Comments on a paper by L. L. Lee [J. Math. Phys. 21, 1055 (1980)]

R. R. Whitehead
Department of Natural Philosophy, The University of Glasgow, Glasgow G12 8QQ, Scotland

(Received 18 July 1980; accepted for publication 30 December 1980)

Analytic expressions are given for the eigenvalues of some matrices arising in connection with Feynman path integrals.

PACS numbers: 02.10.Sp

In a recent paper[1] on the application of functional Laplace transforms to the evaluation of Feynman path integrals Lee had occasion to consider the familiar matrices

$$Q=\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \ddots \\ & -1 & 2 & \\ & & \ddots & \ddots \end{pmatrix},$$

$$R=\begin{pmatrix} 0 & 1 & & \\ -1 & 0 & 1 & \\ & -1 & 0 & \ddots \\ & & \ddots & \ddots \end{pmatrix},$$

and the less familiar matrix

$$P=Q^{-1}RQ^{-1}R.$$

The eigenvalues of these matrices are required in certain manipulations intended to lead to the known expressions for the propagators of simple systems like the harmonic oscillator. Unfortunately Lee obscured his derivations by not noticing that all these matrices have eigenvalues which can be written simply in closed form. (He evaluated them numerically and then showed that the correct numerical values for the propagators followed.)

The eigenvalues of Q are

$$\lambda_n = 2(1-\cos\theta),\quad \theta = n\pi/(N+1),\quad n=1,\ldots,N,$$

which is well known, and those of R are

$$\lambda_n = -2i\cos\theta,$$

where N is the order of the matrix. Both of these may be established simply by reducing the eigenvalue problem to the solution of a three-term recurrence relation, Q and R being tridiagonal.

Similarly for $P=Q^{-1}RQ^{-1}R=(Q^{-1}R)^2$, since the eigenvalue problem

$$Q^{-1}Ry=\mu y$$

can be written as

$$Ry=\mu Qy,$$

we get the recurrence relation

$$(1+\mu)y_{j+1}-2\mu y_j-(1-\mu)y_{j-1}=0.$$

This may be solved subject to the boundary conditions $y_0=y_{N+1}=0$, and we find that

$$\mu_n = i\cot\theta,\quad \theta=n\pi/(N+1),\quad n=1,\ldots,N.$$

The required eigenvalues of P are therefore the squares of these and they are obviously real and negative and come in pairs, as Lee found numerically. On the other hand, Lee's quoted numerical values are very bad approximations to these exact results, the larger eigenvalues being, surprisingly, the worst. It is possible therefore that P has awkward numerical properties.

ACKNOWLEDGMENTS

I wish to thank M. Storm and E. W. Laing for useful comments.

APPENDIX

At the suggestion of the referee I give here some further details of the derivation of the principal result of this paper.

The recurrence relation

$$(1+\mu)y_{j+1}-2\mu y_1-(1-\mu)y_{j-1}=0 \qquad \text{(A1)}$$

has constant coefficients and so is amenable to solution by means of the standard substitution[2]

$$y_j = q^j, \qquad \text{(A2)}$$

where q is some complex number. Inserting (A2) into (A1) gives a quadratic equation for q which may easily be solved giving

$$q=1 \text{ or } (\mu-1)/(\mu+1).$$

The general solution of the recurrence relation is thus

$$y_j = A+B\,[(\mu-1)/(\mu+1)]^j. \qquad \text{(A3)}$$

The recurrence relation corresponds to the eigenvalue problem only if $y_0=y_{N+1}=0$. Setting $j=0$ in (A3) gives immediately $B=-A$. Setting $j=N+1$ gives

$$[(\mu-1)/(\mu+1)]^{N+1}=1, \qquad \text{(A4)}$$

and the eigenvalues are the values of μ for which this is true.

The nontrivial solutions (that is, ones that do not lead to $y_j=0$ for all j) of (A4) are those in which $(\mu-1)/(\mu+1)$ is a complex number of modulus unity. We can thus write

$$(\mu-1)/(\mu+1)=\cos\phi+i\sin\phi,$$

and (A4) is satisfied if $\phi=2n\pi/(N+1)$, $n=1\cdots N$. The required eigenvalues are thus,

$$\mu_n=\frac{1+\cos\phi_n+i\sin\phi_n}{1-\cos\phi_n-i\sin\phi_n},\quad \phi_n=\frac{2n\pi}{N+1},\ n=1\cdots N,$$

which may be rewritten in the form given above. The corresponding eigenvectors are given, from (A3), by

$$y_j^{(n)}=A\,(1-\cos j\phi_n-i\sin j\phi_n).$$

[1] L. L. Lee, J. Math. Phys. **21**, 1055 (1980).
[2] I. Anderson, *A First Course in Combinatorial Mathematics* (Oxford U. P., 1974), p. 42.

 0022-2488/81/071329-01$1.00

Clifford algebras in 2ω dimensions

A. D. Kennedy
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 15 July 1980; accepted for publication 17 October 1980)

We consider the generalization of the Dirac (gamma) algebra to a space-time with dimension 2ω, which is required for the use of the dimensional regularization scheme applied to fermions. In particular we introduce several tricks to facilitate the manipulation of expressions in such algebras, including a polynomial algorithm for taking traces and explicit formulas for Fierz matrices.

PACS numbers: 02.10.Sp

I. INTRODUCTION

Dimensional regularization,[1,2] has become a popular tool for regularizing gauge theories in a manner consistent with their gauge symmetries. This technique requires one to deal with the generalization of the Dirac γ-algebra to $2\omega \equiv n \equiv 4 + \epsilon$ dimensions in order to include fermions into the scheme; such algebras are known as Clifford algebras. The purpose of this paper is to introduce new techniques to replace the efficient algorithms which have been developed specifically for four dimensions.[3,4]

In Sec. II we introduce our notation and discuss the representation of the 2ω-dimensional Clifford algebra on $2^\omega \times 2^\omega$ matrices. Section IIIA introduces and gives proofs for several theorems, mainly concerned with taking traces of products of γ's, culminating in Theorem 5 which provides a systematic method for taking traces.[1] Section IIIB develops a graphical notation which enables most simple traces to be done by inspection; the graphical technique is developed further in Sec. IV. In Sec. IIIC we give theorems which are principally concerned with simplifying expressions involving contracted indices. These lead, in Sec. IIID, to closed-form expressions for the Fierz reordering identities in spaces of arbitrary dimension. These identities are also given in terms of a convenient dual basis which generalizes the form of the usual four-dimensional bilinear spinor covariants. The numerical coefficient matrices are given in an appendix for $\omega = 1$ through 7. Finally in Sec. IV diagrammatic methods are used to find the Clebsch–Gordan series for the Clifford algebra, and a polynomial algorithm for reducing γ-expressions and traces is given.

A number of examples have been inserted into the appropriate sections to illustrate the use of the techniques as applied to some simple problems. It is hoped that these results may prove useful not only in using dimensional regularization but also in other areas, such as supersymmetry.

II. DEFINITIONS

The Clifford algebra in 2ω dimensions is defined by the anticommutation relations

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}\mathbb{1} \quad (1 \leqslant \mu, \nu \leqslant 2\omega),$$

where $\{\gamma_\mu\}$ are the generators of the algebra, $\mathbb{1}$ the unit element of the algebra, and $g_{\mu\nu}$ the $SO(\kappa, 2\omega - \kappa)$ metric tensor. It is usually sufficient to assume $\omega \in N$ and consider only even-dimensional spaces, though of course we "analytically continue" ω to arbitrary complex values in dimensional regularization.[1,2]

We may construct a basis in the algebra by defining

$$\Gamma^{(j)}_{\mu_1 \cdots \mu_j} \equiv \gamma_{[\mu_1} \cdots \gamma_{\mu_j]} = \frac{1}{j!} \delta^{\lambda_1 \cdots \lambda_j}_{\mu_1 \cdots \mu_j} \gamma_{\lambda_1} \cdots \gamma_{\lambda_j},$$

where $0 \leqslant j \leqslant 2\omega$, $1 \leqslant \mu_i \leqslant 2\omega$, $\delta^{\lambda_1 \cdots \lambda_j}_{\mu_1 \cdots \mu_j}$ is the generalized Kronecker delta (or permutation) tensor defined by

$$\delta^{\lambda_1 \cdots \lambda_j}_{\mu_1 \cdots \mu_j} \equiv j! \delta^{[\lambda_1}_{\mu_1} \delta^{\lambda_2}_{\mu_2} \cdots \delta^{\lambda_j]}_{\mu_j}$$

and the $[\cdots]$ indicates antisymmetrization on the enclosed indices as usual.

For each value of j there are $\binom{2\omega}{j}$ distinct values of the antisymmetric subscript $[\mu_1 \cdots \mu_j]$, so our basis has $\sum_{j=0}^{2\omega} \binom{2\omega}{j} = 2^{2\omega}$ members.

If we introduce a Cartan–Killing type of metric on the algebra by $\langle \Lambda, \Lambda' \rangle \equiv \mathrm{tr}[\Lambda \Lambda']$ using the cyclic linear trace functional (Λ, Λ' are arbitrary elements of the algebra) then, as we shall show, our basis is orthogonal. Furthermore, it is clearly complete as it is a maximal set of antisymmetric products of the γ's, and the symmetric part of any Λ may be reduced using the defining anticommutation relation.

For even-dimensional spaces ($\omega \in N$) the algebra is faithfully represented on the complete $2^\omega \times 2^\omega$ matrix algebra, e.g., by the map

$$\mathscr{D}^{(2\omega)} : \gamma_\mu \to \gamma_\mu{}^{(2\omega)},$$

where the matrix $\gamma_\mu{}^{(2\omega)}$ is defined as

$$\gamma_1^{(2)} \equiv \sigma_1, \quad \gamma_1^{(2\omega)} \equiv 1 \times \gamma_1^{(2\omega-2)},$$
$$\gamma_2^{(2)} \equiv \sigma_2, \quad \gamma_2^{(2\omega)} \equiv 1 \times \gamma_2^{(2\omega-2)},$$
$$\gamma_\alpha^{(2\omega)} \equiv \gamma_{\alpha-2}^{(2\omega-2)} \times \sigma_3,$$

($\omega > 1, 2 < \alpha \leqslant 2\omega$) with the σ_i being Pauli matrices. It is easily verified that these matrices satisfy the defining relation of the algebra with the $SO(2\omega)$ metric $g_{\mu\nu} = \delta_{\mu\nu}$; the generalization for $SO(\kappa, 2\omega - \kappa)$ is readily done by inserting the appropriate factors of i. $\mathscr{D}^{(2\omega)}$ is naturally extended to the complete algebra by linearity.

For odd-dimensional spaces the situation is less straightforward; for example the representation[5,6] given by

 0022-2488/81/071330-08$1.00

extending $\mathscr{D}^{(2\omega)}$ to $\mathscr{D}^{(2\omega+1)}$ with

$$\gamma_{2\omega+1} \to \gamma_1^{(2\omega)}\gamma_2^{(2\omega)}\cdots\gamma_{2\omega}^{(2\omega)}$$

is not faithful; also we find that the representation $\mathscr{D}'^{(2\omega+1)}$ $:\gamma_\mu \to -\gamma_\mu^{(2\omega)}$ is inequivalent to $\mathscr{D}^{(2\omega+1)}$.

The trace functional can be represented by the matrix trace, however we shall not use explicit representations in the following.

III. ALGEBRAIC IDENTITIES

A. Trace theorems

We now give the generalizations of some familiar theorems in the Dirac algebra to 2ω dimensions, and we introduce a notation particularly convenient for utilizing them. We shall assume 2ω to be even.

Theorem 1: The trace of an odd number of γ's is zero.

Proof: Consider the quantity ${}^*\gamma \equiv \gamma_1\gamma_2\cdots\gamma_{2\omega}$; as $\omega \in N$ ${}^*\gamma$ anticommutes with all the γ_μ, and further $({}^*\gamma)^2 = \sigma 1$, where $\sigma = \pm 1$. Hence for p odd

$$\begin{aligned}\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_p}] &= \sigma\,\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_p}{}^*\gamma{}^*\gamma] \\ &= -\sigma\,\mathrm{tr}[{}^*\gamma\gamma_{\mu_1}\cdots\gamma_{\mu_p}{}^*\gamma] \\ &= -\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_p}] = 0,\end{aligned}$$

where we first anticommuted ${}^*\gamma$ to the left and then used the cyclic property.□

Theorem 2: $\mathrm{tr}[\Gamma^{(j)}_{\mu_1\cdots\mu_j}] = \delta_{j0}\,\mathrm{tr}[1]$.

Proof: If $j/2 \in N$ and $j \neq 0$, then

$$\mathrm{tr}[\Gamma^{(j)}_{\mu_1\cdots\mu_j}] = \frac{1}{j!}\delta^{\nu_1\cdots\nu_j}_{\mu_1\cdots\mu_j}\mathrm{tr}[\gamma_{\nu_1}\cdots\gamma_{\nu_j}],$$

where by antisymmetry we may take all the subscripts in the trace to be different, hence all the of the γ's anticommute and thus

$$\begin{aligned}\mathrm{tr}[\gamma_{\nu_1}\cdots\gamma_{\nu_j}] &= -\mathrm{tr}[\gamma_{\nu_j}\gamma_{\nu_1}\cdots\gamma_{\nu_{j-1}}] \\ &= -\mathrm{tr}[\gamma_{\nu_1}\cdots\gamma_{\nu_j}] = 0,\end{aligned}$$

again by first anticommuting γ_{ν_j} to the left and then using cyclicity. If $j/2 \notin N$ then using the definition of $\Gamma^{(j)}_{\mu_1\cdots\mu_j}$ as above we obtain the trace of an odd number of γ's which vanishes by Theorem 1. Finally, for $j = 0$ we obtain $\mathrm{tr}[\Gamma^{(0)}] = \mathrm{tr}[1]$ trivially as $\Gamma^{(0)} = 1$ by definition.□

It is reasonable to take $\mathrm{tr}[1] = 2^\omega$, but this not needed as the tr[1]'s cancel in practical calculations.

Theorem 2 is in fact just a special case of

Theorem 3: The basis elements are orthogonal, by which we mean

$$\mathrm{tr}[\Gamma^{(i)}_{\mu_1\cdots\mu_i}\Gamma^{(j)\nu_1\cdots\nu_j}] = (-)^{i(i-1)/2}\delta^{\nu_1\cdots\nu_i}_{\mu_1\cdots\mu_i}\delta_{ij}\,\mathrm{tr}[1].$$

Proof:

$$\mathrm{tr}[\Gamma^{(i)}_{\mu_1\cdots\mu_i}\Gamma^{(j)\nu_1\cdots\nu_j}] = \delta^{\lambda_1\cdots\lambda_i}_{\mu_1\cdots\mu_i}\delta^{\nu_1\cdots\nu_j}_{\rho_1\cdots\rho_j}\mathrm{tr}[\gamma_{\lambda_1}\cdots\gamma_{\lambda_i}\gamma^{\rho_1}\cdots\gamma^{\rho_j}]/i!j!.$$

As the indices $\lambda_1,\ldots,\lambda_i$ and $\rho_1,\ldots,\rho_j$ are antisymmetrized it is easy to see that the trace is

$$\delta_{ij}(-)^{i(i-1)/2}\delta^{\rho_1\cdots\rho_i}_{\lambda_1\cdots\lambda_i}\mathrm{tr}[1]$$

(q.v. Theorem 5), so using the identity

$$\delta^{\lambda_1\cdots\lambda_i}_{\mu_1\cdots\mu_i}\delta^{\rho_1\cdots\rho_i}_{\lambda_1\cdots\lambda_i} = i!\delta^{\rho_1\cdots\rho_i}_{\mu_1\cdots\mu_i}$$

we obtain the required result.□

Theorem 4:

$$(\gamma_{\nu_1}\cdots\gamma_{\nu_p})\gamma_\mu = (-)^p\gamma_\mu\gamma_{\nu_1}\cdots\gamma_{\nu_p} + 2\sum_{i=1}^{p}(-)^{i-p}\left(\prod_{\substack{j=1\\ \neq i}}^{p}\gamma_{\nu_j}\right)g_{\mu\nu_i}.$$

Proof: By straightforward anticommutation

$$\begin{aligned}\gamma_{\nu_1}\cdots\gamma_{\nu_p}\gamma_\mu &= \gamma_{\nu_1}\cdots\gamma_{\nu_{p-1}}\{\gamma_{\nu_p},\gamma_\mu\} - \gamma_{\nu_1}\cdots\gamma_{\nu_{p-1}}\gamma_\mu\gamma_{\nu_p} \\ &= 2g_{\nu_p\mu}\gamma_{\nu_1}\cdots\gamma_{\nu_{p-1}} - [\gamma_{\nu_1}\cdots\gamma_{\nu_{p-1}}\gamma_\mu]\gamma_{\nu_p}.\end{aligned}$$

We now proceed by induction on p: assume the result for $p-1$; then the equation above gives

$$\begin{aligned}\gamma_{\nu_1}\cdots\gamma_{\nu_p}\gamma_\mu &= 2g_{\mu\nu_p}\gamma_{\nu_1}\cdots\gamma_{\nu_{p-1}} \\ &\quad -\left[(-)^{p-1}\gamma_\mu\gamma_{\nu_1}\cdots\gamma_{\nu_{p-1}}\right. \\ &\quad \left.+2\sum_{i=1}^{p-1}(-)^{i-p+1}\left(\prod_{\substack{j=1\\ \neq i}}^{p-1}\gamma_{\nu_j}\right)g_{\mu\nu_i}\right]\gamma_{\nu_p} \\ &= (-)^p\gamma_\mu\gamma_{\nu_1}\cdots\gamma_{\nu_p} + 2\sum_{i=1}^{p}(-)^{i-p}\left(\prod_{\substack{j=1\\ \neq i}}^{p}\gamma_{\nu_j}\right)g_{\mu\nu_i};\end{aligned}$$

as the result holds trivially for $p = 0$ the proof is complete.□

Theorem 5: For p even

$$\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_p}] = \sum_{i=1}^{p-1}(-)^{i+1}g_{\mu_p\mu_i}\mathrm{tr}[\gamma_{\mu_1}\cdots\hat{\gamma}_{\mu_i}\cdots\hat{\gamma}_{\mu_p}],$$

where the circumflex on a γ_μ indicates that it is missing from the trace.

Proof: Using Theorem 4 we have

$$\begin{aligned}(\gamma_{\mu_1}\cdots\gamma_{\mu_{p-1}})\gamma_{\mu_p} &= (-)^{p-1}\gamma_{\mu_p}\gamma_{\mu_1}\cdots\gamma_{\mu_{p-1}} \\ &\quad +2\sum_{i=1}^{p-1}(-)^{i-p+1}\left(\prod_{\substack{j=1\\ \neq i}}^{p-1}\gamma_{\mu_j}\right)g_{\mu_p\mu_i};\end{aligned}$$

so taking the trace and noting that $(-)^p = +1$ gives

$$\begin{aligned}\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_p}] &= -\mathrm{tr}[\gamma_{\mu_p}\gamma_{\mu_1}\cdots\gamma_{\mu_{p-1}}] \\ &\quad +2\sum_{i=1}^{p-1}(-)^{i+1}\mathrm{tr}[\gamma_{\mu_1}\cdots\hat{\gamma}_{\mu_i}\cdots\gamma_{\mu_{p-1}}\hat{\gamma}_{\mu_p}]g_{\mu_i\mu_p}.\end{aligned}$$

As the trace is cyclic our result follows.□

B. Graphical method

Theorem 5, in conjunction with Theorem 1, allows us to evaluate any trace by applying it repeatedly. A convenient notation for using it for hand calculation is to write the *connection*

$$\mathrm{tr}[\gamma_\mu\gamma_\nu] \equiv \underbrace{\gamma_\mu\gamma_\nu} = g_{\mu\nu}\mathrm{tr}[1]$$

exactly analogously to Wick's theorem. Indeed we may introduce the sign convention that

$$\begin{aligned}&\mathrm{tr}[\gamma_{\mu_1}\cdots\underbrace{\gamma_{\mu_p}\cdots\gamma_{\mu_q}}\cdots\gamma_{\mu_r}] \\ &\quad \equiv (-)^{q-p+1}g_{\mu_p\mu_q}\mathrm{tr}[\gamma_{\mu_1}\cdots\hat{\gamma}_{\mu_p}\cdots\hat{\gamma}_{\mu_q}\cdots\gamma_{\mu_r}],\end{aligned}$$

i.e., the sign we would get from anticommuting γ_{μ_q} through to γ_{μ_p}.

With this notation, Theorem 5 may be rewritten as

$$\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_i}\cdots\gamma_{\mu_p}] = \sum_{i=1}^{p-1}\mathrm{tr}[\gamma_{\mu_1}\cdots\underbrace{\gamma_{\mu_i}\cdots\gamma_{\mu_p}}].$$

Proceeding as in Wick's theorem we may apply this formula recursively to each term on the right-hand side so as to obtain eventually

$$\mathrm{tr}[\gamma_{\mu_1}\cdots\gamma_{\mu_p}] = \sum_{\{\text{all connections}\}} \mathrm{tr}[\underbrace{\gamma_{\mu_1}\gamma_{\mu_2}}\underbrace{\gamma_{\mu_3}\cdots\gamma_{\mu_{p-1}}\gamma_{\mu_p}}],$$

where the sum is over all possible pairwise connections. A convenient way to calculate the sign of a given term is to notice that if the connecting lines cross n times, the appropriate sign is $(-)^n$.

Example:

$$\begin{aligned}\mathrm{tr}[\gamma_\mu\gamma_\rho\gamma_\sigma\gamma_\lambda] &= \mathrm{tr}[\underbrace{\gamma_\mu\gamma_\rho}\underbrace{\gamma_\sigma\gamma_\lambda}] \\ &+ \mathrm{tr}[\underbrace{\gamma_\mu\gamma_\rho\gamma_\sigma\gamma_\lambda}] + \mathrm{tr}[\underbrace{\gamma_\mu\gamma_\rho}\underbrace{\gamma_\sigma\gamma_\lambda}] \\ &= (g_{\mu\rho}g_{\sigma\lambda} - g_{\mu\sigma}g_{\lambda\rho} + g_{\mu\lambda}g_{\rho\sigma})\mathrm{tr}[\mathbb{1}].\end{aligned}$$

While this method is fairly reasonable for small traces, the number of terms generated is large if the trace is long: for a trace of p γ's we obtain $(p-1)(p-3)\cdots3\cdot1 \equiv (p-1)!!$ terms (= number of ways of choosing $p/2$ pairs from p objects). Of course this is unavoidable if all the p indices are distinct, however if there are contracted indices or symmetries among the indices—as there almost always are in Feynman diagram calculations for instance—these can be utilized in a systematic manner. We shall discuss this in Sec. IV.

C. Contraction theorems

In a four-dimensional space Chisolm[3] found some useful rules based on the identity

$$\gamma_\alpha\gamma_{\mu_1}\cdots\gamma_{\mu_{2p-1}}\gamma^\alpha = -2\gamma_{\mu_{2p-1}}\cdots\gamma_{\mu_1},$$

which were developed into a powerful systematic reduction algorithm by Kahane.[4] Unfortunately such rules appear to be peculiar to four dimensions, and in general contractions like the one above in 2ω dimensions generate a large number of terms.

Nevertheless such identities may be useful, as we shall show.

Theorem 6:

$$\begin{aligned}&\gamma_\mu\left(\prod_{i=1}^{d}\gamma_{\nu_i}\right)\gamma^\mu \\ &\quad= 2\sum_{i=2}^{d}(-)^{i-d}\gamma_{\nu_i}\left(\prod_{\substack{j=1\\ \neq i}}^{d}\gamma_{\nu_j}\right) + (-)^d(2\omega-2)\left(\prod_{i=1}^{d}\gamma_{\nu_i}\right)\end{aligned}$$

where $d>0$.

Proof: This is really a trivially corollary of Theorem 4.□

Examples:

$$\gamma_\mu\gamma_\rho\gamma^\mu = -(2\omega-2)\gamma_\rho,$$

$$\gamma_\mu\gamma_\rho\gamma_\sigma\gamma^\mu = 2\gamma_\sigma\gamma_\rho + (2\omega-2)\gamma_\rho\gamma_\sigma,$$

$$\gamma_\mu\gamma_\rho\gamma_\sigma\gamma_\lambda\gamma^\mu = -2\gamma_\sigma\gamma_\rho\gamma_\lambda + 2\gamma_\lambda\gamma_\rho\gamma_\sigma - (2\omega-2)\gamma_\rho\gamma_\sigma\gamma_\lambda.$$

We notice that Theorem 6 reduces a string of $d+2$ γ's to d terms each involving d γ's, so if this string occurs in a trace of p γ's we get $d(p-3)!!$ terms instead of $(p-1)!!$. As the trace is cyclic $d \leqslant p/2$ (if we are sensible and contract the shorter way around), so even this weaker identity can save a considerable amount of computation.

A particularly interesting special case of Theorem 6 occurs when we antisymmetrize on the indices $\nu_1,\ldots,\nu_d$, as then

$$\gamma_\mu\Gamma^{(d)}_{\nu_1\cdots\nu_d}\gamma^\mu = (-)^d 2(\omega-d)\Gamma^{(d)}_{\nu_1\cdots\nu_d}.$$

There is an alternative way of deriving this result, however. We introduce the notation of putting a tilde on repeated indices which are *not* to be implicitly summed over, so that $\gamma_{\tilde\mu}\Gamma^{(d)}_{\nu_1\cdots\nu_d}\gamma^{\tilde\mu}$ is one term in the contraction above. In this term either $\tilde\mu$ is equal to one member of the set $\{\nu_1,\ldots,\nu_d\}$ or it is not. In the first case we have

$$\begin{aligned}\gamma_{\tilde\mu}\Gamma^{(d)}_{\nu_1\cdots\nu_d}\gamma^{\tilde\mu} &= -(-)^d\gamma_{\tilde\mu}\gamma^{\tilde\mu}\Gamma^{(d)}_{\nu_1\cdots\nu_d} \\ &= -(-)^d\Gamma^{(d)}_{\nu_1\cdots\nu_d},\end{aligned}$$

and in the second

$$\gamma_{\tilde\mu}\Gamma^{(d)}_{\nu_1\cdots\nu_d}\gamma^{\tilde\mu} = (-)^d\Gamma^{(d)}_{\nu_1\cdots\nu_d}.$$

As $\Gamma^{(d)}_{\nu_1\cdots\nu_d}$ is totally antisymmetric all the ν_i are distinct, hence the first case occurs d times and the second $2\omega-d$ times, hence

$$\begin{aligned}\gamma_\mu\Gamma^{(2)}_{\nu_1\cdots\nu_d}\gamma^\mu &= d\left[-(-)^d\Gamma^{(d)}_{\nu_1\cdots\nu_d}\right] \\ &\quad + (2\omega-d)\left[(-)^d\Gamma^{(d)}_{\nu_1\cdots\nu_d}\right]\end{aligned}$$

as required.

This argument is easily generalized to give

Theorem 7:

$$\Gamma^{(l)}_{\mu_1\cdots\mu_l}\Gamma^{(d)}_{\nu_1\cdots\nu_d}\Gamma^{(l)\mu_1\cdots\mu_l} = \Delta(2\omega,d,l)\Gamma^{(d)}_{\nu_1\cdots\nu_d},$$

where

$$\begin{aligned}\Delta(n,d,l) &\equiv l!(-)^{l(l-1)/2}(-)^{dl} \\ &\quad\times\sum_{p=\max(0,l-n+d)}^{\min(l,d)}(-)^p\binom{n-d}{l-p}\binom{d}{p}.\end{aligned}$$

Example: In a 10-dimensional space $(\omega=5)$[7]

$$\Gamma^{(5)}_{\mu_1\cdots\mu_5}\Gamma^{(4)}_{\nu_1\cdots\nu_4}\Gamma^{(5)\mu_1\cdots\mu_5} = 1440\Gamma^{(4)}_{\nu_1\cdots\nu_4}.$$

Theorem 8: The function $\Delta(n,d,l)$ has the following properties:

$$\Delta(n,d,l) = (-)^{l(l-1)/2}(-)^{d(d-1)/2}\frac{(n-d)!}{(n-l)!}\Delta(n,l,d),$$

$$\Delta(n,n-d,l) = (-)^l\Delta(n,d,l),$$

$$\begin{aligned}\Delta(n,d,n-l) &= (-)^d(-)^{n(n-1)/2}(-)^{l(l-n)} \\ &\quad\frac{(n-l)!}{l!}\Delta(n,d,l).\end{aligned}$$

Proof: For the first identity we use the argument summarized by the following diagram:

$$\begin{array}{ccc} \Gamma^{(l)}_{\mu_1\cdots\mu_l}\Gamma^{(d)}_{\nu_1\cdots\nu_d}\Gamma^{(l)\mu_1\cdots\mu_l}\Gamma^{(d)\nu_1\cdots\nu_d} & & \\ \text{//} & & \text{\textbackslash\textbackslash} \\ \Gamma^{(l)}_{\mu_1\cdots\mu_l}\Delta(n,l,d)\Gamma^{(l)\mu_1\cdots\mu_l} & & \Delta(n,d,l)\Gamma^{(d)}_{\nu_1\cdots\nu_d}\Gamma^{(d)\nu_1\cdots\nu_d} \\ \text{\textbackslash\textbackslash} & & \text{//} \\ & \Delta(n,0,l)\Delta(n,l,d) = \Delta(n,0,d)\Delta(n,d,l), & \end{array}$$

where we have taken $n \equiv 2\omega$ as the dimensionality of spacetime. As

$$\Delta(n,0,k) = (-)^{k(k-1)/2}n!/(n-k)!$$

the result follows immediately. In verifying this result explicitly using the definition of Δ, the identity

$$\binom{n-d}{l-p}\binom{d}{p}\binom{n}{d}=\binom{n-l}{d-p}\binom{l}{p}\binom{n}{l}$$

proves useful.

For the remaining two identities it is convenient to use the anticommuting object ${}^*\gamma$ introduced in the proof of Theorem 1. According to Theorem 9 we have the duality condition

$$\Gamma^{(k)}_{\mu_1\cdots\mu_k}=\frac{(-)^{(n-k)(n-k-1)/2}}{(n-k)!}{}^*\gamma\epsilon_{\mu_1\cdots\mu_n}\times\Gamma^{(n-k)\mu_{k+1}\cdots\mu_n}\det(g)$$

so

$$\begin{aligned}\Gamma^{(l)}_{\mu_1\cdots\mu_l}\Gamma^{(n-d)}_{\nu_1\cdots\nu_{n-d}}\Gamma^{(l)\mu_1\cdots\mu_l}&=\frac{(-)^{d(d-1)/2}}{d!}\det(g)\epsilon_{\nu_1\cdots\nu_n}\\&\times\Gamma^{(l)}_{\mu_1\cdots\mu_l}{}^*\gamma\Gamma^{(d)\nu_{n-d+1}\cdots\nu_n\Gamma^{(l)}\mu_1\cdots\mu_l}\\&=\frac{(-)^{d(d-1)/2}(-)^l}{d!}\det(g)\Delta(n,d,l)\\&\times\epsilon_{\nu_1\cdots\nu_n}{}^*\gamma\Gamma^{\nu_{n-d+1}\cdots\nu_n}\\&=(-)^l\Delta(n,d,l)\Gamma_{\nu_1\cdots\nu_{n-d}};\end{aligned}$$

however the original expression clearly equals

$$\Delta(n,n-d,l)\Gamma_{\nu_1\cdots\nu_{n-d}}$$

too, giving the second identity.

The final identity follows from the preceding two.□

D. Fierz identities

The generalized contraction identity (Theorem 7) has as a useful corollary an explicit formula for 2ω-dimensional Fierz reorderings.

Corollary: The Fierz coefficients $a^{(2\omega)}_{lk}$ in the reordering

$$\begin{aligned}(\Gamma^{(l)\mu_1\cdots\mu_l})_{ab}(\Gamma^{(l)}_{\mu_1\cdots\mu_l})_{cd}&=\sum_{k=0}^{2\omega}a^{(2\omega)}_{lk}(\Gamma^{(k)\nu_1\cdots\nu_k})_{ad}(\Gamma^{(k)}_{\nu_1\cdots\nu_k})_{cb},\end{aligned}$$

or in compact form

$$\Gamma^{(l)}\otimes\Gamma^{(l)}=\sum_{k=0}^{2\omega}a^{(2\omega)}_{lk}\Gamma^{(k)}\times\Gamma^{(k)},$$

are given by

$$a^{(2\omega)}_{lk}=\frac{\Delta(2\omega,k,l)(-)^{k(k-1)/2}}{k!\operatorname{tr}[\mathbb{1}]}.$$

Proof: Multiplying the Fierz equation by

$$(\Gamma^{(j)\rho_1\cdots\rho_j})_{bc}(\Gamma^{(j)}_{\rho_1\cdots\rho_j})_{da}$$

gives

$$\begin{aligned}\operatorname{tr}\left[\Gamma^{(l)\mu_1\cdots\mu_l}\Gamma^{(j)\rho_1\cdots\rho_j}\Gamma^{(l)}_{\mu_1\cdots\mu_l}\Gamma^{(j)}_{\rho_1\cdots\rho_j}\right]&=\sum_{k=0}^{2\omega}a^{(2\omega)}_{lk}\operatorname{tr}\left[\Gamma^{(k)\nu_1\cdots\nu_k}\Gamma^{(j)}_{\rho_1\cdots\rho_j}\right]\\&\times\operatorname{tr}\left[\Gamma^{(k)}_{\nu_1\cdots\nu_k}\Gamma^{(j)\rho_1\cdots\rho_j}\right];\end{aligned}$$

hence

$$\begin{aligned}&\Delta(2\omega,j,l)\Delta(2\omega,0,j)\operatorname{tr}[\mathbb{1}]\\&=\sum_{k=0}^{2\omega}a^{(2\omega)}_{lk}(-)^{k(k-1)/2}\delta^{\nu_1\cdots\nu_k}_{\rho_1\cdots\rho_k}\delta_{jk}\operatorname{tr}[\mathbb{1}]\\&\times(-)^{k(k-1)/2}\delta^{\rho_1\cdots\rho_k}_{\nu_1\cdots\nu_k}\delta_{jk}\operatorname{tr}[\mathbb{1}]\\&\Rightarrow\Delta(2\omega,j,l)(-)^{j(j-1)/2}\frac{(2\omega)!}{(2\omega-j)!}\\&=a^{(2\omega)}_{lk}j!\frac{(2\omega)!}{(2\omega-j)!}\operatorname{tr}[\mathbb{1}].\end{aligned}$$

Example: With $2\omega=10$ the Fierz coefficient $a^{(10)}_{0,10}=-1/116121600$.

In practice the basis tensors $\Gamma^{(l)}_{\mu_1\cdots\mu_l}$ with l near 2ω are more conveniently replaced with their duals constructed using the Levi–Civita tensor $\epsilon_{\mu_1\cdots\mu_{2\omega}}$. This tensor is completely antisymmetric on all its indices and has its sign fixed by

$$\epsilon^{12\cdots2\omega}=1;$$

it thus has the following useful properties:

$$\begin{aligned}\epsilon_{1\cdots2\omega}&=\epsilon^{\mu_1\cdots\mu_{2\omega}}g_{1\mu_1}\cdots g_{2\omega\mu_{2\omega}}=\det(g),\\\epsilon_{\alpha_1\cdots\alpha_{2\omega}}\epsilon^{\beta_1\cdots\beta_{2\omega-p}\alpha_{2\omega-p+1}\cdots\alpha_{2\omega}}&=p!\delta^{\beta_1\cdots\beta_{2\omega-p}}_{\alpha_1\cdots\alpha_{2\omega-p}}\det(g).\end{aligned}$$

In terms of the Levi–Civita tensor ${}^*\gamma$ becomes

$${}^*\gamma\equiv\frac{1}{(2\omega)!}\epsilon^{\mu_1\cdots\mu_{2\omega}}\Gamma^{(2\omega)}_{\mu_1\cdots\mu_{2\omega}},$$

and we introduce the dual basis ${}^*\gamma\Gamma^{(l)}_{\mu_1\cdots\mu_l}$ related to the original basis by the following:

Theorem 9:

$$\Gamma^{(k)}_{\mu_1\cdots\mu_k}=\frac{(-)^{(2\omega-k)(2\omega-k-1)/2}}{(2\omega-k)!}\epsilon_{\mu_1\cdots\mu_{2\omega}}{}^*\gamma\Gamma^{(2\omega-k)\mu_{k+1}\cdots\mu_{2\omega}}\times\det(g).$$

Proof: Multiply by $\Gamma^{(j)}_{\rho_1\cdots\rho_j}$ and take the trace.□

These dual tensors are easily manipulated using the rules that

$$\begin{aligned}({}^*\gamma)^2&=(-)^{\omega(2\omega-1)}\det(g),\\\{{}^*\gamma,\gamma_\mu\}&=0:2\omega\text{ even},\\[{}^*\gamma,\gamma_\mu]&=0:2\omega\text{ odd},\\{}^*\gamma\Gamma^{(l)}_{\mu_1\cdots\mu_l}&=(-)^l\Gamma^{(l)}_{\mu_1\cdots\mu_l}{}^*\gamma:2\omega\text{ even},\\&=\Gamma^{(l)}_{\mu_1\cdots\mu_l}{}^*\gamma:2\omega\text{ odd}.\end{aligned}$$

A useful family of basis vectors consists of a mixture of $\Gamma^{(l)}$ and ${}^*\gamma\Gamma^{(l)}$ for $0\leqslant l<\omega$ (by convention we shall chose $\Gamma^{(\omega)}$ rather than ${}^*\gamma\Gamma^{(\omega)}$ when they differ by a sign). In this basis the Fierz matrix looks like

$$\begin{array}{c|cc}&\Gamma&{}^*\gamma\Gamma\\\hline\Gamma&a_{++}&a_{+-}\\{}^*\gamma\Gamma&a_{-+}&a_{--}\end{array},$$

where it can be shown that

$$(a_{++})_{lj} = a^{(2\omega)}_{lj}, \quad 0 \leqslant l,j \leqslant \omega$$
$$(a_{+-})_{lj} = \sigma^{(2\omega)}_{lj} a^{(2\omega)}_{l\,2\omega-j}, \quad 0 \leqslant l \leqslant \omega < j \leqslant 2\omega$$
$$(a_{-+})_{lj} = \sigma^{(2\omega)}_{jl} a^{(2\omega)}_{2\omega-l\,j}, \quad 0 \leqslant j \leqslant \omega < l \leqslant \omega$$
$$(a_{--})_{lj} = \sigma^{(2\omega)}_{lj}\sigma^{(2\omega)}_{jl} a^{(2\omega)}_{2\omega-l\,2\omega-j} \quad \omega < j,\ l \leqslant 2\omega$$
$$= a^{(2\omega)}_{2\omega-l\,2\omega-j} \quad \text{for } \omega \in N,$$

where

$$\sigma^{(2\omega)}_{lj} \equiv (-)^{l}(-)^{\omega(2\omega-1)}(-)^{j(2\omega-j)}\det(g).$$

The matrices in this mixed basis for $\det(g) = +1$ are given for $\omega = 1,2,3,4,5,6,7$ in the Appendix (only the part a_{++} for the last two; the remainder may be reconstructed using the above symmetries).

IV. REDUCTION ALGORITHMS

In Sec. 2 we showed how to take traces of products of γ matrices. Notice that this, combined with Theorem 3, enables us to put all such expressions Λ into a canonical form, namely

$$\Lambda = \sum_{k=0}^{2\omega} \Lambda^{(k)}_{\mu_1\cdots\mu_k} \Gamma^{(k)\mu_1\cdots\mu_k}$$

(suppressing free indices) where the coefficients $\Lambda^{(k)}$ are obtained using Theorem 3:

$$\Lambda^{(k)}_{\mu_1\cdots\mu_k} = \frac{(-)^{k(k-1)/2}\mathrm{tr}[\Lambda \Gamma^{(k)}_{\mu_1\cdots\mu_k}]}{k!\,\mathrm{tr}[\mathbb{1}]}.$$

A special case is that the scalar (unit matrix) part of Λ is just $\mathrm{tr}[\Lambda]/\mathrm{tr}[\mathbb{1}]$.

Example: To demonstrate the utility of projection techniques we shall find the spinorial generators of SO(n), for N even [strictly speaking of the covering group spin (n)].

The γ_μ transform according to the vector (adjoint) representation of SO(n),

$$\gamma_\mu \to \hat{\gamma}_\mu = \Omega_\mu{}^\nu \gamma_\nu,$$

where the matrices $\Omega_{\hat{\mu}}{}^{\mu}$ preserve the metric $g_{\mu\nu}$:

$$g_{\mu\nu} \to g_{\mu\nu} = \Omega_\mu{}^\rho \Omega_\nu{}^\sigma g_{\rho\sigma};$$

in terms of infinitesimal rotations with parameters

$$\Omega_\mu{}^\nu = \delta_\mu{}^\nu + \omega_\mu{}^\nu$$

this means

$$\omega_{\mu\nu} = -\omega_{\nu\mu}.$$

As the group action preserves the anticommutation relations of the Clifford algebra

$$\{\gamma_\mu,\gamma_\nu\} \to \{\hat{\gamma}_\mu,\hat{\gamma}_\nu\} = g_{\mu\nu}$$

it furnishes an (outer) automorphism; however we have already shown that there is an isomorphism $\mathscr{D}^{(2\omega)}$ of the Clifford algebra onto the complete algebra of $2^\omega \times 2^\omega$ matrices, and hence by a well-known theorem[5] all automorphisms are equivalent to inner automorphisms. This means there is an $S(\Omega)$ in the Clifford algebra such that

$$\hat{\gamma}_\mu = S^{-1}(\Omega)\gamma_\mu S(\Omega),$$

or for infinitesimal rotations

$$\hat{\gamma}_\mu = [\mathbb{1} - s(\omega)]\gamma_\mu[\mathbb{1} + s(\omega)],$$
$$\Rightarrow \omega_\mu{}^\nu \gamma_\nu = [\gamma_\mu, s(\omega)].$$

To solve this equation we multiply both sides by γ^μ and note that ω is antisymmetric

$$\omega_{\mu\nu}\gamma^\mu\gamma^\nu = 2\omega_{\mu\nu}\Gamma^{(2)\mu\nu} = \gamma^\mu\gamma_\mu s(\omega) - \gamma^\mu s(\omega)\gamma_\mu.$$

As the left-hand side is proportional to $\Gamma^{(2)}$ we project out that part

$$2\omega_{\mu\nu}\mathrm{tr}[\Gamma^{(2)\mu\nu}\Gamma^{(2)}_{\alpha\beta}] = n\,\mathrm{tr}[s(\omega)\Gamma^{(2)}_{\alpha\beta}] - \mathrm{tr}[\gamma^\mu s(\omega)\gamma_\mu \Gamma^{(2)}_{\alpha\beta}],$$
$$2\omega_{\mu\nu}\delta^{\mu\nu}_{\beta\alpha}\mathrm{tr}[\mathbb{1}] = n\,\mathrm{tr}[s(\omega)\Gamma^{(2)}_{\alpha\beta}] - \mathrm{tr}[s(\omega)\gamma_\mu\Gamma^{(2)}_{\alpha\beta}\gamma^\mu]$$
$$= 4\mathrm{tr}[s(\omega)\Gamma^{(2)}_{\alpha\beta}]$$

since $\gamma_\mu\Gamma^{(2)}_{\alpha\beta}\gamma^\mu = (n-4)\Gamma^{(2)}_{\alpha\beta}$. Expanding $s(\omega)$ as $\Sigma^n_{k=0} s^{(k)}_{\rho_1\cdots\rho_k}\Gamma^{(k)\rho_1\cdots\rho_k}$ we have

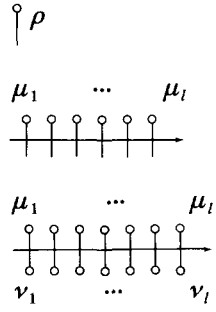
$$\omega_{\alpha\beta} = 2s^{(2)}_{\alpha\beta},$$

or

$$s(\omega) = \tfrac{1}{2}\omega_{\alpha\beta}\Gamma^{(2)\alpha\beta},$$

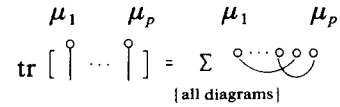
which is the desired result.

Performing a general reduction involves taking 2ω traces half of which are exponential in the length of Λ. We shall now show that this is unnecessary. First of all, we introduced a graphical notation for Theorem 5 rather akin to going from Wick expansions to Feynman diagrams; we write

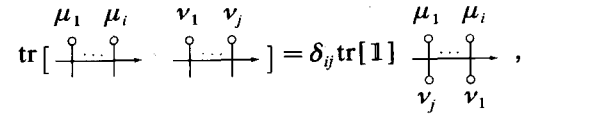


and

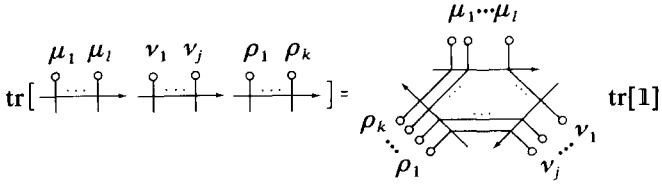
The arrows are needed to specify the correct sign. Theorem 5 now reads:



where the sign is determined as before. Theorem 3 becomes



and similarly we have



where the symbol on the right-hand side stands for

$$\Xi_{\mu_1\cdots\mu_l\nu_1\cdots\nu_j\rho_1\cdots\rho_k}$$

$$\equiv \sum_{\{\text{permutations}\}} (-)^p g_{\mu_1\rho_k} g_{\mu_2\rho_{k-1}}\cdots g_{\mu_s\rho_{k-s+1}} \times g_{\nu_1\mu_l} g_{\nu_2\mu_{l-1}}\cdots g_{\nu_t\mu_{l-t+1}} \times g_{\rho_1\nu_j} g_{\rho_2\nu_{j-1}}\cdots g_{\rho_u\nu_{j-u+1}} / s!t!u!,$$

where clearly

$$\begin{aligned} s+1 &= l-t+1 & s &= \tfrac{1}{2}(l-j+k) \\ t+1 &= j-u+1 \quad\Rightarrow & t &= \tfrac{1}{2}(j-k+l) \\ u+1 &= k-s+1 & u &= \tfrac{1}{2}(k-l+j) \end{aligned}$$

and the summation is over all permutations of the indices within each set $\{\mu_1,\ldots,\mu_l\},\{\nu_i,\ldots,\nu_j\},\{\rho_1,\ldots,\rho_k\}$ with $(-)^p$ being $+(-)$ for even (odd) permutations.

Ξ is similar to the $3j$ symbols arising in the theory of angular momentum, as it tells us how to couple three Γ tensors to produce a scalar. This relation may be rearranged to give the Clebsch–Gordan series for the Clifford algebra.

Theorem 10:

$$\Gamma^{(l)}_{\mu_1\cdots\mu_l}\Gamma^{(j)}_{\nu_1\cdots\nu_j} = \sum_{k=|l-j|}^{l+j} \frac{1}{k!}\Xi_{\mu_1\cdots\mu_l\nu_1\cdots\nu_j\rho_k\cdots\rho_1}\Gamma^{(k)\rho_1\cdots\rho_k}.$$

It is obvious on inspection of the diagramatic representation of Ξ that it vanishes unless l,j,k satisfy the triangle inequality written in the equation above, and also unless $l+j+k$ is even.

This theorem allows us to formulate a polynomial algorithm for the reduction of γ expressions rather than the exponential algorithm of Theorem 5. The method is recursively defined: to reduce an expression we split it into two halves and recursively reduce each half into a sum over basis elements $\Gamma_{(l)}$, which may be combined using Theorem 10. If this algorithm takes $N(p)$ steps to reduce an expression of length p then clearly for large p:

$$N(p) = 2N(p/2) + (p/2)^2,$$

where the final term is the number of steps required to combine the two sets of $p/2$ basis elements together. The leading term in the solution of this equation for $N(p)$ is $p^2/2$, which is clearly a great improvement over $(p-1)!!$ even for quite moderate values of p.

Another advantage of this algorithm is that if factors such as $\not{p}+m$ occur in the expression to be reduced there is no need to distribute the multiplication over the addition; one merely finds the coefficient of the basis elements become a sum of terms.

The result of this algorithm is given in terms of antisymmetrized products of g's, namely Ξ's; however it is almost always the case that the factors constructed out of external momenta and contracted indices have symmetries which may be used to simplify the result *without* expanding all the Ξ's. If all the factors in the original expression are distinct, then one will still obtain an exponential number of terms on expanding these symmetries, but clearly this is unavoidable.

A reduction of the coefficients using the properties of the symmetric group $\mathcal{S}_p$ may presumably be used to exploit all the symmetries systematically.

ACKNOWLEDGMENTS

I would like to thank D. Bailin and J. Sucher for reading various versions of the manuscript, and the Science Research Council for support during the preparation of an early version of this paper.

This research was supported in part by the National Science Foundation.

APPENDIX: FIERZ MATRICES FOR 2ω DIMENSIONS

$2\omega = 2$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	${}^*\gamma\Gamma^{(0)}\times{}^*\gamma\Gamma^{(0)}$
$\Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
$\Gamma^{(1)}\otimes\Gamma^{(1)}$	1	0	1
${}^*\gamma\Gamma^{(0)}\otimes{}^*\gamma\Gamma^{(0)}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

$2\omega = 4$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	$\Gamma^{(2)}\times\Gamma^{(2)}$	${}^*\gamma\Gamma^{(1)}\times{}^*\gamma\Gamma^{(1)}$	${}^*\gamma\Gamma^{(0)}\times{}^*\gamma\Gamma^{(0)}$
$S\otimes S = \mathbb{1}\otimes\mathbb{1} = \Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{8}$	$-\frac{1}{4}$	$\frac{1}{4}$
$V\otimes V = \gamma_\mu\otimes\gamma^\mu = \Gamma^{(1)}\otimes\Gamma^{(1)}$	1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	-1
$T\otimes T = \Gamma^{(2)}\otimes\Gamma^{(2)}$	-3	0	$-\frac{1}{2}$	0	-3
$A\otimes A = {}^*\gamma_\mu\otimes{}^*\gamma\gamma^\mu = {}^*\gamma\Gamma^{(1)}\otimes{}^*\gamma\Gamma^{(1)}$	-1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	1
$P\otimes P = {}^*\gamma\otimes{}^*\gamma = {}^*\gamma\Gamma^{(0)}\otimes{}^*\gamma\Gamma^{(0)}$	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$

$2\omega = 6$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	$\Gamma^{(2)}\times\Gamma^{(2)}$	$\Gamma^{(3)}\times\Gamma^{(3)}$	${}^*\gamma\Gamma^{(2)}\times{}^*\gamma\Gamma^{(2)}$	${}^*\gamma\Gamma^{(1)}\times{}^*\gamma\Gamma^{(1)}$	${}^*\gamma\Gamma^{(0)}\times{}^*\gamma\Gamma^{(0)}$
$\Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{8}$	$\frac{1}{8}$	$-\frac{1}{16}$	$-\frac{1}{48}$	$\frac{1}{16}$	$\frac{1}{8}$	$-\frac{1}{8}$
$\Gamma^{(1)}\otimes\Gamma^{(1)}$	$\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{8}$	0	$-\frac{1}{8}$	$\frac{1}{2}$	$\frac{3}{4}$
$\Gamma^{(2)}\otimes\Gamma^{(2)}$	$-\frac{15}{4}$	$-\frac{5}{4}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$\frac{1}{8}$	$-\frac{5}{4}$	$\frac{15}{4}$
$\Gamma^{(3)}\otimes\Gamma^{(3)}$	-15	0	$-\frac{3}{2}$	0	$-\frac{3}{2}$	0	-15
${}^*\gamma\Gamma^{(2)}\otimes{}^*\gamma\Gamma^{(2)}$	$\frac{15}{4}$	$-\frac{5}{4}$	$\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{5}{4}$	$-\frac{15}{4}$
${}^*\gamma\Gamma^{(1)}\otimes{}^*\gamma\Gamma^{(1)}$	$\frac{3}{4}$	$\frac{1}{2}$	$-\frac{1}{8}$	0	$-\frac{1}{8}$	$-\frac{1}{2}$	$\frac{3}{4}$
${}^*\gamma\Gamma^{(0)}\otimes{}^*\gamma\Gamma^{(0)}$	$-\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$-\frac{1}{48}$	$-\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$

$2\omega = 8$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	$\Gamma^{(2)}\times\Gamma^{(2)}$	$\Gamma^{(3)}\times\Gamma^{(3)}$	$\Gamma^{(4)}\times\Gamma^{(4)}$	${}^*\gamma\Gamma^{(3)}\times{}^*\gamma\Gamma^{(3)}$	${}^*\gamma\Gamma^{(2)}\times{}^*\gamma\Gamma^{(2)}$	${}^*\gamma\Gamma^{(1)}\times{}^*\gamma\Gamma^{(1)}$	${}^*\gamma\Gamma^{(0)}\times{}^*\gamma\Gamma^{(0)}$
$\Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{16}$	$\frac{1}{16}$	$-\frac{1}{32}$	$-\frac{1}{96}$	$\frac{1}{384}$	$\frac{1}{96}$	$-\frac{1}{32}$	$-\frac{1}{16}$	$\frac{1}{16}$
$\Gamma^{(1)}\otimes\Gamma^{(1)}$	$\frac{1}{2}$	$-\frac{3}{8}$	$-\frac{1}{8}$	$\frac{1}{48}$	0	$\frac{1}{48}$	$\frac{1}{8}$	$-\frac{3}{8}$	$-\frac{1}{2}$
$\Gamma^{(2)}\otimes\Gamma^{(2)}$	$-\frac{7}{2}$	$-\frac{7}{4}$	$\frac{1}{4}$	$-\frac{1}{24}$	$\frac{1}{48}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{7}{4}$	$-\frac{7}{2}$
$\Gamma^{(3)}\otimes\Gamma^{(3)}$	-21	$\frac{21}{4}$	$-\frac{3}{4}$	$\frac{3}{8}$	0	$\frac{3}{8}$	$\frac{3}{4}$	$\frac{21}{4}$	21
$\Gamma^{(4)}\otimes\Gamma^{(4)}$	105	0	$\frac{15}{2}$	0	$\frac{3}{8}$	0	$\frac{15}{2}$	0	105
${}^*\gamma\Gamma^{(3)}\otimes{}^*\gamma\Gamma^{(3)}$	21	$\frac{21}{4}$	$\frac{3}{4}$	$\frac{3}{8}$	0	$\frac{3}{8}$	$-\frac{3}{4}$	$\frac{21}{4}$	-21
${}^*\gamma\Gamma^{(2)}\otimes{}^*\gamma\Gamma^{(2)}$	$-\frac{7}{2}$	$\frac{7}{4}$	$\frac{1}{4}$	$\frac{1}{24}$	$\frac{1}{48}$	$-\frac{1}{24}$	$\frac{1}{4}$	$-\frac{7}{4}$	$-\frac{7}{2}$
${}^*\gamma\Gamma^{(1)}\otimes{}^*\gamma\Gamma^{(1)}$	$-\frac{1}{2}$	$-\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{48}$	0	$\frac{1}{48}$	$-\frac{1}{8}$	$-\frac{3}{8}$	$\frac{1}{2}$
${}^*\gamma\Gamma^{(0)}\otimes{}^*\gamma\Gamma^{(0)}$	$\frac{1}{16}$	$-\frac{1}{16}$	$-\frac{1}{32}$	$\frac{1}{96}$	$\frac{1}{384}$	$-\frac{1}{96}$	$-\frac{1}{32}$	$\frac{1}{16}$	$\frac{1}{16}$

$2\omega = 10$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	$\Gamma^{(2)}\times\Gamma^{(2)}$	$\Gamma^{(3)}\times\Gamma^{(3)}$	$\Gamma^{(4)}\times\Gamma^{(4)}$	$\Gamma^{(5)}\times\Gamma^{(5)}$	${}^*\gamma\Gamma^{(4)}\times{}^*\gamma\Gamma^{(4)}$	${}^*\gamma\Gamma^{(3)}\times{}^*\gamma\Gamma^{(3)}$	${}^*\gamma\Gamma^{(2)}\times{}^*\gamma\Gamma^{(2)}$	${}^*\gamma\Gamma^{(1)}\times{}^*\gamma\Gamma^{(1)}$	${}^*\gamma\Gamma^{(0)}\times{}^*\gamma\Gamma^{(0)}$
$\Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{32}$	$+\frac{1}{32}$	$-\frac{1}{64}$	$-\frac{1}{192}$	$\frac{1}{768}$	$\frac{1}{3840}$	$-\frac{1}{768}$	$-\frac{1}{192}$	$\frac{1}{64}$	$\frac{1}{32}$	$-\frac{1}{32}$
$\Gamma^{(1)}\otimes\Gamma^{(1)}$	$\frac{5}{16}$	$-\frac{1}{4}$	$-\frac{3}{32}$	$\frac{1}{48}$	$\frac{1}{384}$	0	$\frac{1}{384}$	$-\frac{1}{48}$	$-\frac{3}{32}$	$\frac{1}{4}$	$+\frac{5}{16}$
$\Gamma^{(2)}\otimes\Gamma^{(2)}$	$-\frac{45}{16}$	$-\frac{27}{16}$	$\frac{13}{32}$	$\frac{1}{32}$	$\frac{1}{128}$	$\frac{1}{384}$	$-\frac{1}{128}$	$\frac{1}{32}$	$-\frac{13}{32}$	$-\frac{27}{16}$	$\frac{45}{16}$
$\Gamma^{(3)}\otimes\Gamma^{(3)}$	$-\frac{45}{2}$	9	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{16}$	0	$\frac{1}{16}$	$-\frac{1}{4}$	$\frac{3}{4}$	-9	$-\frac{45}{2}$
$\Gamma^{(4)}\otimes\Gamma^{(4)}$	$\frac{315}{2}$	$\frac{63}{2}$	$\frac{21}{4}$	$\frac{7}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$-\frac{1}{16}$	$\frac{7}{4}$	$-\frac{21}{4}$	$\frac{63}{2}$	$-\frac{315}{2}$
$\Gamma^{(5)}\otimes\Gamma^{(5)}$	945	0	$\frac{105}{2}$	0	$\frac{15}{8}$	0	$\frac{15}{8}$	0	$\frac{105}{2}$	0	945
${}^*\gamma\Gamma^{(4)}\otimes{}^*\gamma\Gamma^{(4)}$	$-\frac{315}{2}$	$\frac{63}{2}$	$-\frac{21}{4}$	$\frac{7}{4}$	$-\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{7}{4}$	$\frac{21}{4}$	$\frac{63}{2}$	$\frac{315}{2}$
${}^*\gamma\Gamma^{(3)}\otimes{}^*\gamma\Gamma^{(3)}$	$-\frac{45}{2}$	-9	$\frac{3}{4}$	$-\frac{1}{4}$	$\frac{1}{16}$	0	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{4}$	9	$-\frac{45}{2}$
${}^*\gamma\Gamma^{(2)}\otimes{}^*\gamma\Gamma^{(2)}$	$\frac{45}{16}$	$-\frac{27}{16}$	$-\frac{13}{32}$	$\frac{1}{32}$	$-\frac{1}{128}$	$\frac{1}{384}$	$\frac{1}{128}$	$\frac{1}{32}$	$\frac{13}{32}$	$-\frac{27}{16}$	$-\frac{45}{16}$
${}^*\gamma\Gamma^{(1)}\otimes{}^*\gamma\Gamma^{(1)}$	$\frac{5}{16}$	$\frac{1}{4}$	$-\frac{3}{32}$	$-\frac{1}{48}$	$\frac{1}{384}$	0	$\frac{1}{384}$	$\frac{1}{48}$	$-\frac{3}{32}$	$-\frac{1}{4}$	$\frac{5}{16}$
${}^*\gamma\Gamma^{(0)}\otimes{}^*\gamma\Gamma^{(0)}$	$-\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$-\frac{1}{192}$	$-\frac{1}{768}$	$\frac{1}{3840}$	$\frac{1}{768}$	$-\frac{1}{192}$	$-\frac{1}{64}$	$\frac{1}{32}$	$\frac{1}{32}$

$2\omega = 12$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	$\Gamma^{(2)}\times\Gamma^{(2)}$	$\Gamma^{(3)}\times\Gamma^{(3)}$	$\Gamma^{(4)}\times\Gamma^{(4)}$	$\Gamma^{(5)}\times\Gamma^{(5)}$	$\Gamma^{(6)}\times\Gamma^{(6)}$
$\Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{64}$	$\frac{1}{64}$	$-\frac{1}{128}$	$-\frac{1}{384}$	$\frac{1}{1536}$	$\frac{1}{7680}$	$-\frac{1}{46080}$
$\Gamma^{(1)}\otimes\Gamma^{(1)}$	$\frac{3}{16}$	$-\frac{5}{32}$	$-\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{384}$	$-\frac{1}{3840}$	0
$\Gamma^{(2)}\otimes\Gamma^{(2)}$	$-\frac{33}{16}$	$-\frac{11}{8}$	$\frac{13}{32}$	$\frac{1}{16}$	$-\frac{1}{384}$	$\frac{1}{960}$	$-\frac{1}{3840}$
$\Gamma^{(3)}\otimes\Gamma^{(3)}$	$-\frac{165}{8}$	$\frac{165}{16}$	$\frac{15}{8}$	$-\frac{1}{32}$	$\frac{3}{64}$	$-\frac{1}{128}$	0
$\Gamma^{(4)}\otimes\Gamma^{(4)}$	$\frac{1485}{8}$	$\frac{495}{8}$	$-\frac{45}{16}$	$\frac{27}{16}$	$-\frac{17}{64}$	$\frac{1}{64}$	$-\frac{1}{128}$
$\Gamma^{(5)}\otimes\Gamma^{(5)}$	1485	$-\frac{495}{2}$	45	$-\frac{45}{4}$	$\frac{5}{8}$	$-\frac{5}{16}$	0
$\Gamma^{(6)}\otimes\Gamma^{(6)}$	-10395	0	$-\frac{945}{2}$	0	$-\frac{105}{8}$	0	$-\frac{5}{16}$

$2\omega = 14$

	$\Gamma^{(0)}\times\Gamma^{(0)}$	$\Gamma^{(1)}\times\Gamma^{(1)}$	$\Gamma^{(2)}\times\Gamma^{(2)}$	$\Gamma^{(3)}\times\Gamma^{(3)}$	$\Gamma^{(4)}\times\Gamma^{(4)}$	$\Gamma^{(5)}\times\Gamma^{(5)}$	$\Gamma^{(6)}\times\Gamma^{(6)}$	$\Gamma^{(7)}\times\Gamma^{(7)}$
$\Gamma^{(0)}\otimes\Gamma^{(0)}$	$\frac{1}{128}$	$\frac{1}{128}$	$-\frac{1}{256}$	$-\frac{1}{768}$	$\frac{1}{3072}$	$\frac{1}{15360}$	$-\frac{1}{92160}$	$-\frac{1}{645120}$
$\Gamma^{(1)}\otimes\Gamma^{(1)}$	$\frac{7}{64}$	$-\frac{3}{32}$	$-\frac{5}{128}$	$\frac{1}{96}$	$\frac{1}{512}$	$-\frac{1}{3840}$	$-\frac{1}{46080}$	0
$\Gamma^{(2)}\otimes\Gamma^{(2)}$	$-\frac{91}{64}$	$-\frac{65}{64}$	$\frac{43}{128}$	$\frac{25}{384}$	$-\frac{11}{1536}$	$-\frac{1}{7680}$	$-\frac{1}{9216}$	$-\frac{1}{46080}$
$\Gamma^{(3)}\otimes\Gamma^{(3)}$	$-\frac{273}{16}$	$\frac{39}{4}$	$\frac{75}{32}$	$-\frac{1}{4}$	$\frac{1}{128}$	$-\frac{1}{160}$	$-\frac{1}{1280}$	0
$\Gamma^{(4)}\otimes\Gamma^{(4)}$	$\frac{3003}{16}$	$\frac{1287}{16}$	$-\frac{363}{32}$	$\frac{11}{32}$	$-\frac{39}{128}$	$-\frac{19}{640}$	$-\frac{3}{1280}$	$-\frac{1}{1280}$
$\Gamma^{(5)}\otimes\Gamma^{(5)}$	$\frac{15015}{8}$	$-\frac{2145}{4}$	$-\frac{165}{16}$	$-\frac{55}{4}$	$-\frac{95}{64}$	$-\frac{5}{32}$	$-\frac{5}{128}$	0
$\Gamma^{(6)}\otimes\Gamma^{(6)}$	$-\frac{135135}{8}$	$-\frac{19305}{8}$	$-\frac{7425}{16}$	$-\frac{1485}{16}$	$-\frac{405}{64}$	$-\frac{135}{64}$	$-\frac{5}{128}$	$-\frac{5}{128}$
$\Gamma^{(7)}\otimes\Gamma^{(7)}$	-135135	0	$-\frac{10395}{2}$	0	$-\frac{945}{8}$	0	$-\frac{35}{16}$	0

[1] G. 't Hooft and M. Veltman, Nucl. Phys. B **44**, 189 (1972).
[2] G. Leibbrandt, Rev. Mod. Phys. **47**, 849 (1975).
[3] J. S. R. Chisolm, Nuovo Cimento **30**, 426 (1963); Thesis, Cambridge (1952).
[4] J. Kahane, J. Math. Phys. **9**, 1732 (1968).
[5] H. Weyl and R. Brauer, Am. J. Math. **57**, 425 (1935).
[6] D. A. Akyeampong and R. Delbourgo, Nuovo Cimento A **17**, 578 (1973).
[7] S. Bedding, private communication.

On the structure of phase space

M. Moshinsky[a] and T. H. Seligman
Instituto de Física, UNAM, Apdo. Postal 20-364, México 20, D. F., Mexico

(Received 15 July 1980; accepted for publication 16 January 1981)

We consider phase space as the carrier space of canonical transformations and see that this implies, for nonbijective ones, a much subtler structure than the one commonly assumed. Discussing only problems of one degree of freedom, i.e., a phase plane, we are able to clarify this structure by analogy with the better-known situation of conformal mapping in the complex plane. Apart from the usual Riemann sheet concept, an alternate method is developed that involves the irreducible representations of the ambiguity group, i.e., the group of transformations that connects all points that are mapped on a single point by the conformal or canonical transformation. The algebra of variables then becomes a matrix algebra for which bijectiveness is retrieved.

PACS numbers: 02.10.Sp, 03.20. + i

1. INTRODUCTION

One of the fascinating aspects of mathematics and mathematical physics is the introduction of new concepts which initially have a simple formulation, but later become more and more subtle as their deeper meaning is understood better. Thus, for example, the concept of the complex plane envisaged at the beginning of the 19th century in a simple fashion,[1] acquires, through the work of Riemann[2] and later of Klein and Weyl,[3] a great subtlety. One can say that at the present time we have a complex manifold rather than plane, whose structure should allow us to carry out arbitrary conformal transformations in a bijective (i.e., one-to-one onto) fashion.

The above manifolds have been extensively discussed leading to what are now known as Riemann surfaces.[3] Yet a similar problem in phase space—or more correctly, in the phase plane for problems of one degree of freedom—has not received comparable attention. We can, after all, discuss in a somewhat parallel fashion[4–7] the properties of the complex and phase planes as carriers, respectively, of conformal and canonical transformations. Thus the phase plane—and, for more degrees of freedom, the phase space—becomes a manifold with properties that resemble those of a Riemann surface,[4–7] and thus merits a more detailed analysis.

In previous references some of the analysis was carried out with the purpose of understanding better the representation of nonbijective canonical transformations in quantum mechanics. In the present paper we will be interested purely in the classical problem so as to focus more sharply on the properties of phase space. We shall in particular discuss a duality in our characterization of observables in the phase plane: Under nonbijective canonical transformations we can either introduce many-sheeted structures and define the values of the observables in each sheet or we can keep the idea of a single plane but give matrix form to the observables in such a way that they continue to form an algebra. We shall illustrate first these two points of view in relation with conformal transformations, where at least one of them is very familiar.

[a]Member of the Instituto Nacional de Investigaciones Nucleares and El Colegio Nacional.

2. THE STRUCTURE OF THE COMPLEX PLANE

In this section we intend to discuss two very simple conformal transformations,

$$\bar{z} = z^{\kappa}, \quad \kappa \text{ integer}, \tag{2.1}$$

$$\bar{z} = \exp z, \tag{2.2}$$

first in the standard Riemann surface picture and then in a matrix form acting on a basis characterized by the irreducible representations (irreps) of the groups that connect all points z mapped on a single $\bar{z}$. For the conformal transformation (2.1) this group is the cyclic one $\mathscr{C}_{\kappa}$ defined by the operations

$$\mathscr{C}_{\kappa}: z \to z\exp(i2\pi r/\kappa), \quad r = 0,1,\cdots,\kappa - 1, \tag{2.3}$$

while for (2.2) it is the translation group

$$\mathscr{T}: z \to z + i2\pi m, \quad m = 0, \pm 1, \pm 2, \cdots. \tag{2.4}$$

A. The Riemann surface picture associated with the conformal transformations

The conformal transformations (2.1), (2.2) imply no problem if we start with an *entire* function $F(\bar{z})$ and want to express it as an *entire* function of z. We have then

$$f(z) = F(z^{\kappa}), \quad f(z) = F(e^{z}), \tag{2.5}$$

respectively. On the other hand if we start with an entire function

$$f(z) = \sum_{\nu=0}^{\infty} a_{\nu} z^{\nu}, \tag{2.6}$$

we do need to be careful in determining the corresponding function $F(\bar{z})$. To begin with we notice that for the conformal transformation (2.1) the sector of angle $2\pi/\kappa$ in the z plane maps on the full $\bar{z}$ plane as indicated in Fig. 1. To have a bijective mapping we must then introduce a κ-sheeted Riemann surface for the $\bar{z}$ plane, joined in the standard fashion along the heavy line on the positive real axis in this plane. Thus when z is in the first sector it corresponds to $\bar{z}$, but in the second sector it corresponds to $\bar{z}\exp(i2\pi)$, in the third to $\bar{z}\exp(i4\pi)$, etc. Clearly then

$$z = \bar{z}^{1/\kappa}\exp(i2\pi r/\kappa), \quad r = 0,1,\cdots,\kappa - 1, \tag{2.7}$$

when z is in the $(r + 1)$th sector.

 0022-2488/81/071338-07$1.00

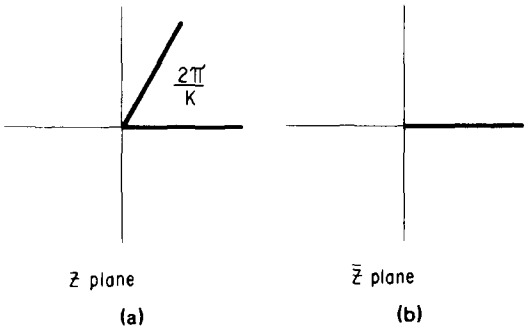


FIG. 1. A sector $2\pi/\kappa$ in the z plane mapped on the full $\bar{z}$ plane when $\bar{z} = \bar{z}^{\kappa}$. The cut connecting the different sheets is marked by the heavy line in the $\bar{z}$ plane.

Turning our attention now to the conformal transformation (2.2), i.e., $\bar{z} = \exp z$, we note that the strip between $-\pi \leqslant \mathrm{Im} z < \pi$ maps on the full $\bar{z}$ plane. As for $\mathrm{Im} z = \pm\pi$, the $\bar{z}$ is real and negative and the cut goes along the real $\bar{z}$ axis from 0 to $-\infty$ as indicated in Figs. 2(a) and 2(b). The $\bar{z}$ manifold then has an infinite number of sheets joined along the heavy line of Fig. 2(b).

If we are in the central strip for z we can write $z = \ln\bar{z}$ but in the other strips, associated with the displacements $2m\pi$, $m = \pm 1, \pm 2, \cdots$, along the imaginary axis, we have then

$$z = \ln[\bar{z}\exp(i2\pi m)] = \ln\bar{z} + i2\pi m. \tag{2.8}$$

B. Matrix representation in the $\bar{z}$ plane associated with conformal transformations

We now proceed to show different ways in which relations such as (2.7) and (2.8) can establish correspondences between functions of z and matrices of functions of $\bar{z}$. For this purpose we note that entire functions of z (the only ones of interest to us here) can be interpreted in two ways: as basis functions, in which case they will be denoted by Greek letters such as $\varphi(z)$, $\psi(z)$, and as operators acting by multiplication on the basis functions when they will be denoted by Latin letters such as $f(z)$, $g(z)$. Thus when $f(z)$ acts on $\varphi(z)$ by multiplication we get a new basis function $\psi(z)$, i.e.,

$$\psi(z) = f(z)\varphi(z). \tag{2.9}$$

The basis functions $\varphi(z) = \Sigma_\nu \alpha_\nu z^\nu$, can be decomposed into their irreducible parts with respect to the group of transformations connecting all points z mapped by the conformal transformation on a single $\bar{z}$. We implement this decomposition first for $\bar{z} = z^\kappa$, κ integer, where the group is the cyclic one $\mathscr{C}_\kappa$ of (2.3). The entire function $\varphi(z)$ can then be decomposed into its irreducible parts $\varphi^\lambda(z)$, $\lambda = 0,1,2,\cdots, \kappa - 1$ with respect to $\mathscr{C}_\kappa$ as[4,8]

$$\varphi(z) = \sum_{\lambda=0}^{\kappa-1} \varphi^\lambda(z), \tag{2.10a}$$

$$\varphi^\lambda(z) = \kappa^{-1} \sum_{r=0}^{\kappa-1} \exp(i2\pi\lambda r/\kappa)\varphi\,[z\exp(-i2\pi r/\kappa)], \tag{2.10b}$$

where $\exp(i2\pi\lambda r/\kappa)$, $r = 0,1,\cdots, \kappa - 1$ are the irreducible representations characterized by $\lambda = 0,1,\cdots, \kappa - 1$ of the operations in the abelian cyclic group $\mathscr{C}_\kappa$.

Clearly then if we write $\nu = n\kappa + \lambda$; $\nu = 0,1,2,\cdots$; $\lambda = 0,1,\cdots, \kappa - 1$; $n = 0,1,2,\cdots$, the $\varphi^\lambda(z)$ become

$$\varphi^\lambda(z) = z^\lambda \Phi^\lambda(z^\kappa), \tag{2.11a}$$

$$\Phi^\lambda(z^\kappa) = \sum_{n=0}^{\infty} \alpha_{n\kappa+\lambda}(z^\kappa)^n, \tag{2.11b}$$

and thus from (2.7) the basis function $\varphi(z)$ corresponds to a vector function in the $\bar{z}$ plane

$$\varphi(z) = \sum_{\lambda=0}^{\kappa-1} z^\lambda \Phi^\lambda(z^\kappa) \leftrightarrow \begin{bmatrix} \sum_\lambda \bar{z}^{\lambda/\kappa}\Phi^\lambda(\bar{z}) \\ \sum_\lambda \bar{z}^{\lambda/\kappa}\exp(2\pi i\lambda/\kappa)\Phi^\lambda(\bar{z}) \\ \cdots\ \cdots\ \cdots\ \cdots \\ \sum_\lambda \bar{z}^{\lambda/\kappa}\exp[2\pi i\lambda(\kappa-1)/\kappa]\Phi^\lambda(\bar{z}) \end{bmatrix} = V\Phi, \tag{2.12}$$

in which the components of the vector represent the corresponding function in the different sheets of the Riemann surface associated with the $\bar{z}$ manifold. We can also write the vector on the right-hand side (rhs) of (2.12) as a product of a $(\kappa\times\kappa)$ matrix

$$V = \|\exp(i2\pi\lambda\lambda'/\kappa)\bar{z}^{\lambda'/\kappa}\|, \quad \lambda,\lambda' = 0,1,\cdots,\kappa - 1, \tag{2.13}$$

(where λ, λ' are respectively the row and column indices) with the vector Φ of components $\Phi^\lambda(\bar{z})$, $\lambda = 0, 1\cdots, \kappa - 1$, i.e.,

$$\Phi = \begin{bmatrix} \Phi^0(\bar{z}) \\ \Phi^1(\bar{z}) \\ \cdots \\ \Phi^{\kappa-1}(\bar{z}) \end{bmatrix}. \tag{2.14}$$

We shall now proceed to derive the matrix associated with an operator $f(z)$ of the form (2.6) both when we use the basis on the rhs of (2.12) as well as when we take (2.14). We note that it would be enough to obtain the matrix representation of z, as then we can get that of any power or linear combination of powers. For the basis on the rhs of (2.12) it is clear from (2.7) that z corresponds to the diagonal matrix

$$z \leftrightarrow \begin{bmatrix} \bar{z}^{1/\kappa} & & \\ & \bar{z}^{1/\kappa}\exp(i2\pi/\kappa) & \\ & & \ddots \\ & & \bar{z}^{1/\kappa}\exp[i2\pi(\kappa-1)/\kappa] \end{bmatrix} \equiv D. \tag{2.15}$$

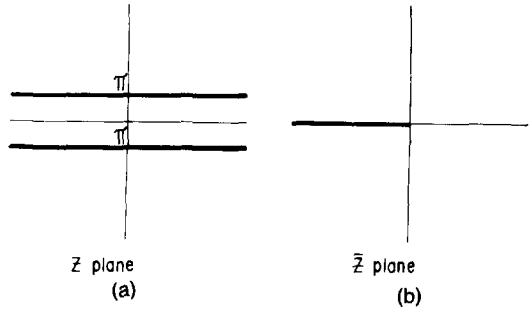


FIG. 2. A strip $-\pi \leqslant z \leqslant \pi$ in the z plane mapped on the full $\bar{z}$ plane when $\bar{z} = \exp z$. The cut connecting the different sheets is marked by the heavy line in the $\bar{z}$ plane.

On the other hand, for the basis Φ of (2.14) we obviously have

$$z \leftrightarrow V^{-1} D V, \tag{2.16}$$

where the reciprocal of the matrix V is given by

$$V^{-1} = \kappa^{-1} \| \bar{z}^{-\lambda/\kappa} \exp(-i2\pi\lambda\lambda'/\kappa) \|; \quad \lambda, \lambda' = 0,1 \cdots, \kappa - 1, \tag{2.17}$$

as can be easily checked. Carrying out the operations in (2.16) we get then finally that

$$z \leftrightarrow \begin{bmatrix} 0 & 0 & \cdots & 0 & \bar{z} \\ 1 & 0 & \cdots & 0 & 0 \\ \cdot\cdot & & \cdots & & \cdot\cdot \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{2.18}$$

This result can also be obtained if we consider

$$\begin{aligned} \psi(z) = z\varphi(z) &= \bar{z}^{1/\kappa} \sum_{\lambda=0}^{\kappa-1} \bar{z}^{\lambda/\kappa} \Phi^{\lambda}(\bar{z}) \\ &= \sum_{\lambda=0}^{\kappa-2} \bar{z}^{(\lambda+1)/\kappa} \Phi^{\lambda}(\bar{z}) + \bar{z}\Phi^{\kappa-1}(\bar{z}) \\ &= \bar{z}\Phi^{\kappa-1}(\bar{z}) + \sum_{\lambda=1}^{\kappa-1} \bar{z}^{\lambda/\kappa} \Phi^{\lambda-1}(\bar{z}) = \sum_{\lambda=0}^{\kappa-1} \bar{z}^{\lambda/\kappa} \Psi^{\lambda}(\bar{z}), \end{aligned} \tag{2.19}$$

where $\{\Psi^{\lambda}(\bar{z}), \lambda = 0,1,\cdots, \kappa - 1\}$ is the vector in the $\bar{z}$ plane associated with $\psi(z)$.

Clearly the basis (2.14) is more convenient than (2.12) as entire functions of z will be mapped on entire $\kappa \times \kappa$ matrix functions of $\bar{z}$. Note in particular that from (2.19), $z^{\kappa} \leftrightarrow \bar{z} I$ (where I is the unit $\kappa \times \kappa$ matrix) as we should expect.

We now turn our attention to the conformal transformation $\bar{z} = \exp(z)$. In analogy to (2.12) and using (2.8) we can establish the correlation of basis functions in the z and $\bar{z}$ manifolds as

$$\varphi(z) \leftrightarrow \begin{bmatrix} \cdots \\ \varphi(\ln\bar{z} + 4\pi i) \\ \varphi(\ln\bar{z} + 2\pi i) \\ \varphi(\ln\bar{z}) \\ \varphi(\ln\bar{z} - 2\pi i) \\ \varphi(\ln\bar{z} - 4\pi i) \\ \cdots \end{bmatrix}, \tag{2.20}$$

where in the latter we have an infinite-dimensional vector whose components are numbered by the integers $m = 0$, ± 1, ± 2, $\cdots$, associated with the different sheets. Clearly then in the basis (2.20) we have the correspondence of operators

$$z \leftrightarrow \begin{bmatrix} \ddots & & & & & \\ & \ln\bar{z} + 4\pi i & & & & \\ & & \ln\bar{z} + 2\pi i & & & \\ & & & \ln\bar{z} & & \\ & & & & \ln\bar{z} - 2\pi i & \\ & & & & & \ln\bar{z} - 4\pi i \\ & & & & & & \ddots \end{bmatrix}, \tag{2.21}$$

where the rhs is a diagonal matrix with the indicated values.

It is now interesting to get the correspondence of z in the $\bar{z}$ plane when we choose as the basis the set of irreducible components of $\varphi(z)$ for the translation group $\mathcal{T}$ whose elements T_m imply the operation

$$T_m \varphi(z) = \varphi(z + 2im\pi). \tag{2.22a}$$

We note first that the irreducible representations of the translation group $\mathcal{T}$ are one-dimensional and of the form[9]

$$T_m \leftrightarrow \exp(i2\pi\lambda m), \quad 0 \leqslant \lambda < 1, \tag{2.22b}$$

with λ characterizing these irreps. Thus the irreducible components $\varphi^{\lambda}(z)$ of the $\varphi(z)$ are given by[8]

$$\varphi^{\lambda}(z) = \sum_{n=-\infty}^{\infty} \exp(i2\pi\lambda n) \varphi(z - i2n\pi), \tag{2.23}$$

which clearly have the property

$$T_m \varphi^{\lambda}(z) = \exp(i2\pi\lambda m) \varphi^{\lambda}(z). \tag{2.24}$$

Furthermore, we could define

$$\Phi^{\lambda} \equiv \exp(-\lambda z) \varphi^{\lambda}(z), \tag{2.25}$$

which implies from (2.24) that $T_m \Phi^{\lambda} = \Phi^{\lambda}$, i.e., it is invariant under translations by $2\pi i m$ and thus can be thought of as a function of e^z, i.e.,

$$\varphi^{\lambda}(z) = e^{\lambda z} \Phi^{\lambda}(e^z). \tag{2.26}$$

We can now express $\varphi(z)$ as[8]

$$\varphi(z) = \int_0^1 \varphi^{\lambda}(z) d\lambda = \int_0^1 e^{\lambda z} \Phi^{\lambda}(e^z) d\lambda = \int_0^1 \bar{z}^{\lambda} \Phi^{\lambda}(\bar{z}) d\lambda, \tag{2.27}$$

and more generally in the different strips of the z plane

$$\varphi(z + i2m\pi) = \int_0^1 \bar{z}^{\lambda} \exp(i2\pi\lambda m) \Phi^{\lambda}(\bar{z}) d\lambda. \tag{2.28}$$

The set of functions

$$\{\Phi^{\lambda}(\bar{z}), \quad 0 \leqslant \lambda < 1\} \tag{2.29}$$

can be considered as a vector of functions of the $\bar{z}$ variable, with a continuous index λ in the interval $0 \leqslant \lambda < 1$ rather than the discrete and finite one appearing in (2.14) for the transformation $\bar{z} = \bar{z}^{\kappa}$.

We now look for the product

$$\begin{aligned} z\varphi(z) &= \int_0^1 z e^{\lambda z} \Phi^{\lambda}(e^z) d\lambda \\ &= \int_0^1 \left[\frac{\partial}{\partial\lambda} e^{\lambda z} \right] \Phi^{\lambda}(e^z) d\lambda \\ &= \int_0^1 \bar{z}^{\lambda} \left[-\frac{\partial}{\partial\lambda} \Phi^{\lambda}(\bar{z}) \right] d\lambda, \end{aligned} \tag{2.30}$$

where in the integration by parts we made use of the fact that from (2.23)

$$\lim_{\lambda \to 1} \varphi^{\lambda}(z) = \varphi^{0}(z). \tag{2.31}$$

If we denote by $\{\Psi^{\lambda}(\bar{z}), 0 \leqslant \lambda < 1\}$ the vector in the $\bar{z}$ plane associated with the function $\psi(z) \equiv z\varphi(z)$ in the z plane, we conclude from (2.30) that

$$\Psi^{\lambda}(\bar{z}) = -\frac{\partial}{\partial\lambda} \Phi^{\lambda}(\bar{z}) = \int_0^1 \left[-\frac{\partial}{\partial\lambda} \delta(\lambda - \lambda') \Phi^{\lambda'}(\bar{z}) \right] d\lambda', \tag{2.32}$$

and we get the correspondence

$$z \leftrightarrow \| -\frac{\partial}{\partial\lambda} \delta(\lambda - \lambda') \|, \tag{2.33}$$

where the rhs is a matrix in the continuous indices

$0 \leqslant \lambda, \lambda' < 1$.

The representation (2.33) of the variable z, on the basis associated with the irreps of $\mathscr{T}$, seems at first sight strange; yet it has all the required properties. For example, for any power z^κ, κ integer, an analysis similar to that carried for z gives

$$z^\kappa \leftrightarrow \| (-1)^\kappa \frac{\partial^\kappa}{\partial \lambda^\kappa} \delta(\lambda - \lambda') \|, \tag{2.34}$$

and thus we have also

$$\exp z \leftrightarrow \| \exp(-\partial/\partial\lambda)\delta(\lambda - \lambda') \|. \tag{2.35}$$

If we apply the rhs of the latter to $\Phi^\lambda(\bar{z})$ we get a new function $\chi(z)$ given by

$$\begin{aligned} \psi(z) &= \int_0^1 \bar{z}^\lambda [\exp(-\partial/\partial\lambda)\Phi^\lambda(\bar{z})] d\lambda \\ &= \int_0^1 \bar{z}^\lambda \Phi^{\lambda-1}(\bar{z}) d\lambda = \bar{z} \int_{-1}^0 \bar{z}^{\lambda'} \Phi^{\lambda'}(\bar{z}) d\lambda' \\ &= e^z \int_{-1}^0 \varphi^{\lambda'}(z) d\lambda' = e^z \int_0^1 \varphi^\lambda(z) d\lambda = e^z \varphi(z), \end{aligned} \tag{2.36}$$

where we made use of the fact that from (2.23) $\varphi^{\lambda+1}(z) = \varphi^\lambda(z)$. Thus we see that we have the relation

$$\| \exp(-\partial/\partial\lambda)\delta(\lambda - \lambda') \| = \bar{z} \| \delta(\lambda - \lambda') \|, \tag{2.37}$$

as we should expect from the fact that for the other representation (2.21) of z the exponentiation of the rhs gives $\bar{z}$ multiplied by the unit matrix.

