain blocks which constitute our system S. Note that the essential fact in this situation is that $F(X)$ is obtained by forming finitely many sums, products and inverses from operators $X_1, X_2, \cdots, X_n$.

To develop these ideas, in the first part of the paper we will study the properties of a gradient which concern linear combinations, products and inverses of differentiable mappings. At the same time, we will be concerned with the "size of the wastebasket term $\Psi(X - X^0)$ in (i)" so as to have an estimate for the accuracy of replacing $F(X) - F(X^0)$ by the term $\sum_{i=1}^n f_i(X_i - X_i^0)$.

Then we will discuss certain applications of our results in the network theory and the theory of feedback systems.

* Received by the editors June 16, 1978, and in revised form September 14, 1978.

† Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York, 11794. This research was supported by the National Science Foundation under Grant MPS7505268.

In the final part of the paper we will give simple rules which permit us to calculate the gradient of performance characteristics usually encountered in systems analysis.

1. Operator-valued functions. We will use the following notations.

Let R^n be the real n-dimensional Euclidean space with norm $|\xi| = (\sum_{i=1}^n \xi_i^2)^{1/2}$.

If H is a Banach space and $n \geqq 1$ is an integer, we let H^n be the n-fold Cartesian product of H. For each $X = (X_1, X_2, \cdots, X_n) \in H^n$ we define the norm $\|X\|'$ by

$$\|X\|' = \left(\sum_{i=1}^n \|X_i\|^2 \right)^{1/2}. \tag{1.1}$$

If $X^0 \in H^n$ and $r > 0$ denote

$$B_r(X^0) = \{X : X \in H^n, \|X - X^0\|' < r\}. \tag{1.2}$$

Finally, if H_1, H_2 are Banach spaces over the same field of scalars, we let $[H_1, H_2]$ be the Banach space of all linear bounded operators $A: H_1 \to H_2$ which is equipped with the customary norm $A = \sup_{\|x\|=1, x \in H_1} \|Ax\|$.

Definition 1.1. Let $\mathscr{L}$, $\mathscr{H}$ be fixed Banach spaces, and let $D \subset \mathscr{L}^n$ be a nonempty open set. Furthermore, let $F: D \to \mathscr{H}$ and let $X^0 \in D$. If

(i) there exist operators $f_i \in [\mathscr{L}, \mathscr{H}]$, $i = 1, 2, \cdots, n$ and

(ii) for some $r > 0$ with $B_r(X^0) \subset D$ there exists a function $\Phi: B_r(0) \to [0, \infty)$ continuous and vanishing at $0 \in \mathscr{L}^n$ such that

$$\|F(X^0 + Z) - F(X^0) - \sum_{i=1}^n f_i(Z_i)\| \leqq \|Z\|' \Phi(Z) \tag{1.3}$$

for all $Z = (Z_1, Z_2, \cdots, Z_n) \in B_r(0)$,

then F will be called differentiable at X^0, and the n-vector

$$\partial F = (f_1, f_2, \cdots, f_n) \in [\mathscr{L}, \mathscr{H}]^n$$

will be called the gradient of F at X^0. Also, any Φ satisfying (1.3) will be called an error function of F on $B_r(0)$.

Observe that from (1.3) it follows that

$$\|F(X^0 + Z) - F(X^0)\| \leqq \sum_{i=1}^n \|f_i\| \cdot \|Z_i\| + \|Z\|' \cdot \Phi(Z) \tag{1.4}$$

for all $Z \in B_r(0)$.

Moreover, from (1.4) we have by (1.1),

$$\|F(X^0 + Z) - F(X^0)\| \leqq (\|\partial F\| + \Phi(Z)) \cdot \|Z\|' \tag{1.5}$$

for all $Z \in B_r(0)$. Hence, F is continuous at X^0.

Before proceeding further, let us emphasize the fact that the operators f_k, $i = 1, 2, \cdots, n$ are determined uniquely by our definition. Indeed, suppose that there exist $\bar{f}_i \in [\mathscr{L}, \mathscr{H}]$, $i = 1, 2, \cdots, n$ and, for some $\bar{r} > 0$ with $B_{\bar{r}}(X^0) \subset D$, a function $\bar{\Phi}: B_{\bar{r}}(0) \to [0, \infty)$, continuous and vanishing at $0 \in \mathscr{L}^n$ such that

$$\left\|F(X^0 + Z) - F(X^0) - \sum_{i=1}^n \bar{f}_i(Z_i)\right\| \leqq \|Z\|' \cdot \bar{\Phi}(Z) \tag{1.6}$$

for all $Z \in B_{\bar{r}}(0)$. Choose an index $1 \leqq j \leqq n$, and let $\varepsilon > 0$. By continuity of Φ and $\bar{\Phi}$ at 0, there exist $r_1 > 0$ and $r_2 > 0$ such that $0 \leqq \Phi(Z) < \varepsilon/2$ for $\|Z\|' < r_1$, and $0 \leqq \bar{\Phi}(Z) < \varepsilon/2$ for $\|Z\|' < r_2$. Put $r_0 = \min [r_1, r_2] > 0$, choose $W \in \mathscr{L}$ with $\|W\| < r_0$ and let $Z = (0, 0, \cdots, 0, W, 0, \cdots, 0) \in \mathscr{L}^n$, where W stands at the jth place. Then $\|Z\|' = \|W\| < r_0$

and (1.3), (1.6) yield by triangular law

$$\|(f_i-\bar{f}_i)W\| \leqq \|Z\|'(\Phi(Z)+\bar{\Phi}(Z)) \leqq \varepsilon\|W\|. \tag{1.7}$$

Since f_i, $\bar{f}_i$ are linear, (1.7) holds for any $W \in \mathscr{L}$. Hence, $\|f_i-\bar{f}_i\| \leqq \varepsilon$, and consequently, $f_i=\bar{f}_i$.

To give an example of a differentiable mapping, let h be a Banach space, and let $H=h$, $\mathscr{L}=[h,h]$. If $A_0 \in [h,h]$ and a_i, $i=1,2,\cdots,n$ are scalars, define $F\colon \mathscr{L}^n \to \mathscr{L}$ by

$$F(X)=A_0+\sum_{i=1}^{n} a_i X_i. \tag{1.8}$$

Then it is easy to see that, for any $X^0 \in \mathscr{L}^n$, F is differentiable at X^0, and the gradient ∂F at X^0 is given by $\partial F=(a_1 I, a_2 I, \cdots, a_n I)$, where I is the identity operator in $[\mathscr{L},\mathscr{L}]$. Also, $\Phi \equiv 0$ is an error function of F on $B_r(0)$ with any $r>0$.

THEOREM 1.1. *Let $D \subset \mathscr{L}^n$ be a nonempty open set, let $F, G\colon D \to \mathscr{H}$ be differentiable at $X^0 \in D$, and let a be a scalar. Then $F+G$ and aF are also differentiable at X^0 and*

$$\partial(F+G)=\partial F+\partial G, \tag{1.9}$$

$$\partial(aF)=a\,\partial F. \tag{1.10}$$

Moreover, if Φ_F and Φ_G is an error function of F and G on $B_r(0)$, respectively, then $\phi_F+\phi_G$ and $|a|\phi_F$ is an error function of $F+G$ and aF on $B_r(0)$ respectively.

Proof. The proof follows trivially from Definition 1.1.

THEOREM 1.2. *Let $D \subset \mathscr{L}^n$ be a nonempty open set, let $F\colon D \to [H_1, H_2]$ be differentiable at $X^0 \in D$, and let $A \in [H_2, H_3]$, $B \in [H_0, H_1]$. Then the mapping $G\colon D \to [H_0, H_3]$, defined by*

$$G(X)=AF(X)B, \tag{1.11}$$

is differentiable at X^0, and for the gradient $\partial G=(g_1, g_2, \cdots, g_n) \in [\mathscr{L}, [H_0, H_3]]^n$ we have

$$g_i(W)=Af_i(W)B \tag{1.12}$$

for all $W \in \mathscr{L}$ and $i=1,2,\cdots,n$, where $(f_1, f_2, \cdots, f_n)=\partial F$. Moreover, if Φ is an error function of F on $B_r(0)$, then $\|A\|\cdot\|B\|\Phi$ is an error function of G on $B_r(0)$.

This assertion is again a trivial consequence of the Definition 1.1.

THEOREM 1.3. *Let $D \subset \mathscr{L}^n$ be a nonempty open set, and let the mappings $F^1\colon D \to [H_1, H_2]$ and $F^2\colon D \to [H_0, H_1]$ be differentiable at $X^0 \in D$. Then the mapping $(F^1 \circ F^2)\colon D \to [H_0, H_2]$, defined by*

$$(F^1 \circ F^2)(X)=F^1(X)F^2(X), \tag{1.13}$$

is differentiable at X_0, and for the gradient $\partial(F^1 \circ F^2)=(g_1, g_2, \cdots, g_n) \in [\mathscr{L}, [H_0, H_2]]^n$ we have

$$g_i(W)=f_i^1(W)F^2(X^0)+F^1(X^0)f_i^2(W) \tag{1.14}$$

for each $W \in \mathscr{L}$, $i=1,2,\cdots,n$, where $\partial F^j=(f_1^j, f_2^j, \cdots, f_n^j)$, $j=1,2$.

Moreover, if Φ^1 and Φ^2 is an error function of F^1 and F^2 on $B_r(0)$, respectively, then the function Φ, defined by

$$\Phi(Z)=\|F^1(X^0)\|\Phi^2(Z)+\|F^2(X^0)\|\Phi^1(Z)+(\|\partial F^1\|+\Phi^1(Z))(\|\partial F^2\|+\Phi^2(Z))\|Z\|' \tag{1.15}$$

is an error function of $F^1 \circ F^2$ on $B_r(0)$.

Proof. Choose $Z \in B_r(0)$. Then we have by (1.13) and (1.14),

$$
\begin{aligned}
m &= \left\| (F^1 \circ F^2)(X^0+Z) - (F^1 \circ F^2)(X^0) - \sum_{i=1}^n g_i(Z_i) \right\| \\
&= \| F^1(X^0)(F^2(X^0+Z) - F^2(X^0)) + (F^1(X^0+Z) - F^1(X^0))F^2(X^0) \\
&\qquad - \sum_{i=1}^n (f_i^1(Z_i)F^2(X^0) + F^1(X^0)f_i^2(Z_i)) \\
&\qquad\qquad + (F^1(X^0+Z) - F^1(X^0))(F^2(X^0+Z) - F^2(X^0)) \| \\
&\leqq \|F^1(X^0)\| \cdot \left\| F^2(X^0+Z) - F^2(X^0) - \sum_{i=1}^n f_i^2(Z_i) \right\| \\
&\qquad + \left\| F^1(X^0+Z) - F^1(X^0) - \sum_{i=1}^n f_i^1(Z_i) \right\| \cdot \|F^2(X^0)\| \\
&\qquad + \|F^1(X^0+Z) - F^1(X^0)\| \cdot \|F^2(X^0+Z) - F^2(X^0)\|.
\end{aligned} \tag{1.16}
$$

Invoking (1.3) and (1.6), it follows that

$$
\begin{aligned}
m &\leqq \|F^1(X^0)\| \Phi^2(Z) \|Z\|' + \|F^2(X^0)\| \Phi^1(Z) \|Z\|' \\
&\qquad + (\|\partial F^1\| + \Phi^1(Z))(\|\partial F^2\| + \Phi^2(Z)) \|Z\|'^2 = \Phi(Z)\|Z\|'.
\end{aligned} \tag{1.17}
$$

Since $g_i \in [\mathscr{L}, [H_0, H_2]]$ by (1.14) for $i = 1, 2, \cdots, n$, and Φ is continuous and vanishing at the origin $0 \in \mathscr{L}^n$, (1.16), (1.17) show that (1.3) holds for $F^1 \circ F^2$. Hence the uniqueness of the gradient concludes the proof.

For the proof of the next theorem we will need the following assertion.

LEMMA 1.1. *Let H_1, H_2 be Banach spaces having the same system of scalars, let ${}^0A \in [H_1, H_2]$ be invertible, and let $0 < \alpha < 1$. If $A \in [H_1, H_2]$ and*

$$\|A - {}^0A\| \leqq \alpha \|{}^0A^{-1}\|^{-1}; \tag{1.18}$$

then A is invertible, $A^{-1} \in [H_2, H_1]$ and

$$\|A^{-1}\| \leqq (1-\alpha)^{-1} \|{}^0A^{-1}\|, \tag{1.19}$$

$$\|A^{-1} - {}^0A^{-1}\| \leqq (1-\alpha)^{-1} \|{}^0A^{-1}\|^2 \cdot \|A - {}^0A\|. \tag{1.20}$$

Proof. First observe that ${}^0A^{-1} \in [H_2, H_1]$ by virtue of the open mapping theorem. If $A \in [H_1, H_2]$ satisfies (1.18), let

$$B = I + (A - {}^0A){}^0A^{-1} \in [H_2, H_2]. \tag{1.21}$$

Then clearly $A = B\,{}^0A$.

Next, choose $y \in H_2$ and define the operator $C_y \colon H_2 \to H_2$ by

$$C_y x = -(A - {}^0A){}^0A^{-1}x + y. \tag{1.22}$$

If $x_1, x_2 \in H_2$, we have by (1.18),

$$
\begin{aligned}
\|C_y x_1 - C_y x_2\| &\leqq \|(A - {}^0A){}^0A^{-1}\| \cdot \|x_1 - x_2\| \\
&\leqq \|A - {}^0A\| \cdot \|{}^0A^{-1}\| \cdot \|x_1 - x_2\| \leqq \alpha \|x_1 - x_2\|.
\end{aligned}
$$

Thus, C_y is a contraction on H_2, and consequently, there exists a unique $x \in H_2$ such that $x = C_y x$, i.e., by (1.22) and (1.21), $Bx = y$. Hence, B is invertible.

Furthermore, let $y_1, y_2 \in H_2$ and put $x_j = B^{-1}y_j$, $j = 1, 2$. Then we have by (1.22),

$$\|x_1 - x_2\| = \|C_{y_1}x_1 - C_{y_2}x_2\| = \|-(A - {}^0A){}^0A^{-1}(x_1 - x_2) + y_1 - y_2\| \leqq \alpha\|x_1 - x_2\| + \|y_1 - y_2\|. \tag{1.23}$$

Hence, from (1.23),

$$\|B^{-1}y_1 - B^{-1}y_2\| = \|x_1 - x_2\| \leqq (1-\alpha)^{-1}\|y_1 - y_2\|. \tag{1.24}$$

Since B^{-1} is linear, (1.24) shows that B^{-1} is bounded and $\|B^{-1}\| \leqq (1-\alpha)^{-1}$.

Since $A = B\,{}^0A$, it follows that A is invertible and $A^{-1} = {}^0A^{-1}B^{-1}$. Thus, $A^{-1} \in [H_2, H_1]$ because ${}^0A^{-1} \in [H_2, H_1]$, $B^{-1} \in [H_2, H_2]$, and $\|A^{-1}\| \leqq \|{}^0A^{-1}\| \cdot \|B^{-1}\| \leqq (1-\alpha)^{-1}\|{}^0A^{-1}\|$ which confirms (1.19).

Finally, we have by (1.19),

$$\begin{aligned}\|A^{-1} - {}^0A^{-1}\| &= \|A^{-1}(A - {}^0A){}^0A^{-1}\| \\ &\leqq \|A^{-1}\| \cdot \|{}^0A^{-1}\| \cdot \|A - {}^0A\| \\ &\leqq (1-\alpha)^{-1}\|{}^0A^{-1}\|^2 \cdot \|A - {}^0A\|,\end{aligned}$$

which proves (1.20).

THEOREM 1.4. *Let $D \subset \mathscr{L}^n$ be a nonempty open set, let the mapping $F\colon D \to [H_1, H_2]$ be differentiable at $X^0 \in D$, and let $F(X^0)$ be invertible. Then there exists $r^* > 0$ such that $F(X)$ is invertible for each $X \in B_{r^*}(X^0)$, and the mapping $G\colon B_{r^*}(X^0) \to [H_2, H_1]$ defined by*

$$G(X) = [F(X)]^{-1} \tag{1.25}$$

is differentiable at X^0. Also, for the gradient $\partial G = (g_1, g_2, \cdots, g_n) \in [\mathscr{L}, [H_2, H_1]]^n$ we have

$$g_i(W) = -[F(X^0)]^{-1}f_i(W)[F(X^0)]^{-1} \tag{1.26}$$

for all $W \in \mathscr{L}$ and $i = 1, 2, \cdots, n$, where $(f_1, f_2, \cdots, f_n) = \partial F$.

Moreover, if Φ is an error function of F on $B_r(0)$, where $r > 0$ is so small that $0 \leqq \Phi(Z) \leqq 1$ for each $Z \in B_r(0)$, and if $0 < \alpha < 1$, then with

$$r^* = \min[r, \alpha(\|\partial F\| + 1)^{-1}\|[F(X^0)]^{-1}\|^{-1}] > 0, \tag{1.27}$$

the function Φ^* defined by

$$\Phi^*(Z) = \|[F(X^0)]^{-1}\|^2\{\Phi(Z) + (1-\alpha)^{-1}\|[F(X^0)]^{-1}\|(\|\partial F\| + 1)^2\|Z\|'\} \tag{1.28}$$

is an error function of G on $B_{r^*}(0)$.

Proof. From Definition 1.1 it follows that, for a given F and $X^0 \in D$, we can always find $r > 0$ such that $0 \leqq \Phi(Z) \leqq 1$ on $B_r(0)$ and Φ satisfies (1.3) for every $Z \in B_r(0)$. Fixing such an $r > 0$, we have by (1.6),

$$\|F(X^0 + Z) - F(X^0)\| \leqq (\|\partial F\| + 1)\|Z\|' \tag{1.29}$$

for every $Z \in B_r(0)$.

Choose now $0 < \alpha < 1$ and define $r^* > 0$ by (1.27). If $Z \in B_{r^*}(0)$, we have by (1.29) and (1.27),

$$\begin{aligned}\|F(X^0 + Z) - F(X^0)\| &\leqq (\|\partial F\| + 1)\|Z\|' \leqq (\|\partial F\| + 1)r^* \\ &\leqq \alpha\|[F(X^0)]^{-1}\|^{-1}.\end{aligned} \tag{1.30}$$

Hence, by Lemma 1.1, $F(X^0 + Z)$ is invertible, $G(X^0 + Z) = [F(X^0 + Z)]^{-1} \in [H_2, H_1]$, and, by (1.20), (1.30),

$$\begin{aligned}\|G(X^0+Z)-G(X^0)\| &\leqq (1-\alpha)^{-1}\|[F(X^0)]^{-1}\|^2\|F(X^0+Z)-F(X^0)\| \\ &\leqq (1-\alpha)^{-1}\|[F(X^0)]^{-1}\|^2(\|\partial F\|+1)\|Z\|'.\end{aligned} \tag{1.31}$$

Moreover, by (1.26), (1.30) and (1.31) we have for such $Z \in B_{r^*}(0)$,

$$\begin{aligned}
&\left\|G(X^0+Z)-G(X^0)-\sum_{i=1}^n g_i(Z_i)\right\| \\
&= \left\|[F(X^0+Z)]^{-1}-[F(X^0)]^{-1}+\sum_{i=1}^n [F(X^0)]^{-1}f_i(Z_i)[F(X^0)]^{-1}\right\| \\
&= \left\|[F(X^0+Z)]^{-1}\{F(X^0+Z)-F(X^0)\}[F(X^0)]^{-1}\right. \\
&\qquad \left. -\sum_{i=1}^n [F(X^0)]^{-1}f_i(Z_i)[F(X^0)]^{-1}\right\| \\
&= \left\|[F(X^0)]^{-1}\left\{F(X^0+Z)-F(X^0)-\sum_{i=1}^n f_i(Z_i)\right\}[F(X^0)]^{-1}\right. \qquad (1.32)\\
&\qquad \left. +([F(X^0+Z)]^{-1}-[F(X^0)]^{-1})(F(X^0+Z)-F(X^0))[F(X^0)]^{-1}\right\| \\
&\leqq \|[F(X^0)]^{-1}\|^2\Phi(Z)\|Z\|'+(1-\alpha)^{-1}\|[F(X^0)]^{-1}\|^3(\|\partial F\|+1)^2\|Z\|'^2 \\
&= \Phi^*(Z)\|Z\|'.
\end{aligned}$$

Since clearly $g_i \in [\mathscr{L}, [H_2, H_1]]$ for $i=1,2,\cdots,n$, and Φ^* is continuous and vanishing at $0 \in \mathscr{L}^n$, (1.3) holds for G and ∂G. Hence, the uniqueness of the gradient concludes the proof.

The rules for the gradient we derived are easy to memorize because they resemble the corresponding rules on partial differentiation (in fact they reduce to them in the particular case of $\mathscr{L}=H_1=H_2=R^1$). For example, (1.14) and (1.26) is an analog of the rule $\partial/\partial x_i(FG)=(\partial F/\partial x_i)G+F(\partial G/\partial x_i)$ and $\partial/\partial x_i(1/F)=-(1/F^2)(\partial F/\partial x_i)$, respectively.

Also, we can easily prove an assertion which is similar to that concerning a total differential and partial derivatives of a function. Indeed, let us define the following concept:

DEFINITION 1.2. Let $\mathscr{L}$, $\mathscr{H}$ be fixed Banach spaces, let $D \subset \mathscr{L}^n$ be a nonempty open set, let $F: D \to \mathscr{H}$ and $X^0 \in D$. If $1 \leqq j \leqq n$, let

$$D_j=\{W: W \in \mathscr{L}, (X_1^0, X_2^0, \cdots, X_{j-1}^0, W, X_{j+1}^0, \cdots, X_n^0) \in D\}. \tag{1.33}$$

We will say that *F possesses a derivative by X_j* at X^0, if the mapping $F_j: D_j \to \mathscr{H}$, defined by

$$F_j(W)=F(X_1^0, X_2^0, \cdots, X_{j-1}^0, W, X_{j+1}^0, \cdots, X_n^0), \tag{1.34}$$

is differentiable at $W=X_j^0$. Then the gradient of F_j (a 1-vector) will be called the derivative of F by X_j at X^0 and denoted by $\delta_j F$.

Clearly, $\delta_j F \in [\mathscr{L}, \mathscr{H}]$, and we have:

THEOREM 1.5. *Let $D \subset \mathscr{L}^n$ be a nonempty set, and let the mapping $F: D \to \mathscr{H}$ be differentiable at $X^0 \in D$. Then, for $j=1,2,\cdots,n$, F possesses a derivative $\delta_j F$ by X_j at X^0, and*

$$\partial F=(\delta_1 F, \delta_2 F, \cdots, \delta_n F). \tag{1.35}$$

Proof. Choose $1 \leqq j \leqq n$. Referring to Definition 1.1, choose $\bar{W} \in \mathscr{L}$ so that $Z = (0, 0, \cdots, 0, \bar{W}, 0, \cdots, 0) \in B_r(0)$. Then we have by (1.3) and (1.34),

$$(1.36) \qquad \begin{aligned} \|F(X^0+Z)-F(X^0)-f_j(\bar{W})\| &= \|F_j(X_j^0+\bar{W})-F_j(X_j^0)-f_j(\bar{W})\| \\ &\leqq \|Z\|'\Phi(Z) = \|\bar{W}\|\Phi(Z). \end{aligned}$$

However, defining $\Phi_j(\bar{W})$ for each $\bar{W}$ with $\|\bar{W}\|<r$ by $\Phi_j(\bar{W}) = \Phi(0, 0, \cdots, 0, \bar{W}, 0, \cdots, 0)$, we see readily that Φ_j is continuous and vanishing at $\bar{W}=0$. Hence, by (1.36), f_j is the gradient of F_j at X_j^0. By uniqueness of a gradient it follows that $f_j = \delta_j F$ and (1.35) is proved.

A converse of Theorem 1.5, of course, does not hold.

Observe also that if F "does not depend on X_j," then $\delta_j F = 0$. In more detail, we have the following obvious proposition:

If $F: D \to \mathscr{H}$ and there exist $1 \leqq j \leqq n$ and $r_0 > 0$ such that $F(X_1^0, X_2^0, \cdots, X_{j-1}^0, W, X_{j+1}^0, \cdots, X_n^0) = F(X_1^0, X_2^0, \cdots, X_n^0)$ for all $W \in \mathscr{L}$ with $\|W - X_j^0\| < r_0$, then $\delta_j F$ exists and is zero.

Before we turn to applications, let us discuss the interpretation of a gradient from the viewpoint of sensitivity analysis.

Consider a system S whose performance characteristic F depends on quantities $X_1, X_2, \cdots, X_n$, which are allowed to vary in a certain vicinity of a nominal point $X^0 = (X_1^0, X_2^0, \cdots, X_n^0)$. If F is differentiable at X^0, then, by virtue of (1.3) and (1.35),

$$(1.37) \qquad F(X^0+Z)-F(X^0) = \sum_{i=1}^{n} (\delta_i F)(Z_i) + \psi(Z)$$

for all $Z \in B_r(0)$ with some $r>0$, where $\|\psi(Z)\| \leqq \|Z\|' \cdot \Phi(Z)$. Thus, for $i = 1, 2, \cdots, n$, we can adopt the operator $\delta_i F$, or possibly its norm, for a measure of sensitivity of F to a variation Z_i of the nominal quantity X_i^0. If, in addition, the function $\Phi(Z)$ is known, we have an estimate for the accuracy of replacing the increment $F(X^0+Z)-F(X^0)$ by the linear combination $\sum_{i=1}^n (\delta_i F)(Z_i)$.

Let us point out the fact that our setting is quite general, since no assumption has been made about the nature of quantities $X_1, X_2, \cdots, X_n$ and values of F.

If, in particular, the X_i's are numbers, i.e., $\mathscr{L} = R^1$, we get the traditional setting for sensitivity analysis. On the other hand, for the X_i's we can take the operators describing certain blocks in our system S. As we shall see below, in this case it is easy to find the gradient of F by using Theorems 1.1 through 1.5.

As far as the nature of values of F is concerned, the following three cases are of interest:

For each $X = (X_1, X_2, \cdots, X_n) \in D$, $F(X)$ is

(i) an operator taking inputs x in a given Banach space H_1 into outputs $y = F(X)x$ in a Banach space H_2,

(ii) the output $y_0 \in H_2$ corresponding to a fixed input $x_0 \in H_1$,

(iii) the value $\mathscr{S}(y_0)$, where $y_0 \in H_2$ is the output corresponding to a fixed nominal input $x_0 \in H_1$, and $\mathscr{S}$ is a fixed functional in $[H_2, R^1]$.

To give an example illustrating the case (iii), assume that $H_2 = [C[0, \infty)]^k$, where $C[0, \infty)$ is the space of all continuous bounded functions on $[0, \infty)$ with the sup norm. If $y = (y_1, y_2, \cdots, y_k) \in H_2$, we can put $\mathscr{S}(y) = y_r(t_0)$, where $1 \leqq r \leqq k$ and $t_0 \geqq 0$ are fixed.

It is clear that, if $F: D \to \mathscr{H}$, then $\mathscr{H} = [H_1, H_2]$ in case (i), $\mathscr{H} = H_2$ in case (ii) and $\mathscr{H} = R^1$ in case (iii).

The case (i) is particularly important, because if the gradient of F at a nominal point X^0 is known, we can immediately establish gradients for the corresponding cases (ii) and (iii). Indeed, if $x_0 \in H_1$ is a nominal input, then for the output $y_0 \in H_2$ we have $y_0 = F(X)x_0$. Thus, if we let $F^1(X) = F(X)x_0$, then $F^1: D \to H_2$ and we have by (1.3)

$$\begin{aligned} &\left\| F^1(X^0+Z) - F^1(X^0) - \sum_{i=1}^n f_i(Z_i)x_0 \right\| \\ &\quad = \left\| \left\{ F(X^0+Z) - F(X^0) - \sum_{i=1}^n f_i(Z_i) \right\} x_0 \right\| \leqq \|Z\|' \cdot \Phi(Z)\|x_0\|. \end{aligned} \tag{1.38}$$

Since $\|x_0\|\Phi(Z)$ is continuous and vanishing at $Z = 0$, (1.38) shows by the uniqueness of gradient that $(f_1^1, f_2^1, \cdots, f_n^1)$ with $f_i^1 \in [\mathscr{L}, H_2]$ defined by

$$f_i^1(W) = f_i(W)x_0, \tag{1.39}$$

$i = 1, 2, \cdots, n$ is the gradient ∂F^1 of F^1 at X^0.

Similarly, if $\mathscr{S} \in [H_2, R_1]$ is fixed, then for $F^2: D \to R^1$ defined by $F^2(X) = \mathscr{S}(F(X)x_0)$ we have by (1.3),

$$\left\| F^2(X^0+Z) - F^2(X^0) - \sum_{i=1}^n (f_i(Z_i)x_0) \right\| \leqq \|Z\|'\Phi(Z)\|x_0\| \cdot \|\mathscr{S}\|.$$

Hence, $(f_1^2, f_2^2, \cdots, f_n^2)$ with $f_i^2 \in [\mathscr{L}, R^1]$ defined by

$$f_i^2(W) = \mathscr{S}(f_i(W)x_0), \tag{1.40}$$

$i = 1, 2, \cdots, n$ is the gradient ∂F^2 of F^2 at X^0.

Finally, let us emphasize the fact that the sensitivity results following from our approach are stronger than those based on the traditional parametrization of system elements. Indeed, suppose that an operator $X_k^0 \in [H, H]$ describing a nomimal system element is replaced by a function $X_k(\cdot): (-a, a) \to [H, H]$ satisfying the condition $X_k(0) = X_k^0$. Then the values $X_k(\alpha)$ fill out only a certain subset of the ball $B_r(X_k^0) \subset [H, H]$, which may reduce to a segment, if $X_k(\alpha)$ is linear in α.

On the other hand, in our approach the allowed variations $X_k$of X_k^0 fill out the entire ball $B_r(X_k^0)$. Thus, our measure of sensitivity takes into account all posssible increments of X_k^0, not only those restricted to segments or the like, and consequently, it provides more information than a measure using parametrization.

2. Applications in network theory.

A. As a first example, let us consider the parallel connection N of 1-ports $N_1, N_2, \cdots, N_n$. Assume that the 1-port N_i has an impedance $X_i^0 \in [H, H]$, $i = 1, 2, \cdots, n$, where H is a fixed Banach space. Our objective is to study the sensitivity of the impedance of N to variations of impedances X_i^0. We are going to show that the following assertion is true:

THEOREM 2.1. *Let* ${}^0X_i \in [H, H]$, $i = 1, 2, \cdots, n$ *be such that*

(a) ${}^0X_i^{-1}$ *exists for each* i,

(b) $Z_0 = (\sum_{i=1}^n {}^0X_i^{-1})^{-1}$ *exists.*

Then there exists $r^* > 0$ *such that*

(i) *the mapping*

$$U(X) = \left(\sum_{i=1}^n X_i^{-1} \right)^{-1} \tag{2.1}$$

is well-defined on $B_{r^*}(X^0)$ *with* $X^0 = ({}^0X_1, {}^0X_2, \cdots, {}^0X_n) \in [H, H]^n$,

(ii) *U is differentiable at X^0 and $\partial U = (u_1, u_2, \cdots, u_n)$, where $u_i \in [[H, H], [H, H]]$ is given by*

$$u_i(W) = Z_0{}^0X_i^{-1}\, W\, {}^0X_i^{-1}Z_0 \tag{2.2}$$

for each $W \in [H, H]$ and $i = 1, 2, \cdots, n$,

(iii) *$\Phi^*: B_{r^*}(0) \to R^1$, defined by*

$$\Phi^*(Z) = 2\|Z_0\|^2\{4nA^3 + \|Z_0\|(\sqrt{n}A^2 + 1)^2\}\|Z\|' \tag{2.3}$$

with $A = \max_{j=1,\cdots,n} \|{}^0X_j^{-1}\|$ is an error function of U at $B_{r^*}(0)$.

Proof. Let $\mathscr{L} = [H, H]$, and, for $j = 1, 2 \cdots, n$, let F_j: $\mathscr{L}^n \to \mathscr{L}$ be defined by $F_j(X) = X_j$. Referring to the example of a mapping defined by (1.9) it follows that F_j is differentiable at X^0 and $\partial F_j = (0, 0, \cdots, 0, I, 0, \cdots, 0)$, where I is the identity in $[\mathscr{L}, \mathscr{L}]$. Also, $\Phi_j \equiv 0$ is the error function of F_j on $B_r(0)$ with any $r > 0$. Note that $\|\partial F_j\| = 1$.

Next, invoking Theorem 1.4 we see that, for $\alpha = \frac{1}{2}$, $F_j(X)$ is invertible for each $X \in B_{r^*_j}(X^0)$, where by (1.27) $r_j^* = \frac{1}{4}\|{}^0X_j^{-1}\|^{-1}$. Thus, putting

$$G_j(X) = [F_j(X)]^{-1} \tag{2.4}$$

for $X \in B_{r^*_j}(X^0)$, it follows that G_j is differentiable at X^0 and $\partial G_j = (0, 0, \cdots, 0, g_j, 0, \cdots, 0)$, where

$$g_j(W) = -{}^0X_j^{-1}W\, {}^0X_j^{-1} \tag{2.5}$$

for each $W \in \mathscr{L}$. Moreover, by (1.28),

$$\Phi_j^*(Z) = 8\|{}^0X_j^{-1}\|^3\|Z\|' \tag{2.6}$$

is an error function of G_j on $B_{r^*_j}(0)$.

Now, putting $\xi_j = \|{}^0X_j^{-1}\|$, $j = 1, 2, \cdots, n$ and $A = \max_j \xi_j$, it follows that with $r_0 = \frac{1}{4}A^{-1}$ the mapping

$$G = \sum_{j=1}^n G_j \tag{2.7}$$

is defined on $B_{r_0}(X^0)$. Hence, by Theorem 1.1, G is differentiable at X^0, and $\partial G = \sum_{j=1}^n \partial G_j$, i.e., $\partial G = (h_1, h_2, \cdots, h_n)$, where

$$h_j(W) = -{}^0X_j^{-1}W\, {}^0X_j^{-1} \tag{2.8}$$

for $j = 1, 2, \cdots, n$ and $W \in \mathscr{L}$. Also, by (2.6),

$$\Phi_0(Z) = 8 \sum_{j=1}^n \xi_j^3 \cdot \|Z\|' \tag{2.9}$$

is an error function of G on $B_{r_0}(0)$. Note that if we put $\rho = \min[r_0, (8\sum_{j=1}^n \xi_j^3)^{-1}]$, then $0 \leqq \Phi(Z) \leqq 1$ on $B_\rho(0)$.

Finally, since $G(X^0) = \sum_{j=1}^n {}^0X_j^{-1}$ is invertible by our hypothesis, by Theorem 1.4 there exists $r^* > 0$ (given by (1.27) with r and F replaced by ρ and G, respectively) such that $G(X)$ is invertible for each $X \in B_{r^*}(0)$. Thus, putting $U(X) = [G(X)]^{-1}$ for $X \in B_{r^*}(X^0)$, it follows that U is differentiable at X^0, and by (1.26), $\partial U = (u_1, u_2, \cdots, u_n)$, where $u_j(W) = [G(X^0)]^{-1}\, {}^0X_j^{-1}\, W\, {}^0X_j^{-1}\, [G(X^0)]^{-1} = Z_0{}^0X_j^{-1}\, W\, {}^0X_j^{-1}Z_0$ for $j = 1, 2, \cdots, n$ and $W \in \mathscr{L}$. Hence, (2.2) is confirmed.

Invoking (1.28) and (2.9), we see that $\Phi^+(Z) = \|Z_0\|^2\{8\sum_{j=1}^n \xi_j^3 + 2\|Z_0\|(\|\partial G\| + 1)^2\} \cdot \|Z\|'$ is an error function of U on $B_{r^*}(0)$. Since $\|\partial G\| \leqq (\sum_{j=1}^n \|{}^0X_j^{-1}\|^4)^{1/2} \leqq \sqrt{n}A^2$ and $\sum_{j=1}^n \xi_j^3 \leqq nA^3$, it follows that $\Phi^+ \leqq \Phi^*$ with Φ^*

being given by (2.3). Hence, Φ^* is an error function of U on $B_{r^*}(0)$ and the proof is complete.

B. Consider now a finite network N, whose variables (voltages and currents) are in a fixed Hilbert space H with inner product $\langle \cdot, \cdot \rangle$. Let the structure of N be described by a finite oriented graph G having $c_2 \geqq 2$ branches, which contains at least one loop, and let the element in branch b_j, $j = 1, 2, \cdots, c_2$ be described by an operator ${}^0X_j \in [H, H]$. Also, assume that there are no mutual couplings between individual branches. Thus, our network N is modeled by a Hilbert network $({}^0\hat{Z}, G)$ over H [3], where the operator ${}^0\hat{Z}: H^{c_2} \to H^{c_2}$ is defined by

$$
{}^0\hat{Z}x = ({}^0X_1x_1, {}^0X_2x_2, \cdots, {}^0X_{c_2}x_{c_2}). \tag{2.10}
$$

Again, our objective will be to study the sensitivity of the admittance of N to variations of elements in branches. (Note that by the admittance A of N we mean the operator $A: H^{c_2} \to H^{c_2}$ which carries a c_2-vector e of branch voltages into a c_2-vector i of branch currents.)

Before solving our task, let us present certain results on Hilbert networks which will be needed.

If d is the $c_2 \times c_1$ incidence matrix of G (rows and columns of d correspond to branches and nodes of G, respectively), let $\tilde{Y}$ be a $c_2 \times c_0$ matrix, whose columns constitute an orthonormal basis in the solution space of the equation $d^T \cdot \xi = 0$, $\xi \in R^{c_2}$. (Note that c_0 is equal to the number of loops in any complete set of linearly independent loops in G). Also, let the operator $Y: H^{c_0} \to H^{c_2}$ be defined by $Yx = \tilde{Y} \cdot [x_k]$, $x = [x_k] \in H^{c_0}$.

Moreover, if the linear space H^c, $c \geqq 1$ is equipped with the inner product $\langle x, y\rangle' = \sum_{j=1}^n \langle x_j, y_j\rangle$, then H^c is again a Hilbert space, and we can show [3] that $Y \in [H^{c_0}, H^{c_2}]$ and $\|Y\| = 1$.

Also, it is known that the admittance A of N exists $\Leftrightarrow$ the operator $Y^*\,{}^0\hat{Z}Y: H^{c_0} \to H^{c_0}$ is invertible. (Here, $Y^* \in [H^{c_2}, H^{c_0}]$ is the adjoint of Y). In this case,

$$
A = Y(Y^*\,{}^0\hat{Z}Y)^{-1}Y^*. \tag{2.11}
$$

Having assembled these preliminaries, we can state a result on sensitivity of A.

THEOREM 2.2. *Let $Y \in [H^{c_0}, H^{c_2}]$ have the same meaning as above, and let $0 < \alpha < 1$. For each $X = (X_1, X_2, \cdots, X_{c_2}) \in [H, H]^{c_2}$, let the operator $F(X) \in [H^{c_2}, H^{c_2}]$ be defined by*

$$
F(X)x = (X_1x_1, X_2x_2, \cdots, X_{c_2}x_{c_2}). \tag{2.12}
$$

Assume that

(a) ${}^0X_i \in [H, H]$, $j = 1, 2, \cdots, c_2$,

(b) *the operator* $Q_0 = Y^*F(X^0)Y \in [H^{c_2}, H^{c_2}]$ *with* $X^0 = ({}^0X_1, {}^0X_2, \cdots, {}^0X_{c_2}) \in [H, H]^{c_2}$ *is invertible. If*

$$
r_\alpha = \alpha(\sqrt{c_2} + 1)^{-1}\|Q_0^{-1}\|, \tag{2.13}
$$

then

(i) *the mapping* $A: B_{r_\alpha}(X^0) \to [H^{c_2}, H^{c_2}]$ *defined by*

$$
A(X) = Y(Y^*F(X)Y)^{-1}Y^* \tag{2.14}
$$

is differentiable at X^0,

(ii) *the gradient* $\partial A = (a_1, a_2, \cdots, a_{c_2}) \in [[H, H], [H^{c_2}, H^{c_2}]]^{c_2}$ *of* A *at* X^0 *is given by*

$$a_j(W) = -A_0(E_jW)A_0 \tag{2.15}$$

for $j = 1, 2, \cdots, c_2$ *and* $W \in [H, H]$, *where*

$$A_0 = YQ_0^{-1}Y^* \tag{2.16}$$

and $E_j \in [[H, H], [H^{c_2}, H^{c_2}]]$ *is defined by*

$$E_jx = (0, 0, \cdots, Wx, 0, \cdots, 0) \tag{2.17}$$

with Wx *standing at the j-th place,*

(iii) Φ_A *defined by*

$$\Phi_A(Z) = (1-\alpha)^{-1}(\sqrt{c_2}+1)^2\|Q_0^{-1}\|^3\|Z\|'$$

is an error function of A *on* $B_{r_\alpha}(0)$.

Proof. From the definition (2.12) of F it follows readily that truly $F(X) \in [H^{c_2}, H^{c_2}]$ for each $X \in [H, H]^{c_2}$, that F is differentiable at X^0, and the gradient $\partial F = (f_1, f_2, \cdots, f_{c_2}) \in [[H, H], [H^{c_2}, H^{c_2}]]^{c_2}$ at X^0 is given by

$$f_j = E_j. \tag{2.18}$$

Also, $\Phi_F \equiv 0$ is the error function of F on $B_r(0)$ with any $r > 0$.

Define now the mapping Q: $[H, H]^{c_2} \to [H^{c_0}, H^{c_0}]$ by

$$Q(X) = Y^*F(X)Y. \tag{2.19}$$

Invoking Theorem 1.2, it follows that Q is differentiable at X^0, and for $\partial Q = (q_1, q_2, \cdots, q_{c_2}) \in [[H, H], [H^{c_0}, H^{c_0}]]^{c_2}$ we have

$$q_j(W) = Y^*(E_jW)Y. \tag{2.20}$$

Also, $\Phi_F \equiv 0$ is the error function of F on $B_r(0)$ with any $r > 0$.

Next, by (2.20) we clearly have $\|q_j(W)\| \leqq \|Y^*\| \cdot \|E_jW\| \cdot \|Y\| \leqq \|W\|$, so that $\|q_j\| \leqq 1$. Hence, $\|\partial Q\| \leqq \sqrt{c_2}$.

Referring to Theorem 1.4 it follows that, with

$$\bar{r}_\alpha = \alpha(\|\partial Q\|+1)^{-1}\|Q_0^{-1}\|^{-1}, \tag{2.21}$$

$Q(X)$ is invertible for each $X \in B_{\bar{r}_\alpha}(X^0)$, and the mapping K: $B_{\bar{r}_\alpha}(X^0) \to [H^{c_0}, H^{c_0}]$, defined by

$$K(X) = [Q(X)]^{-1}, \tag{2.22}$$

is differentiable at X^0. Moreover, for the gradient $\partial K = (k_1, k_2, \cdots, k_{c_2}) \in [[H, H], [H^{c_0}, H^{c_0}]]^{c_2}$ we have by (1.26) and (2.20),

$$k_j(W) = -Q_0^{-1}Y^*(E_jW)YQ_0^{-1}. \tag{2.23}$$

Also, by (1.28)

$$\Phi_K(Z) = (1-\alpha)^{-1}\|Q_0^{-1}\|^3(\|\partial Q\|+1)^2\|Z\|' \tag{2.24}$$

is an error function of K on $B_{\bar{r}_\alpha}(0)$.

However, since $\|\partial Q\| \leqq \sqrt{c_2}$, we have by (2.21),

$$r_\alpha = \alpha(\sqrt{c_2}+1)^{-1}\|Q_0^{-1}\|^{-1} \geqq \bar{r}_\alpha. \tag{2.25}$$

Consequently, the function Φ_K^1 defined by

$$\Phi_K^1(Z) = (1-\alpha)^{-1}\|Q_0^{-1}\|^3(\sqrt{c_2}+1)^2\|Z\|^r \tag{2.26}$$

is an error function of K on $B_{r_\alpha}(0)$.

Finally, let $A\colon B_{r_\alpha}(X^0) \to [H^{c_2}, H^{c_2}]$ be defined by

$$A(X) = YK(X)Y^*. \tag{2.27}$$

Then, by Theorem 1.2, A is differentiable at X^0, and by (2.23), the gradient $\partial A = (a_1, a_2, \cdots, a_{c_2}) \in [[H, H], [H^{c_2}, H^{c_2}]]^{c_2}$ is given by

$$a_j(W) = -YQ_0^{-1}Y^*(E_jW)YQ_0^{-1}Y^* = -A_0(E_jW)A_0 \tag{2.28}$$

(see (2.16)) for $j = 1, 2, \cdots, c_2$ and all $W \in [H, H]$. Hence, claims (i) and (ii) hold.

Moreover, the function $\Phi_A = \|Y\| \cdot \|Y^*\|\Phi_K^1 = \Phi_K^1$ is an error function of A on $B_{r_\alpha}(0)$. Thus, (iii) holds by virtue of (2.26) and the proof is complete.

The Theorem 2.2 has the following interesting interpretation which generalizes the result given in [2, paragraph 3.6].

Let $e_0 \in H^{c_2}$ be a fixed nominal voltage vector which produces a nominal current vector $i_0 = A_0e_0 \in H^{c_2}$ in our network N. Assume that we are interested in the sensitivity of i_0 to variations of the element in the jth branch, i.e., variations of X_j^0. Thus, we have to consider the mapping $i\colon B_{r_\alpha} \to H^{c_2}$ defined by $i(X) = A(X)e_0$.

Recalling (1.39) and (2.15) we see that for the gradient $\partial i = (i_1^*, i_2^*, \cdots, i_{c_2}^*)$ we have

$$i_j^*(W) = -A_0(E_jW)A_0e_0 \tag{2.29}$$

for each $W \in [H, H]$. Hence, if the impedance X_j^0 is changed by increment W, then the nominal current vector i_0 changes by $i_j^*(W)$ with an accuracy better than $\|W\|\Phi(Z)$, where $Z = (0, 0, \cdots, 0, W, 0, \cdots, 0)$.

Now, from (2.29) it is easy to see that $i_j^*(W)$ is the current vector in N which corresponds to the voltage vector $(0, 0, \cdots, 0, -Wi_j^0, 0, \cdots, 0)$, where i_j^0 is the jth component of the nominal current vector i_0. Thus, using physical terminology, the change $i_j^*(W)$ is given by the current distribution in N corresponding to the excitation by a single voltage source $-Wi_j^0$ inserted into the jth branch, provided W is sufficiently small.

3. Applications in the theory of feedback systems. In this part we will briefly discuss applications of our results for establishing the sensitivity of input-output characteristics of a linear feedback system given in Fig. 1.

Let H be a fixed linear space, and let $A_1, A_2\colon H \to H$ be linear operators. If $(u_1, u_2) \in H^2$ (inputs), then a pair $(e_1, e_2) \in H^2$ (errors) will be called a solution of the feedback system $\{A_1, A_2\}$ corresponding to (u_1, u_2), if the equations

$$\begin{aligned} e_1 &= u_1 - y_2, \qquad & y_1 &= A_1e_1, \\ e_2 &= u_2 + y_1, \qquad & y_2 &= A_2e_2 \end{aligned} \tag{3.1}$$

are satisfied. This fact will be symbolized by writing $(u_1, u_2) \to (e_1, e_2)$.

The following result is well known [4], [5]: For any $(u_1, u_2) \in H^2$ there exists a unique solution $(e_1, e_2) \in H^2$ of $\{A_1, A_2\}$ corresponding to $(u_1, u_2) \Leftrightarrow$ the operator $N = I + A_2A_1$ is invertible. In this case,

$$(e_1, e_2) = (N^{-1}(u_1 - A_2u_2), u_2 + A_1N^{-1}(u_1 - A_2u_2)). \tag{3.2}$$

Moreover, we have the assertion [6, p. 315]:

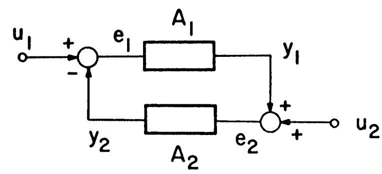

FIG. 1

Let $M = I + A_1A_2$; then M is invertible$\Leftrightarrow N$ is invertible. In this case,

$$A_1N^{-1} = M^{-1}A_1, \qquad N^{-1}A_2 = A_2M^{-1}. \tag{3.3}$$

Using this fact and (3.2), we easily confirm the following proposition:

(i) *If* $(u_1, 0) \to (e_1, e_2)$, *then*

$$(e_1, e_2) = (N^{-1}u_1, A_1N^{-1}u_1). \tag{3.4}$$

(ii) *If* $(0, u_2) \to (e_1', e_2')$, *then*

$$(e_1', e_2') = (-A_2M^{-1}u_2, M^{-1}u_2). \tag{3.5}$$

Since our feedback system $\{A_1, A_2\}$ is linear and enjoys a symmetry exhibited by equations (3.4) and (3.5), we can abridge our attention to transmission characteristics given by operators N^{-1} and A_1N^{-1}. Thus, turning to our sensitivity problem, let us prove the following assertion.

THEOREM 3.1. *Let H be a Banach space, let $A_1, A_2 \in [H, H]$, $a_j = \|A_j\|$, $j = 1, 2$, and let $0 < \alpha < 1$. Furthermore, assume that the operator $N = I + A_2A_1$ is invertible, and let*

$$r_\alpha = \min[1, \alpha(a_1 + a_2 + 1)^{-1}\|N^{-1}\|^{-1}]. \tag{3.6}$$

Then the operator $I + X_2X_1$ is invertible for every $X = (X_1, X_2) \in B_{r_\alpha}(X^0)$ with $X^0 = (A_1, A_2)$, and the mapping E_1: $B_{r_\alpha}(X^0) \to [H, H]$ defined by

$$E_1(X) = (I + X_2X_1)^{-1} \tag{3.7}$$

is differentiable at X^0. If $\partial E_1 = (k_1, k_2) \in [[H, H], [H, H]]^2$, then

$$k_1(W) = -N^{-1}A_2WN^{-1}, \qquad k_2(W) = -N^{-1}WA_1N^{-1} \tag{3.8}$$

for each $W \in [H, H]$. Also,

$$\Phi_1(Z) = \|N^{-1}\|^2\{1 + (1-\alpha)^{-1}(a_1 + a_2 + 1)^2\|N^{-1}\|\}\|Z\|' \tag{3.9}$$

is an error function of E_1 on $B_{r_\alpha}(0)$.

Moreover, if E_2: $B_{r_\alpha}(X^0) \to [H, H]$ is defined by

$$E_2(X) = X_1(I + X_2X_1)^{-1}, \tag{3.10}$$

then E_2 is differentiable at X^0, and for the gradient $\partial E_2 = (h_1, h_2) \in [[H, H], [H, H]]^2$ we have

$$h_1(W) = M^{-1}WN^{-1}, \qquad h_2(W) = -A_1N^{-1}WA_1N^{-1} \tag{3.11}$$

for all $W \in [H, H]$, where $M = I + A_1A_2$. Also,

$$\Phi_2(Z) = \|N^{-1}\|^2\{a_1 + a_2 + (1 + a_1)[1 + (1-\alpha)^{-1}\|N^{-1}\|(a_1 + a_2 + 1)^2]\}\|Z\|' \tag{3.12}$$

is an error function of E_2 on $B_{r_\alpha}(0)$.

Proof. Let $F^j(X)=X_j$, $j=1,2$. Then F^j is differentiable at X^0 and $\partial F^1=(I,0)$, $\partial F^2=(0,I)$, where I is the identity operator in $[[H,H],[H,H]]$. Thus, $\|\partial F^1\|=\|\partial F^2\|=1$. Also, $\Phi\equiv 0$ is the error function of both F^1 and F^2 on $B_r(0)$ with any $r>0$.

Next, put $\tilde{F}=F^2\circ F^1$. By Theorem 1.3, $\tilde{F}$ is differentiable at X^0, and by (1.14) we have for $\partial\tilde{F}=(g_1,g_2)$,

$$g_1(W)=A_2W,\qquad g_2(W)=WA_1 \tag{3.13}$$

for all $W\in[H,H]$. Also, by (1.15), $\tilde{\Phi}(Z)=\|\partial F^1\|\cdot\|\partial F^2\|\cdot\|Z\|'=\|Z\|'$ is an error function of $\tilde{F}$ on $B_r(0)$ with any $r>0$.

Putting $F(X)=I+F^2(X)F^1(X)$ with $I\in[H,H]$, then clearly $\partial F=\partial\tilde{F}$ and the above $\tilde{\Phi}$ is an error function of F on $B_r(0)$ with any $r>0$. Also, observe that, by (3.13), $\|g_1\|\leqq\|A_2\|=a_2$, $\|g_2\|\leqq\|A_1\|=a_1$, and consequently, $\|\partial F\|\leqq(a_1^2+a_2^2)^{1/2}\leqq a_1+a_2$.

Now, let $r=1$. Then $0\leqq\tilde{\Phi}(Z)\leqq 1$ for $Z\in B_1(0)$. Since $F(X^0)=N$ is invertible, Theorem 1.4 shows that $F(X)$ is invertible for each $X\in B_{r^*}(X^0)$, where

$$r^*=\min[1,\alpha(\|\partial F\|+1)^{-1}\|N^{-1}\|^{-1}]. \tag{3.14}$$

Hence, our mapping E_1: $B_{r^*}(X^0)\to[H,H]$ is differentiable at X^0, and for $\partial E_1=(k_1,k_2)$ we have (3.8) by virtue of (1.26). Also, by (1.28),

$$\Phi^*(Z)=\|N^{-1}\|^2\{1+(1-\alpha)^{-1}\|N^{-1}\|(\|\partial F\|+1)^2\}\|Z\|' \tag{3.15}$$

is an error function of E_1 on $B_{r^*}(0)$.

However, since $\|\partial F\|\leqq a_1+a_2$, we have $r^*\geqq r_\alpha$ with r_α being defined by (3.6). Also, by (3.15) and (3.9), $\Phi^*\leqq\Phi_1$ on $B_{r_\alpha}(0)$. Consequently, Φ_1 is an error function of E_1 on $B_{r_\alpha}(0)$ and the first part of our theorem is proven.

Finally, define E_2: $B_{r_\alpha}(X^0)\to[H,H]$ by $E_2(X)=F^1(X)E_1(X)=X_1(I+X_2X_1)^{-1}$. Invoking again Theorem 1.3 it follows that E_2 is differentiable at X^0, and for $\partial E_2=(h_1,h_2)$ we have by (1.14) and (3.3),

$$\begin{aligned}
h_1(W)&=WE_1(X^0)+F^1(X^0)k_1(W)\\
&=WN^{-1}-A_1N^{-1}A_2WN^{-1}\\
&=(I-A_1N^{-1}A_2)WN^{-1}\\
&=(I-M^{-1}A_1A_2)WN^{-1}\\
&=M^{-1}(M-A_1A_2)WN^{-1}=M^{-1}WN^{-1}.
\end{aligned}$$

Similarly,

$$h_2(W)=F^1(X^0)k_2(W)=-A_1N^{-1}WA_1N^{-1},$$

which confirms the formulas (3.11).

Moreover, by (1.15),

$$\Phi_2^*(Z)=\|A_1\|\Phi_1(Z)+\|\partial F^1\|(\|\partial E_1\|+\Phi_1(Z))\|Z\|' \tag{3.16}$$

is an error function of E_2 on $B_{r_\alpha}(0)$. However, since

$$\|\partial F^1\|=1,\qquad \|\partial E_1\|\leqq\|N^{-1}\|^2(\|A_1\|^2+\|A_2\|^2)^{1/2}\leqq\|N^{-1}\|^2(a_1+a_2)$$

and $\|Z\|'\leqq 1$ for $Z\in B_{r_\alpha}(0)$, we see readily that $\Phi_2^*\leqq\Phi_2$ on $B_{r_\alpha}(0)$ with Φ_2 given by (3.12). Hence, Φ_2 is an error function of E_2 on $B_{r_\alpha}(0)$, and our theorem is proven.

The gradients ∂E_1 and ∂E_2 appear as transmission characteristics of input-output systems built from two original feedback systems $\{A_1,A_2\}$. Indeed, consider, for

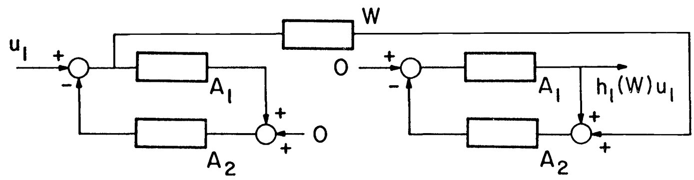

FIG. 2

example, the gradient $\partial E_2 = (h_1, h_2)$. Recalling the formulas (3.11) and (3.4), (3.5), we see easily that the operator $h_1(W) \in [H, H]$ is the input–output characteristic of a system whose block diagram is given in Fig. 2. Similarly, $h_2(W)$ has a meaning apparent from Fig. 3.

Analogous block diagrams can be constructed for the entries k_1, k_2 of ∂E_1.

Thus, these facts generalize the results given in [2, paragraph 6.8].

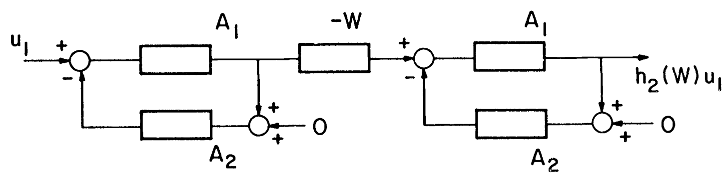

FIG. 3

4. Comments. The sensitivity analysis developed above can be simplified (and mechanized, too) considerably, if the variations of the system elements are assumed to be very small. In other words, under this assumption the quantity $\sum_{i=1}^{n} f_i(Z_i)$ appearing in (1.3) is accepted as a sufficiently good approximation to $F(X^0+Z)-F(X^0)$, and consequently, the value of $\|Z\|'\Phi(Z)$ is neglected. Also, the actual size of the ball $B_r(X^0)$ is immaterial.

Thus, in this case, the gradient ∂F is the only thing that matters. However, finding ∂F is a relatively easy task, provided F has a particular form currently encountered in systems analysis. To be more specific, let us carry out the following consideration.

First of all, referring to the Definition 1.2 of a derivative and Theorem 1.5, we see that our results can be summarized and written in a perhaps more tangible form as follows:

Rule 1. Let F and G be defined on a ball $B_r(X^0) \subset [H, H]^n$, and be differentiable at X^0. If a, b are scalars and A, B are "constant" operators, then the mappings $aF + bG$, AFB, $F \circ G$ are defined on $B_r(X^0)$, differentiable at X^0 and we have

(i) $\delta_i(aF+bG)(\cdot) = a(\delta_i F)(\cdot) + b(\delta_i G)(\cdot)$,

(ii) $\delta_i(AFB)(\cdot) = A(\delta_i F)(\cdot)B$,

(iii) $\delta_i(F \circ G)(\cdot) = (\delta_i F)(\cdot)G(X^0) + F(X^0)(\delta_i G)(\cdot)$.

If, in addition, $F(X^0)$ is invertible, then the mapping $[F]^{-1}(X) = (F(X))^{-1}$ is defined on a ball $B_{\bar{r}}(X^0)$ with $0 < \bar{r} \leqq r$, is differentiable at X^0 and we have

(iv) $\delta_i[F]^{-1}(\cdot) = -[F(X^0)]^{-1}(\delta_i F)(\cdot)[F(X^0)]^{-1}$.

Let us point out the essential fact that forming linear combinations, products and inverses (under the assumption of invertibility at X^0) of mappings differentiable at X^0 yields again mappings differentiable at X^0. On the other hand, the elementary mappings F^j, $j = 1, 2, \cdots, n$ defined by $F^j(X) = X_j$, are differentiable at any point X^0.

Combining these facts, we are led to the following useful result:

Rule 2. Let $X^0 = (X_1^0, X_2^0, \cdots, X_n^0) \in [H, H]^n$, and assume that $F(X^0)$ is obtained from operators $X_1^0, X_2^0, \cdots, X_n^0$ by forming finitely many linear combinations, products and inverses. (We understand that these inverses exist.) Define $F(X)$ for $X=(X_1, X_2, \cdots, X_n)$ by formally replacing X_j^0 in $F(X^0)$ by X_j, $j=1, 2, \cdots, n$. Then there exists $r>0$ such that $F(X)$ is meaningful for each $X \in B_r(X^0) \subset [H, H]^n$, (i.e., the inverses involved exist), and the mapping F is differentiable at X^0. Moreover, the gradient ∂F can be found by applying the above formulas (i) through (iv).

Thus, if the operators X_j^0, $j=1, 2, \cdots, n$ describe linear blocks which constitute a system S, and $F(X^0)$ is a performance characteristic of S, then $F(X^0)$ has the form considered in Rule 2, and consequently, F is differentiable at X^0.

To show how easily the Rule 1 and 2 can be applied, consider a 2-port P given in Fig. 4. Assume that $A_j \in [H, H]$, $j=1, 2, 3$ has the meaning of the admittance of the corresponding 1-port, and that the operator $A_1+A_2+A_3$ is invertible.

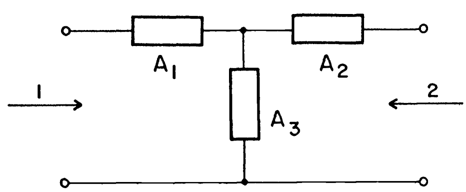

FIG. 4

A simple argument will convince us that the admittance matrix $A=[A_{ik}]$, $i, k = 1, 2$ of P is given by

$$
(4.1)\quad
\begin{aligned}
A_{11}&=A_1(A_1+A_2+A_3)^{-1}(A_2+A_3); \qquad A_{12}=-A_1(A_1+A_2+A_3)^{-1}A_2;\\
A_{21}&=-A_2(A_1+A_2+A_3)^{-1}A_1; \qquad A_{22}=A_2(A_1+A_2+A_3)^{-1}(A_1+A_3).
\end{aligned}
$$

Suppose that we want to establish the sensitivities of entries A_{ik} to variations of admittance A_3. Referring to Rule 2, put $F(X)=X_1+X_2+X_3$, $A_0=A_1+A_2+A_3$ and

$$
(4.2)\quad
\begin{aligned}
G_{11}(X)&=X_1[F(X)]^{-1}(X_2+X_3), \qquad G_{12}(X)=-X_1[F(X)]^{-1}X_2,\\
G_{21}(X)&=-X_2[F(X)]^{-1}X_1, \qquad G_{22}(X)=X_2[F(X)]^{-1}(X_1+X_3).
\end{aligned}
$$

Since we are interested in $\delta_3 G_{ik}$, we treat X_1, X_2 as constants while applying Rule 1. Thus, we have by (iv),

$$
(4.3)\qquad \delta_3[F]^{-1}(W)=-[F(X^0)]^{-1}(\delta_3 F)(W)[F(X^0)]^{-1}=-A_0^{-1}WA_0^{-1}.
$$

Similarly, by (4.2) and (iii),

$$
\begin{aligned}
\delta_3 G_{11}(W)&=A_1\{\delta_3[F]^{-1}(W)(A_2+A_3)+A_0^{-1}\delta_3(X_2+X_3)(W)\}\\
&=A_1\{-A_0^{-1}WA_0^{-1}(A_2+A_3)+A_0^{-1}W\}=A_1A_0^{-1}WA_0^{-1}A_1.
\end{aligned}
$$

Also, by (ii),

$$
\delta_3 G_{12}(W)=-A_1\delta_3[F]^{-1}(W)A_2=A_1A_0^{-1}WA_0^{-1}A_2.
$$

A further easy calculation yields

$$
\begin{aligned}
\delta_3 G_{21}(W)&=A_2A_0^{-1}WA_0^{-1}A_1,\\
\delta_3 G_{22}(W)&=A_2A_0^{-1}WA_0^{-1}A_2.
\end{aligned}
$$

As a final example, let us reconsider the mapping $E_2(X)=X_1(I+X_2X_1)^{-1}$ we discussed in Theorem 3.1. We get readily by Rule 1,

$$\begin{aligned}\delta_1E_2(W)&=\delta_1X_1(W)N^{-1}+A_1\delta_1(I+X_2X_1)^{-1}\\&=WN^{-1}-A_1N^{-1}\delta_1(I+X_2X_1)(W)N^{-1}\\&=WN^{-1}-A_1N^{-1}A_2WN^{-1}=M^{-1}WN^{-1},\end{aligned}$$

and

$$\begin{aligned}\delta_2E_2(W)&=A_1\delta_2(I+X_2X_1)^{-1}(W)\\&=-A_1N^{-1}\delta_2(I+X_2X_1)(W)N^{-1}=A_1N^{-1}WA_1N^{-1}\end{aligned}$$

which agrees with (3.11).

REFERENCES

[1] P. V. KOKOTOVIĆ AND R. S. RUTMAN, *Sensitivity of automatic control systems*, Automat. Remote Control, 26 (1965), pp. 727–749.

[2] W. R. PERKINS AND J. B. CRUZ JR., *Engineering of Dynamic Systems*, John Wiley, New York, 1969.

[3] V. DOLEZAL, *Basic properties of Hilbert networks*, IEEE Trans. Circuits and Systems, CAS-23 (1976), pp. 490–497.

[4] ———, *Feedback systems described by monotone operators*, SIAM J. Control, 17 (1979), pp. 339–364.

[5] C. A. DESOER AND M. VIDYASAGAR, *Feedback Systems: Input–Output Properties*, Academic Press, New York, 1975.

[6] G. F. SIMMONS, *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0005 $01.00/0

ON THE ASYMPTOTIC SPECTRUM OF A CONVOLUTION*

OLOF J. STAFFANS†

Abstract. Let B be a reflexive Banach space, and let $L^2_\infty(R;B)$ denote the set of locally square integrable, B-valued functions on R whose L^2-norm over intervals of length one is bounded. We show that the asymptotic spectrum (i.e. the spectrum of the limit set) of the convolution $\mu * \varphi$ of a finite, scalar-valued measure μ with a function $\varphi \in L^2_\infty(R;B)$ is contained in the asymptotic spectrum of φ. This work is motivated by a recent result on the asymptotic behavior of the bounded solutions of a B-valued, nonlinear Volterra equation, and it strengthens that result.

1. Introduction. In [8] and [9] the author studied an abstract nonlinear Volterra equation of the form

$$\begin{aligned} &x'(t)+\int_{[0,t]} \varphi(t-s)\,d\mu(s)=f(t) \qquad (t\in R^+);\\ &x(0)=x_0 \end{aligned} \tag{1.1}$$

(prime denotes differentiation). Here μ is a given measure, f is a given function, and φ depends in a nonlinear way on the solution x of (1.1) (one gets φ by composing x with a nonlinear function). The functions x, φ and f map $R^+=[0,\infty)$ into a reflexive Banach space B, and μ is a complex measure on R^+ (also operator-valued measures were allowed in [8]). We suppose that we have a solution x of (1.1) which is bounded on R^+, and for which φ in (1.1) is bounded in a sense to be made precise below. As a result of the work in [8] and [9], under certain additional assumptions one has some a priori information on the asymptotic behavior of the function φ in (1.1). The purpose of this work is to combine (1.1) with that information to obtain similar information on the asymptotic behavior of x. The results we obtain are extensions to Banach-space valued equations of the main theorems in [7].

The principal difference between this work and [7] (apart from the fact that we now work in a Banach space) is that we give up the continuous dependence of φ on x in (1.1), which played a crucial role in [7]. As a consequence we have no longer the limit equations derived in [7] at our disposal. This turns out to be only a minor inconvenience, and the main conclusions of this paper, specialized to the scalar-valued case, are only slightly weaker than those of [7] (see § 6).

Here we suppose throughout that the kernel μ in (1.1) is a finite measure. In [7] a somewhat more general kernel was allowed (the sum of a finite measure and a function of bounded variation).

2. Preliminaries. We first recall some background material from [8] and [9].

Throughout we let B be a reflexive Banach space, and let $L^2_\infty(R;B)$ be the set of measurable, locally square-integrable, B-valued functions on R, whose L^2_∞-norm

$$\|\varphi\|_{2,\infty}=\sup_{n\in Z}\left[\int_n^{n+1}|\varphi(s)|^2\,ds\right]^{1/2}$$

is finite. The space $L^2_\infty(R;B)$ is the dual of a Banach space, and most of the time we give it its weak*-topology. We let τ be the left-translation operator: $\tau_t\varphi(s)=\varphi(s+t)$

* Received by the editors February 23, 1978.
† Institute of Mathematics, Helsinki University of Technology, Espoo, Finland.

$(s, t \in R)$. For each $\varphi \in L^2_\infty(R; B)$ we define the limit set $\Gamma(\varphi)$ by

$$\Gamma(\varphi) = \{\psi \in L^2_\infty(R; B) \mid \tau_{t_k}\varphi \to \psi \textit{ weak* in } L^2_\infty(R; B), \textit{for some sequence } t_k \to \infty\}.$$

The weak*-closure of the set $\{\tau_t\varphi \mid t \in R^+\}$ in $L^2_\infty(R; B)$ equals $\{\tau_t\varphi \mid t \in R^+\} \cup \Gamma(\varphi)$, and it is compact and metrizable in the induced weak*-topology.

We denote the distribution Fourier transform of a tempered (complex-valued or B-valued) distribution u by $\hat{u}$. We write $\sigma(u)$ for the support of $\hat{u}$, and call $\sigma(u)$ the spectrum of u. In particular, this is how one defines the spectrum of a function $\varphi \in L^2_\infty(R; B)$. The spectrum $\sigma(\Gamma(\varphi))$ of the limit set $\Gamma(\varphi)$ is defined by

$$\sigma(\Gamma(\varphi)) = \overline{\bigcup_{\psi \in \Gamma(\varphi)} \sigma(\psi)},$$

and it is called the asymptotic spectrum of φ.

The set of all (finite, complete) complex Borel measures on R is denoted $BM(R; C)$, and m stands for the Lebesgue measure (in integrals we replace $dm(t)$ by dt). We write $|\mu|$ for the total variation measure of μ, and $\|\mu\|$ for the total variation of μ on R (i.e. $\|\mu\| = |\mu|(R)$). The statement "for a.e. $[m]$ $t \in R$" should be read "for almost every $t \in R$", with the understanding that "almost every" refers to the Lebesgue measure.

The characteristic function of an interval I is denoted χ_I.

3. On convolutions. We next prove a sequence of auxiliary lemmas.

LEMMA 3.1. *Let $\mu \in BM(R; C)$, $\varphi \in L^2_\infty(R; B)$. Then for a.e. $[m]$ $t \in R$, the function $s \mapsto \varphi(t-s)$ is μ-integrable, and the convolution $\mu * \varphi$ defined a.e. $[m]$ by*

$$\mu * \varphi(t) = \int_R \varphi(t-s)\, d\mu(s)$$

belongs to $L^2_\infty(R; B)$.

Remark 3.1. For each $\mu \in BM(R; C)$ the convolution operator $\varphi \mapsto \mu * \varphi$ in fact maps all mixed L^p-spaces into themselves, but we need only the case mentioned in Lemma 3.1.

Proof of Lemma 3.1. We begin the proof by observing that the function $(t, s) \mapsto \varphi(t-s)$ is $m \otimes \mu$-measurable [8, Lemma 2.2] (here $m \otimes \mu$ is the completed product of m and μ); hence the function $s \mapsto \varphi(t-s)$ is μ-measurable for a.e. $[m]$ $t \in R$ [1, Prop. 22.10].

Before considering the general case we prove a simplified version of Lemma 3.1. Take $B = R$, and suppose that both μ and φ are (real-valued and) nonnegative. Then, by [2, Thm. 7.12], the function $\mu * \varphi$ is defined a.e. $[m]$ (with values in $R^+ \cup \{\infty\}$), and it is m-measurable. Fix $n \in Z$, and compute

$$\int_n^{n+1} \left[\int_R \varphi(t-s)\, d\mu(s)\right]^2 dt \leqq \|\mu\| \int_n^{n+1} \int_R [\varphi(t-s)]^2\, d\mu(s)\, dt$$
$$= \|\mu\| \int_R \int_n^{n+1} [\varphi(t-s)]^2\, dt\, d\mu(s) \leqq 2\|\mu\|^2 \|\varphi\|^2_{2,\infty},$$

where we have used the Schwarz inequality, Fubini's theorem [2, Thm. 7.12], and the obvious inequality

$$\int_n^{n+1} [\varphi(t-s)]^2\, dt \leqq 2\|\varphi\|^2_{2,\infty} \qquad (s \in R).$$

Thus in this case $\mu * \varphi \in L^2_\infty(R; R)$ (and in particular, $\mu * \varphi(t) < \infty$ a.e. $[m]$ so that $s \mapsto \varphi(t-s)$ is μ-integrable for a.e. $[m]$ $t \in R$).

Now consider the general case. It follows from the preceding argument combined with the Schwarz inequality that for each $n \in Z$ the function $(t, s) \mapsto \chi_{[n,n+1]}(t)\varphi(t-s)$ is $m \otimes \mu$-integrable (cf. [1, Thm. 11.10]). This implies that for a.e. $[m]$ $t \in [n, n+1]$ the function $s \mapsto \varphi(t-s)$ is μ-integrable, and that the convolution $\mu * \varphi$, restricted to $[n, n+1]$ is m-measurable [1, Thm. 22.12]. As n was arbitrary, $s \mapsto \varphi(t-s)$ is μ-integrable for a.e. $[m]$ $t \in R$, and $\mu * \varphi$ is m-measurable. Moreover, since obviously $|\mu * \varphi|(t) \leqq |\mu| * |\varphi|(t)$ whenever both sides make sense, one has $\|\mu * \varphi\|_{2,\infty} \leqq 2^{1/2}\|\mu\| \, \|\varphi\|_{2,\infty}$. This shows that $\mu * \varphi \in L^2_\infty(R; B)$, and completes the proof of Lemma 3.1.

LEMMA 3.2. *Let $\varphi \in L^2_\infty(R; C)$, $\eta \in \mathscr{S}(R; C)$. Then $\varphi * \eta \in BUC(R; C)$.*

Here $\mathscr{S}$ is the set of rapidly decreasing, C^∞-functions, and BUC is the set of bounded uniformly continuous functions. The convolution $\varphi * \eta$ can be interpreted as the convolution of a test function with a distribution, or alternatively, it can be defined by the absolutely convergent integral

$$\varphi * \eta(t) = \int_R \varphi(t-s)\eta(s)\, ds \qquad (t \in R).$$

Proof of Lemma 3.2. Using Hölder's inequality it is easy to show that $\varphi * \eta \in L^\infty(R; C)$. Applying the same argument to the derivative $(\varphi * \eta)' = \varphi * \eta'$ one concludes that $(\varphi * \eta)' \in L^\infty(R; C)$. Thus $\varphi * \eta$ is uniformly continuous, and the proof of Lemma 3.2 is complete.

LEMMA 3.3. *Let $\mu \in BM(R; C)$, $\varphi \in L^2_\infty(R; C)$, $\eta \in \mathscr{S}(R; C)$. Then $(\mu * \varphi) * \eta(t) = \mu * (\varphi * \eta)(t)$ $(t \in R)$.*

By Lemmas 3.1 and 3.2, both the double convolutions in Lemma 3.3 can be defined by absolutely convergent integrals:

$$(\mu * \varphi) * \eta(t) = \int_R (\mu * \varphi)(t-s)\eta(s)\, ds,$$

$$\mu * (\varphi * \eta)(t) = \int_R (\varphi * \eta)(t-s)\, d\mu(s).$$

To prove Lemma 3.3 it suffices to apply Fubini's theorem, justified by either Lemma 3.1 or 3.2.

4. The main result. The following two theorems are the key ingredients in the proofs of our new versions of [7, Thms. 3.1 and 3.2] (cf. § 6).

THEOREM 4.1. *Let $\mu \in BM(R; C)$, $\varphi \in L^2_\infty(R; B)$. Then $\sigma(\Gamma(\mu * \varphi)) \subset \sigma(\Gamma(\varphi))$.*

THEOREM 4.2. *Let $\mu \in BM(R; C)$, $\varphi \in L^2_\infty(R; B)$. Moreover, suppose that $\sigma(\Gamma(\varphi))$ is countable, and that $\hat{\mu}(\omega) = 0$ $(\omega \in \sigma(\Gamma(\varphi)))$. Then $\tau_t\mu * \varphi \to 0$ weak* in $L^2_\infty(R; B)$ as $t \to \infty$.*

By Lemma 3.1, $\mu * \varphi \in L^2_\infty(R; B)$, so that $\Gamma(\mu * \varphi)$ and $\sigma(\Gamma(\mu * \varphi))$ are defined.

The conclusion of Theorem 4.2 can be written in two more equivalent ways. Even under the hypothesis of Theorem 4.1 it is true that the distance from $\tau_t\mu * \varphi$ to $\Gamma(\mu * \varphi)$ tends to zero as $t \to \infty$ [9, Lemma 3.1]. Thus the statement $\tau_t\mu * \varphi \to 0$ weak* in $L^2_\infty(R; B)$ is equivalent to $\Gamma(\mu * \varphi) = \{0\}$, which in turn is equivalent to $\sigma(\Gamma(\mu * \varphi)) = \varnothing$. The last formulation is the most convenient one in the proof.

We begin the proofs of Theorems 4.1–4.2 by a reduction to the scalar case.

Proof of Theorems 4.1–4.2, *assuming them to hold when* $B = C$. Suppose that Theorems 4.1–4.2 hold when $B = C$, and let B be an arbitrary complex reflexive Banach space.

In the case of Theorem 4.1, put $A = \sigma(\Gamma(\varphi))$, and in the case of Theorem 4.2, put $A = \varnothing$. We have to show that $\sigma(\Gamma(\mu * \varphi)) \subset A$.

Let B' be the dual of B. For an arbitrary B-valued tempered distribution ξ one has

$$\sigma(\xi) = \overline{\bigcup_{\alpha \in B'} \sigma(\alpha\xi)},$$

where $\alpha\xi$ is the complex-valued distribution one gets by composing ξ with α (see [4, Chap. I, p. 61]). From this one concludes that

$$\sigma(\Gamma(\mu * \varphi)) = \overline{\bigcup_{\substack{\alpha \in B' \\ \xi \in \Gamma(\mu * \varphi)}} \sigma(\alpha\xi)}.$$

Thus, as A is closed, in order to prove Theorems 4.1–4.2 it suffices to show that for each $\alpha \in B'$, $\xi \in \Gamma(\mu * \varphi)$ one has $\sigma(\alpha\xi) \subset A$.

Fix $\alpha \in B'$, $\xi \in \Gamma(\mu * \varphi)$. Then by [8, Lemma 3.6], $\alpha\xi \in \Gamma(\alpha(\mu * \varphi))$. But $\mu * \varphi$ is defined a.e. $[m]$ by an absolutely convergent integral, and so $\alpha(\mu * \varphi) = \mu * \alpha\varphi$. In particular $\alpha\xi \in \Gamma(\mu * \alpha\varphi)$. We claim that $\sigma(\Gamma(\alpha\varphi)) \subset \sigma(\Gamma(\varphi))$. This is true because, by [8, Lemma 3.6] (or rather, its generalization to a reflexive Banach space), every $\psi_\alpha \in \Gamma(\alpha\varphi)$ is of the form $\alpha\psi$ with $\psi \in \Gamma(\varphi)$, and trivially $\sigma(\alpha\psi) \subset \sigma(\psi)$. Now we observe that we can apply the scalar versions of Theorems 4.1–4.2 (in particular, in Theorem 4.2 we have $\hat{\mu}(\omega) = 0$ for each $\omega \in \sigma(\Gamma(\alpha\varphi))$ since $\sigma(\Gamma(\alpha\varphi)) \subset \sigma(\Gamma(\varphi))$), and they give $\sigma(\alpha\xi) \subset A$. This completes the proof of the reduction to the scalar case.

5. Proof of the scalar case. The proofs of the scalar versions of Theorems 4.1–4.2 are based on

LEMMA 5.1. *Let* $\mu \in BM(R; C)$, $\varphi \in L^2_\infty(R; C)$. *Then* $\Gamma(\mu * \varphi) = \{\mu * \psi | \omega \in \Gamma(\varphi)\}$.

In the proof of Lemma 5.1 we make use of the fact that for $\mu \in BM(R; C)$ the convolution operator $\varphi \mapsto \mu * \varphi$ is weakly sequentially continuous from $\mathscr{B}'$ into itself ($\mathscr{B}'$ is the space of bounded distributions; see [3, p. 200] for the definition of $\mathscr{B}'$ as the dual of $\mathscr{D}_{L^1}$). The convolution of $\mu \in BM$ with $\varphi \in \mathscr{B}'$ is defined e.g. as in [3, Thm. XXVI, p. 203] (note that $BM \subset \mathscr{D}'_{L^1}$), and in the special case when $\varphi \in L^2_\infty(R; B)$ it coincides with the convolution defined in Lemma 3.1. One way to describe weak convergence of a sequence $\varphi_n \in \mathscr{B}'$ to $\varphi \in \mathscr{B}'$ is to say that for each $\eta \in \mathscr{D}_{L^1}$, $(\varphi_n - \varphi) * \eta(0) \to 0$. The weak sequential continuity of the convolution operator $\varphi \mapsto \mu * \varphi$ in $\mathscr{B}'$ is an immediate consequence of the facts that $\eta \mapsto \mu * \eta$ maps $\mathscr{D}_{L^1}$ into itself, and that convolution is commutative (if $(\varphi_n - \varphi) * \eta(0) \to 0$ for all $\eta \in \mathscr{D}_{L^1}$, then also $(\varphi_n - \varphi) * (\mu * \eta)(0) \to 0$ for all $\eta \in \mathscr{D}_{L^1}$, and $(\varphi_n - \varphi) * (\mu * \eta)(0) = (\mu * \varphi_n - \mu * \varphi) * \eta(0)$).

Proof of Lemma 5.1. Take $\psi \in \Gamma(\varphi)$. Then $\tau_{t_k}\varphi \to \psi$ weak* in $L^2_\infty(R; C)$ for some sequence $t_k \to \infty$, and so in particular, $\tau_{t_k}\varphi \to \psi$ in $\mathscr{B}'$ (throughout in this proof $\mathscr{B}'$ carries its weak topology as the dual of $\mathscr{D}_{L^1}$). By the sequential continuity of the convolution operator $\varphi \mapsto \mu * \varphi$ from $\mathscr{B}'$ into itself, $\tau_{t_k}(\mu * \varphi) = \mu * \tau_{t_k}\varphi \to \mu * \psi$ in $\mathscr{B}'$. But the set $\{\tau_t(\mu * \varphi) | t \in R^+\} \cup \Gamma(\mu * \varphi)$ is compact in the weak*-topology of $L^2_\infty(R; C)$, and as a compact topology coincides with every weaker Hausdorff topology, we have $\tau_{t_k}(\mu * \varphi) \to \mu * \psi$ weak* in $L^2_\infty(R; C)$. This shows that $\mu * \psi \in \Gamma(\mu * \varphi)$, and we have proved that $\{\mu * \psi | \psi \in \Gamma(\varphi)\} \subset \Gamma(\mu * \varphi)$.

Conversely, pick some $\xi \in \Gamma(\mu * \varphi)$ and take some sequence $t_k \to \infty$ such that $\tau_{t_k}(\mu * \varphi) \to \xi$ weak* in $L^2_\infty(R; C)$. Pass to a subsequence s_k of t_k such that $\tau_{s_k}\varphi \to \psi \in \Gamma(\varphi)$ weak* in $L^2_\infty(R; C)$ (this is possible because of the sequential compactness of $\{\tau_t\varphi \mid t \in R^+\} \cup \Gamma(\varphi)$). By the first part of the proof, $\xi = \mu * \psi$. This shows that $\Gamma(\mu * \varphi) \subset \{\mu * \psi \mid \psi \in \Gamma(\varphi)\}$, and completes the proof of Lemma 5.1.

Remark 5.1. An alternative method to prove Lemma 5.1 without making use of the weak continuity of the convolution operator $\varphi \to \mu * \varphi$ in $\mathscr{B}'$ is to first show that this convolution operator also maps $L^2_1(R; C)$ into itself (for the definition of $L^2_1(R; C)$, see [8]), and that convolution is commutative in the sense that $\mu \in BM(R; C)$, $\varphi \in L^2_\infty(R; C)$, $\eta \in L^2_1(R; C)$ implies $(\mu * \varphi) * \eta = \varphi * (\mu * \eta)$. This immediately yields the weak*-continuity of the map $\varphi \mapsto \mu * \varphi$ from $L^2_\infty(R; C)$ into itself, and [8, Lemma 3.6] can be applied.

Proofs of Theorems 4.1–4.2 *with* $B = C$. First consider Theorem 4.1. By Lemma 5.1, $\Gamma(\mu * \varphi) = \{\mu * \psi \mid \psi \in \Gamma(\varphi)\}$. Thus, in order to prove Theorem 4.1 it suffices to show that $\mu \in BM(R; C)$, $\psi \in L^2_\infty(R; C)$ implies $\sigma(\mu * \psi) \subset \sigma(\psi)$. Fix μ and ψ as above. Let η_n be an approximate identity in $\mathscr{S}$ (i.e. $\eta_n(t) = n\eta(nt)$, where η is some fixed function in $\mathscr{S}$ with $\hat{\eta}(0) = 1$). Then $(\mu * \psi) * \eta_n \to \mu * \psi$ in $\mathscr{S}'$, and thus it suffices to show that $\sigma((\mu * \psi) * \eta_n) \subset \sigma(\psi)$. By Lemmas 3.2, 3.3 and [7, Lemma 1.9], $\sigma((\mu * \psi) * \eta_n) \subset \sigma(\psi * \eta_n)$. But $(\psi * \eta_n)\hat{} = \hat{\psi}\hat{\eta}_n$, and so trivially, $\sigma(\psi * \eta_n) \subset \sigma(\psi)$. This completes the proof of Theorem 4.1.

The proof of Theorem 4.2 is completely analogous to the proof of Theorem 4.1, and it is therefore omitted. The only essential difference is that one applies [7, Lemma 1.12] instead of [7, Lemma 1.9].

6. A Volterra equation. Applying Theorems 4.1–4.2 to (1.1) we get

THEOREM 6.1. (i) *Let* $\varphi \in L^2_\infty(R^+; B)$, $f \in L^2(R^+; B)$, *and suppose that the solution* x *of* (1.1) *is bounded on* R^+. *Then* $\sigma(\Gamma(x')) \subset \sigma(\Gamma(\varphi))$ *and* $\sigma(\Gamma(x)) \subset \sigma(\Gamma(\varphi)) \cup \{0\}$. (ii) *If moreover* $\sigma(\Gamma(\varphi))$ *is countable and* $\hat{\mu}(\omega) = 0$ $(\omega \in \sigma(\Gamma(\varphi))$, *then* $\tau_t x' \to 0$ *weak* in* $L^2_\infty(R; B)$, *and* $x(t+s) - x(t) \to 0$ $(t \to \infty)$ *weakly in* B, *uniformly on compact subsets of* R.

By the solution x of (1.1) we mean the (unique) locally absolutely continuous function on R^+ which satisfies (1.1) a.e. $[m]$ on R^+. We define $\Gamma(x)$ and $\Gamma(\varphi)$ as above, extending x and φ to R by $x(t) = x_0$, $\varphi(t) = 0$ $(t < 0)$. Clearly, if we moreover extend μ and f to R by $d\mu(t) = f(t) = 0$ $(t < 0)$, then (1.1) becomes

$$x'(t) + \mu * \varphi(t) = f(t),$$

and it holds a.e. $[m]$ on R.

In particular, by Lemma 3.1 and the hypothesis of Theorem 6.1, $x' \in L^2_\infty(R; C)$ (so that $\Gamma(x')$ makes sense), and $x \in BUC(R; B)$.

Before proving Theorem 6.1 we state

LEMMA 6.1. *Let* $x \in BUC(R; B)$, $x' \in L^2_\infty(R; B)$, *where* x' *is the (distribution) derivative of* x. *Then* $\sigma(x) \subset \sigma(x') \cup \{0\}$.

The proof of Lemma 6.1 is omitted, as it is quite straightforward (first reduce it to the scalar case as in § 4, then use a smoothing process, and finally apply [7, Lemma 1.10]).

Proof of Theorem 6.1. It is easy to see that $f \in L^2(R; B)$ implies $\Gamma(f) = \{0\}$, and therefore $\Gamma(x') = \Gamma(\mu * \varphi)$. Thus the claims $\sigma(\Gamma(x')) \subset \sigma(\Gamma(\varphi))$ and $\tau_t x' \to 0$ weak* in $L^2_\infty(R; B)$ follow directly from Theorems 4.1–4.2.

In order to verify the inclusion $\sigma(\Gamma(x)) \subset \sigma(\Gamma(\varphi)) \cup \{0\}$ it clearly suffices to show that $\sigma(\Gamma(x)) \subset \sigma(\Gamma(x')) \cup \{0\}$. We claim that each $y \in \Gamma(x)$ is locally absolutely continuous, and satisfies $y' \in \Gamma(x')$. The proof of this claim is an easy modification of the proof of

[8, Lemma 3.6] (choose some sequence t_k such that $\tau_{t_k}x \to y$, pass to a subsequence s_k so that $\tau_{s_k}x' \to z \subset \Gamma(x')$, and observe that $y(t) = y(0) + \int_0^t z(s)\, ds$). Combining the fact that $y' \in \Gamma(x')$ for each $y \in \Gamma(x)$ with Lemma 6.1 we get the desired inclusion $\sigma(\Gamma(x)) \subset \sigma(\Gamma(x')) \cup \{0\}$. In particular, $\sigma(\Gamma(x)) \subset \sigma(\Gamma(\varphi)) \cup \{0\}$, and in the case (ii), $\sigma(\Gamma(x)) \subset \{0\}$.

The only thing left to prove is that $\sigma(\Gamma(x)) \subset \{0\}$ implies that for each $\alpha \in B'$, $\alpha(x(t+s) - x(t)) \to 0$ as $t \to \infty$, uniformly for s in compact sets. This statement essentially concerns complex-valued functions: If $\alpha x = y \in BUC(R; C)$ and $\sigma(\Gamma(y)) \subset \{0\}$, then $y(t+s) - y(t) \to 0$ as $t \to \infty$, uniformly for s in compact sets. Fix $M > 0$. As $t \to \infty$, the distance from $\tau_t y$ to $\Gamma(y)$ tends to zero [6, Lemma 2.1]. This implies that given $\varepsilon > 0$ we can find $T(\varepsilon) > 0$ such that for each $t > T(\varepsilon)$ there exists some $y_t \in \Gamma(x)$ such that $|\tau_t x(s) - y_t(s)| < \varepsilon$ for $|s| \leqq M$ [6, Lemma 2.2]. But as $\sigma(\Gamma(y)) \subset \{0\}$, the functions y_t are necessarily constants, and so

$$|x(t+s) - x(t)| = |\tau_t x(s) - \tau_t x(0)| \leqq 2\varepsilon$$

for $t \geqq T(\varepsilon)$, $|s| \leqq M$. Thus $y(t+s) - y(t) \to 0$ as $t \to \infty$, uniformly for s in compact sets, and the proof of Theorem 6.1 is complete.

Before one can compare Theorem 6.1 with [7, Theorems 3.1 and 3.2] one must apply either [8, Theorem 6.1] or [9, Prop. 5.1] to get an inclusion of the form $\sigma(\Gamma(\varphi)) \subset Z(\mu)$ (of course, a number of assumptions must be added before either of these results apply). The second conclusion of Theorem 6.1(i) then becomes

$$\sigma(\Gamma(x)) \subset \sigma(\Gamma(\varphi)) \cup \{0\}, \quad \textit{and} \quad \sigma(\Gamma(\varphi)) \subset Z(\mu), \tag{6.1}$$

which is quite similar to the conclusion

$$\begin{gathered} y \in \Gamma(x) \Rightarrow \sigma(y) \subset \sigma(g \circ y) \cup \{0\}, \\ g \circ y \in \Gamma(g \circ x), \quad \textit{and} \quad \sigma(\Gamma(g \circ x)) \subset Z(\mu) \end{gathered} \tag{6.2}$$

of [7, Thm. 3.1] (here $g \circ x$ corresponds to φ; (6.2) has been reformulated in a way which makes it correspond more closely to (6.1)). Clearly (6.2) is stronger than (6.1), as the link $g \circ y$ is missing in (6.1). However, the main inclusions of (6.2) are preserved in (6.1). It is only natural that nothing can be said about $g \circ y$ without any continuity assumption on g.

The conclusion of Theorem 6.1(ii) corresponds in a similar way to the conclusion of [7, Thm. 3.2]. In the situation of Theorem 6.1(ii) it is still true that every $y \in \Gamma(x)$ is a constant, but of course, without any continuous dependence of φ on x there is no hope of getting $\varphi(t) \to 0$ $(t \to \infty)$ as in [7, Thm. 3.2].

REFERENCES

[1] N. DINCULEANU, *Integration on Locally Compact Spaces*, Noordhoff, Leyden, 1974.

[2] W. RUDIN, *Real and Complex Analysis*, 2nd ed., McGraw-Hill, New York, 1974.

[3] L. SCHWARTZ, *Théorie des Distributions*, new ed., Hermann, Paris, 1966.

[4] ———, *Théorie des distributions à valeur vectorielles*, Ann. Inst. Fourier (Grenoble), 7 (1957), pp. 1–141; 8 (1958), pp. 1–209.

[5] O. J. STAFFANS, *Positive definite measures with applications to a Volterra equation*, Trans. Amer. Math. Soc., 218 (1976), pp. 219–237.

[6] ———, *Tauberian theorems for a positive definite form, with applications to a Volterra equation*, Ibid., 218 (1976), pp. 239–259.

[7] ———, *On the asymptotic spectra of the bounded solutions of a nonlinear Volterra equation*, J. Differential Equations, 24 (1977), pp. 365–382.

[8] ———, *An asymptotic problem for a positive definite operator-valued kernel*, this Journal, 9 (1978), pp. 855–866.

[9] ———, *On the asymptotic behavior of finite energy solutions of an abstract integral equation*, this Journal, 9 (1978), pp. 867–875.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0006 $01.00/0

ON A CLASS OF FIRST ORDER EVOLUTION INEQUALITIES ARISING IN HEAT CONDUCTION WITH MEMORY*

J. NAUMANN†

Abstract. This paper deals with the evolution problem

$$u'+a_0u+\int_0^t a(t-s)u(s)\,ds+\int_0^t b(t-s)Bu(s)\,ds+\partial\varphi(u)\ni f,\qquad u(0)=u_0,$$

on an arbitrary interval $[0, T]$, where $a_0\in\mathbb{R}$ and $a, b\colon[0, T]\to\mathbb{R}$ are given functions; B is a linear bounded mapping from a Hilbert space V into its dual, while $\partial\varphi$ denotes the subdifferential mapping of a proper, convex and lower semicontinuous functional $\varphi\colon V\to(-\infty, +\infty]$. We prove an existence and uniqueness theorem for a solution u to the above problem. The existence is established by combining the Galerkin method with a regularization of the functional φ. An application of the abstract result to a Volterra integrodifferential equation arising from the theory of heat conduction in materials with memory is also given.

Introduction. Let H be a real Hilbert space with scalar product $(\cdot,\cdot)$ and norm $|\cdot|$. Suppose we are given a second real Hilbert space V with scalar product $((\cdot,\cdot))$ and norm $\|\cdot\|$ such that

$$V\subset H \text{ compactly, densely.}$$

In addition, the space V is assumed to be separable.

We denote by V^* the dual of V, by (x^*, x) the dual pairing between $x^*\in V^*$ and $x\in V$ and by $\|\cdot\|_*$ the dual norm on V^*. In what follows, the space H will be identified with its dual. This yields the continuous and dense imbedding $H\subset V^*$, and in case $h\in H$ and $x\in V$ the dual pairing between these elements coincides with their scalar product in H.

Let $a(\cdot)$ and $b(\cdot)$ be real functions on $[0, T]$ $(0<T<\infty)$, and let B be a mapping from V into V^*. Finally, let $\varphi\colon V\to(-\infty, +\infty]$ be a proper, convex and lower semicontinuous functional with effective domain

$$D(\varphi)=\{x\in V\colon \varphi(x)<+\infty\}$$

and subdifferential mapping

$$\partial\varphi(x)=\{x^*\in V^*\colon \varphi(y)\geqq\varphi(x)+(x^*, y-x)\quad \forall y\in D(\varphi)\}.$$

Let us then look for a function $u\in L^2(0, T; V)$ such that

$$u'+a_0u+\int_0^t a(t-s)u(s)\,ds+\int_0^t b(t-s)Bu(s)\,ds+\partial\varphi(u)\ni f \tag{1}$$

for a.a. $t\in(0, T)$, and

$$u(0)=u_0 \tag{2}$$

where a_0 is a fixed real, while f and u_0 are given data. The derivative $u'=du/dt$ is to be understood in the sense of vector-valued distributions.

Nonlinear Volterra equations with multivalued mappings have been intensively studied in recent years. We refer to the works [1]–[3], [5], [10], [11] where a number of results about the existence, uniqueness and the asymptotic behavior of a solution to

---

* Received by the editors November 4, 1977, and in final revised form June 29, 1978.

† Sektion Mathematik, Humboldt-Universität zu Berlin, 1086 Berlin, Postfach 1297, East Germany.

equations of this type may be found. In these works, the basic idea for proving the existence of a solution consists in replacing the equation under consideration by an equation that involves the Yosida regularization of the multivalued mapping, solving the latter equation and then letting the regularization parameter tend to zero.

The aim of the present paper is to present an existence and uniqueness result for a solution to (1), (2) (§ 1) under conditions which are rigorously motivated by the physical problem we are going to consider in the second section. The proof of the existence of a solution to (1), (2) rests on the Galerkin method combined with the regularization of the functional φ. Although based on the same idea, this approach differs from that used in the above mentioned works at many points. Section 2 is concerned with an application of the abstract result to the history value problem for a Volterra integrodifferential equation which arises from the linearized theory of heat conduction in materials with memory [7]. The Sobolev space setting of the problem under consideration leads in a straightforward manner to an evolution problem of type (1), (2) by noting that the corresponding constitutive equations are only valid for sufficiently small departures from a fixed constant temperature field.