For the sake of completeness we give also the representation of the differential operator d/dz when we go to the $\bar{z}$ plane through the conformal transformations $\bar{z} = z^\kappa, \bar{z} = \exp z$. In the first case we have to consider

$$\begin{aligned} \psi(z) \equiv \frac{d}{dz}\varphi(z) &= \frac{d}{d\bar{z}^{1/\kappa}} \sum_\lambda \bar{z}^{\lambda/\kappa} \Phi^\lambda(\bar{z}) = \kappa \bar{z}^{(\kappa-1)/\kappa} \\ &\times \left[\frac{d}{d\bar{z}} \Phi^0(\bar{z}) + \sum_{\lambda=1}^{\kappa-1} \bar{z}^{(\lambda/\kappa)-1} \left(\frac{\lambda}{\kappa} + \bar{z} \frac{d}{d\bar{z}} \right) \Phi^\lambda(\bar{z}) \right] \\ &= \left[\sum_{\lambda=0}^{\kappa-2} \bar{z}^{\lambda/\kappa} \left(\lambda + 1 + \kappa \bar{z} \frac{d}{d\bar{z}} \right) \Phi^\lambda(\bar{z}) \right] \\ &+ \bar{z}^{(\kappa-1)/\kappa} \kappa \frac{d}{d\bar{z}} \Phi^0(\bar{z}) = \sum_{\lambda=0}^{\kappa-1} \bar{z}^{\lambda/\kappa} \Psi^\lambda(\bar{z}), \end{aligned} \tag{2.38}$$

and thus the vectors

$$\{\Phi^\lambda(\bar{z}), \lambda = 0,1,\cdots,\kappa-1\}, \{\Psi^\lambda(\bar{z}), \lambda = 0,1,\cdots,\kappa-1\},$$

are correlated by the matrix operator

$$\frac{d}{dz} \leftrightarrow \begin{bmatrix} 0 & 1 + \kappa\bar{z}\frac{d}{d\bar{z}} & 0 & \cdots & 0 \\ 0 & 0 & 2 + \kappa\bar{z}\frac{d}{d\bar{z}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \kappa - 1 + \kappa\bar{z}\frac{d}{d\bar{z}} \\ \kappa\frac{d}{d\bar{z}} & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.39}$$

For the conformal transformation $\bar{z} = \exp z$, we have

$$\begin{aligned} \psi(z) \equiv \frac{d}{dz}\varphi(z) &= \bar{z}\frac{d}{d\bar{z}} \int_0^1 \bar{z}^\lambda \Phi^\lambda(\bar{z}) d\lambda \\ &= \int_0^1 \bar{z}^\lambda \left[\left(\lambda + \bar{z}\frac{d}{d\bar{z}} \right) \Phi^\lambda(\bar{z}) \right] d\lambda = \int_0^1 \bar{z}^\lambda \Psi^\lambda(\bar{z}) d\lambda, \end{aligned} \tag{2.40}$$

so that the vectors $\{\Phi^\lambda(\bar{z}), 0 \leqslant \lambda < 1\}$, $\{\Psi^\lambda(\bar{z}), 0 \leqslant \lambda < 1\}$, are related by the matrix operator

$$\frac{d}{dz} \leftrightarrow \| \left(\bar{z}\frac{d}{d\bar{z}} + \lambda \right) \delta(\lambda - \lambda') \|. \tag{2.41}$$

As the commutator of d/dz and z is 1, we must have the unit matrix for the corresponding operators in the $\bar{z}$ plane. This is easily checked using (2.19), (2.39) when $\bar{z} = z^\kappa$, and (2.33), (2.41) when $\bar{z} = \exp z$.

We now proceed to apply a similar analysis in the discussion of the structure of phase space.

3. THE STRUCTURE OF PHASE SPACE

If we have a phase plane whose coordinate and momenta we designate by (q,p), we can consider in it a canonical transformation to $(\bar{q},\bar{p})$ which are functions of q, p such that

$$\{\bar{q},\bar{p}\}_{q,p} \equiv \frac{\partial\bar{q}}{\partial q}\frac{\partial\bar{p}}{\partial p} - \frac{\partial\bar{q}}{\partial p}\frac{\partial\bar{p}}{\partial q} = 1. \tag{3.1}$$

As in the case of the complex plane, any observables, i.e., functions of q, p, can be interpreted in two ways: as basis functions, in which case they will be denoted by Greek letters such as $\varphi(q,p)$, $\psi(q,p)$, or as operators acting by multiplication on the basis function when they will be denoted by Latin letters such as $f(q,p)$. Thus when $f(q,p)$ acts on $\varphi(q,p)$ by multiplication we get a new basis function $\psi(q,p)$, i.e.,

$$\psi(q,p) = f(q,p)\varphi(q,p). \tag{3.2}$$

If we now have a canonical transformation taking us from (q,p) to $(\bar{q},\bar{p})$ we can ask how does a function $f(q,p)$ transform to the new phase space. If the transformation is bijective as, for example, the linear one

$$\bar{q} = aq + bp, \quad \bar{p} = cq + dp, \quad ad - bc = 1, \tag{3.3}$$

then there is no problem because inverting the transformation we see that

$$f(q,p) = f(d\bar{q} - b\bar{p}, -c\bar{q} + a\bar{p}) \equiv F(\bar{q},\bar{p}). \tag{3.4}$$

If, on the other hand, the canonical transformation is nonbijective then we face problems similar to those discussed in the previous section.[4-7] If—as will be the case in the examples to be discussed below—we have several points (q,p) that map on a single $(\bar{q},\bar{p})$, we can investigate the group of canonical transformations relating the (q,p) points. In previous references[4-7] this group has been given the name of ambiguity group and it corresponds to the $\mathscr{C}_\kappa, \mathscr{T}$ discussed in the previous section for conformal transformations. We can then decompose the basis functions $\varphi(q,p)$ into their irreducible parts according to the ambiguity group and these will bring in indices of the λ type of the previous section, associated with the irreps, which for the phase space problem have been given the name of ambiguity spin. Using then the $\varphi^\lambda[q(\bar{q},\bar{p}),p(\bar{q},\bar{p})]$, with the λ's taking an appropriate range of values, we can associate with $f(q,p)$ a matrix $\mathbf{F}(\bar{q},\bar{p})$

in a similar way as was done for the conformal transformation.

We proceed to implement this program for two simple nonbijective canonical transformations which have many points in common with the conformal transformations $\bar{z} = z^\kappa$, $\bar{z} = \exp z$.

A. Canonical transformations and their ambiguity groups

We shall consider here only two nonbijective canonical transformations. One of them will take us from an harmonic oscillator of unit frequency in the variables $(\bar{q},\bar{p})$ to one in the variables (q,p) in which the frequency is $(1/\kappa)$ with κ integer. The other will relate $(\bar{q},\bar{p})$ with the action and angle variables (q,p) of an harmonic oscillator.

Taking units in which the mass, frequency of the oscillator, and a constant of dimension of action are 1, we can write for the first case the Hamiltonians in the variables $(\bar{q},\bar{p})$ and (q,p) as

$$\tfrac{1}{2}(\bar{p}^2 + \bar{q}^2), \quad \tfrac{1}{2}(p^2 + \kappa^{-2}q^2). \tag{3.5a,b}$$

In (3.5b) we can carry out the point transformation $q \to \kappa q$, $p \to \kappa^{-1}p$, and thus the first canonical transformation is defined by the implicit equations[4]

$$\tfrac{1}{2}(\bar{p}^2 + \bar{q}^2) = (1/2\kappa)\,(p^2 + q^2), \tag{3.6a}$$

$$\arctan(\bar{p}/\bar{q}) = \kappa \arctan(p/q). \tag{3.6b}$$

Introducing the observables η, ξ by the definition

$$\eta = (1/\sqrt{2})\,(q - ip), \quad \xi = (1/\sqrt{2})\,(q + ip), \tag{3.7a,b}$$

and similar expression for $\bar{\eta}$, $\bar{\xi}$, we see that (3.6) imply the relations

$$\bar{\eta} = \frac{\eta^\kappa}{\kappa^{1/2}(\eta\xi)^{(\kappa-1)/2}}, \quad \bar{\xi} = \frac{\xi^\kappa}{\kappa^{1/2}(\eta\xi)^{(\kappa-1)/2}}, \tag{3.8}$$

as

$$\eta\xi = \tfrac{1}{2}(p^2 + q^2) = \kappa\bar{\eta}\bar{\xi}, \tag{3.9}$$

From (3.8) the explicit form of the first canonical transformation is

$$\bar{q} = \kappa^{-1/2}(q^2 + p^2)^{(1-\kappa)/2}\sum_s \binom{\kappa}{2s} q^{\kappa-2s}(-1)^s p^{2s};$$

$$\bar{p} = \kappa^{-1/2}(q^2 + p^2)^{(1-\kappa)/2}\sum_s \binom{\kappa}{2s+1} q^{\kappa-2s-1}(-1)^s p^{2s+1}. \tag{3.10a,b}$$

Turning our attention now to the second canonical transformation, it implies that[5]

$$\bar{q} = (2|q|)^{1/2}\cos(qp/|q|), \quad \bar{p} = (2|q|)^{1/2}\sin(qp/|q|), \tag{3.11a,b}$$

as from this it follows that

$$|q| = \tfrac{1}{2}(\bar{p}^2 + \bar{q}^2) = \bar{\eta}\bar{\xi}, \tag{3.12a}$$

$$(qp/|q|) = \arctan(\bar{p}/\bar{q}) = i\ln(\bar{\eta}/\bar{\xi})^{1/2}, \tag{3.12b}$$

and thus, in the present units, $|q|$, $qp/|q|$ are the action and angle variables of the harmonic oscillator of coordinate $\bar{q}$ and momenta $\bar{p}$. Note that the rhs of (3.12a) is positive definite and thus we have equated it to $|q|$ rather than to q. As $(qp/|q|)$ is the canonically conjugate[5] variable to $|q|$, we have taken it as equal to the rhs of (3.12b).

Using the observables $\bar{\eta}$, $\bar{\xi}$ we can also express (3.11) as

$$\bar{\eta} = (|q|)^{1/2}\exp(-iqp/|q|),$$

$$\bar{\xi} = (|q|)^{1/2}\exp(iqp/|q|). \tag{3.13a,b}$$

We now turn our attention to the ambiguity groups associated with the two canonical transformations discussed in this section. For the first one we see from (3.8) that if we carry the transformation

$$\eta \to \eta\exp(i2\pi r/\kappa), \quad \xi \to \xi\exp(-i2\pi r/\kappa),$$

$$r = 0,1,\cdots,\kappa - 1, \tag{3.14}$$

this does not affect the values of $\bar{\eta}$, $\bar{\xi}$. Thus the ambiguity group is again $\mathscr{C}_\kappa$ as was the case for the conformal transformation $\bar{z} = z^\kappa$. The irreps will again be characterized by the index $\lambda = 0,1,\cdots,\kappa - 1$.

For the second canonical transformation we see from (3.13) that the points q, p connected by

$$q \to -q, \; p \to -p, \quad q \to q, \; p \to p + 2m\pi,$$

$$m = 0, \pm 1, \pm 2, \cdots, \tag{3.15}$$

map on the same $\bar{\eta}$, $\bar{\xi}$ and thus also on the same $\bar{q}$, $\bar{p}$. The ambiguity group is then the semidirect product[5] $\mathscr{T} \wedge \mathscr{I}$ of the translation T_m in the momentum variable by $2\pi m$, $m = 0, \pm 1, \pm 2, \cdots$, and of the inversion I in phase space for which $(q,p) \to (-q, -p)$. As is well-known[5] the irreps of this group are two-dimensional, i.e.,

$$T_m \to \begin{bmatrix} \exp(i2\pi\lambda m) & 0 \\ 0 & \exp(-i2\pi\lambda m) \end{bmatrix}, \quad I \to \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{3.16}$$

and characterized by the real index λ in the interval $0 \leqslant \lambda < 1$. In view of the two-dimensional form of the irreps we need besides λ another index, which we denote by $\sigma = \pm 1$, whose two values indicate the row of the representation. Thus the irreducible basis of $\mathscr{T} \wedge \mathscr{I}$ will be a vector characterized by the continuous index $0 \leqslant \lambda < 1$ and discrete one $\sigma' = \pm 1$.

The mapping of the phase spaces associated with the canonical transformations (3.10) and (3.11) is given respectively in Figs. 3 and 4. In Fig. 3(a) we see that the sector $(2\pi/\kappa)$ in the (q,p) plane is mapped on the full $(\bar{q},\bar{p})$ plane of Fig. 3(b) and thus the latter must have κ sheets joined in the cut at the heavy line on the real axis. In Fig. 4(a) we see that the strip marked in the (q,p) plane maps on the full $(\bar{q},\bar{p})$ plane of Fig. 4(b). The strips on the right and left hand side of $q = 0$ map

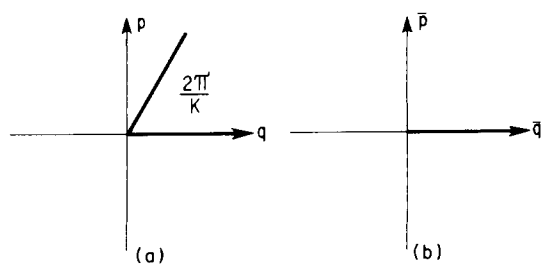


FIG. 3. A sector $2\pi/\kappa$ in the (q,p) phase plane is mapped on the full $(\bar{q},\bar{p})$ plane by the canonical transformation (3.10). The cut connecting the different sheets is the $(\bar{q},\bar{p})$ plane is marked by the heavy line.

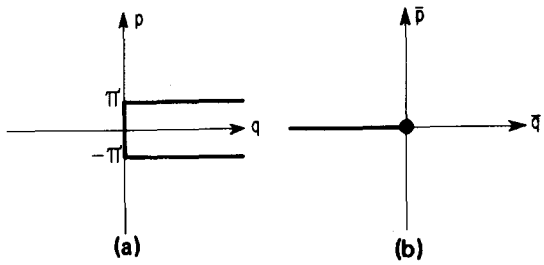


FIG. 4. A strip in the (q,p) phase plane is mapped on the full $(\bar{q},\bar{p})$ plane by the canonical transformation (3.11). The cut connecting the different sheets in the $(\bar{q},\bar{p})$ plane is marked by the heavy line and dot.

two sets of infinitely sheeted manifolds joined along the cut marked by the heavy line along the real axis. The line $q=0$ is, from (3.12a), mapped on the point $\bar{q}=0, \bar{p}=0$, and thus the two manifolds are joined at this point as indicated by the dot there.

We consider now the matrix representation of the observables $f(q,p)$ in the $(\bar{q},\bar{p})$ plane. We shall bypass the Riemann surface picture, as its realization in the phase plane follows straightforwardly from the discussion in the previous section, and work directly on the irreducible basis equivalent to (2.14) and (2.29) for the conformal transformation problem.

B. Matrix representations in the $(\bar{q},\bar{p})$ plane of the observables $f(q,p)$

We start by discussing the representations for the canonical transformation (3.8). We consider as basis only entire functions of q, p which, from (3.7), imply that they are also entire functions of η, ξ, i.e.,

$$\begin{aligned}\varphi(\eta,\xi) &= \sum_{\nu,\nu'=0}^{\infty} \alpha_{\nu\nu'}\eta^{\nu}\xi^{\nu'} \\ &= \sum_{\Lambda,\Lambda'=0}^{\kappa-1}\left(\frac{\eta}{\xi}\right)^{(\Lambda-\Lambda')/2} \\ &\times\left[(\eta\xi)^{(\Lambda+\Lambda')/2}\sum_{n,n'=0}^{\infty}\alpha_{n\kappa+\Lambda,n'\kappa+\Lambda'}(\eta^{\kappa})^{n}(\xi^{\kappa})^{n'}\right] \\ &= \sum_{\lambda=0}^{\kappa-1}\left[\left(\frac{\eta}{\xi}\right)^{1/2}\right]^{\lambda}\Phi^{\lambda}(\bar{\eta},\bar{\xi}).\end{aligned} \tag{3.17}$$

The rhs of (3.17) follows if we replace $\Lambda-\Lambda'$ by λ when $\Lambda-\Lambda'\geqslant 0$ and by $\lambda-\kappa$ if $\Lambda-\Lambda'<0$. It is clear then than the remaining function Φ^{Λ} is invariant under the group $\mathscr{C}_{\kappa}$ and thus it can be written unambiguously as a function of $\bar{\eta}$, $\bar{\xi}$.

The observable $\eta\xi$ is invariant under $\mathscr{C}_{\kappa}$ and thus we expect from (3.9) that $\eta\xi$ corresponds to $\kappa\bar{\eta}\bar{\xi}$ multiplied by the $\kappa\times\kappa$ unit matrix. We postulate a similar relation for any function of $\eta\xi$ and thus, for example, we have

$$(\eta\xi)^{1/2}\leftrightarrow(\kappa\bar{\eta}\bar{\xi})^{1/2}\begin{bmatrix}1&0&0&\cdots&0\\0&1&0&\cdots&0\\0&0&1&\cdots&0\\&&&\ddots&\\0&0&0&\cdots&1\end{bmatrix}. \tag{3.18}$$

On the other hand, for $(\eta/\xi)^{1/2}$ we can consider the function $(\eta/\xi)^{1/2}\varphi(\eta,\xi)$ and from (2.19) and (3.17) immediately conclude that

$$(\eta/\xi)^{1/2}\leftrightarrow\begin{bmatrix}0&0&\cdots&0&(\bar{\eta}/\bar{\xi})^{1/2}\\1&0&\cdots&0&0\\0&1&\cdots&0&0\\\cdot&\cdot&\cdots&\cdot&\cdot\\0&0&\cdots&1&0\end{bmatrix}. \tag{3.19}$$

Thus by multiplication of (3.18) by (3.19) or its inverse we have the correspondences

$$\eta\leftrightarrow\begin{bmatrix}0&0&\cdots&0&\kappa^{1/2}\bar{\eta}\\(\kappa\bar{\eta}\bar{\xi})^{1/2}&0&\cdots&0&0\\\cdot&\cdot&\cdots&\cdots&\cdot\\0&0&\cdots&(\kappa\bar{\eta}\bar{\xi})^{1/2}&0\end{bmatrix},$$

$$\xi\leftrightarrow\begin{bmatrix}0&(\kappa\bar{\eta}\bar{\xi})^{1/2}&\cdots&0\\\cdot&\cdot&\cdots&\cdot\\0&0&\cdots&(\kappa\bar{\eta}\bar{\xi})^{1/2}\\\kappa^{1/2}\bar{\xi}&0&\cdots&0\end{bmatrix}. \tag{3.20a,b}$$

We note that the rhs of (3.20a,b) are Hermitian conjugates to each other as we would like to have in view of the fact in the (q,p) phase space $\xi=\eta^{*}$. Furthermore, we obtain from (3.18), (3.20) the correspondences

$$\kappa^{-1/2}(\eta\xi)^{(1-\kappa)/2}\eta^{\kappa}\leftrightarrow\bar{\eta}I,\quad \kappa^{-1/2}(\eta\xi)^{(1-\kappa)/2}\xi^{\kappa}\leftrightarrow\bar{\xi}I, \tag{3.21}$$

(where I is the $\kappa\times\kappa$ unit matrix) which reflect the defining relations (3.8).

We now pass to the canonical transformation (3.11) and its inverse (3.12). We can again take $\varphi(q,p)$ and decompose it in its irreducible parts according to the $\mathscr{T}\wedge\mathscr{I}$ ambiguity group. It is simpler though to start with basis functions $\varphi[|q|,(qp/|q|)]$ which are invariant under the inversion operation $(q,p)\rightarrow(-q,-p)$. In this case we have only to consider the subgroup $\mathscr{T}$ of $\mathscr{T}\wedge\mathscr{I}$ and the analysis is similar to the one carried in the previous section for $\bar{z}=\exp z$. As in (2.23) we have the irreducible components

$$\begin{aligned}&\varphi^{\lambda}[|q|,(qp/|q|)] \\ &\quad=\sum_{m=-\infty}^{\infty}\exp(i2\pi\lambda m)\varphi[|q|,q(p-2m\pi)/|q|] \\ &\quad=\exp[i\lambda(qp/|q|)]\Phi^{\lambda}(\bar{\eta},\bar{\xi}),\end{aligned} \tag{3.22}$$

where Φ^{λ} in the right-hand side is invariant under the ambiguity group and thus can be expressed as a function of $\bar{\eta}$, $\bar{\xi}$.

In turn we have from (2.27) and (3.12) that

$$\begin{aligned}\varphi[|q|,(qp/|q|)] &= \int_0^1\exp[i\lambda(qp/|q|)]\Phi^{\lambda}(\bar{\eta},\bar{\xi})d\lambda \\ &= \int_0^1[(\bar{\xi}/\bar{\eta})^{1/2}]^{\lambda}\Phi^{\lambda}(\bar{\eta},\bar{\xi})d\lambda,\end{aligned} \tag{3.23}$$

and thus an analysis similar to (2.30)–(2.33) indicates that we have the correspondence

$$\frac{qp}{|q|}\leftrightarrow\|-\frac{1}{i}\frac{\partial}{\partial\lambda}\delta(\lambda-\lambda')\delta_{\sigma\sigma'}\|. \tag{3.24a}$$

On the other hand from the fact that $|q|$ is invariant under the ambiguity group the relation (3.12) translates into

$$|q|\leftrightarrow\bar{\eta}\bar{\xi}\,\|\delta(\lambda-\lambda')\delta_{\sigma\sigma'}\|. \tag{3.24b}$$

In both equations (3.24) $\delta_{\sigma\sigma'}$, $\sigma,\sigma'=\pm 1$ appears be-

cause $(qp/|q|)$, $|q|$ are invariant under the inversion. Had we considered, for example, q, p then the analysis in reference 5, dealing with the inversion operation I only, indicates that all we have to do is to replace $\delta_{\sigma\sigma'}$ by $\sigma\delta_{\sigma\sigma'}$ or, equivalently, by $\epsilon_{\sigma\sigma'}$ where $\epsilon_{+-} = \epsilon_{-+} = 1, \epsilon_{++} = \epsilon_{--} = 0$.

A similar analysis to the one given in the previous sections then indicates that we have also the correspondences

$$|q|^{1/2}\exp(-iqp/|q|)\leftrightarrow\bar{\eta}\|\delta(\lambda-\lambda')\delta_{\sigma\sigma'}\|, \tag{3.25a}$$

$$|q|^{1/2}\exp(iqp/|q|)\leftrightarrow\bar{\xi}\|\delta(\lambda-\lambda')\delta_{\sigma\sigma'}\|, \tag{3.25b}$$

which we expect from (3.13).

Thus we have found the matrix representations in the $(\bar{q},\bar{p})$ phase plane of relevant functions in the (q,p) plane, when these phase spaces are connected by the canonical transformations (3.8) or (3.13). Note though that these matrices do not satisfy in general the Poisson bracket relations with respect to q, p. This would be true for example in the Riemann sheet picture but as the transformation from this picture to the one of irreducible basis *is* a function of q, p, as illustrated in the conformal case by (2.13), it no longer holds in the latter basis. The only exception is for Poisson brackets involving functions invariant under the ambiguity group, such as (3.25), which are then obviously independent of the basis.

4. CONCLUSION

While the examples discussed in the previous section are very simple, it is clear that they provide a general approach to the study of the structure of phase space. What is required for a given canonical transformation is to find first its ambiguity group. We have then to decompose the functions $\varphi(q,p)$ into their irreducible parts with respect to this group and acting on the corresponding vector in the $(\bar{q},\bar{p})$ phase space, to get the matrix form of observables in the latter. Sometimes, as for example, when we go[4] from an oscillator of frequency 1 to that rational frequency (k/κ), where k,κ are relative prime integers, we require matrix observables in *both* phase spaces.

What are the possible applications of the analysis developed in this paper? As already quoted in some of our previous references,[4–7] this type of analysis is fundamental when we want to find the representation in quantum mechanics of nonbijective canonical transformations. Another, and purely classical, application could appear if we consider nonbijective canonical transformations for the Boltzmann equation. But the main application may be the understanding that phase space, as a carrier of canonical transformations, is a more subtle concept than generally envisaged.

[1] A. L. Cauchy, *Oeuvres Completes* (Gauthier-Villars, Paris, 1884).
[2] B. Riemann, *Gesammelte Mathematische Werke* (Dover, New York, 1953).
[3] F. Klein, *On Riemann's Theory of Algebraic Functions* (Dover, New York, 1963); Herman Weyl, *Der Idee der Riemannschen Fläche* (Täubner, Stuttgart, 1955).
[4] P. Kramer, M. Moshinsky, and T. H. Seligman, J. Math. Phys. **19**, 683 (1978).
[5] M. Moshinsky and T. H. Seligman, Ann. Phys. (NY) **114**, 243 (1978).
[6] M. Moshinsky and T. H. Seligman, Ann. Phys. (NY) **120**, 402 (1979); J. Deenen, M. Moshinsky and T. H. Seligman Ann. Phys. (NY) **127**, 458 (1980).
[7] G. García-Calderón and M. Moshinsky, J. Phys. A: Math. Gen. **13**, L 189 (1980).
[8] E. P. Wigner, *Group Theory* (Academic, New York, 1953), pp. 112–23.
[9] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading Mass., 1962), pp. 486–8.

Quantization on the sphere

F. Bayen
Département de Mathématiques, Université de Paris 6, 75230 Paris, France

C. Fronsdal[a)]
Laboratoire de Physique Mathématique Faculté des Sciences, Mirande, 21000 Dijon, France

(Received 13 June 1980; accepted for publication 30 July 1980)

When phase-space is a sphere, then conventional quantization is not available and the Heisenberg algebra is most naturally replaced by so(3). Quantization may be carried out in terms of invariant star-products. This leads to the study of an interesting family of polynomials that are defined in a natural and intrinsic way on the enveloping algebra of sl(2). These polynomials are related to Legendre polynomials by a reordering rule that resembles the relation between Hermite polynomials and normal ordered monomials; they are identified as Pasternack polynomials and related to Jacobi polynomials. New orthogonality properties are found and interpreted in terms of unitary representations of SL(2,$\mathbb{R}$) and SO(3).

PACS numbers: 02.20.Qs, 02.10.Nj

I. INTRODUCTION

1. Conventional, canonical quantization is a procedure that is based on the Heisenberg algebra and relies on the existence of a global system of natural, canonical coordinates; that is, with the possibility of identifying phase-space with $\mathbb{R}^{2n}$. The 2-sphere is at once the simplest and one of the most interesting spaces on which alternative schemes of quantization are called for.

2. On S^2 the obvious choice for the algebra of distinguished observables to replace the Heisenberg algebra is so(3). We identify S^2 with an orbit W of the coadjoint action of $\mathscr{A} = \mathrm{so}(3)$ in the real, vector space dual $\mathscr{A}^*$ of $\mathscr{A}$. If $\{L^A\} A = 1, 2, 3$ is a basis for $\mathscr{A}$, and $\{\lambda_A\}$ the dual basis for $\mathscr{A}^*$, then we denote again by $\{L^A\}$ the coordinates defined on $\mathscr{A}^*$ by the basis $\{\lambda_A\}$. In this way the Lie algebra $\mathscr{A}$ is identified with the vector space $\mathscr{A}^{**}$ of linear functions on $\mathscr{A}^*$. The latter inherits the Lie structure, and this structure defines a Poisson bracket on the space of C^∞ functions on $\mathscr{A}^*$ in the usual way.

3. If g^{-1}_{AB} denote the components of the inverse of the Killing form, then associated with the Casimir element of the enveloping algebra there is as invariant function Q: $\mathscr{A}^* \to \mathbb{R}$ defined by

$$Q = g^{-1}_{AB} L^A L^B. \tag{1.1}$$

To each non-negative value of Q there corresponds a unique orbit of $\mathrm{ad}^*_{\mathscr{A}}$ in $\mathscr{A}^*$, and vice versa. The orbit is S^2 for $Q > 0$ and a single point when $Q = 0$. The Poisson algebra on $C^\infty(\mathscr{A}^*)$ defines a Poisson algebra on $C^\infty(W)$ for each orbit W, and each W is a symplectic space.

4. The essential step in a quantization scheme on S^2 based on $\mathscr{A} = \mathrm{so}(3)$ is the introduction of an invariant $*$-product[1] (an analogue of the product introduced by Moyal[2] for the conventional case). An $\mathscr{A}$-invariant $*$-product on W is an associative product defined on $C^\infty(W)$, denoted $f*g$, with the properties[3]

$$\begin{aligned} &a*f - f*a = i\hbar\{a,f\}, \\ &k*f = kf, \end{aligned} \tag{1.2}$$

for all $a \in \mathscr{A}$, $k \in \mathbb{C}$, $f \in C^\infty(W)$. Here $\{.,.\}$ is the Poisson bracket on W and $\hbar$ is Planck's constant. It was shown[4] that every invariant $*$-product associated with any orbit $W \subset \mathscr{A}^*$ can be constructed by restriction to W of an invariant $*$-product on $\mathscr{A}^*$. Conversely, an invariant $*$-product on $\mathscr{A}^*$ can be restricted to W if and only if the function $\hat{Q}: \mathscr{A}^* \to \mathbb{C}$ defined by

$$\hat{Q} = g^{-1}_{AB} L^A * L^B \tag{1.3}$$

satisfies

$$\hat{Q} * f|_W = \hat{Q}_W f|_W. \tag{1.4}$$

Here $\hat{Q}_W$ is the restriction of $\hat{Q}$ to W and thus a (complex) constant by virtue of (1.2). It is sufficient to consider invariant $*$-products on $\mathscr{A}^*$ that can be restricted to every orbit.

5. For $a = a_A L^A \in \mathscr{A}$ we define the solid Legendre polynomial $P_n(a)$ as follows. Let $t^2 \equiv g^{AB} a_A a_B$, where g^{AB} are the components of the Killing form; then $P_n(a) = (t^2 Q)^{n/2}$ times the ordinary Legendre polynomial with argument $\cos\theta = a(t^2 Q)^{-1/2}$. Hence $P_n(a)$ is homogeneous of degree n in $\{a_A\}$ and also in $\{L^A\}$. Let $P_n(a*)$ denote the same polynomial, after symmetrization in $\{L^A\}$ and replacement of each monomial $L^A L^B \cdots$ by the corresponding $*$-monomial $L^A * L^B * \cdots$. It was shown[5] that every invariant $*$-product on $\mathscr{A}^*$ that satisfies (1.4) is defined in terms of a sequence of constants $c_n \in \mathbb{C}$, with $c_0 = c_1 = 1$, by

$$P_n(a*) = c_n P_n(a) \tag{1.5}$$

together with a function $\hat{Q}: \mathscr{A}^* \to \mathbb{C}$ that is constant on each orbit but otherwise arbitrary.

6. The polynomial $P_n(a*)$ in six variables can be expressed as a $*$-polynomial in a, with coefficients that depend only on t^2, Q and $\hbar$. It turned out, however, to be extraordinarily difficult to evaluate $P_n(a*)$ in that form. It is the main purpose of this paper to give information about these polynomials. It is evident that $P_n(a*)$ is defined (by homogeneity) for all $\hbar$ in terms of its values for $\hbar = -i$. If $\hbar = -i$, then $a*b - b*a = \{a,b\} = [a,b]$ and $P_n(a*)$ can be interpreted as an element $\mathscr{P}_n(a)$ of the enveloping algebra $\mathscr{U}$ of sl(2) by replacing the $*$-product by the product in $\mathscr{U}$. The problem

[a)] Permanent address: Physics Dept., UCLA, Los Angeles, California 90024.

 0022-2488/81/071345-05$1.00

is thus converted into an intrinsic structure problem on $\mathscr{U}$, and the $\mathscr{P}_n(a)$ can be defined within that structure. In Sec. II we give an intrinsic definition of $\mathscr{P}_n(a)$ and show that these polynomials are related to the Pasternack polynomials. In Secs. III and IV we give a number of properties of the $\mathscr{P}_n(a)$, including orthogonality and completeness relations, some of which seem to be new. All these orthogonality relations find a natural interpretation in terms of unitary representations of SL(2,ℝ) and SO(3) (Sec. V).

II. HARMONIC POLYNOMIALS ON sl(2)

1. *Notations*: Letters $a, b, \cdots$ denote elements of sl(2); if $\{L^A\}$ $A = 1,2,3$ is a basis then components a_A of a are defined by $a = a_A L^A$ (summation implied). The field is ℂ. The adjoint action ad(a) is extended to a derivation of the enveloping algebra $\mathscr{U}$ of sl(2) by $\mathrm{ad}(a)f = af - fa$; $a \in \mathrm{sl}(2)$, $f \in \mathscr{U}$. The Killing form is given by

$$g(a,b) = -\tfrac{1}{8}\mathrm{tr}[\mathrm{ad}(a)\mathrm{ad}(b)] = g^{AB}a_A b_B. \tag{2.1}$$

The Laplacian

$$\nabla_a = g^{-1}\left(\frac{\partial}{\partial a}, \frac{\partial}{\partial a}\right) = g^{-1}_{AB}\frac{\partial}{\partial a_A}\frac{\partial}{\partial a_B} \tag{2.2}$$

acts on functions on sl(2) with values in $\mathscr{U}$. The Casimir element

$$\hat{Q} = \tfrac{1}{2}\nabla_a a^2 = g^{-1}_{AB}L^A L^B \in \mathscr{U} \tag{2.3}$$

generates the center of $\mathscr{U}$, and we need two operators associated with it, namely

$$\hat{Q}_{\mathrm{ad}} = \tfrac{1}{2}\nabla_a[\mathrm{ad}(a)]^2 = g^{-1}_{AB}\,\mathrm{ad}(L^A)\,\mathrm{ad}(L^B), \tag{2.4}$$

$$\hat{Q}_l = \tfrac{1}{2}\nabla_a[l(a)]^2 = g^{-1}_{AB}l(L^A)l(L^B). \tag{2.5}$$

The former operates on $\mathscr{U}$; $l(a)$ denotes the vector field defined by left translations on the local Lie group, $\hat{Q}_l$ acts on functions from sl(2) into $\mathscr{U}$.

2. Let $\mathscr{U}_n \subset \mathscr{U}$ denote the subspace that consists of symmetrized, homogeneous polynomials of degree n, so that $\mathscr{U}$, as a vector space, is the direct sum of $\mathscr{U}_n$, $n = 0,1,\ldots,$. The dimension of $\mathscr{U}_n$ is $(n+1)(n+2)/2$. Now $\mathscr{U}_n$ is stable under the adjoint action, and completely reducible: $\mathscr{U}_n = \mathscr{V}^n_n \oplus \mathscr{V}^n_{n-2} \oplus \cdots$. We have

$$\mathscr{U}_n = \mathscr{V}^n_n \oplus \hat{Q}\mathscr{U}_{n-2},\quad \mathscr{V}^n_{n-2} = \hat{Q}\mathscr{V}^{n-2}_{n-2};$$

therefore it is sufficient to study $\mathscr{V}_n \equiv \mathscr{V}^n_n$. The dimension of $\mathscr{V}_n$ is $2n+1$. The vector space $\mathscr{V}_n$, considered as a subspace of $\mathscr{U}_n$, can be characterized in many ways, for example:

(i) If $F \in \mathrm{sl}(2)$ is nilpotent, then $F^n \in \mathscr{V}_n$ and $\mathscr{V}_n$ can thus be generated from F^n by means of the adjoint action.

(ii) $f \in \mathscr{U}_n$ belongs to $\mathscr{V}_n$ iff $\hat{Q}_{ad}f = -4n(n+1)f$.

(iii) $T_{i_1\cdots i_n}L^{i_1}\cdots L^{i_n}$, with T symmetric in all the indices, belongs to $\mathscr{V}_n$ iff T is traceless in the sense that $g^{ij}T_{ijk\cdots} = 0$.

(iv) An invariant function f from sl(2) into $\mathscr{U}_n$ takes values in $\mathscr{V}_n$ iff $\nabla_a f(a) = 0$.

In (iv) an *invariant function* from sl(2) into $\mathscr{U}$ is one that satisfies, for b near the origin,

$$e^b f(a) e^{-b} = f(e^{-b}ae^b), \quad a,b \in \mathrm{sl}(2).$$

(Here we need the extension of the enveloping algebra to its formal closure, composed of formal power series, for the sake of giving a meaning to these expressions.)

3. The function from sl(2) into $\mathscr{U}_n$ defined by $a \mapsto a^n$ enjoys the following properties: (i) it is an invariant function in the sense defined above, and (ii) the vector space generated by all a^n, $a \in \mathrm{sl}(2)$, is $\mathscr{U}_n$. The harmonic polynomial $\mathscr{P}_n(a)$ plays the same role relative to $\mathscr{V}_n$.

Definition 1: We call harmonic polynomial of degree n on sl(2) any invariant function from sl(2) into $\mathscr{V}_n$.

Proposition 2: Up to a multiplicative factor, there exists one and only one harmonic polynomial of degree n, for each $n = 0,1,\cdots$.

Proof: Let $\mathscr{P}_n(a)$ be a harmonic polynomial of degree n, and let Y be a regular element of sl(2). Then, because $a \mapsto \mathscr{P}_n(a)$ is invariant, $\mathscr{P}_n(a)$ is determined by $\mathscr{P}_n(Y)$, and $\mathrm{ad}(Y)\,\mathscr{P}_n(Y) = 0$. Now $\mathscr{V}_n$ is an irreducible sl(2) module defined by the adjoint action; thus $\mathscr{V}_n$ contains precisely a one-dimensional subspace on which ad(Y) vanishes.

Definition 3: We denote by $\mathscr{P}_n(a)$ the unique harmonic polynomial of degree n that is so normalized that it reduces to $[(2n-1)!!/n!]\,a^n$ when a is nilpotent; in particular, $\mathscr{P}_0(a) = 1$ and $\mathscr{P}_1(a) = a$.

4. To evaluate $\mathscr{P}_n(a)$ explicitly one may apply any one of the characterizations (i)–(iv) of $\mathscr{V}_n$ enumerated above.

(i) Let F, G, Y be a basis for sl(2), with F, G nilpotent and Y regular. Then there is a constant d_n such that

$$\mathscr{P}_n(Y) = d_n[\mathrm{ad}(G)]^n F^n;$$

(ii) Again, up to a constant factor, $\mathscr{P}_n(a)$ is determined as being the only invariant solution of

$$\hat{Q}_{\mathrm{ad}}\,\mathscr{P}_n(a) = -4n(n+1)\mathscr{P}_n(a); \tag{2.6}$$

(iii) An explicit formula is[5]

$$\mathscr{P}_n(a) = \frac{(2n-1)!!}{n!}a_{i_1}\cdots a_{i_n}T^{(n)}_{j_1\cdots j_n}{}^{i_1\cdots i_n}L^{j_1}\cdots L^{j_n}, \tag{2.7}$$

where $T^{(n)}$ is the traceless projection operator for symmetric tensors of rank n. This shows that the image of $\mathscr{P}_n(a)$ in the symmetric algebra (by the natural bijection) is the solid Legendre polynomial;

(iv) Finally, $\mathscr{P}_n(a)$ can be evaluated by solving the equation $\nabla_a \mathscr{P}_n(a) = 0$.

5. One finds, with $t^2 = g(a,a)$

$$\mathscr{P}_0(a) = 1, \quad \mathscr{P}_1(a) = a, \quad \mathscr{P}_2(a) = \tfrac{3}{2}a^2 - \tfrac{1}{2}t^2\hat{Q},$$

$$\mathscr{P}_n(a) = \sum_{k=n,n-2,\cdots} a^{n-2k}t^{2k}A^k_n(\hat{Q}), \tag{2.8}$$

but the coefficients $A^k_n(\hat{Q})$ become very complicated as n grows. To study the general properties of these polynomials we shall use another method.

6. *Evaluation of* $\mathscr{P}_n(a)$

The vector field (b) associated with left translations on the local group can be defined by

$$[b - l(b)]e^a = 0; \tag{2.9}$$

therefore

$$[\hat{Q} - \hat{Q}_l]e^a = 0. \tag{2.10}$$

The tangent space at any point of sl(2) can be canonically identified with sl(2), and this allows us to write

$$-l(b)|_a = b + \tfrac{1}{2}[b,a] + \tfrac{1}{12}[[b,a],a] + \cdots$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} B_n x^n b, \tag{2.11}$$

where x is the linear operator $xb = \mathrm{ad}(a)b$ and the coefficients B_n are the Bernoulli numbers. Let $t^2 = g(a,a)$; a calculation yields[6]

$$\hat{Q}_t = \partial_t^2 + \frac{2}{\tan t}\partial_t + \frac{1}{4\sin^2 t}\hat{Q}_{\mathrm{ad}}. \tag{2.12}$$

Here ∂_t is the "radial" derivative; that is, $t\partial_t = a(\partial/\partial a)$ is the generator of the homothety group on sl(2).

Suppose that there is a formal expansion

$$e^a = \sum_n C_n(\hat{Q},t)\mathscr{P}_n(a), \tag{2.13}$$

then (2.10), (2.12), and (2.6) give an equation for C_n; namely

$$\left[\partial_t^2 + \frac{2}{\tan t}\partial_t - \frac{n(n+1)}{\sin^2 t} - \hat{Q}\right] t^n C_n(\hat{Q},t) = 0. \tag{2.14}$$

The function

$$C_n^{\lambda}(t) = c_n \frac{i(2n+1)}{(n+\lambda)!\sin t}\left(\frac{i}{t}\right)^n e^{-i\pi\lambda} Q_n^{\lambda}\left(\frac{1}{i\tan t}\right), \tag{2.15}$$

where Q_n^{λ} denotes a Legendre function of the second kind, is a polynomial of degree $n-2$ in λ^2. A solution to (2.14) is obtained by substituting

$$\lambda^2 \to 1 - \hat{Q} \tag{2.16}$$

in $C_n^{\lambda}(t)$, and every solution to (2.14) that is regular in t near $t = 0$ is of this form, for some choice of the constant c_n.

When the solution to (2.14) determined by (2.15), with $c_n = 1$, $n = 0,1,\cdots$ is substituted for $C_n(\hat{Q},t)$ in (2.13), then one recognizes (see below) an expansion in terms of Pasternack polynomials. Thus, with the identification (2.16)

$$\mathscr{P}_n(a) = (-t)^n(\lambda+1)_n F_n^{\lambda}(a/t). \tag{2.17}$$

In the next sections we have collected some information about these polynomials.

III. PROPERTIES OF THE HARMONIC POLYNOMIALS

1. When the expression (2.7) is evaluated in the symmetric algebra of sl(2) one obtains the solid Legendre polynomial $P_n(a)$. Thus one may write

$$\mathscr{P}_n(a) = \mathrm{sym}P_n(a), \tag{3.1}$$

where sym stands for symmetric ordering. This is analogous to the formula for Hermite polynomials as normal ordered monomials: $H_n \propto :x^n:$ Perhaps (3.1) is directly related to Pasternack's definition[7]:

$$\cosh^{\lambda+1}x\, F_n^{\lambda}\left(\frac{\partial}{\partial x}\right)\cosh^{-\lambda-1}x = P_n(\tanh x).$$

(Here P_n denotes the ordinary Legendre polynomial.)

2. Pasternack gives an explicit formula[8] for F_n^{λ}, as a generalized hypergeometric function, that we may write as

$$\mathscr{P}_n(a) = (-t)^n(\lambda+1)_n \times {}_3F_2\left(\begin{matrix} -n, & n+1, & (\lambda+1+a/t)/2 \\ 1, & \lambda+1; & 1 \end{matrix}\right). \tag{3.2}$$

3. From (2.13), with $C_n(\hat{Q},t)$ given by (2.15) and using well-known recursion relations for associated Legendre polynomials, one finds[9]

$$(n+1)\mathscr{P}_{n+1}(a) = (2n+1)a\mathscr{P}_n(a) + n(\lambda^2 - n^2)t^2\mathscr{P}_{n-1}(a). \tag{3.3}$$

This agrees with a recursion relation given by Pasternack.[10] It was this relation that led us to (2.17). The explicit formula (2.13) was given by Bateman[11] for the case $\lambda = 0$ only, the general case may be new.

4. There is an interesting relationship between the Pasternack polynomials and Bernoulli numbers.[12] Perhaps this is related to the appearance of Bernoulli numbers in Eq. (2.11).

5. The role of harmonic polynomials as the solutions of physical problems was noted by Pasternack[13]; especially noteworthy is their connection with the mean values of r^{-n-2} in hydrogenic atoms.

6. There is a simple relation between harmonic polynomials and Jacobi polynomials that may be exploited to obtain orthogonality relations, including cases that apparently have not been noted previously. Such relations are derived in the next section, and interpreted in terms of the representation theory of SL(2,ℝ) and SO(3) in Sec. V.

7. Finally, there is a finite difference equation satisfied by the Pasternack polynomials,[14] reproduced here in the Appendix, Eq. (A5). In terms of $\mathscr{P}_n(a)$ it reads

$$\Delta(a^2 + \lambda^2 t^2)\Delta\mathscr{P}_n(a) = -4n(n+1)\mathscr{P}_n(a), \tag{3.4}$$

where Δ is the finite difference operator

$$\Delta f(a) = [f(a+it) - f(a-it)]/t^2. \tag{3.5}$$

Comparison with (2.6) is very suggestive, but we do not know how to derive (3.4) directly from (2.6). By group contraction or "classical limit" from (3.4) one obtains Legendre's differential equation for $P_n(a)$.

IV. ORTHOGONALITY RELATIONS AND GENERALIZATIONS

Let $P_n^{(\alpha,\beta)}(y)$ denote the Jacobi polynomials given by Eq. (A1) in the Appendix, and define

$$u_n^{(\alpha,\beta)}(y) = (1-y)^{\alpha/2}(1+y)^{\beta/2}P_n^{(\alpha,\beta)}(y),$$
$$n = 0,1,2,\cdots; \quad \mathrm{Re}\alpha > -1, \quad \mathrm{Re}\beta > -1. \tag{4.1}$$

Let α,β be fixed. Considered as complex-valued functions on $[-1,1]$, this sequence generates a vector space that is dense in $L^2(-1,1)$. One has a biorthogonality relation[15]:

$$(u_m^{(\bar{\alpha},\bar{\beta})}, u_n^{(\alpha,\beta)}) \equiv \int_{-1}^{1} u_m^{(\alpha,\beta)}(y)u_n^{(\alpha,\beta)}(y)\,dy$$
$$= \frac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!\Gamma(n+\alpha+\beta+1)(2n+\alpha+\beta+1)}\delta_{nm}. \tag{4.2}$$

[If $\alpha+\beta = -1$, then for $n = 0$ replace $(2n+\alpha+\beta+1)$ by 1.]

Let $V: L^2(-1,1) \to L^2(\mathbb{R})$ be the Hilbert space isomorphism given by

$$(Vf)(x) = (2\pi)^{-1/2}\int_{\mathbb{R}} \frac{e^{-ixt}}{\cosh t} f(\tanh t)\,dt \tag{4.3}$$

and define

$$v_n^{(\alpha,\beta)} = Vu_n^{(\alpha,\beta)}. \tag{4.4}$$

These functions can be evaluated explicitly in terms of generalized hypergeometric functions, Eqs. (A2)–(A4) of the Appendix. As those equations show, $v_n^{(\alpha,\beta)}(x)/v_0^{(\alpha,\beta)}(x)$ is a polynomial of order n.

Finally, let $G_n^{(\alpha,\beta)}(x)$ be the polynomial (of order n), proportional to $v_n^{(\alpha,\beta)}(x)/v_0^{(\alpha,\beta)}(x)$, having a unit coefficient for the term x^n. Then

$$G_n^{(\alpha,\beta)}(x) = \frac{(2i)^n(\alpha+1)_n}{(n+\alpha+\beta+1)_n}\left(\frac{\alpha+\beta+2}{2}\right)_n \times {}_3F_2\left(\begin{matrix} -n, & n+\alpha+\beta+1, & \tfrac{1}{2}(ix+\alpha+1) \\ \alpha+1, & \tfrac{1}{2}(\alpha+\beta+2); & 1 \end{matrix}\right). \tag{4.5}$$

Using recursion relations for the Jacobi polynomials one derives the recursion relations (A5) for $G_n^{(\alpha,\beta)}(x)$. Comparison with the recursion relation (3.3) for $\mathscr{P}_n(a)$ one makes the following identification:

$$\mathscr{P}_n(a) \underset{\hat{Q}\to 1-\lambda^2}{\longrightarrow} \frac{(2n-1)!!}{n!}\, t^n G_n^{(\lambda,-\lambda)}(a/t). \tag{4.6}$$

We now give orthogonality relations for $G_n^{\lambda}(x) \equiv G_n^{(\lambda,-\lambda)}(x)$. The general case of $G_n^{(\alpha,\beta)}(x)$ is given in the Appendix, Eq. (A5).

Case 1, $0 \leqslant \lambda < 1$ (*Supplementary series*): The following result, which is not new,[16] is easily derived from (4.2):

$$\int_{\mathbb{R}} G_m^{\lambda}(x) G_n^{\lambda}(x) \frac{dx}{\cos\pi\lambda + \cosh\pi x} = \frac{2\lambda(1-\lambda)_n(1+\lambda)_n(n!)^2}{\sin\pi\lambda\,[(2n-1)!!]^2(2n+1)}\delta_{mn}. \tag{4.7}$$

Completeness is also easily verified. We call this case the case of the supplementary series because the correspondence $\hat{Q}\to 1-\lambda^2$ assigns to $\hat{Q}$ the values taken by the Casimir element in the supplementary series of unitary representations of SL(2,$\mathbb{R}$). We next consider the other unitary representations of SL(2,$\mathbb{R}$) and of SO(3).

Case 2, λ *imaginary* (*Principal series*): The result (4.7) still holds; completeness can also be verified.

Case 3, $\lambda = 1,2,\cdots$ (*Discrete series*): In this case the recursion relation (A5) shows that $G_{n+\lambda}^{(\lambda,-\lambda)}(x)$ contains $G_{\lambda}^{\lambda}(x)$ as a factor:

$$G_{n+\lambda}^{\lambda}(x) \equiv G_{n+\lambda}^{(\lambda,-\lambda)}(x) = G_{\lambda}^{\lambda}(x) G_n^{(\lambda,\lambda)}(x), \tag{4.8}$$

$$G_{\lambda}^{\lambda}(x) = \prod_{k=0}^{\lambda-1}[x + i(1-\lambda+2k)]. \tag{4.9}$$

Instead of (4.7)

$$\int_{\mathbb{R}} G_{n+\lambda}^{\lambda}(x) G_{m+\lambda}^{\lambda}(x) \frac{dx}{|\cosh(\pi/2)(x+i\lambda)|^2} = \frac{2^{2-2\lambda-2n} n!(n+\lambda)!(n+\lambda)!(n+2\lambda)!}{[\Gamma(n+\lambda+1/2)]^2(2n+2\lambda+1)}\delta_{mn}. \tag{4.10}$$

Now it is the system $(G_n^{(\lambda,\lambda)}(x))_{n\geqslant 0}$ that is complete. Of course, these polynomials are orthogonal for the measure $|G_{\lambda}^{\lambda}(x)/\cosh(\pi/2)(x+i\lambda)|^2 dx$. Orthogonality is thus verified for all those values of $\hat{Q} = 1-\lambda^2$ that correspond to unitary representations of SL(2,$\mathbb{R}$). It remains only to consider unitary representations of SO(3).

Case 4, $\lambda = 1,2,\cdots$ (*Finite case*): Let us put

$$\lambda = 2l+1,\ l = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots,$$

$$G_n^{\lambda}(ix) = \frac{i^n n!}{(2n-1)!!}\, t_n\left(\frac{x}{2}+l\right), \tag{4.11}$$

$$n = 0,1,\ldots,\lambda-1.$$

Then the polynomials $t_n(z)$ are the "discrete Tchebicheff polynomials".[17] They have the discrete orthogonality property

$$\sum_{k=0}^{N-1} t_m(k)t_n(k) = \frac{\delta_{nm}}{2n+1}(N-n)(N-n+1)\cdots(N+n).$$

Thus

$$\sum_{k=-l}^{+l} G_n^{\lambda}(2ik)\,\overline{G_m^{\lambda}(2ik)} = \frac{1}{2n+1}\left[\frac{n!}{(2n-1)!!}\right]^2(\lambda-n)(\lambda-n+1)\cdots(\lambda+n)\delta_{mn}. \tag{4.12}$$

The system $(G_n(x))_{n=0,1,\ldots,\lambda-1}$ is complete on $\{x = 2ik;\ k = -l, -l+1,\ldots,l\}$. This final result suggests an interpretation of all the orthogonality relations in terms of unitary representations of SO(3) and SL(2,$\mathbb{R}$).

V. INTERPRETATION OF ORTHOGONALITY RELATIONS

Let $a \mapsto T_a$ be a unitary irreducible representation of the compact real subalgebra so(3) of sl(2), by antihermitean matrices of dimension $2l+1$ (= any fixed non-negative integer). Then $\hat{Q}\to 1-\lambda^2 = -4l(l+1)$ and we take $\lambda = 2l+1$. If $t^2 = g(a,a) = 1$, then the spectrum of T_a is the set $\{2ik;\ k = -l, -l+1,\ldots,l\}$, and the "orthogonality" relation (4.12) may be written L — in view of (4.6)—

$$\mathrm{tr}[\mathscr{P}_n(T_a)\mathscr{P}_m(T_a^{\dagger})] = (1/(2n+1))(\lambda-n)(\lambda-n+1)\cdots(\lambda+n)\delta_{nm}. \tag{5.1}$$

In other words, these orthogonality relations express orthogonality in the Hilbert algebra generated by the Hilbert–Schmidt operators $\mathscr{P}_n(T_a)$—here in the finite-dimensional case. We believe that an analogous interpretation of the other orthogonality relations is possible.

ACKNOWLEDGMENTS

We thank M. Flato and G. Pinczon for helpful discussons. One of us (C. F.) thanks M. Flato for hospitality at the University of Dijon.

APPENDIX: ADDITIONAL FORMULAS

Jacobi polynomials are defined by

$$P_n^{\alpha,\beta}(y) = \frac{(\alpha+1)_n}{n!}\,{}_2F_1\left(-n, n+\alpha+\beta+1; \alpha+1; \frac{1-y}{2}\right) \tag{A1}$$

for n integer, α and β complex, $\mathrm{Re}\alpha > -1$ and $\mathrm{Re}\beta > -1$. Eqs. (4.3) and (4.4) give

$$v_n^{(\alpha,\beta)}(x) = (2\pi)^{-1/2} \int_{\mathbb{R}} \frac{e^{-ixt - t(\alpha-\beta)/2}}{(\cosh t)^{(\alpha+\beta+2)/2}} \times P_n^{(\alpha,\beta)}(\tanh t)\, dt, \tag{A2}$$

$$v_n^{(\alpha,\beta)}(x) = \frac{(\alpha+1)_n}{n!} v_0^{(\alpha,\beta)}(x) \times {}_3F_2\left(\begin{matrix} -n, & n+\alpha+\beta+1, & \frac{1}{2}(ix+\alpha+1) \\ \alpha+1, & \frac{1}{2}(\alpha+\beta+2); & 1 \end{matrix}\right) \tag{A3}$$

$$v_0^{(\alpha,\beta)}(x) = (2\pi)^{-1/2} 2^{(\alpha+\beta)/2} B\left(\frac{\alpha+1+ix}{2}, \frac{\beta+1-ix}{2}\right). \tag{A4}$$

The polynomials $G_n^{(\alpha,\beta)}(x)$ defined by Eq. (4.5) are the monic polynomials proportional to $v_n^{(\alpha,\beta)}(x)/v_0^{(\alpha,\beta)}(x)$. They satisfy the folowing recursion relations, derived from recursion relations for $P_n^{(\alpha,\beta)}$ or directly from Eq. (4.5);

$$G_{n+1}^{(\alpha,\beta)}(x) = xG_n^{(\alpha,\beta)}(x) - \gamma_n G_{n-1}^{(\alpha,\beta)}(x), \quad n \geqslant 1, \tag{A5}$$

$$\gamma_n = \begin{cases} \dfrac{n(n+\alpha)(n+\beta)(n+\alpha+\beta)}{(2n+\alpha+\beta)^2-1} & \text{if } \alpha+\beta+1 \neq 0, \\ \frac{1}{4}(n+\alpha)(n-\alpha-1) & \text{if } \alpha+\beta+1=0,\ n>1, \\ -\frac{1}{2}\alpha(\alpha+1) & \text{if } \alpha+\beta+1=0,\ n=1. \end{cases}$$

Finite difference relations can also be derived, for example,

$$[x+i(\alpha+1)]\,[x+i(\beta+1)]G_n^{(\alpha,\beta)}(x+2i) + [x-i(\alpha+1)]\,[x-i(\beta+1)]G_n^{(\alpha,\beta)}(x-2i) - 2[x^2 - 2n(n+\alpha+\beta+1) - (\alpha+1)(\beta+1)] \times G_n^{(\alpha,\beta)}(x) = 0.$$

[1] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. **111**, 61 (Part I); 111 (Part II) (1978).
[2] J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[3] Ref. 1, Part I, Sec. 9.
[4] Ref. 1, Part I, Sec. 11.
[5] Ref. 1, Part I, p. 99.
[6] Ref. 1, Part I, p. 106.
[7] S. Pasternack, Philos. Mag. (7) **28**, 209 (1939).
[8] Ref. 7, p. 210.
[9] C. Fronsdal, Rep. Math. Phys. **15**, 113 (1978).
[10] Ref. 7, p. 213.
[11] H. Bateman, Ann. Math. **35**, 767–68 (1934).
[12] J. Touchard, Canad. J. Math. **8**, 305 (1956); L. Carlitz, Duke Math. J. **26**, 1 (1959).
[13] S. Pasternack, Proc. Nat. Acad. Sci. USA **23**, 91 (1937).
[14] Ref. 7, p. 213.
[15] E. D. Rainville, *Special Functions* (Chelsea, New York, 1960), p. 260.
[16] T. S. Chihara, *An Introduction to Orthogonal Polynomials* (Gordon and Breach, New York (1978), p. 193.
[17] Ref. 16, p. 162.

Irreducible linear–antilinear representations and internal symmetries [a)]

C. Garola and L. Solombrino

Istituto di Fisica e Istituto di Matematica dell'Università, Facoltà di Scienze, Università di Lecce, 73100 Lecce, Italy
Istituto Nazionale di Fisica Nucleare, Italy

(Received 11 November 1980; accepted for publication 31 December 1980)

We reconsider a well-known classification, due to Wigner, of the unitary–antiunitary finite-dimensional irreducible group representations within a somewhat generalized mathematical framework, where, in particular, any algebraically closed field K with an involutory automorphism j is considered in place of the complex field C. We show that each case of the classification can be characterized by the set $\mathscr{U}'^{l}$ of the linear mappings that commute with the given set $\mathscr{U}$, which is now assumed to be an irreducible semigroup of linear and antilinear (i.e., j-semilinear) mappings, and explicitly exhibit $\mathscr{U}'^{l}$. Then, this classification is crossed with the classification of the sets of linear and antilinear mappings that has been obtained in some previous work and that generalizes the old classification of the unitary representations introduced by Frobenius and Schur. We obtain a new classification in which every case can be characterized by the group $\mathscr{U}^{c}$ of the invertible linear and antilinear mappings which commute with the semigroup $\mathscr{U}$. Whenever $\mathscr{U}$ represents the symmetry group of some physical system, $\mathscr{U}^{c}$ may represent "internal" symmetries, so that the problem of finding these symmetries is now related to some natural classification of $\mathscr{U}$. In particular, it turns out that in four cases $\mathscr{U}^{c}$ reduces to its linear part $\mathscr{U}^{cl} = \mathscr{U}'^{l} \setminus \{0\}$, whereas in the remaining nine cases $\mathscr{U}^{c}$ is an extension of $\mathscr{U}^{cl}$ through G_2, which reduces to a semidirect product, but not to a direct one, in three cases and to a direct product in three cases.

PACS numbers: 02.20.Nq

1. INTRODUCTION

According to a classification introduced by Frobenius and Schur in 1906[1] and reported in a number of books on group theory,[2–5] there are three kinds of unitary finite-dimensional complex group representations:

(i) representations which are equivalent to a real representation (*potentially real case*);

(ii) representations which are equivalent to their complex conjugate representation, but not to a real representation (*pseudoreal case*);

(iii) representations which are not equivalent to their complex conjugate representation (*complex case*).

On the other hand, the irreducible finite-dimensional complex group corepresentations (i.e., representations by unitary and antiunitary operators) can be divided into three types according to a classification introduced by Wigner[6]:

Type I: the unitary part of the representation is irreducible;

Type II: the unitary part of the representation reduces to the sum of two equivalent representations;

Type III: the unitary part of the representation reduces to the sum of two equidimensional inequivalent representations.

These classifications have been combined in a remarkable work by Dyson,[7] who has shown, making use of Weyl's general theory of matrix algebras and their commutator algebras[8] and of the theorem of Frobenius[9] that if the classification of Wigner is applied to an irreducible corepresentation and the classification of Frobenius and Schur to the linear part of it, the nine case which might arise actually occur; each case is characterized (with the exception of one case only, which splits further into two subcases) by the canonical form of the algebra A over the real field R generated by the linear part of the representation or, equivalently, by the canonical form of the R-linear commutant of this algebra (commutator algebra of A). Thus, Dyson obtains a new and more detailed classification. This classification has two features which are perhaps unsatisfactory; indeed, (1) it crosses a classification which applies to corepresentations with another classification which applies to unitary representations and (2) the R-linear commutant of the linear part of the representation, which characterizes every possible case, does not seem to have a direct physical meaning.

Here we start from some previous work[10] where the classification of Frobenius and Schur was generalized by ourselves, together with other authors, in such a way that it can be applied to any set $\mathscr{U}$ of (not necessarily linear) mappings of a vector space X over a division ring K with an involutory automorphism j (conjugation); we remark explicitly that we neither assume that X is finite-dimensional nor that a scalar product is defined in X (hence we do not assume that the mappings of $\mathscr{U}$ are unitary). The three cases which can occur can be characterized by sets of equivalent properties of the set $\mathscr{U}$ or of its "linear–antilinear centralizer" or commutant group $\mathscr{U}^{c}$ (i.e., the group of the invertible linear and antilinear mappings of X which commute with $\mathscr{U}$).

In particular, since $\mathscr{U}^{c}$ turns out to be an extension of its linear part $\mathscr{U}^{cl}$ through the abstract two-element group G_2 whenever it does not reduce to $\mathscr{U}^{cl}$, the three cases of our classification can be characterized as follows:

(i) $\mathscr{U}^{c}$ is isomorphic to a semidirect product of $\mathscr{U}^{cl}$ and G_2,

[a)]Research supported by CNR and INFN (Italy).

 0022-2488/81/071350-09$1.00

(ii) $\mathcal{U}^c$ is an extension of $\mathcal{U}^{cl}$ through G_2 but is not isomorphic to any semidirect product of $\mathcal{U}^{cl}$ and G_2,

(iii) $\mathcal{U}^c$ coincides with $\mathcal{U}^{cl}$.

This characterization seems especially interesting, since $\mathcal{U}^c$ has an independent and important physical meaning: when $\mathcal{U}$ represents a given group of symmetries of some physical system, $\mathcal{U}^c$ may represent "internal" symmetries.[11,12]

These results have been deepened in a mathematical work,[13] where we investigated some properties of the morphism associated with the semidirect product and in particular the possibility that the semidirect product reduces to a direct one. This investigation leads to a refined classification, which is sixfold instead of threefold only since the "potentially real" case splits into three subcases and the "pseudoreal" case into two subcases.

Thus, the idea arises to combine these new classifications with the Wigner's classification (which can be applied, however, only under more restrictive assumptions, first of all irreducibility); this, then, is the problem that we are going to discuss here.

In Sec. 2 we give a synthesis of our generalization of the Frobenius and Schur classificaton and of its refinements; we express here these results in a slightly modified form, which makes them more suitable for use in the following in this paper.

In Sec. 3 we generalize the original Wigner classification to irreducible linear–antilinear semigroup representations in a vector space X over an algebraically closed field K with a conjugation (we remark that our operators are no more requested to be unitary or antiunitary) and give the explicit form of the linear commutant $\mathcal{U}'^{l} = \mathcal{U}^{cl} \cup \{0\}$ (the equality holds because of the irreducibility of $\mathcal{U}$) of the representation in each of the four cases that can occur (there are four cases and not three because we also include those representations whose antilinear part is void). Whenever $K = C$, our Proposition 1 has essentially the same content of Dyson's equivalence Theorem I[14]; yet, it gives $\mathcal{U}'^{l}$ as a set of operators in the vector space X over C rather than as a set of matrices in a real vector space, as Dyson does because of his use of Weyl's theory and Frobenius' theorem.

In Sec. 4 we cross this classification with the one that we have obtained when generalizing Frobenius and Schur's; then, 12 cases are abstractly possible. A nontrivial result is that one of them cannot occur; any one of the 11 remaining cases can be characterized by a set of properties of $\mathcal{U}$ or $\mathcal{U}^c$. In particular, it turns out that in four cases $\mathcal{U}^c$ reduces to its linear part $\mathcal{U}^{cl}$ and in seven cases it is an extension of $\mathcal{U}^{cl}$ through the two elements group G_2; in the latter possibility it may be a semidirect product (four cases) or not (three cases). Then, we consider further the more refined classification discussed above; due to the crossing with the Wigner classification, it turns out that only two of the cases split in subcases: In each of these cases $\mathcal{U}^c$ is isomorphic to a semidirect product, and the two subcases refer to the possibility that $\mathcal{U}^c$ reduces or not to a direct product. Thus, we get 13 distinct possibilities, and each of them can actually occur, as we show giving simple examples in Sec. 5[15]; we stress again that any case is characterized by a mathematical object (the linear–antilinear centralizer) which admits a physical interpretation.

Finally, we outline that throughout our work we have made an effort to minimize the mathematical assumptions which are needed to obtain our results; this, besides contributing a welcome simplicity, also helps to pinpoint the essential mathematical features upon which our arguments hang (for instance, we have dispensed with the assumption that a scalar product exists in X everywhere; furthermore, our results do not depend on the field one wants to use as the basic field, provided that this has a conjugation and is algebraically closed).

2. THE LINEAR–ANTILINEAR COMMUTANT

As we have already announced in the Introduction, this section is devoted to giving a synthesis of the classification obtained by the authors of the present paper together with others[10] and of the further refinements that have been recently introduced by the authors themselves in a mathematical work[13]; the presentation of these results will be made in a slightly simplified form. At the end, we add a remark that can be useful in applications: here, we use a bicommutant theorem which was stated by the authors and others in another mathematical work on this subject.[16]

Let us start by introducing some preliminary concepts. First, we give two definitions that will be used throughout the present work.

Definition 1: We call any division ring K endowed with a nonidentical involutory automorphism $j:\alpha \to \overline{\alpha}$ a division ring with a conjugation. We call the subdivision ring of K which consists of the self-conjugate elements of K the j-invariant subring Λ of K.

Definition 2: Let X be a vector space over a division ring K with a conjugation j. We call any mapping of X which is semilinear with respect to j an *antilinear mapping*.

Let $\mathcal{U}$ be any set of mappings of X. We denote the subset of all the linear (antilinear) mappings of $\mathcal{U}$ by $\mathcal{U}^l$ ($\mathcal{U}^a$).

We denote the set of the mappings of X that commute with all the mappings of $\mathcal{U}$ by $\mathcal{U}'$. Hence $\mathcal{U}'^{l}$ (*linear commutant* of $\mathcal{U}$) and $\mathcal{U}'^{a}$ (*antilinear commutant* of $\mathcal{U}$), respectively, are the multiplicative semigroup of all the linear mappings of $\mathcal{U}'$ and the set of all the antilinear mappings of $\mathcal{U}'$.

We call the set of the invertible mappings of the *linear–antilinear commutant* $\mathcal{U}'^{l} \cup \mathcal{U}'^{a}$, that is, the multiplicative group of the linear and antilinear invertible mappings that commute with $\mathcal{U}$, the *linear–antilinear centralizer* (or *commutant group*) of $\mathcal{U}$ and denote it by $\mathcal{U}^c$. Hence, $\mathcal{U}^{cl}$ (*linear centralizer* of $\mathcal{U}$) and $\mathcal{U}^{ca}$ (*antilinear centralizer* of $\mathcal{U}$), respectively, are the subgroup of all the linear mappings of $\mathcal{U}^c$ and the subset of all the antilinear mappings of $\mathcal{U}^c$.

Second, we give a list of the not yet defined symbols that will be used in what follows.

List of symbols:

θ: the mapping from $\mathcal{U}^c$ into the set of the mappings of $\mathcal{U}^{cl}$ into itself such that, for any $A \in \mathcal{U}^c$,

$\theta(A)$: $L \in \mathcal{U}^{cl} \to ALA^{-1} \in \mathcal{U}^{cl}$

$\ltimes$: semidirect product

$\ltimes_\varphi$: the semidirect product associated with the morphism φ

$\times$: direct product

G_2: the abstract two-element group

E: the identity mapping

Third, we convene that any equality of the form $\mathcal{U}^c = \mathcal{U}^{cl} \ltimes_\theta \{E,J\}$, with J antilinear involutory mapping, that will appear in what follows must be intended in the sense of the identifications $(L,E) = L$ and $(L,J) = LJ$ for any $L \in \mathcal{U}^{cl}$.

Now, we come to the classification introduced in Ref. 10. As a preliminary result, we recall that, for any set $\mathcal{U}$ of mappings of X such that $\mathcal{U}^{ca}$ is nonvoid, $\mathcal{U}^c$ is an extension of $\mathcal{U}^{cl}$ through G_2, since $\mathcal{U}^c = \mathcal{U}^{cl} \cup \mathcal{U}^{cl} A$ with $A \in \mathcal{U}^{ca}$. Then, for any set $\mathcal{U}$ (nonnecessarily irreducible) of linear and antilinear mappings of X (the original results are formulated for sets of arbitrary mappings; the restriction to sets of linear and antilinear mappings allows the formulation of our results in terms of matrix representations), one of three mutually exclusive cases occurs, each of which can be characterized by a set of equivalent conditions as follows[17]:

(i) *$\mathcal{U}$ potentially real*: A basis exists in which the matrix representation of $\mathcal{U}$ [18] is self-conjugate (i.e., the matrix elements belong to Λ); equivalently, a mapping $J \in \mathcal{U}^{ca}$ such that $J^2 = E$ exists; equivalently, $\mathcal{U}^c = \mathcal{U}^{cl} \ltimes_\theta \{E,J\}$, with J antilinear involutory mapping (hence $\mathcal{U}^c$ is isomorphic to a semidirect product of $\mathcal{U}^{cl}$ and the abstract two-element group G_2);

(ii) *$\mathcal{U}$ pseudoreal*: The "conjugate" representation of $\mathcal{U}$ (i.e., the representation obtained by conjugation of the matrix elements of $\mathcal{U}$) is equivalent to the matrix representation of $\mathcal{U}$, but no basis exists in which the matrix representation of $\mathcal{U}$ is self-conjugate; equivalently, $\mathcal{U}^{ca}$ is nonvoid but $A^2 \neq E$ for any $A \in \mathcal{U}^{ca}$; equivalently, $\mathcal{U}^{ca}$ is nonvoid but $\mathcal{U}^c$ is not isomorphic to a semidirect product of $\mathcal{U}^{cl}$ and G_2;

(iii) *$\mathcal{U}$ complex*: The conjugate representation is not equivalent to the matrix representation of $\mathcal{U}$; equivalently, $\mathcal{U}^{ca}$ is void.

This classification immediately leads one to wonder about the morphism θ (see the list of symbols above) and the possibility that the semidirect product reduces to a direct one; by investigating these problems we have shown, in particular, that cases (i) and (ii) can be divided in subcases by considering the set $\mathcal{U} \cup \mathcal{U}^c$ or its center $(\mathcal{U} \cup \mathcal{U}^c)^c = \mathcal{U}^c \cap \mathcal{U}^{cc}$.[13] Our main results can be summarized by the following statements:

(a) the set $\mathcal{U} \cup \mathcal{U}^c$ is potentially real [equivalently, apply (i) above, the center $(\mathcal{U} \cup \mathcal{U}^c)^c$ of $\mathcal{U}^c$ contains involutory antilinear mappings] if and only if $\mathcal{U}^c = \mathcal{U}^{cl} \times \{E,J\}$, with J involutory antilinear mapping;

(b) the set $\mathcal{U} \cup \mathcal{U}^c$ is pseudoreal [equivalently, apply (ii) above, the center $(\mathcal{U} \cup \mathcal{U}^c)^c$ of $\mathcal{U}^c$ contains antilinear mappings but none of them is involutory] if and only if $\mathcal{U}^{ca}$ is nonvoid and the mapping $\theta(A)$ is inner for any $A \in \mathcal{U}^{ca}$ and nonidentical for any involutory $J \in \mathcal{U}^{ca}$;

(c) the set $\mathcal{U} \cup \mathcal{U}^c$ is complex [equivalently, apply (iii) above, the center $(\mathcal{U} \cup \mathcal{U}^c)^c$ of $\mathcal{U}^c$ does not contain antilinear mappings] if and only if either $\mathcal{U}^{ca}$ is nonvoid and the mapping $\theta(A)$ is not inner for any $A \in \mathcal{U}^{ca}$ or $\mathcal{U}^{ca}$ is void.

Hence, case (i) can generate three subcases, case (ii) two subcases ($\mathcal{U} \cup \mathcal{U}^c$ cannot be potentially real if $\mathcal{U}$ is pseudoreal), case (iii) no subcase ($\mathcal{U} \cup \mathcal{U}^c$ is complex whenever $\mathcal{U}$ is complex), each of which can be characterized as in (a), (b), (c).

To conclude, we remark that, whenever $\mathcal{U}^{cc}$ coincides with the set $\mathcal{U}_*$ of all the invertible mappings of $\mathcal{U}$, the center $(\mathcal{U} \cup \mathcal{U}^c)^c$ of $\mathcal{U}^c$ coincides with the "invertible center" $\mathcal{U}^c \cap \mathcal{U}$ of $\mathcal{U}$, i.e., with the set of all the invertible mappings contained in the center of $\mathcal{U}$; indeed, we have $(\mathcal{U} \cup \mathcal{U}^c)^c = \mathcal{U}^c \cap \mathcal{U}^{cc} = \mathcal{U}^c \cap \mathcal{U}_* = \mathcal{U}^c \cap \mathcal{U}$ (the last equality holds because $\mathcal{U}^c$ consists of invertible mappings only). This occurs in many interesting cases; for instance, whenever X is finite-dimensional, Λ is commutative and $\mathcal{U}$ is the set of the linear and antilinear mappings of an algebra with identity over Λ (i.e., over R whenever $K = C$) and it is irreducible (the proof of this statement requires the use of the bicommutant theorem of Ref. 12[16] and some simple results about the commutants of irreducible sets of mappings[19]). Thus, in particular, it occurs whenever $\mathcal{U}$ is the set of all the linear and antilinear mappings of an algebra over Λ generated by an irreducible linear–antilinear representation.

3. THE LINEAR COMMUTANT OF IRREDUCIBLE REPRESENTATIONS

Definition 3: Let X be a vector space over a division ring K with a conjugation j. Let $\mathcal{S}$ be any semigroup. We call any homomorphism U from $\mathcal{S}$ into the (multiplicative) semigroup of all the linear and antilinear mappings of X a *linear–antilinear representation* of $\mathcal{S}$ in X. We call U^l the "linear part" of U, i.e., the restriction of U to the subsemigroup $U^{-1}(U(\mathcal{S})^l)$.

When referring to linear–antilinear representations, we will make use of the usual definitions and techniques of representation theory.

Then, the following proposition gives a classification of the irreducible representations which are linear–antilinear and satisfy some simple additional requirement according to the form of their linear commutant:

Proposition 1: Let X be a finite dimensional vector space over an algebraically closed field K with a conjugation j.[20] With reference to Definitions 1,2,3 let $\mathcal{S}$ be a semigroup and let U be an irreducible linear–antilinear representation of $\mathcal{S}$ in the vector space X such that the antilinear part $\mathcal{U}^a$ of $\mathcal{U} = U(\mathcal{S})$ either is void or contains an antilinear mapping A together with its inverse (in the latter case $\mathcal{U}$ necessarily contains the identity mapping E of X).

Then, one of the four mutually exclusive cases described by the rows of Table I occurs, and specifically:

(A) Concerning U: First, we classify the possibilities for U as follows:

(i) $U = U^l$ (hence U^l is irreducible; $\mathcal{U}^a$ is void);

(ii) $U \neq U^l$ (i.e., $\mathcal{U}^a$ is nonvoid) and U^l irreducible;

(iii) $U \neq U^l$ (i.e., $\mathcal{U}^a$ is nonvoid) and U^l reducible.

Whenever $U \neq U^l$, $\mathcal{U}^a = \mathcal{U}^l A = A \mathcal{U}^l$.

TABLE I. Classification of the linear–antilinear irreducible representations by means of their linear commutant.

$\mathscr{U}^a$	$\mathscr{U}^1$		$\mathscr{U}'^1$
$\mathscr{U}^a=\varnothing$	$\mathscr{U}^1$ irreducible		$\mathscr{U}'^1=KE$ $\mathscr{U}'^1=\Lambda E$
$\mathscr{U}^a\neq\varnothing$	$\mathscr{U}^1$ reducible [then, $\mathscr{U}^1(s)=\begin{pmatrix}U_1^1(s) & 0\\ 0 & U_2^1(s)\end{pmatrix}$ and $\mathscr{U}^a=A\mathscr{U}^1$ with $A=\begin{pmatrix}0 & A_{12}\\ A_{21} & 0\end{pmatrix}$]	$U_1^1\not\sim U_2^1$	$\mathscr{U}'^1=\left\{\begin{pmatrix}\alpha E_1 & 0\\ 0 & \bar{\alpha}E_2\end{pmatrix}:\alpha\in K\right\}$
		$U_1^1\sim U_2^1$ (then, $U_2^1=T_{12}^{-1}U_1^1T_{12}$, $A_{12}=\eta^{-1}T_{12}A_{21}T_{12}$)	$\mathscr{U}'^1=\left\{\begin{pmatrix}\alpha E_1 & \beta T_{12}\\ \bar{\beta}\eta T_{12}^{-1} & \bar{\alpha}E_2\end{pmatrix}:\alpha,\beta\in K\right\}$

Here, $\varnothing$ is the empty set, $\sim$ denotes equivalence of representations, E_1, E_2 are the identity mappings of X_1, X_2, respectively.

In the possibility (iii) the linear part U^1 of U reduces in two equidimensional irreducible subrepresentations, $U^1=U_1^1\oplus U_2^1$ (hence the dimension of X must be even), and the corresponding matrix representation[18] of A (hence of $\mathscr{U}^a$) is "off-diagonal."

The possibility (iii) splits further as follows:

(iii′) $U\neq U^1$, $U^1=U_1^1\oplus U_2^1$ with inequivalent U_1^1, U_2^1;

(iii″) $U\neq U^1$, $U^1=U_1^1\oplus U_2^1$ with equivalent U_1^1, U_2^1.

In the case (iii″), for any T_{12} such that $U_1^1=T_{12}U_2^1T_{12}^{-1}$ a unique element $\eta\in K$ exists such that, referring to the matrix representation considered above, $\eta A_{12}=T_{12}A_{21}T_{12}$ (moreover, $\eta\in\Lambda$, $\eta\notin\{\lambda\in K:\lambda=\alpha\bar{\alpha},\alpha\in K\}$).

(B) Concerning $\mathscr{U}'^1$: The linear commutant $\mathscr{U}'^1$ is a division algebra over Λ (hence $\mathscr{U}'^1=\mathscr{U}^{cl}\cup\{0\}$) and the four possibilities listed in (A) are characterized by $\mathscr{U}'^1$ according to Table I. Thus, either $\mathscr{U}'^1$ is isomorphic to Λ (row 2 in Table I) or to K (rows 1 and 3 in Table I), i.e., it is a quadratic algebra over Λ,[21] or, finally, to a division algebra of rank 4 over Λ (row 4 in Table I), i.e., whenever the characteristic of Λ is different from 2, it is a division ring of quaternions over Λ.[22,23]

Proof: *Part* (*A*): We observe that, in cases (ii) and (iii), $\mathscr{U}^a=\mathscr{U}^1A=\mathscr{U}^1A^{-1}=A\mathscr{U}^1=A^{-1}\mathscr{U}^1$; this is indeed an easy consequence of the assumption that $\mathscr{U}$ is a semigroup and of the existence of A and A^{-1} in $\mathscr{U}^a$.[24]

Let us consider case (iii), and let X_1, be an irreducible $\mathscr{U}^1$-invariant subspace of X. Then, the set $X_2=AX_1$ is also a nonzero proper subspace of X (which obviously has the same dimension as X_1) and $\mathscr{U}^aX_2=X_1$, $\mathscr{U}^aX_1=X_2$, $\mathscr{U}^1X_2=X_2$ because of the result obtained above. Thus, X_2 is also a $\mathscr{U}^1$-invariant subspace of X (it is obviously irreducible), $X_1\cap X_2$ is void [since $\mathscr{U}$ is irreducible; indeed $\mathscr{U}(X_1\cap X_2)=X_1\cap X_2$] and $X_1\oplus X_2=X$ (again because of irreducibility). This shows that U^1 can be expressed as the sum of two equidimensional subrepresentations, $U^1=U_1^1\oplus U_2^1$. Hence, any linear mapping $L\in\mathscr{U}^1$ can be written in the form

$$L=\begin{pmatrix}L_{11} & 0\\ 0 & L_{22}\end{pmatrix}.$$

Since $AX_1=X_2$ and $AX_2=X_1$, the mapping A can be written in the form

$$A=\begin{pmatrix}0 & A_{12}\\ A_{21} & 0\end{pmatrix}$$

(where A_{12} and A_{21} are antilinear mappings); hence, we easily get, by making use of the equation $\mathscr{U}^a=A\mathscr{U}^1$, that the matrix representation of $\mathscr{U}^a$ is "off-diagonal."

Now, we easily derive the following observations:

(1) Let two linear mappings $M_{12}:X_2\to X_1$ and $M_{21}:X_1\to X_2$ exist such that $L_{11}=M_{12}L_{22}M_{21}$ for any $L\in\mathscr{U}^1$; then, $M_{21}=M_{12}^{-1}$ (hence the representations U_1^1 and U_2^1 are equivalent);

(2) Let T_{12},M_{12} be linear mappings that implement an equivalence between U_1^1 and U_2^1. Then $M_{12}=\alpha T_{12}$, with $\alpha\in K$.

Observation (1) follows immediately by setting $L=E$; observation (2) follows from the equation

$$L_{11}=T_{12}L_{22}T_{12}^{-1}=M_{12}L_{22}M_{12}^{-1}$$

that holds for any $L\in\mathscr{U}^1$, by making use of the Burnside theorem.[25] Then, let U_1^1 and U_2^1 be equivalent and let their equivalence be implemented by the linear mapping T_{12}. Since $ALA\in\mathscr{U}^1$, we get

$$A_{12}L_{22}A_{21}=T_{12}A_{21}L_{11}A_{12}T_{12}^{-1}\quad\text{for any } L\in\mathscr{U}^1,$$

that is,

$$L_{11}=A_{21}^{-1}T_{12}^{-1}A_{12}L_{22}A_{21}T_{12}A_{12}^{-1}\quad\text{for any } L\in\mathscr{U}^1;$$

hence, because of observation (1), $A_{21}T_{12}A_{12}^{-1}=(A_{21}^{-1}T_{12}^{-1}A_{12})^{-1}$ and, because of observation (2), $\eta A_{21}^{-1}T_{12}^{-1}A_{12}=T_{12}$ with $\eta\in K$, i.e., $\bar{\eta}A_{12}=T_{12}A_{21}T_{12}$. Moreover, by substituting this last equation in the previous one, we get $\eta=\bar{\eta}$, i.e., $\eta\in\Lambda$. Finally, let $\alpha\in K$ and let us set

$$Y_\alpha=\left\{\begin{pmatrix}x_1\\ \alpha T_{12}^{-1}x_1\end{pmatrix}:x_1\in X_1\right\};$$

this is a proper subspace of X, and for any $L\in\mathscr{U}^1$ and $y\in Y_\alpha$ we get

$$Ly=\begin{pmatrix}L_{11}x_1\\ \alpha T_{12}^{-1}L_{11}x_1\end{pmatrix}\in Y_\alpha.$$

Moreover, recalling that the matrix representation of $\mathscr{U}^a$ is "off-diagonal", any $B\in\mathscr{U}^a$ can be written in the form

$$\begin{pmatrix}0 & B_{12}\\ \eta T_{12}^{-1}B_{12}T_{12}^{-1} & 0\end{pmatrix},$$

so that

$$By=\bar{\alpha}\begin{pmatrix}B_{12}T_{12}^{-1}x_1\\ \eta\bar{\alpha}^{-1}T_{12}^{-1}B_{12}T_{12}^{-1}x_1\end{pmatrix};$$

hence, should an $\alpha\in K$ exist such that $\eta=\alpha\bar{\alpha}$, then for any $y\in Y_\alpha$, $By\in Y_\alpha$, so that Y_α would be $\mathscr{U}$-invariant, which contradicts our assumptions.

Part (B): The linear commutant $\mathscr{U}'^1$ obviously is a division algebra over Λ because of the irreducibility of $\mathscr{U}$.

Let $\mathscr{U}^1$ be irreducible. Since X is finite-dimensional and K is algebraically closed, the Burnside theorem applies and $(\mathscr{U}^1)'^1=KE$. Whenever $\mathscr{U}^a$ is void [case (i)], then $\mathscr{U}'^1=(\mathscr{U}^1)'^1=KE$ (it follows that $\mathscr{U}'^1$ is a quadratic algebra over Λ since K has dimension two as a vector space over Λ [26]). Whenever $\mathscr{U}^a$ is nonvoid [case (ii)], then $\mathscr{U}'^1=\Lambda E$.