1. An existence and uniqueness result.

1.1. Statement of the main result. Let $a(\cdot)$ and $b(\cdot)$ be two real functions on $[0, T]$ possessing the following properties:

$$a(\cdot) \in C^1([0, T]); \tag{1.1}$$

$$b(0) > 0, \qquad b(\cdot) \in C^2([0, T]). \tag{1.2}$$

Next, let B be a linear bounded mapping from V into V^* such that

$$\begin{gathered}(Bx, x) + \lambda|x|^2 \geqq b_0\|x\|^2 \quad \forall x \in V, \\ \lambda = \text{const} \geqq 0, \qquad b_0 = \text{const} > 0;\end{gathered} \tag{1.3}$$

$$(Bx, y) = (By, x) \qquad \forall x, y \in V. \tag{1.4}$$

Before stating a further condition let us note that the integral $\int_0^t b(t-s)Bu(s)\,ds$ $(t \in (0, T])$ is well-defined for any $u \in L^1(0, T; V)$ and represents a function in $C([0, T]; V^*)$. Indeed, first of all, it is easy to verify that the function $s \mapsto b(t-s)Bu(s)$ is measurable on $[0, t]$ $(t \in (0, T])$; its integrability follows from that of the function $s \mapsto \|b(t-s)Bu(s)\|_*$ (throughout the paper, measurability and integrability of vector-valued functions are to be understood in the sense of Bochner (cf. e.g. [17])). The continuity of the function $t \mapsto \int_0^t b(t-s)Bu(s)\,ds$ from $[0, T]$ into V^* is seen at once.

We now impose the following additional condition on the function $b(\cdot)$ and the mapping B:

$$\int_0^t \left(\int_0^s b(s-\tau)Bu(\tau)\,d\tau, u(s)\right) ds \geqq 0 \quad \textit{for all } t \in (0, T] \textit{ and any } u \in L^2(0, T; V). \tag{1.5}$$

Sufficient conditions for (1.5) to hold are as follows:

$$(Bx, x) \geqq 0 \qquad \forall x \in V;$$

$$(-1)^k \frac{d^k b}{dt^k}(t) \geqq 0 \quad \forall t \in [0, T] \qquad (k = 0, 1, 2)$$

(cf. [3]; note that the "energy" equality which implies (1.5) is derived in [10] under somewhat weaker conditions on $b(\cdot)$). Conditions on the Laplace transform of $b(\cdot)$ that also guarantee (1.5) may be found in [2; Chap. IV, 4.1], [12].

Finally, given $u \in L^2(0, T; V)$ we define

$$\Phi(u) = \begin{cases} \displaystyle\int_0^T \varphi(u)\,dt & \textit{if } \varphi(u(\cdot)) \in L^1(0, T), \\ +\infty & \textit{otherwise.} \end{cases}$$

The functional Φ is proper, convex and lower semicontinuous on $L^2(0, T; V)$ (cf. [4; Prop. 2.16]). Let $D(\Phi)$ denote the effective domain of Φ, i.e.

$$D(\Phi) = \{u \in L^2(0, T; V) : \Phi(u) < +\infty\}.$$

We now state the main result of our paper.

THEOREM 1. *Let* (1.1)–(1.5) *be satisfied, and let* $a_0 \in \mathbb{R}$ *be fixed. Suppose that the data fulfill the following conditions*:

$$\begin{gathered} f = f_1 + f_2: \qquad f(0) \in H, \\ f_1, f_1' \in L^2(0, T; H), \qquad f_2, f_2', f_2'' \in L^2(0, T; V^*); \end{gathered} \tag{1.6}$$

$$\varphi(x) \geqq \varphi(u_0) \qquad \forall x \in D(\varphi). \tag{1.7}$$

Then there exists exactly one function $u \in C_w([0, T]; V) \cap D(\Phi)$ *such that*

$$u' \in L^\infty(0, T; H); \tag{1.8}$$

$$\begin{gathered} \int_0^T \left(u' + a_0 u + \int_0^t a(t-s)u(s)\,ds + \int_0^t b(t-s)Bu(s)\,ds,\, v-u\right) dt \\ +\Phi(v) - \Phi(u) \geqq \int_0^T (f, v-u)\,dt \qquad \forall v \in D(\Phi); \end{gathered} \tag{1.9}$$

$$u(0) = u_0. \tag{1.10}$$

COROLLARY. *Let the conditions* (1.1)–(1.5) *be satisfied, and let* $a_0 \in \mathbb{R}$.
Further, let $\{f_i, u_{0i}\}$ $(i = 1, 2)$ *be data such that*

$$f_i, f_i' \in L^2(0, T; H), \quad \varphi(x) \geqq \varphi(u_{0i}) \quad \forall x \in D(\varphi) \quad (i = 1, 2).$$

Denote by $u_i \in C_w([0, T]; V) \cap D(\Phi)$ $(i = 1, 2)$ *the solution to* (1.8)–(1.10) *corresponding to* $f = f_i$, $u_0 = u_{0i}$, *respectively. Then*

$$|u_1(t) - u_2(t)|^2 \leqq \text{const}\left(|u_{01} - u_{02}|^2 + \int_0^T |f_1(s) - f_2(s)|^2\,ds\right) \tag{1.11}$$

for all $t \in [0, T]$.

Remark. Since the proof of Theorem 1 we are going to give below does not depend on any convolution argument, the assertion of Theorem 1 continues to hold for Volterra operators with nonconvolution kernels. However, with regard to applications of our result to the theory of materials with memory we have not pursued any extension into this direction.

Further, it is easy to see that the techniques used in the recently developed theory of multivalued Volterra integrodifferential equations (cf. [1]–[3], [5], [10], [11]) do not apply to our case. Theorem 1 thus differs substantially from the results obtained in these works.

Remark. The inequality in (1.9) is equivalent to

$$(1.9^*)\qquad \begin{aligned}&\Big(u'(t)+a_0u(t)+\int_0^t a(t-s)u(s)\,ds+\int_0^t b(t-s)Bu(s)\,ds,\ x-u(t)\Big)+\varphi(x)-\varphi(u(t))\\ &\geqq (f(t), x-u(t)) \text{ for a.a. } t\in(0,T) \text{ and all } x\in D(\varphi)\end{aligned}$$

(cf. [4; Prop. 2.16]).

1.2. Proof of Theorem 1. *Uniqueness.* Let $u, \bar{u}\in C_w([0,T]; V)\cap D(\Phi)$ be two functions that fulfill (1.8)–(1.10). Observing that both u and $\bar{u}$ satisfy (1.9*) we easily find

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}|u(t)-\bar{u}(t)|^2+a_0|u(t)-\bar{u}(t)|^2\\ &\quad+\Big(\int_0^t a(t-s)[u(s)-\bar{u}(s)]\,ds,\ u(t)-\bar{u}(t)\Big)\\ &\quad+\Big(\int_0^t b(t-s)B[u(s)-\bar{u}(s)]\,ds,\ u(t)-\bar{u}(t)\Big)\leqq 0\end{aligned}$$

for a.a. $t\in(0,T)$. Integrating this inequality over $[0,t]$ one gets by virtue of (1.5)

$$\begin{aligned}|u(t)-\bar{u}(t)|^2&\leqq 2|a_0|\int_0^t |u(s)-\bar{u}(s)|^2\,ds\\ &\quad-2\int_0^t\Big(\int_0^s a(s-\tau)[u(\tau)-\bar{u}(\tau)]\,d\tau,\ u(s)-\bar{u}(s)\Big)\,ds\\ &\leqq \text{const}\int_0^t |u(s)-\bar{u}(s)|^2\,ds\end{aligned}$$

for all $t\in[0,T]$. Thus, by Gronwall's Lemma, $u=\bar{u}$.

Existence. Let us begin with some preliminaries which are necessary for our following discussion.

Given $x\in V$ and $\varepsilon>0$ we define

$$\varphi_\varepsilon(x)=\min_{y\in V}\Big\{\frac{1}{2\varepsilon}\|x-y\|^2+\varphi(y)\Big\}$$

(cf. [4; Prop. 2.11]). The functional φ_ε is convex and Fréchet differentiable on the whole of V; its Fréchet derivative $\varphi'_\varepsilon\colon V\to V$ is monotone and Lipschitzian (with Lipschitz constant $1/\varepsilon$). Further, without any loss of generality we may assume that $0\in\partial\varphi(u_0)$ (cf. (1.7); indeed, otherwise we replace φ by $\bar{\varphi}$ where $\bar{\varphi}(x)=\varphi(x)-(u_0^*, x)$ for $x\in D(\varphi)$, $u_0^*\in\partial\varphi(u_0)$ being arbitrary but fixed, while $-u_0^*$ will be identified with a function in $L^2(0,T; V^*)$ and incorporated into the right-hand side f). Then $\varphi'_\varepsilon(u_0)=0$ for all $\varepsilon>0$ which is easily seen when observing that φ'_ε coincides with the Yosida regularization of the subdifferential mapping $\partial\varphi$, the latter being understood in the present situation as a mapping from V into itself.

Next, we introduce a mapping $C_\varepsilon\colon V\to V^*$ by

$$(C_\varepsilon(x), y)=((\varphi'_\varepsilon(x), y))\quad\forall x, y\in V.$$

Setting $(\mathscr{C}_\varepsilon(u))(t)=C_\varepsilon(u(t))$ for any $u\in L^2(0,T; V)$ and for a.a. $t\in(0,T)$ it is easy to check that $\mathscr{C}_\varepsilon(u)\in L^2(0,T; V)$ and that the mapping $\mathscr{C}_\varepsilon$ is monotone and Lipschitzian from $L^2(0,T; V)$ into $L^2(0,T; V^*)$.

Finally, defining for any $u \in L^2(0, T; V)$ and any $\varepsilon > 0$ the functional

$$\Phi_\varepsilon(u) = \min_{v \in L^2(0,T;V)} \left\{ \frac{1}{2\varepsilon} \|u - v\|^2_{L^2(0,T;V)} + \Phi(v) \right\}$$

we have

$$\Phi_\varepsilon(u) = \int_0^T \varphi_\varepsilon(u)\,dt$$

$$= \frac{1}{2\varepsilon} \|u - \mathcal{J}_\varepsilon(u)\|^2_{L^2(0,T;V)} + \Phi(\mathcal{J}_\varepsilon(u))$$

where $\mathcal{J}_\varepsilon = (I + \varepsilon\, \partial\Phi)^{-1}$ (I = identity in $L^2(0, T; V)$),

$$\partial\Phi(u) = \left\{ z \in L^2(0, T; V) \colon \Phi(v) \geqq \Phi(u) + \int_0^T ((z, v-u))\,dt\ \forall v \in D(\Phi) \right\}$$

(cf. [4; Prop. 2.16]).

1.2.1. Approximate solutions. Let $\{w_1, w_2, \cdots\}$ be a system of elements in V such that

(i) *the elements* $w_1, \cdots, w_n$ *are linearly independent for each* n;

(ii) $\overline{\bigcup_{n=1}^\infty V_n} = V$ *where* $V_n = \operatorname{span}\{w_1, \cdots, w_n\}$.

One may assume that $u_0 \in V_{n_0}$ for a certain natural number n_0.

We then consider the following initial value problem for the real functions $g_{n1}, \cdots, g_{nn}$:

$$\begin{aligned}(u_n'(t), w_i) + a_0(u_n(t), w_i) + \left(\int_0^t a(t-s) u_n(s)\,ds, w_i \right) \\ + \left(\int_0^t b(t-s) B u_n(s)\,ds, w_i \right) + (C_\varepsilon(u_n(t)), w_i) \\ = (f(t), w_i) \qquad (i = 1, \cdots, n; \qquad n \geqq n_0, \qquad \varepsilon > 0),\end{aligned} \tag{1.12}$$

$$u_n(0) = u_0 \tag{1.13}$$

where

$$u_n(t) = \sum_{i=1}^n g_{ni}(t) w_i.$$

Observing the equivalence of all norms on V_n the existence of functions $g_{ni} \in C^1([0, T])$ $(i = 1, \cdots, n)$ that satisfy (1.12), (1.13), can be proved by applying the contraction mapping principle within the space $[C([0, T])]^n$ when furnishing it with the usual maximum-norm weighted by e^{-kt} (k being an appropriate positive constant). A somewhat different argument for solving a system of type (1.12), (1.13) is used in [12; pp. 566–567].

1.2.2. A priori estimates. Setting $t = 0$ in (1.12) yields

$$|u_n'(0)| \leqq |f(0) - a_0 u_0| \qquad \forall n \geqq n_0, \quad \forall \varepsilon > 0. \tag{1.14}$$

Next, the function $C_\varepsilon(u_n(\cdot))$ being absolutely continuous from $[0, T]$ into V^*, one concludes from (1.12) that the functions $g_{n,1}', \cdots, g_{nn}'$ are absolutely continuous on $[0, T]$. Since the function $C_\varepsilon(u_n(\cdot))$ is differentiable for a.a. $t \in (0, T)$ (cf. [4;

Appendice]) we may differentiate (1.12) term by term and obtain

$$
\begin{aligned}
&(u_n''(t), w_i)+a_0(u_n'(t), w_i)+a(0)(u_n(t), w_i) \\
&\qquad +\left(\int_0^t a'(t-s)u_n(s)\,ds, w_i\right)+b(0)(Bu_n(t), w_i) \\
&\qquad +\left(\int_0^t b'(t-s)Bu_n(s)\,ds, w_i\right)+\left(\frac{d}{dt}C_\varepsilon(u_n(t)), w_i\right) \\
&= (f'(t), w_i)
\end{aligned}
$$

for a.a. $t \in (0, T)$. Multiplying the ith equation of the latter system by $g_{ni}'(t)$, summing on $1, \cdots, n$, integrating the result over the interval $[0, t]$ ($t \in (0, T]$) and then adding the term $b(0)\lambda |u_n(t)|^2$ to both sides of the equality obtained we get

$$
\begin{aligned}
&|u_n'(t)|^2+\|u_n(t)\|^2 \\
&\leqq c_1\left(1+\int_0^t |u_n'(s)|^2\,ds\right)+c_2\int_0^t (f_2'(s), u_n'(s))\,ds \\
&\quad -c_2\int_0^t\left(\int_0^s a'(s-\tau)u_n(\tau)\,d\tau, u_n'(s)\right)ds \\
&\quad -c_2\int_0^t\left(\int_0^s b'(s-\tau)Bu_n(\tau)\,d\tau, u_n'(s)\right)ds
\end{aligned}
\tag{1.15}
$$

for a.a. $t \in (0, T)$, where we have used the estimate (1.14) and the inequality

$$
\left(\frac{d}{dt}C_\varepsilon(u_n(t)), u_n'(t)\right) \geqq 0 \qquad \textit{for a.a. } t \in (0, T)
$$

(c_1, c_2 are positive constants depending neither on n nor on ε).

The second term on the right hand side of (1.15) can be evaluated by the aid of an integration by parts as follows:

$$
\left|\int_0^t (f_2'(s), u_n'(s))\,ds\right| \leqq \frac{1}{4c_2}\|u_n(t)\|^2+\text{const}\left(1+\int_0^t \|u_n(s)\|^2\,ds\right)
\tag{1.16a}
$$

for all $t \in [0, T]$. Next, the estimate

$$
\begin{aligned}
&\left|\int_0^t\left(\int_0^s a'(s-\tau)u_n(\tau)\,d\tau, u_n'(s)\right)ds\right| \\
&\leqq \text{const}\int_0^t (|u_n'(s)|^2+\|u_n(s)\|^2)\,ds, \qquad t \in [0, T]
\end{aligned}
\tag{1.16b}
$$

is readily seen. Again using integration by parts, we obtain for the last term on the right hand side of (1.15) that

$$
\begin{aligned}
&\left|\int_0^t\left(\int_0^s b'(s-\tau)Bu_n(\tau)\,d\tau, u_n'(s)\right)ds\right| \\
&\leqq \frac{1}{4c_2}\|u_n(t)\|^2+\text{const}\int_0^t \|u_n(s)\|^2\,ds, \qquad t \in [0, T].
\end{aligned}
\tag{1.16c}
$$

Inserting the estimates (16a)–(16c) into (1.15) gives

$$
|u_n'(t)|^2+\|u_n(t)\|^2 \leqq c_3\left[1+\int_0^t (|u_n'(s)|^2+\|u_n(s)\|^2)\,ds\right]
$$

for a.a. $t \in (0, T)$. Hence

$$|u_n'(t)| + \|u_n(t)\| \leqq c_4 \qquad \forall t \in [0, T], \quad \forall n \geqq n_0, \quad \forall \varepsilon > 0 \tag{1.17}$$

($c_k = \text{const} > 0$, $k = 3, 4$).

1.2.3. Passage to limit $n \to \infty$ ($\varepsilon > 0$ arbitrary, but fixed). From (1.17) we conclude (by going to a subsequence if necessary) that

$$\begin{aligned} u_n &\to u_\varepsilon \quad \textit{weakly* in } L^\infty(0, T; V), \\ u_n &\to u_\varepsilon \quad \textit{strongly in } L^2(0, T; H), \\ u_n' &\to u_\varepsilon' \quad \textit{weakly* in } L^\infty(0, T; H), \\ \mathscr{C}_\varepsilon(u_n) &\to w_\varepsilon^* \quad \textit{weakly in } L^2(0, T; V^*) \end{aligned} \tag{1.18}$$

as $n \to \infty$ (note that the second convergence property in (1.18) follows from a compactness theorem (cf. [9; Chap. I, 5])).

Combining the first and the third convergence property in (1.18) with an integration by parts we find

$$u_\varepsilon(0) = u_0 \qquad \forall \varepsilon > 0. \tag{1.19}$$

Let now $\psi \in \mathscr{D}((0, T))$ (= the set of all real infinitely differentiable functions having their support in $(0, T)$) be arbitrary, let $i_0 \geqq n_0$ be an arbitrary natural number, and let a_i ($i = 1, \cdots, i_0$) be arbitrary reals. We multiply the ith equation in (1.12) ($n \geqq i_0$) by $\psi(t)a_i$, sum on $1, \cdots, i_0$ and integrate the result over the interval $[0, T]$. Letting then $n \to \infty$ one obtains by virtue of (1.18)

$$\int_0^T \Big(u_\varepsilon'(t) + a_0 u_\varepsilon(t) + \int_0^t a(t-s) u_\varepsilon(s)\, ds + \int_0^t b(t-s) B u_\varepsilon(s)\, ds \\ + w_\varepsilon^*(t) - f(t), \psi(t) \sum_{i=1}^{i_0} a_i w_i \Big)\, dt = 0$$

or, equivalently,

$$u_\varepsilon'(t) + a_0 u_\varepsilon(t) + \int_0^t a(t-s) u_\varepsilon(s)\, ds + \int_0^t b(t-s) B u_\varepsilon(s)\, ds + w_\varepsilon^*(t) - f(t) = 0 \tag{1.20}$$

for a.a. $t \in (0, T)$.

To show that $w_\varepsilon^* = \mathscr{C}_\varepsilon(u_\varepsilon)$, let $v \in L^2(0, T; V)$ be arbitrary. Taking into account (1.12) one gets

$$\begin{aligned} 0 &\leqq \int_0^T (C_\varepsilon(v) - C_\varepsilon(u_n), v - u_n)\, dt \\ &= \int_0^T (C_\varepsilon(v), v - u_n)\, dt - \int_0^T (C_\varepsilon(u_n), v)\, dt \\ &\quad + \int_0^T \Big(f - u_n' - a_0 u_n - \int_0^t a(t-s) u_n(s)\, ds, u_n \Big)\, dt \\ &\quad - \int_0^T \Big(\int_0^t b(t-s) B u_n(s)\, ds, u_n \Big)\, dt \end{aligned} \tag{1.21}$$

($n \geqq n_0$). Observing the hypothesis (1.5) and using (1.18) as well as (1.19), (1.20) we obtain by taking the lim sup on both sides of (1.21)

$$
\begin{aligned}
0 \leqq & \int_0^T (C_\varepsilon(v), v-u_\varepsilon)\,dt - \int_0^T (w_\varepsilon^*, v)\,dt \\
& + \int_0^T \left(f - u_\varepsilon' - a_0 u_\varepsilon - \int_0^t a(t-s)u_\varepsilon(s)\,ds, u_\varepsilon\right) ds \\
& - \int_0^T \left(\int_0^t b(t-s)Bu_\varepsilon(s)\,ds, u_\varepsilon\right) dt \\
= & \int_0^T (C_\varepsilon(v) - w_\varepsilon^*, v - u_\varepsilon)\,dt.
\end{aligned}
$$

Thus, by a standard device from the theory of monotone operators, $w_\varepsilon^* = \mathscr{C}_\varepsilon(u_\varepsilon)$ and therefore

$$
\begin{aligned}
u_\varepsilon'(t) + a_0 u_\varepsilon(t) + & \int_0^t a(t-s)u_\varepsilon(s)\,ds \\
& + \int_0^t b(t-s)Bu_\varepsilon(s)\,ds + C_\varepsilon(u_\varepsilon(t)) - f(t) = 0
\end{aligned}
\tag{1.20*}
$$

for a.a. $t \in (0, T)$ (and all $\varepsilon > 0$).

1.2.4. Passage to limit $\varepsilon \to 0$. Since the estimate (1.16) continues to hold for the functions u_ε and u_ε', without any loss of generality we may assume that

$$
\begin{aligned}
u_\varepsilon &\to u \quad \textit{weakly* in } L^\infty(0, T; V), \\
u_\varepsilon &\to u \quad \textit{strongly in } L^2(0, T; H), \\
u_\varepsilon' &\to u' \quad \textit{weakly* in } L^\infty(0, T; H)
\end{aligned}
\tag{1.22}
$$

as $\varepsilon \to 0$. By the same device as above, $u(0) = u_0$. Further, it is readily verified that $u \in C_w([0, T]; V)$.

Let $v \in D(\Phi)$ be arbitrary. Multiplying (1.20) by $v(t) - u_\varepsilon(t)$ and integrating over $[0, T]$ yields

$$
\begin{aligned}
\int_0^T \left(u_\varepsilon' + a_0 u_\varepsilon + \int_0^t a(t-s)u_\varepsilon(s)\,ds + \int_0^t b(t-s)Bu_\varepsilon(s)\,ds, v - u_\varepsilon\right) dt \; + \Phi(v) - \Phi_\varepsilon(u_\varepsilon) \\
\geqq \int_0^T (f, v - u_\varepsilon)\,dt.
\end{aligned}
\tag{1.23}
$$

Setting $v = u_0$ in (1.23) we find

$$
\begin{aligned}
\text{const} & \geqq \Phi_\varepsilon(u_\varepsilon) \\
& \geqq \frac{1}{2\varepsilon}\|u_\varepsilon - \mathscr{J}_\varepsilon(u_\varepsilon)\|^2_{L^2(0,T;V)} + T\varphi(u_0)
\end{aligned}
$$

for all $\varepsilon > 0$. Hence

$$
\mathscr{J}_\varepsilon(u_\varepsilon) \to u \quad \textit{weakly in } L^2(0, T; V)
\tag{1.24}
$$

as $\varepsilon \to 0$.

Finally, (1.23) implies

$$\int_0^T \left(u_\varepsilon' + a_0 u_\varepsilon + \int_0^t a(t-s)u_\varepsilon(s)\,ds,\, v - u_\varepsilon\right) dt$$
$$+\int_0^T \left(\int_0^t b(t-s)Bu_\varepsilon(s)\,ds,\, v\right) dt + \Phi(v)$$
$$\geqq \Phi(\mathcal{J}_\varepsilon(u_\varepsilon)) + \int_0^T \left(\int_0^t b(t-s)Bu_\varepsilon(s)\,ds,\, u_\varepsilon\right) dt + \int_0^T (f, v-u_\varepsilon)\,dt.$$

Observing (1.22) and (1.24) and taking the lim inf on both sides of the latter inequality we get $u \in D(\Phi)$ and the desired inequality (1.9).

Proof of the Corollary. Inserting $u = u_i$ $(i = 1, 2)$ in (1.9*) one easily finds

$$\frac{1}{2}\frac{d}{dt}|u_1(t)-u_2(t)|^2 + \left(\int_0^t b(t-s)B[u_1(s)-u_2(s)]\,ds,\, u_1(t)-u_2(t)\right)$$
$$\leqq (\tfrac{1}{2}+|a_0|)|u_1(t)-u_2(t)|^2 + \tfrac{1}{2}|f_1(t)-f_2(t)|^2$$
$$-\left(\int_0^t a(t-s)[u_1(s)-u_2(s)]\,ds,\, u_1(t)-u_2(t)\right)$$

for a.a. $t \in (0, T)$. Integrating this inequality over $[0, t]$ one obtains

$$|u_1(t)-u_2(t)|^2 \leqq |u_{01}-u_{02}|^2 + \int_0^T |f_1(s)-f_2(s)|^2\,ds$$
$$+\mathrm{const}\int_0^t |u_1(s)-u_2(s)|^2\,ds$$

for all $t \in [0, T]$. The inequality (1.11) is now seen at once.

2. Application of the abstract result.

2.1. Formulation of the problem. Let a stationary rigid heat conducting body occupy the bounded three-dimensional domain Ω. Then the law of balance of energy requires that

$$\frac{\partial e}{\partial t} = -\operatorname{div}\mathbf{q} + r \qquad in\ \Omega \times \mathbb{R}. \tag{2.1}$$

Here $e = e(x, t)$ denotes the internal energy, $\mathbf{q} = \mathbf{q}(x, t) = \{q_1(x, t), q_2(x, t), q_3(x, t)\}$ the heat flux and $r = r(x, t)$ the external heat supply ($x = \{x_1, x_2, x_3\}$ is the generic point in $\mathbb{R}^3$).

Inserting $e = e_0 + c(\Theta - \Theta_0)$ ($e_0 = \mathrm{const}$, $c = \mathrm{const} > 0$ specific heat, Θ absolute temperature, Θ_0 a fixed reference temperature) and $\mathbf{q} = -k\nabla\Theta$ ($k = \mathrm{const} > 0$ thermal conductivity) [Fourier's law] into (2.1) we get the classical heat equation. It is well-known that the theory based on this equation predicts an unrealistic result: a thermal disturbance propagates with infinite speed.

Gurtin and Pipkin [7] have proposed a (nonlinear) theory of heat conduction in which the present value of free energy, entropy and heat flux depend on the present value of the absolute temperature, the summed history of the temperature and the summed history of the temperature gradient. In this theory thermal disturbances propagate with finite speeds.

The theory in [7] is linearized as follows. let $T \in \mathbb{R}$ be an arbitrary, but fixed instant (without any loss of generality let $0 < T < \infty$), and let $\Theta_0 = \text{const} > 0$ be fixed (Θ_0 represents a constant equilibrium temperature field). Then the set of all pairs $\{\Theta, \nabla\Theta\}$ which fulfill the condition

$$\sup_{t\in(-\infty,T)} \{|\Theta(x,t)-\Theta_0| + |\nabla\Theta(x,t)|\} \leqq \delta \tag{2.2}$$

($x \in \Omega$ arbitrary, $\delta = \text{const} > 0$ sufficiently small) is used to form a neighborhood of $\{\Theta_0, \Theta_0, 0\}$[1] with respect to a fading memory norm. Differentiating the response functionals that represent e and $\mathbf{q}$, at the point $\{\Theta_0, \Theta_0, 0\}$ and replacing these functionals in the neighborhood under consideration by their first derivatives leads to the following constitutive equations:

$$e = e_0 + c\Theta - \int_0^\infty \alpha'(s)\left[\int_{t-s}^t \Theta(\tau)\, d\tau\right] ds, \tag{2.3}$$

$$\mathbf{q} = \int_0^\infty b'(s)\left[\int_{t-s}^t \nabla\Theta(\tau)\, d\tau\right] ds. \tag{2.4}$$

Here $\alpha(\cdot)$ and $b(\cdot)$ are certain (differentiable) functions on $[0, \infty)$, while $e_0 = \text{const}$, $c = \text{const} > 0$ (cf. [7] for further details).

Without any further reference, in what follows we assume that

$$|\alpha(s)| + |b(s)| \leqq k_0 s^{-\rho} \qquad \forall s \in (0, \infty)$$

($k_0 = \text{const} > 0$, $\rho > 1$; cf. [7; p. 125]). Integrating by parts (for the time being, the boundary Γ of Ω is assumed to be sufficiently smooth) and observing (2.2) as well as the latter assumption on $\alpha(\cdot)$ and $b(\cdot)$ the constitutive equations (2.3), (2.4) take the form

$$e = e_0 + c\Theta + \int_{-\infty}^t \alpha(t-s)\Theta(s)\, ds, \tag{2.3'}$$

$$\mathbf{q} = -\int_{-\infty}^t b(t-s)\nabla\Theta(s)\, ds. \tag{2.4'}$$

We are now going to present the "classical" formulation of the history value problem associated with (2.1) under the constitutive equations (2.3′), (2.4′).

To this end, let

$$r = 0 \qquad \text{on } \Omega \times (0, T),$$

and let ζ denote a given (real) function on $\bar{\Omega} \times (-\infty, 0]$ such that

$$\sup_{t\in(-\infty,0]} \{|\zeta(x,t)| + |\nabla\zeta(x,t)|\} \leqq \delta \qquad (x \in \Omega \textit{ arbitrary}). \tag{2.2*}$$

Next, suppose that $\Gamma = \Gamma_1 \cup \Gamma_2$ where $\Gamma_1 \cap \Gamma_2 = \emptyset$ and either $\operatorname{mes}(\Gamma_k) > 0$ or $\Gamma_k = \emptyset$ ($k = 1, 2$). Let then χ denote a given function on $\Gamma_2 \times (0, T)$.

Define

$$\bar{\Theta}(x,t) = \Theta(x,t) - \Theta_0 \qquad \textit{for } \{x,t\} \in \bar{\Omega} \times (-\infty, T).$$

[1] Cf. [7] for the precise notation of this triplet.

Following [14] *the function* $\bar{\Theta}$ *will be called a solution to the history value problem corresponding to the thermal history* ζ *if*

$$\bar{\Theta} = \zeta \qquad \text{on } \bar{\Omega} \times (-\infty, 0], \tag{2.5}$$

$$\sup_{t \in (0,T)} \{|\bar{\Theta}(x, t)| + |\nabla \bar{\Theta}(x, t)|\} \leqq \delta \qquad (x \in \Omega \text{ arbitrary}), \tag{2.6}$$

$$\begin{aligned} &\frac{\partial \bar{\Theta}}{\partial t} + a_0 \bar{\Theta} + \int_{-\infty}^{t} a(t-s)\bar{\Theta}(s)\, ds - \operatorname{div}\left(\int_{-\infty}^{t} b(t-s)\nabla \bar{\Theta}(s)\, ds\right) \\ &\qquad = 0 \qquad \text{in } \Omega \times (0, T) \end{aligned} \tag{2.7}$$

where

$$a_0 = \alpha(0), \qquad a(\cdot) = \alpha'(\cdot) \qquad (c \text{ normalized to } 1),$$

$$\bar{\Theta}(x, t) = 0 \qquad \text{on } \Gamma_1 \times (0, T), \tag{2.8a}$$

$$q_i(x, t) n_i(x) = \chi(x, t) \qquad \text{on } \Gamma_2 \times (0, T)^2 \tag{2.8b}$$

($n = \{n_1, n_2, n_3\}$ denotes the outward unit normal along Γ,

$$\mathbf{q} = -\int_{-\infty}^{0} b(t-s)\nabla \zeta(s)\, ds - \int_{0}^{t} b(t-s)\nabla \bar{\Theta}(s)\, ds).$$

Remark. Initial boundary value problems for both linear and nonlinear equations arising from the theory of heat conduction in materials with memory have been extensively studied in a more or less classical framework (cf. e.g. the papers [8], [14], [16] and the literature quoted therein). However, all these works only concern the uniqueness and stability of a solution as well as problems of wave propagation. Let us further refer to [13] where under relatively restrictive conditions on the data the existence and uniqueness of a C^2-solution to a one-dimensional, partly nonlinear equation of type (2.7) is proved.

Remark. Condition (2.2*) and satisfying of (2.6) guarantee that the constitutive equations (2.3), (2.4) (and thus (2.3'), (2.4')) make sense as linearization of the general (nonlinear) constitutive relations considered in [7].

Further, without any loss of generality we may assume that $0 < \delta < \Theta_0$. Then (2.6) implies

$$\Theta(x, t) > 0 \qquad \text{for } \{x, t\} \in \Omega \times (0, T).$$

Remark. Let $\gamma: \mathbb{R} \to 2^{\mathbb{R}}$ be a maximal monotone mapping (i.e. γ is monotone and does not possess any proper monotone extension; cf. [4] for details). Then (2.8b) can be replaced by the more general boundary condition

$$q_i(x, t) n_i(x) - \chi(x, t) \in \gamma(\bar{\Theta}(x, t)) \qquad \text{on } \Gamma_2 \times (0, T). \tag{2.8b*}$$

Obviously, (2.8b) is a special case of (2.8b*) (set $\gamma(r) = 0$ for all $r \in \mathbb{R}$).

We briefly note another example for the mapping γ. Let g_i, h_i, k_i ($i = 1, 2$) be real numbers satisfying the following conditions:

$$0 \in [g_1, g_2], \qquad h_1 \leqq h_2, \qquad k_i > 0\ (i = 1, 2).$$

[2] In what follows we use the convention that a repeated subscript means summation over 1, 2, 3.

Let $\chi \equiv 0$, and define

$$\gamma(r)=\begin{cases} g_1 & \text{if } r \leqq h_1+g_1/k_1, \\ k_1(r-h_1) & \text{if } h_1+g_1/k_1 \leqq r \leqq h_1, \\ 0 & \text{if } h_1 \leqq r \leqq h_2, \\ k_2(r-h_2) & \text{if } h_2 \leqq r \leqq h_2+g_2/k_2, \\ g_2 & \text{if } r \leqq h_2+g_2/k_2. \end{cases}$$

Boundary condition (2.8b*) now describes a temperature control through Γ_2 (cf. [6; Chap. I, 2.3]).

Letting formally $k_1 \to +\infty$ and $g_1 \to -\infty$ the corresponding mapping γ is multivalued. We note that this limit case also possesses physical significance.

In what follows we restrict our attention to (2.8a), (2.8b) only for technical simplicity. It is easy to see that Theorem 2, presented below, can be readily extended (with minor modifications) to the case of the more general boundary condition (2.8b*) (cf. [6; Chap. I] for further details).

2.2. Existence and uniqueness of a weak solution to (2.5)–(2.8).

2.2.1. Let h be a fixed influence function, i.e. $h \in C([0,\infty)) \cap L^2(0,\infty)$[3] with $h(t)>0$ for all $t \in [0,\infty)$ and $h(t) \to 0$ monotonically as $t \to \infty$.

Let X be a real Hilbert space with norm $\|\cdot\|_X$. We then denote by $L_h^2(-\infty,0;X)$ the set of all (classes) of (strongly) Bochner-measurable functions $\zeta: (-\infty,0) \to X$ such that

$$\|\zeta\|_{L_h^2(-\infty,0;X)}=\left\{\int_{-\infty}^0 \|\zeta(s)\|_X^2 h^2(-s)\,ds\right\}^{1/2}<+\infty.$$

It is easy to check that $L_h^2(-\infty,0;X)$ (furnished with the usual vector space structure) is complete with respect to the above norm. Let us note that $L_h^2(-\infty,0;X)$ is a "space of histories" in the sense of Coleman and Noll (cf. e.g. Arch. Rational Mech. Anal., 6 (1960), pp. 355–370).

2.2.2. LEMMA. *Let the function $a(\cdot)$ satisfy the following conditions:*

$$\begin{aligned} &a'(t) \text{ exists for all } t \in (0,\infty), \\ &[a'(t_1)]^2 \geqq [a'(t_2)]^2 \qquad \forall 0<t_1<t_2<\infty, \end{aligned} \tag{2.9}$$

$$\int_0^\infty \left[\frac{d^k a}{dt^k}(t)\right]^2 h^{-2}(t)\,dt<+\infty \qquad (k=0,1)^4. \tag{2.10}$$

Let $\zeta \in L_h^2(-\infty,0;X)$. Then the function g,

$$g(t)=\int_{-\infty}^0 a(t-s)\zeta(s)\,ds, \qquad t \in [0,T]$$

is Lipschitz from $[0,T]$ into X and its distributional derivative belongs $L^\infty(0,T;X)$.

Proof. Let $t \in [0,T]$. First of all, it holds

$$|a(t-s)|h^{-1}(-s) \leqq |a(t-s)|h^{-1}(t-s), \qquad s \in (-\infty,0].$$

[3] In [7] a slightly stronger condition is used for technical reasons.

[4] The measurability of $[a'(\cdot)]^2$ follows from its monotonicity. Further, the monotonicity of $[a'(\cdot)]^2$ and (2.10) imply that $a'(\cdot)$ is everywhere finite on $(0,\infty)$.—Note that conditions of type (2.10) are mentioned in [16; p. 198]; cf. also Dafermos, C. M., Arch. Rational Mech. Anal., 37 (1970), pp. 297–308.

Observing (2.9) and (2.10) it is easily seen that the function $s \mapsto a(t-s)h^{-1}(-s)$ is in $L^2(-\infty, 0)$.

Next, the function $s \mapsto a(t-s)\zeta(s)$ is Bochner measurable from $(-\infty, 0)$ into X; its integrability follows from the identity

$$\|a(t-s)\zeta(s)\|_X = (|a(t-s)|h^{-1}(-s))\|\zeta(s)\|_X h(-s)$$

(for a.a. $s \in (-\infty, 0)$) (cf. [17]). Hence, the integral $\int_{-\infty}^0 a(t-s)\zeta(s)\,ds$ is well-defined (even for any real $t \geqq 0$).

By an analogous argument as above, the function $s \mapsto a'(t-s)h^{-1}(-s)$ is in $L^2(-\infty, 0)$. Let now $\bar{t} \in (0, T)$ and $t \in [0, T-\bar{t}]$ be arbitrary. Taking into account (2.9) we get

$$\begin{aligned}\|[a(t+\bar{t}-s)-a(t-s)]\zeta(s)\|_X &\\ = \bar{t}(|a'(t+\sigma\bar{t}-s)|h^{-1}(-s))\|\zeta(s)\|_X h(-s) &\\ \leqq \tfrac{1}{2}\bar{t}\{[a'(t-s)]^2h^{-2}(-s)+\|\zeta(s)\|_X^2 h^2(-s)\}&\end{aligned}$$

for a.a. $s \in (-\infty, 0)$ ($\sigma \in (0, 1)$). Therefore

$$\begin{aligned}\|g(t+\bar{t})-g(t)\|_X &\leqq \int_{-\infty}^0 \|[a(t+\bar{t}-s)-a(t-s)]\zeta(s)\|_X\,ds \\ &\leqq \tfrac{1}{2}\bar{t}\Big(\int_0^\infty [a'(r)]^2h^{-2}(r)\,dr + \int_{-\infty}^0 \|\zeta(s)\|_X^2 h^2(-s)\,ds\Big).\end{aligned}$$

The latter estimate is equivalent for the function g to have its distributional derivative in $L^\infty(0, T; X)$ (cf. e.g. [4; Appendice]).

2.2.3. Set $H = L^2(\Omega)$, and

$$(u, v) = \int_\Omega uv\,dx \quad \textit{for } u, v \in H.$$

In what follows we suppose that the boundary Γ is Lipschitz (cf. [15] for details). Let then $H^m(\Omega)$ $(= W_2^m(\Omega))$ $(m = 1, 2, \cdots)$ denote the usual Sobolev space of real square integrable functions in Ω having generalized partial derivatives of all orders $\leqq m$ in $L^2(\Omega)$ (note that under our above assumption on Γ the Sobolev imbedding theorems and the trace theorems are valid; cf. [15]). $H^m(\Omega)$ is complete with respect to the norm

$$\|u\|_{H^m(\Omega)} = \Big\{\sum_{|l|\leqq m}\int_\Omega (D^l u)^2\,dx\Big\}^{1/2}$$

$$\Big(D^l u = \frac{\partial^{|l|}u}{\partial x_1^{l_1}\,\partial x_2^{l_2}\,\partial x_3^{l_3}}, \qquad |l| = l_1 + l_2 + l_3\Big).$$

Define

$$V = \{u \in H^1(\Omega) \colon u = 0 \textit{ a.e. on } \Gamma_1\}.$$

The imbedding $V \subset H$ is compact and dense (cf. [15]). Next, it is easily checked that

$$K = \{u \in V \colon |u(x)| + |\nabla u(x)| \leqq \delta \textit{ a.e. in } \Omega\}$$

is a closed convex subset of V.

Further, we introduce a linear bounded mapping B from V into V^* by

$$(Bu, v) = \int_\Omega \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i}\,dx \qquad (u, v \in V).$$

The mapping B obviously satisfies the conditions (1.3), (1.4).

2.2.4. In the sequel we suppose that $\chi(\cdot, t) \in L^2(\Gamma_2)$ for all $t \in [0, T]$. By the trace theorem (cf. e.g. [15]), for each $t \in [0, T]$ there exists an element $f_2(t) \in V^*$ such that

$$(f_2(t), v) = \int_{\Gamma_2} \chi(x, t)v(x)\, dS \qquad \forall v \in V.$$

We then have

THEOREM 2. *Let the functions $a(\cdot)$ and $b(\cdot)$ satisfy the following conditions:*

(2.11) $$\text{the restriction of } a(\cdot) \text{ onto } [0, T] \text{ is in } C^1([0, T]),$$

(2.12) $$\begin{gathered} a'(t) \text{ exists for all } t \in (T, \infty), \\ [a'(t_1)]^2 \geqq [a'(t_2)]^2 \qquad \forall 0 < t_1 < t_2 < \infty, \end{gathered}$$

(2.13) $$b(0) > 0,$$

(2.14) $$\begin{gathered} \text{the restriction of } b(\cdot) \text{ onto } [0, T] \text{ is in } C^2([0, T]), \\ (-1)^k \frac{d^k b}{dt^k}(t) \geqq 0 \qquad \forall t \in [0, T] \qquad (k = 0, 1, 2), \end{gathered}$$

(2.15) $$\begin{gathered} b'(t) \text{ exists for all } t \in (T, \infty), \\ [b'(t_1)]^2 \geqq [b'(t_2)]^2 \qquad \forall T < t_1 < t_2 < \infty, \end{gathered}$$

(2.16) $$\int_0^\infty \left(\left[\frac{d^k a}{dt^k}(t)\right]^2 + \left[\frac{d^k b}{dt^k}(t)\right]^2 \right) h^{-2}(t)\, dt < +\infty \qquad (k = 0, 1).$$

Further, let $\zeta \in L_h^2(-\infty, 0; H^2(\Omega))$ such that $(\partial\zeta/\partial n)(s) = 0$[5] *a.e. on Γ_2 (for a.a. $s \in (-\infty, 0)$), and suppose that*

$$f_2, f_2', f_2'' \in L^2(0, T; V^*),$$

$$\left| \left(\int_{-\infty}^0 a(-s)\zeta(s)\, ds + \int_{-\infty}^0 b(-s)B\zeta(s)\, ds + f_2(0), v \right) \right| \leqq \text{const}\, \|v\|_{L^2(\Omega)} \qquad \forall v \in V.$$

Finally, let $u_0 \in K$ be arbitrarily given.

Then there exists exactly one function $u: (-\infty, T] \to V$ such that $u(s) = \zeta(s)$ for a.a. $s \in (-\infty, 0)$ while the restriction of u onto $[0, T]$ satisfies the following conditions:

(2.17) $$u \in C_w([0, T]; V), \qquad u(t) \in K \qquad \forall t \in [0, T],$$

(2.18) $$u' \in L^\infty(0, T; H),$$

(2.19) $$\begin{gathered} \int_0^T \left(u' + a_0 u + \int_{-\infty}^t a(t-s)u(s)\, ds + \int_{-\infty}^t b(t-s)Bu(s)\, ds, v - u \right) dt \\ \geqq \int_0^T (f_2, v - u)\, dt \end{gathered}$$

for all $v \in L^2(0, T; V)$ with $v(t) \in K$ for a.a. $t \in (0, T)$,

(2.20) $$u(0) = u_0.$$

Proof. Define

$$\varphi(v) = \begin{cases} 0 & \text{if } v \in K, \\ +\infty & \text{if } v \in V \backslash K. \end{cases}$$

[5] This assumption in particular implies that $q_i(x, s)n_i(x) = 0$ a.e. on Γ_2 (for all $s \in (-\infty, 0)$).

The functional φ is proper, convex and lower semicontinuous on V.

Next, set

$$f_1(t)=\int_{-\infty}^{0} a(t-s)\zeta(s)\,ds, \qquad \bar{f}_1(t)=\int_{-\infty}^{0} b(t-s)B\zeta(s)\,ds$$

for $t\in[0, T]$. By the above lemma, $f_1\in L^\infty(0, T; H^2(\Omega))$. Further, taking into account that $(\partial\zeta/\partial n)(s)=0$ a.e. on Γ_2 (for a.a. $s\in(-\infty, 0)$) we get the existence of a (uniquely determined) function $z\colon(-\infty, 0)\to H$ such that

$$(B\zeta(s), v)=(z(s), v) \qquad \forall v\in V,$$

$$\|z(s)\|_{L^2(\Omega)}\leqq \text{const}\,\|\zeta(s)\|_{H^2(\Omega)}$$

(for a.a. $s\in(-\infty, 0)$). It is easy to verify that $z\in L_h^2(-\infty, 0; H)$. Our above lemma now implies that the function

$$t\mapsto\int_{-\infty}^{0} b(t-s)z(s)\,ds$$

has its distributional derivative in $L^\infty(0, T; H)$. The definition of $z(s)$ immediately yields

$$(\bar{f}_1(t), v)=\int_{-\infty}^{0} b(t-s)(B\zeta(s), v)\,ds$$

$$=\left(\int_{-\infty}^{0} b(t-s)z(s)\,ds, v\right)$$

for all $t\in[0, T]$ and any $v\in V$.

The assertion now follows from Theorem 1.

Remark 1. Let $\bar{\Theta}$ be a sufficiently regular solution to (2.7), and let $v\in L^2(0, T; V)$ be arbitrary. Multiplying (2.7) by $v(t)-\bar{\Theta}(t)$ and integrating the fourth term on the left hand side by parts one obtains

$$\int_\Omega\left(\frac{\partial\bar{\Theta}}{\partial t}(t)+a_0\bar{\Theta}(t)+\int_{-\infty}^{t} a(t-s)\bar{\Theta}(s)\,ds\right)(v(t)-\bar{\Theta}(t))\,dx$$

$$+\int_\Omega\left(\int_{-\infty}^{t} b(t-s)\frac{\partial\bar{\Theta}}{\partial x_i}(s)\,ds\right)\left(\frac{\partial v}{\partial x_i}(t)-\frac{\partial\bar{\Theta}}{\partial x_i}(t)\right)dx$$

$$-\int_{\Gamma_2} q_i(t)n_i(v(t)-\bar{\Theta}(t))\,dS=0$$

for a.a. $t\in(0, T)$. In order to proceed we let denote $\bar{\Theta}_i(s)=(\partial\bar{\Theta}/\partial x_i)(s)$, $v_i(s)=(\partial v/\partial x_i)(s)$ (for a.a. $s\in(-\infty, T)$; $i=1, 2, 3$). Then

$$\int_\Omega\left(\int_{-\infty}^{t} b(t-s)\frac{\partial\bar{\Theta}}{\partial x_i}(s)\,ds\right)\left(\frac{\partial v}{\partial x_i}(t)-\frac{\partial\bar{\Theta}}{\partial x_i}(t)\right)dx$$

$$=\left(\int_{-\infty}^{t} b(t-s)\bar{\Theta}_i(s)\,ds, v_i(t)-\bar{\Theta}_i(t)\right)$$

$$=\int_{-\infty}^{t} b(t-s)(\bar{\Theta}_i(s), v_i(t)-\bar{\Theta}_i(t))\,ds$$

$$=\left(\int_{-\infty}^{t} b(t-s)B\bar{\Theta}(s)\,ds, v(t)-\bar{\Theta}(t)\right).$$

Hence, $\bar{\Theta}$ satisfies the equation in (2.19) (for any $v \in L^2(0, T; V)$).

Remark 2. In addition to the above assumptions on ζ, let $\zeta(s) \in K$ for a.a. $s \in (-\infty, 0)$. Then for the function u obtained in Theorem 2 the constitutive equations (2.3), (2.4) (and thus (2.3′), (2.4′)) make sense as linearization of the general response functionals in [7] (i.e. our discussion remains within the range of validity of the linearized theory in [7]).

Remark 3. In addition to our above conditions on $a(\cdot)$ and $b(\cdot)$ suppose that

$$f_1'(t) = \int_{-\infty}^{0} a'(t-s)\zeta(s)\, ds, \qquad \bar{f}_1'(t) = \int_{-\infty}^{0} b'(t-s)B\zeta(s)\, ds$$

for all $t \in [0, T]$ (this can be easily verified when both $a'(\cdot)$ and $b'(\cdot)$ are uniformly continuous on $[0, \infty)$). Then

$$\|u'(t)\|_{L^2(\Omega)} + \|u(t)\|_{H^1(\Omega)} \leqq c(1 + \|\zeta\|_{L^2_h(-\infty,0;H^2(\Omega))})$$

for a.a. $t \in (0, T)$ (the constant being dependent on T). This estimate is readily deduced from the proof of Theorem 1.

Remark 4. Let us now suppose that $u(t) \in C^1(\Omega)$ for a certain (fixed) $t \in (0, T)$ (u being the solution to (2.17)–(2.20)).

Define

$$\Omega_{0,t} = \{x \in \Omega \colon |u(x, t)| + |\nabla u(x, t)| < \delta\}.$$

Let $\psi \in \mathscr{D}(\Omega_{0,t})$ (= set of all infinitely differentiable functions in $\mathbb{R}^3$ having their support in $\Omega_{0,t}$) be arbitrarily chosen. Then there exists a real $0 < \delta^* < \delta$ such that $|u(x, t)| + |\nabla u(x, t)| \leqq \delta^*$ for all $x \in \operatorname{supp}(\psi)$. Next, the inequality in (2.19) implies

$$\left(u'(s) + a_0 u(s) + \int_{-\infty}^{s} a(s-\tau)u(\tau)\, d\tau + \int_{-\infty}^{s} b(s-\tau)Bu(\tau)\, d\tau,\, v - u(s)\right)$$

$$\geqq \int_{\Gamma_2} \chi(x, s)[v(x) - u(x, s)]\, dS$$

for all $s \in (0, T)$ and any $v \in K$ (here we have fixed a representative in the class u'). We now insert $v = u(t) \pm \varepsilon\psi$ ($\varepsilon > 0$ sufficiently small) into the latter inequality ($s = t$). Observing that

$$\left(\int_{-\infty}^{t} b(t-s)Bu(s)\, ds,\, \psi\right) = \int_{\Omega} \left[\int_{-\infty}^{t} b(t-s)\frac{\partial u}{\partial x_i}(s)\, ds\right]\frac{\partial \psi}{\partial x_i}\, dx$$

it is easily seen that $u(t)$ satisfies (2.7) in the sense of distributions in $\Omega_{0,t}$. Furthermore, it holds $u = 0$ on $\Gamma_1 \times (0, T)$ (in the trace sense) and $u(0) = u_0$ a.e. in Ω.

Finally, in addition to our above assumption we now suppose that $u(t) \in H^2(\Omega)$ and $\Omega_{0,t} = \Omega$ (for the value t under consideration). One then obtains

$$\int_{\Gamma_2} (q_i(t)n_i - \chi(t))v\, dS = 0 \qquad \forall v \in V.$$

(Note that the functions n_i are measurable on Γ and bounded by 1; cf. [15].) If $\Gamma = \Gamma_2$ the latter relation implies $q_i(t)n_i = \chi(t)$ a.e. on Γ (cf. [15]).

Addition. After submitting this paper the author became familiar with a preprint of the paper by V. Barbu and M. A. Malik, *Semilinear integro-differential equations in Hilbert spaces*. This work is concerned with the existence, uniqueness and asymptotic behavior of solutions to integrodifferential equations within a Hilbert space. Although the class of integrodifferential equations considered by these authors essentially coincides with that we have studied in § 1, the theory developed in the above mentioned paper does not apply to our situation: (1) cannot be reformulated as an equation in H (indeed, the extension of the functional φ onto H is proper and convex but need not be lower semi-continuous on H). Furthermore, we dispense with a condition of type (iii) used in the above mentioned paper.

Acknowledgment. The author is indebted to the referees for their kind advice when preparing this paper.

REFERENCES

[1] V. BARBU, *Nonlinear Volterra equations in a Hilbert space*, this Journal, 6 (1975), pp. 728–741.
[2] ———, *Nonlinear semigroups and differential equations in Banach spaces*, Ed. Acad., Bucharest, Noordhoff Int. Publ., Leyden 1976.
[3] ———, *Nonlinear Volterra integro-differential equations in Hilbert spaces*, Conf. Sem. Mat. Bari, No. 143, 1976.
[4] H. BRÉZIS, *Opérateurs maximaux monotones et semigroupes de contractions dans les espaces de Hilbert*, Math. Studies 5, North Holland, 1973.
[5] M. G. CRANDALL, S.-O. LONDEN AND J. A. NOHEL, *An abstract nonlinear Volterra integro-differential equation*, MRC Techn. Summary Report, Univ. Wisconsin, 1975.
[6] G. DUVAUT AND J. L. LIONS, *Les inéquations en mécanique et en physique*, Dunod, Paris, 1972.
[7] M. E. GURTIN AND A. C. PIPKIN, *A general theory of heat conduction with finite wave speed*, Arch. Rational Mech. Anal., 31 (1968), pp. 113–126.
[8] R. P. HERRMANN AND R. R. NACHLINGER, *On uniqueness and wave propagation in nonlinear heat conductors with memory*, J. Math. Anal. Appl., 50 (1975), pp. 530–547.
[9] J. L. LIONS, *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod, Gauthier-Villars, Paris, 1969.
[10] S.-O. LONDEN, *On an integral equation in a Hilbert space*, this Journal, 8 (1977), pp. 950–970.
[11] ———, *On an integrodifferential Volterra equation with a maximal monotone mapping*, Report-HTKK-MAT-A89 (1976), Helsinki Univ. Techn.
[12] R. C. MACCAMY, *Stability theorems for a class of functional-differential equations*, SIAM J. Appl. Math., 30 (1976), pp. 557–576.
[13] ———, *An integro-differential equation with application in heat flow*, Quart. Appl. Math., 35 (1977), pp. 1–19.
[14] R. R. NACHLINGER AND J. W. NUNZIATO, *Stability of uniform temperature fields in linear heat conductors with memory*, Int. J. Engrg. Sci., 14 (1976), pp. 693–701.
[15] J. NEČAS, *Les méthodes directes en théorie des équations elliptiques*, Academia, Prague, 1967.
[16] J. W. NUNZIATO, *On heat conduction in materials with memory*, Quart. Appl. Math., 29 (1971), pp. 187–203.
[17] K. YOSIDA, *Functional analysis*, Springer; Berlin-Göttingen-Heidelberg, 1965.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0007 $01.00/0

PERTURBATION THEORY FOR DIVISORS OF OPERATOR POLYNOMIALS*

I. GOHBERG†, P. LANCASTER‡, AND L. RODMAN¶

Abstract. This paper is devoted to perturbation theory for polynomial divisors of operator polynomials

$$L(\lambda, \mu) = I\lambda^l + A_{l-1}(\mu)\lambda^{l-1} + \cdots + A_1(\mu)\lambda + A_0(\mu),$$

where $A_0(\mu), \cdots, A_{l-1}(\mu)$ are operators acting in some Banach space and depending analytically (or continuously) on μ in some domain of the complex plane. Local behavior of the divisors is studied. For the case that the Banach space is finite dimensional, the global behavior of the divisors is studied, including description of the singularities.

Introduction. The classical problem of damped oscillations demands detailed study of differential equations of the form

$$\text{(A)} \qquad \ddot{x} + \mu B\dot{x} + Cx = 0$$

where B, C are constant Hermitian positive definite matrices and μ is a nonnegative parameter. It is natural to study this equation as a problem of analytic perturbation of that for harmonic oscillations:

$$\ddot{x} + Cx = 0.$$

In this way, one is led to consider the dependence of solutions of (A) on μ and, hence, to the dependence on μ of eigenvalues and generalized eigenvectors for the matrix polynomial

$$L(\lambda, \mu) = I\lambda^2 + \mu B\lambda + C.$$

This may be described as a spectral perturbation problem.

However, M. G. Krein and H. Langer [12] pioneered a second approach to the solution of (A). For example, if μ is small in the sense

$$\mu < \inf_{x \in \mathbb{C}_n, x \neq 0} \{2(Cx, x)^{1/2}/(Bx, x)\},$$

any solution of (A) can be written in the form

$$\text{(B)} \qquad x(t) = e^{tZ_+}x_1 + e^{tZ_-}x_2$$

where Z_+, Z_- are solutions of the matrix equation $Z^2 + \mu BZ + C = 0$ and have their spectra in the upper and lower halves of the complex plane, respectively. These matrices depend on μ and the problem is now to study such matrix valued functions $Z_+(\mu)$, $Z_-(\mu)$ for which

$$Z_+(0) = iC^{1/2}, \qquad Z_-(0) = -iC^{1/2}.$$

Both polynomials $I\lambda - Z_\pm(\mu) = L_{1,2}(\lambda, \mu)$ are evidently right divisors of $L(\lambda, \mu)$, and this formulation is a problem in the perturbation of divisors of polynomials. In this case, there are special restrictions imposed on the divisors.

* Received by the editors December 22, 1977, and in revised form June 20, 1978.

† Department of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel, and Department of Pure Mathematics, Weizmann Institute of Science, Rehovot, Israel.

‡ Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

¶ Department of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel.

In developing further some results of R. J. Duffin [2], Krein and Langer also show that solutions of (A) take the form (B) in the case when μ is sufficiently large. Thus, the condition for strong damping is

$$\mu > \sup_{x \in \mathbb{C}_n, x \neq 0} \{2(Cx, x)^{1/2}/(Bx, x)\},$$

and implies that $L(\lambda, \mu)$ has right divisors of first degree which yield solutions of the differential equation in the form (B). Our investigations will throw some light on the existence and behavior of divisors of $L(\lambda, \mu)$ as μ ranges over all complex values.

More generally, this paper is devoted to analytic perturbation theory for divisors of operator polynomials

$$L(\lambda, \mu) = I\lambda^l + A_{l-1}(\mu)\lambda^{l-1} + \cdots + A_1(\mu)\lambda + A_0(\mu)$$

where $A_0(\mu), \cdots, A_{l-1}(\mu)$ depend analytically on μ in some domain Ω of the complex plane. In particular, suppose that

$$L_1(\lambda) = I\lambda^k + B_{k-1}\lambda^{k-1} + \cdots + B_1\lambda + B_0$$

is a right divisor of L at $\mu = \mu_0$ and that L_1 has its spectrum in the interior of a bounded closed contour Γ. Suppose, in addition, that the quotient $L_2(\lambda) = L(\lambda, \mu_0)L_1^{-1}(\lambda)$ has its spectrum outside Γ. Then $L_1(\lambda)$ is called a Γ-spectral divisor of $L(\lambda, \mu_0)$. One result of this paper shows that there will always exist Γ-spectral divisors for $L(\lambda, \mu)$ in a neighborhood of μ_0. This divisor, $L_1(\lambda, \mu)$ can be considered as a perturbation of $L_1(\lambda) = L_1(\lambda, \mu_0)$. If the coefficients $A_j(\mu)$ are merely continuous it turns out that $L_1(\lambda, \mu)$ has coefficients continuous in μ. Similarly, the coefficients of $L_1(\lambda, \mu)$ will inherit analyticity from those of $L(\lambda, \mu)$. These results, among others, are developed in the first chapter for the infinite dimensional case. The main tools are theorems about divisors from our paper [6]. In that paper a correspondence is developed between divisors and certain subspaces known as "supporting subspaces." In general, this dependence on μ of the original polynomial and of this supporting subspace completely determines the dependence of the divisor on μ.

The first section of this paper has, in priciple, a local character. In the second the global properties of divisors are to be investigated and attention is confined to the finite-dimensional case. The nature of the singularities of divisors is explored as well as the existence of divisors which do not admit analytic continuation. The simplest example of the latter is $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$, which is an isolated divisor of $\begin{bmatrix} \lambda^2 & 0 \\ \mu & \lambda^2 \end{bmatrix}$ at $\mu = 0$.

Section 1 consists of a detailed investigation of the direct and inverse dependence of divisors and supporting subspaces. This involves the use of a metric on the set of subspaces of a Banach space which is very close to the "gap" of M. G. Krein, M. A. Krasnoselskii and D. P. Milman (see [11] and Gohberg–Markus [8]). Sections 1.1 and 1.2 have a preliminary character and in §§ 1.3 and 1.4 the continuous and analytic dependence of divisors, respectively, are studied.

Section 2, as mentioned above, deals only with the finite dimensional case. A very important role is played here by theorems of Baumgärtel on analytic perturbation theory for matrices [1]. In § 2.1 the necessary corollaries of Baumgärtel's results are developed. In the remaining sections applications are made to continuations of divisors in the large and to the study of singularities of the divisors.

From the point of view of *systems theory* the paper is concerned with systems (T, X, Y) with frequency response function equal to the inverse of a monic matrix polynomial

L. The authors' earlier work has developed and explored the one-to-one correspondence between factorizations L_1L_2 of L and certain invariant subspaces of T. The relevant results are summarized in § 1.1. The contributions of this paper include the behavior of these factorizations and subspaces when the coefficients of the polynomial depend analytically on a parameter, in both the finite dimensional and Banach space context.

1. Local perturbations.

1.1. Algebraic preliminaries. Let $\mathscr{X}$ be a complex Banach space and $\mathscr{L}(\mathscr{X})$ the algebra of bounded linear operators on $\mathscr{X}$. A *monic operator polynomial* (more briefly, a m.o.p.) is a function from the complex numbers to $\mathscr{L}(\mathscr{X})$ of the form

$$L(\lambda)=I\lambda^l+A_{l-1}\lambda^{l-1}+\cdots+A_1\lambda+A_0 \tag{1}$$

where $A_0, A_1, \cdots, A_{l-1}$ and the identity I are members of $\mathscr{L}(\mathscr{X})$. The symbol $\mathscr{X}^r$ denotes the direct sum of r copies of $\mathscr{X}$. It has been shown in the authors' earlier papers [4], [5] and [6] that spectral information about L can be concentrated in certain pairs of operators. They are defined as follows: Operators $X\in\mathscr{L}(\mathscr{X}^l, \mathscr{X})$ and $T\in\mathscr{L}(\mathscr{X}^l)$ form a *standard pair* for the m.o.p. L of (1) if

$$A_0X+A_1XT+\cdots+A_{l-1}XT^{l-1}+XT^l=0$$

and the operator Q defined by

$$Q=\begin{bmatrix} X \\ XT \\ \vdots \\ XT^{l-1}\end{bmatrix} \tag{2}$$

is invertible.

An important example of a standard pair consists of an operator $X=[I\ 0\ 0\cdots 0]$ and $T=C_L$, the *first companion operator* for L defined by

$$C_L=\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & I \\ -A_0 & -A_1 & \cdots & -A_{l-2} & -A_{l-1}\end{bmatrix}.$$

In this case the operator Q of (2) is the identity operator on $\mathscr{X}^l$.

Note that, for any $X\in\mathscr{L}(\mathscr{X}^l, \mathscr{X})$ and $T\in\mathscr{L}(\mathscr{X}^l)$ l satisfying only $\sum_{i=0} A_iXT^i=0$, and with Q defined by (2), we have $QT=C_LQ$. Thus, X, T a standard pair implies that T is similar to C_L.

The second example of a standard pair (to be exploited in § 2) is confined to finite dimensional spaces $\mathscr{X}$. In this case, let J be a matrix in Jordan normal form representing C_L. Then J is $ln\times ln$ and a $n\times ln$ matrix X for which X, J form a standard pair will necessarily have the chains of generalized eigenvectors (Jordan chains) for L displayed as its columns. Such a pair is described as *canonical.*

In earlier papers the authors have studied the correspondence between monic right divisors of L (i.e., m.o.p. L_1 for which $L=L_2L_1$ for some m.o.p. L_2) and certain invariant subspaces of operators T belonging to standard pairs. For future reference we present this result and the complementary result concerning representation of quotient polynomials.

First define an operator $Q_k: \mathscr{X}^l \to \mathscr{X}^k$ $(1 \leqq k \leqq l)$ in terms of a standard pair X, T by

$$Q_k = \begin{bmatrix} X \\ XT \\ \vdots \\ XT^{k-1} \end{bmatrix}. \tag{3}$$

PROPOSITION 1. *Let L be a m.o.p. on $\mathscr{X}$ of degree l with standard pair X, T. Let $\mathscr{M}$ be an invariant subspace of T on which Q_k, viewed as an operator from $\mathscr{M}$ to $\mathscr{X}^k$, is invertible. Then $\mathscr{M} = \mathscr{I}m\, R_\Gamma$ where R_Γ is the Riesz projector corresponding to T and Γ:*

$$L_1(\lambda) = I\lambda^k - XT^k(W_1 + W_2\lambda + \cdots + W_k\lambda^{k-1}) \tag{4}$$

where $W_1, \cdots, W_k \in \mathscr{L}(\mathscr{X}, \mathscr{M})$ are defined by

$$[W_1 \cdots W_k] = (Q_k|_{\mathscr{M}})^{-1}. \tag{5}$$

Conversely, if Γ is a contour of regular points and $\mathscr{M} = \mathscr{I}m\, R_\Gamma$ is the supporting subspace $\mathscr{M}$ of T such that $Q_k|_{\mathscr{M}}$ is invertible and (4) *holds.*

The invariant subspace $\mathscr{M}$ of T associated uniquely with a divisor L_1 is called a *supporting subspace* for L with respect to T.

There is an asymmetry about the notion of a standard pair X, T which is removed by the introduction of *standard triples*. For a full discussion see [6], but for our purposes observe that X, T, Y form a standard triple iff X, T are a standard pair, $Y \in \mathscr{L}(\mathscr{X}, \mathscr{X}^l)$ and

$$XT^{r-1}Y = \begin{cases} 0, & r = 1, 2, \cdots, l-1, \\ I, & r = l. \end{cases}$$

This idea is useful in the description of quotient operator polynomials.

A standard triple is used in the following description of the quotient L_2 associated with the right divisor L_1 of Proposition 1. First define an operator $R_{l-k}: \mathscr{X}^{l-k} \to \mathscr{X}^l$ in terms of operators T, Y of a standard triple by

$$R_{l-k} = [Y, TY, \cdots, T^{l-k}Y]. \tag{6}$$

PROPOSITION 2. *Let L be a m.o.p. with standard triple X, T, Y. Let $L = L_2L_1$ where L_1, L_2 are m.o.p. of degrees k and $l-k$ respectively, and $\mathscr{M}$ be the supporting subspace of L_1 with respect to T. Then*

$$\mathscr{X}^l = \mathscr{M} \oplus \mathscr{I}m\, R_{l-k} \tag{7}$$

and the map $R: \mathscr{X}^{l-k} \to \mathscr{I}m\, R_{l-k}$ generated by R_{l-k} is invertible.

For $1 \leqq j \leqq l-k$ define $Z_k: \mathscr{I}m\, R_{l-k} \to \mathscr{X}$ by

$$R^{-1} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{l-k} \end{bmatrix}$$

and let $P \in \mathscr{L}(\mathscr{X}^l)$ be the projection on $\mathscr{I}m\, R_{l-k}$ along $\mathscr{M}$. Then

$$L_2(\lambda) = I\lambda^{l-k} - (Z_1 + Z_2\lambda + \cdots + Z_{l+k}\lambda^{l-k-1})PT^{l-k}PY. \tag{8}$$

Observe that the standard pair $X=[I\ 0\cdots 0]$, C_L generates the triple X, C_L, Y with $Y=[0\ 0\cdots 0\ I]^T$ (where T denotes the transpose) and that, in this case

$$\mathcal{I}m\, R_{l-k}=\{x\in\mathscr{X}^l : x_1=x_2=\cdots=x_k=0\}. \tag{9}$$

Note also that the operator Q_k of (3) has the representation $Q_k=[I_k, 0]$ when using this standard triple, and $\mathcal{M}$ is a supporting subspace iff it is C_L invariant and is a direct complement for $\mathcal{I}m\, R_{l-k}$ as defined in (9).

There is a class of divisors of a m.o.p. L which are of particular interest because of the simple relationship of their spectra to the spectrum of L. To define these more precisely some familiar concepts are needed. First, the *resolvent set* of L, written Res (L), is the set of complex numbers λ for which $L(\lambda)$ has an inverse in $\mathscr{L}(\mathscr{X})$. Then the *spectrum* of L, written $\sigma(L)$, is the complement of Res (L) in $\mathbb{C}$. Each point of Res (L) is called a *regular point* for L. If Γ is a closed contour in the complex plane consisting of regular points of L, then a monic right divisor L_1 of L is a Γ*-spectral* (*right*) *divisor* of L if $L=L_2L_1$ and $\sigma(L_1)$, $\sigma(L_2)$ are inside and outside Γ, respectively. For such divisors a supporting subspace can be described as the image of a Riesz projector. The fundamental result in this direction is Theorem 19 of [6] which is now presented, but see also Theorems 23 and 24 of [5].

PROPOSITION 3. *Let L be a m.o.p. with standard triple X, T, Y and let L_1 be a Γ-spectral right divisor of L having associated supporting subspace $\mathcal{M}$ with respect to T. Then $\mathcal{M}=\mathcal{I}m\, R_\Gamma$ where R_Γ is the Riesz projector corresponding to T and Γ:*

$$R_\Gamma=\frac{1}{2\pi i}\oint_\Gamma (I\lambda-T)^{-1}\,d\lambda.$$

Conversely, if Γ is a contour of regular points and $\mathcal{M}=\mathcal{I}m\, R_\Gamma$ is the supporting subspace for the monic right divisor L_1 of L with respect to T, then L_1 is a Γ-spectral right divisor.

1.2. Topological preliminaries. Denote by $\mathscr{A}$ the class of all subspaces of $\mathscr{X}^l$. The ideas of an "opening" between subspaces and of the "spherical metric" are now to be introduced. The first of these originates in the work of Krein, Krasnoselskii and Milman [11] and led to the development of the second by Gohberg and Markus [8]. Good descriptions of the subject can be found in the works of Gohberg and Krein [3] and of Kato [10].

If $x\in\mathscr{X}_l$ and J is any subset of $\mathscr{X}_l$, let $\rho(x, J)$ denote the distance from x to J. If $\mathcal{M}\in\mathscr{A}$, then $S_\mathcal{M}$ will denote the unit sphere in $\mathcal{M}$. The *opening* between subspaces $\mathcal{M}, \mathcal{N}\in\mathscr{A}$ can now be defined as

$$\theta(\mathcal{M},\mathcal{N})=\max\{\sup_{x\in S_\mathcal{M}}\rho(x,\mathcal{N}),\ \sup_{x\in S_\mathcal{N}}\rho(x,\mathcal{M})\}.$$

If $P_\mathcal{M}$, $P_\mathcal{N}$ are any two projections in $\mathscr{L}(\mathscr{X})$ for which $\mathcal{I}m\, P_\mathcal{M}=\mathcal{M}$ and $\mathcal{I}m\, P_\mathcal{N}=\mathcal{N}$ then

$$\theta(\mathcal{M},\mathcal{N})\leqq\|P_\mathcal{M}-P_\mathcal{N}\| \tag{10}$$

applies. To see this let $x\in S_\mathcal{M}$, so that

$$\|x-P_\mathcal{N}x\|=\|P_\mathcal{M}x-P_\mathcal{N}x\|\leqq\|P_\mathcal{M}-P_\mathcal{N}\|$$

and, consequently $\rho(x,\mathcal{N})=\inf_{y\in\mathcal{N}}\|x-y\|\leqq\|P_\mathcal{M}-P_\mathcal{N}\|$. Similarly, for any $y\in S_\mathcal{N}$ it is found that $\rho(y,\mathcal{M})\leqq\|P_\mathcal{M}-P_\mathcal{N}\|$. Then (10) follows from the definition of θ.

The opening cannot be used to define a metric on $\mathscr{A}$ since, in general, it does not satisfy the triangle inequality axiom. However, the *spherical metric* $\tilde{\theta}$ obtained by modifying the definition is indeed a metric. If $\mathcal{M}, \mathcal{N} \in \mathcal{A}$ define

$$\tilde{\theta}(\mathcal{M}, \mathcal{N}) = \max\{\sup_{x \in S_{\mathcal{N}}} \rho(x, S_{\mathcal{M}}), \sup_{x \in S_{\mathcal{M}}} \rho(x, S_{\mathcal{N}})\}.$$

The opening and the spherical metric are then related by the inequalities:

$$\theta(\mathcal{M}, \mathcal{N}) \leqq \tilde{\theta}(\mathcal{M}, \mathcal{N}) \leqq 2\theta(\mathcal{M}, \mathcal{N}) \tag{11}$$

for any $\mathcal{M}, \mathcal{N} \in \mathcal{A}$.

LEMMA 1. *Let $\mathcal{M}, \mathcal{M}_1 \in \mathcal{A}$ and $\mathcal{M} \oplus \mathcal{M}_1 = \mathcal{X}^l$. If $\mathcal{N} \in \mathcal{A}$ and $\tilde{\theta}(\mathcal{M}, \mathcal{N})$ is sufficiently small, then*

$$\mathcal{N} \oplus \mathcal{M}_1 = \mathcal{X}^l \tag{12}$$

and

$$\tfrac{1}{2}\tilde{\theta}(\mathcal{M}, \mathcal{N}) \leqq \|P_{\mathcal{M}} - P_{\mathcal{N}}\| \leqq C\tilde{\theta}(\mathcal{M}, \mathcal{N}) \tag{13}$$

where $P_{\mathcal{M}}$ ($P_{\mathcal{N}}$) projects $\mathcal{X}^l$ onto $\mathcal{M}$ (onto $\mathcal{N}$) along $\mathcal{M}_1$ and C is a constant depending on $\mathcal{M}$ and $\mathcal{M}_1$ but not on $\mathcal{N}$.

Proof. The decomposition (12) follows from Theorem 2 of [8]. In order to prove (13) observe that, by use of (11) it is sufficient to prove that

$$\theta(\mathcal{M}, \mathcal{N}) \leqq \|P_{\mathcal{M}} - P_{\mathcal{N}}\| \leqq C\theta(\mathcal{M}, \mathcal{N}),$$

and the left inequality is just (10) which holds for *any* $\mathcal{M}, \mathcal{N}$.

To establish the right-hand inequality in (13) two preliminary remarks are needed. First note that for any $x \in \mathcal{M}$, $y \in \mathcal{M}_1$, $x = P_{\mathcal{M}}(x+y)$ so that

$$\|x+y\| \geqq \|P_{\mathcal{M}}\|^{-1}\|x\|. \tag{14}$$

It is claimed that, for $\theta(\mathcal{M}, \mathcal{N})$ small enough,

$$\|z+y\| \geqq \tfrac{1}{2}\|P_{\mathcal{M}}\|^{-1}\|z\| \tag{15}$$

for all $z \in \mathcal{N}$ and $y \in \mathcal{M}_1$.

Without loss of generality assume $\|z\| = 1$. Suppose $\theta(\mathcal{M}, \mathcal{N}) < \delta$ and let $x \in \mathcal{M}$. Then, using (14) we obtain

$$\|z+y\| \geqq \|x+y\| - \|z-x\| \geqq \|P_{\mathcal{M}}\|^{-1}\|x\| - \delta.$$

But then $x = (x-z)+z$ implies $\|x\| \geqq 1-\delta$ and so

$$\|z+y\| \geqq \|P_{\mathcal{M}}\|^{-1}(1-\delta) - \delta$$

and, for δ small enough, (15) is established.

The second remark is that, for any $x \in \mathcal{X}^l$,

$$\|x - P_{\mathcal{M}}x\| \leqq C_0\rho(x, \mathcal{M}) \tag{16}$$

for some constant C_0. This follows from the topological equivalence of the complementary subspace $\mathcal{M}_1$ of $\mathcal{M}$ (in which $x - P_{\mathcal{M}}x$ lies) and the factor space $\mathcal{X}^l | \mathcal{M}$ with its usual norm ($\|x + \mathcal{M}\| = \rho(x, \mathcal{M})$).

Now for any $x \in S_{\mathcal{N}}$, by use of (16),

$$\|(P_{\mathcal{M}} - P_{\mathcal{N}})x\| = \|x - P_{\mathcal{M}}x\| \leqq C_0\rho(x, \mathcal{M}) \leqq C_0\theta(\mathcal{M}, \mathcal{N}).$$

Then, if $\omega \in \mathscr{X}^l$, $\|\omega\| = 1$, and $\omega = y + z$, $y \in \mathcal{N}$, $z \in \mathcal{M}_1$,

$$\|(P_{\mathcal{M}} - P_{\mathcal{N}})\omega\| = \|(P_{\mathcal{M}} - P_{\mathcal{N}})y\| \leqq \|y\| C_0 \theta(\mathcal{M}, \mathcal{N}) \leqq 2C_0 \|P_{\mathcal{M}}\| \theta(\mathcal{M}, \mathcal{N})$$

and the last inequality follows from (15). This completes the proof of the lemma. □

Let $\mathscr{P}_k$ be the class of all monic operator polynomials of degree k acting on $\mathscr{X}$. It is easily verified that the function σ defined on $\mathscr{P}_k \times \mathscr{P}_k$ by

$$\sigma\left(I\lambda^k + \sum_{i=0}^{k-1} \lambda^i B_i, \quad I\lambda^k + \sum_{i=0}^{k-1} \lambda^i B_i'\right) = \sum_{i=0}^{k-1} \|B_i - B_i'\| \tag{17}$$

is a metric and $\mathscr{P}_k$ will now be viewed as the corresponding metric space.

Consider the set $\mathscr{W} \subset \mathscr{A} \times \mathscr{P}_l$ consisting of pairs $\{\mathcal{M}, L(\lambda)\}$ for which $\mathcal{M}$ is an invariant subspace in $\mathscr{X}^l$ for the companion operator C_L of L. This set $\mathscr{W}$ will be called the *supporting set* and is provided with the topology induced from $\mathscr{A} \times \mathscr{P}_l$.

Then define the subset $\mathscr{W}_k \subset \mathscr{W}$ consisting of all pairs $\{\mathcal{M}, L(\lambda)\}$ where $L(\lambda) \in \mathscr{P}_l$ and $\mathcal{M}$ is a supporting subspace (with respect to C_L) associated with a monic right divisor of L of degree k. The set $\mathscr{W}_k$ is called the *supporting set* of order k.

THEOREM 1. *$\mathscr{W}_k$ is open in $\mathscr{W}$.*

Proof. Define the subspace $\mathscr{Y}_{l-k}$ of $\mathscr{X}^l$ by the condition $x = (x_1, \cdots, x_l) \in \mathscr{Y}_{l-k}$ iff $x_1 = \cdots = x_k = 0$. We have seen ((7), (9)) that, if $\mathcal{M}$ is a supporting subspace for L, with respect to C_L, then $\mathcal{M} \oplus \mathscr{Y}_{l-k} = \mathscr{X}^l$. Conversely, if $\mathcal{M}$ is an invariant subspace for C_L and $\mathcal{M} \oplus \mathscr{Y}_{l-k} = \mathscr{X}^l$, then it is apparent that the operator $Q_k = [I_k, 0]$ of Proposition 1 is invertible on $\mathcal{M}$ so that $\mathcal{M}$ is a supporting subspace for L. In other words $(\mathcal{M}, L(\lambda)) \in \mathscr{W}$ is a member of $\mathscr{W}_k$ iff $\mathcal{M} \oplus \mathscr{Y}_{l-k} = \mathscr{X}^l$.

Now let $(\mathcal{M}, L) \in \mathscr{W}_k$ and let $(\hat{\mathcal{M}}, \hat{L}) \in \mathscr{W}$ be in an ε-neighborhood of $(\mathcal{M}, \mathscr{L})$. Then certainly $\tilde{\theta}(\mathcal{M}, \hat{\mathcal{M}}) < \varepsilon$. By choosing ε small enough it can be guaranteed, using Lemma 1, that $\hat{\mathcal{M}} \oplus \mathscr{Y}_{l-k} = \mathscr{X}^l$. Consequently, $\hat{\mathcal{M}}$ is a supporting subspace for $\hat{L}$ and $(\hat{\mathcal{M}}, \hat{L}) \in \mathscr{W}_k$. □

We now wish to show that, if $\mathcal{M}$ corresponds to a Γ-spectral divisor of L (Proposition 3), then there is a neighborhood of $(\mathcal{M}, L)$ in $\mathscr{W}_k$ consisting of pairs $(\hat{\mathcal{M}}, \hat{L})$ for which the spectral property is retained. A pair $(\mathcal{M}, L)$ for which $\mathcal{M}$ corresponds to a Γ-spectral divisor of L will be said to be *Γ-spectral.*

LEMMA 2. *Let $(\mathcal{M}, L) \in \mathscr{W}$ be Γ-spectral. Then there is a neighborhood of $(\mathcal{M}, \mathscr{L})$ in $\mathscr{W}$ consisting of Γ-spectral pairs.*

Proof. Let $L_0 \in \mathscr{P}_l$ and be close to L in the σ norm. Then C_{L_0} will be close to C_L and, since each isolated part of the spectrum of L is upper semicontinuous [10, Thm. IV.3.12], L_0 can be chosen to that $I\lambda - C_{L_0}$ is invertible for all $\lambda \in \Gamma$. Let Δ be a closed contour which does not intersect with Γ and contains in its interior exactly those parts of the spectrum of L which are outside Γ. Then we may assume L_0 chosen in such a way that the spectrum of L_0 also consists of two parts; one part inside Γ and the other part inside Δ.

Define $\mathcal{N}$, $\mathcal{N}_0$, $\mathcal{M}_0$ to be the images of the Riesz projectors determined by (Δ, L), (Δ, L_0) and (Γ, L_0), respectively. Then

$$\mathcal{M} \oplus \mathcal{N} = \mathcal{M}_0 \oplus \mathcal{N}_0 = \mathscr{X}^l,$$

$(\mathcal{M}_0, L_0)$ is Γ-spectral, and if $L_0 \to L$ then $\mathcal{M}_0 \to \mathcal{M}$.

Now let $(\mathcal{M}_1, L_0) \in \mathscr{W}$ be in a neighborhood N of $(\mathcal{M}, L)$. We are finished if we can show that for N small enough, $\mathcal{M}_1 = \mathcal{M}_0$. First, N can be chosen small enough for *both* L_0 to be chosen close to L (whence $\mathcal{M}_0$ to $\mathcal{M}$) *and* $\mathcal{M}$ to be so close to $\mathcal{M}_1$, that $\mathcal{M}_1$ and $\mathcal{M}_0$ are arbitrarily close. Thus the first statement of Lemma 1 will apply and, if N is chosen small enough, then $\mathcal{M}_1 \oplus \mathcal{N}_0 = \mathscr{X}^l$.

Now $\mathcal{M}_1$ invariant under C_{L_0} implies the invariance of $\mathcal{M}_1$ under the projector

$$P_0 = \frac{1}{2\pi i} \int_\Gamma (\lambda I - C_{L_0})^{-1} \, d\lambda.$$

Applying this projector to the decompositions $\mathcal{M}_1 \oplus \mathcal{N}_0 = \mathcal{M}_0 \oplus \mathcal{N}_0 = \mathcal{X}^l$ it is found that $\mathcal{M}_1 = \mathcal{M}_0$. $\square$

THEOREM 2. *Let $(\mathcal{M}, L) \in \mathcal{W}_k$ be Γ-spectral. Then there is a neighborhood of $(\mathcal{M}, L)$ in $\mathcal{W}$ consisting of Γ-spectral pairs from $\mathcal{W}_k$.*

Proof. This follows immediately from Theorem 1 and Lemma 2. $\square$

Note that in the discussion of this theorem it is not essential for Γ to be fixed. For example, let the coefficients of L depend continuously on a parameter ζ and suppose that there is a Γ-spectral divisor at ζ_0 containing the part $\sigma(L_1(\zeta_0))$ of $\sigma(L(\zeta_0))$. Then provided that the set $\sigma(L_1(\zeta))$, for each fixed ζ in a neighborhood of ζ_0, is separated from its complement in $\sigma(L(\zeta))$, a spectral divisor with respect to a ζ-dependent contour Γ_ζ can be constructed. In some cases it will be advantageous to consider Γ_ζ-spectral divisors of this kind.

1.3. The continuous dependence of the supporting subspace on the coefficients. Define a map $F_k: \mathcal{W}_k \to \mathcal{P}_{l-k} \times \mathcal{P}_k$ in the following way: The image of $(\mathcal{M}, L) \in \mathcal{W}_k$ is to be the pair of m.o.p. (L_2, L_1) where L_1 is the right divisor of L associated with $\mathcal{M}$ and L_2 is the quotient obtained on division of L on the right by L_1. It is evident that F_k is one-to-one and surjective so that the map F_k^{-1} exists.

Introduce the topology generated by (17) to $\mathcal{P}_{l-k} \times \mathcal{P}_k$ and define the space $\mathcal{P}_0$ to be the disconnected union: $\mathcal{P}_0 = \bigcup_{k=1}^{l-1} (\mathcal{P}_{l-k} \times \mathcal{P}_k)$. In view of Theorem 1, the space $\mathcal{W}_0 = \bigcup_{k=1}^{l-1} \mathcal{W}_k$ is also a disconnected union of its subspaces $\mathcal{W}_1, \cdots, \mathcal{W}_{l-1}$. The map F between topological spaces $\mathcal{W}_0$ and $\mathcal{P}_0$ can now be defined by the component maps $F_1, \cdots, F_{l-1}$, and F will be invertible.

If X_1, X_2 are metric spaces with metrics ρ_1, ρ_2, respectively, the map $G: X_1 \to X_2$ is called *locally Lipschitz continuous* if, for every $x \in X_1$, there is a deleted neighborhood U_x of x for which

$$\sup_{y \in U_x} \frac{\rho_2(Gx, Gy)}{\rho_1(x, y)} < \infty.$$

THEOREM 3. *The maps F and F^{-1} are locally Lipschitz continuous.*

Proof. It is sufficient to prove that F_k and F_k^{-1} are locally Lipschitz continuous for $k = 1, 2, \cdots, l-1$. Let $(\mathcal{M}, L) \in \mathcal{W}_k$ and

$$F_k(\mathcal{M}, L) = (L_2, L_1).$$

Then (Proposition 1) L_1 has the representation of (4), and (Proposition 2) L_2 has the representation of (8). Note also the remarks immediately following Proposition 2.

To prove the required continuity properties of $F_k: \mathcal{W}_k \to \mathcal{P}_{l-k} \times \mathcal{P}_k$ it is necessary to estimate the distance between pairs (L_2, L_1), $(\hat{L}_2, \hat{L}_1)$ in $\mathcal{P}_{l-k} \times \mathcal{P}_k$ using the topology determined by the metric σ of (11). Note first that if $P_{\mathcal{M}}$ is the projection on $\mathcal{M}$ along $\mathcal{Y}_{l-k}$ then $P = I - P_{\mathcal{M}}$ (the projector appearing in (8)), and that $(Q_k|_{\mathcal{M}})^{-1} = P_{\mathcal{M}} C_{kl}$ in (5), where $C_{kl}: \mathcal{X}^k \to \mathcal{X}^l$ is the imbedding of $\mathcal{X}^k$ onto the first k components of $\mathcal{X}^l$. Then observe that in the representations (4) and (8) the coefficients of L_1 and L_2 can be uniformly bounded in some neighborhood of $(\mathcal{M}, L)$. Using standard procedures it is then easily seen that, in order to establish the continuity required of F_k, it is sufficient to verify the assertion: For a fixed $(\mathcal{M}, L) \in \mathcal{W}_k$ there exist positive constants δ and C such that, for any $\mathcal{N} \in \mathcal{A}$ satisfying $\tilde{\theta}(\mathcal{M}, \mathcal{N}) < \delta$, it follows that $\mathcal{N} \oplus \mathcal{Y}_{l-k} = \mathcal{X}^l$ and

$$\|P_{\mathcal{M}} - P_{\mathcal{N}}\| \leqq C\tilde{\theta}(\mathcal{M}, \mathcal{N}),$$

where $P_{\mathcal{N}}$ is the projector of $\mathcal{X}^l$ on $\mathcal{N}$ along $\mathcal{Y}_{l-k}$. But this conclusion can be drawn from Lemma 1 and so the Lipschitz continuity of F_k follows.

To establish the local Lipschitz continuity of F_k^{-1} we consider a fixed $(L_2, L_1) \in \mathcal{P}_{l-k} \times \mathcal{P}_k$. It is apparent that the polynomial $L = L_2 L_1$ will be a Lipschitz continuous function of L_2 and L_1 in a neighborhood of the fixed pair. To examine the behavior of the gap between supporting subspaces associated with neighboring pairs we observe an explicit construction for $P_{\mathcal{M}}$, the projection on $\mathcal{M}$ along $\mathcal{Y}_{l-k}$ (associated with $(L_2 - L_1)$), obtained in [4] (see particularly equations (59) and (36)). In fact, $P_{\mathcal{M}}$ has the representation

$$P_{\mathcal{M}} \begin{bmatrix} I & 0 \\ F & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} P_1 C_{L_1}^k \\ \vdots \\ P_1 C_{L_1}^{l-1} \end{bmatrix} \tag{18}$$

with respect to the decomposition $\mathcal{X}^l = \mathcal{X}^k \oplus \mathcal{X}^{l-k}$, where $P_1 = [I\ 0 \cdots 0]$. The local Lipschitz continuity of $\mathcal{I}m\,(P_{\mathcal{M}})$ as a function of L_2 and L_1 is apparent from this representation.

Given the appropriate continuous dependence of L and $\mathcal{M}$ the conclusion now follows from the left-hand inequality of (13) (Lemma 1). □

Note that if $L_0(\lambda) = I\lambda^l$ then for $1 \leqq k \leqq l-1$, $I\lambda^k$ is a divisor of L_0 with associated supporting subspace $\mathcal{X}^k$. For any $\gamma > 0$ define

$$\mathcal{D}_k(\gamma) = \{(\mathcal{M}, L) \in \mathcal{W}_k : \|P_{\mathcal{M}} - P_{\mathcal{X}^k}\| \leqq \gamma \text{ and } \sigma(L(\lambda), I\lambda^l) \leqq \gamma\}$$

and then $\mathcal{D}(\gamma) = \bigcup_{k=1}^{l-1} \mathcal{D}_k(\gamma)$. It is claimed that a Lipschitz constant $C = C(\gamma)$ can be chosen for F which will be valid for all $(\mathcal{M}, L) \in \mathcal{D}(\gamma)$. Such a result is generally true for arbitrarily locally Lipschitz continuous functions defined on a compact domain $\mathcal{D}$ and is therefore true in the case of a finite dimensional space $\mathcal{X}$. However, it follows from the proof of Theorem 3 that such a constant can be found for F and *any* Banach space $\mathcal{X}$.

Similarly, it is possible to choose a local Lipschitz constant $C = C(\gamma)$ for F^{-1} which will apply at every $(L_2, L_1) \in E(\gamma)$ where $E(\gamma) = \bigcup_{k=1}^{l-1} E_k(\gamma)$ and

$$E_k(\gamma) = \{(L_2, L_1) \in \mathcal{P}_{l-k} \times \mathcal{P}_k : \sigma(I\lambda^{l-k}, L_2(\lambda)) \leqq \gamma \text{ and } \sigma(I\lambda^k, L_1(\lambda)) \leqq \gamma\}.$$

COROLLARY 1. *Let $L(\lambda, \mu) = \sum_{i=0}^{l} A_i(\mu)\lambda^i$ be a m.o.p. whose coefficients $A_i(\mu)$ depend continuously on a parameter μ in an open set $\mathcal{D}$ of the complex plane. Assume that there exists for each $\mu \in \mathcal{D}$ a monic right divisor $L_1(\lambda, \mu) = \sum_{i=0}^{k} B_i(\mu)\lambda^i$ of L with quotient $L_2(\lambda, \mu) = \sum_{j=0}^{l-k} C_j(\mu)\lambda^j$. Let $\mathcal{M}(\mu)$ be the supporting subspace of L with respect to C_L and corresponding to L_1.*

(i) *If $\mathcal{M}(\mu)$ is continuous on $\mathcal{D}$ (in the metric $\tilde{\theta}$) then both L_1 and L_2 are continuous on $\mathcal{D}$.*

(ii) *If one of L_1, L_2 is a continuous function of μ on $\mathcal{D}$ then the second is also continuous on $\mathcal{D}$, as is $\mathcal{M}(\mu)$.*

Proof. Part (i) follows immediately from the continuity of $(\mathcal{M}, L)$ and of F since (L_2, L_1) is the image of the composition of two continuous functions.

For part (ii) suppose first that L_1 is continuous in μ on $\mathcal{D}$. Then, using (18), it is clear that $\mathcal{M}(\mu) = \mathcal{I}m\,(P_{\mathcal{M}(\mu)})$ depends continuously on μ. The continuity of L_2 follows as in part (i).

Given the continuity of L_2 on $\mathscr{D}$ consider the m.o.p. L^* defined on $\mathscr{X}^*$ by

$$L^*(\lambda, \mu) = \sum_{i=0}^{l} A_i(\mu)^* \lambda^i$$

with L_1^*, L_2^* defined similarly. Then, $L^*(\lambda, \mu) = L_1^*(\lambda, \mu) L_2^*(\lambda, \mu)$ and, since norms are invariant under the operation of taking adjoints, $L_2^*(\lambda, \mu)$ will be continuous in μ on $\mathscr{D}$. Applying the preceding argument the continuous dependence of L_1^* on μ, and hence of L_1, is obtained. The continuity of $\mathscr{M}(\mu)$ on $\mathscr{D}$ then follows from the continuity of F^{-1} proved in the theorem. □

COROLLARY 2. *Let $L(\lambda, \mu)$ be a m.o.p. in λ depending continuously on μ in an open set $\mathscr{D}$ of the complex plane. Let Γ be a closed contour in the complex plane and, for every $\mu \in \mathscr{D}$, let $L_1(\lambda, \mu)$ be a Γ-spectral (right) divisor of L with (left) quotient $L_2(\lambda, \mu)$. Then L_1 and L_2 are continuous functions of μ on $\mathscr{D}$.*

Proof. We have for the supporting subspace $\mathscr{M}(\mu)$ of $L(\lambda, \mu)$ corresponding to $L_1(\lambda, \mu)$:

$$\mathscr{M}(\mu) = \mathscr{I}m \frac{1}{2\pi i} \oint_\Gamma (I\lambda - C_{L(\cdot, \mu)})^{-1} \, d\lambda.$$

The integral determines a family of projectors which depend continuously on μ. Consequently the images, $\mathscr{M}(\mu)$, will also be continuous function on $\mathscr{D}$. Then the conclusion follows from part (i) of Corollary 1. □

1.4. Analytic perturbations. In this section we consider a m.o.p. L whose coefficients depend analytically on a parameter μ. Thus, we write

$$L(\lambda, \mu) = I\lambda^l + \sum_{i=0}^{l-1} A_i(\mu)\lambda^i \tag{19}$$

where the coefficients $A_i(\mu)$, $i = 1, 2, \cdots, l-1$ are analytic operator-valued functions defined on a connected domain Ω of the complex plane and with values in $\mathscr{L}(\mathscr{X})$. For this purpose the following definition will be required: A family $\mathscr{M}(\mu)$, $\mu \in \Omega$, of subspaces of $\mathscr{X}$ is an *analytic family* in Ω if, for every $\mu_0 \in \Omega$, there exists a neighborhood $U(\mu_0) \subset \Omega$ such that $\mathscr{M}(\mu) = A(\mu)\mathscr{M}_0$ for every $\mu \in U(\mu_0)$, where $A(\mu)$ is an invertible analytic operator-valued function with values in $\mathscr{L}(\mathscr{X})$, and $\mathscr{M}_0$ is a fixed subspace (depending on μ_0) of $\mathscr{X}$. This definition is due to M. A. Shubin [14]; see also the paper of Gohberg and Leiterer [7].

The following result is due to Shubin [14] and will play an important role in the sequel:

PROPOSITION 4 (M. A. Subin). *If $\mathscr{M}(\mu)$ is an analytic family of complemented subspaces in domain Ω, then there exists an analytic projector-valued function $P(\mu)$ such that $\mathscr{M}(\mu) = \mathscr{I}m\, P(\mu)$ for every $\mu \in \Omega$.*

In the applications that we will make of Proposition 4 the subspaces in question will be families of supporting subspaces for a matrix polynomial and, as described in Proposition 2, these are always complemented.

A variant of Lemma 1 will also be needed:

LEMMA 3. *Let $\mathscr{M}, \mathscr{M}_1 \in \mathscr{A}$ and $\mathscr{M} \oplus \mathscr{M}_1 = \mathscr{X}^l$ and let P be any projector onto $\mathscr{M}$. If Q is a projector and $\|P - Q\|$ is sufficiently small, then $\mathscr{I}m\, Q \oplus \mathscr{M}_1 = \mathscr{X}^l$.*

Like the first conclusion of Lemma 1, this is an easy consequence of Theorem 2 of the paper [8] of Gohberg and Markus.

THEOREM 4. *Let $L(\lambda, \mu)$ be a m.o.p. of the form* (19) *with coefficients analytic in Ω. For each $\mu \in \Omega$ let $L_1(\lambda, \mu)$ be a monic right divisor of $L(\lambda, \mu)$ with supporting subspace $\mathcal{M}(\mu)$ with respect to $C_{L(\cdot, \mu)}$, and let $L_2(\lambda, \mu)$ be the corresponding left quotient.*

(i) *If $\mathcal{M}(\mu)$ forms an analytic family in Ω, then $L_1(\lambda, \mu)$ and $L_2(\lambda, \mu)$ are analytic in Ω.*

(ii) *If either $L_1(\lambda, \mu)$ or $L_2(\lambda, \mu)$ is analytic in Ω, then $\mathcal{M}(\mu)$ forms an analytic family in Ω and $L_2(\lambda, \mu)$ or $L_1(\lambda, \mu)$, respectively, is analytic in Ω.*

Proof. (i) Let $\mathcal{M}(\mu)$ form an analytic family in Ω and $\mu_0 \in \Omega$. Then, by Proposition 4, there is a projector-valued analytic function $P(\mu)$ such that $\mathcal{M}(\mu) = \mathcal{Im}\, P(\mu)$, $\mu \in \Omega$. By use of Lemma 3 it is deduced that there is a neighborhood $\mathcal{V}(\mu_0)$ in which $\mathcal{M}(\mu)$ is a supporting subspace for $L(\lambda, \mu)$.

Using the representations (4) and (8) for L_1 and L_2 it can be seen that, in order to prove the analytic dependence of these m.o.p. on μ, it is sufficient to prove that the projector $P_{\mathcal{M}(\mu)}$ on $\mathcal{M}(\mu)$ along $\mathcal{Y}_{l-k}$ depends analytically on μ. To this end, define an analytic operator-valued function $A(\mu)$ on $\mathcal{X}^l = \mathcal{M}(\mu_0) \oplus \mathcal{Y}_{l-k}$ by:

$$A(\mu)|_{\mathcal{M}(\mu_0)} = P(\mu)|_{\mathcal{M}(\mu_0)}, \qquad A(\mu)|_{\mathcal{Y}_{l-k}} = I. \tag{20}$$

It is claimed that, in a neighborhood of μ_0, $A(\mu)$ is an invertible map onto $\mathcal{X}^l$. To see this, first use Lemma 3 to establish a neighborhood of μ_0 so that, with μ in this neighborhood,

$$\mathcal{M}(\mu_0) \oplus \operatorname{Ker} P(\mu) = \mathcal{M}(\mu) \oplus \mathcal{Y}_{l-k} = \mathcal{X}^l.$$

If $x \in \operatorname{Ker} A(\mu)$ and $x = x_1 + x_2$, $x_1 \in \mathcal{M}(\mu_0)$, $x_2 \in \mathcal{Y}_{l-k}$, then $x_2 = 0$ and $P(\mu)x_1 = 0$. Thus $x_1 \in \mathcal{M}(\mu_0) \cap \operatorname{Ker} P(\mu) = \{0\}$ for μ close enough to μ_0. Thus, $\operatorname{Ker} A(\mu) = 0$.

For the surjective property we show that $\mathcal{X}^l = \mathcal{M}(\mu) \oplus \mathcal{Y}_{l-k} \subset \mathcal{Im}\, A(\mu)$. That $\mathcal{Y}_{l-k} \subset \mathcal{Im}\, A(\mu)$ is obvious. Then let $y \in \mathcal{M}(\mu)$ and $y = y_1 + y_2$ where $y_1 \in \mathcal{M}(\mu_0)$, $y_2 \in \operatorname{Ker} P(\mu)$. Then $A(\mu)y_1 = P(\mu)y_1 = P(\mu)y = y$ so that $\mathcal{M}(\mu) \subset \mathcal{Im}\, A(\mu)$ also.

If P_0 is the projector on $\mathcal{M}(\mu_0)$ along $\mathcal{Y}_{l-k}$ then $A(\mu)P_0A^{-1}(\mu)$ is a projector with image $\mathcal{M}(\mu)$ and kernel $\mathcal{Y}_{l-k}$. Thus,

$$P_{\mathcal{M}(\mu)} = A(\mu)P_0A^{-1}(\mu).$$

The analyticity of $P_{\mathcal{M}(\mu)}$ and hence of $L_1(\lambda, \mu)$, $L_2(\lambda, \mu)$, in a neighborhood of μ_0 follows from this relation. Since μ_0 is an arbitrary member of Ω conclusion (i) is obtained.

(ii) Suppose that $L_1(\lambda, \mu)$ is an analytic function of μ in Ω. Then it is clear from the representation (18) that $P_{\mathcal{M}(\mu)}$ is analytic in Ω. It follows that $\mathcal{M}(\mu)$ is an analytic family in Ω and, using part (i), that $L_2(\lambda, \mu)$ is analytic in μ in Ω.

Finally, assume it is given that $L_2(\lambda, \mu)$ depends analytically on μ in Ω and write $L_2(\lambda, \mu) = I\lambda^{l-k} + \sum_{i=0}^{l-k-1} C_i(\mu)\lambda^i$. Taking adjoints in (19), we find that

$$L_1^*(\lambda, \mu) = I\lambda^l + \sum_{i=0}^{l-1} A_i(\mu)^*\lambda^i$$

has a right divisor

$$L_2^*(\lambda, \mu) = I\lambda^{l-k} + \sum_{j=0}^{l-k-1} C_j(\mu)^*\lambda^j.$$

Now this divisor has supporting subspace $\mathcal{M}(\mu)^\perp$ in $(\mathcal{X}^*)^l$ with respect to C_L^* (X, T, Y a standard triple for L implies Y^*, T^*, X^* a standard triple for L^*; cf. Remark 3 following Theorem 9 of [4]). The argument of the preceding paragraph can now be applied to deduce that $\mathcal{M}(\mu)^\perp$ is an analytic family in Ω and that L_1^* is an analytic in Ω. Corresponding results for $\mathcal{M}(\mu)$ and L_1 follow and the theorem is proved. □

THEOREM 5. *Let $L(\lambda, \mu)$ be a m.o.p. of the form* (19) *with coefficients analytic in Ω. Assume that, for each $\mu \in \Omega$, there is a separated part of the spectrum of L, say σ, and a corresponding Γ_μ-spectral divisor $L_1(\lambda, \mu)$ of $L(\lambda, \mu)$. Assume also that, for each $\mu_0 \in \Omega$, there is a neighborhood $U(\mu_0)$ such that, for each $u \in U(\mu_0)$, σ_0 is inside Γ_{μ_0}.*

Then $L_1(\lambda, \mu)$ and the corresponding left quotient $L_2(\lambda, \mu)$ are analytic functions of μ in Ω.

Proof. The hypotheses imply that, for $\mu \in U(\mu_0)$, the Riesz projector $P(\mu)$ can be written

$$P(\mu) = \frac{1}{2\pi i} \oint_{\Gamma_\mu} \{I\lambda - C_{L(\cdot,\mu)}\}^{-1} \, d\lambda = \frac{1}{2\pi i} \oint_{\Gamma_{\mu_0}} \{I\lambda - C_{L(\cdot,\mu)}\}^{-1} \, d\lambda$$

from which we deduce the analytic dependence of P on μ in $U(\mu_0)$. Let $\mathcal{M}(\mu) = \mathcal{I}m\, P(\mu)$ be the corresponding supporting subspace of L and use (20) to define an invertible operator function $A(\mu)$ in a neighborhood of μ_0 with the property that $\mathcal{M}(\mu) = A(\mu)\mathcal{M}(\mu_0)$. Then $\mathcal{M}(\mu)$ is seen to be an analytic family in Ω and the conclusion follows from part (i) of Theorem 4. □

It is evident that, in the case of a Γ_μ-spectral divisor, the projector $P(\mu)$, and hence $\mathcal{M}(\mu)$ and the divisor $L_1(\lambda, \mu)$ are uniquely defined for each μ. This is the case whether the continuation is analytic, as in Theorem 5, or merely continuous, as in Corollary 1 of Theorem 3.

2. Global analytic perturbations in the finite dimensional case. In § 1 direct advantage was taken of the assumed dependence of the coefficients of L on μ by relating divisors to corresponding subspaces of the operator C_L, in which the coefficients of L are explicitly displayed. Thus the appropriate standard pair for that analysis was $X = [I\ 0 \cdots 0]$, C_L (as introduced in § 1.1). However, this standard pair is used at the cost of leaving the relationship of divisors with invariant subspaces obscure and, in particular, giving no direct line of attack on the *continuation* of divisors from a point to neighboring points.

In this chapter we shall need more detailed information on the behavior of supporting subspaces and, for this purpose, *canonical* pairs X_μ, J_μ will be used. Here the linearization J_μ is relatively simple and its invariant subspaces are easy to describe. The necessity of using canonical pairs means, however, that our attention will be confined to the finite dimensional case.

2.1. Preliminaries. Let $A_0(\mu), \cdots, A_{l-1}(\mu)$ be analytic functions on a connected domain Ω taking values in the linear space of complex $n \times n$ matrices. With I the $n \times n$ identity matrix consider the matrix-valued function

$$L(\lambda, \mu) = I\lambda^l + \sum_{i=0}^{l-1} A_i(\mu)\lambda^i. \tag{21}$$

Using techniques developed by the authors in [4] and [5] it is possible to construct for each $\mu \in \Omega$ a canonical pair of matrices X_μ, being $n \times ln$, and J_μ, which is $ln \times ln$, for $L(\lambda, \mu)$. The matrix J_μ is in Jordan normal form and will be supposed to have r Jordan cells $J_\mu^{(i)}$ of size q_i, $i = 1, \cdots, r$ with associated eigenvalues $\lambda_i(\mu)$, (not necessarily distinct). In general, r and the q_i depend on μ. We write $J_\mu = \operatorname{diag}\{J_\mu^{(1)}, \cdots, J_\mu^{(r)}\}$. The columns of X_μ determine the Jordan chains for L at μ. Partition X_μ as follows:

$$X_\mu = [X_\mu^{(1)} \cdots X_\mu^{(r)}]$$

where $X_\mu^{(i)}$ has q_i columns and each submatrix represents one Jordan chain.

As outlined in § 1.1, the pair X_μ, J_μ forms an important example of a standard pair and can be used to form a standard triple X_μ, J_μ, Y_μ. Using Proposition 1 we may therefore couple the study of divisors of $L(\lambda, \mu)$ with the existence of certain invariant subspaces of J_μ ([4] and [5]), and we are now to determine the nature of the dependence of divisors and corresponding subspaces on μ via the fine structure of the canonical pairs X_μ, J_μ.

In order to achieve our objective, some important ideas developed in the monograph of Baumgärtel [1] are needed. The first, a "global" theorem concerning invariance of the Jordan structure of J_μ follows from Theorem 1 of Section V.7 of [1]. It is only necessary to observe that J_μ is also a Jordan form for the first companion matrix

$$C_L(\mu) = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & & \\ \vdots & & & & \\ 0 & 0 & \cdots & 0 & I \\ -A_0(\mu) & -A_2(\mu) & & \cdots & -A_{l-1}(\mu) \end{bmatrix} \tag{22}$$

to obtain Proposition 5.

PROPOSITION 5. *The matrix valued functions X_μ, J_μ can be defined on Ω in such a way that, for some countable set S_1 of isolated points in Ω, the following statements hold:*

(a) *For every $\mu \in \Omega \backslash S_1$ the number, r, of Jordan cells in J_μ and their sizes $q_1, \cdots, q_r$ are independent of μ.*

(b) *The eigenvalues $\lambda_i(\mu)$, $i = 1, \cdots, r$, are analytic functions in $\Omega\backslash S_1$ and may have algebraic branch points at some points of S_1.*

(c) *The blocks $X_\mu^{(i)}$, $i = 1, \cdots, r$, of X_μ are analytic functions of μ on $\Omega\backslash S_1$ which may also be branches of analytic functions having algebraic branch points at some points of S_1.*

The set S_1 is associated with Baumgärtel's "Hypospaltpunkte" and consists of points of discontinuity of J_μ in Ω as well as all branch points associated with the eigenvalues.

The next proposition is derived primarily from Baumgärtel's Theorem 1, Section IX.1. It describes the local behavior of canonical matrices in a neighborhood of a point $\mu_0 \in S_1$. Note that r has the same meaning as in Proposition 5.

PROPOSITION 6. *For any $\mu_0 \in S_1$ there is a deleted neighborhood N_{μ_0} in which the following statements hold:*

(a) *The number of distinct eigenvalues in N_{μ_0} is constant.*

(b) *There exist disjoint sets of integers $T_1, \cdots, T_s$ for which $\bigcup_{j=1}^s T_j = \{1, 2, \cdots, r\}$ and all eigenvalues $\lambda_i(\mu)$, $i \in T_j$, are determined by the distinct branches of an analytic function having μ_0 as a branch point of multiplicity m_j, the cardinality of T_j.*

(c) *If $k, l \in T_j$ then $q_k = q_l$ (sizes of the Jordan cells).*

(d) *If $\mu \in N_{\mu_0}$, then $\{X_\mu^{(i)} : i \in T_j\}$ is determined by the distinct branches of a (matrix-valued) analytic function having μ_0 as a branch point of multiplicity m_j.*

(e) *For $\mu \in N_{\mu_0} \cup \{\mu_0\}$ and $i \in T_j$ there are expansions:*

$$\lambda_i(\mu) = \mu_0 + \sum_{\alpha=1}^{\infty} a_{\alpha j}(\mu - \mu_0)^{\alpha/m_j},$$

$$X_\mu^{(i)} = \sum_{j=0}^{\infty} B_{\alpha j}(\mu - \mu_0)^{\alpha/m_j}.$$

The least common multiple m of $m_1, m_2, \cdots, m_s$ is called the *branch multiplicity* of μ_0.

Let $\hat{\lambda}_j(\mu)$, $j = 1, 2, \cdots, t$ denote all the *distinct* eigenvalue functions defined on Ω and let

$$S_2 = \{\mu \in \Omega \backslash S_1 : \hat{\lambda}_i(\mu) = \hat{\lambda}_j(\mu),\ 1 \leqq i, j \leqq t,\ i \neq j\}.$$

(Note the following relations: $t \leqq s \leqq r \leqq ln$, $\sum_{i=1}^{s} m_i = r$, $\sum_{j=1}^{r} q_j = ln$.) The canonical matrix J_μ then has the same set of invariant subspaces for every $\mu \in \Omega \backslash (S_1 \cup S_2)$. The set S_2 consists of multiple points (mehrfache Punkte) in the terminology of Baumgärtel. The set $S_1 \cup S_2$ is described as the *exceptional set* of $L(\lambda, \mu)$ in Ω and is countable, having its limit points (if any) on $\partial\Omega$.

Example 2.1. Let $L(\lambda, \mu) = \lambda^2 + \mu\lambda + \omega^2$ for some constant ω, and $\Omega = \mathbb{C}$. Then

$$J_{\pm 2\omega} = \begin{bmatrix} \mp\omega & 1 \\ 0 & \mp\omega \end{bmatrix}$$

and, for $\mu \neq \pm 2\omega$, $J_\mu = \operatorname{diag}\{\lambda_1(\mu), \lambda_2(\lambda)\}$ where $\lambda_{1,2}$ are the zeros of $\lambda^2 + \mu\lambda + \omega^2$. Here, $S_1 = \{2\omega, -2\omega\}$, $S_2 = \phi$.

Example 2.2. Let

$$L(\lambda, \mu) = \begin{bmatrix} (\lambda-1)(\lambda-\mu) & 0 \\ \mu & (\lambda-2)(\lambda-\mu^2) \end{bmatrix}$$

and $\Omega = \mathbb{C}$. The canonical Jordan matrix is found for every $\mu \in \mathbb{C}$ and hence the sets S_1 and S_2. For $\mu \notin \{0, 2, \pm 1, \pm\sqrt{2}\}$,

$$J_\mu = \operatorname{diag}\{\mu, \mu^2, 1, 2\}.$$

Then

$$J_0 = \operatorname{diag}\{0, 0, 1, 2\}, \qquad J_2 = \operatorname{diag}\left\{\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, 1, 4\right\},$$

$$J_{-1} = \operatorname{diag}\left\{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, -1, 2\right\}, \qquad J_1 = \operatorname{diag}\left\{\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, 2\right\}.$$

$$J_{\pm\sqrt{2}} = \operatorname{diag}\left\{\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, 1, \pm\sqrt{2}\right\}.$$

It follows that $S_1 = \{\pm 1, 2, \pm\sqrt{2}\}$, $S_2 = \{0\}$ and, for $\mu \in \Omega \backslash S_1$,

$$X_\mu = \begin{bmatrix} (\mu-2)(\mu-1) & 0 & 1-\mu^2 & 0 \\ 1 & 1 & \mu & 1 \end{bmatrix}.$$

2.2. Analytic divisors. As in equation (21), let $L(\lambda, \mu)$ be defined for all $(\lambda, \mu) \in \mathbb{C} \times \Omega$. Suppose that for some $\mu_0 \in \Omega$ the m.o.p. $L(\lambda, \mu_0)$ has a right divisor $L_1(\lambda)$. The possibility of extending $L_1(\lambda)$ to a continuous (or an analytic) family of right divisors $L_1(\lambda, \mu)$ of $L(\lambda, \mu)$ is to be investigated. It turns out that this may not be possible, in which case $L_1(\lambda)$ will be described as an isolated divisor, and that this can only occur if μ_0 is in the exceptional set of L in Ω. In contrast, we have:

THEOREM 6. *If $\mu_0 \in \Omega \backslash (S_1 \cup S_2)$, then every monic right divisor $L_1(\lambda)$ of $L(\lambda, \mu_0)$ can be extended to an analytic family $L_1(\lambda, \mu)$ of monic right divisors of $L(\lambda, \mu)$ in the domain $\Omega \backslash (S_1 \cup S_3)$ where S_3 is an, at most, countable subset of isolated points of $\Omega \backslash S_1$ and $L_1(\lambda, \mu)$ has poles or removable singularities at the points of S_3.*

Proof. Let $\mathcal{M}$ be the supporting subspace of divisor $L_1(\lambda)$ of $L(\lambda, \mu_0)$ with respect to the canonical matrix J_{μ_0}. Then, if L_1 has degree k, define the map $Q_k(\mu)$ on $\Omega\backslash S_1$ to be the restriction of

$$\begin{bmatrix} X_\mu \\ X_\mu J_\mu \\ \vdots \\ X_\mu J_\mu^{k-1} \end{bmatrix}$$

to $\mathcal{M}$. It follows from Proposition 1 that $Q_k(\mu_0)$ is invertible. Since (by Proposition 5) $Q_k(\mu)$ is analytic on $\Omega\backslash S_1$ it follows that $Q_k(\mu)$ is invertible on a domain $(\Omega\backslash S_1)\backslash S_3$ where S_3 is, at most, a countable subset of isolated points of $\Omega\backslash S_1$. Furthermore, as noted in the preceding section, the invariant subspaces of J_μ are invariant for $\mu \in \Omega\backslash S_1$. Hence, by the converse of the theorem just cited there exists a family of divisors $L_1(\lambda, \mu)$ of $L(\lambda, \mu)$ for each $\mu \in \Omega\backslash(S_1 \cup S_3)$ each divisor having the same supporting subspace $\mathcal{M}$ with respect to J_μ. By part (i) of Theorem 4, it follows that this family is analytic in $\Omega\backslash(S_1 \cup S_3)$.

An explicit representation of $L_1(\lambda, \mu)$ is obtained in the following way (cf. Proposition 1). Take a fixed basis in $\mathcal{M}$ and for each μ in $\Omega\backslash S_1$ represent $Q_k(\mu)$ as a matrix defined with respect to this basis and the natural basis for $\mathscr{X}^k$. Then, for $\mu \notin S_3$ define $ln \times n$ matrix-valued functions $W_1(\mu), \cdots, W_k(\mu)$ by

$$[W_1(\mu) \cdots W_k(\mu)] = RQ_k^{-1}(\mu)$$

where R is the matrix (independent of μ) representing the embedding of $\mathcal{M}$ into $\mathscr{X}^l$. The divisor L_1 has the form:

$$L_1(\lambda, \mu) = I\lambda^k - X_\mu J_\mu^k\{W_1(\mu) + W_2(\mu)\lambda + \cdots + W_k(\mu)\lambda^{k-1}\}.$$

The nature of the singularities of $L_1(\lambda, \mu)$ is apparent from this representation. □

Note that more can be said about the orders of the poles of $L_1(\lambda, \mu)$. First, it is clear that if $L_1(\lambda, \mu)$ has a pole at μ_1 then the order of the pole cannot exceed the order of the zero of $\det Q_k(\mu)$ at μ_1. More precisely, there is a factorization of $Q_k(\mu)$ valid in a neighborhood U of μ_1 of the form

$$Q_k(\mu) = P(\mu)D(\mu)Q(\mu),$$

where

$$D(\mu) = \operatorname{diag}\{(\mu - \mu_1)^{k_1}, \cdots, (\mu - \mu_1)^{k_n}\}$$

and D, Q are invertible in U (see Gohberg–Sigal [9], for example). Consequently, the order of a pole of L_1 at μ_1 (if any) cannot exceed $\max_{1 \leqq i \leqq h} k_i$.

An important special case of the theorem is

COROLLARY 1. *If* $\det L(\lambda, \mu)$ *has ln distinct zeros for every* $\mu \in \Omega$, *and if* $\mu_0 \in \Omega$, *then every monic right divisor of* $L(\lambda, \mu_0)$ *can be extended to a family of monic right divisors for* $L(\lambda, \mu)$ *which is analytic on* $\Omega\backslash(S_1 \cup S_3)$.

Proof. Under these hypotheses S_1 and S_2 are empty and the conclusion follows. □

Example 2.3. Let

$$L(\lambda, \mu) = \begin{bmatrix} \lambda^2 & 0 \\ -\lambda\mu & (\lambda - 1)^2 \end{bmatrix}.$$

Then a Jordan matrix for L does not depend on μ, i.e. for every $\mu \in \mathbb{C}$,

$$J_\mu = J = \operatorname{diag}\left\{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right\}.$$

We have S_1 and S_2 both empty and

$$X_\mu = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \mu & 1 & 0 \end{bmatrix}.$$

The subspace $\mathcal{M}$ spanned by the first two component vectors e_1 and e_2 is invariant under J and $X_\mu|_{\mathcal{M}} = \begin{bmatrix} 1 & 0 \\ 0 & \mu \end{bmatrix}$. Thus, for $\mu \neq 0$, $\mathcal{M}$ is a supporting subspace for $L(\lambda, \mu)$. Since the corresponding divisor is $I\lambda - X_\mu J_\mu (X_\mu|_{\mathcal{M}})^{-1}$ it follows that

$$L(\lambda, \mu) = I\lambda - \begin{bmatrix} 0 & \mu^{-1} \\ 0 & 0 \end{bmatrix},$$

and $S_3 = \{0\}$.

COROLLARY 2. *If the divisor $L_1(\lambda)$ of Theorem 6 is, in addition, a Γ-spectral divisor for some closed contour Γ then it has a* unique *analytic extension to a family $L_1(\lambda, \mu)$ of monic right divisors defined on $\Omega \backslash (S_1 \cup S_3)$.*

Proof. Note that, as remarked after the proof of Theorem 5, $L_1(\lambda)$ can be continued locally *in a unique way* to an analytic family of divisors. Thus, the continuations determined by Theorems 5 and 6 coincide. □

Example 2.4. Consider the example used in the Introduction:

$$L(\lambda, \mu) = I\lambda^2 + \mu B\lambda + C$$

with B, C positive definite matrices and $\Omega = \mathbb{C}$. At $\mu = 0$ there is a spectral divisor $L_1(\lambda) = I\lambda - iC^{1/2}$ associated with the eigenvalues of $I\lambda^2 + C$ in the upper half-plane. Corollary 2 implies the existence of a unique analytic family of divisors $L_1(\lambda, \mu)$ with $L_1(\lambda, 0) = L_1(\lambda)$ defined on the whole of $\mathbb{C}$ with the exception of an at most countable set of isolated points in $\mathbb{C}$.

2.3. Singularities of analytic divisors. The next theorem and example show that divisors of L may or may not exist at points of S_1, the set of exceptional points introduced in § 2.1.

THEOREM 7. *Let $\mu_0 \in \Omega$ be a branch point of at least one eigenvalue function of $L(\lambda, \mu)$ and let m be the branch multiplicity of μ_0. Let $L_1(\lambda, \mu)$ be an analytic family of right monic divisors of $L(\lambda, \mu)$ (obtained as in Theorem 6 from some $L_1(\lambda, \mu_1)$, $\mu_1 \in \Omega \backslash (S_1 \cup S_2)$); then, in a deleted neighborhood $U(\mu_0)$, L_1 has one of the following representations: Either $L_1(\lambda, \mu)$ is an analytic function of the form*

$$(23) \qquad L_1(\lambda, \mu) = I\lambda^k + \sum_{j=1}^{k-1} B_j(\mu)\lambda^j, \qquad B_j(\mu) = \sum_{i=0}^{\infty} C_{ij}(\mu - \mu_0)^{i/m}$$

in which case $L_1(\lambda, \mu_0)$ is a right divisor of L at μ_0, or, for some positive integer p

$$(24) \qquad L_1(\lambda, \mu) = I\lambda^k + \sum_{j=1}^{k-1} B_j(\mu)\lambda^j, \qquad B_j(\mu) = \sum_{i=0}^{\infty} C_{ij}(\mu - \mu_0)^{(i-p)/m}.$$

Proof. Let $L_1(\lambda, \mu)$ be a right monic divisor of $L(\lambda, \mu)$ in $U(\mu_0)$ of degree k. As established in the preceding section, a supporting subspace $\mathcal{M}$ which is independent of μ can be associated with this divisor. Using this subspace define the map $Q_k(\mu)$,

$\mu \in U(\mu_0)$ to be the restriction to $\mathcal{M}$ of

$$\begin{bmatrix} X_\mu \\ X_\mu J_\mu \\ \vdots \\ X_\mu J_\mu^{k-1} \end{bmatrix}.$$

Then Proposition 6 implies that X_μ, J_μ are analytic functions with algebraic branch points at μ_0 and that $Q_k(\mu)$ has a series expansion in nonnegative powers of $(\mu-\mu_0)^{1/m}$.

The argument of Theorem 6 can now be followed to obtain (23), provided $\lim_{\mu\to\mu_0} \det Q_k(\mu) \neq 0$. If, on the other hand, this limit is zero then there is a smallest integer p such that $\lim_{\mu\to\mu_0} (\mu-\mu_0)^{-p} Q_k(\mu) \neq 0$ and (24) holds in this case. □

If $\mathcal{L}(\mu)$ is the supporting subspace of divisor $L_1(\lambda, \mu)$, $\mu \in U(\mu_0)$, associated with the companion operator then, by use of Corollary 1 of Theorem 3 it is clear that $\lim_{\mu\to\mu_0} \mathcal{L}(\mu)$ will exist in the first case of Theorem 7 but not in the second.

Example 2.5. Let

$$L(\lambda, \mu) = \begin{bmatrix} \lambda^2 & 1 \\ \mu & \lambda^2 \end{bmatrix}$$

and $\Omega = \mathbb{C}$. Then for $\mu \neq 0$, $J_\mu = \operatorname{diag}\{\mu^{1/4}, i\mu^{1/4}, -\mu^{1/4}, -i\mu^{1/4}\}$ and $S_1 = \{0\}$, $S_2 = \phi$. For $\mu \neq 0$,

$$X_\mu = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -\mu^{1/2} & \mu^{1/2} & -\mu^{1/2} & \mu^{1/2} \end{bmatrix}.$$

Let $\mathcal{M}$ be the invariant subspace of J_μ spanned by e_1 and e_2 and, for $\mu \neq 0$, $X_\mu|_{\mathcal{M}}$ is invertible so that $\mathcal{M}$ is a supporting subspace for a divisor. It is found that

$$L_1(\lambda, \mu) = I\lambda - \begin{bmatrix} \frac{1}{2}(1+i)\mu^{1/4} & \frac{1}{2}(-1+i)\mu^{-1/4} \\ \frac{1}{2}(-1+i)\mu^{1/4} & \frac{1}{2}(1+i)\mu^{1/4} \end{bmatrix}$$

for $\mu \neq 0$, and there is no monic right divisor for L at $\mu = 0$.

Example 2.6. Let $L(\lambda, \mu)$ be as in Example 2.2 and $\mathcal{M}$ be spanned by e_2 and e_3. If $\mu \neq \pm 1$, $\mathcal{M}$ is an invariant subspace of J_μ and $X_\mu|_{\mathcal{M}}$ is invertible so that $\mathcal{M}$ is a supporting subspace. The corresponding divisor is

$$L_1(\lambda, \mu) = I\lambda - \begin{bmatrix} 1 & 0 \\ \mu & \mu^2 \end{bmatrix}$$

and can be continued to the whole of $\mathbb{C}$.

2.4. Isolated and nonisolated divisors. As before, let $L(\lambda, \mu)$ be a monic matrix polynomial of degree l with coefficients depending analytically on μ in Ω, and let $L_1(\lambda)$ be a monic right divisor of degree k of $L(\lambda, \mu_0)$, $\mu_0 \in \Omega$. The divisor $L_1(\lambda)$ is said to be *isolated* if there is a neighborhood $U(\mu_0)$ of μ_0 such that $L(\lambda, \mu)$ has no family of right monic divisors $L_1(\lambda, \mu)$ of degree k which (a) depends continuously on μ in $U(\mu_0)$ and (b) has the property that $\lim_{\mu\to\mu_0} L_1(\lambda, \mu) = L_1(\lambda)$.

Theorem 6 shows that if, in the definition, we have $\mu_0 \in \Omega \backslash (S_1 \cup S_2)$ then monic right divisors cannot be isolated. We demonstrate the existence of isolated divisors by means of examples.

Example 2.7. Let $C(\mu)$ be any matrix depending analytically on μ in a domain Ω with the property that for $\mu = \mu_0 \in \Omega$, $C(\mu_0)$ has a square root and for $\mu \neq \mu_0$, μ in a neighborhood of μ_0, $C(\mu)$ has no square root. The prime example here is $\mu_0 = 0$,

$C(\mu)=\begin{bmatrix} 0 & 0 \\ \mu & 0 \end{bmatrix}$. Then define $L(\lambda, \mu)=I\lambda^2-C(\mu)$. It is easily seen that if $L(\lambda, \mu)$ has a right divisor $I\lambda-A(\mu)$ then $L(\lambda, \mu)=I\lambda^2-A^2(\mu)$ and hence that $L(\lambda, \mu)$ has a monic right divisor iff $C(\mu)$ has a square root. Thus, under the hypotheses stated, $L(\lambda, \mu)$ has an isolated divisor at μ_0.

It will now be shown that there are some cases where divisors which exist at points of the exceptional set can, nevertheless, be extended analytically. An example of this situation is provided by taking $\mu_0=0$ in Example 2.2.

THEOREM 8. *Let $\mu_0 \in S_2$ and $L_1(\lambda)$ be a nonisolated right monic divisor of degree k of $L(\lambda, \mu_0)$. Then $L_1(\lambda)$ can be extended to an analytic family $L_1(\lambda, \mu)$ of right monic divisors of degree k of $L(\lambda, \mu)$.*

Proof. In effect, this result provides an extension of the conclusions of Theorem 6. The statements concerning the singularities also carry over but are omitted for brevity.

Note first that with a divisor $L_1(\lambda, \mu)$ there are associated supporting subspaces $\mathcal{N}(\mu)$ ($\mathcal{M}(\mu)$) with respect to J_μ (with respect to $C_{L(\mu)}$) and these are connected by the relationship $\mathcal{N}(\mu)=Q^{-1}(\mu)\mathcal{M}(\mu)$, where

$$Q(\mu)=\begin{bmatrix} X_\mu \\ X_\mu J_\mu \\ \vdots \\ X_\mu J_\mu^{l-k} \end{bmatrix}$$

(refer to § 1.1). We have seen in Corollary 1 of Theorem 3 that if L_1 depends continuously on μ then so does $\mathcal{M}$ and it follows from the above observation (and Proposition 5) that $\mathcal{N}$ will also depend continuously on μ.

Since $L_1(\mu)$ is not isolated, there is a sequence $\{\mu_n\}$ in $\Omega\backslash(S_1 \cup S_2)$ such that $\mu_n \to \mu_0$ and there are monic right divisors $L_1(\lambda, \mu_n)$ of $L(\lambda, \mu_n)$ such that $\lim_{n\to\infty} L_1(\lambda, \mu_n)= L_1(\lambda)$. Since the correspondence between divisors L_1 and supporting subspaces $\mathcal{N}$ is continuous, it follows that $\lim_{n\to\infty}\mathcal{N}(\mu_n)=\mathcal{N}(\mu_0)$. Since $\mathcal{N}(\mu_n)$ is invariant under J_{μ_n} and all invariant subspaces of J_μ are independent of μ in $\Omega\backslash(S_1 \cup S_2)$ it follows that $\mathcal{N}(\mu_0)$ is invariant under J_{μ_n} for each n.

From this point, the proof of Theorem 6 can be applied to obtain the required conclusion. □

The following special case is important:

COROLLARY 3. *If the elementary divisors of $L(\lambda, \mu)$ are linear for each fixed $\mu \in \Omega$, then every nonisolated right monic divisor $L_1(\lambda)$ of $L(\lambda, \mu_0)$ can be extended to an analytic family of right monic divisors for $L(\lambda, \mu)$ on $\Omega\backslash S_3$.*

Proof. In this case S_1 is empty and so the conclusion follows from Theorems 6 and 8. □

Our final theorem will describe the continuations of divisors which can be made from divisors at points of S_1 which are known to be nonisolated. The statement of Theorem 9 is disarmingly simple. In order to complete the proof, however, it is necessary to introduce some notions which are familiar in the theory of functions of several complex variables. We first introduce some of these concepts and two preliminary lemmas. In order to maintain some continuity in our argument the more technical aspects of proof are relegated to the Appendix.

Write $z=(z_1, \cdots, z_\beta)$ for a typical point of $\mathbb{C}^\beta$. A complex-valued function f defined on a domain in $\mathbb{C}^\beta$ is said to be *meromorphic* in a neighborhood of the point $a \in \mathbb{C}^\beta$ if there exist two functions $g_1(z)$ and $g_2(z)$ analytic in a neighborhood of a such that $f(z)g_1(z)=g_2(z)$ holds in some neighborhood of a which is contained in the domains of f, g_1 and g_2. The point a is (i) a *regular* point of f if $g_1(a)\neq 0$, (ii) a *pole* of f if $g_1(a)=0$ and $g_2(a)\neq 0$, (iii) an *ambiguous point* of f if $g_1(a)=g_2(a)=0$.

Suppose that the domain of f contains a neighborhood U of the point $a\in\mathbb{C}^\beta$. The set $B\subset U$ is an *analytic set* in U if there exists a function f, analytic in U, such that $f(z)=0$ for all $z\in B$. A set $V\subset U$ is a *variety* in the neighbourhood U of a if V is the intersection of finitely many analytic sets in U.

LEMMA 4. *Let $f_1,\cdots,f_\alpha$ be meromorphic functions defined on a neighborhood of the origin in $\mathbb{C}^\beta$ and assume that the origin is an ambiguous point of each function. Assume that there is a sequence $\{z^{(n)}\}$ in this neighborhood with the following properties*: (a) *$z^{(n)}$ is a regular point of f_i, $i=1,\cdots,\alpha$, $n=1,2,\cdots$.* (b) *$z^{(n)}\to 0$ as $n\to\infty$.* (c) *For some complex numbers $b_1,\cdots,b_\alpha$, $f_i(z^{(n)})\to b_i$ as $n\to\infty$.* (d) *With $z^{(n)}=(z_{1n},\cdots,z_{\beta n})$, there is an i_0 such that infinitely many z_{i_0n} are different from zero.* (e) *$\{z^{(n)}\}$ is contained in some variety X in a neighborhood of the origin.*

Then there exist complex valued functions $g_1,\cdots,g_\beta$ of the single complex variable ζ which are analytic in a neighborhood V of $\zeta=0$ and for which:

(i) $(g_1(\zeta),\cdots,g_\beta(\zeta))\in X$ *when* $\zeta\in V$,

(ii) $(g_1(\zeta),\cdots,g_\beta(\zeta))$ *is a regular point of* f_i, $i=1,\cdots,\alpha$, *when* $\zeta\in V\backslash\{0\}$, *and* $\lim_{\zeta\to 0}f_i(g_1(\zeta),\cdots,g_\beta(\zeta))=b_i$, $i=1,\cdots,\alpha$,

(iii) $g_{i_0}(\zeta)\neq 0$ *in* V,

(iv) $g_1(0),\cdots,g_\beta(0))=0$.

The proofs of this lemma and the next are deferred to the Appendix. Now let $\mathscr{A}$ be any topological space and, for z in some domain of $\mathbb{C}^\beta$, let $\mathscr{M}(z)$ be an analytic family in $\mathscr{A}$. Then the family $\mathscr{M}(z)$ is a *complete covering* for $a\in\mathscr{A}$ if $\mathscr{M}(0)=a$ and, for any neighborhood V of the origin in $\mathbb{C}^\beta$, $\{\mathscr{M}(z): z\in V\}$ contains a neighborhood of a.

LEMMA 5. *Let J be a $p\times p$ complex matrix and $\mathscr{M}$ an invariant subspace of J. There exists an analytic family of subspaces $\mathscr{M}(\varepsilon)\subset\mathbb{C}^p$, $(\varepsilon\in\mathbb{C}^\beta)$, which is a complete covering of $\mathscr{M}$ (in the metric $\tilde{\theta}$) and a variety $X\subset\mathbb{C}^\beta$ such that for any neighborhood V of zero in $\mathbb{C}^\beta$, the set $S=\{\mathscr{M}(\varepsilon):\varepsilon\in X\cap V\}$ consists of subspaces invariant under J and contains all subspaces in a neighborhood of $\mathscr{M}$ which are invariant under J.*

THEOREM 9. *Let $\mu_0\in S_1$ and assume that there exists a monic right divisor $L_1(\lambda)$ of $L(\lambda,\mu_0)$ which is not isolated. Then either,*

(a) *there is a family of monic right divisors $L_1(\lambda,\mu)$ of $L(\lambda,\mu)$ which is analytic in a neighborhood of μ_0, for which $L_1(\lambda,\mu_0)=L_1(\lambda)$, and for which μ_0 is a regular point*; *or*

(b) *there is a family of monic right divisors $L_1(\lambda,\mu)$ of $L(\lambda,\mu)$ which is analytic in a deleted neighborhood of μ_0, for which $\lim_{\mu\to\mu_0}L_1(\lambda,\mu)=L_1(\lambda)$, and for which μ_0 is an algebraic branch point.*

Proof. Since $L_1(\lambda)$ is not isolated, there is a sequence $\{L_1(\lambda,\mu_n)\}$ of monic right divisors for $\{L(\lambda,\mu_n)\}$ $(n=1,2,\cdots)$ where $\mu_n\to\mu_0$ as $n\to\infty$ ($\mu_n\neq\mu_0$ for any n) and $L_1(\lambda,\mu_n)\to L_1(\lambda)$. We may also suppose that $\{\mu_n\}\subset\Omega\backslash(S_1\cup S_2)$.

Let X_μ, J_μ be canonical matrices for $L(\lambda,\mu)$ defined in a deleted neighborhood of μ_0 in such a way that they have an algebraic branch point at μ_0 with branch multiplicity μ. It follows, as in Theorem 7, that X_μ, J_μ may be viewed as analytic functions of $\nu=(\mu-\mu_0)^{1/u}$ in some neighborhood U_1 of μ_0, and where convenient we may write X_ν, J_ν for X_μ, J_μ. Note that, although X_{μ_0}, J_{μ_0} exist, they are not necessarily canonical matrices for $L(\lambda,\mu_0)$. However, in a *deleted* neighborhood of μ_0 all matrices J_μ have the same invariant subspaces. We may assume that the sequence $\{\mu_n\}$ is in U_1.

Now let $\mathscr{L}_n$ be the supporting subspace for divisor $L_1(\lambda,\mu_n)$ with respect to J_{μ_n}. Since the metric space of subspaces is compact there is a convergent subsequence of $\{\mathscr{L}_n\}_{n=1}^\infty$. Without loss of generality we may assume that $\{\mathscr{L}_n\}$ is convergent with

$\lim_{n\to\infty} \mathcal{L}_n = \mathcal{L}$ and then it is easily seen that $\mathcal{L}$ is invariant under J_μ for $\mu \neq \mu_0$, $\mu \in U_1$.

Using Lemma 5 we assert the existence of an analytic family of subspaces $\mathcal{M}(\varepsilon)$, ($\varepsilon \in \mathbb{C}^\beta$), a variety X, and a neighborhood U_2 of the origin in $\mathbb{C}^\beta$ for which $\mathcal{M}(0) = \mathcal{L}$, the set of subspaces $\mathcal{M}(\varepsilon)$ for which $\varepsilon \in X \cap U_2$ consists of J_μ-invariant subspaces ($\mu \in U_1 \backslash \{\mu_0\}$), and $\mathcal{L}_n \in S$, $n = 1, 2, \cdots$. Note that there exists a sequence $\{\varepsilon^{(n)}\} \supset \mathbb{C}^\beta$ such that $\varepsilon^{(n)} \to 0$ as $n \to \infty$ and $\mathcal{L}_n = \mathcal{M}(\varepsilon^{(n)})$.

There is a neighborhood U_4 of the origin in $\mathbb{C}^\beta$ in which a family of basis vectors $x_1(\varepsilon), \cdots, x_{nk}(\varepsilon)$ for $\mathcal{M}(\varepsilon)$ exists and depends analytically on ε. Now form the matrix $V(\nu, \varepsilon)$ which is the representation (formed using this basis) of the restriction to $\mathcal{M}(\varepsilon)$ of

$$\begin{bmatrix} X_\nu \\ X_\nu J_\nu \\ \vdots \\ X_\nu J_\nu^{k-1} \end{bmatrix}.$$

Here, k is the degree of $L_1(\lambda)$ (and hence of each $L_1(\lambda, \mu_n)$). If U_3 is the neighborhood of the origin in the γ-plane corresponding to U_1 in the μ-plane then $V(\nu, \varepsilon)$ is an analytic function in the neighborhood $U = U_3 \times U_4$ of the origin in $\mathbb{C} \times \mathbb{C}^\beta$. Furthermore, det $V(\mu_n, \varepsilon^{(n)}) \neq 0$ for each n implies that det $V(\nu, \varepsilon) \not\equiv 0$ in U. Hence we may construct the polynomial

$$L_1(\lambda, \nu, \varepsilon) = I\lambda^k - X_\nu J_\nu^k (W_1 + W_2\lambda + \cdots + W_k\lambda^{k-1})$$

where

$$[W_1 \cdots W_k] = [V(\nu, \varepsilon)]^{-1},$$

and it is evident that $L_1(\lambda, \nu_n, \varepsilon^{(n)}) = L_1(\lambda, \mu_n)$ where $\nu_n = (\mu_n - \mu_0)^{1/u}$.

Furthermore $L_1(\lambda, \nu, \varepsilon)$ is a meromorphic function in U and does not have a pole at the origin since

$$\lim_{n\to\infty} L_1(\lambda, \nu_n, \varepsilon^{(n)}) = L_1(\lambda).$$

Let $f_i(\nu, \varepsilon)$, $i = 1, \cdots, \alpha$, be all the elements of $L_1(\lambda, \nu, \varepsilon)$ having an ambiguous point at (0, 0) in the neighborhood U. Since $\lim_{n\to\infty} f_i(\nu_n, \varepsilon_n) = b_i$ exists for $i = 1, \cdots, \alpha$ (b_i is the corresponding element of $L_1(\lambda)$) Lemma 4 applies where ν plays the role of the i_0 coordinate in hypothesis (d) and $X \cap U$ (with X as in Lemma 5) plays the role of the variety X in Lemma 4. Hence there exists a family of divisors $L_1(\lambda, \zeta)$ which is analytic in a neighborhood of the origin in a ζ-plane and for which $\lim_{\zeta\to 0} L_1(\lambda, \zeta) = L_1(\lambda)$. We write $\nu = g(\zeta) \not\equiv 0$ and g is analytic in a neighborhood of $\zeta = 0$.

Now we have power series expansions $\nu = \sum_{i=0}^\infty \alpha_i \zeta^i$ and

$$L_1(\lambda, \zeta) = \sum_{i=0}^\infty D_i(\lambda)\zeta^i. \tag{25}$$

Inverting the series for ν, write $\zeta = \sum_{i=0}^\infty c_i \nu^{i/p}$ where p is the least integer for which $\alpha_p \neq 0$. Finally set $\nu = (\mu - \mu_0)^{1/u}$ and substitute in (25) to obtain

$$L_1(\lambda, \mu) = \sum_{i=0}^\infty \hat{D}_i(\lambda)(\mu - \mu_0)^{i/k}$$

where $k = up$. The conclusion follows: $L_1(\lambda, \mu)$ is an analytic family of right divisors of $L(\lambda, \mu)$ in a neighborhood of μ_0, with the possible exception of a branch point at $\mu = \mu_0$ when the branch multiplicity is a divisor of up. □

It remains an open question whether the local continuation of $L_1(\lambda)$ obtained in this theorem has a global extension of the kind established in Theorem 6.

Appendix. We prove here Lemmas 4 and 5.

Proof of Lemma 4. The proof consists of two parts; in the first part an auxiliary result is proved.

(a) Let $B \subset U_1$ be a variety in a neighborhood U_1 of the point $0 \in \mathbb{C}_\beta$. Suppose that there exists a sequence of points $a_n = (z_{1n}, \cdots, z_{\beta n}) \in B\backslash\{0\}$ such that $a_n \to 0$, and let i_0 be an index for which infinitely many $z_{i_0 n}$ are nonzero.

Let $C_1, C_2, \cdots, C_\gamma$ be varieties in the neighborhood U_1 of zero such that

$$a_n \notin C_1 \cup C_2 \cup \cdots \cup C_\gamma; \qquad n = 1, 2, \cdots.$$

Then there exist scalar functions $f_1(\zeta), \cdots, f_\beta(\zeta)$ analytic in a neighborhood U_2 of zero of the ζ-plane with the following properties:

(i) $(f_1(0), \cdots, f_\beta(0)) = 0$;

(ii) $(f_1(\zeta), \cdots, f_\beta(\zeta)) \subset B$ for $\zeta \in U_2$;

(iii) $f_{i_0}(\zeta) \neq 0$ for $\zeta \in U_2\backslash\{0\}$;

(iv) $(f_1(\zeta), \cdots, f_\beta(\zeta)) \notin C_1 \cup C_2 \cup \cdots \cup C_\gamma$ for $\zeta \in U_2\backslash\{0\}$.

To prove this statement, we shall use definitions and notation from R. C. Gunning and H. Rossi [15].

By joining the variety $\{z \in \mathbb{C}_\beta \mid z_{i_0} = 0\}$ to the set $\{C_1, \cdots, C_\gamma\}$, we can omit property (iii). Without loss of generality suppose that $C_1 \cup C_2 \cup \cdots \cup C_\gamma \subset B$ (taking $C_j \cap B$ in place of C_j, if necessary) and B is irreducible (taking the irreducible branch of B in which there are infinitely many a_n, see [15, Thm. 15, Chap. II E]). We suppose also that $a_n \in U_1$, $n = 1, 2, \cdots$; that $C_1, C_2, \cdots, C_\gamma$ are irreducible; that $0 \in \bigcap_{i=1}^{\gamma} C_i$; and that $C_i \neq \{0\}$; $i = 1, \cdots, \gamma$. Then $\dim C_i < \dim B$ (see [16, Thm. 9(a), p. 53]). Let P_i be the set of all β-ples $(a_1, \cdots, a_\beta)$ such that $\sum_{j=1}^{\beta} a_j z_j \in \operatorname{id} C_i$; $i = 1, \cdots, \gamma$ (for the definition of id C_i see [15, Chap. II E]); let $P_{\gamma+1}$ be the set of all β-ples such that $\sum a_j z_j \in \operatorname{id} B$. Then P_i is a proper linear subspace in $\mathbb{C}_\beta$ (because if $P_i = \mathbb{C}_\beta$ then $C_i = \{0\}$ or $B = \{0\}$). Let $f(z) = \sum_{i=1}^{\beta} a_i z_i$ be a linear form such that $(a_1, \cdots, a_\beta) \notin \bigcup_{i=1}^{\gamma+1} P_i$. By definition of P_i, $f(z) \notin \bigcup_{i=1}^{\gamma} (\operatorname{id} C_i) \cup (\operatorname{id} B)$. Then [15, Thm. 14, Chap. III C]:

$$\dim (B \cap V(f)) = \dim B - 1;$$

$$\dim (C_1 \cap V(f) = \dim C_1 - 1.$$

Now use induction on $\dim B$. (Note that the induction hypothesis is applicable to the varieties $B \cap V(f)$ and $C_1 \cap V(f), \cdots, C_\gamma \cap V(f)$, because $\dim (C_i \cap V(f)) < \dim (B \cap V(f))$, and therefore there exists a sequence $a'_n \in B \cap V(f)$, $a'_n \to 0$, such that $a'_n \notin \bigcup_{i=1}^{\alpha} (C_i \cap V(f))$. The functions $f_1(\zeta), \cdots, f_\beta(\zeta)$ which satisfy (i)–(iv) for the varieties $C_j \cap V(f)$, $j = 1, \cdots, \gamma$, $B \cap V(f)$, satisfy (i)–(iv) also for the original varieties C_j, $j = 1, \cdots, \gamma$, and B.)

So, we have to consider only the case $\dim B = 1$, and in this case we can suppose that $\gamma = 0$ (i.e. there is no $C_1, \cdots, C_\gamma$). Now (i)–(iv) follow from the local description of one dimensional varieties (see Lemma 3.3 in J. Milnor [17]).

(b) Now we shall prove Lemma 4 itself. Let $h_{i1}(z)$ and $h_{i2}(z)$ be analytic functions in a neighborhood of zero such that

$$f_i(z) h_{i1}(z) = h_{i2}(z), \qquad i = 1, \cdots, \alpha.$$

Then $h_{i1}(0) = h_{i2}(0) = 0$ (because $z = 0$ is an ambiguous point). Without loss of generality we can suppose that $b_1 = b_2 = \cdots = b_\alpha = 0$. Introduce new complex variables $\zeta =$

$(\zeta_1, \cdots, \zeta_\alpha) \in \mathbb{C}_\alpha$ and define analytic sets

$$X_i = \{(z, \zeta) \in \mathbb{C}_{\beta+\alpha} \mid h_{i2}(z) - \zeta_i \cdot h_{i1}(z) = 0\}, \qquad i = 1, 2, \cdots, \alpha.$$

Write $b_{in} = f_i(z^{(n)})$; and $\tilde{X} = \{(z, \zeta) \in \mathbb{C}_{\beta+\alpha} \mid z \in X\}$. Then for $n = 1, 2, \cdots$, $(z_{1n}, z_{2n}, \cdots, z_{\beta n}, b_{1n}, b_{2n}, \cdots, b_{\alpha n}) \in (X_1 \cap X_2 \cap \cdots \cap X_\alpha) \cap \tilde{X}$, and $(z_{1n}, z_{2n}, \cdots, z_{\beta n}, b_{1n}, b_{2n}, \cdots, b_{\alpha n}) \to 0$. Applying part (a) with

$$C_i = \{(z, \zeta) \in \mathbb{C}_{\beta+\alpha} \mid h_{i1}(z) = 0\}, \qquad i = 1, 2, \cdots, \alpha,$$

we find $\alpha + \beta$ functions $g_1(\eta), \cdots, g_{\alpha+\beta}(\eta)$, analytic in a neighborhood of zero, such that for η in some neighborhood U of zero in the η-plane:

$$(g_1(\eta), \cdots, g_{\beta+\alpha}(\eta)) \in X_1 \cap X_2 \cap \cdots \cap X_\alpha \cap \tilde{X},$$

and for $\eta \in U \backslash \{0\}$,

$$(g_1(\eta), \cdots, g_{\beta+\alpha}(\eta)) \notin C_1 \cup \cdots \cup C_\alpha,$$

$g_i(0) = 0$, for $i = 1, 2, \cdots, \beta + \alpha$, and $g_{i_0}(\eta) \neq 0$. Then for η in this deleted neighborhood $(g_1(\eta), \cdots, g_\beta(\eta))$ is a regular point of $f_1(z), \cdots, f_\alpha(z)$ and so

$$f_i(g_1(\eta), \cdots, g_\beta(\eta)) = g_{\beta+i}(\eta); \qquad i = 1, \cdots, \alpha,$$

is well defined, and

$$\lim_{\eta \to 0} f_i(g_1(\eta), \cdots, g_\beta(\eta)) = 0.$$

The condition (i) of Lemma 4 follows from the definition of $\tilde{X}$. □

Proof of Lemma 5. Let $q = \dim \mathcal{M}$ and let $v_1, v_2, \cdots, v_q$ be a basis for $\mathcal{M}$; we regard v_i as vectors written in the standard basis. The matrix $W = (v_1, v_2, \cdots, v_q)$ has independent columns. Suppose, for example, that the q upper rows in W are independent. Then $\mathcal{M}$ is spanned by the columns of a matrix W' of the form

$$W' = \begin{bmatrix} & I & \\ v_1' & v_2' \cdots & v_q' \end{bmatrix},$$

where v_i' are $(p-q)$-dimensional vectors.

Since $\mathcal{M}$ is invariant under J, we have $JW' = W'A$ for some matrix A. Consider the analytic family of matrices

$$W'(\varepsilon) = \begin{bmatrix} & I & \\ v_1'(\varepsilon) & v_2'(\varepsilon) \cdots & v_q'(\varepsilon) \end{bmatrix}$$

where

$$v_i'(\varepsilon) = \begin{bmatrix} v_{i1}' + \varepsilon_{i1} \\ v_{i2}' + \varepsilon_{i2} \\ \vdots \\ v_{i,p-q}' + \varepsilon_{i,p-q} \end{bmatrix}$$

(here v_{ij}' are the coordinates of v_i'). So $W'(\varepsilon)$ depends on $q(p-q)$ complex parameters $\varepsilon = (\varepsilon_{ij}; j = 1, \cdots, p-q; i = 1, \cdots, q)$.

Let $\mathcal{M}(\varepsilon)$ be the subspace spanned by the columns of $W'(\varepsilon)$. Clearly, $\mathcal{M}(\varepsilon)$ is a complete covering of $\mathcal{M}$. The subspace $\mathcal{M}(\varepsilon)$ is invariant under J if and only if $JW'(\varepsilon) = W'(\varepsilon)A(\varepsilon)$ for some matrix $A(\varepsilon)$. Let J_q be the matrix consisting of the upper q rows of J. Then the equality $JW'(\varepsilon) = W'(\varepsilon)A(\varepsilon)$ gives $A(\varepsilon) = J_q W'(\varepsilon)$. So the necessary and sufficient condition for $\mathcal{M}(\varepsilon)$ to be invariant under J is $JW'(\varepsilon) = W'(\varepsilon)J_q W'(\varepsilon)$. This condition defines the desired (algebraic) variety in a neighborhood of zero in ε-space. □

REFERENCES

[1] H. BAUMGÄRTEL, *Endlich-dimensionale Analytische Störungstheorie*, Akademie-Verlag, Berlin, 1972.

[2] R. J. DUFFIN, *A minimax theory for overdamped networks*, J. Rational Mech. Anal., 4 (1955), pp. 221–233.

[3] I. GOHBERG AND M. G. KREIN, *The basic propositions on defect numbers, root numbers and indices of linear operators*, Uspehi Mat. Nauk, 12 (1957), pp. 43–118 (in Russian); Amer. Math. Soc. Transl., 13 (1960), pp. 185–264.

[4] I. GOHBERG, P. LANCASTER AND L. RODMAN, *Spectral analysis of matrix polynomials, I Canonical forms and divisors*, J. Linear Algebra Appl., 20 (1978), pp. 1–44.

[5] ———, *Spectral analysis of matrix polynomials, II Resolvent forms and spectral divisors*, Ibid., 21 (1978), pp. 65–88.

[6] ———, *Representations and divisibility of operator polynomials*, Canad. J. Math., 30 (1978), pp. 1045–1069.

[7] I. GOHBERG AND Y. LEITERER, *On cocycles, operator functions and families of subspaces*, Mat. Issled., 8 (1973), pp. 23–55. (In Russian.)

[8] I. GOHBERG AND A. S. MARKUS, *Two theorems on the opening of subspaces of Banach space*, Upsehi Mat. Nauk, 14 (1959), pp. 135–140. (In Russian.)

[9] I. GOHBERG AND E. I. SIGAL, *An operator generalization of the logarithmic residue theorem and the theorem of Rouché*, Math. USSR-Sb., 13 (1971), pp. 603–625.

[10] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.

[11] M. G. KREIN, M. A. KRASNOSEL'SKII AND D. P. MILMAN, *On the defect numbers of linear operators in Banach space and on some geometric problems*, Sbornik Trud. Inst. Mat. Akad. Nauk Ukr. SSR, 11 (1948), pp. 97–112. (In Russian.)

[12] M. G. KREIN AND H. LANGER, *On some mathematical principles of linear theory of damped vibrations of continua*, Proc. Int'l. Symp. in Appl. of the Theory of Functions in Continuum Mechanics, 1965, pp. 283–322. (In Russian.)

[13] P. LANCASTER, *A fundamental theorem on lambda-matrices with applications I, Ordinary differential equations with constant coefficients*, J. Linear Algebra Appl., 18 (1977), pp. 189–211.

[14] M. A. SHUBIN, *On holomorphic families of subspaces in Banach space*, Mat. Issled., 5 (19), pp. 153–165. (In Russian.)

[15] R. C. GUNNING AND H. ROSSI, *Analytic Functions of Several Complex Variables*, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[16] R. C. GUNNING, *Lectures on Complex Analytic Varieties*, Princeton University Press, Princeton, NJ, 1970.

[17] J. MILNOR, *Singular Points of Complex Hypersurfaces*, Princeton University Press and The University of Tokyo Press, Princeton, NJ, 1968.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0008 $01.00/0

ON ASYMPTOTIC AVERAGE PROPERTIES OF ZEROS OF ORTHOGONAL POLYNOMIALS*

PAUL G. NEVAI† AND JESUS S. DEHESA‡

Abstract. Distribution of zeros of orthogonal polynomials is investigated by means of the coefficients of the three-term recurrence relation which generates the orthogonal polynomials.

1. Introduction. The purpose of this paper is to study certain asymptotic average properties of zeros of orthogonal polynomials given by a three-term recurrence relation. The exploration of these properties is a very relevant problem in many-body physics, for instance in investigating the nuclear level density in nuclear physics and in searching for the density of electronic states of disordered systems in solid state physics. The reason is as follows: the Hamiltonian operator H which completely describes any physical system appears as a real and symmetric matrix. The conventional way to obtain its eigenvalues and eigenvectors is tridiagonalization of H and then diagonalization of the resulting Jacobi matrix. In fact, it is often possible, after some physical approximations, to write the Hamiltonian of the given physical system as a Jacobi matrix with very large dimension n, and then one has "only" to diagonalize it to get its eigensolutions. Much attention has recently been concentrated in obtaining certain average properties of the eigenvalue spectrum of the Jacobi Hamiltonian of the system such as its eigenvalue density. This problem is equivalent to investigating the distribution of zeros of orthogonal polynomials by means of the coefficients of their three-term recurrence relation.

Let $\{\alpha_n\}_{n=0}^{\infty}$ and $\{\gamma_n > 0\}_{n=0}^{\infty}$ be given sequences of real numbers. Let the system of polynomials $\{p_n(x)\}_{n=0}^{\infty}$ be defined by the three-term recurrence relation

$$xp_{n-1}(x)=\frac{\gamma_{n-1}}{\gamma_n}p_n(x)+\alpha_{n-1}p_{n-1}(x)+\frac{\gamma_{n-2}}{\gamma_{n-1}}p_{n-2}(x), \tag{1}$$

$p_{-1}=0$, $p_0=\gamma_0$ and $n=1, 2, \cdots$. It is well known that there exists a positive measure $d\alpha$ acting on the real line such that the support of $d\alpha$ is infinite and

$$\int_{-\infty}^{\infty} p_n(x)p_m(x)\, d\alpha(x)=\delta_{nm}. \tag{2}$$

(See e.g., [2, p. 60].) The polynomials $p_n(x)$ and the coefficients in the recurrence formula (1) can easily be expressed in terms of the measure $d\alpha$. Let π_n be the set of all polynomials of degree at most n. Define the Christoffel function $\lambda_n(d\alpha, x)$ by

$$\lambda_n(d\alpha, x)=\min_{\substack{P\in\pi_{n-1}\\ P(x)=1}}\int_{-\infty}^{\infty} P^2(t)\, d\alpha(t). \tag{3}$$

Then

$$p_n^2(x)=\lambda_{n+1}^{-1}(d\alpha, x)-\lambda_n^{-1}(d\alpha, x) \tag{4}$$

* Received by the editors May 9, 1978 and in revised form November 2, 1978.

† Department of Mathematics, Ohio State University, Columbus, Ohio 43210. The work of this author was supported in part by the National Science Foundation under Grants MCS 77-06687 and MCS 78-01868.

‡ Departmento de Fisica Nuclear, Facultad de Ciencias, Universidad de Granada, Granada, Spain.

and

$$\alpha_n = \int_{-\infty}^{\infty} x p_n^2(x)\, d\alpha(x), \qquad \frac{\gamma_{n-1}}{\gamma_n} = \int_{-\infty}^{\infty} x p_{n-1}(x) p_n(x)\, d\alpha(x). \tag{5}$$

For this reason we will write $\alpha_n(d\alpha)$, $\gamma_n(d\alpha)$ and $p_n(d\alpha, x)$ instead of α_n, γ_n and $p_n(x)$ where $d\alpha$ means a positive measure for which (2) is satisfied. This notation is also justified by the fact that if $d\alpha$ is an arbitrary positive measure on the real line such that the support of $d\alpha$ is infinite and $x^k \in L^2(d\alpha)$ for $k = 0, 1, \cdots$ and $p_n(x)$, α_n and γ_n are defined by (3), (4) and (5) then both the recurrence formula (1) and the orthogonality relation (2) will be satisfied. In the following such a measure $d\alpha$ will be called a weight.

The zeros of $p_n(d\alpha, x)$, which are real and simple, will be denoted by $x_{kn}(d\alpha)$: $x_{1n}(d\alpha) > x_{2n}(d\alpha) > \cdots > x_{nn}(d\alpha)$.

2. Main result. Our main result is the following

THEOREM 1. *Let $\varphi\colon \mathbb{R}^+ \to \mathbb{R}^+$ be a nondecreasing function such that for every fixed $t \in \mathbb{R}$*

$$\lim_{x \to +\infty} \frac{\varphi(x+t)}{\varphi(x)} = 1.$$

Assume that there exists two numbers a and $b \geqq 0$ such that the coefficients in the recurrence relation (1) *satisfy*

$$\lim_{n \to \infty} \frac{\alpha_n(d\alpha)}{\varphi(n)} = a$$

and

$$\lim_{n \to \infty} \frac{2\gamma_{n-1}(d\alpha)}{\gamma_n(d\alpha)\varphi(n)} = b.$$

Then for every nonnegative integer M

$$\lim_{n \to \infty} \frac{\sum_{k=1}^n [x_{kn}(d\alpha)]^M}{\int_0^n [\varphi(t)]^M\, dt} = K_M(a, b)$$

where K_M is defined by

$$K_M(a, b) = \begin{cases} a^M & \text{for } b = 0, \\ \dfrac{1}{\pi} \displaystyle\int_{a-b}^{a+b} \frac{t^M}{\sqrt{b^2 - (t-a)^2}}\, dt & \text{for } b > 0. \end{cases} \tag{6}$$

The case $\varphi \equiv 1$ of this theorem has been proved in [3]. In order to prove Theorem 1 we will need two lemmata.

In the following, the fundamental polynomials of Lagrange interpolation corresponding to $\{x_{kn}(d\alpha)\}_{k=1}^n$ will be denoted by $l_{kn}(\delta\alpha, x)$:

$$l_{kn}(d\alpha, x) = \frac{p_n(d\alpha, x)}{p_n'(d\alpha, x_{kn}(d\alpha))[x - x_{kn}(d\alpha)]}$$

for $k = 1, 2, \cdots, n$. The numbers $\lambda_{kn}(d\alpha)$ defined by

$$\lambda_{kn}(d\alpha) = \lambda_n(d\alpha, x_{kn}(d\alpha))$$

$(k=1,2,\cdots,n)$ are called Christoffel numbers. The identity

$$\int_{-\infty}^{\infty} P(x)\,d\alpha(x)=\sum_{k=1}^{n}\lambda_{kn}(d\alpha)P(x_{kn}(d\alpha)),$$

which is true for every polynomial P of degree at most $2n-1$, is called the Gauss–Jacobi quadrature formula. Sometimes we will omit unnecessary indices and parameters, e.g. λ_{kn} stands for $\lambda_{kn}(d\alpha)$.

LEMMA 2. *Let M be a fixed nonnegative integer. Let α be an arbitrary weight. Define $Z_n(M, d\alpha)$ by*

$$Z_n(M,d\alpha)=\sum_{k=1}^{n}[x_{kn}(d\alpha)]^M-\sum_{k=1}^{n-1}\int_{-\infty}^{\infty}x^Mp_k^2(d\alpha,x)\,d\alpha(x).$$

Then the inequality

$$(7)\qquad |Z_n(M,d\alpha)|\leqq\frac{(M-1)M}{2}\left[\max_{0\leqq j\leqq n+M-1}|\alpha_j(d\alpha)|+2\max_{1\leqq j\leqq n+M-1}\frac{\gamma_{j-1}(d\alpha)}{\gamma_j(d\alpha)}\right]^M$$

holds for every $n=1,2,\cdots$.

Proof. Since

$$\sum_{k=1}^{n-1}p_k^2(x)=\sum_{k=1}^{n}\frac{l_{kn}^2(x)}{\lambda_{kn}},$$

and

$$\lambda_{kn}=\int_{-\infty}^{\infty}l_{kn}^2(x)\,d\alpha(x).$$

(See e.g., [2, p. 25].) We can write $Z_n(M, d\alpha)$ in the form

$$Z_n(M,d\alpha)=\sum_{k=1}^{n}\int_{-\infty}^{\infty}(x_{kn}^M-x^M)\frac{l_{kn}^2(x)}{\lambda_{kn}}\,d\alpha(x),$$

that is

$$Z_n(M,d\alpha)=\sum_{j=0}^{M-1}\sum_{k=1}^{n}x_{kn}^{M-1-j}\int_{-\infty}^{\infty}x^j(x_{kn}-x)\frac{l_{kn}^2(x)}{\lambda_{kn}}\,d\alpha(x).$$

It is well known that

$$(x-x_{kn})l_{kn}(x)=\frac{\gamma_{n-1}}{\gamma_n}\lambda_{kn}p_{n-1}(x_{kn})p_n(x)$$

(see e.g., [2, p. 114].) Therefore we obtain

$$(8)\qquad -Z_n(M,d\alpha)=\frac{\gamma_{n-1}}{\gamma_n}\sum_{j=0}^{M-1}\sum_{k=1}^{n}p_{n-1}(x_{kn})x_{kn}^{M-1-j}\int_{-\infty}^{\infty}x^jp_n(x)l_{kn}(x)\,d\alpha(x).$$

Expand $l_{kn}(x)$ into a Fourier series in $\{p_l(x)\}$. Since $l_{kn}(x)$ is a polynomial of degree $n-1$ we have

$$l_{kn}(x)=\sum_{l=0}^{n-1}\int_{-\infty}^{\infty}l_{kn}(t)p_l(t)\,d\alpha(t)p_l(x).$$

By the Gauss–Jacobi quadrature formula

$$\int_{-\infty}^{\infty}l_{kn}(t)p_l(t)\,d\alpha(t)=\lambda_{kn}p_l(x_{kn}).$$

Thus we get

$$l_{kn}(x)=\lambda_{kn}\sum_{l=0}^{n-1}p_l(x_{kn})p_l(x).$$

Putting this representation for $l_{kn}(x)$ in (8) we obtain

$$-Z_n(M,d\alpha)=\frac{\gamma_{n-1}}{\gamma_n}\sum_{j=0}^{M-1}\sum_{l=0}^{n-1}\sum_{k=1}^{n}\lambda_{kn}p_l(x_{kn})p_{n-1}(x_{kn})x_{kn}^{M-1-j}$$
$$\cdot\int_{-\infty}^{\infty}x^jp_n(x)p_l(x)\,d\alpha(x).$$

It follows from the orthogonality properties that

$$\int_{-\infty}^{\infty}x^jp_n(x)p_l(x)\,d\alpha(x)=0$$

for $j+l<n$. Consequently

$$-Z_n(M,d\alpha)=\frac{\gamma_{n-1}}{\gamma_n}\sum_{j=1}^{M-1}\sum_{l=n-j}^{n-1}\sum_{k=1}^{n}\lambda_{kn}p_l(x_{kn})p_{n-1}(x_{kn})x_{kn}^{M-1-j}\int_{-\infty}^{\infty}x^jp_n(x)p_l(x)\,d\alpha(x).$$

In this triple sum $j+n+l\leqq 2n+M-2<2(n+M)-1$. Hence, by the Gauss–Jacobi quadrature formula

$$\int_{-\infty}^{\infty}x^jp_n(x)p_l(x)\,d\alpha(x)=\sum_{m=1}^{n+M}\lambda_{m,n+M}x^j_{m,n+M}p_n(x_{m,n+M})p_l(x_{m,n+M}).$$

Therefore we get the formula

$$-Z_n(M,d\alpha)=\frac{\gamma_{n-1}}{\gamma_n}\sum_{j=1}^{M-1}\sum_{l=n-j}^{n-1}\sum_{k=1}^{n}\sum_{m=1}^{n+M}\lambda_{kn}\lambda_{m,n+M}x_{kn}^{M-1-j}x^j_{m,n+M}\cdot$$
$$\cdot\,p_l(x_{kn})p_{n-1}(x_{kn})p_n(x_{m,n+M})p_l(x_{m,n+M}). \tag{9}$$

If $M=0,1$ then by (9), $Z_n(M,d\alpha)=0$ for every n so that (y) is certainly true. Now let $M>1$. Let X_n denote the zero of $p_n(x)$ with largest absolute value. Since $|X_n|$ is an increasing function of n we obtain from (9) that

$$|Z_n(M,d\alpha)|\leqq\frac{\gamma_{n-1}}{\gamma_n}|X_{n+M}|^{M-1}\sum_{j=1}^{M-1}\sum_{l=n-j}^{n-1}\left\{\sum_{k=1}^{n}\lambda_{kn}|p_l(x_{kn})|\right.$$
$$\left.\cdot\,|p_{n-1}(x_{kn})|\sum_{m=1}^{n+M}\lambda_{m,n+M}|p_n(x_{m,n+M})|\,|p_l(x_{m,n+M})|\right\}.$$

By Schwarz' inequality, the expression between braces is not greater than 1. Hence

$$|Z_n(M,d\alpha)|\leqq\frac{\gamma_{n-1}}{\gamma_n}|X_{n+M}|^{M-1}\sum_{j=1}^{M-1}\sum_{l=n-j}^{n-1}1$$
$$=\frac{(M-1)M}{2}\frac{\gamma_{n-1}}{\gamma_n}|X_{n+m}|^{M-1}. \tag{10}$$

The next step is to get an estimate for $|X_n|$. It follows from the recurrence formula that

$$x\sum_{j=0}^{n-1}p_j^2(x)=\sum_{j=0}^{n-1}\alpha_jp_j^2(x)+2\sum_{j=0}^{n-2}\frac{\gamma_j}{\gamma_{j+1}}p_j(x)p_{j+1}(x)+\frac{\gamma_{n-1}}{\gamma_n}p_{n-1}p_n(x).$$

Putting here $x = x_{kn}$ $(k = 1, 2, \cdots, n)$ and using Schwarz' inequality we obtain

$$|x_{kn}| \leqq \max_{0 \leqq j \leqq n-1} |\alpha_j| + 2 \max_{1 \leqq j \leqq n-1} \frac{\gamma_{j-1}}{\gamma_j} \qquad (k = 1, 2, \cdots, n). \tag{11}$$

Hence

$$|X_{n+M}| \leqq \max_{0 \leqq j \leqq n+M-1} |\alpha_j| + 2 \max_{1 \leqq j \leqq n+M-1} \frac{\gamma_{j-1}}{\gamma_j}. \tag{12}$$

Now (7) follows from (10) and (12).

LEMMA 3. *Let the conditions of Theorem* 1 *be satisfied. Let M be an arbitrary nonnegative integer. Then*

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)}\right]^M p_n^2(d\alpha, x)\, d\alpha(x) = K_M(a, b) \tag{13}$$

where K_M is defined by (6).

Proof. First let $b = 0$. Then, since

$$\left[\frac{x}{\varphi(n)}\right]^M = a^M + \sum_{j=1}^{M} \binom{M}{j} \left[\frac{x}{\varphi(n)} - a\right]^j a^{M-j},$$

the lemma will be proved if we can show that for every natural integer j

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)} - a\right]^j p_n^2(x)\, d\alpha(x) = 0. \tag{14}$$

If $j = 1$ then by (5) this is clearly true. Now let $j = 2$. Then by the recurrence formula

$$\left[\frac{x}{\varphi(n)} - a\right] p_n(x) = \frac{\gamma_n}{\varphi(n)\gamma_{n+1}} p_{n+1}(x) + \left[\frac{\alpha_n}{\varphi(n)} - a\right] p_n(x) + \frac{\gamma_{n-1}}{\varphi(n)\gamma_n} p_{n-1}(x).$$

Hence we get

$$\int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)} - a\right]^2 p_n^2(x)\, d\alpha(x) = \left(\frac{\gamma_n}{\varphi(n)\gamma_{n+1}}\right)^2 + \left(\frac{\alpha_n}{\varphi(n)} - a\right)^2 + \left(\frac{\gamma_{n-1}}{\varphi(n)\gamma_n}\right)^2$$

which implies (14) for $j = 2$. If $j > 2$ by the Gauss–Jacobi quadrature formula

$$\int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)} - a\right]^j p_n^2(x)\, d\alpha(x) = \sum_{k=1}^{n+j} \lambda_{k,n+j} \left[\frac{x_{k,n+j}}{\varphi(n)} - a\right]^j p_n^2(x_{k,n+j}).$$

Then

$$\begin{aligned} \left|\int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)} - a\right]^j p_n^2(x)\, d\alpha(x)\right| &\leqq \max_{1 \leqq k \leqq n+j} \left|\frac{x_{k,n+j}}{\varphi(n)} - a\right|^{j-2} \\ &\quad \cdot \sum_{k=1}^{n+j} \lambda_{k,n+j} \left[\frac{x_{k,n+j}}{\varphi(n)} - a\right]^2 p_n^2(x_{k,n+j}) \\ &= \max_{1 \leqq k \leqq n+j} \left|\frac{x_{k,n+j}}{\varphi(n)} - a\right|^{j-2} \int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)} - a\right]^2 p_n^2(x)\, d\alpha(x). \end{aligned}$$

Now applying (11) we obtain (14) for every j. If $b > 0$, define the numbers $\alpha_{nk}(d\alpha)$ by

$$\alpha_{nk}(d\alpha) = \begin{cases} \gamma_{n-1}(d\alpha)/\gamma_n(d\alpha) & \text{for } k = n-1 \\ \alpha_n(d\alpha) & \text{for } k = n \\ \gamma_n(d\alpha)/\gamma_{n+1}(d\alpha) & \text{for } k = n+1. \end{cases}$$

Apply the recurrence formula repeatedly to get

$$x^M p_n(x) = \sum_{\substack{-1 \leq k_i \leq 1 \\ i=1,2,\cdots,M}} \alpha_{n,n+k_1} \alpha_{n+k_1,n+k_1+k_2} \cdots \alpha_{n+k_1+\cdots+k_{M-1},n+k_1+\cdots+k_M} p_{n+k_1+\cdots+k_M}(x)$$

for $n > M$. Thus

$$\int_{-\infty}^{\infty} x^M p_n^2(x)\, d\alpha(x) = \sum_{\substack{-1 \leq k_i \leq 1 \\ i=1,2,\cdots,M \\ \Sigma k_i = 0}} \alpha_{n,n+k_1} \alpha_{n+k_1,n+k_1+k_2} \cdots \alpha_{n+k_1+\cdots+k_{M-1},n}.$$

Divide both sides by $\varphi(n)^M$ and let $n \to \infty$. We see that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left[\frac{x}{\varphi(n)}\right]^M p_n^2(x)\, d\alpha(x) \tag{15}$$

exists and it depends on a and b but not on $\{\alpha_n\}$, $\{\gamma_n\}$ and φ. Therefore if we put $\alpha_n^* = a$ and $\gamma_n^*/\gamma_{n+1}^* = b/2$ for $n = 0, 1, 2, \cdots$ and $\varphi^* \equiv 1$ then (15) must equal

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} x^M p_n^2(d\alpha^*, x)\, d\alpha^*(x)$$

where α^* denotes the corresponding weight function. It is easy to show that

$$p_n(d\alpha^*, x) = U_n\left(\frac{x-a}{b}\right)$$

where

$$U_n(x) = \frac{\sin (n+1)\theta}{\sin \theta}, \qquad x = \cos \theta$$

is the Chebyshev polynomial of second kind and

$$\alpha^*(x) = \frac{2}{b^2\pi} \int_{a-b}^{x} \sqrt{b^2 - (t-a)^2}\, dt, \qquad a - b \leq x \leq a + b,$$

with supp $(d\alpha^*) = [a - b, a + b]$. Since

$$U_n^2(x) = \frac{1 - T_{2n+2}(x)}{2(1-x^2)}$$

where $T_n(x) = \cos n\theta$ $(x = \cos \theta)$, we obtain

$$\int_{-\infty}^{\infty} x^M p_n^2(d\alpha^*, x)\, d\alpha^*(x) = \frac{1}{\pi} \int_{a-b}^{a+b} \frac{x^M}{\sqrt{b^2 - (x-a)^2}}\, dx - \frac{1}{\pi} \int_{-1}^{1} (bx+a)^M T_{2n+2}(x) \frac{dx}{\sqrt{1-x^2}}.$$

Hence

$$\int_{-\infty}^{\infty} x^M p_n^2(d\alpha^*, x)\, d\alpha^*(x) = \frac{1}{\pi} \int_{a-b}^{a+b} \frac{x^M}{\sqrt{b^2 - (x-a)^2}}\, dx$$

whenever $2n + 2 > M$. Consequently, the lemma is also true for $b > 0$.

Proof of Theorem 1. By Lemma 3

$$\sum_{j=0}^{n-1} \int_{-\infty}^{\infty} x^M p_j^2(x)\, d\alpha(x) = \sum_{j=0}^{n-1} \varphi(j)^M [K_M(a, b) + \varepsilon_j]$$

where $\varepsilon_j \to 0$ when $j \to \infty$. Simple estimation yields

$$\lim_{n\to\infty} \frac{\sum_{j=0}^{n-1} \varepsilon_j \varphi(j)^M}{\sum_{j=0}^{n-1} \varphi(j)^M} = 0$$

and

$$\min_{n\to\infty} \frac{\sum_{j=0}^{n-1} \varphi(j)^M}{\int_0^n [\varphi(t)]^M \, dt} = 1.$$

Hence

$$\sum_{j=0}^{n-1} \int_{-\infty}^{\infty} x^M p_j^2(x) \, d\alpha(x) = (K_M(a, b) + \delta_n) \int_0^n [\varphi(t)]^M \, dt$$

with $\lim_{n\to\infty} \delta_n = 0$. By the assumptions in Theorem 1

$$\limsup_{n\to\infty} \frac{1}{\varphi(n)} \left[\max_{0 \leq j \leq n+M-1} |\alpha_j| + 2 \max_{1 \leq j \leq n+M-1} \frac{\gamma_{j-1}}{\gamma_j} \right] < \infty$$

and

$$\lim_{n\to\infty} \frac{\varphi(n)^M}{\int_0^n [\varphi(t)]^M \, dt} = 0.$$

Hence by Lemma 2

$$\sum_{j=1}^{n} x_{jn}^M = (K_M(a, b) + \rho_n) \int_0^n [\varphi(t)]^M \, dt$$

($\lim_{n\to\infty} \rho_n = 0$) what was to be proved.

3. Systems of orthogonal polynomials. In the following we are going to apply Theorem 1 to six systems of orthogonal polynomials for which the coefficients in the recurrence relation (1) are known. Let us note that the recently published book [7] of T. S. Chihara contains many other interesting examples of orthogonal polynomials given by a recurrence relation.

3A. The modified Lommel polynomials [4] $R_n^{(\nu)}(x)$ ($\nu > 0$) satisfy (1) with

$$\alpha_n = -n \quad \text{and} \quad \frac{\gamma_{n-1}}{\gamma_n} = \frac{\nu}{2}.$$

Hence by putting $\varphi(t) \equiv t$ we obtain

$$\lim_{n\to\infty} \frac{1}{n^{M+1}} \sum_{k=1}^{n} x_{kn}^M = \frac{(-1)^M}{M+1}.$$

3B. The Carlitz polynomials [6] $F_n^{(k)}(x)$ ($k > 0$) satisfy (1) with

$$\alpha_n = (k^2+1)(2n+1)^2 \quad \text{and} \quad \frac{\gamma_{n-2}}{\gamma_{n-1}} = k(2n-2)\sqrt{(2n-1)(2n-3)}.$$

Let $\varphi(t) \equiv t^2$. We get

$$\lim_{n\to\infty} n^{1/(2M+1)} \sum_{k=1}^{n} x_{kn}^M = \frac{1}{(2M+1)\pi} \int_{4(k-1)^2}^{4(k+1)^2} \frac{t^M}{\sqrt{64k^2 - (4k^2+4-t)^2}} \, dt.$$

3C. The Charlier polynomials [1] $C_n(x, \beta)$ $(\beta > 0)$ are defined by (1) with

$$\alpha_n = n + \beta \quad \text{and} \quad \frac{\gamma_{n-1}}{\gamma_n} = \sqrt{\beta n}.$$

If we put $\varphi(t) \equiv t$ we see that

$$\lim_{n\to\infty} \frac{1}{n^{M+1}} \sum_{k=1}^{n} x_{kn}^M = \frac{1}{M+1}.$$

3D. The Meixner polynomials [5] $M_n(x, \delta, \varepsilon)(\delta \in \mathbb{R}, \varepsilon > 0)$ satisfy (1) with

$$\alpha_n = \delta(2n+\varepsilon) \quad \text{and} \quad \frac{\gamma_{n-2}}{\gamma_{n-1}} = \sqrt{(1+\delta^2)(n-1)(n+\varepsilon-2)}.$$

Letting $\varphi(t) \equiv t$ we get

$$\lim_{n\to\infty} \frac{1}{n^{M+1}} \sum_{k=1}^{n} x_{kn}^M = \frac{1}{(M+1)\pi} \int_{2\delta-2\sqrt{1+\delta^2}}^{2\delta+2\sqrt{1+\delta^2}} \frac{t^M}{\sqrt{4+4t\delta-t^2}}\, dt.$$

3E. The Hermite polynomials [1] $h_n(x)$ satisfy (1) with

$$\alpha_n = 0 \quad \text{and} \quad \frac{\gamma_{n-1}}{\gamma_n} = \sqrt{\frac{n}{2}}.$$

Let $\varphi(t) \equiv \sqrt{2t}$. We obtain

$$\lim_{n\to\infty} \frac{1}{n^{1+(M/2)}} \sum_{k=1}^{n} x_{kn}^M = \frac{2^{1+(M/2)}}{\pi(M+2)} \int_{-1}^{1} \frac{t^M}{\sqrt{1-t^2}}\, dt.$$

3F. The Laguerre polynomials [1] $l_n^{(\beta)}(x)(\beta > -1)$ may be defined by (1) with

$$\alpha_n = 2n+1+\beta \quad \text{and} \quad \frac{\gamma_{n-1}}{\gamma_n} = \sqrt{n(n+\beta)}.$$

Thus by putting $\varphi(t) \equiv 2t$ we get

$$\lim_{n\to\infty} \frac{1}{n^{M+1}} \sum_{k=1}^{n} x_{kn}^M = \frac{2^M}{\pi(M+1)} \int_0^2 \frac{t^M}{\sqrt{1-(t-1)^2}}\, dt.$$

Since our result is new only if the function φ is different from a constant, we did not include examples when the coefficients in the recurrence relation (1) converge to finite limits.

4. Other forms. The function $K_M(a, b)$ defined by (6) can be given in other forms which are easy to compute.

$$K_M(a,b) = \sum_{j=0}^{[M/2]} b^{2j} a^{M-2j} 2^{-2j} \binom{2j}{j} \binom{M}{2j} = a^M {}_2F_1(-M/2, (1-M)/2; 1; b^2/a^2).$$

REFERENCES

[1] A. ERDÉLYI, ed., *Bateman Manuscript Project—Higher Transcendental Functions*, vols. 1 and 2, McGraw-Hill, New York, 1953.
[2] G. FREUD, *Orthogonal Polynomials*, Pergamon Press, New York, 1971.
[3] P. NEVAI, *Orthogonal polynomials*, Memoirs of the Amer. Math. Soc., to appear.

[4] D. J. Dickinson, *On certain polynomials associated with orthogonal polynomials*, Bol. Un. Mat. Ital. (3), 13 (1958), pp. 116–124.

[5] J. Meixner, *Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzengenden Function*, J. London Math. Soc., 9 (1934), pp. 6–13.

[6] L. Carlitz, *Some orthogonal polynomials related to elliptic functions*, Duke Math. J., 27 (1960), pp. 443–460.

[7] T. S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York, 1978.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0009 $01.00/0

COUNTABLY INFINITE TIME-VARYING ELECTRICAL NETWORKS*

A. H. ZEMANIAN†

Abstract. An existence and uniqueness theorem is established for the currents and voltages in a nonlinear time-varying countably infinite electrical network whose parameters are restricted to inductors and capacitors except possibly in certain branches, called joints, which are allowed to have any kind of electrical parameters. The LC form of the network allows the basic determining equation of the network to be obtained in normal form, and this in turn leads to substantially stronger results than those obtained in prior works. Similar results are obtained for linear time-varying RLC networks as well.

1. Introduction. This work is a sequel to a prior paper [5] in which a number of existence and uniqueness theorems were established for the dynamic responses of countably infinite nonlinear time-varying electrical networks. That paper followed in turn some papers on countably infinite linear time-invariant networks [2], [3], [4]. The scope of the present work lies between [5] on the one hand and [2], [3], and [4] on the other. It discusses nonlinear time-varying LC networks and also linear time-varying RLC networks.

The determining differential equations for the dynamic responses are obtained herein in normal form; this could not be done under the generality assumed in [5]. As a result, the conclusions of the present work are substantially stronger than those of [5]. In particular, existence and uniqueness theorems could be stated in [5] only for networks whose chords are sufficiently close to open circuits and whose limb branches are sufficiently close to short circuits. (See either [4] or [5] for the meanings of "chords", "limb branches", "joints", etc.) These conditions are imposed by the requirement that the network's elements satisfy Lipschitz conditions whose Lipschitz coefficients are sufficiently small. Moreover, for very complicated networks, these restrictions can be quite severe. To state this another way, consider the class of all electrical networks with a given infinite graph. [5] presents a local analysis around those degenerate networks whose branches—other than the joints—are either short circuits or open circuits. In contrast to this, the present work is global in scope in that no restrictions on the Lipschitz coefficients are imposed; that is, our networks can now be far different from the aforementioned degenerate case.

We use the notations and terminology of [5] without repeating their definitions.

2. Networks with nonlinear inductors and capacitors. Let N be a countably infinite electrical network that satisfies

Conditions 1. A full set $\mathscr{L}$ of limbs and a full set $\mathscr{J}$ of joints can be so chosen that every limb branch is a series connection of a capacitor and a voltage source, either but not both of which may be zero, and every chord is a parallel connection of an inductor and a current source, either but not both of which may be zero. (The joints may be combinations of any network parameters including resistors and conductors.) There is no mutual coupling. Furthermore, each capacitor is in general nonlinear, time-varying, and defined by an equation of the form

$$v(t)=\gamma(q(t),t)$$

where v and q are respectively the voltage drop and the charge on the capacitor, $\gamma(\cdot,\cdot)$

* Received by the editors July 6, 1978, and in final revised form November 20, 1978.

† Department of Electrical Engineering, State University of New York at Stony Brook, Stony Brook, New York 11794. This research was supported by the National Science Foundation under Grant MCS 75-05268-A03.

maps $R^1 \times [0, \infty)$ into R^1, and $\gamma(q(\cdot), \cdot)$ is an integrable function on $[0, T)$ for $T > 0$ whenever q is a continuous function on $[0, T]$. Also, γ satisfies a Lipschitz condition with respect to its first argument uniformly with respect to its second argument; that is, for every $\xi, \eta \in R^1$,

$$|\gamma(\xi, t) - \gamma(\eta, t)| \leqq P_0 |\xi - \eta| \tag{2.1}$$

where the constant P_0 does not depend upon t. Similarly, each inductor is in general nonlinear, time-varying, and defined by an equation of the form

$$i(t) = \lambda(\phi(t), t)$$

where i and ϕ are respectively the current and flux linkages in the inductor. $\lambda(\cdot, \cdot)$ maps $R^1 \times [0, \infty)$ into R^1, $\lambda(\phi(\cdot), \cdot)$ is an integrable function on $[0, T)$ for $T > 0$ whenever ϕ is a continuous function on $[0, T]$, and, for every $\xi, \eta \in R^1$,

$$|\lambda(\xi, t) - \lambda(\eta, t)| \leqq P_0 |\xi - \eta| \tag{2.2}$$

where the constant P_0 does not depend upon t. Finally, every voltage source v^s and current source i^s is a continuous mapping of $[0, \infty)$ into R^1. This ends Conditions 1.

Now, in accordance with [5; § 2], let $N = \bigcup_{p=1}^{\infty} N_p$ be a partition of N into the finite subnetworks N_p corresponding to a choice of $\mathscr{L}$ and $\mathscr{J}$ for which Conditions 1 are satisfied. Fix upon a particular N_p and index its limb branches by $\nu = 1, \cdots, n$ and its chords by $\mu = n+1, \cdots, n+m$. Let

$$x(t) = [q_1(t), \cdots, q_n(t), \phi_{n+1}(t), \cdots, \phi_{n+m}(t)]^T$$

be the vector of charges $q_\nu(t)$ on its limb capacitors and flux linkages $\phi_\mu(t)$ in its chord inductors. Since $q_\nu(t) = \int_0^t i_\nu(\omega)\, d\omega + q_\nu(0)$ and $\phi_\mu(t) = \int_0^t v_\mu(\omega)\, d\omega + \phi_\mu(0)$, where $i_\nu(t)$ and $v_\mu(t)$ are respectively the current in the νth limb branch and the voltage on the μth chord, it follows that in the present case equations (5.1) and (5.2) of [5] take on the forms

$$\dot{q}_\nu(t) - \sum_{\mu}^{\lambda,\nu} (\pm)\lambda_\mu(\phi_\mu(t), t) = h_\nu(t) + \sum_{k}^{J,\nu} (\pm) j_k(t) + \sum_{\mu}^{I,\nu} (\pm) i_\mu^s(t) \tag{2.3}$$

and

$$\dot{\phi}_\mu(t) - \sum_{\nu}^{\gamma,\mu} (\pm)\gamma_\nu(q_\nu(t), t) = e_\mu(t) + \sum_{k}^{J,\mu} (\pm) w_k(t) + \sum_{\nu}^{V,\mu} (\pm) v_\nu^s(t) \tag{2.4}$$

where the dot denotes the derivatives with respect to time. All the symbols herein are defined in [5]. These two equations taken together can be written in matrix form as follows.

$$\dot{x}(t) = w(x(t), t) + u(t) \tag{2.5}$$

$u(t)$ is the $(n+m) \times 1$ vector corresponding to the right-hand sides of (2.3) and (2.4) and is a known quantity whenever the currents and voltages in $\bigcup_{s=1}^{p-1} N_s$ are known and all the initial conditions at $t = 0$ and all the joint currents and voltages are given. $w(\cdot, t)$ is the matrix defined by equation (5.3) of [5] but having in the present case only capacitor and inductor terms. Thus, we now have a normal form for the determining differential equations. One integration yields

$$x(t) = x(0) + \int_0^t w(x(\omega), \omega)\, d\omega + \int_0^t u(\omega)\, d\omega \tag{2.6}$$

where $x(0)$ is the vector of initial conditions on the limb-capacitor charges and chord-inductor flux linkages in N_p. So, given a dynamic response in $\bigcup_{s=1}^{p-1} N_s$, a dynamic response for N_p can be determined if a vector $x(t)$ can be found that satisfies (2.6).

To this end, define the operator A by the equation

$$(Ax)(t) = x(0) + \int_0^t w(x(\omega), \omega)\, d\omega + \int_0^t u(\omega)\, d\omega. \tag{2.7}$$

In view of (2.6), we seek a fixed point for A, given $x(0)$ and $u(t)$.

For any $T > 0$, let $C[0, T]$ denote the space of continuous functions f on the compact interval $[0, T]$ supplied with the topology induced by the norm

$$\|f\|_C = \max_{0 \leqq t \leqq T} |f(t)|.$$

Furthermore, let $V[0, T]$ be the Banach space of $(n+m) \times 1$ vectors x whose elements x_k are members of $C[0, T]$ and whose norm is

$$\|x\|_V = \sum_{k=1}^{n+m} \|x_k\|_C.$$

By our assumption in Conditions 1 concerning the integrability of $\gamma(q(\cdot), \cdot)$ and $\lambda(\phi(\cdot), \cdot)$, it follows that A is a mapping of $V[0, T]$ into $V[0, T]$ for every T. We shall show that, for any arbitrarily chosen but fixed T, some power of A is a contraction on $V[0, T]$.

For the sake of a more concise notation we set $\beta_\nu(x_\nu(t), t) = \gamma_\nu(q_\nu(t), t)$ for $\nu = 1, \cdots, n$ and $\beta_\mu(x_\mu(t), t) = \lambda_\mu(\phi_\mu(t), t)$ for $\mu = n+1, \cdots, n+m$. In the following the indices k and j are restricted to the integers from 1 to $n+m$. The kth component of $w(x(t), t)$ can be written as

$$\sum_j^k (\pm)\beta_j(x_j(t), t)$$

where the summation is over those columns in $w(\cdot, t)$ that contain nonzero terms in the kth row. It will also be helpful to us to denote the kth component x_k of any vector x by either $(x)_k$ or $[x]_k$.

Assume that the vectors $x, y \in V[0, T]$ satisfy $x(0) = y(0)$. In view of (2.7), we may then write for any positive integer s

$$\begin{aligned} &[A^{s+1}x(t) - A^{s+1}y(t)]_k \\ &\quad = \int_0^t [w(A^s x(\omega), \omega) - w(A^s y(\omega), \omega)]_k\, d\omega \\ &\quad = \int_0^t \sum_j^k (\pm)[\beta_j((A^s x)_j(\omega), \omega) - \beta_j((A^s y)_j(\omega), \omega)]\, d\omega. \end{aligned} \tag{2.8}$$

According to (2.1) and (2.2), there is a constant P_j for every j such that for all $\xi, \eta \in R^1$

$$|\beta_j(\xi, t) - \beta_j(\eta, t)| \leqq P_j |\xi - \eta|. \tag{2.9}$$

We let $P = \max\{P_j : j = 1, \cdots, n+m\}$.

LEMMA 1. *For $x, y \in V[0, T]$, $0 \leqq t \leqq T$, and $s = 1, 2, \cdots$, the magnitude of the k-th component of $(A^s x)(t) - (A^s y)(t)$ is bounded by*

$$\frac{(Pt)^s}{s!}(n+m)^{s-1}\|x - y\|_V. \tag{2.10}$$

Proof. We use an inductive argument. The magnitude of (2.8) is bounded by

$$\int_0^t \sum_j^k |\beta_j((A^s x)_j(\omega), \omega) - \beta_j((A^s y)_j(\omega), \omega)|\, d\omega$$

which by (2.9) is dominated by

$$\int_0^t \sum_j^k P|(A^s x)_j(\omega) - (A^s y)_j(\omega)|\, d\omega.$$

So, if Lemma 1 is true for some s, the last expression is bounded by

$$\int_0^t \sum_j^k P\frac{(P\omega)^s}{s!}(n+m)^{s-1}\|x-y\|_V\, d\omega \leqq \frac{(Pt)^{s+1}}{(s+1)!}(n+m)^s\|x-y\|_V.$$

Hence, Lemma 1 is true for $s+1$ if it is true for s.

It is also true for $s=1$. Indeed, in this case we have

$$\begin{aligned}|\beta_j(x_j(\omega), \omega) - \beta_j(y_j(\omega), \omega)| &\leqq P|x_j(\omega) - y_j(\omega)| \\ &\leqq P\|x_j - y_j\|_C\end{aligned}$$

so that we can write

$$\begin{aligned}|(Ax)_k(t) - (Ay)_k(t)| &\leqq \int_0^t \sum_j^k |\beta_j(x_j(\omega), \omega) - \beta_j(y_j(\omega), \omega)|\, d\omega \\ &\leqq Pt \sum_j^k \|x_j - y_j\|_C \\ &\leqq Pt\|x-y\|_V.\end{aligned}$$

This completes the proof.

LEMMA 2. *If A maps a complete metric space into itself and if A^s is a contraction for some positive integer s, then A has a unique fixed point in that space.*

Proof. If A^s is a contraction, then there is a unique x in the metric space for which $x = A^s x$. Thus, $Ax = A(A^s)x = A^s(Ax)$. By the uniqueness of the fixed point of A^s, $Ax = x$. Therefore, x is a fixed point of A.

To show the uniqueness of the fixed point of A, let $y = Ay$. Then, $y = Ay = A^2 y = \cdots = A^s y$. Therefore, a fixed point of A is also a fixed point of A^s. But, A^s has only one fixed point x. Therefore, $y = x$. Q.E.D.

In view of Lemma 1, we may write

$$\begin{aligned}\|A^s x - A^s y\|_V &= \sum_k \|(A^s x)_k - (A^s y)_k\|_C \\ &= \sum_k \max_{0\leqq t\leqq T} |(A^s x)_k(t) - (A^s y)_k(t)| \\ &\leqq \sum_k \max_{0\leqq t\leqq T} \frac{(Pt)^s}{s!}(n+m)^{s-1}\|x-y\|_V \\ &\leqq \frac{(PT)^s}{s!}(n+m)^s\|x-y\|_V.\end{aligned}$$

We continue to assume that T is any arbitrary but fixed positive number. Since the factorial function tends to infinity faster than the powers of any constant, A^s is a contraction on $V[0, T]$ for some sufficiently large s. So, by Lemma 2, A has a unique fixed point in $V[0, T]$. Since T can be chosen arbitrarily, it follows that (2.6) has a unique continuous solution $x(t)$ on $[0, \infty)$ satisfying the given initial condition $x(0)$.