Let U^1 be reducible [case (iii)]. Let $M\in(\mathscr{U}^1)'^1$; then, we get, using matrix representations with respect to the decomposition $X=X_1\oplus X_2$,

$$\text{for any } L\in\mathscr{U}^1,\quad \begin{cases}L_{11}M_{11}=M_{11}L_{11},\\ L_{22}M_{22}=M_{22}L_{22},\end{cases}\quad \begin{cases}L_{11}M_{12}=M_{12}L_{22},\\ L_{22}M_{21}=M_{21}L_{11}.\end{cases}$$

Since U^1_1 and U^1_2 are irreducible, from the Burnside theorem we obtain $M_{11}=\alpha E_1$, $M_{22}=\delta E_2$ with $\alpha,\delta\in K$. Furthermore, according to the Schur lemma, M_{12} either is zero or is an isomorphism; in the latter case M_{12} implements an equivalence of U^1_1 and U^1_2 (a similar statement holds for M_{21}). Then, let us consider the two possibilities.

Let U^1_1 and U^1_2 be not equivalent [case (iii')]. Then $M_{12}=M_{21}=0$; hence

$$(\mathscr{U}^1)'^1=\left\{\begin{pmatrix}\alpha E_1 & 0\\ 0 & \delta E_2\end{pmatrix}:\alpha,\delta\in K\right\}.$$

Since $\mathscr{U}^a=A\mathscr{U}^1$, any $M\in(\mathscr{U}^1)'^1$ also belongs to $\mathscr{U}'^1$ if and only if $AM=MA$; thus, we get through a straightforward calculation $\delta=\bar{\alpha}$, and $\mathscr{U}'^1$ has the form reported in the third row of Table I (hence $\mathscr{U}'^1$ is isomorphic to K).

Let U^1_1 and U^1_2 be equivalent [case (iii")] and let $T_{12}:X_2\to X_1$ be a linear mapping which implements the equivalence between them; then, bearing in mind the observation (2) stated above, we get $M_{12}=\beta T_{12}$ and, analogously, $M_{21}=\gamma T_{12}^{-1}$ with $\beta,\gamma\in K$.

Therefore in this case

$$(\mathscr{U}^1)'^1=\left\{\begin{pmatrix}\alpha E_1 & \beta T_{12}\\ \gamma T_{12}^{-1} & \delta E_2\end{pmatrix}:\alpha,\beta,\gamma,\delta\in K\right\}.$$

As above, for any $M\in(\mathscr{U}^1)'^1$ we get with a straightforward calculation that $M\in\mathscr{U}'^1$ if and only if

$$\begin{cases}\alpha A_{12}=\bar{\delta}A_{12},\\ \bar{\alpha}A_{21}=\delta A_{21},\end{cases}\quad \begin{cases}\bar{\gamma}A_{12}T_{12}^{-1}=\beta T_{12}A_{21},\\ \gamma A_{12}T_{12}^{-1}=\bar{\beta}T_{12}A_{21}.\end{cases}$$

From the first pair of equations we get $\delta=\bar{\alpha}$; from the second pair, by making use of the equation $\eta A_{12}=T_{12}A_{21}T_{12}$, we get $\gamma=\eta\bar{\beta}$ (we recall that $\eta\in\Lambda$). Thus, $\mathscr{U}'^1$ has the form reported in the fourth row in Table I (hence, the division ring $\mathscr{U}'^1$, K being a vector space of dimension two over Λ, is a division algebra of rank 4 over Λ; therefore, $\mathscr{U}'^1$ is a division ring of quaternions over Λ if the characteristic of Λ is not 2.[27] ■

Remark 1: If we refer to the Wigner classification of the irreducible corepresentations over C of a group,[6] we see that the representations belonging to the Wigner types I, II, and III, respectively, fall in the cases (ii), (iii"), and (iii').

Remark 2: By making use of the observation (2) in the proof, one easily gets in the case (iii") that a linear mapping $T'_{12}:X_2\to X_1$ implements the equivalence of U^1_1 and U^1_2 if and only if $T'_{12}=\alpha T_{12}$, with $\alpha\in K$. Thus, the element $\eta'\in K$ such that $\eta' A_{12}=T'_{12}A_{21}T'_{12}$ is connected to η by the equation $\eta'=\alpha\bar{\alpha}\eta$, with $\alpha\in K$.

Furthermore, one can easily verify by a direct calculation that the conditions:

(a) $\eta\in\Lambda$,

(b) for any $\alpha\in K,\eta\neq\alpha\bar{\alpha}$

are the necessary and sufficient conditions in order to make invertible any nonzero mapping of the form

$$\begin{pmatrix}\alpha E_1 & \beta T_{12}\\ \beta\eta T_{12}^{-1} & \bar{\alpha}E_2\end{pmatrix},\quad \text{with } \alpha,\beta\in K.$$

Then, in our case, one could also prove that the properties (a) and (b) hold by making use of the irreducibility of $\mathscr{U}$, which implies that $\mathscr{U}'^1\setminus\{0\}$ is a group.[28]

Remark 3: We observe that, in the case (iii"), the algebra $\mathscr{U}'^1$ over Λ admits as a basis, for any $\epsilon\in K\setminus\Lambda$, the family

$$\left(\begin{pmatrix}E_1 & 0\\ 0 & E_2\end{pmatrix},\begin{pmatrix}\epsilon E_1 & 0\\ 0 & \bar{\epsilon}E_2\end{pmatrix},\begin{pmatrix}0 & T_{12}\\ \eta T_{12}^{-1} & 0\end{pmatrix},\begin{pmatrix}0 & \epsilon T_{12}\\ \bar{\epsilon}\eta T_{12}^{-1} & 0\end{pmatrix}\right).$$

Moreover, if ϵ is such that $\bar{\epsilon}=-\epsilon$ (this is always possible if the characteristic of Λ, hence of K, is not 2 [29]), then $\mathscr{U}'^1$ is a division ring of quaternions over Λ of type $(\epsilon^2,0,\eta)$. In particular, whenever $K=C$, one can choose $\epsilon=i$ (imaginary unit) and $\eta=-1$ (see Remark 2) so that $\mathscr{U}'^1$ is of type $(-1,0,-1)$, i.e., it is a division ring of quaternions in the usual sense.[30]

4. THE LINEAR–ANTILINEAR COMMUTANT OF IRREDUCIBLE REPRESENTATIONS

In the following proposition we cross the classification reported in Sec. 2 with the generalization of the Wigner classification obtained in Sec. 3.

Proposition 2: Let X be a finite-dimensional vector space over an algebraically closed field K with a conjugation j. With reference to Definitions 1–3, let $\mathscr{S}$ be a semigroup and let U be an irreducible linear–antilinear representation of $\mathscr{S}$ in the vector space X such that the antilinear part $\mathscr{U}^a$ of $\mathscr{U}=U(S)$ either is void or contains an invertible mapping together with its inverse. Then, one of the mutually exclusive cases corresponding to the squares of Table II occurs, each of them being characterized by the linear–antilinear centralizer (or commutant group) $\mathscr{U}^c$ of $\mathscr{U}$ [where $\mathscr{U}^c=(\mathscr{U}'^1\cup\mathscr{U}'^a)\setminus\{0\}$] as shown in Table II. (In Table II, J is an antilinear involutory mapping, Ω is a division algebra of rank 4 over Λ, and we set $\Lambda_*=\Lambda\setminus\{0\}$, $K_*=K\setminus\{0\}$, $\Omega_*=\Omega\setminus\{0\}$. Furthermore, the symbols U^1_1, U^1_2, E_1, E_2, T_{12}, η, $\sim$, have the meaning defined in Proposition 1; $\approx$ means group isomorphism. The explanation of the other symbols used in Table II is given in Definitions 1–3 and in the *list of symbols*, introduced in Sec. 2. We recall that, whenever $K=C$, Λ is the real field R, and Ω is the real quaternion field Q.)

TABLE II. Classification of the irreducible linear–antilinear representations by means of the commutant group.

	$\mathscr{U}^{a}=\emptyset$		$\mathscr{U}^{a}\neq\emptyset$	
	$\mathscr{U}^{1}$ irreducible		$\mathscr{U}^{1}$ reducible $(U^{1}=U^{1}_{1}\oplus U^{1}_{2})$	
			$U^{1}_{1}\not\sim U^{1}_{2}$	$U^{1}_{1}\sim U^{1}_{2}$
	$\mathscr{U}'^{1}=KE$	$\mathscr{U}'^{1}=\Lambda E$	$\mathscr{U}'^{1}=\left\{\begin{pmatrix}\alpha E_1 & 0\\ 0 & \bar{\alpha}E_2\end{pmatrix}:\alpha\in K\right\}\approx K$	$\mathscr{U}'^{1}=\left\{\begin{pmatrix}\alpha E_1 & \beta T_{12}\\ \bar{\beta}\eta T_{12}^{-1} & \bar{\alpha}E_2\end{pmatrix}:\alpha,\beta\in K\right\}\approx\Omega$
$\mathscr{U}$ is *potentially real* [see Sec. 2, case (i)]	$\mathscr{U}^{c}=K.\otimes_{\theta}\{E,J\}$ $\approx K.\otimes G_2\not\simeq K.\times G_2$ [$\mathscr{U}\cup\mathscr{U}^{c}$ is *complex*, see Sec. 2, (c)]	$\mathscr{U}^{c}=\Lambda.\times\{E,J\}$ $\approx\Lambda.\times G_2$ [$\mathscr{U}\cup\mathscr{U}^{c}$ is *potentially real*, see Sec. 2, (a)]	$\mathscr{U}^{c}=\mathscr{U}^{c1}\times\{E,J\}$ $\approx K.\times G_2$ [$\mathscr{U}\cup\mathscr{U}^{c}$ is *potentially real*, see Sec. 2, (a)]	$\mathscr{U}^{c}=\mathscr{U}^{c1}\times\{E,J\}$ $\approx\Omega.\times G_2$ [$\mathscr{U}\cup\mathscr{U}^{c}$ is *potentially real*, see Sec. 2, (a)]
			$\mathscr{U}^{c}=\mathscr{U}^{c1}\otimes_{\theta}\{E,J\}$ $\approx K.\otimes G_2\not\simeq K.\times G_2$ [$\mathscr{U}\cup\mathscr{U}^{c}$ is *complex*, see Sec. 2, (c)]	$\mathscr{U}^{c}=\mathscr{U}^{c1}\otimes_{\theta}\{E,J\}$ $\approx\Omega.\otimes G_2\not\simeq\Omega.\times G_2$ [$\mathscr{U}\cup\mathscr{U}^{c}$ is *pseudoreal*, see Sec. 2, (b)]
$\mathscr{U}$ is *pseudoreal*, [see Sec. 2, case (ii)]	For any $A\in\mathscr{U}^{ca}$, $\mathscr{U}^{c}=K.E\cup K.A\not\simeq K.\otimes G_2$ $\theta(A)$ is not identical [$\mathscr{U}\cup\mathscr{U}^{c}$ is *complex*, see Sec. 2, (c)]	For any $A\in\mathscr{U}^{ca}$, $\mathscr{U}^{c}=\Lambda.E\cup\Lambda.A\not\simeq\Lambda.\otimes G_2$ $\theta(A)$ is identical [$\mathscr{U}\cup\mathscr{U}^{c}$ is *pseudo real*, see Sec. 2, (b)]	For any $A\in\mathscr{U}^{ca}$, $\mathscr{U}^{c}=\mathscr{U}^{c1}\cup\mathscr{U}^{c1}A\not\simeq K.\otimes G_2$ $\theta(A)$ is not identical [$\mathscr{U}\cup\mathscr{U}^{c}$ is *complex*, see Sec. 2, (c)]	
$\mathscr{U}$ is *complex* (see Sec. 2, case (iii))	$\mathscr{U}^{c}=K.E$	$\mathscr{U}^{c}=\Lambda.E$	$\mathscr{U}^{c}=\mathscr{U}^{c1}\approx K.$	$\mathscr{U}^{c}=\mathscr{U}^{c1}\approx\Omega.$

Proof: First, we observe that $\mathscr{U}^c = (\mathscr{U}'^1 \cup \mathscr{U}'^a) \setminus \{0\}$ since $\mathscr{U}$ is irreducible[28] and that, for any $A \in \mathscr{U}^{ca}$, $\mathscr{U}^{ca} = A\mathscr{U}^{cl}$.[31] Moreover, we remark that, for any $A \in \mathscr{U}^c$ the mapping $\theta(A): L \in \mathscr{U}'^1 \to ALA^{-1} \in \mathscr{U}'^1$ is an automorphism of the division ring $\mathscr{U}'^1$. Then, we consider the four mutually exclusive cases than can occur, according to Proposition 1.

(i) Let $\mathscr{U}'^1 = KE$. Whenever $\mathscr{U}^{ca}$ is nonvoid, for any $A \in \mathscr{U}^{ca}$ and $L = \alpha E \in \mathscr{U}'^1$ we get $ALA^{-1} = \bar{\alpha}E$. Therefore, the mapping $\theta(A)$ coincides, up to a trivial isomorphism, with the automorphism j of K; hence, $\theta(A)$ is not inner. Then, the first column of Table II immediately follows by comparison with the results reported in Sec. 2.

(ii) Let $\mathscr{U}'^1 = \Lambda E$. For any $A \in \mathscr{U}^c$ and $L = \lambda E \in \mathscr{U}'^1$ we get $ALA^{-1} = \lambda E = L$; therefore, the mapping $\theta(A)$ is the identity. Then, the second column of Table II follows again by comparison with the results reported in Sec. 2.

(iii') Let

$$\mathscr{U}'^1 = \left\{ \begin{pmatrix} \alpha E_1 & 0 \\ 0 & \bar{\alpha} E_2 \end{pmatrix} : \alpha \in K \right\};$$

then, $\mathscr{U}'^1$ is isomorphic to K and the mapping $\theta(A)$ either is the identity for any $A \in \mathscr{U}^{ca}$ or is not inner. If $\theta(A)$ is the identity, $\mathscr{U}^c$ is commutative and, since K is algebraically closed, for any $A \in \mathscr{U}^{ca}$ a mapping $L_A \in \mathscr{U}^{cl}$ such that $L_A^2 = A^2$ exists, so that $\mathscr{U}^c$ must be isomorphic to the direct product of $\mathscr{U}^{cl}$ and G_2[32]; if $\theta(A)$ is not inner, one of the remaining cases listed in the third column of Table II occurs (see again the results reported in Sec. 2).

(iii") Let

$$\mathscr{U}'^1 = \left\{ \begin{pmatrix} \alpha E_1 & \beta T_{12} \\ \bar{\beta}\eta T_{12}^{-1} & \bar{\alpha} E_2 \end{pmatrix} : \alpha, \beta \in K \right\};$$

then, $\mathscr{U}'^1$ is a simple algebra of finite rank over the field Λ, and, for any $A \in \mathscr{U}^c$, the center ΛE of $\mathscr{U}'^1$ is invariant under $\theta(A)$. Thus, because of a known corollary of the Skolem–Noether theorem,[33] $\theta(A)$ is inner. Hence, the last column of Table II follows by comparison with the results reported in Sec. 2, except for the impossibility of the case "$\mathscr{U}$ pseudoreal," which requires a supplementary discussion.

Let us prove that a mapping $J \in \mathscr{U}^{ca}$ such that $J^2 = E$ necessarily exists if $\mathscr{U}'^1$ has the form given above. Indeed, by making use of $\mathscr{U}^{ca}$ being equal to $A\mathscr{U}^{cl}$ for any $A \in \mathscr{U}^{ca}$ and of $\theta(A)$ being inner, we get that for any $L \in \mathscr{U}^{cl}$ a mapping $A_L \in \mathscr{U}^{ca}$ exists such that $\theta(L) = \theta(A_L)$. Now, let $\epsilon \in K \setminus \Lambda$, let us set

$$L = \begin{pmatrix} \epsilon E_1 & 0 \\ 0 & \bar{\epsilon} E_2 \end{pmatrix},$$

and let us consider the mapping $\theta(A_L)$. The $\theta(A_L)$-invariant subfield of $\mathscr{U}'^1$ is

$$\mathscr{K} = \{ \alpha_1 E + \alpha_2 L : \alpha_1, \alpha_2 \in \Lambda \}.$$

This field is isomorphic to K. Since $\theta(A_L)(A_L^2) = A_L^2$, it follows $A_L^2 \in \mathscr{K}$; thus, $\mathscr{K}$ being algebraically closed, a square root M_L of A_L^2 exists in $\mathscr{K}$ and $A_L M_L = M_L A_L$ [since $\mathscr{K}$ is $\theta(A_L)$-invariant]. Therefore, a mapping $J = M_L^{-1} A_L \in \mathscr{U}^{ca}$ exists such that $J^2 = E$,[34] so that the case in the second line of Table II never occurs. ■

Remark 4: We observe that the case $\mathscr{U}^c \simeq \Omega_* \times G_2$ can only occur if $\dim X \geqslant 4$. Indeed, let $\dim X = 2$ and $\mathscr{U}'^1 \simeq \Omega_*$. Should $\mathscr{U}^c$ be isomorphic to $\Omega_* \times G_2$, then a basis of X over K would exist (see Sec. 2, statement (a)) in which $\mathscr{U}'^1$ is represented by the ring of matrices

$$\left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : \alpha, \beta, \gamma, \delta \in \Lambda \right\},$$

which is not a division ring.

5. EXAMPLES

Now, we want to show with simple examples that all the cases listed in Table II actually occur.

In what follows, we will always set $K = C$, so that j is complex conjugation and $\Lambda = R$.

As a premise, we would like to prove two remarks which hold in this case.

Remark 5: We recall that whenever $\mathscr{U}$ is a semigroup of mappings which satisfies the requirements specified in Proposition 1, $\mathscr{U}'^1$ is isomorphic to R, or to C, or to Q.

Let $\mathscr{U}'^1 = RE$. Then, the statements (a) a mapping $J \in \mathscr{U}^{ca}$ exists such that $J^2 = E$ [see Sec. 2, case (i)] and (b) its converse [see Sec. 2, cases (ii) and (iii)] can be characterized respectively as follows:

(a') for any $B \in \mathscr{U}^{ca}$, $B^2 \in R^+ E$,

(b') for any $B \in \mathscr{U}^{ca}$, $B^2 \in R^- E$

(here, R^+ and R^- respectively are the positive and the negative reals). Analogously, whenever $\mathscr{U}'^1$ is isomorphic to C (we call ϕ the isomorphism $\phi: \mathscr{U}'^1 \to C$) and for any $B \in \mathscr{U}^{ca}$ the mapping $\theta(B): L \in \mathscr{U}'^1 \to BLB^{-1} \in \mathscr{U}'^1$ is not the identity [i.e., case (c) of Sec. 2 occurs], then the statements (a) and (b) above can be characterized respectively as follows:

(a") for any $B \in \mathscr{U}^{ca}$, $\phi(B^2) \in R^+$,

(b") for any $B \in \mathscr{U}^{ca}$, $\phi(B^2) \in R^-$.

Remark 6: The cases in which $\mathscr{U}^{ca}$ is nonvoid and a mapping $J = \mathscr{U}^{ca}$ such that $J^2 = E$ does not exist can only occur either if $\dim X \geqslant 2$ (whenever $\mathscr{U}'^1 = RE$ or $\mathscr{U}'^1 = CE$) or if $\dim X \geqslant 4$ (whenever

$$\mathscr{U}'^1 = \left\{ \begin{pmatrix} \alpha E_1 & 0 \\ 0 & \bar{\alpha} E_2 \end{pmatrix} : \alpha \in C \right\}).$$

Whenever $\mathscr{U}'^1$ is isomorphic to R, the proof of Remark 5 is immediate. Whenever $\mathscr{U}'^1$ is isomorphic to C, the proof can be carried out through the following steps: (α) for any $B \in \mathscr{U}^{ca}$, $\phi(B^2) \in R$ [since B^2 is $\theta(B)$-invariant]; (β) if a $B \in \mathscr{U}^{ca}$ exists such that $\phi(B^2) \in R^+$, then the image of the square of any mapping of $\mathscr{U}^{ca}$ is in R^+ (since $\mathscr{U}^{ca} = B\mathscr{U}^{cl}$); ($\gamma$) statement (v) of Proposition 2 in Ref. 13 is equivalent to the statement that a mapping $B \in \mathscr{U}^{ca}$ exists such that $\phi(B^2) \in R^+$.

Remark 5 can then be used to prove Remark 6 (some easy additional considerations are required whenever

$$\mathscr{U}'^1 = \left\{ \begin{pmatrix} \alpha E_1 & 0 \\ 0 & \bar{\alpha} E_2 \end{pmatrix} : \alpha \in C \right\}).$$

We now come back to the problem of giving an example for each of the cases listed in Table II. We consider separate-

ly the four columns in the table and, for any row, we choose an example with $\dim X$ as small as possible (see Sec. 4, Remark 4 and, above, Remark 6). Furthermore, the sets of mappings are given in matrix form (we represent any antilinear mapping by a matrix followed by the conjugation j [18]).

(i) First column of Table II ($\mathscr{U}^{a}=\emptyset$):

row 1, X one-dimensional:

$$\mathscr{U}=R,^{35}$$

$$\mathscr{U}^{c}=C_{\bullet}\cup C_{\bullet}\,j;$$

row 2, X two-dimensional (see Remark 6):

$$\mathscr{U}=\mathscr{U}^{1}=\left\{\begin{pmatrix}\alpha & \beta\\ -\bar{\beta} & \bar{\alpha}\end{pmatrix}:\ \alpha,\beta\in C\right\}^{35}$$

(we observe that $\mathscr{U}^{1}$ is a division ring isomorphic to Q),

$$\mathscr{U}^{cl}=C_{\bullet}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad \mathscr{U}^{ca}=C_{\bullet}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}j;$$

row 3, X one-dimensional:

$$\mathscr{U}=C,^{35}$$

$$\mathscr{U}^{c}=C_{\bullet}.$$

(ii) Second column of Table II:

row 1, X one-dimensional:

$$\mathscr{U}=R\cup R\alpha j\quad (0\neq\alpha\in C),^{35}$$

$$\mathscr{U}^{c}=\mathscr{U}\setminus\{0\};$$

row 2, X two-dimensional (see Remark 6):

$$\mathscr{U}=\left\{\begin{pmatrix}\alpha & \beta\\ -\bar{\beta} & \bar{\alpha}\end{pmatrix}:\ \alpha,\beta\in C\right\},\quad \mathscr{U}^{a}=\mathscr{U}^{1}j$$

$$\mathscr{U}^{cl}=R_{\bullet}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad \mathscr{U}^{ca}=R_{\bullet}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}j;$$

row 3, X one-dimensional:

$$\mathscr{U}=C\cup Cj,^{35}$$

$$\mathscr{U}^{c}=R_{\bullet}.$$

(iii) Third column of Table II:

row 1, first case: X two-dimensional:

$$\mathscr{U}^{1}=\left\{\begin{pmatrix}\alpha & 0\\ 0 & \bar{\alpha}\end{pmatrix}:\ \alpha\in C\right\},\quad \mathscr{U}^{a}=\mathscr{U}^{1}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}j,$$

$$\mathscr{U}^{c}=\mathscr{U}\setminus\{0\};$$

row 1, second case: X two-dimensional:

$$\mathscr{U}^{1}=\left\{\begin{pmatrix}r & 0\\ 0 & s\end{pmatrix}:\ r,s\in R\right\},\quad \mathscr{U}^{a}=\mathscr{U}^{1}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}j$$

$$\mathscr{U}^{cl}=\left\{\begin{pmatrix}\alpha & 0\\ 0 & \bar{\alpha}\end{pmatrix}:\ \alpha\in C_{\bullet}\right\},\quad \mathscr{U}^{ca}=\mathscr{U}^{cl}j;$$

row 2, X four-dimensional (see Remark 6):

$$\mathscr{U}^{1}=\left\{\begin{pmatrix}\alpha & \beta & 0 & 0\\ -\bar{\beta} & \bar{\alpha} & 0 & 0\\ 0 & 0 & \gamma & \delta\\ 0 & 0 & -\bar{\delta} & \bar{\gamma}\end{pmatrix}:\ \alpha,\beta,\gamma,\delta\in C\right\},$$

$$\mathscr{U}^{a}=\mathscr{U}^{1}\begin{pmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix}j,$$

$$\mathscr{U}^{cl}=\left\{\begin{pmatrix}\alpha & 0 & 0 & 0\\ 0 & \alpha & 0 & 0\\ 0 & 0 & \bar{\alpha} & 0\\ 0 & 0 & 0 & \bar{\alpha}\end{pmatrix}:\ \alpha\in C_{\bullet}\right\},$$

$$\mathscr{U}^{ca}=\mathscr{U}^{cl}\begin{pmatrix}0 & 1 & 0 & 0\\ -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & -1 & 0\end{pmatrix}j;$$

row 3, X two-dimensional:

$$\mathscr{U}^{1}=\left\{\begin{pmatrix}\alpha & 0\\ 0 & \beta\end{pmatrix}:\ \alpha,\beta\in C\right\},\quad \mathscr{U}^{a}=\mathscr{U}^{1}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}j,$$

$$\mathscr{U}^{c}=\mathscr{U}^{cl}=\left\{\begin{pmatrix}\alpha & 0\\ 0 & \bar{\alpha}\end{pmatrix}:\ \alpha\in C_{\bullet}\right\}.$$

(iv) Fourth column of Table II:

row 1, first case, X four-dimensional (see Remark 4):

$$\mathscr{U}^{1}=\left\{\begin{pmatrix}\alpha & \beta & 0 & 0\\ -\bar{\beta} & \bar{\alpha} & 0 & 0\\ 0 & 0 & \bar{\alpha} & -\bar{\beta}\\ 0 & 0 & \beta & \alpha\end{pmatrix}:\ \alpha,\beta\in C\right\},$$

$$\mathscr{U}^{a}=\mathscr{U}^{1}\begin{pmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\\ 1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\end{pmatrix}j,$$

$$\mathscr{U}^{cl}=\left\{\begin{pmatrix}\alpha & 0 & 0 & \beta\\ 0 & \alpha & \beta & 0\\ 0 & -\bar{\beta} & \bar{\alpha} & 0\\ -\bar{\beta} & 0 & 0 & \bar{\alpha}\end{pmatrix}:\ \alpha,\beta\in C\right\}\setminus\{0\},$$

$$\mathscr{U}^{ca}=\mathscr{U}^{cl}\begin{pmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\\ 1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\end{pmatrix}j;$$

row 1, second case, X two-dimensional: see the example for column 2, row 2, and interchange $\mathscr{U}\setminus\{0\}$ with $\mathscr{U}^{c}$ [16];

row 3, X two-dimensional: see the example for column 1, row 2, and interchange $\mathscr{U}\setminus\{0\}$ with $\mathscr{U}^{c}$.[16]

ACKNOWLEDGMENT

We wish to thank Professor Renato Ascoli for all his help and suggestions and, in particular, for his contribution to the revised version of the present paper.

[1] G. Frobenius and I. Schur, Berl. Berichte, 186 (1906).
[2] I. Smirnov, *Linear Algebra and Group Theory*, edited by R. A. Silberman (McGraw-Hill, New York, 1961), Chap. 8, Problems.
[3] E. P. Wigner, *Group Theory* (Academic, New York, 1959), Sec. 24.
[4] M. Hammermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962), Chap. 5.

[5]G. Ya Lyubarskii, *The Application of the Group Theory in Physics* (Pergamon, Oxford, 1960), Sec. 23.
[6]See Ref. 3, Sec. 26.
[7]F. J. Dyson, "The Threefold Way. Algebraic Structure of Symmetry Groups and Ensembles in Quantum Mechanics," J. Math. Phys. **3**, 1199 (1962).
[8]H. Weyl, *The Classical Groups, Their Invariants and Representations* (Princeton U. P. Princeton, N. J., 1939).
[9]G. Frobenius, J. Reine Angew. Math. **81**, 59 (1878).
[10]R. Ascoli, C. Garola, L. Solombrino, and G. Teppati, "Real versus Complex Representations and Linear-Antilinear Commutant," in *Physical Reality and Mathematical Description*, edited by C. P. Enz and J. Mehra (Reidel, Dordrecht, 1974), p. 239.
[11]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, Mass., 1976), §2-4 and §3-2.
[12]R. Ascoli, C. Garola, L. Solombrino, and G. Teppati, "Vector Spaces over Fields with a Conjugation and Linear-Antilinear Commutants," Rend. Mat. Ser. VI **10**(1), 129 (1977), note 4.
[13]C. Garola and L. Solombrino, "Commutation in Vector Spaces over Division Rings with a Conjugation," to appear in *Linear Algebra and Its Applications* (1980).
[14]See Ref. 7, Sec. III.
[15]We remark that no trivial correspondence between our classification and Dyson's appears (for instance, if $\mathscr{U}^{\prime 1}$ is isomorphic to Q a representation which Dyson classifies as "pseudoreal" is bound to be complex according to our classification; moreover, the distinction between Dyson's cases CC1 and CC2 is irrelevant from our point of view).
[16]See Ref. 12, §6, Theorem 5.
[17]We advice the reader that the terminology introduced here has not the same meaning that it has in Dyson's paper (see Ref. 7), where the "potentially real," "pseudoreal," and "complex" case, respectively, occur whenever $\mathscr{U}^1$ (and not the whole of $\mathscr{U}$) is potentially real, pseudoreal, complex in the sense defined above.
[18]See Ref. 10, Sec. 4.
[19]See Ref. 13, Sec. 3, Definition 4 and Proposition 5.
[20]According to the classification introduced in a previous work (see Ref. 12, §4, Theorem 3), K is of type I since it is commutative.
[21]N. Bourbaki, *Eléments de mathématique, Algébre* (Hermann, Paris, 1970), Chaps. 1–3, A III, §2, n°3.
[22]See Ref. 21, A III, §2, n°5.
[23]Whenever $K = C$, $\mathscr{U}^{\prime 1}$ is isomorphic to R (real field) or to C or to Q (real quaternions). These isomorphisms are well known, see Ref. 7 and R. Ascoli and G. Teppati, "On the Commutant of Irreducible Sets of Semilinear Operators," Nuovo Cimento B **65**(2), 229 (1970).
[24]See Ref. 13, Sec. 2, Proposition 1, first remark in the proof.
[25]See, for instance S. Lang, *Algebra* (Addison-Wesley, New York, 1967), Chap. XVII, §3, Corollary 1, first statement in the proof.
[26]See Ref. 12, §2, Theorem 1.
[27]N. Bourbaki, *Eléments de mathématique, Algèbre*, Chap. 8: "Modules et anneaux semi-simple" (Hermann, Paris, 1973), §11, n°2, Proposition 1.
[28]See Ref. 13, Sec. 3, Proposition 5.
[29]See Ref. 12, Footnote 5.
[30]See Ref. 19, A III, §2, n°5, Example 2 and Ref. 23, Chap. XVIII, Exercises.
[31]See Ref. 13, Sec. 2, Proposition 1.
[32]See Ref. 13, Sec. 2, Proposition 4.
[33]See Ref. 27, §10, n°1, Corollaire.
[34]See also Ref. 13, Sec. 2, Proposition 2.
[35]See Ref. 13, Sec. 5.

Matrix orthogonal polynomials on the unit circle

Jeffrey S. Geronimo
Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia 30332

(Received 17 September 1980; accepted for publication 21 January 1981)

The properties of matrix orthogonal polynomials on the unit circle are investigated beginning with their recurrence formulas. The techniques of scattering theory and Banach algebras are used in the investigation. A matrix generalization of a theorem of Baxter is proved.

PACS numbers: 02.30.Gp

I. INTRODUCTION

Recently Geronimo and Case[1,2] have used the techniques of scattering theory and Banach algebras to study the properties of scalar orthogonal polynomials on the unit circle (SOPUC) and real line (SOPRL). These techniques have provided a unified basis for studying the properties of SOPUC and SOPRL and they have shed light on the close connection between these two different systems of polynomials.

This article is the first in a series of two articles in which we use the techniques of scattering theory and Banach algebras to study the properties of matrix orthogonal polynomials, on the unit circle (MOPUC) and, on the real line (MOPRL). Many of the properties of MOPUC and their relation to the planar-least squares problem have been investigated by Delsarte *et al.*,[3-5] and we shall call upon their work (especially Ref. 3) many times. Our main concern in this and the following article will be to investigate the properties of MOPUC and MOPRL beginning with the recurrence formulas satisfied by these systems. As in the scalar case the techniques of scattering theory and Banach algebras provide a unified basis for studying the properties of MOPUC and MOPRL and exhibit the close connection between these two systems.

In order to help motivate our discussion we begin in Sec. II with a Hermitian matrix-valued spectral function $\rho(\theta)$ on the interval $[-\pi,\pi]$. Two dual sets of matrix polynomials orthonormal with respect to $d\rho(\theta)$ are developed and the recurrence formulas they satisfy are derived. The recurrence formulas are used to obtain the Christoffel-Darboux formulas. In Sec. III we take the recurrence formulas as fundamental. The techniques of Banach algebras and scattering theory are introduced and used to discuss the properties of the Szegö–Jost function. A number of other useful solutions to the recurrence formulas are introduced including polynomial matrices of the second kind. Finally in Sec. IV it is shown that the matrix polynomials are orthogonal with respect to some spectral function. Also discussed is a generalization of a theorem of Baxter[6] which exhibits the close connection between the decay of the coefficients in the recurrence formulas and the decay of the Fourier coefficients of the weight. Explicit integral representations of many of the functions introduced in Sec. III are derived.

II. PRELIMINARIES

Let $\rho(\theta)$ be a bounded, nondecreasing $l\times l$ Hermitian matrix on the interval $[-\pi,\pi]$ with an infinite number of points of increase.[7] By nondecreasing it is meant $\rho(\theta_1)\leqslant\rho(\theta_2)$ for all $\theta_1\leqslant\theta_2$. (The notation $A\leqslant B$ for Hermitian matrices A and B means $B-A$ is nonnegative definite.)

Let $\{C_k\}_{k=-\infty}^{\infty}$ be the matrix Fourier coefficients associated with $\rho(\theta)$, i.e.,

$$C_k=(1/2\pi)\int_{-\pi}^{\pi}e^{-ik\theta}d\rho(\theta),\quad k=0,\pm1,\pm2,\cdots. \tag{2.1}$$

Since $\rho(\theta)$ is Hermitian it follows that

$$C_{-k}=C_k^{\dagger}. \tag{2.2}$$

Using the C_k a square block-Töeplitz matrix $T_n(\rho)$ of order $l(n+1)$ can be constructed as follows:

$$T_n(\rho)=\begin{bmatrix} C_0 & C_{-1} & C_{-2} & \cdots & C_{-n} \\ C_1 & C_0 & C_{-1} & \cdots & C_{1-n} \\ \vdots & & & & \\ C_n & C_{n-1} & C_{n-2} & \cdots & C_0 \end{bmatrix},$$

$$n=0,1,2,\cdots. \tag{2.3}$$

From the properties of $\rho(\theta)$ one deduces that $T_n(\rho)$ is positive definite. Consequently

$$D_n(\rho)=\det[T_n(\rho)]>0. \tag{2.4}$$

It is now shown that $\rho(\theta)$ determines, up to a constant right unitary factor, a set of matrix polynomials $\{\phi^{R}(z,n)\}$ with the following properties (for alternative developments see Youla and Kazanjian[8] and Delarte *et al.*[3]).

(a) $\phi^{R}(z,n)$ is a polynomial in z of degree n with the leading coefficient a $l\times l$ nonsingular matrix.

(b) $(1/2\pi)\int_{-\pi}^{\pi}\phi^{R}(e^{i\theta},n)^{\dagger}d\rho(\theta)\phi^{R}(e^{i\theta},m)=I\delta_{n,m}\quad z=e^{i\theta}$.

To see this, assume that (a) and (b) hold and write

$$\phi^{R}(e^{i\theta},n)=\sum_{m=0}^{n}K(n,m)e^{im\theta}. \tag{2.5}$$

Multiplying on the left by $e^{-ip\theta}d\rho(\theta)$ and integrating gives

$$\sum_{m=0}^{n}C_{p-m}K(n,m)=K^{\dagger}(n,n)^{-1}\delta_{n,p},\quad 0\leqslant p\leqslant n, \tag{2.6}$$

or

$$\sum_{m=0}^{n}C_{p-m}H(n,m)=I\delta_{n,p},\quad 0\leqslant p\leqslant n, \tag{2.7}$$

where I is a $l\times l$ identity matrix and

$$H(n,m)=K(n,m)K^{\dagger}(n,n). \tag{2.8}$$

 0022-2488/81/071359-07$1.00

The above equations are equivalent to

$$\begin{bmatrix} C_0 & C_{-1} & C_{-2} & \cdots & C_{-n} \\ C_1 & C_0 & C_{-1} & \cdots & C_{-n+1} \\ \vdots & & & & \\ C_n & C_{n-1} & C_{n-2} & \cdots & C_0 \end{bmatrix} \begin{bmatrix} H(n,0) \\ H(n,1) \\ \vdots \\ H(n,n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ I \end{bmatrix}, \tag{2.9}$$

where all the entries in the above equation are $l \times l$ matrices. In order to solve the above equation it is necessary and sufficient that the determinant of the $l(n+1) \times l(n+1)$ matrix on the lhs of the above equation not vanish. From Eqs. (2.3) and (2.4) one finds that this condition is satisfied. Therefore,

$$\begin{bmatrix} H(n,0) \\ H(n,1) \\ \vdots \\ H(n,n) \end{bmatrix} = T_n^{-1}(\rho) \begin{bmatrix} 0 \\ 0 \\ \vdots \\ I \end{bmatrix}. \tag{2.10}$$

Multiplying the above equation on the left by the $l \times l(n+1)$ matrix $[0,0,\cdots,I]$ and using (2.8) gives

$$K(n,n)K^{\dagger}(n,n) = H(n,n) = [0,0,\cdots,I]\,T_n^{-1}(\rho) \begin{bmatrix} 0 \\ 0 \\ \vdots \\ I \end{bmatrix}. \tag{2.11}$$

Since $T_n^{-1}(\rho)$ is positive definite, the above equation determines $K(n,n)$ up to a right unitary factor. Let us call $\{\phi^{R}(z,n)\}$ the right orthonormal matrix polynomials associated with $\rho(\theta)$. To find the left orthonormal matrix polynomials $\{\phi^{L}(z,n)\}$, change ϕ^{R} to ϕ^{L} in (a) and change (b) to

(b1) $(1/2\pi)\int_{-\pi}^{\pi} \phi^{L}(e^{i\theta},n)d\rho(\theta)\phi^{L}(e^{i\theta},m)^{\dagger} = I\delta_{n,m}$.

Writing

$$\phi^{L}(e^{i\theta},n) = \sum_{m=0}^{n} J(n,m)e^{im\theta}, \tag{2.12}$$

and following steps similar to the ones given above yield

$$[L(n,n),L(n,n-1)\cdots L(n,0)] = [I,0,\cdots,0]\,T_n^{-1}(\rho), \tag{2.13}$$

where

$$L(n,m) = J^{\dagger}(n,m)J(n,m). \tag{2.14}$$

Multiplying (2.13) on the right by the $l(n+1) \times l$ matrix $[I,0,\cdots,0]'$ and using (2.14) gives

$$J^{\dagger}(n,n)J(n,n) = L(n,n) = [I,0,0,\cdots]\,T_n^{-1} \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{2.15}$$

which determines $J(n,n)$ up to a left unitary factor. The useful relation

$$L(n,0) = [I,0,\cdots,0]\,T_n^{-1}(\rho)[0,0,\cdots,I]' = H(n,0), \tag{2.16}$$

can be deduced by multiplying (2.10) on the left by $[I,0,\cdots,0]$ and (2.13) on the right by $[I,0,\cdots,0]'$.

Beginning with $\phi^{R}(z,n)$ one can derive the recurrence formulas satisfied by these polynomials by noting that

$$z^{n-1}[(1/z^n)\phi^{R}(z,n) - \phi^{L}(1/z^*,n)^{\dagger}J^{\dagger}(n,n)^{-1}K(n,0)] = \sum_{i=0}^{n-1} \phi^{R}(z,i)B(i,n-1), \quad n = 1,2,\cdots. \tag{2.17}$$

Multiplying on the left by $\phi^{R}(1/z^*,m)^{\dagger}d\rho(\theta)$, $m < n$, gives, using (b) and (b1),

$$K^{\dagger}(n-1,n-1)K^{\dagger}(n,n)^{-1}\delta_{m,n-1} = B(m,n-1). \tag{2.18}$$

Therefore,

$$\phi^{R}(z,n)K^{\dagger}(n,n) - z^n\phi^{L}(1/z^*,n)^{\dagger}J^{\dagger}(n,n)^{-1}K(n,0)K^{\dagger}(n,n) = z\phi^{R}(z,n-1)K^{\dagger}(n-1,n-1). \tag{2.19}$$

The dual relation

$$J^{\dagger}(n,n)\phi^{L}(z,n) - J^{\dagger}(n,n)J(n,0)K^{\dagger}(n,n)^{-1}z^n\phi^{R}(1/z^*,n)^{\dagger} = J^{\dagger}(n-1,n-1)z\phi^{L}(z,n-1) \tag{2.20}$$

follows in a similar manner.

Following Delsarte *et al.*,[3] it is possible to rewrite (2.19) and (2.20) in terms of a single parameter by defining

$$E_n = J^{\dagger}(n,n)^{-1}H(n,0)K^{\dagger}(n,n)^{-1} = J^{\dagger}(n,n)^{-1}L(n,0)K^{\dagger}(n,n)^{-1}, \tag{2.21}$$

where (2.16) has been used. Substituting the above equation into (2.19) and equating coefficients of z^n yields

$$I - E_n^{\dagger}E_n = K(n,n)^{-1}K(n-1,n-1)K^{\dagger}(n-1,n-1) \times K^{\dagger}(n,n)^{-1}, \tag{2.22}$$

which is positive definite. Equation (2.22) implies

$$\|E_n\|_s < 1, \tag{2.23}$$

where the norm used is the spectral norm (i.e., the magnitude of the largest eigenvalue). Since $K(n,n)$ is defined only up to a right unitary factor, it is convenient to choose

$$(I - E_n^{\dagger}E_n)^{1/2} = K^{\dagger}(n-1,n-1)K^{\dagger}(n,n)^{-1}, \tag{2.24}$$

where the Hermitian square root of $I - E_n^{\dagger}E_n$ is implied. Note that the following identities are easy consequences of (2.23):

$$E_n(I - E_n^{\dagger}E_n)^{1/2} = (I - E_nE_n^{\dagger})^{1/2}E_n, \tag{2.25}$$

$$E_n(I - E_n^{\dagger}E_n)^{-1} = (I - E_nE_n^{\dagger})^{-1}E_n. \tag{2.26}$$

Note that (2.21) and (2.24) determine E_n only up to left and right unitary factors independent of n. If the spectral function is chosen so that C_0 is equal to I and if $\phi^{L}(z,0) = I = \phi^{R}(z,0)$ then the $\{E_n\}$ are uniquely determined.

Using the above equations one can recast (2.18) and (2.19) to read

$$\phi^{L}(z,n) - E_n\tilde{\phi}^{R}(z,n) = z(I - E_nE_n^{\dagger})^{1/2}\phi^{L}(z,n-1) \tag{2.27}$$

and

$$\tilde{\phi}^{R}(z,n) - E_n^{\dagger}\phi^{L}(z,n) = (I - E_n^{\dagger}E_n)^{1/2}\tilde{\phi}^{R}(z,n-1), \tag{2.28}$$

where

$$\tilde{\phi}^{R}(z,n) = z^n\phi^{R}(1/z^*,n)^{\dagger}. \tag{2.29}$$

As in the scalar case, the above equations can be written in the compact form

$$\tilde{\Phi}(z,n) = \tilde{C}(z,n)\tilde{\Phi}(z,n-1), \tag{2.30}$$

with

$$\tilde{\Phi}(z,n-1) = \begin{pmatrix} \phi^{L}(z,n-1) \\ \tilde{\phi}^{R}(z,n-1) \end{pmatrix} \tag{2.31a}$$

and

$$\tilde{C}(z,n) = B(E_n)D(z). \tag{2.32a}$$

Here

$$B(E_n) = \begin{bmatrix} (I-E_nE_n^\dagger)^{-1/2} & E_n(I-E_n^\dagger E_n)^{-1/2} \\ E_n^\dagger(I-E_nE_n^\dagger)^{-1/2} & (I-E_n^\dagger E_n)^{-1/2} \end{bmatrix}, \tag{2.33}$$

and

$$D(z) = \begin{bmatrix} zI & 0 \\ 0 & I \end{bmatrix}. \tag{2.34}$$

As noted by Delsarte *et al.*,[3] $B(E_n)$ has the following interesting properties:

$$B(-E_n) = B(E_n)^{-1}, \quad B^\dagger(E_n)JB(E_n) = J, \tag{2.35}$$

with

$$J = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}. \tag{2.36}$$

The last property classified $B(E_n)$ as a J-unitary matrix.

Equation (2.30) can be recast in the equivalent but more symmetric form

$$\Phi(z,n) = C(z,n)\Phi(z,n-1), \quad n = 1,2,\cdots, \tag{2.37}$$

with

$$\Phi(z,n) = \begin{pmatrix} \phi^{L}(z,n) \\ \phi^{R}(1/z^*,n)^\dagger \end{pmatrix}, \tag{2.31b}$$

and

$$C(z,n) = \begin{bmatrix} z(I-E_nE_n^\dagger)^{-1/2} & E_n(I-E_n^\dagger E_n)^{-1/2}z^{n-1} \\ E_n^\dagger(I-E_nE_n^\dagger)^{-1/2}z^{-n+1} & (I-E_n^\dagger E_n)^{-1/2}z^{-1} \end{bmatrix}. \tag{2.32b}$$

The following is an easy consequence of the J-unitarity of $B(E_n)$:

Theorem 1: Let $\tilde{\Phi}^{(1)}(z_1,n)$ and $\tilde{\Phi}^{(2)}(z_2,n)$ be solutions of (2.30). Then

$$\tilde{\Phi}(1/z_1^*,n)^\dagger \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \tilde{\Phi}^{(2)}(z_2,n) = \tilde{\Phi}^{(1)}(1/z_1^*,n-1)^\dagger \begin{pmatrix} -(z_2/z_1)I & 0 \\ 0 & I \end{pmatrix} \tilde{\Phi}^{(2)}(z_2,n-1). \tag{2.38}$$

Two important special cases are given below.

Corollary 1:

$$L[\tilde{\Phi}^{(1)},\tilde{\Phi}^{(2)}] \equiv \tilde{\Phi}^{(1)}(1/z^*,n)^\dagger \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \tilde{\Phi}^{(2)}(z,n), \tag{2.39a}$$

and

$$L[\Phi^{(1)},\Phi^{(2)}] \equiv \Phi^{(1)}(1/z^*,n)^\dagger \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \Phi^{(2)}(z,n), \tag{2.39b}$$

are independent of n.

Proof: Equation (2.39a) is obtained by setting $z_1 = z_2 = z$ in (2.38). Equation (2.39b) can be proved using either (2.39a) and (2.29) or (2.32b).

Corollary 2 (Christoffel–Darboux formulas):

$$\tilde{\phi}^{R}(z_1,n)^\dagger\tilde{\phi}^{R}(z_2,n) - \phi^{L}(z_1,n)^\dagger\phi^{L}(z_2,n) = (1-z_1^*z_2)\sum_{i=0}^{n-1}\phi^{L}(z_1,i)^\dagger\phi^{L}(z_2,i), \tag{2.40a}$$

$$\tilde{\phi}^{L}(z_1,n)\tilde{\phi}^{L}(z_2,n)^\dagger - \phi^{R}(z_1,n)\phi^{R}(z_2,n)^\dagger = (1-z_1z_2^*)\sum_{i=1}^{n-1}\phi^{R}(z_1,i)\phi^{R}(z_2,i)^\dagger. \tag{2.40b}$$

Proof: Setting $\tilde{\Phi}^{(1)} = \tilde{\Phi}^{(2)}$ and $z_1 = 1/z_1^*$ in (2.38) gives

$$\tilde{\phi}^{R}(z_1,n)^\dagger\tilde{\phi}^{R}(z_2,n) - \phi^{L}(z_1,n)^\dagger\phi^{L}(z_2,n) = \tilde{\phi}^{R}(z_1,n-1)^\dagger\tilde{\phi}^{R}(z_2,n-1) - z_1^*z_2\phi^{L}(z_1,n-1)^\dagger\phi^{L}(z_2,n-1) \tag{2.41}$$

$$= \tilde{\phi}^{R}(z_1,n-1)^\dagger\tilde{\phi}^{R}(z_2,n-1) - \phi^{L}(z_1,n-1)^\dagger\phi^{L}(z_2,n-1) + (1-z_1^*z_2)\phi^{L}(z_1,n-1)^\dagger\phi^{L}(z_2,n-1). \tag{2.42}$$

Equation (2.40a) is obtained by iterating (2.42) down and using the fact that

$$\tilde{\phi}^{R}(z_1,0)^\dagger\tilde{\phi}^{R}(z_2,0) = \phi^{L}(z_1,0)^\dagger\phi^{L}(z_2,0) \tag{2.43}$$

[see (2.11) and (2.14)]. Equation (2.40b) can be obtained by setting $\tilde{\Phi}^{(1)} = \tilde{\Phi}^{(2)}$, $z_1 = 1/z_1^*$ and $z_2 = 1/z_2^*$ in Eq. (2.38), then multiplying by $-(z_1z_2^*)^n$ and following procedures similar to those that led to (2.43). From (2.40) one can easily show (see Delsarte *et al.*[3]) that $\tilde{\phi}^{L}(z,n)$ and $\tilde{\phi}^{R}(z,n)$ are nonsingular for $|z| \leqslant 1$.

For $z_1 = 1/z_2^*$ in (2.40) one has

$$\tilde{\phi}^{R}(1/z^*,n)^\dagger\tilde{\phi}^{R}(z,n) = \phi^{L}(1/z^*,n)^\dagger\phi^{L}(z,n). \tag{2.44}$$

III. THE RECURRENCE FORMULAS

Now let us turn things around and begin with the recurrence formulas. Thus given a sequence $\{E_n\}_{n=1}^{\infty}$ of complex $l \times l$ matrices satisfying Eq. (2.23), one can uniquely construct, using the recurrence formulas (2.37) and initial condition

$$\phi^{R}(z,0) = \phi^{L}(z,0) = I, \tag{3.1}$$

a dual sequence of polynomial matrices $\{\phi^{R}(z,n)\}_{n=0}^{\infty}$ and $\{\phi^{L}(z,n)\}_{n=0}^{\infty}$. Furthermore, Delsarte *et al.*[3] have shown that these polynomial matrices are orthogonal with respect to a unique spectral function $\rho(\theta)$.

To proceed further, it is convenient at this point to introduce the techniques of Banach algebras. Let $v(n)$ be a real even function of n with the following properties:

$$v(0) = 1, \quad v(n) \geqslant 1, \quad \forall n,$$

$$v(n) \leqslant v(m)v(n-m), \quad \forall n,m,$$

and

$$\lim_{n\to\infty}[v(n)]^{1/n} = R \geqslant 1. \tag{3.2}$$

Let a_v be the space of functions such that if $g \in a_v$, then

$$g(z) = \sum_{m=-\infty}^{\infty} c(n) z^n, \quad 1/R \leqslant |z| \leqslant R, \tag{3.3}$$

with

$$\|g\|_v = \sum_{n=-\infty}^{\infty} v(n)|c(n)| < \infty. \tag{3.4}$$

Let a_v^+ and a_v^- denote those functions in a_v of the form

$$g(z) = \sum_{k=0}^{\infty} c(k) z^k, \tag{3.5}$$

and

$$h(z) = \sum_{k=-\infty}^{0} b(k) z^k, \tag{3.6}$$

respectively.

If $\|g\|_v$ is the norm on a_v, a_v^+, and a_v^-, then a_v, and a_v^+, and a_v^- are Banach algebras. We will denote by a the Banach algebra where $v(n) = 1$ for all n. It is obvious that

$$a_v \subset a. \tag{3.7}$$

Let A_v be the class of $l \times l$ matrices such that if $G \in A_v$, then the entries in G are in a_v and

$$G(z) = \sum_{k=-\infty}^{\infty} C(k) z^k, \quad 1/R \leqslant |z| \leqslant R, \tag{3.8}$$

with

$$\|G\|_v = \sum_{k=-\infty}^{\infty} v(k)|C(k)| < \infty. \tag{3.9}$$

Here $C(k)$ is an $l \times l$ matrix. The matrix norm used above can be any Banach space norm with the property

$$|AB| \leqslant |A|\,|B|, \tag{3.10}$$

where A and B are $l \times l$ matrices. For convenience the Hilbert–Schmidt norm, i.e.,

$$|B| = \left(\sum_{ij} |b_{ij}|^2\right)^{1/2} = |B^\dagger|, \tag{3.11}$$

will be used.

A_v^+ and A_v^- will denote the collection of functions in A_v of the form

$$G(z) = \sum_{k=0}^{\infty} C(k) z^k \tag{3.12}$$

and

$$H(z) = \sum_{k=\infty}^{0} B(k) z^k, \tag{3.13}$$

respectively.

If $\|G\|_v$ is the norm on A_v, A_v^+, and A_v^-, then A_v, A_v^+, and A_v^- are Banach algebras. We will denote by A the Banach algebra where $v(n) = 1$ for all n. Again it is clear that

$$A_v \subset A. \tag{3.14}$$

Returning to the recurrence formulas, it will be assumed throughout the rest of the discussion that

$$\sum_{n=0}^{\infty} v(n)|E_n| < \infty. \tag{3.15}$$

It is useful to define

$$\hat{\phi}^{R}(z,n) = \alpha(n)^{-1} \tilde{\phi}^{R}(z,n) \tag{3.16a}$$

and

$$\hat{\phi}^{L}(z,n) = \tilde{\phi}^{L}(z,n) \beta(n)^{-1}, \tag{3.16b}$$

where

$$\alpha(n) = \prod_{i=1}^{n} (I - E_i^\dagger E_i)^{1/2} \equiv (I - E_1^\dagger E_1)^{1/2}(I - E_2^\dagger E_2)^{1/2} \cdots (I - E_n^\dagger E_n)^{1/2} \tag{3.17a}$$

and

$$\beta(n) = \prod_{i=1}^{n} (I - E_i E_i^\dagger)^{1/2} \equiv (I - E_n E_n^\dagger)^{1/2}(I - E_{n-1} E_{n-1}^\dagger)^{1/2} \cdots (I - E_1 E_1^\dagger)^{1/2}, \tag{3.17b}$$

with

$$\alpha(0) = \beta(0) = I.$$

Equations (2.23) and (3.15) imply $0 < |\alpha(\infty)| < \infty$ and $0 < |\beta(\infty)| < \infty$. (See Delsarte *et al.*[4]) Using the above equations, (2.30) can be recast as

$$\hat{\phi}^{R}(z,n) = \hat{\phi}^{R}(z,n-1) + \alpha(n-1)^{-1} E_n^\dagger \beta^\dagger(n-1) z^n \hat{\phi}^{L}(1/z^*, n-1)^\dagger \tag{3.18}$$

and

$$\hat{\phi}^{L}(z,n) = \hat{\phi}^{L}(z,n-1) + z^n \hat{\phi}^{R}(1/z^*, n-1)^\dagger \alpha^\dagger(n-1) E_n^\dagger \beta(n-1)^{-1}. \tag{3.19}$$

These recurrence formulas lead to the following theorem.

Theorem 2: For every $n > 0$

$$\|\hat{\phi}^{R,L}(1/z^*, n)^\dagger\|_v \leqslant \prod_{i=1}^{n} [1 + v(i)|E_i^\dagger| \gamma(i-1)], \tag{3.20}$$

and for $n > m > 0$

$$\|\hat{\phi}^{R,L}(z,n) - \hat{\phi}^{R,L}(z,m)\|_v \leqslant \sum_{k=m}^{n-1} v(k+1)|E_{k+1}^\dagger|\; \|\hat{\phi}^{R,L}(1/z^*, k)^\dagger\|_v \gamma(k), \tag{3.21}$$

with

$$\gamma(i) = \max(|\alpha(i)^{-1}|\;|\beta(i)|,\; |\alpha(i)|\;|\beta(i)^{-1}|), \tag{3.22}$$

and

$$\phi^{R,L} = \phi^{R}, \phi^{L}. \tag{3.23}$$

Proof: Equation (3.21) is an immediate consequence of (3.18) and (3.19) while (3.20) follows from (3.18) and (3.19) by induction.

It is clear that the bounds on $\hat{\phi}^{R,L}(z,n)$ are the same as above.

Corollary 3: If (3.15) holds then $\hat{\phi}^{R}(z,n)$ and $\hat{\phi}^{L}(z,n)$ are bounded in norm for all n and there exists $\hat{f}_+^{L} \in A_v^+$ and $\hat{f}_+^{R} \in A_v^+$ such that

$$\|\hat{\phi}^{R,L}(z,n) - \hat{f}_+^{R,L}(z)\|_v \to 0, \tag{3.24}$$

as $n \to \infty$.

In analogy with scalar polynomials orthonormal on the unit circle (Géronimo and Case[1]),

$$f_+^{L}(z) = \alpha(\infty) \hat{f}_+^{L}(z) \tag{3.25a}$$

and

$$f_+^{R}(z) = \hat{f}_+^{R}(z) \beta(\infty) \tag{3.25b}$$

will be called, respectively, the left and right Szegö–Jost functions. Since the convergence of $\tilde{\phi}^{R,L}(z,n)$ to $\hat{f}_{+}^{R,L}(z)$ is uniform in the disk of radius $R \geqslant 1$, $f_{+}^{R,L}(z)$ are continuous on the disk and analytic in its interior. This, plus the fact that $\tilde{\phi}^{R}(z,n)$ and $\tilde{\phi}^{L}(z,n)$ are nonsingular for $|z| \leqslant 1$, implies that $f_{+}^{R}(z)$ and $f_{+}^{L}(z)$ are nonsingular for $|z| \leqslant 1$.

A second set of matrix polynomials $\{\Psi(z,n)\}_1^\infty$

$$\Psi(z,n) = \begin{pmatrix} \psi^{L}(z,n) \\ -\psi^{R}(1/z^*,n)^\dagger \end{pmatrix}, \tag{3.26}$$

satisfying (2.37b) with initial conditions

$$\psi^{L}(z,1) = (I - E_1 E_1^\dagger)^{-1/2}(z - E_1),$$
$$\psi^{R}(z,1) = (z - E_1)(I - E_1^\dagger E_1)^{-1/2} \tag{3.27}$$

can also be constructed. These will be called matrix polynomials of the second kind.

Theorem 3: For every $n > 1$

$$\|\hat{\psi}^{R,L}(1/z^*,n)^\dagger\|_v \leqslant \prod_{i=1}^{n} [1 + v(i)|E_i|\gamma(i-1)], \tag{3.28}$$

and for every $n > m > 1$

$$\|\hat{\psi}^{R,L}(z,n) - \hat{\psi}^{R,L}(z,n)\|_v \leqslant \sum_{k=m}^{n-1} v(k+1)|E_{k+1}^\dagger|\gamma(k)\|\hat{\psi}^{R,L}(1/z^*,k)\|_v. \tag{3.29}$$

Proof: Use procedures analogous to those used in the proof of Theorem 2.

Corollary 4: If (3.15) holds then $\tilde{\psi}^{R,L}(z,n)$ are bounded in norm for all n and there exist $\hat{f}_{+\alpha}^{R,L}$ in A_v^+ such that

$$\|\hat{\psi}^{R,L}(z,n) - \hat{f}_{+\alpha}^{R,L}(z)\|_v \to 0, \tag{3.30}$$

as $n \to \infty$.

Again $\hat{f}_{+\alpha}^{R,L}(z)$ have extensions that are continuous on and analytic within the circle of radius R.

Theorem 4: If (3.15) holds then there exist two solutions of (2.37b),

$$\Psi_{+}(z,n) = \begin{pmatrix} \psi_{+}(z,n) \\ \phi_{+}(z,n) \end{pmatrix}, \tag{3.31}$$

and

$$\Psi_{-}(z,n) = \begin{pmatrix} \phi_{-}(z,n) \\ \psi_{-}(z,n) \end{pmatrix}, \tag{3.32}$$

with $\psi_{+}(z,n), \phi_{+}(z,n) \in A_{+}$ and $\phi_{-}(z,n), \psi_{-}(z,n) \in A_{-}$ satisfying the boundary conditions

$$\|\psi_{\pm}(e^{i\theta},n) - Ie^{\pm in\theta})\| \to 0 \tag{3.33}$$

and

$$\|\phi_{\pm}(e^{i\theta},n)\| \to 0 \tag{3.34}$$

as $n \to \infty$.

Proof: Inverting (2.37b), then iterating upwards and using the boundary conditions yields the following equations for $\psi_{+}(z,n)$ and $\phi_{+}(z,n)$:

$$\psi_{+}(z,n) = z^n \prod_{i=n+1}^{\infty} (I - E_i E_i^\dagger)^{-1/2} - z^n \sum_{i=n+1}^{\infty} \prod_{j=n+1}^{i} (I - E_j E_j^\dagger)^{-1/2} E_i \phi_{+}(z,i) \tag{3.35}$$

and

$$\phi_{+}(z,n) = -\bar{z}^n \sum_{i=n+1}^{\infty} \prod_{j=n+1}^{i} (I - E_j^\dagger E_j)^{-1/2} E_j^\dagger \psi_{+}(z,i). \tag{3.36}$$

Substituting (3.36) into (3.35) then using successive approximations yield the bound

$$\|\psi_{+}(e^{i\theta},n)\| \leqslant D(n+1) \exp\left(\sum_{i=n+1}^{\infty} |E_i| \right)^2, \tag{3.37}$$

where

$$D(n+1) = \prod_{j=n+1}^{\infty} |(I - E_j E_j^\dagger)^{-1/2}|\,|(I - E_j^\dagger E_j)^{-1/2}|. \tag{3.38}$$

This leads to

$$\|\phi_{+}(e^{i\theta},n)\| \leqslant D(n+1) \sum_{i=n+1}^{\infty} |E_j^\dagger| \exp\left(\sum_{j=i+1}^{\infty} |E_i| \right)^2 \tag{3.39}$$

and

$$\|\psi_{+}(e^{i\theta},n) - e^{in\theta} \prod_{i=n+1}^{\infty} (I - E_i E_i^\dagger)^{-1/2}\| \leqslant D(n+1)\left(\sum_{i=n+1}^{\infty} |E_i| \right)^2 \exp\left(\sum_{j=i+1}^{\infty} |E_i| \right)^2. \tag{3.40}$$

Letting $z \to 1/z$, $\psi_{+} \to \psi_{-}$, $\phi_{+} \to \phi_{-}$, and $E_n \to E_n^\dagger$ in (3.35) and (3.36) gives equations satisfied by ψ_{-} and ϕ_{-}. These in turn lead to inequalities for ψ_{-} and ϕ_{-} similar to those above.

Since $\Psi_{+}(z,n)$ and $\Psi_{-}(x,n)$ are linearly independent, one can write

$$\Phi(z,n) = \Psi_{-}(z,n)A(z) + \Psi_{+}(z,n)B(z), \quad |z| = 1, \tag{3.41}$$

where each component of $\Psi_{-}(z,n)$ and $\Psi_{+}(z,n)$ is multiplied by $A(z)$ and $B(z)$ respectively. Matrices A and B can be evaluated using (2.39b). Thus

$$A(z) = L[\Psi_{-},\Phi] = \psi_{-}(1/z^*,n)^\dagger \phi^{R}(1/z^*,n)^\dagger - \phi_{-}(1/z^*,n)^\dagger \phi^{L}(z,n) = f_{+}^{R}(z). \tag{3.42}$$

Here Corollary 1, (3.33), (3.34), and (3.25) have been used. Likewise

$$B(z) = -L[\Psi_{+},\Phi] = \psi_{+}(1/z^*,n)^\dagger \phi^{L}(z,n) - \phi_{+}(1/z^*,n)^\dagger \phi^{R}(1/z^*,n)^\dagger = f_{+}^{L}(1/z^*)^\dagger. \tag{3.43}$$

For polynomials of the second kind one has

$$\Psi(z,n) = \Psi_{+}(z,n)f_{+\alpha}^{L}(1/z^*)^\dagger - \Psi_{-}(z,n)f_{+\alpha}^{R}(z), \quad |z| = 1, \tag{3.44}$$

where

$$f_{+\alpha}^{L}(1/z^*)^\dagger = \psi_{+}(1/z^*,n)^\dagger \psi^{L}(z,n) + \phi_{+}(1/z^*,n)^\dagger \psi^{R}(1/z^*,n)^\dagger \tag{3.45}$$

and

$$f_{+\alpha}^{R}(z) = \psi_{-}(1/z^*,n)^\dagger \psi^{R}(1/z^*,n)^\dagger + \phi_{-}(1/z^*,n)^\dagger \psi^{L}(z,n). \tag{3.46}$$

Similarly

$$\Psi_{+}(z,n) = \tfrac{1}{2}[\Psi(z,n)f_{+}^{L}(z) + \Phi(z,n)f_{+\alpha}^{L}(z)], \tag{3.47}$$

and

$$\Psi_{-}(z,n) = \tfrac{1}{2}\left[\Phi(z,n) f^{R}_{+\alpha}(1/z^{*})^{\dagger} - \Psi(z,n) f^{R}_{+}(1/z^{*})^{\dagger}\right], \quad (3.48)$$

where (3.42,43,45,46) and (2.44) have been used.

IV. THE SPECTRAL FUNCTION

With the above material one can prove the following theorem.

Theorem 5 (orthogonality): If (3.15) holds then

$$(1/2\pi)\int_{-\pi}^{\pi} \phi^{L}(e^{i\theta},n)\, d\rho(\theta)\, \phi^{L}(e^{i\theta},m)^{\dagger} = I\delta_{n,m}$$

and

$$(1/2\pi)\int_{-\pi}^{\pi} \phi^{R}(e^{i\theta},n)^{\dagger}\, d\phi(\theta)\, \phi^{R}(e^{i\theta},m) = I\delta_{n,m}, \quad (4.1)$$

where

$$d\rho(\theta) = \sigma(\theta)\, d\theta, \quad (4.2)$$

almost everywhere with

$$\sigma(\theta) = \left[f^{L}_{+}(e^{i\theta}) f^{L}_{+}(e^{i\theta})^{\dagger}\right]^{-1} = \left[f^{R}_{+}(e^{i\theta})^{\dagger} f^{R}_{+}(e^{i\theta})\right]^{-1} > 0. \quad (4.3)$$

Furthermore, if $\det\left[f^{L}_{+}(z)\right] \neq 0$ and $\det\left[f^{R}_{+}(z)\right] \neq 0$ $|z| \leqslant R$, then $\sigma(z)$, and $\sigma(z)^{-1} \in A_{v}$. Note that (4.3) gives a right and left factorization of σ.

Proof: Consider the following integral:

$$A = (1/2\pi)\int_{-\pi}^{\pi} \psi_{+}(e^{i\theta},n) f^{L}_{+}(e^{i\theta})^{-1} \phi^{L}(e^{i\theta},m)^{\dagger}\, d\theta, \quad n \geqslant m. \quad (4.4)$$

Since $\psi_{+}(e^{i\theta},n)$ and $f^{L}_{+}(e^{i\theta})^{-1}$ are analytic inside the unit circle and continuous on it, and $\phi^{L}(e^{i\theta},n)^{\dagger}$ is a polynomial in $1/z$, (4.4) may be evaluated using the residue theorem. From the recurrence formulas and (3.1) one finds

$$\psi_{+}(z,n) = z^{n} \prod_{i=n+1}^{\infty} (I - E_i E_i^{\dagger})^{-1/2} + O(z^{n+1}),$$

$$f^{L}_{+}(z) = \prod_{i=1}^{\infty} (I - E_i E_i^{\dagger})^{-1/2} + O(z),$$

and

$$\phi^{L}(1/z^{*},m)^{\dagger} = 1/z^{m} \prod_{i=1}^{m} (I - E_i E_i^{\dagger})^{-1/2} + O(1/z^{m-1}). \quad (4.5)$$

Therefore,

$$A = \delta_{n,m} I. \quad (4.6)$$

Alternatively, solving (3.41) for $\psi_{+}(z,n)$ and using the result in (4.3) yield

$$\begin{aligned} A = {} & (1/2\pi)\int_{-\pi}^{\pi} \phi^{L}(e^{i\theta},n)\left[f^{L}_{+}(e^{i\theta})^{\dagger}\right]^{-1} f^{L}_{+}(e^{i\theta})^{-1} \\ & \times \phi^{L}(e^{i\theta},m)^{\dagger}\, d\theta \\ & - (1/2\pi)\int_{-\pi}^{\pi} \phi_{-}(e^{i\theta},n) f^{R}_{+}(e^{i\theta})\left[f^{L}_{+}(e^{i\theta})^{\dagger}\right]^{-1} \\ & \times f^{L}_{+} e^{i\theta})^{-1} \phi^{L}(e^{i\theta},m)^{\dagger}\, d\theta. \end{aligned} \quad (4.7)$$

From (2.44)

$$f^{R}_{+}(e^{i\theta})\left[f^{L}_{+}(e^{i\theta})^{\dagger}\right]^{-1} = \left[f^{R}_{+}(e^{i\theta})^{\dagger}\right]^{-1} f^{L}_{+}(e^{i\theta}). \quad (4.8)$$

Therefore the second integral in (4.7) becomes

$$= (1/2\pi)\int_{-\pi}^{\pi} \phi_{-}(e^{i\theta},n)\left[f^{R}_{+}(e^{i\theta})^{\dagger}\right]^{-1} \phi^{L}(e^{i\theta},m)^{\dagger}\, d\theta. \quad (4.9)$$

Taking the Hermitian conjugate of (4.9), then evaluating it using the residue theorem and the recurrence formulas, one finds that it is equal to zero. Thus

$$\begin{aligned} & (1/2\pi)\int_{-\pi}^{\pi} \phi^{L}(e^{i\theta},n)\left[f^{L}_{+}(e^{i\theta}) f^{L}_{+}(e^{i\theta})^{\dagger}\right]^{-1} \phi^{L}(e^{i\theta},m)^{\dagger}\, d\theta \\ & = I\delta_{n,m}, \quad n \geqslant m. \end{aligned}$$

If $n \leqslant m$ in (4.1), the result follows by taking the Hermitian conjugate. The dual relation for right orthogonal polynomials can be obtained by using the procedures described above. Since $f^{L}_{+}(e^{i\theta})$ and $f^{R}_{+}(e^{i\theta})$ are nonsingular on the unit circle, Eq. (4.3) implies $\sigma(\theta) > 0$. If $\det\left[f^{L}_{+}(z)\right]$ and $\det\left[f^{R}_{+}(z)\right] \neq 0$, $|z| \leqslant R$, the Wiener–Levy theorem and Corollary 3 imply σ and σ^{-1} are elements of A_{v}.

Theorem 5 shows how tightly the decay of the Fourier coefficients of the weight is linked to the decay of the coefficients in the recurrence formula. The converse is also true.

Theorem 6: Let $R = 1$ in (3.2) and

$$(1/2\pi)\int_{-\pi}^{\pi} d\rho(\theta) = I, \quad (4.10)$$

with

$$d\rho(\theta) = \sigma(\theta)\, d\theta,$$

almost everywhere. Then (3.15) holds if and only if $\sigma(\theta) > 0$, and $\|\sigma\|_{v}$ is finite.

This is the matrix generalization of a theorem of Baxter.[6] Before proving Theorem 6 let us consider another useful matrix generalization of a result of Baxter.

Theorem 7: Let $\sigma(\theta) > 0$, and $\|\sigma\|_{v}$ be finite. Furthermore, let the matrix functions

$$g^{1,2}_{n}(e^{i\theta}) = \sum_{m=0}^{n} \hat{G}^{1,2}(m) e^{im\theta}, \quad (4.11)$$

and

$$f^{1,2}(e^{i\theta}) = \sum_{m=0}^{\infty} \hat{F}^{1,2}(m) e^{im\theta}, \quad f^{1,2} \in A_{v}^{+}, \quad (4.12)$$

be related by the equation

$$\begin{aligned} & (1/2\pi)\int_{-\pi}^{\pi} g^{1}_{n}(e^{i\theta})\sigma(\theta) e^{-ik\theta}\, d\theta \\ & = (1/2\pi)\int_{-\pi}^{\pi} f^{1}(e^{i\theta})\sigma(\theta) e^{-ik\theta}\, d\theta, \quad 0 \leqslant k \leqslant n, \end{aligned} \quad (4.13)$$

and

$$\begin{aligned} & (1/2\pi)\int_{-\pi}^{\pi} \sigma(\theta) g^{2}_{n}(e^{i\theta}) e^{-ik\theta} \\ & = (1/2\pi)\int_{-\pi}^{\pi} \sigma(\theta) f^{2}(e^{i\theta}) e^{-ik\theta}, \quad 0 \leqslant k \leqslant n. \end{aligned} \quad (4.14)$$

Then there exists an integer N and a constant M, both depending on $\sigma(\theta)$ such that for all $n \geqslant N$

$$\|g^{1,2}_{n} - f^{1,2}(n)\|_{v} \leqslant M \|f^{1,2} - f^{1,2}(n)\|_{v}. \quad (4.15)$$

Here

$$f^{1,2}(n) = \sum_{m=0}^{n} \hat{F}^{1,2}(m)e^{im\theta}. \tag{4.16}$$

Proof: Hirschman[9] has proved this result in the case $v(n) = 1$ for all n and a minor modification of his techniques gives the theorem.

Proof of Theorem 6: That the conditions are necessary is a consequence of Theorem 5. To prove sufficiency note that (3.16), (3.17), and (4.1) imply

$$(1/2\pi)\int_{-\pi}^{\pi} \sigma(\theta)\hat{\phi}^{L}(e^{i\theta},n)e^{-ik\theta}\,d\theta = \delta_{k,0}, \quad 0 \leqslant k \leqslant n. \tag{4.17}$$

Since $\sigma(\theta) > 0$ and $\sigma \in A_v$ one can factorize σ (Gohberg and Krein,[10] Gohberg[11]). Thus

$$\sigma = DD^{\dagger} = B^{\dagger}B, \tag{4.18}$$

where

$$B, B^{-1}, D, D^{-1} \in A_v^{+}, \quad \text{and} \quad B^{\dagger}, B^{\dagger -1}, D^{\dagger}, D^{\dagger -1} \in A_v^{-}. \tag{4.19}$$

Therefore,

$$(1/2\pi)\int_{-\pi}^{\pi} \sigma(\theta)B^{-1}B^{\dagger}(0)^{-1}e^{-ik\theta}\,d\theta = \delta_{k,0}, \tag{4.20}$$

where

$$B^{\dagger}(0) = (1/2\pi)\int_{-\pi}^{\pi} B^{\dagger}\,d\theta. \tag{4.21}$$

Using Theorem 7 one finds that $\hat{\phi}^{L}(e^{i\theta},n) \to B^{-1}B^{\dagger}(0)^{-1}$ in norm. Likewise, $\hat{\phi}^{R}(e^{i\theta},n) \to D^{\dagger}(0)D^{-1}$ in norm. Substituting (3.16) into (2.27), then multiplying on the right by $DD^{\dagger}(0)$ yields

$$\begin{aligned}\hat{\phi}^{L}(e^{i\theta},n-1)^{\dagger}DD^{\dagger}(0)^{-1} &= \hat{\phi}^{L}(e^{i\theta},n)^{\dagger}DD^{\dagger}(0)^{-1} \\ &\quad - e^{-in\theta}\beta^{\dagger}(n)^{-1}E_n\alpha(n)\hat{\phi}^{R}(e^{i\theta},n)DD^{\dagger}(0)^{-1}.\end{aligned} \tag{4.22}$$

Noting that the coefficients of $e^{-in\theta}$ on both sides of (4.22) are equal to zero, one can write

$$\begin{aligned}\|\hat{\phi}^{L}(e^{i\theta},n-1)^{\dagger}DD^{\dagger}(0)^{-1}\|_v &\leqslant \|\hat{\phi}^{L}(e^{i\theta},n)^{\dagger}DD^{\dagger}(0)^{-1}\|_v \\ &\quad + v(n)\gamma(n)|E_n|\,\|\hat{\phi}^{R}(e^{i\theta},n)DD^{\dagger}(0)^{-1}\|_v \\ &\quad - 2v(n)\gamma(n)|E_n|\,\|\hat{\phi}^{R}(0,n)D(0)D^{\dagger}(0)^{-1}\|_v.\end{aligned} \tag{4.23}$$

Therefore,

$$\begin{aligned}&\sum_{m=1}^{n} v(m)\gamma(m)|E_m|\,[2\|\hat{\phi}^{R}(0,n)D(0)D^{\dagger}(0)^{-1}\|_v \\ &\quad - \|\hat{\phi}^{R}(e^{i\theta},m)DD^{\dagger}(0)^{-1}\|_v] + \|DD^{\dagger}(0)^{-1}\|_v \\ &\quad \leqslant \|\hat{\phi}^{L}(e^{i\theta},n)^{\dagger}DD^{\dagger}(0)^{-1}\|_v.\end{aligned} \tag{4.24}$$

Since $\hat{\phi}^{L}(e^{i\theta},n)^{\dagger}$ converges in norm, the lhs of (4.24) is finite as $n \to \infty$. The term in the bracket on the rhs converges to one while the norm convergence of $\hat{\phi}^{R}(e^{i\theta},n)$ and $\hat{\phi}^{L}(e^{i\theta},n)$ imply $\gamma(n)$ is nonzero for all n. Thus the theorem is proved.

Using Theorem 5, the following expressions for $\psi^{L}(z,n)$ and $\psi^{R}(z,n)$ can now be obtained:

$$\psi^{L}(z,n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{e^{i\theta}+z}{e^{i\theta}-z}\right)[\phi^{L}(e^{i\theta},n) - \phi^{L}(z,n)]\sigma(\theta)\,d\theta, \tag{4.25}$$

and

$$\psi^{R}(z,n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{e^{i\theta}+z}{e^{i\theta}-z}\right)\sigma(\theta)[\phi^{R}(e^{i\theta},n) - \phi^{R}(z,n)]\,d\theta. \tag{4.26}$$

With these expressions and Corollaries 3 and 4, $f^{L}_{+\alpha}(z)$ and $f^{R}_{+\alpha}(z)$ can be represented as

$$f^{L}_{+\alpha}(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{e^{i\theta}+z}{e^{i\theta}-z}\right)\sigma(\theta)\,d\theta f^{L}_{+}(z), \quad |z| < 1 \tag{4.27}$$

and

$$f^{R}_{+\alpha}(z) = \frac{f^{R}_{+}(z)}{2\pi}\int_{-\pi}^{\pi}\left(\frac{e^{i\theta}+z}{e^{i\theta}-z}\right)\sigma(\theta)\,d\theta, \quad |z| < 1. \tag{4.28}$$

Integral representation for $\psi_{\pm}(z,n)$ and $\phi_{\pm}(z,n)$ can now be derived using (4.27) and (4.28) in (3.47) and (3.48), respectively. For example,

$$\begin{aligned}\psi_{+}(z,n) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{e^{i\theta}+z}{e^{i\theta}-z}\right)\phi^{L}(e^{i\theta},n) \\ &\quad \times\sigma(\theta)\,d\theta f^{L}_{+}(z), \quad |z| < 1.\end{aligned} \tag{4.29}$$

and

$$\begin{aligned}\psi_{-}(z,n) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{z+e^{i\theta}}{z-e^{i\theta}}\right)\phi^{R}(e^{i\theta},n)^{\dagger} \\ &\quad \times\sigma(\theta)\,d\theta f^{R}_{+}(1/z^{*})^{\dagger}, \quad |z| > 1.\end{aligned} \tag{4.30}$$

Since $\sigma \in A_v$, the series given by the above integrals is uniformly convergent for $|z| = 1$.

V. CONCLUSIONS

The techniques of scattering theory and Banach algebras have been used to study matrix orthogonal polynomials on the unit circle. Using these techniques, the close connection between the decay of the coefficients in the recurrence formula and the decay of the Fourier coefficients of the weight has been exhibited.

ACKNOWLEDGMENT

This work has been supported in part by NSF Grant MCS-8002731.

[1] J. S. Geronimo and K. M. Case, J. Math. Phys. **20**, 299 (1979).
[2] J. S. Geronimo and K. M. Case, Trans. Am. Math. Soc. **258**, 467 (1980).
[3] P. Delsarte, Y. V. Genin, and Y. G., Kamp, IEEE Trans. Circuits and Systems **25**, 149 (1978).
[4] P. Delsarte, Y. V. Genin, and Y. G. Kamp, SIAM J. Appl. Math. **36**, 34 (1979).
[5] P. Delsarte, Y. V. Genin, and Y. G. Kamp, IEEE Trans. Inform. Theory **26**, 465 (1979).
[6] G. Baxter, Trans. Amer. Math. Soc. **99**, 471 (1961).
[7] M. Rosenberg, Duke Math. J. **31**, 291 (1964).
[8] D. C. Youla and N. N. Kazanjian, IEEE Trans. Circuits and Systems **25**, 57 (1978).
[9] I. I. Hirschman, Jr., Duke Math. J. **34**, 403 (1967).
[10] I. C. Gohberg and M. G. Krein, Am. Math. Soc. Transl. Ser. 2 **14**, 217 (1960).
[11] I. C. Gohberg, Am. Math. Soc. Transl. Ser. 2 **49**, 130 (1966).

Universal formats for nonlinear ordinary differential systems

Edward H. Kerner
Sharp Physics Laboratory, University of Delaware, Newark, Delaware 19711

(Received 15 October 1980; accepted for publication 13 February 1981)

It is shown that very general nonlinear ordinary differential systems (embracing all that arise in practice) may, first, be brought down to polynomial systems (where the nonlinearities occur only as polynomials in the dependent variables) by introducing suitable new variables into the original system; second, that polynomial systems are reducible to "Riccati systems," where the nonlinearities are quadratic at most; third, that Riccati systems may be brought to elemental universal formats containing purely quadratic terms with simple arrays of coefficients that are all zero or unity. The elemental systems have representations as novel types of matrix Riccati equations. Different starting systems and their associated Riccati systems differ from one another, at the final elemental level, in order and in initial data, but not in format.

PACS numbers: 02.30.Hg, 03.20. + i

I. INTRODUCTION

The object of the present paper is to show how ordinary differential systems of arbitrary order

$$\dot{\xi}_i = X_i(\xi_1,\xi_2,...,\xi_k,t), \quad i = 1,2,...,k, \tag{1}$$

having a great diversity of nonlinearity of the X_i, and covering all cases arising in practice, may be reduced down to *Riccati systems*

$$\dot{x}_i = A_i + B_{i\alpha}x_\alpha + C_{i\alpha\beta}x_\alpha x_\beta, \tag{2}$$
$$i = 1,2,...,n, \quad n > k, \quad A,B,C \text{ constant}$$

(summation on repeated Greek indices understood), having only *quadratic* nonlinearity. Finally, reductions of Eq. (2) to *elemental Riccati systems* (ERS)

$$\dot{z}_i = E_{i\alpha\beta}z_\alpha z_\beta, \tag{3}$$
$$i = 1,2,...,p \quad p(n) > n,$$

are achieved, wherein each E_{ijk} is either 0 or 1 in a definite and simple pattern. The differing Riccati systems (2) for differing A,B,C are all subsumed in the same ERS of Eq. (3) upon assigning in the latter suitably differing initial data.

The cost of reduction to the universal ERS formats is an appreciable increase in order, $p > n > k$. This is owing to the repeated and systematic use of an ancient and very simple device, that of introducing appropriate collectives of variables and parameters as new variables.

Notwithstanding this cost, the point remains, first, that it is the quadratic nonlinearity alone which may be regarded as the ultimate one; and, second, that very general systems, Eq. (1), may not only be "quadraticized," Eq. (2), but elementally quadraticized, Eq. (3). This does not, of course, in itself solve nonlinear problems as yet. It does, however, so rephrase them into sufficiently simple and universal terms that something approaching a general theory of differential nonlinearity may be vizualized, whereas, otherwise, there is only a potpourri of special cases, disconnected, incoherent, and intractable.

As so many natural laws and models of natural phenomena have their statements directly in the form of Eq. (1), it is scarcely necessary to belabor how ubiquitous and refractory they are. Let us recall only a few examples of classical and of current interest, such as: the many-body problem of Newtonian or Hamiltonian dynamics[1]; the Fermi–Pasta–Ulam problem[2] of nonlinear yet non-energy-equipartitioning conservative oscillators; the nonconservative oscillators like that of Van der Pol, exhibiting limit-cycles, or like that of Lorenz showing "chaotic" behavior[3]; the Painlevé transcendents that stem out of similarity solutions to solitonic wave equations[4]; the Lane–Emden equations[5] for stellar polytropes, and as well the Fermi–Thomas equation for atomic potentials; the Navier–Stokes equation of fluid dynamics[6] (when the velocity field is Fourier-analyzed spatially and the Fourier series is at first truncated); the nonlinear rate equations of chemical kinetics,[7] which comprise one of the significant statements of the nonequilibrium behavior of matter (including oscillatory behavior of Zhabotinsky and other types), and which rival or exceed in complexity the Navier–Stokes equation in important situations of high biochemical diversity; and the similarly structured rate-laws in models of ecological dynamics, such as those due to Volterra.[8] Several of these examples are at the outset in the form of Riccati systems, Eq. (2).

II. REDUCTION TO RICCATI FORMAT

We may first notice in regard to general differential systems of type (1) that the near-trivial introduction of an additional variable formally erases the distinction between autonomous and nonautonomous problems, according to

$$\dot{\xi}_i = X_i(\xi_1,\xi_2,...,\xi_k,\xi_{k+1}), \quad \dot{\xi}_{k+1} = 1.$$

In this view the linear system

$$\dot{\xi}_i = F_{i\alpha}(t)\xi_\alpha + F_i(t) \tag{4}$$

falls into

$$\dot{\xi}_i = F_{i\alpha}(\xi_{k+1})\xi_\alpha + F_i(\xi_{k+1}), \quad \dot{\xi}_{k+1} = 1, \tag{5}$$

which may be considered in present context to be nonlinear. This reclassification of Eq. (4) reflects and respects the fact that (4) suffers fundamentally from the same difficulty as general nonlinear $\dot{\xi}_i = X_i$, namely, that, in order to compute $\ddot{\xi}_i, \dddot{\xi}_i, \cdots$ (by way of formal construction of power-series solutions), one does not generally have a simple recursive rule for the computation; the binary products $F_{i\alpha}\xi_\alpha$ in fact have the same character as the products $C_{ijk}x_jx_k$ in the Riccati sys-

 0022-2488/81/071366-06$1.00

tem (2), and (5) may indeed be made over into a Riccati system on this basis, using methods described below.