This argument can be applied recursively to $N_1, N_2, N_3, \cdots$ to obtain the following existence and uniqueness theorem.

THEOREM 1. *Let N be a countably infinite electrical network satisfying Conditions 1. Then, any assignment of the voltages and currents in all the joints as continuous functions for $0 \leqq t < \infty$ (in conformity with the joints' parameters), any assignment of the initial charges at $t = 0$ on all the limb capacitors, and any assignment of the initial flux linkages at $t = 0$ in all the chord inductors uniquely determines under Ohm's law and Kirchhoff's node and loop laws the voltages and currents in all the branches of N as continuous functions for $0 \leqq t < \infty$.*

This result is stronger than Theorems 1 and 2 of [5] in that no restrictions on the constants P_0 in the Lipschitz conditions (2.1) and (2.2) need now be imposed. The meaning of this is discussed in the Introduction. However, our present theorem does not allow resistors or conductors in the limb branches and chords in contrast to our prior theorems.

3. Linear networks. We now round out all our prior results with an existence and uniqueness theorem for linear time-varying infinite networks. When N consists only of linear parameters (actually, this assumption can be relaxed for the joints), the determining equations can be written in normal form even when N contains resistors and conductors so long as certain matrices related to those resistors and conductors are invertible. We now assume that N is a countably infinite electrical network that satisfies

Conditions 2. A full set $\mathscr{L}$ of limbs and a full set $\mathscr{J}$ of joints can be so chosen that every limb branch is either a voltage source, a linear resistor, a linear capacitor, or a series connection of any two or all three of these elements and that every chord is either a current source, a linear conductor, a linear inductor, or a parallel connection of any two or all three of these elements. (As before, the joints may be any combination of linear or nonlinear parameters.) There is no mutual coupling. The voltages and currents in the resistors, conductors, capacitors, and inductors are related by the following four equations respectively.

$$v(t) = r(t)i(t),$$

$$i(t) = g(t)v(t),$$

$$v(t) = \gamma(t)\left[\int_0^t i(\omega)\, d\omega + q(0)\right],$$

$$i(t) = \lambda(t)\left[\int_0^t v(\omega)\, d\omega + \phi(0)\right].$$

Here, r, g, γ, and λ are continuous mappings of $[0, \infty)$ into R^1. Once again, every voltage source v^s and current source i^s is a continuous mapping of $[0, \infty)$ into R^1. This ends Conditions 2.

As in the preceding section, we let $x(t)$ be the vector of limb-capacitor charges $q_\nu(t)$ and chord-inductor flux linkages $\phi_\mu(t)$ in some fixed subnetwork N_p. Then, the determining equations for N_p are

$$\text{(3.1)} \quad \dot{q}_\nu(t) - \sum_\mu^{g,\mu} (\pm) g_\mu(t)\dot{\phi}_\mu(t) - \sum_\mu^{\lambda,\nu} (\pm)\lambda_\mu(t)\phi_\mu(t) = h_\nu(t) + \sum_k^{J,\nu} (\pm) j_k(t) + \sum_\mu^{I,\nu} (\pm) i_\mu^s(t),$$

$$\text{(3.2)} \quad \dot{\phi}_\mu(t) - \sum_\nu^{r,\mu} (\pm) r_\nu(t)\dot{q}_\nu(t) - \sum_\nu^{\gamma,\mu} (\pm)\gamma_\nu(t) q_\nu(t) = e_\mu(t) + \sum_k^{J,\mu} (\pm) w_k(t) + \sum_\nu^{V,\mu} (\pm) v_\nu^s(t).$$

In matrix form this becomes

$$\dot{x}(t)-R(t)\dot{x}(t)-K(t)x(t)=u(t). \tag{3.3}$$

More specifically, $R(t)$ is the matrix obtained by setting to zero the capacitor and inductor terms in the matrix corresponding to equation (5.3) of [5]. Also, $K(t)$ is the matrix obtained by setting to zero the resistor and capacitor terms in the same matrix. $u(t)$ is the vector representing the right-hand sides of (3.1) and (3.2). By Conditions 2, $R(t)$ and $K(t)$ are continuous for $0 \leqq t < \infty$. If $I-R(t)$ is invertible for each $t \geqq 0$, then $[I-R(t)]^{-1}$ is also continuous for $0 \leqq t < \infty$ and we may write

$$\dot{x}(t)=[I-R(t)]^{-1}K(t)x(t)+[I-R(t)]^{-1}u(t). \tag{3.4}$$

(Now, I is the identity matrix.) If $u(t)$ is also continuous on $[0, \infty)$, this differential equation has, as is well known [1; p. 20 and p. 74], a unique continuous solution on $[0, \infty)$ for any given vector $x(0)$. This result applied in turn to $N_1, N_2, \cdots$ yields our last theorem.

THEOREM 2. *Let N be a countably infinite electrical network satisfying Conditions 2. For each N_p assume that for every $t \geqq 0$ the determinant of $I-R(t)$ does not vanish. Then, the conclusion of Theorem* 1 *holds once again.*

4. A final remark. Mutual coupling can be incorporated into the analysis of §§ 2 and 3 in much the same way as that indicated in § 7 of [5].

REFERENCES

[1] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

[2] A. H. ZEMANIAN, *The complete behavior of certain infinite networks under Kirchhoff's node and loop laws*, SIAM J. Appl. Math., 30 (1976), pp. 278–295.

[3] ———, *Infinite electrical networks*, Proc. IEEE, 64 (1976), pp. 6–17.

[4] ———, *The limb analysis of countably infinite electrical networks*, Combinatorial Theory, Ser. B, 24 (1978), pp. 76–93.

[5] ———, *Countably infinite nonlinear time-varying active electrical networks*, this Journal, 10 (1979), pp. 944–960.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0010 $01.00/0

SOME SINGULAR NONOSCILLATORY BOUNDARY VALUE PROBLEMS*

PHILIP HARTMAN†

Abstract. This paper concerns singular boundary value problems involving a nonlinear second order equation, say, $y''=F(t, y, y', \lambda)$ for $-\infty \leqq \alpha < t < \omega \leqq \infty$, in which it is specified that a desired solution does not vanish or has exactly N zeros. Existence is proved by considering a family of solutions $y=y(t, \lambda)$ having $N(\lambda)$ zeros, $0 \leqq N(\lambda) < \infty$, and examining the discontinuities of $N(\lambda)$. The novelty of the paper lies in the simplicity of the method and proofs, and the fact that singularities occur at both $t=\alpha$ and $t=\omega$ (even if one or both are finite). The method is applied to problems arising in the theory of (1) the interreaction of elementary particles and (2) superconductivity.

1. Introduction. This note concerns a singular boundary value problem associated with a nonlinear scalar differential equation

$$(p(t, \lambda)y')' = q(t, \lambda)y + f(t, y, y', \lambda) \tag{1.1}$$

(which may or may not depend on λ) on a t-interval $(-\infty) \leqq \alpha < t < \omega (\leqq \infty)$, and with prescribed boundary behavior at $t=\alpha$, e.g.,

$$y(\alpha+)=0, \tag{1.2}$$

with prescribed boundary behavior at $t=\omega$, e.g.,

$$\lim_{t\to\omega} y(t) \quad \text{exists (finite)} \tag{1.3}$$

or $y(t)$ is bounded or $y(t)$ is exponentially small at $\omega=\infty$, and with a prescribed side condition of the type

$$y(t) \quad \text{has exactly } k \text{ zeros on } \alpha < t < \omega. \tag{1.4}$$

This is the type of problem treated in [4] and our principal arguments below will be similar to those of [4]. They do not however involve the introduction of any auxiliary differential equation $x' = \pi(t, x, \lambda)$ as in [4, § 7] so that the resulting theorems are more general and are easier to apply.

My return to this type of problem was prompted by a question raised by my colleague Professor G. Domokos of the Physics Department concerning the boundary value problem (see [9])

$$x^2 \frac{d}{dx}\left[(1-x)\frac{dy}{dx}\right] = 2y(1-y^2), \tag{1.5}$$

$$y(+0)=0 \quad \text{and} \quad \lim_{x\to 1} y(x) \quad \text{exists (finite)}, \tag{1.6}$$

and $y(x) \neq 0$ on $0<x<1$. This problem arose in considerations of interactions of elementary particles. A similar type of problem

$$y'' + y'/t - \nu^2 y/t^2 = y(y^2-1) \quad \text{for } 0<t<\infty, \quad \nu>0, \tag{1.7}$$

$$y(+0)=0, \qquad y(\infty)=1, \quad \text{and} \quad 0<y(t)<1 \quad \text{for } t>0, \tag{1.8}$$

arose in Abrikosov's [1] discussion of superconductivity. (For other occurrences of (1.7)–(1.8) in physics literature, see [7, pp. 151–152], for references to E. Abrahams

* Received by the editors March 13, 1978.

† Department of Mathematics, Johns Hopkins University, Baltimore, Maryland 21218. This work was supported by the National Science Foundation under Grant MPS75-15733.

and T. Tsuneto, to L. P. Pitaevski, to V. L. Ginzburg and L. D. Landau, and to E. Gross.) Actually, the problems involving (1.5) and (1.7) are of a different nature (and the latter is simpler) because the factors $\pm(1-y^2)$ of y on the right are of different signs. Kametka [7] proved existence and uniqueness for a problem more general than (1.7)–(1.8) (cf. § 4 below), and Iwano [6] gave another existence proof for (1.7)–(1.8).

Section 2 deals with general existence theorems for problems of the type (1.1)–(1.4). The procedure of § 2 is applied in §§ 3 and 3.1 to problems more general than (1.4)–(1.6) and in § 4 to a problem more general than (1.7)–(1.8) (and the problem treated by Kametka). Kametka's uniqueness theorem is also generalized in Theorem 4.2.

2. A general existence theorem. We state a very general existence theorem (Theorem 2.1) with a simple proof. But, in any particular case, there remains the problem of verifying the applicability of the theorem. Following the statement of the theorem, we consider the latter problem, i.e., we discuss various conditions sufficient for the hypotheses of the theorem. In §§ 3 and 4, we illustrate the use of the theorem by applying it to the problem (1.4)–(1.6) and (1.7)–(1.8) and generalizations. We begin by enumerating some hypotheses which may occur in some of the results below.

(H_1) $p(t, \lambda)>0$, $q(t, \lambda)$ are continuous on $(\alpha, \omega)\times[\Lambda_0, \Lambda_N]$ and $f(t, y, y', \lambda)$ is continuous on $(\alpha, \omega)\times R^2\times[\Lambda_0, \Lambda_N]$, where $-\infty\leqq\alpha<\omega\leqq\infty$ and $-\infty<\Lambda_0<\Lambda_N<\infty$. Solutions of (1.1) are uniquely determined by initial conditions.

(H_2) For fixed λ, the linear differential equation

$$(p(t, \lambda)u')' = q(t, \lambda)u \tag{2.1}$$

is nonoscillatory at $t=\omega$ and, in fact, there exists a continuous $b(\lambda)$ such that $\alpha<b(\lambda)<\omega$ and (2.1) is disconjugate on $[b(\lambda), \omega)$. Let $u=u_0(t, \lambda)$ and $u_1(t, \lambda)$ be principal and nonprincipal solutions of (2.1) at ω, positive on $[b(\lambda), \omega)$; cf., e.g., [3, pp. 350–361]. We also make the trivial assumption that $u_1(t, \lambda)$ is continuous in (t, λ), but we do not make the more severe requirement of [4] that the principal solution $u_0(t, \lambda)$ is continuous in (t, λ); cf. [4, § 4].

(H_3) $y(t)=y(t, \lambda)$ is a solution of (1.1), for fixed λ, on $\alpha<t<\omega$ such that $y(t, \lambda)$, $y'(t, \lambda)$ are continuous in (t, λ),

$$|y(t, \lambda)|+|y'(t, \lambda)|\neq 0, \tag{2.2}$$

$$\lim_{t\to\omega} y(t, \lambda)/u_1(t, \lambda) = c_1(\lambda) \quad \text{exists (finite).} \tag{2.3}$$

The following assumption which is rather natural will *not* be needed below.

(H_4) If $c_1(\lambda)=0$ in (2.3), then

$$\lim_{t\to\omega} y(t, \lambda)/u_0(t, \lambda) = c_0(\lambda)\neq 0 \quad \text{exists (finite).} \tag{2.4}$$

When $c_1(\lambda)\neq 0$ in (2.3), then $y(t)=y(t, \lambda)\neq 0$ for t near ω. If, in (H_1), (H_3) and (H_4), $\alpha>-\infty$ and (α, ω) can be replaced by $[\alpha, \omega)$, then $y(t, \lambda)$ has a finite number of zeros. We shall need the following assumption below.

(H_5) $y(t)=y(t, \lambda)$ has a finite number $N(\lambda)$ of zeros on $\alpha<t<\omega$ and there exists a positive integer N such that

$$N(\Lambda_0)+N\leqq N(\Lambda_N). \tag{2.5}$$

Definition. When $c_1(\lambda)=0$ in (2.3), we call λ a *principal value* and $y(t, \lambda)$ a *principal solution* of (1.1). Otherwise we apply the adjective "nonprincipal" to λ and $y(t, \lambda)$.

(H_6) If $\lambda=\lambda_0$ is a nonprincipal solution of (1.1), then $N(\lambda)$ is continuous at $\lambda=\lambda_0$.

(H_7) If $\lambda=\lambda_0$ is a principal value, then

$$N(\lambda_0)\leqq N(\lambda)\leqq N(\lambda_0)+1 \quad \text{for small } |\lambda-\lambda_0|. \tag{2.6}$$

Of course, the first inequality is trivially implied by (2.2).

THEOREM 2.1. (i) *Assume* (H_1)–(H_3) *and* (H_5)–(H_6). *Then there exists at least one principal value* λ_0, $\Lambda_0\leqq\lambda_0<\Lambda_N$, *with* $N(\lambda_0)\leqq N(\Lambda_0)$. (ii) *If, in addition* (H_7) *holds, then for* $k=0,1,\cdots,N-1$, *there exists a principal value* $\lambda=\lambda_k$ *on the interval* $\Lambda_0\leqq\lambda<\Lambda_N$ *with* $N(\lambda_k)=N(\Lambda_0)+k$.

Proof of (i). The existence of a principal value λ_0 is clear, since every discontinuity point of the interger-valued function $N(\lambda)$ is a principal value by (H_6). It remains to show that λ_0 can be chosen so that $N(\lambda_0)\leqq N(\Lambda_0)$. This is obvious if $\lambda=\Lambda_0$ is a discontinuity point of $N(\lambda)$. If not, let

$$\lambda_0=\sup\{\lambda: N(\sigma)\leqq N(\Lambda_0) \text{ for } \Lambda_0\leqq\sigma\leqq\lambda<\Lambda_N\}.$$

Clearly $N(\lambda)$ is not continuous at $\lambda=\lambda_0$ and $N(\lambda_0)\leqq N(\sigma)\leqq N(\Lambda_0)$ for $\sigma(<\lambda_0)$ arbitrarily near λ_0; cf. the sentence following (2.6).

Proof of (ii). We shall show that $N(\lambda_0)=N(\Lambda_0)$. In fact, $N(\lambda_0)\leqq N(\Lambda_0)<N(\Lambda_N)$ implies that $\lambda_0<\Lambda_N$, so that there exist $\lambda(>\lambda_0)$ arbitrarily near to λ_0 such that $N(\lambda)\geqq N(\Lambda_0)+1$. By (2.6), $N(\lambda_0)+1\geqq N(\lambda)\geqq N(\Lambda_0)+1$. Consequently $N(\lambda_0)=N(\Lambda_0)$. This gives the existence of λ_0 with the desired properties. The proofs for the existence $\lambda_1,\cdots,\lambda_{N-1}$ are similar.

We now give sufficient conditions for some of the hypotheses above. In general, we shall assume that $t=\alpha$ is "harmless", say, in the sense that $\alpha>-\infty$ and (α,ω) can be replaced by $[\alpha,\omega)$ or more generally that

(H_8) There exists a continuous $a(\lambda)$ such that $\alpha<a(\lambda)<\omega$ and $y(t)=y(t,\lambda)\neq 0$ on $(\alpha, a(\lambda))$.

PROPOSITION 2.1 ON (H_6). *Assume* (H_1)–(H_3), (H_8) *and that* (2.3) *holds uniformly on* $\Lambda_0\leqq\lambda\leqq\Lambda_N$. *Then* ($H_6$) *holds.*

For the assumption of uniformity in (2.3) implies that $c_1(\lambda)$ is continuous. Also, if $c_1(\lambda_0)\neq 0$, then there exists a t-value $T(\lambda_0)$ such that $y(t,\lambda)\neq 0$ for $t\geqq T(\lambda_0)$ and small $|\lambda-\lambda_0|$. Hence (H_6) follows from (H_8).

(H_9). For small $\varepsilon>0$, there exists a continuous function $b_\varepsilon(\lambda)$ such that $\alpha<b_\varepsilon(\lambda)<\omega$ and that any solution $y=y(t)$ of (1.1) satisfying

$$|y(t)|\leqq\varepsilon u_1(t,\lambda) \quad \text{for } t\geqq b_\varepsilon(\lambda) \tag{2.7}$$

has at most one zero on $b_\varepsilon(\lambda)\leqq t<\omega$.

For example, if there exists a continuous $q_{1\varepsilon}(t,\lambda)$ such that

$$|f(t,y,y',\lambda)|\leqq q_{1\varepsilon}(t,\lambda)|y| \quad \text{for } |y|\leqq\varepsilon u_1(t,\lambda),\ t\geqq b_\varepsilon(\lambda), \tag{2.8}$$

and that

$$(p(t,\lambda)y')'=[q(t,\lambda)-q_{1\varepsilon}(t,\lambda)]y \tag{2.9}$$

is disconjugate on $b_\varepsilon(\lambda)\leqq t<\omega$, then ($H_9$) holds. This is the case if, for example,

$$\int_{b_\varepsilon(\lambda)}^{\omega}|q(t,\lambda)-q_{1\varepsilon}(t,\lambda)|\left(\int_{b_\varepsilon(\lambda)}^{t}ds/p(s,\lambda)\right)dt<1, \tag{2.10}$$

cf., e.g., [3, Thm. 5.1, p. 345], with $m(t)=t-a$.

PROPOSITION 2.2 ON (H_7). *Assume the conditions of Proposition* 2.1 *and hypothesis* (H_9). *Then* (H_7) *holds.*

For if $\lambda = \lambda_0$ is a principal value and (2.3) holds uniformly on $[\Lambda_0, \Lambda_N]$, then one can suppose that (2.7) holds for $y(t) = y(t, \lambda)$ and small $|\lambda - \lambda_0|$.

The condition (2.7) and/or (2.8) plays down the dependence of f on y'. We can modify this in the following way.

(H_9') For small $\varepsilon > 0$, there exists a continuous function $b_\varepsilon(\lambda)$ such that $\alpha < b_\varepsilon(\lambda) < \omega$ and that any solution $y(t)$ of (1.1) satisfying

$$|y(t)| \leqq \varepsilon u_1(t, \lambda), \qquad |y'(t)| \leqq \varepsilon u_1'(t, \lambda) \quad \text{for } t \geqq b_\varepsilon(\lambda) \tag{2.11}$$

has at most one zero on $b_\varepsilon(\lambda) \leqq t < \omega$.

For example, if there exist continuous $p_{1\varepsilon}(t, \lambda)$, $q_{1\varepsilon}(t, \lambda)$ such that (2.11) implies

$$|f(t, y, y', \lambda)| \leqq p_{1\varepsilon}(t, \lambda)|y'| + q_{1\varepsilon}(t, \lambda)|y| \tag{2.12}$$

and that

$$\begin{gathered}(p(t, \lambda)P_\varepsilon(t, \lambda)y')' = [q(t, \lambda) - q_{1\varepsilon}(t, \lambda)]y/P_\varepsilon(t, \lambda), \\ \text{where } P_\varepsilon(t, \lambda) = \exp\left[-\int_{b_\varepsilon(\lambda)}^{t} p_{1\varepsilon}(s, \lambda)\, ds\right],\end{gathered} \tag{2.13}$$

is disconjugate on $b_\varepsilon(\lambda) < t < \omega$, then ($H_9'$) holds.

PROPOSITION 2.2$'$ ON (H_7). *Assume the conditions of Proposition* 2.1, *hypothesis* (H_9'), *that* $u_1'(t, \lambda) > 0$ *on* $b_\varepsilon(\lambda) \leqq t < \omega$, *and that*

$$\lim_{t \to \omega} y_1'(t, \lambda)/u_1'(t, \lambda) = c_1(\lambda) \tag{2.14}$$

uniformly on $[\Lambda_0, \Lambda_N]$. *Then* (H_7) *holds.*

It is often of interest to have Theorem 2.1 (ii) applicable with N arbitrary. It is easy to give sufficient conditions for (2.5) with large N.

PROPOSITION 2.3 ON (2.5) IN (H_5). *Assume* (H_1), (H_2) *and that* $f(t, y, y', \lambda) = f(t, y, \lambda)$ *is independent of* y', *with* $\Lambda_0 \leqq \lambda \leqq \Lambda_N$ *replaced by* $\lambda \geqq \Lambda_0$. *Suppose that there is a closed t-interval* $[a, b] \subset (\alpha, \omega)$ *such that*

$$f(t, y, \lambda)/y \to \infty \quad \text{as } |y| \to \infty \tag{2.15}$$

uniformly for $a \leqq t \leqq b$ *and large* $\lambda > 0$. *Suppose also that there exists a sequence* $\Lambda_0, \Lambda_1, \cdots$ *of positive* λ*-values such that*

$$\sup_{a \leqq t \leqq b} (|y(t, \Lambda_n)| + |y'(t, \Lambda_n)|) \to \infty \quad \text{as } n \to \infty. \tag{2.16}$$

Then

$$N(\Lambda_n) \to \infty \quad \text{as } n \to \infty. \tag{2.17}$$

This is a consequence of [5, Cor. 2.1]. For simplicity, we have stated proposition 2.3 under the assumption that f is independent of y'. Corresponding results follow without this assumption from [5, Cor. 6.1].

We turn to sufficient conditions for hypothesis (H_3).

Remark. Condition (2.2) holds if $y(t, \lambda) \not\equiv 0$ for fixed λ, and the initial value problem (1.1) and

$$y(t_0) = 0, \qquad y'(t_0) = 0, \tag{2.18}$$

has the unique solution $y \equiv 0$ for all t_0, λ. For example, suppose that $f(t, 0, 0, \lambda) = 0$ and that

$$f(t, y, y', \lambda) = O(|y| + |y'|) \quad \text{as } |y| + |y'| \to 0 \tag{2.19}$$

uniformly on t-compacts of (α, ω) for every fixed λ.

Sufficient conditions for the validity of (2.3) [or (2.3) uniformly with respect to λ] can be obtained in some cases from the theory of asymptotic integration. We illustrate this under the condition that

$$y(t, \lambda) = O(u_1(t, \lambda)) \quad \text{as } t \to \omega \tag{2.20}$$

holds for every λ [or holds uniformly with respect to λ]. We might remark that estimates such as (2.20) can sometimes be obtained from simple theorems on asymptotic integration or by the use of Lyapounov functions (as in § 3 below).

PROPOSITION 2.4 ON (2.3) IN (H_3). *Let $u_1(t, \lambda)$ be as in* (H_2) *but let $u = u_0(t, \lambda)$ be the unique principal solution of* (2.1) *defined by*

$$u_0(t, \lambda) = u_1(t, \lambda) \int_t^\omega ds/p(s, \lambda)u_1(s, \lambda) \quad \textit{for } b(\lambda) \leqq t < \omega. \tag{2.21}$$

Assume (2.20) *[uniformly with respect to λ]. For every constant $K > 0$, assume the existence of a $\gamma_K(t, \lambda)$ continuous on $[b(\lambda), \omega) \times [\Lambda_0, \Lambda_N]$ such that*

$$|f(t, y, y', \lambda)| \leqq \gamma_K(t, \lambda)|y| \quad \textit{if } |y| \leqq Ku_1(t, \lambda), \quad b(\lambda) \leqq t < \omega, \tag{2.22}$$

and that

$$\int^\omega u_0(t, \lambda)u_1(s, \lambda)\gamma_K(s, \lambda)\, dt < \infty \tag{2.23}$$

converges [uniformly with respect to λ]. Then (2.3) *and*

$$y'(t, \lambda)/y(t, \lambda) = u_1'(t, \lambda)/u_1(t, \lambda) + o(1/p(t, \lambda)u_0(t, \lambda)u_1(t, \lambda)), \tag{2.24}$$

as $t \to \omega$, hold [uniformly with respect to λ]. Also, if $c_1(\lambda) = 0$ in (2.3), *then* (2.4) *and*

$$y'(t, \lambda)/y(t, \lambda) = u_0'(t, \lambda)/u_0(t, \lambda) + o(1/p(t, \lambda)u_0(t, \lambda)u_1(t, \lambda)), \tag{2.25}$$

as $t \to \omega$, hold [uniformly on λ-compacts where $c_1(\lambda) = 0$].

This can be deduced by the methods of Hartman and Wintner applied to the asymptotic integration of linear second order equations (see [3, pp. 375–380]) after the change of dependent variables $y = zu_1(t, \lambda)$ in (1.1) to obtain the second order equation in z,

$$(pu_1^2z')' = u_1 f(t, u_1 z, (u_1 z)', \lambda)$$

or, equivalently, the first order system for (z, v)

$$z' = v/pu_1^2, \qquad v' = u_1 f(t, u_1 z, (u_1 z)', \lambda),$$

where

$$|f(t, u_1 z, (u_1 z)', \lambda)| \leqq \gamma_K(t, \lambda)u_1(t, \lambda)|z| \quad \text{if } |z| \leqq K,\ b(\lambda) \leqq t < \omega.$$

3. On the problem (1.5)–(1.6). We shall prove the following theorem.

THEOREM 3.1. *For $k = 0, 1, \cdots$,* (1.5) *has a solution $y = y_k(x)$ satisfying*

$$\lambda_k = \lim_{x \to +0} y_k(x)/x^2 > 0 \quad \textit{exists (finite)}, \tag{3.1}$$

$$(-1)^k \lim_{x \to 1} y_k(x) > 0 \quad \textit{exists (finite)}, \tag{3.2}$$

(3.3) $$y_k(x) \quad \textit{has exactly } k \textit{ zeros on } 0<x<1.$$

Note that $y=-y_k(x)$ is also a solution of (1.5).

Make the change of independent variables

(3.4) $$x=1-e^{-t}, \quad \text{so that} \quad dt=dx/(1-x),$$

and (1.5), where $0<x<1$, goes over into

(3.5) $$y''=r(t)y\,g(y), \quad \text{where } 0<t<\infty,$$

(3.6) $$r(t)=2e^t/(e^t-1)^2 \quad \text{and} \quad g(y)=1-y^2.$$

It is easily seen that Theorem 3.1 is a particular case of the following (with $\pm y_{\pm 1}=1$, $\sigma=1$, $C_1=2$, $a=2$, $K=2$, $\gamma=\frac{1}{2}$):

THEOREM 3.2. *Let* $y_1>0>y_{-1}$, $\sigma\geqq 0$, *and* $C_1>1$, a, K, γ *be positive constants satisfying*

(3.7) $$4Ka>1$$

and either

(3.8) $$0<a<2, \quad \textit{or } a=2 \quad \textit{and} \quad \sigma>\tfrac{11}{12}, \quad \textit{or } 2^{1/2}\gamma^{1/2}/y_1>1.$$

Let $r(t)\in C^1(0,\infty)$ *and* $g(y)\in C^0(R^1)$. *Assume that* $r(t)$ *satisfies*

(3.9) $$r'(t)\leqq 0 \quad \textit{for } t>0,$$

(3.10) $$r'/r\sim -2/t \quad \textit{and} \quad r\sim a/t^2, \quad \textit{as } t\to +0,$$

(3.11) $$\int_{+0} t^{-1}|t^2r(t)-a|\,dt<\infty,$$

(3.12) $$\int^{\infty} r(t)t\{\max_{|y|\leqq Ct}|g(y)|\}\,dt<\infty \quad \textit{for all } C>0.$$

Assume that $g(y)$ *satisfies*

(3.13) $$g(0)=1 \quad \textit{and} \quad 0<g(y)\leqq 1 \quad \textit{on } (y_{-1}, y_1),$$

(3.14) $$|g(y)-g(z)|\leqq C_1|y-z| \quad \textit{for } 0\leqq y, z\leqq 1/C_1,$$

(3.15) $$g(y)\geqq \sigma\eta \quad \textit{for } 0<y\leqq(1-\eta)y_1 \textit{ and small } \eta>0,$$

(3.16) $$|g(y)|\leqq C_1|y-y_{\pm 1}| \quad \textit{for } |y-y_{\pm 1}|\leqq 1/C_1,$$

(3.17) $$\int_0^{y_1} sg(s)\,ds=-\int_{y_{-1}}^0 sg(s)\,ds=\gamma/2>0,$$

(3.18) $$-|y|g(y)\geqq K|y-y_{1[\textit{or } -1]}| \quad \textit{for } y>y_1\ [\textit{or } y<y_{-1}].$$

Finally, assume that solutions of (3.5) *are uniquely determined by initial conditions. Then, for* $k=0, 1, \cdots$, (3.5) *has a solution* $y_k(t)$ *satisfying*

(3.19) $$\lambda_k=\lim_{t\to 0} y_k(t)/t^{\mu}>0 \quad \textit{exists (finite)},$$

where $2\mu=1+(1+4a)^{1/2}>2$,

(3.20) $$d_k=(-1)^k \lim_{t\to\infty} y_k(t)>0 \quad \textit{exists (finite)},$$

and $y_k(t)$ *has exactly* k *zeros on* $t>0$.

For generalizations, see § 3.1 below.

Remarks. If the proviso in (3.14) is replaced by $0 \geqq y$, $z \geqq -1/C_1$ and (3.15) is replaced by "$g(y) \geqq \sigma\eta$ *for* $0 > y \geqq (1-\eta)y_{-1}$ *and small* $\eta > 0$," then the assertion remains valid with $\lambda_k < 0$ and $d_k < 0$ in (3.19), (3.20). Condition (3.17) holds, for example, if $y_{-1} = -y_1$ and $g(y)$ is even on $(-y_1, y_1)$. Condition (3.8) is used below only to assure that (3.47) below holds for small $\eta > 0$. The conditions (3.11) and (3.14) are only used in the asymptotic integration of (3.5) at $t = +0$ (i.e., of (3.22) below at $s = \infty$) in parts (a) and (i) and can be replaced by other sets of conditions on $t^2r(t) - a$ and $g(y)$. The following corollary follows by extending the definition of g for $y \leqq 0$ by $g(y) = g(-y)$.

COROLLARY 3.1. *The assertion of Theorem* 3.2 *remains valid for* $k = 0$ *if* $g(y) \in C^0\{y \geqq 0\}$ *satisfies* (3.7)–(3.8) *and* (3.12)–(3.18) *only for* $y \geqq 0$.

Proof of Theorem 3.2. (a) For every λ, (3.5) has a unique solution $y = y(t, \lambda)$ for small $t > 0$; $y(t, \lambda)$ and $y'(t, \lambda)$ are continuous in (t, λ) and satisfy

$$y = \lambda t^{\mu} + o(t^{\mu}), \qquad y' = \lambda\mu t^{\mu-1} + o(t^{\mu-1}), \quad \text{as } t \to +0. \tag{3.21}$$

In order to see this, make the change of independent variables

$$t = e^{-s}, \quad \text{so that } ds = -dt/t,$$

while (3.5) and $0 < t < \infty$ go over into

$$d^2y/ds^2 + dy/ds - ay = H(t, y) \quad \text{and} \quad \infty > s > 0, \tag{3.22}$$

where

$$H(t, y) = (t^2r(t) - a)yg(y) + ay(g(y) - 1).$$

The condition (3.11) is equivalent to

$$\int^{\infty} |t^2r(t) - a|\,|ds| < \infty, \quad \text{where } t = e^{-s}, \quad |ds| = dt/t. \tag{3.23}$$

Since the characteristic numbers of the linear differential operator on the left of (3.22) are $-\mu < 0$ and $-\frac{1}{2} + (\frac{1}{4} + a)^{1/2} > 0$, it follows from results of Perron and Hartman–Wintner that (3.23) and the hypothesis (3.14) imply the existence, for every λ, of a unique solution $y = Y(s; \lambda)$ of (3.22) for large s satisfying

$$e^{\mu s}Y \to \lambda \quad \text{and} \quad e^{\mu s}\,dY/ds \to -\lambda\mu \quad \text{as } s \to \infty, \tag{3.24}$$

uniformly on λ-compacts and, furthermore, $Y(s; \lambda)$ and $dY(s; \lambda)/ds$ are locally uniformly Lipschitz continuous in (s, λ); cf. [3, Chap. X, §§ 8 and 13, and p. 321] for earlier references. Part (i) below can also be modified to give another proof.

In what follows, we let $y(t, \lambda) = Y(s; \lambda)$ and consider only $\lambda \geqq 0$. Thus (3.21) holds uniformly on λ-compacts and $y(t, \lambda)$, $y'(t, \lambda)$ are continuous in (t, λ) for small $t > 0$ and $\lambda \geqq 0$.

(b) *On* (H_1), (H_2), (H_8). In order to apply the results of § 2, we identify (3.5) with (1.1), so that $(\alpha, \omega) = (0, \infty)$, $f = r(t)yg(y)$ is independent of (y', λ), $p(t, \lambda) \equiv 1$ and $q(t, \lambda) \equiv 0$. Hence (H_1) and (H_2) hold with $u_0(t, \lambda) \equiv 1$, $u_1(t, \lambda) \equiv t$, and arbitrary continuous (say, constant) $b(\lambda) > 0$ for $\lambda \geqq 0$. Also, (H_8) holds for suitable (small) $a(\lambda) > 0$, by virtue of the uniformity of (3.21) on λ-compacts.

(c) *On* (H_3), (H_6). Introduce the Lyapunov function

$$E(t, \lambda) = r(t)G(y) + y'^2 \quad \text{with } y = y(t, \lambda), \tag{3.25}$$

and, by (3.13), (3.17), (3.18),

$$G(y) = -2\int_{y_1}^{y} sg(s)\,ds \geqq 0 \quad \text{for all } y. \tag{3.26}$$

In fact,

$$G(y) = 0 \quad \text{if and only if} \quad y = y_{\pm 1}. \tag{3.27}$$

By (3.9) and (3.17),

$$E'(t, \lambda) = r'(t)G(y) \leqq 0. \tag{3.28}$$

Thus, $E(t, \lambda) \sim \gamma r(t)$ as $t \to +0$ and $E(t, \lambda) \geqq 0$ is nonincreasing in t. Hence $y(t, \lambda)$, $y'(t, \lambda)$ exist (and are continuous) for $t > 0$, $\lambda \geqq 0$. Also (2.2) holds for $\lambda > 0$ by virtue of the remark concerning (2.19).

It also follows that $E(t, \lambda)$ is uniformly bounded as $t \to \infty$ on λ-compacts. Consequently,

$$y'(t, \lambda),\ y(t, \lambda)/t = O(1) \quad \text{as } t \to \infty \tag{3.29}$$

uniformly on λ-compacts. Thus $y = y(t, \lambda)$ satisfies a linear equation

$$y'' = r_1(t, \lambda)y, \quad \text{where } r_1(t, \lambda) = r(t)g(y(t, \lambda)), \tag{3.30}$$

and the integral

$$\int^{\infty} t|r_1(t, \lambda)|\,dt < \infty \tag{3.31}$$

converges uniformly on λ-compacts, by (3.12) and (3.29). It follows from standard theorems on asymptotic integration (due to Bôcher [2]) that

$$c_1(\lambda) = \lim_{t\to\infty} y'(t, \lambda) = \lim_{t\to\infty} y(t, \lambda)/t \quad \text{exists (finite)} \tag{3.32}$$

uniformly on λ-compacts, and that

$$c_1(\lambda) = 0,\ \lambda > 0 \quad \Rightarrow \quad \lim_{t\to\infty} y(t, \lambda) = c_0(\lambda) \neq 0 \quad \text{exists (finite);} \tag{3.33}$$

cf., e.g., [3, Chap. XI, § 9]. Consequently (H_3) holds for $\lambda > 0$. Also, by Proposition 2.1, (H_6) holds for $\lambda > 0$.

(d) *On* (H_9), (H_7). The linear second order equation (3.30) is nonoscillatory at $t = \infty$. In fact, the uniformity of (3.31) shows that there exists a continuous $b_1(\lambda) > 0$ such that

$$\int_{b_1(\lambda)}^{\infty} t|r_1(t, \lambda)|\,dt < 1.$$

Hence (3.30) is disconjugate on $[b_1(\lambda), \infty)$; cf. the remark concerning (2.10). Thus (H_9) and, by Proposition 2.2, (H_7) hold on $\lambda > 0$.

Thus Theorem 3.2 follows from Theorem 2.1 if we verify that

$$N(\lambda) = 0 \quad \text{for small } \lambda > 0, \tag{3.34}$$

$$N(\lambda) \to \infty \quad \text{as } \lambda \to \infty; \tag{3.35}$$

that is, (H_5) holds with $N(\Lambda_0) = 0$ for small $\Lambda_0 > 0$ and arbitrary N for large Λ_N. In order to verify (3.34)–(3.35), we need some information about $y'(t, \lambda)$.

(e) *On* $y'(t, \lambda)$. A solution $y(t)$ of (3.5) is convex ($y'' \geqq 0$) on the intervals where $0 \leqq y \leqq y_1$ or $y \leqq y_{-1}$ and concave ($y'' \leqq 0$) where $y_{-1} \leqq y \leqq 0$ or $y \geqq y_1$. Also y' has its local extrema where $y'' = 0$ (i.e., where $y = 0$, $y_{\pm 1}$). For $\lambda > 0$, the convexity of $y(t, \lambda)$ when $0 < y < y_1$ implies that there exists a least positive t-value $t_1 = t_1(\lambda)$ such that

$$0 < y(t, \lambda) < y_1 \quad \text{for } 0 < t < t_1 \text{ and } y(t_1, \lambda) = y_1. \tag{3.36}$$

There exists a *finite* sequence $0 < t_1 < t_2 < \cdots$, $t_j = t_j(\lambda)$, such that

$$y(t_{4j+1}, \lambda) = y(t_{4j+2}, \lambda) = y_1, \qquad y(t_{4j+3}, \lambda) = y'(t_{4j+4}, \lambda) = y_{-1} \tag{3.37}$$

for $j = 0, 1, \cdots$. By (3.25), (3.27), and (3.28),

$$|y'(t_j, \lambda)| \geqq |y'(t_{j+1}, \lambda)|. \tag{3.38}$$

It is easy to see that the positive maxima and negative minima of $y'(t, \lambda)$ occur at the points $t = t_1, t_2, \cdots$. Hence

$$|y'(t, \lambda)| \leqq |y'(t_j, \lambda)| \quad \text{for } t \geqq t_j. \tag{3.39}$$

Also, by the convexity of $y(t, \lambda)$ on $(0, t_1]$,

$$|y'(t, \lambda)| \leqq y'(t_1, \lambda) \quad \text{for } 0 < t < \infty. \tag{3.40}$$

(f) *On* (3.34). By (3.40), $|y_1 - y(t, \lambda)| \leqq y'(t_1, \lambda)(t - t_1)$ for $t \geqq t_1$. An integration of (3.30) over $[t_1, t]$ shows that

$$|y'(t, \lambda) - y'(t_1, \lambda)| \leqq y'(t_1, \lambda) \int_{t_1}^{\infty} h(s, \lambda)\, ds, \tag{3.41}$$

where

$$h(t, \lambda) = r(t) t |y(t, \lambda) g(y(t, \lambda))| / |y(t, \lambda) - y_1|.$$

The ratio $|y|/|y - y_1|$ is bounded when $|y| = |y(t, \lambda)|$ is large, while $|g(y)|/|y - y_1|$ is bounded when $|y - y_1|$ is small, by (3.16). Hence there is a constant $C > 0$ such that

$$h(t, \lambda) \leqq C r(t) t \max_{|y| \leqq Ct} |g(y)| \quad \text{for } t \geqq 1 \text{ and small } \lambda > 0.$$

Since $y(t, \lambda) \to y(t, 0) \equiv 0$ as $\lambda \to 0$ uniformly on t-compacts, it follows that $t_1 = t_1(\lambda) \to \infty$ as $\lambda \to +0$. Also, by the uniformity in (3.32) and $c_1(0) = 0$, (3.41) implies that $0 < y'(t_1, \lambda) \to 0$ as $\lambda \to +0$; cf. (3.52) below. Hence the factor of $y'(t_1, \lambda)$ on the right of (3.41) is small for small $\lambda > 0$, by (3.12). Thus $y'(t, \lambda) \geqq y'(t_1, \lambda)/2 > 0$ for $t \geqq t_1$ and small $\lambda > 0$. Consequently, $y(t, \lambda) \geqq y_1$ for $t \geqq t_1$ and small $\lambda > 0$. This gives (3.34).

(g) *"Strategy" for the proof of* (3.35). In part (i) below, we verify that

$$t_1 = t_1(\lambda) \to 0 \quad \text{as } \lambda \to \infty. \tag{3.42}$$

We fix an $\eta > 0$ so small that $0 < \eta < a$, $0 < \eta < 1$, (3.15) holds, and

$$c_1 \equiv K(a - \eta) > \tfrac{1}{4}; \tag{3.43}$$

cf. (3.7). For $\eta > 0$, there exists a $T = T(\eta) > 0$ such that

$$(a - \eta)/t^2 \leqq r(t) \leqq (a + \eta)/t^2 \quad \text{for } 0 < t \leqq T, \tag{3.44}$$

$$-r'/r^2 \geqq (2 - \eta) t/a \quad \text{for } 0 < t \leqq T. \tag{3.45}$$

Define $\tau = \tau(\eta)$ by

$$\tau(\eta) = \tfrac{1}{2} + (\tfrac{1}{4} + \sigma(a - \eta)\eta)^{1/2} = 1 + \sigma a \eta + O(\eta^2) \quad \text{as } \eta \to 0. \tag{3.46}$$

Condition (3.8) assures that if $\eta > 0$ is fixed sufficiently small, then

(3.47) $$c_2 \equiv \tfrac{1}{2} + (a-\eta)^{1/2} \max\{\gamma^{1/2}/y_1, (1-\eta)\tau(\eta)/(a+\eta)^{1/2}\} - (\tfrac{1}{4}+a+\eta)^{1/2} > 0.$$

Let $N > 1$ be a fixed positive integer. It will be shown that if λ is sufficiently large, then $t_2(\lambda), \cdots, t_N(\lambda)$ exist, so that $N(\lambda) \geqq N/4$ for large λ, hence (3.35) holds. In the course of this proof, we shall show that there are constants $\gamma_1, \gamma_2, \cdots$ (independent of N and λ) such that

(3.48) $$t_j < t_{j+1} \leqq \gamma_j t_j \quad \text{for } j = 1, \cdots, N-1 \text{ and large } \lambda,$$

so that, by (3.42),

$$t_{j+1} \leqq \gamma_1 \gamma_2 \cdots \gamma_j t_1 \leqq T \quad \text{if } \lambda \text{ is large}.$$

(h) *Majorant for* $y'(t, \lambda)$. For $0 \leqq y(t, \lambda) \leqq y_1$, the function $r_1(t, \lambda)$ in (3.30) satisfies $0 \leqq r_1(t, \lambda) \leqq b/t^2$ for some $b < 0$ and, say, $0 < t \leqq 1$; cf. (3.10) and (3.13). The solutions of $u'' = bu/t^2$ are linear combinations of t^τ, $\tau = \frac{1}{2} \pm (\frac{1}{4}+b)^{1/2}$; so that a principal solution at $t = 0$ is $u = t^\nu$, $\nu = \frac{1}{2} + (\frac{1}{4}+b)^{1/2}$. It follows from a comparison theorem of Hartman–Wintner that the principal solution $y = y(t, \lambda)$ of (3.30) satisfies

$$0 < y'/y \leqq u'/u = \nu/t \quad \text{on } 0 < t \leqq t_1^0(\lambda), \quad \nu = \tfrac{1}{2} + (\tfrac{1}{4}+b)^{1/2},$$

where $t_1^0(\lambda) = \min(1, t_1(\lambda))$; cf. [3, pp. 358–359]. In particular,

(3.49) $$0 < ty'(t, \lambda) \leqq \nu y_1 \quad \text{for } \lambda > 0, \quad 0 < t \leqq t_1^0(\lambda) = \min(1, t_1(\lambda)).$$

Note also that if $t_1, \cdots, t_j$ and $\gamma_1, \cdots, \gamma_{j-1}$ exist and $t_1(\lambda) \leqq 1$, then (3.39) implies that

(3.50) $$t_j |y'(t_j, \eta)| \leqq \gamma_1 \gamma_2 \cdots \gamma_{j-1} t_1 y'(t_1, \lambda) \leqq (\lambda_1 \cdots \lambda_{j-1}) \nu y_1.$$

(i) *On* (3.42). By convexity, (3.40) and (3.49), $y'(t, \lambda) \leqq \nu y_1/t_1^0$ for $0 < t \leqq t_1^0$. Thus

$$0 < y(t, \lambda) \leqq 1/(16 C_1 \nu(a+1)) < 1/C_1 \quad \text{for } 0 < t \leqq t_1^0/(16 C_1 \nu y_1 (a+1)).$$

Let $t^* > 0$ be a small number, independent of λ, to be fixed below and let $t_0 = t_0(\lambda) = \min(t^*, t_1^0(\lambda)/(16 C_1 \nu y_1(a+1))), 1/(16 C_1 \nu y_1 (a+1))$.

The proof of (3.42) will depend on a re-examination of the asymptotic integration of (3.22) in part (a) above. It will be shown that $y(t, \lambda) > t^\mu \lambda/2$ for $0 < t \leqq t_0$ and large λ. This implies (3.42), for otherwise we are led to the contradiction that $t_0 \geqq \text{const.} > 0$ and $1/C_1 > y(t_0, \lambda) > t_0^\mu \lambda/2$ hold for some arbitrarily large λ.

Let $2\mu = 1 + (4a+1)^{1/2} > 1$ and $2m = 1 - (4a+1)^{1/2} < 0$, so that $\mu + m = 1$ and $\mu - m = (1+4a)^{1/2} > 1$. Rewrite the left side of (3.22) as $(d/ds + m)(d/ds + \mu)$, and (3.22) as

$$\frac{d}{ds}\left[e^{(m-\mu)s} \frac{d}{ds}(e^{\mu s} y)\right] = e^{ms} H(t, y), \qquad t = e^{-s}.$$

Two quadratures give

$$e^{\mu s} y(e^{-s}, \lambda) = \lambda + \int_s^\infty e^{(\mu-m)v}\left[\int_v^\infty e^{(m-\mu)u} e^{\mu u} H(t, y)\, du\right] dv,$$

$$= \lambda + (\mu - m)^{-1} \int_s^\infty e^{\mu u} H(t, y)[1 - e^{(\mu-m)(s-u)}]\, du, \quad t = e^{-u}.$$

Equivalently, we have

$$t^{-\mu} y(t, \lambda) = \lambda + (\mu - m)^{-1} \int_0^t w^{-\mu-1} H(w, y(w, \lambda))[1 - (w/t)^{\mu-m}]\, dw.$$

For $0<y\leqq y_1$, the last term $ay(g(y)-1)$ of $H(t,y)$ is nonpositive by (3.13), so that $H(t,y)\leqq|t^2r(t)-a|y$ if $y=y(t,\lambda)$ and $0<t\leqq t_1(\lambda)$. Hence

$$t^{-\mu}y(t,\lambda)\leqq\lambda+(\mu-m)^{-1}\int_0^t w^{-\mu}y(w,\lambda)w^{-1}|w^2r(w)-a|\,dw.$$

Thus, by Gronwall's inequality, $0<t^{-\mu}y(t,\lambda)\leqq 2\lambda$ for $0<t\leqq t_0\leqq t^*$ if t^* satisfies

$$\exp\left[(\mu-m)^{-1}\int_0^{t^*} t^{-1}|t^2r(t)-a|\,dt\right]\leqq 2.$$

On the other hand, if $y=y(t,\lambda)$ and $0<t\leqq t_0$, then $|H(t,y)|\leqq|t^2r(t)-a|y+C_1y^2a$. Since $\mu-m>1$,

$$t^{-\mu}y(t,\lambda)\geqq\lambda-2\lambda\int_0^t w^{-1}|w^2r(w)-a|\,dw-C_1a\int_0^t w^{-\mu-1}y^2(w,\lambda)\,dw.$$

By (3.11), it can be supposed that t^* satisfies

$$2\lambda\int_0^{t^*} w^{-1}|w^2r(w)-a|\,dw<\lambda/4.$$

An integration by parts gives

$$C_1a\int_0^t w^{-\mu-1}y^2\,dw\leqq 2C_1a\int_0^t y'yw^{-\mu}\,dw/\mu\leqq 4C_1a\lambda y(t,\lambda),$$

since $0<yw^{-\mu}\leqq 2\lambda$, $y'\geqq 0$, and $\mu>1$. For $0<t\leqq t_0$, we have $4C_1a\lambda y\leqq\lambda/4$ and so, by the last three displays, $t^{-\mu}y(t,\lambda)>\lambda/2$. This completes the proof of (3.42).

(j) *Minorants for* $t_j|y'(t_j,\lambda)|$. Introduce the Lyapunov function

$$F(t,\lambda)=G(y)+y'^2/r(t)\quad\text{with } y=y(t,\lambda)\tag{3.51}$$

and $G(y)$ given by (3.26), so that

$$F'(t,\lambda)=-r'(t)y'^2/r^2(t)\geqq 0.\tag{3.52}$$

It follows from $G(y_{\pm1})=0$ that

$$|y'(t_j,\lambda)/r^{1/2}(t_j)|\leqq|y'(t_{j+1},\lambda)/r^{1/2}(t_{j+1})|,\tag{3.53}$$

$$y'^2(t_j,\lambda)/r(t_j)\geqq G(y(t,\lambda))+y'^2(t,\lambda)/r(t)\quad\text{for } 0<t\leqq t_j.$$

Since $y'^2(t,\lambda)/r(t)\sim\lambda^2\mu^2t^{2\mu}/a\to 0$ as $t\to 0$ by (3.10) and (3.21), the last inequality gives

$$y'^2(t_j,\lambda)/r(t_j)\geqq G(0)=\gamma,\tag{3.54}$$

so that

$$t_j|y'(t_j,\lambda)|\geqq\gamma^{1/2}(a-\eta)^{1/2}\quad\text{if } 0<t_j\leqq T.$$

If $\sigma>0$, we can obtain another minorant. For $0<y\leqq(1-\eta)y_1$ and $0<t\leqq T$, (3.15) and (3.44) give $r(t)g(y)\geqq\sigma(a-\eta)\eta/t^2$. A principal solution at $t=0$ of $u''=\sigma(a-\eta)\eta/t^2$ is $u=t^\tau$, where $\tau=\tau(\eta)$ is defined in (3.46). The comparison theorem of Hartman–Wintner on principal solutions shows that the principal solution $y(t,\lambda)$ of (3.30) satisfies $y'/y\geqq\tau/t$ if $0<t\leqq(1-\eta)y_1$ and $0<t\leqq T$. If $t^*=t^*(\lambda)$ is the t-value, $0<t^*<t_1$, where $y(t,\lambda)=(1-\eta)y_1$ and if $t_1\leqq T$, then

$$y'(t_1,\lambda)\geqq y'(t^*,\lambda)\geqq(1-\eta)\tau y_1/t^*\geqq(1-\eta)\tau y_1/t_1.$$

Hence, by (3.44),

$$y'^2(t_1, \lambda)/r(t_1) \geqq (1-\eta)^2\tau^2 y_1^2/(a+\eta) \quad \text{for large } \lambda.$$

In view of (3.53), this inequality holds if t_1 is replaced by any t_j which exists. Hence, by (3.44), $0 < t_j \leqq T$ implies that

$$t_j|y'(t_j, \lambda)| \geqq c_3 y_1 \quad \text{if } c_3 = (1-\eta)\tau[(a-\eta)/(a+\eta)]^{1/2},$$

so that $c_3 = 1 + \eta(\sigma a - 1 - 1/a) + O(\eta^2)$ as $\eta \to 0$.

Combining this with our first minorant (following (3.54)), we have, if $0 < t_j \leqq T$,

$$c_2 = c_4 + \tfrac{1}{2} - (\tfrac{1}{4} + a + \eta)^{1/2} > 0. \tag{3.56}$$

Note that the inequality $c_2 > 0$ in (3.47) is equivalent to

$$c_2 = c_4 + \tfrac{1}{2} - (\tfrac{1}{4} + a + \eta)^{1/2} > 0. \tag{3.56}$$

(k) *On* t_2. We now show that if λ is large, then $t_2 = t_2(\lambda)$ exists. Introduce the dependent variable $z = y - y_1$, so that (3.5) becomes

$$z'' - [r(t)yg(y)/(y-y_1)]z = 0, \qquad z = y - y_1.$$

If $y \geqq y_1$, then, by (3.18) and (3.44), the coefficient of z satisfies $-r(t)yg(y)/(y-y_1) \geqq K(a-\eta)/t^2$ for $0 < t \leqq T$. By (3.43), a solution of the equation

$$u'' + c_1 u/t^2 = 0, \qquad c_1 = K(a-\eta) > \tfrac{1}{4},$$

vanishing at $t = t_1$ is $u = t^{1/2} \sin[(c_1 - \frac{1}{4})^{1/2} \log(t/t_1)]$, which also has a zero at $t = t_1 \exp[(c_1 - \frac{1}{4})^{-1/2}\pi]$. By the Sturm comparison theorem $z = y - y_1$ has a zero at some t_2, $t_1 < t_2 \leqq \gamma_1 t_1$, where $\gamma_1 = \exp[(c_1 - \frac{1}{4})^{-1/2}\pi]$, provided that $\gamma_1 t_1 \leqq T$, which is the case for large λ.

This argument also shows that if $\lambda > 0$ is large and $(t_1, \cdots, t_{2k+1})$ and $(\gamma_1, \cdots, \gamma_{2k})$ exist, then t_{2k+2} exists and $t_{2k+1} < t_{2k+2} \leqq \gamma_{2k+1} t_{2k+1}$ with $\gamma_{2k+1} = \gamma_1$ for $k = 0, 1, \cdots$.

(l) *On* $t_{3/2}$. Let $t_{3/2} \in (t_1, t_2)$ be the unique t-value where $z = y - y_1$ has its maximum. We wish to find a minorant for $z(t_{3/2})$. If

$$z(t_{3/2}) \geqq 1/C_1 \tag{3.57}$$

does not hold, where C_1 is the constant in (3.16), then

$$(1/4t^2 <) - r(t)yg(y)/(y-y_1) \leqq c_{21}/t^2, \quad \text{where } c_{21} = C_1(y_1 + 1/C_1)(a+\eta)$$

on $[t_1, t_2]$ if $y = y(t, \lambda)$. Thus the differential equation $u'' + c_{21}u/t^2 = 0$ is a Sturm majorant for the differential equation for z. If u is the solution such that $u(t_1) = z(t_1) = 0$ and $u'(t_1) = z'(t_1) = y'(t_1, \lambda)$, then

$$u = At^{1/2} \sin[(c_{21} - \tfrac{1}{4})^{1/2} \log(t/t_1)], \qquad A = t_1^{1/2} y'(t_1, \lambda)(c_{21} - \tfrac{1}{4})^{-1/2}.$$

Since $u = At^{1/2}$ at $t = t_1 \exp[\pi/2(c_{21} - \frac{1}{4})^{1/2}]$, the Sturm comparison theorem implies that

$$z(t_{3/2}) \geqq t_1 y'(t_1, \lambda)(c_{21} - \tfrac{1}{4})^{1/2} \exp[\pi/4(c_{21} - \tfrac{1}{4})^{1/2}].$$

In view of $t_1 y'(t_1, \lambda) \geqq y_1$ by convexity,

$$z(t_{3/2}) \geqq c_{22} > 0 \quad \text{for large } \lambda, \tag{3.58}$$

where c_{22} is the constant (independent of N and λ) given by

$$c_{22} = \min(1/C_1,\ y_1(c_{21} - \tfrac{1}{4})^{1/2} \exp[\pi/4(c_{21} - \tfrac{1}{4})^{1/2}]). \tag{3.59}$$

The same argument shows that if $\lambda > 0$ is large, $t_1, \cdots, t_{2k+2}$ exist, and $t = t_{2k+3/2}$ is the point where $z = y - y_1$ has a maximum or minimum on (t_{2k+1}, t_{2k+2}), then

$$|z(t_{2k+3/2})| \geqq c_{22} > 0 \quad \text{for large } \lambda. \tag{3.60}$$

(m) *On* $t_{5/2}$. We now show that, for sufficiently large λ, $y(t, \lambda)$ has a first zero, say at $t_{5/2} = t_{5/2}(\lambda)$ $(>t_2)$. To this end, we use (3.13) and (3.42) to compare (3.30) with its Sturm minorant $u'' = (a+\eta)u/t^2$ for $0 < t \leqq T$. If $u = t^{1/2}(a_1 t^{\theta} + a_2 t^{-\theta})$, $\theta = (\frac{1}{4} + a + \eta)^{1/2}$, is the solution satisfying the same initial conditions as $y(t, \lambda)$ at $t = t_2$, then the Sturm comparison theorem implies that $y(t, \lambda)$ vanishes (for $t > t_2$) before u does. It is readily verified that

$$-2\theta t^{\theta+1/2} a_1 = -y'(t_2)t_2 - (\theta - \tfrac{1}{2})y_1, \qquad 2\theta t^{-\theta+1/2} a_2 = -y'(t_2)t_2 + (\theta + \tfrac{1}{2})y_1,$$

where $y'(t_2) = y'(t_2, \lambda)$, and that $u = 0$ at $t = (-a_2/a_1)^{1/(2\theta)}$. Thus $t_{5/2}$ exists and

$$t_{5/2} \leqq (-a_2/a_1)^{1/(2\theta)} = t_2[(-y'(t_2)t_2 + (\theta + \tfrac{1}{2})y_1)/(-y'(t_2)t_2 - (\theta - \tfrac{1}{2})y_1]^{1/(2\theta)}$$

or, by (3.50), (3.55) and (3.56),

$$t_2 < t_{5/2} \leqq \gamma_{3/2} t_2 \quad \text{for large } \lambda, \tag{3.61}$$

where $\gamma_{3/2} = [(\gamma_1 \nu + \theta + \frac{1}{2})/c_2]^{1/(2\theta)}$.

The same argument shows that if λ is large and $t_1, \cdots, t_{2k+2}$ and $\gamma_1, \cdots, \gamma_{2k+1}$ exist, then y has a first zero $t = t_{2k+5/2} > t_{2k+2}$ and

$$t_{2k+2} < t_{2k+5/2} \leqq \gamma_{2k+3/2} t_{2k+2}, \tag{3.62}$$

where $\gamma_{2k+3/2} = [(\gamma_1\gamma_2 \cdots \gamma_{2k+1}\nu + \theta + \frac{1}{2})/c_2]^{1/(2\theta)}$.

(n) *On* t_3. It is clear from a convexity argument that t_3 exists when $t_{5/2}$ exists and that

$$t_3 - t_{5/2} \leqq y_{-1}/y'(t_{5/2}, \lambda). \tag{3.63}$$

In order to find a minorant of $|y'(t_{5/2}, \lambda)|$, note that (3.51)–(3.52) give $\gamma + y'^2(t_{5/2}, \lambda)/r(t_{5/2}) \geqq y'^2(t_2, \lambda)/r(t_2)$. Thus, by (3.54),

$$y'^2(t_{5/2}, \lambda)/r(t_{5/2}) \geqq y'^2(t_2, \lambda)/r(t_2) - y'^2(t_1, \lambda)/r(t_1).$$

The right side is

$$F(t_2, \lambda) - F(t_1, \lambda) = -\int_{t_1}^{t_2} r'(t) y'^2(t, \lambda)\, dt/r^2(t).$$

By (3.45) and (3.48),

$$F(t_2, \lambda) - F(t_1, \lambda) \geqq [(2-\eta)t_2/(a\gamma_1)] \int_{t_1}^{t_2} y'^2(t, \lambda)\, dt$$

for large λ. By Schwarz's inequality,

$$\int_{t_1}^{t_2} y'^2(t, \lambda)\, dt \geqq \left(\int_{t_1}^{t_2} |y'(t, \lambda)|\, dt\right)^2 \Big/ (t_2 - t_1) = [2z(t_{3/2})]^2/(t_2 - t_1).$$

By (3.58), for large λ,

$$y'^2(t_{5/2}, \lambda)/r(t_{5/2}) \geqq 4(2-\eta)z^2(t_{3/2})/(a\gamma_1) \geqq 4(2-\eta)c_{22}^2/(a\gamma_1),$$

which implies that

$$t_{5/2}|y'(t_{5/2},\lambda)|\geqq 2[(2-\eta)(a-\eta)/a\gamma_1]^{1/2}c_{22}\equiv 1/c_{23}.$$

Hence, by (3.63), $t_3-t_{5/2}\leqq|y_{-1}|t_{5/2}c_{23}$, so that $t_3\leqq\gamma_2 t_2$ with $\gamma_2=\gamma_{3/2}(1+|y_{-1}|c_{23})$.

The same argument shows that if λ is large, $t_1,\cdots,t_{2k+2}$ and $\gamma_1,\cdots,\gamma_{2k+1}$ exist, as do $t_{2k+5/2}$ and $\gamma_{2k+3/2}$ (cf. (3.62)), then $t_{2k+5/2}|y'(t_{2k+5/2},\lambda)|\geqq 1/c_{2,2k+3}$, where $c_{2,2k+3}=2[(2-\eta)(a-\eta)/(a\gamma_{2k+1})]^{1/2}c_{22}$, and that t_{2k+3} exists and $t_{2k+2}<t_{2k+3}\leqq\gamma_{2k+2}t_{2k+2}$ with $\gamma_{2k+2}=\gamma_{2k+3/2}(1+|y_{\pm1}|c_{2,2k+3})$ and $\pm1=(-1)^{k+1}$.

(o) *Completion of the proof.* The ends of paragraphs (k) and (n) show that the "strategy" in (g) has been carried out. Hence (3.35) holds and Theorem 3.1 follows.

3.1. Generalization of Theorem 3.2 (added September 6, 1978). After this paper was accepted for publication, the manuscript [9] became available to me. It suggests the consideration of the more general equation

$$y''=r(t)yg(t,y). \tag{3.64}$$

THEOREM 3.3. *Let $y_1>0>y_{-1}$, $\sigma\geqq 0$ and $C_1>1$, a, K, γ be positive constants satisfying* (3.7)–(3.8). *Let $r(t)\in C^1(0,\infty)$ satisfy* (3.9)–(3.11). *Let $g(t,y)$, $\partial g(t,y)/\partial t\in C^0([0,\infty)\times R^1)$ and $g(y)=g(0,y)$ satisfy* (3.13), (3.15), (3.17),

$$\int^{\infty} r(t)t\{\max_{|y|\leqq Ct}|g(t,y)|\}\,dt<\infty \quad \textit{for all } C>0, \tag{3.65}$$

$$G(t,y)\equiv\gamma-2\int_0^y sg(t,s)\,ds\geqq 0 \quad \textit{for } t>0,\ -\infty<y<\infty, \tag{3.66}$$

$$\partial g(t,y)/\partial t\leqq 0 \quad \textit{for } t>0,\ -\infty<y<\infty, \tag{3.67}$$

$$\gamma r'(t)-2\int_0^y s(\partial[r(t)g(t,s)]/\partial t)\,ds\leqq 0 \quad \textit{for } t>0,\ -\infty<y<\infty, \tag{3.68}$$

$$|g(t,y)-g(t,z)|\leqq C_1|y-z| \quad \textit{for } 0<t,\ y,\ z\leqq 1/C_1. \tag{3.69}$$

Let $y=u_{\pm1}(t)\in C^0[0,\infty)$ satisfy $y_{-1}\leqq u_{-1}(t)<0<u_1(t)\leqq y_1$, $u_{\pm1}(0)=y_{\pm1}$, $\pm u_{\pm1}(t)$ are nonincreasing;

$$g(t,y)\geqq 0 \quad \textit{for } t\geqq 0 \textit{ and } u_{-1}(t)\leqq y\leqq u_1(t), \tag{3.70}$$

$$g(t,y)\leqq 0 \quad \textit{for } t\geqq 0 \textit{ and } y\geqq u_1(t) \textit{ or } y\leqq u_{-1}(t); \tag{3.71}$$

$\pm u_{\pm1}(t)\in C^2(0,1/C_1]$ and, for $0<t\leqq 1/C_1$, $\pm u''_{\pm1}(t)\geqq 0$ and

$$|g(t,y)|\leqq C_1|y-u_{\pm1}(t)| \quad \textit{for } |y|\leqq 1+\max|y_{\pm1}|, \tag{3.72}$$

$$-|y|g(t,y)\geqq K|y-u_{1[or\,-1]}(t)| \quad \textit{for } y\geqq u_1(t)[\textit{or } y\leqq u_{-1}(t)]. \tag{3.73}$$

Assume also that solutions of (3.64) *are uniquely determined by initial conditions. Then the conclusions of Theorem* 3.2 *hold if* (3.5) *is replaced by* (3.64).

Actually, the differential equations in [9] are of the form (3.64), where

$$g(t,y)=g(y)-v(t). \tag{3.74}$$

For this case, we have

COROLLARY 3.2. *Let $v(t)\in C^1[0,\infty)$,*

$$v(0)=0 \quad \textit{and} \quad 0\leqq v\leqq 1,\ v'\geqq 0,\ (rv)'\leqq 0 \quad \textit{for } t>0. \tag{3.75}$$

Then, in the case (3.74) *of* (3.64) *in Theorem* 3.3, (3.65) *and* (3.69) *can be replaced by* (3.12) *and* (3.14), *while* (3.66)–(3.68) *become redundant.*

Proof of Theorem 3.3. This proof is similar to that of Theorem 3.2. We only indicate the perhaps-not-so-obvious modifications, referring to steps (a), (b), $\cdots$ of the proof of Theorem 3.2 (and not mentioning steps requiring inessential changes).

(a) This goes as above, except that $H(t, y)$ is replaced by

$$\begin{aligned} H(t, y) &= (t^2r(t)-a)yg(t, y)+ay(g(t, y)-1) \\ &= (t^2r(t)-a)yg(t, y)+ay(g(y)-1)+ay(g(t, y)-g(0, y)). \end{aligned} \tag{3.76}$$

We also use (3.69) in place of (3.14) and the fact that $g(t, y)-g(0, y)=O(t)$ as $t\to 0$ uniformly for small $|y|$.

(c) Similar to above, except that (3.25), (3.26) and (3.28) are replaced by (3.66),

$$E(t, \lambda)=r(t)G(t, y)+y'^2 \quad \text{with } y=y(t, \lambda); \tag{3.77}$$

so that, by (3.64) and (3.68),

$$E'(t, \lambda)\leqq 0. \tag{3.78}$$

Also, (3.30) is replaced by

$$y''=r_1(t, \lambda)y, \quad \text{where } r_1(t, \lambda)=r(t)g(t, y(t, \lambda)). \tag{3.79}$$

(e) should be replaced by: a solution $y(t)$ of (3.64) is convex ($y''\geqq 0$) on the t-intervals where $0<y(t)<u_1(t)$ or $y(t)<u_{-1}(t)$ and concave ($y''\leqq 0$) where $u_{-1}(t)< y(t)<0$ or $y(t)>u_1(t)$. Also y' has local extrema where $y=0$ or $y(t)=u_{\pm 1}(t)$. There exists a $t_1=t_1(\lambda)>0$ for $\lambda>0$ such that

$$0<y(t, \lambda)<u_1(t) \quad \text{for } 0<t<t_1 \text{ and } y(t_1, \lambda)=u_1(t_1). \tag{3.80}$$

Hence, there exists a finite sequence $0<t_1<t_2<\cdots$, $t_j=t_j(\lambda)$ and $\lambda=0$, such that

$$\begin{aligned} y(t_k, \lambda)&=u_1(t_k) \quad \text{for } k=4j+1, 4j+2 \quad \text{and} \\ y(t_k, \lambda)&=u_{-1}(t_k) \quad \text{for } k=4j+3, 4j+4. \end{aligned} \tag{3.81}$$

Omit the inequalities (3.38)–(3.40).

(f) It will be shown that, for small $\lambda>0$, $y'(t, \lambda)>0$ for $t>0$. Suppose, if possible, that $y'(t^*, \lambda)=0$ for some small $\lambda>0$, $t_1<t^*$, and $y'(t, \lambda)>0$ on $(0, t^*)$. Note that y is increasing for $0<t<t^*$, convex for $0<t\leqq t_1$ and concave for $t_1\leqq t\leqq t^*$. Also $u_1(t^*)\leqq u_1(t_1)$ since u_1 is nonincreasing. Determine t_* by the equation $y'(t_1, \lambda)(t_*-t_1)+u_1(t_1)= u_1(t^*)$. Then $0<t_*\leqq t_1$, and so

$$y(t, \lambda)-u_1(t^*)\leqq y'(t_1, \lambda)(t-t_*)\leqq y'(t_1, \lambda)t \quad \text{for } t_1\leqq t\leqq t^*. \tag{3.82}$$

Quadratures of (3.64) give

$$y(t, \lambda)=y(t^*, \lambda)+\int_t^{t^*} (s-t)r(s)y(s, \lambda)g(s, y(s, \lambda))\, ds, \tag{3.83}$$

$$y'(t, \lambda)=y'(t_1, \lambda)+\int_{t_1}^{t} r(s)y(s, \lambda)g(s, y(s, \lambda))\, ds. \tag{3.84}$$

Since $y(t, \lambda)\to y(t, 0)=0$ and $y'(t, \lambda)\to 0$ as $\lambda\to 0$ uniformly on t-compacts, it follows that $t_1(\lambda)\to\infty$ as $\lambda\to 0$. From (3.66) and (3.77)–(3.78), there exists a constant $C>0$ such that $|y(t, \lambda)|\leqq Ct+1$ for $t\geqq t_1(\lambda)$ and small $\lambda>0$. Thus, if $0<\varepsilon<\frac{1}{2}$, then (3.65), (3.82) and (3.83) imply that, for sufficiently small $\lambda > 0$,

$$u(t_1) = y(t_1, \lambda) \geqq y(t^*, \lambda) - \varepsilon y(t^*, \lambda).$$

Hence, $y(t^*, \lambda) \leqq u_1(t_1)/(1-\varepsilon) \leqq 2u_1(t_1)$. Thus $t = t^*$ in (3.84) gives, for some constant C,

$$0 \geqq y'(t_1, \lambda)\left[1 - C\int_{t_1}^{\infty} sr(s)\, ds\right].$$

This inequality is impossible for small $\lambda > 0$, by (3.65). This proves (3.34).

(h) As above through (3.49); omit (3.50).

(i) As above, with $H(t, y)$ defined by (3.76). Note that $g(y) = g(0, y)$ satisfies (3.13), so that (3.67) implies that

$$g(t, y) \leqq g(0, y) \leqq 1 \quad \text{for } y_{-1} \leqq y \leqq y_1. \tag{3.85}$$

(j) Replace the four relations (3.51)–(3.54) by the following three:

$$F(t, \lambda) = G(t, y) + y'^2/r(t) \quad \text{with } y = y(t, \lambda), \tag{3.86}$$

$$F'(t, \lambda) = -2\int_0^y s[\partial g(t, s)/\partial t]\, ds - r'y'^2/r^2 \geqq -r'y'^2/r^2 \geqq 0, \tag{3.87}$$

$$F(t, \lambda) \geqq F(+0, \lambda) = \gamma \quad \text{for } t > 0. \tag{3.88}$$

If $t_j = t_j(\lambda) > 0$ exists and is small for large λ, then $y(t_j, \lambda)$ is nearly u_1. Hence $G(t_j, y(t_j, \lambda)) \geqq 0$ is small, by (3.17) and $g(s) = g(0, s)$, so that $F(t_j, \lambda)$; hence $y'^2(t_j, \lambda)/r(t_j)$, is nearly γ. Hence the unnumbered inequality following (3.54) is valid (even though (3.54) need not be). This is the first minorant obtained in (j).

An analogue of the second minorant is obtained as in (j) above. For, if $\sigma > 0$, (3.15) gives $g(t, y) = g(y) + O(t) \geqq \sigma\eta + O(t) \geqq \sigma^*\eta$ for any σ^*, $0 < \sigma^* < \sigma$, and $0 < t \leqq T_1(\eta)$. Thus we obtain analogues of the second and third inequalities following (3.54).

Using (3.77) and (3.78) (instead of (3.86) and (3.88)), the arguments in the last paragraph show that

$$|y'(t_{j+1}, \lambda)| \leqq (1 + \eta^2)|y'(t_j, \lambda)| \tag{3.89}$$

for small t_{j+1}. Thus we obtain an analogue of (3.55) with σ replaced by σ^* in c_3.

(k) Instead of the first differential equation and $y \geqq y_1$, use

$$z'' - [r(t)yg(t, y) - u_1''(t)]z/(y - u_1(t)) = 0, \qquad z = y - u_1(t),$$

and $y \geqq u_1(t)$. Note that $u_1'' \geqq 0$.

(l) Let $t_{3/2} \in (t_1, t_2)$ be the unique t-value where $y(t, \lambda)$ has a maximum. We seek a minorant for $z(t_{3/2})$, where $z(t) = y(t, \lambda) - u_1(t_2)$. Thus z satisfies

$$z'' - [r(t)yg(t, y)/(y - u_1(t_2)]z = 0.$$

If (3.57) does not hold, we obtain an analogue of the display following (3.57) in which $y - y_1$ is replaced by $y(t, \lambda) - u_1(t_2) \geqq y(t, \lambda) - u_1(t)$. The arguments proceed as in (l) above in which we use the solution $u(t)$ of the Sturm majorant $u'' + c_{21}u/t^2 = 0$ determined by the initial conditions $u = z = 0$, $u' = z' = y'(t_2, \lambda)$ at $t = t_2$ (rather than $t = t_1$).

(n) Here, (3.63) holds since $y_{-1} \leqq u_1(t) < 0$. By (3.86) and (3.87),

$$F(t_{5/2}, \lambda) = \gamma + y'^2(t_{5/2}, \lambda)/r(t_{5/2}) \geqq F(t_2, \lambda)$$

and so, by (3.88),

$$y'^2(t_{5/2}, \lambda)/r(t_{5/2}) \geqq F(t_2, \lambda) - F(t_1, \lambda).$$

The argument can now be completed as above by virtue of (3.87).

4. On Abrikosov's problem. Below we state and prove Theorem 4.1 which implies existence for a solution of the boundary value problem

$$y'' + y'/t + (1 - \nu^2/t^2)y = f(t, y)y, \qquad \nu > 0, \tag{4.1}$$

$$\lambda_0 = \lim_{t\to 0} y(t)/t^\nu > 0 \quad \text{exists (finite)}, \tag{4.2}$$

$$0 < y(t) < 1 \quad \text{for } t > 0, \tag{4.3}$$

under the hypotheses that $f(t, y) \in C^0((0, \infty) \times [0, 1])$ satisfies: solutions of (4.1) are uniquely determined by initial conditions,

$$f(t, 0) \leqq \text{const.} < 1, \qquad f(t, 1) \equiv 1, \tag{4.4}$$

$$f(t, y) \geqq 0 \quad \text{for } 0 \leqq y \leqq 1 \text{ and small } t > 0, \tag{4.5}$$

$$f(t, y) \leqq \psi(t) + \phi(s) \quad \text{for small } t > 0,\ 0 \leqq y \leqq s, \tag{4.6}$$

$$|f(t, y) - f(t, z)| \leqq (\psi(t) + \phi(s))|y - z| \quad \text{for small } t > 0,\ 0 \geqq y,\ z \leqq s, \tag{4.7}$$

where $\phi(s)$ is a nondecreasing function,

$$\int_{+0} \psi(t)\, dt < \infty \quad \text{and} \quad \int_{+0} \phi(s) s^{-1}\, ds < \infty. \tag{4.8}$$

In addition, (4.3) implies that

$$y(t) \to 1 \quad \text{as } t \to \infty \tag{4.9}$$

if, for every $\theta \in (0, 1)$, there exists an $\eta = \eta(\theta)$ such that

$$f(t, y) \leqq \eta(\theta) < 1 \quad \text{when } 0 \leqq y \leqq \theta < 1. \tag{4.10}$$

Note that if $f \equiv 0$, (4.1) reduces to Bessel's equation with a principal solution $y = J_\nu(t) \sim t^\nu/(2^\nu \Gamma(1+\nu))$ as $t \to 0$. In Kametka's theorem [7], it is assumed that $f(t, y)$ is independent of t and that $F(y) = (1 - f(t, y))y \in C^2[0, 1]$, $F(0) = F(1) = 0$ and $d^2F/dy^2 < 0$ on $(0, 1)$. The assertion above can be generalized as follows:

THEOREM 4.1. (*Existence*). *In the differential equation*

$$(p(t)y')' + r(t)y' + q(t)y = yf(t, y), \tag{4.11}$$

let $p > 0$, q, $r \in C^0(0, \infty)$ *and* $f(t, y) \in C^0((0, \infty) \times [0, 1])$ *and let solutions be uniquely determined by initial conditions. Assume that*

$$(p(t)u')' + r(t)u' + q(t)u = 0 \tag{4.12}$$

is nonoscillatory at $t = 0$ *and*

$$(p(t)u')' + r(t)u' + [q(t) - f(t, 0)]u = 0 \tag{4.13}$$

is not disconjugate on $(0, \infty)$; (ii) *if* $u = u_0(t)$ *is a principal solution of* (4.12) *at* $t = 0$, *positive for small* $t > 0$, *then*

$$u_0(t) \to 0 \quad \textit{as } t \to \infty \tag{4.14}$$

and, for every $\lambda \geqq 0$, (4.11) *has a solution* $y(t, \lambda)$ *for small* $t>0$ *satisfying*

$$y(t, \lambda)=(\lambda+o(1))u_0(t) \quad \text{as } t \to 0, \tag{4.15}$$

$y(t, \lambda)$ *and* $y'(t, \lambda)$ *are continuous in* (t, λ), *and* $y(t, 0) \equiv 0$; (iii) *on compact subintervals of* $0<t<\infty$,

$$f(t, y) \geqq q(t) \quad \text{for small } 1-y \geqq 0; \tag{4.16}$$

(iv) $f(t, y)$ *satisfies*

$$f(t, y) \geqq 0 \quad \text{for } 0 \leqq y \leqq 1 \text{ and small } t>0. \tag{4.17}$$

Then there exists a $\lambda_0>0$ *such that* $y(t)=y(t, \lambda_0)$ *satisfies* (4.3).

Remark 1. Concerning the assumption (ii), it is of interest to make the following remark: Suppose that (4.11) is of the form

$$y''+(g(t)+a/t)y'+(h(t)-b/t^2)y=f(t, y)y, \tag{4.18}$$

where $g(t), h(t) \in C^0(0, \infty)$ and a, b are constants such that the indicial equation

$$\rho^2+(a-1)\rho-b=0 \tag{4.19}$$

has real roots and that the larger, say $\rho=\nu$, is positive. If the roots of (4.19) are simple and

$$\int_{\theta t}^{t}|g(s)|\,ds \to 0 \quad \text{as } t \to 0 \text{ for } 0<\theta<1, \quad \int_{+0}|h(t)|t\,dt<\infty, \tag{4.20}$$

then the linear equation (4.18) with $f \equiv 0$ has a principal solution $y=u_0(t)$ at $t=0$ satisfying $u_0(t) \sim t^\nu$ and $u_0'(t) \sim \nu t^{\nu-1}$ as $t \to +0$. Also if $f(t, y)$ satisfies (4.6)–(4.8), then hypothesis (ii) holds. In fact, $y(t, \lambda)$ is uniquely determined by (4.15) and, in addition,

$$y'(t, \lambda)=[\lambda+o(1)]u_0'(t) \quad \text{as } t \to 0. \tag{4.21}$$

If $\rho=\nu$ is a double root of (4.19), then the same holds if (4.20) and (4.8) are replaced by

$$\int_{+0}|g(t)|\,dt<\infty \quad \text{and} \quad \int_{+0}|h(t)\log t|t\,dt<\infty$$

$$\int_{+0}\psi(t)|\log t|t\,dt<\infty \quad \text{and} \quad \int_{+0}\phi(s)|\log s|s^{-1}\,ds<\infty.$$

See Prevatt [8] or, under the condition $\int_{+0}|g|\,dt<\infty$ in (4.20), [3, pp. 304–314].

Remark 2. The condition (iii) can be replaced by $f(t, 1)=q(t)$ for $t>0$, in which case $y(t) \equiv 1$ is a solution of (4.11).

In many cases, $y(t)=y(t, \lambda_0)$ satisfies (4.9) as well as (4.3). Sufficient conditions for this are given by

PROPOSITION 4.1. *Let* $p>0$, $q, r \in C^0(0, \infty)$ *and* $f \in C^0((0, \infty)\times[0, 1])$ *and* $y(t)$ *a solution of* (4.11) *satisfying* $0<y(t)<1$ *for large* t. *Then* (4.9) *holds provided that* $F(t, \theta)=\max\{f(t, y): 0 \leqq y \leqq \theta<1\}$ *is such that*

$$F(t, \theta) \leqq q(t) \quad \text{for large } t \tag{4.22}$$

and the linear differential equation

$$(p(t)y')'+r(t)y'+[q(t)-F(t, \theta)]y=0 \tag{4.23}$$

is oscillatory at $t=\infty$ *for every* $\theta \in (0, 1)$.

Kametka's [7] uniqueness Theorem 7 is contained in part (β) of the following theorem (and Remark 1 and Proposition 4.1 above).

THEOREM 4.2 (*Uniqueness*). *Let* $p>0$, $q, r \in C^0(0,\infty)$ *and* $f \in C^0((0,\infty)\times\{y>0\})$. *Let* $y(t)$, $z(t)$ *be solutions of* (4.11) *for* $t>0$.

(α) *If* $yf(t, y)-q(t)y$ *is strictly increasing in* y (*for fixed* t) *and* $y(t)-z(t)\to 0$ *as* $t\to 0$ *and* $t\to\infty$, *then* $y(t)\equiv z(t)$.

(β) *If* $f(t, y)$ *is strictly increasing in* y (*for fixed* t), $y(t)>z(t)$ *for small* $t>0$, *and*

$$\limsup p(t)E(t)[y'(t)z(t)-y(t)z'(t)]\geqq 0, \qquad t\to 0, \tag{4.24}$$

where

$$E(t)=\exp\int_1^t [r(s)/p(s)]\,ds,$$

then $y(t)/z(t)$ *is an increasing function of* $t>0$ (*so that* $y/z\to 1$ *as* $t\to\infty$ *cannot hold*).

Proof of Theorem 4.1. We use the arguments in the proof of Theorem 2.1, but we modify the concept of "principal and nonprincipal" to fit the present situation.

(a) *Extension of* f. We extend the definition of $f(t, y)$, or rather $yf(t, y)$, as follows:

$$y[f(t, y)-q(t)]=0 \quad \text{and} \quad y[f(t, y)-q(t)]=f(t, 1)-q(t) \tag{4.25}$$

for $y<0$ and $y>1$, so that $yf(t, y)$ is continuous for $t>0$ and all y. Also $y(t, \lambda)$ exists for all $t>0$ as a solution of (4.11).

(b) *Principal values.* A λ-value $\lambda_0>0$ is called principal if $0<y(t, \lambda_0)<1$ for all $t>0$ and, correspondingly, $y(t, \lambda_0)$ is called a principal solution. If $\lambda_0>0$ is nonprincipal, then either $y(t_0, \lambda_0)=0$ or $y(t_0, \lambda_0)=1$ for some $t_0>0$.

(c) *On* $y(t_0, \lambda_0)=0$. If $y(t_0, \lambda_0)=0$ for some positive t_0 and λ_0, then $y'(t_0, \lambda_0)<0$, by the uniqueness of the solution $y\equiv 0$ of (4.11). Thus $y=y(t, \lambda_0)$ satisfies

$$(p(t)y')'+r(t)y'=0$$

for $t\geqq t_0$, by (4.25), and has no zeros for $t>t_0$. Thus, if $N(\lambda)$ is the number of zeros of $y(t, \lambda)$ on $t>0$, then $0\leqq N(\lambda)\leqq 1$ and $N(\lambda)$ is continuous at $\lambda=\lambda_0$ if $N(\lambda_0)=1$.

(d) *On* $y(t_0, \lambda_0)=1$. If, for some $\lambda_0>0$, there is at least $t_0>0$ such that $y(t_0, \lambda_0)=1$, then for small $t_0-t>0$,

$$(p(t)y')'+r(t)y'=y[f(t, y)-q(t)]\geqq 0$$

by condition (iii). A convexity argument implies therefore that $y'(t_0, \lambda)>0$. Hence, for $t\geqq t_0$, $y=y(t, \lambda)$ satisfies

$$(p(t)y')'+r(t)y'=f(t, 1)-q(t)\geqq 0 \quad \text{for } t\geqq t_0,$$

by (4.25), and hence $y'(t, \lambda)>0$ for $t>t_0$. Consequently $N(\lambda_0)=0$ and $N(\lambda)$ is continuous at $\lambda=\lambda_0$.

Thus $N(\lambda)$ is continuous at $\lambda=\lambda_0$ if λ_0 is a nonprincipal solution. By the proof of Theorem 2.1, it is clear that Theorem 4.1 follows if we verify that

$$N(\lambda)=1 \quad \text{for small } \lambda>0, \tag{4.26}$$

$$N(\lambda)=0 \quad \text{for large } \lambda>0. \tag{4.27}$$

(e) *On* (4.26). By assumption (i), (4.13) has a solution ($\not\equiv 0$) with a pair of zeros, say $t=t_1, t_2$ with $0<t_1<t_2<\infty$. It follows that if $\theta>0$ is sufficiently small, then

$$(p(t)u')'+r(t)u'+[q(t)-f(t, 0)-\theta]u=0$$

has a solution ($\not\equiv 0$) with a pair of zeros on $(t_1/2, 2t_2)$. By continuity, $y(t, \lambda) \to 0$ as $\lambda \to 0$ uniformly on t-compacts of $0 < t < \infty$. In particular, $f(t, y(t, \lambda)) \leqq f(t, 0) + \theta$ on $[t_1/2, 2t_2]$ for small λ. If we write (4.11) as

$$(4.28) \qquad (p(t)y')' + r(t)y' + [q(t) - f(t, y)]y = 0,$$

it follows from Sturm's comparison theorem, that $y(t, \lambda)$ has a zero on $(t_1/2, 2t_2)$ for small $\lambda > 0$.

(f) *On* (4.27). Let $t_1 > 0$ be so small that $u_0(t) > 0$ on $(0, t_1]$ and, by (iv), $f(t, y) \geqq 0$ for $0 \leqq y \leqq 1$, $0 < t \leqq t_1$. If $0 < y(t, \lambda) \leqq 1$ on $0 < t \leqq t_1$, then (4.28) and the comparison theorem of Hartman and Wintner (cf. [3, Cor. 6.5, pp. 358–359]) imply that $y'(t, \lambda)/y(t, \lambda) \geqq u_0'(t)/u_0(t)$ for $0 < t \leqq t_1$, and hence that $y(t, \lambda) \geqq u_0(t)y(s, \lambda)/u_0(s)$ for $0 < s < t \leqq t_1$. If $s \to 0$, (4.15) gives $y(t, \lambda) \geqq \lambda u_0(t)$ for $0 < t \leqq t_1$. But then $y(t_0, \lambda) = 1$ for large λ and some $0 < t_0 \leqq t_1$, so that (4.27) holds. This completes the proof of Theorem 4.1.

Proof of Proposition 4.1. Let $0 < y(t) < 1$ for large t. Since the equation (4.23) is oscillatory at $t = \infty$ and $y(t)$ has no zeros for large t, it follows that lim sup $y(t) = 1$ as $t \to \infty$. If (4.9) fails to hold, then there is a θ such that $0 \leqq \liminf y(t) < \theta < 1$ as $t \to \infty$. It follows that there exist arbitrarily large values of $t = t_0$ such that $y(t_0) < \theta$ and $y'(t_0) < 0$. Hence a convexity argument employing (4.22) and (4.23) shows that $y(t) < y(t_0) < \theta$ for $t > t_0$. This is a contradiction, and so (4.9) holds.

Proof of Theorem 4.2. *On* (α). The proof follows usual maximum principal arguments and will be omitted; cf. [3, Ex. 4.6, p. 427 and p. 574].

On (β). Rewrite (4.11) in the form

$$(P(t)y')' + Q(t)y = G(t, y)y,$$

where $P = pE$, $Q = qE$ and $G = fE$. Thus

$$[P(y'z - yz')]' = yz[G(t, y) - G(t, z)]$$

so that, by (4.24) and $G(t, y(t)) - G(t, z(t)) > 0$ for small $t > 0$,

$$P(y'z - yz') \geqq \int_0^t y(s)z(s)[G(s, y(s)) - G(s, z(s))]\, ds$$

or equivalently

$$(y/z)' \geqq [P(t)z^2(t)]^{-1} \int_0^t y(s)z(s)[G(s, y) - G(s, z)]\, ds.$$

Thus $(y/z)' > 0$ for small $t > 0$ and, in fact, for all $t > 0$.

REFERENCES

[1] A. A. Abrikosov, *On the magnetic properties of superconductors of the second group*, Soviet Physics JETP, 5 (1957), pp. 1174–1182.

[2] M. Bôcher, *On regular singular points of linear differential equations whose coefficients are not necessarily analytic*, Trans. Amer. Math. Soc., 1 (1900), pp. 40–53.

[3] P. Hartman, *Ordinary Differential Equations*, Hartman, Baltimore, 1973.

[4] ———, *Boundary value problems for second order, ordinary differential equations involving a parameter*, J. Differential Equations, 12 (1972), pp. 194–212.

[5] ———, *On boundary value problems for superlinear second order differential equations*, Ibid., 26 (1977), pp. 37–52.

[6] M. IWANO, *Applications of Nagumo–Hukuhara theory on the boundary value problems for nonlinear ordinary differential equations to Abrikosov problem and Falkner–Skan problem*, Ann. Mat. Pura Appl., 113 (1977), pp. 303–392.
[7] Y. KAMETKA, *On a nonlinear Bessel equation*, Publ. Res. Inst. Math. Sci., 8 (1972/3), pp. 151–200.
[8] T. W. PREVATT, *Applications of exponential dichotomies to asymptotic integration and the spectral theory of ordinary differential operators*, J. Differential Equations, 17 (1975), pp. 444–460.
[9] G. DOMOKOS AND B. POMORIŠAC, *A statistical model of inclusive distributions*, preprint, June, 1978.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0011 $01.00/0

ON THE PASSAGE THROUGH RESONANCE*

JAN SANDERS†

Abstract. The phenomenon of "passage through resonance" is studied from the asymptotic point of view. Using averaging techniques we'll describe this process and prove the validity of the approximations obtained, on their natural time-scale. The theory is then applied to a model problem, proposed by Kevorkian (1974).

1. Introduction. The aim of this paper is to study the phenomenon of "passage through resonance" from the asymptotic point of view. Using averaging techniques we'll describe this process and prove the validity of the approximations obtained, on their natural time-scale.

The organization of the paper is as follows:

Section 2: The differential equations under consideration here are introduced, together with a formal theory of averaging.

Section 3: The proof of validity of this method (i.e. averaging) is given.

Section 4: The behavior of the solution near the so-called resonance manifold is considered here.

Section 5: Since there remains a gap between the two regions for which estimates have been obtained, a bridge between the results of the two preceding sections is established.

Section 6: The inner-outer vector field is introduced.

Section 7: The final estimates for the composite expansion are given. This section concludes the general theory.

Section 8: The theory can be immediately applied to a model problem, proposed by Kevorkian (1974). This problem is stated and as much information as possible without numerical analysis is given here.

The reader is warned that the terms inner and outer do not refer to the time-variable, but to the space-variables. Kevorkian's (1974) paper may provide the reader with some of the intuition behind the theory developed here.

2. The differential equations and formal averaging. (If the reader does not have a nontrivial example in mind of the problem posed in this section, he is advised to read § 8 first, where he will find one.)

Consider the following local representation of a one-parameter family of C^∞-vector fields, with "small" parameter $\varepsilon \in (0, \varepsilon_0]$, on $T^m \times M^n \times P^r$, where T^m is the m-torus, M^n and P^r are m- and r-dimensional manifolds respectively. P^r is serving as a kind of control space, in the sense that the vector field restricted to P^r is a vector field on P^r (it forms an independent subsystem).

$$\begin{aligned}
&\dot\theta = X(\mathfrak{m}, u)+\varepsilon^2 R_1(\theta, \mathfrak{m}, u) && \theta \in \mathbb{R}^m\\
(2.1)\quad &\dot{\mathfrak{m}} = \varepsilon \sum_{i=0}^{m} \mathfrak{X}_0^i(\theta_i, \mathfrak{m}, u)+\varepsilon^2 \sum_{i=0}^{m} \mathfrak{X}_1^i(\theta_i, \mathfrak{m}, u)+\varepsilon^3 R_2(\theta, \mathfrak{m}, u) && \mathfrak{m} \in \mathbb{R}^n\\
&\dot u = \varepsilon W(u) && u \in \mathbb{R}^r.
\end{aligned}$$

$\mathfrak{X}_j^0$, $j=0, 1$, is understood to be of the form $\mathfrak{X}_j^0 = \mathfrak{X}_j^0(\mathfrak{m}, u)$, The dot means differentiation with respect to a parameter, called time.

---

* Received by the editors September 26, 1977 and in revised form September 10, 1978.

† Mathematisch Institut, Rijksuniversiteit Utrecht, Netherlands.

We refer to the θ_i's, $i=1,\cdots,m$, as the angular variables. The $\mathfrak{X}^i_j(\cdot,\mathfrak{m},u)\colon S^1\to\mathbb{R}^n$, $i=1,\cdots,m$, $j=0,1$, are restricted to be finite Fourier-series, with mean value zero. This because of their nice product rules.

In the sequel we shall need to know how to handle the combination tones $\theta_i\pm\theta_j$, $i,j=1,\cdots,m$, and therefore we add their differential equations to (2.1). We will refer to this extended equation as (2.1)*. This may change m into some $m^*>m$, but otherwise doesn't change the form of the equations. Take a compact domain $D\subset\mathbb{R}^{n+r}$ and let $(\theta_0,\mathfrak{m}_0,u_0)\in T^{m^*}\times D$ be the initial values for (2.1)*. Next we assume a splitting $\mathbb{R}^k\oplus\mathbb{R}^l$ of $\mathbb{R}^n$, such that for $\mathfrak{m}=(w,x)\in\mathbb{R}^k\oplus\mathbb{R}^l$, one has $X(\mathfrak{m},u)=X(x,u)$, and, furthermore, the second order averaged equation (to be defined in § 3) restricted to $T^{m^*}\times\mathbb{R}^l\times\mathbb{R}^r$ is an independent subsystem.

We shall restrict ourselves to the case $l=1$, since this keeps the discussion as simple as possible. The assumption is satisfied in the model problem. This simplifies matters because we don't have to choose new coordinates when studying the flow in the neighborhood of the resonance manifold (cf. § 4). Writing

$$\mathfrak{X}^i_j(\theta_i,\mathfrak{m},u)=\begin{pmatrix}Y^i_j(\theta_i,\mathfrak{m},u)\\ Z^i_j(\theta_i,\mathfrak{m},u)\end{pmatrix}\qquad \begin{matrix}i=1,\cdots,m\\ j=0,1\end{matrix}\tag{2.2}$$

one has the equations:

$$\begin{aligned}
\dot\theta&=X(x,u)+\varepsilon^2R_1(\theta,w,x,u)\\
\dot w&=\varepsilon Y^0_0(w,x,u)+\varepsilon\sum_{i=1}^m Y^i_0(\theta_i,w,x,u)+\varepsilon^2R_2(\theta,w,x,u)\\
\dot x&=\varepsilon Z^0_0(x,u)+\varepsilon\sum_{i=1}^m Z^i_0(\theta_i,w,x,u)+\varepsilon^2\sum_{i=0}^m Z^i_1(\theta_i,w,x,u)+\varepsilon^3R_3(\theta,w,x,u)\\
\dot u&=\varepsilon W(u).
\end{aligned}\tag{2.3}$$

What one would like to do now, in order to simplify this system, is to average over the angular variables and forget the remainder terms R_i, $i=1,2,3$. This can be done as follows: define a domain $D_{\delta(\varepsilon)}\subset\mathbb{R}^{n+r}$ (where δ is some order function of ε) and a map $\Phi_\varepsilon\colon T^{m^*}\times D_{\delta(\varepsilon)}\to T^{m^*}\times D$, such that there is vector field on $T^{m^*}\times D_{\delta(\varepsilon)}$ of the form (2.1)*, but with $\mathfrak{X}^i_j=0$ for $i=1,\cdots,m$; $j=0,1$, which is Φ_ε-related to (2.3).

(Two vector fields v and $\tilde v$ are ϕ-related iff $v(\phi(x))=d\phi(x)\tilde v(x)$; if ϕ is invertible, $\tilde v$ is merely the pull-back of v along ϕ.)

Usually one finds Φ_ε by substituting some formal development in the definition of Φ_ε-relatedness and equating terms with the same order of magnitude in ε. In general this procedure does not lead to a unique result (neither in the map, nor in the vector field), which probably accounts for the vast number of different theories on this subject.

Having obtained this vector field, one forgets about the remainder terms. The resulting vector field is usually called the averaged vector field.

Trying to define Φ_ε everywhere on $T^{m^*}+\mathbb{R}^{n+r}$ is likely to produce trouble, namely the small-divisor problem, which is also called resonance. It follows from the formal computations that resonance occurs iff one of the X_i becomes zero. Therefore we introduce some concepts in order to define a nice domain $D_{\delta(\varepsilon)}$, before presenting the formulae for Φ_ε and the averaged vector field. First we assume X to have the following structure:

$$X_i(x,u)=X^{\#}_i(x)\Omega_i(u).\tag{2.4}$$

Call

$$\mathcal{N}_i = \{(\mathfrak{m}, u) \in \mathbb{R}^{n+r} \mid X_i(x, u) = 0\} \tag{2.5}$$

the "*i-th resonance manifold*" (a manifold only in the sense that it is the solution of some equations; it may not be locally Euclidean everywhere). Condition (2.4) assures us that the resonance manifolds are not slowly moving, which makes matters simpler.

Define $\mathcal{M}_i(\delta(\varepsilon))$ by

$$\mathcal{M}_i(\delta(\varepsilon)) = \left\{ (\mathfrak{m}, u) \in \mathbb{R}^{n+r} \,\middle|\, \frac{\delta(\varepsilon)}{|X_i(x, u)|} = o(1) \text{ for } \varepsilon \downarrow 0 \right\} \tag{2.6}$$

where δ is some order function of ε with $\delta(\varepsilon) = o(1)$ and $\varepsilon/\delta^2(\varepsilon) = o(1)$ as $\varepsilon \downarrow 0$ (δ lies somewhere between $\varepsilon^{1/2}$ and 1).

Now let D_δ be the intersection of the local coordinate domain and $\bigcap_{i=1}^{m^*} \mathcal{M}_i(\delta(\varepsilon))$. D_δ may be disconnected.

In the sequel it is assumed that only the zeros of $X_i^{\#}$ are of importance and that the zeros of the Ω_i, if they exist at all, cannot become important due to the initial conditions and the time-scale of interest to us.

We introduce for $(\phi, y, z, u) \in T^{m^*} \times D_\delta$ the map

$$\Phi_\varepsilon : \begin{pmatrix} \phi \\ y \\ z \\ u \end{pmatrix} \mapsto \begin{pmatrix} \phi + \varepsilon \sum_{i=1}^{m} u_{(0)}^i(\phi_i, y, z, u) \\ y + \varepsilon \sum_{i=1}^{m} v_{(0)}^i(\phi_i, y, z, u) \\ z + \varepsilon \sum_{i=1}^{m} w_{(0)}^i(\phi_i, y, z, u) + \varepsilon^2 \sum_{i=1}^{m^*} w_{(1)}^i(\phi_i, y, z, u) \\ u \end{pmatrix} \tag{2.7}$$

The reason for the appearance of the $w_{(1)}^i$'s is the following: In the $\delta(\varepsilon)$-neighborhood of $\mathcal{N}_i$ the second variable y is slowly varying with respect to the natural time-scale $1/\sqrt{\varepsilon}$, and therefore can be approximated by a constant.

The only remaining problem is to approximate the first and third component of the solution in this neighborhood.

To do this we need sufficient accuracy of the approximate initial values provided by the outer expansion. It follows from the estimates in the next section that one needs to know the third variable very accurately in order to get a reasonable approximation to the first (here one needs the splitting assumption). This cannot be done without the $w_{(1)}^i$'s, as the reader may want to verify.

Then, differentiating the relations

$$\begin{aligned} \theta &= \phi + \varepsilon \sum_i u_{(0)}^i(\phi_i, y, z, u) \\ w &= y + \varepsilon \sum_i v_{(0)}^i(\phi_i, y, z, u) \quad \text{etc.} \end{aligned} \tag{2.8}$$

we obtain the following equations:

(2.9)
$$\begin{aligned}
\dot{\phi} &= X(z,u)+\frac{\varepsilon^2}{X^3(z,u)}R_1(\phi,y,z,u;\varepsilon) && \phi\in\mathbb{R}^{m^*}\\
\dot{y} &= \varepsilon Y^0_{(0)}(y,z,u)+\frac{\varepsilon^2}{X^2(z,u)}R_2(\phi,y,z,u;\varepsilon) && y\in\mathbb{R}^k\\
\dot{z} &= \varepsilon Z^0_{(0)}(z,u)+\varepsilon^2\bar{Z}^0_{(1)}(z,u)+\frac{\varepsilon^3}{X^4(z,u)}R_3(\phi,y,z,u;\varepsilon) && z\in\mathbb{R}^l\\
\dot{u} &= W(u) && u\in\mathbb{R}^r
\end{aligned}$$

where the R_i, $i=1,\cdots,3$, are uniformly bounded in all variables on $T^{m^*}\times D_\delta\times(0,\varepsilon_0]$ and

(2.10)
$$\frac{1}{X^n(z,u)}=\sum_{i=1}^{m^*}\frac{1}{|X_i(z,u)|^n}.$$

Here we have incorporated the assumptions on the averaged equations which characterized the splitting of $\mathbb{R}^n$. Introducing the following notation:

(2.11)
$$\mathfrak{v}^i_{(j)}(\phi_i,y,z,u)=\begin{pmatrix} v^i_{(j)}(\phi_i,y,z,u)\\ w^i_{(j)}(\phi_i,y,z,u)\end{pmatrix}$$
$$\frac{\partial}{\partial y}=\left(\frac{\partial}{\partial y_1},\cdots,\frac{\partial}{\partial y_k}\right),\qquad \frac{\partial}{\partial z}=\left(\frac{\partial}{\partial z_1},\cdots,\frac{\partial}{\partial z_l}\right),$$
$$\frac{\partial}{\partial u}=\left(\frac{\partial}{\partial u_1},\cdots,\frac{\partial}{\partial u_r}\right),\qquad \frac{\partial}{\partial \mathfrak{m}}=\left(\frac{\partial}{\partial y_1},\cdots,\frac{\partial}{\partial y_k},\frac{\partial}{\partial z_1},\cdots,\frac{\partial}{\partial z_l}\right)$$

we see that (2.9) holds when the following choices have been made:

(2.12)
$$\mathfrak{v}^i_{(0)}(\phi_1,y,z,u)=\frac{1}{X_i(z,u)}\int^{\phi_i}\mathfrak{X}^i_0(\psi_i,y,z,u)\,d\psi_i$$
$$u^i_{(0)}(\phi_i,y,z,u)=\frac{1}{X_i(z,u)}\int^{\phi_i}w^i_{(0)}(\psi_i,y,z,u)\frac{\partial}{\partial z}X(z,u)\,d\psi_i$$
$$\mathfrak{X}^0_{(0)}=\mathfrak{X}^0_0,\qquad \mathfrak{X}^0_{(0)}=\begin{pmatrix}Y^0_{(0)}\\ Z^0_{(0)}\end{pmatrix}$$
$$\bar{V}^{0,i}(y,z,u)=\int_{S^1}\mathfrak{v}^i_{(0)}(\phi_i,y,z,u)\frac{\nabla}{\partial\mathfrak{m}}Z^i_{(0)}(\phi_i,y,z,u)\,d\phi_i$$
$$\bar{Z}^0_{(1)}(z,u)=Z^0_1(y,z,u)+\sum_{i=1}^{m}\bar{V}^{0,i}(y,z,u)$$
$$w^i_{(1)}=\frac{1}{X_i}\int^{\phi_i}\Bigg\{w^i_{(0)}\frac{\partial}{\partial z}Z^0_0+Z^i_1+\sum_{\substack{j,k=1\\ \theta_j+\theta_k=\psi_i}}^{m}\left[u^j_{(0)}\frac{\partial}{\partial\phi_k}Z^k_0+\mathfrak{v}^j_{(0)}\frac{\partial}{\partial\mathfrak{m}}Z^k_0\right]$$
$$-\mathfrak{X}^0_0\frac{\partial}{\partial\mathfrak{m}}w^i_{(0)}-W\frac{\partial}{\partial u}w^i_{(0)}-\bar{V}^{0,i}\Bigg\}\,d\psi_i.$$

Here the integrals are taken in such a way that $u^i_{(0)}$ and $\mathfrak{v}^i_{(0)}$ have mean value zero.

We will now proceed to define approximations for a solution of the initial value problem (2.3) starting in some domain D_1 (that is, the distance to any resonance manifold is $O_S(1)$), going through a resonance manifold until it is again in (another component) of D_1. This is the so-called "*passage through resonance*". The neighborhood of a resonance manifold is called the *inner domain* and its complement the *outer domain*.

In the outer domain the natural time scales are 1 and $1/\varepsilon$. In the inner domain it is $1/\sqrt{\varepsilon}$, or, if there is a saddlepoint, $1/\sqrt{\varepsilon}\log 1/\varepsilon$ depending on how close to this saddlepoint one permits the solution to come. All estimates in the $\sqrt{\varepsilon}$-neighborhood of the resonance manifold will be on a time-scale $\nabla 1/\sqrt{\varepsilon}$, but the Gronwall-type estimate allows one to take the logarithmic factor into account.

3. Estimates for the outer expansion. Now consider the "*outer equation*":

$$\begin{aligned}
&\dot{\phi} = X(z, u)\\
&\dot{y} = \varepsilon Y^0_{(0)}(y, z, u)\\
&\dot{z} = \varepsilon Z^0_{(0)}(z, u) + \varepsilon^2 \bar{Z}^0_{(1)}(z, u)\\
&\dot{u} = \varepsilon W(u).
\end{aligned} \tag{3.1}$$

The solution of this equation is called the *outer expansion*. In this section we are going to estimate the difference between the solutions of (2.3) and an outer expansion with appropriate initial conditions, as long as they are in the outer domain. Therefore the time-scale of validity has to be at least $1/\varepsilon$. Suppose one knows the initial conditions for (2.3) within some order of accuracy. Then the first problem is to show how one can use these approximate values to solve (2.9) or (3.1) without losing too much accuracy.

Let $(\theta_0, w_0, x_0, u_0)$ be the exact initial conditions and $(\bar{\theta}_0, \bar{w}_0, \bar{x}_0, u_0)$ some approximation with:

$$(\bar{\theta}_0, \bar{w}_0, \bar{x}_0, u_0) \in D_{\delta_0(\varepsilon)}$$

$$\left.\begin{aligned}
&\|\theta_0 - \bar{\theta}_0\| = O(\delta_1(\varepsilon)) && \delta_0(\varepsilon) = O(1)\\
&\|w_0 - \bar{w}_0\| = O(\delta_2(\varepsilon)) && \delta_i(\varepsilon) = o(\delta_0),\ i = 1, \cdots, 3\\
&\|x_0 - \bar{x}_0\| = O(\delta_3(\varepsilon)) && \frac{\varepsilon}{\delta_0^2(\varepsilon)} o(1)
\end{aligned}\right\} \quad \text{as } \varepsilon \downarrow 0. \tag{3.2}$$

(In the end we will consider the special case where all initial conditions are known exactly; this case follows from the present discussion.) Since u_0 is always known exactly, we will restrict our attention to the other variables. This is easily done, since Φ_ε, restricted to P^r, is the identity.

Define $\bar{\Phi}_\varepsilon$ to be

$$\bar{\Phi}_\varepsilon(\bar{\theta}_0, \bar{w}_0, \bar{x}_0) = \left(\theta_0, \bar{w}_0, \bar{x}_0 - \varepsilon \sum_{i=1}^{m} w^i_{(0)}(\bar{\theta}_{i0}, \bar{w}_0, \bar{x}_0, u_0)\right) \tag{3.3}$$

and let $(\bar{\phi}_0, \bar{y}_0, \bar{z}_0)$ be the image of $(\bar{\theta}_0, \bar{w}_0, \bar{x}_0)$ under $\bar{\Phi}_\varepsilon$.

One has, by abuse of notation:

$$\Phi_\varepsilon(\bar{\Phi}_\varepsilon(\bar{\theta}_0, \bar{w}_0, \bar{x}_0)) - \Phi_\varepsilon(\Phi_\varepsilon^{-1}(\theta_0, w_0, x_0))$$

$$= \begin{pmatrix} \bar{\theta}_0 + \varepsilon \sum_{i=1}^{m} u^i_{(0)}(\bar{\phi}_{i0}, \bar{y}_0, \bar{z}_0, u_0) - \theta_0 \\ \bar{w}_0 + \varepsilon \sum_{i=1}^{m} v^i_{(0)}(\bar{\phi}_{i0}, \bar{y}_0, \bar{z}_0, u_0) - w_0 \\ \bar{x}_0 - \sum_{i=1}^{m} w^i_{(0)}(\bar{\theta}_{i0}, \bar{w}_0, \bar{x}_0, u_0) + \varepsilon \sum_{i=1}^{m} w^i_{(0)}(\bar{\phi}_{i0}, \bar{y}_0, \bar{z}_0, u_0) \\ + \varepsilon^2 \sum_{i=1}^{m^*} w^i_{(1)}(\bar{\phi}_{i0}, \bar{y}_0, \bar{z}_0, u_0) - x_0 \end{pmatrix} \tag{3.4}$$

$$+ \begin{pmatrix} O(\delta_1) + O\left(\frac{\varepsilon}{\delta_0^2}\right) \\ O(\delta_2) + O\left(\frac{\varepsilon}{\delta_0}\right) \\ O(\delta_3) + O\left(\frac{\varepsilon^2}{\delta_0^3}\right) \end{pmatrix} \quad \text{as } \varepsilon \downarrow 0 \text{ on } D_{\delta_0}.$$

Thus $\bar{\Phi}_\varepsilon$ approximates Φ_ε^{-1} in a certain sense. The mean-value theorem gives the following estimate for the difference between $(\bar{\phi}_0, \bar{y}_0, \bar{z}_0)$ and $(\phi_0, y_0, z_0) = \Phi_\varepsilon^{-1}(\theta_0, w_0, x_0)$:

$$\begin{aligned} \|\phi_0 - \bar{\phi}_0\| &= O(\delta_1) + O\left(\frac{\varepsilon}{\delta_0^2}\right) \\ \|y_0 - \bar{y}_0\| &= O(\delta_2) + O\left(\frac{\varepsilon}{\delta_0}\right) \\ \|z_0 - \bar{z}_0\| &= O(\delta_3) + O\left(\frac{\varepsilon^2}{\delta_0^3}\right) + O\left(\frac{\varepsilon\delta_1}{\delta_0}\right) + O\left(\frac{\varepsilon\delta_2}{\delta_0^2}\right). \end{aligned} \tag{3.5}$$

If we started with exact conditions, (3.5) would read:

$$\begin{aligned} \|\phi_0 - \bar{\phi}_0\| &= O\left(\frac{\varepsilon}{\delta_0^2}\right) \\ \|y_0 - \bar{y}_0\| &= O\left(\frac{\varepsilon}{\delta_0}\right) \\ \|z_0 - \bar{z}_0\| &= O\left(\frac{\varepsilon}{\delta_0}\right). \end{aligned} \tag{3.6}$$

This completes the estimates for the initial conditions.

We now turn to the solutions.

Let $(\phi, y, z, u)(\cdot)$ be the solution of (2.9) with initial conditions (ϕ_0, y_0, z_0, u_0) and let $(\bar{\phi}, \bar{y}, \bar{z}, u)(\cdot)$ be the solution of (3.1) with initial conditions $(\bar{\phi}_0, \bar{y}_0, \bar{z}_0, u_0)$. Note that $\Phi_\varepsilon(\phi, y, z, u)(t) = (\theta, w, x, u)(t)$ since (2.9) and (2.3) are Φ_ε-related and have unique solutions. (This is a general line of argument: Let v, $\bar{v}$ and ϕ be as in § 2; suppose x is an orbit of $\bar{v}$, i.e. $\dot{x}(t)=\bar{v}(x(t))$. Define $\phi(x)(\cdot)$ by $(\phi(x))(t)=\phi(x(t))$ and put $y=\phi(x)$. Then: $\dot{y}(t)=d\phi(x(t))\bar{v}(x(t))=v(\phi(x(t)))=v(\phi(x)(t))=v(y(t))$. Thus y is an orbit of v. Let $\tilde{y}$ be another orbit of v with initial conditions $\tilde{y}(0)=\phi(x(0))$. Then, by uniqueness, $\tilde{y}=y$.)

We now make the following

ASSUMPTION. *Let* $n\in\mathbb{N}$, $n\geqq 2$. *Then*

$$\int_0^t \frac{\varepsilon\, d\tau}{X^n(\bar{z}(\tau), u(\tau))}\leqq C\left(\frac{1}{X^{n-1}(\bar{z}_0, u_0)}+\frac{1}{X^{n-1}(\bar{z}(t), u(t))}\right). \tag{3.7}$$

This assumption has to be made to insure that the solution enters (leaves) the resonance fast enough. For suppose $Z^0_{(0)}$ is zero or very small; then (3.7) is not likely to hold, because the natural time-scale for $\bar{z}$ will be much longer than $1/\varepsilon$ in that case.

Of course one can develop a theory for such cases by changing the assumption and putting the appropriate time-scale in formula (3.7).

Note that the problem with $Z^0_{(0)}=0$ and $\bar{Z}^0_{(1)}\neq 0$ is not only of theoretical interest, but it has an application in the spin-orbit resonances in 2-body problems with low eccentricity (for the equations, see Kyner (1969)).

In § 8 we show that it is not always impossible to verify (3.7) in concrete situations; this is because $\bar{z}$ and u are known explicitly, at least formally.

LEMMA. *With the notations as above and with assumption* (3.7) *satisfied, the following estimate holds*:

$$\begin{aligned}
&\|\theta(t)-\bar{\phi}(t)\|=O\left(\frac{\varepsilon}{\delta^2}\right)+O\left(\frac{\delta_1}{\delta_0}\right)+O\left(\frac{\delta_2}{\delta_0^2}\right)+O\left(\frac{\delta_3}{\varepsilon}\right)+O\left(\frac{\varepsilon}{\delta_0^3}\right)\\
&\|w(t)-\bar{y}(t)\|=O\left(\frac{\varepsilon}{\delta}\right)+O(\delta_2)+O(\delta_3)\\
&\|x(t)-\bar{z}(t)\|=O\left(\frac{\varepsilon}{\delta}\right)+O\left(\frac{\varepsilon\delta_2}{\delta_0^2}\right)+O(\delta_3)\\
&\left\|x(t)-\bar{z}(t)-\varepsilon\sum_{i=1}^m w^i_{(0)}(\bar{\phi}(t),\bar{y}(t),\bar{z}(t),u(t))\right\|\\
&\qquad=O\left(\frac{\varepsilon^2}{\delta^3}\right)+O\left(\frac{\varepsilon\delta_1}{\delta_0}\right)+O\left(\frac{\varepsilon\delta_2}{\delta_0^2}\right)+O(\delta_3)
\end{aligned} \tag{3.8}$$

as long as $0\leqq\varepsilon t\leqq L$ *and* $(\bar{\phi},\bar{y},\bar{z},u)(t)\in D_\delta$.

Proof. In this proof appear a number of C_i, $i=1,2,\cdots$, all ε-independent. These will be denoted by the universal constant C to simplify notation. Suppose $\tilde{\theta}$ is some function from $\mathbb{R}\to[0,1]$ and let

$$\xi_{\tilde{\theta}}=\tilde{\theta}z+(1-\tilde{\theta})\bar{z}. \tag{3.9}$$

Then

$$\begin{aligned}\xi_{\tilde{\theta}}(0)&=\tilde{\theta}(0)z(0)+(1-\tilde{\theta}(0))\bar{z}(0)=\bar{z}(0)+\tilde{\theta}(0)(z(0)-\bar{z}(0))=\bar{z}(0)+O(\delta_3),\\
&=\bar{z}(0)+o(1)\quad\text{as } \varepsilon\downarrow 0.\end{aligned} \tag{3.10}$$

Thus there is a nonempty interval $[0,\tau^*)$, with $0<\varepsilon\tau^*\leqq L$, such that

$$\frac{1}{X(\xi_{\tilde{\theta}}(t),u(t))}<\frac{2}{X(\bar{z}(t),u(t))},\qquad t\in[0,\tau^*) \tag{3.11}$$

because $\bar{z}(0) \in D_\delta$ and $1/X$ behaves more or less nicely there. Suppose there is some θ_0 such that, if $\xi_0 = \theta_0 z + (1-\theta_0)\bar{z}$, then

$$\frac{1}{X(\xi_0(\tau^*), u(\tau^*))} = \frac{2}{X(\bar{z}(\tau^*), u(\tau^*))}. \tag{3.12}$$

Combining (2.9), (3.1) and the mean-value theorem, one obtains the following estimate:

$$z(t) = z(0) + \int_0^t \Big\{ \varepsilon Z_{(0)}^0(z(\sigma), u(\sigma)) + \varepsilon^2 \bar{Z}_{(1)}^0(z(\sigma), u(\sigma)) + \frac{\varepsilon^3}{X^4(z(\sigma), u(\sigma))} R_3(\phi(\sigma), y(\sigma), z(\sigma), u(\sigma); \varepsilon) \Big\} \, d\sigma,$$

$$\bar{z}(t) = \bar{z}(0) + \int_0^t \{ \varepsilon Z_{(0)}^0(\bar{z}(\sigma), u(\sigma)) + \varepsilon^2 \bar{Z}_{(1)}^0(\bar{z}(\sigma), u(\sigma)) \} \, d\sigma,$$

which implies

$$\begin{aligned}
\|z(t) - \bar{z}(t)\| \leqq \|z(0) - \bar{z}(0)\| &+ \int_0^t \Big\{ \varepsilon \|Z_{(0)}^0(z(\sigma), u(\sigma)) - Z_{(0)}^0(\bar{z}(\sigma), u(\sigma))\| \\
&+ \varepsilon^2 \|\bar{Z}_{(1)}^0(z(\sigma), u(\sigma)) - \bar{Z}_{(1)}^0(\bar{z}(\sigma), u(\sigma))\| \\
&+ \varepsilon^3 \frac{1}{X^4(z(\sigma), u(\sigma))} \|R_3(\phi(\sigma), y(\sigma), z(\sigma), u(\sigma); \varepsilon)\| \Big\} \, d\sigma \\
\leqq \|z(0) - \bar{z}(0)\| &+ \int_0^t C \Big\{ \Big(\varepsilon + \frac{\varepsilon^2}{X^2(\xi_{\theta^*}(\sigma), u(\sigma))} \Big) \|z(\sigma) - \bar{z}(\sigma)\| \\
&+ \frac{\varepsilon^3}{X^4(z(\sigma), u(\sigma))} \Big\} \, d\sigma \\
\leqq \|z(0) - \bar{z}(0)\| &+ C\varepsilon \int_0^t \|z(\sigma) - \bar{z}(\sigma)\| \, d\tau + \int_0^t \frac{C\varepsilon^3}{X^4(\bar{z}(\sigma), u(\sigma))} \, d\sigma.
\end{aligned} \tag{3.13}$$

In proving (3.13) we have used implicitly the following estimates:

$$\begin{aligned}
&\frac{\varepsilon}{X^2(\xi_{\theta^*}(\sigma), u(\sigma))} = O\Big(\frac{\varepsilon}{X^2(\bar{z}(\sigma), u(\sigma))} \Big) = O\Big(\frac{\varepsilon}{\delta^2} \Big) = o(1) \quad \text{as } \varepsilon \downarrow 0 \\
&\qquad\qquad (\tilde{\theta} = \theta^* \text{ in } (3.11)) \\
&\frac{1}{X^4(z(\sigma), u(\sigma))} = O\Big(\frac{1}{X^4(\bar{z}(\sigma), u(\sigma))} \Big) \qquad (\tilde{\theta}(t) = 1 \text{ in } (3.11)).
\end{aligned} \tag{3.14}$$

The estimate for the derivative of $Z_{(1)}^0$ follows from the definition of $Z_{(1)}^0$ in (2.12).

Here we use the fact that $Z_{(0)}^0$ and $\bar{Z}_{(1)}^0$ do not depend on y. The existence of θ^* follows from the mean-value theorem. We apply Gronwall's lemma to (3.13) to get

$$\|z(t) - \bar{z}(t)\| \leqq \|z(0) - \bar{z}(0)\| e^{C\varepsilon t} + \int_0^t \frac{C\varepsilon^3}{X^4(\bar{z}(\sigma), u(\sigma))} e^{(t-\sigma)} \, d\sigma. \tag{3.15}$$

Using (3.7) gives

$$\|z(t) - \bar{z}(t)\| \leqq C\Big(\|z(0) - \bar{z}(0)\| + \frac{\varepsilon^2}{X^3(z(t), u(t))} + \frac{\varepsilon^2}{\delta_0^3} \Big) \tag{3.16}$$

and thus

$$\|z(\tau^*)-\bar{z}(\tau^*)\| = O\left(\frac{\delta_1\varepsilon}{\delta_0}\right)+O\left(\delta_2\frac{\varepsilon}{\delta^2}\right)+O(\delta_3)+O\left(\frac{\varepsilon^2}{\delta_0^3}\right)+O\left(\frac{\varepsilon^2}{\delta^3}\right)$$
$$= o(1) \quad \text{as } \varepsilon \downarrow 0. \tag{3.17}$$

Applying (3.17) to (3.12) we see that

$$\begin{aligned}
&\frac{2}{X(\bar{z}(\tau^*), u(\tau^*))}\\
&\quad = \frac{1}{X(\xi_0(\tau^*), u(\tau^*))}\\
&\quad \leqq \frac{1}{X(\bar{z}(\tau^*), u(\tau^*))} + \left|\frac{1}{X(\xi_0(\tau^*), u(\tau^*))} - \frac{1}{X(\bar{z}(\tau^*), u(\tau^*))}\right|\\
&\quad \leqq \frac{1}{X(\bar{z}(\tau^*), u(\tau^*))} + \frac{C\|z(\tau^*)-\bar{z}(\tau^*)\|}{X^2(\bar{z}(\tau^*), u(\tau^*))}\\
&\quad \leqq \frac{1}{X(\bar{z}(\tau^*), u(\tau^*))}\left(1+\frac{C\|z(\tau^*)-\bar{z}(\tau^*)\|}{X(\bar{z}(\tau^*), u(\tau^*))}\right)\\
&\quad = \frac{1}{X(\bar{z}(\tau^*), u(\tau^*))}(1+o(1)) \quad \text{as } \varepsilon \downarrow 0,
\end{aligned} \tag{3.18}$$

which contradicts the existence of ξ_0 and τ^*. Thus (3.18) implies that (3.15) is valid on $0 \leqq \varepsilon t \leqq L$.

The only difficulty in this proof was that in order to estimate $\|z(t)-\bar{z}(t)\|$ one needed an estimate for $1/(X^n(z(t), u(t)))$ in terms of $1/(X^n(\bar{z}(t), u(t)))$, which was easy enough, once one had an estimate for $\|z(t)-\bar{z}(t)\|$, etc.

The remaining estimates follow easily:

$$\begin{aligned}
&\|\phi(t)-\bar{\phi}(t)\|\\
&\quad \leqq \|\phi_0-\bar{\phi}_0\| + C_0\int_0^t \|z(\tau)-\bar{z}(\tau)\|\,d\tau + C\varepsilon\int_0^t \frac{\varepsilon}{X^3(z(\tau), u(\tau))}\,d\tau\\
&\quad \leqq \|\phi_0-\bar{\phi}_0\| + C\left(\|z_0-\bar{z}_0\|t+\frac{\varepsilon^2}{\delta_0^3}t\right) + C\varepsilon\int_0^t \frac{\varepsilon}{X^3(\bar{z}(\tau), u(\tau))}\,d\tau\\
&\quad \leqq \|\phi_0-\bar{\phi}_0\| + \frac{C}{\varepsilon}\left(\|z_0-\bar{z}_0\|+\frac{\varepsilon^2}{\delta_0^3}\right)+\frac{C\varepsilon}{\delta^2}.
\end{aligned} \tag{3.19}$$

Using (3.5) in (3.19) gives:

$$\|\phi(t)-\bar{\phi}(t)\| = O\left(\frac{\delta_3}{\varepsilon}\right)+O\left(\frac{\delta_1}{\delta_0}\right)+O\left(\frac{\delta_2}{\delta_0^2}\right)+O\left(\frac{\varepsilon}{\delta_0^3}\right)+O\left(\frac{\varepsilon}{\delta^2}\right). \tag{3.20}$$

In the same way one obtains

$$\|y(t)-\bar{y}(t)\| = O(\delta_2)+O(\delta_3)+O\left(\frac{\varepsilon}{\delta_0}\right)+O\left(\frac{\varepsilon}{\delta}\right). \tag{3.21}$$

The derived estimate (3.8) now follows from (2.7) (using (3.11) and applying the triangle inequality).

Suppose the initial conditions are known exactly and $\delta_0 = 1$; then (3.8) becomes

$$\|\theta(t) - \bar{\phi}(t)\| = O\left(\frac{\varepsilon}{\delta^2}\right)$$

$$\|w(t) - \bar{y}(t)\| = O\left(\frac{\varepsilon}{\delta}\right) \tag{3.22}$$

$$\|x(t) - \bar{z}(t)\| = O\left(\frac{\varepsilon}{\delta}\right)$$

$$\left\|x(t) - \bar{z}(t) - \varepsilon \sum_{i=1}^{m} w^{i}_{(0)}(\bar{\phi}(t), \bar{y}(t), \bar{z}(t), u(t))\right\| = O\left(\frac{\varepsilon^2}{\delta^3}\right).$$

This is the case of a solution starting in D_1, coming to the region of resonance.

4. Estimates for the inner expansion. The results in the preceding section hold in D_δ. That leaves us with the question: what happens if the solution does not stay in D_δ, but enters the inner domain?

To answer this question we introduce the inner vector field. To estimate the difference between the exact solution and the inner expansion (i.e., the solution of the inner vector field) is not easy. We do it in two steps: first in a $\sqrt{\varepsilon}$-neighborhood of a resonance manifold, then in the next section in δ-neighborhood minus the $\sqrt{\varepsilon}$-neighborhood.

The first step is straightforward:

Suppose we have already constructed an approximation that is entering the $\sqrt{\varepsilon}$-neighborhood of $\mathcal{N}_j$. Then we can average over all angles except θ_j. Since we know that we are working in a $\sqrt{\varepsilon}$-neighborhood of $\mathcal{N}_j$, we scale the coordinate transversal to $\mathcal{N}_j$ and develop the vector field by Taylor-expansion. Throwing away the higher order terms, we call the remaining equation the "*inner vector field*" and its solutions *inner expansions.*

So let $D^{\hat{j}}_\delta$ denote the intersection of the local coordinate domain and $\bigcap_{i\neq j}^{m^*} \mathcal{M}_i(\delta)$ and define

$$\Psi_\varepsilon : T^{m^*} \times D^{\hat{j}} \to T^{m^*} \times D$$

$$\begin{pmatrix} \phi \\ y \\ z \\ u \end{pmatrix} \mapsto \begin{pmatrix} \phi + \varepsilon \sum_{i\neq j}^{m} u^{i}_{(0)}(\phi_i, y, z, u) \\ y + \varepsilon \sum_{i\neq j}^{m} v^{i}_{(0)}(\phi_i, y, z, u) \\ z + \varepsilon \sum_{i\neq j}^{m} w^{i}_{(0)}(\phi_i, y, z, u) \\ u \end{pmatrix} \tag{4.1}$$

where $u^{i}_{(0)}$, $v^{i}_{(0)}$ and $w^{i}_{(0)}$ are defined as in (2.11) and (2.12).

This transformation is intermediate between first and second order averaging and has to be used since first order averaging doesn't give the desired results. It is not obvious why first order averaging does not suffice. The technical reasons for its failure will appear only in the next section.

The Ψ_ε-related vector field is:

$$\begin{aligned}\dot\phi &= X(z,u)+\varepsilon^2 R_1(\phi, y, z, u;\varepsilon)\\ \dot y &= \varepsilon Y_0^0(y,z,u)+\varepsilon Y_0^j(\phi_j, y, z, u)+\varepsilon^2 R_2(\phi, y, z, u;\varepsilon)\\ \dot z &= \varepsilon Z_0^0(z,u)+\varepsilon Z_0^j(\phi_j, y, z, u)+\varepsilon^2 R_3(\phi, y, z, u;\varepsilon)\\ \dot u &= \varepsilon W(u)\end{aligned} \tag{4.2}$$

where the R_i are, as usual, uniformly bounded and

$$X_j(z,u)|_{\mathcal{N}_j}=0. \tag{4.3}$$

Let z_0^j be a zero of $X_j^{\#}$ and scale

$$z-z_0^j=\sqrt{\varepsilon}\zeta. \tag{4.4}$$

We have $\zeta=O(1)$, since we are in a $\sqrt{\varepsilon}$-neighborhood of $\mathcal{N}_j$. We assume that $(\partial X_j/\partial z)(z_0^j, u)\neq 0$ (cf. (3.7)).

Following the outline in the Introduction we write down the inner equations:

$$\begin{aligned}\dot\phi &= X(z_0^j, u_0)+\sqrt{\varepsilon}\zeta\frac{\partial X}{\partial z}(z_0^j,u_0)\\ \dot y &= 0\\ \dot\zeta &= \sqrt{\varepsilon}\tilde Z^j(\phi_j, y_0, z_0^j, u_0)\\ \dot u &= \varepsilon W(u)\end{aligned} \tag{4.5}$$

where $\tilde Z^j = Z_0^0+Z_0^j$. The initial conditions u_0 and y_0 used in (4.5) are initial with respect to the moment of entering the $\sqrt{\varepsilon}$-neighborhood. In the sequel both z and ζ are used, but one should keep in mind that (4.4) always holds, although not explicitly written down on every occasion.

Let $(\phi, y, z, u)(\cdot)$ be the solution of (4.2) with initial conditions (ϕ_0, y_0, z_0, u_0) and let $(\bar\phi, \bar y, \bar\zeta, u)$ be the solution of (4.5) with initial conditions $(\bar\phi, \bar y_0, \bar\zeta_0, u_0)$.

(Since the vector field is autonomous, one can always translate in time and call any t_0 zero; thus the zero-time in this section need not be the zero-time in another section.)

Using again Gronwall's lemma, we see that the following estimate holds:

$$\begin{aligned}&\left(\|\phi_j(t)-\bar\phi_j(t)\|+\|y(t)-\bar y(t)\|+\frac{1}{\sqrt{\varepsilon}}\|z(t)-\bar z(t)\|\right)\\ &\quad\leqq\left(\|\phi_{j0}-\bar\phi_{j0}\|+\|y_0-\bar y_0\|+\frac{1}{\sqrt{\varepsilon}}\|z_0-\bar z_0\|+C\varepsilon t\right)e^{C\sqrt{\varepsilon}t}.\end{aligned} \tag{4.6}$$

Moreover

$$\|y(t)-\bar y(t)\|\leqq C(\|y_0-\bar y_0\|+\varepsilon t). \tag{4.7}$$

These estimates can be obtained by a slight modification of the method of analysis given in Sanders (1978).

Let (θ, w, x, u) be the solution of (2.3) with initial conditions $(\theta_0, w_0, x_0, u_0)$ and let (ϕ, y, z, u) be the solution of (4.2) with the same initial conditions; then it is not difficult to see that the difference between these solutions is of $O(\varepsilon)$.

Together with (4.6) this gives us the desired estimate for the difference between (θ_j, w, x, u) and $(\bar\phi_j, \bar y, \bar z, u)$, on a time scale $0\leqq\sqrt{\varepsilon}t\leqq L$. It is not clear how to extend this time-scale using Gronwall-type estimates, other than with logarithmic factors in ε.

5. Extension of the time-scale for the inner expansion. Were it not for the fact that $\varepsilon/\delta^2 = o(1)$ as $\varepsilon \downarrow 0$, the preceding analysis would have given the complete picture of what is happening. The problem remains to extend the validity of the inner (or the outer, or some other) expansion to a $\delta(\varepsilon)$-neighborhood of the resonance manifold. The difficulty of extending the validity of the inner expansion is in the fact that its natural time-scale is of $O(1/\sqrt{\varepsilon})$ which is not enough to go through the larger δ-neighborhood.

Following an idea due to Eckhaus (1976), we prove in this section the validity of the inner expansion on the larger time scale δ/ε.

It follows from (4.2) by a straightforward argument that one has the following estimates on its solutions:

$$\begin{aligned}
&\zeta(t) = O\left(\frac{\delta}{\sqrt{\varepsilon}}\right)\\
&u(t) = u_0 + O(\delta) \quad \text{on } t \in \left[0, \frac{\delta}{\varepsilon}L\right]\\
&y(t) = y_0 + O(\delta).
\end{aligned}\tag{5.1}$$

Using (5.1) we expand (4.2) to get:

$$\begin{aligned}
&\dot{\phi} = X(z_0^i, u_0) + \sqrt{\varepsilon}\zeta\frac{\partial X}{\partial z}(z_0^i, u_0) + (\varepsilon\zeta^2 + \varepsilon^2)R_1(\phi, y, \zeta, u; \varepsilon)\\
&\dot{y} = \varepsilon R_2(\phi, y, \zeta, u; \varepsilon)\\
&\dot{\zeta} = \sqrt{\varepsilon}\tilde{Z}^i(\phi_j, y_0, z_0^i, u_0) + \sqrt{\varepsilon}\delta R_3(\phi, y, \zeta, u; \varepsilon)\\
&\dot{u} = \varepsilon W(u).
\end{aligned}\tag{5.2}$$

Differentiating ϕ_j gives

$$\begin{aligned}
\ddot{\phi}_j &= \sqrt{\varepsilon}\frac{\partial X_j}{\partial z}(z_j^0, u_0)\dot{\zeta} + (\sqrt{\varepsilon}\delta\dot{\zeta} + \varepsilon^2)R_{1,j}(\phi, y, \zeta, u; \varepsilon)\\
&= \varepsilon\frac{\partial X_j}{\partial z}(z_j^0, u_0)\tilde{Z}^i(\phi_j, y_0, z_0^i, u_0) + \varepsilon\delta R_{1,j}(\phi, y, \zeta, u; \varepsilon).
\end{aligned}\tag{5.3}$$

(Here we have used the "one-and-a-half-order averaging").

Integrating yields

$$\begin{aligned}
\frac{1}{2}\dot{\phi}_j^2 - \frac{1}{2}\dot{\phi}_j(0)^2 &= \varepsilon\int_{\phi_j(0)}^{\phi_j}\frac{\partial X}{\partial z}\tilde{Z}^i(\psi_j)\,d\psi_j + \varepsilon\delta\int_0^t R_{1,j}\dot{\phi}_j(\tau)\,d\tau\\
&= \varepsilon H(\phi_j, \dot{\phi}_j(0), y_0, z_0^i, u_0)\\
&\quad + \varepsilon\delta\int_0^t \dot{\phi}_j(\tau)R_{1,j}(\phi(\tau), y(\tau), \zeta(\tau), u(\tau); \varepsilon)\,d\tau.
\end{aligned}\tag{5.4}$$

Note: H is not a function on the circle but on its covering space, since Z^0 need not be zero.

Consider the equation:

$$\tfrac{1}{2}\dot{\phi}_j^2 - \tfrac{1}{2}\dot{\phi}_j(0)^2 = \varepsilon H(\phi_j, \dot{\phi}_j(0), y_0, z_0^i, u_0).\tag{5.5}$$

Let ϕ^* be the solution of (5.5) with $\phi^*(0)=\bar{\phi}_j(0)$ and let t^* be the solution of

$$\dot{t}^{*2}=1+\frac{\varepsilon\delta\int_0^t \dot{\phi}^*(t^*(\tau))R_{1,j}(\phi^*(t^*(\tau)), y(\tau), \zeta(\tau), u(\tau); \varepsilon)\,d\tau}{\frac{1}{2}\dot{\phi}_j(0)^2+\varepsilon H(\phi^*(t^*), \phi_j(0), y_0, z_0^j, u_0)} \tag{5.6}$$

with initial condition such that $\phi^*(t^*(0))=\phi_j(0)$.

Define

$$\tilde{\phi}_j(t)=\phi^*(t^*(t)); \tag{5.7}$$

then

$$\dot{\tilde{\phi}}_j \doteq \dot{t}^*\frac{d\phi^*}{dt^*} \tag{5.8}$$

and

$$\begin{aligned}\frac{1}{2}\dot{\tilde{\phi}}_j^2(t)&=\left\{1+\frac{\varepsilon\delta\int_0^t \dot{\phi}^*(t^*(\tau))R_{1,j}(\phi^*(t^*(\tau)), y(\tau), \zeta(\tau), u(\tau); \varepsilon)\,d\tau}{\frac{1}{2}\dot{\phi}_j(0)^2+\varepsilon H(\phi^*(t^*(t)), \phi_j(0), y_0, z_0^j, u_0)}\right\}\left(\frac{1}{2}\dot{\phi}_j(0)^2+\varepsilon H\right)\\ &=\frac{1}{2}\dot{\phi}(0)^2+\varepsilon H(\phi^*(t^*(t)), \phi_j(0), y_0, z_0^j, u_0)\\ &\quad+\varepsilon\delta\int_0^t \dot{\phi}^*(t^*(\tau))R_{1,j}(\phi^*(t^*(\tau)), \cdots,)\,d\tau\\ &=\frac{1}{2}\dot{\phi}_j(0)^2+\varepsilon H(\tilde{\phi}_j(t), \phi_j(0), y_0, z_0^j, u_0)+\varepsilon\delta\int_0^t \dot{\tilde{\phi}}(\tau)R_{1,j}(\tilde{\phi}(\tau), y(\tau), \cdots,)\,d\tau.\end{aligned} \tag{5.9}$$

$\tilde{\phi}_j$ therefore obeys (5.4) and has as initial condition $\tilde{\phi}_j(0)=\phi^*(t^*(0))=\phi_j(0)$.

Furthermore

$$|\phi^*(t^*(t))-\phi^*(t)|\leqq\int_t^{t^*(t)}|\dot{\phi}^*(\tau)|\,d\tau\leqq C\delta(\varepsilon)|t^*(t)-t|. \tag{5.10}$$

Assume

$$\frac{\varepsilon}{\frac{1}{2}\dot{\phi}_j(0)^2+\varepsilon H(\phi^*(t^*), \phi_j(0), y_0, z_0^j, u_0)}=O(1) \tag{5.11}$$

which is more or less equivalent to $\varepsilon/|\dot{\phi}_j^2|=O(1)$. Thus $1/\zeta^2=O(1)$, implying that one has to stay outside the $\sqrt{\varepsilon}$-neighborhood of the resonance manifold.

From (5.6), using (5.11), we obtain the following estimate:

$$|t^*(t)-t|\leqq C\left(\frac{\delta^4}{\varepsilon^2}+\left|\frac{\phi_{j0}-\bar{\phi}_{j0}}{\dot{\phi}_{j0}}\right|\right)\quad\text{with } t\in\left[0, \frac{\delta}{\varepsilon}L\right], \tag{5.12}$$

which implies:

$$|\tilde{\phi}_j(t)-\phi^*(t)|\leqq C\left(\frac{\delta^5}{\varepsilon^2}+\frac{\delta}{|\dot{\phi}_{j0}|}|\phi_{j0}-\bar{\phi}_{j0}|\right). \tag{5.13}$$

It is easy to see that on the one hand if we have a solution of (4.5) with initial conditions $(\bar{\phi}_0, \bar{y}_0, \bar{\zeta}_0, u_0)$, then its ϕ_j-component can be identified with ϕ^*; while on the other hand $\tilde{\phi}_j$ has initial conditions $\tilde{\phi}_j(0)=\phi_j(0)$ and $\dot{\tilde{\phi}}_j(0)=\dot{\phi}_j(0)$, which implies that $\tilde{\phi}_j$ is equal to the ϕ_j-component of the solution of (4.2) with initial conditions (ϕ_0, y_0, z_0, u_0) by uniqueness of the solutions of this equation.

We now have two estimates ((5.14) and (5.15)) for the difference between ϕ_j and $\bar{\phi}_j$, one for the ingoing and one for the outgoing solution. (We call a solution ingoing if it enters the resonance region from outside, i.e. from D_δ.) The first estimate, assuming $|\dot{\phi}_{j0}| \sim \delta$, is

$$|\phi_j(t) - \bar{\phi}_j(t)| \leqq C\left(\frac{\delta^5}{\varepsilon^2} + |\phi_{j0} - \bar{\phi}_{j0}|\right) \tag{5.14}$$

while the second, assuming $|\dot{\phi}_{j0}| \sim \sqrt{\varepsilon}$, is:

$$|\phi_j(t) - \bar{\phi}_j(t)| \leqq C\left(\frac{\delta^5}{\varepsilon^2} + \frac{\delta}{\sqrt{\varepsilon}}|\phi_{j0} - \bar{\phi}_{j0}|\right). \tag{5.15}$$

Using (4.5) and (5.2) we obtain

$$\begin{aligned} |\zeta(t) - \bar{\zeta}(t)| &\leqq |\zeta_0 - \bar{\zeta}_0| + \sqrt{\varepsilon}\int_0^t C|\phi_j(\tau) - \bar{\phi}_j(\tau)|\, d\tau + C\sqrt{\varepsilon}\delta t \\ &\leqq |\zeta_0 - \bar{\zeta}_0| + C\frac{\delta}{\sqrt{\varepsilon}}\left(\frac{\delta^5}{\varepsilon^2} + \left\{1, \frac{\delta}{\sqrt{\varepsilon}}\right\}|\phi_{j0} - \bar{\phi}_{j0}|\right) + C\frac{\delta^2}{\sqrt{\varepsilon}}. \end{aligned} \tag{5.16}$$

Here $\{1, \delta/\sqrt{\varepsilon}\}$ means 1 for the ingoing and $\delta/\sqrt{\varepsilon}$ for the outgoing estimate.

Note that $\delta^2/\sqrt{\varepsilon}$ is small compared with $\delta^6/(\varepsilon^2\sqrt{\varepsilon})$, since $\varepsilon/\delta^2 = o(1)$. Our final estimate is the easiest:

$$\|y(t) - \bar{y}(t)\| \leqq C(\delta + \|y_0 - \bar{y}_0\|). \tag{5.17}$$

The connection of the approximations found here with the original equation has been discussed at the end of the preceding section.

6. The inner-outer vector field. The inner-outer vector field may have little to do with the original equations, since it is the inner expansion of the outer vector field and the outer expansion of the inner vector field.

This means that one is "assuming" two contradictory facts at once: that one is near a resonance manifold and far away from all of them.

But the flow of the inner-outer vector field (the inner-outer expansion) does approximate the inner expansion in the outer region and the outer expansion near the resonance manifold. Thus it is possible to make composite expansions like:

$$x_C = x_I + x_O - x_{IO} \tag{6.1}$$

where x_I, x_O and x_{IO} are the inner, outer and inner-outer expansion. In the outer domain the estimates run as follows:

$$\|x - x_C\| \leqq \|x - x_O\| + \|x_I - x_{IO}\| = o(1) \quad \text{as } \varepsilon \downarrow 0 \tag{6.2}$$

and in the inner domain:

$$\|x - x_C\| \leqq \|x - x_I\| + \|x_O - x_{IO}\| = o(1) \quad \text{as } \varepsilon \downarrow 0 \tag{6.3}$$

where x is the exact solution and all approximations have the right initial conditions (or asymptotic behavior).

The inner-outer equation is:

$$\begin{aligned}
\dot\phi &= X(z_0^j, u_0)+\sqrt{\varepsilon}\zeta\frac{\partial X}{\partial z}(z_0^j, u_0)\\
\dot y &= 0\\
\dot\zeta &= \sqrt{\varepsilon}Z^0(z_0^j, u_0)\\
\dot u &= \varepsilon W(u).
\end{aligned}\tag{6.4}$$

Integrating (6.4) one shows that $\phi_{IO}(t)$ is a parabola which takes an extremal value at the resonance manifold under consideration. (All the inner-outer talk is about one passage through one resonance; this is not a global inner-outer field, as is indicated by the presence of z_0^j in (6.4).)

To get the right inner-outer expansions we proceed as follows: (x_O^{in} denotes the ingoing outer solution, etc.).

Let $x_O^{\text{in}}(0)=x(0)$ ($x(0)$ is the given initial condition of the exact solution). Suppose at some τ, $x_0(\tau)\in\mathcal{N}_j$. Let $x_{IO}^{\text{in}}(\tau)=x_O^{\text{in}}(\tau)$; this determines $x_{IO}(0)$. Then let $x_I(0)=x_{IO}^{\text{in}}(0)$. All estimates are given in the next section, proving that this is a possible way to proceed. On $[0, \tau)$ the composite expansion is defined by

$$x_C^{\text{in}}(t)=x_I(t)+x_O^{\text{in}}(t)-x_{IO}^{\text{in}}(t).\tag{6.5}$$

On the next interval $[\tau, \bar\tau)$, where $\bar\tau$ is to be defined below, let

$$x_C^I(t)=x_I(t).\tag{6.6}$$

Suppose at some time $\tilde t$ the inner solution has a distance of $O_S(1)$ to the resonance manifold.

Let $x_{IO}^{\text{out}}(\tilde t)=x_I(\tilde t)$. Now $\bar\tau$ is defined by $x_{IO}^{\text{out}}(\bar\tau)\in\mathcal{N}_j$. Let $x_O(\bar\tau)=x_{IO}(\bar\tau)$. Then the outgoing composite expansion is, on $[\bar\tau, T)$, given by

$$x_C^{\text{out}}(t)=x_I(t)+x_O^{\text{out}}(t)-x_{IO}^{\text{out}}(t).\tag{6.7}$$

Collecting all our results we can prove that the composite expansion obtained by

$$x_C(t)=\begin{cases} x_C^{\text{in}}(t) & 0\leqq t<\tau\\ x_C^I(t) & \tau\leqq t<\bar\tau\\ x_C^{\text{out}}(t) & \bar\tau\leqq t<T\end{cases}\tag{6.8}$$

is in fact an approximation of the exact solution.

7. Estimates for the composite expansion. We will now give the estimates for (6.4). Using the notation adopted there, the solution $(\bar\phi, \bar y, \bar z)$ in § 3 will read $(\theta_0^{\text{in}}, w_0^{\text{in}}, x_0^{\text{in}})$ here.

First using (3.22) we obtain

$$\begin{aligned}
\|\theta(t)-\theta_O^{\text{in}}(t)\| &= O\left(\frac{\varepsilon}{\delta_{\text{in}}^2}\right)\\
\|w(t)-w_O^{\text{in}}(t)\| &= O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right)\qquad 0\leqq t\leqq\hat\tau<\tau\\
|x(t)-x_O^{\text{in}}(t)| &= O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right)
\end{aligned}\tag{7.1}$$

where δ_{in} is an order function as defined in (2.6). Using (3.22) again, estimates the difference between the inner and the ingoing inner-outer solution:

$$
\begin{aligned}
&\|\theta_I(t)-\theta_{IO}^{\text{in}}(t)\| = O\left(\frac{\varepsilon}{\delta_{\text{in}}^2}\right) \\
&\|w_I(t)-w_{IO}^{\text{in}}(t)\| = O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right) \qquad 0 \leqq t \leqq \hat{\tau} < \tau. \\
&|x_I(t)-x_{IO}^{\text{in}}(t)| = O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right)
\end{aligned} \tag{7.2}
$$

The estimate for the composite expansion now follows from (6.2).

We know that the outer and the ingoing inner-outer solution are equal at $t=\tau$. Using (4.6) and (4.7) we get the following backward estimates:

$$
\begin{aligned}
&\Big(|\theta_{IO}(t)-\theta_{IO}^{\text{in}}(t)| + \|w_O(t)-w_{IO}^{\text{in}}(t)\| \\
&\qquad + \frac{1}{\sqrt{\varepsilon}}|x_0(t)-x_{IO}^{\text{in}}(t)|\Big) = O(\sqrt{\varepsilon}) \quad \text{for } \sqrt{\varepsilon}|t-\tau| \leqq L.
\end{aligned} \tag{7.3}
$$

Using (5.15), (5.16) and (5.17) we have

$$
\begin{aligned}
&|\theta_{j,O}(t)-\theta_{j,IO}^{\text{in}}(t)| = O\left(\frac{\delta_{\text{in}}^5}{\varepsilon^2}\right) \\
&\|w_O(t)-w_{IO}^{\text{in}}(t)\| = O(\delta_{\text{in}}) \qquad \frac{\varepsilon}{\delta}|t-\tau| \leqq L. \\
&|x_O(t)-x_{IO}^{\text{in}}(t)| = O\left(\frac{\delta_{\text{in}}^6}{\varepsilon^2}\right)
\end{aligned} \tag{7.4}
$$

Combining (7.1), (7.2) and (7.4), we have, at the moment $\hat{\tau}$ of entering the δ_{in}-neighborhood of W_j, the following estimate

$$
\begin{aligned}
&|\theta_j(\hat{\tau})-\theta_{j,I}(\hat{\tau})| = O\left(\frac{\varepsilon}{\delta_{\text{in}}^2}\right) + O\left(\frac{\delta_{\text{in}}^5}{\varepsilon^2}\right) \\
&\|w(\hat{\tau})-w_I(\hat{\tau})\| = O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right) + O(\delta_{\text{in}}) \\
&|x(\hat{\tau})-x_I(\hat{\tau})| = O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right) + O\left(\frac{\delta_{\text{in}}^6}{\varepsilon^2}\right).
\end{aligned} \tag{7.5}
$$

Using (5.14), (5.16) and (5.17) gives

$$
\begin{aligned}
&|\theta_j(t)-\theta_{j,I}(t)| = O\left(\frac{\delta_{\text{in}}^5}{\varepsilon^2}\right) + O\left(\frac{\varepsilon}{\delta_{\text{in}}^2}\right) \\
&\|w(t)-w_I(t)\| = O(\delta_{\text{in}}) + O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right) \\
&|x(t)-x_I(t)| = O\left(\frac{\delta_{\text{in}}^6}{\varepsilon^2}\right) + O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right)
\end{aligned} \tag{7.6}
$$

until the solution reaches the $\sqrt{\varepsilon}$-neighborhood of $\mathcal{N}_j$; here the estimates become, using (4.6) and (4.7):

$$\begin{aligned}
&|\theta_j(t)-\theta_{j,I}(t)| = O\left(\frac{\sqrt{\varepsilon}}{\delta_{\text{in}}}\right)+O\left(\frac{\delta_{\text{in}}^6}{\varepsilon^2\sqrt{\varepsilon}}\right)\\
&\|w(t)-w_I(t)\| = O(\delta_{\text{in}})+O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right)\\
&|x(t)-x_I(t)| = O\left(\frac{\varepsilon}{\delta_{\text{in}}}\right)+O\left(\frac{\delta_{\text{in}}^6}{\varepsilon^2}\right).
\end{aligned} \tag{7.7}$$

We can do the same thing for the difference between the outer solution and the inner-outer solution, but that has already been estimated in (7.4).

At this point we choose δ_{in}; in fact we can take different δ_{in}'s for each of the components, since we work with a composite expansion.

Thus we get

$$\begin{aligned}
&|\theta_j(t)-\theta_{j,C}(t)| = O(\varepsilon^{1/14})\\
&\|w(t)-w_C(t)\| = O\left(\frac{\sqrt{\varepsilon}}{\eta(\varepsilon)}\right), \quad \text{with } \eta(\varepsilon)=o(1) \text{ as } \varepsilon\downarrow 0,\\
&|x(t)-x_C(t)| = O(\varepsilon^{4/7}).
\end{aligned} \tag{7.8}$$

Again using (5.15), (5.16) and (5.17) yields:

$$\begin{aligned}
&|\theta_j(t)-\theta_{j,I}(t)| = O\left(\frac{\delta_{\text{out}}^5}{\varepsilon^2}\right)+O(\varepsilon^{1/14})\\
&\|w(t)-W_I(t)\| = O(\delta_{\text{out}})+O\left(\frac{\sqrt{\varepsilon}}{\eta(\varepsilon)}\right)\\
&|x(t)-x_I(t)| = O(\varepsilon^{4/7})+O\left(\frac{\delta_{\text{out}}^6}{\varepsilon^2}\right)+O\left(\frac{\delta_{\text{out}}^2}{\sqrt{\varepsilon}}\varepsilon^{1/14}\right).
\end{aligned} \tag{7.9}$$

The difference between the inner solution and the outgoing inner-outer solution is in the outer domain estimated by using (3.18)

$$\begin{aligned}
&|\theta_{j,I}(t)-\theta_{j,IO}^{\text{out}}(t)| = O\left(\frac{\varepsilon}{\delta_{\text{out}}^2}\right)\\
&\|w_I(t)-w_{IO}^{\text{out}}(t)\| = O\left(\frac{\varepsilon}{\delta_{\text{out}}}\right)\\
&|x_I(t)-x_{IO}^{\text{out}}(t)| = O\left(\frac{\varepsilon}{\delta_{\text{out}}}\right).
\end{aligned} \tag{7.10}$$

In the inner domain the difference between the outgoing outer solution and the outgoing inner-outer solution is given by (4.6) and (5.15), (5.16) and (5.17):

$$\begin{aligned}
&|\theta_{j,O}^{\text{out}}(t)-\theta_{j,IO}^{\text{out}}(t)| = O\left(\frac{\delta_{\text{out}}^5}{\varepsilon^2}\right)\\
&\|w_O^{\text{out}}(t)-w_{IO}^{\text{out}}(t)\| = O(\delta_{\text{out}})\\
&|x_O^{\text{out}}(t)-x_{IO}^{\text{out}}(t)| = O\left(\frac{\delta_{\text{out}}^6}{\varepsilon^2}\right).
\end{aligned} \tag{7.11}$$

Therefore if the solution leaves the δ_{out}-neighborhood of $\mathcal{N}_j$, one has at this moment $\tilde{\tau}$ (combining (7.9), (7.10) and (7.11)):

$$
\begin{aligned}
|\theta_j(\tilde{\tau})-\theta_{j,O}^{\text{out}}(\tilde{\tau})| &= O\left(\frac{\delta_{\text{out}}^5}{\varepsilon^2}\right)+O\left(\frac{\varepsilon}{\delta_{\text{out}}^2}\right)+O(\varepsilon^{1/14})\\
\|w(\tilde{\tau})-w_O^{\text{out}}(\tilde{\tau})\| &= O(\delta_{\text{out}})+O\left(\frac{\varepsilon}{\delta_{\text{out}}}\right)+O\left(\frac{\sqrt{\varepsilon}}{\eta(\varepsilon)}\right)\\
|x(\tilde{\tau})-x_O^{\text{out}}(\tilde{\tau})| &= O\left(\frac{\delta_{\text{out}}^6}{\varepsilon^2}\right)+O\left(\frac{\varepsilon}{\delta_{\text{out}}}\right)+O(\varepsilon^{4/7})+O\left(\frac{\delta_{\text{out}}^2}{\sqrt{\varepsilon}}\varepsilon^{1/14}\right).
\end{aligned}
\tag{7.12}
$$

Using (3.16), (3.17) and (3.13), we have:

$$
\begin{aligned}
|\theta_j(t)-\theta_{j,O}^{\text{out}}(t)| &= O\left(\frac{\delta_{\text{out}}^4}{\varepsilon^2}\right)+O\left(\frac{\varepsilon}{\delta_{\text{out}}^3}\right)+O\left(\frac{\varepsilon^{1/14}}{\delta_{\text{out}}}\right)+\cdots\\
\|w(t)-w_O^{\text{out}}(t)\| &= O(\delta_{\text{out}})+O\left(\frac{\varepsilon}{\delta_{\text{out}}}\right)+O\left(\frac{\sqrt{\varepsilon}}{\eta(\varepsilon)}\right)\\
&\quad +O\left(\frac{\delta_{\text{out}}^6}{\varepsilon^2}\right)+O(\varepsilon^{4/7})+O\left(\frac{\delta_{\text{out}}^2}{\sqrt{\varepsilon}}\varepsilon^{1/14}\right)\\
|x(t)-x_O^{\text{out}}(t)| &= O\left(\frac{\varepsilon}{\delta_{\text{out}}}\right)+O\left(\frac{\delta_{\text{out}}^6}{\varepsilon^2}\right)+O(\varepsilon^{4/7})+O\left(\frac{\delta_{\text{out}}^2}{\sqrt{\varepsilon}}\varepsilon^{1/14}\right).
\end{aligned}
\tag{7.13}
$$

Combining (7.10) and (7.13) and choosing δ_{out} optimally:

$$
\begin{aligned}
\|w(t)-w_C(t)\| &= O\left(\frac{\sqrt{\varepsilon}}{\eta(\varepsilon)}\right) \quad \text{for any } \eta \text{ s.t. } \eta(\varepsilon)=o(1) \text{ as } \varepsilon \downarrow 0\\
|x(t)-x_C(t)| &= O(\varepsilon^{11/21}).
\end{aligned}
\tag{7.14}
$$

This procedure does not give a nice estimate for the angular variable(s), although there seems to be no fundamental reason against this.

Here we conclude the more theoretical part of this paper and turn to the application promised in the Introduction.

Remark. More experienced readers may wonder what happened to the more classical methods like matched asymptotic expansions; those readers are strongly encouraged to provide (as an "exercise") approximations and estimates of validity along those classical lines.

The reason for the approach given here is that it does not depend upon explicit realizations of solutions (to be expanded in the matching process). Furthermore, the estimates for the difference between the inner, respectively outer, solution and the inner-outer solutions were already obtained in earlier sections. One of the difficulties in the matched asymptotic expansion approach is that one needs an expansion of the inner solution, which is not explicitly known. A glance at Kevorkian's 1974 paper may convince the reader that it is indeed possible to overcome this difficulty, if only he is willing to pay a price in mathematical elegance.

8. The model problem. In order to illustrate the theory developed in the preceding sections, we look at a problem of reentry roll resonance, studied by Kevorkian (1974). In this paper, Kevorkian gives a formal treatment of the resonance problem with multiple-time-scale expansions. Lacking proof techniques, this was a most difficult thing to do, since he had to match terms with small divisors.

The reader is advised to study this paper for two reasons:

1°. It gives a lot of background information and references on the problem of reentry roll resonance.

2°. It is a nice example of a complete treatment of a problem based on good intuition.

The only thing Kevorkian doesn't show is that the phenomenon of sustained resonance (on a time-scale $1/\varepsilon$) is exhibited by the model problem.

Having introduced the basic equations, Kevorkian restricts his attention to the following model problem:

$$(8.1)\qquad \begin{aligned} &\ddot{y}+(p^2+\omega^2)y=0 && y(0)=0\\ & && \dot{y}(0)=\chi_0\\ &\dot{p}=\varepsilon\omega^2 y\sin\psi && p(0)=p_0\\ &\dot{\omega}=\tfrac{1}{2}\varepsilon\omega && \omega(0)=\omega_0\\ &\dot{\psi}=\sqrt{2}p && \psi(0)=0\\ & && \alpha=\frac{\omega_0}{p_0}<1. \end{aligned}$$

We extend this system with

$$(8.2)\qquad \dot{\xi}=(p^2+\omega^2)^{1/2}\qquad \xi(0)=0.$$

These equations are not in the standard form (2.1); therefore we make the following transformation:

$$(8.3)\qquad \begin{aligned} &y=a\sin(\xi+\phi) && \phi(0)=0\\ &\dot{y}=a(p^2+\omega^2)^{1/2}\cos(\xi+\phi) && a(0)=\chi_0 p_0(1+\alpha^2)^{1/2}. \end{aligned}$$

(The $(y, \dot{y})$-notation is a bit sloppy, but one gets tired introducing new symbols all the time.)

The induced equations are:

$$(8.4)\qquad \begin{aligned} &\dot{\psi}=\sqrt{2}p\\ &\dot{\xi}=(p^2+\omega^2)^{1/2}\\ &\dot{a}=-\varepsilon a(p^2+\omega^2)^{-1}\omega^2\cos^2(\xi+\phi)\{\tfrac{1}{2}+ap\sin(\xi+\phi)\sin\psi\}\\ &\dot{\phi}=\varepsilon\omega^2(p^2+\omega^2)^{-1}\sin(\xi+\phi)\cos(\xi+\phi)\{\tfrac{1}{2}+ap\sin(\xi+\phi)\sin\psi\}\\ &\dot{p}=\varepsilon\omega^2 a\sin(\xi+\phi)\sin\psi\\ &\dot{\omega}=\tfrac{1}{2}\varepsilon\omega. \end{aligned}$$

There are two fast variables ψ and ξ and four slow ones: a, ϕ, p and ω, where ω is the only one that doesn't depend on the others.

Equation (8.4) is still not in standard form; to put (8.4) in standard form we write the right hand side in the form of finite Fourier-series and then introduce the combination angles:

(8.5)
$$\begin{aligned}
\dot\psi &= \sqrt{2}p\\
\dot\xi &= (p^2+\omega^2)^{1/2}\\
\dot a &= -\varepsilon a\omega^2(p^2+\omega^2)^{-1}\{\tfrac14(1+2\cos 2(\xi+\phi))+\tfrac18 ap[\cos((3\xi-\psi)+3\phi)\\
&\qquad -\cos((3\xi+\psi)+\phi)+\cos((\xi-\psi)+\phi)-\cos((\xi+\psi)+\phi)]\}\\
\dot\phi &= \varepsilon\omega^2(p^2+\omega^2)^{-1}\{\tfrac14\sin 2(\xi+\phi)+\tfrac18 ap[\sin((\xi+\psi)+\phi)\\
&\qquad -\sin((\xi-\psi)+\phi)-\sin((3\xi+\psi)+3\phi)+\sin((3\xi-\psi)+3\phi)]\}\\
\dot p &= \tfrac12\varepsilon\omega^2 a(\cos((\xi-\psi)+\phi)-\cos((\xi+\psi)+\phi))\\
\dot\omega &= \tfrac12\varepsilon\omega.
\end{aligned}$$

Now define

(8.6)
$$\begin{aligned}
\theta_1 &= \xi & w_1 &= a\\
\theta_2 &= 3\xi-\psi & w_2 &= \phi\\
\theta_3 &= 3\xi+\psi & x &= p\omega^{-1}\\
\theta_4 &= \xi-\psi & u &= \omega\\
\theta_5 &= \xi+\psi. & &
\end{aligned}$$

From (8.5) it follows that

(8.7)
$$\begin{aligned}
\dot\theta_1 &= u(1+x^2)^{1/2}\\
\dot\theta_2 &= u(3(1+x^2)^{1/2}-\sqrt{2}x)\\
\dot\theta_3 &= u(3(1+x^2)^{1/2}+\sqrt{2}x)\\
\dot\theta_4 &= u((1+x^2)^{1/2}-\sqrt{2}x)\\
\dot\theta_5 &= u((1+x^2)^{1/2}+\sqrt{2}x)\\
\dot w_1 &= -\varepsilon w_1(1+x^2)^{-1}\{\tfrac14(1+\cos(2\theta_1+2w_2))+\tfrac18 uw_1x[\cos(\theta_2+3w_2)\\
&\qquad -\cos(\theta_3+3w_2)+\cos(\theta_4+w_2)-\cos(\theta_5+w_2)]\}\\
\dot w_2 &= \varepsilon(1+x^2)^{-1}\{\tfrac14\sin(2\theta_1+2w_2)+\tfrac18 uw_1x[\sin(\theta_2+3w_2)\\
&\qquad -\sin(\theta_3+3w_2)-\sin(\theta_4+w_2)+\sin(\theta_5+w_2)]\}\\
\dot x &= \tfrac12\varepsilon\{uw_1[\cos(\theta_4+w_2)-\cos(\theta_5+w_2)]-x\}\\
\dot u &= \tfrac12\varepsilon u.
\end{aligned}$$

Equation (8.7) is in standard form. The initial conditions for (8.7) are:

(8.8)
$$\begin{aligned}
\theta(0) &= 0\\
w_1(0) &= \chi_0 p_0(1+\alpha^2)^{1/2}\\
w_2(0) &= 0\\
x(0) &= 1/\alpha\\
u(0) &= \omega_0.
\end{aligned}$$

Using the notation of § 2, one has

$$\Omega_i(u) = u, \qquad i = 1, \cdots, 5,$$

$$X^{\#}(x) = \begin{pmatrix} (1+x^2)^{1/2} \\ 3(1+x^2)^{1/2} - \sqrt{2}x \\ 3(1+x^2)^{1/2} + \sqrt{2}x \\ (1+x^2)^{1/2} - \sqrt{2}x \\ (1+x^2)^{1/2} + \sqrt{2}x \end{pmatrix},$$

$$Y_0^0(w, x) = \begin{pmatrix} -\frac{1}{4}w_1(1+x^2)^{-1} \\ 0 \end{pmatrix}, \qquad Z_0^0(x) = -\frac{1}{2}x,$$

$$Y_0^1(\theta_1, w, x) = \frac{1}{4(1+x^2)} \begin{pmatrix} -w_1 \cos(2\theta_1 + 2w_2) \\ \sin(2\theta_1 + 2w_2) \end{pmatrix}, \qquad Z_0^1 = 0,$$

(8.9)

$$Y_0^2(\theta_2, w, x, u) = \frac{uw_1 x}{8(1+x^2)} \begin{pmatrix} -w_1 \cos(\theta_2 + 3w_2) \\ \sin(\theta_2 + 3w_2) \end{pmatrix}, \qquad Z_0^2 = 0,$$

$$Y_0^3(\theta_3, w, x, u) = \frac{uw_1 x}{8(1+x^2)} \begin{pmatrix} w_1 \cos(\theta_3 + 3w_2) \\ -\sin(\theta_3 + 3w_2) \end{pmatrix}, \qquad Z_0^3 = 0,$$

$$Y_0^4(\theta_4, w, x, u) = \frac{uw_1 x}{8(1+x^2)} \begin{pmatrix} -w_1 \cos(\theta_4 + w_2) \\ -\sin(\theta_4 - w_2) \end{pmatrix}, \qquad Z_0^4 = \frac{1}{2}uw_1 \cos(\theta_4 + w_2),$$

$$Y_0^5(\theta_5, w, x, u) = \frac{uw_1 x}{8(1+x^2)} \begin{pmatrix} +w_1 \cos(\theta_5 + w_2) \\ \sin(\theta_5 - w_2) \end{pmatrix}, \qquad Z_0^5 = -\frac{1}{2}uw_1 \cos(\theta_5 + w_2).$$

It is clear that on the time-scale of interest, namely $1/\varepsilon$, $\Omega_i(u) = O_S(1)$; the only $X^{\#}$-component which can become zero is $X_4^{\#}$, where one has $z_0^4 = 1$. That this is the only resonance is a special feature of the model problem.

The extended system has five more angular variables

(8.10)

$$\begin{aligned} \theta_6 &= 5\xi - \psi \\ \theta_7 &= 5\xi + \psi \\ \theta_8 &= 2\xi - \psi \\ \theta_9 &= 2\xi + \psi \\ \theta_{10} &= \psi \end{aligned}$$

and does not have any new resonances.

We will now construct the outer expansion:

The only thing one has to compute to get the outer expansion is $\bar{Z}_{(1)}^0$. This turns out to be zero, thus satisfying all assumptions on its structure.

The outer equations for the model problem are:

(8.11)

$$\begin{aligned} \dot{\phi}_4 &= (1+z^2)^{1/2} - \sqrt{2}z & \phi_4^{\text{in}}(0) &= 0 \\ \dot{y}_1 &= \tfrac{1}{4}\varepsilon y_1(1+z^2)^{-1} & y_1^{\text{in}}(0) &= \chi_0 p_0(1+\alpha^2)^{1/2} \\ \dot{y}_2 &= 0 & y_2^{\text{in}}(0) &= 0 \\ \dot{z} &= -\tfrac{1}{2}\varepsilon z & z^{\text{in}}(0) &= 1/\alpha \\ \dot{u} &= \tfrac{1}{2}\varepsilon u & u^{\text{in}}(0) &= \omega_0. \end{aligned}$$

We do not write down the equations for all the other angular variables.

We compute the ingoing outer solution explicitly:

$$\theta_{4O}^{\text{in}}(t) = (1-\sqrt{2})p_0 t - \frac{2}{\varepsilon}p_0\left\{\sqrt{1+\alpha^2} - \sqrt{1+\alpha^2 e^{\varepsilon t}} + \log\left(\frac{1+\sqrt{1+\alpha^2 e^{\varepsilon t}}}{1+\sqrt{1+\alpha^2}}\right)\right\}$$

$$w_{1O}^{\text{in}}(t) = \chi_0 p_0 (1+\alpha)^{1/2}\left(\frac{1+\alpha^2}{1+\alpha^2 e^{\varepsilon t}}\right)^{1/4}$$

(8.12)
$$w_{2O}^{\text{in}}(t) = 0$$

$$x_O^{\text{in}}(t) = \frac{1}{\alpha}e^{-(1/2)\varepsilon t}, \qquad u(t) = \omega_0 e^{(1/2)\varepsilon t}$$

where τ is given by the equation

(8.13)
$$x_O^{\text{in}}(\tau) = 0$$

and thus

(8.14)
$$\varepsilon\tau = \log\frac{1}{\alpha^2}.$$

One has to check assumption (3.7). It suffices to show that:

(8.15)
$$\int_0^t \frac{\varepsilon}{X_4^n(x_O^{\text{in}}(\tau), u(\tau))}\, d\tau \leqq C\left(\frac{1}{X_4^{n-1}(x_O^{\text{in}}(t), u(t))} + \frac{1}{X_4^{n-1}(x_O^{\text{in}}(0), u_0)}\right).$$

The following estimate holds on $[1, \infty)$:

(8.16)
$$\frac{1}{\sqrt{2}}(1+x) \geqq \sqrt{2}x - (1+x^2)^{1/2} \geqq (\sqrt{2}-1)(x-1).$$

Using (8.12) and (8.15), it is not difficult to show that (8.15) holds.
For the outgoing solution one can use the estimate:

(8.17)
$$1+x \geqq (1+x^2)^{1/2} - \sqrt{2}x \geqq \frac{1}{\sqrt{2}}(1-x) \qquad x \in [0, 1].$$

In the original coordinates for the model problem one has

(8.18)
$$y_O^{\text{in}}(t) = \chi_0 p_0 (1+\alpha^2)^{1/2}\left(\frac{1+\alpha^2}{1+\alpha^2 e^{\varepsilon t}}\right)^{1/4} \sin p_0\left(t - \frac{2}{\varepsilon}\left\{\sqrt{1+\alpha^2} - \sqrt{1+\alpha^2 e^{\varepsilon t}} + \log\left(\frac{1+\sqrt{1+\alpha^2 e^{\varepsilon t}}}{1+\sqrt{1+\alpha^2}}\right)\right\}\right).$$

We will now turn our attention to the inner expansion.
The inner equations are:

(8.19)
$$\begin{aligned}
\dot{\phi}_4 &= -\frac{\sqrt{\varepsilon}}{\sqrt{2}}\zeta u(\tau) \\
\dot{y} &= 0 \\
\dot{\zeta} &= \tfrac{1}{2}\sqrt{\varepsilon}(-1 + u(\tau) w_{1O}^{\text{in}}(\tau)\cos\phi_4) \\
\dot{u} &= \tfrac{1}{2}\varepsilon u,
\end{aligned}$$

where

(8.20)
$$\begin{aligned}
u(\tau) &= p_0 \\
w_{1O}^{\text{in}}(\tau) &= \chi_0 p_0 (1+\alpha^2)^{1/2}\left(\frac{1+\alpha^2}{1+\alpha}\right)^{1/4}.
\end{aligned}$$

Thus one has

$$\ddot{\phi}_4+\frac{\varepsilon}{2\sqrt{2}}u(\tau)(u(\tau)w_{1O}^{\text{in}}(\tau)\cos\phi_4-1)=0. \tag{8.21}$$

There are critical orbits (in the original coordinates) iff $u(\tau)w_{1O}^{\text{in}}(\tau)>1$. In that case one is elliptic, one hyperbolic. Since the solution of (8.19) is not an elementary function of time, one has to be content with numerical results.

Having done the numerical work, one may proceed with constructing the outgoing solution, as described in § 6.

The difficulty here is that one has to know the asymptotic behavior of the inner solution and thus the numerical integration has to be done on a time scale that is $1/\sqrt{\varepsilon}$ times the natural time-scale of the inner equation. The same difficulty occurs when one starts integrating since the initial conditions are given for outside the resonance manifold. We are now going to compute these initial conditions:

The inner-outer equations are given by

$$\begin{aligned}\dot{\phi}_4&=-\frac{\sqrt{\varepsilon}}{\sqrt{2}}\zeta\\ \dot{y}&=0\\ \dot{\zeta}&=-\tfrac{1}{2}\sqrt{\varepsilon}\\ \dot{u}&=\tfrac{1}{2}\varepsilon u.\end{aligned} \tag{8.22}$$

The initial conditions have to be determined from the value of the outer expansion at τ. Thus:

$$\begin{aligned}\theta_{4IO}^{\text{in}}(\tau)&=(1-\sqrt{2})\frac{2p_0}{\varepsilon}\log\frac{1}{\alpha}-\frac{2p_0}{\varepsilon}\left\{\sqrt{1+\alpha^2}-\sqrt{2}+\log\left(\frac{1+\sqrt{2}}{1+\sqrt{1+\alpha^2}}\right)\right\}\\ w_{1IO}^{\text{in}}(\tau)&=\chi_0p_0(1+\alpha^2)\left(\frac{1+\alpha^2}{2}\right)^{1/4}\\ w_{2IO}^{\text{in}}(\tau)&=0\\ \zeta_{IO}^{\text{in}}(\tau)&=0\\ u(\tau)&=p_0.\end{aligned} \tag{8.23}$$

Integrating backwards, we get

$$\begin{aligned}\zeta_{IO}^{\text{in}}(t)&=-\tfrac{1}{2}\sqrt{\varepsilon}(t-\tau) &&\Rightarrow \zeta_I(0)=\tfrac{1}{2}\sqrt{\varepsilon}\tau\\ & &&\Rightarrow x_I(0)=1+\log\frac{1}{\alpha}\\ w_{1IO}^{\text{in}}(t)&=\chi_0p_0(1+\alpha^2)^{1/2}\left(\frac{1+\alpha^2}{2}\right)^{1/4} &&\Rightarrow w_{1I}(0)=\chi_0p_0(1+\alpha_2)\left(\frac{1+\alpha^2}{2}\right)^{1/4}\\ w_{2IO}^{\text{in}}(t)&=0 &&\Rightarrow w_{2I}(0)=0\end{aligned} \tag{8.24}$$

$$\theta_{4IO}^{\text{in}}(t)=(1-\sqrt{2})p_0\tau-\frac{2p_0}{\varepsilon}\left\{\sqrt{1+\alpha^2}-\sqrt{2}+\log\left(\frac{1+\sqrt{2}}{1+\sqrt{1+\alpha^2}}\right)\right\}-\frac{p_0}{2\sqrt{2}}\varepsilon(t-\tau)^2,$$

which implies

$$\theta_{4I}(0)=(1-\sqrt{2})p_0\tau-\frac{2p_0}{\varepsilon}\left\{\sqrt{1+\alpha^2}-\sqrt{2}+\log\left(\frac{1+\sqrt{2}}{1+\sqrt{1+\alpha^2}}\right)\right\}-\frac{p_0\varepsilon}{2\sqrt{2}}\tau^2. \tag{8.25}$$

This is as far as one can carry through the analytic treatment if one is interested in the time-behavior of the solution. If one is interested in the orbits one may go further, using the fact that the inner vector field has an integral with a rather easy asymptotic behavior. But this analysis should not be trusted too much, since the validity of the approximation is only proven on the natural time-scale of the inner expansion. Then it may happen that the "approximating" orbit goes out of resonance, while the exact solution stays in resonance.

Comparing our results with the numerical work as represented in Figs. 1, 2 and 3 in Kevorkian (1974) we can make the following remarks.

The theory applies to the phenomena in Fig. 1 and Fig. 3, but not to Fig. 2.

In Fig. 2 we see that $p\sim\omega$ on a time-scale $1/\varepsilon$, with oscillations of period $1/\sqrt{\varepsilon}$.

Since our estimates are valid on $1/\sqrt{\varepsilon}$, the theory does not apply to this picture.

The period of the oscillations of p around ω suggests that the solution is oscillating around the elliptic orbit (which may be slowly moving on $1/\varepsilon$) and does not stay near the hyperbolic one, as Kevorkian claims (cf. however Lewin and Kevorkian (1978)).

Anyway, it is not difficult to see that the solution cannot stay near the hyperbolic on $1/\varepsilon$ in the model problem, using the equations for the slow variables in the inner expansion.

In conclusion we state that the problem of finding approximations valid on a time-scale $1/\varepsilon$ and describing Fig. 2 in Kevorkian (1974) is still open.

REFERENCES

W. ECKHAUS (1976), Private communication.

J. KEVORKIAN (1974), *On a model for reentry roll resonance*, SIAM J. Appl. Math., 26, pp. 638–669.

W. T. KYNER (1969), *Passage through resonance*; *Periodic Orbits, Stability and Resonances*, G. E. O. Giacaglia, ed., Reidel Publishing Company, Dordrecht, Holland, 1970.

L. LEWIN AND J. KEVORKIAN (1978), *On the problem of sustained resonance*, SIAM J. Appl. Math., 35, pp. 738–754.

J. A. SANDERS (1977), *Ai higher order resonances really interesting*? Celestial Mech., 16 (1978), pp. 421–440.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0012 $01.00/0

A REMARK ON A PAPER BY PAYNE AND PHILLIPPIN*

CATHERINE BANDLE†

Abstract. Some isoperimetric inequalities for convex domains are derived and then used to compare different bounds for the electrostatic capacity.

1. Introduction. In [2] Payne and Phillippin established a maximum principle for certain combinations of harmonic functions and their derivatives. As an application they derived among others, isoperimetric bounds for the electrostatic capacity. The aim of this note is to show that in some special cases those bounds follow directly from isoperimetric inequalities by Pólya and Szegö [3]. Our arguments are based on simple geometrical inequalities which are contained in the next section.

2. Geometrical inequalities. Let $D \subset \mathbb{R}^3$ be a bounded *convex* domain and denote by $R_1(P)$ and $R_2(P)$ the principal radii of curvature at the point $P \in \partial D$. We set $K = (R_1R_2)^{-1}$ for the Gaussian and $H = \frac{1}{2}(R_1^{-1} + R_2^{-1})$ for the mean curvature. By M we denote the mean curvature integral $\oint_{\partial D} H\,ds$, ds being the area element of ∂D. If h stands for the support function of ∂D and r, $\theta \in [-\pi/2, \pi/2]$, $\varphi \in [0, 2\pi)$ are the polar coordinates, then the volume V of D can be expressed as [1]

$$V = \frac{1}{3}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} \frac{h}{K} \sin\theta \, d\theta \, d\varphi. \tag{1}$$

In addition we have

$$M = \int_0^{2\pi}\int_{-\pi/2}^{\pi/2} h \sin\theta \, d\theta \, d\varphi \tag{2}$$

and for the surface area A of ∂D

$$A = \frac{1}{2}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} h(R_1 + R_2) \sin\theta \, d\theta \, d\varphi. \tag{3}$$

From (1) and (2) it follows that

$$V \geqq \frac{M}{3K_{\max}}. \tag{4}$$

By applying Minkowski's inequality [1], $M^2 \geqq 4\pi A$, we get

$$K_{\max} \geqq \frac{\sqrt{4\pi A}}{3V}. \tag{5}$$

Let R be the radius of the sphere with the same volume as D, that is $(4\pi/3)R^3 = V$. From (5) and the isoperimetric inequality $A \geqq 4\pi R^2$, we conclude that

$$K_{\max} \geqq 1/R^2. \tag{6}$$

Another consequence of (2) and (3) is

$$M \leqq \sqrt{K_{\max}}A. \tag{7}$$

In all inequalities (4)–(7) *the equality sign holds if and only if D is a sphere.*

* Received by the editors November 13, 1978.

† Mathematisches Institut des Universität Basel, CH-4051 Basel, Switzerland.

3. Bounds for the electrostatic capacity. Let $D \subset \mathbb{R}^3$ be a convex bounded domain and n be its inner normal. The exterior capacity of the conductor D is

$$C = \frac{1}{4\pi} \oint_{\partial D} \frac{\partial h}{\partial n}\, ds,$$

where h is the solution of the boundary value problem

$$\Delta h = 0 \quad \text{in } \mathbb{R}^3 - D, \qquad h = 1 \quad \text{on } \partial D \quad \text{and} \quad h = O(|x|^{-1}) \quad \text{as } |x| \to \infty.$$

In [3] it is shown that the following pair of inequalities holds for C:

$$R \leqq C \leqq M/4\pi. \tag{8}$$

On both sides the equality sign is attained if and only if D is a sphere.

In view of (6) we have

$$C \geqq K_{\max}^{-1/2}$$

and by the inequality between the geometric and the arithmetic mean, $K_{\max}^{1/2} \leqq H_{\max}$, we find Payne and Phillippin's estimate [2]

$$C \geqq H_{\max}^{-1}.$$

From (4) and (8) we conclude that

$$4\pi C \leqq 3\, V K_{\max} \leqq 3\, V H_{\max}^2$$

and from (7) and (8) it follows that

$$4\pi C \leqq A K_{\max}^{1/2} \leqq A H_{\max}.$$

These two results have also been derived by Payne and Phillippin [2], but they didn't have to make the assumption that D is convex.

REFERENCES

[1] W. Blaschke, *Kreis und Kugel*, Leipzig, 1916.

[2] L. E. Payne and G. A. Phillippin, *On some maximum principles involving harmonic functions and their derivatives*, SIAM J. Math. Analysis, 10 (1979), pp. 96–104.

[3] G. Pólya and G. Szegö, *Isoperimetric Inequalities in Mathematical Physics*, Princeton, NJ, 1951.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0013 $01.00/0

CONFORMAL MAPPING OF THE DOMAIN EXTERIOR TO A THIN REGION*

DOREL HOMENTCOVSCHI†

Abstract. This paper gives the asymptotic expansion of the function $Z = F(z, \varepsilon)$ which maps conformally the domain exterior to a thin region into the plane Z with a cut on the real axis. The function $F(z, \varepsilon)$ is represented as a superposition of singularities on the segment $[\alpha(\varepsilon), \beta(\varepsilon)]$ inside the thin region.

The singularity intensities $f(x, \varepsilon)$ and $g(x, \varepsilon)$ satisfy a system of linear integral equations. This system is asymptotically integrated by using the method of Handelsman and Keller (*Axially symmetric potential flow around a slender body*, J. Fluid Mech., 28, (1967) pp. 131–142). The explicit solution is constructed for some classes of domains.

1. Introduction. This paper derives the asymptotic development of the conformal mapping function $Z = F(z, \varepsilon)$ of the domain D, exterior to a thin region in the plane z, into the plane Z which has the cut $[X_1, X_2]$ on the real axis.

Once the function $F(z, \varepsilon)$ is determined, a series of boundary-value problems can be resolved for the domain D. Thus, the motion of an inviscid incompressible fluid (having the complex velocity $V \cdot e^{-i\theta}$ at infinity) around a thin profile is characterized by the complex potential [1],

$$F_1(z) = V\cos\theta\, Z - iV\sin\theta\,\{(Z-X_1)(Z-X_2)\}^{1/2} + \frac{\Gamma}{2\pi i}\log\left[Z - \frac{X_1+X_2}{2} + \{(Z-X_1)(Z-X_2)\}^{1/2}\right]$$

where $Z = F(z, \varepsilon)$ and Γ is the circulation about the profile. If the domain D consists in a homogeneous and isotropic dielectric, the thin body is a conducting one and the electrostatic field at infinity is homogeneous, then the corresponding electrostatic complex potential is $F_2(x) = -iF_1(z)$. The function F is also useful in elasticity (e.g. for the study of crack problems), in the examination of ground water flow problems, etc.

The solutions of the hydrodynamic and electrostatic problems mentioned above by methods which use asymptotic expansions similar to those used in this paper were given by Geer and Keller [2] and Geer [3].

The advantage of our way of approach is that given by the knowledge of the conformal mapping function; by means of this function one can solve a series of boundary-value problems for harmonic functions in the domain D, the separate treatment of each problem being no longer necessary.

Here, we shall try to represent the function $F(z, \varepsilon)$ as a superposition of singularities distributed along the segment $[\alpha(\varepsilon), \beta(\varepsilon)]$ of the profile chord. The determination of the intensities $f(x, \varepsilon)$ and $g(x, \varepsilon)$, which characterize the singularities, reduces to the solving of a system of integral equations of Fredholm type and first kind. In order to solve this system we use the method of Handelsman and Keller [4]. As a result we obtain both the asymptotic expansions of the functions f and g and the limits $\alpha(\varepsilon)$ and $\beta(\varepsilon)$ of the segment containing the singularities.

In the case of any thin region the method presented permits one to obtain the first three approximations of the function $F(z, \varepsilon)$. For symmetric regions, the complete asymptotic expansion is given in § 3. When the curve C has equations of the form

$$Y_+(x) + Y_-(x) = P(x); \qquad Y_+(x) - Y_-(x) = \sqrt{1-x^2}\cdot Q(x)$$

* Received by the editors January 27, 1978, and in revised form September 27, 1978.
† Institute of Mathematics, Bucharest, Romania.

where P and Q are polynomials, this paper yields the explicit solution of the system of integral equations. Formulae which permit the explicit calculation of the function $F(z, \varepsilon)$ are also derived.

2. The case of an arbitrary thin region. We consider a thin region whose contour C is described by the equations

$$y = \varepsilon Y_{\pm}(x), \qquad -1 \leqq x \leqq 1, \tag{2.1}$$

where ε is a small parameter. We assume that $Y_+(x) \geqq 0$ and $Y_-(x) \leqq 0$. The equations of the two arcs of the curve C are assumed to be of the form

$$\begin{aligned} Y_+(x) + Y_-(x) &= 2(1-x^2) \cdot S_1(x), \\ Y_+(x) - Y_-(x) &= 2\sqrt{1-x^2} \cdot D_1(x), \qquad -1 \leqq x \leqq 1, \end{aligned} \tag{2.2}$$

where $S_1(x)$ and $D_1(x)$ are twice differentiable functions on the interval $[-1, 1]$. We shall attempt to find the asymptotic expansion, for small values of ε, of the function $Z = F(z, \varepsilon)$, which represents conformally the domain D, external to the curve C, into the plane Z which has a cut $[X_1, X_2]$ on the real axis. We shall norm the function $F(z, \varepsilon)$ by the conditions

$$F(\infty, \varepsilon) = \infty; \qquad \left(\frac{dF}{dz}\right)_{z=\infty} = 1.$$

We represent the function $F(z, \varepsilon)$ as a superposition of singularities placed on the segment $(\alpha(\varepsilon), \beta(\varepsilon))$ of the Ox axis contained inside the curve C. The functions $\alpha(\varepsilon), \beta(\varepsilon)$ assumed of the form

$$\begin{aligned} \alpha(\varepsilon) &= -1 + \alpha_2 \varepsilon^2 + O(\varepsilon^3), \\ \beta(\varepsilon) &= 1 - \beta_2 \varepsilon^2 + O(\varepsilon^3) \end{aligned}$$

will be determined at the same time with resolving of the problem.

We set

$$\begin{aligned} F(z, \varepsilon) = z &+ \frac{1}{\pi} \int_\alpha^\beta \sqrt{(\beta - t)(t - \alpha)} \frac{f(t, \varepsilon)}{t - z} \, dt \\ &+ \sqrt{(z-\alpha)(z-\beta)} \frac{1}{i\pi} \int_\alpha^\beta \frac{g(t, \varepsilon)}{\sqrt{(\beta - t)(t-\alpha)}} \frac{dt}{t - z}, \end{aligned} \tag{2.3}$$

$f(t, \varepsilon)$ and $g(t, \varepsilon)$ being twice differentiable functions on the segment $[\alpha, \beta]$.

On the curve C, the conditions

$$\operatorname{Im}\{F(x + i\varepsilon Y_{\pm}(x), \varepsilon)\} = 0 \tag{2.4}$$

must be fulfilled. These relations lead to the following system of integral equations for determining the functions $f(t, \varepsilon)$ and $g(t, \varepsilon)$

$$\begin{aligned} &\varepsilon Y_{\pm}(x) + \operatorname{Im}\left\{\frac{1}{\pi} \int_\alpha^\beta \sqrt{(\beta-t)(t-\alpha)} \frac{f(t, \varepsilon)}{t - x - i\varepsilon Y_{\pm}} \, dt\right\} \\ &+ \operatorname{Im}\left\{\sqrt{(x + i\varepsilon Y_{\pm} - \alpha)(x + i\varepsilon Y_{\pm} - \beta)} \right. \\ &\qquad \left. \cdot \frac{1}{i\pi} \int_\alpha^\beta \frac{g(t, \varepsilon)}{\sqrt{(\beta - t)(t - \alpha)}} \frac{dt}{t - x - i\varepsilon Y_{\pm}}\right\} = 0. \end{aligned} \tag{2.5}$$

We consider the asymptotic expansions of the integral equations (2.5) for small values of the parameter ε, in which the ε^3-order terms are neglected. To obtain these asymptotic expansions we use relations (A.5) and (A.7) where we set $k=3$. Thus, we get

$$
\begin{aligned}
&\varepsilon Y_\pm(x)+f(x,\varepsilon)\cdot \operatorname{Re}\{\pm H_\pm(x,\varepsilon)-\varepsilon Y_\pm(x)\}+f'(x,\varepsilon)\cdot\operatorname{Re}\{\pm i\varepsilon H_\pm(x,\varepsilon)Y_\pm\}\\
&\quad+\frac{f''(x,\varepsilon)}{2}\operatorname{Re}\{\mp H_\pm(x,\varepsilon)\varepsilon^2 Y_\pm^2(x)\}\\
&\quad+\varepsilon Y_\pm(x)\frac{1}{\pi}\int_\alpha^\beta \sqrt{(\beta-t)(t-\alpha)}\frac{f(t,\varepsilon)-f(x,\varepsilon)}{(t-x)^2}\,dt \qquad (2.6)\\
&\quad+g(x)-\frac{g''(x)}{2}\varepsilon^2Y_\pm^2(x)+\operatorname{Re}\left\{\mp iH_\pm(x,\varepsilon)\frac{1}{\pi}\int_\alpha^\beta\frac{g(t)}{\sqrt{(\beta-t)(t-\alpha)}}\frac{dt}{t-x}\right\}\\
&\quad+\operatorname{Re}\left\{\pm\varepsilon Y_\pm(x)H_\pm(x,\varepsilon)\cdot\frac{1}{\pi}\int_\alpha^\beta\frac{g(t,\varepsilon)-g(x,\varepsilon)}{\sqrt{(\beta-t)(t-\alpha)}}\frac{dt}{(t-x)^2}\right\}+O(\varepsilon^3)=0,
\end{aligned}
$$

the integrals which occur in these equations being considered as principal values in the Cauchy sense.

Here, we have denoted

$$H_\pm(x,\varepsilon)=\sqrt{(\beta-x-i\varepsilon Y_\pm)(x+i\varepsilon Y_\pm-\alpha)},$$

the determination of this function being the positive one for $\alpha<x<\beta$ and $\varepsilon=0$.

We shall develop the functions $H_\pm(x,\varepsilon)$ and all the expressions in which $\alpha(\varepsilon)$ and $\beta(\varepsilon)$ appear into Taylor series. From the sum and difference of the relations thus obtained we find finally the following asymptotic forms of the integral equations (2.5):

$$
\begin{aligned}
&\{-2\varepsilon(1-x^2)S_1(x)+2\varepsilon^2S_1(x)\cdot D_1(x)\}\cdot f(x,\varepsilon)\\
&\quad+4\varepsilon^2x(1-x^2)S_1(x)D_1(x)f'(x,\varepsilon)\\
&\quad-2\varepsilon^2(1-x^2)^2S_1(x)D_1(x)\cdot f''(x,\varepsilon)+2\varepsilon(1-x^2)S_1(x)\\
&\quad\cdot\frac{1}{\pi}\int_{-1}^{+1}\sqrt{1-t^2}\frac{f(t,\varepsilon)-f(x,\varepsilon)}{(t-x)^2}\,dt+2g(x,\varepsilon)\\
&\quad-\varepsilon^2\{(1-x^2)^2S_1^2(x)+(1-x^2)D_1^2(x)\}g''(x,\varepsilon)-2\varepsilon xD_1(x) \qquad (2.7)\\
&\quad\cdot\frac{1}{\pi}\int_{-1}^{+1}\frac{g(t,\varepsilon)}{\sqrt{1-t^2}}\frac{dt}{t-x}\\
&\quad+2\varepsilon(1-x^2)D_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g(t,\varepsilon)-g(x,\varepsilon)}{\sqrt{1-t^2}(t-x)^2}\,dt\\
&\quad+O(\varepsilon^3)=-2\varepsilon(1-x^2)S_1(x);
\end{aligned}
$$

$$
\begin{aligned}
&\Big\{2\sqrt{1-x^2}-2\varepsilon\sqrt{1-x^2}D_1(x)\\
&\quad+\varepsilon^2\frac{(1-x^2)S_1^2(x)+D_1^2(x)+(x-1)\alpha_2-(x+1)\beta_2}{\sqrt{1-x^2}}\Big\}f(x,\varepsilon)\\
&\quad+2\varepsilon^2x\sqrt{1-x^2}\{(1-x^2)S_1^2(x)+D_1^2(x)\}f'(x,\varepsilon)\\
&\quad-\varepsilon^2\sqrt{1-x^2}\{(1-x^2)^2S_1^2(x)+(1-x^2)D_1^2(x)\}f''(x,\varepsilon) \qquad (2.8)
\end{aligned}
$$

$$+2\varepsilon\sqrt{1-x^2}D_1(x)\frac{1}{\pi}\int_{-1}^{+1}\sqrt{(1-t^2)}\frac{f(t,\varepsilon)-f(x,\varepsilon)}{(t-x)^2}\,dt$$
$$-2\varepsilon^2(1-x^2)\sqrt{1-x^2}S_1(x)D_1(x)g''(x,\varepsilon)$$
$$-2\varepsilon x\sqrt{1-x^2}S_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g(t,\varepsilon)}{\sqrt{1-t^2}}\frac{dt}{t-x}$$
$$+2\varepsilon(1-x^2)\sqrt{1-x^2}S_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g(t,\varepsilon)-g(x,\varepsilon)}{\sqrt{1-t^2}}\frac{dt}{(t-x)^2}+O(\varepsilon^3)$$
$$=-2\varepsilon\sqrt{1-x^2}D_1(x);$$

$f'(x,\varepsilon)$, $f''(x,\varepsilon)$, etc. stand for the derivatives with respect to x.