A few examples will now be considered to show how, by other simple adjoinders of new variables (besides ξ_{k+1}), quite diverse differential systems may be cast first of all into *polynomial ordinary differential systems* (PODS)

$$\dot{\eta}_i = P_i(\eta_1,\eta_2,...,\eta_l), \quad i = 1,2,...,l, \quad l > k$$

(P_i a polynomial in its arguments).

Following that, the reduction of PODS to Riccati systems will be described.

Consider three particles in nonrelativistic motion under Yukawa forces with potentials

$$V_{12} = g_{12}\,[\exp(-\lambda_1 r_{12})]/r_{12}, \quad etc,$$

and equations of motion

$$m_1\ddot{\mathbf{r}}_1 = g_{12}(\mathbf{r}_1 - \mathbf{r}_2)(1/r_{12}^3 + \lambda_1/r_{12}^2)\exp(-\lambda_1 r_{12}) + g_{31}(\mathbf{r}_1 - \mathbf{r}_3)(1/r_{31}^3 + \lambda_3/r_{31}^2)\exp(-\lambda_3 r_{31}), \quad etc, \tag{6}$$

where *etc* stands for the remaining potentials and equations of motion obtained by cyclic permutation of 1,2,3. This, of course, stands for a little prototype of the many-body problem, whose generality may also encompass more particles, other sorts of central as well as noncentral forces, many-body forces, dissipative forces, etc., without harm to the thrust of the following type of construction.

Write

$$u_1 = 1/r_{12}, \quad u_2 = 1/r_{23}, \quad u_3 = 1/r_{31},$$
$$r_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2},$$
$$w_1 = \exp(-\lambda_1 r_{12}), \quad w_2 = \exp(-\lambda_2 r_{23}),$$
$$w_3 = \exp(-\lambda_3 r_{31}),$$

and then Eqs. (6) go over into the polynomial system

$$\begin{aligned}
&\dot{\mathbf{r}}_1 = \mathbf{v}_1,\\
&m_1\dot{\mathbf{v}}_1 = g_{12}w_1(u_1^3 + \lambda_1 u_1^2)(\mathbf{r}_1 - \mathbf{r}_2) + g_{31}w_3(u_3^3 + \lambda_3 u_3^2)(\mathbf{r}_1 - \mathbf{r}_3),\\
&\dot{u}_1 = -u_1^3(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2),\\
&\dot{w}_1 = -\lambda_1 w_1 u_1(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2),\\
&etc\,.
\end{aligned} \tag{7}$$

We have also

$$\frac{d}{dt}\left(u_1 - \frac{1}{r_{12}}\right) = -(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2)\left(u_1^3 - \frac{1}{r_{12}^3}\right)$$
$$= -(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2) \times\left(u_1^2 + \frac{u_1}{r_{12}} + \frac{1}{r_{12}^2}\right)\left(u_1 - \frac{1}{r_{12}}\right),$$

$$\frac{d}{dt}\left[w_1 - \exp - \lambda_1 r_{12})\right]$$
$$= -\lambda_1(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2)\left(w_1 u_1 - \frac{\exp(-\lambda_1 r_{12})}{r_{12}}\right)$$
$$= -\lambda_1(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2) \times\left[w_1\left(u_1 - \frac{1}{r_{12}}\right) + \frac{1}{r_{12}}\left[w_1 - \exp(-\lambda_1 r_{12})\right]\right],$$

etc,

or

$$\frac{d}{dt}\begin{pmatrix} u_1 - \dfrac{1}{r_{12}} \\ w_1 - \exp(-\lambda_1 r_{12}) \end{pmatrix} = -(\mathbf{v}_1 - \mathbf{v}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2)\begin{pmatrix} u_1^2 + \dfrac{u_1}{r_{12}} + \dfrac{1}{r_{12}^2} & 0 \\ \lambda_1 w_1 & \dfrac{\lambda_1}{r_{12}} \end{pmatrix} \times\begin{pmatrix} u_1 - \dfrac{1}{r_{12}} \\ w_1 - \exp(-\lambda_1 r_{12}) \end{pmatrix} \quad etc,$$

telling that the initial data

$$\left.\begin{aligned} &u_1 - 1/r_{12} = 0\\ &w_1 - \exp(-\lambda_1 r_{12}) = 0 \end{aligned}\right\} \quad etc \quad \text{at } t = t_0, \tag{8}$$

persist for all times, and that, upon these data, Eqs. (7) encompass Eqs. (6) with perfect fidelity.

Thus is achieved, through systematic use of the old device of bringing in suitable new variables, a recasting of the original problem (6) in $\mathbf{r}_i,\mathbf{v}_i$ over into its faithful rerepresentation as the PODS (7) in $\mathbf{r}_i,\mathbf{v}_i,u_i,w_i$. with the proviso of the initial data (8). Absent this proviso, the system (7), being of higher order than (6), overreaches itself in stating, as it stands alone, quite a lot more than (6). The proviso then remedies this overstatement, trimming (7) to exact faithfulness to (6).

Consider next the following somewhat contrived illustrative problem. A plane pendulum, with an arm of negligible mass compared to its bob, has a temperature oscillation impressed upon it, producing a prescribed alteration of its length say,

$$r = \alpha + \beta\,\mathrm{sn}(\gamma t, k),$$

with sn being the Jacobian elliptic function of modulus k. The equation of motion for this arm-variable pendulum is

$$\frac{d}{dt}(r^2\dot{\theta}) = -gr\sin\theta$$

or, using the above device of making the system formally autonomous,

$$\begin{aligned}
&\dot{\theta} = \omega,\\
&\dot{\omega} = -\frac{2\beta\gamma\omega\,\mathrm{cn}\gamma\tau\,\mathrm{dn}\gamma\tau}{\alpha + \beta\,\mathrm{sn}\gamma\tau} - \frac{g\sin\theta}{\alpha + \beta\,\mathrm{sn}\gamma\tau},\\
&\dot{\tau} = 1.
\end{aligned} \tag{9}$$

Introduce new variables

$$\xi = \mathrm{sn}\gamma\tau, \quad \eta = \mathrm{cn}\gamma\tau, \quad \zeta = \mathrm{dn}\gamma\tau,$$
$$u = \sin\theta, \quad v = \cos\theta, \quad w = (\alpha + \beta\xi)^{-1},$$

to bring the PODS

$$\begin{aligned}
&\dot{\theta} = \omega, \quad \dot{\omega} = -2\beta\gamma\omega\eta\zeta w - guw,\\
&\dot{\tau} = 1, \quad \dot{\xi} = \gamma\eta\zeta,\\
&\dot{\eta} = -\gamma\xi\zeta, \quad \dot{\zeta} = -\gamma k^2\xi\eta,\\
&\dot{u} = v\omega, \quad \dot{v} = -u\omega, \quad \dot{w} = -\beta\gamma\omega^2\eta\zeta.
\end{aligned} \tag{10}$$

The overstatement here, as earlier, is remedied by remarking that (abbreviating the sn,cn,dn as s,c,d)

$$\frac{d}{dt}\begin{pmatrix}\xi-\mathrm{s}\\ \eta-\mathrm{c}\\ \zeta-\mathrm{d}\end{pmatrix} = -\frac{1}{2}\gamma\begin{pmatrix}0 & -(\zeta+\mathrm{d}) & -(\eta+\mathrm{c})\\ \zeta+\mathrm{d} & 0 & \xi+\mathrm{s}\\ k^2(\eta+\mathrm{c}) & k^2(\xi+\mathrm{s}) & 0\end{pmatrix}\begin{pmatrix}\xi-\mathrm{s}\\ \eta-\mathrm{c}\\ \zeta-\mathrm{d}\end{pmatrix},$$

$$\frac{d}{dt}\begin{pmatrix}u-\sin\theta\\ v-\cos\theta\end{pmatrix} = \begin{pmatrix}0 & \omega\\ -\omega & 0\end{pmatrix}\begin{pmatrix}u-\sin\theta\\ v-\cos\theta\end{pmatrix},$$

$$\frac{d}{dt}\left(w-\frac{1}{\alpha+\beta\xi}\right) = -\beta\gamma\eta\zeta\left(w+\frac{1}{\alpha+\beta\xi}\right)\left(w-\frac{1}{\alpha+\beta\xi}\right),$$

so that if we impose the initial data

$$\left.\begin{aligned}&\xi-\mathrm{s}=\eta-\mathrm{c}=\zeta-\mathrm{d}=0\\ &u-\sin\theta=v-\cos\theta=w-(\alpha+\beta\xi)^{-1}=0\end{aligned}\right\}\quad \text{at } t=t_0,$$

then these relations are maintained for all times. And so, upon these data and none other, the enlarged system (10) comprises the starting system (9) faithfully. The starting system has been "polynomialized."

The examples have been chosen to be a little involved in order to illuminate and buttress the contention being made that even quite complex starting differential systems will yield to reduction to PODS. In particular the lesson is clear that the presence of transcendental functions in the initial differential system is no bar to polynomialization. The process illustrated above of effectively "differentiating-away" the transcendents, as with $\mathrm{sn}\gamma\tau$, $\sin\theta$,$\exp(-\lambda r)$, is evidently very far-reaching. One can see quite easily that, so long as the initial system involves transcendental functions which are themselves controlled by differential equations of finite order of a wide variety (including the circular, hyperbolic, elliptic, hypergeometric, Bessel, etc., functions of classical analysis—but excluding the like of the gamma function which obeys no differential equation of finite order), then the scheme of introducing new variables and differentiating-away the transcendents becomes practicable. While further characterization of the polynomializable differential systems is not simple, their scope certainly spans rational and irrational functions, and certainly an appreciable class of transcendents (including also rational or irrational functions of transcendental functions and vice-versa, as well as many kinds of transcendental functions of transcendental functions). The reader may find it interesting to try to build differential systems from the familiar functions of classical analysis (excluding the gamma or like functions failing to satisfy finite-order differential equations) which are *not* polynomializable.

When one thinks at the outset of computing from $\dot{\xi}_i = X_i$ the succession $\ddot{\xi}_i, \dddot{\xi}_i, \cdots$ toward thereby building a formal solution $\xi_i(t)$ as, say, a Taylor series, one recognizes that the differentiating-away process that produces $\dot{\eta}_i = P_i$ is simply of the nature of a systematic means for handling the peculiarities of the assorted special functions entering X_i through, in effect, extruding the peculiarities into that much simpler condition represented in their assorted individual defining differential equations, that is, through exploiting the simple recursive relationships between the functions and their derivatives. Both the "gross" structure of X_i considered as a function of other functions, as well as the "fine" structure of these other functions, will in very many cases be capable of being differentiated-away.

We will now take it for granted that the redescription of an initial differential system by means of an equivalent autonomous PODS has been performed, and examine whether and how PODS may be reduced down to quadratic or Riccati format.

Let us begin with a simple example,

$$\dot{\eta} = a_0 + a_1\eta + a_2\eta^2 + a_3\eta^3, \tag{11}$$

going a step beyond the well-known classical Riccati equation

$$\dot{\eta} = a_0 + a_1\eta + a_2\eta^2,$$

where the a_i are prescribed constants. Using again the device of bringing in additional variables, let $\zeta = \eta^2$, giving the joint system

$$\begin{aligned}&\dot{\eta} = a_0 + a_1\eta + a_2\zeta + a_3\zeta\eta,\\ &\dot{\zeta} = 2a_0\eta + 2a_1\zeta + 2a_2\eta\zeta + 2a_3\zeta^2.\end{aligned} \tag{12}$$

This is a Riccati system in that only terms up to quadratic degree enter the right-hand members. As previously, we have an overstatement in this second-order system, again remediable by computing

$$\frac{d}{dt}(\zeta-\eta^2) = (2a_1 + 2a_3\zeta)(\zeta-\eta^2),$$

and insisting on initial data

$$\zeta-\eta^2 = 0 \quad \text{at} \quad t=t_0$$

that perpetuates $\zeta = \eta^2$ for all time, ensuring that Eq. (12) holds perfectly to Eq. (11). The decrease in polynomial degree has been traded for an increase in order, with appropriately restricted initial data.

In this same vein, for the general PODS we have

$$\begin{aligned}\dot{\eta}_i &= P_i(\eta|r) = \text{polynomial in } \eta \text{ of degree } r \text{ at the highest}\\ &= \Phi^0_i + \Phi^1_{i\alpha}\eta_\alpha + \Phi^2_{i\alpha\beta}\eta_a\eta_\beta + \Phi^3_{i\alpha\beta\gamma}\eta_\alpha\eta_\beta\eta_\gamma + \Phi^4_{i\alpha\beta\gamma\delta}\eta_\alpha\eta_\beta\eta_\gamma\eta_\delta + \cdots + \underbrace{\Phi^r_{i\alpha\beta\cdots\rho\sigma}\eta_\alpha\eta_\beta\cdots\eta_\rho\eta_\sigma}_{r\text{-fold monomial}},\end{aligned} \tag{13}$$

$$i = 1,2,\ldots,l,$$

and may introduce $x_{ij} = \eta_i\eta_j$ to give

$$\dot{\eta}_i = \Phi^0_i + \Phi^1_{i\alpha}\eta_\alpha + \Phi^2_{i\alpha\beta}x_{\alpha\beta} + \Phi^3_{i\alpha\beta\gamma}x_{\alpha\beta}\eta_\gamma + \Phi^4_{i\alpha\beta\gamma\delta}x_{\alpha\beta}\eta_\gamma\eta_\delta + \cdots + \Phi^r_{i\alpha\beta\cdots\rho\sigma}x_{\alpha\beta}\eta_\gamma\eta_\delta\cdots\eta_\rho\eta_\sigma, \tag{14}$$

and additionally

$$\begin{aligned}\dot{x}_{ij} &= \Phi^0_i\eta_j + \Phi^1_{i\alpha}x_{\alpha j} + \Phi^2_{i\alpha\beta}x_{\alpha\beta}\eta_j + \Phi^3_{i\alpha b\gamma}x_{\alpha\beta}x_{\gamma j}\\ &\quad + \Phi^4_{i\alpha\beta\gamma\delta}x_{\alpha\beta}\eta_\gamma x_{\delta j} + \cdots + \Phi^r_{i\alpha\beta\gamma\cdots\rho\sigma}x_{\alpha b}\eta_\gamma\eta_\delta\cdots\eta_\rho x_{\sigma j} + [*],\end{aligned} \tag{15}$$

where [*] means the same terms as first written but with i and j interchanged.

The introduction of the x's into $\dot{\eta}_i, \dot{x}_{ij}$ is at the outset ambiguous, and the ambiguity has been resolved in the particular way shown for the sake of a certain convenience. Namely, when we calculate

$$\frac{d}{dt}(x_{ij} - \eta_i \eta_j)$$
$$= \phi^{1}_{i\alpha}(x_{\alpha j} - \eta_\alpha \eta_j) + \phi^{2}_{i\alpha\beta\gamma} x_{\alpha\beta}(x_{\gamma j} - \eta_\gamma \eta_j)$$
$$+ \cdots + \phi^{r}_{i\alpha\beta\gamma\cdots\rho\sigma} x_{\alpha\beta}\eta_\gamma \cdots \eta_\rho (x_{\sigma j} - \eta_\sigma \eta_j)$$
$$+ [*],$$

we have that, if initial data

$$x_{ij} - \eta_i \eta_j = 0 \quad \text{at } t = t_0$$

are prescribed, then the connection $x_{ij} = \eta_i \eta_j$ is maintained for all time. And then the joint system Eqs. (14) and (15) in all the $\lambda = l + \frac{1}{2}l(l+1)$ variables

$$\eta_1, \eta_2, \ldots, \eta_l;\ x_{11}, x_{12}, \ldots, x_{1l};\ x_{22}, x_{23}, \ldots, x_{2l};\ \ldots;\ x_{ll},$$

or, say, $u_1, u_2, \ldots, u_\lambda$ is of structure

$$\dot{u}_i = P_i(u|r-1), \tag{16}$$

while being exactly equivalent to the starting system, Eqs. (13). So quite generally a decrease of polynomial degree is achieved, albeit at the expence of an increase in order (together with restricted initial data).

By repeating this type of process $r-2$ times in all, with due increase in order along the way and with due restrictions on initial data, we have the result that *very general starting systems may, following their preliminary reduction to polynomial systems, be brought down to a Riccati system*

$$\dot{x}_i = A_i + B_{i\alpha} x_\alpha + C_{i\alpha\beta} x_\alpha x_\beta, \tag{17}$$

which is, under suitably restricted initial data, exactly equivalent to the starting system.

As will be at once evident, the Riccati system cannot, on the above basis, be reduced any further. The quadratic nonlinearity stands as the final intrinsic one. If, however, one persists in grouping $x_j x_k$ as a single new variable, linear equations plus new Riccati equations follow, and continuing in this way produces a linear system of infinite order, as will elsewhere be discussed.

Certain kinds of Riccati systems have been discussed[9] from the viewpoint of exact linearization procedures, akin more or less to that for the classical elementary Riccati equation, but none so far has proven to be comprehensive.

III. CONCLUSION: ELEMENTAL RICCATI SYSTEMS

The reduction scheme that has been advanced, running from general differential systems to PODS to Riccati systems, admits further simple reduction to elemental or "canonical" forms of Riccati systems.

It may be noticed first that Riccati systems are formally reducible, in many ways, to systems with *only* quadratic terms. The simplest way is to adjoin one more variable to Eq. (17),

$$\dot{x}_i = A_i x_{n+1}^2 + B_{i\alpha} x_\alpha x_{n+1} + C_{i\alpha\beta} x_\alpha x_\beta,$$
$$\dot{x}_{n+1} = 0$$

with the initial datum $x_{n+1} = 1$. That is, linear terms may always be quadraticized, though not conversely. It is sufficient then to deal always with

$$\dot{x}_i = D_{i\alpha\beta} x_\alpha x_\beta \tag{18}$$

of arbitrary order, or, in a condensed notation,

$$\dot{\mathbf{x}} = \mathbf{x} \circ \mathbf{x}, \quad (\mathbf{a} \circ \mathbf{b})_i \equiv D_{i\alpha\beta} a_\alpha b_\beta,$$

where D_{ijk} can always be counted as having a prescribed symmetric part in j,k, plus an arbitrary skew-symmetric part in j,k. The Riccati problem (18) is then characterized by commutativity of the binary operation $\circ$ (when D is taken purely symmetric) but by nonassociativity $\mathbf{a}(\mathbf{b}\circ\mathbf{c}) \neq (\mathbf{a}\circ\mathbf{b})\circ\mathbf{c}$ in general.

Rewrite Eq. (18) as

$$\dot{x}_i = (D_{i\alpha\beta} x_\beta) x_\alpha$$
$$\equiv X_{i\alpha} x_\alpha, \tag{19}$$

and consider the outer product

$$U_{ijkl} \equiv D_{ijk} x_l$$

with

$$X_{ij} = U_{ij\lambda\lambda}.$$

Then

$$\dot{U}_{ijkl} = D_{ijk} X_{l\sigma} x_\sigma$$
$$= D_{ijk} x_\sigma X_{l\sigma}$$

or

$$\dot{U}_{ijkl} = U_{ijk\sigma} U_{l\sigma\lambda\lambda}, \tag{20}$$

with initial data $U_{ijkl}(0) = D_{ijk} x_l(0)$. The quadruple indices to U here may be understood to run, for example, in dictionary order 1111,1112,···, which may in turn be designated anew as 1,2,···, so that Eq. (20) is itself a Riccati system

$$\dot{U}_a = E_{a\xi\eta} U_\xi U_\eta, \tag{21}$$

with all E_{abc} zero or unity in accordance with Eq. (20).

The integration of the Riccati system (19) is thus made to depend on that of the *elemental Riccati system*, Eq. (20) or (21).

The ERS of Eq. (20) may also be seen as a type of matrix Riccati system as follows. Write Eq. (20) as

$$U_{ijkl} = U_{ijk\sigma} \delta_{\sigma\nu} \delta_{\lambda\mu} U_{l\nu\lambda\mu},$$

and regroup as

$$U^{(jk)}_{il} = U^{(jk)}_{i\sigma} \Delta^{(\nu\lambda)}_{\sigma\mu} U^{(\nu\lambda)}_{l\mu}, \tag{22}$$
$$(\Delta^{(\nu\lambda)}_{\sigma\mu} \equiv \delta_{\sigma\nu}\delta_{\lambda\mu}).$$

Now designating the collection of square matrices $U^{(11)}, U^{(12)}, \cdots$ and $\Delta^{(11)}, \Delta^{(12)}, \cdots$ as $U^1, U^2, \cdots$ and $\Delta^1, \Delta^2, \cdots$ (going, say, in dictionary order for the double superscripts jk), Eq. (22) is

$$\dot{U}^i = U^i \Delta^\gamma \tilde{U}^\gamma.$$

Also, under linear transformation

$$U_{ijkl} = \Gamma_{jk\alpha\beta} V_{\alpha\beta il},$$

Eq. (20) goes over to

$$\dot{V}_{abil} = V_{abi\sigma} \Gamma_{\sigma\lambda\rho\tau} V_{\rho\tau l\lambda},$$

or to

$$\dot{V}^{(ab)}_{il} = V^{(ab)}_{i\sigma} \Gamma^{(\rho\tau)}_{\sigma\lambda} V^{(\rho\tau)}_{l\lambda},$$

viz.,

$$\dot{V}^i = V^i \Gamma^\gamma \tilde{V}^\gamma,$$

while, under the quite different transformation

$$U_{ijkl} = \Lambda_{ki\alpha\beta} W_{\alpha\beta jl},$$

Eq. (20) reads

$$\dot{W}_{abjl} = W_{abj\sigma} \Lambda_{\lambda l \mu\nu} W_{\mu\nu\sigma\lambda},$$

$$\dot{W}^{(ab)}_{jl} = W^{(ab)}_{j\sigma} W^{(\mu\nu)}_{\sigma\lambda} \Lambda^{(\mu\nu)}_{\lambda l},$$

$$\dot{W}^i = W^i W^\gamma \Lambda^\gamma.$$

Since the "direction" of $\mathbf{W} \equiv (W^1, W^2, \cdots)$ remains fixed (all derivatives of W^i being proportional to W^i), we may take

$$W^i(t) = W^i(0) Z(t) \equiv Q^i Z,$$

and deal with the scalar Z according to

$$\dot{Z} = Z Q^\gamma Z \Lambda^\gamma,$$

$$Z(0) = I.$$

Similar scalar equations follow by similar means for U^i and V^i above.

The simple case here of Z commuting with all Λ^i brings

$$\dot{Z} = Z Q^\gamma \Lambda^\gamma Z$$

with solution

$$Z = (I - Q^\gamma \Lambda^\gamma t)^{-1}.$$

But, as examination shows, this is of little use in view of the severe restrictions that come to be placed upon D_{ijk}.

The theory of matrix Riccati equations[9] appears so far to be limited to single matrix equations of a particularly simple type, and is of no evident bearing in connection with the present matrix Riccati systems in U^i, V^i, W^i or the single matrix Riccati equation in Z.

Finally, another sort of reduction of general Riccati systems (17) to elemental form may be considered in an additional differentiating-away procedure. In (17) place $\dot{x}_i = y_i$, so that

$$\dot{y}_i = (B_{i\alpha} + 2C_{i\alpha\beta} x_\beta) y_\alpha$$

$$(C_{ijk} = C_{ikj}), \quad i = 1,2,\ldots,n.$$

Calling

$$Y_{ij} \equiv B_{ij} + 2C_{ij\beta} x_\beta,$$

we have

$$\dot{Y}_{ij} = 2C_{ij\beta} y_\beta,$$

or

$$\dot{Y}_{ij} = Z_{ij\beta} y_\beta, \quad \dot{Z}_{ijk} = 0.$$

In all, the Riccati system

$$\dot{y}_i = Y_{i\alpha} y_\alpha, \quad \dot{Y}_{ij} = Z_{ij\beta} y_\beta, \quad \dot{Z}_{ijk} = 0, \tag{23}$$

follows from Eq. (17), with initial data

$$y_i(0) = A_i + B_{i\alpha} x_\alpha(0) + C_{i\alpha\beta} x_\alpha(0) x_\beta(0),$$

$$Y_{ij}(0) = B_{ij} + 2C_{ij\beta} x_\beta(0),$$

$$Z_{ijk}(0) = 2C_{ijk},$$

and with x_i computed from y_i by quadrature.

A convenient redesignation of y, Y, Z is

$$y_1, y_2, \ldots, y_n \equiv z_1, z_2, \ldots, z_n,$$

$$Y_{11}, Y_{12}, \ldots, Y_{1n}; \ldots; \ldots Y_{nn} \text{ in dictionary order}$$

$$\equiv z_{n+1}, z_{n+2}, \ldots, z_{2n}; \ldots; \ldots z_{n+n^2},$$

$$Z_{111}, Z_{112}, \ldots, Z_{11n}; \ldots; \ldots Z_{nnn} \text{ in dictionary order}$$

$$\equiv z_{n+n^2+1}, \ldots, z_{n+n^2+n}; \ldots; \ldots z_{n+n^2+n^3}.$$

Then the system (23) is

$$\dot{z}_i = E^*_{i\alpha\beta} z_\beta z_\alpha, \quad i = 1,2,\ldots,n+n^2+n^3,$$

with coefficients

$$E^*_{ijk} = \delta_{1j}\delta_{ni+1,k} + \delta_{2j}\delta_{ni+2,k} + \cdots + \delta_{nj}\delta_{ni+n,k}.$$

For example, for a starting Riccati system of second order ($n = 2$)

$$E^*_{ijk} = \delta_{1j}\delta_{2i+1,k} + \delta_{2j}\delta_{2i+2,k}$$

$$i,j,k = 1,2,\ldots,14,$$

and the square matrices $[E^*_i]_{jk}$ have unity in the $(1,2i+1)$ and $(2,2i+2)$ positions, and zeroes elsewhere.

This completes the statement of the second form of reduction to an *elemental Riccati system*.

In summary, it has been shown, through a chain of elementary steps, that the reduction scheme

very general nonlinear systems→

polynomial systems→

Riccati systems→

elemental Riccati systems,

performs a universalization of nonlinear differential systems.

The reduction to elementary universal formats in Eqs. (20) and (24) is, of course, bought at the price of appreciable increase in order of the ERS. So long as one wishes to survey the general situation of initial Riccati systems (17) bearing a great burden of general parameters A_i, B_{ij}, C_{ijk}, the shifting of this burden into the initial data of structurally simple ERS clearly must cost this price. It is to say simply that great parametrization may be viewed in terms of great order.

ACKNOWLEDGMENTS

My thanks go to the U. S. Department of Energy and to the University of Delaware Center for Advanced Study for partial support of this work, and to Professor Robert N. Hill for valuable criticism of an earlier version.

[1] E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., Cambridge, 1937).

[2] E. Fermi, J. Pasta, and S. Ulam, in *Collected Works of Enrico Fermi*, Vol. 2 (Chicago U. P., Chicago, 1965).

[3] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1972); J. B. McLaughlin and P. C. Martin, Phys. Rev. A **12**, 186 (1975); O. E. Rossler, in *Synergetics, A Workshop*, edited by H. Haken (Springer, New York, 1977); P. K. C. Wang, J. Math. Phys. **21**, 398 (1980).

[4] M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103, (1977).

[5] S. Chandrasekhar, *Introductiont to the Study of Stellar Structure* (Dover, New York, 1957).

[6] D. C. Leslie, *Developments in the Theory of Turbulence* (Clarendon, Oxford, 1973); J. B. McLaughlin and P. C. Martin, Phys. Rev. A **12**, 186 (1975).

[7]G. R. Gavalas, *Nonlinear Differential Equations of Chemically Reacting Systems* (Springer, New York, 1968); E. H. Kerner, Bull. Math. Biophy. **34**, 243 (1972); B. Chance, E. K. Pye, A. K. Ghosh, and B. Hess, Eds., *Biological and Biochemical Oscillators* (Academic, New York, 1973); H. Haken, Rev. Mod. Phys. **47**, 67 (1975).

[8]V. Volterra, *Lecons sur la théorie mathematique de la lutte pour la vie* (Gauthier-Villars, Paris, 1931); E. H. Kerner, *Gibbs Ensemble: Biological Ensemble* (Gordon and Breach, New York, 1972).

[9]W. T. Reid, *Riccati Differential Systems* (Academic, New York, 1972), treats linearizable matrix Riccati equations which are special cases of the more general Riccati systems of Eq. (17); E. H. Kerner, Bull. Math. Biophy. **34**, 243 (1972); I. Pereyra, thesis (University of Delaware, 1980); see also L. Markus, Contr. Theory Non-Linear Oscillations **5**, 185 (1960), and R. E. Beck and B. Kolman, J. Diff. Eq. **24**, 466 (1977), for certain algebraic aspects of Riccati systems, and of higher-degree matrix differential equations.

Treatment of some singular potentials by change of variables in Wiener integrals

Ph. Blanchard and M. Sirugue [a)]
Fakultät für Physik, Universität Bielefeld, D-4800 Bielefeld, West Germany

(Received 5 January 1981; accepted for publication 13 February 1981)

Using a regularization of the classical Kepler problem, we show that a change of process and a change of time in stochastic differential equations allow a treatment of singular potentials in Wiener integrals.

PACS numbers: 02.30.Jr, 02.50.Ey, 03.65.Db

1. INTRODUCTION

It has been known since the thirties that in some sense the Coulomb problem is equivalent to the harmonic oscillator problem. See for example Ref. 1. More recently it has been used by J. Schwinger in his lectures to solve the problem of the hydrogen atom. Heuristically it suffices to observe the close relationship between the Laguerre and Hermite polynomials.[2] More generally it has been demonstrated by D. Fivel[3] that a solution ψ of the radial Schrödinger equation for a spherically symmetric potential of the type

$$V(r) = (g/r) + (G/r)u(r)$$

and of energy E, can be transformed into a solution φ of the radial Schrödinger equation

$$-\frac{\hbar^2}{2m}\varphi'' + \left[\frac{1}{2}Ky^2 + Gu(y^2) + \frac{\hbar^2}{2m}\frac{\lambda(\lambda+1)}{y^2}\right]\varphi = \epsilon\varphi,$$

where one has made the following substitutions

$$r = y^2, \quad \psi(r) = y^{1/2}\varphi(y),$$

and the identifications

$$-8E = K, \quad \epsilon = -4g, \quad \lambda = 2l + \tfrac{1}{2}.$$

See also, F. Goded.[4]

Moreover, it should be noted that these transformations have made it possible to find the discrete spectrum of the hydrogen atom from the one of the harmonic oscillator. The three-dimensional Coulomb problem was related to the two-dimensional harmonic oscillator, for instance, in Ref. 5 and to the three-dimensional oscillator in Ref. 6. The connection between the nonrelativistic quantum mechanical eigenvalue problem and a Lie algebra eigenvalue problem has been extensively developed. See, e.g., A. O. Barut.[7]

A similar correspondence exists also at classical level as has been noted by P. Kustaanheimo and E. Stiefel[8] and was used to regularize the Kepler problem in R^3 by mapping R^3 into R^4 and by an appropriate change of time. In R^4 the equations of motion of the Kepler problem are also linear differential equations with constant coefficients, thus remaining completely regular at the center of motion. In 1979 I. H. Duru and H. Kleinert[9] used these transformations to express the Green's function of the hydrogen atom calculated as a Feynman path integral in terms of the Green's function of the four-dimensional harmonic oscillator.

However, the mathematical theory of Feynman path integral[10–12] does not allow one to deal rigorously with the case of a singular potential like the Coulomb potential. This is the reason why in this paper we treat the corresponding imaginary-time Schrödinger equation which can be solved by the Wiener integral. In the past decade the method of associating stochastic processes with quantum theories for imaginary times has become more and more important. See Ref. 13 and the references therein. In the first section we briefly review the results known about the connection between stochastic differential equations and elliptic equations and Ito's formulas of stochastic calculus. We also recall the formulas allowing a stochastic change of time.

The second section deals with the well-known results of A. Hurwitz about the composition of quadratic forms.

The third section gives the main result of our paper, namely that through a change of variables and the introduction of a stochastic time one can map the stochastic differential equation associated with a harmonic-like potential in R^N into the stochastic differential equation associated with a Coulomb-like potential in R^n. Finally, special cases are considered, in particular the pure harmonic oscillator in four dimensions.

Our purpose is to stress that the method we describe is especially well suited to the treatment of perturbations of the Coulomb potential as was the case in classical mechanics.[8] In a later publication we intend to elaborate this point.

2. SOME RESULTS ABOUT STOCHASTIC CALCULUS

In this section first we briefly describe the well-known results about the connection between the elliptic equations and the stochastic differential equations (see, e.g., Ref. 14 for more details).

Given an equation of the type

$$\left\{-\sum b_{ik}b_{jk}\frac{\partial^2}{\partial x_i \partial x_j} + V(x)\right\}\psi = \frac{\partial\psi}{\partial t},$$

for which there exists a *strictly* positive solution ϕ_0, one makes the change of function

$$\psi(x) = \phi_0(x)\phi(x),$$

and ϕ satisfies the parabolic equation

$$\frac{\partial\phi}{\partial t} = -\sum b_{ik}b_{jk}\frac{\partial^2}{\partial x_i \partial x_j}\phi + 2\sum b_{ik}b_{jk}\frac{\partial F}{\partial x_i}\frac{\partial\phi}{\partial x_j},$$

where $F = \ln\phi_0$.

The solution of this equation with initial condition φ (with appropriate regularity properties) is given by

$$\phi(t,x) = E_{tx}\{\varphi(\xi(t))\},$$

[a)] On leave of absence from the CPT CNRS (Marseille).

 0022-2488/81/071372-05$1.00

where ξ is a stochastic process solution of the stochastic differential equation

$$d\xi_i(t) = 2\left(\sum_{j,k} b_{ik} b_{jk} \frac{\partial F}{\partial x_j}\right) dt + \sum b_{ij} dw_j$$
$$= a_i\, dt + \sum b_{ij} dw_j,$$

where the w_i are independent Wiener processes.

Ito's lemma (formula) allows us to consider a change of variables, viz. if u is a C^2 function then $u(\xi_i(t)) = \zeta(t)$ defines a stochastic process which satisfies the stochastic differential equation (*Ito's lemma*):

$$d\zeta(t) = A\, dt + \sum B_j dw_j,$$

where

$$A = \sum_{k=1}^{N} \frac{\partial u}{\partial x_k} a_k + \frac{1}{2} \sum_{i=1}^{N} \frac{\partial^2 u}{\partial x_i \partial x_k} \sum_{j=1}^{N} b_{ij} b_{kj},$$

$$B_j = \sum_{k=1}^{N} \frac{\partial u}{\partial x_k} b_{kj}.$$

In what follows we shall need a change of time and this can be made through the introduction of a family τ_t of stochastic times defined by

$$t = \int_0^{\tau_t} ds\, g((\zeta(s)),$$

where g is a positive continuous function for which

$$\int_0^{\infty} g(\zeta(s))\, ds = \infty,$$

with probability one. Introducing the random process

$$\eta(s) = \zeta(\tau_s),$$

the new random process $\eta(s)$ satisfies the following stochastic differential equation:

$$d\eta(s) = \frac{A}{g(\eta(s))} ds + \sum \frac{B_j}{[g(\eta(s))]^{1/2}} dw_j.$$

See Ref. 14 for a proof of this result.

3. THE THEOREM OF A. HURWITZ ABOUT COMPOSITION OF QUADRATIC FORMS

Let us consider the following problem: For which value of N does there exist a formula

$$(x_1^2 + \cdots x_N^2)(y_1^2 + \cdots y_N^2) = z_1^2 + \cdots + z_N^2, \tag{3.1}$$

where the z_i are homogeneous bilinear forms in x and y? This problem has been solved by A. Hurwitz[15] for quadratic forms over the complex field. The theorem of A. Hurwitz states that this is only possible if $N = 1,2,4$, or 8. The case $N = 1$ is trivial and we will discuss the other cases briefly. For a proof see Refs. 15 and 16.

The assumption about z implies that

$$z = B(x)\, y, \tag{3.2}$$

$$z_i = \sum_{k,l} (B^k)_i^l x_k\, y_l = \sum_l B(x)_i^l\, y_l, \tag{3.3}$$

and (3.1) is equivalent to the condition on the matrices B^k

$${}^tB^k B^k = 1, \tag{3.4}$$

$${}^tB^k B^l + {}^tB^l B^k = 0, \quad k \neq l, \tag{3.5}$$

as a special case for $x = y$

$$\sum_i z_i^2 = \left(\sum_j x_j^2\right)^2, \tag{3.6}$$

and the matrix $B(x)$ satisfies the condition

$${}^tB(x)B(x) = |x|^2. \tag{3.7}$$

For $N = 2$ one has explicitly the matrix $B(x)$,

$$B(x) = \begin{vmatrix} x_1 & -x_2 \\ x_2 & x_1 \end{vmatrix}. \tag{3.8}$$

This matrix has the following geometrical interpretation.

Let $z = x_1 + ix_2$ be the plane of the trajectory of a moving particle. For regularization purposes Levi-Civita[17] has introduced a parameter plane $w = x_1 + ix_2$ mapped onto the physical plane conformally by the transformation

$$z = w^2, \quad \begin{aligned} y_1 &= x_1^2 - x_2^2 \\ y_2 &= 2x_1 x_2. \end{aligned} \tag{3.9}$$

Accordingly, parabolic coordinates are introduced in the plane of motion. A conical section centered at the origin of the w-plane is transformed into a conical section of the z-plane having one focus at the origin. For this reason the transformation is a very practical method for the discussion of the Kepler problem.

For $N = 4$ a generalization of this transformation is en by

$$B(x) = \begin{vmatrix} x_1 & -x_2 & -x_3 & x_4 \\ x_2 & x_1 & -x_4 & -x_3 \\ x_3 & x_4 & x_1 & x_2 \\ x_4 & -x_3 & x_2 & -x_1 \end{vmatrix}. \tag{3.10}$$

Notice that $\Sigma_l B(x)_4^l x_l = 0$, the mapping $x \to z$ is actually a mapping of R^4 into R^3. This is exactly what we want to study motions in R^3. R^4 is mapped into R^3 in Levi-Civita's style. That is to say, distances are squared and angles at the origin are doubled. For $N = 8$ a matrix satisfying the required properties is, for example,

$$B(x) = \begin{vmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ -x_2 & x_1 & x_4 & -x_3 & x_6 & -x_5 & x_8 & -x_7 \\ -x_3 & -x_4 & x_1 & x_2 & x_7 & -x_8 & -x_5 & x_6 \\ -x_4 & x_3 & -x_2 & x_1 & -x_8 & -x_7 & x_6 & x_5 \\ -x_5 & -x_6 & -x_7 & x_8 & x_1 & x_2 & x_3 & -x_4 \\ -x_6 & x_5 & x_8 & x_7 & -x_2 & x_1 & -x_4 & -x_3 \\ -x_7 & -x_8 & x_5 & -x_6 & -x_3 & x_4 & x_1 & x_2 \\ -x_8 & x_7 & -x_6 & -x_5 & x_4 & x_3 & -x_2 & x_1 \end{vmatrix}. \tag{3.11}$$

Again we remark that

$$\sum_{l=1}^{8} B(x)_i^l x_l = 0, \quad i = 2,3,...,8.$$

Hence $x \to z$ is a mapping from R^8 into R.

Remark 3.1: The matrix $B(x)$ has the following basic properties: the elements are linear homogeneous functions of the x_i; $B(x)$ is orthogonal in the following sense: the scalar product of two different rows vanishes and each row has the norm $|x^2| = x_1^2 + \cdots + x_N^2$.

Remark 3.2: For $N = 4$ the matrix $B(x)$ is connected with the rule of multiplication of quaternions and for $N = 8$ with the multiplication of octonions.

4. TRANSFORMATION OF THE STOCHASTIC DIFFERENTIAL EQUATION

Let us consider the following elliptic equation,

$$\frac{\partial u}{\partial t} = -\frac{1}{2m}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2}u + \frac{m}{2}\Omega\left(\sum_{i=1}^{N} x_i^2\right)u, \tag{4.1}$$

which is the Schrödinger equation (with imaginary time) for a potential which is rotationally invariant in N dimensions.

Let us assume that

$$\Omega(z) \to \infty, \quad z \to +\infty,$$

which ensures that the spectrum of the Hamiltonian is purely discrete. Moreover, we assume that the ground state ϕ_0 is a positive nonzero function (see Ref. 18 for the conditions on Ω).

Our first observation is that ϕ_0 is rotationally invariant, hence

$$\phi_0(x) = \Phi_0\left(\sum_{i=1}^{N} x_i^2\right). \tag{4.2}$$

Consequently the drift term in the associated stochastic differential equation is of the form

$$A_i = -F\left(\sum_{i=1}^{N} x_i^2\right)x_i, \tag{4.3}$$

and the stochastic differential equation can be rewritten as

$$d\eta_i(t) = -F\left(\sum_{i=1}^{N}\eta_i^2\right)\eta_i(t)\,dt + \frac{1}{(2m)^{1/2}}dw_i(t),$$
$$i = 1,2,...,N, \tag{4.4}$$

where the w_i are independent Wiener processes.

Now let us consider n matrices B^i from R^N to R^N with real entries, such that

$$^tB^iB^i = 1, \tag{4.5}$$
$$^tB^iB^j + {}^tB^jB^i = 0, \quad \forall i \neq j.$$

We have discussed in Sec. 3 the possible choices of N and n. Nevertheless, these matrices allow us to define a transformation from $R^N \to R^n$

$$y_i(x) = \sum_{k,l=1}^{N}(B^k)_i^l x_k x_l. \tag{4.6}$$

Then if we define a new stochastic process $\zeta_i(t)$ for $i = 1,...n$

$$\zeta_i(t) = y_i(\eta(t)), \tag{4.7}$$

Ito's formulas allow us to write the stochastic differential equation satisfied by the ζ_i. Explicitly,

$$\frac{\partial y_i}{\partial x_k} = 2\sum_{l=1}^{N}(B^l)_i^k x_l.$$

Hence

$$\sum_{k=1}^{N} A_k\frac{\partial y_i}{\partial x_k} = -2F\left(\sum_{i=1}^{N}x_i^2\right)\sum_{k=1}^{N}\sum_{l=1}^{N}(B^k)_i^l x_k x_l$$
$$= -2F\left(\sum_{j}^{N}x_j^2\right)y_i,$$

whereas

$$\frac{\partial^2 y_i}{\partial x_k \partial x_l} = 2(B^k)_i^l.$$

Hence

$$\sum_{k=1}^{N}\frac{\partial^2 y_i}{\partial x_k^2} = 2c_i.$$

Consequently

$$d\zeta_i(t) = (-2F\{[\Sigma\zeta_j(t)^2]^{1/2}\}\zeta_i(t) + 2c_i)\,dt$$
$$+ \frac{2}{\sqrt{m}}\sum_{k,l=1}^{N}(B^k)_i^l\eta_k(t)\,dw_l(t). \tag{4.8}$$

We introduce now a stochastic time τ_t (see Ref. 14) through the formula

$$t = \int_0^{\tau_t} ds\left[\sum_{i=1}^{n}\zeta_i(s)^2\right]^{1/2}. \tag{4.9}$$

Then if we define a new stochastic process Θ_i

$$\Theta_i(s) = \zeta_i(\tau_s), \quad i = 1,2,...,n, \tag{4.10}$$

it satisfies

$$d\Theta_i(s) = \left\{-\frac{2F\{[\Sigma\Theta_i(s)^2]^{1/2}\}}{[\Sigma\Theta_i(s)^2]^{1/2}}\Theta_i(s) + \frac{2c_i}{[\Sigma\Theta_i(s)^2]^{1/2}}\right\}ds$$
$$+ \frac{1}{m^{1/2}}\frac{2}{[\Sigma\Theta_i(s)^2]^{1/4}}\sum_{k,l}^{N}(B^k)_i^l\eta_k(\tau_s)dw_l(s). \tag{4.11}$$

Equation (4.11) can be simplified by the following remark. Let ξ_i be the stochastic processes whose differentials are given by

$$d\xi_i(s) = \frac{1}{m^{1/2}}\frac{2}{[\Sigma\Theta_i(2)^2]^{1/4}}\sum_{k,l}(B^k)_i^l\eta_k(\tau_s)\,dw_l. \tag{4.12}$$

They are independent Wiener processes.

Indeed, let h be a smooth function. Then

$$H(t,x) = E_{tx}\{h[\xi(t)]\},$$

satisfies the differential equation

$$\frac{\partial}{\partial t}H(t,x) = -\frac{2}{m}\sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2}H(t,x), \tag{4.13}$$

with the initial condition $\lim_{t\to 0} H(t,x) = h(x)$. Hence if we specialize h to be

$$h(x) = \exp(i\sum\lambda_i x_i)$$

we get

$$H(t,x) = \exp\left(-\frac{2t}{m}\sum\lambda_i^2 + i\sum\lambda_i x_i\right). \tag{4.14}$$

Hence the result.

Consequently one writes Eq. (4.11) as

$$d\Theta_i(s) = \left\{ -2\frac{F\{[\Sigma\Theta_i(s)^2]^{1/2}\}}{[\Sigma\Theta_i{}^2(s)]^{1/2}}\Theta_i(s) + \frac{2c_i}{[\Sigma\Theta_i(s)^2]^{1/2}}\right\} ds + \frac{2}{m^{1/2}} d\xi_i. \tag{4.15}$$

Notice that for $N = 2$ and $N = 4$ the c_i's are zero but for $N = 8$, $C_1 = 16$. Solutions of differential equations of this form have been called "distorted Brownian motion." See, e.g., Ref. 13.

Now let us consider a C^2-bounded function φ on R^3. Then

$$\psi(s, y) = E_{sy}\{\varphi[\Theta_i(s)]\}, \tag{4.16}$$

satisfies the parabolic equation

$$\frac{\partial\psi}{\partial s} - 2\frac{F[(\Sigma_{i=1}^n y_i^2)^{1/2}]}{(\Sigma_{i=1}^n y_i^2)^{1/2}}\sum_{j=1}^n y_j\frac{\partial\psi}{\partial y_j} + \frac{2}{m}\sum_{i=1}^n\frac{\partial^2\psi}{\partial y_i^2} = 0, \tag{4.17}$$

which actually corresponds to a (imaginary time) Schrödinger equation with the spherically symmetric potential given, up to a constant, by

$$V(R) = (1/2m')(\nabla A + A\cdot A), \quad m' = m/4, \tag{4.18}$$

A being the drift in (4.3).

Explicitly from (4.15) we have the new potential (in the case $N = 2,4$)

$$V(R) = \frac{1}{m'}\left[2F(R)^2 - F'(R) - \frac{2F(R)}{R}\right], \tag{4.19}$$

which corresponds to the old one

$$v(r^2) = \frac{1}{m}[F(r^2)r^2 - 2F'(r^2)r^2 - 4F(r^2)]. \tag{4.20}$$

As a special case we recover the pure Coulomb potential if F is a constant, i.e., if $v(r^2)$ is a harmonic potential.

We can summarize our results in the following.

Theorem 4.1: To any stochastic process η in R^N ($N = 1,2,4,8$) the solution of the system of stochastic differential equations is

$$d\eta_i(t) = -F\left[\sum_{i=1}^N \eta_i(t)^2\right]\eta_i(t)dt + \frac{1}{(2m)^{1/2}} dw_i,$$

where the w_i's are independent Wiener processes and F such that the solution is unique. One can associate a stochastic process Θ in R^n ($n = 1,2,3,1$) by the following relation

$$\Theta(s) = B(\eta_{\tau_s})\eta(\tau_s),$$

where $B(x)$ is an $N \times N$ matrix whose entries are linear and homogeneous in x_i and satisfy

$${}^tB(x)B(x) = \sum_{i=1}^N x_i^2,$$

and the stochastic time τ_s is defined by

$$s = \int_0^{\tau_s} dt\left[\sum_{i=1}^N \eta_i(t)^2\right]^{1/2}.$$

The random process Θ in R^n satisfies the stochastic differential equations

$$d\Theta_i(s) = \left\{ -2\frac{F\{[\Sigma\Theta_i(s)^2]^{1/2}\}}{[\Sigma\Theta_i(s)^2]^{1/2}}\Theta_i(s) + \frac{c_i}{[\Sigma\Theta_i(s)^2]^{1/2}}\right\} ds + \frac{2}{\sqrt{m}} d\xi_i,$$

where ξ_i are independent Wiener processes.

Corolary 4.2: For $F \equiv \omega$ (a positive constant) and $N = 4$ the process Θ in R^3 is associated with the pure Coulomb problem with a mass $m' = m/4$ and a coupling constant $g = 2\omega$.

Remark 4.1: The previous result can be applied to the radial (imaginary time) Schrödinger equation if we remark that the centrifugal potential in R^N ($N \geqslant 2$) is of the type

$$k(k + N - 2)/r^2,$$

and can be absorbed in the solution through a change of the drift term. Explicitly one has the equation

$$\frac{\partial}{\partial t} U(t,r) = \frac{1}{2m}\left[-\frac{\partial^2}{\partial r^2} + 2\ln\Omega + \frac{2\alpha + N - 1}{r}\frac{\partial}{\partial r}\right]U(t,r).$$

Then we are reduced to the previous case for one-dimensional systems.

Remark 4.2: In Ref. 8 the kernel of the transformation (4.6) has been explicitly given for $N = 2$ or 4. The counter image of a point in R^3 is a circle in R^4 of radius $\sqrt{r}$. Nevertheless it is shown that one can choose an inverse if we use the auxiliary condition

$$x_4dx_1 - x_3dx_2 + x_2dx_3 - x_1dx_4 = 0.$$

Then we can compare the two systems of stochastic differential equations and derive explicitly an expression for the Radon–Nykodim derivative of the corresponding measures; each of them is equivalent to the Wiener measure through a well-known formula of the Feynman–Kac type (see, e.g., Ref. 18).

Remark 4.3: As is pointed out in Ref. 14, any homogeneous Markov process can be obtained from a Wiener process by means of a random time substitution and a state transformation. A similar remark was made for the Coulomb problem many years ago by V. Fock,[1] which showed the equivalence of the bound-state problem of the Coulomb problem to force free motion on the surface of a 4-dimensional sphere.

ACKNOWLEDGMENTS

It gives us special pleasure to express our thanks to Professors Sergio Albeverio and Ludwig Streit for helpful discussions. One of us (M. S.) wants to express his deep appreciation to the members of the Fakultät für Physik der Universität Bielefeld for their kind hospitality and the USP Mathematisierung der Universität Bielefeld for giving full support to a most stimulating workshop where this work was initiated.

[1] V. Fock, Z. Phys. **98**, 145 (1935).
[2] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, 1939), Vol. XXIII.

[3]D. Fivel, Phys. Rev. **142**, 129–1235 (1966).
[4]F. Goded, Nuovo Cimento **26**, 1346 (1962).
[5]F. Ravndal and T. Toyoda, Nucl. Phys. B **3**, 372 (1967).
[6]V. A. Dulock and H. V. McIntosh, Pacific J. Math. **19**, 39–55 (1966).
[7]A. O. Barut, *Dynamical groups and generalized symmetries in quantum theory* (University of Canterbury, New Zealand, 1971).
[8]P. Kustaanheimo and E. Stiefel, J. Reine Angew. Math. **218**, 204 (1965).
[9]I. H. Duru and H. Kleinert, Phys. Lett. B **84**, 185–188 (1979).
[10]S. Albeverio and R. Høegh-Krohn, "Mathematical Theory of Feynman Path Integrals," in *Lecture Notes in Mathematics* 523 (Springer, New York, 1976).
[11]C. Dewitt–Morette, A. Maheshwari, and B. Nelson, Physics Reports No. 5, 255–372 (1979).
[12]Ph. Combe, R. Høegh-Krohn, R. Rodriguez, M. Sirugue, and M. Sirugue-Collin, Commun. Math. Phys. **77**, 269 (1980).
[13]L. Streit, "Energy forms: Schrödinger Theory Processes" in *Stochastic Differential Equations in Physics* (Les Houches, 1980), to appear in Physics Reports.
[14]I. I. Gihman and A. V. Skorohod, *Stochastic differential equations* (Springer, 1972).
[15]A. Hurwitz, Math. Werke II, 565–571.
[16]T. Y. Lam, *The algebraic theory of quadratic forms* (Benjamin, New York, 1973).
[17]T. Levi-Civita, *Sur la résolution qualitative du problème restreint des trois corps* (Opere Matematice, Bologna, 1956), Vol. 2.
[18]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I–IV* (Academic, New York, 1972–1978).

Jump conditions for fields that have infinite, integrable singularities at an interface

Robert C. Costen
NASA Langley Research Center, Hampton, Virginia 23665

(Received 20 July 1979; accepted for publication 6 February 1981)

Jump conditions are derived for a basic set of first order partial differential equations whose fields have infinite, integrable singularities, as well as finite discontinuities, at a moving and deforming interface. These basic formulas are then applied to Maxwell's electrodynamic equations and yield the jump conditions that hold when the electric and magnetic fields have such singularities. These jump conditions are shown to be generalizations of formulas previously derived for double charge layers in electrostatics and for an interface with surface magnetization density in magnetostatics. The basic formulas are also used to obtain jump conditions for the wave equation and other second order partial differential equations whose fields have finite discontinuities at an interface. A similar application to second order vector identities yields a new set of *jump identities*. These identities show that the normal and tangential components of the jump in a vector field are kinematically interdependent; e.g., shock waves and vortex sheets are kinematically linked—a fact that may be significant for shock-slip flows in aerodynamics. The jump identities also indicate the fields that must be measured at an epoch in order to calculate the instantaneous growth/decay rate of a propagating discontinuity, such as an atmospheric front. A feature of the derivation is that the field jumps and surface densities are mathematically defined as continuous and differentiable functions of three-dimensional space and time that assume physical values on the physical interface. This approach is simpler and more general than previous approaches that define jumps and surface densities only on the physical interface because the brackets (jumps) now commute with *all* derivatives, instead of with only the tangential and displacement derivatives. Boundary value problems with propagating infinite singularities in the electrodynamic fields are presented as examples.

PACS numbers: 02.30.Jr, 03.50.De

I. INTRODUCTION

The jump conditions that hold for fields having step function discontinuities at an interface do not, in general, apply if the fields have coincident Dirac δ function or higher order integrable singularities. For example, if curl $\mathbf{p} = 0$, the familiar jump condition $\hat{n} \times [\mathbf{p}] = 0$ does not usually apply if the surface density $\mathbf{P} \equiv \lim_{\Delta n \to 0} \int_{\Delta n} \mathbf{p}\, dn$ is definite. Such a surface density could arise from a discontinuity in the potential function for $\mathbf{p}$, and several restricted examples of this type have been treated by Stratton.[1] The purpose of this investigation is to derive more general jump conditions for wider applications in physics and mechanics.

In our derivation, the field jumps and surface densities are treated as continuous and differentiable functions of three-dimensional space and time that assume physical values on the physical interface. This approach, which is justified in a subsequent section, is simpler and more general than previous treatments[2] that restrict the mathematical definition of these entities to the physical interface itself. Elementary examples are given in Appendices B and C.

Derivations of jump conditions at moving and deforming surfaces may be done either in four-dimensional space–time,[3,4] or in three-dimensional space by utilizing moving regions of integration[5,6]; we will use the latter three-dimensional approach. The relevant characteristics of moving surfaces and the sign convention are summarized in Appendix A.

II. DERIVATION OF BASIC JUMP CONDITIONS

Derivation: Consider equations of the form

$$\lambda = -\frac{\partial \theta}{\partial t}, \tag{1a}$$

$$\mathbf{m} = \operatorname{grad} \psi, \tag{1b}$$

$$\mathbf{b} = -\frac{\partial \mathbf{w}}{\partial t}, \tag{1c}$$

$$\mathbf{j} = \operatorname{curl} \mathbf{p}, \tag{1d}$$

$$\xi = \operatorname{div} \mathbf{q}, \tag{1e}$$

where the fields $\theta, \psi, \mathbf{w}, \mathbf{p}, \mathbf{q}$ have step-function (H-function) and δ-function singularities at a moving and deforming interface S that has local unit normal $\hat{n}$ and speed of displacement N [Eqs. (A2),(A3)], as shown in Fig. 1. As a consequence of the movement of S and of the space and time derivatives in Eqs. (1), the fields $\lambda, \mathbf{m}, \mathbf{b}, \mathbf{j}, \xi$ have H-function, δ-function, and δ'-function singularities on S, where δ' denotes $d\delta(u)/du$. We shall write, for brevity,

$$(\theta, \psi, \mathbf{w}, \mathbf{p}, \mathbf{q}) = O(\delta) \tag{2a}$$

and

$$(\lambda, \mathbf{m}, \mathbf{b}, \mathbf{j}, \xi) = O(\delta'), \tag{2b}$$

where the highest order singularity is indicated and the existence of the lower order singularities is presumed.

In order to derive the corresponding jump conditions, Eqs. (1), with the exception of (1c) which is essentially repet-

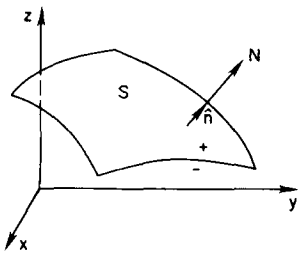


FIG. 1. Moving and deforming interface S with unit normal $\hat{n}$ and speed of displacement N. Positive and negative sides of S are defined relative to $\hat{n}$, as shown.

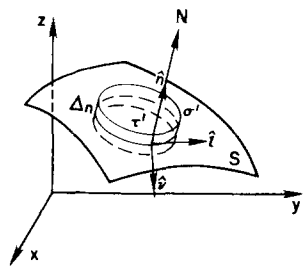


FIG. 2. Volume of integration τ' that intersects interface S. The bounding surface σ' consists of a right quasicylinder of length Δn and its two quasiplanar end surfaces, all with outward unit normal $\hat{v}$. On the circuit line l, where σ' intersects S, the unit vectors $\hat{v}$, $\hat{l}$, and $\hat{n}$ are mutually orthogonal, as shown.

itive, are expressed in integral form

$$\iiint \lambda \, d\tau = -\frac{d}{dt}\iiint \theta \, d\tau + \oiint \theta \, \mathbf{U}\cdot\hat{v} \, d\sigma, \tag{3a}$$

$$\iiint \mathbf{m} \, d\tau = \oiint \psi \hat{v} \, d\sigma, \tag{3b}$$

$$\iiint \mathbf{j} \, d\tau = \oiint \hat{v}\times \mathbf{p} \, d\sigma, \tag{3c}$$

$$\iiint \xi \, d\tau = \oiint \mathbf{q}\cdot\hat{v} \, d\sigma, \tag{3d}$$

where volume τ and its surrounding surface σ, with outward unit normal $\hat{v}$, are taken to be moving with the mathematical velocity field $\mathbf{U}(\mathbf{r},t)$, which is distinct from any physical field, and where Gauss' divergence theorem and the transport theorem[2] have been used. Since the second integral in Eq. (3a) is purely a function of t, it is appropriate to take its time derivative. We adhere to Hilbert's view that natural laws should be expressed in integral form; in regular regions, Eqs. (3) reduce to (1), and at a singular interface they give the jump conditions that we shall now derive.

We choose τ to be the curvilinear pill box τ' that intersects the interface S, as shown in Fig. 2, and whose height Δn is everywhere bisected by S and may be arbitrarily small; τ' is further taken to move with S through normal displacements by specifying that $\mathbf{U}(\mathbf{r},t)$ satisfies the condition

$$\mathbf{U}\cdot\hat{n} = N. \tag{4}$$

Subject to these conditions and to Eqs. (2), Eqs. (3) become in the limit as $\Delta n \to 0$

$$\iint \Lambda \, dS = -\frac{d}{dt}\iint \Theta \, dS + \iint [\theta] N \, dS + \oint \Theta \, \mathbf{U}\cdot(\hat{l}\times\hat{n}) dl, \tag{5a}$$

$$\iint \mathbf{M} \, dS = \iint [\psi]\hat{n} \, dS + \oint \Psi(\hat{l}\times\hat{n}) \, dl, \tag{5b}$$

$$\iint \mathbf{J} \, dS = \iint \hat{n}\times[\mathbf{p}] \, dS + \oint (\hat{l}\times\hat{n})\times \mathbf{P} \, dl, \tag{5c}$$

$$\iint \Xi \, dS = \iint [\mathbf{q}]\cdot\hat{n} \, dS + \oint \mathbf{Q}\cdot(\hat{l}\times\hat{n}) \, dl, \tag{5d}$$

where the capitalized fields are surface densities that correspond to the lower case volumetric densities and are defined, representatively, by

$$\Theta \equiv \lim_{\Delta n \to 0} \int_{\Delta n} \theta \, dn \tag{6a}$$

for the fields of Eq. (2a) and by the more general relation

$$\iint \Lambda \, dS \equiv \lim_{\Delta n \to 0} \iiint \lambda \, d\tau' \tag{6b}$$

for the fields of (2b).

We now assume that all of the integrands in Eqs. (5) are continuous and differentiable functions of three-dimensional space and time. (This assumption will be examined in Sec. III.) The second term in Eq. (5a) may then be transformed by the kinematic theorem[6]

$$\frac{d}{dt}\iint \Theta \, dS = \iint \left\{\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\Theta - 2\Omega N\Theta\right\} dS + \oint \Theta \, \mathbf{U}\cdot(\hat{l}\times\hat{n}) \, dl, \tag{7}$$

where Ω is the mean curvature of S [Eq. (A4)], and the line integrals in Eqs. (5b)–(5d) may be transformed by Stokes' theorem[7] (actually due to Ampere, Kelvin, and Hankel[2]); then Eqs. (5) become

$$\iint \left\{-\Lambda - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\Theta + 2\Omega N\Theta + N[\theta]\right\} dS = 0, \tag{8a}$$

$$\iint \left\{-\mathbf{M} + \hat{n}[\psi] + \hat{n}2\Omega\Psi + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)\Psi\right\} dS = 0, \tag{8b}$$

$$\iint \{-\mathbf{J} + \hat{n}\times[\mathbf{p}] - ((\hat{n}\times\mathbf{P})\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\mathrm{curl}\,\mathbf{P}) - \hat{n}\times\mathrm{grad}(\hat{n}\cdot\mathbf{P})\} dS = 0, \tag{8c}$$

$$\iint \{-\Xi + \hat{n}\cdot[\mathbf{q}] + \hat{n}\cdot\mathrm{curl}(\hat{n}\times\mathbf{Q})\} dS = 0, \tag{8d}$$

where the third term in Eq. (8c) is expressed for brevity in Cartesian notation. The integrands in Eqs. (8) are assumed to be continuous, and since these equations hold for any area on S the integrands must vanish:

$$\Lambda = N[\theta] + 2\Omega N\Theta - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\Theta, \tag{9a}$$

$$\mathbf{M} = \hat{n}[\psi] + \hat{n}2\Omega\Psi + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)\Psi, \tag{9b}$$

$$\mathbf{B} = N[\mathbf{w}] + 2\Omega N\mathbf{W} - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\mathbf{W}, \tag{9c}$$

$$\mathbf{J} = \hat{n}\times[\mathbf{p}] - ((\hat{n}\times\mathbf{P})\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\text{curl}\,\mathbf{P}) - \hat{n}\times\text{grad}(\hat{n}\cdot\mathbf{P}), \tag{9d}$$

$$\Xi = \hat{n}\cdot[\mathbf{q}] + \hat{n}\cdot\text{curl}(\hat{n}\times\mathbf{Q}), \tag{9e}$$

where Eq. (9c) is a vector form of (9a). Equations (9) are the jump conditions for Eqs. (1) when the fields have singularities given by (2) and surface densities defined by (6). They may be used to obtain by inspection the jump conditions for other first order partial differential equations whose dependent variables are $O(\delta)$ at an interface, as illustrated in Sec. V, and for second order equations whose variables are $O(H)$, as illustrated in Sec. IV. Equations (9) are amenable to simple checks, as shown in Appendix B.

Physical interpretation: Many of the terms in Eqs. (9) are easily interpreted from a physical viewpoint. In Eq. (9a) for the surface *partial-time-derivative* density Λ, the term $N[\theta]$ represents the rate of accumulation of Θ due to the "snowplow" action; the term $2\Omega N\Theta$ accounts for an expanding/receding bulge on the interface that tends to cause local dilution/concentration of Θ; and the term $(\partial/\partial t + N\partial/\partial n)\Theta$ represents the time rate of change of Θ following the moving interface along its local normal. In Eq. (9b) for the surface *gradient* density $\mathbf{M}$, the terms $\hat{n}[\psi]$ and $\hat{n}2\Omega\Psi$ are the normal contributions from the discontinuity in ψ and from the curvature of the interface, respectively; and the term $(\nabla - \hat{n}\partial/\partial n)\Psi$ is the tangential contribution.

In Eq. (9d) for the surface *curl* density $\mathbf{J}$, the term $((\hat{n}\times\mathbf{P})\cdot\nabla)\hat{n}$ represents contributions from the curvature and twist of the interface along the direction of $\hat{n}\times\mathbf{P}$; and $\hat{n}(\hat{n}\cdot\text{curl}\,\mathbf{P})$ is the normal contribution of curl $\mathbf{P}$. The term $-\hat{n}\times\text{grad}(\hat{n}\cdot\mathbf{P})$ may be interpreted by taking the scalar line integral of $\mathbf{p}$ about a quasirectangular circuit that penetrates interface S, as shown in Fig. 3, and dividing by Δl; as $\Delta n\to 0$, the parallel sides contribute $\hat{n}\times[\mathbf{p}]$, and the perpendicular sides give $-\hat{n}\times\text{grad}(\hat{n}\cdot\mathbf{P})$ as $\Delta l\to 0$ also.

In Eq. (9e) for the surface *divergence* density Ξ, the term $\hat{n}\cdot[\mathbf{q}]$ is the contribution from the discontinuity in $\mathbf{q}$; the term $\hat{n}\cdot\text{curl}(\hat{n}\times\mathbf{Q})$ may be rewritten by Stokes' and Gauss' theorems as $\text{div}_2(\mathbf{Q} - \hat{n}Q_n)$ and represents the two-dimensional divergence of the tangential vector $(\mathbf{Q} - \hat{n}Q_n)$. The linking of $\hat{n}\cdot[\mathbf{q}]$ with $(\hat{n}\times\mathbf{Q})$ in Eq. (9e) and of $\hat{n}\times[\mathbf{p}]$ with $\hat{n}\cdot\mathbf{P}$ in (9d) will gain further physical significance in Sec. IV.

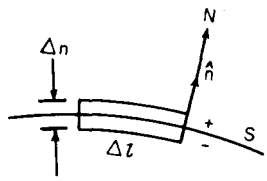


FIG. 3. Circuit of integration on a plane locally normal to interface S. The height Δn of the circuit is bisected by S, and in the limiting process $\Delta n\to 0$ before $\Delta l\to 0$.

Note from the combinations of terms in Eqs. (9) that only two types of derivatives occur: (a) spatial derivatives that are tangent to S; and (b) the displacement derivative $(\partial/\partial t + N\partial/\partial n)$ that follows S. In order for the terms to be *individually* meaningful, the surface densities, field jumps, and $\hat{n}, N, \Omega$ must all be regarded as continuous and differentiable fields, in agreement with the assumption preceding Eq. (7). This concept will be developed in the next section.

III. CONTINUITY AND DIFFERENTIABILITY OF FIELD JUMPS AND SURFACE DENSITIES

We shall show that field jumps and surface densities can be viewed as continuous and differentiable functions of $(\mathbf{r},t)$ through mathematical continuation and induction. Instead of Eqs. (2), take

$$(\theta',\psi',\mathbf{w}',\mathbf{p}',\mathbf{q}') = O(H) \tag{10a}$$

and

$$(\lambda',\mathbf{m}',\mathbf{b}',\mathbf{j}',\xi') = O(\delta); \tag{10b}$$

then Eqs. (9) become

$$\Lambda' = N[\theta'], \tag{11a}$$

$$\mathbf{M}' = \hat{n}[\psi'], \tag{11b}$$

$$\mathbf{B}' = N[\mathbf{w}'], \tag{11c}$$

$$\mathbf{J}' = \hat{n}\times[\mathbf{p}'], \tag{11d}$$

$$\Xi' = \hat{n}\cdot[\mathbf{q}']. \tag{11e}$$

From a physical viewpoint, the jumps $[\theta'],[\psi'],[\mathbf{w}'],[\mathbf{p}'],[\mathbf{q}']$ and the surface densities $\Lambda',\mathbf{M}',\mathbf{B}',\mathbf{J}',\Xi'$ are specified only on the physical interface S; however, we may mathematically continue the definitions of these quantities throughout $(\mathbf{r},t)$ by the following abstractions:

(1) The physical interface S is viewed as one of a manifold of moving and deforming mathematical surfaces defined by $\phi(x,y,z,t) = \text{const}$, where ϕ is continuous and differentiable in $(\mathbf{r},t)$. Then, as defined by Eqs. (A2)–(A4), $\hat{n}$, N, and Ω are also continuous and differentiable in $(\mathbf{r},t)$.

(2) The jump in a physical field $\mathbf{v}$ is given by $[\mathbf{v}] = \mathbf{v}^+ - \mathbf{v}^-$, where $\mathbf{v}^+$ is physically specified throughout $(\mathbf{r}^+,t)$ and $\mathbf{v}^-$ throughout $(\mathbf{r}^-,t)$, and where the $\pm$ spaces are separated by the physical interface S. By mathematical continuation across S, both $\mathbf{v}^+$ and $\mathbf{v}^-$ may be defined as continuous and differentiable functions throughout $(\mathbf{r},t)$, whence the jump $[\mathbf{v}] = \mathbf{v}^+(\mathbf{r},t) - \mathbf{v}^-(\mathbf{r},t)$ is continuous and differentiable in $(\mathbf{r},t)$. (This abstraction is consistent with the classical treatment of boundary value problems, where the component mathematical solutions are often defined beyond their regions of physical validity.) It follows that

$$\nabla*[\mathbf{v}] = \nabla*\mathbf{v}^+ - \nabla*\mathbf{v}^- = [\nabla*\mathbf{v}], \tag{12a}$$

where ∇ is the three-dimensional gradient and $*$ represents any appropriate vector or scalar product, and

$$\frac{\partial[\mathbf{v}]}{\partial t} = \frac{\partial\mathbf{v}^+}{\partial t} - \frac{\partial\mathbf{v}^-}{\partial t} = \left[\frac{\partial\mathbf{v}}{\partial t}\right]. \tag{12b}$$

Therefore the brackets (jumps) commute with *all* derivatives. (Previous treatments[2] admitted such commutability only for the tangential and displacement derivatives; commutations of brackets with the normal and partial time derivatives were excluded.) Because of these abstractions and

Eqs. (11), the surface densities $\Lambda', \mathbf{M}', \mathbf{B}', \mathbf{J}', \Xi'$ are also continuous and differentiable in $(\mathbf{r},t)$.

We further infer from Eqs. (11) that for any vector field $\mathbf{v}$ or scalar field χ of $O(\delta)$, the corresponding surface densities $\mathbf{V}$ and X may be represented by the forms

$$\mathbf{V} = N[\mathbf{u}] + \hat{n}[\eta] + \hat{n}\times[\mathbf{s}], \tag{13a}$$

$$\mathrm{X} = N[\mu] + \hat{n}\cdot[\mathbf{k}]. \tag{13b}$$

Continuing inductively, we reconsider Eqs. (9) subject to (2). The surface densities $\Theta, \Psi, \mathrm{W}, \mathbf{P}, \mathbf{Q}$ may now be expressed as in Eqs. (13), and we conclude that the surface densities $\Lambda, \mathbf{M}, \mathbf{B}, \mathbf{J}, \Xi$ are also continuous and differentiable in $(\mathbf{r},t)$. Therefore, field jumps and surface densities that are physically defined only on an interface may be represented mathematically by continuous and differentiable fields. Examples of such representations are given in Appendices B and C.

IV. JUMP CONDITIONS FOR SECOND ORDER PARTIAL DIFFERENTIAL EQUATIONS

Jump identities: A specialized form of Eqs. (1) is given by

$$\gamma = -\frac{\partial \eta}{\partial t}, \tag{14a}$$

$$\mathbf{g} = \mathrm{grad}\eta, \tag{14b}$$

$$\mathbf{f} = -\frac{\partial \mathbf{v}}{\partial t}, \tag{14c}$$

$$\mathbf{c} = \mathrm{curl}\mathbf{v}, \tag{14d}$$

$$\epsilon = \mathrm{div}\mathbf{v}, \tag{14e}$$

By differentiating and combining Eqs. (14) in appropriate pairs, we obtain

$$\mathrm{grad}\gamma = -\frac{\partial \mathbf{g}}{\partial t}, \tag{15a}$$

$$\mathrm{curl}\,\mathbf{f} = -\frac{\partial \mathbf{c}}{\partial t}, \tag{15b}$$

$$\mathrm{div}\,\mathbf{f} = -\frac{\partial \epsilon}{\partial t}, \tag{15c}$$

$$\mathrm{grad}\epsilon = \mathrm{curl}\mathbf{c} + \nabla^2\mathbf{v}, \tag{15d}$$

which are second order identities in $\mathbf{v}$ and η. Reference to Eqs. (1) and (2a) shows that jump conditions can be obtained for Eqs. (15) provided $(\gamma, \mathbf{g}, \mathbf{f}, \mathbf{c}, \epsilon) = O(\delta)$ and $(\eta, \mathbf{v}) = O(H)$, whence the jump conditions for Eqs. (14) become

$$\Gamma = N[\eta], \tag{16a}$$

$$\mathbf{G} = \hat{n}[\eta], \tag{16b}$$

$$\mathbf{F} = N[\mathbf{v}], \tag{16c}$$

$$\mathbf{C} = \hat{n}\times[\mathbf{v}], \tag{16d}$$

$$E = \hat{n}\cdot[\mathbf{v}]. \tag{16e}$$

The jump conditions for Eqs. (15) are found by inspection from Eqs. (1) and (9):

$$\hat{n}[\gamma] + \hat{n}2\Omega\Gamma + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)\Gamma = N[\mathbf{g}] + 2\Omega N\mathbf{G} - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\mathbf{G}, \tag{17a}$$

$$\hat{n}\times[\mathbf{f}] - ((\hat{n}\times\mathbf{F})\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\mathrm{curl}\mathbf{F}) - \hat{n}\times\mathrm{grad}(\hat{n}\cdot\mathbf{F}) = N[\mathbf{c}] + 2\Omega N\mathbf{C} - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\mathbf{C}, \tag{17b}$$

$$\hat{n}\cdot[\mathbf{f}] + \hat{n}\cdot\mathrm{curl}(\hat{n}\times\mathbf{F}) = N[\epsilon] + 2\Omega NE - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)E, \tag{17c}$$

$$\hat{n}[\epsilon] + \hat{n}2\Omega E + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)E = \hat{n}\times[\mathbf{c}] - ((\hat{n}\times\mathbf{C})\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\mathrm{curl}\mathbf{C}) + \left[\frac{\partial \mathbf{v}}{\partial n}\right]. \tag{17d}$$

Substituting Eqs. (16) into (17) and using Eqs. (12) and (A5) give

$$\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)[\eta] = -[\gamma] + N\hat{n}\cdot[\mathbf{g}], \tag{18a}$$

$$\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)[\mathbf{v}] = -[\mathbf{f}] + \hat{n}N[\epsilon] - N\hat{n}\times[\mathbf{c}] + \hat{n}2\Omega N(\hat{n}\cdot[\mathbf{v}]) + N\left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)(\hat{n}\cdot[\mathbf{v}]) + N((\hat{n}\times(\hat{n}\times[\mathbf{v}]))\cdot\nabla)\hat{n} - \hat{n}N(\hat{n}\cdot\mathrm{curl}(\hat{n}\times[\mathbf{v}])), \tag{18b}$$

$$\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)(\hat{n}\times[\mathbf{v}]) = -\hat{n}\times[\mathbf{f}] + N[\mathbf{c} - \hat{n}c_n] + 2\Omega N\hat{n}\times[\mathbf{v}] + ((\hat{n}\times[\mathbf{v}])\cdot\nabla)(N\hat{n}) + \hat{n}\times\mathrm{grad}(N\hat{n}\cdot[\mathbf{v}]), \tag{18c}$$

$$\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)(\hat{n}\cdot[\mathbf{v}]) = -\hat{n}\cdot[\mathbf{f}] + N[\epsilon] + 2\Omega N\hat{n}\cdot[\mathbf{v}] - \hat{n}\cdot\mathrm{curl}(N\hat{n}\times[\mathbf{v}]), \tag{18d}$$

$$\left[\frac{\partial \mathbf{v}}{\partial n}\right] = \hat{n}[\epsilon] - \hat{n}\times[\mathbf{c}] + \hat{n}2\Omega(\hat{n}\cdot[\mathbf{v}]) + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)(\hat{n}\cdot[\mathbf{v}]) + ((\hat{n}\times(\hat{n}\times[\mathbf{v}]))\cdot\nabla)\hat{n} - \hat{n}(\hat{n}\cdot\mathrm{curl}(\hat{n}\times[\mathbf{v}])), \tag{18e}$$

where Eq. (18b) results from combining a vector form of Eq. (18a) with (18e). When one or more of the fields $\gamma, \mathbf{g}, \mathbf{f}, \mathbf{c}, \epsilon$ are known explicitly, Eqs. (18) provide a variety of relations, as illustrated in Sec. V.

The jump identities (18) have the following physical interpretations:

(a) Equations (18c,d) show that the normal and tangential components of the jump in a vector field $\mathbf{v}$ are kinematically interdependent. We conclude, for example, that shock waves and vortex sheets are kinematically linked—a fact that may be significant for shock-slip flows in aerodynamics.

(b) Equations (18) indicate the quantities that must be measured at an epoch in order to calculate the instantaneous growth/decay rate of a propagating discontinuity, such as an atmospheric front.

(c) Equations (18) are seen to be identities subject only to Eqs. (12), (14), and (A4)–(A8); hence their validity transcends the derivation given, and they apply to fields η and $\mathbf{v}$ that have interface singularities of *arbitrary* order. Equations (18)

also remain valid when all of the brackets are removed. Then they apply individually to the $\pm$ sides of the interface; also, in the absence of an interface, they give the normal and displacement derivatives for continuous fields relative to a moving reference surface.

Jump conditions for other equations: The same procedure may be used to obtain jump conditions for other second order equations. Several representative equations, including Poisson's equation, the diffusion equation, and the wave equation, may be written

$$\nabla^2\varphi = \xi, \tag{19a}$$

$$\text{curl curl}\mathbf{u} = \mathbf{j}, \tag{19b}$$

$$\nabla^2\theta - \frac{1}{k}\frac{\partial\theta}{\partial t} = 0, \tag{19c}$$

$$\nabla^2\psi - \frac{1}{\omega_c^2}\frac{\partial^2\psi}{\partial t^2} = 0, \tag{19d}$$

where k and ω_c are constants. If $(\varphi,\mathbf{u},\theta,\psi) = O(H)$ and $(\xi,\mathbf{j}) = O(\delta')$, the corresponding jump conditions are given by

$$\left[\frac{\partial\varphi}{\partial n}\right] = \Xi, \tag{20a}$$

$$\hat{n}\times[\text{curl}\mathbf{u}] - ((\hat{n}\times(\hat{n}\times[\mathbf{u}]))\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\text{curl}(\hat{n}\times[\mathbf{u}])) = \mathbf{J}, \tag{20b}$$

$$\left[\frac{\partial\theta}{\partial n}\right] + \frac{N}{k}[\theta] = 0, \tag{20c}$$

$$\omega_c^2\left[\frac{\partial\psi}{\partial n}\right] + N\left[\frac{\partial\psi}{\partial t}\right] - 2\Omega N^2[\psi] + \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)(N[\psi]) = 0. \tag{20d}$$

Jump condition (20a) is familar from electrostatics. When $N = \omega_c$, Eq. (20d) becomes

$$\left(\frac{\partial}{\partial t} + \omega_c\frac{\partial}{\partial n}\right)[\psi] - \Omega\omega_c[\psi] = 0, \tag{20e}$$

and this jump condition will be applied in Appendix C.

V. APPLICATIONS IN ELECTRODYNAMICS

Jump conditions: Equations for the electrodynamic potentials φ and $\mathscr{A}$ may be written

$$\mathscr{E} + \text{grad}\varphi = -\frac{\partial\mathscr{A}}{\partial t}, \tag{21a}$$

$$\mathscr{B} = \text{curl}\mathscr{A}, \tag{21b}$$

$$-\frac{1}{\omega_c^2}\frac{\partial\varphi}{\partial t} = \text{div}\mathscr{A}, \tag{21c}$$

where ω_c is a constant. The jump conditions for Eqs. (21) are obtained by inspection from Eqs. (1), (2), and (9):

$$\mathbf{E} + \hat{n}[\varphi] + \hat{n}2\Omega\Phi + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)\Phi = N[\mathscr{A}] + 2\Omega N\mathbf{A} - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\mathbf{A}, \tag{22a}$$

$$\mathbf{B} = \hat{n}\times[\mathscr{A}] - ((\hat{n}\times\mathbf{A})\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\text{curl}\mathbf{A}) - \hat{n}\times\text{grad}(\hat{n}\cdot\mathbf{A}), \tag{22b}$$

$$N[\varphi] + 2\Omega N\Phi - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\Phi = \omega_c^2\hat{n}\cdot[\mathscr{A}] + \omega_c^2\hat{n}\cdot\text{curl}(\hat{n}\times\mathbf{A}), \tag{22c}$$

where $(\varphi,\mathscr{A}) = O(\delta)$ and $(\mathscr{E},\mathscr{B}) = O(\delta')$ and where sans serif symbols represent surface densities for the corresponding capital script symbols. [The surface densities Φ and $\mathbf{A}$ can be generated by the sudden creation of an electric dipole in free space with idealized charge and current density sources of the form $\rho_s = -\delta(x)\delta(y)\delta'(z)H(t)$ and $\mathscr{J}_s = \hat{z}\delta(x)\delta(y)\delta(z)\delta(t)$.]

The equations for the electric and magnetic fields are given by

$$\text{curl}\mathscr{E} = -\frac{\partial\mathscr{B}}{\partial t}, \tag{23a}$$

$$\text{div}\mathscr{B} = 0, \tag{23b}$$

$$\mathscr{J} - \text{curl}\mathscr{H} = -\frac{\partial\mathscr{D}}{\partial t}, \tag{23c}$$

$$\rho = \text{div}\mathscr{D}, \tag{23d}$$

and the corresponding jump conditions are

$$\hat{n}\times[\mathscr{E}] - ((\hat{n}\times\mathbf{E})\cdot\nabla)\hat{n} + \hat{n}(\hat{n}\cdot\text{curl}\mathbf{E}) - \hat{n}\times\text{grad}(\hat{n}\cdot\mathbf{E}) = N[\mathscr{B}] + 2\Omega N\mathbf{B} - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\mathbf{B}, \tag{24a}$$

$$\hat{n}\cdot[\mathscr{B}] + \hat{n}\cdot\text{curl}(\hat{n}\times\mathbf{B}) = 0, \tag{24b}$$

$$\mathbf{K} - \hat{n}\times[\mathscr{H}] + ((\hat{n}\times\mathbf{H})\cdot\nabla)\hat{n} - \hat{n}(\hat{n}\cdot\text{curl}\mathbf{H}) + \hat{n}\times\text{grad}(\hat{n}\cdot\mathbf{H}) = N[\mathscr{D}] + 2\Omega N\mathbf{D} - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\mathbf{D}, \tag{24c}$$

$$\omega = \hat{n}\cdot[\mathscr{D}] + \hat{n}\cdot\text{curl}(\hat{n}\times\mathbf{D}), \tag{24d}$$

where $(\mathscr{E},\mathscr{B},\mathscr{D},\mathscr{H}) = O(\delta)$ and $(\mathscr{J},\rho) = O(\delta')$ and where $\mathbf{K}$ (ω) is the conventional symbol for surface current (charge) density. (The surface densities $\mathbf{E},\mathbf{B},\mathbf{D},\mathbf{H}$ can be generated statically, as will be discussed later in this section; they can also be generated dynamically, as shown in Appendix C.) The equation for conservation of charge is given by

$$\text{div}\mathscr{J} = -\frac{\partial\rho}{\partial t}. \tag{25}$$

and its jump condition is

$$\hat{n}\cdot[\mathscr{J}] + \hat{n}\cdot\text{curl}(\hat{n}\times\mathbf{K}) = N[\rho] + 2\Omega N\omega - \left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\omega, \tag{26}$$

where $(\mathscr{J},\rho) = O(\delta)$. For physical interpretations of the terms in jump conditions (22), (24), and (26), the reader is referred to Sec. II.

In applications of these jump conditions, the orders of the field singularities must be compatible; e.g., jump conditions (22a,b) and (24a,b) are not simultaneously applicable because the singularities in $\mathscr{E},\mathscr{B}$ have different orders unless $\Phi = \mathbf{A} = 0$. Similarly, jump conditions (24c,d) and (26) are not simultaneously applicable unless $\mathbf{H} = \mathbf{D} = 0$.

Agreement with previous formulas: When $\mathbf{K}$ and ω are the only definite surface densities, Eqs. (24) and (26) reduce to the familiar jump conditions.[6] As mentioned before, Stratton[1] considers electric and magnetic surface densities and

TABLE I.

Row	Electrodynamic equations of the form (14a,b)	Corresponding substitutions for jump identity (18a) ζ	γ	g
1	$\left\{ \begin{array}{l} c_c^2 \,\mathrm{div}.\mathscr{A} = -\dfrac{\partial\varphi}{\partial t} \\ -\mathscr{E} - \dfrac{\partial.\mathscr{A}}{\partial t} = \mathrm{grad}\varphi \end{array} \right\}$	φ	$c_c^2\,\mathrm{div}.\mathscr{A}$	$-\mathscr{E} - \dfrac{\partial.\mathscr{A}}{\partial t}$
2	$\mathrm{div}.\mathscr{J} = -\dfrac{\partial\rho}{\partial t}$	ρ	$\mathrm{div}.\mathscr{J}$	$\mathrm{grad}\rho$

derives jump conditions for the static case. We shall now show that under the same conditions Eqs. (24) reduce to his formulas. In static electricity and magnetism, when $(\varphi,.\mathscr{A}) = O(H)$ and $(\mathscr{E},\mathscr{B}) = O(\delta)$, jump conditions (22) reduce to

$$\mathbf{E} = -\hat{n}[\varphi], \tag{27a}$$

$$\mathbf{B} = \hat{n}\times[.\mathscr{A}], \tag{27b}$$

$$\hat{n}\cdot[.\mathscr{A}] = 0. \tag{27c}$$

Equations (27a,b) provide the connection between Stratton's notation and our own.