In order to solve equations (2.7) and (2.8) we shall look for the functions $f(x,\varepsilon)$ and $g(x,\varepsilon)$ in the form

$$f(x,\varepsilon)=\sum_{j=1}^{3} f_j(x)\varepsilon^j+O(\varepsilon^4),\qquad g(x,\varepsilon)=\sum_{j=1}^{3} g_1(x)\varepsilon^j+O(\varepsilon^4), \tag{2.9}$$

where the functions $f_j(x)$ and $g_j(x)$ have to be determined. Inserting (2.9) into (2.7) and (2.8) and equating coefficients of ε on the two sides of these equations we obtain

$$f_1(x)=-D_1(x), \tag{2.10}$$

$$g_1(x)=-(1-x^2)S_1(x), \tag{2.11}$$

$$f_2(x)=D_1(x)f_1(x)-D_1(x)\cdot\frac{1}{\pi}\int_{-1}^{+1}\sqrt{1-t^2}\frac{f_1(t)-f_1(x)}{(t-x)^2}\,dt$$
$$+xS_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g_1(t)}{\sqrt{1-t^2}}\frac{dt}{t-x} \tag{2.12}$$
$$-(1-x^2)S_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g_1(t)-g_1(x)}{\sqrt{1-t^2}}\frac{dt}{(t-x)^2};$$

$$g_2(x)=(1-x^2)S_1(x)f_1(x)-(1-x^2)S_1(x)\frac{1}{\pi}\int_{-1}^{+1}\sqrt{1-t^2}\frac{f_1(t)-f_1(x)}{(t-x)^2}\,dt$$
$$+xD_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g_1(t)}{\sqrt{1-t^2}}\frac{dt}{t-x} \tag{2.13}$$
$$-(1-x^2)D_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g_1(t)-g_1(x)}{\sqrt{1-t^2}}\frac{dt}{(t-x)^2};$$

$$f_3(x)=\frac{\beta_2(x+1)-\alpha_2(x-1)-(1-x^2)S_1^2(x)-D_1^2(x)}{2(1-x^2)}f_1(x)$$
$$-x\{(1-x^2)S_1^2(x)+D_1^2(x)\}f_1'(x)$$
$$+\{(1-x^2)^2S_1^2(x)+(1-x^2)D_1^2(x)\}\frac{f_1''(x)}{2}$$
$$+(1-x^2)S_1(x)D_1(x)g_1''(x)+D_1(x)f_2(x) \tag{2.14}$$

$$-D_1(x)\frac{1}{\pi}\int_{-1}^{+1}\sqrt{1-t^2}\frac{f_2(t)-f_2(x)}{(t-x)^2}\,dt$$
$$+xS_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g_2(t)}{\sqrt{1-t^2}}\frac{dt}{t-x}-(1-x^2)S_1(x)$$
$$\cdot\frac{1}{\pi}\int_{-1}^{+1}\frac{1}{\sqrt{1-t^2}}\frac{g_2(t)-g_2(x)}{(t-x)^2}\,dt;$$

$$\begin{aligned}g_3(x)= &-S_1(x)D_1(x)f_1(x)-2x(1-x^2)S_1(x)D_1(x)f_1'(x)\\ &+(1-x^2)^2S_1(x)D_1(x)f_1''(x)\\ &+\{(1-x^2)^2S_1^2(x)+(1-x^2)D_1^2(x)\}\frac{g_1''(x)}{2}+(1-x^2)S_1(x)f_2(x)\\ &-(1-x^2)S_1(x)\frac{1}{\pi}\int_{-1}^{+1}\sqrt{1-t^2}\frac{f_2(t)-f_2(x)}{(t-x)^2}\,dt\\ &+xD_1(x)\frac{1}{\pi}\int_{-1}^{+1}\frac{g_2(t)}{\sqrt{1-t^2}}\frac{dt}{t-x}\\ &-(1-x^2)D_1(x)\cdot\frac{1}{\pi}\int_{-1}^{+1}\frac{1}{\sqrt{1-t^2}}\frac{g_2(t)-g_2(x)}{(t-x)^2}\,dt.\end{aligned}\tag{2.15}$$

The functions f_1, f_2, g_1, g_2, g_3 are bounded for $x\in[-1,+1]$. We shall determine the constants α_2 and β_2 such that the function $f_3(x)$ is integrable on the interval $[-1, 1]$. From (2.14) we have

$$\begin{aligned}\beta_2&=\tfrac{1}{2}D_1^2(1),\\ \alpha_2&=\tfrac{1}{2}D_1^2(-1).\end{aligned}\tag{2.16}$$

Let $F_3(z,\varepsilon)$ be the conformal mapping function obtained by using the functions $f(t,\varepsilon)$ and $g(t,\varepsilon)$ determined above. From relation (2.3) we have

$$F(z,\varepsilon)-F_3(z,\varepsilon)=O(\varepsilon^4)$$

such that the function $F_3(z,\varepsilon)$ represents an uniform approximation of the function $F(z,\varepsilon)$ in the domain exterior to the curve C.

Further on, we shall calculate the explicit solution of the problem in the case of equations of the curve C expressed by polynomial functions.

Let

$$\begin{aligned}D_1(x)&=\sum_{k=0}^{N}d_kx^k,\\ S_1(x)&=\sum_{k=0}^{M}s_kx^k.\end{aligned}\tag{2.17}$$

From relations (2.10) and (2.11) we have

$$\begin{aligned}f_1(x)&=-\sum_{k=0}^{N}d_kx^k\equiv\sum_{k=0}^{N_1}f_{1,k}x^k \qquad (N_1=N),\\ g_1(x)&=-\sum_{k=0}^{M+2}(s_k-s_{k-2})x^k\equiv\sum_{k=0}^{M_1}g_{1,k}x^k \qquad (M_1=M+2)\end{aligned}\tag{2.18}$$

where $s_k=0$ for $k<0$ and $k>M$.

Similarly, we set $d_k = 0$ for $k < 0$ and $k > M$, $f_{1,k} = 0$ if $k < 0$ and $k > N_1$, $g_{1,k} = 0$ for $k < 0$ and $k > M_1$, etc.

For the functions f_2 and g_2, after some calculation from (2.12) and (2.13), we obtain

$$
\begin{aligned}
f_2(x) = &-\sum_{p=0}^{N+N_1} x^p \sum_{j=0}^{p} (j+1) d_{p-j} \sum_{k=j}^{N_1}{}' \frac{(k-j-3)!!}{(k-j)!!} f_{1,k} \\
&+ \sum_{p=0}^{M+M_1} x^p \left\{ \sum_{j=1}^{p} s_{p-j} \sum_{k=j}^{M_1}{}' \frac{(k-j-1)!!}{(k-j)!!} g_{1,k} \right. \\
&\left. - \sum_{j=0}^{p} (s_{p-j} - s_{p-j-2}) \cdot (j+1) \sum_{k=j+2}^{M_1}{}' \frac{(k-j-3)!!}{(k-j-2)!!} g_{1,k} \right\};
\end{aligned} \tag{2.19}
$$

$$
\begin{aligned}
g_2(x) = &-\sum_{p=0}^{M+N_1+2} x^p \sum_{j=0}^{p} (j+1)(s_{p-j} - s_{p-j-2}) \sum_{k=j}^{N_1}{}' \frac{(k-j-3)!!}{(k-j)!!} f_{1,k} \\
&+ \sum_{p=1}^{N+M_1} x^p \left\{ \sum_{j=1}^{p} d_{p-j} \sum_{k=j}^{M_1}{}' \frac{(k-j-1)!!}{(k-j)!!} g_{1,k} \right. \\
&\left. - \sum_{j=0}^{p} (j+1)(d_{p-j} - d_{p-j-2}) \sum_{k=j+2}^{M_1}{}' \frac{(k-j-3)!!}{(k-j-2)!!} g_{1,k} \right\}.
\end{aligned} \tag{2.20}
$$

$\sum_{k=j}' a_k$ stands for the sum $a_j + a_{j+2} + a_{j+4} + \cdots$.

Thus, we can write

$$f_2(x) = \sum_{k=0}^{N_2} f_{2,k} x^k, \tag{2.19'}$$

$$g_2(x) = \sum_{k=0}^{M_2} g_{2,k} x^k \tag{2.20'}$$

where $N_2 = \max\{N + N_1; M + M_1\}$, $M_2 = \max\{M + N_1 + 2; N + M_1\}$ and the coefficients $f_{2,k}$ and $g_{2,k}$ are obtained by matching (2.19)', (2.20)' with expressions (2.19) and (2.20), respectively. Further on the functions $f_3(x)$ and $g_3(x)$ result from relations (2.14) and (2.15) in the form:

$$
\begin{aligned}
f_3(x) = &\frac{1}{2} \sum_{p=0}^{2N+N_1-2} x^p \sum_{j+l=p} b_l f_{1,j} - \frac{1}{2} \sum_{p=0}^{2M+N_1} x^p \sum\sum_{j+l+k=p} s_j s_l f_{1,k} \\
&- \sum_{p=0}^{2M+N_1+2} x^p \sum\sum_{j+l+k=p} k(s_j - s_{j-2}) s_l f_{1,k} - \sum_{p=0}^{2N+N_1} x^p \sum\sum_{j+l+k=p} k d_j d_l f_{1,k} \\
&+ \frac{1}{2} \sum_{p=0}^{2M+N_1+2} x^p \sum\sum_{j+l+k=p} (s_j - s_{j-2})(s_l - s_{l-2})(k+2)(k+1) f_{1,k+2} \\
&+ \frac{1}{2} \sum_{p=0}^{2N+N_1} x^p \sum\sum_{j+l+k=p} (d_j - d_{j-2}) d_l (k+2)(k+1) f_{1,k+2} \\
&+ \sum_{p=0}^{M+M_1+N} x^p \sum\sum_{j+l+k=p} (s_j - s_{j-2}) d_l (k+2)(k+1) g_{1,k+2} + \hat{f}_3(x);
\end{aligned} \tag{2.21}
$$

$$
\begin{aligned}
g_3(x) = &- \sum_{p=0}^{M+N+N_1} x^p \sum\sum_{j+l+k=p} s_j d_l f_{1,k} - 2 \sum_{p=0}^{M+N+N_1+2} x^p \sum\sum_{j+l+k=p} k \\
&\cdot (s_j - s_{j-2}) d_l f_{1,k} + \sum_{p=0}^{M+N+N_1+2} x^p \sum\sum_{j+l+k=p} (s_j - s_{j-2})(d_l - d_{l-2}) \\
&\cdot (k+2)(k+1) f_{1,k+2} + \frac{1}{2} \sum_{p=0}^{2M+M_1+2} x^p \sum\sum_{j+l+k=p} (s_j - s_{j-2})(s_l - s_{l-2}) \\
&\cdot (k+2)(k+1) g_{1,k+2} \\
&+ \frac{1}{2} \sum_{p=0}^{2N+M_1} x^p \sum\sum_{j+l+k=p} (d_j - d_{j-2}) d_l (k+2)(k+1) g_{1,k+2} + \hat{g}_3(x);
\end{aligned}
\tag{2.22}
$$

where

$$
b_0 = \sum_{k=1}^{N} \sum_{j=0}^{2k} d_j d_{2k-j},
$$

$$
b_1 = \sum_{k=1}^{N-1} \sum_{j=0}^{2k+1} d_j d_{2k+1-j},
$$

$$
b_j = b_{j-2} - \sum_{k=0}^{j} d_k d_{j-k}, \qquad (j = 2, 3, \cdots).
$$

The functions $\hat{f}_3(x)$ and $\hat{g}_3(x)$ can be obtained from relations (2.19) and (2.20), respectively, by replacing M_1 and N_1 by M_2 and N_2 and $f_{1,k}$ and $g_{1,k}$ by $f_{2,k}$ and $g_{2,k}$, respectively.

Now, let

$$
\tilde{f}(x) = \sum_{k=0}^{P} \tilde{f}_k x^k,
$$

$$
\tilde{g}(x) = \sum_{k=0}^{Q} \tilde{g}_k x^k.
$$

From (2.3) we have

$$
\begin{aligned}
\tilde{F}(z) = z &+ \sum_{k=0}^{P} \tilde{f}_k \frac{1}{\pi} \int_\alpha^\beta \sqrt{(\beta - t)(t - \alpha)} \frac{t^k}{t - z} \, dt \\
&+ \sqrt{(z-\alpha)(z-\beta)} \sum_{k=0}^{Q} \tilde{g}_k \frac{1}{i\pi} \int_\alpha^\beta \frac{t^k}{\sqrt{(\beta - t)(t - \alpha)}} \frac{dt}{t - z}.
\end{aligned}
$$

The integrals in the above relation can be analytically estimated, leading finally to the following expression for the function $\tilde{F}(z)$:

$$
\begin{aligned}
\tilde{F}(z) = z &+ i \sum_{k=0}^{Q} \tilde{g}_k z^k - \sum_{k=0}^{P} \tilde{f}_k \gamma_{k+1} - \sum_{l=1}^{P+1} z^l \sum_{k=l-1}^{P} \tilde{f}_k \gamma_{k+1-l} \\
&+ \sqrt{(z-\alpha)(z-\beta)} \left\{ \sum_{k=0}^{P} \tilde{f}_k z^k - i \sum_{l=0}^{Q-1} z^l \sum_{k=l+1}^{Q} \tilde{g}_k \cdot \delta_{k-l-1} \right\}
\end{aligned}
\tag{2.23}
$$

where $\delta_j = \Delta_j(0)$ and $\gamma_j = \Gamma_j(0)$.

Thus, if $S_1(x)$ and $D_1(x)$ are polynomial functions the asymptotic expansion of the solution $F(z)$ is obtained in the explicit form as a sum of the polynomial in z and the polynomial multiplied by $\{(z-\alpha)(z-\beta)\}^{1/2}$.

As an application we consider the case of the ellipse

$$y = \pm\varepsilon\sqrt{1-x^2}, \qquad -1 \leqq x \leqq 1.$$

The function conformally mapping (exactly) the domain exterior to this ellipse onto the Z plane with a cut on the OX axes is

$$F(z, \varepsilon) = \frac{1}{1-\varepsilon}(z - \varepsilon\sqrt{z^2 - 1 + \varepsilon^2}).$$

In this case (2.17)–(2.22) give

$$f(x, \varepsilon) = -(\varepsilon + \varepsilon^2 + \varepsilon^3) + O(\varepsilon^4),$$

and from equation (2.23) we have

$$F_3(z, \varepsilon) = (1 + \varepsilon + \varepsilon^2 + \varepsilon^3)z - (\varepsilon + \varepsilon^2 + \varepsilon^3)\sqrt{z^2 - 1 + \varepsilon^2 - \frac{\varepsilon^4}{4}}.$$

We can write

$$F(z, \varepsilon) - F_3(z, \varepsilon) = \frac{\varepsilon^4}{1-\varepsilon}(z - \sqrt{z^2 - 1 + \varepsilon^2})$$

$$- \frac{\varepsilon^5 + \varepsilon^6 + \varepsilon^7}{4(\sqrt{z^2 - 1 + \varepsilon^2} + \sqrt{z^2 - 1 + \varepsilon^2 - 0.25\varepsilon^4}} = O(\varepsilon^4),$$

hence the function $F_3(z, \varepsilon)$ is an uniformly approximation of the conformal mapping of the domain exterior to the considered ellipse.

3. The case of the symmetrical domain. If the curve C is symmetrical with respect to the Ox axis, the theory given in the preceding section can be applied, by setting $S_1(x) \equiv 0$. However the results obtained of Geer and Keller [2] and Geer [3] permit us to get in this case the completely asymptotic expansion of the function $F(z, \varepsilon)$.

In order to be able to use the asymptotic expansions given by the above-mentioned authors, we shall denote by

$$y = \pm\varepsilon\sqrt{S(x)}, \qquad 0 \leqq x \leqq 1, \tag{3.1}$$

the equation of the (symmetric) curve (C). The function $S(x)$ is analytical on the range [0, 1] and in the neighborhood of the bound 0 and 1 admits the expansion

$$S(x) = \sum_{n=1} d_n x^n; \qquad S(x) = \sum_{n=1} \tilde{d}_n (1-x)^n. \tag{3.2}$$

The function which performs the conformal mapping shall be represented in the form

$$F(z, \varepsilon) = z + \frac{1}{\pi}\int_\alpha^\beta \frac{f(t, \varepsilon)}{\sqrt{(\beta - t)(t - \alpha)}} \frac{dt}{t - z}, \tag{3.3}$$

where $f(t, \varepsilon)$ is a function analytical on $[\alpha, \beta]$ and which remains to be deduced. α and β are real parameters of the form

$$\alpha(\varepsilon) = \sum_{n=1}^{\infty} \alpha_{2n}\varepsilon^{2n}; \qquad \beta(\varepsilon) = 1 - \sum_{n=1}^{\infty} \beta_{2n}\varepsilon^{2n}. \tag{3.4}$$

They also must be determined. The integral equation satisfied by the function $f(t, \varepsilon)$ is

$$\varepsilon\sqrt{S(x)}+\operatorname{Im}\left\{\frac{1}{\pi}\int_\alpha^\beta \frac{f(t,\varepsilon)}{\sqrt{(\beta-t)(t-\alpha)}}\frac{dt}{t-x-i\varepsilon\sqrt{S(x)}}\right\}=0. \tag{3.5}$$

We shall write the relation (3.5) in the form

$$\varepsilon\pi\sqrt{S(x)}+\frac{\varepsilon\sqrt{S(x)}}{4\varepsilon S(x)+\varepsilon^3(S'(x))^2}(\varepsilon S'(x)I^0+2I^1)=0, \tag{3.6}$$

pointing out the integral operators $I^p(x, \varepsilon)$

$$\begin{aligned} I^0(x,\varepsilon)&=\int_{\alpha(\varepsilon)}^{\beta(\varepsilon)}\frac{2(x-\xi)+\varepsilon^2S'(x)}{(x-\xi)^2+\varepsilon^2S(x)}\frac{F(\xi)\,d\xi}{\sqrt{(\beta-\xi)(\xi-\alpha)}},\\ I^1(x,\varepsilon)&=\int_{\alpha(\varepsilon)}^{\beta(\varepsilon)}\frac{2S(x)-S'(x)(x-\xi)}{(x-\xi)^2+\varepsilon^2S(x)}\frac{F(\xi)\,d\xi}{\sqrt{(\beta-\xi)(\xi-\alpha)}} \end{aligned} \tag{3.7}$$

defined in [3].

The asymptotic expansions of these operators have the form

$$I^p(x,\varepsilon)=\sum_{q=0}^\infty \varepsilon^q L_q^p(F(x)), \tag{3.8}$$

where L_p^q, $q=0, 1$, are linear operators. Their expressions are obtained from (4.4)–(4.7) from [3] setting $C(x)\equiv 0$.

The use of the asymptotic expansions (3.8) into (3.6) leads to the following form of the integral equation

$$\pi\varepsilon(4S(x)+\varepsilon^2(S'(x))^2+\varepsilon S'(x)\sum_{j=0}^\infty \varepsilon^j L_j^0(f(x,\varepsilon))+2\sum_{j=0}^\infty \varepsilon^j L_j^1(f(x,\varepsilon))=0. \tag{3.9}$$

This relation suggests the search for the solution as the following asymptotic expansion

$$f(x,\varepsilon)=\sum_{n=1}^\infty f_n(x)\cdot\varepsilon^n. \tag{3.10}$$

Introducing (3.10) into (3.9) we get

$$\begin{aligned} &4\pi S(x)+\pi\varepsilon^2(S'(x))^2+S'(x)\sum_{n=1}^\infty \varepsilon^n\sum_{j=0}^{n-1}L_j^0f_{n-j}(x)\\ &\qquad+2\sum_{n=0}^\infty\varepsilon^n\sum_{j=0}^n L_j^1f_{n-j+1}(x)=0; \end{aligned} \tag{3.11}$$

and therefore

$$\begin{aligned} f_1(x)&=-\sqrt{x(1-x)S(x)},\\ f_2(x)&=-\sqrt{x(1-x)S(x)}\frac{1}{\pi}\int_0^1\frac{f_1(\xi)-f_1(x)}{\sqrt{\xi(1-\xi)}(\xi-x)^2}d\xi,\\ f_3(x)&=-\frac{1}{4\pi}\sqrt{\frac{x(1-x)}{S(x)}}(\pi(S'(x))^2+L_0^0f_2(x)+L_1^0f_1(x)+2L_1^1f_2(x)+2L_2^1f_1(x))\\ f_{n+1}(x)&=-\frac{1}{4\pi}\sqrt{\frac{x(1-x)}{S(x)}}\left(S'(x)\sum_{j=0}^{n-1}L_j^0f_{n-j}(x)+2\sum_{j=1}^n L_j^1f_{n-j+1}(x)\right). \end{aligned} \tag{3.12}$$

Similarly with the reasonings from [3], in order that the functions $f_n(x)$ be analytic on the range $[-1, 1]$ it is necessary that $L_q^p f_n$ are analytic functions, whence it follows the following asymptotic expansions for α and β.

$$\alpha(\varepsilon)=\tfrac{1}{4}d_1\varepsilon^2-\tfrac{1}{16}d_1d_2\varepsilon^4+\tfrac{1}{64}d_1(d_1d_3+2d_2^2)\varepsilon^6+\cdots,$$
$$\beta(\varepsilon)=1-\tfrac{1}{4}\tilde{d}_1\varepsilon^2+\tfrac{1}{16}\tilde{d}_1\tilde{d}_2\varepsilon^4-\tfrac{1}{64}\tilde{d}_1(\tilde{d}_1\tilde{d}_3+2\tilde{d}_2^2)\varepsilon^6-\cdots.$$

We remark that the leading terms from the relations (3.12) coincide with the corresponding terms deduced in the previous section.

Appendix. The asymptotic expansion of a Cauchy-type integral. We give here the asymptotic expansion of the integral

$$G(x+i\varepsilon Y(x))=\frac{1}{\pi}\int_\alpha^\beta \frac{\mu(t)h(t)}{t-x-i\varepsilon Y(x)}\,dt \tag{A.1}$$

for small values of the parameter ε. α and β are two real numbers and $\mu(t)$ is an integrable weight function. We assume that the function $Y(x)$ keeps a constant sign over the interval $[-1, 1]$.

We start with the identity

$$\begin{aligned}&\frac{1}{\pi}\int_\alpha^\beta \mu(t)\left\{h(t)-\sum_{j=0}^{k-1} h^{(j)}(x)\frac{(t-x)^j}{j!}\right\}\\ &\cdot\left\{\frac{1}{t-x-i\varepsilon Y}-\sum_{l=0}^{k-1}\frac{(i\varepsilon Y)^l}{(t-x)^{l+1}}-\frac{(i\varepsilon Y)^k}{(t-x)^k(t-x-i\varepsilon Y)}\right\}dt=0,\end{aligned} \tag{A.2}$$

valid for any integer k. From this we have

$$\begin{aligned}\frac{1}{\pi}\int_\alpha^\beta \frac{\mu(t)h(t)}{t-x-i\varepsilon Y}\,dt&=\frac{1}{\pi}\int_\alpha^\beta\frac{\mu(t)h(t)}{t-x}\,dt\\ &+\sum_{j=0}^{k-1}\frac{h^{(j)}(x)}{j!}\cdot\frac{1}{\pi}\int_\alpha^\beta \mu(t)(t-x)^j\cdot\left\{\frac{1}{t-x-i\varepsilon Y}-\frac{1}{t-x}\right\}dt\\ &+\sum_{j=1}^{k-1}\frac{(i\varepsilon Y)^j}{\pi}\int_\alpha^\beta\frac{\mu(t)h_k(t,x)}{(t-x)^{j+1}}\,dt\\ &+(i\varepsilon Y)^k\frac{1}{\pi}\int_\alpha^\beta\frac{\mu(t)h_k(t,x)\,dt}{(t-x)^k(t-x-i\varepsilon Y)}.\end{aligned} \tag{A.3}$$

In relation (A.3) we have denoted

$$h_k(t,x)=h(t)-\sum_{j=0}^{k-1}\frac{h^{(j)}(x)}{j!}(t-x)^j.$$

As the last integral on the right-hand side of the relation (A.3) is a bounded quantity, we have

$$\begin{aligned}\frac{1}{\pi}\int_\alpha^\beta\frac{\mu(t)h(t)}{t-x-i\varepsilon Y}\,dt&=\frac{1}{\pi}\int_\alpha^\beta\frac{\mu(t)h(t)}{t-x}\,dt+\sum_{j=0}^{k-1}\frac{h^{(j)}(x)}{j!}\\ &\cdot\frac{1}{\pi}\int_{\alpha-x}^{\beta-x}\mu(t+x)t^j\left\{\frac{1}{t-i\varepsilon Y}-\frac{1}{t}\right\}dt\\ &+\sum_{j=1}^{k-1}(i\varepsilon Y)^j\frac{1}{\pi}\int_\alpha^\beta\frac{\mu(t)h_k(t,x)}{(t-x)^{j+1}}\,dt+O(\varepsilon^k).\end{aligned} \tag{A.4}$$

Further on, we consider two particular cases:

(a) If $\mu(t)=\{(\beta-t)(t-\alpha)\}^{1/2}$ then from (A.4) we have

$$
\begin{aligned}
\frac{1}{\pi}\int_\alpha^\beta &\sqrt{(\beta-t)(t-\alpha)}\,\frac{h(t)\,dt}{t-x-i\varepsilon Y}\\
&=\frac{1}{\pi}\int_\alpha^\beta \sqrt{(\beta-t)(t-\alpha)}\,\frac{h(t)\,dt}{t-x}\\
&\quad+\sqrt{(x+i\varepsilon Y-\alpha)}\cdot\sqrt{(x+i\varepsilon Y-\beta)}\sum_{j=0}^{k-1}\frac{h^{(j)}(x)}{j!}(i\varepsilon Y)^j\\
&\quad-\sum_{j=1}^{k-1}(i\varepsilon Y)^j\sum_{r=j-1}^{k-1}\Gamma_{r-j+1}\frac{h^{(r)}(x)}{r!}\\
&\quad+\sum_{j=1}^{k-1}(i\varepsilon Y)^j\frac{1}{\pi}\int_\alpha^\beta\sqrt{(\beta-t)(t-\alpha)}\,\frac{h_k(t,x)}{(t-x)^{j+1}}\,dt+O(\varepsilon^k).
\end{aligned}\tag{A.5}
$$

The determination of the square root is the positive one for $Y=0$ and $x>\beta$.

In relation (A.5) we have also denoted

$$
\Gamma_r=\Gamma_r(x)=\frac{1}{2^r}\sum_{p=0}^{r}\frac{(2p-3)!!}{p!}\,\frac{[2(r-p)-3]!!}{(r-p)!}(\alpha-x)^p(\beta-x)^{r-p},\qquad r=0,1,\cdots.\tag{A.6}
$$

(In this formula we set $(-1)!!=1$ and $(-3)!!=-1$.)

(b) When $\mu(t)=\{(\beta-t)(t-\alpha)\}^{-1/2}$, the relation (A.4) yields

$$
\begin{aligned}
\frac{1}{\pi}\int_\alpha^\beta &\frac{h(t)}{\sqrt{(\beta-t)(t-\alpha)}}\,\frac{dt}{t-x-i\varepsilon Y}\\
&=\frac{1}{\pi}\int_\alpha^\beta \frac{h(t)}{\sqrt{(\beta-t)(t-\alpha)}}\,\frac{dt}{t-x}\\
&\quad-\frac{1}{\sqrt{(x+i\varepsilon Y-\alpha)(x+i\varepsilon Y-\beta)}}\sum_{j=0}^{k-1}(i\varepsilon Y)^j\frac{h^{(j)}(x)}{j!}\\
&\quad+\sum_{j=1}^{k-2}(i\varepsilon Y)^j\sum_{r=j+1}^{k-1}\frac{h^{(r)}(x)}{s!}\Delta_{r-j-1}\\
&\quad+\sum_{j=1}^{k-1}(i\varepsilon Y)^j\frac{1}{\pi}\int_\alpha^\beta\frac{h_k(t,x)}{\sqrt{(\beta-t)(t-\alpha)}}\,\frac{dt}{(t-x)^{k+1}}+O(\varepsilon^k).
\end{aligned}\tag{A.7}
$$

where we have denoted

$$
\begin{aligned}
\Delta_r&=\Delta_r(x)\\
&=\frac{1}{2^r}\sum_{p=0}^{r}\frac{(2p-1)!!}{p!}\,\frac{[2(r-p)-1]!!}{(r-p)!}(\alpha-x)^p(\beta-x)^{r-p}.
\end{aligned}\tag{A.8}
$$

using the same conventions as in (A.5) and (A.6).

Acknowledgment. I would like to thank Professors J. B. Keller and J. F. Geer for their useful comments on an earlier version of this paper.

REFERENCES

[1] C. JACOB, *Introduction mathématique à la mécanique des fluides*, Gauthier-Villars, Bucarest-Paris, 1959.

[2] J. F. GEER AND J. B. KELLER, *Uniform asymptotic solutions for potential flow around a thin airfoil and the electrostatic potential about a thin conductor*, SIAM J. Appl. Math., 16 (1968), pp. 75–101.

[3] J. F. GEER, *Uniform asymptotic solutions for the two dimensional potential field about a slender body*, Ibid., 26 (1974) pp. 539–553.

[4] R. A. HANDELSMAN AND J. B. KELLER, *Axially symmetric potential flow around a slender body*, J. Fluid Mech., 28 (1967), pp. 131–142.

[5] C. JACOB, *Sur la résolution explicite du problème plan de Dirichlet pour certaines domaines canoniques*, Bull. Math. Soc. Sci. Math. R. S. Roumanie, 10 (1966), no. 58, pp. 13–26.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0014 $01.00/0

SOME WIRTINGER-LIKE INEQUALITIES*

JERRY M. FEINBERG†

Abstract. This paper extends a variational inequality [G. Hardy, J. Littlewood and G. Polya, *Inequalities*, Cambridge University Press, Cambridge, 1967; p. 182] for real valued functions and their derivatives to those functions on an interval with zero boundary conditions or with zero integral. Theorem 4.2, the main inequality, has an important application to differential geometry [J. M. Feinberg, *The isoperimetric inequality for doubly-connected minimal surfaces in* R^n, J. Analyse Math., 32 (1977), pp. 249–278]. A discrete version of the inequality is also derived and some applications are provided.

Introduction. Let a real valued piecewise C^1 function f be defined on a finite closed interval $[a, b]$ with $f(a)=f(b)$. We are interested in inequalities between $\int_a^b |df/dt|^\lambda\, dt$ and $\int_a^b |f|^\lambda\, dt$ subject to various restrictions on λ and f. Our main result (Theorem 3.3) is, if $\mu \geqq 1$ is the quotient of two odd integers, and $\int_a^b f^\mu\, dt = 0$, $f \not\equiv 0$, then

$$\int_a^b \left|\frac{df}{dt}\right|^\mu dt \geqq K(\mu)\cdot\left(\frac{4}{b-a}\right)^\mu \cdot \int_a^b |f|^\mu\, dt$$

where

$$K(\mu)=\begin{cases} 1, & \mu = 1, \\ (\mu-1)\cdot\left(\dfrac{\pi}{\mu}\csc\dfrac{\pi}{\mu}\right)^\mu, & \mu > 1. \end{cases}$$

Equality can occur only if $\mu > 1$ and f is a given function depending on μ; for $\mu = 1$, equality never occurs.

The case $\mu = 1$ may be extended to vector valued functions f, and it is shown that the inequality still holds with the same constant. We will also derive a discrete analogue of this main result, proving (Theorem 4.4):

$$\left(\sum_{i=1}^n \beta_i\right)\cdot\left(\sum_{i=1}^n \|v_{i+1}-v_i\|\right) \geqq 4\cdot \sum_{i=1}^n \beta_i\|v_i\|$$

for $\beta_i \in R^+ \cup \{0\}$, $v_i \in R^N$ with $v_{n+1}=v_1$, and $\sum_{i=1}^n \beta_i v_i = \mathbf{0}$. The final section of this paper will be devoted to the proof of this discrete analogue and some of its consequences.

The inequality $\int_a^b (df/dt)^2\, dt \geqq [2\pi/(b-a)]^2 \int_a^b f^2\, dt$ if $f(a)=f(b)$ and $\int_a^b f\, dt = 0$ is known as the Wirtinger inequality [6, p. 185]. Related results have been proved by Bellman [3, p. 140] for the $2k$ powers of f and df/dt, k an integer, with f subject to $\int_a^b f\, dt = 0$, and by Beesack [2, p. 21] for the $2k$ powers with f subject to $\int_a^b f^{2k-1}\, dt = 0$. Discrete analogues have also been proved by Fan, Taussky, and Todd [4, p. 73].

1. Notation and basic results. We will use the following notation:

λ; any real number $\geqq 1$;

μ, ν; real numbers $\geqq 1$ which are the quotients of two odd integers;

$[a, b]$; a bounded closed interval $\in R$;

f; a real valued function on $[a, b]$;

β_i; nonnegative real numbers, $i = 1, 2, \cdots, n$ and $\beta = \sum_{i=1}^n \beta_i$;

$v = (v_1, v_2, \cdots, v_N) \in R^N$ and $\|v\| = (\sum_{i=1}^N v_i^2)^{1/2}$.

* Received by the editors July 6, 1977 and in revised form October 13, 1978.

† Department of Mathematics, Brandeis University, Waltham, Massachusetts 02154. This research was supported by the Fannie and John Hertz Foundation and Stanford University. This work was completed while the author was a visiting member at the Courant Institute of Mathematical Sciences, New York University.

We define a class $A(\lambda)$ of admissable real valued functions as follows: $f\in A(\lambda)$ if $f(t)$ is continuous on $[a, b]$, $df(t)/dt$ is defined and continuous except at finitely many points in $[a, b]$, df/dt is absolutely integrable with $f(t)=f(a)+\int_a^t df(s)/ds$, and $df/dt\in L^\lambda[a, b]$.

The number μ, being the quotient of two odd integers, is of the most general form which permits

$$(-a)^\mu=-(a^\mu)\quad \forall a\in R. \tag{1.1}$$

LEMMA 1.1. *If $f\in A(\mu)$ and $\int_a^b f^\mu\,dt=0$, then $\exists\tau\in[a, b]$ such that $f(\tau)=0$.*

Proof. By the definition of μ, f^μ has the same sign as f. Since f is continuous, no zero on the interval $[a, b]$ would imply f is entirely of one sign, and hence the same for f^μ. This is in violation of the condition $\int_a^b f^\mu\,dt=0$. Q.E.D.

If $f\in A(\mu)$ with $f(a)=f(b)$ and $\int_a^b f^\mu\,dt=0$, we may extend f periodically beyond $[a, b]$. By defining $\hat f(t)=f(\tau+t-a)$ where $f(\tau)=0$, we find that $\hat f\in A(\mu)$, $\hat f(a)=\hat f(b)=0$, and $\int_a^b \hat f^\mu\,dt=\int_a^b f^\mu\,dt=0$. Furthermore, $\int_a^b|\hat f|^\mu\,dt=\int_a^b|f|^\mu\,dt$ and $\int_a^b|d\hat f/dt|^\mu\,dt=\int_a^b|df/dt|^\mu\,dt$. Hence any inequality relating the μ powers of $|f|$ and $|df/dt|$ for $f\in A(\mu)$ with $\int_a^b f^\mu\,dt=0$ and $f(a)=f(b)$, need be proven only for such functions with $f(a)=f(b)=0$.

2. Lower bounds. The existence of lower bounds for the ratio $\int_a^b|df/dt|^\mu\,dt/\int_a^b|f|^\mu\,dt$ for $f\in A(\mu)$ subject to the restriction $\int_a^b f^\mu\,dt=0$ may be established in the following manner.

LEMMA 2.1. *Let g and h be real valued L^1 functions and assume that g is periodic with period $b-a$. If*

$$k(t)=\int_a^b g(t+x)h(x)\,dx,$$

then

$$\int_a^b|k(t)|\,dt\leqq\int_a^b|g(t)|\,dt\cdot\int_a^b|h(t)|\,dt.$$

Proof. By a standard inequality, we have

$$|k(t)|\leqq\int_a^b|g(t+x)|\cdot|h(x)|\,dx$$

and hence

$$\begin{aligned}\int_a^b|k(t)|\,dt&\leqq\int_a^b\int_a^b|g(t+x)|\cdot|h(x)|\,dx\,dt\\&=\int_a^b\left\{\int_a^b|g(t+x)|\,dt\right\}|h(x)|\,dx\\&=\int_a^b\left\{\int_a^b|g(t)|\,dt\right\}|h(x)|\,dx\quad\text{since } g \text{ is periodic}\\&=\int_a^b|g(t)|\,dt\cdot\int_a^b|h(x)|\,dx.\end{aligned}$$

Q.E.D.

If $f\in A(1)$ with $f(a)=f(b)$ and $\int_a^b f\,dt=0$, apply Lemma 2.1 with $g=df/dt$ and $h(t)=t-(b+a)/2$. (Note that f can be extended periodically and that df/dt is defined and can be extended periodically except for at most a finite number of points which make no difference in Lemma 2.1.) Then

$$\begin{aligned}
k(t) &= \int_a^b \frac{df}{dt}(t+x) \cdot \left(x - \frac{b+a}{2}\right) dx \\
&= \int_a^b x \frac{df}{dt}(t+x)\, dx - \frac{b+a}{2} \int_a^b \frac{df}{dt}(t+x)\, dx \\
&= xf(t+x) \Big|_{x=a}^{b} - \int_a^b f(t+x)\, dx - \frac{b+a}{2} f(t+x) \Big|_{x=a}^{b} \\
&= (b-a)f(t+a) - 0 + 0 \\
&= (b-a)f(t+a)
\end{aligned}$$

since f is periodic and $\int_a^b f(t)\, dt = 0$. Then, again by the periodicity of f, $\int_a^b |k(t)|\, dt = (b-a)\int_a^b |f(t+a)|\, dt = (b-a)\int_a^b |f(t)|\, dt$. An easy calculation shows that $\int_a^b |h'(t)|\, dt = (b-a)^2/4$. Lemma 2.1 yields $(b-a) \cdot \int_a^b |f(t)|\, dt \leqq \int_a^b |df/dt|\, dt \cdot (b-a)^2/4$ and we conclude

THEOREM 2.1. *If $f \in A(1)$ with $f(a) = f(b)$, then*

$$\int_a^b \left|\frac{df}{dt}\right| dt \geqq \frac{4}{b-a} \int_a^b |f|\, dt \quad \textit{if} \quad \int_a^b f\, dt = 0. \tag{2.1}$$

LEMMA 2.2. *Let $\mu, \nu \geqq 1$ be quotients of odd integers. If $\int_a^b |df/dt|^\mu\, dt \geqq K \cdot \int_a^b |f|^\mu\, dt$ for all $f \in A(\mu)$ with $f(a) = f(b)$ and $\int_a^b f^\mu\, dt = 0$, then $\int_a^b |dg/dt|^{\mu\nu}\, dt \geqq K^\nu/\nu^{\mu\nu} \cdot \int_a^b |g|^{\mu\nu}\, dt$ for all $g \in A(\mu\nu)$ with $g(a) = g(b)$ and $\int_a^b g^{\mu\nu}\, dt = 0$.*

Proof. First note that $\mu \cdot \nu \geqq 1$ is also the quotient of odd integers and that $[g(t)]^\nu$ is well defined. Now assume $g \in A(\mu\nu)$, $g(a) = g(b)$, and $\int_a^b g^{\mu\nu}\, dt = 0$. Define $f(t) = [g(t)]^\nu$. Clearly $f(a) = f(b)$ and $\int_a^b f^\mu\, dt = 0$. Since $df/dt = \nu \cdot g^{\nu-1} \cdot dg/dt$, we have $|df/dt|^\mu = \nu^\mu |g|^{\mu(\nu-1)} |dg/dt|^\mu$. Using the Hölder inequality [1, p. 21] with $p = \nu/(\nu-1) > 1$ and $q = \nu$, we have

$$\int_a^b |g|^{\mu(\nu-1)} \left|\frac{dg}{dt}\right|^\mu dt \leqq \left\{\int_a^b (|g|^{\mu(\nu-1)})^{\nu/(\nu-1)}\, dt\right\}^{(\nu-1)/\nu} \cdot \left\{\int_a^b \left(\left|\frac{dg}{dt}\right|^\mu\right)^\nu dt\right\}^{1/\nu}.$$

Then

$$\int_a^b \left|\frac{df}{dt}\right|^\mu dt \leqq \nu^\mu \left\{\int_a^b |g|^{\mu\nu}\, dt\right\}^{(\nu-1)/\nu} \cdot \left\{\int_a^b \left|\frac{dg}{dt}\right|^{\mu\nu} dt\right\}^{1/\nu} < \infty \tag{2.2}$$

hence df/dt is L^μ and $f \in A(\mu)$.

By assumption, now,

$$\int_a^b \left|\frac{df}{dt}\right|^\mu dt \geqq K \int_a^b |f|^\mu\, dt,$$

or

$$\int_a^b \left|\frac{df}{dt}\right|^\mu dt \geqq K \int_a^b |g|^{\mu\nu}\, dt.$$

Combining this with (2.2) gives

$$\left\{\int_a^b |g|^{\mu\nu}\,dt\right\}^{(\nu-1)/\nu}\cdot\left\{\int_a^b \left|\frac{dg}{dt}\right|^{\mu\nu} dt\right\}^{1/\nu}\cdot \nu^{\mu} \geqq K\int_a^b |g|^{\mu\nu}\,dt$$

or

$$\int_a^b \left|\frac{dg}{dt}\right|^{\mu\nu} dt \geqq \frac{K^{\nu}}{\nu^{\mu\nu}}\cdot\int_a^b |g|^{\mu\nu}\,dt. \tag{2.3}$$

Q.E.D.

THEOREM 2.2. *If $f\in A(\mu)$ with $f(a)=f(b)$, then*

$$\int_a^b \left|\frac{df}{dt}\right|^{\mu} dt \geqq \left[\frac{4}{\mu(b-a)}\right]^{\mu}\cdot\int_a^b |f|^{\mu}\,dt \quad \text{if} \quad \int_a^b f^{\mu}\,dt=0. \tag{2.4}$$

Proof. In Lemma 2.2, let $\mu=1$ and $\nu=\mu$. Theorem 2.1 allows us to use $K=4/(b-a)$. Q.E.D.

Unfortunately, this method gives no indication of when equality may occur or the form of the extremal functions should equality occur. In the next section, we will show, in fact, that these lower bounds are never attained.

3. Precise lower bounds. In this section, we will derive the best possible lower bounds for the ratio $\int_a^b|df/dt|^{\mu}\,dt/\int_a^b|f|^{\mu}\,dt$ for $f\in A(\mu)$ with $f(a)=f(b)$ and $\int_a^b f^{\mu}\,dt=0$. We begin with some results for less restricted functions and for arbitrary $\lambda\geqq 1$.

LEMMA 3.1. *If $f\in A(1)$ with $f\geqq 0$, $f\not\equiv 0$, and $f(a)=0$, then $\int_a^b|df/dt|\,dt>1/(b-a)\int_a^b f\,dt$. Equality never occurs, but the constant $1/(b-a)$ is the best possible.*

Proof. Let $M=\max_{t\in[a,b]}f(t)=f(\xi)$ for some $\xi\in[a,b]$. Then $\int_a^b f\,dt<M\cdot(b-a)$, and this inequality is strict since $f(a)=0$. We have

$$\begin{aligned}\int_a^b\left|\frac{df}{dt}\right|dt &= \int_a^{\xi}\left|\frac{df}{dt}\right|dt+\int_{\xi}^b\left|\frac{df}{dt}\right|dt\\ &\geqq \int_a^{\xi}\left|\frac{df}{dt}\right|dt\\ &\geqq \left|\int_a^{\xi}\frac{df}{dt}\,dt\right|\\ &= |f(\xi)-f(a)|\\ &= M.\end{aligned}$$

Hence

$$\int_a^b\left|\frac{df}{dt}\right|dt\Big/\int_a^b f\,dt>\frac{M}{M(b-a)}=\frac{1}{b-a}.$$

Equality never occurs, but the sequence of functions

$$f_n(t)=\begin{cases} n(t-a), & \text{if } a\leqq t\leqq a+\dfrac{1}{n},\\[2ex] 1, & \text{if } a+\dfrac{1}{n}\leqq t\leqq b\end{cases}$$

for $n>1/(b-a)$ shows that the constant $1/(b-a)$ is the best possible. Q.E.D.

LEMMA 3.2. *If $f \in A(\lambda)$, $\lambda > 1$, with $f \geqq 0$, $f \not\equiv 0$, and $f(a)=0$, then*

$$\int_a^b \left|\frac{df}{dt}\right|^\lambda dt \geqq \left(\frac{1}{b-a}\right)^\lambda \cdot (\lambda-1) \cdot \left(\frac{\pi}{\lambda}\csc\frac{\pi}{\lambda}\right)^\lambda \cdot \int_a^b f^\lambda\, dt.$$

Equality occurs only for multiples of a certain hyper-elliptic curve $y(t)$ whose inverse is given by

$$t = a + (b-a)\cdot\left(\frac{\lambda}{\pi}\sin\frac{\pi}{\lambda}\right)\cdot\int_0^y \frac{d\omega}{(1-\omega^\lambda)^{1/\lambda}}, \qquad 0 \leqq y \leqq 1. \tag{3.1}$$

Proof. The proof given in [6, p. 182] for $\lambda = 2k$, an even integer, is valid for any real number $\lambda > 1$. Q.E.D.

For simplicity, we define $K(\lambda)$ as follows:

$$K(\lambda) = \begin{cases} 1, & \text{if } \lambda = 1, \\ (\lambda-1)\left(\dfrac{\pi}{\lambda}\csc\dfrac{\pi}{\lambda}\right)^\lambda, & \text{if } \lambda > 1. \end{cases} \tag{3.2}$$

LEMMA 3.3.

$$\lim_{\lambda \to 1^+} K(\lambda) = 1 = K(1).$$

The proof is an elementary exercise in calculus. We may summarize Lemmas 3.1 and 3.2 as follows:

THEOREM 3.1. *If $f \in A(\lambda)$, $\lambda \geqq 1$, with $f \geqq 0$, $f \not\equiv 0$, and $f(a)=0$, then*

$$\int_a^b \left|\frac{df}{dt}\right|^\lambda dt \geqq \left(\frac{1}{b-a}\right)^\lambda \cdot K(\lambda) \cdot \int_a^b f^\lambda\, dt. \tag{3.3}$$

For $\lambda > 1$, equality occurs only for multiples of a curve given by (3.1). *For $\lambda = 1$, equality never occurs, but the inequality is the best possible.*

LEMMA 3.4. *If $\lambda \geqq -1$ and $x, y > 0$, then*

$$\frac{1}{x^\lambda} + \frac{1}{y^\lambda} \geqq \frac{2^{\lambda+1}}{(x+y)^\lambda}.$$

Equality holds if and only if $x = y$.

Proof. This is the inequality between the means. See [1, p. 16] with $\alpha = (\frac{1}{2}, \frac{1}{2})$ and $t = -\lambda$ and $t = 1$. Q.E.D.

THEOREM 3.2. *If $f \in A(\lambda)$, $\lambda \geqq 1$, with $f \geqq 0$, $f \not\equiv 0$, and $f(a) = f(b) = 0$, then*

$$\int_a^b \left|\frac{df}{dt}\right|^\lambda dt \geqq \left(\frac{2}{b-a}\right)^\lambda \cdot K(\lambda) \cdot \int_a^b f^\lambda\, dt. \tag{3.4}$$

For $\lambda > 1$, equality can occur only for multiples of a function $f(t)$ given by

$$f(t) = \begin{cases} y(-a+2t), & \text{if } a \leqq t \leqq \dfrac{a+b}{2}, \\ y(a+2b-2t), & \text{if } \dfrac{a+b}{2} \leqq t \leqq b \end{cases}$$

where $y(t)$ is given by (3.1). *For $\lambda = 1$, equality never occurs, but the inequality is the best possible.*

Proof. $f \geqq 0$ implies $\int_a^x f^\lambda\,dt$ is an increasing continuous function of x. By the intermediate value theorem, we may choose $\xi \in (a, b)$ such that $\int_a^\xi f^\lambda\,dt = \int_\xi^b f^\lambda\,dt = \frac{1}{2}\int_a^b f^\lambda\,dt$.

Then

$$\begin{aligned}
\frac{\int_a^b |df/dt|^\lambda\,dt}{\int_a^b f^\lambda\,dt} &= \frac{\int_a^\xi |df/dt|^\lambda\,dt + \int_\xi^b |df/dt|^\lambda\,dt}{\int_a^\xi f^\lambda\,dt + \int_\xi^b f^\lambda\,dt} \\
&= \frac{1}{2}\left\{\frac{\int_a^\xi |df/dt|^\lambda\,dt}{\int_a^\xi f^\lambda\,dt} + \frac{\int_\xi^b |df/dt|^\lambda\,dt}{\int_\xi^b f^\lambda\,dt}\right\} \\
&\geqq \frac{1}{2}\cdot K(\lambda)\cdot\left\{\frac{1}{(\xi-a)^\lambda} + \frac{1}{(b-\xi)^\lambda}\right\} \qquad \text{(by Theorem 3.1 since } f(a)=0=f(b)) \\
&\geqq \frac{1}{2}\cdot K(\lambda)\cdot\frac{2^{\lambda+1}}{(b-a)^\lambda} \qquad \text{(by Lemma 3.4)} \\
&= \left(\frac{2}{b-a}\right)^\lambda \cdot K(\lambda).
\end{aligned}$$

For $\lambda > 1$, equality occurs if and only if $\xi - a = b - \xi$, i.e. $\xi = (a+b)/2$, and f or its reflection is the hyper-elliptic curve given by (3.1) on each of the intervals $[a, (a+b)/2]$ and $[(a+b)/2, b]$. For $\lambda = 1$, equality never occurs, but the sequence of functions

$$f_n(t) = \begin{cases} n(t-a), & \text{if } a \leqq t \leqq a + \dfrac{1}{n}, \\ 1, & \text{if } a + \dfrac{1}{n} \leqq t \leqq b - \dfrac{1}{n}, \\ -n(t-b), & \text{if } b - \dfrac{1}{n} \leqq t \leqq b \end{cases}$$

for $n \geqq 2/(b-a)$ shows that the constant $2/(b-a)$ is the best possible. Q.E.D.

LEMMA 3.5. *If $x_l, y_l \geqq 0$, $\lambda \geqq 1$, the sums exist, and $\sum_l y_l \neq 0$, then*

$$\sum_l x_l \geqq \sum_l (x_l \cdot y_l) / \sum_l y_l \tag{3.5}$$

$$\left(\sum_l x_l\right)^\lambda \geqq \sum_l x_l^\lambda. \tag{3.6}$$

Proof. The inequality (3.5) is trivial. Equality can hold only if $x_l = y_l = 0$ for all but one l. Inequality (3.6) is Jensen's inequality [1, p. 18]. Equality holds in (3.6) if $\lambda = 1$; if $\lambda > 1$, equality holds only if $x_l = 0$ for all but one l. Q.E.D.

THEOREM 3.3. *If $f \in A(\mu)$ where $\mu \geqq 1$ is the quotient of two odd integers, if $f \not\equiv 0$ and if*

$$f(a) = f(b) \quad \textit{and} \quad \int_a^b f^\mu\,dt = 0, \tag{3.7}$$

then

$$\int_a^b \left|\frac{df}{dt}\right|^\mu dt \geqq \left(\frac{4}{b-a}\right)^\mu \cdot K(\mu) \cdot \int_a^b |f|^\mu\,dt. \tag{3.8}$$

For $\mu > 1$, equality can occur only for multiples and translates of a function $f(t)$ given by

$$f(t)=\begin{cases} y(-3a+4t), & \text{if } a \leqq t \leqq \frac{3a+b}{4}, \\ y(3a+2b-4t), & \text{if } \frac{3a+b}{4} \leqq t \leqq \frac{a+b}{2}, \\ -y(-a-2b+4t), & \text{if } \frac{a+b}{2} \leqq t \leqq \frac{a+3b}{4}, \\ -y(a+4b-4t), & \text{if } \frac{a+3b}{4} \leqq t \leqq b \end{cases} \tag{3.9}$$

where $y(t)$ is given by (3.1). *For $\mu = 1$, equality never occurs, but the inequality is the best possible.*

Proof. By Lemma 1.1, we may assume that $f(a)=f(b)=0$. Let $S^+ = \{t \in (a, b) \mid f(t) > 0\}$ and $S^- = \{t \in (a, b) \mid f(t) < 0\}$. Since f is continuous, S^+ and S^- are open sets, each the union of a countable number of open intervals, $S^+ = \bigcup_{i=1}^{\infty} P_i$, $S^- = \bigcup_{j=1}^{\infty} Q_j$. We define

$$A_i = \int_{P_i} |f|^{\mu} \, dt = \int_{P_i} f^{\mu} \, dt$$

and

$$B_j = \int_{Q_j} |f|^{\mu} \, dt = \int_{Q_j} (-f)^{\mu} \, dt = -\int_{Q_j} f^{\mu} \, dt$$

where the last equality follows from (1.1).

Equation (3.7) requires that

$$\sum_i A_i = \sum_j B_j \tag{3.10}$$

and Theorem 3.2 for f on A_i and $-f$ on B_j implies

$$\int_{P_i} \left|\frac{df}{dt}\right|^{\mu} dt \geqq \left(\frac{2}{l(P_i)}\right)^{\mu} \cdot K(\mu) \cdot A_i, \tag{3.11}$$

$$\int_{Q_j} \left|\frac{df}{dt}\right|^{\mu} dt \geqq \left(\frac{2}{l(Q_j)}\right)^{\mu} \cdot K(\mu) \cdot B_j \tag{3.12}$$

where $l(P_i)$ = length of the interval P_i, and similarly for $l(Q_j)$.
Hence

$$\frac{\int_a^b |df/dt|^{\mu} \, dt}{\int_a^b |f|^{\mu} \, dt} = \frac{\sum_i \int_{P_i} |df/dt|^{\mu} \, dt + \sum_j \int_{Q_j} |df/dt|^{\mu} \, dt}{\sum_i \int_{P_i} |f|^{\mu} \, dt + \sum_j \int_{Q_j} |f|^{\mu} \, dt}$$

$$\geqq 2^{\mu} \cdot K(\mu) \cdot \frac{\sum_i A_i/[l(P_i)]^{\mu} + \sum_j B_j/[l(Q_j)]^{\mu}}{\sum_i A_i + \sum_j B_j} \quad \text{(from (3.11) and (3.12))} \tag{3.13}$$

$$\geqq 2^{\mu} \cdot K(\mu) \cdot \frac{\sum_i A_i \cdot 1/\sum_i [l(P_i)]^{\mu} + \sum_j B_j \cdot 1/\sum_j [l(Q_j)]^{\mu}}{\sum_i A_i + \sum_i B_j},$$

(from (3.5) with $x_j = A_i/[l(P_i)]^{\mu}$, $y_i = [l(P_i)]^{\mu}$) and then

$$x_j = \frac{B_j}{[l(Q_j)]^{\mu}},\ y_j = [l(Q_j)]^{\mu})$$

$$\geqq 2^{\mu} \cdot K(\mu) \cdot \frac{1}{2} \cdot \left(\frac{1}{\sum_i [l(P_i)]^{\mu}} + \frac{1}{\sum_j [l(Q_j)]^{\mu}}\right) \qquad \text{(from (3.10))}$$

$$\geqq 2^{\mu-1} \cdot K(\mu) \cdot \left(\frac{1}{[\sum_i l(P_i)]^{\mu}} + \frac{1}{[\sum_j l(Q_j)]^{\mu}}\right) \qquad \text{(from (3.6))} \tag{3.14}$$

$$\geqq 2^{\mu-1} \cdot K(\mu) \cdot \frac{2^{\mu+1}}{[\sum_i l(P_i) + \sum_j l(Q_j)]^{\mu}} \qquad \text{(from Lemma 3.4)} \tag{3.15}$$

$$\geqq 4^{\mu} \cdot K(\mu) \cdot \frac{1}{(b-a)^{\mu}} = \left(\frac{4}{b-a}\right)^{\mu} \cdot K(\mu). \tag{3.16}$$

For equality to occur in (3.14), we must have only one P_i and one Q_j. For equality in (3.15) and (3.16), we must have $l(P_1) = l(Q_1) = \frac{1}{2}(b-a)$. For equality in (3.13), f must have the form of (3.1). Finally, $\int_a^b f^{\mu}\,dt = 0$ forces $\max_{P_1} f = -\min_{Q_1} f$. If $f(a) \neq 0$, then define $\hat{f}(t) = f(\tau + t - a)$ where τ is defined in Lemma 1.1. Then $\hat{f}$ has the form (3.9), and f is a translate of the periodic extension of $\hat{f}$. For $\mu = 1$, equality never occurs, but the sequence of functions

$$f_n(t) = \begin{cases} n(t-a), & \text{if } a \leqq t \leqq a + \dfrac{1}{n}, \\ 1, & \text{if } a + \dfrac{1}{n} \leqq t \leqq \dfrac{a+b}{2} - \dfrac{1}{n}, \\ -n\left(t - \dfrac{a+b}{2}\right), & \text{if } \dfrac{a+b}{2} - \dfrac{1}{n} \leqq t \leqq \dfrac{a+b}{2} + \dfrac{1}{n}, \\ -1, & \text{if } \dfrac{a+b}{2} + \dfrac{1}{n} \leqq t \leqq b - \dfrac{1}{n}, \\ n(t-b), & \text{if } b - \dfrac{1}{n} \leqq t \leqq b \end{cases}$$

for $n > 4/(b-a)$ shows that the constant $4/(b-a)$ is the best possible. Q.E.D.

4. Vector valued functions and discrete analogues for $\mu = 1$. In this section, we first establish an extension of Theorem 3.3 to vector valued functions for the case $\mu = 1$. We then use this extended theorem to prove some discrete analogues of the Wirtinger inequality for vectors.

DEFINITION. $f \in A'$ if $f \in A(1)$, $f(a) = f(b)$, and $\int_a^b f\,dt = 0$.

It is clear that A' is a vector space of functions. Furthermore, from Theorem 3.3, if $f \in A'$ and $f \not\equiv 0$, then

$$\int_a^b \left|\frac{df}{dt}\right| dt > \frac{4}{b-a} \int_a^b |f|\,dt.$$

We now use, and for completeness provide a direct proof of the following theorem originally due to A. Zygmund [7, p. 117].

THEOREM 4.1. *Let V and W be vector spaces of L^1 functions and let T be a linear transformation of V into W. If for all $f \in V$ with $f \not\equiv 0$,*

$$\int_a^b |(Tf)(\xi)|\,d\xi > M \int_c^d |f(x)|\,dx \tag{4.1}$$

where M is independent of f; then for any $f_1, \cdots, f_N \in V$, at least one of which $\not\equiv 0$,

$$\int_a^b \left[\sum_{j=1}^N (Tf_j)^2(\xi)\right]^{1/2} d\xi > M \int_c^d \left[\sum_{j=1}^N f_j^2(x)\right]^{1/2} dx. \tag{4.2}$$

Proof. Let $\mathbf{f} = (f_1, \cdots, f_N)$ and let $S = S^{N-1}(1)$ be the unit $N-1$ dimensional sphere in R^N. Let $\mathbf{e} = (e_1, \cdots, e_N) \in S$ and consider

$$F(x) \equiv \sum_{j=1}^N e_j f_j(x) = \mathbf{e} \cdot \mathbf{f}(x).$$

Then $F \in V$ and since T is linear,

$$TF = \sum_{j=1}^N e_j T(f_j) = \mathbf{e} \cdot (T\mathbf{f}).$$

From (4.1), if $\mathbf{e} \cdot \mathbf{f} \not\equiv 0$, we have

$$\int_a^b |\mathbf{e} \cdot (T\mathbf{f})(\xi)|\,d\xi > M \int_c^d |\mathbf{e} \cdot \mathbf{f}(x)|\,dx.$$

We consider $\mathbf{e}$ to be a variable point of the sphere, integrate both sides of the above inequality over S, and then reverse the orders of integration to produce

$$\int_a^b \left\{\int_S |\mathbf{e} \cdot (T\mathbf{f})(\xi)|\,dS\right\} d\xi > M \int_c^d \left\{\int_S |\mathbf{e} \cdot \mathbf{f}(x)|\,dS\right\} dx.$$

The strict inequality prevails provided $\mathbf{e} \cdot \mathbf{f} \not\equiv 0$ for all $\mathbf{e} \in S$. This is clearly equivalent to our hypothesis $\mathbf{f} \not\equiv \mathbf{0}$.

We may simplify the inner integrals on both sides above by observing that if $\mathbf{a} \in R^N$ is a constant vector and $\mathbf{e} \in S$ varies, then by symmetry

$$\int_S |\mathbf{e} \cdot \mathbf{a}|\,dS = \|\mathbf{a}\| \cdot K,$$

where $\|\mathbf{a}\| = [\sum_{j=1}^N a_j^2]^{1/2}$ is the Euclidean norm of $\mathbf{a}$ and

$$K = \int_S |\mathbf{e} \cdot \mathbf{e}_1|\,dS = \frac{4}{N-1} \cdot \frac{\pi^{(N-1)/2}}{\Gamma\left(\frac{N-1}{2}\right)} > 0.$$

Thus

$$\int_a^b K\|(T\mathbf{f})(\xi)\|\,d\xi > M \int_c^d K\|\mathbf{f}(x)\|\,dx$$

which becomes (4.2) after the cancellation of $K > 0$. Q.E.D.

Using Theorem 4.1 with $V = A'$, $Tf = f'$, $c = a$, $d = b$, and $M = 4/(b-a)$, we have

THEOREM 4.2. *If $f_j \in A'$, $j = 1, 2, \cdots, N$, and at least one $f_j \not\equiv 0$, then*

$$\int_a^b \left[\sum_{j=1}^N \left(\frac{df_j}{dt}\right)^2\right]^{1/2} dt > \frac{4}{b-a} \int_a^b \left[\sum_{j=1}^N f_j^2\right]^{1/2} dt. \tag{4.3}$$

Equality never occurs.

The vector form of Theorem 4.2 is

THEOREM 4.3. *Let* $\mathbf{f}=(f_1,\cdots,f_N)$ *define a map of* $[a,b]$ *into* R^N. *If* $f_j\in A(1)$, $j=1,2,\cdots,N$, $\mathbf{f}(a)=\mathbf{f}(b)$, *and* $\mathbf{f}\not\equiv 0$, *then*

$$\int_a^b \mathbf{f}(t)\,dt=\mathbf{0} \text{ implies } \int_a^b \left\|\frac{d\mathbf{f}}{dt}\right\| dt>\frac{4}{b-a}\int_a^b \|\mathbf{f}\|\,dt. \tag{4.4}$$

Equality never occurs.

Application. We provide a geometrical interpretation of Theorem 4.3. Let $a=0$, $b=2\pi$ and consider $\mathbf{f}$ to be a map of the unit circle parameterized by θ. This map is well defined and the image C is a closed curve in R^N since $\mathbf{f}(0)=\mathbf{f}(2\pi)$. The term $\int_0^{2\pi}\mathbf{f}(\theta)\,d\theta$ is 2π times the "average" position of the curve C. The term

$$\frac{4}{2\pi}\int_0^{2\pi}\|\mathbf{f}(\theta)\|\,d\theta=4\cdot\frac{1}{2\pi}\int_0^{2\pi}\|\mathbf{f}(\theta)-\mathbf{0}\|\,d\theta$$

is just 4 times the "average" distance from the "average" position ($\mathbf{0}$ in this case) for the curve. It is important to note that in both cases "average" means the average with respect to the parameter θ of the domain circle.

The term $\int_0^{2\pi}\|d\mathbf{f}/d\theta\|\,d\theta$ is precisely the arc length of the closed curve. Since by translating R^N we can always assume that the "average" position of a closed curve is at the origin, we interpret (4.4) as follows:

The arc length of a closed curve in R^N is always greater than four times the "average" distance of the curve from its "average" position.

For the special case of a plane closed curve ($N=2$), we expect that the "average" distance from the "average" position is $\sqrt{A/\pi}$ where A is the area enclosed by the curve. (For $f(\theta)=(r\cos\theta, r\sin\theta)$ mapping the circle into a circle, this is exact.) Then (4.4) implies

$$4\cdot\sqrt{A/\pi}<L \quad\text{or}\quad L^2>\frac{16}{\pi}A,$$

an isoperimetric-type inequality.

In fact, Theorem 4.2 is used to prove an isoperimetric inequality for minimal surfaces. A minimal surface may be thought of as the shape assumed by the surface of a soap film spanning a wire loop. More precisely, a minimal surface is a surface with vanishing mean curvature (in the sense of differential geometry). Such surfaces in R^n are determined by n holomorphic functions $\phi_1,\cdots,\phi_n$ satisfying certain conditions.

Theorem 4.2 enters the proof of the isoperimetric inequality as follows: If $\int_0^{2\pi}\phi_j(r,\theta)\,d\theta=0$ for some fixed r, we split ϕ_j into its real and imaginary parts U_j+iV_j and conclude

$$\int_0^{2\pi}U_j(r,\theta)\,d\theta=\int_0^{2\pi}V_j(r,\theta)\,d\theta=0.$$

If we do this for $\phi_1,\cdots,\phi_n$, we have produced $2n$ functions U_j, $V_j\in A'$ and can apply Theorem 4.2 with $N=2n$. For each holomorphic function ϕ_j,

$$U_j^2+V_j^2=|U_j+iV_j|^2=|\phi_j|^2$$

and

$$\left(\frac{\partial U_j}{\partial\theta}\right)^2+\left(\frac{\partial V_j}{\partial\theta}\right)^2=\left|\frac{\partial U_j}{\partial\theta}+i\frac{\partial V_j}{\partial\theta}\right|^2=\left|\frac{\partial\phi_j}{\partial\theta}\right|^2=\left|iz\frac{d\phi_j}{dz}\right|^2$$

where only the last equality requires ϕ_j holomorphic.

We have proved

COROLLARY 4.1. *Let ϕ_j, $j=1,\cdots,n$ be n holomorphic functions on a domain containing the circle $|z|=r$. If at least one $\phi_j\not\equiv 0$ and if*

$$\int_0^{2\pi}\phi_j(r\,e^{i\theta})\,d\theta=0$$

for each j, then

$$\int_0^{2\pi}\left[\sum_{j=1}^n\left|r\frac{d\phi_j}{dz}(r\,e^{i\theta})\right|^2\right]^{1/2}d\theta>\frac{2}{\pi}\int_0^{2\pi}\left[\sum_{j=1}^n|\phi_j(r\,e^{i\theta})|^2\right]^{1/2}d\theta.$$

This corollary is used to provide a lower bound for the arc lengths of certain closed curves on minimal surfaces. This lower bound leads to the desired isoperimetric inequality. For details, see [5].

To derive a discrete analogue of Theorem 3.3, we let $v_1, v_2,\cdots,v_n\in R^N$, $\beta_1,\beta_2,\cdots,\beta_n\in R^+$, and assume

$$\sum_{i=1}^n\beta_i v_i=\mathbf{0}. \tag{4.5}$$

Let $\beta=\sum_{i=1}^n\beta_i$, and for $\varepsilon<\frac{1}{2}\min_i\beta_i$ define $f_{j,\varepsilon}(t)$, $j=1,2,\cdots,N$, as follows:

$$f_{j,\varepsilon}(t)=\begin{cases} v_{1,j}, & \text{if } \varepsilon\leqq t\leqq\beta_1-\varepsilon,\\ v_{2,j}, & \text{if } \beta_1+\varepsilon\leqq t\leqq\beta_1+\beta_2-\varepsilon,\\ \vdots & \\ v_{n,j}, & \text{if } \beta_1+\cdots+\beta_{n-1}+\varepsilon\leqq t\leqq\beta_1+\cdots+\beta_{n-1}+\beta_n-\varepsilon,\\ \text{linear in between} & \end{cases}$$

where $v_{i,j}$ is the jth component of the vector v_i. "Linear in between" includes between $v_{n,j}$ and $v_{n+1,j}\equiv v_{1,j}$, i.e. $\beta-\varepsilon\leqq t\leqq\beta$ and $0\leqq t\leqq\varepsilon$ so that $f_{j,\varepsilon}(0)=f_{j,\varepsilon}(\beta)$. Then

$$\begin{aligned}\int_0^\beta f_{j,\varepsilon}\,dt&=\sum_{i=1}^n v_{i,j}(\beta_i-2\varepsilon)+\sum_{i=1}^n\tfrac{1}{2}\cdot(2\varepsilon)\cdot(v_{i,j}+v_{i+1,j})\\&=\sum_{i=1}^n\beta_i v_{i,j}-2\varepsilon\sum_{i=1}^n v_{i,j}+\varepsilon\sum_{i=1}^n v_{i,j}+\varepsilon\sum_{i=1}^n v_{i,j}\\&=\left\{\sum_{i=1}^n\beta_i v_i\right\}_j\\&=0.\end{aligned}$$

Hence $f_{j,\varepsilon}\in A'$ for $j=1,2,\cdots,N$.

$$\int_0^\beta\left[\sum_{j=1}^N f_{j,\varepsilon}^2\right]^{1/2}dt=\sum_{i=1}^n(\beta_i-2\varepsilon)\|v_i\|+O(\varepsilon)\to\sum_{i=1}^n\beta_i\|v_i\|$$

as $\varepsilon\to0$ and

$$\begin{aligned}\int_0^\beta\left[\sum_{j=1}^N\left(\frac{df_{j,\varepsilon}}{dt}\right)^2\right]^{1/2}dt&=\sum_{i=1}^n\int_{-\varepsilon}^{\varepsilon}\left[\sum_{j=1}^N\left(\frac{v_{i+1,j}-v_{i,j}}{2\varepsilon}\right)^2\right]^{1/2}\\&=\sum_{i=1}^n\frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}\|v_{i+1}-v_i\|\,dt\\&=\sum_{i=1}^n\|v_{i+1}-v_i\|\end{aligned}$$

which is independent of β_i and ε, and where $v_{n+1} \equiv v_1$. Applying Theorem 4.2 to $\{f_{j,\varepsilon}\}$ and letting $\varepsilon \to 0$, we obtain

$$\sum_{i=1}^{n} \beta_i \cdot \sum_{i=1}^{n} \|v_{i+1} - v_i\| \geqq 4 \sum_{i=1}^{n} \beta_i \|v_i\| \quad \text{if } \sum_{i=1}^{n} \beta_i v_i = \mathbf{0}. \tag{4.6}$$

If we adjoin an additional vector v_a to the v_i's, along with $\beta_a = 0$, only the second sum on the left-hand side of (4.6) is altered. By the triangle inequality, $\|v_{j+1} - v_a\| + \|v_a - v_j\| \geqq \|v_{j+1} - v_j\|$ and hence the left-hand side is increased. Consequently, the β_i need be only nonnegative, not just positive. We conclude

THEOREM 4.4. *If $v_i \in R^N$, $\beta_i \in R^+ \cup \{0\}$ for $i = 1, 2, \cdots, N$, then*

$$\sum_{i=1}^{n} \beta_i \cdot \sum_{i=1}^{n} \|v_{i+1} - v_i\| \geqq 4 \cdot \sum_{i=1}^{n} \beta_i \|v_i\| \tag{4.7}$$

if $\sum_{i=1}^{n} \beta_i v_i = \mathbf{0}$ and $v_{n+1} = v_1$. Equality occurs in (4.7) only if

$$v_i = \begin{cases} v_1, & \text{if } i = 1, \cdots, j, \\ -v_1, & \text{if } i = j+1, \cdots, n \end{cases}$$

and $\sum_{i=1}^{j} \beta_i = \sum_{i=j+1}^{n} \beta_i$, or a cyclic permutation of these values.

COROLLARY 4.2. *For $v_i \in R^N$, we have*

$$\sum_{i=1}^{n} \|v_{i+1} - v_i\| \geqq \frac{4}{n} \sum_{i=1}^{n} \|v_i\| \quad \text{if } \sum_{i=1}^{n} v_i = \mathbf{0}. \tag{4.8}$$

Equality holds only if n is even, $v_i = v_1$ for $i = 1, \cdots, n/2$, $v_i = -v_1$ for $i = n/2+1, \cdots, n$, or a cyclic permutation of these values.

Proof. (4.8) follows directly from (4.7) with $\beta_i = 1$, $i = 1, 2, \cdots, n$. Q.E.D.

The article of Fan, Taussky, and Todd [4, p. 84] includes the following: If $z_i \in \mathbb{C}$ and $\sum_{i=1}^{n} z_i = 0$, then

$$\sum_{i=1}^{n} |z_{i+1} - z_i| \geqq \frac{2}{n-1} \sum_{i=1}^{n} |z_i|.$$

This is a weaker result than Corollary 4.2 for all values of n ($n \geqq \frac{1}{2}$).

We briefly consider a closed curve in R^N with parameter t (representing time) chosen so that

(1) The curve spends an equal nonzero time at each of the n points $v_1, \cdots, v_n \in R^N$ and

(2) The curve spends virtually no time at all travelling a straight line path from v_i to v_{i+1}.

The left-hand side of (4.8) is the arc length of the curve and the right-hand side is four times the time-average distance of the curve from the time-average position, the origin. We have thus recovered a special case of the application presented after Theorem 4.3.

COROLLARY 4.3. *If $v_i \in R^N$, $\|v_i\| \geqq R$ for $i = 1, 2, \cdots, n$, and if the zero vector is contained in the closed convex hull of the v_i, then*

$$\sum_{i=1}^{n} \|v_{i+1} - v_i\| \geqq 4R. \tag{4.9}$$

Proof. The zero vector in the closed convex hull implies the existence of $\{\beta_i\}$, $0 \leqq \beta_i \leqq 1$, $\sum_{i=1}^{n} \beta_i = 1$, such that $\mathbf{0} = \sum_{i=1}^{n} \beta_i v_i$. Using $\|v_i\| \geqq R$ in the right-hand side of (4.7), (4.9) follows immediately. Q.E.D.

Application 1. It is desired to devise a "global communications network" for a group of planets in space. The network must be able to handle all of the possible calls at any time.

Suppose the planets have coordinates $v_1, v_2, \cdots, v_n \in R^N$ and the ith planet has population $\beta_i \geqq 0$. By a translation of the coordinate system, we can assume that the "center of population" is at the origin, or that $\sum_{i=1}^n \beta_i v_i = \mathbf{0}$. Two possible methods for constructing the communications network, neither of which is generally optimal, are:

Method 1. Use a single cable of $\frac{1}{2}\sum_{i=1}^n \beta_i$ wires to connect all the planets together. We may connect the planets to each other in any order. Choose the order which minimizes the total distance, and by re-labeling if necessary, assume that this order is planet 1, planet 2, $\cdots$, planet n. Since the total population of the planets is $\sum_{i=1}^n \beta_i$, and since it takes two people for each conversation, $\frac{1}{2}\sum_{i=1}^n \beta_i$ wires in the cable will handle all possible conversations. The total wire length required for this method is at least

$$L_1 = \tfrac{1}{2}\left(\sum_{i=1}^{n} \beta_i\right) \cdot \left(\sum_{i=1}^{n-1} \|v_{i+1} - v_i\|\right) \tag{4.10}$$

or

Method 2. Use a single cable of β_i wires to connect the ith planet with the population center, the origin, in a straight line. Since the population of the ith planet is β_i, all possible conversations are allowed, but a horrendous switching problem (which we ignore) may occur at the population center. The total wire length required for this method is

$$L_2 = \sum_{i=1}^{n} \beta_i \|v_i\|. \tag{4.11}$$

To compare the two methods, we will use Theorem 4.4 and the following Lemma.

LEMMA 4.1. *If $w_i \in R^N$ and $\sum_{i=1}^n w_i = \mathbf{0}$, then*

$$\sum_{i=1}^{n-1} \|w_i\| \geqq \frac{1}{2} \sum_{i=1}^{n} \|w_i\|.$$

Proof. $w_n = -\sum_{i=1}^{n-1} w_i$ implies

$$\begin{aligned} \|w_n\| &= \left\|\sum_{i=1}^{n-1} w_i\right\| \\ &\leqq \sum_{i=1}^{n-1} \|w_i\|. \end{aligned}$$

Adding $\sum_{i=1}^{n-1} \|w_i\|$ to each side and dividing by 2 gives the required result. Q.E.D.

Since $\sum_{i=1}^n (v_{i+1} - v_i) = \mathbf{0}$ (where $v_{n+1} = v_1$), we apply Lemma 4.1 to produce

$$\sum_{i=1}^{n-1} \|v_{i+1} - v_i\| \geqq \frac{1}{2} \sum_{i=1}^{n} \|v_{i+1} - v_i\|.$$

Hence

$$\begin{aligned} L_1 &= \frac{1}{2}\left(\sum_{i=1}^{n} \beta_i\right) \cdot \left(\sum_{i=1}^{n-1} \|v_{i+1} - v_i\|\right) \\ &\geqq \frac{1}{2}\left(\sum_{i=1}^{n} \beta_i\right) \cdot \frac{1}{2}\left(\sum_{i=1}^{n} \|v_{i+1} - v_i\|\right) \\ &\geqq \frac{1}{4} \cdot 4 \cdot \sum_{i=1}^{n} \beta_i \|v_i\| \qquad \text{by (4.7)} \\ &= L_2. \end{aligned}$$

Hence $L_1 \geqq L_2$ with equality only for two equally populated planets, in which case the two methods coincide.

Application 2. A similar application can be made to the travelling salesman problem. Let each person in the universe desire one unit of a commodity which comes in a container. Assume that the universe has adopted strict "throw-away" container laws and hence all "empties" must be carried away by the salesman. Further assume that the "empties" are as costly to transport as the "fulls" (consider liquid oxygen in heavy metal pressure cylinders as an example).

Method 1 requires the salesman to carry $\sum_{i=1}^{n} \beta_i$ units for his grand tour (loop) of the universe, switching β_i "fulls" for β_i "empties" at the ith planet. Method 2 requires stockpiling all the units at a central warehouse at the population center of the universe, the origin. The salesman then makes short shuttle trips *back and forth* to each planet carrying the "fulls" one way and the "empties" the other. (We proceed by ignoring the potential construction cost of a warehouse in the middle of nowhere.)

If the transportation cost per container is proportional to the distance travelled, then Method 1 costs

$$C_1 = \left(\sum_{i=1}^{n} \beta_i \right) \cdot \left(\sum_{i=1}^{n-1} \|v_{i+1} - v_i\| \right) = 2L_1$$

and Method 2 costs

$$C_2 = \sum_{i=1}^{n} \beta_i \cdot 2 \cdot \|v_i\| = 2L_2.$$

L_1 and L_2 are defined in Application 1, and it is demonstrated there that $L_1 \geqq L_2$. We conclude $C_1 \geqq C_2$ with equality only for a universe composed of two equally populated planets, in which case the two methods coincide.

Application 3. Suppose a body at rest is acted upon by a set of forces $v_1, v_2, \cdots, v_n$. The total magnitude of these forces is $\sum_{i=1}^{n} \|v_i\| = M_1$. Since the body is at rest, we have $\sum_{i=1}^{n} v_i = \mathbf{0}$. Consider the system of "differences forces" $w_i = v_{i+1} - v_i$ where $v_{n+1} = v_1$. The total magnitude of these forces is $\sum_{i=1}^{n} \|v_{i+1} - v_1\| = M_2$. Since $\sum_{i=1}^{n} (v_{i+1} - v_i) = \mathbf{0}$ always, the body acted upon by the difference forces also remains at rest. Then (4.8) expresses an inequality between M_1 and M_2, or $M_2 \geqq (4/n)M_1$.

REFERENCES

[1] E. F. BECKENBACH AND R. BELLMAN, *Inequalities*, Springer-Verlag, Berlin, 1961.

[2] P. R. BEESACK, *Extensions of Wirtinger's inequality*, Trans. Roy. Soc. Canada, (3), 53 (1959), pp. 21–30.

[3] R. BELLMAN, *A note on periodic functions and their derivatives*, J. London Math. Soc., 18 (1943), pp. 140–142.

[4] K. FAN, O. TAUSSKY AND J. TODD, *Discrete analogues of inequalities of Wirtinger*, Monatsh. Math. Physik, 59 (1955), pp. 73–90.

[5] J. M. FEINBERG, *The isoperimetric inequality for doubly-connected minimal surfaces in R^n*, J. Analyse Math., 32 (1977), pp. 249–278.

[6] G. HARDY, J. LITTLEWOOD AND G. POLYA, *Inequalities*, Cambridge University Press, Cambridge, 1967.

[7] A. ZYGMUND, *Two notes on inequalities*, J. Mathematical Phys., 21 (1942), pp. 117–123.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0015 $01.00/0

GROWTH, OSCILLATION AND COMPARISON THEOREMS FOR SECOND ORDER LINEAR DIFFERENCE EQUATIONS*

WILLIAM T. PATULA†

Abstract. This paper studies homogeneous and nonhomogeneous second order linear difference equations. Comparison theorems based on the coefficients are proven for the homogeneous equation. Existence of so called r-type solutions for the forced equation is established. Oscillation and nonoscillation properties of the forced equation based on the forcing term and the associated homogeneous equation are discussed.

1. Introduction. In this paper, we continue the investigation begun in [8] of second order linear difference equations, where now we include both the homogeneous and nonhomogeneous cases, equations (1) and (2) respectively.

$$c_n x_{n+1} + c_{n-1} x_{n-1} = b_n x_n, \qquad c_n > 0, \quad n \geqq 0. \tag{1}$$

$$c_n z_{n+1} + c_{n-1} z_{n-1} = b_n z_n + f_n, \qquad c_n > 0, \quad n \geqq 0. \tag{2}$$

In [6] and [7], numerical techniques were developed to approximate solutions of (1) and (2), assuming the existence of an r-type solution (see § 3) of (2) and recessive and dominant solutions of (1). In this paper and in [8], we establish the existence of these solutions, under certain hypotheses, and prove other related results also. We feel that some of this theory might be useful in developing new or improved convergence schemes along the lines of those presented in [6] and [7].

A brief sketch of the contents of the paper is as follows. In § 2, we study growth and comparison properties of solutions of (1) and (2). In § 3, we establish some general existence theorems for r-type solutions and, applying the results of § 2, state specific criteria guaranteeing the existence of such solutions. In § 4, we investigate the oscillation and nonoscillation of solutions of (2).

We first make some preliminary remarks. Equation (1) has two linearly independent solutions, u and v say, which satisfy Abel's formula; namely, $c_n(u_{n+1}v_n - u_n v_{n+1}) = 1$, for all n. If $\hat{z}$ is a particular solution of (2), then any solution z of (2) has the form $z = \hat{z} + \alpha u + \beta v$, for some constants α and β. A nontrivial solution x of (1) or (2) will be called oscillatory if for any n, there exists a $k \geqq n$ such that $x_k x_{k+1} \leqq 0$. If one solution of (1) oscillates, all solutions of (1) oscillate. See [2, p. 221]. Thus for (1), a solution being oscillatory (nonoscillatory) is equivalent to the equation being oscillatory (nonoscillatory). This is not the case for (2), however. The second paragraph following Corollary 7 contains an example of a nonhomogeneous equation (2) which has both an oscillatory solution and a nonoscillatory solution. For our purposes, (2) being totally oscillatory means that all solutions oscillate. If we say that (2) is totally nonoscillatory, we mean that no solution oscillates. As mentioned, these two possibilities are not exhaustive when discussing (2). In an attempt to avoid ambiguity, we will usually assume the oscillation or nonoscillation of (1) is known and then state whether all, none or some of the solutions of (2) oscillate. For other properties and definitions concerning (1) and (2), we refer to the remarks in § 1 of [8] and to the books [1] and [2] in general.

2. Uniqueness and comparison theorems. In [3], it was shown that if any nontrivial solution x of the homogeneous equation (1) can have at most one value $x_k = 0$, then any two values x_n, x_m, $n \neq m$, uniquely determine a solution of (1). See also [8, Lemma 1]. For the forced equation (2), we make the following observation.

* Received by the editors February 13, 1978, and in final revised form December 18, 1978.
† Department of Mathematics, Southern Illinois University, Carbondale, Illinois 62901.

LEMMA 1. *Any nontrivial solution x of* (1) *can have at most one value $x_k = 0$ if and only if any two values z_n, z_m, $m \neq n$, uniquely determine a solution z of* (2).

Proof. (Only if). Suppose there are two solutions, z^1 and z^2, of (2) such that $z^1_m = z^2_m$ and $z^1_n = z^2_n$. Then the quantity $(z^1 - z^2)$ is a solution of (1) such that $(z^1 - z^2)_n = (z^1 - z^2)_m = 0$. This means $(z^1 - z^2)_n = 0$, for all n, or that $z^1 = z^2$.

(If). Suppose x is a solution of (1) such that $x_m = x_n = 0$, $m \neq n$. Define a solution z of (2) by setting $z_m = 0$ and $z_n = 0$. This is well defined, by hypothesis. Let $z^1 = z + x$. Then $z^1_m = z_m$ and $z^1_n = z_n$. However, this means $z^1_n = z_n$, for all n, or that $x_n = 0$, for all n. That is, x is the trivial solution.

In [7], the following lemma was proven.

LEMMA 2 ([7, Lemma 1]). *Suppose $|b_n| \geqq c_n + c_{n-1}$. If a nontrivial solution x of* (1) *satisfies $|x_{N+1}| \geqq |x_N|$, for some N, then $|x_{n+1}| \geqq |x_n|$, for all $n \geqq N$. In particular, x can have at most one value $x_k = 0$.*

We remark that if $x_{N+1} \geqq x_N \geqq 0$ and if $b_n \geqq c_n + c_{n-1}$, then $x_{n+1} \geqq x_n$, for all $n \geqq N$; i.e., the absolute value signs in the preceding lemma can all be omitted. One of the important aspects of Lemma 2 is that it provides a class of equations to which Lemma 1 can be applied.

As an analog of Lemma 2 for the forced equation (2), we have the following result.

THEOREM 1. *Suppose there is an integer $k \geqq 0$ such that for $n = k, \cdots, j$, we have* (a), *$z_k \geqq 0$ and $z_k \geqq z_{k-1}$, and* (b), *$b_n \geqq c_n + c_{n-1}$ and $f_n \geqq 0$. Then* (i) *$z_{n+1} \geqq z_n \geqq 0$, for $n = k, \cdots, j$.*

Suppose in addition to (a) *and* (b) *at least one of the following conditions holds; namely* (c), *$f_k > 0$,* (d) *$b_k > c_k + c_{k-1}$ and $z_k > 0$, or* (e) *$z_k > z_{k-1}$. Then* (ii) *$z_{n+1} > z_n$, for $n = k, \cdots, j$.*

Finally, suppose that (a) *and* (b) *are true for all $n \geqq k$ and we have either* (f) *$\sum^\infty f_n/c_n = \infty$, or* (g) *the existence of a sequence $\{\varepsilon_n\}$, $\varepsilon_n \geqq 0$, such that*

$$b_n \geqq (1 + \varepsilon_n)c_n + c_{n-1} \quad \textit{and} \quad \sum^\infty \varepsilon_n = \infty. \tag{3}$$

Then (iii) *$z_n \to \infty$ as $n \to \infty$.*

Proof. From (2), for $n = k$ we have

$$\begin{aligned} z_{n+1} &= (b_n z_n)/c_n - (c_{n-1} z_{n-1})/c_n + f_n/c_n \\ &\geqq (b_n/c_n - c_{n-1}/c_n) z_n + f_n/c_n, \end{aligned} \tag{4}$$

or

$$z_{n+1} \geqq z_n. \tag{5}$$

Since (5) is true for $n = k$, by induction we can conclude (5) holds for $n = k, \cdots, j$. This proves (i), and (ii) follows in a similar fashion.

For (iii), since $z_{n+1} - z_n \geqq f_n/c_n$, for $n \geqq k$, then $z_{n+k} - z_k \geqq \sum_{j=0}^{n-1} f_{j+k}/c_{j+k}$. Thus, if $\sum^\infty f_n/c_n = \infty$, $z_n \to \infty$ as $n \to \infty$. Finally, if condition (3) is satisfied, by (4) and (5) we have $z_{n+1} \geqq (1 + \varepsilon_n) z_n$, which again implies $z_n \to \infty$ as $n \to \infty$. This completes the proof.

COROLLARY 1. *Suppose $b_n \geqq c_n + c_{n-1}$, for $k \leqq n \leqq j$. Consider the solution z of* (2) *defined by $z_k = z_j = 0$, $j > k + 1$. If $f_n \geqq 0$ ($\leqq 0$), then $z_n \leqq 0$ ($\geqq 0$), $k \leqq n \leqq j$. If $f_{k+1} > 0$ (< 0), then $z_n < 0 (> 0)$, $k < n < j$.*

Proof. First, assume $f_n \geqq 0$, $k \leqq n \leqq j$. By Lemmas 1 and 2, z is well defined. If $z_{k+1} > 0$, Theorem 1 implies $z_j \neq 0$, a contradiction. Thus $z_{k+1} \leqq 0$. If $z_{k+2} > 0$, the arguments of Theorem 1 again imply $z_j \neq 0$, a contradiction. Continuing in this fashion, we conclude that $z_n \leqq 0$, $k \leqq n \leqq j$.

Suppose $f_{k+1} > 0$. If $z_{k+1} = 0$, after replacing n by $(k+1)$ in (4), we see that $z_{k+2} > 0$, a contradiction. Thus $z_{k+1} < 0$. Continuing in this fashion, we conclude $z_n < 0$, $k < n < j$.

If $f_n \leqq 0$, multiply (2) by -1 and reapply the preceding argument.

COROLLARY 2. *For all n sufficiently large, if $b_n \geqq c_n + c_{n-1}$, $f_n \geqq 0 (\leqq 0)$ and $\sum^\infty f_n/c_n = \infty(-\infty)$, then there exists a solution z of* (2) *such that $z_n \to \infty$ $(-\infty)$, as $n \to \infty$.*

Proof. The nonnegative case follows directly from Theorem 1 by a proper choice of initial conditions. For the nonpositive case, multiply (2) by -1 and reapply Theorem 1.

Corollary 2 proves the existence of at least one solution which diverges. It can happen that all solutions of (2) diverge. Consider the example

$$z_{n+1} + z_{n-1} = 2z_n + 1. \tag{6}$$

Linearly independent solutions of the associated homogeneous equation are $u_n = 1$ and $v_n = n$. A particular solution $\hat{z}$ of (6) is $\hat{z} = n(n+1)/2$. Since any solution z of (6) has the form $z = \hat{z} + \alpha u + \beta v$, it is clear that no choice of α or β will yield a solution z of (6) which is bounded. In this particular case, for every solution z, $z_n \to \infty$ as $n \to \infty$. We remark that under the hypothesis $b_n \geqq c_n + c_{n-1}$, this type of behavior cannot occur for the homogeneous equation (1). That is, if $b_n \geqq c_n + c_{n-1}$ for all n, then (1) has a bounded solution. See [8, Thm. 2].

We would next like to prove two comparison results for homogeneous equations. The first theorem is similar to Theorem 1 in [5]. See also [8, Thm. 4]. In addition to (1), we consider the equation

$$r_n w_{n+1} + r_{n-1} w_{n-1} = d_n w_n, \qquad r_n > 0. \tag{7}$$

THEOREM 2. *Suppose $b_n \geqq c_n + c_{n-1}$, $d_n \geqq r_n + r_{n-1}$, $c_n \geqq r_n$ and $b_n \leqq d_n$ for all n. If $(w_1 - x_1) \geqq (w_0 - x_0) \geqq 0$ and $x_1 \geqq x_0 \geqq 0$, then $(w_n - x_n) \geqq (w_{n-1} - x_{n-1}) \geqq 0$, for all n. In particular, $w_n \geqq x_n$, for all n.*

Proof. By hypothesis, $w_1 \geqq w_0 \geqq 0$, and so the remark following Lemma 2 implies that $w_n \geqq 0$, for all n. Consider (7) with several quantities added and subtracted.

$$\begin{aligned} c_n w_{n+1} + (r_n - c_n) w_{n+1} + c_{n-1} w_{n-1} + (r_{n-1} - c_{n-1}) w_{n-1} \\ = b_n w_n + (d_n - b_n) w_n. \end{aligned} \tag{8}$$

Since $(r_n - c_n) \leqq 0$ and $(d_n - b_n) \geqq 0$, for all n, we may write

$$c_n w_{n+1} + c_{n-1} w_{n-1} \geqq b_n w_n.$$

Hence

$$c_n (w_{n+1} - x_{n+1}) + c_{n-1}(w_{n-1} - x_{n-1}) \geqq b_n (w_n - x_n)$$

or

$$c_n (w_{n+1} - x_{n+1}) + c_{n-1}(w_{n-1} - x_{n-1}) = b_n (w_n - x_n) + \varepsilon_n, \qquad \varepsilon_n \geqq 0. \tag{9}$$

Define $z_n = (w_n - x_n)$. Since (9) is of the form (2) with $z_1 \geqq z_0 \geqq 0$, Theorem 1 implies $z_{n+1} \geqq z_n \geqq 0$, or that $w_{n+1} - x_{n+1} \geqq w_n - x_n \geqq 0$. This proves the theorem.

Following the definition given in [7], nontrivial solutions u and v of (1) are said to be recessive and dominant, respectively, if $u_n/v_n \to 0$, as $n \to \infty$. For existence and growth properties of recessive and dominant solutions of (1), see [8, Thms. 1 and 2].

Under the hypothesis of Theorem 2, dominant and recessive solutions exist for (1) and (7). Theorem 2 compares dominant solutions of (1) and (7). It is also possible to compare recessive solutions.

THEOREM 3. *Suppose $b_n \geqq c_n + c_{n-1}$, $d_n \geqq r_n + r_{n-1}$, $c_n \geqq r_n$ and $b_n \leqq d_n$, for all n. Suppose $x_0 \geqq w_0 > 0$, where x and w are the recessive solutions of* (1) *and* (7), *respectively.*

Then $x_n \geqq w_n$, *for all* n.

Proof. Since the equations are linear and since recessive solutions are unique except for a constant factor, we may assume $x_0 = w_0 = 1$. Let x^j be a sequence of solutions of (1) defined by

$$x_0^j = 1 \quad \text{and} \quad x_j^j = 0. \tag{10}$$

These are well-defined by the introductory remarks of § 2 and by Lemma 2. In [3], it is shown that $1 \geqq x_k^j \geqq x_{k+1}^j > 0$, for $k+1<j$, $x_k^{j+1} \geqq x_k^j$ for $k \leqq j$, and the solution x, defined by $x_k = \lim_{j\to\infty} x_k^j$, is the recessive solution. Similarly, for (7) we define a sequence of solutions w^j by

$$w_0^j = 1 \quad \text{and} \quad w_j^j = 0. \tag{11}$$

Using x^j and w^j, form (9) again and define $z_k^j = w_k^j - x_k^j$. Then $z_0^j = z_j^j = 0$. By Corollary 1, $z_k^j \leqq 0$ or $x_k^j \geqq w_k^j \geqq 0$, $0 \leqq k \leqq j$. This means $\lim_{j\to\infty} x_k^j \geqq \lim_{j\to\infty} w_k^j$, or that $x_k \geqq w_k$, for any k. This proves the theorem.

Notice that in comparing solutions of (1) and (7), a dominant solution of (7) is greater than or equal to a corresponding dominant solution of (1) while a recessive solution of (7) is less than or equal to a corresponding recessive solution of (1).

1. *r*-Type solutions. We define a particular solution z of (2) to be an r-type solution if $z_n/v_n \to 0$, as $n \to \infty$, where v is a dominant solution of (1). The definition, but not the name, was originally presented in [6]. In [6], numerical techniques were developed to approximate solutions of (2) based on the existence of an r-type solution of (2) and recessive and dominant solutions of (1). We will prove an existence theorem for r-type solutions and then use Theorems 2 and 3 to obtain specific criteria.

We will assume the existence of recessive and dominant solutions of (1). Note that two sufficient conditions implying the existence of such solutions are (i), (1) is non-oscillatory, which includes the case $b_n \geqq c_n + c_{n-1}$, or (ii), $b_n \leqq -c_n - c_{n-1}$. See [8].

THEOREM 4. *If* (1) *has a recessive solution* u *and a dominant solution* v *such that* $\sum^\infty f_i u_i$ *exists and* $(u_n/v_n)\sum_{i=1}^n f_i v_i \to 0$ *as* $n \to \infty$, *then* (2) *has an r-type solution.*

Proof. We may assume u and v are linearly independent solutions of (1) such that $c_n(u_n v_{n+1} - v_n u_{n+1}) = 1$, for all n. By the variation of constants formula, the general solution z of (2) has the form

$$\begin{aligned} z_n &= a u_n + b v_n - \sum_{i=1}^n f_i(u_n v_i - v_n u_i) \\ &= u_n\Big(a - \sum_{i=1}^n f_i v_i\Big) + v_n\Big(b + \sum_{i=1}^n f_i u_i\Big). \end{aligned} \tag{12}$$

If we let $b = -\sum_1^\infty f_i u_i$, then

$$z_n/v_n = (u_n/v_n)a - (u_n/v_n)\sum_{i=1}^n f_i v_i + \Big(\sum_{i=1}^\infty f_i u_i - \sum_{i=1}^n f_i u_i\Big). \tag{13}$$

From (13) and the hypotheses, we see that $z_n/v_n \to 0$, as $n \to \infty$.

COROLLARY 3. *If* $\sum_{i=1}^\infty f_i x_i$ *exists for every solution* x *of* (1) *and if* (1) *has recessive and dominant solutions, then* (2) *has an r-type solution.*

An example of when the conclusion of Theorem 4 does not hold is as follows. Let $b_n = 2$ and $c_n = 1$, so that $u_n = 1$ and $v_n = n$. Let $f_n = 1/n$. Then $\sum^\infty u_n f_n = \sum^\infty 1/n$ does not exist. In addition, $(u_n/v_n)\sum^n f_i v_i = (1/n)\sum^n (1/i)\ i = 1$, for all n. Equation (12) becomes

$$z_n = (a-n) + n\left(b + \sum^n 1/i\right),$$

or

$$z_n/n = (a-n)/n + b + \sum^n 1/i.$$

It follows that for any solution z of (2) and for any dominant solution v of (1), we have that $z_n/v_n \to \infty$, as $n \to \infty$. Therefore (2) does not have an r-type solution. The example following Corollary 2 also does not have any r-type solutions.