On pp. 188–92, Stratton[1] treats a fixed interface between nondispersive media with zero net surface charge, but across which φ is discontinuous, as shown:

$$[\varphi] = \tau/\epsilon_0. \tag{28}$$

He determines that the jump conditions are

$$\hat{n}\cdot[\mathscr{E}] = 0, \tag{29a}$$

$$\hat{n}\times[\mathscr{E}] = -\epsilon_0{}^{-1}\hat{n}\times\mathrm{grad}\tau. \tag{29b}$$

Under these conditions, we may set

$$\mathbf{E} = -\hat{n}\tau/\epsilon_0, \quad \mathscr{D} = \epsilon_0\mathscr{E},$$

$$\mathbf{D} = -\hat{n}\tau, \quad \omega = \mathscr{A} = \mathscr{B} = \mathbf{B} = 0,$$

and Eqs. (24a,d) reduce, respectively, to (29a,b).

On pp. 247–50, Stratton[1] considers a fixed interface devoid of surface current, but across which $\mathscr{A}$ is discontinuous due to a surface density of magnetization **M**, as shown:

$$\hat{n}\cdot[.\mathscr{A}] = 0, \tag{30a}$$

$$\hat{n}\times[.\mathscr{A}] = \mu_0\mathbf{M} - \mu_0\hat{n}(\hat{n}\cdot\mathbf{M}). \tag{30b}$$

TABLE II.

Row	Electrodynamic equations of the form (14c–e)	Corresponding substitutions for jump identities (18b–e) v	f	c	ϵ
1	$\left\{ \begin{array}{l} \mathscr{E} + \mathrm{grad}\varphi = -\dfrac{\partial.\mathscr{A}}{\partial t} \\ \mathscr{B} = \mathrm{curl}.\mathscr{A} \\ -\dfrac{1}{c_c^2}\dfrac{\partial\varphi}{\partial t} = \mathrm{div}.\mathscr{A} \end{array} \right\}$	$\mathscr{A}$	$\mathscr{E} + \mathrm{grad}\varphi$	$\mathscr{B}$	$-\dfrac{1}{c_c^2}\dfrac{\partial\varphi}{\partial t}$
2	$\left\{ \begin{array}{l} \mathrm{curl}\mathscr{E} = -\dfrac{\partial\mathscr{B}}{\partial t} \\ 0 = \mathrm{div}.\mathscr{B} \end{array} \right\}$	$\mathscr{B}$	$\mathrm{curl}\mathscr{E}$	$\mathrm{curl}.\mathscr{B}$	0
3	$-\dfrac{\partial\mathscr{B}}{\partial t} = \mathrm{curl}\mathscr{E}$	$\mathscr{E}$	$-\dfrac{\partial\mathscr{E}}{\partial t}$	$-\dfrac{\partial\mathscr{B}}{\partial t}$	$\mathrm{div}\mathscr{E}$
4	$\left\{ \begin{array}{l} \mathscr{J} - \mathrm{curl}.\mathscr{H} = -\dfrac{\partial\mathscr{D}}{\partial t} \\ \rho = \mathrm{div}.\mathscr{D} \end{array} \right\}$	$\mathscr{D}$	$\mathscr{J} - \mathrm{curl}\mathscr{H}$	$\mathrm{curl}\mathscr{D}$	ρ
5	$\mathscr{J} + \dfrac{\partial\mathscr{D}}{\partial t} = \mathrm{curl}\mathscr{H}$	$\mathscr{H}$	$-\dfrac{\partial\mathscr{H}}{\partial t}$	$\mathscr{J} + \dfrac{\partial\mathscr{D}}{\partial t}$	$\mathrm{div}\mathscr{H}$
6	$-\dfrac{\partial\rho}{\partial t} = \mathrm{div}\mathscr{J}$	$\mathscr{J}$	$-\dfrac{\partial\mathscr{J}}{\partial t}$	$\mathrm{curl}\mathscr{J}$	$-\dfrac{\partial\rho}{\partial t}$

He determines that the jump conditions are

$$\hat{n}\cdot[\mathscr{B}] = -\mu_0\hat{n}\cdot\mathrm{curl}(\hat{n}\times\mathbf{M}), \tag{31a}$$

$$\hat{n}\times[\mathscr{B}] = -\mu_0\hat{n}\times\mathrm{grad}(\hat{n}\cdot\mathbf{M}), \tag{31b}$$

where jumps in the volumetric magnetization $\mathscr{M}$ have been deleted. Under these conditions, we may set

$$\mathbf{B} = \mu_0\mathbf{M} - \mu_0\hat{n}(\hat{n}\cdot\mathbf{M}), \quad \mathscr{H} = (\mathscr{B}/\mu_0) - \mathscr{M},$$

$$\mathbf{H} = (\mathbf{B}/\mu_0) - \mathbf{M} = -\hat{n}(\hat{n}\cdot\mathbf{M}), \quad \mathbf{K} = \varphi = \mathscr{D} = \mathbf{D} = 0,$$

and Eqs. (24b,c) reduce, respectively, to (31a,b). Therefore, Eqs. (24) may be regarded as dynamic generalizations of the static jump conditions given by Stratton.[1]

Relations from the jump identities: The jump identities for the electrodynamic equations are obtained by inspection from either Eqs. (14a,b) and (18a), as summarized in Table I, or from Eqs. (14c–e) and (18b–e), as summarized in Table II. In contrast with jump conditions (22), (24), and (26), all of the jump identities are simultaneously applicable because they hold for field singularities of arbitrary order at the interface. All are not independent, however, as shown by the following examples: Jump identity (18c) for Table II, row 1, becomes

$$\begin{aligned}&\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)(\hat{n}\times[\mathscr{A}])\\ &= -\hat{n}\times[\mathscr{E} + \mathrm{grad}\varphi] + N[\mathscr{B} - \hat{n}\mathscr{B}_n]\\ &\quad + 2\Omega N\hat{n}\times[\mathscr{A}] + ((\hat{n}\times[\mathscr{A}])\cdot\nabla)(N\hat{n})\\ &\quad + \hat{n}\times\mathrm{grad}(N\hat{n}\cdot[\mathscr{A}]).\end{aligned} \tag{32}$$

This equation agrees with jump condition (24a); however, Eq. (32) is more general for the reasons given in Sec. IV. The jump identity (18d) for Table II, row 4, becomes

$$\begin{aligned}&\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)(\hat{n}\cdot[\mathscr{D}])\\ &= -\hat{n}\cdot[\mathscr{J} - \mathrm{curl}\mathscr{H}] + N[\rho] + 2\Omega N\hat{n}\cdot[\mathscr{D}]\\ &\quad - \hat{n}\cdot\mathrm{curl}(N\hat{n}\times[\mathscr{D}]);\end{aligned} \tag{33}$$

this equation agrees with jump condition (26), but, again, is more general.

VI. CONCLUSIONS

Jump conditions are derived for a basic set of first order partial differential equations whose fields have infinite, integrable singularities, as well as finite discontinuities, at a moving and deforming interface. These basic formulas are then applied to Maxwell's electrodynamic equations and yield the jump conditions that hold when the electric and magnetic fields have such singularities. These jump conditions are shown to be generalizations of formulas previously derived for double charge layers in electrostatics and for an interface with surface magnetization density in magnetostatics.

The basic formulas are also used to obtain jump conditions for the wave equation and other second order partial differential equations whose fields have finite discontinuities at an interface. Such jump conditions for the electrodynamic potentials are then used to solve elementary boundary value problems that have propagating infinite singularities in the electrodynamic fields.

A similar application to second order vector identities yields a new set of *jump identities*. These identities show that the normal and tangential components of the jump in a vector field are kinematically interdependent; e.g., shock waves and vortex sheets are kinematically linked—a fact that may be significant for shock-slip flows in aerodynamics. The jump identities also indicate the fields that must be measured at an epoch in order to calculate the instantaneous growth-/decay rate of a propagating discontinuity, such as an atmospheric front.

A feature of the derivation is that the field jumps and surface densities are treated as continuous and differentiable functions of three-dimensional space and time that assume physical values on the physical interface. This approach is supported by mathematical induction and continuation and is illustrated by the boundary value problems mentioned earlier and by other elementary examples. This approach is simpler and more general than previous treatments that define jumps and surface densities only on the physical interface because the brackets (jumps) now commute with *all* derivatives, instead of with only the tangential and displacement derivatives.

APPENDIX A: MOVING SURFACES

Take the interface S to be defined by

$$\phi(x,y,z,t) = 0, \tag{A1}$$

where ϕ is continuous and differentiable. This surface is one of a manifold of moving and deforming surfaces given by $\phi(x,y,z,t) = \text{const}$. The unit normal $\hat{n}$, speed of displacement N, and mean curvature Ω are given by[2,8]

$$\hat{n} = |\mathrm{grad}\phi|^{-1}\mathrm{grad}\phi, \tag{A2}$$

$$N = -|\mathrm{grad}\phi|^{-1}\frac{\partial\phi}{\partial t}, \tag{A3}$$

$$\Omega = -\tfrac{1}{2}\mathrm{div}\hat{n}. \tag{A4}$$

These formulas conform to the sign convention of $N > 0$ for movement in the sense of $\hat{n}$ and $\Omega > 0$ for concavity in the sense of $\hat{n}$. As a consequence of Eqs. (A2),(A3), we have the identities

$$\left(\frac{\partial}{\partial t} + N\frac{\partial}{\partial n}\right)\hat{n} + \left(\nabla - \hat{n}\frac{\partial}{\partial n}\right)N = 0, \tag{A5}$$

$$\mathrm{curl}\hat{n} = \hat{n}\times\frac{\partial\hat{n}}{\partial n}, \tag{A6}$$

$$\frac{\partial\hat{n}}{\partial n} = -\hat{n}\times\mathrm{curl}\hat{n}, \tag{A7}$$

$$\hat{n}\cdot\mathrm{curl}\hat{n} = \hat{n}\cdot\frac{\partial\hat{n}}{\partial n} = \hat{n}\cdot\frac{\partial\hat{n}}{\partial t} = 0. \tag{A8}$$

APPENDIX B: SIMPLE CHECK OF THE BASIC JUMP CONDITIONS

Equations (9) are amenable to simple checks. For example, to check (9a) in spherical coordinates (r,ϑ,ψ), we may take

$$\theta = f(r,\vartheta,\psi,t) + g(r,\vartheta,\psi,t)H(r - ct) + h(r,\vartheta,\psi,t)\delta(r - ct),$$

where c is a constant, where the interface is the expanding sphere $r - ct = 0$, and where f,g,h are continuous and differentiable functions. It follows that

$$[\theta] = g$$

and by (6a) and (A1)–(A4)

$$\Theta = h, \quad N = c,$$
$$\hat{n} = \hat{r}, \quad \Omega = -r^{-1}.$$

Then, by (1a), we have

$$\lambda = -\frac{\partial f}{\partial t} - \frac{\partial g}{\partial t} H(r - ct) + \left(cg - \frac{\partial h}{\partial t}\right)\delta(r - ct) + ch\delta'(r - ct),$$

whence

$$[\lambda] = -\frac{\partial g}{\partial t}$$

and by (6b)

$$\Lambda = cg - \frac{2ch}{r} - \left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial r}\right)h.$$

Note that $[\theta], \Theta, \hat{n}, N, \Omega, \Lambda$ are continuous and differentiable functions, in agreement with Sec. III. Substitution of these quantities shows that (9a) is indeed satisfied. The remainder of Eqs. (9) have similar checks.

APPENDIX C: DYNAMIC BOUNDARY VALUE PROBLEMS WITH ELECTRIC AND MAGNETIC SURFACE DENSITIES

Formulation: We seek solutions of Maxwell's equations in free space where the electric and magnetic fields have infinite, integrable singularities on the expanding spherical surface $r - ct = 0$ (the light sphere) and vanish outside it. The corresponding potentials φ and $\mathscr{A} = \hat{z}\mathscr{A}_z$ have finite discontinuities on the light sphere and vanish outside it. The differential equations are given by

$$\nabla^2 \begin{Bmatrix} \varphi \\ \mathscr{A}_z \end{Bmatrix} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\begin{Bmatrix} \varphi \\ \mathscr{A}_z \end{Bmatrix} = 0, \tag{C1a}$$

$$-\frac{\partial \varphi}{\partial t} = c^2 \frac{\partial \mathscr{A}_z}{\partial z}, \tag{C1b}$$

the surface relations are given by [Eqs. (A1)–(A8)]

$$\phi = r - ct, \quad N = c,$$
$$n_i = \frac{x_i}{r}, \quad \frac{\partial n_i}{\partial x_j} = \frac{1}{r}(\delta_{ij} - n_i n_j),$$
$$\Omega = -\frac{1}{r}, \quad \frac{\partial \hat{n}}{\partial t} = \frac{\partial \hat{n}}{\partial n} = \mathrm{curl}\hat{n} = 0, \tag{C2}$$

and the jump conditions by [Eqs. (20e),(22c),(C2)]

$$\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial r}\right)\begin{Bmatrix} [\varphi] \\ [\mathscr{A}_z] \end{Bmatrix} + \frac{c}{r}\begin{Bmatrix} [\varphi] \\ [\mathscr{A}_z] \end{Bmatrix} = 0, \tag{C3a}$$

$$[\varphi] = \frac{cz}{r}[\mathscr{A}_z]. \tag{C3b}$$

Sudden deposition of a point charge: The jump conditions (C3) are satisfied by $[\varphi] = -1/r$ and $[\mathscr{A}_z] = -1/cz$, and since, for $r > ct$, $\varphi = \mathscr{A}_z = 0$, we try, unsuccessfully, to set $\varphi = 1/r$ and $\mathscr{A}_z = 1/cz$ for $r < ct$, which violates Eqs. (C1). However, on $r - ct = 0$, we may replace z with $\zeta \equiv (c^2t^2 - x^2 - y^2)^{1/2}$, and this suggests that we choose

$$[\varphi] = -1/r, \quad [\mathscr{A}_z] = -1/c\zeta,$$
$$\varphi = 1/r, \quad \mathscr{A}_z = 1/c\zeta \quad (r < ct),$$
$$\varphi = \mathscr{A}_z = 0 \quad (r > ct), \tag{C4}$$

where a constant common factor of dimension meter-volts is understood. These fields satisfy both the jump conditions and the differential equations. (Physically interpreted, solutions (C4) are the fields excited by the sudden deposition of a point charge at the end of a semi-infinite wire with idealized charge and current density sources of the form $\rho_s = 4\pi\epsilon_0\delta(x)\delta(y)\delta(z)H(t)$ and $\mathscr{J}_s = -\hat{z}4\pi\epsilon_0\delta(x)\delta(y)H(z)\delta(t)$.) Note that, except for isolated singularities, the jumps $[\varphi]$ and $[\mathscr{A}_z]$ are continuous and differentiable throughout $(\mathbf{r},t)$, in agreement with Sec. III.

The jumps in the electric and magnetic fields are given by [Eqs. (12),(21),(C4)]

$$[\mathscr{E}] = -\mathrm{grad}[\varphi] - \hat{z}\frac{\partial}{\partial t}[\mathscr{A}_z] = -\left(\frac{x}{r^3}, \frac{y}{r^3}, \frac{z}{r^3} + \frac{ct}{\zeta^3}\right),$$
$$[\mathscr{B}] = \mathrm{curl}(\hat{z}[\mathscr{A}_z]) = (-y, x, 0)/c\zeta^3,$$

and the surface densities are given by [Eqs. (22),(C4)]

$$\mathbf{E} = -\hat{n}[\varphi] + \hat{z}c[\mathscr{A}_z] = \left(\frac{x}{r^2}, \frac{y}{r^2}, \frac{z}{r^2} - \frac{1}{\zeta}\right)$$
$$\mathbf{B} = \hat{n} \times \hat{z}[\mathscr{A}_z] = (-y, x, 0)/c\zeta r,$$

whence

$$\mathbf{E} = -c\hat{n} \times \mathbf{B}, \quad \mathbf{B} = c^{-1}\hat{n} \times \mathbf{E},$$
$$\hat{n} \cdot \mathbf{E} = \hat{n} \cdot \mathbf{B} = 0,$$
$$\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial r}\right)\begin{Bmatrix} \mathbf{E} \\ \mathbf{B} \end{Bmatrix} + \frac{c}{r}\begin{Bmatrix} \mathbf{E} \\ \mathbf{B} \end{Bmatrix} = 0,$$
$$\Omega\mathbf{E} + (\mathbf{E} \cdot \nabla)\hat{n} = \Omega\mathbf{B} + (\mathbf{B} \cdot \nabla)\hat{n} = 0,$$
$$\hat{n} \cdot \mathrm{curl}\mathbf{E} = 0, \quad \hat{n} \cdot \mathrm{curl}\mathbf{B} = \frac{1}{c}\left(\frac{1}{r^2} + \frac{1}{\zeta^2}\right)$$

on $r - ct = 0$. Therefore, (24a,b) reduce to the familiar jump conditions

$$\hat{n} \times [\mathscr{E}] - c[\mathscr{B}] = 0, \quad \hat{n} \cdot [\mathscr{B}] = 0$$

on $r - ct = 0$; but (24c,d) with $\mathbf{K} = \omega = 0$ retain surface density terms, as shown:

$$\hat{n} \times [\mathscr{B}] + c^{-1}[\mathscr{E}] + \hat{n}(\hat{n} \cdot \mathrm{curl}\mathbf{B}) = 0,$$
$$\hat{n} \cdot [\mathscr{E}] + c\hat{n} \cdot \mathrm{curl}\mathbf{B} = 0.$$

The normal electric component $\mathscr{E}_n$ is discontinuous on the light sphere despite the absence of surface charge, and this discontinuity is maintained by the surface density $\mathbf{B}$. We conclude that the surface densities $\mathbf{E}$ and $\mathbf{B}$, as included in our jump conditions, are essential to the physics of this elementary problem.

Gradual charging of a point electric dipole: Another solution of (C1),(C3), similarly obtained, is given by

$$[\mathscr{A}_z] = -1/r, \quad [\varphi] = -c^2zt/r^3,$$
$$\mathscr{A}_z = 1/r, \quad \varphi = c^2zt/r^3 \quad (r < ct),$$
$$\mathscr{A}_z = \varphi = 0 \quad (r > ct),$$

where a constant common factor of dimension volt-seconds is understood. [Physically interpreted, this solution gives the fields generated by the gradual charging of a point electric dipole with charge and current density sources of the form $\rho_s = -4\pi\epsilon_0 c^2 \delta(x)\delta(y)\delta'(z)tH(t)$ and $\mathscr{J}_s = \hat{z}4\pi\epsilon_0 c^2 \delta(x)\delta(y)\delta(z)H(t)$.] The jumps $[\mathscr{E}]$ and $[\mathscr{B}]$, the surface densities **E** and **B**, and their relationships may be obtained as in the preceding example.

[1] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).
[2] C. Truesdell and R. Toupin (with appendix by J. L. Ericksen), Handb. Phys. **3-1**, 491–529 (1960).
[3] R. K. Luneburg (with supplementary notes by M. Herzberger), *Mathematical Theory of Optics* (Univ. of California Press, Berkeley, 1964).
[4] R. C. Costen, NASA TN D-2618 (1965).
[5] T. Y. Thomas, Math. Mag. **22**, 169–89 (1949).
[6] R. C. Costen and D. Adamson, Proc. IEEE **53**, 1181–96 (1965).
[7] H. B. Phillips, *Tensor Analysis* (Wiley, New York, 1957), pp. 72–4.
[8] A. J. McConnell, *Applications of Tensor Analysis* (Dover, New York, 1957), p. 206.

Distributionlike representations of $*$-algebras

Françoise Mathot
Institut de Physique Théorique, Université Catholique de Louvain, Chemin du Cyclotron, 2, B-1348 Louvain-la-Neuve, Belgium

(Received 22 October 1980; accepted for publication 23 January 1981)

Non-self-adjoint representations of $*$-algebras in a Hilbert space give rise by an extension and transposition procedure to representations in larger spaces, such as distribution spaces. Those new representations provide examples of operators of nested Hilbert spaces which would be very singular operators when considered in the Hilbert space only.

PACS numbers: 02.30.Tb, 02.10. – v

INTRODUCTION

In this paper we exhibit a construction of representations of $*$-algebras by "singular" operators in a Hilbert space. By singular we mean that their domain can be nondense or even reduced to zero. Nevertheless from a mathematical point of view those operators are well-defined objects, namely operators of a nested Hilbert space[1–3] constructed in a canonical way around the Hilbert space. We get in this way representations of $*$-algebras by operators acting in spaces larger than the usual Hilbert space, analogous to "distribution spaces."

Beginning with a state on a $*$-algebra $\mathcal{A}$, we get by the usual GNS construction[4] a Hilbert space $\mathscr{H}$ with a dense domain $\mathscr{D}$, a cyclic vector Ω, and a representation $\Pi(\mathcal{A})$ by continuous operators from $\mathscr{D}$ into $\mathscr{D}$ (with the graph topology). If we want to construct a representation of $\mathcal{A}$ in a space larger than $\mathscr{H}$, the first thing that comes to mind is to consider the transposed representation Π' defined on the dual space $\mathscr{D}'$ by $\langle \Pi'(A) f | g \rangle = \langle f | \Pi(A) g \rangle$, where $A \in \mathcal{A}$, $f \in \mathscr{D}'$, $g \in \mathscr{D}$ ($\langle \cdot | \cdot \rangle$ is the dual pairing between $\mathscr{D}$ and $\mathscr{D}'$ which extends the scalar product of $\mathscr{H}$). However, by the GNS construction we get a Hermitian representation of $\mathcal{A}$ (i.e., $\Pi(\mathrm{A}^*) \subset \Pi(\mathrm{A})^* \ \forall \mathrm{A} \in \mathcal{A}$, where the last * denotes the Hilbertian adjoint), which implies that the new representation Π' is just an extension of Π to $\mathscr{D}'$ [$\Pi'(A)$ restricted to $\mathscr{D}$ coincides with $\Pi(\mathrm{A}^*)$].

So as operators in $\mathscr{H}$ the $\Pi'(A)$'s are defined on a dense domain ($\mathscr{D}$) and can be made continuous from $\mathscr{D}$ into $\mathscr{H}$. This is a well-known type of unbounded operators.

A more interesting case occurs if the representation Π is not self-adjoint (Sec. III). We can then construct the adjoint representation Π^*[4] which is an extension of Π on a larger domain $\mathscr{D}^*$ still dense in $\mathscr{H}$, by continuous operators of $\mathscr{D}^*$. But this time Π^* need not be a Hermitian representation (as the extension of a symmetric operator need not to be symmetric in general).

Considering then the transposed representation $\Pi^{*\prime}$, which acts in the dual space $\mathscr{D}^{*\prime}$ (Sec. IV); we get a new representation which is not an extension of Π^*. The operators $\Pi^{*\prime}(A)$ are continuous maps from $\mathscr{D}^{*\prime}$ into itself, but they are not necessarily defined as operators in $\mathscr{H}$. Although we can apply them to elements of $\mathscr{D}^*$, they "jump" over $\mathscr{H}$ and the resulting vector will be in $\mathscr{D}^{*\prime}$.

On the other hand, at each stage of the construction of those representations, we emphasize the fact that there is a natural nested Hilbert space associated to the GNS construction, constructed around the GNS Hilbert space. The various representations obtained are then described as operators in that nested Hilbert space, which provides a better way of controlling their continuity properties and the maximal domain to which they can be extended. One of the aims of introducing nested Hilbert spaces,[1–3] was that they allow one to handle unbounded operators and even more singular objects by considering them as bounded operators between different Hilbert spaces. Very singular operators such as quantized fields at a point have been successfully treated in this framework.[5]

The representation $\Pi^{*\prime}$ we construct in this paper provides another example of well-defined operators in the nested Hilbert space which are very "bad" when viewed as operators in the Hilbert space only.

In the last section we discuss in detail the simple case where the algebra is generated by one symmetric, non-self-adjoint operator. In this example the domains $\mathscr{D}$, $\mathscr{D}^*$ and the representations Π, Π^*, $\Pi^{*\prime}$ can be computed explicitly.

Further applications, such as the field algebra and the universal enveloping algebra of a Lie algebra, are briefly discussed.

I. A NATURAL STRUCTURE ASSOCIATED TO A $*$-ALGEBRA

(1.1) Let $\mathcal{A}$ be a $*$-algebra and ω a state of $\mathcal{A}$. By the GNS construction,[4] there exists a Hilbert space $\mathscr{H}$, a dense domain $\mathscr{D}$ in $\mathscr{H}$, a closed representation $\Pi(\mathcal{A})$ of $\mathcal{A}$ by (closed) operators in $\mathscr{H}$, all defined at least on $\mathscr{D}$, and a strongly cyclic vector $\Omega \in \mathscr{H}$ such that

$$\omega(A) = (\Omega, \Pi(A)\Omega), \tag{1}$$

for every $A \in \mathcal{A}$. By "closed representation" we mean that the domain $\mathscr{D}$ is complete with respect to the topology defined by the set of graph norms:

$$\|f\|_A^2 = \|f\|^2 + \|\Pi(A)f\|^2 = \|[1 + \Pi(A)^*\Pi(A)]^{1/2} f\|^2, \tag{2}$$

where $f \in \mathscr{D}$ and A runs over $\mathcal{A}$.

Taking each one of those norms into consideration separately, we can complete $\mathscr{D}$ with respect to it and get a Hilbert space $\mathscr{H}_A$ which coincides with the domain of the closed operator $\Pi(A)$. We have then

$$\mathscr{D} = \bigcap_{A \in \mathcal{A}} \mathscr{H}_A. \tag{3}$$

 0022-2488/81/071386-04$1.00

Each of the $\mathscr{H}_A$'s is continuously embedded in $\mathscr{H}$ and is also unitarily isomorphic to $\mathscr{H}$ by the unitary operators $[1 + \Pi(A)^*\Pi(A)]^{\pm 1/2}$. Moreover, the set $\{\mathscr{H}_A | A \in \mathscr{A}\}$ possesses a lattice structure. Given two spaces $\mathscr{H}_A$ and $\mathscr{H}_B$ there always exists $C \in \mathscr{A}$ such that $\mathscr{H}_C \subseteq \mathscr{H}_A \cap \mathscr{H}_B$. Assuming that $\mathscr{A}$ has an identity element, the actual C is $1 + A^*A + B^*B$.

(1.2) We can also consider the completion of $\mathscr{D}$ with respect to the following norms:

$$\|f\|_{\underline{A}} \equiv \|[1 + \Pi(A)^*\Pi(A)]^{-1/2} f\|, \tag{4}$$

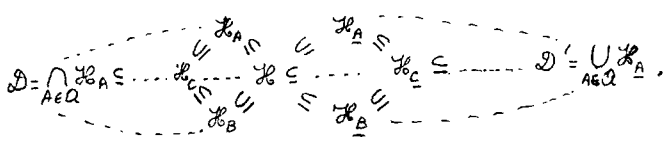
for all $A \in \mathscr{A}$.

In this way, we get Hilbert spaces $\mathscr{H}_{\underline{A}}$ larger than $\mathscr{H}$ and each pair $(\mathscr{H}_A, \mathscr{H}_{\underline{A}})$ is a dual pair with respect to the scalar product of $\mathscr{H}$: given $f \in \mathscr{H}_A$ and $g \in \mathscr{H}_{\underline{A}}$ we can define an inner product

$$\langle f | g \rangle \equiv ([1 + \Pi(A)^*\Pi(A)]^{1/2} f, [1 + \Pi(A)^*\Pi(A)]^{-1/2} g), \tag{5}$$

which extends the scalar product of $\mathscr{H}$ to the various pairs and also to the dual pair $(\mathscr{D}, \mathscr{D}')$ where $\mathscr{D}' = \bigcup_{A \in \mathscr{A}} \mathscr{H}_{\underline{A}}$.

Finally, with each $*$-algebra provided with a state ω, we have associated in a natural way the following structure:



(6)

This type of structure, called a nested Hilbert space or a partial inner product space, has been studied in Refs. 1–3.

II. REPRESENTATION OF THE ALGEBRA

(2.1) The representation $\Pi(\mathscr{A})$ we get by the GNS construction is a representation by continuous operators on $\mathscr{D}$ with respect to the set of graph norms (2). Actually, given $A \in \mathscr{A}$, for every $B \in \mathscr{A}$, $\exists C \in \mathscr{A}$ such that for every $f \in \mathscr{D}$,

$$\|\Pi(A) f\|_B \leqslant k \|f\|_C, \tag{7}$$

with some constant k.

It is easy to check that a suitable choice of C is $C = 1 + A^*A + (BA)^*BA$. This relation (7) shows how the representation operators act in the structure (6).

Indeed (7) says that $\Pi(\mathscr{A})$ can be extended to a bounded operator between $\mathscr{H}_C$ and $\mathscr{H}_B$. So for every $\mathscr{H}_B$ there always exists a $C \in \mathscr{A}$ such that $\Pi(A)$ is bounded from $\mathscr{H}_C$ into $\mathscr{H}_B$. In that way we identify the unbounded operators $\Pi(A)$ in $\mathscr{H}$ with a family of bounded operators between pairs $(\mathscr{H}_C, \mathscr{H}_B)$. These bounded operators will be denoted by $\Pi(A)_{BC}$. They obviously satisfy the coherence relation

$$\Pi(A)_{DF} = 1_{DB} \Pi(A)_{BC} 1_{CF}, \tag{8}$$

if $\mathscr{H}_B \subseteq \mathscr{H}_D, \mathscr{H}_F \subseteq \mathscr{H}_C$ and where 1_{DB} represents the injection map from $\mathscr{H}_B$ into $\mathscr{H}_D$.

(2.2) Since we have begun with a $*$-algebra $\mathscr{A}$, the GNS construction gives us a $*$-representation, or Hermitian representation, i.e., $\Pi(A^*) \subseteq \Pi(A)^* \forall A \in \mathscr{A}$ (where the last star denotes the adjoint in the Hilbertian sense). [If $A = A^*$, $\Pi(A)$ is thus a symmetric operator in $\mathscr{H}$ but not necessarily self-adjoint.] This fact allows us to derive more continuity properties for the representation.

Consider a pair $(\mathscr{H}_C, \mathscr{H}_B)$ between the elements of which $\Pi(A)$ is bounded. Let $f \in \mathscr{H}_B$ and $g \in \mathscr{H}_C$. The inner product (5) is well defined between $f \in \mathscr{H}_B$ and $\Pi(A)_{BC} g_c \in \mathscr{H}_C$. Denoting by $\Pi(A)^t$ the adjoint of $\Pi(A)$ with respect to the inner product (5), i.e.,

$$\langle f | \Pi(A) g \rangle = \langle \Pi(A)^t f | g \rangle, \tag{9}$$

we have by duality that $\Pi(A)^t$ is a bounded operator from $\mathscr{H}_B$ into $\mathscr{H}_C$. (Note: This adjoint operation t is denoted by $*$ in Refs. 1, 2, but for us $*$ means the involution in $\mathscr{A}$ and also the Hilbertian adjoint of an operator.)

Restricting (9) to f and g in $\mathscr{D}$, we see that $\Pi(A)^t \restriction_{\mathscr{D}} = \Pi(A^*)$, which means that $\forall B, \exists C$ such that $\Pi(A^*)$ can be extended to a bounded operator from $\mathscr{H}_B$ into $\mathscr{H}_C$. But $A^* \in \mathscr{A}$, $\Pi(A^*)$ possesses already the boundedness property given by (7), $\forall B \in \mathscr{A}$, $\exists D \in \mathscr{A}$ such that $\Pi(A^*)$ is bounded from $\mathscr{H}_D$ into $\mathscr{H}_B$.

Interchanging the role of A and A^*, since $*$ is an involution, we get for $\Pi(A)$ the following property: Choosing any Hilbert space in the set $\{\mathscr{H}_A, \mathscr{H}_{\underline{A}}\}$, we can always find another one such that $\Pi(A)$ is bounded from this one into the chosen one and we can find a third one such that $\Pi(A)$ is bounded from the chosen one into the third one. Because of that, $\Pi(A)$ can be applied to any vector in $\mathscr{D}'$ and is of the type of operator called "good operators" in Ref. 7, i.e., operators which map continuously $\mathscr{D}$ into itself (considered with the projective limit topology) and which map continuously $\mathscr{D}'$ into itself (with the inductive limit topology).

In other words, if we denote by $I = \{A, \underline{A} | A \in \mathscr{A}\}$ the set of elements which indexes the lattice of Hilbert spaces and if we define[1] $J(\Pi(A)) \subset I \times I$ as the set of pairs of indices such that $\Pi(A)$ is bounded between the corresponding pairs of Hilbert spaces, the "good operators" have the property that $\mathrm{pr}_1 J(\Pi(A)) = I$ and $\mathrm{pr}_2 J(\Pi(A)) = I$ where pr_1 and pr_2 denote the projection on the first (respectively the second) variable in the Cartesian product $I \times I$.

III. NON-SELF-ADJOINT REPRESENTATIONS

(3.1) Let us go back to $\Pi(\mathscr{A})$ defined on $\mathscr{D}$. We can associate with it another representation Π^* called the adjoint representation of Π, defined on the following domain:

$$\mathscr{D}^* = \bigcap_{A \in \mathscr{A}} D(\Pi(A)^*), \tag{10}$$

by

$$\Pi^*(A) f \equiv \Pi(A^*)^* f, \quad \forall f \in \mathscr{D}^*. \tag{11}$$

Since $\mathscr{D}^* \supseteq \mathscr{D}$ and $\Pi^* \restriction_{\mathscr{D}}$ coincide with Π (because Π is a $*$-representation), we have Π^* is an extension of Π. Some representations called self-adjoint representations[4] are such that $\Pi = \Pi^*$. Those representations come from a particular type of states called Riesz states, characterized in Ref. 8. Examples are also found in Ref. 9. However, from now on we shall assume Π to be non-self-adjoint, which is generally the case, and see what can be said about Π^*.

(3.2) First, Π^* is a closed representation so that we may consider on $\mathscr{D}^*$ the graph norms ($A\in\mathscr{A}$)

$$^*\|g\|_A^2 = \|g\|^2 + \|\Pi^*(A)g\|^2, \tag{12}$$

which coincide with (2) if $g\in\mathscr{D}$.

$\mathscr{D}^*$ is complete with respect to the topology defined by this set of norms when A varies in $\mathscr{A}$ and $\mathscr{D}$ appears as a closed subspace of $\mathscr{D}^*$ in that topology. We can repeat the construction given in Sec. I, completing $\mathscr{D}^*$ with respect to each graph-norm and getting a lattice of Hilbert spaces $\{\mathscr{K}_A | A\in\mathscr{A}\}$. For each $A\in\mathscr{A}$, $\mathscr{H}_A$ appears as a closed subspace of $\mathscr{K}_A$. Taking the dual spaces $\mathscr{K}_{\underline{A}}$, $A\in\mathscr{A}$, we get again a structure similar to (6) with dual pairs with respect to the scalar product of $\mathscr{H}$. We have indeed $\mathscr{H}\equiv\mathscr{H}_1 = \mathscr{K}_1 = \mathscr{K}$, because $\mathscr{D}\subset\mathscr{D}^*\subset\mathscr{H}$ and $\mathscr{H}$ is the completion of $\mathscr{D}^*$ in the initial Hilbert space norm. Since $\mathscr{H}_A\subset\mathscr{K}_A$, we have for the dual spaces $\mathscr{K}_{\underline{A}}\subseteq\mathscr{H}_{\underline{A}}$ and finally $\mathscr{D}^{*\prime}\subset\mathscr{D}'$ as a closed subspace.

So we get that Π^* is a representation acting in the space $\mathscr{D}^{*\prime}$ which is another partial inner product space but constructed around the same initial Hilbert space $\mathscr{H}$.

(3.3) In that structure, Π^* possesses continuity properties similar to (7), i.e., $\forall B\in\mathscr{A}$, $\exists C\in\mathscr{A}$ such that, $\forall f\in\mathscr{D}^*$,

$$^*\|\Pi^*(A)f\|_B \leqslant k\,^*\|f\|_C, \tag{13}$$

which means that $\Pi^*(A)$ can be extended to a bounded operator between $\mathscr{K}_C$ and $\mathscr{K}_B$. [Let us note that the pairs (C,B) such that Π^* is bounded from $\mathscr{K}_C$ into $\mathscr{K}_B$ are exactly the ones such that Π was bounded from $\mathscr{H}_C$ into $\mathscr{H}_B$.]

Nevertheless, even if Π is a $*$-representation [$\Pi(A^*)\subseteq\Pi(A)^*$], Π^* may fail to be a $*$-representation[4] just as the extension of a symmetric operator may fail to be symmetric itself. In that case we cannot derive further continuity properties (between $\mathscr{K}_B$ and $\mathscr{K}_C$) as we did for Π by duality with respect to the scalar product of $\mathscr{H}$. As operator in $\mathscr{D}^{*\prime}$, the domain of $\Pi^*(A)$ is $\mathscr{K}_A$ (dense in $\mathscr{H}$). In terms of indices we have $\mathrm{pr}_2 J(\Pi^*(A)) = I$, but not the same relation with pr_1, i.e., $\Pi^*(A)$ is *not* a good operator.

IV. A SINGULAR REPRESENTATION OF $\mathscr{A}$

In the case where Π^* is not a $*$-representation [i.e., $\Pi^*(A^*)\not\subseteq\Pi^*(A)^*$ at least for some $A\in\mathscr{A}$], we can consider the set of operators $\{\Pi^*(A)^t | A\in\mathscr{A}\}$ where t means the transposed operator with respect to the scalar product (5), as defined in (9). $\Pi^*(A)^t$ is not an extension of $\Pi^*(A^*)$ as it was the case for the representation Π, but, putting $\Pi^{*t}(A)\equiv\Pi^*(A)^t, \forall A\in\mathscr{A}$, we get an antirepresentation Π^{*t} of $\mathscr{A}$ by operators of $\mathscr{D}^{*\prime}$. Indeed, since $\Pi^{*t}(A)$ is the transposed operator of $\Pi^*(A)$, we get that $\forall B\in\mathscr{A}$, $\exists C\in\mathscr{A}$ such that $\Pi^{*t}(A)$ is bounded from $\mathscr{K}_{\underline{B}}$ into $\mathscr{K}_{\underline{C}}$. Since this is true for every $\mathscr{K}_{\underline{B}}$, $\Pi^{*t}(A)$ is everywhere defined on $\mathscr{D}^{*\prime}$.

Let us note that it can happen that $\Pi^{*t}(A)$ is not defined as an operator in $\mathscr{H}$ because every element f of $\mathscr{H}$ is mapped by $\Pi^{*t}(A)$ into an element of $\mathscr{K}_{\underline{C}}$ for some C, i.e., in a larger space. So the domain of $\Pi^{*t}(A)$ in $\mathscr{H}$ can be very small or even reduced to $\{0\}$. Nevertheless, as an operator in the space $\mathscr{D}^{*\prime}$, $\Pi^{*t}(A)$ is a suitable mathematical object, everywhere defined. Indeed this time we have $\mathrm{pr}_1 J(\Pi^{*t}(A)) = I$.

V. EXAMPLES

(5.1) Let $\mathscr{A}$ be the $*$-algebra generated by one single Hermitian generator A, and let $\Pi(\mathscr{A})$ be its representation by a differential operator in the space $\mathscr{H} = L^2([0,1])$ [we could do similar things with $L^2([0,\infty)]$. Consider the domain

$$\mathscr{D} = \{f\in L^2([0,1]) | f^{(n)}\in L^2([0,1]),$$
$$f^{(n)}(0) = f^{(n)}(1) = 0, \quad n = 1,2,\cdots\}, \tag{14}$$

and $\Pi(A) = (id/dx)$ restricted to $\mathscr{D}$. $\mathscr{D}$ can be considered as the intersection of a countable set of Hilbert spaces $\mathscr{H}_n$ defined by

$$\mathscr{H}_n = [f\in L^2([0,1]) | f^{(j)}\in L^2([0,1]),$$
$$f^{(j)}(0) = f^{(j)}(1) = 0, j = 1,2,\cdots,n], \tag{15}$$

which can be seen as the domain of $\Pi(A^n)$. As an operator in $\mathscr{H}$, $\Pi(A)$ is a closed symmetric operator (with domain $\mathscr{H}_1$).

(5.2) According to the criterion given in Ref. 4 for algebras generated by a single operator, the representation Π will be self-adjoint iff $\Pi(A)$ is essentially self-adjoint. This is not the case here because $\Pi(A)^*$ is a proper extension of $\Pi(A)$. Indeed $D(\Pi(A)^*) = \{f\in L^2([0,1]) | f^{(1)}\in L^2([0,1])\}$ is strictly larger than $\mathscr{H}_1$.

We can define the adjoint representation

$$\Pi^*(A) = \Pi(A)^* = i\frac{d}{dx} \text{ on the domain}$$
$$\mathscr{D}^* = \{f\in L^2([0,1]), f^{(n)}\in L^2([0,1]), n = 1,2\cdots\}, \tag{16}$$

which is strictly larger than $\mathscr{D}$.

We have $\mathscr{D}^* = \cap_{n=1}^{\infty}\mathscr{K}_n$,

where

$$\mathscr{K}_n = \{f\in L^2([0,1]) | f^{(j)}\in L^2([0,1]), j = 1,2,\cdots,n\} \tag{17}$$

is the domain of $\Pi^*(A^n)$. For every $n\geqslant 1$, $\Pi^*(A)$ is bounded from $\mathscr{K}_{n+1}$ into $\mathscr{K}_n$ and $\Pi^*(A^j)$ is bounded from $\mathscr{K}_{n+j}$ into $\mathscr{K}_j$. As in Sec. III we have a non-self-adjoint representation Π with a strict extension Π^*. Moreover Π^* is not a $*$-representation [i.e., $\Pi^*(A^*)\not\subset\Pi^*(A)^*$] because $\Pi(A)^*$ is not a symmetric operator; $\Pi^*(A^*)^* = \Pi^*(A)^* = \Pi(A)^{**} = \Pi(A$ defined on $\mathscr{H}_1$ again. So we have $\Pi = \Pi^{**}\subset\Pi^*$, the first two defined on $\mathscr{D}$ and the third one on $\mathscr{D}^*$.

The next step is then to construct the antirepresentation Π^{*t}.

(5.3) For any $n\geqslant 0$, the dual space of $\mathscr{K}_n$ with respect to $\mathscr{H}$ is

$$\mathscr{K}_{-n} = \left\{F = \sum_{i\leqslant n}\frac{d^i G_i}{dx^i} | G_i\in L^2([0,1])\right\}, \tag{18}$$

and the dual space of $\mathscr{D}^*$.

$$\mathscr{D}^{*\prime} = \left\{F = \sum_{0\leqslant i\leqslant N}\frac{d^i G_i}{dx^i} | G_i\in L^2([0,1]), N<\infty\right\}. \tag{19}$$

We then construct $\Pi^{*t}(A) = \Pi^*(A)^t$ by duality. For any $f\in\mathscr{K}_{n+1}$ and any $F\in\mathscr{K}_{-n}$ we have

$$\langle\Pi^{*t}(A)F | f\rangle = \langle F | \Pi^*(A)f\rangle. \tag{20}$$

The rhs is well defined since, for $f\in\mathscr{K}_{n+1}$, $\Pi^*(A)f\in\mathscr{K}_n$. We already see that, for every $n\geqslant 0$, $\Pi^{*t}(A)$ will be bounded from $\mathscr{K}_{-n}$ into $\mathscr{K}_{-(n+1)}$. Similarly $\Pi^{*t}(A^j)$ is bounded

from $\mathscr{K}_{-n}$ into $\mathscr{K}_{-(n+j)}$. Moreover, since we have a continuous injection from every $\mathscr{K}_n$ into $\mathscr{H}$, $\Pi^{*t}(A)$ is bounded from any $\mathscr{K}_n$ into $\mathscr{K}_{-1}$. In particular, as an operator in $\mathscr{H}$, $\Pi^{*t}(A)$ has its domain reduced to $\{0\}$.

If we choose f and F in $\mathscr{D}^*$, we can compute the explicit form of $\Pi^{*t}(A)$ on $\mathscr{D}^*$:

$$\Pi^{*t}(A) = i\frac{d}{dx} + i\delta(x) - i\delta(x-1), \tag{21}$$

which extends by continuity between pairs $(\mathscr{K}_{-n}, \mathscr{K}_{-(n+1)})$. In particular, we see that every vector in a space smaller than $\mathscr{K}_{-1}$ is mapped automatically into $\mathscr{K}_{-1}$ (jumping over $\mathscr{H}$) because of the delta-function terms.

Those operators, which are very singular objects when regarded in $\mathscr{H}$, are now well defined as operators on the larger space $\mathscr{D}^{*\prime}$, and their continuity properties can be controlled. Similary an explicit form of $\Pi^{*t}(A^m)$ on $\mathscr{D}^*$ would involve, besides the term $i(d/dx)$, other terms consisting of δ and derivatives $\delta^{(j)}$ of order $\leqslant m$. This operator can be extended to a bounded operator between pairs $(\mathscr{K}_{-n}, \mathscr{K}_{-(n+m)})$ and every vector in a space smaller than $\mathscr{K}_{-m}$ is mapped into $\mathscr{K}_{-m}$.

(5.4) There are other examples where we can expect to get distributionlike representations. For instance, if $\mathscr{A} = \mathscr{S}$, the Borchers algebra[10] or field algebra (or tensor algebra over Schwartz space), a Wightman state ω on it gives rise in general to a non-self-adjoint representation of the fields. (That is why in general we have to distinguish between the notions of "weak" commutant and "strong" commutant[11] for the fields.)

If $f = f^* \in \mathscr{S}$, the representation operator $\Pi(f)$ [more commonly denoted $A(f)$] is a symmetric operator. We can then construct the adjoint representation $A^{*t}(f)$ on a larger domain $\mathscr{D}^*$ and the transposed $A^{*t}(f)$ acting in $\mathscr{D}^{*\prime}$. In practice, it might be hard to give an explicit description of $\mathscr{D}^*$ and to find the explicit form of the fields, but this depend of course of the Wightman state ω chosen.

A similar situation occurs when we look at representations of a Lie algebra and its universal enveloping algebra on a dense invariant domain $\mathscr{D}$ other than the Gårding domain[12] (on this last domain we get a self-adjoint representation). The generators of the Lie algebra will be represented by symmetric, non-self-adjoint operators and similarly many elements of the enveloping algebra will not admit self-adjoint extensions [for these last ones, it is already true on the Gårding domain (Ref. 12, Chap. 11, pp. 323, 330)]. If we deal with a nonintegrable representation, even the Nelson operator Δ will not be essentially self-adjoint.

This is a typical situation where we can get the domain $\mathscr{D}^*$ strictly larger than $\mathscr{D}$, and a non-Hermitian extension Π^* which can be transposed to the dual space.

ACKNOWLEDGMENTS

I would like to thank Professor J.-P. Antoine for reading and improving the manuscript.

[1] A. Grossmann, "Elementary properties of nested Hilbert spaces," Commun. Math. Phys. **2**, 1–30 (1966).
[2] J. P. Antoine and A. Grossmann, "Partial inner product spaces I. General properties," J. Funct. Anal. **23**, 369–78 (1976); "II. Operators," J. Funct. Anal. **23**, 379–91 (1976).
[3] J. P. Antoine, "Partial inner product spaces, III. Compatibility relations revisited," J. Math. Phys. **21**, 268–79 (1980); "IV. Topological considerations," J. Math. Phys. **21**, 2067–79 (1980).
[4] R. T. Powers, "Self-adjoint algebras of unbounded operators," Commun. Math. Phys. **21**, 85–124 (1971).
[5] A. Grossmann, "Fields at a point," Commun. Math. Phys. **4**, 203–16 (1967).
[6] F. Debacker-Mathot, "Integral decomposition of unbounded operator families," Commun. Math. Phys. **71**, 47–58 (1980).
[7] J. P. Antoine, F. Debacker-Mathot, and J. R. Fontaine, "Group representations and operator algebras in partial inner product spaces," *Proceedings of the Third International Colloquium on Group Theoretical Methods in Physics, Marseille* 1974, edited by H. Bacry and A. Grossman (Faculty of Sciences of the University of Nijmegen, The Netherlands, 1974), Vol. I, 138–52.
[8] S. Gudder and W. Scruggs, "Unbounded representations of *-algebras," Pacific J. Math **70**, 369 (1977).
[9] A. Inoue, "On a class of unbounded operator algebras II," Pacific J. Math **66**, 411–31 (1976).
[10] H. J. Borchers, "On structure of the algebra of field operators," Nuovo Cimento **24**, 214 (1962).
[11] H. J. Borchers and J. Yngvason, "On the algebra of field operators. The weak commutant and integral decomposition of states," Commun. Math. Phys. **42**, 231–52 (1975).
[12] A. O. Barut and R. Raczka, *Theory of group representations and applications* (PWN-Polish Scientific, Warsaw, 1977).

A class of solutions of Calapso–Guichard equations

Dipankar Ray
Department of Physics, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 8 November 1979; accepted for publication 8 October 1980)

A class of solutions is indicated for Calapso–Guichard equations of surface theory.

PACS numbers: 02.40. + m

1. INTRODUCTION

For the problem of imbedding Guichard surfaces of the second kind[1] in the three-dimensional Euclidean space, Calapso[2] obtained the following equations:

$$[\theta_x + \tan\theta\xi_x]_x - [\theta_t - \cot\theta\xi_t]_t + (\cos\theta + h\sin\theta)(-\sin\theta + h\cos\theta) = 0, \tag{1.1a}$$

$$h_x = (h - \tan\theta)\xi_x, \tag{1.1b}$$

$$h_t = (h + \cot\theta)\xi_t, \tag{1.1c}$$

where the intrinsic metric on the Guichard surface is given by

$$dS_1^2 = w^{1^2} + w^{2^2},$$
$$w^1 = e^{\xi}\sin\theta dt,$$
$$w^2 = e^{\xi}\cos\theta dx.$$

The imbedding of this surface on E^3 is given by the second fundamental form, or the extrinsic cuvature:

$$dS_2^2 = \pi^{1^2} + \pi^{2^2},$$
$$\pi^1 = (\cos\theta + h\sin\theta)dt,$$
$$\pi^2 = (-\sin\theta + h\cos\theta)dx,$$

and $\theta_x \equiv \partial\theta/\partial x, \theta_t \equiv \partial\theta/\partial t, \theta_{xx} \equiv \partial^2\theta/\partial x^2$, and so on.

Gurses and Nutku[3] have noted that Eqs. (1.1) form a generalization of the sine-Gordon equation and have reformulated (1.1) as an inverse scattering problem. In the present note we seek to obtain explicit solutions of (1.1) under the assumption:

$$h \neq \text{const}, \quad \theta = \theta(h). \tag{1.2}$$

For $h = \text{const}$, (1.1) can be put into the form of the sine-Gordon equation and hence need not be discussed here.

2. SOLUTIONS

To obtain the solutions we consider the following cases one by one.

Case I: $h = \tan\theta$. (2.1)

From (1.1b), (1.2) and (2.1) we note that h and θ are functions of t only. (1.1) reduces to

$$\xi_{xx} + \frac{1}{h}\left(\frac{1}{h}\xi_t\right)_t = 0, \tag{2.2a}$$

$$\frac{h_t}{h + 1/h} = \xi_t. \tag{2.2b}$$

Integrating (2.2b) and putting into (2.2a), we get the solution of (1.1) as

$$\xi = -ax^2 + kx + a\alpha^2 + b\alpha + c,$$
$$h = (e^{2(a\alpha^2 + b\alpha + c)} - 1)^{1/2}, \tag{2.3}$$
$$\theta = \tan^{-1}h.$$

α is defined through

$$\int \frac{dx}{(e^{2(a\alpha^2 + b\alpha + c)} - 1)^{1/2}} = t,$$

a, b, c and k are constants.

Case II: $h = -\cot\theta$. (2.4)

Here, h and θ are functions of x only. Proceeding as in case (I) we get the solution as:

$$\xi = -a't^2 + k't + a'\beta^2 + b'\beta + c',$$
$$h = (e^{2(a'\beta^2 + b'\beta + c')} - 1)^{1/2}, \tag{2.5}$$
$$\theta = -\cot^{-1}h.$$

β is defined through:

$$\int \frac{d\beta}{(e^{2(a'\beta^2 + b'\beta + c')} - 1)^{1/2}} = x.$$

a', b', c' and d' are constants.

Case III: $h_x = 0, h \neq \tan\theta$. (2.6)

Here, in view of (1.1b), (1.2), and (2.6), h, θ, and ξ are all functions of t only. (1.1a) now reduces to:

$$(\theta_h - \cot\theta/(h + \cot\theta))h_{tt} + (\theta_h - \cot\theta/(h + \cot\theta))h_t^2 = \sin\theta\cos\theta(h + \cot\theta)(h - \tan\theta),$$

which gives the complete solutions of (1.1), subject to (1.2), and (2.6) as θ is any arbitrary function of h such that

$$(h + \cot\theta)\theta_h - \cot\theta \neq 0.$$

h is given by

$$\int \frac{(\theta_h - \cot\theta/(h + \cot\theta))\,dh}{\left(A + 2\int \sin\theta\cos\theta(h + \cot\theta)(h - \tan\theta)(\theta_h - \cot\theta/(h + \cot\theta))\,dh\right)^{1/2}} = t. \tag{2.7}$$

 0022-2488/81/071390-03$1.00

ξ is given by

$$\int \frac{dh}{h+\cot\theta} = \xi.$$

A is a constant.

Case IV: $h_t = 0$, $h + \cot\theta \neq 0$. (2.8)

Here, in view of (1.1c), (1.2) and (2.6), h, θ and ξ are functions of x only. Proceeding as in Case III, we get the complete solution of (1.1), subject (1.2) and (2.8) θ is an arbitrary function of h such that

$$\theta_h + \tan\theta/(h-\tan\theta) \neq 0.$$

h is given by

$$t = \int \left(\frac{(\theta_h + \tan\theta/(h-\tan\theta))dh}{A + 2\int \sin\theta\cos\theta\,(h+\cot\theta)(h-\tan\theta)(\theta_h - \cot\theta/(h+\cot\theta))\,dh} \right)^{1/2}. \tag{2.9}$$

ξ is given by

$$\xi = \int \frac{dh}{h+\cot\theta}.$$

Case V: $h_x \neq 0$, $h_t \neq 0$. (2.10)

Define

$$y = \int \left(\frac{1}{h-\tan\theta} + \frac{1}{h+\cot\theta} \right) dh. \tag{2.11}$$

From (1.1b) and (2.11), the condition that $\partial\xi_x/\partial t = \partial\xi_t/\partial x$ give

$$y = F + G, \tag{2.12}$$

where

$$F = F(x) \quad \text{and} \quad G = G(t).$$

(1.1a) reduces to

$$2U_y f + U f_F = 2V_y g + V g_G, \tag{2.13}$$

where

$$f = F_x^2 - 1, \quad g = G_t^2 - 1, \tag{2.14a}$$

$$U = \theta_y + \frac{\tan\theta h_y}{h-\tan\theta}, \quad V = U - h. \tag{2.14b}$$

We shall show that g as a function of G satisfies a differential equation of the form

$$C g_{GGG} + D g_{GG} + M g_G + N = 0, \tag{2.15}$$

where C, D, M, and N are constants, not all of them are zero. The proof is as follows.

If $U = 0$, then (2.13) reduces to

$$2V_y g + V g_G = 0, \tag{2.16}$$

which is valid for a continuous range of values of F and G, in that case. Since F can be varied with G so as to keep y constant, we see from (2.16) that, if $U = 0$, then g satisfies an equation of the form (2.15) where $C = 0$, $D = 0$. On the other hand, if $U \neq 0$, dividing (2.13) by U and differentiating with respect to F, we get

$$\left(\frac{2U_y}{U}\right)_y f = \left(\frac{2V_y}{U}\right)_y g + \left[\frac{2V_y}{U} + \left(\frac{V}{U}\right)_y\right] g_G + (V/U) g_{GG}. \tag{2.17}$$

If $(U_y/U)_y = 0$, then, by arguments similar to the case of $U = 0$, we see that g satisfies an equation of the form (2.15) with $C = 0$. On the other hand, if $(U_y/U)_y \neq 0$, then dividing (2.17) by $(U_y/U)_y$ and differentiating with respect to F, we see that g satisfies an equation of the form (2.15).

Therefore, in either case, g satisfies equation of the form (2.15). So, either

$$g = \mu e^{-KG} + \mu_1 e^{-K_1 G} + \mu_2 e^{-K_2 G},$$

or

$$g = \mu e^{-KG} + e^{-K_1 G}(\mu_1 + \mu_2 G),$$

or

$$g = e^{-KG}(\mu + \mu_1 G + \mu_2 G^2),$$

where $\mu, \mu_1, \mu_2, K, K_1$ and K_2 are constants which could be real or complex, but g is real. Similarly, either

$$f = \lambda e^{-LF} + \lambda_1 e^{-L_1 F} + \lambda_2 e^{-L_2 F},$$

or

$$f = \lambda e^{-LF} + e^{-L_1 F}(\lambda_1 + \lambda_2 F),$$

or

$$f = e^{-LF}(\lambda + \lambda_1 F + \lambda_2 F^2),$$

where $\lambda, \lambda_1, \lambda_2, L, L_1$ and L_2 are constants which would be real or complex, but f is real.

Putting these various possible forms of g and f into (2.13) and using (2.14b), we see that the only combination of f and g that can satisfy (2.13) and (2.14) is

$$f = \lambda e^{KF}, \quad g = \mu e^{-KG}, \tag{2.18}$$

where λ, μ, and K are now real constants.

Putting (2.18) into (2.13) and integrating

$$\lambda U e^{(K/2)y} = \mu V e^{-(K/2)y} + l, \tag{2.19}$$

where l is a constant, (2.11) can be rewritten as

$$h_y = (h - \tan\theta)(h + \cot\theta)/(\tan\theta + \cot\theta). \tag{2.20a}$$

Using (2.14b) and (2.20a), we rewrite (2.19) as

$$\lambda\left(\theta_y + \frac{\tan\theta h_y}{h-\tan\theta}\right) e^{(k/2)y} = \mu\left(\theta_y - \frac{\cot\theta h_y}{h+\cot\theta}\right) \times e^{-(K/2)y} + l. \tag{2.20b}$$

λ, μ, K, and l are constants.

From (2.14a) and (2.18), y is given by

$$y = F + G,$$

where

$$\int \frac{dF}{(\lambda e^{KF}+1)^{1/2}} = x, \quad \int \frac{dG}{(\mu e^{-KG}+1)^{1/2}} = t, \quad (2.20c)$$

and ξ is given by

$$\xi = \int \frac{h_x dx}{h - \tan\theta} + \int \frac{h_t dt}{h + \cot\theta}. \quad (2.20d)$$

It can be checked by direct substitution that (2.20) satisfies (1.1). Hence (2.20) contains all the solutions of (1.1) for the case under consideration. Since (2.20a) and (2.20b) form coupled first order differential equations for h and θ as functions of y, it can be solved numerically. Since y is given by (2.20c), ξ can then be determined from (2.20d).

3. CONCLUSION

Summarily all the solutions of (1.1) subject to (1.2) are given by (2.3), (2.5), (2.7), (2.9),and (2.20). While (2.3), (2.5), (2.7),and (2.9) are in completely integrated form, in Eq. (2.20) we have a coupled first order differential equation to be solved numerically. It can also be noted that the equations imbedding a Guichard surface of first kind can be obtained from (1.1) by the transformation $\theta \to i\theta$, $h \to ih$, $t \to it$, $x \to x$ and $\xi \to \xi$. The present work can also be transformed to the case of Guichard surface of the first kind without difficulty by using the same transformation.

[1]C. Guichard, C. R. Acad. Sci. **130**, 159 (1900).
[2]P. Calapso, Ann. Mat. Ser. III **11**, 201 (1905).
[3]M. Gürses and Y. Nutku, Princeton University, Joseph Henry Laboratory preprint.

New nonlinear evolution equations from surface theory

Metin Gürses [a)]
Max Planck Institut für Physik and Astrophysik, Garching 8046, München, Germany
Physics Department, Middle East Technical University, Ankara, Turkey

Yavuz Nutku
Mathematics Department, Boğaziçi University, Istanbul, Turkey

(Received 26 August 1980; accepted for publication 8 October 1980)

We point out that the connection between surfaces in three-dimensional flat space and the inverse scattering problem provides a systematic way for constructing new nonlinear evolution equations. In particular we study the imbedding for Guichard surfaces which gives rise to the Calapso–Guichard equations generalizing the sine-Gordon (SG) equation. Further, we investigate the geometry of surfaces and their imbedding which results in the Korteweg–deVries (KdV) equation. Then by constructing a family of applicable surfaces we obtain a generalization of the KdV equation to a compressible fluid.

PACS numbers: 02.40 + m, 02.30.Jr

I. INTRODUCTION

The Korteweg–deVries and sine-Gordon equations are classic examples of nonlinear evolution equations which form completely integrable systems. They belong to a class of partial differential equations which can be solved by the inverse scattering method of Gardner, Greene, Kruskal, and Miura[1] and Zakharov and Shabat.[2] At present we do not have a set of partial differential equations encompassing all such systems. Thus, given a set of equations we must first check whether or not they can be formulated as an inverse scattering problem according to the general framework provided by Lax[3] and Ablowitz, Kaup, Newell, and Segur,[4] (LAKNS). It is therefore of interest to construct new equations generalizing the KdV and SG equations and in this paper we shall present such new systems. We shall verify that these generalizations are viable by showing that the new equations form LAKNS systems. For this purpose we shall in each case obtain the linear equations of the scattering problem for which the new nonlinear equations act as integrability conditions. Finding the explicit form of the potentials for the appropriate inverse scattering problem is also the first step toward the solution of these equations. But a discussion of solutions will be postponed to a future paper because here we shall be concerned solely with the problem of constructing new nonlinear evolution equations.

We have obtained the new equations by exploiting the one-to-one correspondence between LAKNS systems and the classical theory of surfaces in three-dimensional space. For the SG equation where the underlying surface is pseudospherical this is very familar territory,[5] while for the general case this correspondence has been discussed by Crampin Pirani and Robinson[6] at the level of connection. The "soliton connection" is a flat linear connection in a principal fiber bundle with structure group SL $(2,R)$. Its relationship to the problem of imbedding surfaces in three-dimensional Euclidean space arises from the fact that the Gauss–Codazzi equations are in this case equivalent to Cartan's equations of structure for SO (3).[7] This correspondence suggests that the soliton connection can be given a richer structure at the level of Riemannian metrics. We shall discuss this problem in Sec. II where we shall formulate the LAKNS equations in terms of the first and second fundamental forms of the surface. Then, quite generally, the linear equations to be solved by the inverse scattering method are the Weingarten equations and their integrability conditions which result in the nonlinear evolution equations consist of the Gauss–Codazzi equations.

There are advantages to be derived from the recognition of the metric level appropriate to the soliton connection. First of all it will enable us to correct misleading statements in the literature to the effect that all LAKNS systems correspond to pseudospherical surfaces.[8] The validity of such a statement requires the use of the equations of motion in the definition of the connection, but this is an identity which does not yield any new information about metric structure. On the other hand, if we can properly identify the surface at the metric level we can consider its generalizations in classical differential geometry and such surfaces will provide new examples of completely integrable systems. As an illustration of this point, in Sec. III we shall consider the problem of imbedding surfaces which are applicable to quadrics. These are known as surfaces of Guichard[9] and they are generalizations of pseudospherical surfaces. The Gauss–Codazzi equations for Guichard surfaces, first obtained by Calapso,[10] provide an attractive generalization of the SG equation.[11] We shall give an updated derivation of the Calapso–Guichard equations (CG) and cast them into the form of an inverse scattering problem.

The equivalence between LAKNS systems and surface theory at the metric level can be exploited in a systematic way in order to construct new nonlinear evolution equations. We shall discuss this process in Sec. II and in Sec. IV apply it to the KdV equation to obtain a generalization of this equation,

[a)] Alexander von Humboldt fellow.

 0022-2488/81/071393-06$1.00

$$\lambda_t + 2u\lambda_x + 2\kappa u_x = 0,$$
$$u_t + 6uu_x + u_{xxx} - \lambda_x u_{xx} = 0, \tag{1.1}$$

where we have introduced a new field λ and κ is constant. This pair of coupled partial differential equations can be interpreted as the continuity and Euler equations for a compressible fluid. In deriving these equations the first step was to investigate the family of surfaces for which the imbedding problem results in the KdV equation. It appears that this question has not been asked before even though it is of basic interest. We shall obtain the expression for the first and second fundamental forms characterizing the geometry of surfaces underlying the KdV equation. Then we shall consider the problem of constructing applicable surfaces, that is, surfaces with the same intrinsic geometry as that appropriate to the KdV equation but with a different imbedding into three-dimensional flat space. The resulting Gauss–Codazzi equations (1.1) are new nonlinear evolution equations. We shall formulate them as an inverse scattering problem and conclude by pointing out an alternative derivation of these equations. The connection 1-form appropriate to the KdV equation can in particular be subjected to gauge transformations belonging to R, an abelian subgroup of SL (2,R), which is also familiar as the group of scale transformations leaving the KdV equation invariant. If we let the "scale parameter" become a function of position and time while preserving the connection 1-form appropriate to this abelian subgroup, we obtain a new realization of the solition connection. Equations (1.1) are the conditions for this new connection to have zero curvature.

II. SURFACES

We shall briefly review the LAKNS equations of the inverse scattering method and the Weingarten and Gauss–Codazzi equations of the imbedding problem in order to point out the one-to-one correspondence between them. The rest of this section is devoted to the construction of applicable surfaces, which provides a systematic way for generalizing familar examples of LAKNS systems.

We consider a moving frame in a three-dimensional flat space M with Euclidean or Lorentzian signature. At a point P in M we have

$$dP = \omega^i \mathbf{e}_i, \tag{2.1}$$

$\mathbf{e}_i$, $i = 1,2,3$, are the basis vectors and ω^i the basis 1-forms. A surface S is defined by

$$\omega^3 = 0, \tag{2.2}$$

accordingly $\mathbf{e}_3$ is the normal vector to the surface. We shall let the metric of M be of the form

$$ds^2 = \mathrm{diag}(1,\epsilon,\eta), \quad \epsilon = \pm 1, \quad \eta = \pm 1 \tag{2.3}$$

so that there is no loss of generality in always defining the surface by Eq. (2.2). Cartan's equations of structure are

$$d\mathbf{e}_i = \omega_i{}^k \mathbf{e}_k, \tag{2.4}$$

where $\omega_i{}^k$ are connection 1-forms and their integrability conditions become

$$\Theta^i{}_k = 0, \tag{2.5}$$

where

$$\Theta^i{}_k = d\omega^i{}_k + \omega^i{}_j \wedge \omega^j{}_k \tag{2.6}$$

are the curvature 2-forms. We can write these equations in a more familar form if we introduce the first and second fundamental forms of the surface. That is, we consider a surface with the intrinsic metric

$$ds_1{}^2 = (\omega^1)^2 + \epsilon(\omega^2)^2 \tag{2.7}$$

and the Riemannian connection on S is given by the 1-form $\omega^1{}_2$ where

$$d\omega^\alpha + \omega^\alpha{}_\beta \wedge \omega^\beta = 0 \tag{2.8}$$

with the Greek indices ranging over two values only. Finally we have

$$d\omega^\alpha{}_\beta = K\omega^\alpha \wedge \omega^\beta, \tag{2.9}$$

where K is the Gaussian curvature of the surface. We shall now consider the imbedding problem and to this end we let

$$-ds_2{}^2 = \omega^1 \otimes \pi^1 + \epsilon\omega^2 \otimes \pi^2 \tag{2.10}$$

denote the extrinsic curvature, or the second fundamental form of S. The Gauss–Codazzi equations for imbedding S in M are the same as Eqs. (2.5)–(2.6) with the identification

$$\omega^1{}_3 = \pi^1,$$
$$\omega^2{}_3 = \pi^2, \tag{2.11}$$

and therefore we have

$$d\omega^1{}_2 - \epsilon\eta\pi^1 \wedge \pi^2 = 0,$$
$$d\pi^1 + \omega^1{}_2 \wedge \pi^2 = 0, \tag{2.12}$$
$$d\pi^2 - \epsilon\omega^1{}_2 \wedge \pi^1 = 0.$$

Finally, from Eq. (2.2) which defines the surface, we have the condition

$$\omega^1 \wedge \pi^1 + \epsilon\omega^2 \wedge \pi^2 = 0. \tag{2.13}$$

In the imbedding problem the equations of structure (2.4) are also known as Weingarten equations and together with their integrability conditions which are the Gauss–Codazzi equations (2.12) they constitute the fundamental equations of the subject.

The Gauss–Codazzi equations for imbedding S in M can be written as Cartan's equations for SL (2,R)

$$d\theta^i + \tfrac{1}{2}c_j{}^i{}_k \theta^j \wedge \theta^k = 0, \tag{2.14}$$

where $c_j{}^i{}_k$ are the structure constants of SL (2,R). Hereafter Latin indices will stand for SL (2,R) values and range over $i = 0,1,2$. Equations (2.14) will be identical to Eqs. (2.12) provided we let

$$\omega^1{}_2 = -(2i/\epsilon^{\frac{1}{2}})\theta^0,$$
$$\pi^1 = i\eta^{\frac{1}{2}}(\theta^1 + \theta^2), \tag{2.15}$$
$$\pi^2 = \eta^{\frac{1}{2}}\epsilon^{\frac{1}{2}}(-\theta^1 + \theta^2),$$

since

$$c_1{}^0{}_2 = 1,$$
$$c_0{}^1{}_1 = -c_0{}^2{}_2 = 2, \tag{2.16}$$

are the nonvanishing structure constants of SL (2,R). In the

form which will be most useful for our purposes we can express this result as follows:

Given the first and second fundamental forms of the surface as in Eqs. (2.7) and (2.10) we can construct an SL (2,R) valued connection 1-form

$$\Gamma = \begin{pmatrix} \theta^0 & \theta^1 \\ \theta^2 & -\theta^0 \end{pmatrix}, \tag{2.17}$$

where θ^i are given by Eqs. (2.15), which has a vanishing curvature 2-form

$$\Theta = d\Gamma + \Gamma \wedge \Gamma = 0 \tag{2.18}$$

by virtue of the Gauss–Codazzi equations (2.12).

When we turn to the problem of solitons, we find that the soliton connection is an SL (2,R) valued 1-form as in Eqs. (2.17), where[6]

$$\begin{aligned} \theta^0 &= -(A dt + i\zeta dx), \\ \theta^1 &= -(B dt + q dx), \\ \theta^2 &= -(C dt + r dx), \end{aligned} \tag{2.19}$$

and A,B,C,q,r are functions of t and x while ζ is a constant. The condition that its curvature 2-form should vanish, [cf. Eqs. (2.18)], reduces to the LAKNS equations

$$\begin{aligned} A_x &= qC - rB, \\ B_x - 2i\zeta B &= q_t - 2Aq, \\ C_x + 2i\zeta C &= r_t + 2Ar, \end{aligned} \tag{2.20}$$

where here and in the following subscripts such as x, t denote partial differentiation with respect to the coordinates. In both problems we have SL (2,R)-valued flat connection forms but there is one further condition which must be satisfied before we can establish their equivalence. In Eqs. (2.19) ζ, which will be the eigenvalue in the inverse scattering method, must be a constant. Starting with the connection given by Eqs. (2.17) we can always perform an SL (2,R) gauge transformation

$$\Gamma' = \Sigma \Gamma \Sigma^{-1} + \Sigma d\Sigma^{-1}, \tag{2.21}$$

where

$$\Sigma = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}, \quad \alpha\delta - \beta\gamma = 1, \tag{2.22}$$

to cast it into the form given in Eqs. (2.19) and vice versa. The condition (2.18) is gauge-invariant and therefore we have a one-to-one correspondence between LAKNS systems and the classical theory of surfaces imbedded in a three-dimensional flat space M.

We shall now consider the problem of constructing applicable surfaces. These surfaces will carry the same intrinsic metric as in Eqs. (2.7) but a different imbedding. The choice of the new expression for the second fundamental form (which via the Gauss–Codazzi equations leads to new equations of motion) is best understood by turning to Eqs. (2.21). In these equations we have the transformation rule for the connection under a change of gauge. We have already remarked that the soliton connection is a flat connection and this property is invariant under gauge transformations. It will be useful for our purposes to specialize to transformations belonging to R^+ which is an abelian subgroup of SL (2,R), where

$$\Sigma = \begin{pmatrix} e^{\lambda} & 0 \\ 0 & e^{-\lambda} \end{pmatrix} \tag{2.23}$$

is a typical element. In this case Eqs. (2.21) reduce to

$$\begin{aligned} \theta'^0 &= \theta^0 + d\lambda, \\ \theta'^1 &= e^{2\lambda}\theta^1, \\ \theta'^2 &= e^{-2\lambda}\theta^2, \end{aligned} \tag{2.24}$$

and through the identifications (2.15) we find that this change of gauge corresponds to the simultaneous rotations

$$\begin{aligned} \omega'^{\alpha} &= (\Lambda^{-1})^{\alpha}{}_{\beta}\omega^{\beta}, \\ \pi'^{\alpha} &= \Lambda^{\alpha}{}_{\beta}\pi^{\beta}, \end{aligned} \tag{2.25}$$

where

$$\Lambda = \begin{pmatrix} \cosh 2\lambda & -(i/\epsilon^{\frac{1}{2}})\sinh 2\lambda \\ i\epsilon^{\frac{1}{2}}\sinh 2\lambda & \cosh 2\lambda \end{pmatrix}. \tag{2.26}$$

is a position-dependent rotation. The transformation law for the connection 1-form on S is given by

$$\omega^1{}_2 \to \omega^1{}_2 + 2id\lambda, \tag{2.27}$$

which follows from Eqs. (2.15) or directly from Eqs. (2.25). Since the equations of motion are obtained from the gauge-invariant condition that the curvature should vanish they are not affected by gauge transformations such as the one implicitly defined by Eqs. (2.25). If, however, we undo the rotation in Eqs. (2.25) for $\{\omega^{\alpha}\}$ only, we shall find an applicable surface where

$$\begin{aligned} \omega^{\alpha} &\to \omega^{\alpha}, \\ \pi^{\alpha} &\to \Lambda^{\alpha}{}_{\beta}\pi^{\beta}, \end{aligned} \tag{2.28}$$

and λ will now enter as a new field into the equation of motion. In this case $\omega^1{}_2$ remains invariant and we have

$$\begin{aligned} \theta^0 &\to \theta^0, \\ \theta^1 &\to e^{2\lambda}\theta^1, \\ \theta^2 &\to e^{-2\lambda}\theta^2, \end{aligned} \tag{2.29}$$

in place of Eqs. (2.24). The requirement that the curvature of the new connection be zero results in

$$\begin{aligned} \Theta^0 &= 0, \\ \Theta^1 + 2d\lambda \wedge \theta^1 &= 0, \\ \Theta^2 - 2d\lambda \wedge \theta^2 &= 0, \end{aligned} \tag{2.30}$$

which are new equations of motion. We note that the Θ^0 component of curvature remains unchanged when we consider applicable surfaces. Therefore if we want to obtain new field equations by using this method, we must start with a connection where Θ^0 vanishes identically. In Sec. IV we shall apply this construction to the KdV equation.

III. CALAPSO–GUICHARD EQUATIONS

The prototype of a completely integrable system where surface theory plays a paramount role is the SG equation.

The Gauss–Codazzi equations for imbedding pseudospherical surfaces in E^3 reduce to the SG equation and there are generalizations of this equation which utilize the same intrinsic geometry.[5,12] In contrast, the 2-surface behind the Calapso–Guichard (CG) equations is a quadric. So we consider an intrinsic metric on the surface with the basis 1-forms

$$\omega^1 = e^{\xi}\sin\vartheta dt,$$
$$\omega^2 = e^{\xi}\cos\vartheta dx, \tag{3.1}$$

which differs from the metric of a pseudospherical surface by a conformal factor. The Riemannian connection on the surface is given by the 1-form

$$\omega^1{}_2 = (\vartheta_x + \tan\vartheta\,\xi_x)dt + (\vartheta_t - \cot\vartheta\,\xi_t)dx \tag{3.2}$$

which follows from Eqs. (2.8) and (3.1). The imbedding of this surface in E^3 is given by the second fundamental form [cf. Eq. (2.10)], where

$$\pi^1 = (\cos\vartheta + h\sin\vartheta)dt,$$
$$\pi^2 = (-\sin\vartheta + h\cos\vartheta)dx \tag{3.3}$$

are the basis 1-forms. We can now verify the fundamental property of Guichard surfaces which is responsible for the parametrization used in Eqs. (3.3): If we consider two Guichard surfaces S,S' which have the same extrinsic curvature, then their principal radii of curvature ρ_i,ρ_i' where

$$\pi^1 = \rho_1\omega^1 = \rho_1'(\omega^1)',$$
$$\pi^2 = \rho_2\omega^2 = \rho_2'(\omega^2)'$$

satisfy the relation

$$\tfrac{1}{2}(\rho_1\rho_1' + \rho_2\rho_2') = -1. \tag{3.4}$$

The surface S' is said to be the associate of S and they are related by the transformations

$$e^{i\vartheta'} = [(1 + ih)/(1 - ih)]e^{-i\vartheta},$$
$$e^{\xi'} = e^{-\xi}(1 + h^2), \tag{3.5}$$
$$h' = h.$$

Historically Eq. (3.4) has been the starting point of the investigations on the surfaces of Guichard. Using Eqs. (3.2) and (3.3) in Eqs. (2.12) we obtain the Gauss–Codazzi equations for imbedding S in E^3

$$(\vartheta_x + \tan\vartheta\,\xi_x)_x - (\vartheta_t - \cot\vartheta\,\xi_t)_t$$
$$+ (\cos\vartheta + h\sin\vartheta)(-\sin\vartheta + h\cos\vartheta) = 0,$$
$$h_x = (h - \tan\vartheta)\xi_x,$$
$$h_t = (h + \cot\vartheta)\xi_t. \tag{3.6}$$

These are the Calapso–Guichard equations. We may also call them sine-Guichard equations because they are obtained for Guichard surfaces of the second kind. Needless to say, there is also a sinh-Guichard system which generalizes the sinh-Gordon equation

$$(\vartheta_x + \coth\vartheta\xi_x)_x + (\vartheta_y + \tanh\vartheta\,\xi_y)_y$$
$$+ (\cosh\vartheta + h\sinh\vartheta)(\sinh\vartheta + h\cosh\vartheta) = 0,$$
$$h_x = (h + \coth\vartheta)\xi_x,$$
$$h_y = (h + \tanh\vartheta)\xi_y. \tag{3.7}$$

These equations are Gauss–Codazzi equations for the surfaces of Guichard of the first kind. We shall not consider them any further in this paper except to note that every statement concerning the sine-Guichard equations can be translated into an analoguous one about the sinh-Guichard equations.

We shall now consider the formulation of CG equations as an inverse scattering problem. For this purpose we need to transform these equations into a form whereby the correspondence limit of our results with those of Ablowitz, *et al.*[4] will become manifest. Hence we shall first introduce null coordinates.

$$u = t - x, \quad v = t + x, \tag{3.8}$$

and rewrite Eqs. (3.6) in this new coordinate system. We find

$$(\vartheta_u - \cot2\vartheta\,\xi_u - \csc2\vartheta\,\xi_v)_v$$
$$+ (\vartheta_v - \cot2\vartheta\,\xi_v - \csc2\vartheta\,\xi_u)_u$$
$$+ (\cos\vartheta + h\sin\vartheta)(-\sin\vartheta + h\cos\vartheta) = 0,$$
$$h_v = (h + \cot2\vartheta)\xi_v + \csc2\vartheta\,\xi_u,$$
$$h_u = (h + \cot2\vartheta)\xi_u + \csc2\vartheta\,\xi_v, \tag{3.9}$$

which has a symmetric dependence on the null coordinates but the lack of Lorentz covariance of the CG equations has resulted in long expressions. We note that either in Eqs. (3.6) or in Eqs. (3.9) we can completely eliminate ξ to obtain two coupled partial differential equations for ϑ and h. This form of the equations is useful in the formulation of the initial value problem but we shall keep ξ for ease of handling the equations. Finally, we shall remark that in CG equations h behaves in a manner similar to a stereographic variable. In particular the introduction of stereographic angle where

$$h = \tan(\phi/2) \tag{3.10}$$

simplifies the calculations. This relation is suggested, for example, by Eqs. (3.5) which now reduce to the transformations

$$\vartheta' = \phi - \vartheta,$$
$$\phi' = \phi, \tag{3.11}$$

relating a Guichard surface S to its associate S'.

The formulation of CG equations as Cartan's equations for SL (2,R), [cf. Eq. (2.14)] follows from Eqs. (3.2), (3.3) and the identification in Eqs. (2.15). In the null coordinate system the 1-forms are given by

$$\theta^0 = -\tfrac{1}{2}i(\vartheta_u - \cot2\vartheta\,\xi_u - \csc2\vartheta\,\xi_v)du$$
$$+ \tfrac{1}{2}i(\vartheta_v - \cot2\vartheta\,\xi_v - \csc2\vartheta\,\xi_u)dv,$$
$$\theta^1 = \tfrac{1}{2}i(1 + ih)e^{-i\vartheta}du + \tfrac{1}{2}i(1 - ih)e^{i\vartheta}dv,$$
$$\theta^2 = -\overline{\theta^1}, \tag{3.12}$$

where bar denotes complex conjugation. However, as we mentioned in Sec. II, the connection 1-form Γ obtained from Eqs. (3.12) is not directly in the canonical form of a soliton connection. In order to achieve the canonical form required by the inverse scattering method we must perform a gauge

transformation, as in Eqs. (2.21). We find that

$$\Sigma = \begin{pmatrix} \alpha & \alpha \\ -1/2\alpha & 1/2\alpha \end{pmatrix}, \tag{3.13}$$

where

$$\alpha^2 = [(\zeta + f)/(1 + ih)]e^{i\vartheta}, \tag{3.14}$$

with

$$f^2 = \zeta^2 - \tfrac{1}{4}(1 + h^2) \tag{3.15}$$

the desired transformation matrix. ζ is so far an arbitrary constant which will eventually be interpreted as the eigenvalue in the inverse scattering problem. From Eqs. (2.21) and (2.19) we can now determine the potentials which appear in the inverse scattering problem. The results are

$$\begin{aligned}
q &= -\tfrac{1}{2}i(2\vartheta_v - \cot 2\vartheta\, \xi_v - \csc 2\vartheta\, \xi_u) + Hh_v + if,\\
r &= -\tfrac{1}{2}i(2\vartheta_v - \cot 2\vartheta\, \xi_v - \csc 2\vartheta\, \xi_u) + Hh_v - if,\\
A &= [i/(1 + h^2)][\zeta(1 - h^2) - 2ihf]\cos 2\vartheta\\
&\quad + [i/(1 + h^2)][2\zeta h + i(1 - h^2)f]\sin 2\vartheta,\\
B &= -\tfrac{1}{2}i(\csc 2\vartheta\, \xi_v + \cot 2\vartheta\, \xi_u) + Hh_u\\
&\quad + [1/(1 + h^2)][2\zeta h + i(1 - h^2)f]\cos 2\vartheta\\
&\quad - [1/(1 + h^2)][\zeta(1 - h^2) - 2ihf]\sin 2\vartheta,\\
C &= -\tfrac{1}{2}i(\csc 2\vartheta\, \xi_v + \cot 2\vartheta\, \xi_u) + Hh_u\\
&\quad - [1/(1 + h^2)][2\zeta h + i(1 - h^2)f]\cos 2\vartheta\\
&\quad + [1/(1 + h^2)][\zeta(1 - h^2) - 2ihf]\sin 2\vartheta,
\end{aligned} \tag{3.16}$$

where

$$H = \tfrac{1}{2}(\zeta h + if)/[f(1 + h^2)]. \tag{3.17}$$

We have thus cast the Calapso–Guichard equations into a form whereby solutions can be obtained by a direct application of the inverse scattering method.

IV. SURFACE THEORY AND KdV EQUATION

The theory of surfaces has not played a significant role in the case of KdV equation. In this respect historically the approach to the SG and KdV equations has led through very different routes. But in this paper we have emphasized the one-to-one correspondence between surface theory and LAKNS systems, a distinguished member of which is the KdV equation. It is therefore natural to ask about the nature of the surface, i.e., its intrinsic geometry and imbedding in M, which gives rise to the KdV equation.

In order to investigate the geometry underlying a given LAKNS system it is necessary to proceed in a direction opposite to that familiar from surface theory. Thus we shall start with the soliton connection of AKNS and finally obtain the first and second fundamental forms of the surface. The AKNS connection for KdV equation is given by the 1-forms.

$$\begin{aligned}
\theta^0 &= (-4i\zeta^3 + 2i\zeta u - u_x)dt + i\zeta dx,\\
\theta^1 &= (4\zeta^2 u + 2i\zeta u_x - 2u^2 - u_{xx})dt + u dx,\\
\theta^2 &= (-4\zeta^2 + 2u)dt - dx,
\end{aligned} \tag{4.1}$$

and the requirement that its curvature 2-form should vanish results in

$$u_t + 6uu_x + u_{xxx} = 0, \tag{4.2}$$

which is the KdV equation. The connection 1-form in Eqs. (4.1) is not particularly useful for our purposes but we are free to perform SL $(2,R)$ transformations to write it in any gauge we desire. There is in fact a well-known gauge where simple expressions for the connection 1-form are obtained:

$$\begin{aligned}
\theta^0 &= -u_x dt,\\
\theta^1 &= dx - 2u dt,\\
\theta^2 &= -u dx + (2u^2 + u_{xx})dt.
\end{aligned} \tag{4.3}$$

and this gauge is interesting because the new θ^0 is manifestly the connection of a Riemannian metric. Hence we can identify the first fundamental form of the surface

$$ds_1^2 = 32u dt^2 + 16 dt dx, \tag{4.4}$$

where

$$\begin{aligned}
\omega^1 &= dx + 2(u + 2)dt,\\
\omega^2 &= dx + 2(u - 2)dt
\end{aligned} \tag{4.5}$$

are the orthonormal basis 1-forms. We can verify that with

$$\epsilon = -1 \tag{4.6}$$

the Riemannian connection 1-form $\omega^1{}_2$ is consistent with θ^0 given by Eq. (4.3) and the definition in Eq. (2.15); and furthermore the conditions (2.13) are satisfied identically. Further comparison of Eqs. (2.15) and (4.3) leads to

$$\eta = -1 \tag{4.7}$$

and equations (4.6) and (4.7) determine the nature of the imbedding problem for the KdV equation. From Eq. (2.9) we find that

$$K = -\tfrac{1}{4}u_{xx} \tag{4.8}$$

is the Gaussian curvature of the surface. It remains to identify the second fundamental form of this surface and from Eqs. (2.15) and (4.3) we find

$$ds_2^2 = 2dx^2 - 8u dt dx + 8(u^2 - 4K)dt^2. \tag{4.9}$$

The imbedding problem of a surface with the first and second fundamental forms given by Eqs. (4.4) and (4.9) is equivalent to the KdV equation. We shall now consider a new family of surfaces applicable to this surface and obtain a generalization of KdV equation.

The KdV equation has well-known scale invariance properties. That is, it remains invariant under the transformations

$$\begin{aligned}
u &\to e^{-4\lambda}u,\\
x &\to e^{2\lambda}x,\\
t &\to e^{6\lambda}t,
\end{aligned} \tag{4.10}$$

where λ is an arbitrary constant parameter. A familiar consequence of this invariance is the similarity solution of the KdV equation which is known as cnoidal waves. The origin of this invariance can be traced to the gauge transformations the KdV soliton connection may be subjected to. In particular we can check that under the transformations (4.10) the soliton connection defined by the 1-forms in Eqs. (4.3) transforms according to Eqs. (2.24). Hence scale transformations leaving the KdV equation invariant can alternatively be regarded as gauge transformations of the KdV connection 1-form for R. In order to construct applicable surfaces to the

KdV-surface we shall proceed as in Sec. II and promote the "scale-parameter" λ to become a function of x and t. We shall further let the KdV connection 1-form (4.3) suffer the transformations (2.28) so that

$$\begin{aligned}\theta^0 &= -u_x dt,\\ \theta^1 &= e^{\lambda}(dx - 2udt),\\ \theta^2 &= e^{-\lambda}[-udx + (2u^2 + u_{xx})dt],\end{aligned} \tag{4.11}$$

and from Eqs. (2.30) we find

$$\begin{aligned}&\lambda_t + 2u\lambda_x = 0,\\ &u_t + 6uu_x + u_{xxx} - \lambda_x u_{xx} = 0,\end{aligned} \tag{4.12}$$

which are new equations of motion. These equations are also scale-invariant under the transformations (4.10).

Equations (4.12) can be further extended to Eqs. (1.1) by scaling the connection 1-forms (4.11) according to

$$\begin{aligned}\tilde{\theta}^0 &= (1-\kappa)\theta^0,\\ \tilde{\theta}^1 &= (1-\kappa)\theta^1,\\ \tilde{\theta}^2 &= \theta^2,\end{aligned} \tag{4.13}$$

where $\kappa \neq 1$ is an arbitrary constant. Such a transformation does not belong to SL $(2,R)$ and therefore it does not describe an invariance property of the field equations. Rather, it leads to a new result which consists of the introduction of an arbitray constant into the field equations. This process is consistent because $\Theta^0 = 0$ is once again identically satisfied. The remaining conditions on the curvature result in Eqs. (1.1) and we note that the KdV limit of these equations is given by

$$\lambda \to 0, \quad \kappa \to 0 \tag{4.14}$$

which are both necessary. Finally we shall remark that for $\kappa \neq 0$ the first of Eqs. (1.1) can be written as

$$\rho_t + (2u\rho)_x = 0, \tag{4.15}$$

with

$$\rho = \kappa e^{(1/\kappa)\lambda} \tag{4.16}$$

which is a continuity equation without source terms.

We shall now formulate Eqs. (1.1) as an inverse scattering problem. Once again we must subject the connection 1-forms (4.11) to a gauge transformation in order to cast them into the form of Eqs. (2.19). In this case the required SL $(2,R)$ transformation is given by

$$\Sigma = \begin{pmatrix} 1 & \beta \\ i\zeta & 1 + i\zeta\beta \end{pmatrix}, \tag{4.17}$$

where

$$\beta = i\zeta(e^{\lambda} - 1)/(ue^{-\lambda} - \zeta^2(1-\kappa)e^{\lambda}). \tag{4.18}$$

From Eqs. (2.21) and (2.19) we can identify the LAKNS potentials which turn out as follows:

$$\begin{aligned}A &= -(1 + 2i\zeta\beta)(1-\kappa)u_x - 2i\zeta(1-\kappa)u(1 + i\zeta\beta)e^{\lambda}\\ &\quad - \beta(2u^2 + u_{xx})e^{-\lambda},\\ B &= 2i\zeta(1-\kappa)u_x - 2\zeta^2(1-\kappa)ue^{\lambda} + (2u^2 + u_{xx})e^{-\lambda},\\ C &= -2(1-\kappa)\beta(1 + i\zeta\beta)u_x + \beta_t\\ &\quad - 2(1-\kappa)u(1 + i\zeta\beta)^2 e^{\lambda} - \beta^2(2u^2 + u_{xx})e^{-\lambda},\\ q &= \zeta^2(1-\kappa)e^{\lambda} - ue^{-\lambda},\\ r &= (1-\kappa)(1 + i\zeta\beta)^2 e^{\lambda} + u\beta^2 e^{-\lambda} + \beta_x,\end{aligned} \tag{4.19}$$

and solutions can be obtained by application of the inverse scattering method.

There is another aspect of Eqs. (1.1) which is useful from the stand point of constructing solutions. Namely, the scale invariance of these equations allows them to be brought into the form of a single nonlinear ordinary differential equation. We shall introduce

$$\begin{aligned}v &= x^2 u,\\ z &= x^3/t,\end{aligned} \tag{4.20}$$

which together with λ from a set of scale-invariant variables. In terms of these quantities Eqs. (1.1) reduce to

$$\begin{aligned}&[27z^3 v''' + z(18v + 24 - z)v' - 12v(2 + v)](6v - z)\\ &\quad - 3\kappa(6v' - 4v)(27z^3 v'' - 18z^2 v' + 18zv) = 0,\end{aligned} \tag{4.21}$$

where prime denotes derivative with respect to z. For $\kappa = 0$ the content of this equation is not different from the corresponding result for the KdV equation.

AKNOWLEDGMENTS

We thank R. Benguria, R. Güven, and F. Lund for interesting conversations. The research reported in this paper has been supported in part by the Turkisch Scientific and Technical Research Council. One of us (M.G) would like to thank Alexander von Humboldt foundation for the grant of a fellowship.

[1] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[2] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1972) [Sov. Phys. JETP **34**, 62 (1972)].
[3] P. D. Lax, Comm. Pure Appl. Math. **21**, 467 (1968).
[4] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973); Stud. Appl. Math. **53**, 249 (1974).
[5] F. Lund and T. Regge, Phys. Rev. D **14**, 1524 (1976); F. Lund, Phys. Rev. Lett. **38**, 1175; F. Lund, Ann. Phys. (N.Y.) **115**, 251 (1978), where the connection between surface theory and completely integrable systems is particularly emphasized.
[6] M. Crampin, F.A.E. Pirani, and D. C. Robinson, Lett. Math. Phys. **2**, 15 (1977) and the references contained therin.
[7] L. Auslander, *Differential Geometry* (Harper and Row, New York, 1967).
[8] R. Sasaki Nuc. Phys. B **154**, 343 (1979).
[9] C. Guichard, Comptes Rendus **130**, 159 (1900).
[10] P. Calapso, Annali di Mat. Ser. III **11**, 201 (1905).
[11] An incomplete list of references to Guichard surfaces following [10] is L. P. Eisenhart, Annali di Mat. Ser. III, **22**, 1 (1914); P. Calapso, Annali. di Mat. Ser. III **29**, 84 (1920); L. P. Einsenhart, *Transformations of Surfaces* (Chelsea, New York, 1962), 2nd ed.; Note the sign mistake in Eq. (22*) of Eisenhart's paper which is carried to his book.
[12] J. Corones, J. Math. Phys. **19**, 341 (1978).

Numerical evaluation of integrals containing a spherical Bessel function by product integration

D. R. Lehman [a] and William C. Parke
Department of Physics, The George Washington University, Washington, D. C. 20052

L. C. Maximon
National Bureau of Standards, Washington, D. C. 20234

(Received 28 October 1980; accepted for publication 6 February 1981)

A method is developed for numerical evaluation of integrals with k-integration range from 0 to ∞ that contain a spherical Bessel function $j_l(kr)$ explicitly. The required quadrature weights are easily calculated and the rate of convergence is rapid—only a relatively small number of quadrature points is needed—for an accurate evaluation even when r is large. The quadrature rule is obtained by the method of product integration. With the abscissas chosen to be those of Clenshaw–Curtis and the Chebyshev polynomials as the interpolating polynomials, quadrature weights are obtained that depend on the spherical Bessel function. An inhomogenous recurrence relation is derived from which the weights can be calculated without accumulation of roundoff error. The procedure is summarized as an easily implementable algorithm. Questions of convergence are discussed and the rate of convergence demonstrated for several test integrals. Alternative procedures are given for generating the integration weights and an error analysis of the method is presented.

PACS numbers: 02.60.Jh

I. INTRODUCTION

A recurring problem in many areas of physics is the evaluation of integrals such as

$$I_l(r) = \int_0^\infty k^2\, dk\, j_l(kr)\phi_l(k), \tag{1}$$

where $j_l(kr)$ is a spherical Bessel function of the first kind.[1] The function $\phi_l(k)$ is usually known only numerically, but is well-behaved, i.e., it has threshold and asymptotic behavior constrained by the physics of the problem. Such integrals arise in nuclear physics when calculating charge-form factors, in atomic and molecular physics when working with variational functionals, in many types of scattering problems, etc. This report derives from our need to evaluate integrals as in Eq. (1) for a study concerning the shell structure of nuclear three-body wave functions for the $A = 6$ bound states.[2] Though integrals of this type do arise frequently in physical problems, there does not seem to be a method published for evaluating them *directly* by numerical methods.[3] Certainly, $j_l(kr)$ can be decomposed as[1]

$$j_l(kr) = u_l(kr)\cos kr + v_l(kr)\sin kr, \tag{2}$$

and then one of the methods for evaluating integrals of the form

$$I'_l(r) = \int_0^\infty k^2\, dk \begin{Bmatrix}\cos kr\\ \sin kr\end{Bmatrix} f_l(kr) \tag{3}$$

can be used, but a great amount of facility is lost by not dealing directly with Eq. (1).[3] The method we propose overcomes the two main difficulties in evaluating numerically integrals like $I(r)$: (1) the infinite range of integration and (2) the highly oscillatory nature of $j_l(kr)$ when r becomes large. The second difficulty in particular thwarts all conventional quadrature rules. Our approach is to apply the *approximate product-integration method* directly to Eq. (1).

The product-integration method is now almost 35 years old[4] and has been applied in a number of situations.[5] One of its earliest applications was in the numerical solution of integral equations by Young[6] and not long ago it was the technique proposed by Kim[7] for solving the three-body Faddeev integral equations with local potentials. Lately, Sloan and Smith,[8] and Sloan[5] have studied the convergence properties, error estimates, and choice of abscissas in applying the product-integration method to various integrals. The power of the method is readily apparent in reviewing its past applications.[5] The essence of the product-integration rule is to approximate an integral as follows:

$$\int_a^b g(k)f(k)\, dk \approx \sum_{i=0}^{N} w_i f(k_i), \tag{4}$$

where the weights w_i are exact if $f(k)$ is any linear combination of a chosen set of functions $\{\varphi_i\}_0^N$. Generally, the integrand contains an "offensive," but analytically known and factorable part, $g(k)$, which is used to generate a set of weights $\{w_i\}_0^N$ that contain it implicitly. One of the important aspects with respect to convergence is the choice of $\{k_i\}_0^N$. In fact, the convergence theorems that have been established[5] can yield practical means for choosing the set of abscissas $\{k_i\}_0^N$. Specifically, it is known[8] that for the Clenshaw–Curtis[9] abscissas, $k_i = \cos i\pi/N$, $0 \leqslant i \leqslant N$, if the weights w_i are chosen to make Eq. (4) (with $b = -a = 1$) exact when $f(k)$ is any polynomial of degree $\leqslant N$, then the sum in Eq. (4) converges to the exact result as $N \to \infty$ for all continuous functions $f(k)$ provided $g(k) \in L_p[-1,1]$ for

[a] Guest Worker, National Bureau of Standards.

 0022-2488/81/071399-15$1.00

some $p > 1$. Furthermore, the sum of the absolute values of the weights converges to the least possible value

$$\lim_{N\to\infty}\sum_{i=0}^{N}|w_i| = \int_{-1}^{1}|g(k)|\,dk. \tag{5}$$

These results provide a solid framework within which to work.

With the functions φ_i, chosen to be the Chebyshev polynomials[10] of the first kind, T_i, our objective is to apply the product-integration method to Eq. (1) and obtain the weights $w_i(r)$ [dependent on $j_l(kr)$] with minimal (physical) assumptions about the threshold and asymptotic behavior of $\phi_l(k)$. By a judicious choice of mapping from $[0,\infty)$, to $[-1,1]$ and explicit removal of the threshold and part or all of the asymptotic behavior from $\phi_l(k)$, we find inherent simplicity in the mathematics in that the w_i can be expressed analytically in terms of known functions. Moreover, the weights are easily computed directly from an inhomogeneous recurrence relation. The strengths of the method will become more apparent later in the paper, but at this stage it is already clear that given the weights w_i, the integration is easily done and convergence is guaranteed. We emphasize in this paper the choice of abscissas, methods for obtaining the weights (together with our "best"choice method), applications to test integrals to study and demonstrate the rate of convergence, and the exact analytical results for the w_i, including sum rules.

The layout of the paper is as follows: Sec. II contains a description of the product-integration method applied to Eq. (1) and mentions the possible methods for calculating the weights. This section also includes brief comments about convergence; Sec. III describes in detail our preferred method for computing the weights w_i, i.e., recurrence for the weights directly; Sec. IV comprises applications of the method to various test integrals; and Sec. V presents a summary and conclusions. Two appendices follow the main text and give alternative methods for computing the weights. In Appendix A, an indirect recurrence method is given including an error analysis. Appendix B holds closed form expressions for the weights.

II. FORMALISM FOR ABSCISSAS AND WEIGHTS

We are interested in integrals of the type

$$I_l(r) = \int_0^\infty k^2\,dk\,j_l(kr)\phi_l(k), \tag{1}$$

where $\phi_l(k)$ comes from a physical problem such that its threshold and asymptotic behavior are

$$\phi_l(k) \underset{k\to 0}{\longrightarrow} C_0^l k^l \tag{6}$$

and

$$\phi_l(k) \underset{k\to 0}{\longrightarrow} C_\infty^l / k^{m+2}, \tag{7}$$

where C_0^l and C_∞^l are constants and $m \geqslant 0$. (The method given below for evaluating Eq. (1) in fact requires only that $\phi_l(k)$ goes to zero equal to or faster than $1/k^2$.) In applying the product-integration method, we map onto the interval $[-1,1]$ by

$$x = \frac{k^2-\alpha^2}{k^2+\alpha^2} \tag{8}$$

and inversely by

$$k = \alpha\left(\frac{1+x}{1-x}\right)^{1/2} \geqslant 0, \tag{9}$$

where α is a scale parameter. With this transformation and with Chebyshev polynomials used to expand $\phi_l[\alpha((1+x)/(1-x))^{1/2}]$, the quadrature weights can be obtained analytically. A simple transformation given below takes the quadrature weights on $[-1,1]$ to those on $[0,\infty)$. Besides the change of variable, we divide out the threshold and part or all of the asymptotic behavior of $\phi_l(k)$, making the result a rather slowly varying function for fitting purposes

$$\frac{k^l}{(k^2+\alpha^2)^{1+(n+l)/2}}\left[\frac{(k^2+\alpha^2)^{1+(n+l)/2}}{k^l}\phi_l(k)\right],$$

where $n \leqslant m$ with m as defined in Eq. (7). After these two steps, Eq. (1) can be written as

$$I_l(r) = \int_{-1}^{1} g_{ln}^{\alpha}(r,x) f_{ln}(x)\,dx, \tag{10}$$

with

$$g_{ln}^{\alpha}(r,x) \equiv \alpha^3 \frac{(1+x)^{(l+1)/2}}{(1-x)^{(3-n)/2}}\, j_l\left[\alpha r\left(\frac{1+x}{1-x}\right)^{1/2}\right], \tag{11}$$

and

$$f_{ln}(x) \equiv \frac{1}{\alpha^{n+2}2^{1+(n+l)/2}}\frac{(k^2+\alpha^2)^{1+(n+l)/2}}{k^l}\phi_l(k). \tag{12}$$

Eqs. (10)–(12) are the starting point for the product-integration method.

The approximate product-integration is defined as

$$I_l(r) \approx I_l^N(r) = \sum_{i=0}^{N} w_i(l,n;\alpha;r) f_{ln}(x_{Ni}), \tag{13}$$

where $N+1$ is the number of quadrature points, w_i are the quadrature weights, and x_{Ni} are the abscissas. In this application, the weights carry knowledge about the oscillatory Bessel function. As the fitting polynomials, the Chebyshev polynomials of the first kind are chosen because of their powerful fitting characteristics, especially to smooth functions like $f_{ln}(x)$. We write

$$f_{ln}(x) \approx \sum_{i=0}^{N} b_{Ni}^{ln} T_i(x), \tag{14}$$

and require that

$$f_{ln}(x_{Nj}) = \sum_{i=0}^{N} b_{Ni}^{ln} T_i(x_{Nj}), \tag{15}$$

where the x_{Nj} are the Clenshaw–Curtis abscissas[9]

$$x_{Nj} = \cos j\pi/N \qquad 0 \leqslant j \leqslant N. \tag{16}$$

The Clenshaw–Curtis abscissas are chosen because they are easy to compute, they permit us to use the discrete orthogonality properties of the Chebyshev polynomials, they introduce economy when N is doubled since the $N+1$ previous evaluations of $f_{ln}(x)$ can be used again, and they permit application of the Sloan–Smith convergence theorems.[8] A drawback in their use is the inclusion of the end points of the

integration although this is easily overcome if it becomes a nuisance (see below and Sec. IV). With this choice,

$$I_l(r) = \int_{-1}^{1} g^{\alpha}_{ln}(r,x) f_{ln}(x)\, dx \approx \sum_{i=0}^{N} b^{\,ln}_{Ni} a_i(l,n;\alpha;r), \tag{17}$$

where

$$a_i(l,n;\alpha;r) = \int_{-1}^{1} g^{\alpha}_{ln}(r,x) T_i(x)\, dx. \tag{18}$$

The $b^{\,ln}_{Ni}$ are obtained from Eq. (15) by use of the discrete orthogonality of the $T_i(x_{Nj}) = \cos ij\pi/N$,

$$\sum_{j=0}^{N} \alpha_{Nj} T_k(x_{Nj}) T_i(x_{Nj}) = N\delta_{ik}/2\alpha_{Ni}, \tag{19}$$

where δ_{ik} is the Kronecker delta and

$$\alpha_{N0} = \alpha_{NN} = \tfrac{1}{2}, \quad \alpha_{Ni} = 1 \quad \text{for} \quad 1 \leqslant i \leqslant N-1, \tag{20}$$

thus,

$$b^{\,ln}_{Ni} = \alpha_{Ni} \frac{2}{N} \sum_{j=0}^{N} \alpha_{Nj} T_i(x_{Nj}) f_{ln}(x_{Nj}). \tag{21}$$

When Eq. (21) is substituted into Eq. (17) and the result compared with Eq. (13), the formula for the weights is obtained:

$$w_i = \alpha_{Ni} \frac{2}{N} \sum_{j=0}^{N} \alpha_{Nj} T_j(x_{Ni}) a_j(l,n;\alpha;r). \tag{22}$$

Clearly, the problem of generating the weights amounts to evaluating the integrals $a_j(l,n;\alpha;r)$, Eq. (18).

To set up the product-integration method, it was natural to work on $[-1,1]$. Nevertheless, the evaluation of the $a_j(l,n;\alpha;r)$ is more transparent on $[0,\infty)$. The a_j integrals become

$$a_j(l,n;\alpha;r) = \alpha^3 2^{1+(l+n)/2} \int_0^{\infty} j_l(\mu k) T_j\left(\frac{k^2-1}{k^2+1}\right) \times \frac{k^{l+2}\, dk}{(k^2+1)^{1+(l+n)/2}}, \tag{23}$$

with $\mu = \alpha r$. In the schemes for computing the a_j, we restrict $(l+n)/2$ to be an integer $\geqslant 0$. For the recurrence scheme outlined in Appendix A, the integral in Eq. (23) is generalized to

$$\mathscr{H}^{l}_{jq}(\mu) = \int_0^{\infty} j_l(\mu k) T_j\left(\frac{k^2-1}{k^2+1}\right) \frac{k^{l+2} dk}{(k^2+1)^{q+1}} \tag{24}$$

and a recurrence scheme worked out for the $\mathscr{H}^{l}_{jq}(\mu)$. The a_j are obtained for $q = (l+n)/2 = \text{integer} \geqslant 0$. Exact expressions for the w_i can be obtained by substitution of Eq. (23) into Eq. (22) and summing over j as described in Appendix B. The latter procedure serves as the starting point for deriving inhomogeneous recurrence relations for the w_i themselves—the method we advocate for practical computation of the w_i with the least loss of accuracy. These weight-recurrence relations are derived in the next section.

The weights on $[-1,1]$ are certainly the quantities to generate by the product-integration scheme presented here, but weights on $[0,\infty)$ are more useful for direct evaluation of Eq. (1). When we define

$$I_l(r) \approx \sum_{i=0}^{N} W_i(l,n;\alpha;r) \phi_l(k_{Ni}) \tag{25}$$

with

$$k_{Ni} = \alpha \cot i\pi/2N, \tag{26}$$

the weights on $[0,\infty)$ are given by

$$W_i(l,n;\alpha;r) = \frac{w_i(l,n;\alpha;r)}{(1-x_{Ni})^{1+n/2}(1+x_{Ni})^{l/2}}. \tag{27}$$

The interval end points do not pose a problem since

$$\lim_{k\to 0} \phi_l(k) \to C^{l}_{0} k^{l} = C^{l}_{0} \alpha^{l} \left(\frac{1+x}{1-x}\right)^{l/2} \tag{28}$$

and

$$\lim_{k\to\infty} \phi_l(k) \to C^{l}_{\infty} k^{-m-2} = C^{l}_{\infty} \alpha^{-m-2} \left(\frac{1-x}{1+x}\right)^{1+m/2}. \tag{29}$$

Moreover, in practice, if $m > n$, then the $i = 0$ term (weight W_0) can be dropped since its contribution is zero. Otherwise, for $m = n$, the $i = 0$ term contributes

$$w_0 C^{l}_{\infty} \alpha^{-n-2} 2^{-1-(l+n)/2}.$$

Similarly, at the origin for $l \neq 0$, the $i = N$ term contributes

$$w_N C^{l}_{0} \alpha^{l} 2^{-1-(l+n)/2}.$$

Now that the scheme has been laid out, it is appropriate to raise questions about convergence; in particular, is the condition of the Sloan–Smith theorem[8] satisfied? The theorem guarantees convergence of $I^{N}_{l}(r)$ to $I_l(r)$ as $N \to \infty$ provided

$$\int_{-1}^{1} |g^{\alpha}_{ln}(r,x)|^{p}\, dx < \infty \tag{30}$$

for some $p > 1$, or applied to the present case,

$$\int_0^{\infty} |j_l(\mu k)|^{p} \frac{k^{p(l+1)+1}\, dk}{(k^2+1)^{p(l+n)/2-p+2}} < \infty. \tag{31}$$

For the l values considered here, $l = 0$, 1, and 2, one can show that Eq. (31) is satisfied as follows:

$$l = 0,2 \quad n \geqslant 2 \quad (n = 2,4,6,\ldots), \qquad l = 1 \quad n \geqslant 1 \quad (n = 1,3,5,\ldots). \tag{32}$$

Since Eq. (31) is a sufficiency statement, it can be violated and convergence still is possible. That convergence can still occur will be demonstrated explicitly in Sec. IV for test integrals in the cases

$$l = 0 \quad n = 0; \qquad l = 1 \quad n = -1; \qquad l = 2 \quad n = -2,0 \tag{33}$$

Therefore, the product-integration method as applied here to integrals of the type given in Eq. (1) is on a sound footing and convergence is guaranteed for a large range of n values. We are now ready to get to the practical matter of computing the w_i.

III. CALCULATION OF THE WEIGHTS

In this section we derive the details of our recommended procedure for calculating the weights w_i defined in Sec. II. The final algorithm for generating the weights is in fact quite simple and straightforward. We summarize it, therefore, at the end of this section for the reader who may wish to

implement it immediately.

From Sec. II, Eqs. (18) and (22), the weights can be expressed as

$$w_i(l,n;\alpha;r) = \alpha_{Ni}\frac{2}{N}\sum_{j=0}^{N}\alpha_{Nj}T_j(x_{Ni})a_i(l,n;\alpha;r),$$

where

$$\alpha_{Ni} = \begin{cases} \frac{1}{2} & i = 0,N, \\ 1 & 0<i<N, \end{cases} \tag{22}$$

$$x_{Ni} = \cos(i\pi/N), \tag{16}$$

$$a_j(l,n;\alpha;r) = \int_{-1}^{1} g_{ln}^{\alpha}(r,x)T_j(x)\,dx. \tag{18}$$

Thus, we have

$$w_i(l,n;\alpha;r) = \alpha_{Ni}\frac{2}{N}\int_{-1}^{1} dx\, g_{ln}^{\alpha}(r,x)\sum_{j=0}^{N}\alpha_{Nj}T_j(x_{Ni})T_j(x). \tag{34}$$

The sum appearing in this last expression may be evaluated in closed form. The derivation is given in Appendix B [Eqs. (1)–(4)]. The result needed here is

$$\sum_{j=0}^{N}\alpha_{Nj}T_j(x_{Ni})T_j(x) = \tfrac{1}{2}(-1)^i\frac{\sin N\phi\,\sin\phi}{\cos(i\pi/N)-\cos\phi}, \tag{35}$$

where

$$x = \cos\phi, \quad 0\leqslant\phi\leqslant\pi. \tag{36}$$

Defining

$$\beta = i\pi/N, \tag{37}$$

$$C = \alpha_{Ni}(-1)^i/N, \tag{38}$$

we then have

$$w_i(l,n;\alpha;r) = C\int_{-1}^{1} dx\, g_{ln}^{\alpha}(r,x)\frac{\sin N\phi\,\sin\phi}{\cos\beta-\cos\phi}. \tag{39}$$

We note here that the integrand has in fact no singularity at $\phi=\beta$ since $\sin N\beta = 0$. However, with the aim of writing a difference equation, we write Eq. (39) in the form

$$w_i(l,n;\alpha;r) = C\int_{-1}^{1} dx\, g_{ln}^{\alpha}(r,x)\frac{\sin N\phi\,\sin\phi-\sin N\beta\,\sin\beta}{\cos\beta-\cos\phi}. \tag{40}$$

We now define the function

$$V_j(\beta) = C\int_{-1}^{1} dx\, g_{ln}^{\alpha}(r,x)\frac{\sin j\phi\,\sin\phi-\sin j\beta\,\sin\beta}{\cos\beta-\cos\phi}, \tag{41}$$

in which β and C retain their previous definitions, i.e., they are functions solely of i and N. The only dependence on the index j is that shown explicitly above in the numerator of the integrand. For $j=N$, the function $V_j(\beta)$ is identical to the desired weights

$$V_N(\beta) = w_i(l,n;\alpha;r). \tag{42}$$

For $j\neq N$, the $V_j(\beta)$ are not equal to the weights for a different N, and, in particular, $\sin j\beta\neq 0$ for $j\neq N$. The integral as defined above, has, however, no singularity at $\phi=\beta$ for any j (which is the reason for making the subtraction in the numerator of the integrand).

We can now write a difference equation for the $V_j(\beta)$. Using

$$\sin(j+1)\phi + \sin(j-1)\phi = 2\sin j\phi\cos\phi$$

(and similarly with ϕ replaced by β), we obtain the recurrence relation which constitutes the basis for our calculation of the weights

$$V_{j+1}(\beta) - 2\cos\beta V_j(\beta) + V_{j-1}(\beta) = G_j, \tag{43}$$

where

$$G_j = -2C\int_{-1}^{1} dx\, g_{ln}^{\alpha}(r,x)\sin j\phi\sin\phi. \tag{44}$$

Equation (43) is a second-order, linear, inhomogeneous difference equation with constant coefficients (β does not depend on j). Note that $V_j(\beta)$ is defined for all j, although w_i was defined only for $N>0$.

To implement the recurrence relation, we will need

(a) The initial values $V_0(\beta)$ and $V_1(\beta)$;

(b) A means of calculating the G_j.

[In order to obtain $V_N(\beta)$ we need G_j only for $0\leqslant j\leqslant N-1$]. The form of the difference equation itself [Eq. (43)] is independent of the particular kernel $g_{ln}^{\alpha}(r,x)$. The values obtained for the functions $V_j(\beta)$ and for the weights $w_i(l,n;\alpha;r)$ depend on the kernel via the values $V_0(\beta)$, $V_1(\beta)$, and the set of inhomogeneous terms, $\{G_j\}_0^{N-1}$.

Since the difference equation has constant coefficients, we can in fact write the solution directly, as a sum. However, for computational purposes, it is simpler and quicker to compute $V_N(\beta)$ from the recurrence relation. We therefore leave details of the solution to the difference equation for Appendix B and proceed to the evaluation of the initial value $V_1(\beta)$, after noting that $V_0(\beta)=0$.

From the expression (41) for the $V_j(\beta)$, we have

$$V_1(\beta) = C\int_{-1}^{1} dx\, g_{ln}^{\alpha}(r,x)(\cos\beta+\cos\phi). \tag{45}$$

To proceed further we must specify the function g_{ln}^{α}. From Eq. (11),

$$g_{ln}^{\alpha}(r,x)\equiv\alpha^3\frac{(1+x)^{(l+1)/2}}{(1-x)^{(3-n)/2}}j_l\left[\alpha r\left(\frac{1+x}{1-x}\right)^{1/2}\right]. \tag{11}$$

We now change back from the variable x to the variable k. Using Eq. (8) and (36), we have

$$x = \frac{k^2-\alpha^2}{k^2+\alpha^2} = \cos\phi, \quad 0\leqslant k<\infty \tag{46}$$

$$(1-x^2)^{1/2} = \frac{2k\alpha}{k^2+\alpha^2} = \sin\phi, \tag{47}$$

giving

$$g_{ln}^{\alpha}(r,x)dx = \frac{2^{q+1}\alpha^{n+2}k^{l+2}}{(k^2+\alpha^2)^{q+1}}j_l(kr)dk, \tag{48}$$

where $q\equiv(l+n)/2$. We assume further that n is always chosen so that q is a non-negative integer. Thus we have

$$V_1(\beta) = C\cdot 2^{q+1}\alpha^{n+2}\int_0^{\infty}\frac{k^{l+2}j_l(kr)}{(k^2+\alpha^2)^{q+1}}\times\left(1+\cos\beta-\frac{2\alpha^2}{k^2+\alpha^2}\right)dk. \tag{49}$$

As shown in Appendix A [Eqs. (A2) to (A18)], this integral may be evaluated in terms of the Bessel function of the third kind $k_l(z)$. The result is

$$V_1(\beta) = 2C\alpha^3 \frac{(\alpha r)^q}{q!}\left((1+\cos\beta)k_{l-q}(\alpha r) - \frac{\alpha r}{q+1}k_{l-q-1}(\alpha r)\right). \quad (50)$$

The functions $k_l(z)$ appearing here may be calculated quite easily from the recursion [Eq. (A14)] in the form

$$k_{l+1}(z) = \frac{(2l+1)}{z}k_l(z) + k_{l-1}(z) \quad (51)$$

and the starting values

$$k_{-1}(z) = k_0(z) = (\pi/2z)e^{-z}. \quad (52)$$

Next we consider the calculation of the inhomogeneous term G_j. From Eqs. (44), (47), and (48), we have

$$G_j = -C2^{q+3}\alpha^{n+3}\int_0^\infty \frac{k^{l+3}j_l(kr)}{(k^2+\alpha^2)^{q+2}}\sin j\phi\, dk. \quad (53)$$

Using the raising operator for the functions $j_l(kr)$, we write[11]

$$(l/r - \partial/\partial r)j_l(kr) = kj_{l+1}(kr), \quad (54)$$

and then

$$k^l j_l(kr) = \mathscr{T}_l j_0(kr) = \mathscr{T}_l\left\{\frac{\sin kr}{kr}\right\}, \quad (55)$$

where

$$\mathscr{T}_l \equiv \left(\frac{l-1}{r} - \frac{\partial}{\partial r}\right)\left(\frac{l-2}{r} - \frac{\partial}{\partial r}\right)\cdots\left(-\frac{\partial}{\partial r}\right) \quad l>0, \quad (56)$$

with $\mathscr{T}_0 \equiv 1$. Next, we use

$$e^{i\phi} = (k+i\alpha)/(k-i\alpha) \quad (57)$$

from Eqs. (46) and (47), and thus substitute, in Eq. (53),

$$\sin j\phi = \frac{1}{2i}\left[\left(\frac{k+i\alpha}{k-i\alpha}\right)^j - \left(\frac{k+i\alpha}{k-i\alpha}\right)^{-j}\right]. \quad (58)$$

Finally, we obtain

$$G_j = C2^{q+1}\alpha^{n+3}\mathscr{T}_l\left\{\frac{1}{r}J\right\}, \quad (59)$$

where

$$J = \int_{-\infty}^{\infty} e^{ikr}\frac{(k+i\alpha)^{j-q-1}}{(k-i\alpha)^{j+q+1}}dk - \alpha^2\int_{-\infty}^{\infty} e^{ikr}\frac{(k+i\alpha)^{j-q-2}}{(k-i\alpha)^{j+q+2}}dk - \int_{-\infty}^{\infty} e^{ikr}\frac{(k+i\alpha)^{-j-q-1}}{(k-i\alpha)^{-j+q+1}}dk + \alpha^2\int_{-\infty}^{\infty} e^{ikr}\frac{(k+i\alpha)^{-j-q-2}}{(k-i\alpha)^{-j+q+2}}dk. \quad (60)$$

The integrals appearing here may be expressed in terms of terminating confluent hypergeometric functions. The derivation is given in Appendix B [Eqs. (B16) to (B22)]. The results needed for the integrals above are

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} e^{ikr}(k+i\alpha)^n(k-i\alpha)^m\, dk = 0 \quad (61)$$

for $m,n = 0,1,2...$

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} e^{ikr}\frac{(k+i\alpha)^n}{(k-i\alpha)^{m+1}}dk = \frac{e^{-\alpha r}(ir)^{m-n}}{(m-n)!}{}_1F_1(-n;m-n+1;2\alpha r) \quad (62)$$

for $0\leqslant n\leqslant m$, and

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} e^{ikr}\frac{dk}{(k+i\alpha)^{n+1}(k-i\alpha)^{m+1}} = \frac{e^{-\alpha r}(-1)^m(n+m)!}{(2i\alpha)^{m+n+1}m!n!}{}_1F_1(-m;-n-m;2\alpha r) \quad (63)$$

for $m, n = 0,1,2,...$ where ${}_1F_1(-m;-m;2\alpha r) \equiv \Sigma_{j=0}^m (2\alpha r)^j/j!$. In evaluating the expressions (60) for J, we must distinguish three regions:

(1) $j \geqslant q+2$

(2) $j = q+1$

(3) $j \leqslant q$.

(1) $j \geqslant q+2$: In this case, the last two integrals in the expression (60) for J are zero from Eq. (61). From Eqs. (59), (60), and (62) we have

$$G_j = -C\pi\alpha^{n+3}(-1)^q 2^{q+2}\mathscr{T}_l\left[\left(\frac{e^{-\alpha r}}{r}\right)\times\left\{\frac{1}{(2q+1)!}r^{2q+1}{}_1F_1(-j+q+1;2q+2;2\alpha r) + \frac{\alpha^2}{(2q+3)!}r^{2q+3}{}_1F_1(-j+q+1;2q+4;2\alpha r)\right\}\right]. \quad (64)$$

(2) $j = q+1$: In this case, we have, from Eqs. (59), (60), (62), and (63),

$$G_j = -C\pi\alpha^{n+3}(-1)^q 2^{q+2}\mathscr{T}_l\left[\left(\frac{e^{-\alpha r}}{r}\right)\times\left\{\frac{1}{(2q+1)!}r^{2q+1} + \frac{\alpha^2}{(2\alpha)^{2q+3}}[1 - {}_1F_1(-(2q+2);-(2q+2);2\alpha r)]\right\}\right]. \quad (65)$$

(3) $j \leqslant q$ (needed only if $q>0$): In this case, we have, from Eqs. (59), (60), and (63),

$$G_j = \frac{-C\pi\alpha^{n+3}(-1)^j 2^{q+2}}{(2\alpha)^{2q+1}}\mathscr{T}_l\left[\left(\frac{e^{-\alpha r}}{r}\right)\times\left\{\frac{(2q)!}{(q-j)!(q+j)!}[{}_1F_1(-q+j;-2q;2\alpha r) - {}_1F_1(-q-j;-2q;2\alpha r)] - \tfrac{1}{4}\frac{(2q+2)!}{(q-j+1)!(q+j+1)!}\times[{}_1F_1(-q+j-1;-2q-2;2\alpha r) - {}_1F_1(-q-j-1;-2q-2;2\alpha r)]\}\right]. \quad (66)$$

The equations (64), (65), and (66) express G_j in terms of terminating confluent hypergeometric functions of the form ${}_1F_1(-j;m;2\alpha r)$ in which $j\geqslant 0$ and either $m>0$ or $m\leqslant -j$. These functions may be computed easily from the recurrence relation[12]

$$(c-a)F(a-1;c;x) + (2a-c+x)F(a;c;x) - aF(a+1;c;x) = 0. \quad (67)$$

Writing

$$F_j \equiv {}_1F_1(-j;m;x), \tag{68}$$

we have

$$(m+j)F_{j+1} = (2j+m-x)F_j - jF_{j-1} \tag{69}$$

and initial values

$$F_0 = 1,$$
$$F_1 = 1 - x/m \tag{70}$$

for $m \neq 0$ and ${}_1F_1(0;0;2\alpha r) \equiv 1$. Note, from the asymptotic expansion for the confluent hypergeometric function[13] ${}_1F_1(-j;m;x)$ with m and x fixed and $j \to \infty$,

$$F_j \sim \frac{(m-1)!e^{\alpha r}}{\sqrt{\pi}\,(2\alpha r)^{(m/2)-(1/4)}} \cos\left[2(2\alpha r j)^{1/2} - \frac{\pi}{2}(m-\tfrac{1}{2})\right], \tag{71}$$

that for large j, the functions F_j are oscillatory and slowly decreasing. Thus computer implementation of this recursion for the F_j will encounter no problems due to the buildup of round off errors.

Although we have now presented all of the expressions which are essential for the calculation of the weights, we give here the expressions for the functions G_j for the cases $l = 0$, 1, and 2 explicitly. [For higher l values, one may use the general expression for G_j in Eq. (65) after performing the indicated differentiations, or one may use a direct recurrence relation on the weights for raising the l value. This relation is derived in Appendix B.] On taking the derivatives indicated in the operator $\mathcal{T}_l$, we clearly have several choices for expressing the derivatives of the confluent hypergeometric function in terms of similar functions. Referring to the expressions (64), (65), and (66) for G_j, we use both[14]

$$\frac{d}{dx}[x^{c-1}F(a;c;x)] = (c-1)x^{c-2}F(a;c-1;x) \tag{72}$$

(provided $c \neq 1$) and

$$\frac{d}{dx}F(a;c;x) = \frac{a}{c}F(a+1;c+1;x) \tag{73}$$

(provided $c \neq -1$), depending on which provides a result which has a simpler form. We indicate the specific l value as a superscript $G_j^{(l)} = G_j$. Defining the common factor that appears throughout by

$$\mathcal{N} \equiv -C \cdot 4\pi\alpha^3 e^{-\alpha r}(\alpha r)^n 2^q, \tag{74}$$

we have for $j \geq q+2$

$$G_j^{(0)} = \mathcal{N}(-1)^q\left\{\frac{1}{(2q+1)!}F(-j+q+1;2q+2;2\alpha r) + \frac{(\alpha r)^2}{(2q+3)!}F(-j+q+2;2q+4;2\alpha r)\right\}, \tag{75}$$

$$G_j^{(1)} = \mathcal{N}(-1)^q\left\{\frac{(\alpha r+1)}{(2q+1)!}F(-j+q+1;2q+2;2\alpha r) - \frac{1}{(2q)!}F(-j+q+1;2q+1;2\alpha r) + \frac{(\alpha r)^2(\alpha r+1)}{(2q+3)!}F(-j+q+2;2q+4;2\alpha r) - \frac{(\alpha r)^2}{(2q+2)!}F(-j+q+2;2q+3;2\alpha r)\right\}, \tag{76}$$

$$G_j^{(2)} = \mathcal{N}(-1)^q\left\{\frac{((\alpha r)^2+3\alpha r+3)}{(2q+1)!}F(-j+q+1;2q+2;2\alpha r) + \frac{(2q-3)}{(2q)!}F(-j+q+1;2q+1;2\alpha r) - (q+j)\frac{2\alpha r}{(2q+1)!}F(-j+q+1;2q+2;2\alpha r) + (\alpha r)^2\frac{((\alpha r)^2+3\alpha r+3)}{(2q+3)!}F(-j+q+2;2q+4;2\alpha r) - (\alpha r)^2\frac{(2\alpha r+3)}{(2q+2)!}F(-j+q+2;2q+3;2\alpha r) + \frac{(\alpha r)^2}{(2q+1)!}F(-j+q+2;2q+2;2\alpha r)\right\}. \tag{77}$$

For $j = q+1$, we have

$$G_j^{(0)} = \mathcal{N}(-1)^q\left\{\frac{1}{(2q+1)!} + \tfrac{1}{4}\frac{1}{(2\alpha r)^{2q+1}}[1 - F(-2q-2;-2q-2;2\alpha r)]\right\}, \tag{78}$$

$$G_j^{(1)} = \mathcal{N}(-1)^q\left\{\frac{(\alpha r-2q)}{(2q+1)!} + \tfrac{1}{4}\frac{(\alpha r+1)}{(2\alpha r)^{2q+1}}[1 - F(-2q-2;-2q-2;2\alpha r)] + \tfrac{1}{4}\frac{1}{(2\alpha r)^{2q}}F(-2q-1;-2q-1;2\alpha r)\right\}, \tag{79}$$

$$G_j^{(2)} = \mathcal{N}(-1)^q\left\{\frac{[(\alpha r)^2+(1-4q)\alpha r+4q(q-1)]}{(2q+1)!} + \tfrac{1}{4}\frac{1}{(2\alpha r)^{2q+1}}[[(\alpha r)^2+3\alpha r+3] \times [1 - F(-2q-2;-2q-2;2\alpha r)] + 2\alpha r(2\alpha r+3)F(-2q-1;-2q-1;2\alpha r) - 4(\alpha r)^2F(-2q;-2q;2\alpha r)]\}. \tag{80}$$

For $j \leq q$, we have

$$G_j^{(0)} = \mathcal{N}\frac{(-1)^j}{(2\alpha r)^{2q+1}}\left\{\frac{(2q)!}{(q-j)!(q+j)!}[F(-q+j;-2q;2\alpha r) - F(-q-j;-2q;2\alpha r)] - \tfrac{1}{4}\frac{(2q+2)!}{(q-j+1)!(q+j+1)!} \times [F(-q+j-1;-2q-2;2\alpha r) - F(-q-j-1;-2q-2;2\alpha r)]\}, \tag{81}$$

$$G_j^{(1)} = \mathcal{N}\frac{(-1)^j}{(2\alpha r)^{2q+1}}\left\{\frac{(\alpha r+1)(2q)!}{(q-j)!(q+j)!}[F(-q+j;-2q;2\alpha r) - F(-q-j;-2q;2\alpha r)] - \frac{(\alpha r+1)(2q+2)!}{4(q-j+1)!(q+j+1)!} \times [F(-q+j-1;-2q-2;2\alpha r) - F(-q-j-1;-2q-2;2\alpha r)] - \frac{2\alpha r(2q-1)!}{(q-j)!(q+j)!}[(q-j)F(-q+j+1;-2q+1;2\alpha r) - (q+j)F(-q-j+1;-2q+1;2\alpha r)]$$

$$+\tfrac{1}{4}\frac{(2\alpha r)(2q+2)!}{(q-j+1)!(q+j+1)!}$$
$$\times[(q-j+1)F(-q+j;-2q-1;2\alpha r)$$
$$-(q+j+1)F(-q-j;-2q-1;2\alpha r)]\},\qquad(82)$$

$$G_j^{(2)}=\frac{\mathcal{N}(-1)^j}{(2\alpha r)^{2q+1}}\Bigg\{\frac{[(\alpha r)^2+3\alpha r+3](2q)!}{(q-j)!(q+j)!}$$
$$\times[F(-q+j;-2q;2\alpha r)$$
$$-F(-q-j;-2q;2\alpha r)]$$
$$+\frac{2\alpha r(2\alpha r+3)(2q-1)!}{(q-j)!(q+j)!}$$
$$\times[(q-j)F(-q+j+1;-2q+1;2\alpha r)$$
$$-(q+j)F(-q-j+1;-2q+1;2\alpha r)]$$
$$+\frac{4(\alpha r)^2(2q-2)!}{(q-j)!(q+j)!}$$
$$\times[(q-j)(q-j-1)F(-q+j+2;-2q+2;2\alpha r)$$
$$-(q+j)(q+j-1)F(-q-j+2;-2q+2;2\alpha r)]$$
$$-\frac{[(\alpha r)^2+3\alpha r+3](2q+2)!}{4(q-j+1)!(q+j+1)!}$$
$$\times[F(-q+j-1;-2q-2;2\alpha r)$$
$$-F(-q-j-1;-2q-2;2\alpha r)]$$
$$+\frac{2\alpha r(2\alpha r+3)(2q+1)!}{4(q-j+1)!(q+j+1)!}[(q-j+1)$$
$$\times F(-q+j;-2q-1;2\alpha r)$$
$$-(q+j+1)F(-q-j;-2q-1;2\alpha r)]$$
$$-\frac{(2\alpha r)^2(2q)!}{4(q-j+1)!(q+j+1)!}[(q-j+1)(q-j)$$
$$\times F(-q+j+1;-2q;2\alpha r)$$
$$-(q+j+1)(q+j)F(-q-j+1;-2q;2\alpha r)]\}.\qquad(83)$$

A. Summary of algorithm

We now give a brief resume of our algorithm for calculating the weights.

(1) The functions F_j, defined in Eq. (68), are calculated from the recursion (69) and the initial values (70). For each of the four positive values of m: $m=2q+1$; $2q+2$; $2q+3$; and $2q+4$, they are calculated for $j=0,1,\ldots,N-q-2$. For $m=0,-1,-2,\cdots,-2q-2$, they are calculated for $j=0,1,2,\cdots,|m|$.

(2) The factorials, $m!$, are computed by recursion for $m=0,1,2,\ldots,2q+3$.

(3) The functions G_j are computed for $l=0,1$, and 2 from the functions F_j using Eqs. (75) to (83) for $j=1,2,\ldots,N-1$.

(4) The functions k_l are computed from the recursion relations and starting values given by Eqs. (51) and (52) for $l=-1,0,1,2,\ldots,|q-1|+1$.

(5) For each value of i, $i=0,1,2,\ldots,N$, the starting value $V_1(\beta)$ is computed from the functions k_l using Eq. (50) and $V_0(\beta)=0$, with $\beta=i\pi/N$ as defined in Eq. (37).

(6) The functions $V_j(\beta)$ are computed for $j=0,1,\ldots,N$ from the recursion of Eq. (43) and the starting values $V_0(\beta)$ and $V_1(\beta)$.

(7) The weights w_i for each i are equal to $V_N(\beta)$ (still with $\beta=i\pi/N$).

IV. APPLICATIONS

Now we apply the purposed product-integration method to some test integrals. Our purpose is to examine the method's rate of convergence as N (the number of abscissas) increases. Under the conditions mentioned in Secs. I and II, we are guaranteed that the method will converge, but the rate of convergence to a given accuracy will depend on the given value of r and the choices made for the scale-transformation parameter α along with the asymptotic-behavior index n. Specific emphasis is given to the role of α and how to choose it. In this development, the test integrals used involve $j_l(kr)$ for $0\leqslant l\leqslant 2$, but it is clear that this is not a limitation. The weights and abscissas needed were generated using the algorithm given at the end of Sec. III. Double-precision arithmetic ($\sim$15 significant figures) was used throughout. A sample set of weights and abscissas is given in Table I.

The sample set of weights in Table I can be used to illustrate constraints they must satisfy due to the fact that the integration rule is exact for

$$\phi_l(k)=k^l/(k^2+\beta^2)^p\quad(p\geqslant 1+l/2),\qquad(84)$$

when α and N are chosen such that $\alpha=\beta$ and $N\geqslant p-1-(n+l)2\geqslant 0$. The exactness of the integration rule under the previous conditions is equivalent to the fact that Eq. (14) is an identity for all x under these conditions and not just for the Clenshaw–Curtis abscissas

$$f_{ln}(x)\equiv\sum_{i=0}^{N}{}''\, b_{Ni}^{ln}T_i(x).\qquad(14')$$

In fact, the b_{Ni}^{ln} can be found analytically

$$b_{No}^{ln}=\frac{\alpha^l 2^{1+(n+l)/2}}{(2\alpha)^{2p}}\binom{2[p-1-(n+l)/2]}{p-1-(n+l)/2},$$
$$p\geqslant 1+(n+l)/2\qquad(85)$$

$$b_{Ni}^{ln}\equiv 0\quad p=1+(n+l)/2,\quad i>0\qquad(86)$$

$$b_{Ni}^{ln}=\frac{\alpha^l 2^{2+(n+l)/2}}{(2\alpha)^{2p}}(-1)^i\binom{2[p-1-(n+l)/2]}{p-1-\frac{(n+l)}{2}-i},$$
$$p>1+\frac{(n+l)}{2},\quad i>0.\qquad(87)$$

TABLE I. Sample set of abscissas and weights ($N=8$, $l=0$, $r=1.0$, $\alpha=0.5$, $n=0$).

x_{Ni}	$w_i(\ell,n;\alpha;r)$	k_{Ni}	$W_i(\ell,n;\alpha;r)$
-1.0000 0000	0.2894 9668 (-2)[a]	0.0000 0000	0.1447 4834 (-2)
-0.9238 7953	-0.4084 8926 (-2)	0.9945 6184 (-1)	-0.2123 2580 (-2)
-0.7071 0678	0.1516 8968 (-1)	0.2071 0678	0.8885 7755 (-2)
-0.3826 8343	0.1306 0362 (-1)	0.3340 8932	0.9445 6633 (-2)
-0.0000 0000	0.5887 3310 (-1)	0.5000 0000	0.5887 3310 (-1)
0.3826 8343	0.8044 1714 (-1)	0.7483 0288	0.1303 0869
0.7071 0678	0.2645 0971	1.2071 0678	0.9030 9265
0.9238 7953	0.1352 4790	2.5136 6975	1.7767 6126
1.0000 0000	-0.8974 3972 (-1)	∞	b

[a] (-2) $\equiv 10^{-2}$, etc.

[b] Whether this weight enters and its contribution depends on the asymptotic behavior of ϕ_ℓ -- see Eqs. (27) and (29) and the discussion immediately following those equations.

TABLE II. Results for $I_0^t(r)$.

	r = 1.0 Exact = 0.4502 2246 2499 8651		
	α = 0.2 n = 0	α = 0.5 n = 0	α = 1.0 n = 0
N = 2	0.13	0.30	0.37
4	0.30	0.449	0.57
8	0.433	0.4507 6	0.442
16	0.4523	0.4502 218	0.4503 5
32	0.4502 213	0.4502 2246 250	0.4502 2244 0
40	0.4502 213	0.4502 2246 2499 853	0.4502 2246 34
		α = 0.5 n = 2	
N = 2		0.52	
4		0.447	
8		0.4501 2	
16		0.4502 2255	
32		0.4502 2246 2499 68	
40		0.4502 2246 2499 858	
	$r = 10^{-4}$ Exact = 0.9624 1921 3196 6023		
		α = 0.5 n = 0	
N = 2		0.41	
8		0.957	
16		0.9624 06	
32		0.9624 1921 300	
40		0.9624 1921 303	
	r = 15.0 Exact = 0.1692 4693 5215 6365 (-2)		
		α = 0.5 n = 0	
N = 2		0.94 (-3)	
4		0.195 (-2)	
8		0.1685 (-2)	
16		0.1692 493 (-2)	
32		0.1692 4693 5237 (-2)	
40		0.1692 4693 5215 659 (-2)	
	r = 50.0 Exact = 0.1510 6744 5473 8440 (-6)		
		α = 0.5 n = 0	
N = 2		0.66 (-10)	
4		0.15 (-8)	
8		0.44 (-7)	
16		0.1510 31 (-6)	
32		0.1510 6748 0 (-6)	
40		0.1510 6744 539 (-6)	

As simple examples, let $N = 8$, $l = 0$, $n = 0$, $\alpha = 0.5$ and $r = 1$. Then, for $p = 1$, we have

$$b_{Ni}^{00} = \frac{\delta_{i0}}{2\alpha^2}, \tag{88}$$

$$I_0(r) \equiv I_0^N(r) = \frac{1}{2\alpha^2} a_0(0,0;\alpha;r) \tag{89}$$

$$= \alpha \mathscr{H}_{00}^0(\alpha r) \tag{90}$$

$$= \frac{\pi e^{-\alpha r}}{2r} \tag{91}$$

$$= \frac{1}{2\alpha^2} \sum_{i=0}^{N} w_i(0,0;\alpha;r), \tag{92}$$

and for $p = 2$,

$$I_0(r) = \frac{1}{4\alpha^4}[a_0(0,0;\alpha;r) - a_1(0,0;\alpha;r)] \tag{93}$$

$$= \frac{1}{2\alpha}[\mathscr{H}_{00}^0(\alpha r) - \mathscr{H}_{10}^0(\alpha r)] \tag{94}$$

$$= \frac{\pi e^{-\alpha r}}{4\alpha} \tag{95}$$

$$= \frac{1}{2\alpha^4} \sum_{i=0}^{N} \sin^2(i\pi/2N) w_i(0,0;\alpha;r). \tag{96}$$

Obviously, the last two equalities for each case are sum rules for the weights and can be generalized [see Appendix B, Eqs. (B41)–(B44)]. By use of Table I, we find

$$\left| \frac{1}{2\alpha^2} \sum_{i=0}^{N} w_i(0,0;\alpha;r) - \frac{\pi e^{-\alpha r}}{2r} \right| < 3 \times 10^{-10} \tag{97}$$

and

$$\left| \frac{1}{2\alpha^4} \sum_{i=o}^{N} \sin^2\left(\frac{i\pi}{2N}\right) w_i(0,0;\alpha;r) - \frac{\pi e^{-\alpha r}}{4\alpha} \right| < 5 \times 10^{-9}, \tag{98}$$

a nice check on the consistency of the weights. At the same time, integrands like Eq. (84) can be used as a check in computer codes since they yield results accurate to the precision of the computation. To test the rate of convergence of our method, we must choose integrands more complicated than Eq. (84).

As our three basic test integrals, we choose

$$I_0^t(r) = \int_0^\infty k^2\, dk\, j_0(kr) \frac{1}{(k^2 + a^2)(k^2 + b^2)}, \tag{99}$$

$$I_1^t(r) = \int_0^\infty k^2\, dk\, j_1(kr) \frac{k}{(k^2 + a^2)^2(k^2 + b^2)}, \tag{100}$$

and

$$I_2^t(r) = \int_0^\infty k^2\, dk\, j_2(kr) \frac{k^2}{(k^2 + a^2)^2(k^2 + b^2)}, \tag{101}$$

all of which can be evaluated in closed form (the superscript t represents "test"). Specifically, we chose $a = 0.232$ and $b = 1.4$. Results for various tests with these integrals are given in Tables II–IV.

The first point we note in scanning Tables II–IV for fixed α and n is that the rate of convergence to a result accurate to ~ 10 significant figures is not greatly affected by the value of r for a wide range of r. Especially notable are the

TABLE III. Results for $I_1^t(r)$.

	r = 1.0 Exact = 0.1602 4405 8201 8119		
	α = 0.2	α = 0.4	α = 0.6
N = 8	0.20	0.1606	0.23
14	0.1629	0.1601 7	0.1604 5
20	0.1598 0	0.1602 431	0.1602 36
30	0.1601 9	0.1602 4406 4	0.1602 4405 93
40	0.1602 428	0.1602 4405 8207	0.1602 4405 8202 7
	r = 15.0 Exact = 0.1242 8338 1560 5883 (-1)		
	α = 0.2	α = 0.4	α = 0.6
N = 8	0.1230 (-1)	0.1243 0 (-1)	0.1179 (-1)
14	0.1242 41 (-1)	0.1242 79 (-1)	0.1241 4 (-1)
20	0.1243 3 (-1)	0.1242 8347 (-1)	0.1242 824 (-1)
30	0.1242 79 (-1)	0.1242 8338 118 (-1)	0.1242 8338 24 (-1)
40	0.1242 8329 (-1)	0.1242 8338 1560 29 (-1)	0.1242 8338 1560 569 (-1)

TABLE IV. Results for $I_2^{\mathrm{t}}(r)$.

r = 1.0 n = -2 Exact = 0.9359 5083 9114 4211 (-1)						
	α = 0.2		α = 0.4		α = 0.6	
N = 8	0.20	(0)	0.989	(-1)	0.11	(0)
14	0.11	(0)	0.9344	(-1)	0.974	(-1)
20	0.950	(-1)	0.9358 5	(-1)	0.9358 6	(-1)
30	0.9344	(-1)	0.9359 5090	(-1)	0.9359 5069	(-1)
40	0.9358 1	(-1)	0.9359 5084 0	(-1)	0.9359 5083 9178	(-1)
r = 15.0 N = 30 Exact = 0.3690 3053 8135 4993 (-2)						
	α = 0.2		α = 0.6			
n = -2	0.3689 8	(-2)	0.3690 3054 7	(-2)		
0	0.3691 3	(-2)	0.3690 3053 92	(-2)		

TABLE V. Results for $I_0^{\mathrm{st}}(r,r')$.

r' = 0.1 r = 1.0 Exact = 0.9491 6163 7440			
	α = 0.2 n = 0	α = 0.5 n = 0	α = 1.0 n = 0
N = 2	0.3	0.951	0.74
4	0.78	0.9490	1.04
8	0.9475	0.9482	0.9486
16	0.9482	0.9496	0.9489 2
32	0.9492 4	0.9489 1	0.9492 1
40	0.9496 0	0.9490 2	0.9491 2
		α = 0.5 n = 2	
N = 2		0.952	
4		0.9489	
8		0.9471	
16		0.9507	
32		0.9475	
40		0.9483	
r' = 1.0 r = 1.0 Exact = 0.5778 6367 4895			
		α = 0.5 n = 0	
N = 2		0.81	
4		0.62	
8		0.591	
16		0.5726	
32		0.5761	
64		0.5797	
128		0.5785 1	
256		0.5781 5	

$l = 0$ results when $r = 50$. Moreover, computations with $N \cong 20$ yield results accurate to $\gtrsim 5$ or 6 significant figures which is quite adequate for many applications. It should also be noted that most of these results are for n values outside the convergence sufficiency condition of Sloan and Smith as mentioned in the Introduction [see Eqs. (32) and (33)].

The second aspect to observe is that removing the asymptotic behavior from $\phi_l(k)$ by an appropriate choice of n does improve covergence. In the two examples given, ($l = 0$, $n = 2$, $r = 1$) and ($l = 2$, $n = 0$, $r = 15.0$), this is readily apparent. It occurs because the function fitted with the Chebyshev polynomials becomes more slowly varying, if not essentially flat, and is thus easily represented by fewer polynomials.

Perhaps more outstanding in Tables II–IV with respect to convergence is the role played by the scale parameter α. The choice of α for $N = 40$ can mean the difference between a result accurate to 5 or 6 significant figures as opposed to 13 or 14 significant figures. This is because α controls the distribution and range of abscissas. Half of the abscissas (k_{Ni}) lie below α and half above. Thus, for the test integrals, it is nearly optimal to choose α at the maximum of the function $k^2\phi_l^{\mathrm{t}}(k) \geqslant 0$, since then the integrand is sampled equally on either side. For example, $k^2\phi_0^{\mathrm{t}}(k)$ has its maximum at $k = 0.57$ and $\alpha = 0.5$ leads to excellent convergence. Also, $k^2\phi_1^{\mathrm{t}}(k)$ has its maximum at $k = 0.37$, and $\alpha = 0.4$ or 0.6 work very well. Since the distribution of abscissas, k_{Ni}, is proportional to α, the last finite abscissa, $k_{N,N-1}$, is 50% larger when $\alpha = 0.6$ than when $\alpha = 0.4$. Clearly, the scale parameter α plays a critical role in the rate of convergence, or said dfferently, its choice determines the minimal number of terms in the Chebyshev fit that will be necessary to obtain a given accuracy in representing the integrand $\phi_l(k)$.

Finally, we push the method to demonstrate its limitations. The special test (st) integral chosen is

$$I_0^{\mathrm{st}}(r,r') = \int_0^\infty k^2\,dk\, j_0(kr)\,\frac{\cos kr'}{(k^2+a^2)^2}, \tag{102}$$

where $a = 0.5$. As can be seen in Table V, the method is not as powerful when $\phi_0(k)$ is highly oscillatory in nature. Choosing $n = 2$ does not help at all; in fact, it makes matters worse since as $k \to \infty$ $k^4\phi_0(k) \to \cos kr' \neq$ const. In such circumstances, it should be kept in mind that many identities exist between spherical Bessel functions and trigonometric functions. In our case, we can use

$$j_0(kr)\cos kr' = (1 + r'/r)j_0(k\,[r + r']) + (1 - r'/r)j_0(k\,[r - r']) \tag{103}$$

and apply the product-integration method to each term on the right-hand side [Note: with $\alpha = a$, $n = 0$, and $N \geqslant 1$, the integration rule is exact for each term and $I_0^{\mathrm{st}}(r,r')$ can be obtained accurate to the precision of the computation.]

V. SUMMARY AND CONCLUSION

A method has been described for numerically evaluating integrals from 0 to ∞ that contain a spherical Bessel function, $j_l(kr)$, by the method of product-integration. The integration is approximated by a summation over terms that are products of a weight and the factored integrand [$j_l(kr)$ removed] evaluated at the appropriate abscissas. The weights contain the spherical Bessel function implicitly. Generation of the weights is achieved by choosing for the quadrature-rule abscissas the Clenshaw–Curtis points and expanding the integrand in terms of Chebyshev polynomials. The combination of Clenshaw–Curtis abscissas with Chebyshev polynomials makes the mathematics elegant and permits derivation of an inhomogeneous recurrence relation that is easily solved for the weights without roundoff errors. Several test integrals are evaluated using the weights generated by an algorithm based on the recurrence relation given in the text. We conclude that the method is extremely powerful in evaluating integrals over an infinite range containing a spherical Bessel function because the quadrature weights are easily calculated and the rate of convergence is rapid (few quadrature points are needed) for an accurate result even when r is large.

ACKNOWLEDGMENTS

The work of D. R. Lehman is supported in part by the U. S. Department of Energy. D. R. L. extends his thanks to Ian Sloan for introducing him to the product-integration method and suggesting its possible applicability to integrals

of the type discussed in this paper. The authors are grateful to Daniel W. Lozier and Donald Orser for modifying and running our computer codes through their arbitrary precision arithmetic precompilers to test the accuracy of our results.

APPENDIX A: RECURRENCE SCHEMES FOR PRODUCT-INTEGRATION WEIGHTS

In calculating a set of weights corresponding to the Clenshaw–Curtis abscissa points, we have used several alternative methods. The most efficient and accurate of these is presented in Sec. III. For the benefit of the reader, we present in this Appendix an alternative method for generating the same weights. The presentation allows us to discuss several interesting properties of the weights and related functions, as well as some sum rules and a "hidden precision" feature of the weights.

As shown in Sec. II, the integral

$$I_l = \int_0^\infty k^2\,dk\, j_l(kr)\phi_l(k) \tag{1}$$

can be approximated as a finite sum over weighted values of ϕ_l

$$I_l \approx \sum_{i=0}^{N} W_i(l,n;\alpha;r)\phi_l(k_{Ni}), \tag{25}$$

where

$$W_i(l,n;\alpha;r) = \alpha^3\left(\frac{k_{Ni}^2+\alpha^2}{\alpha^2}\right)^{q+1}\left(\frac{\alpha}{k_{Ni}}\right)^l \frac{2}{N}\alpha_{Ni}\sum_{j=0}^{N}\alpha_{Nj} \times\cos\left(\frac{ij\pi}{N}\right)\mathscr{H}^l_{jq}(\alpha r), \tag{A1}$$

$$\mathscr{H}^l_{jq}(\mu) = \int_0^\infty j_l(\mu k)T_j\left(\frac{k^2-1}{k^2+1}\right)\frac{k^{l+2}\,dk}{(k^2+1)^{q+1}}, \tag{24}$$

and

$$k_{Ni} = \alpha\cot(i\pi/2N). \tag{26}$$

with $q=(n+l)/2$ and $\mu=\alpha r$. The factors $(k^2+\alpha^2)^{q+1}$ and k^l which appear in the weights were chosen to factor out part or all of the asymptotic and all of the threshold behavior of the integrand function $\phi_l(k)$, making the resulting function smoother near its argument limit points, and thus easier to fit with the Chebyshev polynomials. The index q should be taken near $m/2$ [see Eq. (7)] for this smoothing purpose. Correspondingly, the parameter α is best taken in the region of k where $\phi_l(k)$ begins its asymptotic behavior.

The functions $\mathscr{H}^l_{jq}$ which appear in the expression (A1) for the weights can be generated by a simple recurrence scheme. Consider the integral

$$\mathscr{H}^l_{0q} \equiv \int_0^\infty j_l(\mu k)\frac{k^{l+2}\,dk}{(k^2+1)^{q+1}} \tag{A2}$$

$$= \mu^{2q-l-1}\int_0^\infty j_l(z)\frac{z^{l+2}\,dz}{(z^2+\mu^2)^{q+1}}. \tag{A3}$$

Using[12]

$$\frac{d}{dz}[z^{l+1}j_l(z)] = z^{l+1}j_{l-1}(z), \tag{A4}$$

the integral reduces to

$$\mathscr{H}^l_{0q}(\mu) = \frac{\mu^{2q-l-1}}{2^l q(q-1)\ldots(q-l)} \times\int_0^\infty \frac{\cos z\,dz}{(z^2+\mu^2)^{q-l}} \quad \text{for} \quad q>l>0 \tag{A5}$$

$$= \frac{\mu^{2q-l-1}}{2^q q!}\int_0^\infty \frac{j_{l-q}(z)z^{l-q+2}\,dz}{(z^2+\mu^2)} \quad \text{for} \quad l\geqslant q\geqslant 0. \tag{A6}$$

Now

$$\int_0^\infty \frac{\cos z\,dz}{(z^2+\mu^2)^{n+1}} = \frac{(-1)^n}{n!}\left(\frac{d}{2\mu d\mu}\right)^n \times\int_0^\infty \frac{\cos z\,dz}{z^2+\mu^2} = \frac{(-1)^n}{n!}\left(\frac{d}{2\mu d\mu}\right)^n\left(\frac{\pi e^{-\mu}}{2\mu}\right), \tag{A7}$$

which is directly related to Bessel functions of the third kind[15]

$$k_n(z) \equiv \left(\frac{\pi}{2z}\right)^{1/2}K_{n+1/2}(z) = -\frac{\pi}{2}i^l h_l^{(1)}(iz) = (-1)^n z^n\left(\frac{d}{zdz}\right)^n\left(\frac{\pi e^{-z}}{2z}\right) = \frac{\pi e^{-z}}{2z}\sum_{p=0}^{n}\frac{(n+p)!}{(n-p)!p!2^p}\frac{1}{z^p}, \tag{A8}$$

i.e.,

$$\int_0^\infty \frac{\cos z\,dz}{(z^2+\mu^2)^{n+1}} = \frac{1}{n!(2\mu)^n}k_n(z). \tag{A9}$$

Note the following properties of the functions $k_n(z)$[11]:

$$k_{-1}(z) = k_0(z) = \pi e^{-z}/2z, \tag{A10}$$

$$k_{-n}(z) = k_{n-1}(z), \tag{A11}$$

$$\left(-\frac{d}{dz}+\frac{n}{z}\right)k_n(z) = k_{n+1}(z), \tag{A12}$$

$$\left(-\frac{d}{dz}-\frac{n+1}{z}\right)k_n(z) = k_{n-1}(z), \tag{A13}$$

and

$$k_{n+1}(z) = k_{n-1}(z) + \frac{2n+1}{z}k_n(z). \tag{A14}$$

As for the integral in (A6), we use[11]

$$j_l(z) = e^{iz}R_l(z) + (-1)^l e^{-iz}R_l^*(z), \tag{A15}$$

where

$$R_l(z) = \frac{(-i)^{l+1}}{2z}\sum_{p=0}^{l}\frac{i^p(l+p)!}{(l-p)!p!2^p}\frac{1}{z^p} \tag{A16}$$

and find, by contour integration

$$\int_0^\infty \frac{j_l(z)z^{l+2}\,dz}{(z^2+\mu^2)} = \frac{\pi e^{-\mu}}{2\mu}\mu^{l+1}\sum_{p=0}^{l}\frac{(p+l)!}{(p-l)!p!(2\mu)^p}. \tag{A17}$$

With the identity Eq. (A11), the two cases in (A5) and (A6) become

$$\mathscr{H}^l_{0q}(\mu) = \int_0^\infty j_l(\mu k)\frac{k^{l+2}\,dk}{(k^2+1)^{q+1}} = \frac{\mu^q}{2^q q!}k_{l-q}(\mu) \tag{A18}$$

for all $l \geqslant 0$ and $q \geqslant 0$. Using the recurrence relations (A14), the expression (A18) determines $\mathscr{H}^l_{0q}(\mu)$ with errors no larger than approximately 10^{-D}, where D is the number of digits carried in the calculation.

For higher Chebyshev index i in $\mathscr{H}^l_{iq}(\mu)$, we can use the recurrence relation for the Chebyshev polynomials[11]

$$T_{i+1}(z) = 2zT_i - T_{i-1} \tag{A19}$$

and the starting values

$$T_0(z) = 1\,,$$
$$T_1(z) = z\,, \tag{A20}$$

giving

$$\mathscr{H}^l_{(i+1)q}(\mu) = \int_0^\infty j_l(\mu k)\left[2\left(1 - \frac{2}{k^2+1}\right)T_i - T_{i-1}\right] \times \frac{k^{l+2}\,dk}{(k^2+1)^{q+1}} \tag{A21}$$

or

$$\mathscr{H}^l_{(i+1)q}(\mu) = 2\mathscr{H}^l_{iq}(\mu) - 4\mathscr{H}^l_{i(q+1)}(\mu) - \mathscr{H}^l_{(i-1)q}(\mu)\,, \tag{A22}$$

$$\mathscr{H}^l_{1q}(\mu) = \mathscr{H}^l_{0q}(\mu) - 2\mathscr{H}^l_{0(q+1)}(\mu). \tag{A23}$$

The relations (A18), (A22), and (A23) form an attractive scheme for calculating the $\mathscr{H}$'s, and thus the needed weights for the expression (A1). As we shall see, however, the procedure is limited by error accumulation having its origin in the polynomial expansion of the oscillatory Chebyshev functions

$$T_i(\cos\theta) = \cos i\theta \tag{A24}$$

into powers of $z = \cos\theta$.

As a check of the numerical calculation of the $\mathscr{H}$'s, we note the sum rule:

$$\begin{aligned}\sum_{q=0}^\infty \mathscr{H}^l_{0q}(\mu) &= \int_0^\infty j_l(\mu k)\sum_{q=0}^\infty \frac{1}{(k^2+1)^q}\frac{dk}{(1+k^2)} \\ &= \int_0^\infty j_l(\mu k)k^l\,dk = \frac{1}{\mu^{l+1}}\int_0^\infty j_l(z)z^l\,dz \\ &= \frac{1}{\mu^{l+1}}\int_0^\infty z^{2l}\left[\left(-\frac{d}{zdz}\right)^l j_0(z)\right]dz \\ &= \frac{(2l-1)!!}{\mu^{l+1}}\int_0^\infty j_0(z)\,dz \\ &= \frac{(2l-1)!!}{\mu^{l+1}}\frac{\pi}{2}.\end{aligned} \tag{A25}$$

Convergence to this limit is slow, since

$$\lim_{q\to\infty}\mathscr{H}^l_{0q}(\mu) \sim \left(\frac{\mu}{2}\right)^l e^{-\mu}\frac{\pi^{1/2}}{4}\frac{1}{q^{3/2}}. \tag{A26}$$

For $i \neq 0$, we use Eq. (A23) to find

$$\sum_{q=0}^\infty \mathscr{H}^l_{1q}(\mu) = 2k_l - \frac{(2l-1)!!}{\mu^{l+1}}\frac{\pi}{2}. \tag{A27}$$

Similarly, from (A22)

$$\sum_{q=0}^\infty \mathscr{H}^l_{2q}(\mu) = \frac{\mu}{2}k_{l-1} - 4k_l + \frac{(2l-1)!!}{\mu^{l+1}}\frac{\pi}{2}, \tag{A28}$$

etc.

The errors in the calculated $\mathscr{H}$'s using the relations (A22) and (A23) are dominated by cancellation due to the Chebyshev oscillations. This property is explicit in the Chebyshev expansion[16]

$$\begin{aligned}T_i\left(\frac{k^2-1}{k^2+1}\right) &= {}_2F_1\left(i, -i; \tfrac{1}{2}; \frac{1}{k^2+1}\right) \\ &= \sum_{p=0}^i \frac{(i)_p(-i)_p}{(1/2)_p p!}\frac{1}{(k^2+1)^p} \\ &\equiv \sum_{p=0}^i t_p \frac{1}{(k^2+1)^p}.\end{aligned} \tag{A29}$$

The Chebyshev polynomial itself is bounded by 1,

$$|T_i(z)| = |\cos i\theta| \leqslant 1,$$

where $z = \cos\theta$, while the coefficients t_p in the series above may individually greatly exceed one. The largest such coefficient, for fixed, but large i, reaches the approximate value:

$$t_p|_{\max} \approx (-1)^p \times 0.47 \times (5.828)^i / i^{1/2} \tag{A30}$$

when p is the integer closest to $i/\sqrt{2}$. If the T_i are calculated carrying D decimals using a recurrence relation reconstructing the series in (A29), then cancellation errors will leave an error in the result of the order

$$\delta T_i \approx (5.8)^i \times 10^{-D} \tag{A31}$$

[we neglect the weakly varying part $i^{-1/2}$, so that (A31) overestimates the error].

The functions $\mathscr{H}^l_{iq}$ calculated from (A22) will also have the Chebyshev errors incorporated. For μ of the order of one, these will be

$$\delta\mathscr{H}^l_{iq} \approx (5.8)^i \times 10^{-D} \tag{A32}$$

[with $\mu \gg 1$ or $\ll 1$, additional errors arise from the higher powers of μ affecting the maximum term in the polynomial series generated by (A22)]. The estimate (A32) has been verified by double, quadruple, and arbitrary precision computer calculation of the $\mathscr{H}$'s. Table VI gives a sample calculation agreeing with the estimated error of Eq. (A32).

Clearly, the errors in $\mathscr{H}^l_{iq}$ will affect the errors of the weights calculated from (A1). If we let

$$\delta\mathscr{H}^l_{iq} = a\epsilon^i \times 10^{-D} \tag{A33}$$

TABLE VI. Computational errors of $\bar{\mathscr{H}}^l_{iq}(\mu)$[a].

D	$\bar{\mathscr{H}}^0_{35,2}(0.5)$	$\delta\bar{\mathscr{H}}^0_{35,2}(0.5)$	$\bar{\mathscr{H}}^0_{40,0}(0.5)$	$\delta\bar{\mathscr{H}}^0_{40,0}(0.5)$
35	0.1297 6530 (-4)	5 (-9)	0.2760 617	3 (-5)
43	0.1297 6009 5729 7255 1 (-4)	5 (-17)	0.2760 6114 0654 1257 1	3 (-13)
62	0.1297 6009 5729 7255 2 266 0450 4940 5302 37 (-4)	5 (-36)	0.2760 6114 0654 1257 2 058 3812 1109 3282 73	3 (-32)

[a] $\bar{\mathscr{H}}^l_{iq}(\mu) \equiv \mathscr{H}^l_{iq}(\mu)/\mathscr{H}^0_{00}(\mu)$.

and

$$w_i(l,n;\alpha;r) = \alpha^3 \frac{2^{q+2}}{N} \alpha_{Ni} \sum_{j=0}^{N} \alpha_{Nj} \cos\left(\frac{ij\pi}{N}\right) \mathscr{H}^l_{jq}(\alpha r), \quad (22)$$

then these weights will have error estimated by

$$\delta w_i = \alpha_{Ni} \frac{2^{q+2}}{N} a 10^{-D} \left(\frac{\epsilon^2 - 1}{2}\right) \times \left\{ \frac{(-1)^i \epsilon^N - 1}{\epsilon^2 - 2\epsilon \cos(i\pi/N) + 1} \right\} \quad (A34)$$

$$\approx \frac{6.4 \times 2^q}{N} 10^{0.765N - D}. \quad (A35)$$

Note that the error of every weight is dominated by the error of the highest i value of $\mathscr{H}^l_{iq}$, namely, $i = N$. Despite this fact, the errors in the evaluation of an integral by (A1) may be many orders of magnitude less than the errors of the weights. We should expect this to occur, since rapidly oscillating Chebyshev terms are not important in approximating a smooth integrand function.

By extraction of threshold and asymptotic behavior in ϕ_l, the function

$$f_l(k) = \frac{1}{2^{q+1}} \left(\frac{\alpha}{k}\right)^l \left(\frac{k^2 + \alpha^2}{\alpha^2}\right)^{q+1} \phi_l(k) \quad (A36)$$

is slowly varying for small and large k. In the Chebyshev expansion

$$f_l(k) \cong \sum_{i=0}^{N} \alpha_{Ni} b^l_{Ni} T_i\left(\frac{k^2 - \alpha^2}{k^2 + \alpha^2}\right), \quad (A37)$$

where

$$b^l_{Ni} = \frac{2}{N} \sum_{j=0}^{N} \alpha_{Nj} f_l(k_{Ni}) \cos\frac{ij\pi}{N}, \quad (A38)$$

we expect that an f_l which is smoothly varying over all k will be accurately represented by only a few b terms. Now

$$I_l \approx \sum_{i=0}^{N} W_i \phi_l(k_{Ni}) = \sum_{i=1}^{N} w_i f_l(k_{Ni}) = \sum_{i,j=0}^{N} \alpha_{Ni} w_i b^l_{Nj} \cos\frac{ij\pi}{N} \quad (A39)$$

so that the errors in I_l due to w_i become

$$\delta I_l \cong \frac{2^{q+2}}{N} a \times 10^{-D} \left(\frac{\epsilon^2 - 1}{2}\right) \sum_{j=0}^{N} \alpha_{Nj} b^l_{Nj} \times \sum_{i=0}^{N} \alpha_{Ni} \left[\frac{(-1)^i \epsilon^N - 1}{\epsilon^2 - 2\epsilon\cos(i\pi/N) + 1}\right] \cos\frac{ij\pi}{N}$$

$$\cong 2^{q+1} a \times 10^{-D} \sum_{j=0}^{N} \alpha_{Nj} b^l_{Nj} \epsilon^j. \quad (A40)$$

The errors in δI_l are not of order ϵ^N as they were for the $\mathscr{H}$'s, but rather ϵ^m, where m is the highest order Chebyshev term contributing to the f_l expansion (A38). This remarkable error cancellation effect gives a hidden precision to the weights computed by this method. Table VII shows an example of this phenomena.

Closed form analytic expressions for the $\mathscr{H}$'s are useful in checking calculations. By using (A2), (A18), and (A29), we have

$$\mathscr{H}^l_{kq}(\mu) = \sum_{p=0}^{k} \frac{(k)_p (-k)_p \mu^{q+p}}{(\frac{1}{2})_p p! 2^{q+p} (q+p)!} k_{q+p-l-1}(\mu). \quad (A41)$$

TABLE VII. Error cancellation effect[a].

Sample Weights w_i: $\ell = 0$, $r = 1.0$, $\alpha = 0.5$, and $N = 40$

i	$w_i(D = 35)$	$w_i(D = 62)$	error
0	- 0.1759 0430 (-1)	- 0.1759 0406 3350 4658 (-1)	~ 1 (-8)
20	0.9857 6077 (-2)	0.9857 6103 2627 723 (-2)	~ 1 (-8)
40	0.1106 2870 (-3)	0.1106 3166 0775 78 (-3)	~ 1 (-8)

$I^t_0(r)$: [D = 35] = 0.4502 2246 2499 8697 170
[D = 62] = 0.4502 2246 2499 8697 173
[exact] = 0.4502 2246 2499 8651 0932
|[D = 35] - [D = 62]| ~ 1 (-18)!

[a]Weights calculated using method of Appendix A.

With (A8), the following examples are offered:

$$\mathscr{H}^0_{00} = \frac{\pi}{2} e^{-\mu}/\mu$$

$$\mathscr{H}^0_{01} = \frac{\pi}{4} e^{-\mu}$$

$$\mathscr{H}^0_{02} = \frac{\pi}{16}(1+\mu)e^{-\mu}$$

$$\mathscr{H}^0_{03} = \frac{\pi}{96}(3+3\mu+\mu^2)e^{-\mu}$$

$$\mathscr{H}^0_{04} = \frac{\pi}{768}(15+15\mu+6\mu^2+\mu^3)e^{-\mu}$$

$$\mathscr{H}^0_{05} = \frac{\pi}{7680}(105+105\mu+45\mu^2+10\mu^3+\mu^4)e^{-\mu}$$

$$\mathscr{H}^0_{10} = \frac{\pi}{2\mu}(1-\mu)e^{-\mu}$$

$$\mathscr{H}^0_{11} = \frac{\pi}{8}(1-\mu)e^{-\mu}$$

$$\mathscr{H}^0_{12} = -\frac{\pi}{48}\mu^2 e^{-\mu}$$

$$\mathscr{H}^0_{13} = -\frac{\pi}{128}\left(1+\mu-\tfrac{2}{3}\mu^2+\tfrac{1}{3}\mu^3\right)e^{-\mu}$$

$$\mathscr{H}^0_{14} = -\frac{\pi}{128}\left(1+\mu+\tfrac{1}{2}\mu^2+\tfrac{1}{6}\mu^3+\tfrac{1}{30}\mu^4\right)e^{-\mu}$$

$$\mathscr{H}^0_{20} = \frac{\pi}{2\mu}(1-3\mu+\mu^2)e^{-\mu}$$

$$\mathscr{H}^0_{21} = -\frac{\pi}{2}\mu\left(1-\tfrac{1}{3}\mu\right)e^{-\mu}$$

$$\mathscr{H}^0_{30} = \frac{\pi}{2\mu}\left(1-5\mu+4\mu^2-\tfrac{2}{3}\mu^3\right)e^{-\mu}$$

$$\mathscr{H}^1_{01} = \frac{\pi}{4} e^{-\mu}$$

$$\mathscr{H}^1_{11} = \frac{\pi}{4}\left(1-\frac{\mu}{2}\right)e^{-\mu}. \quad (A42)$$

APPENDIX B: CLOSED FORM EXPRESSIONS FOR THE WEIGHTS

In this appendix we present closed form analytic expressions for the weights and weight sum rules. In the pro-

cess we will derive several integrals needed in Sec. III for the development of a recurrence scheme for generating the weights, explain why the forms for the weights following from Sec. III do not cause computer error buildup, and give a recurrence relation for generating arbitrarily high l-value weights.