Although Corollary 3 covers several cases, it has at least one deficiency in that one is required to know something about the behavior of solutions of the homogeneous equation (1). It would be better to have sufficient conditions which guarantee the existence of an r-type solution based only on the coefficients c_n, b_n and f_n. Using Theorems 2 and 3, we can state the following.

THEOREM 5. *For all n if $b_n \geqq c_n + c_{n-1}$, $c_n \geqq a > 0$, $b_n \leqq d$, $d \geqq 2a$ and if $\sum^\infty |f_n(s)^n| < \infty$, where*

$$s = \begin{cases} [d + (d^2 - 4a^2)^{1/2}]/2a, & \text{if } d > 2a, \\ (n)^{1/n}, & \text{if } d = 2a, \end{cases}$$

then (2) *has an r-type solution.*

Proof. Since $b_n \geqq c_n + c_{n-1}$, (1) is nonoscillatory and has a nonincreasing recessive solution u and a nondecreasing dominant solution v, both of which are positive. See [8, Thm. 2]. Consider the equation

$$ax_{n+1} + ax_{n-1} = dx_n. \tag{14}$$

If $d > 2a$, then (14) has a dominant solution $v_n^1 = ([d + (d^2 - 4a^2)^{1/2}]/2a)^n$ and a recessive solution $u_n^1 = ([d - (d^2 - 4a^2)^{1/2}]/2a)^n$. By Theorems 2 and 3 and by a proper choice of initial conditions, we can assume $v_n^1 \geqq v_n \geqq u_n \geqq u_n^1 > 0$. By hypothesis $\sum^\infty |f_n v_n^1| < \infty$, which means the hypotheses of Corollary 3 are satisfied, and the result follows. The argument is trivial if $d = 2a$.

The hypotheses of Corollary 3 in general assume the existence of the sum $\sum^\infty x_n f_n$. For some remarks concerning the existence of this sum, we refer to the discussion following Theorem 8.

4. Nonoscillation and oscillation. In this section we present some theorems dealing with oscillation and nonoscillation of solutions of (2). Usually we will assume the oscillation or nonoscillation of (1) is known and, based on f_n, will attempt to determine the behavior of solutions of (2). We refer to § 1 for our general definitions and comments concerning oscillation and nonoscillation of solutions of (1) and (2). We will also assume that for any n, there exists a $k \geqq n$ such that $f_k \neq 0$. That is, we are excluding the case where f_n could eventually become identically equal to zero.

If x and z are solutions of either (1) or (2), we define $W(x, z)(n) = c_n(x_{n+1}z_n - z_{n+1}x_n)$. We first make the following observation, which is basically an extension of Abel's formula of § 1 to the nonhomogeneous equation (2).

LEMMA 3. *If x is a nontrivial solution of* (1) *and z a solution of* (2), *then for any $n > k$,*

$$W(x, z)(n) = -\left(\sum_{j=k+1}^{n} x_j f_j\right) + W(x, z)(k). \tag{15}$$

Proof. $W(x, z)(n) - W(x, z)(n-1) = c_n(x_{n+1}z_n - z_{n+1}x_n) - c_{n-1}(x_n z_{n-1} - z_n x_{n-1}) = z_n(b_n x_n - c_{n-1}x_{n-1}) - x_n(b_n z_n + f_n - c_{n-1}z_{n-1}) - c_{n-1}x_n z_{n-1} + c_{n-1}z_n x_{n-1} = -x_n f_n$. The result follows.

THEOREM 6. *For some nontrivial solution x of* (1) *and some solution z of* (2), *suppose the quantity* $W(x, z)(n)$ *is eventually of one sign* ($\geqq 0$ *or* $\leqq 0$). *Then* (1) *nonoscillatory if and only if z is a nonoscillatory solution of* (2), *which is equivalent to stating that* (1) *is oscillatory if and only if z is an oscillatory solution of* (2).

Proof. (Only if). Suppose (1) is nonoscillatory and $c_n(x_{n+1}z_n - z_{n+1}x_n) \geqq 0$, for $n \geqq k$. We may assume k is large enough so that x_n is of one sign, say positive, for $n \geqq k$. Then $c_n x_{n+1} z_n \geqq c_n z_{n+1} x_n$ or $x_{n+1}z_n/x_n \geqq z_{n+1}$, for $n \geqq k$. Let n_1 be the first integer $\geqq k$ such that $z_{n_1} \leqq 0$, if such an integer exists. Then $z_n \leqq 0$, for $n \geqq n_1$. If $z_n \not\equiv 0$, for $n \geqq n_1$, there exists an integer $n_2 \geqq n_1$ such that $z_{n_2} < 0$, which means $z_n < 0$, $n \geqq n_2$. If $z_n = 0$ for all $n \geqq n_1$ then $f_n = 0$ for $n \geqq n_1$, which possibility we are excluding. If n_1 does not exist, then $z_n > 0$, for $n \geqq k$. In either case, z is nonoscillatory. The arguments are similar if we had assumed $c_n(x_{n+1}z_n - z_{n+1}x_n) \leqq 0$ or $x_n < 0$.

(If). Next, assume z is nonoscillatory, say positive, for $n \geqq k$, and assume $c_n(x_{n+1}z_n - z_{n+1}x_n)$ is of one sign, say $\geqq 0$, for $n \geqq k$. However, suppose x is an oscillatory solution of (1). Since c_n is positive, $x_{n+1} \geqq x_n z_{n+1}/z_n$. Choose a value $n_1 \geqq k$ such that x_{n_1} is positive. One such value must exist, since (1) cannot have a nontrivial oscillatory solution which is $\leqq 0$. (However, (2) can.) Then the previous inequality implies that $x_n > 0$, for $n \geqq n_1$, a contradiction. A similar argument holds if $c_n(x_{n+1}z_n - z_{n+1}x_n) \leqq 0$ or if z_n is eventually negative. This proves the theorem.

COROLLARY 4. *If* (1) *is nonoscillatory and* f_n *is eventually of one sign* ($\geqq 0$ or $\leqq 0$), *then every solution of* (2) *is nonoscillatory.*

Proof. Let x be any solution of (1) and z be any solution of (2). We may choose k large enough so that $x_n f_n$ is of one sign for all $n \geqq k$. Thus all the terms of the sum in (15) will be of one sign and the term $W(x, z)(k)$ is merely a constant. This means that eventually $W(x, z)(n)$ will be of one sign, and the result follows from Theorem 6.

COROLLARY 5. *If* (1) *is oscillatory* (*nonoscillatory*) *and if there exists a solution x of* (1) *such that* $\sum^{\infty} x_n f_n = +\infty$ *or* $-\infty$, *then every solution of* (2) *is oscillatory* (*nonoscillatory*).

Proof. The hypothesis implies that for any solution z of (2), the quantity $W(x, z)(n)$ in (15) must eventually be of one sign.

COROLLARY 6. *Suppose* f_n *has the form* $a_n x_n$, *where* a_n *is of one sign* ($\geqq 0$ *or* $\leqq 0$) *and* x_n *is a solution of* (1). *If* (1) *is oscillatory* (*nonoscillatory*), *then every solution of* (2) *is oscillatory* (*nonoscillatory*).

Proof. The argument is the same as the one for Corollary 5.

COROLLARY 7. *For all n sufficiently large, if* $b_n \leqq -c_n - c_{n-1}$ *and* $f_n = (-1)^n a_n$, *where* a_n *is of one sign, then every solution of* (2) *oscillates.*

Proof. The hypothesis $b_n \leqq -c_n - c_{n-1}$ implies that every solution of (1) oscillates and eventually alternates in sign. See [8, § 4]. Thus $x_n f_n = x_n(-1)^n a_n$ will eventually be of one sign. The proof of Corollary 5 now applies.

We next consider several examples. The conclusion of Corollary 4 may no longer be true if f_n is allowed to change sign. Let $b_n = 1$, $c_n = \frac{1}{2}$, and $f_n = 4(-1)^{n+1}$. Linearly independent solutions of (1) are $u_n = 1$ and $v_n = n$. A solution of the forced equation is $z_n = 1 + 2(-1)^n$, which clearly oscillates.

Parts of Corollaries 5, 6 and 7 assume (1) is oscillatory and state sufficient conditions for every solution of (2) to oscillate. However, (1) being oscillatory does not always imply that every solution of (2) must oscillate. For example, let $c_n = 1$, $b_n = -2$ and $f_n = (4n^2 - 2)/(n^3 - n)$. Then (1) has the oscillatory solutions $(-1)^n$ and

$n(-1)^n$ while $z_n = 1/n$ is a nonoscillatory solution of (2). In this example, $z_n = 1/n$ is clearly a unique nonoscillatory solution of (2). The example actually illustrates a general result.

THEOREM 7. *If* (1) *is oscillatory and if f_n is eventually of one sign* ($\geqq 0$ *or* $\leqq 0$), *then any nonoscillatory solution of* (2) *must eventually be of the same sign as f_n.*

Proof. Suppose f_n is eventually of one sign, say $\geqq 0$. Suppose also that (1) is oscillatory and z is a nonoscillatory solution of (2) such that $z_n < 0$, for $n \geqq k$, for some k. Rewriting (2) we have

$$c_n z_{n+1} + c_{n-1} z_{n-1} = (b_n + f_n/z_n) z_n.$$

This is now a homogeneous equation of the form (1) with a nonoscillatory solution z and with $(b_n + f_n/z_n) \leqq b_n$. Therefore, by a comparison theorem of Fort [2, p. 222], we conclude that $c_n x_{n+1} + c_{n-1} x_{n-1} = b_n x_n$ is nonoscillatory, a contradiction. Note the term $G_2(n)$ in [2, p. 222] has the form $G_2(n-1) = b_n - c_n - c_{n-1}$ and the term $K_2(n) = c_n$.

As previously mentioned, the example preceding Theorem 7 shows that if (1) is oscillatory, it does not have to be the case that all solutions of (2) oscillate. In a similar fashion, the example following Corollary 7 indicates that if (1) is nonoscillatory, all solutions of (2) need not be nonoscillatory. We next investigate the number of possible nonoscillatory (oscillatory) solutions of (2) under the assumption that (1) is oscillatory (nonoscillatory).

THEOREM 8. *Suppose $\sum^\infty x_n f_n$ exists, for every solution x of* (1). *If* (1) *is oscillatory, then* (2) *has at most one nonoscillatory solution. If* (1) *is nonoscillatory, then* (2) *has at most one oscillatory solution.*

Proof. Suppose (1) is oscillatory, and suppose $\hat{z}$ is a nonoscillatory solution of (2). Consider any other solution z of (2) of the form

$$z = \hat{z} + cu, \qquad c \neq 0, \tag{16}$$

where u is a solution of (1). We must show z oscillates.

Let v be a solution of (1) which is linearly independent of u such that $W(u, v)(n) = 1$. Then

$$W(z, v) = W(\hat{z} + cu, v) = W(\hat{z}, v) + cW(u, v). \tag{17}$$

From (15) and the hypothesis, $\lim_{n\to\infty} W(\hat{z}, v)(n)$ exists. If the limit is nonzero, since v oscillates, Theorem 6 would imply $\hat{z}$ oscillates, a contradiction. Thus $W(\hat{z}, v)(n) \to 0$, as $n \to \infty$. But this means that in (17) $W(z, v)(n) \to c \neq 0$, as $n \to \infty$. Since v oscillates, Theorem 6 implies z oscillates.

Next, assume (1) is nonoscillatory and suppose $\hat{z}$ is an oscillatory solution of (2). Choose z, c, u and v as above. Based on the preceding argument, we must have $\lim_{n\to\infty} W(\hat{z}, v) = 0$. Again, this implies that in (17) $\lim_{n\to\infty} W(z, v)(n) = c \neq 0$, which by Theorem 6 means that z is nonoscillatory. This proves the theorem.

In view of Theorem 8 (and Corollary 3), it would be interesting to know when the hypothesis concerning the convergence of $\sum^\infty x_n f_n$ is satisfied for any solution x of (1). If $f \in l^2$ and all solutions of (1) are square summable, in which case we say (1) is limit circle, then obviously $\sum^\infty x_n f_n$ exists for any solution x of (1). Sufficient conditions which guarantee that (1) is limit circle can be found in [4, Thm. 12]. Note that $f \in l^p$, for any $p > 0$, iff $\sum^\infty |f|^p < \infty$.

Another situation insuring the existence of $\sum^\infty x_n f_n$ would be when $f \in l^1$ and all solutions of (1) are bounded. (By (12), this would also imply that all solutions of (2) are bounded). We conclude by making several observations concerning the boundedness of all solutions of (1).

THEOREM 9. *If* $b_n - c_n - c_{n-1} \leqq 0$, (1) *is nonoscillatory and* $\sum^\infty 1/c_n < \infty$, *then all solutions of* (1) *are bounded.*

Proof. For the proof see [8, Theorem 3].

Consider the equation

$$c_n x_{n+1} + c_{n-1} x_{n-1} = (b_n + g_n) x_n. \tag{18}$$

THEOREM 10. *If* $\sum^\infty |g_n| < \infty$ *and all solutions of* (1) *are bounded, then all solutions of* (18) *are bounded.*

Proof. For the proof see [4, Lemma 2].

THEOREM 11. *Suppose* $\sum^\infty |b_n| < \infty$. *If the sequence* $\{c_n\}$ *is eventually either nondecreasing or nonincreasing and bounded below by a positive constant, then all solutions of* (1) *are bounded.*

Proof. Consider the equation

$$c_n u_{n+1} + c_{n-1} u_{n-1} = 0. \tag{19}$$

We may write

$$u_{n+1} = -(c_{n-1}/c_n) u_{n-1} = (-1)^2 (c_{n-1}/c_n)(c_{n-3}/c_{n-2}) u_{n-3},$$

or

$$|u_{n+1}| = (c_{n-1}/c_n)(c_{n-3}/c_{n-2}) \cdots (c_k/c_{k+1}) |u_k|, \tag{20}$$

where $k = N$ or $N+1$. If c_n is nondecreasing for $n \geqq N$, we have $|u_{n+1}| \leqq |u_k|$, for any $n > N$, which means the solution u is bounded.

Next, suppose c_n is nonincreasing and bounded below by a positive constant, say ε, for $n \geqq N$. From (20), we have

$$|u_{n+1}| \leqq (c_k/c_n)|u_k| \leqq (c_k/\varepsilon)|u_k|,$$

which again implies that u is bounded. Since all solutions of (19) are bounded and since $\sum^\infty |b_n| < \infty$, Theorem 10 implies that all solutions of $c_n x_{n+1} + c_{n-1} x_{n-1} = (0 + b_n) x_n$ are also bounded. This proves the theorem.

Acknowledgments. We are grateful to the referee for several helpful suggestions and comments.

REFERENCES

[1] F. V. ATKINSON, *Discrete and Continuous Boundary Value Problems*, Academic Press, New York, 1964.

[2] T. FORT, *Finite Differences and Difference Equations in the Real Domain*, Oxford University Press, London, 1948.

[3] P. HARTMAN AND A. WINTNER, *On linear difference equations of the second order*, Amer. J. Math., 72 (1950), pp. 124–128.

[4] D. HINTON AND R. LEWIS, *Spectral analysis of second order difference equations*, J. Math. Anal. Appl., 63 (1978), pp. 421–438.

[5] F. W. J. OLVER, *Bounds for the solutions of second order linear difference equations*, J. Res. Nat. Bur. Standards Sect. B, 71 (1967), pp. 161–166.

[6] ———, *Numerical solution of second-order linear difference equations*, J. Res. Nat. Bur. Standards Sect. B, 71 (1967), pp. 111–129.

[7] F. W. J. OLVER AND D. J. SOOKNE, *Note on backward recurrence algorithms*, Math. Comput., no. 120, 26 (1972), pp. 941–947.

[8] W. T. PATULA, *Growth and oscillation properties of second order linear difference equations*, this Journal, 10 (1979), pp. 55–61.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0016 $01.00/0

LOWER BOUNDS FOR THE FUNDAMENTAL FREE MEMBRANE EIGENVALUE*

J. LADEVÈZE† AND P. LADEVÈZE†

Abstract. New lower bounds for the fundamental eigenvalue in the free membrane problem are derived in terms of elementary characteristics of the domain. In the first part, the case of strongly starshaped domains will be dealt with. The second part will introduce a method of estimate for domains which are unions of strongly sharshaped domains.

1. Introduction. The lower bounds for the fundamental eigenvalue in the free membrane problem established in this paper are explicitly defined as functions of elementary characteristics of the domain (area, diameter, $\cdots$). For a large class of domains, they improve the bounds obtained in [1], [3], [4] and [6].

To construct a priori estimates, one generally uses the inverse of the fundamental frequency, called the Poincaré constant. Thus in the following, the results will refer to the Poincaré constant.

First of all, the case of strongly starshaped domains is investigated. A domain Ω is said to be *strongly starshaped* if the interior of the set of points of Ω for which Ω is sharshaped is not empty. The interior of this set, the so-called *starshaping domain* of Ω, is obviously a convex domain. In particular, the starshaping domain of a convex domain is the domain itself. Two bounds, (P_{*1}) and (P_{*2}), of the Poincaré constant are obtained; it can be noticed that (P_{*2}) is optimal, i.e., it is the exact value for convex domains.

Furthermore, we establish the inequalities (Π), (Π') and (Π'') in connection with the Poincaré inequality; beyond their own interest, they will prove useful to obtain upper bounds of the Steklov constant in [5].

Finally, we give a method leading to bounds of the Poincaré constant for domains which are the union of strongly starshaped domains.

The proofs and the results are given in the two-dimensional case; they can easily be extended to the three-dimensional one.

2. Definitions and notations. $\mathscr{E}_2$ denotes the usual two-dimensional oriented Euclidean point space while E_2 denotes its associated vector space. i_2 is to represent the $\pi/2$-rotation operator on E_2.

Ω is assumed to be a bounded domain of $\mathscr{E}_2$ with C^0 boundary $\partial\Omega$; M is the current point of Ω. Moreover, Ω satisfies the segment property; therefore the function set $C^1_*(\Omega)$ is dense in $H^1(\Omega)$. Ω', with boundary $\partial\Omega'$, denotes a subdomain of Ω.

α is a nonzero real function defined on Ω which belongs to $H^1(\Omega)$. α is said to satisfy the condition $\mathscr{C}(\Omega')$ provided its mean value over Ω' is zero.

The first nonzero eigenvalue of the Neumann problem relative to the Laplacian operator for functions of $H^1(\Omega)$ with zero mean value over Ω' is defined by:

$$\mu_2[\mathscr{C}(\Omega')] = \inf_{\alpha\in H^1(\Omega)/\mathscr{C}(\Omega')} \frac{\|\operatorname{grad}\alpha\|^2_{L^2(\Omega)}}{\|\alpha\|^2_{L^2(\Omega)}}.$$

Therefore, the Poincaré constant corresponding to the condition $\mathscr{C}(\Omega')$ is: $\mu_2^{-1/2}[\mathscr{C}(\Omega')]$.

* Received by the editors May 26, 1978, and in revised form November 20, 1978.

† Institut de Mécanique Théorique et Appliquée, Université Pierre et Marie Curie, 4, place Jussieu, Paris, France.

Notations and geometric characteristics.

$\|\cdot\|_K$ represents the $L^2(K)$-norm.

For a strongly starshaped domain Ω, Ω^* denotes its so-called starshaping domain. We note:

A (resp. A^*, A') = measure (Ω) (resp. Ω^*, Ω');

L (resp. L^*, L') = measure ($\partial\Omega$) (resp. $\partial\Omega^*, \partial\Omega'$);

Φ (resp. Φ^*, Φ') = diameter (Ω) (resp. Ω^*, Ω').

Remark. In the proofs, the function α is assumed to belong to $C^1_*(\Omega)$; the final results are obtained by density.

3. Upper bound ($\mathbf{P}_{*1}$). We begin with the identity

$$(1)\quad \int_{\Omega\times\Omega^*}[\alpha(M)-\alpha(M^*)]^2\,\underline{dM}\,\underline{dM^*}=A^*\int_\Omega \alpha^2(M)\,\underline{dM}+A\int_{\Omega^*}\alpha^2(M^*)\,\underline{dM^*}$$

which holds for functions α *which satisfy the condition* $\mathscr{C}(\Omega)$ *or* $\mathscr{C}(\Omega^*)$.

$\underset{\sim}{\alpha}(\Omega^*)$ denotes the mean value of α and Ω^* and we introduce the following function:

$$\mathring{\alpha}(\Omega^*)=\alpha-\underset{\sim}{\alpha}(\Omega^*);$$

Ω' being the interior of $\Omega-\Omega^*$, we get from (1)

$$\|\alpha\|^2_\Omega=\frac{1}{A^*}\int_{\Omega'\times\Omega^*}[\alpha(M')-\alpha(M^*)]^2\,\underline{dM'}\,\underline{dM^*}+\left(2-\frac{A}{A^*}\right)\|\mathring{\alpha}(\Omega^*)\|^2_{\Omega^*}-A\underset{\sim}{\alpha}^2(\Omega^*)$$

and therefore the bound

$$(2)\quad \|\alpha\|^2_\Omega\leqq\frac{1}{A^*}\int_{\Omega'\times\Omega^*}[\alpha(M')-\alpha(M^*)]^2\,\underline{dM'}\,\underline{dM^*}+\max\left\{0,2-\frac{A}{A^*}\right\}\|\mathring{\alpha}(\Omega^*)\|^2_{\Omega^*}.$$

3.1. Estimate of the first contribution in (2). In a first stage, we introduce on $\Omega\times\Omega^*$ a system of coordinates which plays an essential part in this work.

M' (resp. M^*) is the current point of Ω' (resp. Ω^*). O is a point of $\mathscr{E}_2$ taken as the origin while H is its orthogonal projection on the line (M', M^*) (see Fig. 1).

Let us denote ω_H as follows:

$$\omega_H=\{H\,|\,H\in\mathscr{E}_2-\{0\};\,H=\mathrm{Proj}_{(M',M^*)}(0),\,M'\in\Omega',\,M^*\in\Omega^*\}$$

and the interior domain of ω_H by Ω_H.

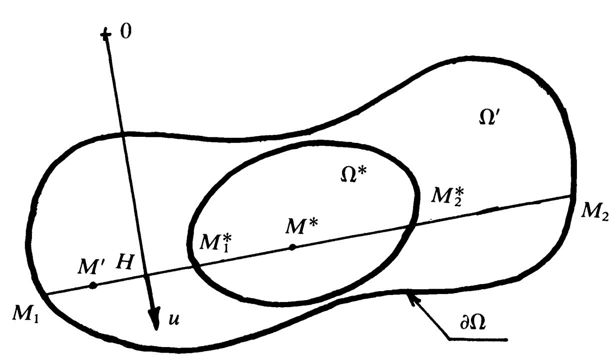

FIG. 1

H is now to be the current point of Ω_H and we set: $u(H)=OH/|OH|$. M_1 (resp. M_1^*) and M_2 (resp. M_2^*) are the intersection points of the boundary $\partial\Omega$

(resp. $\partial\Omega^*$) with the line $(H, i_2u(H))$. We note:

$$
\begin{aligned}
&x_1(H) = HM_1 \cdot i_2u(H),\\
&x_2(H) = HM_2 \cdot i_2u(H) \qquad (x_1(H) \leqq x_2(H)),\\
&\mathcal{T}(H) =]x_1(H), x_2(H)[,\\
&x_1^*(H) = HM_1^* \cdot i_2u(H),\\
&x_2^*(H) = HM_2^* \cdot i_2u(H) \qquad (x_1^*(H) \leqq x_2^*(H)),\\
&\mathcal{T}^*(H) =]x_1^*(H), x_2^*(H)[\subset \mathcal{T}(H).
\end{aligned}
$$

With these notations, we can define the following coordinate system:

$$
\begin{aligned}
F\colon (H, x, x^*) &\mapsto (M, M^*) = (OH + xi_2u(H), OH + x^*i_2u(H)),\\
\Sigma &\to \Omega \times \Omega^*
\end{aligned}
$$

where Σ is the manifold: $\Sigma = \{(H, x, x^*) \mid H \in \Omega_H, x \in \mathcal{T}(H), x^* \in \mathcal{T}^*(H)\}$.

Furthermore $F(\Sigma)$ is equal to $\Omega \times \Omega^*$ modulo a set of zero measure. Setting Σ' as the manifold $\{(H, x', x^*) \mid H \in \Omega_H, x' \in \mathcal{T}(H) - \overline{\mathcal{T}^*(H)}, x^* \in \mathcal{T}^*(H)\}$ we also get that $F(\Sigma')$, up to a set of zero measure, is equal to $\Omega' \times \Omega^*$.

Consequently, we get:

$$
\begin{aligned}
&\int_{\Omega' \times \Omega^*} [\alpha(M') - \alpha(M^*)]^2 \, \underline{dM'} \, \underline{dM^*}\\
&\quad = \int_{\Omega_H} \frac{1}{|OH|} \underline{dH} \int_{\mathcal{T}(H) - \overline{\mathcal{T}^*(H)}} dx' \int_{\mathcal{T}^*(H)} dx^* [\hat{\alpha}(H, x') - \hat{\alpha}(H, x^*)]^2 |x' - x^*|.
\end{aligned} \tag{3}
$$

We set:

$$
\begin{aligned}
&\delta(H) = x_2(H) - x_1(H),\\
&\varepsilon(H) = \frac{x_2^*(H) - x_1^*(H)}{x_2(H) - x_1(H)} \leqq 1,\\
&X = \frac{2}{\delta(H)}[x - \tfrac{1}{2}(x_1(H) + x_2(H))],\\
&a(H) = \frac{1}{\delta(H)}\{[x_1^*(H) + x_2^*(H)] - [x_1(H) + x_2(H)]\},\\
&\mathcal{O}^*(H) =]a(H) - \varepsilon(H), a(H) + \varepsilon(H)[\subset]-1, +1[.
\end{aligned}
$$

Using these notations, we obtain

$$
\int_{\Omega' \times \Omega^*} [\alpha(M') - \alpha(M^*)]^2 \, \underline{dM'} \, \underline{dM^*} = \int_{\Omega_H} \frac{1}{|OH|} \underline{dH} \frac{\delta^3(H)}{8} \mathcal{A}
$$

where

$$
\mathcal{A} = \int_{-1}^{1} [1 - g(X')] \, dX' \int_{-1}^{1} g(x^*) \, dX^* |X' - X^*| [\tilde{\alpha}(H, X') - \tilde{\alpha}(H, X^*)]^2
$$

and g is the characteristic function of the interval $\mathcal{O}^*(H)$ defined on $]-1, +1[$. $\mathcal{A}$ can be written as follows:

$$\mathscr{A} = \int_{-1}^{1} \int_{-1}^{1} |X - X'|[\tilde{\alpha}(H, X) - \tilde{\alpha}(H, X')]^2 \tfrac{1}{2}\{g(X')[1 - g(X)] + g(X)[1 - g(X')]\}\, dX\, dX'$$

$$= \int_{-1}^{1} dX \int_{-1}^{X} dX' |X - X'| \{g(X')[1 - g(X)] + g(X)[1 - g(X')]\} \int_{X'}^{X} f(\eta)\, d\eta \int_{X'}^{X} f(\tau)\, d\tau$$

in which $f(\eta) = \partial\tilde{\alpha}/\partial\eta\ (H, \eta)$.

Permuting the integration order and using the Cauchy–Schwarz inequality, we find the estimate:

$$\mathscr{A} = \int_{-1}^{1} \int_{-1}^{1} f(\eta) f(\tau) K(\eta, \tau)\, d\eta\, d\tau \leqq \int_{-1}^{1} f^2(\eta) G(\eta)\, d\eta \cdot \left[\int_{-1}^{1} \int_{-1}^{1} \frac{K^2(\eta, \tau)}{G(\eta) G(\tau)}\, d\eta\, d\tau\right]^{1/2}$$

where

$$K(\eta, \tau) = \int_{\max\{\eta, \tau\}}^{1} dX \int_{-1}^{\min\{\eta, \tau\}} dX' |X - X'| \{g(X')[1 - g(X)] + g(X)[1 - g(X')]\}$$

and

$$G(X) = \int_{-1}^{1} |X - X'| g(X')\, dX'.$$

The quantity $\int_{-1}^{1} \int_{-1}^{1} K^2(\eta, \tau)/(G(\eta) G(\tau))\, d\eta\, d\tau$, considered as a function of $\varepsilon(H)$ and $a(H)$, can be uniformly bounded with respect to $a(H)$. We have established that its maximum is obtained for $|a(H)| = 1 - \varepsilon(H)$; we get:

$$\left[\int_{-1}^{1} \int_{-1}^{1} \frac{K^2(\eta, \tau)}{G(\eta) G(\tau)}\, d\eta\, d\tau\right]^{1/2} \leqq \varphi(\varepsilon(H))$$

in which φ is a function given in the Appendix.

However, the upper bound of the function φ on $]0, 1[$:

$$\varphi(\nu) \leqq 2\ \mathrm{Log}\,(1/\nu) + 2$$

has been computed.

We therefore deduce the following bound for the first contribution in (2):

$$\begin{aligned} &\frac{1}{A^*} \int_{\Omega' \times \Omega^*} [\alpha(M') - \alpha(M^*)]^2\, \underline{dM'}\, \underline{dM^*} \\ &\leqq \frac{1}{A^*} \int_{\Omega_H} \frac{\delta^2(H)}{4|OH|} \underline{dH} \int_{\mathscr{T}(H) \times \mathscr{T}^*(H)} |\mathrm{grad}_M\, \alpha|^2 \varphi(\varepsilon(H)) |x - x^*|\, dx\, dx^* \\ &= \frac{1}{A^*} \int_{\Omega \times \Omega^*} |\mathrm{grad}_M\, \alpha|^2 \frac{\delta^2}{4}(H) \varphi(\varepsilon(H))\, \underline{dM}\, \underline{dM^*} \\ &\leqq \frac{\Phi^2}{2} \left\{1 + \sup_{M \in \Omega} \left[\frac{1}{A^*} \int_{\Omega^*} \mathrm{Log}\,(1/\varepsilon(H))\, \underline{dM^*}\right]\right\} \|\mathrm{grad}\ \alpha\|_{\Omega}^2. \end{aligned} \tag{4}$$

In the Appendix, explicit bounds of $\sup_{M \in \Omega} [(1/A^*) \int_{\Omega^*} \mathrm{Log}\,(1/\varepsilon(H))\, \underline{dM^*}]$ are given.

3.2. Estimate of the norm $\|\mathring{\alpha}(\Omega^*)\|_{\Omega^*}$. As Ω^* is a convex domain and the function $\mathring{\alpha}(\Omega^*)$ has a zero mean value over Ω^*, we can use the result obtained by Payne and Weinberger [7]:

$$\|\mathring{\alpha}(\Omega^*)\|^2_{\Omega^*} \leqq \frac{\Phi^{*2}}{\pi^2}\|\text{grad }\alpha\|^2_{\Omega^*}. \tag{5}$$

Remark. Noting that

$$2A^*\|\mathring{\alpha}(\Omega^*)\|^2_{\Omega^*} = \int_{\Omega^*\times\Omega^*} [\alpha(M)-\alpha(M')]^2 \, \underline{dM} \, \underline{dM'},$$

we can apply the method already used and we obtain the following bound:

$$\|\mathring{\alpha}(\Omega^*)\|^2_{\Omega^*} \leqq \frac{\sqrt{3}}{8}\Phi^{*2}\|\text{grad }\alpha\|^2_{\Omega^*}.$$

This result is close to that of Payne and Weinberger (the constant $1/\pi^2$ being changed into $\sqrt{3}/8$).

3.3. Upper bound ($\mathbf{P_{*1}}$). From the previous estimates (2), (4), (5), we get for the functions satisfying the condition $\mathscr{C}(\Omega)$ or $\mathscr{C}(\Omega^*)$

$$\|\alpha\|_\Omega \leqq P_{*1}\|\text{grad }\alpha\|_\Omega$$

$$\text{with } P_{*1} = \Phi\left[\tfrac{1}{2}E + \frac{1}{\pi^2}\left(\frac{\Phi^*}{\Phi}\right)^2 \max\left\{0, 2-\frac{A}{A^*}\right\}\right]^{1/2}, \tag{P_{*1}}$$

$$\text{where } E = 1 + \min\left\{\text{Log}\left[(\sqrt{2}+1)\frac{A}{A^*}\right], \text{Log}\left[\frac{\Phi L^*}{2A^*}\right]\right\}.$$

3.4. Comments. The use of the bound (P_{*1}) for an L-shaped domain (see Fig. 2) gives for $l/e > \frac{3}{2}$:

$$P_{*1} = l\left[1 + \text{Log}\left(2^{3/2}\frac{l}{e}\right)\right]^{1/2}.$$

This result is similar to the one obtained by using the upper bound [4; 3.3].

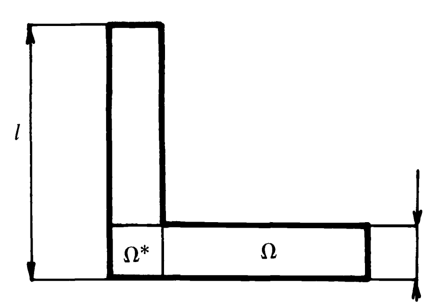

FIG. 2

It can also be compared with the one obtained in [1; 5.5]. This last estimate gives, for $l/e = 500$, a bound of the Poincaré constant greater than $10^3\, l$; with (P_{*1}), we get $2.86\, l$.

Furthermore, the estimate P_{*1} weakly depends on the ratio l/e for practical values (in structural design) of this ratio. For $l/e = 5$, we get $P_{*1} = 1.91\, l$, whereas for $l/e = 500$, we have $2.86\, l$ which is of the same magnitude.

4. Upper bounds (Π) and (P_{*2}). Beyond its own interest, the bound (Π) plays a major part in § 5 and in the paper [5] dealing with upper bounds in the Steklov constant. Furthermore, it leads to an upper bound (P_{*2}) of the Poincaré constant.

Let Ω' and Ω'' be two bounded domains in which M' and M'' are respectively the current points. We define the domain:

$$\Lambda(\Omega', \Omega'') = \{M \mid M = \lambda M' + (1-\lambda)M'', \lambda \in [0, 1], M' \in \Omega', M'' \in \Omega''\}.$$

4.1. Estimate. We have:

$$\int_{\Omega'\times\Omega''} [\alpha(M') - \alpha(M'')]^2 \, \underline{dM'} \, \underline{dM''}$$
$$\leqq \int_0^1 \int_0^1 d\lambda \, d\mu \int_{\Omega'\times\Omega''} |M'M''|^2 f(\lambda M' + (1-\lambda)M'') f(\mu M' + (1-\mu)M'') \, \underline{dM'} \, \underline{dM''}$$

in which $f(M) = |\mathrm{grad}_M \, \alpha|$.

From the Cauchy–Schwarz inequality, we get:

$$\int_{\Omega'\times\Omega''} [\alpha(M') - \alpha(M'')]^2 \, \underline{dM'} \, \underline{dM''}$$
$$\leqq \left[\int_0^1 d\lambda \left[\int_{\Omega'\times\Omega''} |M'M''|^2 f^2(\lambda M' + (1-\lambda)M'') \, \underline{dM'} \, \underline{dM''} \right]^{1/2} \right]^2.$$

A simple change of variable gives:

$$\int_{\Omega'\times\Omega''} |M'M''|^2 f^2(\lambda M' + (1-\lambda)M'') \, \underline{dM'} \, \underline{dM''}$$
$$\leqq \Phi^2(\Lambda(\Omega', \Omega'')) \int_{\Lambda(\Omega',\Omega'')} f^2(M) \, \underline{dM} \cdot \inf \left\{ \frac{A''}{\lambda^2}, \frac{A'}{(1-\lambda)^2} \right\}$$

where $\Phi(\Lambda(\Omega', \Omega''))$ is the diameter of $\Lambda(\Omega', \Omega'')$. We find:

$$\int_{\Omega'\times\Omega''} [\alpha(M') - \alpha(M'')]^2 \, \underline{dM'} \, \underline{dM''}$$
$$\leqq \left[\Phi(\Lambda(\Omega', \Omega'')) \|\mathrm{grad} \, \alpha\|_{\Lambda(\Omega',\Omega'')} \int_0^1 \inf \left\{ \frac{A''^{1/2}}{\lambda}, \frac{A'^{1/2}}{1-\lambda} \right\} d\lambda \right]^2.$$

Then we obtain:

$$\int_{\Omega'\times\Omega''} [\alpha(M') - \alpha(M'')]^2 \, \underline{dM'} \, \underline{dM''} \leqq \Pi^2 \|\mathrm{grad} \, \alpha\|^2_{\Lambda(\Omega',\Omega'')}$$

(Π)
$$\text{with } \Pi = \Phi(\Lambda(\Omega', \Omega''))$$
$$\cdot \left\{ A'^{1/2} \, \mathrm{Log} \left(1 + \frac{A''^{1/2}}{A'^{1/2}}\right) + A''^{1/2} \, \mathrm{Log} \left(1 + \frac{A'^{1/2}}{A''^{1/2}}\right) \right\}.$$

4.2. Applications. 1. Let Ω' be an arbitrary subdomain of the starshaping domain Ω^* of Ω. From the previous bound (Π), we derive the following estimate:

$$\|\alpha\|_{\Omega'} \leqq \Phi \left\{ \left(\frac{A'}{A}\right)^{1/2} \mathrm{Log} \left[1 + \left(\frac{A}{A'}\right)^{1/2} \right] + \mathrm{Log} \left[1 + \left(\frac{A'}{A}\right)^{1/2} \right] \right\} \|\mathrm{grad} \, \alpha\|_{\Omega}$$

which holds for a function α which satisfies $\mathscr{C}(\Omega)$.

It can be noticed that the constant involved in the last inequality can be bounded by:

$$(\Pi') \qquad \Pi' = \Phi\left(\frac{A'}{A}\right)^{1/2}\left\{1+\operatorname{Log}\left[1+\left(\frac{A}{A'}\right)^{1/2}\right]\right\}.$$

2. Let $\underset{\sim}{\alpha}(\Omega')$ (resp. $\underset{\sim}{\alpha}(\Omega'')$) be the mean value of the function α over Ω' (resp. Ω''). From the identity

$$\underset{\sim}{\alpha}(\Omega')-\underset{\sim}{\alpha}(\Omega'')=\frac{1}{A'A''}\int_{\Omega'\times\Omega''}[\alpha(M')-\alpha(M'')]\,\underline{dM'}\,\underline{dM''}$$

we get the estimate:

$$(\Pi'') \qquad \begin{aligned} &|\underset{\sim}{\alpha}(\Omega')-\underset{\sim}{\alpha}(\Omega'')| \leqq \Pi''\|\operatorname{grad}\alpha\|_{\Lambda(\Omega',\Omega'')} \\ &\text{with } \Pi''=\Phi(\Lambda(\Omega',\Omega'')) \\ &\cdot\left\{A''^{-1/2}\operatorname{Log}\left[1+\left(\frac{A''}{A'}\right)^{1/2}\right]+A'^{-1/2}\operatorname{Log}\left[1+\left(\frac{A'}{A''}\right)^{1/2}\right]\right\}. \end{aligned}$$

3. Let Ω be a strongly starshaped domain. For a function α *satisfying the condition* $\mathscr{C}(\Omega)$ *or* $\mathscr{C}(\Omega^*)$, we also deduce from the estimates (Π) and (2), the following upper bound:

$$\|\alpha\|_\Omega \leqq P_{*2}\|\operatorname{grad}\alpha\|_\Omega$$

$$(P_{*2}) \text{ with } P_{*2}=\left\{\Phi^2\left[\left(\frac{A-A^*}{A^*}\right)^{1/2}\operatorname{Log}\left\{1+\left(\frac{A^*}{A-A^*}\right)^{1/2}\right\}+\operatorname{Log}\left\{1+\left(\frac{A-A^*}{A^*}\right)^{1/2}\right\}\right]^2 + \max\left\{0,2-\frac{A}{A^*}\right\}\mu_{2*}^{-1}[\mathscr{C}(\Omega^*)]\right\}^{1/2}.$$

Remarks 1. The constant P_{*2} is optimal in the sense that it is the exact value for convex domains.

2. The quantity $\mu_{2*}^{-1/2}[\mathscr{C}(\Omega^*)]$ can of course be bounded by Φ^*/π.

3. For an L-shaped domain (see Fig. 2), we obtain (for $l/e>\frac{3}{2}$):

$$P_{*2}=l[1+(e/l)^2]^{1/2}\left[\lambda\operatorname{Log}\left(1+\frac{1}{\lambda}\right)+\operatorname{Log}(1+\lambda)\right]$$

with $\lambda=[2(l/e-1)]^{1/2}$.

P_{*2} asymptotically behaves as $\frac{1}{2}l\operatorname{Log}(2l/e)$ $(l/e\to\infty)$; this is not as good a result as the asymptotic behavior of P_{*1}: $l(\operatorname{Log}(2^{3/2}l/e))^{1/2}$.

5. Extension to more general domains.

5.1. Generalization of the upper bound (P_{*1}). Ω is a strongly sharshaped domain; $\underset{\sim}{\alpha}(\Omega^*)$ denotes the mean value of α over Ω^* and Ω', the interior of $\Omega-\Omega^*$.

From the inequality

$$\|\alpha\|_\Omega^2 \leqq \frac{2}{A^*}\int_{\Omega'\times\Omega^*}[\alpha(M')-\alpha(M^*)]^2\,\underline{dM'}\,\underline{dM^*}+\left(2-\frac{A}{A^*}\right)\|\mathring{\alpha}(\Omega^*)\|_{\Omega^*}^2+2A\underset{\sim}{\alpha}^2(\Omega^*)$$

we derive

$$(6) \qquad \|\alpha\|_\Omega^2 \leqq 2P_{*1}^2\|\operatorname{grad}\alpha\|_\Omega^2+2A\underset{\sim}{\alpha}^2(\Omega^*).$$

5.2. Upper bound of the Poincaré constant for more general domains. Ω is assumed to be the union of n strongly starshaped domains Ω_i such that the set $\Omega_i \cap \Omega_{i+1}$ is not empty for each i belonging to $\{1, 2, \cdots, n-1\}$. We denote:

$\Omega_{i,i+1} = \Omega_i \cap \Omega_{i+1}$

Ω_i^*: starshaping domain of Ω_i.

For the sake of simplicity, we shall develop the method only for $n = 2$ (see Fig. 3).

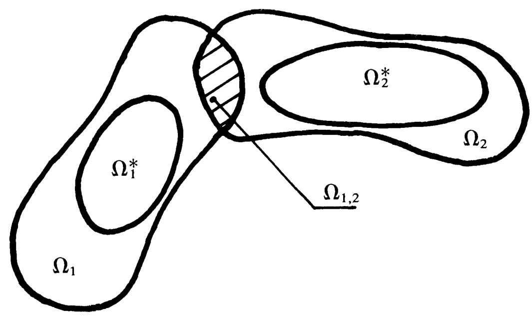

FIG. 3

From (6), we have:

$$\|\alpha\|_\Omega^2 \leqq 2 \sum_{i=1}^{2} P_{*1}^2(i) \|\text{grad}\, \alpha\|_{\Omega_i}^2 + 2 \sum_{i=1}^{2} A_i \underline{\alpha}^2(i)$$

where $P_{*1}(i)$ is the constant P_{*1} for the domain Ω_i and $\underline{\alpha}(i)$ is the mean value of α over Ω_i^*.

To estimate the quantities $\underline{\alpha}(i)$, we use the identities

$$\underline{\alpha}(i) - \underline{\alpha}(\Omega_{1,2}) = \frac{1}{A_i^* A_{1,2}} \int_{\Omega_i^* \times \Omega_{1,2}} [\alpha(M) - \alpha(M')]\, \underline{dM}\, \underline{dM'} \qquad (i = 1, 2).$$

Using the methods developed to get the bounds (P_{*1}) or (II), we obtain:

(7) $$|\underline{\alpha}(i) - \underline{\alpha}(\Omega_{1,2})|^2 \leqq C_i^2 \|\text{grad}\, \alpha\|_{\Omega_i}^2 \qquad (i = 1, 2).$$

The quantity $\sum_{i=1}^2 A_i \underline{\alpha}^2(i)$ can be written as follows:

$$\sum_{i=1}^{2} \{A_i[\underline{\alpha}(i) - \underline{\alpha}(\Omega_{1,2})]^2 + A_i \underline{\alpha}^2(\Omega_{1,2}) - 2A_i[\underline{\alpha}(i) - \underline{\alpha}(\Omega_{1,2})]\underline{\alpha}(\Omega_{1,2})\}.$$

The value of $\underline{\alpha}(\Omega_{1,2})$ can be chosen; taking the value which minimizes the previous quantity, we get

$$\underline{\alpha}(\Omega_{1,2}) = \frac{1}{A_1 + A_2} \sum_{i=1}^{2} A_i[\underline{\alpha}(i) - \underline{\alpha}(\Omega_{1,2})].$$

Consequently, we derive the estimate

$$\|\alpha\|_\Omega^2 \leqq 2 \sum_{i=1}^{2} P_{*1}^2(i) \|\text{grad}\, \alpha\|_{\Omega_i}^2 + 4 \frac{A_1 A_2}{A_1 + A_2} [C_1^2 \|\text{grad}\, \alpha\|_{\Omega_1}^2 + C_2^2 \|\text{grad}\, \alpha\|_{\Omega_2}^2].$$

Therefore, we have for a function α *satisfying* $\mathscr{C}(\Omega)$:

(8) $$\|\alpha\|_\Omega^2 \leqq 4 \max_{i=1,2} \left[P_{*1}^2(i) + \frac{2A_1 A_2}{A_1 + A_2} C_i^2 \right] \|\text{grad}\, \alpha\|_\Omega^2.$$

Example. Ω is the union of two convex domains Ω_1 and Ω_2. By using the estimate

(II), we obtain

$$C_i = \Phi_i\left\{A_{1,2}^{-1/2}\,\mathrm{Log}\left[1+\left(\frac{A_{1,2}}{A_i}\right)^{1/2}\right]+A_i^{-1/2}\,\mathrm{Log}\left[1+\left(\frac{A_i}{A_{1,2}}\right)^{1/2}\right]\right\}.$$

Thus, for α satisfying $\mathscr{C}(\Omega)$, we get the result

$$\|\alpha\|_\Omega \leqq P\|\mathrm{grad}\,\alpha\|_\Omega$$

with

$$P^2 = 4\max_{i=1,2}\left[\frac{1}{\pi^2}\Phi_i^2+2\frac{A_1A_2}{A_1+A_2}C_i^2\right].$$

If we apply this bound in the case of an L-shaped domain, the union of two identical rectangles with sides of lengths l and e (see Fig. 2), we obtain:

$$P = 2l(1+\varepsilon^4)^{1/2}\left\{\frac{1}{\pi^2}+\left[\left(1+\frac{1}{\varepsilon}\right)\mathrm{Log}\,(1+\varepsilon)+\mathrm{Log}\,(1/\varepsilon)\right]^2\right\}^{1/2}$$

in which $\varepsilon=(e/l)^{1/2}$. Asymptotically $(l/e\to\infty)$, the previous bound P behaves as $l\,\mathrm{Log}\,(l/e)$; this result is similar to the asymptotic behavior of P_{*2}.

We have summarized, in the next table, the numerical results on the estimates P_{*1}, P_{*2} and P for two values of the ratio l/e $(l/e>\frac{3}{2})$.

TABLE 1

l/e	P_{*1}/l	P_{*2}/l	P/l
5	1.91	2.24	8.37
500	2.86	4.47	34.3

For the special kind of domain considered here, we find that P_{*1} always gives the lowest bound of the Poincaré constant.

Appendix.

A.1. Function φ. The function φ is defined on $]0,1[$ by

(A.1) $$\varphi(\nu)=(F(\nu))^{1/2},$$

where

$$F(\nu)=\tfrac{1}{4}(2-\nu)^4\left(\mathrm{Log}\left(\frac{2-\nu}{\nu}\right)\right)^2+\left[\tfrac{1}{4}(2-\nu)^2+2\nu\left(\pi-2-\frac{\nu}{3}\right)\right]\tfrac{1}{2}(2-\nu)^2\,\mathrm{Log}\left(\frac{2-\nu}{\nu}\right)$$
$$+(1-\nu)[\tfrac{1}{8}\{3\nu^2-5(2-\nu)^2\}+((\pi-2)/6)\nu(-4\nu^2+26\nu-28)]+2(1-\nu)^2[-(\pi-\tfrac{8}{3})(\nu^2/2)$$
$$+2(\pi-4\lambda+\pi\,\mathrm{Log}\,2)+\nu(\pi^2-4-8\lambda+2\pi\,\mathrm{Log}\,2)+\nu^2(\pi/2)(\pi-\tfrac{8}{3})]$$

and λ is the numerical constant defined by:

$$\lambda=\int_0^1 \mathrm{Arctan}\,(y)\frac{dy}{y}=\sum_{n=0}^{\infty}(-1)^n\frac{1}{(2n+1)^2}\simeq 0.916.$$

A numerical computation shows that on $]0, 1[$

$$\varphi(\nu)-2\,\mathrm{Log}\,(1/\nu)\leqq 1.995(1-\nu)<2.$$

A.2. Estimates of $\sup_{M\in\Omega}\,[(1/A^*)\int_{\Omega^*}\mathrm{Log}\,(1/\varepsilon(H))\,\underline{dM^*}]$.
A.2.1. M belongs to Ω'.

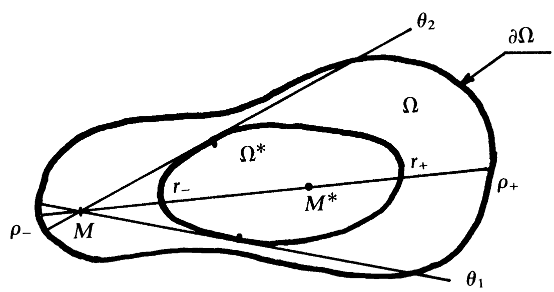

FIG. A.1

$$C(M)=\frac{1}{A^*}\int_{\Omega^*}\mathrm{Log}\,(1/\varepsilon(H))\,\underline{dM^*}$$

$$=\mathrm{Log}\,(1/\eta)+\frac{1}{A^*}\int_{\Omega^*}\mathrm{Log}\left(\frac{\eta}{\varepsilon(H)}\right)\underline{dM^*}$$

in which η is an arbitrary real positive constant to be determined later.

We have

$$C(M)\leqq\mathrm{Log}\,(1/\eta)+\frac{1}{A^*}\int_{\Omega^*}\mathrm{Log}^+\left(\frac{\eta}{\varepsilon(H)}\right)\underline{dM^*}$$

where

$$f^+(x)=\begin{cases}f(x) & \text{if } f(x)>0,\\ 0 & \text{if } f(x)\leqq 0.\end{cases}$$

We introduce the polar coordinates (r, θ) of center M (see Fig. A.1). Ω^* is included in the sector: $\theta_1\leqq\theta\leqq\theta_2$ and the boundary $\partial\Omega^*$ is represented by the equations

$$\begin{cases}r=r_+(\theta)\\ r=r_-(\theta)\end{cases},\qquad \theta\in\,]\theta_1, \theta_2[.$$

$r=\rho_+(\theta)$ is the equation of the part of the boundary $\partial\Omega$ inside the sector $\theta_1\leqq\theta\leqq\theta_2$; for the part of the boundary $\partial\Omega$ in the diametrally opposite sector of the first one, the equation is: $r=\rho_-(\theta)$. Furthermore, we have the relations

$$\rho_++\rho_-=\delta,$$

$$r_+-r_-=\delta^*=\varepsilon\delta.$$

With these notations, we get the following estimate:

$$C(M)\leqq\mathrm{Log}\,\frac{1}{\eta}+\frac{1}{A^*}\int_{\theta_1}^{\theta_2}\mathrm{Log}^+\left(\frac{\eta}{\varepsilon}\right)\frac{r_+^2-r_-^2}{2}\,d\theta$$

$$\leqq\mathrm{Log}\,\frac{1}{\eta}+\frac{1}{A^*}\int_{\theta_1}^{\theta_2}\varepsilon\,\mathrm{Log}^+\left(\frac{\eta}{\varepsilon}\right)\frac{\delta}{2}(r_++r_-)\,d\theta.$$

Since for $\eta \leqq e$:

$$\sup_{\varepsilon \in]0,1]} \left[\varepsilon \operatorname{Log}^+ \left(\frac{\eta}{\varepsilon}\right)\right] = \frac{\eta}{e},$$

we have the bound (for $\eta \leqq e$)

$$C(M) \leqq \operatorname{Log}\frac{1}{\eta} + \frac{\eta}{2e}\,\frac{1}{A^*}\int_{\theta_1}^{\theta_2} \delta(r_+ + r_-)\, d\theta.$$

We then establish:

A.2.1.1. First estimate.

$$\int_{\theta_1}^{\theta_2} \delta(r_+ + r_-)\, d\theta \leqq \Phi \int_{\theta_1}^{\theta_2} (r_+ + r_-)\, d\theta \leqq \Phi L^*.$$

We obtain for $\eta \leqq e$:

$$C(M) \leqq \operatorname{Log}\frac{1}{\eta} + \frac{1}{2e}\,\frac{\Phi L^*}{A^*}\eta.$$

Minimization over η gives $\eta_1 = 2eA^*/(\Phi L^*)$.

From the isoperimetric inequality: $A^* \leqq L^{*2}/4\pi$ and from $L^* \leqq \pi\Phi^*$ (Ω^* is convex), one verifies that $4A^*/(\Phi L^*) \leqq 1$, so $\eta_1 \leqq e/2$.

We conclude:

$$C(M) \leqq \operatorname{Log}\left(\frac{\Phi L^*}{2A^*}\right). \tag{A.2}$$

A.2.1.2. Second estimate. We have the estimate:

$$\int_{\theta_1}^{\theta_2} \delta(r_+ + r_-)\, d\theta \leqq \int_{\theta_1}^{\theta_2} 2\delta\rho_+\, d\theta.$$

From the inequality:

$$2\delta\rho_+ \leqq (\sqrt{2}+1)[\rho_+^2 + (\delta - \rho_+)^2] = (\sqrt{2}+1)(\rho_+^2 + \rho_-^2) \quad \forall \rho_+ \in [0, \delta]$$

we obtain:

$$\int_{\theta_1}^{\theta_2} \delta(r_+ + r_-)\, d\theta \leqq 2(\sqrt{2}+1)\int_{\theta_1}^{\theta_2} \frac{\rho_+^2 + \rho_-^2}{2}\, d\theta \leqq 2(\sqrt{2}+1)A.$$

We get for $\eta \leqq e$:

$$C(M) \leqq \operatorname{Log}\frac{1}{\eta} + \frac{\sqrt{2}+1}{e}\,\frac{A}{A^*}\eta.$$

The minimization of the above mentioned bound gives for η the value

$$\eta_2 = \frac{e}{\sqrt{2}+1}\,\frac{A^*}{A}.$$

Checking that $\eta_2 < e$, we conclude

$$C(M) \leqq \operatorname{Log}\left[(\sqrt{2}+1)\frac{A}{A^*}\right]. \tag{A.3}$$

A.2.2. *M* belongs to Ω^*.

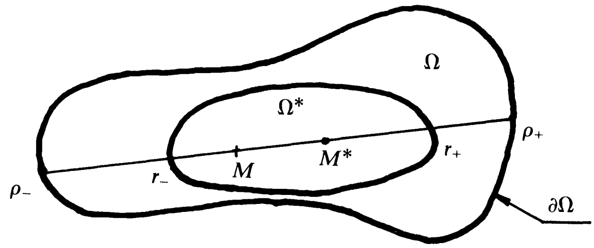

FIG. A.2

In a similar way as the one used in A.2.1, we get

$$C(M) \leqq \operatorname{Log}\frac{1}{\eta} + \frac{1}{A^*}\int_0^\pi \operatorname{Log}^+\left(\frac{\eta}{\varepsilon}\right)\frac{r_+^2 + r_-^2}{2}\,d\theta.$$

A.2.2.1. First estimate. From $r_+^2 + r_-^2 \leqq (r_+ + r_-)^2 = \varepsilon\delta(r_+ + r_-)$, we obtain:

$$C(M) \leqq \operatorname{Log}\frac{1}{\eta} + \frac{\Phi}{2A^*}\int_0^\pi \varepsilon \operatorname{Log}^+\left(\frac{\eta}{\varepsilon}\right)(r_+ + r_-)\,d\theta.$$

Using the inequality $\int_0^\pi (r_+ + r_-)\,d\theta \leqq L^*$, we obtain the same estimate as (A.2).

A.2.2.2. Second estimate. From

$$r_+^2 + r_-^2 \leqq (r_+ + r_-)^2 = \delta^{*2} = \varepsilon^2\delta^2 = \varepsilon^2(\rho_+ + \rho_-)^2 \leqq 2\varepsilon^2(\rho_+^2 + \rho_-^2)$$

we get:

$$C(M) \leqq \operatorname{Log}\frac{1}{\eta} + \frac{1}{A^*}\int_0^\pi \varepsilon^2 \operatorname{Log}^+\left(\frac{\eta}{e}\right)(\rho_+^2 + \rho_-^2)\,d\theta.$$

For $\eta \leqq \sqrt{e}$, we have:

$$\sup_{\varepsilon \in]0,1]}\left[\varepsilon^2 \operatorname{Log}^+\left(\frac{\eta}{\varepsilon}\right)\right] = \frac{1}{2e}\eta^2.$$

Therefore,

$$\begin{aligned} C(M) &\leqq \operatorname{Log}\frac{1}{\eta} + \frac{1}{A^*}\frac{\eta^2}{e}\int_0^\pi \tfrac{1}{2}(\rho_+^2 + \rho_-^2)\,d\theta \\ &\leqq \operatorname{Log}\frac{1}{\eta} + \frac{A}{A^*}\frac{\eta^2}{e}. \end{aligned}$$

The minimum of the last quantity is obtained for $\eta_3 = [(e/2)(A^*/A)]^{1/2}$, (with $\eta_3 < \sqrt{e}$). We find:

$$C(M) \leqq \tfrac{1}{2}\operatorname{Log}\left(\frac{2A}{A^*}\right). \tag{A.4}$$

A.2.3. Conclusions. From results (A.2), (A.3), (A.4), we get for

$$E = \sup_{M \in \Omega}\left[\frac{1}{A^*}\int_{\Omega^*} \operatorname{Log}(1/\varepsilon(H))\,\underline{dM^*}\right],$$

the following bounds:

$$E \leqq \operatorname{Log}\left(\frac{\Phi L^*}{2A^*}\right),$$

(A.5)

$$E \leqq \operatorname{Log}\left[(1+\sqrt{2})\frac{A}{A^*}\right].$$

REFERENCES

[1] J. H. BRAMBLE AND L. E. PAYNE, *Bounds in the Neumann problem for second order uniformly elliptic operators*, Pacific J. Math., 12 (1962), pp. 823–833.

[2] ———, *Bounds for solutions of second order elliptic partial differential equations*, Contributions to Differential Equations, 1 (1963), pp. 95–127.

[3] J. R. KUTTLER AND V. G. SIGILLITO, *Inequalities for membrane and Stekloff eigenvalues*, J. Math. Anal. Appl., 23 (1968), pp. 148–160.

[4] J. LADEVÈZE AND P. LADEVÈZE, *Majorations de la constante de Poincaré relative au problème de la membrane—Domaines étoilés*, Z. Angew. Math. Phys., 29 (1978), pp. 670–683.

[5] ———, *Majorations de constantes de Steklov*, Ibid., 29 (1978), pp. 684–692.

[6] L. E. PAYNE, *Isoperimetric inequalities and their applications*, SIAM Rev., 9 (1967), pp. 453–488.

[7] L. E. PAYNE AND H. F. WEINBERGER, *An optimal Poincaré inequality for convex domains*, Arch. Rational Mech. Analysis, 5 (1960), pp. 286–292.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0017 $01.00/0

ON THE POINCARÉ CRITERION FOR ASYMPTOTIC STABILITY*

R. C. CHURCHILL† AND J. F. SELGRADE‡

Abstract. This work discusses an n dimensional generalization of the Poincaré criterion for asymptotic stability of a periodic solution to an ordinary differential equation. We define a function along a trajectory which measures the projection, normal to the flow, of a solution to the linearized equations. When this function satisfies certain monotone conditions, we show that a compact invariant set with recurrence is an orbitally asymptotically stable periodic solution. Our result is then used to analyze a 3-dimensional system of equations proposed by J. S. Griffith to model a cellular process for control of gene expression by positive feedback.

We discuss an n-dimensional generalization of the following well-known result.

POINCARÉ CRITERION. *Let F: $R^2 \to R^2$ be a C^1-vector field, and let γ: $R \to R^2$ be a periodic integral curve of F. If $\int_\gamma \operatorname{div}(F) < 0$, then γ is asymptotically orbitally stable, with asymptotic phase* [6].

The usual proof amounts to the observation that the integral in question has the same sign as the nontrivial characteristic exponent of the associated linearized equations. However, we will see that by interpreting the integral in a slightly different way, the result admits a useful generalization.

1. Statement of the theorem. Let F: $R^n \to R^n$ be a C^1-vector field, let ϕ^t: $R^n \to R^n$ be the associated flow, let $T(R^n) \approx R^n \times R^n$ be the tangent bundle of R^n, and let $D\phi^t$: $T(R^n) \to T(R^n)$ be the tangent flow, i.e., the flow induced by the linearized equations. For $p \in R^n$ write $\phi^t(p)$ as $p \cdot t$, and for $(p, w) \in T(R^n)$ write $D\phi^t(p)w$ as $w : t$. Also, let $\langle , \rangle$ denote the usual inner product on R^n, let $T^1(R^n)$ denote the unit sphere bundle of R^n, and let $E^\perp$ denote the orthogonal complement of any subbundle $E \subset T(R^n)$.

THEOREM 1. *In the notation above, assume $X \subset R^n$ is a compact connected invariant set, and that $F \mid X$ does not vanish. Next, let $E \subset T_X(R^n)$ be the line bundle generated by F, and for $p \in X$, $(p, w) \in E^\perp \cap T^1_p(R^n)$, $t \in R$ define*

$$V(t) = V(p, w; t) = \langle F(p \cdot t), F(p \cdot t)\rangle\langle w : t, w : t\rangle - \langle F(p \cdot t), w : t\rangle^2.$$

Now assume:

(a) *Each $p \in X$ is negatively Poisson stable*; *and*

(b) *There is a real number $s < 0$ and a continuous function $(p, w) \to \varepsilon(p, w) > 0$, $(p, w) \in T^1_p(R^n) \cap E^\perp$, such that $V(p, w; t) > V(p, w; 0) + \varepsilon(p, w)$ for all $t \leqq s$.*

Then X is a periodic orbit, and the nontrivial characteristic exponents of X have negative real parts. In particular, X is asymptotically orbitally stable with asymptotic phase.

Remarks. (i) If X is already known to be a periodic orbit, then (a) obviously holds.

(ii) Modifications of (b) are available; one will be discussed in § 3.

(iii) If $\dot{V} < 0$, then (b) obviously holds; this is the first thing to check in applications.

(iv) $V(p, w; t)$ is simply

$$|(w : t)^\perp|^2 |F(p \cdot t)|^2, \tag{1}$$

* Received by the editors October 17, 1978.

† Department of Mathematics, Hunter College, CUNY, New York, New York 10021.

‡ Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650.

where the "$\perp$" denotes perpendicular projection onto $E^{\perp}$. The theorem can in fact be formulated on any Riemannian manifold. Notice from (1) that

$$V > 0. \tag{2}$$

(v) For $a = (a_1, \cdots, a_n)^T$ write $a_{ij} = (a_i, a_j)^T$, and set $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then $V(t)$ may be written as

$$\sum_{\substack{i=1 \\ i<j}}^{n-1} \langle (w: t)_{ij}, -J(F(p \cdot t))_{ij} \rangle^2. \tag{3}$$

(vi) For $n = 3$, $V(t)$ can also be expressed as

$$|F(p \cdot t) \times (w: t)|^2, \tag{4}$$

where $\times$ denotes the usual cross product.

(vii) For $n = 2$ one computes that $\dot{V} = 2 \operatorname{div}(F) V$; hence in this case

$$V(t) = V(0)\, e^{2 \int_0^t \operatorname{div}(F)} \tag{5}$$

Suppose X is already known to be a periodic orbit, of primitive period $\tau > 0$, and that $\int_0^\tau \operatorname{div}(F) = \rho < 0$. Let $M = \inf \{e^{-2 \int_0^s \operatorname{div}(F)}: 0 \leqq s \leqq \tau\} > 0$, and for $t < 0$ let $k = k(t)$ be the greatest integer in $-t/\tau$. Then from (5) we have

$$\begin{aligned} V(t) &= V(0)\, e^{-2 \int_t^0 \operatorname{div}(F)} \\ &= V(0)\, e^{-2 \int_t^{-k\tau} \operatorname{div}(F) - 2 \int_{-k\tau}^0 \operatorname{div}(F)} \\ &\geqq V(0) M e^{-2k\rho}. \end{aligned}$$

Since $k \to \infty$ as $t \to -\infty$, we conclude that (b) must hold, while (a) follows from (i). Theorem 1 thus includes the Poincaré criterion.

(viii) Considered as an existence theorem for periodic orbits, Theorem 1 is similar in spirit to Theorem VII.4, p. 116, of [10], a result concerning continuous flows on uniform spaces.

2. Proof of Theorem 1. The proof is based on the following result found in [3]:

THEOREM 2. *Let X be as in Theorem 1, and suppose for all $p \in X$, $(p, w) \in T(R^n) - E$, the set $\{|(w: t)^{\perp}|: t \leqq 0\}$ is unbounded. Then X is a periodic orbit with nontrivial characteristic exponents having negative real parts.*

Therefore we must show that hypotheses (a) and (b) of Theorem 1 imply the unboundedness of $\{|(w: t)^{\perp}|: t \leqq 0\}$ for any $(p, w) \in T_X(R^n) - E$. For this, suppose we set

$$E_1^{\perp} = \{(p, w) \in T_X(R^n): |w| = 1,\ w \in E^{\perp}\}, \tag{6}$$

and we fix some $(p, w) \in E_1^{\perp}$. It suffices to consider only such w because of the invariance of E and linearity.

By (b) of Theorem 1 we can choose an open neighborhood $N \subset E_1^{\perp}$ of (p, w) and a constant $\varepsilon > 0$ such that

$$V(q, y; t) > V(q, y; 0) + \varepsilon, \qquad (q, y) \in N, \quad t \leqq s. \tag{7}$$

Moreover, by shrinking N if necessary we can also assume that

$$|F(q)|^2 > |F(r)|^2 - (\varepsilon/2), \qquad q, r \in \pi(N), \tag{8}$$

where $\pi\colon T(R^n) \to R^n$ is the usual projection.

If $q \in \pi(N)$ is arbitrary, and if there is a $t < s$ such that $q \cdot t \in \pi(N)$, then (8) implies

$$|F(q)|^2 > |F(q \cdot t)|^2 - (\varepsilon/2). \tag{9}$$

But then choose $(q, y) \in E_1^\perp$, and from (1) and (7) we obtain

$$|(y:t)^\perp|^2 |F(q \cdot t)|^2 > |y^\perp|^2 |F(q)|^2 + \varepsilon;$$

since $|y^\perp| = |y| = 1$, (8) then gives

$$|(y:t)^\perp|^2 |F(q \cdot t)|^2 > |F(q \cdot t)|^2 + (\varepsilon/2). \tag{10}$$

Letting $K = \sup\{|F(x)| : x \in \pi(N)\}$, which we may assume is finite by further restricting N if necessary, we can thus conclude

$$|(y:t)^\perp|^2 > 1 + (\varepsilon/2K) = \beta > 1, \tag{11}$$

where β is independent of $(q, y) \in N$.

Now use (a) of Theorem 1 to choose a sequence $t_j \searrow -\infty$, with $t_1 < s$, $t_{j+1} < t_j + s$, $j \geqq 1$, such that $p \cdot t_j \in \pi(N)$. Equation (11) then gives

$$\begin{aligned} |(w:t_{j+1})^\perp|^2 &= |((w:t_j):(t_{j+1}-t_j))^\perp|^2 \\ &= |((w:t_j)^\perp:(t_{j+1}-t_j))^\perp|^2 \\ &= |(((w:t_j)^\perp/|(w:t_j)^\perp|):(t_{j+1}-t_j))^\perp|^2 |(w:t_j)^\perp|^2 \\ &> \beta|(w:t_j)^\perp|^2 > \cdots > \beta^{j+1}|w| = \beta^{j+1}. \end{aligned}$$

Since $\beta > 1$, this gives the unboundedness of $\{|(w:t)^\perp| : t \leqq 0\}$, and completes the proof. Q.E.D.

Notice that the preceding proof made no use of the compactness of X; that hypothesis was needed for Theorem 2.

3. Modifications. Hypothesis (b) of Theorem 1 requires a fairly detailed knowledge of V for *all* large negative t; too stringent a hypothesis in many applications. Here we offer an alternate method for verifying the assumptions of Theorem 2, which we state using the notation of the previous section.

THEOREM 3. *For each $(p, w) \in E_1^\perp$, assume there is a $\delta > 0$ and a sequence $t_j \searrow -\infty$ such that*:

(a) $|F(p \cdot t_j)|^2 > |F(p \cdot t_{j+1})|^2 - (\delta/2)$; *and*

(b) *for any $(p \cdot t_j, y) \in E_1^\perp$,*

$$V(p \cdot t_j, y; t_{j+1} - t_j) > V(p \cdot t_j, y; 0) + \delta.$$

Then $\{|(w:t)^\perp| : t \leqq 0\}$ is unbounded.

Proof. For $(p \cdot t_j, y) \in E_1^\perp$, (b) and (1) imply that

$$|(y:(t_{j+1}-t_j))^\perp|^2 |F(p \cdot t_{j+1})|^2 > |F(p \cdot t_j)|^2 + \delta.$$

Hence from (a) we get

$$|(y:(t_{j+1}-t_j))^\perp|^2 |F(p \cdot t_{j+1})|^2 > |F(p \cdot t_{j+1})|^2 + (\delta/2).$$

This is the analogue of (10) of the previous proof, and by choosing $L =$

$\sup\{|F(x)|: x \in X\}$, we can then get the analogue of (11), i.e.

$$|(y:(t_{j+1}-t_j))^{\perp}|^2 > 1 + (\delta/2L) = \gamma > 1. \tag{12}$$

Notice that linearity then implies

$$|(y:(t_{j+1}-t_j))^{\perp}|^2 \geqq \gamma|y|^2 \geqq \gamma|y^{\perp}|^2$$

for *any* $(p, y) \in E^{\perp}$, and so

$$\begin{aligned}|(w:t_{j+1})^{\perp}|^2 &= |((w:t_j)^{\perp}:(t_{j+1}-t_j))^{\perp}|^2\\ &> \gamma|(w:t_j)^{\perp}|^2 > \cdots > \gamma^{j+1},\end{aligned}$$

from which the theorem follows. Q.E.D.

From the proof of Theorem 1 (but now using compactness) one can see that the hypothesis of Theorem 3 will hold if p is negatively Poisson stable, and if $V(q, y; t)$ increases as $t \to -\infty$ for each $(q, y) \in E_1^{\perp}$, q in the negative half-orbit of p.

4. An application. Let $F = (F_1, F_2, F_3)$ be a C^1-vector field on R^3, and for $x = (x_1, x_2, x_3) \in R^3$ consider the system

$$\dot{x} = F(x). \tag{13}$$

Suppose $p \cdot t$ is a solution to (13) and $F(p) \neq 0$. Along the orbit of $(p, w) \in R^3 \times R^3$, (3) implies that $V(p, w; t) = V_1(p, w; t) + V_2(p, w; t) + V_3(p, w; t)$, where

$$\begin{aligned}V_1 &\equiv \langle (w:t)_{12}, -J(F(p\cdot t))_{12}\rangle^2\\ V_2 &\equiv \langle (w:t)_{13}, -J(F(p\cdot t))_{13}\rangle^2\\ V_3 &\equiv \langle (w:t)_{23}, -J(F(p\cdot t))_{23}\rangle^2.\end{aligned} \tag{14}$$

A straightforward computation gives the following time-dependent system for V_1, V_2, V_3:

$$\begin{aligned}\dot{V}_1 &= 2V_1\left(\frac{\partial F_1}{\partial x_1}+\frac{\partial F_2}{\partial x_2}\right) + 2V_1^{1/2}V_2^{1/2}\frac{\partial F_2}{\partial x_3} - 2V_1^{1/2}V_3^{1/2}\frac{\partial F_1}{\partial x_3}\\ \dot{V}_2 &= 2V_2\left(\frac{\partial F_1}{\partial x_1}+\frac{\partial F_3}{\partial x_3}\right) + 2V_2^{1/2}V_3^{1/2}\frac{\partial F_1}{\partial x_2} + 2V_1^{1/2}V_2^{1/2}\frac{\partial F_3}{\partial x_2}\\ \dot{V}_3 &= 2V_3\left(\frac{\partial F_2}{\partial x_2}+\frac{\partial F_3}{\partial x_3}\right) + 2V_2^{1/2}V_3^{1/2}\frac{\partial F_2}{\partial x_1} - 2V_1^{1/2}V_3^{1/2}\frac{\partial F_3}{\partial x_1}\end{aligned} \tag{15}$$

where $\partial F_i/\partial x_j$ is evaluated at $p \cdot t$ and where $V_j^{1/2}$ is the (possibly negative) inner product in (14), $j = 1, 2, 3$. Compare this system to the equation for $n = 2$ discussed in (vii): $\dot{V} = 2\,\mathrm{div}\,(F)V$.

Consider the system

$$\begin{aligned}\dot{x}_1 &= \frac{x_3}{1+x_3} - \alpha x_1\\ \dot{x}_2 &= x_1 - \beta x_2\\ \dot{x}_3 &= x_2 - \gamma x_3,\end{aligned} \tag{16}$$

$\alpha > 0$, $\beta > 0$, $\gamma > 0$, which was proposed by J. S. Griffith [5] to model a cellular process for control of gene expression by positive feedback. In [5] it is shown that the first octant $\mathscr{H} = \{(x_1, x_2, x_3): x_j \geqq 0, j = 1, 2, 3\}$ is invariant under the resulting positive-time flow, and that all positive half-orbits within this octant are bounded. Notice that the origin is a critical point, which we denote by θ. If $\alpha\beta\gamma < 1$, then it is easy to see that the equations admit precisely one other critical point in $\mathcal{H}$, which we denote by P.

We use Theorem 1 to prove the following theorem. Similar results can be found in [8], [9], and [11].

THEOREM 4. *Suppose in* (16) *we require*

$$\alpha \geqq \tfrac{1}{2}, \quad \beta \geqq 1, \quad \gamma \geqq \tfrac{1}{2}, \quad \alpha\beta\gamma < 1. \tag{17}$$

Then, with the exception of θ, every orbit within or entering $\mathcal{H}$ must be positively asymptotic to P.

Proof. Let X be a minimal set in $\mathcal{H}$ which is not a critical point. If $p \in X$ and $(p, w) \in E^{\perp} \cap T_p^1(R^3)$, then (15) gives

$$\begin{aligned} \dot{V}_1 &= -2(\alpha+\beta)V_1 - 2V_1^{1/2}V_3^{1/2}(1+x_3)^{-2} \\ \dot{V}_2 &= -2(\alpha+\gamma)V_2 + 2V_1^{1/2}V_2^{1/2} \\ \dot{V}_3 &= -2(\beta+\gamma)V_3 + 2V_2^{1/2}V_3^{1/2}. \end{aligned} \tag{18}$$

Now write $g = (1+x_3)^{-2}$, and note that $0 < g \leqq 1$ in $\mathcal{H}$; from (18) we then have

$$\begin{aligned} \dot{V} &= \dot{V}_1 + \dot{V}_2 + \dot{V}_3 \\ &= -2(\alpha+\beta)V_1 - 2V_1^{1/2}V_3^{1/2}g - 2(\alpha+\gamma)V_2 + 2V_1^{1/2}V_2^{1/2} - 2(\beta+\gamma)V_3 + 2V_2^{1/2}V_3^{1/2}. \end{aligned}$$

But (17) implies $-2(\alpha+\beta)V_1 \leqq -3V_1$, $-2(\alpha+\gamma)V_2 \leqq -2V_2$, and $-2(\beta+\gamma)V_3 \leqq -3V_3$; hence

$$\begin{aligned} \dot{V} &\leqq -3V_1 - 2V_2 - 3V_3 + 2V_1^{1/2}V_2^{1/2} + 2V_2^{1/2}V_3^{1/2} - 2gV_1^{1/2}V_3^{1/2} \\ &= -2(V_1 + V_3 + gV_1^{1/2}V_3^{1/2}) - (V_1 + V_2 - 2V_1^{1/2}V_2^{1/2}) - (V_2 + V_3 - 2V_2^{1/2}V_3^{1/2}) \\ &= -2(V_1 + V_3 + gV_1^{1/2}V_3^{1/2}) - (V_1^{1/2} - V_2^{1/2})^2 - (V_2^{1/2} - V_3^{1/2})^2. \end{aligned}$$

Since $0 < g \leqq 1$, it follows that $V_1 + V_3 + gV_1^{1/2}V_3^{1/2} > 0$ if $V_1^{1/2}$ and $V_3^{1/2}$ are not both zero. From (2) we see that $V_1^{1/2} = V_2^{1/2} = V_3^{1/2} = 0$ is impossible, hence $\dot{V} < 0$. By our minimality assumption, (a) of Theorem 1 holds; and (b) holds by virtue of (iii). X is therefore an asymptotically stable periodic orbit.

Since positive half-orbits of (16) in the first octant are bounded, the ω-limit sets of such orbits are nonempty and compact. By the preceding argument, the minimal subsets of such an invariant set are either critical points or asymptotically stable periodic orbits. However, in the second case it is a simple matter to see that the periodic orbit must coincide with the original ω-limit set. We thus conclude:

Every orbit intersecting the first octant is either

(a) *positively asymptotic to an attracting periodic orbit, or*

(b) *contains a critical point in its ω-limit set.*

Linearizing (16) at a critical point gives the matrix

$$A(x_3) = \begin{pmatrix} -\alpha & 0 & g \\ 1 & -\beta & 0 \\ 0 & 1 & -\gamma \end{pmatrix}.$$

Using the Routh–Hurwitz criterion [2], it is then a simple matter to see that P is asymptotically stable, and that θ admits two eigenvalues with negative real parts, and one positive eigenvalue. Moreover, it is easy to see that the eigenspace corresponding to this positive eigenvalue will pass through the interior of $\mathcal{H}$. In contrast, we will show that the 2-dimensional local stable manifold of θ meets $\mathcal{H}$ only at θ.

If $u=(u_1, u_2, u_3)$ and $v=(v_1, v_2, v_3) \in R^3$, write $u \geqq v$ if $u_j \geqq v_j$, $j=1,2,3$, and $u>v$ if $u_j>v_j$, $j=1, 2, 3$. Because all off-diagonal entries of $A(0)$ are nonnegative, and because none of the coordinate planes or lines are invariant under this matrix, it follows [7] that $e^{A(0)t}v>0$ for all $v \geqq 0$, $v \neq 0$, and $t>0$. The tangent flow at θ thus carries $\mathcal{H}$ into int $\mathcal{H}$. Moreover, $e^{A(0)}$ cannot have two linearly independent positive eigenvectors [4, p. 63]. Using this fact and the invariance of $\mathcal{H}$, it is clear that the invariant plane of linearized solutions asymptotic to the origin intersects $\mathcal{H}$ only at θ. Since the stable manifold of θ is tangent to this plane at θ and since $\mathcal{H}$ is positively invariant, we conclude that there is no orbit in $\mathcal{H}$ having θ in its ω-limit set.

Let $p \in \mathcal{H}$, $p \neq \theta$, and note that the ω-limit set $\omega(p)$ of p is not empty. From (a) and (b), together with the previous paragraph we thus conclude that $\omega(p)=P$ or $\omega(p)$ is an asymptotically stable periodic orbit. But in the second case P and $\omega(p)$ can be connected by an arc; and, since the basins of attraction for P and $\omega(p)$ are open and disjoint, there must be a point q on this arc which is not in either basin of attraction. In fact, by choosing q on the boundary of one of these basins, it follows that $\omega(q)$ is not an attractor. But then $\omega(q)$ contains θ, which is impossible. We conclude that $\omega(p)=P$, and our argument is complete. Q.E.D.

REFERENCES

[1] R. C. CHURCHILL, J. FRANKE AND J. SELGRADE, *A geometric criterion for hyperbolicity of flows*, Proc. Amer. Math. Soc., 62 (1977), pp. 137–143.

[2] W. A. COPPEL, *Stability and Asymptotic Behavior of Differential Equations*, D. C. Heath and Co., Boston, 1965.

[3] J. FRANKE AND J. SELGRADE, *A computer method for verification of asymptotically stable periodic orbits*, this Journal, 10 (1979), pp. 614–628.

[4] F. R. GANTMACHER, *The Theory of Matrices II*, Chelsea, New York, 1959.

[5] J. S. GRIFFITH, *Mathematics of cellular control processes, II. Positive feedback to one gene*, J. Theoret. Biol., 20 (1968), pp. 209–216.

[6] J. K. HALE, *Ordinary Differential Equations*, Wiley-Interscience, New York, 1969.

[7] R. H. MARTIN, *Asymptotic stability and critical points for nonlinear quasimonotone parabolic systems*, J. Differential Equations, to appear.

[8] H. G. OTHMER, *The qualitative dynamics of a class of biochemical control circuits*, J. Math. Biol., 3 (1976), pp. 53–78.

[9] J. F. SELGRADE, *Mathematical analysis of a cellular control process with positive feedback*, SIAM J. Appl. Math., 36 (1979), pp. 219–229.

[10] G. SELL, *Topological Dynamics and Ordinary Differential Equations*, Van Nostrand Reinhold, New York, 1971.

[11] J. J. TYSON AND H. G. OTHMER, *The dynamics of feedback control circuits in biochemical pathways*, Progr. Theoret. Biol., to appear.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0018 $01.00/0

A STUDY OF PC^1 HOMEOMORPHISMS ON SUBDIVIDED POLYHEDRONS*

M. KOJIMA† AND R. SAIGAL‡

Abstract. In this paper we consider the problem of establishing conditions when a given piecewise continuously differentiable mapping is a homeomorphism of a closed convex polyhedral set. These conditions are a generalization of the ones used by Gale–Nikaido and are similar in spirit to those of Mas-Colell. For the special case when the mapping is piecewise linear, we give an apparently new sufficiency condition for the mapping to be a homeomorphism of R^n. The results are further extended to include the case when the Jacobians may be singular.

1. Introduction. Let S be a closed convex polyhedral subset of R^n, the n-dimensional Euclidean space, and let Σ be a class of closed convex polyhedral subsets of S which partition S. A function F from S into S is called piecewise continuously differentiable (PC^1 for short) on the subdivided polyhedron (S, Σ) if it is continuous, and for each piece σ in Σ, $F_\sigma \equiv F|\sigma$ (the restriction of F to σ) is a continuously differentiable mapping. The problem we consider in this paper is that of establishing conditions under which F maps S homeomorphically onto $F(S)$; i.e., F is one to one and onto.

One of the early works establishing such a result is that of Gale and Nikaido [5], which is often used to establish the uniqueness of solutions. Their result states that if $S = \{x: a_i \leqq x_i \leqq b_i\}$ and F is a continuously differentiable mapping from S into R^n, then if, for all x, the Jacobian matrix $DF(x)$ of F has all principal minors positive, F maps S homeomorphically onto $F(S)$. H. Scarf [21] has conjectured that since in the nonlinear complementarity problem, such a strong requirement on the Jacobian can be considerably weakened (see, for example, Corollary 2.6, Saigal and Simon [19]), such a weakening should be possible for the hypothesis of the Gale–Nikaido theorem. This was verified by Mas-Colell [12]. He further generalized the result to the case when S is a compact convex polyhedron, and showed that such a result would be false for nonconvex objects. The proof of [12] involved the use of degree theoretic arguments (a possibility of which had been foreseen by H. Scarf). Independently, Garcia and Zangwill [6] again verified this conjecture, using the norm-coerciveness theorem [15, 5.3.8]. Their result is on a rectangle S, but a slight weakening of the requirement on the derivatives was achieved. In this paper, we further generalize this result. In one generalization, using degree theoretic arguments similar to those of [12], we establish the result for PC^1 mappings. In the other, we find conditions under which this result holds, when the derivatives may be singular. Under a similar hypothesis involving negative determinants, we show that our approach fails for PC^1 cases.

In case the restriction to each piece in Σ of the mapping F is affine, we call it a piecewise linear mapping and, for brevity, PL. Considerable attention has been paid to the study of such mappings (see, for example, Eaves and Scarf [3], Fujisawa and Kuh [4], and Ohtsuki, Fujisawa and Kumagai [14]), as well as to the problem of generating PL approximations (see, for example, Kojima [8], [9], Saigal [18]). In addition, several authors have contributed to the conditions under which such mappings are onto (see, for example, Chien and Kuh [1], Rheinboldt and Vandergraft [15]). Also, a set of

* Received by the editors May 9, 1978, and in revised form December 12, 1978. This work was supported in part by Grant MCS77-03472 from the National Science Foundation.

† Department of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan. The work of this author was done while visiting Northwestern University.

‡ Northwestern University, Department of Industrial Engineering and Management Sciences, Evanston, Illinois 60201.

conditions under which the mapping is a homeomorphism are developed in [4] and [14]. In this paper, we present a sufficiency condition which is weaker than that of [4] (see Kojima and Saigal [11]). By an example, we show that it is not necessary, and that any condition put only on all subsets of the Jacobians of the pieces cannot be necessary and sufficient.

After presenting the terminology and notation in § 2, in § 3 we calculate the local degree of certain PC^1 mappings. In § 4 we prove the extension of the Gale–Nikaido theorem for PC^1 mappings, and show by a counter example that the appropriate negative condition on certain minors of the Jacobian is not sufficient to guarantee a homeomorphism. In § 5 we prove a sufficiency condition under which a PL mapping is a homeomorphism and in § 6 we present two PL mappings which are homeomorphisms. One of these mappings is generated by the Samelson–Thrall–Wesler [20] partition theorem, and the other by the recent result of Kojima and Saigal [10] relating to the linear complementarity problem with negative principal minors. The later example is presented in the hope that it will help to generate conditions ensuring homeomorphisms with the hypothesis that certain minors of the Jacobian are negative. Finally, in § 7, we show how our results can be extended to include the case when the appropriate minors of the Jacobian may be zero.

2. Notation and definitions. In this section we present the notation and definitions that will be needed in the subsequent sections. In particular, we establish some properties of subdivided polyhedrons and functions on them.

By a bounded polyhedron, we represent the convex combination of a finite collection of points. Also, given a set τ, we represent by H_τ the subspace spanned by τ, i.e., $H_\tau = \{y: y = \sum_{i=1}^r \lambda_i x_i, \sum_{i=1}^r \lambda_i = 1, x_i \in \tau\} - \tau$, and thus the origin is contained in H_τ. A convex set of the form $\{x + \lambda y: \lambda \geqq 0\}$ is called a half-line. We will call the convex hull of a finite collection of points and half-lines in R^n a convex polyhedral set. The dimension of a set is the dimension of the subspace spanned by the set.

The interior $\mathring{\sigma}$, and the boundary $\partial\sigma$ of a set σ, are the relative interior and boundary of the set in the affine subspace $H_\sigma + \sigma$. Also, a subset τ of σ is called a face of σ if for every x and y in σ $0 < \lambda < 1$ and $(1-\lambda)x + \lambda y$ in τ imply x and y are in τ. It can be readily confirmed that the faces of convex polyhedral sets are also convex polyhedral sets. For an n-dimensional set σ, a $(n-1)$-dimensional face is called a facet.

Now, given a set S and a finite class Σ of nonempty subsets of S, we say (S, Σ) is a subdivided polyhedron of dimension n if:

a) elements of Σ are n-dimensional convex and polyhedral, and are called pieces;
b) any two members of Σ are either disjoint, or meet on a common face;
c) the union of the pieces in Σ is S.

We say that (S, Σ) is a subdivided compact polyhedron if S is compact, and a subdivided convex polyhedron if S is convex. Also, the dimension of (S, Σ) is the dimension of the set S.

Let (S, Σ) be a subdivided compact convex polyhedron of dimension n, with S in R^n. Then, there exists an extension Σ' of Σ such that (R^n, Σ') is a subdivided polyhedron. This can be observed by defining the projection mapping:

$$\|x - P(x)\| = \min_{y \in S} \|x - y\| \tag{2.1}$$

and noting that Σ' is generated by adding the closure of $P^{-1}(\mathring{\tau})$ for τ, a face of some σ in Σ to those already in Σ (see Fig. 2.1).

Now, let $F: S \to R^n$ be a continuous function on a subdivided polyhedron (S, Σ). We say F is PC^1, i.e., piecewise continuously differentiable, on (S, Σ), if for each piece σ

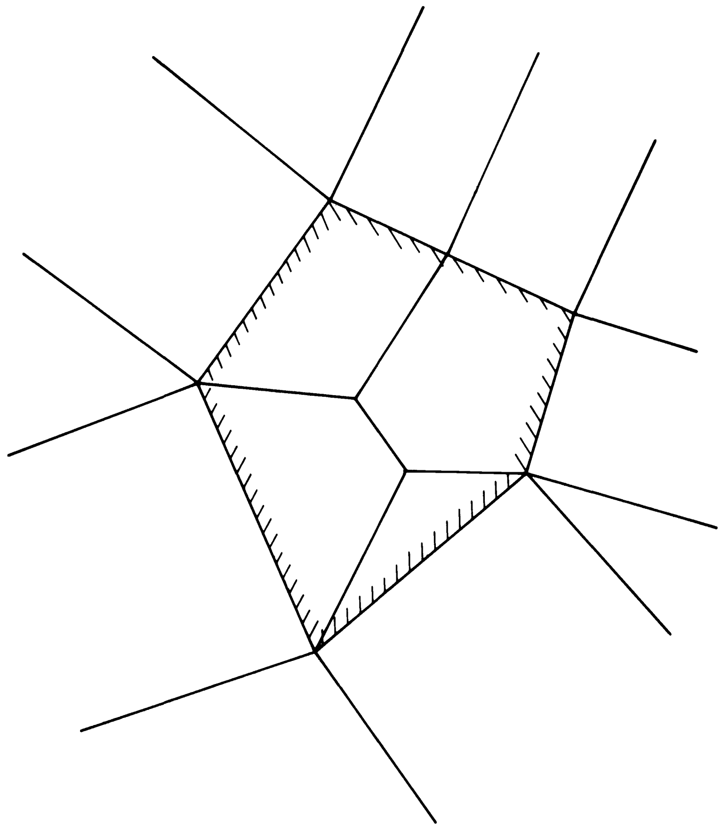

FIG. 2.1

in Σ there exists an open set B_σ containing σ such that $F_\sigma = F|\sigma$ can be extended to B_σ continuously differentiably. In particular, it is called piecewise linear if F_σ is affine, i.e., $F_\sigma(x) = A_\sigma x - a_\sigma$ for some $n \times n$ matrix A_σ and n vector a_σ.

Now, given a subdivided compact convex polyhedron (S, Σ) and a mapping $F: S \to R^n$ which is PC^1 on (S, Σ), there exists a PC^1 extension to (R^n, Σ') when the subdivision Σ is extended by the projection mapping (2.1). This mapping is $F \circ P$: $R^n \to R^n$, and as can be readily verified, it is PC^1 on (S, Σ').

3. Local degree of PC^1 mappings. In this section we consider subdivided polyhedron (R^n, Σ), and a PC^1 mapping $F: R^n \to R^n$. Our aim is to get sufficient conditions which establish the local degree of such a mapping. We now have a lemma, which will then be used to prove the main result:

LEMMA 3.1. *Let $\hat{x}$ be such that $DF_\sigma(\hat{x})$ is nonsingular for all σ containing $\hat{x}$. Then, there exist positive numbers α and ε such that*

$$(3.1) \qquad \|F(x) - F(\hat{x})\| \geqq \alpha \|x - \hat{x}\| \quad \textit{for all } x \in B_\varepsilon(\hat{x}) = \{x: \|x - \hat{x}\| \leqq \varepsilon\}.$$

Proof. Let $\sigma_1, \sigma_2, \cdots, \sigma_k$ be all the pieces of Σ which contain $\hat{x}$. Then, there is a $\delta > 0$ such that

$$B_\delta(\hat{x}) \subset \bigcup_{i=1}^{k} \sigma_i.$$

Let $\sigma' \in \{\sigma_i : i = 1, \cdots, k\}$. For each $x \in B_\delta(\hat{x}) \cap \sigma'$, we have

$$\|F(x) - F(\hat{x})\| \geqq \|DF_{\sigma'}(\hat{x})(x - \hat{x})\| - o(\|x - \hat{x}\|)$$

and since $DF_{\sigma'}(\hat{x})$ is nonsingular, there is $\alpha' > 0$ such that

$$\|F(x) - F(\hat{x})\| \geqq 2\alpha' \|x - \hat{x}\| - o(\|x - \hat{x}\|).$$

Hence, there is a $\delta > \varepsilon' > 0$ and $0 < \alpha'' < \alpha'$ such that

$$\|F(x) - F(\hat{x})\| \geqq \alpha'' \|x - \hat{x}\| \quad \text{for all } x \in B_{\varepsilon'}(\hat{x}) \cap \sigma';$$

now letting α be the smallest α'' and ε be the smallest ε', we have our result.

LEMMA 3.2. *Let F be continuously differentiable, and H a subspace of R^n of dimension $\leqq n - 1$. Then $F(H + v)$ contains no open set, for all $v \in R^n$.*

Proof. Since F is Lipschitz continuous, the proof of the lemma follows trivially.

Given a continuous mapping F and an open bounded set C with $y \notin \partial C$, following Ortega and Rheinboldt [15], we define deg (F, C, y) to be the degree of F with respect to C at y.

We now prove our main theorem:

THEOREM 3.3. *Let $\hat{x} \in R^n$, such that* det $(DF_\sigma(\hat{x}))$ *is positive (negative) for every σ containing $\hat{x}$. Then, there exists $\varepsilon > 0$ such that* deg $(F, B_\delta(\hat{x}), F(\hat{x})) \geqq +1$ $(\leqq -1)$ *for each δ in* $(0, \varepsilon)$.

Proof. By Lemma 3.1, we have $\varepsilon > 0$, $\alpha > 0$ satisfying the hypothesis of the lemma. We shall now show the theorem for the case when det $(DF_\sigma(\hat{x})) > 0$ for all σ containing $\hat{x}$.

Let $0 < \delta < \varepsilon$, $B = B_\delta(\hat{x})$, $y = F(\hat{x})$ and ∂B the boundary of B. Then, from Lemma 3.1,

$$\|F(x) - F(\hat{x})\| \geqq \alpha\delta \quad \text{for all } x \in \partial B \text{ and some } \alpha > 0.$$

Since F is continuous, there exists a $\beta > 0$ such that

$$\|F(x) - q\| \geqq \alpha\delta/2 \quad \text{for all } x \text{ in } \partial B \text{ and } q \text{ in hull } \{F(B_\beta(\hat{x}))\}.$$

Let $q \in F(B_\beta(\hat{x}))$. Now consider the mapping $G(x) = F(x) + y - q$, and the homotopy $H: B \times [0, 1] \to R^n$ defined by $H(x, t) = (1 - t)G(x) + tF(x)$. Then, for (x, t) in $\partial B \times [0, 1]$ we have

$$\|H(x, t) - y\| = \|F(x) - (ty + (1 - t)q)\| \geqq \frac{\alpha\delta}{2}.$$

Hence, by the homotopy invariance theorem [15, 6.2.2], deg $(G, B, y) =$ deg (F, B, y), or deg $(F, B, q) = (F, B, y)$.

Since $DF_\sigma(x)$ is nonsingular for all x in $B \cap \sigma$ and σ containing $\hat{x}$, using the inverse function theorem, it can be established that $F(B_\beta(\hat{x}))$ contains an open ball U. Also, from Lemma 3.2, the image $F(\tau)$ of a proper face τ of any piece σ contains no open ball.

Thus, we can choose q in U such that $\hat{B} = \{x \in B : F(x) = q\}$ does not intersect any facet of a piece σ. Hence, as the $\hat{B}$ is a set of isolated points, and the local degree of each $x \in \hat{B}$ is $+1$, we have, from the fact that the degree of a mapping is the sum of local degrees (from the decomposition of domain [15, 6.2.7]), that

$$\deg(G, B, y) \geqq 1$$

and we have our result.

For a continuously differentiable mapping F, if det $DF(x) > 0$ at some x, using the inverse function theorem [15, 5.2.1] there is an open bounded set U such that deg $(F, U, F(x)) = +1$. We observe that this stronger conclusion for the above result is not possible. For this, consider the piecewise linear function of Fig. 3.1. The degree of this mapping at 0 is 2, and the determinant of Jacobian of the linear mapping on each piece is positive.

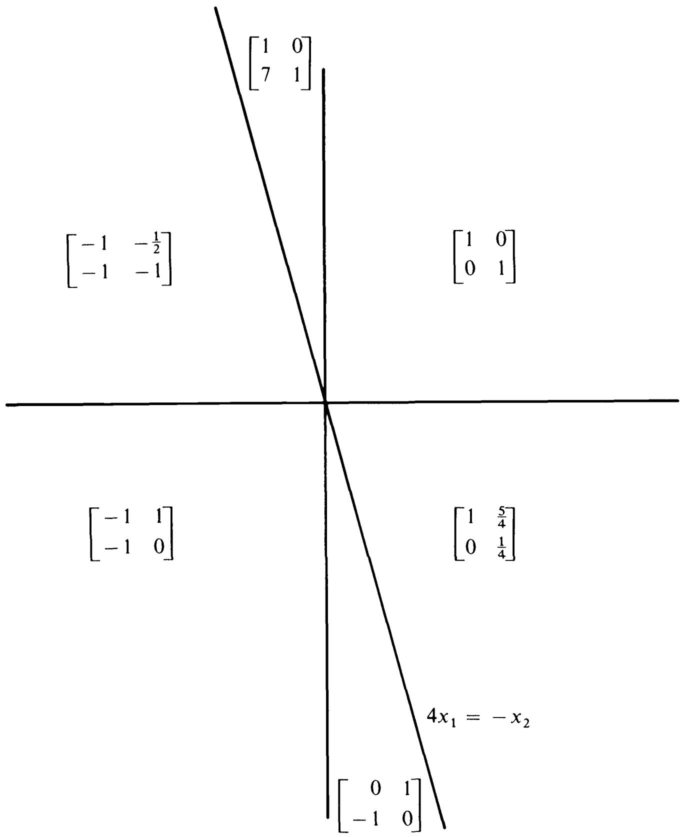

FIG. 3.1

4. PC^1 homeomorphisms of compact convex polyhedrons. Let (S, Σ) be a subdivided compact convex polyhedron, and let $F: S \to R^n$ be a PC^1 mapping. In this section we consider the conditions on F and S under which F maps S homeomorphically onto $F(S)$, i.e., $F(x) = y$ has a unique solution for each $y \in F(S)$. The results presented in this section are in the spirit of the recent extension of the Gale–Nikaido theorem [5] by Mas-Collel [12] (see also Garcia and Zangwill [6]).

Let $P: R^n \to S$ be the projection mapping (2.1), and let $G: R^n \to R^n$ be the mapping

$$G(x) = F \circ P(x) + x - P(x). \tag{4.1}$$

We observe that G is a PC^1 mapping on the subdivided polyhedron (R^n, Σ).

We now state our condition, which is the same as the one used by Mas-Collel [12] (compare also with condition (ii), Corollary 2.6 of Saigal and Simon [19]).

Condition 4.1. Let x in S lie in a face T of S. Also, let x be an element of σ where σ in Σ is a piece such that $\dim \sigma \cap T = \dim T$, H_T be the subspace spanned by T and P_T the projection mapping of R^n onto H_T. Then, the linear mapping $P_T \circ DF_\sigma(x): H_T \to H_T$ has positive determinant.

Under this condition, the following can be proved as is done for Lemma 1 in [12] (see also [19, Lemma 3.4]).

LEMMA 4.2. *Let x be arbitrary, and lie in the pieces σ_i, $i = 1, \cdots, k$, in Σ'. Then* $\det DG_\sigma(x) > 0$ *for each* $\sigma = \sigma_i$, $i = 1, \cdots, k$.

We now prove our main theorem.

THEOREM 4.3. *Let (S, Σ) be a subdivided compact convex polyhedron, and let $F: S \to R^n$ be a PC^1 mapping. Also, let F and S satisfy Condition 4.1. Then F maps S homeomorphically onto $F(S)$.*

Proof. Extend Σ to a subdivision Σ' using the mapping (2.1), and let the mapping G of (4.1) be the corresponding PC^1 extension of F.

Now, from Theorem 2.3, since the Condition 4.1 implies that the determinant of G is positive in each piece, for each x in R^n, there exists an open ball B such that

$$\deg(G, B, F(x)) \geqq 1. \tag{4.2}$$

Let A be a $n \times n$ positive definite matrix, and consider the homotopy

$$H(x, t) = (1-t)Ax + t(G(x) - y) \tag{4.3}$$

for any $y \in F(S)$. We now show that $H^{-1}(0)$ is bounded, and thus the degree of $G - y$ is $+1$ since it is homotopic to a map of degree $+1$. But, this is true, since for sufficiently large x, $x^TAx > 0$ and

$$x^TG(x) - x^Ty = x^Tx - x^T(F \circ P(x) - P(x) - y) > 0$$

since $F \circ P(x) - P(x) + y$ is bounded.

Now, using the decomposition of domain [15, 6.2.7] and (4.2), we conclude that, for each y in $F(S)$, $\{x : F(x) = y\}$ is a singleton, and we are done.

Note. This theorem is false if the property of positive determinants is replaced by negative determinants. A counterexample for a PL mapping is given in Fig. 4.1. This demonstrates that such an extension for C^1 mappings involving $\Sigma = \{S\}$ may also be hard, and conjecture that in this case, the result is true (see also [12]).

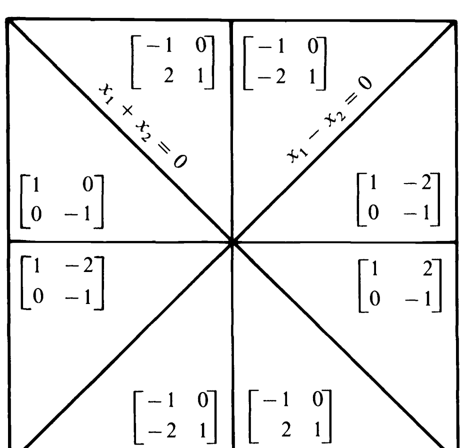

FIG. 4.1

5. On PL homeomorphisms of R^n. In this section we give a set of sufficient conditions for a piecewise linear function in R^n to be a homeomorphism. Let (R^n, Σ) be a subdivided polyhedron, and let

$$F: R^n \to R^n$$

be piecewise linear on this subdivision, i.e., PC^1 with affine on each piece of Σ. Since Σ contains a finite number of pieces, outside some compact region, points of R^n lie in some unbounded piece in Σ. Let these unbounded pieces be numbered $\sigma_1, \sigma_2, \cdots, \sigma_k$ for some k, and let $F|\sigma_i(x) = A_i x - a_i$ for some $n \times n$ matrices A_i, and n-vectors a_i. Then, we can prove:

THEOREM 5.1. *Assume that the Jacobian matrix of each piece of linearity of F has a positive determinant. Also, let there exist a matrix B such that $(1-t)B + tA_i$ is nonsingular for each $t \in [0, 1]$ and $i = 1, \cdots, k$. Then, F is a homeomorphism.*

Proof. Let y be arbitrary. Then, consider the homotopy

$$H(x, t) = (1-t)Bx + t(F(x) - y), \qquad t \in [0, 1]. \tag{5.1}$$

We claim that $H^{-1}(0)$ has no unbounded component. This is true, since the contrary implies that for some σ_i, we can find a sequence $(x^p, t_p) \in H^{-1}(0)$, $p = 1, 2, \cdots$, such that $x^p \in \sigma_i$ and $\|x^p\| \to \infty$. Also, on some subsequences $x^p/\|x^p\| \to x^*$, $t_p \to t^*$, $t^* \in [0, 1]$ and $x^* \neq 0$. Hence, from (5.1), $(1-t_p)Bx^p + t_p(A_i x^p - a^i) - t_p y = 0$. Dividing by $\|x^p\|$ and taking limits, we get

$$(1-t^*)Bx^* + t^* A_i x^* = 0$$

which is a contradiction. Now, to see that it is one to one and onto, we observe that since $H^{-1}(0)$ is bounded for each y, and $\det(B) > 0$, from the homotopy invariance theorem [15, 6.2.2] the degree of $F(x) - y$ is $+1$ for all y. The result then follows from Theorem 3.3.

The onto part of the theorem also follows from the works of several authors, including Chien and Kuh [1], Rheinboldt and Vandergraft [16]. The sufficiency condition of Theorem 5.1 is weaker than that of Fujisawa and Kuh [4], see Saigal and Kojima [11]. In Fig. 5.1 we present a homeomorphism satisfying the conditions of our

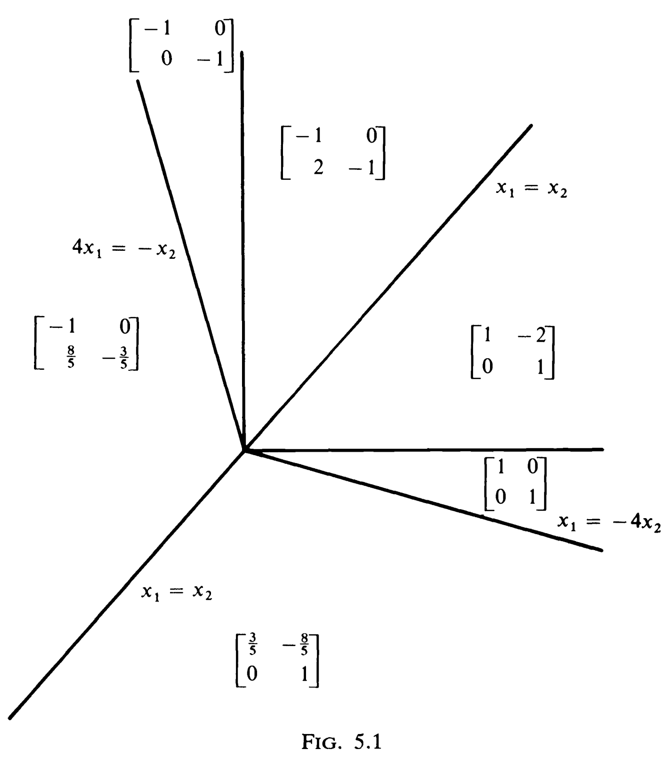

FIG. 5.1

theorem with $B=\begin{bmatrix}1 & -1\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$. Also, since I and $-I$ appear as Jacobians of the pieces of linearity, no linear transform of it will satisfy the condition of [4], though there is a linear transform for which the homeomorphism of [4, Fig. 7] will satisfy the condition. Now consider the example of Fig. 5.2. This is a homeomorphism which does not satisfy the

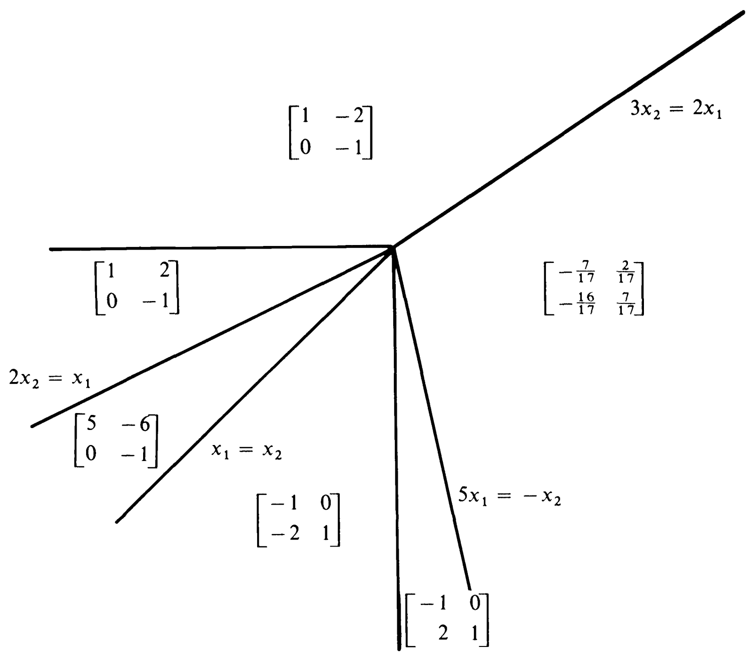

FIG. 5.2

condition of Theorem 5.1, and is thus a counterexample to the necessity of our condition. To see this, note that the matrices

$$\begin{bmatrix}-1 & 0\\ -2 & 1\end{bmatrix},\quad \begin{bmatrix}1 & -2\\ 0 & -1\end{bmatrix},\quad \begin{bmatrix}1 & 2\\ 0 & -1\end{bmatrix},\quad \begin{bmatrix}-1 & 0\\ 2 & 1\end{bmatrix}$$

are Jacobians of the pieces of linearity of the nonhomeomorphisms of Fig. 4.1, and, for these, thus, there is no matrix B satisfying the conditions of Theorem 5.1. We also observe that this example is also a counterexample to any set of necessary and sufficient conditions put on all subsets of the Jacobians of the pieces of linearity.

Theorem 5.1 is true if the property of positive determinants is replaced by negative determinants. Also, if the unbounded pieces satisfy the conditions of Theorem 5.1, it can be readily shown that $\{x: f(x)=y\}$ has an odd number of elements, if each of its elements lies interior to some piece.

A corollary to Theorem 5.1 is the following result which can also be considered as an explanation of the boundary condition 4.1. Let (S, Σ) be a subdivided compact polyhedron, with $F: S\to R^n$ a piecewise linear function. Then, we can prove

COROLLARY 5.3. *Let (S, Σ) admit an extension (R^n, Σ) such that F can be extended to F' on R^n with $F'|\sigma'$ affine, and F' satisfying the conditions of Theorem* 5.1. *Then F maps S homeomorphically onto $F(S)$.*

Two applications of this corollary are given in the next section.

6. Two PL homeomorphisms. We now present two PL homeomorphisms; one satisfies the sufficiency condition of Fujisawa and Kuh [4] while the other does not. The first homeomorphism is constructed by the use of a matrix which has all principal minors positive, and establishes the sufficiency part of the Samelson–Thrall–Wesler [20] partition theorem. The other is constructed by using a matrix which has all principal minors negative. In the process of the construction, we will prove the main theorem of Kojima and Saigal [10], and this can be considered a degree theoretic proof of the same. We now introduce the necessary notation.

Let $I=\{1,\cdots,n\}$ and U and V be $n\times n$ nonsingular matrices. Now, for any $J\subseteq I$ let $W_J=(W_1,\cdots,W_n)$ be the $n\times n$ matrix with

$$W_j=\begin{cases}U_j, & j\in J,\\ V_j, & j\notin J.\end{cases}\tag{6.1}$$

Also, let $\operatorname{pos}(A)=\{y\colon y=Ax,\ x\geqq 0\}$ represent the cone generated by a matrix A. For $J\subseteq I$, let $\sigma(J)=\{x\colon x_j\geqq 0,\ j\in J \text{ and } x_j\leqq 0,\ j\notin J\}$, and by $\Sigma=\{\sigma(J)\colon J\subseteq I\}$. In this case, (R^n,Σ) is a subdivided polyhedron. Now, define the PL mapping $F\colon R^n\to R^n$ by

$$F(x)=\sum_{j\in J}U_jx_j+\sum_{j\notin J}V_jx_j\tag{6.2}$$

for x in $\sigma(J)$.

6.1. The first homeomorphism. We now prove our first homeomorphism theorem:

THEOREM 6.1. *Let U, V, W_J, $J\subseteq I$, be defined as above, and let* $\det(W_J)>0$ *for each J. Then F is a PL homeomorphism of R^n onto R^n, on the subdivided polyhedron (R^n,Σ).*

Proof. On each piece of Σ, $F(x)=W_Jx$ for $x\in\sigma(J)$. Also, $\det(U^{-1}W_J)=\det(U^{-1})\det(W_J)>0$ for all J. By choosing $\bar J\subseteq J$, $\bar J\subseteq I$, we can show that each principal minor of $U^{-1}W_J$ is positive. Hence, for each J, we have $\det((1-t)U+tW_J)=\det(U)(\det((1-t)I+tU^{-1}W_J)>0$ since $U^{-1}W_J$ has all principal minors positive (see Lemma 3.1.1, Saigal [17]). Hence, the result follows from Theorem 5.1.

As a corollary of this theorem, we prove the sufficiency part of the Samelson–Thrall–Wesler [20] partition theorem.

COROLLARY 6.2. *Let U, V, W_J, $J\subseteq I$, be defined as above, and* $\det(U)>0$, *with* $\det(W_J)=(-1)^{|J|}$ *when $|J|$ is the number of elements in J. Then, the collection of cones $\Delta=\{\operatorname{pos}(W_J)\colon J\subseteq I\}$ partitions R^n.*

Proof. Define

$$\bar W_J=\begin{cases}W_j, & j\in J,\\ -W_j, & j\notin J\end{cases}$$

and we note that the mapping $F(x)=\bar W_Jx$, $x\in\sigma(J)$ is a PL mapping. Also, since $\det(\bar W_J)>0$, $F(x)$ is a PL homeomorphism from Theorem 6.1. This corollary follows by observing that the cones of Δ are images of the cones of Σ.

6.2. The second homeomorphism. In this section we consider $U=E$ (the identity matrix) and $V\nleq 0$ a matrix having all principal minors negative. Then Kojima and Saigal [10] have shown that F defined by (6.2) is not a homeomorphism of R^n on the subdivided polyhedron (R^n,Σ). In this section, we will show that there exists a PL-homeomorphism G of $Q=R^n\backslash R^n_+$ onto Q such that $F\circ G$ is the identity mapping on Q (where R^n_+ is the nonnegative orthant).

Now, define $\Sigma' = \Sigma \backslash \sigma(I)$. Then (Q, Σ') is a subdivided polyhedron (which is not convex). Define $\hat{F}$ as the restriction of F(as defined by 6.2)) to Q. We now state some preliminary results.

LEMMA 6.3. *Let $V \nleq 0$ and have all principal minors negative. Then, there is a $d > 0$ such that $Vd > 0$.*

Proof. For the proof see [10, Lemma 2.1].

LEMMA 6.4. *Let V have all principal minors negative. Then all proper principal minors of V^{-1} are positive.*

Proof. For the proof see [10, Lemma 4.1].

Now, for $d > 0$ such that $Vd > 0$, consider the homotopy:

$$H(x, t) = (1-t)V(x+d) + t[\hat{F}(x) + Vd]. \tag{6.3}$$

LEMMA 6.5. $H^{-1}(0) \cap \partial Q \times [0, 1] = \emptyset$.

Proof. Assume the contrary, that there is a $(x, t) \in \partial Q \times [0, 1]$ with $(x, t) \in H^{-1}(0)$. Then, $x \geqq 0$ with $J = \{j : x_j > 0\}$, $|J| < n$. Thus, if $J \neq \emptyset$, $(1-t)Vx + tW_J x = -Vd$, or multiplying by V^{-1}, we get

$$(1-t)x + tV^{-1}W_J x = -d. \tag{6.4}$$

Now, let A be the principal minor of V^{-1} in $V^{-1}W_J$. Then, from (6.4) we can conclude that $A\bar{x} < 0$, $\bar{x} > 0$ has a solution. But, from Lemma 6.4, A has all positive principal minors, which leads to a contradiction, [6]. Also, $J \neq \emptyset$, since the contrary implies that $x = -d$.

We are now ready to prove our main result.

LEMMA 6.6. *$\{x : \hat{F}(x) = -Vd\}$ is a singleton.*

Proof. Assume the contrary. Then, since Lemma 6.5 implies that in $H^{-1}(0)$ no solution inside Q lies on a component intersecting ∂Q, there must be an unbounded component inside Q.

But, since $\hat{F}(x) = W_J x$ for some $J \subseteq I$, and $V^{-1}W_J$ has all positive principal minors, using arguments of Theorem 5.1, we get a contradiction. Thus, the result follows.

THEOREM 6.7. *For any $y \in Q$, $T = \{x : F(x) = y\}$ is a singleton. Also, $T \subset Q$.*

Proof. For any $y \in Q$, $y \ngeqq 0$ and thus $T \cap \sigma(I) = \emptyset$. Hence, for each x in T, $\det DF(x)$ is the same as the determinant of some principal minor of V, and so $\det DF(x) < 0$. Hence, from Theorem 3.3, deg $(F, B, y) \leqq -1$ for some neighborhood B of x.

Now consider the homotopy

$$H(x, t) = F(x) + (1-t)Vd - ty;$$

we note that $H(x, 0) = 0$ has a unique solution $x = -d$, from Lemma 6.6. Hence the degree of $H(x, 0)$ is -1. Also, $H^{-1}(0) \subset Q \times [0, 1]$ is bounded, and hence by the homotopy invariance theorem, [15, 6.2.2], $H(x, 1)$ has degree -1. Since deg $(\hat{F}, B, y) =$ deg $(F, B, y) \leqq -1$, the result follows.

We now prove the main result of this section.

THEOREM 6.8. *Let $V \nleq 0$ and have all principal minors negative, $U = E$ and F be as defined by (6.2). Then there exists a PL homeomorphism G on a subdivision of Q such that $F \circ G$ is the identity on Q.*

Proof. Let Δ be as in Corollary 6.2 and let Δ' be the collection of polyhedrons of the type $\sigma_i = \sigma \cap \{x : x_i \leqq 0\}$, $i = 1, \cdots, n$, and $\sigma \in \Delta$. Then it is readily confirmed that (Q, Δ') is a subdivided polyhedron. Define $G: Q \to Q$ by $y \to \{x : F(x) = y\}$. This is well defined by Theorem 6.7. Also G is PL, and for y in pos (W_J), $J \neq I$, $G(y) = W_J^{-1}y$, and that is a homeomorphism of Q onto $G(Q)$.

We now give an example of such a mapping G. Let $V = \begin{bmatrix} -1 & 2 \\ 4 & -6 \end{bmatrix}$. For this case

$$W_\varnothing = V, \qquad W_{\{1\}} = \begin{bmatrix} 1 & 2 \\ 0 & -6 \end{bmatrix}, \qquad W_{\{2\}} = \begin{bmatrix} -1 & 0 \\ 4 & 1 \end{bmatrix},$$

$$W_\varnothing^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & \frac{1}{2} \end{bmatrix}, \qquad W_{\{1\}}^{-1} = \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & -\frac{1}{6} \end{bmatrix}, \qquad W_{\{2\}}^{-1} = \begin{bmatrix} -1 & 0 \\ 4 & 1 \end{bmatrix}.$$

The pieces of linearity of the mapping G are given in Fig. 6.1.

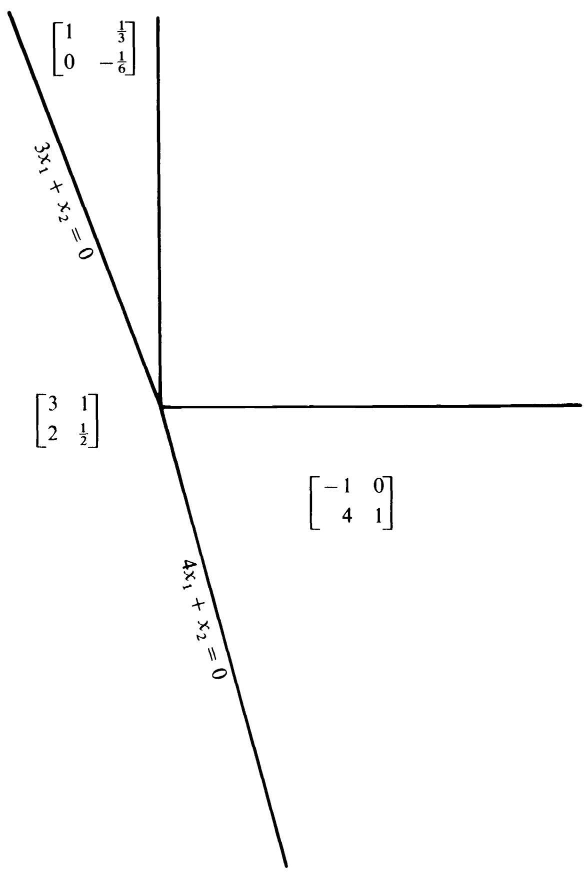

FIG. 6.1

Also, as can be readily confirmed, G, in R^2, has a PL extension onto R^2 which is also a homeomorphism of R^2. For the above example, if one added $R^2_+ = \sigma(I)$ to the set Δ', and extended the mapping G to $\hat{G}$ by

$$\hat{G}(y) = \begin{cases} G(y), & y \in Q, \\ Wy, & y \in \sigma(I), \end{cases}$$

where $W = \begin{bmatrix} -1 & \frac{1}{3} \\ 4 & -\frac{1}{6} \end{bmatrix}$ (the matrix consisting of the nontrivial columns of $W_{\{i\}}$), $\hat{G}$ maps R^2 homeomorphically onto R^2.

We conjecture that G has such an extention in n-dimensional Euclidean space as well, but see no way to prove this.

7. Extensions when the Jacobians may be singular. Our aim in this section is to extend the results of §§ 3 and 4 to cases when the Jacobians of the mappings may be singular. Our main assumption is that for any y in R^n, the sets of the type $\{x: F(x)=y$ and $DF(x)$ is singular$\}$ are finite. We then show that the results of § 3 can be extended, and thus a further extension of the Gale–Nikaido theorem [5] is obtained.

We consider a subdivided polyhedron (R^n, Σ) and consider a PC^1 mapping $F: R^n \to R^n$ on it. Then, an extension of the Lemma 3.1 is the following.

LEMMA 7.1. *Let $\hat{x}$ be in R^n and $\sigma_1, \cdots, \sigma_k$ be the pieces in which it lies. Suppose that $\{x \in \sigma_i: F(x)=F(\hat{x})$ and $DF_{\sigma_i}(x)$ is singular$\}$ has at most a finite number of elements, for each i. Then, for each $\varepsilon_0>0$, there is a $0<\varepsilon<\varepsilon_0$ such that*

$$\|F(x)-F(\hat{x})\|>0 \quad \text{if } \|x-\hat{x}\|=\varepsilon.$$

Proof. Let $\varepsilon_0>0$, $y=F(\hat{x})$ and $X=U_i\{x \in \sigma_i: F(x)=y$ and $DF_{\sigma_i}(x)$ is singular$\}$. Since X is finite, there is a positive number $\delta<\varepsilon_0$ such that $B\equiv B_\delta(\hat{x}) \subset U_i\sigma_i$ and $\partial B \cap X=\emptyset$. Hence, for each x in ∂B, we have either $F(x)\neq y$ or $F(x)=y$ and $DF_\sigma(x)$ nonsingular. In the former case, by the continuity of F, there exists $\gamma(x)>0$ such that $y \notin F(B_{\gamma(x)}(x))$, and in the latter case, by Lemma 3.1, a $\gamma(x)>0$ such that $y \notin F[B_{\gamma(x)}(x)\backslash\{x\}]$. Let $V=\bigcup_{x\in\partial B} \operatorname{int}(B_{\gamma(x)}(x))$, and V is an open set in R^n with $\partial B \subset V$ and $F(x)\neq y$ for all $x \in V\backslash\partial B$.

Hence we can choose $0<\varepsilon<\delta$ with the required property.

We now use Lemma 7.1 to compute the local degree of a mapping.

THEOREM 7.2. *For every piece σ in Σ and y in R^n, let*

(7.1) $$\det DF(x) \geqq 0 \quad \text{for all } x \in \sigma.$$

(7.2) $$\{x \in \sigma: DF_\sigma(x) \text{ is singular}\} \quad \text{contains no open set.}$$

(7.3) $$\{x \in \sigma: F(x)=y \text{ and } DF_\sigma(x) \text{ is singular}\} \quad \text{has at most a finite number of elements.}$$

Then, for every x in R^n and $\varepsilon_0>0$, there is $0<\varepsilon<\varepsilon_0$ such that

$$\deg(F, B_\varepsilon(\hat{x}), F(\hat{x})) \geqq 1.$$

Proof. Let x in R^n and $\varepsilon_0>0$. By Lemma 7.1, there is a positive number $\varepsilon<\varepsilon_0$ such that for $B=B_\varepsilon(\hat{x})$, $y=F(\hat{x})$ we have

$$\|F(x)-y\|>0 \quad \text{for all } x \in \partial B.$$

Using arguments identical to those of Theorem 3.2, we have our result.

We note that if, in (7.1), we assumed that the $\det(Df_\sigma(x)) \leqq 0$, then, by an identical argument, we could establish that $\deg(F, y, B) \leqq -1$.

We now weaken the hypothesis of Condition 4.1 so that we can obtain a further generalization of the Mas-Colell [12] generalization of the Gale–Nikaido theorem [5].

Consider a PC^1 mapping $F: S \to R^n$ on the subdivided compact convex polyhedron (S, Σ). We now state our condition.

Condition 7.3. Let T be a face of S and σ a piece in Σ such that the dimension of $\tau=\sigma \cap T$ is the same as the dimension of T. Then

(7.4) $$P_T \circ DF_\sigma(x): H_T \to H_T \quad \text{has nonnegative determinant for each } x \text{ in } \tau.$$

(7.5) $$\{x \in \tau: P_T \cdot DF_\sigma(x) \text{ is singular}\} \quad \text{has at most a finite number of elements.}$$

We now show that if Condition 7.3 is satisfied, then the conditions of Theorem 7.2 are satisfied.

LEMMA 7.4. *If F satisfies Condition* 7.3, *then the mapping G defined by* (4.1) *satisfies the conditions of Theorem* 7.2.

Proof. Let (R^n, Σ') be the subdivision on which G, defined by (4.1), is PC^1, and let $\sigma \in \Sigma'$ be an unbounded piece. Then there exists a face F of S and $\bar{\sigma} \in \Sigma$ such that if $\tau = \bar{\sigma} \cap T$, $\dim \tau = \dim T$, and $P_T(x) \in \tau$ for all $x \in \sigma$; and

$$DG_\sigma(x) = P_T DF_{\bar{\sigma}}(P_T(x)) + I - P_T$$

Now, by using the same argument as in the proof of Lemma 1 of [12], we have (7.1), and

$$\det DG_\sigma(x) = 0 \quad \text{iff } P_T DF_{\bar{\sigma}}(P_T(x)) \text{ is singular.} \tag{7.6}$$

If $\dim H_T = 0$, then $DG_\sigma(x) = I$ for all $x \in \sigma$ and (7.2) holds. Now we take $\dim H_T \geqq 1$. Now, assume the set $\{x \in \sigma: DG_\sigma(x) \text{ is singular}\}$ contains and open set X. Then, the projection $P_T(X)$ of X into H_T is open and $P_T DF_{\bar{\sigma}}(P_T(x))$ is singular on $P_T(X)$. Since $\dim H_T \geqq 1$, this contradicts (7.5). Thus we have shown (7.2). It follows from (7.5) that there exists a finite number of points $x^1, x^2, \cdots, x^m$ in τ such that $P_T DF_{\bar{\sigma}}(x)$ has a positive determinant if $x \in \tau$ and $x \neq x^i$, $i = 1, \cdots, m$. Let $y \in R^n$ and $Y = \{x \in \sigma: G(x) = y \text{ and } DG_\sigma(x) \text{ is singular}\}$. By (7.6) we obtain

$$\begin{aligned} Y &\subset \bigcup_{i=1}^{m} \{x \in \sigma: P_T(x) = x^i,\ F(x^i) + x^i - x = y\} \\ &\subset \bigcup_{i=1}^{m} \{x \in R^n: F(x^i) + x^i - x = y\} \end{aligned}$$

and we see that $\{x \in R^n: F(x^i) + x^i - x = y\}$ has at most one element, and thus (7.3) follows.

Thus, we obtain the following theorem:

THEOREM 7.5. *If F satisfies Condition* 7.3, *then F maps S homeomorphically onto* $F(S)$.

Proof. The theorem follows directly from Theorem 7.2, Lemma 7.4, and the argument used in the proof of Theorem 4.3.

Postscript. Recently it was brought to our attention that G. Chichilinsky, M. Hirsch and H. Scarf has also verified the extension of the Gale–Nikaido theorem as considered in [12]. In addition, Y. Kawamura has extended the homeomorphism theorem of Fujisawa and Kuh [4] to the case where the functions are Lipschitz continuous.

REFERENCES

[1] M. J. CHIEN AND E. S. KUH, *Solving piecewise-linear equations for resistive networks*, Circuit Theory and Appl., 4 (1976), pp. 3–4.

[2] B. C. EAVES AND R. SAIGAL, *Homotopies for computation of fixed points on unbounded regions*, Math. Programming, 3 (1972), pp. 225–237.

[3] B. C. EAVES AND H. SCARF, *The solution of systems of piecewise linear equations*, Math. Operations Res., 1 (1976), pp. 1–27.

[4] T. FUJISAWA AND E. S. KUH, *Piecewise linear theory of nonlinear networks*, SIAM J. Appl. Math., 22 (1972), pp. 307–328.

[5] D. GALE AND H. NIKAIDO, *The Jacobian matrix and the global univalence of mappings*, Math. Ann., 159 (1965), pp. 81–93.

[6] C. B. GARCIA AND W. I. ZANGWILL, *On univalence and P-matrices*, Report 7737, Center for Mathematical Studies in Business and Economics, University of Chicago, July 1977.

[7] M. W. HIRSCH, *Differential Topology*, Springer-Verlag, New York, 1976.
[8] M. KOJIMA, *Studies on PL approximations of PC^1-mappings in fixed points and complementarity theory*, Math. Operations Res., 3 (1968), pp. 17–36.
[9] M. KOJIMA, H. NISHINO AND N. ARIMA, *A PL Homotopy for finding all roots of a polynomial*, Math. Programming, 16 (1979), pp. 37–62.
[10] M. KOJIMA AND R. SAIGAL, *On the number of solutions for a class of linear complementarity problems*, Math. Programming, submitted.
[11] ———, *On the relationship between conditions that insure a PL mapping in a homeomorphism*, Math. Operations Res., to appear.
[12] A. MAS-COLELL, *Homeomorphisms of compact convex sets and the Jacobian matrix*, manuscript, Universität Bonn, April 1977, this Journal, 10 (1979), pp. 1105–1109.
[13] J. MILNOR, *Topology from the Differentiable Viewpoint*, The University Press of Virginia, Charlottesville, VA, 1965.
[14] T. OHTSUKI, T. FUJISAWA AND S. KUMAGAI, *Existence theorems and a solution algorithm for piecewise linear resister networks*, this Journal, 8 (1977), pp. 69–99.
[15] J. ORTEGA AND W. RHEINBOLDT, *Iterative Solutions of Nonlinear Equations on Several Variables*, Academic Press, New York, 1970.
[16] W. C. RHEINBOLDT AND J. S. VANDERGRAFT, *On piecewise affine mappings in R^n*, SIAM J. Appl. Math., 29 (1975), pp. 680–689.
[17] R. SAIGAL, *On paths generated by fixed point algorithms*, Math. Operations Res., 1 (1976), pp. 359–380.
[18] ———, *On piecewise linear approximations to smooth mappings*, Ibid., to appear.
[19] R. SAIGAL AND C. B. SIMON, *Generic properties of the complementarity problem*, Math. Programming, 4 (1973), pp. 324–335.
[20] H. SAMELSON, R. M. THRALL AND O. WESLER, *A partition theorem for Euclidean n-space*, Proc. Amer. Math. Soc., 9 (1958), pp. 805–807.
[21] H. SCARF, private letter, April 1975.
[22] ———, *The approximation of fixed points of a continuous mapping*, SIAM J. Appl. Math., 15 (1967), pp. 1328–1343.
[23] J. T. SCHWARTZ, *Nonlinear Functional Analysis*, Gordon and Breach, New York, 1969.
[24] M. J. TODD, *Orientation in complementary pivoting*, Math. Operations Res., 1 (1976), pp. 54–66.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0019 $01.00/0

WEAK SOLUTIONS OF $(p(x)u'(x))'+g(x)u'(x)+qu(x)=f$ WITH $q, f \in H_{-1}[a, b]$, $0<p(x)\in L_\infty[a, b]$, $g(x)\in L_\infty[a, b]$ AND $u \in H_1[a, b]$*

R. E. WHITE†

Abstract. In this paper we consider weak solutions $u(x)\in H_1[a, b]$ of the differential equation $(p(x)u'(x))'+g(x)u'(x)+qu(x)=f$ where q and f may be distributional derivatives of elements in $L_2[a, b]$. The linear initial value and linear boundary value $B_1u\equiv\alpha_1u(a)+\beta_1u'(a)=0$, $B_2u\equiv\alpha_2u(b)+\beta_2u'(b)=0$, $\alpha_1, \alpha_2>0$, $\beta_1\leqq 0\leqq\beta_2$ problems are studied. Also the semilinear boundary value problem where $f=f(x, u)$ and $B_1u=b_1(u)$ and $B_2u=b_2(u)$ is studied.

1. Introduction. Consider the ordinary differential equation $(p(x)u'(x))'+g(x)u'(x)+qu(x)=f$ where $0<p(x)\in L_\infty[a, b]$, $g(x)\in L_\infty[a, b]$, $q, f\in H_{-1}[a, b]$ and $u\in H_1[a, b]$. $H_{-1}[a, b]$ is the set of distributions (see L. Schwartz [8]) which are distributional derivatives of functions in $L_2[a, b]$ and $L_2[a, b]$ is the set $L_2(a, b)$ functions with distributional values at a and b. Consequently, q or f may be a delta "function" or a function of the form $(x-a)^{-1-\alpha}$ where $\alpha>\frac{1}{2}$. A particular example of $qu(x)$ is $\delta(x-x_0)u(x)=\delta(x-x_0)u(x_0)$ where $u\in H_1[a, b]$. $H_1[a, b]$ is the set of functions in $L_2[a, b]$ whose distributional derivatives are also in $L_2[a, b]$ and hence $H_1[a, b]$ is a subspace of the Sobolev space $H_1(a, b)$ which is continuously embedded into $C[a, b]$. The general definition of $qu(x)$ will place $qu(x)\in H_{-1}[a, b]$. Thus, by a solution of the differential equation we shall mean a distributional solution.

The advantages of considering weak (distributional) solutions are three fold: First, we may consider more general coefficients and inhomogeneous term. Second, the conditions for existence are relaxed and hence easier to establish. Third, because of the previous two, nonlinear problems are more readily dealt with.

Since our solutions are to be in $H_1[a, b]$, they are also in $C[a, b]$ and hence physically meaningful. Physical situations in which ordinary differential equations are of the above form occur in many problems. Perhaps the simplest example is the steady state ideal string problem with a point force at x_0. The differential equation is $-u''=a\delta(x-x_0)$ where a is proportional to the magnitude of the force. The derivation of this problem may be found in I. Stakgold [11]. This type of problem with inhomogeneous term $\delta(x-x_0)$ also occurs in the definition of the Green's function for Sturm–Louville problems. Another example comes from the steady state heat conduction problem between two plates with a heat source between them. The differential equations is $-(p(x)u')'=a\delta(x-x_0)$ where $p(x)=c_1(1-H(x-x_0))+c_2H(x-x_0)$, $H(x)\equiv 0$ for $x<0$ and $\equiv 1$ for $x\geqq 0$ and is called the Heavyside function, c_1 is the thermal conducting of the bottom plate and c_2 is the thermal conductivity of the top plate. This problem is discussed for $c_1=c_2$ in [8]. In general many problems of a diffusion nature in steady state and in one dimension are described by such equations. For example see D. D. Joseph and E. M. Sparrow [4] and the interesting references pretaining to chemical reactions with diffusion, neutron diffusion and heat diffusion. In our case we are interested in problems with composite or inhomogeneous media where the physical properties have discontinuities or nonlinearities.

Examples of nonlinearities in heat conduction include (i) heat loss due to radiation and which is proportional to $u-u_0$ and (for large $u-u_0$) to $u^4-u_0^4$ and (ii) heat loss due to heat convection to the adjacent fluid and is proportional (in some cases) to $(u-u_0)^{5/4}$. Also these losses may occur at just the boundaries or at discrete points in the media. In

* Received by the editors August 3, 1976, and in final revised form November 8, 1978.

† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650.

the latter cases they are proportional to $\delta(x-x_0)(u(x)-u_0(x))$, $\delta(x-x_0)(u(x)^4-u_0^4(x))$ and $\delta(x-x_0)(u(x)-u_0(x))^{5/4}$, respectively. The equations with nonlinear boundary conditions and inhomogeneous term with smooth $(C^{2+\alpha})$ coefficients have been considered in H. B. Keller [5]. The solutions are constructed via a monotone sequence of solutions of the linear problems. We shall also use monotone sequence to construct a solution of these equations with "badly" behaved coefficients. Nonlinearities may also occur in the conductivity term, $p(x, u)$. If the conductivity is independent of x, then the problem may be dealt with by a change of dependent variable given by Kirchoff's transform $V \equiv \int_0^u p(u')\, du'$. If p depends on both x and u, then other methods must be used such as described in R. E. White [15].

We shall consider the linear initial value, the linear boundary value and the nonlinear inhomogeneous boundary value problems. Section two contains the preliminaries. This includes the notion of average value of a locally integrable function and some spaces of functions or distributions associated with this notion of value. The average value of the derivative at the boundary is needed since any solution that is in $H_1(a, b)$ may not have continuity at a or b of $u'(x)$. Section three is a study of the initial value problem via the Volterra integral equation. To a certain degree this follows the classical arguments. In § 4 we study uniqueness of the boundary value problem with boundary conditions $B_1u=\alpha_1u(a)+\beta_1u'(a)=b_1$, $B_2u=\alpha_2u(b)+\beta_1u'(b)=b_2$, $\alpha_1, \alpha_2>0$ and $\beta_1\leqq 0\leqq\beta_2$. We do this by using the generalization given in R. E. White [14] of the classical strong maximum principle. Because derivatives appear in the boundary condition, it is necessary to use the strong as opposed to the weak maximum principle. In § 5 we establish existence to the differential equation with the above boundary conditions. The classical variation of parameters formula still holds. Finally in § 6 we establish the existence to the semilinear boundary value problem where $f=f(x, u)$, $B_1u=b_1(u)$ and $B_2u=b_2(u)$. We construct the solution by monotone methods and use the generalized strong maximum principle to establish the monotonicity of the corresponding integral operator.

These results generalize many of the classical theorems in ordinary differential equations. In general see P. Hartman [3] and the references mentioned in the sections of this paper. The proofs given in this paper essentially follow the classical proofs. The differences arise when we must deal with average value of the derivative and when q and f are distributional derivatives of $L_2(a, b)$ functions. In our proofs we have tried to stress these differences.

2. Preliminaries. In the initial conditions and in the boundary conditions we need the value of an integrable function to exist in some sense. We define the average, right average and left average values at a point for integrable functions. The average value of integrable functions coincides with symmetric value of certain distributions as defined in P. Antosik [1]. In order to give a characterization of symmetric value, we need to define a δ-sequence: $\delta_n \in C_c^\infty(a, b)$ is a δ-sequence about $x \in (a, b)$ if and only if (i) support $\delta_n \subset (x-1/n, x+1/n)$, (ii) $\delta_n(x) \geqq 0$ and (iii) $\int_a^b \delta_n(y)\, dy = 1$. $\{\delta_n\}$ is even about x if and only if each δ_n is even about x. The symmetric value at x of a distribution $u \in \mathscr{D}'(a, b)$ may be defined to be the unique limit of $u(\delta_n)$ for all even δ-sequence $\{\delta_n\}$. The right and left values of an integrable function coincides with the right and left value of distribution as defined by S. Lojasiewicz [7]. The left value at x for example, may be defined as the unique limit of $u(\delta_n)$ for all δ-sequences about x with support $\delta_n \subset (x-1/n, x)$. Finally these values may be defined in terms of value sets for distributions as defined in R. E. White [14]. Let $u \in \mathscr{D}'(a, b)$ and $x \in (a, b)$. The value set of u at x is defined by $\mathscr{D}^0(u, x) \equiv \bigcap_{\varepsilon>0} \overline{\{u(\phi) \mid \phi \in C_c^\infty(a, b),\ \text{support } \phi \subset (x-\varepsilon, x+\varepsilon),\ \phi(x) \geqq 0, \int_a^b \phi(y)\, dy = 1\}}$ where the closure is in the two point compactification of the real line, $[-\infty, \infty]$. A distribution is defined to have value at x if $\mathscr{D}^0(u, x)$ is a singleton. Examples include $\mathscr{D}^0(H, 0)=[0, 1]$, $\mathscr{D}^0(\delta(x-x_0), x_0)=[0, \infty]$ and $\mathscr{D}^0(f(x), x)=\{f(x)\}$ when $f(x)$ is continuous at x. The notion of value set of elements in $H_{-1}[a, b]$ will be used later in this paper.

By demanding that square integrable functions have right average value at $x=a$ and left average value at $x=b$ and introducing an appropriate norm we obtain a complete subspace, $L_2[a, b] \subsetneq L_2(a, b)$, of the square integrable functions. Thus one can use the Neumann series to obtain solution to a Volterra integral equation which eventually yields a solution to the initial value problem. This solution and its derivative will have right average value at $x=a$ and left average value at $x=b$.

DEFINITIONS. Let $u \in L_1(a, b)$, $(S_n u)(x) \equiv n/2 \int_{x-1/n}^{x+1/n} u(y)\,dy$, $(L_n u)(x) \equiv n \int_{x-1/n}^{x} u \cdot (y)\,dy$ and $(R_n u)(x) \equiv n \int_x^{x+1/n} u(y)\,dy$. If the limits exists, then u has the following values:

Average value of u at $x = u(x) = \lim_{n\to\infty} S_n u(x)$.

Left average value of u at $x = u(x-) = \lim_{n\to\infty} L_n u(x)$.

Right average value of u at $x = u(x+) = \lim_{n\to\infty} R_n u(x)$.

Examples. 1. Let $u = H(x) =$ the Heavyside function. Then clearly $H(0)=\frac{1}{2}$.

2. Let u be given by the graph shown in Fig. 1. Then $u(0+) = \lim_{n\to\infty} n \int_0^{1/n} u(y)\,dy = \lim_{n\to\infty} n(\frac{1}{2})1/n = \frac{1}{2}$.

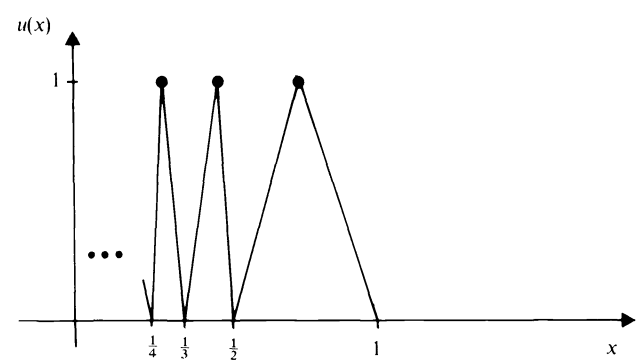

FIG. 1

3. If $\lim_{y\to x} u(y)$ exist, then the value of u exist at x and is this limit.

The first theorem states some of the routine properties of these values. The proofs are nearly the same as those given in [1] and [2]. For the purpose of completeness, we sketch the proofs. Recall that a distributional derivative of $u \in L_1(a, b)$ or $\varepsilon\mathscr{D}'(a, b)$, u', is a map from $C_c^\infty(a, b) \to R$ given by $u'(\phi) \equiv u(-\phi') = \int_a^b u(x)(-\phi'(x))\,dx$ where $\phi \in C_c^\infty(a, b)$.

THEOREM 1. *Let $u, v \in L_1(a, b)$. Then we have the following*:

1. *The average value of u at $x \in (a, b)$ is equivalent to the symmetric value* (*see* [1]) *of u at x.*
2. *The average left and right values at b and a are equivalent to the left and right values* (*see* [2]) *at b and a.*
3. *If the value at $x \in (a, b)$ exist for u and v, then $\alpha u + \beta v$ has value at x and $(\alpha u + \beta v)(x) = \alpha u(x) + \beta v(x)$ for $\alpha, \beta \in \mathbb{R}$. Also this is valid for right and left values.*
4. *If u is continuous at $x \in (a, b)$ and the value of v exist at x, then uv has value at x and $(uv)(x) = u(x) \cdot v(x)$. This also holds for right and left values.*

5. *If u, u', v, $v' \in L_2(a, b)$ and all have values at x, then $(uv)'$ has value at x and $(uv)'(x) = u'(x)v(x) + u(x)v'(x)$. This also holds for right and left values.*
6. *If $u' = 0$, i.e. $u'(\phi) = 0$ for all $\phi \in \mathscr{D}(\Omega)$, and u has zero value (or right or left values) at $x_0 \in (a, b)$, then $u = 0$.*

Sketch of proofs. 1. One should use the characterization given in [1] of symmetric value as the unique limit of $\int_a^b u(y)\delta_n(x-y)\,dy$ where $\{\delta_n\}$ is any even delta sequence.

2. In a similar manner, one should use the characterization given in [2].

3. This is proved by the direct application of the definitions.

4. If we write $u(y)v(y) = (u(y) - u(x))v(y) + u(x)v(y)$ and integrate according to the definition, then the desired result follows from the continuity of u at x.

5. The derivatives are distributional derivatives and it is not difficult to show that $(uv)' = u'v + uv'$. By part 4, $(u'v)(x) = u'(x)v(x)$ and $(uv')(x) = u(x)v'(x)$. By part 3, $(uv)'(x)$ exists and $(uv)'(x) = u'(x)v(x) + u(x)v'(x)$.

6. Since u is a distribution and $u' = 0$, $u = \text{constant}$, i.e. $u(\phi) = \int_a^b u(y)\phi(y)\,dy = \text{constant}.\ \int_a^b \phi(y)\,dy$ for all $\phi \in \mathscr{D}(\Omega)$. Thus $u(x) = \text{constant}$ for all $x \in [a, b]$ and so $\text{constant} = u(x_0) = 0$.

We shall use the following spaces. Note that the square brackets indicate that certain distributions have average values at a and b.

DEFINITIONS.

$$L_2[a, b] = \{u \in L_2(a, b) \mid u(a+), u(b-) \text{ exist}\} \quad \text{with norm}$$

$$u \mapsto \|u\|_{L_2[a,b]} = \|u\|_{L_2(a,b)} + \sup_m |(L_m u)(b)| + \sup_m |R_m u(a)|,$$

$$H_1[a, b] = \{u \in L_2[a, b] \mid u' \in L_2[a, b]\} \quad \text{with norm}$$

$$u \mapsto \|u\|_{L_2[a,b]} + \|u'\|_{L^2[a,b]},$$

$$H_{-1}[a, b] = \{v \in \mathscr{D}'(a, b) \mid v = u',\ u \in L_2[a, b]\} \quad \text{with norm}$$

$$v \mapsto \|u\|_{L_2[a,b]},\ u' = v,\ u(a+) = 0.$$

THEOREM 2. *$L_2[a, b]$, $H_1[a, b]$ and $H_{-1}[a, b]$ are complete.*

Proof. Consider $u_k \in L_2[a, b]$ and let u_k be Cauchy in $L_2[a, b]$. Then u_k is Cauchy in $L_2(a, b)$ and so must converge to $u \in L_2(a, b)$. We must show that u has right and left values at a and b. Consider the right value at a. Since $\sup_m |R_m u_k(a)|$ is Cauchy, $|R_m u_k(a) - R_m u_l(a)| \leqq \sup_m |R_m(u_k - u_l)(a)| < \varepsilon$ for $k, l \geqq N$. By taking limits as $m \to \infty$, we obtain $|u_k(a+) - u_l(a+)| \leqq \varepsilon$ for $k, l \geqq N$. Thus $u_k(a+)$ is Cauchy in $\mathbb{R}$ and so must converge to a real number, d. Also if one takes the limit as $l \to \infty$, we obtain $|R_m u_k(a) - R_m u(a)| \leqq \varepsilon$ for $k \geqq N$ and all m. Now we show that $d = \lim_{m\to\infty} R_m u(a) = u(a+)$.

$$\begin{aligned} |d - R_m u(a)| &\leqq |d - u_k(a+)| + |u_k(a+) - R_m u_k(a)| + |R_m u_k(a) - R_m u(a)| \\ &\leqq \varepsilon + |u_k(a+) - R_m u_k(a)| + \varepsilon, \qquad k \leqq N \text{ and all } m. \end{aligned}$$

Fix $k \geqq N$ and note that $R_m u_k(a) \to u_k(a+)$ as $m \to \infty$. Thus there is an M such that $|d - R_m u(a)| \leqq 3\varepsilon$ when $m \geqq M$. This shows that $u(a+)$ exists and is the limit of $u_k(a+)$. A similar argument yields that $u_k(b-) \to u(b-)$ as $k \to \infty$. Consequently $L_2[a, b]$ is complete.

In order to show that $H_1[a, b]$ is complete, we will use the fact that the Sobolev space $H_1(a, b)$ with norm $u \to \|u\|_{L_2(a,b)} + \|u'\|_{L_2(a,b)}$ is complete. If u_k is Cauchy in $H_1[a, b]$, it is Cauchy in $H_1(a, b)$ and hence must converge to u in $H_1(a, b)$. Now applying the arguments given in the proof that $L_2[a, b]$ is complete yields that $u(a+)$, $u'(a+)$, $u(b-)$ and $u'(b-)$ exist and are limits of $u_k(a+)$, $u'_k(a+)$, $u_k(b-)$ and $u'_k(b-)$.

Finally consider $H_{-1}[a, b]$ and let v_n be Cauchy in $H_{-1}[a, b]$. Let $u_n \in L_2[a, b]$ such that $u'_n = v_n$ and $u_n(a+) = 0$. Then u_n is Cauchy in $L_2[a, b]$ and so $u_n \to u$ in $L_2[a, b]$. $0 = u_n(a+) \to u(a+)$ and thus $u(a+) = 0$. Also since $u_n \to u$ in $L_2(a, b) u'_n \to u'$ in $\mathscr{D}'(a, b)$. Thus $v_n \to v \equiv u'$ in $H_{-1}[a, b]$ and consequently $H_{-1}[a, b]$ is complete.

Notation. If $v \in H_{-1}[a, b]$ and $u \in L_2[a, b]$ such that $u(a+) = 0$ and $u' = v$, then we will denote u by either $\int_a^x v$ or V or V_a. It is important to note that convergence in any of these spaces implies convergence of the respective values.

We will need the following maps from these spaces into one another.

Definitions. Let $L_\infty[a, b] = \{u \in L_\infty(a, b) \mid u$ is left continuous at b and u is right continuous at $a\}$ with norm $u \to \|u\|_{L_\infty(a,b)} + \sup_m |R_m u(a+)| + \sup_m |L_m u(b-)|$

1. $L_\infty[a, b] \times L_2[a, b] \to L_2[a, b]$ by $(f, u) \mapsto fu$ (pointwise multiplication).
2. $H_{-1}[a, b] \times H_1[a, b] \to H_{-1}[a, b]$ by $(q, u) \mapsto qu \equiv (Qu - \int_a^x Qu')'$.
3. Let $p, g \in L_\infty[a, b]$, $q \in H_{-1}[a, b]$, $u \in H_1[a, b]$
   $H_1[a, b] \to H_{-1}[a, b]$ by $u \mapsto Lu \equiv (pu')' + gu' + qu$.

THEOREM 3. *The above maps are well defined and continuous.*

Proofs. 1. Since $f \in L_\infty(a, b)$ and $u \in L_2(a, b)$, $fu \in L_2(a, b)$. The values of fu exist because of part 4 of Theorem 1. Since $(f, u) \to fu$ is continuous from $L_\infty(a, b) \times L_2(a, b) \to L_2(a, b)$, it suffices to consider whether or not the product of values is continuous. Let $f_n(a+) \to f(a+)$ and $u_n(a+) \to u(a+)$. Clearly $(f_n u_n)(a+) = f_n(a+) u_n(a+)$ and $(fu)(a+) = f(a+) u(a+)$. Since the product of numbers is continuous, the desired conclusion holds.

2. The definition of $qu \equiv (Qu - \int_a^x Qu')'$ is from the integration by parts formula. Since $u \in H_1[a, b]$, u is continuous on $[a, b]$ and thus Qu has values at a and b. Also $Qu \in L_2(a, b)$ and hence $Qu \in L_2[a, b]$. Since $Q \in L_\infty(a, b)$ and $u' \in L_2(a, b)$, $Qu' \in L_2(a, b)$ and so $\int_a^x Qu'$ is continuous. Thus $qu \in H_{-1}[a, b]$. Note that $(Qu)(a+) = Q(a+)u(a+) = 0$ and $\int_a^a Qu' = 0$ and so $\int_a^x qu = Qu - \int_a^x Qu'$. Let $q_n \to q$ in $H_{-1}[a, b]$ and $u_n \to u$ in $H_1[a, b]$. Thus $u_n \to u$ in $L_\infty[a, b]$ and $Q_n u_n \to Qu$ in $L_2[a, b]$ and so, by part one of this theorem, $Q_n u_n \to Qu$ in $L_2[a, b]$. Since $Q_n \to Q$ in $L_2(a, b)$ and $u'_n \to u'$ in $L_2(a, b)$, $Q_n u'_n \to Qu'$ in $L_1(a, b)$ and so $\int_a^x Q_n u'_n \to \int_a^x Qu'$ in $C[a, b]$ and hence in $L_2[a, b]$. Consequently $Q_n u_n - \int_a^x Q_n u'_n \to Qu - \int_a^x Qu'$ in $L_2[a, b]$. Thus $q_n u_n \to qu$ in $H_{-1}[a, b]$.

3. Since $pu' \in L_2[a, b]$, $(pu')' \in H_{-1}[a, b]$. Also $gu' \in L_2[a, b]$ which is a subset of $H_{-1}[a, b]$ and $qu \in H_{-1}$. Thus $qu \in H_{-1}[a, b]$. The continuity of L follows from the previous parts of this theorem.

3. Initial value problem. In this section we shall consider the initial value problem. In the classical case the initial value problem is solved via a Volterra integral equation in u. The existence of values at a and b follows in the classical case from the smoothness of the coefficients and inhomogeneous term. Both these steps fail in our case. We develop a Volterra integral equation in u' in place of u and the values are attached to the spaces we work with. Throughout the remainder of the paper, we assume $0 < m \leqq p \leqq M < \infty$, $p, g, Q \in L_\infty[a, b]$, $f, q \in H_{-1}[a, b]$ and $u \in H_1[a, b]$. This is a further restriction on q.

$$
\begin{aligned}
&Lu \equiv (pu')' + gu' + qu = f,\\
(1)\qquad &u(a+) = \text{given real number},\\
&u'(a+) = \text{given real number}.
\end{aligned}
$$

DEFINITION. *$u \in H_1[a, b]$ is a solution to* (1) if and only if in addition to satisfying the initial conditions, for all $\phi \in \mathcal{D}(a, b) = C_c^\infty(a, b)$

$$((pu')' + gu' + qu)(\phi) = f(\phi).$$

Remark. We may write $g(x)u'(x) = (\int_a^x g(y)u'(y)\,dy + g(a)u'(a))'$ and $qu = (\int_a^x qu)' = (Q(x)u(x) - \int_a^x Q(y)u'(y)\,dy)'$. Thus the differential equation becomes $(p(x)u'(x) + \int_a^x g(y)u'(y)\,dy + g(a)u'(a) + Q(x)u(x) - \int_a^x Q(y)u'(y)\,dy)'(\phi) = (F(x))'(\phi)$. Since the difference of derivatives of two distributions being zero is characterized by the distributions differing by a constant (see L. Schwartz [9]), u satisfying the differential equation is characterized by the equality

$$p(x)u'(x) + \int_a^x g(y)u'(y)\,dy + g(a+)u'(a+) + Q(x)u(x) - \int_a^x Q(y)u'(y)\,dy$$
$$= F(x) + \text{Constant}.$$

The solution for (1) will be found by solving a Volterra integral equation for u' and then integrating u' to solve for u. Note that we may write $u(x) = \int_a^x u'(y)\,dy + u(a+)$. It is necessary to consider an integral equation for u' in place of u because of the lack of smoothness of q and f. By integrating (1) and solving for $u'(x)$ we obtain the following integral equation.

$$\begin{aligned} &v - Tv = h \quad \text{where } v = u' \in L_2[a, b], \\ &Tv \equiv \int_a^x \frac{-g(y) + Q(y) - Q(x)}{p(x)} v(y)\,dy, \\ &h \equiv \frac{p(a+)u'(a+) + F(x) + Q(x)u(a+)}{p(x)}. \end{aligned} \tag{2}$$

Note that $h \in L_2[a, b]$ since $p^{-1} \in L_\infty[a, b]$ and $F, Q \in L_2[a, b]$. Also

$$K(x, y) = (-g(y) + Q(y) - Q(x))/p(x) \in L_2((a, b) \times (a, b)),$$

$$\lim_{\substack{(x,y)\to(a,a)\\ y \geqq x \geqq a}} K(x, y) = K(a, a) \quad \text{and} \quad \lim_{\substack{(x,y)\to(b,b)\\ b \geqq y \geqq x}} K(x, y) = K(b, b).$$

We now state a general theorem which includes the integral equation (2). Since we are working in $L_2[a, b]$ and not in $L_2(a, b)$, the proof of this theorem is a little more complicated than the theorem in $L_2(a, b)$. See F. Smithies [10] for the proof in $L_2(a, b)$.

THEOREM 4. *Let $K(x, y) \in L_2((a, b) \times (a, b))$, $\lim_{(x,y)\to(a,a), x \geqq y \geqq a} K(x, y) = K(a, a)$ and $\lim_{(x,y)\to(b,b), b \geqq x \geqq y} K(x, y) = K(b, b)$. If $T: L_2[a, b] \to L_2[a, b]$ is defined by $Tv = \int_a^x K(x, y)v(y)\,dy$, then the integral equation $v - Tv = h \in L_2[a, b]$ has one and only one solution $v = \sum_{n=0}^\infty T^n h \in L_2[a, b]$.*

Proof. Since $L_2[a, b]$ is complete, it suffices to show that the series $\sum_{n=0}^\infty T^n h$ converges absolutely in $L_2[a, b]$. Note that $Tv \in L_2[a, b]$ when $v \in L_2[a, b]$. Certainly $Tv \in L_2(a, b)$. The values of Tv exist because $\lim_{(x,y)\to(a,a), x \geqq y \geqq a} K(x, y) = K(a, a)$ and $\lim_{(x,y)\to(b,b), b \geqq x \geqq y} K(x, y) = K(b, b)$. In order to show absolute convergence in $L_2[a, b]$, we will need the following inequality which was proved in Smithies [10]: for $n = 1, 2, 3, \cdots$, $K^{n+1}(x, y) \leqq \|K\|^{n-1}_{L_2((a,b)\times(a,b))}((n-1)!)^{-1/2} k_1(x) \cdot k_2(y)$ where $k_1(x) = (\int_a^x K(x, y)^2\,dy)^{1/2}$, $k_2(y) = (\int_y^b K(x, y)^2\,dx)^{1/2}$ and $K^{n+1}(x, y)$ is the $n+1$th iterated kernal, i.e. $T^{n+1}v = \int_a^x K^{n+1}(x, y)v(y)\,dy$. Clearly, $k_1(x)$ has left and right values at b

and a. Thus

$$\|T^{n+1}h\|_{L_2[a,b]} \leqq \|K\|^{n-1}_{L_2((a,b)\times(a,b))}((n-1)!)^{-1/2}\left\{\left\|k_1(x)\int_a^x k_2(y)h(y)\,dy\right\|_{L_2(a,b)} + \sup_m \left|R_m\left(k_1(x)\int_a^x k_2(y)h(y)\,dy\right)(a+)\right| + \sup_m \left|L_m\left(k_1(x)\int_a^x k_2(y)h(y)\,dy\right)(b-)\right|\right\}.$$

Since $k_2(y), h(y) \in L_2(a, b)$, $k_2(y)h(y) \in L_1(a, b)$ and $\int_a^x k_2(y)h(y)\,dy$ is continuous. Thus $k_1(x)\int_a^x k_1(y)h(y)\,dy$ is in $L_2(a, b)$. Also since $k_1(x)$ have values at a and b, $k_1(x)\int_a^x k_2(y)h(y)\,dy$ has values at a and b. Thus the $\{\cdot\} \leqq M < \infty$ and consequently $\sum_{n=0}^{\infty} \|T^n h\|_{L_2(a,b)} \leqq \|h\|_{L_2[a,b]} + \|Th\|_{L_2[a,b]} + \sum_{n=1}^{\infty}((n-1)!)^{-1/2}\|K\|_{L_2((a,b)\times(a,b))}M$. Since $L_2[a, b]$ is complete, $\sum_{n=0}^{\infty} T^n h$ converges in $L_2[a, b]$ to v. Since T: $L_2[a, b] \to L_2[a, b]$ is continuous, $v - Tv = h$. Clearly v is unique.

THEOREM 5. *Let v be the solution of* (2). *Then $u(x) \equiv \int_a^x v(y)\,dy + u(a+)$ is the one and only solution of* (1).

Proof. Certainly $u(x)$ is the one and only one solution of $u(a+) = u(a+)$ and $u' = v$. Thus if suffices to show v satisfies the equation

$$p(x)v(x) + \int_a^x g(y)v(y)\,dy + g(a)u'(a) + Q(x)\left(\int_a^x v(y)\,dy + u(a+)\right) - \int_a^x Q(y)v(y)\,dy = F(x) + \text{constant}$$

and the initial condition $v(a+) = u'(a+)$. The initial condition holds because $v = Tv + h$ and by Theorem 1, $v(a) = T(v)(a+) + h(a+) = 0 + (p(a+)u'(a+) + 0)/(p(a+)) = u'(a+)$. Since $v - Tv = h$, we have $p(x)v(x) - \int_a^x(-g(y) + Q(y) - Q(x))v(y)\,dy = p(a+)u'(a+) + F(x) + Q(x)u(a+)$. Thus $p(x)v(x) + \int_a^x g(y)v(y)\,dy + Q(x)\int_a^x v(y)\,dy - \int_a^x Q(y)v(y)\,dy = F(x) + \text{constant}$.

Remarks. 1. In our considerations of the initial value problem, it is not necessary to demand that the solution or its derivative have value at b. We could replace $H_1[a, b]$ by $H_1[a, b)$ with the obvious norm.

2. Also it is not necessary to have $u \in H_1[a, b)$. We could have considered $H_1^{loc}[a, b)$ where u, u' are square integrable on (a, b − (1/n)) for all n and have values at a. With the proper choice of seminorms, this becomes a Fréchet space and one can proceed in the usual manner to solve the Volterra integral equation.

4. Uniqueness for the boundary value problem. This section contains a discussion of maximum principles and the question of uniqueness for the boundary value problem

$$\begin{aligned} &Lu = f \\ &B_1 u \equiv \alpha_1 u(a+) + \beta_1 u'(a+) = \text{given}, \qquad \alpha_1 > 0, \quad \beta_1 \leqq 0 \\ &B_2 u = \alpha_2 u(b-) + \beta_2 u'(b-) = \text{given}, \qquad \alpha_2 > 0, \quad \beta_2 \geqq 0. \end{aligned} \tag{3}$$

There are two versions of the maximum principle, the strong and the weak. The *strong maximum principle* states if $u \in H_1(a, b)$ is a weak solution to $Lu = f \in H_{-1}(a, b)$, $u \geqq 0$, $u \neq$ constant, $\mathscr{D}^0(q, x) \subset [-\infty, 0]$ and $\mathscr{D}^0(f, x) \subset [0, \infty]$ for all $x \in (a, b)$, then $u(x) < \sup_{y \in (a,b)} u(y)$ for all $x \in (a, b)$. This is proved in R. E. White [14] for the nonclassical case, i.e. $0 < m \leqq p \leqq M$, p, g, $Q \in L_\infty[a, b]$, and the proof for the classical case can be found in M. Protter and H. Weinberger [13]. The *weak maximum principle* states if $u \in H_1(a, b)$ is a weak solution to $Lu = f \in H_{-1}(a, b)$, $u \geqq 0$, $u \neq$ constant, $\mathcal{D}^0(q, x) \subset [-\infty, 0]$ and $\mathcal{D}^0(f, x) \subset [0, \infty]$ for all $x \in (a, b)$, then $\sup_{y \in (a,b)} u(y) = u(a)$ or $u(b)$. The weak maximum principle was proved in G. Stampacchia [12] for elliptic operator of more than one variable with badly behaved coefficients or inhomogeneous term. It is shown in R. E. White [14] that these conditions on the coefficients cannot be relaxed and still have either the strong or weak maximum principles.

In G. Stampacchia [12] the weak maximum principle was used to establish uniqueness for the Dirichlet problem and existence for semilinear problems. The weak maximum principle can also be used to solve semilinear problems via monotone sequences of iterates for solutions to linear problems. For example see Hendrik J. Kuiper [6]. It is our intention to develop similar arguments for the Robin-type boundary conditions as stated above for the one dimensional case. However, in this case we need the strong maximum principle as well as the notion of average value for the derivatives at the boundary. In order to do this we shall make use of Theorem 5 in [14] which states if $u \in H_1(a, b)$ is a weak solution to $Lu = f \in H_{-1}(a, b)$, $\mathcal{D}^0(q, x) \subset [-\infty, 0]$ and $\mathcal{D}^0(f, x) \subset [m(x), \infty]$ when $m(x) > 0$ for all $x \in (a, b)$, then $u \neq$ constant cannot have a nonnegative maximum in the interior of $[a, b]$.

For the remainder of the paper we shall assume $\mathcal{D}^1(p, x) \subset [-K, \infty] \forall x \in (a, b)$, $K < \infty$.

THEOREM 6. *Let $u \in H_1[a, b]$, u is not constant and satisfies $Lu = f$ where $\mathcal{D}^0(q, x) \subset [-\infty, 0]$, $\mathcal{D}^0(f, x) \subset [0, \infty]$ for all $x \in (a, b)$. If u attains a nonnegative maximum at a, then $u'(a+) < 0$. Also if u attains a nonnegative maximum at b, then $u'(b-) > 0$.*

Proof. Since $\mathcal{D}^0(q, x) \subset [-\infty, 0]$ and $u \geqq 0$, it suffices to prove the theorem for $q \equiv 0$. Let $d \in (a, b)$ such that $u(d) < M = u(a)$. Because $Q \in L_\infty[a, b]$ we may choose, as in the proof of Theorem 6 in [14], $\alpha > 0$ of $z(x) \equiv e^{\alpha(x-a)} - 1$ so that $\mathcal{D}^0(Lz, x) \subset [m(x), \infty]$ for all $x \in (a, d)$ and $m(x) > 0$. Let $0 < \gamma < (M - u(d))/(z(d))$. Now $\mathcal{D}^0(L(u + \gamma z), x) = \mathcal{D}^0(Lu + \gamma Lz, x) \subset \mathcal{D}^0(Lu, x) + \gamma \mathcal{D}^0(Lz, x) \subset [m(x), \infty]$ for all $x \in (a, d)$. Thus by Theorem 5 in [14] the nonnegative maximum of $u + \gamma z$ is attained at either a or d. But $(u + \gamma z)(d) = u(d) + \gamma z(d) < M = u(a)$. Thus $u + \gamma z$ must be nonincreasing on $[a, d]$. By Theorem 2 in [14] $\mathcal{D}^1(u + \gamma z, x) \subset [-\infty, 0]$ for all $x \in [a, d]$. By part 6 of Theorem 1 in [14] and the fact that $\mathcal{D}^1(\gamma z, x) = \{\gamma z'(x)\}$, we have $\mathcal{D}^1(u + \gamma z, x) = \mathcal{D}^1(u, x) + \gamma z(x)$. At $x = a$, $u'(a+)$ exists and so $u'(a+) + \gamma z'(a) \leqq 0$. Since $\gamma z'(a) > 0$, $u'(a+) < 0$.

A similar argument yields the desired conclusion of $u'(b-) > 0$.

THEOREM 7. *Let $u \in H_1[a, b]$ be at solution of* (3) *and $\mathcal{D}^0(q, x) \subset [-\infty, 0]$. If $\mathcal{D}^0(f, x) \subset [0, \infty] (\subset [-\infty, 0])$ for all $x \in (a, b)$, $B_1 u \leqq 0 (\geqq 0)$ and $B_2 u \leqq 0 (\geqq 0)$, then $u(x) \leqq 0 (\geqq 0)$ for $x \in (a, b)$.*

Proof. Because $B_1 u \leqq 0$ or $B_2 u \leqq 0$, u cannot be a positive constant. By Theorem 6 in [14], the strong maximum principle, u can only assume a nonnegative maximum at a or b. If it is at a, $B_1 u \leqq 0$ implies $\alpha_1 u(a+) \leqq -\beta_1 u'(a+) < 0$ which implies that $u(a+) < 0$. A similar argument gives that $u(b-) < 0$ if the maximum is at b. Thus $u(x) \leqq 0$ for all $x \in (a, b)$.

COROLLARY 1. *If $\mathcal{D}^0(q, x) \subset [-\infty, 0]$ and* (3) *has a solution in $H_1[a, b]$, then it is unique.*

Proof. Let u, v be the two solutions. Then $L(u - v) = 0$, $B_1(u - v) = 0$, and $B_2(u - v) = 0$. Apply the theorem to conclude that $0 \leqq u - v \leqq 0$.

COROLLARY 2. *Let $u \in H_1[a, b]$ be a solution of* (3) *and $\mathcal{D}^0(q, x) \subset [-\infty, 0]$. If $\mathcal{D}^0(f - qM, x) \subset [-\infty, 0]$, $\mathcal{D}^0(f - qN, x) \subset [0, \infty]$, $\alpha_1 M \leqq B_1 u \leqq \alpha_1 N$ and $\alpha_2 M \leqq B_2 u \leqq \alpha_2 N$, then $M \leqq u \leqq N$.*

Proof. In order to show $u \leqq N$ consider $L(u-N)=f-qN$, $B_1(u-N)=B_1u-\alpha_1 N$ and $B_2(u-N)=B_2u-\alpha_2 N$. Apply the theorem to conclude that $u \leqq N$. In order to show $M \leqq u$ consider the boundary value problem for $-u+M$.

5. Existence for the boundary value problem. In this section we show that the classical variation of parameters formula extends to this more general problem.

First we shall consider (3) with homogeneous boundary conditions $B_1u=0=B_2u$. By Theorem 5 we may choose $u, v \in H_1[a,b]$ to be solutions of the initial value problems $Lu=0$, $u(a+)=\beta_1$, $u'(a+)=-\alpha_1$ and $Lv=0$, $v(b-)=\beta_2$, $v'(b-)=-\alpha_2$. In these cases $pu'=-\int_a^x gu'-Qu+\int_a^x Qu'+\text{constant}$ and $pv'=-\int_a^x gv'-Qv+\int_a^x Qv'+\text{constant}$. Thus by part 4 of Theorem 1 $pW \equiv p(uv'-vu')=-u\int_a^x gv'+v\int_a^x gu'+u\int_a^x Qv'-v\int_a^x Qu'+\text{constant}$ is an element of $H_1[a,b]$. Also the Lagrange identity holds:

$$\begin{aligned}
0 = uLv - vLu &= (u(pv')'+guv'+quv)-(v(pu')'+gu'v+quv)\\
&= \left(upv'-\int_a^x pv'u'\right)'-\left(vpu'-\int_a^x pu'v'\right)'+g(uv'-u'v)\\
&= (p(uv'-u'v))'+\frac{g}{p}p(uv'-u'v)\\
&= (pW)'+\frac{g}{p}(pW).
\end{aligned}$$

Thus, $pW=C\exp -\int_a^x (g/p)$. If $W=0$, then $B_1v=0=B_2v$ and $B_1u=0=B_2u$. By Corollary 1 of Theorem 7, $u=v=0$ which is a contradiction. Thus $C\neq 0$ and so $pW\neq 0$.

The classical variation of parameters or Green's function may now be defined. Note that the integrals below are defined because the integrands are a product of a distribution in $H_{-1}[a,b]$ and a function in $H_1[a,b]$.

DEFINITION. Let $f\in H_{-1}[a,b]$ and define $G: H_{-1}[a,b]\to H_1[a,b]$ by

$$Gf \equiv v(x)\int_a^x \frac{u}{pW}f-u(x)\int_b^x \frac{v}{pW}f.$$

THEOREM 8. *Gf is a solution to the boundary value problem $Lu=f\in H_{-1}[a,b]$, $B_1u=0=B_2u$. Moreover G is a continuous linear operator.*

Proof. Note $u/pW=uC\exp\int_a^x (g/p)$ and $(u/pW)'=u'C\exp\int_a^x (g/p)+uC\exp\int_a^x (g/p)\cdot g/p$.

$$\begin{aligned}
(Gf)' &= \left(v\int_a^x \frac{u}{pW}f\right)'-\left(u\int_b^x \frac{v}{pW}f\right)'\\
&= \left(-v\int_a^x\left(\frac{u}{pW}\right)'F_a\right)'+u\int_b^x\left(\frac{V}{pW}\right)'F_b, \qquad F_a=-F_b+F_a(b)\\
&= -v'\int_a^x\left(\frac{u}{pW}\right)'F_a+u'\int_b^x\left(\frac{v}{pW}\right)'F_b+\frac{F_a(b)}{p}.
\end{aligned}$$

Thus $Gf\in H_{+1}[a,b]$ and G is continuous. It is easy to see that $p(Gf)'+\int_a^x g(Gf)'+Q(Gf)-\int_a^x Q(Gf)'=(-\int_a^x (u/pW)'F_a)\cdot Lv+(\int_b^x (v/pW)'F_b)\cdot Lu+F_a$. Since $Lv=0=Lu$, the differential equation is satisfied. $B_1Gf=\alpha_1(Gf)(a+)+\beta_1(Gf)'(a+)=\alpha_1(+u(a+)\int_b^a (v/pW)'F_b)+\beta_1u'(a+)\int_b^a (v/pW)'F_b=0$. $B_2Gf=\alpha_2(Gf)(b-)+\beta_2(Gf)'(b-)=\alpha_2(-v(b-)\int_a^b (u/pW)'F_a)+\beta_2(-v'(b-)\int_a^b (u/pW)'F_a)=0$.

COROLLARY. *Equation* (3) *has a solution.*

Proof. Let $B_1u = b_1$, $B_2u = b_2$. Clearly one can choose c, d such that $u + cx + d = w$ satisfies $B_1w = 0 = B_2w$. Then $Lw = Lu + gc + (cx + d)q \in H_{-1}[a, b]$. Thus $u = w - (cx + d) = G(f + gc + (cx + d)q) - (cx + d)$ is the solution of BV.

6. A nonlinear boundary value problem. In this section we will examine the following nonlinear boundary value problem

$$(4) \qquad Lu = f(x, u), \quad B_1u = b_1(u) \in \mathbb{R}, \quad B_2u \in b_2(u) \in \mathbb{R} \quad \text{and} \quad u \in H_1[a, b].$$

This is the one variable version of the classical nonlinear problem discussed in H. B. Keller [5] and of the weak nonlinear equation with Dirichlet boundary conditions. In our case the Robin-type boundary conditions with "badly" behaved coefficients and more general $f(x, u)$ are consider. In particular, $f(x, u)$ is an operator from a subset of $H_1[a, b]$ into $H_{-1}[a, b]$. In some sense, in Theorem 9 $f(x, u)$ is nonincreasing in u. In certain examples of $f(x, u)$, the chord method may be used to relax this restriction.

DEFINITION. Let $K \equiv \{u \in H_1[a, b] \mid M \leqq u \leqq N\}$. $f: (a, b) \times K \to H_{-1}[a, b]$ *is of type* D if and only if the following hold:

(i) When $u(x) \leqq v(x)$ for all $x \in [a, b]$, then $\mathscr{D}^0(f(x, u) - f(x, v), x) \subset [0, \infty]$.

(ii) $u_n \uparrow u$ uniformly implies $f(x, u_n) \to f(x, u)$ in $H_{-1}[a, b]$.

(iii) There exists $C > 0$ such that $\sup_{u \in K} \|f(x, u)\|_{H_{-1}[a,b]} \leqq C < \infty$.

Examples. 1. Let $f(x, u) = -\int_a^x u(y)\, dy$. The three conditions are easy to verify.

2. Let $f(x, u) = k(x, u(x))$ where $k \in L_1((a, b) \times (M, N))$, k is for almost all $x \in (a, b)$ continuous in u, k is for all $u \in (M, N)$ measurable in x and k is for almost all $x \in (a, b)$ nonincreasing in u. Since $k(x, u(x)) = (\int_a^x k(y, u(y))\, dy)'$ and $\int_a^x k(y, u(y))\, dy$ is continuous, $f(x, u) \in H_{-1}[a, b]$. Condition (i) is easy to verify. Condition (ii) follows from the monotone convergence theorem. The third condition is a consequence of

$$\int_a^x k(y, u(y))\, dy \leqq \|k\|_{L_1((a,b)\times(M,N))}$$

and

$$\|f(x, u)\|_{H_{-1}[a,b]} = \left\| \int_a^x k(y, u(y))\, dy \right\|_{L_2[a,b]}$$
$$\leqq [(b - a) + c_1 + c_2] \|k\|_{L_1((a,b)\times(M,N))}$$

where C_1 and C_2 are a result of $\sup |L_m|$, $\sup |R_m|$ terms in the $L_2[a, b]$ norm.

3. Let $f(x, u) = rh(x, u)$ where $r \in H_{-1}[a, b]$,

$$R(x) \equiv \int_a^x r \equiv \sum_{i=1}^k \Delta R_i H(x - x_i) + \bar{R}(x) \quad \text{with}$$

$$\bar{R}(x) \equiv \begin{cases} R(x), & x \neq x_i = \text{points where } R(x) \text{ is discontinuous}, \\ 0, & x = x_i, \end{cases}$$

$$\Delta R_i = R(x_i+) - R(x_i-) > 0,$$

$R(x)$ is nondecreasing and bounded.

$0 \leqq h(x, u) \in C([a, b] \times [M, N])$ and is nondecreasing in u. Thus

$$f(x, u) = \int_0^x rh(x, u(x))\, dx = \sum_{i=1}^k \int_a^x \delta(x - x_i) h(x, u(x))\, dx + \int_a^x \bar{R}'(\bar{x}) h(\bar{x}, u(\bar{x}))\, d\bar{x}$$
$$= \sum_{i=1}^{i(x)} h(x_i, u(x_i)) + \int_a^x \bar{R}'(\bar{x}) h(\bar{x}, u(\bar{x}))\, d\bar{x},$$

$i(x) \equiv$ largest integer such that $x_i \leqq x$.

Since R is nondecreasing, $\bar{R}'$ exist and is integrable. The three conditions of the above are easily verified.

Remark. Let $\tilde{K} = \{u \in C([a, b]) \mid u_n \in K, u_n \uparrow u \text{ in } L_2[a, b]\}$. Then we may extend $f: (a, b) \times K \to H_1[a, b]$ to $\tilde{f}: (a, b) \times \tilde{K} \to H_1[a, b]$ so that the same conditions hold for the extension $\tilde{f}$. Note that in our examples the extension is independent of the choice of $\{u_n\} \subset K$. We shall make no notation to distinguish between f, K and $\tilde{f}$, $\tilde{K}$.

ASSUMPTIONS.

A1. *Let* $f: (a, b) \times K \to H_{-1}[a, b]$ *be of type* D.

A2. *For all* $u \in K$ *suppose* $\mathscr{D}^0(f(x, u) - qM, x) \subset [-\infty, 0]$, $\mathscr{D}^0(f(x, u) - qN, x) \subset [0, \infty]$ *and* $\mathscr{D}^0(q, x) \subset [-\infty, 0]$ *for all* $x \in (a, b)$.

A3. *Let* $b_1, b_2 \in C(M, N)$ *be such that they are nondecreasing and* $\alpha_1 M \leqq b_1(u) \leqq \alpha_1 N$ *and* $\alpha_2 M \leqq b_2(u) \leqq \alpha_2 N$.

A4. *There exists* $u_0 \in K$ *such that* $Tu_0 \equiv G(f(x, u_0) + gc + (cx + d)q) - (cx + d) \geqq u_0$ *where c and d are as in the corollary to Theorem* 8.

THEOREM 9. *If assumptions* A1–A4 *hold, then the following hold*:

1. $u_n \in K$, $n = 0, 1, 2, \cdots$ *and* $u_{n+1} \equiv Tu_n$, $n = 0, 1, 2 \cdots$.
2. $u_{n+1} \geqq u_n$.
3. $u_n \to u \in K$ *in* $H_1[a, b]$.
4. *u is a solution to* (4).

Proofs. 1. By Corollary 2 of Theorem 7 any solution will satisfy $M \leqq u_n(x) \leqq N$. The Corollary to Theorem 8 yields that $B_1 u_{n+1} = b_1 u_n$ and $B_2 u_{n+1} = b_2 u_n$.

2. This proof is by mathematical induction. A4 states that $u_1 = Tu_0 \geqq u_0$. Assume $u_n \geqq u_{n-1}$. Then $L(u_{n+1} - u_n) = Lu_{n+1} - Lu_n = f(x, u_n(x)) - f(x, u_{n-1}(x))$. Since $f(x, u)$ of type D(i), $B_1(u_{n+1} - u_n) \geqq 0$ and $B_2(u_{n+1} - u_n) \geqq 0$, we have by Corollary 2 of Theorem 7 that $u_{n+1} - u_n \geqq 0$.

3. In order to prove this, we need to make use of the fact that $H_1(a, b)$ is continuously imbedded into $C^{0+1/2}[a, b]$, the Hölder continuous functions with exponent $\frac{1}{2}$. $C^{0+1/2}[a, b]$ is compactly imbedded into the continuous functions with sup norm. Since $f(x, u)$ is of type D(iii) $\|f(y, u_n(y))\|_{H_{-1}[a,b]} \leqq C < \infty$. Since $G: H_{-1}[a, b] \to H_1[a, b]$ is continuous, $\{Tu_{n-1} = u_n\}$ is a bounded sequence in $H_1[a, b]$. Thus there is a subsequence u_{n_i} that converges in $C[a, b]$. Since $u_{n+1} \geqq u_n$, u_n converges to $u \in C[a, b]$. Since $f(x, u)$ is of type D(ii), $f(x, u_n(x))$ converges to $f(x, u(x))$ in $H_{-1}[a, b]$. Because G is continuous, $u_{n+1} = Tu_n$ converges to Tu in $H_1[a, b]$. Since $u_{n+1} \to u$ in $C[a, b]$, $Tu = u \in K$.

4. We must show $LTu = f(x, u(x))$. since $Tu_n \to Tu$ in $H_1[a, b]$ and L is continuous, $LTu_n \to LTu$ in $H_{-1}[a, b]$. Now $LTu_n = f(x, u_n(x)) \to f(x, u(x))$ in $H_{-1}[a, b]$. Thus $LTu = f(x, u(x))$.

Suppose that $f(x, u)$ is "increasing" in u, i.e. condition D(i) does not hold. If $\tilde{f}(x, u) = \tilde{k}(x, u(x))$ of example 2 and $\tilde{k}$ has the property of $\mathscr{D}^0(\tilde{k}(x, u(x)) - \tilde{k}(x, v(x)) - R(u(x) - v(x)), x) \subset [-\infty, 0]$ for all $x \in (a, b)$, $R \in \mathbb{R}$ and $u \leqq v$, then we may define

$f(x, u) \equiv \tilde{f}(x, u) - Ru$. Thus if initially we were considering the differential equation $\tilde{L}u = (pu')' + gu' + \tilde{q}u = \tilde{f}(x, u)$, we could consider the equation $Lu = \tilde{L}u - Ru = \tilde{f}(x, u) - Ru = f(x, u)$. Now provided $f(x, u)$ satisfied the assumptions of Theorem 9, a solution $Lu = f(x, u)$ would also be a solution of $\tilde{L}u = \tilde{f}(x, u)$. A similar substitution may be made if $b_i'(u) \geqq -R$, $i = 1, 2$. This method is called the chord method and it may also be applied if $\tilde{f}(x, u) = r\tilde{h}(x, u)$ and $\tilde{h}$ also has the property of $\mathscr{D}^0(\tilde{h}(x, u(x)) - \tilde{h}(x, u(x)) - R(u(x) - v(x)), x) \subset [-\infty, 0]$ for all $x \in (a, b)$ and $u \leqq v$. In this case we define $f(x, u) = \tilde{f}(x, u) - rRu$. An example of this variation of the chord method is given at the end of this section.

One advantage of the monotone method is that it constructs the solution. Another is that one can often also construct the solution by a decreasing sequence of iterates. This results in bounding the solution between known iterates. In order to do this, we need a $v_0 \in K$ such that $v_1 \equiv Tv_0 \leqq v_0$. Then define $v_{n+1} \equiv Tv_n$, $n = 0, 1, 2, \cdots$. One can show that $v_n \downarrow v \in H_1[a, b]$ and v is a solution to the semilinear problem. If $u_0 \leqq v_0$, then by using the maximum principle one can show $u_n \leqq v_n$ and $n = 0, 1, 2, \cdots$ and hence $u \leqq v$. These results are summarized in Theorem 10.

THEOREM 10. *Let $A1, A2, A3$ and $A4$ hold. If there exists $v_0 \in K$ such that $Tv_0 \leqq v_0$ and $u_0 \leqq v_0$, then the following statements are true:*

1. *$u_n \uparrow u \in K$ and $v_n \downarrow v \in K$ in $H_1[a, b]$,*
2. *u, v are solutions to* (4),
3. *$u_n \leqq v_n$, $n = 0, 1, 2, \cdots$ and $u \leqq v$,*
4. *Suppose $U_0, V_0 \in K$, $U_{n+1} \equiv TV_n$, $V_{n+1} = TV_n$, $u_0 \leqq U_0 \leqq V_0 \leqq v_0$, $TU_0 \geqq U_0$ and $TV_0 \leqq V_0$, then $U_n \uparrow U \in K$, $V_n \downarrow V \in K$* in $H_1[a, b]$, *U, V are solutions to* (5) *and $u \leqq U \leqq V \leqq v$.*

Example. Consider $u''(x) = \delta(x - \frac{1}{2})(u^4(x) - 1^4)$ on $[-1, 1]$ with boundary conditions $u(-1) = 0$ and $u(1) = 0$. If we are interested in values of $u(x)$ between 0 and 1, we note that $u^4 - 1^4$ is increasing and therefore $\delta(x - \frac{1}{2})(u^4 - 1^4)$ is not of type D. Therefore we consider using a variation of the chord method. Thus consider $u'' - 4\delta(x - \frac{1}{2})u = \delta(x - \frac{1}{2})(u^4 - 4u - 1)$ with $u(-1) = 0$ and $u(1) = 0$. In this case $h(u) = u^4 - 4u - 1$ and is nonincreasing for $0 \leqq u(x) \leqq 1$. Thus $\mathscr{D}^0(\delta(x - \frac{1}{2})(h(u(x)) - h(v(x)), x) = \{0\}$ if $x \neq \frac{1}{2}$ and $= (h(u(\frac{1}{2}) - h(v(\frac{1}{2})) \cdot \mathscr{D}^0(\delta(x - \frac{1}{2}), \frac{1}{2}) = [-\infty, 0]$ when $u \leqq v$.

In order to solve the nonlinear problem, we must first construct the Green's function for $Lu = u'' - 4\delta(x - \frac{1}{2})u$. It is easy to see that $u = (1 - H(x - \frac{1}{2}))(-x - 1) + H(x - \frac{1}{2})(-7x + 2)$ is a solution of the initial value problem $Lu = 0$, $u(-1) = 0$ and $u'(-1) = -1$ and that $v \equiv (1 - H(x - \frac{1}{2}))(-3x + 2) + H(x - \frac{1}{2})(1 - x)$ is a solution of the initial value problem $Lv = 0$, $v(1) = 0$ and $v'(1) = -1$. From these u and v we may construct the Green's operator $Gf \equiv v \int_{-1}^{x} f(u/pW) - u \int_{1}^{x} f(v/pW)$ where $pW = 5$. Gf satisfies $LGf = f$, $(Gf)(-1) = 0$ and $(Gf)(1) = 0$.

Now we may consider $Tu \equiv Gf(x, u) = G(\delta(x - \frac{1}{2})h(u))$ defined on $K \equiv \{u \in H_1[a; b] | 0 \leqq u \leqq 1\}$. It is not difficult to show that $u_0 = 0$ and $v_0 = 1$ will work in Theorem 10. In fact, if $w(x) = \gamma$, then $Tw = (1 - H(x - \frac{1}{2}))(h(\gamma)/10)(x + 1) + H(x - \frac{1}{2})((h(\gamma)(-3))/10)(x - 1)$. If $\gamma \geqq .4$ and we choose $v_0(x) = \gamma$, then $Tv_0 \leqq v_0$. Thus any solution must be larger than $u_1(x) = (Tu_0)(x) = (To)(x) = (1 - H(x - \frac{1}{2})) \cdot \frac{1}{10}(x + 1) + H(x - \frac{1}{2})((-3/10)(x - 1)$ and less than $v_1(x) = (Tv_0)(x) = (T.4)(x) = (1 - H(x - \frac{1}{2}))2.5744/10(x + 1) + H(x - \frac{1}{2})2.5744/10(-3)(x - 1)$. Additional accuracy may be obtained by calculating $u_2 = Tu_1$ and $v_2 = Tv_1$. In this problem we are also able to determine the $H_1[a, b]$ solution for $0 \leqq u \leqq 1$ by examining the equation to the left and right of $x = \frac{1}{2}$. The solution is $u(x) = (1 - H(x - \frac{1}{2}))a(x + 1) + H(x - \frac{1}{2})(-3a)(x - 1)$ where a is the positive root of $a \frac{481}{16} + 4a - 1 = 0$.

REFERENCES

[1] P. ANTOSIK, *The symmetric value of a distribution at a point*, unpublished, 1974.

[2] P. ANTOSIK, J. MIKUSIŃSKI AND R. SIKORSKI, *Theory of Distributions, The Sequential Approach*, Elsevier Scientific Publishing Co., Amsterdam, PWN—Polish Scientific Publishers, Warsawa, 1973.

[3] P. HARTMAN, *Ordinary Differential Equations*, Wiley–Interscience, New York, 1964.

[4] D. D. JOSEPH AND E. M. SPARROW, *Nonlinear diffusion induced by nonlinear sources*, Quart. App. Math., 28 (1970), pp. 327–342.

[5] H. B. KELLER, *Elliptic boundary value problems suggested by nonlinear diffusion processes*, Arch. Rational Mech. Anal., 35 (1969), pp. 366–381.

[6] H. J. KUIPER, *Some nonlinear boundary value problems*, this Journal, 7 (1976), pp. 551–564.

[7] S. LOJASIEWICZ, *Sur la valeur et la limite d'une distribution d'un point*, Studia Math., 16 (1957), pp. 1–36.

[8] M. N. ÖZISIK, *Boundary Value Problems of Heat Conduction*, International Textbook Co., Scranton, 1968.

[9] L. SCHWARTZ, *Théorie des distributions*, Hermann, Paris, 1966.

[10] F. SMITHIES, *Integral Equations*, Cambridge University Press, London, 1958.

[11] I. STAKGOLD, *Boundary Value Problems of Mathematical Physics*, vol. I. Macmillan, New York, 1967–68.

[12] G. STAMPACCHIA, *Equations elliptiques de second ordre a coefficients discontinuous*, Les Presses de l'Université de Montréal, Montréal, 1966.

[13] H. WEINBERGER AND M. PROTTER, *Maximum Principles in Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1967.

[14] R. E. WHITE, *Value set at $x \in \Omega$ for an arbitrary distribution with applications to local extrema of $f \in C(\Omega)$ and a maximum principle for ordinary differential equations*, this Journal, to appear.

[15] ———, *Weak solution of $(r(x, u)u')' = (Fu)'$ with $r(0, u(0))u'(0) = ku(0)$, $r(L, u(L)u'(L) = hu(L)$ and h, k are suitable elements of* $[0, \infty]$, J. Math. Anal. Appl., to appear.

SIAM J. MATH. ANAL.
Vol. 10, No. 6, November 1979

0036-1410/79/1006-0020 $01.00/0

CORRIGENDUM AND ADDENDUM: MAXIMUM PRINCIPLES AND BOUNDS IN SOME INHOMOGENEOUS ELLIPTIC BOUNDARY PROBLEM*

RENÉ P. SPERB†

THEOREM 2 must read: *If u is a solution of* (1.1), (1.2) *where the geodesic curvature k_g of D is nonnegative, and g, h, ρ satisfy* (2.7) *then Φ takes its maximum at a critical point of u or at $\Gamma_1 \cap \Gamma_2$.*

Here $\Gamma_1 \cap \Gamma_2$ denotes a point on ∂D separating pieces of ∂D where $u = 0$ and those where $\partial u/\partial n = 0$. The reason for including the possibility that Φ takes its maximum at $\Gamma_1 \cap \Gamma_2$ is that $|\nabla u|$ is in general *not* continuous at $\Gamma_1 \cap \Gamma_2$. We can see this as follows. Take $\rho = 1$ for simplicity, i.e. u is a solution now of

$$\Delta u + 1 = 0 \quad \text{in } D, \qquad u = 0 \quad \text{on } \partial D. \tag{1}$$

Assume that $|\nabla u|$ is continuous at $\Gamma_1 \cap \Gamma_2$. Since on Γ_1 we have $\partial u/\partial s = 0$ and on Γ_2, $\partial u/\partial n = 0$ it follows that $|\nabla u| = 0$ at $\Gamma_1 \cap \Gamma_2$. Hence Φ must take its maximum at a critical point of u (inside D). Then, proceeding analogously as in the calculations following (3.16) (with P in the same sense as in the paper and Q a point of Γ_1) we find from equation (3.5) e.g. that for $\tau = \max_{\partial D} |\nabla u|$ we have

$$\tau^2 \leqq \frac{|D|^2}{2}, \qquad |D| = \text{diameter of } D. \tag{2}$$

Now on the other hand we have by Green's identity

$$A = -\oint_{\partial D} \frac{\partial u}{\partial n}\, ds \leqq \tau L_{\Gamma_1}, \tag{3}$$

i.e.

$$\tau \geqq \frac{A}{L_{\Gamma_1}}, \tag{4}$$

where A = area of D, L_{Γ_1} = length of Γ_1. Clearly, (2) and (4) contradict each other if L_{Γ_1} is small enough. Therefore, $|\nabla u|$ cannot be continuous at $\Gamma_1 \cap \Gamma_2$ for any $\Gamma_2 \neq \varnothing$.

In the applications on page 815 and later, $\Gamma_2 = \varnothing$ has to be assumed everywhere.

It seems in general a difficult task to describe the exact behavior of the solution in the neighborhood of $\Gamma_1 \cap \Gamma_2$ for an arbitrary domain D. Also bounds for $|\nabla u|$ and $u_{\max}$ seem to be rather hard to get if $\Gamma_2 \neq \varnothing$.

Acknowledgment. I would like to thank Catherine Bandle (University of Basle) for pointing out that the inequality (3.4) in the above paper can become wrong if Γ_1 is small, which led to the conclusion that $|\nabla u|$ may be discontinuous on ∂D.

* This Journal, 8 (1977), pp. 871–878. Received by the editors August 31, 1978, and in revised form September 18, 1978.

† Department of Pharmacology, Biocenter, University of Basle, Basle, Switzerland.