Starting with the expression Eq. (22) for the weights, we may explicitly sum over the j index using the identity

$$\sum_{k=0}^{N} \alpha_{Nk} T_k(x) T_k(y) = \sum_{k=0}^{N} \alpha_{Nk} \cos k\phi \cos k\theta = \tfrac{1}{2}\left\{ \frac{\sin N\phi \cos N\theta \sin\phi - \cos N\phi \sin N\theta \sin\theta}{\cos\theta - \cos\phi} \right\}, \quad \text{(B1)}$$

where $x = \cos\phi$, $y = \cos\theta$. This relation can be derived by expressing the cosine factors as exponentials using the Euler relation

$$\cos k\theta = \tfrac{1}{2}(e^{ik\theta} + e^{-ik\theta}) \quad \text{(B2)}$$

and then summing with the identity:

$$\sum_{k=0}^{N} x^k = \frac{1 - x^{N+1}}{1 - x}. \quad \text{(B3)}$$

Taking $\phi = j\pi/N$, and $z = \cot(\theta/2)$, (B1) gives

$$\sum_{k=0}^{N} \alpha_{Nk} \cos\frac{jk\pi}{N} \cos k\theta = \frac{(-1)^{j+1}}{4i}(1 + z_j^2)\frac{z}{(z^2 - z_j^2)} \times \left[\left(\frac{z+i}{z-i}\right)^N - \left(\frac{z-i}{z+i}\right)^N \right], \quad \text{(B4)}$$

where $z_j \equiv \cot(j\pi/2N)$. Note that as $\theta \to l\pi/N$ (l an integer), the right hand side is not singular, but

$$\text{RHS} \to N\delta_{jl}/2\alpha_{Nj}. \quad \text{(B5)}$$

Now using (A15), the weights can be expressed as

$$w_j = \frac{2^{q+2}\alpha^3}{N}\alpha_{Nj}\frac{(-1)^j(1 + z_j^2)}{4i} \times \int_0^\infty e^{i\mu z} R_l(\mu z) \frac{z^{l+3}(z+i)^{N-q-1}}{(z_j^2 - z^2)} \frac{dz}{(z-i)^{N+q+1}}, \quad \text{(B6)}$$

where we assume $N > q$ to eliminate the $z = -i$ pole. (This is easily done, as N can be freely choosen and we expect q to be small on physical grounds.) With the relation

$$e^{-\mu} R_l(i\mu) = -(-i)^l k_l(\mu)/\pi \quad \text{(B7)}$$

which follows from (A8) and (A16), we have the following contour integral form for the weights:

$$w_j = \alpha^3(-1)^{j+q}\alpha_{Nj}\frac{2^{q+1}}{2\pi i} \times \oint_{z=1} \frac{k_l(\mu z) z^{l+3}(z+1)^{N-q-1} dz}{(1 + y_j^2(z^2 - 1))(z-1)^{N+q+1}}, \quad \text{(B8)}$$

with the contour taken around the integrand pole at $z = 1$ alone, and where

$$y_j = \sin(j\pi/2N). \quad \text{(B9)}$$

The above form, Eq. (B8), gives a differential expression for the weights

$$w_j = \alpha^3 \frac{2^{q+1}}{N}\alpha_{Nj}(-1)^{j+q}\frac{(1 + z_j^2)}{(N+q)!}\left(\frac{d}{dz}\right)^{N+q} \times \left\{ z^{l+1} k_l(\mu z)(1+z)^{N-q-1} \frac{z^2}{(z_j^2 + z^2)} \right\}_{z=1}. \quad \text{(B10)}$$

For example, for $l = 0$, this relation leads to

$$w_j(0,0;\alpha;r) = \alpha^3 \frac{\pi(-1)^{j+1}}{N^2} e^{-\alpha r} \sum_{k=0}^{N-1} \binom{N}{k+1} \frac{(-1)^k (2\alpha r)^k}{k!} \cdot \left\{ 1 + \sum_{p=1}^{N-k-1} \frac{(-N+k+1)_p}{(k+1)_p} 2^p \cos\frac{j\pi}{2N} \times \sin^{p-2}\frac{j\pi}{2N} \sin\left[\left(\frac{j\pi}{2N} - \frac{\pi}{2}\right)(p+1) \right] \right\} \quad \text{(B11)}$$

for $0 < j < N$ where

$$\binom{N}{k} \equiv \frac{N!}{(N-k)!k!}, \quad (a)_p = a(a+1)\cdots(a+p-1),$$

and the end point expressions become

$$w_0(0,0;\alpha;r) = \alpha^3 \frac{\pi}{2\alpha r} e^{-\alpha r}\left(1 - \frac{\alpha r}{2N}\right) {}_1F_1(1-N;2;2\alpha r), \quad \text{(B12)}$$

$$w_N(0,0;\alpha;r) = \alpha^3 \frac{\pi}{4N}(-1)^{N-1} e^{-\alpha r} {}_1F_1(1-N;2;2\alpha r). \quad \text{(B13)}$$

These forms reveal the general analytic structure of the weights, namely, an exponential factor in αr, a polynomial in αr, and an oscillatory part in the index i for $1 \leqslant i \leqslant N-1$. For the purposes of numerical calculation, (B11) suffers from cancellation errors again arising from the Chebyshev oscillations, as we discussed in Appendix A. (For example, with 16 digit arithmetic and $N = 30$ only six significant figures survive cancellation errors.) For $l > 0$, we can use the relation (A12) and (B8) to derive

$$w_i(l+1,q;\alpha;r) = \frac{1}{\alpha}\left(-\frac{d}{dr} + \frac{l}{r} \right) w_i(l,q;\alpha;r) \quad \text{(B14)}$$

and thus find analytic forms for higher values of l.

More directly, starting with the contour integral form for the weights, we can find an exact expression for w_i in powers of μ and $\sin(i\pi/2N)$. Using (B8) and expanding the factor $[1 + y_i^2(z^2 - 1)]^{-1}$ into powers of y_i^2, we have

$$w_i(l,q;\alpha;r) = \alpha^3 \frac{2^{q+1}}{N}\alpha_{Ni}(-1)^{i+q} \sum_{p=0}^{N+q} \{ (-1)^p \times y_i^{2p} v_{(N+q-p)}^{(N-q+p-1)l} + v_{(N+q-p-1)}^{(N-q+p)l} \}, \quad \text{(B15)}$$

where

$$v_m^{kl} = \frac{1}{2\pi i}\oint_{z=1} z^{l+1} k_l(\mu z)(1+z)^k \frac{dz}{(z-1)^{m+1}} \quad \text{(B16)}$$

$$= \frac{1}{m!}\left(\frac{d}{dz}\right)^m [z^{l+1} k_l(\mu z)(1+z)^k]_{z=1}. \quad \text{(B17)}$$

For $l = 0$, the integral (B16) is a terminating confluent hypergeometric function. The result below can be found using the differential form (B17)

$$v_m^{k0}(\mu) = \frac{\pi e^{-\mu}}{2\mu} 2^k \frac{(-\mu)^m}{m!} {}_2F_0\left(-m, -k; -\frac{1}{2\mu}\right) \quad \text{(B18)}$$

$$= \frac{\pi e^{-\mu}}{2\mu}\binom{k}{m}2^{k-m}{}_1F_1(-m;k-m+1;2\mu) \quad k\geqslant m\geqslant 0 \tag{B19}$$

$$= \frac{\pi e^{-\mu}}{2\mu}\frac{(-\mu)^{m-k}}{(m-k)!}{}_1F_1(-k;m-k+1;2\mu) \quad 0\leqslant k\leqslant m. \tag{B20}$$

If k is negative, we have

$$\frac{1}{2\pi i}\oint_{z=1}\frac{\pi e^{-\mu z}}{2\mu}\frac{dz}{(1+z)^{n+1}(z-1)^{m+1}} = \frac{\pi e^{-\mu}}{2\pi}\frac{(-1)^m(n+m)!}{m!n!}{}_1F_1(-m;-n-m;2\alpha r) \tag{B21}$$

or

$$\frac{1}{2\pi i}\oint_{z=1}\frac{e^{-\mu z}dz}{(z+1)^{n+1}(z-1)^{m+1}} = e^{-\mu}\frac{(-1)^m}{2^{m+n+1}}\frac{(n+m)!}{m!n!}{}_1F_1(-m;-n-m;2\alpha r). \tag{B22}$$

These cases are used in Sec. III. In (B16), we assume $k\geqslant 0$, $m\geqslant 0$.

For higher l, we again use (B14). Taking

$$v_m^{k0}(\mu)\equiv k_0(\mu)\vec{v}_m^{k0}(\mu) \tag{B23}$$

allows one to express derivatives of $\vec{v}_m^{k0}$ in a particularly simple form

$$\left(-\frac{d}{d\mu}\right)\vec{v}_m^{k0}(\mu) = \vec{v}_{(m-1)}^{k0}(\mu), \tag{B24}$$

with

$$v_0^{k0}(\mu) = 2^k k_0(\mu) \tag{B25}$$

and

$$v_{-1}^{k0}(\mu) = 0. \tag{B26}$$

Now applying (B14)

$$v_m^{kl}(\mu) = \mu^l\left(-\frac{d}{\mu d\mu}\right)^l v_m^{k0}(\mu) \tag{B27}$$

we can write

$$v_m^{kl}(\mu) = k_l(\mu)\vec{v}_m^{k0}(\mu) + \sum_{s=1}^{l}\sum_{t=0}^{s-1}\binom{l}{s} \times \frac{(s-1+t)!}{2^t t!(s-1-t)!}\frac{k_{l-s}(\mu)\vec{v}_{(m-s+t)}^{k0}(\mu)}{\mu^{s+t}}. \tag{B28}$$

This result, together with (B15), constitutes a closed analytic form for the weights. For example, for $q=0$, $l=0$,

$$\begin{aligned} &w_i(0,0;\alpha;r) \\ &= \alpha^3\frac{2}{N}\alpha_{Ni}(-1)^i\left(\frac{\pi e^{-\alpha r}}{2\alpha r}\right)\{(2N-\alpha r)F(-N+1;2;2\alpha r) \\ &\quad + \sum_{p=1}^{N}(-1)^p\sin^{2p}\left(\frac{i\pi}{2N}\right) \\ &\quad \times\binom{N+p-1}{N-p}2^{2p-1}[F(-N+p;2p;2\alpha r) \\ &\quad + \frac{(N+p)}{(N-p+1)}F(-N+p+1;2p;2\alpha r)]\}. \end{aligned} \tag{B29}$$

The forms above give the weights as a series of the form

$$\sum_{p=0}^{N+q}F_p(\alpha r)\sin^{2p}\left(\frac{j\pi}{2N}\right),$$

where $F_p(\alpha r)$ is a polynomial in αr times $(e^{-\alpha r}/\alpha r)$. The method of Sec. III, in contrast, leads to a series for the weights in the form

$$\sum_{p=1}^{N-1}G_p(\alpha r)\sin\left(\frac{pi\pi}{N}\right),$$

where $G_p(\alpha r)$ is also a polynomial in αr times an exponential $(e^{-\alpha r}/\alpha r)$. This can be seen as follows. The difference equation (43), having constant coefficients, allows a homogeneous solution in the form

$$V_j^{(0)}(\beta) = A\cos j\beta + B\sin j\beta \ (\beta = i\pi/N), \tag{B30}$$

where A and B are constant and we have taken i so that $0<i<N$. Thus the solutions are oscillatory functions of j, rather than exponentially increasing and exponentially decreasing functions of j. For $i=0$, we have $\cos\beta = 1$ and

$$V_j^{(0)} = A + Bj, \tag{B31}$$

while for $i=N$, $\cos\beta = -1$ and

$$V_j^{(0)} = (-1)^j(A+Bj). \tag{B32}$$

Here, although one of the solutions of the homogeneous equation increases with j, the increase is only linear. We therefore do not expect (and indeed have not found) any problems in the computer implementation of the recurrence relation for V_j. Such problems, which generally appear if the desired solution is an exponentially decreasing function, are due to the introduction, through roundoff error, of a small component of the exponentially increasing solution.

We next observe that the Wronskian of the difference equation for $V_j(\beta)$, defined by

$$\mathscr{W} = u_{j+1}v_j - u_j v_{j+1}, \tag{B33}$$

where u_j and v_j are linearly independent solutions of the homogeneous solution ($u_j = \sin j\beta$, $v_j = \cos j\beta$), is a constant

$$\mathscr{W} = \sin(j+1)\beta\cos j\beta - \sin j\beta\cos(j+1)\beta = \sin\beta. \tag{B34}$$

We can therefore write $V_N(\beta)$ quite simply as a sum. For $i\neq 0, N$

$$\begin{aligned} V_N(\beta) = {} & \frac{\sin N\beta}{\sin\beta}\sum_{j=1}^{N}G_j\cos j\beta - \frac{\cos N\beta}{\sin\beta}\sum_{j=1}^{N}G_j\sin j\beta \\ & - V_0(\beta)\sin(N-1)\beta/\sin\beta - V_1(\beta)\sin N\beta/\sin\beta, \end{aligned} \tag{B35}$$

which reduces, because of $\sin N\beta = 0$ and $V_0(\beta) = 0$, to

$$w_i = V_N(\beta) = \frac{(-1)^N}{\sin\beta}\sum_{j=1}^{N-1}G_j\sin j\beta. \tag{B36}$$

For $i=0$, the Wronskian is one and

$$\begin{aligned} V_N(\beta) &= N\sum_{j=1}^{N}G_j - \sum_{j=1}^{N}jG_j + V(\beta) = [V_1(\beta) - V_0(\beta)]N \\ &= \sum_{j=1}^{N-1}(N-j)G_j + NV_1(\beta). \end{aligned} \tag{B37}$$

Finally, for $i=N$ the Wronskian is minus one and

$$V_N(\beta) = (-1)^{N-1}\sum_{j=1}^{N-1}(-1)^j(N-j)G_j + (-1)^{N-1}NV_1(\beta). \tag{B38}$$

In spite of the fact that such analytic and closed forms can be found for the weights, it is simpler and faster to compute the weights from the recurrence relation than to compute them from the above sums.

Once the weights $l = 0$ and $l = 1$ have been generated, weights for higher l values can be generated recursively using

$$w_i(l+1,q;\alpha;r) = w_i(l-1,q-1;\alpha;r) + w_i(l-1,q;\alpha;r) + ((2l+1)/\alpha r)w_i(l,q;\alpha;r). \tag{B39}$$

This can be shown starting with the contour integral form (B8) for the general weight and using the recursion relation for the k_l functions [Eq. (A14)] in the form

$$z^{l+4}k_{l+1}(\mu z) = \{(z+1)(z-1)+1\}z^{l+2}k_{l-1}(\mu z) + \frac{2l+1}{\mu}z^{l+3}k_l(\mu z). \tag{B40}$$

Finally, as a check of a calculation for the weights, sum rules can be used (although they are not foolproof!). They follow simply by specializing the integrand function $f_l(k)$. For the choices

$$f_l(k) = 1 \tag{B41}$$

and

$$f_l(k) = \alpha^2/(k^2+\alpha^2), \tag{B42}$$

for example, we find from Eqs. (1), (A36), and (A39)

$$\sum_{i=0}^{N} w_i(l,q;\alpha;r) = 2\alpha^3\frac{\mu^q}{q!}k_{q-l-1}(\alpha r) \tag{B43}$$

and

$$\sum_{i=o}^{N} w_i(l,q;\alpha;r)\sin^2\left(\frac{i\pi}{2N}\right) = \alpha^3\frac{\mu^{q+1}}{(q+1)!}k_{q-l}(\alpha r). \tag{B44}$$

[1]H. A. Antosiewicz, in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, AMS 55, Washington, D.C., 1964—Ninth Printing, 1970), p. 437.
[2]D. R. Lehman and W. C. Parke (to be published).
[3]M. Blakemore, G. A. Evans, and J. Hyslop, J. Comput. Phys. **22**, 352 (1976).
[4]R. E. Beard, J. Inst. Actu. **73**, 356 (1947).
[5]I. H. Sloan, J. Math. Phys. **21**, 1032 (1980) and references therein.
[6]A. Young, Proc. Roy. Soc. London Ser. A. **224**, 561 (1954).
[7]Y. E. Kim, J. Math. Phys. **10**, 1491 (1969).
[8]I. H. Sloan and W. E. Smith, Numerische Math. **30**, 415 (1978); W. E. Smith and I. H. Sloan, SIAM J. Numer. Anal. **17**, 1 (1980); I. H. Sloan and W. E. Smith, Numerischhe Math. **34**, 387 (1980).
[9]C. W. Clenshaw and A. R. Curtis, Numerische Math. **2**, 197 (1960). For an earlier application of the Clenshaw–Curtis quadrature method, see R. Piessens and M. Branders, J. Appl. Math. **1**, 55 (1955).
[10]Urs. W. Hochstrasser, in Ref. 1 above, p. 774.
[11]See Ref. 1 above, p. 439.
[12]L. J. Slater, in Ref. 1 above, p. 506.
[13]L. J. Slater, in Ref. 1 above, p. 508.
[14]H. Bateman, *Higher Transcedenental Functions*, Vol. 1 (McGraw-Hill, New York, 1953), pp. 254–55.
[15]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1945), 2nd ed., p. 434.
[16]Urs. W. Hochstrasser, in Ref. 1 above, p. 779.

Equivalent Lagrangians: Multidimensional case [a)]

S. Hojman and H. Harleston
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Apdo. Postal 70-543, México 20, D. F., Mexico

(Received 27 May 1980; accepted for publication 17 September 1980)

We generalize a theorem known for one-dimensional nonsingular equivalent Lagrangians (L and $\bar{L}$) to the multidimensional case. In particular, we prove that the matrix Λ, which relates the left-hand sides of the Euler–Lagrange equations obtained from L and $\bar{L}$, is such that the trace of all its integer powers are constants of the motion. We construct several multidimensional examples in which the elements of Λ are functions of position, velocity, and time, and prove that in some cases equivalence prevails even if $\det\Lambda = 0$.

PACS numbers: 03.20. + i

I. INTRODUCTION

The aim of this work is to generalize some results obtained by Currie and Saletan[1] in the study of one-dimensional nonsingular equivalent Lagrangians to the multidimensional case and illustrate them with the explicit construction of nontrivial multidimensional examples.

A Lagrangian $L = L(q^i,\dot{q}^i,t)$ is said to be nonsingular when

$$\det W \neq 0, \quad W_{ij} = \frac{\partial^2 L}{\partial\dot{q}^i\partial\dot{q}^j}.$$

We say that two Lagrangians L and $\bar{L}$ are s-equivalent when the sets of all solutions to the differential equations obtained from them coincide. Notice that it is not necessary for the equations of motion obtained from both Lagrangians to be *exactly* the same, so that situations more interesting than $\bar{L} \equiv \rho L - dF(q,t)/dt$ $(\rho \neq 0)$ arise.

We are interested in different Lagrangian descriptions of a system that give rise to the same set of solutions because, after all, the trajectories (and *not* the equations of motion themselves) are related to observation and experiment, at least within the realm of classical physics[2,3]:

Let

$$L_s \equiv \frac{d}{dt}\frac{\partial L}{\partial\dot{q}^s} - \frac{\partial L}{\partial q^s}, \quad \bar{L}_r \equiv \frac{d}{dt}\frac{d\bar{L}}{\partial\dot{q}^r} - \frac{\partial\bar{L}}{\partial q^r};$$

we will prove that s-equivalence implies

$$\bar{L}_r \equiv \Lambda_r{}^s(q,\dot{q},t)L_s,$$

and that $\mathrm{tr}(\Lambda)^k$ are constants of the motion for any positive integer k, generalizing a theorem given in Ref. 1, for one dimension. Furthermore, we will show through some examples that even when $\det\Lambda = 0$, s-equivalence prevails; i.e., it is possible that L being regular gives rise to an s-equivalent Lagrangian $\bar{L}$ which may be singular. As a matter of fact, for the two-dimensional examples we obtain s-equivalence even when the rank of Λ is zero.

We will work out examples where the elements of Λ are functions of position, velocity, and time (not just numbers), also avoiding the repeated use of the one-dimensional results of Ref. 1. These Lagrangians represent the two- and three-dimensional harmonic oscillators and are of fourth degree rather than quadratic.

We should mention that there is a closely related (though more general) problem called "the inverse problem of the calculus of variations," which we will not discuss here. It consists essentially in trying to find all the Lagrangians that upon variation will give rise to a given system of differential equations. A large amount of very interesting work in that direction has been published lately.[4–12]

The impact of the equivalent description of classical systems at the quantum level is still not completely understood. The study of the problem in gauge as well as field theories also remains to be done.

In Sec. III we prove the theorem of multidimensional s-equivalence, in particular that $\mathrm{tr}(\Lambda)^k$ are constant for any positive integer k.

In Sec. IV we work out three examples of s-equivalent Lagrangians in some detail, and, finally, in Sec. V we discuss briefly what has been done and point out some problems which remain to be solved.

II. THE ONE-DIMENSIONAL CASE

The problem of determining whether a (second-order) differential equation can be derived from a variational principle (the inverse problem of the calculus of variations) was solved by Darboux in 1891[13] (see also Ref. 14). He proved that in the case of one second-order differential equation for one variable it is always possible to construct infinite Lagrangians that will yield the desired equation upon variation, and provided a way to construct them. The two-dimensional case was treated by Douglas.[15] In the work presented here, however, we will not consider the inverse problem but will be mostly interested in dealing with the problem of equivalent Lagrangians in the case of dynamical systems with n degrees of freedom.

The study of equivalent nonsingular Lagrangians for the one-dimensional case in classical mechanics was both proposed and solved in Ref. 1. We will briefly review the main concepts and results here to facilitate the understanding of the multidimensional case, in which we are mainly interested. We will use a slightly different wording and ap-

[a)] Part of this work was presented by one of the authors (H. H.) as a thesis in partial fulfillment of the requirements to obtain a B.Sc. in Physics at Universidad Nacional Autónoma de México.

 0022-2488/81/071414-06$1.00

proach but both the ideas and results are equivalent to those of Ref. 1.

Two Lagrangians $L(q,\dot{q},t)$ and $\bar{L}(q,\dot{q},t)$ will be called s-equivalent iff the two sets of all solutions to the Euler–Lagrange equations obtained from them coincide or, in other words, when every solution to the equations of motion corresponding to one of the Lagrangians satisfies the equations of motion obtained from the other Lagrangian, and conversely. The task of finding the general relationship between s-equivalent (one-dimensional) Lagrangians was undertaken in Ref. 1, and it was proved that s-equivalence implies that the left-hand side of the equations of motion are "proportional" to each other, the proportionality factor (which must neither vanish nor become infinite) being a function of the coordinate q, the velocity $\dot{q}$, and time t, then, the condition for $L(q,\dot{q},t)$ and $\bar{L}(q,\dot{q},t)$ to be s-equivalent can be written as

$$\frac{d}{dt}\frac{\partial \bar{L}}{\partial \dot{q}} - \frac{\partial \bar{L}}{\partial q} = f(q,\dot{q},t)\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q}\right], \tag{1}$$

with

$$f = \bar{L}_{\dot{q}\dot{q}}/L_{\dot{q}\dot{q}}, \quad 0 \neq f \neq \infty. \tag{2}$$

Currie and Saletan proved that f is a constant of the motion, and, given any constant of the motion of the dynamics due to L, they provided a way to construct $\bar{L}$ explicitly.

An interesting example of the above mentioned theorem is

$$L = \tfrac{1}{2}\dot{x}^2 - \tfrac{1}{2}x^2, \tag{3}$$

$$\bar{L} = \tfrac{1}{2}x\dot{x}^2\cos t - \tfrac{1}{6}\dot{x}^3\sin t - \tfrac{1}{3}x^3\cos t, \tag{4}$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = \ddot{x} + x = 0, \tag{5}$$

$$\frac{d}{dt}\frac{\partial \bar{L}}{\partial \dot{x}} - \frac{\partial \bar{L}}{\partial x} = (x\cos t - \dot{x}\sin t)(\ddot{x} + x) = 0; \tag{6}$$

therefore,

$$f = x\cos t - \dot{x}\sin t \tag{7}$$

and

$$\dot{f} = 0, \tag{8}$$

as it can be easily verified. The set of all solutions to the equation of motion (6), i.e.,

$$(x\cos t - \dot{x}\sin t) = 0, \tag{9}$$

and/or

$$(\ddot{x} + x) = 0, \tag{10}$$

can be written as

$$x(t;A,\alpha) = A\sin(t+\alpha), \tag{11}$$

A,α being constant, which is the general solution to Eq. (6). Therefore, it is not necessary to require $f \neq 0$ to prove the equivalence of L and $\bar{L}$; i.e., the solutions to the equation of motion $f(q,\dot{q},t) = 0$ are particular solutions of Eq. (6) for $\alpha = 0$.

The fact that one still has s-equivalence even when considering $f = 0$ constitutes a slight (but significant, especially for quantum purposes) generalization of the cases considered in the literature. We will elaborate this point further in the multidimensional case.

III. THE MULTIDIMENSIONAL CASE

Even though the one-dimensional case is of great interest in itself, it is appealing to consider more realistic cases where the number of dimensions could be appropriate to describe nature. In particular, when discussing the impact of these results in quantum theory, it is relevant to ask whether interesting examples exist in higher dimensions.

We will discuss in this section the generalization of the one dimensional problem to n dimensions.

Consider the Lagrangian $L = L(q^i,\dot{q}^i,t)$, $i = 1,\ldots,n$, such that

$$\det\left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \equiv \det(W_{ij}) \neq 0.\right) \tag{12}$$

Define

$$L_r \equiv \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^r}\right) - \frac{\partial L}{\partial q^r} = W_{rs}\ddot{q}^s + \frac{\partial^2 L}{\partial \dot{q}^r \partial q^s}\dot{q}^s - \frac{\partial L}{\partial q^r}. \tag{13}$$

(We drop the explicit time dependence of L, but all the results which will be obtained hold for an arbitrary explicitly time-dependent Lagrangian as well.)

The Lagrangian $\bar{L} = \bar{L}(q^i,\dot{q}^i,t)$ (with $\det \bar{W} \neq 0$) is said to be subordinate to L iff

$$\{L_r = 0\} \Rightarrow \{\bar{L}_s = 0\}. \tag{14}$$

We will now prove that if $\bar{L}$ is subordinate to L, then L is subordinate to $\bar{L}$ and that

$$\bar{L}_r = \Lambda_r{}^s(q,\dot{q},t)L_s, \tag{15}$$

with $\det\Lambda \neq 0$. Moreover, the trace of all integer powers of Λ are constants of the motion.

The equations of motion for L and $\bar{L}$ read

$$W_{rs}\ddot{q}^s + \frac{\partial^2 L}{\partial \dot{q}^r \partial q^s}\dot{q}^s - \frac{\partial L}{\partial q^r} = 0, \tag{16}$$

$$\bar{W}_{rs}\ddot{q}^s + \frac{\partial^2 \bar{L}}{\partial \dot{q}^r \partial q^s}\dot{q}^s - \frac{\partial \bar{L}}{\partial q^r} = 0. \tag{17}$$

Let U and $\bar{U}$ be the inverse matrices to W and $\bar{W}$ respectively; then from Eq. (17)

$$\ddot{q}^s = \bar{U}^{sp}\left(\frac{\partial \bar{L}}{\partial q^p} - \frac{\partial^2 \bar{L}}{\partial \dot{q}^p \partial q^t}\dot{q}^t\right), \tag{18}$$

and we have assumed that Eq. (16) implies Eq. (17). Therefore, the expression (18) can be inserted back in Eq. (16), from which we get

$$W_{rs}\bar{U}^{sp} + \left(\frac{\partial^2 \bar{L}}{\partial q^p} - \frac{\partial^2 \bar{L}}{\partial \dot{q}^p \partial q^t}\dot{q}^t\right) = \frac{\partial L}{\partial q^r} - \frac{\partial^2 L}{\partial \dot{q}^r \partial \dot{q}^s}\dot{q}^s, \tag{19}$$

and L_r can be written as

$$L_r = W_{rs}\bar{U}^{sp}\left(\bar{W}_{pt}\ddot{q}^t + \frac{\partial^2 \bar{L}}{\partial \dot{q}^p \partial q^t}\dot{q}^t - \frac{\partial \bar{L}}{\partial q^p}\right) = W_{rs}\bar{U}^{sp}\bar{L}_p, \tag{20}$$

i.e.,

$$\bar{L}_p = \Lambda_p{}^r L_r, \tag{21}$$

with

$$\Lambda_p{}^r \equiv \bar{W}_{ps}(q,\dot{q},t)U^{sr}(q,\dot{q},t), \tag{22}$$

and $\det\Lambda \neq 0$, i.e.,

$$(\{L_r = 0\} \Rightarrow \{\bar{L}_s = 0\}) \Rightarrow (\{L_r = 0\} \Leftrightarrow \{\bar{L}_s = 0\}). \tag{23}$$

We will now get a differential equation for Λ that will allow us to prove that $\mathrm{tr}(\Lambda)^m$ is constant for any integer m. Equation (19) can be written as

$$\frac{\partial^2 \bar{L}}{\partial \dot{q}^s \partial q^t}\dot{q}^t - \frac{\partial \bar{L}}{\partial q^s} = \Lambda_s{}^r\left(\frac{\partial^2 L}{\partial \dot{q}^r \partial q^t}\dot{q}^t - \frac{\partial L}{\partial q^r}\right); \tag{24}$$

when using the definition (22). Differentiating with respect to $\dot{q}^u$, one gets

$$\begin{aligned}\frac{\partial^3 \bar{L}}{\partial \dot{q}^u \partial \dot{q}^s \partial q^t}\dot{q}^t + \frac{\partial^2 \bar{L}}{\partial \dot{q}^s \partial q^u} - \frac{\partial^2 \bar{L}}{\partial \dot{q}^u \partial q^s} \\ = \frac{\partial \Lambda_s{}^r}{\partial \dot{q}^u}\left(\frac{\partial^2 L}{\partial \dot{q}^r \partial q^t}\dot{q}^t - \frac{\partial L}{\partial q^r}\right) \\ + \Lambda_s{}^r\left(\frac{\partial^3 L}{\partial \dot{q}^r \partial \dot{q}^u \partial q^t}\dot{q}^t + \frac{\partial^2 L}{\partial \dot{q}^r \partial q^u} - \frac{\partial^2 L}{\partial \dot{q}^u \partial q^r}\right),\end{aligned} \tag{25}$$

and, using the equation of motion (16), one gets

$$\frac{\partial^2 L}{\partial \dot{q}^r \partial q^t}\dot{q}^t - \frac{\partial L}{\partial q^r} = -W_{rt}\ddot{q}^t. \tag{26}$$

Using the fact that

$$\Lambda_r{}^s W_{st} = \bar{W}_{rt}, \tag{27}$$

one can prove that

$$\frac{\partial \Lambda_s{}^r}{\partial \dot{q}^u}\frac{\partial^2 L}{\partial \dot{q}^r \partial \dot{q}^t} = \frac{\partial \Lambda_s{}^r}{\partial \dot{q}^t}\frac{\partial^2 L}{\partial \dot{q}^r \partial \dot{q}^u}. \tag{28}$$

Furthermore,

$$\begin{aligned}\Lambda_s{}^r \frac{\partial^3 L}{\partial \dot{q}^r \partial \dot{q}^u \partial q^t}\dot{q}^t - \frac{\partial^3 \bar{L}}{\partial \dot{q}^u \partial \dot{q}^s \partial q_t}\dot{q}^t \\ = -\frac{\partial \Lambda_s{}^r}{\partial q^t}\dot{q}^t \frac{\partial^2 L}{\partial \dot{q}^r \partial \dot{q}^u}.\end{aligned} \tag{29}$$

Therefore, Eq. (25) becomes

$$\begin{aligned}\frac{\partial^2 \bar{L}}{\partial \dot{q}^s \partial q^u} - \frac{\partial^2 \bar{L}}{\partial \dot{q}^u \partial q^s} = -\frac{d\Lambda_s{}^r}{dt}W_{ru} \\ + \Lambda_s{}^r\left(\frac{\partial^2 L}{\partial \dot{q}^r \partial q^u} - \frac{\partial^2 L}{\partial \dot{q}^u \partial q^r}\right);\end{aligned} \tag{30}$$

i.e., defining

$$T_{ru} \equiv \frac{\partial^2 L}{\partial \dot{q}^r \partial q^u} - \frac{\partial^2 L}{\partial \dot{q}^u \partial q^r} = -T_{ur}, \tag{31a}$$

$$\bar{T}_{su} \equiv \frac{\partial^2 \bar{L}}{\partial \dot{q}^s \partial q^u} - \frac{\partial^2 \bar{L}}{\partial \dot{q}^u \partial q^s} = -\bar{T}_{us}, \tag{31b}$$

one gets, finally,

$$\dot{\Lambda} = -\bar{T}U + \Lambda T U. \tag{32}$$

It can be directly proved that

$$(\mathrm{tr}(\Lambda)^m)^{\cdot} = 0, \tag{33}$$

for any integer m, using the fact that T and $\bar{T}$ are antisymmetric, U and $\bar{W}$ are symmetric, and $\Lambda = \bar{W}U$. Of course, not all these constants are functionally independent because, for any $n \times n$ matrix Λ, $\mathrm{tr}(\Lambda)^k$ can be written algebraically in terms of $\mathrm{tr}(\Lambda)^i$, $i = 1,\ldots,n$ for $k \geqslant n+1$. Furthermore, it may happen that even the trace of the first n powers of Λ are not functionally independent. This result agrees with the one obtained when discussing the one-dimensional case[1] and can be interpreted by saying that all the invariants (or eigenvalues) of Λ are constants of the motion.

The theorem we have just proved will be helpful in the construction of examples in Sec. V. It is worth noting that the result is a natural generalization of the analogous theorem discussed in Sec. II.

We should anticipate that in some of the examples the condition $\det\Lambda \neq 0$ may be relaxed without losing s-equivalence.

We will now define as l as the difference between $\bar{L}$ and ρL ($\rho \neq 0$):

$$l = \bar{L}(q,\dot{q},t) - \rho L(q,\dot{q},t). \tag{34}$$

We have that $l_r = \bar{L}_r - \rho L_r$, but $\bar{L}_r = \Lambda_r{}^s L_s$, so that

$$l_r = (\Lambda_r{}^s - \rho\delta_r{}^s)L_s \equiv \Omega_r{}^s L_s, \tag{35}$$

where

$$\Omega = \Lambda - \rho I. \tag{36}$$

We have that $\det\Lambda \neq 0$, but this fact does not imply $\det\Omega \neq 0$ so that, although $\{L_s = 0\}$ implies $\{l_r = 0\}$, $\{l_r = 0\}$ does not imply $\{L_s = 0\}$ in general; i.e., the equations of motion of l are valid whenever the ones for L are, but not conversely. For instance, if L can be written as

$$L = L(q^i,\dot{q}^i,t) = L_1(q^a,\dot{q}^a,t) + L_2(q^b,\dot{q}^b,t), \tag{37}$$

such that L_1 and L_2 have no variables in common, then

$$l_1 = \lambda_1 L_1 \quad \text{and} \quad l_2 = \lambda_2 L_2, \quad -\rho \neq \lambda_1 \neq 0 \neq \lambda_2 \neq -\rho,$$

are two examples of such an l; i.e.,

$$\bar{L}_1 = \rho L + \lambda_1 L_1 \quad \text{and} \quad \bar{L}_2 = \rho L + \lambda_2 L_2 \tag{38}$$

are both equivalent to L and

$$\{L_r = 0\} \Rightarrow \{l_{1a} = 0\} \quad \text{but} \quad \{l_{1a} = 0\} \not\Rightarrow \{L_r = 0\},$$

$$\{L_r = 0\} \Rightarrow \{l_{2b} = 0\} \quad \text{but} \quad \{l_{2b} = 0\} \not\Rightarrow \{L_r = 0\}.$$

One explicit example is

$$L = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2) - \tfrac{1}{2}(x^2 + y^2), \tag{39}$$

$$l_1 = \tfrac{1}{2}\lambda_1(\dot{x}^2 - x^2), \quad l_2 = \tfrac{1}{2}\lambda_2(\dot{y}^2 - y^2), \tag{40}$$

$$\bar{L}_2 = \tfrac{1}{2}[\dot{x}^2 - \dot{y}) - (x^2 - y^2)], \quad \rho = 1,\ 2\ \ \lambda_1 = 0,\ 2$$

$$\lambda_2 = 2. \tag{41}$$

[The case $l = -dF(q,t)/dt$ is, of course, trivial with $l_r = 0$ being identities.]

It is interesting to note that in some of the examples we will consider $\det\Lambda = 0$ as a possible equation of motion, and in spite of this fact we will still obtain s-equivalence; i.e., $\det\bar{W}$ may vanish even if $\det W \neq 0$ and this will not give rise to solutions of the equations of motion $\bar{L}_r = 0$, not already contained in the set of all solutions of $L_s = 0$. A theorem discussing in which cases this situation is possible is currently under investigation.

IV. EXAMPLES

Using the results obtained in Sec. III, we will now construct three examples of s-equivalent Lagrangians; two for

the bidimensional simple harmonic oscillator (BSHO), and another one for the same mechanical system, but in three dimensions (TSHO).

Example 1: Consider the usual Lagrangian associated with the BSHO:

$$L = T - V = \tfrac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2) - \tfrac{1}{2}(q_1^2 + q_2^2); \tag{42}$$

the related equations of motion for L are

$$L_i \equiv \ddot{q}_i + q_i = 0, \quad i = 1,2. \tag{43}$$

Then, since Eq. (35) is an identity, linear in the accelerations $(\ddot{q}^r)$, it can be separated into two identities: one for the terms containing accelerations and another one for the rest of the terms; that is,

$$\frac{\partial^2 l}{\partial \dot{q}^r \partial \dot{q}^t} \equiv \Omega_{rt}, \tag{44a}$$

$$\frac{\partial^2 l}{\partial \dot{q}^r \partial q^t}\dot{q}^t + \frac{\partial^2 l}{\partial \dot{q}^r \partial t} - \frac{\partial l}{\partial q^r} \equiv \Omega_r{}^s q_s, \tag{44b}$$

where we have used L given in Eq. (42).

Since the trace of Ω and Ω^2 must be constant, and since Ω is symmetric, according to Eq. (44a) we propose Ω to have the following structure:

$$\Omega = \begin{pmatrix} E & C \\ C & E \end{pmatrix}, \tag{45}$$

where $E = \tfrac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2 + q_1^2 + q_2^2)$ is the mechanical energy of the BSHO and the quantity C is a constant of motion whose dependence on q^i and $\dot{q}^i$ is to be determined from Eq. (44).

With this choice of Ω we can rewrite Eq. (44) explicitly; thus,

$$\frac{\partial^2 l}{\partial \dot{q}_1 \partial \dot{q}_1} \equiv E, \tag{46a}$$

$$\frac{\partial^2 l}{\partial \dot{q}_2 \partial \dot{q}_2} \equiv E, \tag{46b}$$

$$\frac{\partial^2 l}{\partial \dot{q}_1 \partial \dot{q}_2} \equiv C, \tag{46c}$$

$$\frac{\partial^2 l}{\partial \dot{q}_1 \partial q_1}\dot{q}_1 + \frac{\partial^2 l}{\partial \dot{q}_1 \partial q_2}\dot{q}_2 + \frac{\partial^2 l}{\partial \dot{q}_1 \partial t} - \frac{\partial l}{\partial q_1} \equiv Eq_1 + Cq_2, \tag{46d}$$

$$\frac{\partial^2 l}{\partial \dot{q}_2 \partial q_1}\dot{q}_1 + \frac{\partial^2 l}{\partial \dot{q}_2 \partial q_2}\dot{q}_2 + \frac{\partial^2 l}{\partial \dot{q}_2 \partial t} - \frac{\partial l}{\partial q_2} \equiv Cq_1 + Eq_2. \tag{46e}$$

One solution to these equations is

$$C = \dot{q}_1\dot{q}_2 + q_1q_2, \tag{47}$$

$$l = \tfrac{1}{24}(\dot{q}_1^4 + \dot{q}_2^4) + \tfrac{1}{4}\dot{q}_1^2\dot{q}_2^2 + \tfrac{1}{4}(q_1^2 + q_2^2)(\dot{q}_1^2 + \dot{q}_2^2) + q_1q_2\dot{q}_1\dot{q}_2 - \tfrac{3}{4}q_1^2q_2^2 - \tfrac{1}{8}(q_1^4 + q_2^4). \tag{48}$$

The equations of motion obtained from l are then

$$l_1 = E(\ddot{q}_1 + q_1) + C(\ddot{q}_2 + q_2) = 0, \tag{49a}$$

$$l_2 = C(\ddot{q}_1 + q_1) + E(\ddot{q}_2 + q_2) = 0, \tag{49b}$$

or

$$l_r = \Omega_r{}^s L_s = 0, \quad r,s = 1,2, \tag{50}$$

which is the expected result, with Ω given in (45) and L_s given in (43).

On calculating the determinant of Ω, we obtain

$$\det\Omega = (E^2 - C^2) = (E - C)(E + C), \tag{51}$$

which is always a quantity greater than or equal to zero. In order to see this clearly, define the vectors

$$\mathbf{u}_i = (1/\sqrt{2})(\dot{q}_i, q_i), \quad i = 1,2, \tag{52}$$

with which we have

$$E = \mathbf{u}_1^2 + \mathbf{u}_2^2, \tag{53a}$$

$$C = 2\mathbf{u}_1 \cdot \mathbf{u}_2, \tag{53b}$$

so that

$$\det\Omega = (E - C)(E + C) = (\mathbf{u}_1 - \mathbf{u}_2)^2(\mathbf{u}_1 + \mathbf{u}_2)^2 \geqslant 0. \tag{54}$$

It is interesting to note that if

$$\det\Omega = 0, \tag{55}$$

the solutions to this Eq. (55) are particular solutions to Eqs. (43) since $\det\Omega = 0$ implies $\mathbf{u}_1 = \mathbf{u}_2$ and/or $\mathbf{u}_1 = -\mathbf{u}_2$. If $\mathbf{u}_1 = \mathbf{u}_2$, then $E = C$, the rank of Ω is 1 whenever $E \neq 0$, and Eqs. (49) both reduce to $2(\dot{q}_1^2 + q_1^2)(\ddot{q}_1 + q_1) = 0$; that is, we have two equations whose solutions are contained in the set of solutions to Eqs. (49). If $\mathbf{u}_1 = -\mathbf{u}_2$, then $E = -C$, the rank of Ω is 1 (whenever $E \neq 0$), and again we have two equations whose solutions are contained in the set of solutions to Eqs. (49). Finally, if $\mathbf{u}_1 = \mathbf{u}_2$ *and* $\mathbf{u}_1 = -\mathbf{u}_2$, then $q_i = \dot{q}_i = 0$, and we have the oscillator at rest. Thus, we have s-equivalence even when we relax the condition $\det\Omega \neq 0$.

If we now consider relation (34), i.e.,

$$\bar{L} = \rho L + l \quad (\rho > 0), \tag{56}$$

the equations of motion related to $\bar{L}$ are

$$\begin{pmatrix} \bar{L}_1 \\ \bar{L}_2 \end{pmatrix} = \begin{pmatrix} E + \rho & C \\ C & E + \rho \end{pmatrix}\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{57}$$

where

$$\Lambda = \begin{pmatrix} E + \rho & C \\ C & E + \rho \end{pmatrix}, \tag{58}$$

and we have

$$\det\Lambda = (E^2 - C^2) + 2\rho E + \rho^2 > 0, \tag{59}$$

so that we have s-equivalence among L and $\bar{L}$, because Λ is invertible. The Hamiltonian obtained from the Lagrangian (48) is

$$h = \tfrac{1}{8}[\dot{q}_1^4 + \dot{q}_2^4 + q_1^4 + q_2^4] + \tfrac{3}{4}(q_1^2q_2^2 + \dot{q}_1^2\dot{q}_2^2) + \tfrac{1}{4}(q_1^2 + q_2^2)(\dot{q}_1^2 + \dot{q}_2^2) + q_1q_2\dot{q}_1\dot{q}_2,$$

where

$$\dot{q}_1 = \tfrac{1}{2}(S_1 + S_2 + T_1 + T_2),$$

$$\dot{q}_2 = \tfrac{1}{2}(S_1 - S_2 + T_1 - T_2),$$

with

$$S_1 = \{3(p_1 + p_2) + [(q_1 + q_2)^6 + 9(p_1 + p_2)^2]^{1/2}\}^{1/3},$$

$$S_2 = \{3(p_1 - p_2) + [(q_1 - q_2)^6 + 9(p_1 - p_2)^2]^{1/2}\}^{1/3},$$

$$T_1 = \{3(p_1 + p_2) - [(q_1 + q_2)^6 + 9(p_1 + p_2)^2]^{1/2}\}^{1/3},$$

$$T_2 = \{3(p_1 - p_2) - [(q_1 - q_2)^6 + 9(p_1 - p_2)^2]^{1/2}\}^{1/3},$$

$$p_1 = \frac{\partial l}{\partial \dot{q}_1}, \quad p_2 = \frac{\partial l}{\partial \dot{q}_2}.$$

Example 2: If we now chose Ω to be given by

$$\Omega = \begin{pmatrix} C & E \\ E & C \end{pmatrix}, \tag{60}$$

Eqs. (54) take the form

$$\frac{\partial^2 l}{\partial \dot{q}_1 \partial \dot{q}_1} \equiv C, \tag{61a}$$

$$\frac{\partial^2 l}{\partial \dot{q}_2 \partial \dot{q}_2} \equiv C, \tag{61b}$$

$$\frac{\partial^2 l}{\partial \dot{q}_1 \partial \dot{q}_2} \equiv E, \tag{61c}$$

$$\frac{\partial^2 l}{\partial \dot{q}_1 \partial q_1}\dot{q}_1 + \frac{\partial^2 l}{\partial \dot{q}_1 \partial q_2}\dot{q}_2 + \frac{\partial^2 l}{\partial \dot{q}_1 \partial t} - \frac{\partial l}{\partial q_1} \equiv Cq_1 + Eq_2, \tag{61d}$$

$$\frac{\partial^2 l}{\partial \dot{q}_2 \partial q_1}\dot{q}_1 + \frac{\partial^2 l}{\partial \dot{q}_2 \partial q_2}\dot{q}_2 + \frac{\partial^2 l}{\partial \dot{q}_2 \partial t} - \frac{\partial l}{\partial q_2} \equiv Eq_1 + Cq_2, \tag{61e}$$

One solution to these equations is

$$l = \tfrac{1}{6}\dot{q}_1\dot{q}_2(\dot{q}_1^2 + \dot{q}_2^2) + \tfrac{1}{2}\dot{q}_1\dot{q}_2(q_1^2 + q_2^2) + \tfrac{1}{2}q_1q_2(\dot{q}_1^2 + \dot{q}_2^2) - \tfrac{1}{2}q_1q_2(q_1^2 + q_2^2). \tag{62}$$

The equations of motion obtained from this Lagrangian are then

$$l_1 = C(\ddot{q}_1 + q_1) + E(\ddot{q}_2 + q_2) = 0, \tag{63a}$$

$$l_2 = E(\ddot{q}_1 + q_1) + C(\ddot{q}_2 + q_2) = 0, \tag{63b}$$

which is again the expected result for Ω given in (60).

With arguments similar to those stated in Example 1, we conclude here that we also have s-equivalence, even when we relax the condition $\det\Omega \neq 0$. Also, with relation (34) ($\bar{L} = \rho L + 1$, $\rho > 0$) we obtain $\bar{L}$ which is s-equivalent to L. The Hamiltonian obtained from the Lagrangian (62) is

$$h = \tfrac{1}{2}(q_1q_2 + \dot{q}_1\dot{q}_2)(\dot{q}_1^2 + \dot{q}_2^2 + q_1^2 + q_2^2),$$

where

$$\dot{q}_1 = \tfrac{1}{2}(S_1 + S_2 + T_1 + T_2),$$

$$\dot{q}_2 = \tfrac{1}{2}(S_1 - S_2 + T_1 - T_2),$$

with

$$S_1 = \{3(p_1 + p_2) + [(q_1 + q_2)^6 + 9(p_1 + p_2)^2]^{1/2}\}^{1/3},$$

$$S_2 = \{-3(p_1 - p_2) + [(q_1 - q_2)^6 + 9(p_1 - p_2)^2]^{1/2}\}^{1/3},$$

$$T_1 = \{3(p_1 + p_2) - [(q_1 + q_2)^6 + 9(p_1 + p_2)^2]^{1/2}\}^{1/3},$$

$$T_2 = \{-3(p_1 - p_2) - [(q_1 - q_2)^6 + 9(p_1 - p_2)^2]^{1/2}\}^{1/3},$$

$$p_1 = \frac{\partial l}{\partial \dot{q}_1}, \quad p_2 = \frac{\partial l}{\partial \dot{q}_2}.$$

Example 3: Consider now the usual Lagrangian associated with the TSHO

$$L = \tfrac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2) - \tfrac{1}{2}(q_1^2 + q_2^2 + q_3^2). \tag{64}$$

The related equations of motion fo L are now

$$L_i \equiv \ddot{q}_i + q_i = 0, \quad i = 1,2,3. \tag{65}$$

The method we used in the last two examples for constructing both s-equivalent Lagrangians for the BSHO case is also useful in the three-dimensional case, although the calculations become much more tedious. Nevertheless, we can take advantage of the symmetry in the Lagrangian obtained in the first example and given by Eq. (48) from which we will construct the three-dimensional s-equivalent Lagrangian by just symmetrizing this expression to include the third coordinate (q_3). After adding the necessary terms, we obtain

$$\begin{aligned} l = {} & \tfrac{1}{24}(\dot{q}_1^4 + \dot{q}_2^4 + \dot{q}_3^4) + \tfrac{1}{4}(\dot{q}_1^2\dot{q}_2^2 + \dot{q}_1^2\dot{q}_3^2 + \dot{q}_2^2\dot{q}_3^2) \\ & + \tfrac{1}{4}(q_1^2 + q_2^2 + q_3^2)(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2) \\ & + q_1q_2\dot{q}_1\dot{q}_2 + q_1q_3\dot{q}_1\dot{q}_3 + q_2q_3\dot{q}_2\dot{q}_3 \\ & - \tfrac{3}{4}(q_1^2q_2^2 + q_1^2q_3^2 + q_2^2q_3^2) \\ & - \tfrac{1}{8}(q_1^4 + q_2^4 + q_3^4). \end{aligned} \tag{66}$$

The equations of motion obtained from this Lagrangian are then:

$$l_1 = E(\ddot{q}_1 + q_1) + C_{12}(\ddot{q}_2 + q_2) + C_{13}(\ddot{q}_3 + q_3) = 0, \tag{67a}$$

$$l_2 = C_{21}(\ddot{q}_1 + q_1) + E(\ddot{q}_2 + q_2) + C_{23}(\ddot{q}_3 + q_3) = 0, \tag{67b}$$

$$l_3 = C_{31}(\ddot{q}_1 + q_1) + C_{32}(\ddot{q}_2 + q_2) + E(\ddot{q}_3 + q_3) = 0, \tag{67c}$$

which can be written concisely in the form

$$\begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \begin{pmatrix} E & C_{12} & C_{13} \\ C_{21} & E & C_{23} \\ C_{31} & C_{32} & E \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{67'}$$

where $E = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 + q_1^2 + q_2^2 + q_3^2)$ is the total energy of the oscillator and the quantities C_{ij} are all constants and are given by

$$C_{ij} = q_iq_j + \dot{q}_i\dot{q}_j. \tag{68}$$

In this case we then have

$$\Omega = \begin{pmatrix} E & C_{12} & C_{13} \\ C_{21} & E & C_{23} \\ C_{31} & C_{32} & E \end{pmatrix}. \tag{69}$$

We can define again the vectors $\mathbf{u}_i$ as in the previous two examples [cf. Eq. (52)]. If $E = 0$, then $\mathbf{u}_i = 0$, for every i, and we have the oscillator at rest. Now, taking $E > 0$, the determinant of Ω can be written:

$$\det\Omega = E^3 \begin{vmatrix} 1 & 2\mathbf{u}_1\cdot\mathbf{u}_2/E & 2\mathbf{u}_1\cdot\mathbf{u}_3/E \\ 2\mathbf{u}_2\cdot\mathbf{u}_1/E & 1 & 2\mathbf{u}_2\cdot\mathbf{u}_3/E \\ 2\mathbf{u}_3\cdot\mathbf{u}_1/E & 2\mathbf{u}_3\cdot\mathbf{u}_2/E & 1 \end{vmatrix}. \tag{70}$$

Using the inequalities

$$(\mathbf{u}_i - \mathbf{u}_j)^2 + \mathbf{u}_k^2 \geqslant 0,$$

$$-(\mathbf{u}_i + \mathbf{u}_j)^2 - \mathbf{u}_k^2 \leqslant 0,$$

where $i \neq j \neq k \neq i; i,j,k = 1,2,3$, it is easy to show that $-1 \leqslant 2\mathbf{u}_i\cdot\mathbf{u}_j/E \leqslant 1$, which means we can define the unitary vectors $\hat{v}_i$ such that

$$\hat{v}_i\cdot\hat{v}_i = 1, \tag{71a}$$

$$\hat{v}_i\cdot\hat{v}_j = 2\mathbf{u}_i\cdot\mathbf{u}_j/E, \quad i,j = 1,2,3. \tag{71b}$$

This fact allows us to write the determinant of Ω as the square of a scalar triple product of unitary vectors times E cubed, that is,

$$\det\Omega = E^3[\hat{v}_1 \times \hat{v}_2\cdot\hat{v}_3]^2, \tag{72}$$

which is obviously greater than or equal to zero, since $E > 0$. With this in mind, and using relation (34) ($\bar{L} = \rho L + l$,

$\rho > 0$), once again, we have

$$\Lambda = \begin{pmatrix} E+\rho & C_{12} & C_{13} \\ C_{21} & E+\rho & C_{23} \\ C_{31} & C_{32} & E+\rho \end{pmatrix}, \tag{73}$$

from which we get

$$\det\Lambda = \det\Omega + (3E^2 - C_{12}^2 - C_{13}^2 - C_{23}^2)\rho + 3E\rho^2 + \rho^3 > 0,$$

because $E^2 \geqslant C_{ij}^2$ due to the fact that $-1 \leqslant C_{ij}/E \leqslant 1$. This means the family of Lagrangians $\bar{L}$ is s-equivalent to L, because Λ is invertible.

V. CONCLUSIONS AND OUTLOOK

We have generalized the results of Ref. 1 to the case of several dimensions. The theorem which states that the trace of all integer powers of Λ are constants of the motion allows one to construct examples in which the elements of Λ are functions of q_i, $\dot{q}_i$, and t (without reiterating the one-dimensional procedure). A particularly interesting result is that it is possible to retain s-equivalence even when $\det\Lambda$ vanishes, and this fact is important when the quantum theory is considered. For the two-dimensional examples we have shown that s-equivalence prevails even when the rank of Λ is zero. A general theorem which would allow one to determine in which cases this is possible is currently under investigation.[16]

Some problems which remain unsolved are:

(i) the general solution of Eqs. (35) (for l and Ω) in any number of dimensions;

(ii) the consequences in quantum theory of the fact that classical realistic systems can be described equally well in more than one way be means of s-equivalent Lagrangians[2,3];

(iii) the study of the problem discussed here within the realm of gauge and field theories.[17]

ACKNOWLEDGMENTS

We would like to thank Luis de la Peña, Santiago López de Medrano, and Michael P. Ryan Jr. for fruitful discussions and encouragement. We also thank the referee for useful suggestions.

[1]D. F. Currie and E. J. Saletan, J. Math. Phys. **7**, 967 (1966).
[2]G. Marmo and E. J. Saletan, Hadronic J. **1**, 955 (1978).
[3]S. Hojman and R. Montemayor, Hadronic J. **3**, 1644 (1980).
[4]R. M. Santilli, Ann. Phys. (N. Y.) **103**, 354 (1977).
[5]R. M. Santilli, Ann. Phys. (N. Y.) **103**, 409 (1977).
[6]R. M. Santilli, Ann. Phys. (N. Y.) **105**, 227 (1977).
[7]R. M. Santilli, *Foundations of Theoretical Mechanics. I* (Springer, New York, 1978).
[8]R. M. Santilli, Hadronic J. **1**, 223 (1978).
[9]R. M. Santilli, Hadronic J. **1**, 574 (1978).
[10]Y. Gelman and E. J. Saletan, Nuovo Cimento B **18**, 62 (1973).
[11]G. Marmo and E. J. Saletan, Nuovo Cimento B **40**, 67 (1977).
[12]D. G. Currie and E. J. Saletan, Nuovo Cimento B **9**, 143 (1972).
[13]G. Darboux, *Leçons sur la théorie générale des surfaces* (Gauthier-Villars, Paris, 1891).
[14]O. Bolza, *Lectures on the Calculus of Variations* (Dover, New York, 1961).
[15]J. Douglas, "Solution of the Inverse Problem of the Calculus of Variations," Trans. Am. Math. Soc. **50**, 71 (1941).
[16]Some work along these lines has been done, for the one-dimensional case, by S. López de Medrano and M. Lara Aparicio (private communication).
[17]S. Hojman and L. Shepley, in preparation.

Decomposition of vector fields and mixed dynamics

R. Vilela Mendes and J. Taborda Duarte
CFMC-Instituto Nacional de Investigação Cientifica, Av. Gama Pinto, 2-1699 Lisboa Codex, Portugal

(Received 16 July 1980; accepted for publication 20 January 1981)

Some theorems are proved concerning the decomposition of vector fields into gradient and Hamiltonian components. A constructive method to carry out one of the decompositions is applied to some three- and four- dimensional dynamical models.

PACS numbers: 03.20. + i

1. DECOMPOSITION OF VECTOR FIELDS

A classical dynamical system is defined by the couple (M,X), M being a differentiable manifold and X a C^r vector field. The study of simple ways to describe general vector fields will lead therefore to a parametrization and classification of classical dynamical systems. A step in this direction was taken by Roels[1] who proved that in a two-dimensional symplectic manifold every vector field is locally the sum of a Hamiltonian and a gradient field. Our purpose in this paper is to obtain similar decompositions for N-dimensional manifolds. The proof of the main result uses the following local lemma.

Lemma : Let R^N (N even) be endowed with the canonical scalar product. Then on every compact neighborhood Ω there are $N-1$ nondegenerate 2-forms α_i with the properties :

(a) $d\alpha_i = 0$,

(b) $\underbrace{\alpha_i \wedge \cdots \wedge \alpha_i}_{N/2} = (N/2)!v$ (v volume form on R^N),

(c) $*\alpha_i = \dfrac{1}{(N/2-1)!}\underbrace{\alpha_i \wedge \cdots \wedge \alpha_i}_{N/2-1}$,

(d) $*\alpha_i \wedge \alpha_j = 0, \quad i \neq j$,

and such that given a C^∞ 2-form η, there are $N-1$ C^∞ functions a_i and a 2-form α satisfying locally:

(1) $\delta\alpha = \delta\eta$,

(2) $\alpha = \sum_{i=1}^{N-1} a_i \alpha_i$.

For the proof one uses an Euclidean coordinate system. In these coordinates a constructive recipe for a set of 2-forms α_i is

$$\begin{aligned}
\alpha_1 &= dx^1 \wedge dx^2 + dx^{i_{13}} \wedge dx^{i_{14}} + \cdots + dx^{i_{1(N-1)}} \wedge dx^{i_{1N}},\\
\alpha_2 &= dx^1 \wedge dx^3 + dx^{i_{23}} \wedge dx^{i_{24}} + \cdots + dx^{i_{2(N-1)}} \wedge dx^{i_{2N}},\\
&\cdots\\
\alpha_{N-1} &= dx^1 \wedge dx^N + dx^{i_{(N-1)3}} \wedge dx^{i_{(N-1)4}} + \cdots\\
&\quad + dx^{i_{(N-1)(N-1)}} \wedge dx^{i_{(N-1)N}},
\end{aligned} \tag{1.1}$$

where in α_p the numbers $1, p+1, i_{p3}, i_{p4}, \cdots, i_{pN}$ are an even permutation of $1 \cdots N$, and no elementary 2-form $dx^i \wedge dx^j$ appears more than once in (1.1). For example, for $N = 4$ and 6, one has

$$\begin{aligned}
\alpha_1 &= dx^1 \wedge dx^2 + dx^3 \wedge dx^4,\\
\alpha_2 &= dx^1 \wedge dx^3 + dx^4 \wedge dx^2,\\
\alpha_3 &= dx^1 \wedge dx^4 + dx^2 \wedge dx^3,\\
\alpha_1 &= dx^1 \wedge dx^2 + dx^3 \wedge dx^4 + dx^5 \wedge dx^6,\\
\alpha_2 &= dx^1 \wedge dx^3 + dx^2 \wedge dx^5 + dx^4 \wedge dx^6,\\
\alpha_3 &= dx^1 \wedge dx^4 + dx^2 \wedge dx^6 + dx^3 \wedge dx^5,\\
\alpha_4 &= dx^1 \wedge dx^5 + dx^2 \wedge dx^4 + dx^3 \wedge dx^6,\\
\alpha_5 &= dx^1 \wedge dx^6 + dx^2 \wedge dx^3 + dx^4 \wedge dx^5.
\end{aligned} \tag{1.2}$$

It is straightforward to check that for general N the forms constructed according to (1.1) are nondegenerate and satisfy the conditions (a)–(d). Clearly for $N \geqslant 6$ one does not obtain a unique set.

To prove the lemma, one should now check that given a smooth 2-form η it is possible to find $N-1$ functions a_i such that

$$\delta \sum_{i=1}^{N-1} a_i \alpha_i = \delta\eta. \tag{1.3}$$

In Euclidean coordinates the codifferential of a 2-form β reads $\delta\beta = \partial_j \beta_{ij}\, dx^i$. Therefore, from the knowledge of the α_i forms (1.1), one writes (1.3) as a simple first-order partial differential system. To avoid the introduction of cumbersome index notation we will merely illustrate this for $N = 4$ and $N = 6$:

$$\begin{vmatrix} \partial_2 & \partial_3 & \partial_4 \\ -\partial_1 & -\partial_4 & \partial_3 \\ \partial_4 & -\partial_1 & -\partial_2 \\ -\partial_3 & \partial_2 & -\partial_1 \end{vmatrix} \begin{vmatrix} a_1 \\ a_2 \\ a_3 \end{vmatrix} = \sum_j \partial_j \eta_{ij},$$

$$\begin{vmatrix} \partial_2 & \partial_3 & \partial_4 & \partial_5 & \partial_6 \\ -\partial_1 & \partial_5 & \partial_6 & \partial_4 & \partial_3 \\ \partial_4 & -\partial_1 & \partial_5 & \partial_6 & -\partial_2 \\ -\partial_3 & \partial_6 & -\partial_1 & -\partial_2 & \partial_5 \\ \partial_6 & -\partial_2 & -\partial_3 & -\partial_1 & -\partial_4 \\ -\partial_5 & -\partial_4 & -\partial_2 & -\partial_3 & -\partial_1 \end{vmatrix} \begin{vmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{vmatrix} = \sum_j \partial_j \eta_{ij}.$$

The general rule for writing the matrix of partial derivatives in

$$\sum_{j=1}^{N-1} M(\partial)_{ij} a_j = \sum_{j=1}^{N} \partial_j \eta_{ij} \quad (i = 1 \cdots N) \tag{1.4}$$

is that, if the form $dx^i \wedge dx^j$ occurs in α_r, then the i, r and j, r matrix elements are ∂_j and $-\partial_i$.

By a smooth truncation of $\Sigma \partial_j \eta_{ij}$ outside the neighborhood Ω, one replaces the system (1.4) by

$$\sum_{j=1}^{N-1} M(\partial)_{ij} a_j = u_i \quad (i = 1 \cdots N), \tag{1.5}$$

where the u_i are $C_0^\infty(R^N)$ functions that coincide with $\Sigma \partial_j \eta_{ij}$ in Ω. The solutions of (1.5) will also coincide with solutions of (1.4) in Ω.

The system (1.5) has $N-1$ unknowns a_j and N equations. However, not all equations are independent because by construction

$$\sum_{i=1}^{N} \xi_i M(\xi)_{ij} = 0 \quad (j = 1 \cdots N-1).$$

Eliminating one of the rows in $M(\partial)_{ij}$, one is led to a system of $N-1$ equations with $N-1$ unknowns,

$$\sum_{j=1}^{N-1} \tilde{M}(\partial)_{ij} a_j = u_i \quad (i = 1 \cdots N-1), \tag{1.6}$$

whose $\det \tilde{M}(\xi)$ is not identically zero.

The existence of a fundamental solution ($\tilde{M}(\partial)E = \delta 1$) follows from the existence of a fundamental solution for the differential operator with constant coefficients $\det \tilde{M}(\partial)$.[2] From the fundamental solution, by convolution with the C_0^∞ functions u_i, one finally proves the existence of C^∞ solutions a_j to (1.4) in Ω.

For $N = 4$ the lemma is equivalent to the statement that there is a self-dual α such that $\delta\alpha = \delta\eta$. In this case one can apply the Hodge–de Rham theorem to write $\eta = d\beta + \delta\gamma$, and choosing $\alpha = d\beta + {}^*d\beta$, one proves the assertion in a coordinate free manner. Unfortunately, we could not find a similar proof for higher dimensions.

On the other hand, the coordinatewise proof of the lemma and in particular the system (1.4) provides a constructive method to obtain in practice the decomposition of vector fields whose existence is asserted in the following theorem.

Theorem 1: Given an N-dimensional (N even) C^∞-manifold we can find for every $x \in M$ a nbd Ω of x, a Riemannian metric $\tilde{g}$, and $N-1$ symplectic forms $\tilde{\alpha}_i$ on Ω such that every vector field X defined on the nbd can be decomposed into one gradient and $N-1$ Hamiltonian fields.

Let $\phi: \Omega \to R^N$ be a chart around x such that $\phi(x) = 0$. Defining $\tilde{g}$ as the pullback by ϕ of the Euclidean metric, and $\tilde{\alpha}_i$ as $\phi^*(\alpha_i)$, we observe that the 2-forms $\tilde{\alpha}_i$ have the same properties as the forms α_i in the lemma.

Let $\tilde{g}^b: \mathfrak{X}(\Omega) \to \Omega^1(\Omega)$ be the isomorphism from the vector fields onto the 1-forms induced by $\tilde{g}$, and $\tilde{g}^\#$ its inverse. By the Hodge–de Rham theorem and Poincaré's lemma,

$$\tilde{g}^b(X) = dS + \delta\tilde{\eta}.$$

Hence $X = \tilde{g}^\#(dS) + \tilde{g}^\#(\delta\tilde{\eta})$, and $\tilde{g}^\#(dS)$ is a gradient vector field. For $\delta\tilde{\eta}$ we can write $\delta\tilde{\eta} = \delta(\Sigma_i^{N-1} b_i \tilde{\alpha}_i)$, where b_i are C functions defined on Ω, simply by carrying $\tilde{\eta}$ to R^N by the chart and applying the lemma. We have then

$$\tilde{g}^\#(\delta\tilde{\eta}) = \sum_i^{N-1} \tilde{g}^\#(\delta(b_i \tilde{\alpha}_i)).$$

It remains to prove that each $\tilde{g}^\#(\delta(b_i\tilde{\alpha}_i))$ is a Hamiltonian vector field for the symplectic form $\tilde{\alpha}_i$. This follows from a computation that uses the properties of the $\tilde{\alpha}_i$ and the equalities $i(Z)\alpha = \tilde{*}(\tilde{*}\alpha \wedge \tilde{g}^b(Z))$, and

$$\tilde{*} i(Z)\tilde{v} = -(-1)^N \tilde{g}_b(Z),$$

where $\tilde{*}$ and $\tilde{v}$ are the Hodge star and the volume form associated to $\tilde{g}$:

$$\begin{aligned} i(\tilde{g}^\# \delta(b_i\tilde{\alpha}_i))\tilde{\alpha}_i &= -i(\tilde{g}^\# \tilde{*}(db_i \wedge \tilde{*}\tilde{\alpha}_i))\tilde{\alpha}_i \\ &= -\tilde{*}(\tilde{*}\tilde{\alpha}_i \wedge \tilde{*}(db_i \wedge \tilde{*}\tilde{\alpha}_i)) \\ &= \frac{-1}{(N/2-1)!} \tilde{*}(\tilde{\alpha}_i \wedge \cdots \wedge \tilde{\alpha}_i \wedge i(\tilde{g}^\# db_i)\tilde{\alpha}_i) \\ &= \frac{-1}{(N/2)!} \tilde{*} i(\tilde{g}^\# db_i)(\tilde{\alpha}_i \wedge \cdots \wedge \tilde{\alpha}_i) = -\tilde{*} i(\tilde{g}^\# db_i)\tilde{v} \\ &= db_i \end{aligned}$$

The theorem states that, locally at least, one can erect a system of $N-1$ symplectic forms that together with the metric form a fixed framework enabling us to decompose any smooth motion into elementary gradient and Hamiltonian components. This is the situation that seems to be the most useful for the applications. However, there exists a different decomposition problem when for a given vector field one is allowed to choose either a metric or a symplectic form adapted to that particular vector field. The following results are almost trivial consequences of the "flow box" theorem.[3]

Theorem 2: Let X be a vector field on a Riemannian manifold M_g. Then for each $p \in M_g$ there is a neighborhood Ω of p and a symplectic form ω_X on Ω such that X is decomposed in, at most, one gradient and one Hamiltonian vector field.

Proof: Take $p \in M$. Either $X(p) \neq 0$ or $X(p) = 0$. If $X(p) \neq 0$ by the flow box theorem there is a neiborhood Ω and a local diffeomorphism $\phi: \Omega \to R^N$, $\phi(y) = (y_1, \ldots, y_N)$ such that $\phi_*(X) = \partial/\partial y_1$. Then $\phi_*(X)$ is Hamiltonian for the canonical symplectic form $\omega = \Sigma_{i=1}^{N/2} dy_{2i-1} \wedge dy_{2i}$ in R^N. Then X is Hamiltonian in Ω for $\phi^*\omega = \omega_X$.

If $X(p) = 0$, take X_g any gradient vector field (for the metric g) such that $X_g(p) \neq 0$. Then $Y = X + X_g$ does not vanish at p and we can apply the above argument so that Y is Hamiltonian, i.e., $X = X_g - Y$ as stated.

Theorem 2′: Let X be a vector field on a symplectic manifold M_ω. Then for each $p \in M$, there is a neighborhood and a Riemannian metric g_X on Ω such that X is decomposed in, at most, one gradient and one Hamiltonian vector field.

Proof: Take $p \in M$. Either $X(p) \neq 0$ or $X(p) = 0$. If $X(p) \neq 0$ by the flow box theorem we can find a nbd of p and a metric g_X on Ω such that X be gradient. If $X(p) = 0$, we choose a Hamiltonian vector field X_ω such that $X_\omega(p) \neq 0$. Then we take $Y = X + X_\omega$ and apply again the same argument.

Although simpler than those of Theorem 1, these decompositions are probably of little practical value because to find the flow box coordinate system is equivalent to finding the orbits. Hence, to write such a decomposition should not be much simpler than to find an exact solution of the equations of motion.

The results in this paper imply that general classical motions are mixed Hamiltonian or mixed Hamiltonian-gradient systems. Besides the obvious parametrization usefulness of such decompositions they may also provide new ways of studying classical systems, for example by developing a perturbative theory of gradient deformations of Hamilton-

ian systems. Preliminary results in this direction indicate that at least it is then simple to establish necessary conditions for the existence of constants of motion in dissipative systems.[7]

2. EXAMPLES

The proofs of Theorem 1 and the lemma provide a constructive method to obtain the corresponding decomposition once a vector field X is given. In Euclidean coordinates the function S of the gradient part is obtained as a solution of the Poisson equation

$$\Delta S = \mathrm{div} X,$$

and with the choice of the symplectic forms (1.1) the Hamiltonian functions are obtained by solving the differential system (1.4). Using this method on finds:

= for the van der Pol oscillator,

$$\dot{x} = y = \frac{\partial S}{\partial x} + \frac{\partial H}{\partial y}, \quad S = \alpha\left(\frac{x^2}{2} - \frac{x^4}{12}\right),$$

$$\dot{y} = \alpha(1 - x^2)y - x = \frac{\partial S}{\partial y} - \frac{\partial H}{\partial x},$$

$$H = \frac{y^2}{2} - \alpha\left(x - \frac{x^3}{3}\right)y + \frac{x^2}{2}.$$

= for Rossler's model for "hyperchaos,"[4]

$$\dot{x} = -y - z = \frac{\partial S}{\partial z} + \frac{\partial H}{\partial y} + \frac{\partial H'}{\partial z},$$

$$\dot{y} = x + \frac{y}{4} + w = \frac{\partial S}{\partial y} - \frac{\partial H}{\partial x} - \frac{\partial H'}{\partial w},$$

$$\dot{z} = 3 + xz = \frac{\partial S}{\partial z} + \frac{\partial H}{\partial w} - \frac{\partial H'}{\partial x},$$

$$\dot{w} = -\frac{z}{2} + 0.05w = \frac{\partial S}{\partial w} - \frac{\partial H}{\partial z} + \frac{\partial H'}{\partial y},$$

with

$$S = \tfrac{1}{6}x^3 + 0.3w^2/2 \quad H = -\tfrac{1}{2}(x^2 + y^2) + \tfrac{1}{4}z^2 + 3w,$$
$$H' = -\tfrac{1}{2}x^2 z - \tfrac{1}{2}z^2 - \tfrac{1}{2}w^2 - \tfrac{1}{4}yw.$$

Systems of odd dimensionality N may always be imbedded into a $(N+1)$-dimensional manifold, the same methods become applicable and one obtains:

= for the Lorenz model[5]

$$\dot{x} = -\sigma x + \sigma y = \frac{\partial S}{\partial x} + \frac{\partial H}{\partial y},$$

$$\dot{y} = -xz + rx - y = \frac{\partial S}{\partial y} - \frac{\partial H}{\partial x},$$

$$\dot{z} = xy - bz = \frac{\partial S}{\partial z},$$

with

$$S = \tfrac{1}{2}\sigma x^2 - \tfrac{1}{2}y^2 - \tfrac{1}{2}bz^2 + xyz,$$
$$H = \tfrac{1}{2}y^2(\sigma - z) + x^2(z - \tfrac{1}{2}r).$$

= for the Gause–Lotka–Volterra equations (3 species)[6]

$$\dot{x} = x(1 - x - \alpha y - \beta z) = \frac{\partial S}{\partial x} + \frac{\partial H}{\partial y} + \frac{\partial H'}{\partial z},$$

$$\dot{y} = y(1 - \beta x - y - \alpha z) = \frac{\partial S}{\partial y} - \frac{\partial H}{\partial x},$$

$$\dot{z} = z(1 - \alpha x - \beta y - z) = \frac{\partial S}{\partial z} - \frac{\partial H'}{\partial x},$$

with

$$S = \tfrac{1}{2}(x^2 + y^2 + z^2) - \tfrac{1}{6}(2 + \alpha + \beta)(x^3 + y^3 + z^3),$$
$$H = -\tfrac{1}{2}(\alpha + \beta)xy^2 + y(\tfrac{1}{2}\beta x^2 + \alpha xz),$$
$$H' = -\tfrac{1}{2}(\alpha + \beta)xz^2 + z(\tfrac{1}{2}\alpha x^2 + \beta xy).$$

AKNOWLEDGMENTS

The authors are grateful to Professor Claude Bruter for bringing to their attention the potential value of generalizing Roels' Theorem.

[1] J. Roels, C. R. Acad. Sci. Ser. A **278**, 29 (1974).
[2] L. Hörmander, *Linear Partial Differential Operators* (Springer, Berlin, 1969), Chap. III.
[3] F. Dumortier, "Singularities of vector fields," IMPA math. monogr. No. 32, Rio de Janeiro, 1978, p. 8.
[4] O. E. Rossler, Phys. Lett.A **71**, 155 (1979).
[5] E. N. Lorenz; J. Atmos. Sci. **20**, 130 (1963).
[6] R. M. May, W. J. Leonard, SIAM J. Appl. Math. **29**, 243 (1975).
[7] J. T. Duarte and R. V. Mendes, "Deformation of Hamiltonian dynamics and constants of motion in dissipative systems" preprint CFMC, E-15/80.

An alternative derivation of transition amplitudes for time-dependent harmonic oscillators

H. E. De Meyer [a] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, R.U.G., Krijgslaan 271-S9, B-9000 Gent, Belgium

(Received 28 October 1980; accepted for publication 30 January 1981)

It is shown that the relations which link the x and p operators of a time-dependent harmonic oscillator to the x and p operators of the corresponding time-independent oscillator can be cast into a generalized Bogoliubov transformation on the standard harmonic oscillator boson creation and annihilation operators. Once this fact has been recognized, various techniques can be invoked to derive in a short way transition amplitudes.

PACS numbers: 03.65Db

I. INTRODUCTION

Utilizing the exact quantum mechanical Green's function, Landovitz *et al.*[1] were able to obtain a closed expression for the transition amplitudes in the case of an arbitrary time-dependent harmonic oscillator. Their derivation of that expression is mainly based on mathematical properties of the usual harmonic oscillator eigenstates, and hence on those of the Hermite polynomials. It appeared to us that the rather tedious computations resemble closely similar ones carried out by Tanabe,[2] in order to calculate transformation matrix elements for the linear boson Bogoliubov transformation, well-known in the theoretical developments of superfluidity and superconductivity. Since also the transition amplitudes of Ref. 1 exhibit striking similarities with Bogoliubov transformation matrix elements, our first concern has been to demonstrate that the relationship between the time-dependent and the usual time-independent quantum harmonic oscillator can be expressed as a generalized complex Bogoliubov transformation on boson operators. A proof of this statement is presented in Sec. II.

Once this underlying Bogoliubov transformation property is recognized, we have at our disposal a variety of very powerful techniques which enable one to deduce almost immediately the time-dependent harmonic oscillator transition amplitudes. Indeed, as alternatives of the laborious method of Tanabe,[2] we mention the recursive method of Rashid[3] and an operator based technique developed by Witschel[4] and by Kelemen.[5] The complex Bogoliubov transformation has also been studied by Tikochinsky[6] and by the present authors,[7] who extended it to the para-Bose situation. In Sec. III of this paper we rederive in slightly different form the expression for the transition amplitudes by means of results established by Tikochinsky.[6] Finally, the extension of the theory envisaged by Landovitz *et al.*[1] is discussed in the context of Witschel's theoretical work.[4]

II. THE BOLOGIUBOV TRANSFORMATION

It has been shown by Landovitz *et al.*[1] that the Hamiltonian H, defined as

$$H = f(t)p^2/2M + g(t)\tfrac{1}{2}M\omega_0^2 x^2, \tag{2.1}$$

with $f(0) = g(0) = 1$, accepts solutions for the operators $x_+(t)$ and $p_+(t)$ which can be expressed as

$$x_+ = a(t)x + b(t)p, \tag{2.2}$$

$$p_+ = c(t)x + d(t)p, \tag{2.3}$$

where $a(0) = d(0) = 1$ and $b(0) = c(0) = 0$. For given functions $f(t)$ and $g(t)$, the corresponding functions $a(t)$, $b(t)$, $c(t)$, and $d(t)$ are the solutions of a set of coupled linear homogeneous differential equations of first order.[1] Furthermore, the condition

$$[x_+, p_+] = [x,p] = i\hbar \tag{2.4}$$

implies that

$$a(t)d(t) - b(t)c(t) = 1, \tag{2.5}$$

a condition which was proved to be satisfied for all times.[1]

We now introduce a boson creation operator a^+ and a boson annihilation operator a in the usual way by setting

$$x = (\hbar/2M\omega_0)^{1/2}(a + a^+), \tag{2.6}$$

$$p = -i(M\omega_0\hbar/2)^{1/2}(a - a^+), \tag{2.7}$$

where

$$[a,a^+] = 1. \tag{2.8}$$

Equivalently, for the time-dependent problem we define operators b and b^+ by means of the analogs of (2.6) and (2.7), i.e.,

$$x_+ = (\hbar/2M\omega_0)^{1/2}(b + b^+), \tag{2.9}$$

$$p_+ = -i(\tfrac{1}{2}M\omega_0\hbar)^{1/2}(b - b^+), \tag{2.10}$$

where also

$$[b,b^+] = 1. \tag{2.11}$$

By direct substitution of (2.6), (2.7), (2.9), and (2.10) into (2.2) and (2.3), it is straightforward to obtain the following relations expressing b and b^+ in terms of the boson operators a and a^+:

$$b = \lambda_T a + \mu_T a^+ \qquad (\lambda_T, \mu_T \in C), \tag{2.12}$$

$$b^+ = \mu_T^* a + \lambda_T^* a^+. \tag{2.13}$$

Herein, λ_T and μ_T are given by

$$\lambda_T = \tfrac{1}{2}\{a(t) + d(t) + (i/M\omega_0)[c(t) - M^2\omega_0^2 b(t)]\}, \tag{2.14}$$

$$\mu_T = \tfrac{1}{2}\{a(t) - d(t) + (i/M\omega_0)[c(t) + M^2\omega_0^2 b(t)]\}. \tag{2.15}$$

[a] Qualified researcher N.F.W.O. (Belgium).

 0022-2488/81/071423-02$1.00

It has to be noticed that (2.12) and (2.13) are compatible with each other. Also, the subscript T has been introduced for referring to the notation of Tikochinsky.[6] Furthermore, since we learn from (2.14), (2.15), and (2.5) that

$$|\lambda_T|^2 - |\mu_T|^2 = 1, \tag{2.16}$$

for all times, a property which is in accordance with (2.8) and (2.11), Eq. (2.12), or equivalently (2.13), can be viewed as a complex linear Bogoliubov transformation. For this type of transformation it was shown by the present authors[7] that (2.12) and (2.13) can be written as

$$b = e^{S_c} e^{-i\sigma N} a e^{i\sigma N} e^{-S_c}, \tag{2.17}$$

$$b^+ = e^{S_c} e^{-i\sigma N} a^+ e^{i\sigma N} e^{-S_c}, \tag{2.18}$$

where the operator S_c and the number operator N are given by

$$S_c = -\tfrac{1}{2}x(a^{+2} e^{i\phi} - a^2 e^{-i\phi}) = -S_c^+, \tag{2.19}$$

$$N = a^+ a, \tag{2.20}$$

and where the parameters x, σ, and ϕ have to be chosen such that the equalities

$$\lambda_T = e^{i\sigma} \cosh x, \tag{2.21}$$

$$\mu_T = e^{i(\sigma+\phi)} \sinh x, \tag{2.22}$$

are satisfied. In fact, the operator $e^{S_c} e^{-i\sigma N}$, acting in the space of the harmonic oscillator Fock states, plays the same role as the classical Green's function derived by Landovitz *et al.*[1] in the coordinate representation. As a consequence the time-dependent harmonic oscillator transition amplitude from a state $|m\rangle$ to a state $|n\rangle$, previously denoted by a_{nm}, can be calculated as follows:

$$a_{nm} = \langle n | e^{S_c} e^{-i\sigma N} | m \rangle. \tag{2.23}$$

III. TRANSITION AMPLITUDES

The explicit evaluation of the matrix elements a_{nm}, defined by (2.23), has been carried out by Tikochinsky[6] and by the present authors in the more general para-Bose situation.[7] However, restricting the latter results to the pure Bose case (which corresponds to setting $h_0 = \frac{1}{2}$ in Ref. 7), there is perfect agreement with Tikochinsky's result which reads

$$a_{nm} = \left(\frac{m!n!}{|\lambda_T|}\right)^{1/2} \left(\frac{\mu_T^*}{2}\right)^{(m-n)/2} (\lambda_T)^{-(m+n)/2} \times \sum_k \frac{(-1)^k}{k!(n-2k)!(k+(m-n)/2)!} \left(\frac{|\mu_T|}{2}\right)^{2k} \quad \text{if } |m-n| = \text{even}, \tag{3.1}$$

$$a_{nm} = 0 \quad \text{otherwise.}$$

In fact, this expression can be obtained very easily with the help of various powerful techniques. So for instance, Rashid's[3] method which is based on a set of recursion relations for the coefficients a_{nm} leads almost immediately to (3.1), and this without the necessity, as it is the case in Landovitz's derivation, of knowing explicitly the harmonic oscillator eigenstates. In a similar way, the introduction of Hermite polynomials can also be avoided completely by using a technique first outlined by Witschel,[4] and later worked out in more detail by Kelemen.[5] Again, the result (3.1) follows almost immediately. In order to compare the result (3.1) to the corresponding one of Landovitz,[1] let us recall that the following expressions have been introduced there (a subscript L is added in order to refer to the Landovitz notation):

$$\sigma_L = 1/M\omega_0 b(t), \tag{3.2}$$

$$\mu_L = 1 + i\sigma_L [a(t) - d(t)] - \sigma_L^2 [1 - a(t)d(t)], \tag{3.3}$$

$$\lambda_L = 1 - i\sigma_L [a(t) + d(t)] + \sigma_L^2 [1 - a(t)d(t)]. \tag{3.4}$$

The reader can easily convince himself that on account of the relations (2.5), (2.14), and (2.15), the following equalities hold:

$$\mu_L/2\sigma_L = i\mu_T^*, \tag{3.5}$$

$$\lambda_L/2\sigma_L = -i\lambda_T^*. \tag{3.6}$$

As a consequence, the expression (3.1) can be written in the alternative form

$$a_{nm} = (2\sigma/|\lambda_L|)^{1/2} (m!n!)^{1/2} / 2^{(m+n)/2} (-i)^{-n} \times (\mu_L)^{(m-n)/2} (\lambda_L)^{-(m+n)/2} |\mu_L|^n \times \tfrac{1}{2} \sum_l [1 + (-1)^{m+l}] \frac{(-1)^{(n-l)/2} 2^{2l}}{l![(m-l)/2]![(n-l)/2]!} \times \left(\frac{\sigma}{|\mu_L|}\right)^l, \tag{3.7}$$

$$(|m-n| = \text{even})$$

where we have also replaced the summation index k by $l = n - 2k$. It has to be noticed that the expression (3.7) for a_{nm} does not agree completely with the result of Landovitz *et al.*[1] In fact, the obvious discrepancy is contained in a phase factor and is therefore harmless, since only $P_{nm} = |a_{nm}|^2$, the transition probability can be attributed a physical meaning.

IV. DISCUSSION

By the use of the very powerful techniques developed in the context of the linear Bogoliubov transformation, we have retrieved the results recently obtained for the transition amplitudes of time-dependent harmonic oscillators in a less laborious way. Moreover, our alternative approach to the problem allows the extension of the theory to Hamiltonians with terms, linear in the x and p-operators. Indeed, the generalized Bogoliubov transformation

$$b = \lambda a + \mu a^+ + \tau,$$

for real λ, μ, and τ has already been studied in detail by Witschel,[4] whereas the extension to complex coefficients is straightforward and proceeds in the same way as in Ref. 7.

[1] L. F. Landovitz, A. M. Levine, and W. M. Schreiber, J. Math. Phys. **21**, 2159 (1980).
[2] K. Tanabe, J. Math. Phys. **14**, 618 (1973).
[3] M. A. Rashid, J. Math. Phys. **16**, 378 (1975).
[4] W. Witschel, Z. Phys. B **21**, 313 (1975).
[5] A. Keleman, Z. Phys. A **274**, 109 (1975).
[6] Y. Tikochinsky, J. Math. Phys. **19**, 270 (1978).
[7] H. E. DeMeyer and G. Vanden Berghe, J. Phys. A: Math. Gen. **13**, 161 (1980).

Symmetries of the Schrödinger bundle

V. Aldaya and J.A. de Azcárraga
Dpto. De Física Teórica, Facultad de C. Físicas, Burjasot, Valencia, Spain

(Received 16 July 1980; accepted for publication 13 November 1980)

We exhibit in this paper the invariance of the Schrödinger Lagrangian density under the eleven-parameter group $\tilde{\mathcal{G}}_{(m)}$, central extension of the Galilei group $\mathcal{G}$. As a result, the quantum mechanical probability $\int \rho\, d^3x$ turns out to be the conserved charge associated with the central generator of $\tilde{\mathcal{G}}_{(m)}$, and the continuity equation is simply the expression of the conservation of the corresponding Noether current.

PACS numbers: 03.65.Fd

1. INTRODUCTION AND RESULTS

It is well known that the invariance of the Schrödinger equation under Galilean transformations is accomplished by making use of the fact that in quantum mechanics a global phase is irrelevant in the definition of the wavefunctions and that, accordingly, the space of states is[1] the space of rays. Thus, the Schrödinger equation is covariant, i.e., the Schrödinger equation in the new Galilean frame of reference ($\mathbf{r}' = \mathcal{R}\mathbf{r} + \mathbf{v}t + \mathbf{a}$, $t' = t + \tau$) is obtained by simply adding the appropriate primes if the new wavefunction is defined by

$$[U(\tau, \mathbf{a}, \mathbf{v}, \mathcal{R})\Psi](\mathbf{r}', t') = \exp \mathrm{im}(\tfrac{1}{2}\mathbf{v}^2 t + \mathbf{v}\cdot\mathcal{R}\mathbf{x})\Psi(\mathbf{r}, t). \tag{1}$$

The *projective* or *ray*[2] representation of the Galilei group $\mathcal{G}$ given by (1) comes from a representation of $\tilde{\mathcal{G}}_{(m)}$, the eleven parameter group which is the central extension of $\mathcal{G}$ by the so-called phase group $\mathcal{O}$, $\mathcal{G} = \tilde{\mathcal{G}}_{(m)}/\mathcal{O}$. This representation is defined by ($\hbar = 1$)

$$[U(\theta, \tau, \mathbf{a}, \mathbf{v}, \mathcal{R})\Psi](\mathbf{r}', t') = \exp \mathrm{im}(\tfrac{1}{2}\mathbf{v}^2 t + \mathbf{v}\cdot\mathcal{R}\mathbf{x} + \theta/m)\Psi(\mathbf{r}, t), \tag{2}$$

where $\theta \in \mathcal{O}$; it is simple to check that (2) fulfills the composition law for $\tilde{\mathcal{G}}_{(m)}$, which is given by[3]

$$(\theta'; \tau', \mathbf{a}', \mathbf{v}', \mathcal{R}')*(\theta; \tau, \mathbf{a}, \mathbf{v}, \mathcal{R}) = (\theta + \theta' + \xi_m(g', g); g'g) \tag{3a}$$

$$g'g = (\tau', \mathbf{a}', \mathbf{v}', \mathcal{R}')*(\tau, \mathbf{a}, \mathbf{v}, \mathcal{R}) = (\tau' + \tau, \mathbf{a}' + \mathcal{R}'\mathbf{a} + \tau\mathbf{v}', \mathbf{v}' + \mathcal{R}'\mathbf{v}, \mathcal{R}'\mathcal{R}), \tag{3b}$$

where (3b) is the Galilei group $\mathcal{G}$ composition law and

$$\xi_m(g', g) = m(\tfrac{1}{2}\mathbf{v}'^2\tau + \mathbf{v}'\cdot\mathcal{R}'\mathbf{a}) \tag{3c}$$

is the factor system (local exponent) of the central extension.[4] The fact that this ξ_m cannot be eliminated is due to the special structure of the Galilei group; in contrast with the Poincaré case, for which the projective representations are reduced to ordinary ones because the extension is trivial, the elements of $H_0^2(\mathcal{G}, U(1))$—the second cohomology group—are parametrized by a real number (the mass m), and the projective representations of $\mathcal{G}$ given by (1) cannot be reduced to ordinary ones since $\xi_m \not\simeq 0$.[5]

Given the relevance of $\tilde{\mathcal{G}}_{(m)}$ in nonrelativistic dynamics it seems only natural to analyze the invariance of the Schrödinger Lagrangian to obtain the charges associated with the different Noether currents. Following the usual procedure (see, e.g., Ref. 6) we take as the Schrödinger bundle the trivial[7] bundle η_s (E, π, M) with coordinate system $(t, \mathbf{x}; \Psi, \Psi^*)$, $(t, x^i) \in M$, and define the Lagrangian $\mathcal{L}_s$ on the bundle $J^1(E)$ of the 1-jets of E, parametrized by $(t, x^i, \Psi, \Psi^*, \Psi_t, \Psi_t^*, \Psi_i, \Psi_i^*)$. In this formalism, the ordinary wavefunction is a cross section of η_S, and the Schrödinger Lagrangian on $J^1(E)$ is given by

$$\mathcal{L}_S = -(1/2m)\Psi_i^*\Psi^i + i\Psi^*\Psi_t; \tag{4}$$

the application of the ordinary variational principle[6] to (4) leads to the Schrödinger equation.

As may be read from (2), we can realize $\tilde{\mathcal{G}}_{(m)}$ on the Schrödinger bundle by the set of *vector fields* acting on E (note the presence of m in K_i)

$$\begin{aligned} P_{(t)} &= -\frac{\partial}{\partial t}, \quad P_{(i)} = -\frac{\partial}{\partial x^i}, \quad J_{(i)} = -\epsilon_{ij}{}^k x^j \frac{\partial}{\partial x^k}, \\ K_{(i)} &= -t\frac{\partial}{\partial x^i} - imx_i\Psi\frac{\partial}{\partial \Psi} + imx_i\Psi^*\frac{\partial}{\partial \Psi^*}, \\ X_{(\theta)} &= -i\Psi\frac{\partial}{\partial \Psi} + i\Psi^*\frac{\partial}{\partial \Psi^*}, \end{aligned} \tag{5}$$

which obey the commutation relations of $\tilde{\mathcal{G}}_{(m)}$, namely

$$\begin{aligned} &[J_{(i)}, J_{(k)}] = \epsilon_{ij}{}^k J_{(k)}, \\ &[J_{(i)}, K_{(j)}] = \epsilon_{ij}{}^k K_{(k)}, \quad [J_{(i)}, P_{(j)}] = \epsilon_{ij}{}^k P_{(k)} \\ &[J_{(i)}, P_{(t)}] = [K_{(i)}, K_{(j)}] = [P_{(i)}, P_{(j)}] = [P_{(i)}, P_{(t)}] = 0, \\ &[K_{(i)}, P_{(t)}] = P_{(i)} \quad [X_{(\theta)}, \text{ any other}] = 0, \\ &[K_{(i)}, P_{(j)}] = m\delta_{ij} X_{(\theta)} \end{aligned} \tag{6}$$

(The commutation relations for the Galilei group are the same except that $[K_i, P_j] = 0$ and that there is no $X_{(\theta)}$). To check the invariance of (4) under $\tilde{\mathcal{G}}_{(m)}$ the 1-jet prolongation of the vector fields on E to $J^1(E)$ is required. The result is given by[8]

$$\begin{aligned} \overline{P}^1_{(t)} &= P_{(t)}, \quad \overline{P}^1_{(i)} = P_{(i)}, \\ \overline{J}^1_{(i)} &= J_{(i)} = \epsilon_{ij}{}^k\Psi^j\frac{\partial}{\partial \Psi^k} - \epsilon_{ij}{}^k\Psi^{*j}\frac{\partial}{\partial \Psi^{*k}}, \\ \overline{K}^1_{(i)} &= K_{(i)} + (\Psi_i - imx_i\Psi_t)\frac{\partial}{\partial \Psi_t} \\ &\quad - im(\delta_{ik}\Psi + x_i\Psi_k)\frac{\partial}{\partial \Psi_k} \end{aligned}$$

 0022-2488/81/071425-03$1.00

$$+(\Psi_i^* + imx_i\Psi_t^*)\frac{\partial}{\partial\Psi_t^*}$$
$$+ im(\delta_{ik}\Psi^* + x_i\Psi_k^*)\frac{\partial}{\partial\Psi_k^*},$$
$$\bar{X}^1_{(\theta)} = X_{(\theta)} - i\Psi_t\frac{\partial}{\partial\Psi_t} - i\Psi_i\frac{\partial}{\partial\Psi_i} + i\Psi_t^*\frac{\partial}{\partial\Psi_t^*} + i\Psi_i^*\frac{\partial}{\partial\Psi_i^*}, \tag{7}$$

where the bar and the superindex 1 indicate 1-jet prolongation. With the help of (7) we may compute the Lie derivative $L_{\bar{X}^1}\mathscr{L}_S$ of (4) for the different vector fields. It is a straightforward task to see that $\mathscr{L}_S$ is strictly invariant under spatial and temporal translations, rotations, and under the one parameter group generated by $X_{(\theta)}$, though semi-invariant under the Galilean boosts.[9] Explicitly, one gets

$$L_{\bar{X}^1}\mathscr{L}_S = 0, \tag{8}$$

except for the boosts, for which one obtains

$$L_{\bar{K}^1_i}\mathscr{L}_S = \frac{i}{2}(\Psi_i\Psi^* + \Psi\Psi_i^*) \equiv \frac{i}{2}\frac{D}{Dx^k}(\delta_i^k\Psi\Psi^*), \tag{9}$$

where the last term of (9) constitutes the definition of a "formal derivative" D/Dx^k which on cross sections is the total derivative with respect to x^k. The right-hand side of (9) does not modify the equations of motion as derived from the variational principle, so that there is invariance under the "Galilean" boosts of $\bar{\mathscr{G}}_{(m)}$. For instance, for a finite transformation of parameter V, the usual Lagrangian (for simplicity in one dimension) is modified by the addition of the term

$$-\frac{i}{2}V\frac{d}{dx}(\Psi^*\Psi), \tag{10}$$

and thus the invariance of the Schrödinger equation is guaranteed. Note, however, that this is not the case if one considers the truly Galilean boosts of $\mathscr{G}$, which clearly do not have components on $\partial/\partial\Psi$, $\partial/\partial\Psi^*$ (i.e., on the fiber part of E). Thus, the Schrödinger Lagrangian is *not* invariant under the Galilei group $\mathscr{G}$; this is already illustrated by the fact that (1) is not an ordinary representation of $\mathscr{G}$ (note that $\mathscr{G}$ is *not* a subgroup of $\bar{\mathscr{G}}_{(m)}$) though not always clearly realized.

Once the invariance (or semi-invariance) of $\mathscr{L}_S$ under $\bar{\mathscr{G}}_{(m)}$ has been proved, we may proceed to evaluate the Noether current and the associated Noether charges. The expression of conserved Noether current associated with the general vector field on E

$$X = X^t\frac{\partial}{\partial t} + X^i\frac{\partial}{\partial x^i} + X_\Psi\frac{\partial}{\partial\Psi} + X_{\Psi^*}\frac{\partial}{\partial\Psi^*}, \tag{11}$$

is given by

$$j^{0,i} = -\mathscr{L}X^{0,i} + (X^t\Psi_t + X^j\Psi_j - X_\Psi)\frac{\partial\mathscr{L}}{\partial\Psi_{t,i}} + (\Psi_t^*X^t + \Psi_j^*X^j - X_{\Psi^*})\frac{\partial\mathscr{L}}{\partial\Psi^*_{t,i}}, \tag{12}$$

when $L_{\bar{X}^1}\mathscr{L}_S = 0$ and with the additional term $-\Delta^\mu$ ($\mu = 0, i$) when $L_{\bar{X}^1}\mathscr{L}_S = (D/Dx^\mu)\Delta^\mu$;[6] (12) is understood to be restricted to cross sections which are solutions of the Euler–Lagrange equations associated with $\mathscr{L}_S$, i.e., the Schrödinger wavefunctions. A straightforward application of (12) leads to the customary results for the energy, momentum, and rotations which, omitting the expression of the spatial components of the currents for the sake of brevity, take the form

$$\begin{aligned} j^0_{(0)} &= -(1/2m)\Psi_i^*\Psi^i \equiv \Psi^*(\mathbf{p}^2/2m)\Psi = \mathscr{H}, \\ j^0_{(i)} &= -i\Psi^*\Psi_i \equiv \Psi^* p_i\Psi = \mathscr{P}_i, \\ j^0_{(i)} &= -i\Psi^*\epsilon_{ij}{}^k x^j\Psi_k \equiv \Psi^*\epsilon_{ij}{}^k x^j p_k\Psi = J_i. \end{aligned} \tag{13}$$

For the boosts we get

$$j^0_{(i)} = i\Psi^*(imx_i\Psi - t\Psi_i) \equiv \Psi^*(tp_i - mx_i)\Psi = K_i. \tag{14}$$

Note that although the charge density receives no contribution from (9), that term does contribute to the spatial components $j^i_{(i)}$ of the current, guaranteeing in this way its conservation.

Finally, we compute the expression of the Noether current associated with $X_{(\theta)}$, the central vector field of $\bar{\mathscr{G}}_{(m)}$. One immediately gets from (12)

$$\begin{aligned} j^0_{(\theta)} &= i\Psi^* i\Psi = -\Psi^*\Psi \equiv -\rho, \\ j^i_{(\theta)} &= (i/2m)(\Psi^*\Psi^i - \Psi^{i*}\Psi). \end{aligned} \tag{15}$$

Equation (15) shows that the charge density associated with the central element of $\bar{\mathscr{G}}_{(m)}$ is nothing but the nonrelativistic probability density, and the expression of the conservation of the corresponding current,

$$\frac{\partial\rho}{\partial t} + \nabla\mathbf{j} = 0, \quad \mathbf{j} = \frac{1}{2im}[\Psi^*(\nabla\Psi) - (\nabla\Psi^*)\Psi], \tag{16}$$

is the well-known continuity equation of wave mechanics. Thus the total probability appears in Galiean wave mechanics as the conserved quantity associated with one of the generators of the symmetry group of $\bar{\mathscr{G}}_{(m)}$. In all, formulas (13)–(15) provide a realization of the infinitesimal generator of $\bar{\mathscr{G}}_{(m)}$ on the space of cross sections of η_S, which [adding the appropriate $\hbar$ which we have been taking up to now as the unity, but that might have been inserted from the beginning in the exponential of (2)] is clearly given by

$$\begin{aligned} &H = i\hbar\frac{\partial}{\partial t}, \quad P_i = -i\hbar\frac{\partial}{\partial x^i}, \quad K_i = -i\hbar t\frac{\partial}{\partial x^i} - mx_i, \\ &J_i = -\hbar\epsilon_{ij}{}^k x^j\frac{\partial}{\partial x^k}, \quad X_\theta = -I. \end{aligned} \tag{17}$$

Clearly the finite transformations of the wavefunctions are obtained by exponentiating (17), and it is trivial to check that the infinitesimal generators of (17) satisfy the commutation rules of (6) once a factor $i\hbar$ is added in all the r.h.s. terms.

2. CONCLUSION AND COMMENTS

In the above paragraphs it has been shown how the extended Galilei symmetry leads to the quantum mechanical probability density whose associated "charge" is the central element of the extended Galilei algebra. The fact that a probability density appears from a phase-gauge transformation is only natural; what makes it interesting is that it comes from a transformation of $\bar{\mathscr{G}}_{(m)}$ [see (2)], and that this extension of

the Galilei group is necessary if the Schrödinger Lagrangian is to be invariant under the symmetry transformations of nonrelativistic mechanics. Thus, and via the Noether theorem, the probability conservation is a consequence of the extended Galilei invariance and X_θ is the central generator of the Lie algebra of $\bar{\mathcal{G}}_{(m)}$ [see (17)]. It is not surprising that the added element of the algebra is a central one: the total probability should not be affected by the action of the remaining kinematical transformations of $\bar{\mathcal{G}}_{(m)}$.

Moreover, the above results are consistent with the usual Galilean invariance of Newtonian classical mechanics. Indeed, the conserved charge associated with expression (14) gives the classical law of motion, in Ehrenfest's sense, of a free particle of mass m, and the charge corresponding to the central generator (15)—the unity for a normalized wave function—is irrelevant in the classical limit. The transition from $\bar{\mathcal{G}}_{(m)}$ to $\mathcal{G}$ is thus accomplished by averaging over densities for the classical approximation, i.e., it may be done by going from quantum to classical Galilean mechanics. This, by the way, provides a group-theoretical definition of the classical limit. Reciprocally, it may be noted that it is not possible to define a faithful action of $\bar{\mathcal{G}}_{(m)}$ on the configuration space (t, q_i)—which could have led to a $\bar{\mathcal{G}}_{(m)}$-invariant classical mechanics—and that it is therefore necessary to perform an extension of the configuration space to allow the phase part of $\bar{\mathcal{G}}_{(m)}$ to act in a nontrivial way.[10]

To conclude, we might mention that the above situation is in sharp contrast with its relativistic counterpart.[11] This can be traced to the different cohomological structure of the Galilei and the Poincaré groups. As already mentioned, $H^2(\bar{\mathcal{P}}^\uparrow_+, \mathcal{U}(1))$ is trivial and thus—as has been known since the work of Wigner[12]—the phase part cannot be added to the Poincaré group in a nontrivial way. One might say that, in this sense, $\bar{\mathcal{G}}_{(m)}$ is more a quantum group than $\bar{\mathcal{P}}^\uparrow_+ \otimes \mathcal{U}(1)$.

[1]The concept of ray was introduced by H. Weyl; see *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), pp. 4, 20, 74, and 180 (original German edition 1931).

[2]V. Bargmann, Ann. Math. **59**, 1 (1954).

[3]J. M. Lévy-Leblond in *Group theory and its Applications*, Vol. II, edited by E. M. Loebl (Academic, New York, 1971). This review paper contains a fairly complete discussion and list of references on the Galilei groups.

[4]See, e.g., A. G. Kurosh, *The Theory of Groups*, Vol. II (Chelsea, New York, 1955); L. Michel in *Group Theoretical Concepts and Methods in Elementary Particle Physics* (Gordon and Breach, New York, 1964) and Refs. 2 and 3.

[5]Throughout this paper we omit the consideration of spin, which is naturally incorporated by the Galilei group as has been particularly stressed by Lévy-Leblond (see, e.g. Ref. 3 and references therein). A nonzero spin may be easily incorporated by adding the representative $D^s(R)^\alpha_\beta$ of the rotation R acting on $\Psi^\beta(\alpha, \beta = 1, ..., 2s + 1)$ in (1) and (2), but this is irrelevant for the purposes of this paper.

[6]V. Aldaya and J. A. de Azcárraga, J. Math. Phys. **19**, 1869 (1978).

[7]All bundles on $\mathbb{R}^4$ are trivial.

[8]These expressions may be obtained using formulas (2.3)–(2.7) of Ref. 6.

[9]It is also possible to obtain exact invariance for the boosts by taking as Langragian density

$$\mathcal{L}_S = -\frac{1}{2m}\Psi_i^*\Psi^i + \frac{i}{2}(\Psi^*\Psi_t - \Psi_t^*\Psi),$$

which in the more familiar formulation in terms of the Schrödinger field Ψ corresponds to taking $\mathcal{L}_S$ Hermitian.

[10]The extended Galilei group does, nevertheless, show up in *classical* mechanics, see J. M. Lévy-Leblond, Commun. Math. Phys. **12**, 64 (1969); and in another context, M. Pauri and G. M. Prosperi, J. Math. Phys. **7**, 366 (1966); **9**, 1146 (1968).

[11]The fact is, of course, that relativistic probability densities require in general a reinterpretation by means of quantum field theory.

[12]E. P. Wigner, Ann. Math. **40**, 149 (1939).

Bound state perturbation theory for the one-space and one-time dimension Klein–Gordon equation

C. K. Au
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208

Y. Aharonov
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208 and Department of Physics, University of Tel Aviv, Tel Aviv, Israel

(Received 8 December 1980; accepted for publication 27 February 1981)

We present a perturbation theory for an arbitrary bound state in the one-space and one-time dimension Klein–Gordon equation in the presence of a scalar potential and a vector (fourth component only) potential by reducing it to a Ricatti equation with the method of logarithmic perturbation expansions. All corrections to the energies and wavefunctions, including corrections to the positions of the nodes in excited states, are expressed in quadratures in a hierarchical scheme, without the use of either the Green's function or the sum over intermediate states.

PACS numbers: 03.65.Ge

I. INTRODUCTION

Recently, in a series of papers,[1–3] we have presented perturbation theories for an arbitrary bound state in static potentials for the Schrödinger equation and the Dirac equation. In central field problems[3] and in problems reducible to one dimension in the case of the Schrödinger equation,[1,2] we have shown that all corrections to the energies and wavefunctions, including corrections to the positions of the nodes in excited states, can be expressed in quadratures in a hierarchical scheme, without the use of either the Green's function or the sum over intermediate states. This is achieved through the reduction of the differential equations involved to a Ricatti equation, followed by a perturbation expansion. In the case of the Schrödinger equation,[1,4,5] this is equivalent to performing a perturbation expansion on the logarithmic derivative of the wavefunction instead of on the wavefunction itself. In the case of the Dirac equation,[3,6] this is equivalent to carrying out a perturbation expansion on the ratio between the radial parts of the small and large components of the Dirac spinor. We emphasize that in the case of excited states where the wavefunctions possess nodes,[1,3,4] the zeros must be factored out first. We have also shown that the first-order perturbation iteration method (FOPIM), first introduced by Hirschfelder,[7] can be incorporated into this perturbation approach to yield accelerated convergence, if convergence exists.[3,8] In nonrelativistic problems that are not reducible to one dimension, we show that, for the ground state, the method of logarithmic perturbation leads to a hierarchy of equations that determines the corrections to the energy and wave function for each order. In this hierarchy, the equation for the ith-order correction is isomorphic to the equation for the first-order correction. Moreover, these equations have the same form as Gauss' law in classical electrodynamics.[2] As an application, we have shown that this method can be used to obtain the corrections to the energy and the logarithm of the wavefunction of the ground state of a hydrogen atom in a multipole field or a linear combination of static multipole fields to any order in perturbation theory.[9]

In this paper, we would like to extend similar techniques to the one-space and one-time dimension Klein–Gordon equation. We shall assume that it is possible to solve this equation with a certain scalar potential and a fourth component vector potential. The problem of a charged spinless boson in a central field is reducible to this form. We then consider the change in the energy and in the wavefunction as a perturbation is introduced to the fourth component vector potential or to the scalar potential. This will be developed in Secs. II and III. In Sec IV, we conclude by mentioning some possible applications of the presently developed techniques.

II. PERTURBATION IN THE FOURTH COMPONENT VECTOR POTENTIAL

The single-particle one-space and one-time dimension Klein–Gordon equation in the presence of a fourth component vector potential V and a scalar potential S can be written as[10,11] in natural units $\hbar = c = 1$:

$$[E - V(x)]^2\psi(x) = \left[-\frac{d^2}{dx^2} + m^2 + 2mS(x)\right]\psi(x). \tag{2.1}$$

We assume that, for a certain V_0 and S_0, the above eigenvalue problem is solvable so that the energy eigenvalue E_0 and the corresponding wavefunction ψ_0 are known for a particular state. We shall consider the correction to the energy and wavefunction as a perturbation λV_1 is introduced to the potential V_0. In the following section, we consider the same corrections as a perturbation ηS_1 is introduced to the scalar potential S_0. For the sake of brevity, and yet without sacrificing clarity of the essense of our method, we shall limit our detailed discussions to the ground state where the wavefunction does not contain any zero and to the first excited where the wavefunction possesses one zero. The generalization to an arbitrary excited bound state with a finite number of nodes is straightforward, and the mechanism is similar to what has been reported previously.[1,3]

In the absence of any degeneracy, the wavefunction ψ can be taken as real. In order that the charge density be normalizable, we require that ψ vanishes as x approaches

 0022-2488/81/071428-05$1.00

$\pm\infty$. In the case of the ground state, ψ does not contain any zero, and so its logarithm is regular. In analogy to the nonrelativistic Schrödinger equation, we define

$$\psi(x) = \exp[-G(x)] \tag{2.2a}$$

and

$$g(x) = \frac{d}{dx}G(x). \tag{2.2b}$$

Equation (2.1) is then transformed to the Ricatti form

$$(E-V)^2 = g' - g^2 + m^2 + 2mS, \tag{2.3}$$

where a prime denotes a derivative with respect to its argument. The unperturbed Klein–Gordon equation in Ricatti form is

$$(E_0 - V_0)^2 = g_0' - g_0^2 + m^2 + 2mS_0. \tag{2.4}$$

In the presence of a perturbation to the fourth component vector potential, λV_1, Eq. (2.3) becomes

$$(E - V_0 - \lambda V_1)^2 = g' - g^2 + m^2 + 2mS_0. \tag{2.5}$$

We seek a perturbative solution to Eq. (2.5) by expanding the eigenvalue E and the logarithmic derivative of the wavefunction g in power series in λ:

$$E = E_0 + \lambda E_1 + \lambda^2 E_2 + \cdots = \Sigma \lambda^i E_i, \tag{2.6}$$

and

$$g = g_0 + \lambda g_1 + \lambda^2 g_2 + \cdots = \Sigma \lambda^i g_i. \tag{2.7}$$

We next define

$$B_0 \equiv E_0 - V_0, \tag{2.8}$$

$$B_1 \equiv E_1 - V_1, \tag{2.9}$$

$$B_i \equiv E_i \quad \text{for all } i \geqslant 2. \tag{2.10}$$

On comparing coefficients of various powers in λ, we obtain

$$B_0^2 = g_0' - g_0^2 + m^2 + 2mS_0, \tag{2.11}$$

which is the unperturbed Klein–Gordon equation for the ground state in the Ricatti form, and

$$2B_0B_1 = g_1' - 2g_0g_1, \tag{2.12}$$

$$2B_0B_k + \sum_{j=1}^{k-1}(B_jB_{k-j} + g_jg_{k-j}) = g_k' - 2g_0g_k \tag{2.13}$$

for all $k \geqslant 2$. We then observe, similar to the nonrelativistic case in the Schrödinger equation, that the square of the unperturbed wavefunction serves as an integration factor to this hierarchy of equations,

$$2B_0B_1e^{-2G_0} = [g_1e^{-2G_0}]', \tag{2.14}$$

and for $k \geqslant 2$

$$\left[2B_0B_k + \sum_{j=1}^{k-1}(B_jB_{k-j} + g_jg_{k-j})\right]e^{-2G_0} = [g_ke^{-2G_0}]'. \tag{2.15}$$

Equation (2.15) can be brought to a form similar to (2.14) if we define an effective k th-order perturbation potential V_k by

$$B_0V_k \equiv -\tfrac{1}{2}\sum_{j=1}^{k-1}(B_jB_{k-j} + g_jg_{k-j}). \tag{2.16}$$

Equations (2.14) and (2.15) can then be rewritten as

$$2(E_0 - V_0)(E_k - V_k)e^{-2G_0} = [g_ke^{-2g_0}]' \tag{2.17}$$

for all $k \geqslant 1$.

We now readily see that this hierarchy of equations can be solved in quadrature. On integrating Eq. (2.17) from $-\infty$ to $+\infty$, the right-hand side vanishes according to the boundary conditions and the integration yields

$$E_k = \frac{\int_{-\infty}^{+\infty}(E_0 - V_0)V_k\psi^2(x)\,dx}{\int_{-\infty}^{+\infty}(E_0 - V_0)\psi^2(x)\,dx}. \tag{2.18}$$

If we normalize the charge of the Klein–Gordon particle to unity, then the integral in the denominator of (2.18) is equal to m, the mass of the Klein-Gordon particle in equation. In this case, the k th-order correction to the energy is given by

$$E_k = \int_{-\infty}^{+\infty}\frac{(E_0 - V_0)V_k\psi^2(x)\,dx}{m}. \tag{2.19}$$

Having obtained E_k, Eq. (2.17) can now be readily integrated to yield a solution for g_k:

$$g_k(x) = e^{2G_0(x)}\int_{-\infty}^{x}2(E_0 - V_0)(E_k - V_k)e^{-2G_0}\,dx. \tag{2.20}$$

Since V_k is defined by E_j and g_j where $j \leqslant k-1$, it is apparent that the perturbative solution can be obtained in this hierarchical scheme.

It may appear worthwhile to show that E_1, as given by Eq. (2.19), is the same as that obtained in standard perturbation theory in terms of the two-component wave function formalism that leads to the first-order Klein-Gordon equation.[10,11] As is well known, the two-component isospinor wavefunction Ψ can be written in terms of the Klein-Gordon wavefunction ψ as

$$\Psi = \begin{pmatrix}\phi\\ \chi\end{pmatrix} = \frac{\psi}{2}\begin{pmatrix}1 + (E-V)/m\\ 1 - (E-V)/m\end{pmatrix}. \tag{2.21}$$

In terms of this isospinor, the Klein-Gordon equation can be written in Hamiltonian form

$$H\Psi = E\Psi, \tag{2.22}$$

where the operator H is identified as

$$H = (\tau_3 + i\tau_2)\left(\frac{-1}{2m}\frac{d^2}{dx^2} + S\right) + \tau_3 m + V, \tag{2.23}$$

τ_1, τ_2, τ_3 being the Pauli isospin matrices, and the scalar product is defined in general as

$$\langle\Psi\,|\,\Psi'\rangle \equiv \int \Psi^\dagger\tau_3\Psi'. \tag{2.24}$$

Thus, the normalization that fixes the charge of the Klein-Gordon particle to be unity is equivalent to $\langle\Psi\,|\,\Psi\rangle = 1$ or

$$\int\{[1 + (E-V)/m]^2 - [1 - (E-V)/m]^2\}\psi^2/4 = 1 \tag{2.25}$$

or to

$$\int(E - V)\psi^2\,dx = m, \tag{2.26}$$

which is the condition used to replace the denominator of Eq. (2.18) by the mass of the Klein-Gordon particle in terms of the unperturbed solution. Then, as is well known, the first-order energy shift is given by

$$E_1 = \langle\Psi_0|H_1|\Psi_0\rangle, \tag{2.27}$$

where Ψ_0 is the isospinor corresponding to the unperturbed solution. Here, $H_1 = V_1 \mathbb{1}$, where $\mathbb{1}$ is the unit matrix. Then Eqs. (2.27), (2.24), and (2.21) together will lead to Eq. (2.19) for $k = 1$.

Once the g_k's are obtained, they can be integrated to given the G_k's, the correction to the logarithm of the wavefunction. The integration constants here are additive constants to the logarithm of the wavefunction and are hence multiplicative constants to the wavefunction that can be fixed by normalization of the charge density.

As discussed in earlier papers,[3,8] an alternative approach to the perturbation is the first-order perturbation iteration method (FOPIM). From the knowledge of E_1 and g_1, we construct the function

$$g_0^{\mathrm{I}} \equiv g_0 + \lambda g_1 \tag{2.28}$$

and

$$E_0^{\mathrm{I}} = E_0 + \lambda E_1. \tag{2.29}$$

We then seek a potential V_0^{I} that will satisfy Eq. (2.24)

$$(E_0^{\mathrm{I}} - V_0^{\mathrm{I}})^2 \equiv (g_0^{\mathrm{I}})' - (g_0^{\mathrm{I}})^2 + m^2 + 2mS. \tag{2.30}$$

This is an algebraic equation and V_0^{I} can be solved. However, because of the quadratic nature, there will be two possible solutions for V_0^{I}. The correct choice is the one that approaches V_0 as λ approaches zero. Having found this V_0^{I}, the new unperturbed potential, the perturbation can be chosen as

$$V_1^{\mathrm{I}} = V_0 + \lambda V_1 - V_0^{\mathrm{I}}. \tag{2.31}$$

It is not hard to show that from Eqs. (2.4), (2.12), (2.28), (2.29), and (2.30) that V_1^{I} is of order λ^2. We thus succeed in reducing a problem with a perturbation of order λ to the one with order λ^2. This process can be continued. The next step will reduce the perturbation to order λ^4.

We not turn to the excited states. For simplicity, we consider the first excited state where the wavefunction contains one node and give the corrections to an arbitrary order. For an arbitrary excited bound state where the wavefunction contains a finite number of, but more than one, zeros, we give the expressions for the first-order corrections. One then can use the FOPIM to generate all higher corrections.

In the case of the first excited state where the wavefunction has one node we can write

$$\psi = (x - a)e^{-G}, \tag{2.32}$$

where a is the position of the node. The Klein-Gordon equation is then transformed to the Ricatti form:

$$[g^2 - g' + (E - V)^2 - m^2 - 2mS](x - a) = 2g. \tag{2.33}$$

The unperturbed equation is

$$[g_0^2 - g_0' + (E_0 - V_0)^2 - m^2 - 2mS_0](x - a_0) = 2g_0, \tag{2.34}$$

where E_0 is the zeroth-order energy eigenvalue, and the zeroth-order eigenfunction is

$$\psi_0 = (x - a_0)e^{-G_0}, \tag{2.35}$$

where a_0 is the nodal position in the absence of any perturbation. In the presence of a perturbation λV_1, we seek the solution of E and G in power series of λ as in Eq. (2.6) and (2.7). In addition, the nodal position a is also expanded in powers of λ:

$$a = a_0 + \lambda a_1 + \lambda^2 a_2 + \cdots = \Sigma \lambda^i a_i. \tag{2.36}$$

On substituting Eqs. (2.36) and (2.6)–(2.10) in Eq. (2.33) and using (2.34), we obtain, after multiplying throughout by $(x - a_0)e^{-2G_0}$,

$$[a_1 e^{-2G_0}]' - [g_1(x - a_0)^2 e^{-2G_0}]' + 2B_0 B_1 (x - a_0)^2 e^{-2G_0} = 0, \tag{2.37}$$

and for all $i \geqslant 2$,

$$a_i [e^{-2G_0}]' - [g_i (x - a_0)^2 e^{-2G_0}]' + 2B_0 (E_i - V_i)(x - a_0)^2 e^{-2G_0} = 0, \tag{2.38}$$

where the effective ith-order perturbation potential is defined by

$$2B_0 V_i (x - a_0) \equiv \sum_{k=1}^{i-1} \Big\{ a_{i-k} \Big[2g_0 g_k - g_k' + 2B_0 B_k + \sum_{j=1}^{k-1} (B_j B_{k-j} + g_j g_{k-j}) \Big] - [B_k B_{i-k} + g_k g_{i-k}](x - a_0) \Big\}, \tag{2.39}$$

and the summation is understood to be zero if the upper limit of the running index is smaller than the lower limit. It is then trivial to obtain the corrections to the energy, the nodal position, and g by integrating Eqs. (2.36) and (2.37) from $-\infty$ to $+\infty$, a_0 and x in a hierarchical scheme.

In the case of an arbitrary excited bound state where the wavefunction prossesses N zeros, we write the wavefunction as

$$\psi = \Big[\prod_{\mu=1}^{N} (x - a_\mu) \Big] e^{-G}. \tag{2.40}$$

Then the Klein-Gordon equation in Ricatti form becomes

$$[g^2 - g' + (E - V)^2 - m^2 - 2mS] \prod_{\mu} (x - a_\mu) = 2g \sum_{\nu} \prod_{\mu \neq \nu} (x - a_\mu) - \sum_{\sigma,\nu} \prod_{\mu \neq \sigma,\nu} (x - a_\mu). \tag{2.41}$$

Here, we only give the first-order corrections. The higher-order corrections can be brought to forms analogous to Eqs. (2.37) and (2.38) by keeping track of the indices. Alternatively, one can use the first-order perturbation iteration method. The first-order corrections are given by the following equation analogous to (2.36):

$$\sum_{\mu} a_{\mu 1} \Big[\prod_{\nu \neq \mu} (x - a_{\nu 0})^2 e^{-2G_0} \Big]' - \Big[g_1 \prod_{\mu} (x - a_{\mu 0})^2 e^{-2G_0} \Big]' = -2(E_1 - V_1)(E_0 - V_0) \prod_{\mu} (x - a_{\mu 0})^2 e^{-2G_0}, \tag{2.42}$$

from which the first-order energy correction E_1 can be obtained by integrating from $-\infty$ to $+\infty$, the first-order correction to the μth node $a_{\mu 1}$ can be obtained by integrating from $-\infty$ to $a_{\mu 0}$, and then g_1 can be obtained by integrating from $-\infty$ to x after E_1 and the $a_{\mu 1}$'s have been obtained. As in the case of the ground state, the additive constant to G from the integration of g is fixed by normalization of the charge density.

III. PERTURBATION IN THE SCALAR POTENTIAL

We now consider the situation where the perturbation ηS_1 is in the scalar potential S. In the case of the ground state where the wavefunction does not contain any zero, it can be written as in Eqs. (2.2a) and (2.2b), and the unperturbed equation is the same as Eq. (2.4). In the presence of the additional scalar potential ηS_1, the Klein-Gordon equation in Ricatti form becomes

$$(E - V_0)^2 = g' - g^2 + m^2 + 2mS_0 + 2m\eta S_1. \tag{3.1}$$

In analogy to Eqs. (2.6) and (2.7), we seek the solution in power series in η:

$$E = E_0 + \eta E_1 + \eta^2 E_2 + \cdots = \Sigma \eta^i E_i \tag{3.2}$$

and

$$g = g_0 + \eta g_1 + \cdots = \Sigma n^i g_i. \tag{3.3}$$

On comparing coefficients of various powers in η, we obtain

$$(E_0 - V_0)^2 = g_0' - g_0^2 + m^2 + 2mS_0, \tag{3.4}$$

which is the unperturbed Eq. (2.4), and

$$2E_1(E_0 - V_0) = g_1' - 2g_0 g_1 + 2mS_1, \tag{3.5}$$

$$2E_i(E_0 - V_0) + \sum_{j=1}^{i-1}(E_j E_{i-j} + g_j g_{j-j}) = g_i' - 2g_0 g_i \tag{3.6}$$

for all $i \geqslant 2$, We then readily observe, as in the previous section, that the square of the unperturbed wavefunction, e^{-2G_0}, acts as an integration factor to this hierarchy of equation:

$$2E_1(E_0 - V_0)e^{-2G_0} = [g_1 e^{-2G_0}]' + 2mS_1 e^{-2G_0}, \tag{3.7}$$

from which we obtain

$$E_1 = \int_{-\infty}^{+\infty} S_1 e^{-2G_0}\, dx, \tag{3.8}$$

on using the normalization condition (2.26), and

$$g_1 = e^{2G_0}\int_{-\infty}^{x} 2[E_1(E_0 - V_0) - mS_1]e^{-2G_0}\, dx. \tag{3.9}$$

The higher-order corrections are given by

$$2E_i(E_0 - V_0)e^{-2G_0} = [g_i e^{-2G_0}]' + 2mS_i e^{-2G_0}, \tag{3.10}$$

where we identify the ith order effective scalar perturbation potential S_i by

$$2\,mS_i \equiv -\sum_{j=1}^{i-1}(E_j E_{i-j} + g_j g_{i-j}). \tag{3.11}$$

In terms of the two-component wavefunction Ψ in Eq. (2.21), the first-order correction to the energy is given by Eq. (2.27), where $H_1 = (\tau_3 + i\tau_2)S_1$, which can be easily shown to lead to the same result as Eq. (3.8).

We now turn to the first excited state whose wavefunction is written as in Eq. (2.32). The Klein-Gordon equation in Ricatti form is given by Eq. (2.33). We then expand the energy E and g in power series in η as in Eqs. (3.2) and (3.3). In addition, the nodal position a is also expanded in a power series of η analogous to Eq. (2.36):

$$a = a_0 + \eta a_1 + \cdots = \Sigma \eta^i a_i. \tag{3.12}$$

On executing procedures similar to the last section, we obtain the following for the first-order corrections:

$$[a_1 e^{-2G_0}]' - [g_1(x - a_0)^2 e^{-2G_0}]' = -2[E_1(E_0 - V_0) - mS_1](x - a_0)^2 e^{-2G_0}, \tag{3.13}$$

from which the corrections E_1, a_1, and g_1 can be obtained by integrating from $-\infty$ to $+\infty$, a_0 and x. The higher-order corrections are given by

$$a_i[e^{-2G_0}]' - [g_i(x - a_0)^2 e^{-2G_0}]' = -2[E_i(E_0 - V_0) - mS_i](x - a_0)^2 e^{-2G_0}, \tag{3.14}$$

where the effective ith-order scalar perturbation S_i is defined by

$$\begin{aligned} 2mS_i(X - a_0) \equiv \sum_{k=1}^{i-1}\Big\{ & a_{i-k}\Big[2g_0 g_k - g_k' + 2(E_0 - V_0)E_k \\ & + \sum_{j=1}^{k-1}(E_j E_{k-j} + g_j g_{k-j})\Big] \\ & - [E_k E_{i-k} + g_k g_{i-k}](x - a_0)\Big\}. \end{aligned} \tag{3.15}$$

Equation (3.14) can readily be integrated to give the corrections E_i, a_i, and g_i with the same sets of integration limits. It is then obvious that Eqs. (3.13)–(3.15) can be integrated in a hierarchical scheme to yield corrections to any order in perturbation theory.

For an arbitrary excited bound state, we only give the equation for the first-order corrections:

$$\sum_{\mu} a_{\mu 1}\left[\prod_{\nu \neq \mu}(x - a_{\nu 0})^2 e^{-2G_0}\right]' - \left[g_1 \prod_{\mu}(x - a_{\mu 0})^2 e^{-2G_0}\right]' = -2[E_1(E_0 - V_0) - mS_1]\Pi(x - a_{\mu 0})^2 e^{-2G_0}, \tag{3.16}$$

which is analogous to Eq. (2.41) and can be solved in quadrature by integration in an analogous manner.

IV. CONCLUDING REMARKS

In this paper we have presented a bound state perturbation theory for the one-space and one-time dimension Klein-Gordon equation in the presence of a scalar potential and a fourth component vector potential by reducing it to a Ricatti equation with the method of logarithmic perturbation expansions. The problem of a charged spinless boson in a central field is reducible to this form. Our results are thus applicable to the study of the perturbative corrections to the energies and wavefunctions of bound states in pionic atoms due to a screened Coulomb potential or due to a finite-sized but spherical nucleus. We have shown that it is possible to obtain all preturbative corrections in quadrature in a hierarchical scheme without the use of either the Green's function or the sum over the intermediate states, especially in the latter where negative energy states are also involved. The computation of the higher energy corrections in closed quadratures can be used to identify sum rules, as reported by us previously.[1,2]

ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation under Grant Nos. PHY-79-01053 and ISP-80-11451.

[1]Y. Aharonov and C. K. Au, Phys. Rev. Lett. **42**, 1582 (1979); **43**, 176(E) (1979); **44**, 619(E) (1980).
[2]C. K. Au and Y. Aharonov, Phys. Rev. A **20**, 2245 (1979).
[3]C. K. Au and G. W. Rogers, Phys. Rev. A **22**, 1820 (1980).
[4]V. S. Polikanov, Zh. Eksp. Teor. Fiz. **52**, 1326 (1967) [Sov. Phys. JETP **25**, 882 (1967)]; Theoret. Math. Phys. (USSR) **24**, 230 (1975).
[5]R. J. Price, Proc. Phys. Soc. London **67**, 383 (1954).
[6]A. I. Mikhailov and V. S. Polikanov, Zh. Eksp. Teor. Fiz. **54**, 175 (1968) [Sov. Phys. JETP **27**, 95 (1968)].
[7]J. O. Hirshfelder, J. Chem. Phys. **39**, 2009 (1963).
[8]C. K. Au, Phys. Lett. A **77**, 221 (1980).
[9]C. K. Au and Y. Aharonov, Phys. Rev. A **22**, 328 (1980).
[10]H. Feshbach and F. Villars, Rev. Mod. Phys. **30**, 24 (1958).
[11]Also see, for example, G. Baym, *Lectures on Quantum Mechanics* (Benjamin, New York, 1969).

Formulation of variational principles via Lagrange multipliers

Kalman Kalikstein
Hunter College CUNY, Physics Department, New York, New York 10021

(Received 2 October 1980; accepted for publication 9 January 1981)

Recent work on variational principles in mathematical physics enables one to construct, in a novel and systematic way, stationary expressions for a wide class of functionals $P(\mathbf{E})$, where $\mathbf{E}$ is an unknown (vector) function whose defining equation cannot be solved exactly. The method involves the use of Lagrange multipliers, which can be a constant λ, a function $\mathbf{F}(x,y,z)$, and a dyadic operator Γ, to account for each of the equations (constraints) that define $\mathbf{E}$. As illustrations, we consider vector and scalar scattering problems, and eigenvalue problems such as the determination of resonant frequencies and propagation constants of waveguides.

PACS numbers: 03.65.Ge, 03.80. + r, 84.40.Sr, 02.30.Jr

I. INTRODUCTION

Variational formulations, such as the Rayleigh–Ritz principle, have been employed to determine the discrete eigenvalues for bound state problems. Variational principles have also been used in waveguide[1] and quantum-mechanical scattering,[2–4] acoustical and optical diffraction,[5] and neutron diffusion problems.[6] Different procedures, such as Rumsey's reaction concept,[7,8] Cairo and Kahan's transpose-operator and field technique,[9] have been used in a systematic derivation of variational principles in electromagnetic eigenvalue, scattering, diffraction, and radiation problems. In addition, Morishita and Kumagai[10] have derived by means of Hamilton's principle variational expressions for waveguide eigenvalue problems and the impedance matrix of an n-port waveguide junction.

The extremum values of a functional whose arguments are constrained are normally determined through the use of Lagrange undertermined multipliers. This suggested to Spruch and his co-workers that one can construct variational principles for functionals $P(\mathbf{E})$ [11] and for the function $\mathbf{E}$ itself[12] by incorporating in the variational principle constraints on $\mathbf{E}$ via Lagrange undetermined multipliers. For example, if $\mathbf{E}$ represents a normalized electric field which obeys the Maxwell wave equation, then the normalization constraint will require a constant multiplier, while the wave equation, which is a contraint at each point in space, will require an undetermined function.

The method of constraints is applicable to a very wide class of problems and has the advantage of being systematic and, in general, straightforward. As illustrations, we will derive here Schwinger's principle for the scattering of electromagnetic waves in waveguides,[1] variational expressions for the scattering amplitude of scalar waves due to potentials and to perturbations on reflecting surfaces,[9,13] and stationary expressions for the resonance frequencies of cavities and propagation constants of waveguides.[9,14]

II. VARIATIONAL EXPRESSIONS FOR SCATTERING PROBLEMS

A. Schwinger's variational principle for waveguides

Here we derive Schwinger's variational principle for the scattering of an electromagnetic wave by a dielectric obstacle in a waveguide. The electric field $\mathbf{E}$ satisfies the wave equation

$$-\nabla\times(\mu^{-1}\nabla\times\mathbf{E})+(\omega^2/c^2+W)\mathbf{E}=-\mathscr{L}\mathbf{E}=0, \quad (1)$$

where μ, ω, and c are the relative permeability, angular frequency and speed of light respectively. W is defined by

$$W=\omega^2[\epsilon(x,y,z)-1]/c^2, \quad (2)$$

where ϵ is the relative permittivity. The quantum-mechanical analog of W is proportional to the potential. Let $\mu=1$ and, for simplicity, assume that the obstacle is symmetric with respect to a plane perpendicular to the direction of propagation z. The problem is then analyzed in terms of even and odd standing waves. We will consider only the odd case. For H modes in a homogeneous waveguide the odd component of $\mathbf{E}$ satisfies the condition $\mathbf{E}_0=0$ at $z=0$, and has the asymptotic form for $z\sim\infty$

$$\mathbf{E}_0(x,y,z)\sim\mathbf{e}(x,y)[\sin kz-\tan\eta_0\cos kz], \quad (3)$$

where $\mathbf{e}(x,y)$ is the form function of the propagating mode and $\mathbf{e}$ is normalized such that $\int\mathbf{e}\cdot\mathbf{e}dS=1$, and the surface integration dS is over the guide's cross section. k is the propagation constant, and η_0 is the odd phase shift. Now

$$(-\nabla\times\nabla\times+\omega^2/c^2)\mathbf{G}_0(\mathbf{r},\mathbf{r}')=-\mathscr{L}^{(0)}\mathbf{G}_0$$
$$=-\mathbf{I}\delta(\mathbf{r}-\mathbf{r}'), \quad (4a)$$

where

$$\mathscr{L}^{(0)}=\mathscr{L}+W, \quad (4b)$$

$\mathbf{I}$ is the idemfactor, and $\mathbf{G}_0$ is the odd dyadic Green's function given by Schwinger[1] for TE_{n0} modes in rectangular guide and by Tai[15] for the general case. The Green's function satisfies the conditions

$$\mathbf{G}_0(0,\mathbf{r}')=0,$$

and

$$\mathbf{G}_0\sim-\mathbf{e}(x,y)\mathbf{e}(x',y')[\sin kz_<\cos kz_>]/k, \quad z\sim\infty. \quad (5)$$

$\mathbf{E}_0$ can be expressed in the integral equation form

$$\mathbf{E}_0=\mathbf{e}\sin kz+\int\mathbf{G}_0(\mathbf{r},\mathbf{r}')\cdot W(\mathbf{r}')\mathbf{E}_0(\mathbf{r}')d\tau', \quad (6)$$

where $d\tau$ is an element of volume and the z integration is from 0 to ∞. Asymptotically, from Eqs. (5) and (6),

 0022-2488/81/071433-05$1.00

$$\mathbf{E}_0 \sim \mathrm{e}\sin kz - (\mathrm{e}/k)\left[\int \mathrm{e}\sin kz' \cdot W\mathbf{E}_0 d\tau'\right]\cos kz. \tag{7}$$

Therefore, from Eqs. (3) and (7),

$$\tan\eta_0 = k^{-1}\int \mathrm{e}\sin kz \cdot W\mathbf{E}_0 d\tau$$

$$= k^{-1}\int \mathrm{e}\sin kz \cdot \mathscr{L}^{(0)}\mathbf{E}_0 d\tau. \tag{8}$$

By the way,[1]

$$\tan\eta_0 = -i(Z_{11} - Z_{12}), \tag{9}$$

where Z_{11} and Z_{12} are elements of the impedance network of the obstacle.

The problem has one constraint, Eq. (1), and the variational principle for $\tan\eta_0$ can be written down routinely[16]

$$(k\tan\eta_0)_{\mathrm{VAR}} = \int \mathrm{e}\sin kz \cdot \mathscr{L}^{(0)}\mathbf{E}_{0t} d\tau - \int \mathbf{F}_t \cdot \mathscr{L}\mathbf{E}_{0t} d\tau, \tag{10}$$

where

$$\mathbf{E}_{0t} = \mathbf{E}_0 + \delta\mathbf{E}_0 \quad \text{and} \quad \mathbf{F}_t = \mathbf{F} + \delta\mathbf{F}$$

are the trial electric field and trial Lagrange multiplier function. $\mathbf{E}_{0t}(x,y,0) = 0$ and asymptotically it has the same form as $\mathbf{E}_0$ in Eq. (3) except η_0 is replaced by η_{0t}. $\mathbf{F}$ is determined by the requirement that the first-order variation of Eq. (10) be zero. One obtains

$$\int \mathrm{e}\sin kz \cdot \mathscr{L}^{(0)}\delta\mathbf{E}_0 d\tau - \int \mathbf{F} \cdot \mathscr{L}\delta\mathbf{E}_0 d\tau = 0. \tag{11}$$

$\mathbf{E}_0$ can be written

$$\mathbf{E}_0 = \sum_{n=1}^{\infty} \mathbf{e}_n(x,y)h_n(z), \quad \text{at} \quad z\sim\infty \quad \mathbf{E}_0 = \mathbf{e}(x,y)h_1(z),$$

where $\mathbf{e} \equiv \mathbf{e}_1$ and $\mathbf{e}_n$ for $n \neq 1$ are the form factors of the propagating and evanescent modes, respectively. Assume that $\mathbf{F}$ is of the form

$$\mathbf{F} = \sum_{n=1}^{\infty} \mathbf{e}_n(x,y)f_n(z), \quad \text{at infinity} \quad \mathbf{F} = \mathbf{e}(x,y)f_1(z).$$

Integrating Eq. (11) by parts, it follows, since the modes are orthogonal, that

$$-\int \mathscr{L}\mathbf{F}\cdot\delta\mathbf{E}_0 d\tau - \left[(\sin kz \frac{d}{dz}\delta h_1 - \Sigma f_n \frac{d}{dz}\delta h_n)\right.$$

$$\left. - (\delta h_1 \frac{d}{dz}\sin kz - \Sigma\delta h_n \frac{d}{dz}f_n)\right]_0^{\infty} = 0. \tag{12}$$

Now $\delta h_n = 0$ at $z = 0$ since $\mathbf{E}_{0t}$ satisfies the same boundary conditions as $\mathbf{E}_0$, and therefore the term in the square bracket will vanish at the lower limit if $f_n = 0$ at $z = 0$. Asymptotically, only the δh_1 term survives and is proportional to $\cos kz$, or $d(\delta h_1)/dz$ to $\sin kz$, therefore the term in the square bracket will vanish at the upper limit ($z = \infty$) if $(\sin kz - f_1)$ is asymptotically of the form $\cos kz$. It follows from Eq. (12) and the above boundary conditions that

$$\mathscr{L}\mathbf{F} = 0 \quad \text{and} \quad \mathbf{F} = \mathbf{E}_0. \tag{13}$$

The expression in Eq. (10), with $\mathbf{F}_t$ replaced by $\mathbf{E}_{0t}$, is the analog of Kohn's variational principle in quantum mechanics.[2]

We will now deduce Schwinger's variational principle from Kohn's variational principle. Using Eqs. (4b) and (6), and $\mathbf{F}_t = \mathbf{E}_{0t}$, the term $\int \mathbf{F}_t \cdot \mathscr{L}\mathbf{E}_{0t} d\tau$ in Eq. (10) is equal to

$$\int\left[\mathbf{E}_{0t}\cdot(\mathscr{L}^{(0)} - W)(\mathrm{e}\sin kz + \int \mathbf{G}_0 \cdot W\mathbf{E}_{0t}\, d\tau')\right]d\tau$$

$$= -\int \mathbf{E}_{0t}\cdot W\mathrm{e}\sin kz\, d\tau + \int \mathbf{E}_{0t}\cdot W\mathbf{E}_{0t}\, d\tau$$

$$- \int\left[\mathbf{E}_{0t}\, W\cdot\int \mathbf{G}_0\cdot W\mathbf{E}_{0t} d\tau'\right] d\tau. \tag{14}$$

The relationships

$$\mathscr{L}^{(0)}\mathrm{e}\sin kz = 0 \text{ and } \mathscr{L}^{(0)}\mathbf{G}_0 = \mathbf{I}\delta(\mathbf{r} - \mathbf{r}')$$

were used in the derivation. Therefore, Eq. (10) reduces to Schwinger's variational expression for $\tan\eta_0$

$$(k\tan\eta_0)_{\mathrm{VAR}} = -\int \mathbf{E}_{0t}\cdot W\mathbf{E}_{0t}\, d\tau$$

$$+ \int\left[\mathbf{E}_{0t}\, W\cdot\int \mathbf{G}_0\cdot W\mathbf{E}_{0t}\, d\tau'\right]d\tau$$

$$+ 2\int \mathbf{E}_{0t}\cdot W\mathrm{e}\sin kz\, d\tau. \tag{15}$$

Kohn's and Schwinger's variational principles produce identical results for identical trial functions. Omitting the subscripts t and VAR, transfering the last term in Eq. (15) to the left side, and dividing the resulting equation by $[\int \mathbf{E}_0\cdot W\mathrm{e}\sin kz\, d\tau]^2$ we are led to Schwinger's stationary expression for $\cot\eta_0$ (p. 51 of Ref. 1), in which normalization does not enter,

$$\frac{\cot\eta_0}{k} = \frac{i}{k}\frac{1}{(Z_{11} - Z_{12})}$$

$$= \frac{\int W\mathbf{E}_0^2 d\tau - \int\int W(\mathbf{r})\mathbf{E}_0(\mathbf{r})\cdot\mathbf{G}_0(\mathbf{r},\mathbf{r}')\cdot W(\mathbf{r}')\mathbf{E}(\mathbf{r}')\, d\tau' d\tau}{[\int \mathrm{e}\sin kz\cdot W\mathbf{E}_0\, d\tau]^2}. \tag{16}$$

B. Scattering amplitude

We can now derive Schwinger's variational principle for the scattering amplitude by mere inspection. Consider the scattering of scalar waves (sound, quantum mechanical) by a potential. The wave equation has the form

$$[\nabla^2 + k^2 - U(\mathbf{r})]\psi(\mathbf{r}) = 0, \tag{17}$$

where U is proportional to the potential. Asymptotically

$$\psi(\mathbf{r}) \sim \exp(i\mathbf{k}_i\cdot\mathbf{r}) + (1/r)e^{ikr}f(\theta,\phi), \quad r\sim\infty, \tag{18}$$

where $f(\theta,\phi)$ is the scattering amplitude and $\mathbf{k}_i$ is of magnitude k and points in the direction of the incident wave. The solution of Eq. (17) can be expressed in the integral equation form

$$\psi(\mathbf{r}) = \exp(i\mathbf{k}_i\cdot\mathbf{r}) + \int \psi(\mathbf{r}')G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')d\tau', \tag{19}$$

where the Green's function G satisfies the equation

$$(\nabla^2 + k^2)G = \delta(\mathbf{r} - \mathbf{r}'), \tag{20}$$

and is given by

$$G = -\frac{1}{4\pi}\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|}. \tag{21}$$

For large r,

$$k\,|\mathbf{r}-\mathbf{r}'| \sim kr - \mathbf{k}_S\cdot\mathbf{r}', \tag{22}$$

where $\mathbf{k}_S$ is of magnitude k and points in the direction of the scattered wave. Thus,

$$f(\theta,\phi) = -(4\pi)^{-1}\int \exp(-i\mathbf{k}_S\cdot\mathbf{r}')U(\mathbf{r}')\psi(\mathbf{r})\,d\tau'. \tag{23}$$

Comparing Eqs. (6) and (8) with (19) and (23), respectively, we notice that ψ corresponds to $\mathbf{E}_0$,exp$(-i\mathbf{k}_i\cdot\mathbf{r})$ to esinkz, U to W, and $4\pi f$ to $-k\tan\eta_0$. Therefore from Eq. (16), the variational expression for f can be written

$$\frac{1}{4\pi f} = \frac{-\int\psi^\dagger(\mathbf{r})U(\mathbf{r})\psi(\mathbf{r})\,d\tau + \int\int\psi^\dagger(\mathbf{r})U(\mathbf{r})G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\psi(\mathbf{r}')\,d\tau d\tau'}{\int\psi^\dagger(\mathbf{r})U(\mathbf{r})\exp(i\mathbf{k}_i\cdot\mathbf{r})\,d\tau\int\exp(-i\mathbf{k}_S\cdot\mathbf{r}')U(\mathbf{r}')\psi(\mathbf{r}')\,d\tau'}. \tag{24}$$

The dagger denotes the adjoint (complex conjugate transpose), and $\psi^\dagger$ is the adjoint solution of the wave equation. It is given by

$$\psi^\dagger(\mathbf{r}) = \exp(-i\mathbf{k}_S\cdot\mathbf{r}) + \int G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\psi^\dagger(\mathbf{r}')\,d\tau', \tag{25}$$

and corresponds to a plane wave incident along the $-\mathbf{k}_S$ direction, In subsection A, there was no need for an adjoint function since $\mathbf{E}_0$ was real. In a similar fashion one can derive the stationary expression for the transmission amplitude in one dimensional scattering.[13]

C. Scattering at a surface

Here we consider the scattering of a scalar plane wave by a reflecting surface. The wave equation satisfies the Helmholtz equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0, \tag{26}$$

and the boundary conditions are either of the Dirichlet type

$$\psi = 0 \quad \text{on the surface } S \tag{27}$$

or the Neumann type

$$\frac{\partial\psi}{\partial n} = 0 \quad \text{on the surface } S, \tag{28}$$

where the derivative is outwardly directed and normal to the surface. For the Dirichlet case we have

$$\psi(\mathbf{r}) = \exp(i\mathbf{k}_i\cdot\mathbf{r}) - \oint\oint G(\mathbf{r},\mathbf{r}')\frac{\partial}{\partial n'}\psi(\mathbf{r}')dS', \tag{29}$$

where the integration is over the closed surface S and G is given, as before, by

$$G = -\frac{1}{4\pi}\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}. \tag{30}$$

From Eqs. (18), (29), and (30) it follows that the scattering amplitude is defined by

$$f(\theta,\phi) = \frac{1}{4\pi}\int\exp(-i\mathbf{k}_S\cdot\mathbf{r}')\frac{\partial}{\partial n'}\psi(\mathbf{r}')dS'. \tag{31}$$

The variational expression for $f(\theta,\phi)$ has the form

$$(4\pi f)_{\mathrm{VAR}} = \oint\exp(-i\mathbf{k}_S\cdot\mathbf{r}')\frac{\partial}{\partial n'}\psi_t(\mathbf{r}')\,dS + \int\mathbf{F}_t^\dagger(\mathbf{r}')(\nabla'^2 + k^2)\psi_t(\mathbf{r}')\,d\tau', \tag{32}$$

where ψ_t and $F_t^\dagger$ are the trial function and trial Lagrange multiplier. The requirement that the first order variation of f be zero gives

$$\oint\exp(-i\mathbf{k}_S\cdot\mathbf{r}')\frac{\partial}{\partial n'}(\delta\psi)dS' + \int F^\dagger(\nabla'^2 + k^2)\,\delta\psi d\tau' = 0. \tag{33}$$

Integrating by parts, one obtains

$$\oint\exp(-i\mathbf{k}_S\cdot\mathbf{r}')\frac{\partial}{\partial n'}(\delta\psi)dS' + \int\delta\psi(\nabla'^2 + k^2)F^\dagger d\tau' + \oint_{S+S_0}\left(F^\dagger\frac{\partial}{\partial n'}\delta\psi - \delta\psi\frac{\partial}{\partial n'}F^\dagger\right)dS' = 0. \tag{34}$$

The integration in the last term of Eq. (34) is over the closed reflecting surface S and a closed surface S_0 at infinity. It follows that

$$(\nabla^2 + k^2)F^\dagger = 0$$

and

$$F^\dagger = -\psi^\dagger, \tag{35}$$

if ψ_t satisfies the same boundary conditions as ψ on the surface S. Replacing $F_t^\dagger$ by $-\psi_t^\dagger$ and substituting the integral equation(29) for ψ_t in Eq. (32), one obtains

$$(4\pi f)_{\mathrm{VAR}} = 2\oint\exp(-i\mathbf{k}_S\cdot\mathbf{r}')\frac{\partial}{\partial n'}\psi_t(\mathbf{r}')dS' - \oint\oint\frac{\partial}{\partial n}\psi_t^\dagger(\mathbf{r})G(\mathbf{r},\mathbf{r}')\frac{\partial}{\partial n'}\psi_t(\mathbf{r}')dSdS', \tag{36}$$

where $\psi^\dagger$ is the adjoint solution of the Helmholtz equation. In the derivation we took account of $(\nabla^2 + k^2)\exp(-i\mathbf{k}_S\cdot\mathbf{r}) = 0$ and $(\nabla^2 + k^2)G = \delta(\mathbf{r}-\mathbf{r}')$. The above variational expression for f can be made homogeneous in ψ. As in Eq. (15), this is accomplished by introducing a multiplying factor and demanding that the expression be stationary with respect to that factor. Namely, by dividing the second term on the right-hand side of Eq. (36) by the first term times its adjoint,

$$\oint\exp(-i\mathbf{k}_S\cdot\mathbf{r}'\frac{\partial}{\partial n'}\psi(\mathbf{r}')\,dS' \times\oint\left[\frac{\partial}{\partial n}\psi^\dagger(\mathbf{r})\right]\exp(i\mathbf{k}_i\cdot\mathbf{r})\,dS, \tag{37}$$

we obtain Schwinger's normalized stationary expression for f[5,9,13]

$$\frac{1}{4\pi f} = \frac{\oint\oint(\partial/\partial n)\psi^\dagger(\mathbf{r})G(\mathbf{r},\mathbf{r}')(\partial/\partial n')\psi(\mathbf{r}')\,dSdS'}{\oint\exp(-i\mathbf{k}_S\cdot\mathbf{r}')\frac{\partial}{\partial n'}\psi(\mathbf{r}')\,dS'\oint[(\partial/\partial n)\psi^\dagger(\mathbf{r})]\exp(i\mathbf{k}_i\cdot\mathbf{r})\,dS}. \tag{38}$$

Using the same approach, one can derive the scattering amplitude for the Neumann case. The extension to the scattering of electromagnetic waves, which are vectorial in character, by conducting surfaces (Chap. 7 of Ref. 9) is straightforward.

III. VARIATIONAL EXPRESSIONS FOR ELECTROMAGNETIC EIGENVALUE PROBLEMS

The prescription presented in Secs. I and II will be used to derive Berk's[14,9] stationary expressions for the resonance frequencies and propagation constants in waveguides in terms of the electric field **E**. The field satisfies the equation

$$-\nabla\times(\mu^{-1}\nabla\times\mathbf{E})+\epsilon(\omega/c)^2\mathbf{E}=[-\mathscr{H}_0+\epsilon(\omega/c)^2]\mathbf{E}=0, \tag{39}$$

where

$$\mathscr{H}_0=\nabla\times\mu^{-1}\nabla\times .$$

E is normalized such that

$$\int\mathbf{E}^\dagger\cdot\rho\mathbf{E}d\tau=1, \tag{40}$$

where ρ is a weight factor. The variational principle for an arbitrary, vector or scalar Hermitian operator Λ can now be expressed in the form

$$\begin{aligned}\left(\int\mathbf{E}^\dagger\cdot\Lambda\,\mathbf{E}\,d\tau\right)_{\mathrm{VAR}}&=\int\mathbf{E}_t^\dagger\cdot\Lambda\,\mathbf{E}_t\,d\tau+\lambda_t\left[\int\mathbf{E}_t^\dagger\cdot\rho\mathbf{E}_t\,d\tau-1\right]\\&+\int\mathbf{F}_t^\dagger\cdot[H_0-\epsilon(\omega/c)^2]\mathbf{E}_t\,d\tau\\&+\int\mathbf{F}_t\cdot[(\mathscr{H}_0-\epsilon\omega^2/c^2)\mathbf{E}_t]^\dagger\,d\tau,\end{aligned} \tag{41}$$

where $\mathbf{E}_t$, $\mathbf{F}_t$, and λ_t are the trial electric field, constraint function, and constraint multiplier, respectively. The preceding variational principle for the normalized (bound state function) is more complicated than the one, Eq. (10), for the scattering wavefunction. The reason is that the normalization condition (40) contains $\mathbf{E}^\dagger$, and, therefore, the last term in the right-hand side of Eq. (41) must be included to account for the wave equation obeyed by $\mathbf{E}^\dagger$. λ and **F** are determined by the requirement that $\delta\int\mathbf{E}^\dagger\cdot\Lambda\,\mathbf{E}d\tau$ vanishes to first-order in $\delta\,\mathbf{E}$. Neglecting second-order terms, it follows that

$$[\mathscr{H}_0-\epsilon(\omega/c)^2]\mathbf{F}+\Lambda\,\mathbf{E}+\lambda\rho\mathbf{E}=0, \tag{42}$$

if the condition

$$[(\mathscr{H}_0-\epsilon\omega^2/c^2)\,\delta\mathbf{E}]^\dagger\cdot\mathbf{F}=\delta\mathbf{E}^\dagger\cdot[\mathscr{H}_0-\epsilon\omega^2/c^2]\mathbf{F} \tag{43}$$

is satisfied, i.e., if μ and the ϵ are Hermitian tensors. Multiplying Eq. (42) with $\mathbf{E}^\dagger$, and, integrating, one obtains

$$\lambda=-\int\mathbf{E}^\dagger\cdot\Lambda\,\mathbf{E}\,d\tau. \tag{44}$$

Thus, Eq. (42) becomes

$$[\mathscr{H}_0-\epsilon(\omega/c)^2]\mathbf{F}=\rho\mathbf{E}\int\mathbf{E}^\dagger\cdot\Lambda\,\mathbf{E}\,d\tau-\Lambda\,\mathbf{E}. \tag{45}$$

Equation (41) in conjunction with λ_t and $\mathbf{F}_t$, trial estimates of λ and **F** which satisfy Eqs. (44) and (45), respectively, define the variational principle. If $\mathbf{E}_t$ has the same normalization as **E**, the term involving λ_t disappears. This variational principle for an arbitrary Hermitian operator[16] was originally derived (by another procedure) by Schwartz[17] and Delves.[18]

In general, **F** differs from **E** but if

$$\Lambda\,\mathbf{E}=\alpha\rho\mathbf{E}, \tag{46}$$

where α is an eigenvalue and **E** an eigenfunction of Λ, then Eq. (45) becomes

$$(\mathscr{H}_0-\epsilon\omega^2/c^2)\mathbf{F}=0,\quad\text{and}\quad\mathbf{F}=\mathbf{E}. \tag{47}$$

The variational principle reduces then to

$$\begin{aligned}\left(\int\mathbf{E}^\dagger\cdot\Lambda\,\mathbf{E}\,d\tau\right)_{\mathrm{VAR}}&=\int\mathbf{E}_t^\dagger\cdot\Lambda\,\mathbf{E}_t\,d\tau\\&+2\int\mathbf{E}_t^\dagger\cdot(\mathscr{H}_0-\epsilon\omega^2/c^2)\mathbf{E}_t\,d\tau.\end{aligned} \tag{48}$$

It is assumed that ω is known experimentally essentially exactly or is estimated variationally. The following examples serve as illustrations of (48).

A. Resonance frequency of a cavity

Here it is assumed that ω is not known. Let

$$\Lambda=\nabla\times\mu^{-1}\nabla\times=\mathscr{H}_0,\quad\text{and}\quad\rho=\epsilon, \tag{49}$$

then $\alpha=(\omega/c^2)$. Replacing Λ in the rhs of (48) by $\epsilon(\omega/c)^2$, one obtains

$$\begin{aligned}\left(\int\mathbf{E}^\dagger\cdot\Lambda\,\mathbf{E}\,d\tau\right)_{\mathrm{VAR}}&=(\omega^2/c^2)_{\mathrm{VAR}}\\&=2\int\mathbf{E}_t^\dagger\cdot\mathscr{H}_0\mathbf{E}_t\,d\tau-(\omega/c)^2.\end{aligned} \tag{50}$$

Since $(\omega^2)_{\mathrm{VAR}}$ differs by a second-order term from ω^2, one has

$$(\omega^2/c^2)_{\mathrm{VAR}}=\int\mathbf{E}_t^\dagger\cdot\mathscr{H}_0\mathbf{E}_t\,d\tau. \tag{51}$$

For $\mathbf{E}_{\tan}=0$ on conducting surfaces, one obtains Berk's stationary expression

$$(\omega/c)^2=\int(\nabla\times\mathbf{E})^\dagger\cdot(\nabla\times\mathbf{E})d\tau. \tag{52}$$

Equation (51) is equivalent to Rayleigh–Ritz's principle, and, of course, could have been derived in a much simpler way.

B. Propagation constant of a waveguide

Consider a waveguide containing a medium whose permittivity and permeability are Hermitian tensors. The medium is inhomogeneous in the directions transverse to the direction of paropagation z. **E** can be written

$$\mathbf{E}=\mathbf{e}(x,y)e^{-j\gamma z}, \tag{53}$$

where γ is the propagation constant. Let

$$\rho=1,\quad\text{and}\quad\Lambda=j\frac{\partial}{\partial z}; \tag{54}$$

then $\alpha=\gamma$ and

$$\left(\int \mathbf{E}^{\dagger}\cdot\Lambda\,\mathbf{E}\,d\tau\right)_{\mathrm{VAR}} = \gamma_{\mathrm{VAR}}$$
$$= \gamma + 2\int \mathbf{E}_t^{\dagger}\cdot(\mathscr{H}_0 - \epsilon\omega^2/c^2)\mathbf{E}_t\,d\tau. \tag{55}$$

Since γ_{VAR} differs by a second-order term from γ, it follows that

$$\int \mathbf{E}\cdot(\mathscr{H}_0 - \epsilon\omega^2/c^2)\mathbf{E}\,d\tau = 0 \tag{56}$$

is a stationary expression. The above result reduces to Berk's stationary expression for the propagation constant

$$\gamma^2\int(\mathbf{i}_z\times\mathbf{e}^{\dagger})\cdot\mu^{-1}(\mathbf{i}_z\times\mathbf{e})\,dS$$
$$-j\gamma\int[(\nabla_T\times\mathbf{e}^{\dagger})\cdot\mu^{-1}(\mathbf{i}_z\times\mathbf{e}) - (\mathbf{i}_z\times\mathbf{e}^{\dagger})\cdot\mu^{-1}(\nabla_T\times\mathbf{e})]\,dS$$
$$+\int(\nabla_T\times\mathbf{e}^{\dagger})\cdot\mu^{-1}(\nabla_T\times\mathbf{e})\,dS - (\omega/c)^2\int\mathbf{e}\cdot\epsilon\mathbf{e}\,dS = 0, \tag{57}$$

by writing

$$\nabla = \nabla_T + \mathbf{i}_z\frac{\partial}{\partial z} = \nabla_T - j\mathbf{i}_z\frac{\partial}{\partial z}, \tag{58}$$

where $\mathbf{i}_z$ is a unit vector in the z direction and ∇_T is the transverse component of the gradient. The variational principle can be extended to Λ not a Hermitian operator.

ACKNOWLEDGMENT

I would like to thank Mr. Alonso Sepulveda for helpful comments. The work has been supported in part by a grant from the City University of New York Faculty Research Grant.

[1] J. Schwinger and D. S. Saxon, *Discontinuities in Waveguides*(Gordon and Breach, New York, 1968).
[2] W. Kohn, Phys. Rev. **74**, 1763 (1948).
[3] J. M. Blatt and J. D. Jackson, Phys. Rev. **76**, 18 (1949).
[4] B. A. Lippman and J. Schwinger, Phys. Rev. **79**, 469 (1950).
[5] H. Levine and J. Schwinger, Phys. Rev. **74**, 958 (1948); **75**, 1423 (1949).
[6] R. E. Marshak, Phys. Rev. **71**, 688 (1947).
[7] V. H. Rumsey, Phys. Rev. **94**, 1483 (1954).
[8] R. F. Harrington, *Time-Harmonic Electromagnetic Fields* (McGraw-Hill, New York, 1961), Chap. 7.
[9] L. Cairo and T. Kahan, *Variational Techniques in Electromagnetism* (Gordon and Breach, New York, 1965), Chaps. 4–7.
[10] K. Morishita and N. Kumagai, IEEE Trans. **MTT-25**, 34 (1977).
[11] E. Gerjuoy, A. R. P. Rau, and L. Spruch, Phys. Rev. A **8**, 662 (1973).
[12] E. Gerjuoy, A. R. P. Rau, L. Rosenberg, and L. Spruch, J. Math. Phys. **16**, 1104 (1975).
[13] B. L. Moiseiwitsch, *Variational Principles* (Wiley, New York, 1966), Chap. 5.
[14] A. D. Berk, IRE Trans. Antennas Propagation AP-4, 104 (1956).
[15] C. T. Tai, *Dyadic Green's Functions in Electromagnetic Theory* (McGraw-Hill, New York, 1960).
[16] E. Gerjuoy, A. R. P. Rau, and L. Spruch, J. Math. Phys. **13**, 1797 (1972).
[17] C. Schwartz, Ann. Phys. (N.Y.) **2**, 156, 170 (1959).
[18] L. M. Delves, Nucl. Phys. **41**, 479 (1963).

Integrated rotational parameter as a constraint for the variational lower bound of the diffraction peak

I. A. Sakmar [a), b)]
Université de Nice, Laboratoire de Physique Théorique, Parc Valrose 06034 Nice Cedex, France [c)]

(Received 17 June 1980; accepted for publication 23 January 1981)

The contributions of the imaginary parts of the partial waves to an angle integral of the rotation parameter R is taken as a third constraint in addition to the total cross section σ^{T} and $\sigma_{el,im}$, in the variational calculation of the lower bound for the diffraction peak.

PACS numbers: 03.80. + r, 11.80.Et

I. INTRODUCTION

Unitarity in the form of the positivity and boundedness of the imaginary parts of the partial waves together with the constraints of total and elastic cross sections were used by MacDowell and Martin[1] in 1964 to obtain a lower bound on the diffraction peak with the Lagrange multipliers method. Since then more work has been done on the subject,[2-8] including recently, on the extension to the spin case.

For the pion-nucleon scattering in addition to the total and elastic cross sections good polarization measurements over the entire angle region are available. It would be nice if one could use one more input to improve the bound. As constraints, both the total and elastic cross sections are gobal quantities, that is quantities with no z dependence, or integrated over the angle. This is needed because of the nature of the method used. Therefore an additional input has to be of the same nature, so that we have to use an integrated polarization or some combination of it.

However, with the integral of the polarization one encounters several problems. The integral of the differential cross section reduces to a single series because of the orthogonality of the Legendre polynomials in the case of the spin-non-flip amplitude and of the derivatives of the Legendre polynomials with the weight factor $\sin^2\theta$ in the case of the spin-flip amplitude. For the polarization one has the product of P_l's with dP_l/dz 's which do not have an orthogonality relation. The second point is that the contribution of the imaginary parts of the partial waves to $\sigma^{T}$and σ_{el} are separable from those of the real parts and also are positive. In the case of σ_{el} these contributions are less than those of the entire partial waves. This makes it possible to write an inequality like $\sigma_{el,im} \leqslant \sigma_{el}$. For the integrated polarization the imaginary parts of the partial waves appear as products with the real parts, so that their contribution is not separable. Also the integral is not positive, neither can we write an inequality like the one given above.

The rotation parameter is a better candidate as a constraint, because in this case the contributions of the imaginary parts of the partial waves can be separated from those of the real parts. The integral is still not positive and an inequality in the above sense can still not be written. In spite of these difficulties we have considered this problem because it exhibits novel features which one does not have in the two constraint case. We assume the contribution of the imaginary parts of partial waves to a certain integral of the rotation parameter to have been calculated from the phase shifts. But after it is found it becomes a global constraint and is treated as a parameter. As such the results become an internal consistency check, or in the opposite direction, a bound on the contributions of the imaginary parts of the partial waves. We also recover the formulas for the spin 0-$\frac{1}{2}$ scattering case, which are the same as the ones for the scalar case.

In Sec. II we define our quantities, specify the first two constraints, and evalutate the third constraint as a single series.

In Sec. III we apply the Lagrange multipliers method and find two coupled equations whose solutions give us the imaginary parts of the l_+ and l_- amplitudes in terms of Lagrange multipliers and the wave number l. Unitarity is imposed in the form of boundedness and positiveness and constraints are expressed in terms of the Lagrange multipliers alone.

Finally in the conclusion we summarize and discuss the results.

II. SETUP AND CONSTRAINTS

The spin-non-flip (f) and spin-flip (g) amplitudes of the spin 0-$\frac{1}{2}$ scattering process are given by

$$f = \frac{\sqrt{s}}{k}\sum_l [(l+1)f_{l+} + f_{l-}]P_l(z), \tag{1}$$

$$g = \frac{\sqrt{s}}{k}\sum_l (f_{l+} - f_{l-})P'_l(z). \tag{2}$$

with these definitions the differential cross section is

$$\frac{d\sigma}{d\Omega} = \frac{1}{s}(|f|^2 + |g|^2\sin^2\theta). \tag{3}$$

The total and elastic cross sections are

$$\sigma^{T} = \frac{4\pi}{k^2}\sum_l [(l+1)a_{l+} + la_{l-}], \tag{4}$$

$$\sigma_{el} = \frac{4\pi}{k^2}\sum_l [(l+1)|f_{l+}|^2 + l|f_{l-}|^2]. \tag{5}$$

Here

$$a_{l+} = \mathrm{Im} f_{l+}, \tag{6}$$

[a)] On leave from the Applied Mathematics Department, University of Western Ontario, London, Ontario, Canada.

[b)] Research supported partially by the National Research Council of Canada.

[c)] Equipe de Recherche Ass. C.N.R.S.

 0022-2488/81/071438-05$1.00

$$a_{l-} = \mathrm{Im} f_{l-}. \tag{7}$$

The expression of σ^{T} is obtained from the "optical theorem"

$$\sigma^{T} = (4\pi/k\sqrt{s})\,\mathrm{Im} f(z=1), \tag{8}$$

and σ_{el} by integrating Eq. (3) and using the orthogonality relations of the Legendre Polynomials and their derivatives. Only the contributions of a_{l+} and a_{l-} to σ_{el} are taken as constraint. Hence one defines

$$\sigma_{el,im} = \frac{4\pi}{k^2}\sum_{l}[(l+1)|a_{l+}|^2 + l|a_{l-}|^2] < \sigma_{el}. \tag{9}$$

The quantity to be extremized is

$$\frac{dA}{dt}\Big|_{t=0} = \frac{1}{2k^2}\frac{\sqrt{s}}{k}\sum_{l}^{\frac{1}{2}} l(l+1)[(l+1)a_{l+} + la_{l-}], \tag{10}$$

where $A \equiv \mathrm{Im} f$.

We consider now

$$RD = 2\mathrm{Re}(g^{*}f)\sin\theta, \tag{11}$$

where R is the rotation parameter and $D \equiv d\sigma/d\Omega$, the differential cross section.

$$g^{*}f = \frac{s}{k^2}\sum_{l}\sum_{n}(f^{*}_{l+} - f^{*}_{l-})[(n+1)f_{n+} + nf_{n-}] \times P'_{l}(z)P_{n}(z). \tag{12}$$

We form $(1-z^2)g^{*}f$ and integrate it over z. The z integral is

$$\int_{-1}^{+1}(1-z^2)P'_{l}P_{n}\,dz = \int_{-1}^{+1}(-lzP_{l} + lP_{l-1})P_{n}\,dz = \frac{l(l+1)}{2l+1}\left(-\int_{-1}^{+1}P_{l+1}P_{n}\,dz + \int_{-1}^{+1}P_{l-1}P_{n}\,dz\right)$$
$$= \frac{l(l+1)}{2l+1}\frac{2}{2n+1}(-\delta_{n,l+1} + \delta_{n,l-1}), \tag{13}$$

and we find

$$\frac{k^2}{s}\int_{-1}^{+1}(1-z^2)g^{*}f\,dz$$
$$= \sum_{l}\frac{l(l+1)}{2l+1}2\left(-\frac{l+2}{2l+3}f^{*}_{l+}f_{(l+1)+} - \frac{l+1}{2l+3}f^{*}_{l+}f_{(l+1)-} + \frac{l+2}{2l+3}f^{*}_{l-}f_{(l+1)+} + \frac{l+2}{2l+3}f^{*}_{l-}f_{(l+1)+}\right.$$
$$+ \frac{l+1}{2l+3}f^{*}_{l-}f_{(l+1)-} + \frac{l}{2l-1}f^{*}_{l+}f_{(l-1)+} + \frac{l-1}{2l-1}f^{*}_{l+}f_{(l-1)-} - \frac{l}{2l-1}f^{*}_{l-}f_{(l-1)+}$$
$$\left. - \frac{l-1}{2l-1}f^{*}_{l-}f_{(l-1)-}\right). \tag{14}$$

Contribution from imaginary parts of the partial waves to the real part of this quantity is

$$\left(\mathrm{Re}\frac{k^2}{s}\int_{-1}^{+1}g^{*}f(1-z^2)\,dz\right)_{i} = \sum_{l=1}^{\infty} -2\frac{l(l+1)(l+2)}{(2l+1)(2l+3)}a_{l+}a_{(l+1)+} + 2\frac{l(l+1)(l+2)}{(2l+1)(2l+3)}a_{l-}a_{(l+1)-}$$
$$- 2\frac{l(l+1)^2}{(2l+1)(2l+3)}a_{l+}a_{(l+1)-}$$
$$+ 2\frac{l(l+1)^2}{(2l+1)(2l+3)}a_{l-}a_{(l+1)-} + 2\frac{l^2(l+1)}{(2l-1)(2l+1)}a_{l+}a_{(l-1)+}$$
$$+ 2\frac{(l-1)l(l+1)}{(2l-1)(2l+1)}a_{l+}a_{(l-1)-} - 2\frac{l^2(l+1)}{(2l-1)(2l+1)}a_{l-}a_{(l-1)+}$$
$$- 2\frac{(l-1)l(l+1)}{(2l-1)(2l+1)}a_{l-}a_{(l-1)-} \equiv r. \tag{15}$$

As we mentioned in the Introduction, only products of imaginary parts of the partial waves appear in the real part of the quantity on the left side. The subscript i stands for the contribution from these imaginary parts. Had we taken Im instead of Re of the integral (which is the case of polarization rather than the rotation parameter), we would have cross terms only of the real and imaginary parts of the partial waves.

We defined our third constraint as r. We also define $\sigma^{T} \equiv t$, $\sigma_{el,im} \equiv h$ and $dA/dt\,|_{t=0} \equiv u$. Here t and h are the other two constraints.

III. VARIATIONS

We now apply the Lagrange multipliers method by forming $u - \alpha t - \beta h - \gamma r$ and taking its derivatives with respect to a_{l+} and a_{l-}. The calculations are lengthy, but after redefining the Lagrange multipliers one finds for $l \neq 0$,

$$a_{l+} = \alpha - \beta l(l+1)$$
$$+ \gamma\left(\frac{l+2}{(2l+1)(2l+3)}a_{(l+1)+} - 2\frac{(l+1)^2}{(2l+1)(2l+3)}a_{(l+1)-}\right.$$
$$\left. + \frac{l}{(2l-1)(2l+1)}a_{(l-1)+} + 2\frac{(l-1)l}{(2l-1)(2l+1)}a_{(l-1)-}\right), \tag{16}$$

$$a_{l-} = \alpha - \beta l(l+1)$$
$$+ \gamma\left(2\frac{(l+1)(l+2)}{(2l+1)(2l+3)}a_{(l+1)+} - \frac{(l+1)}{(2l+1)(2l+3)}a_{(l+1)-}\right.$$
$$\left. - \frac{l^2}{(2l-1)(2l+1)}a_{(l-1)+} - \frac{(l-1)}{(2l-1)(2l+1)}a_{(l-1)-}\right). \tag{17}$$

The solutions of these two coupled equations are

$$a_{l+} = \alpha + 4\beta\gamma^2 - \beta\,[l(l+1) - 2\gamma l\,], \tag{18}$$

$$a_{l-} = \alpha + 4\beta\gamma^2 - 2\beta\gamma - \beta\,[l(l+1) + 2\gamma l\,]. \tag{19}$$

Now the unitarity is imposed in the form

$$0 \leqslant a_{l+} \leqslant 1,\ 0 \leqslant a_{l-} \leqslant 1.$$

As for the scalar case we first consider $\alpha + 4\beta\gamma^2 > 1$ and distinguish three regions (Fig. 1)

(1) Regions where a_{l+} and a_{l-} are 1.

(2) Regions where a_{l+} and a_{l-} are given by Eqs. (18) and (19).

(3) Regions where a_{l+} and a_{l-} are zero.

The boundaries of these regions will be shown by l_+ and l_- between the first two regions, and by L_+ and L_- between the last two regions respectively, for a_{l+} and a_{l-}. These are obtained from Eqs. (18) and (19) by setting them equal to 1 or 0. The results are

$$l_+ = \gamma - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{\Delta_l}\,, \qquad l_- = -\gamma - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{\Delta_l}\,, \tag{20}$$

where $\Delta_l = 20\gamma^2 - 4\gamma + 1 + 4(\alpha - 1)/\beta$, and

$$L_+ = \gamma - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{\Delta_L}\,, \qquad L_- = -\gamma - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{\Delta_L}\,, \tag{21}$$

where $\Delta_L = 20\gamma^2 - 4\gamma + 1 + 4\alpha/\beta$.

These are the boundaries imposed by the unitarity, written in terms of the Lagrange mulipliers.

From L_+ one sees that if $\gamma < 0$, we must chose the $+$ sign of the square root. Also we must have $\frac{1}{2}\sqrt{\Delta_L} > |\gamma| + \frac{1}{2}$. From L_- one sees that if $\gamma < 0$ and if we chose the $-$ sign of the square root, we have

$$L_- = |\gamma| - \tfrac{1}{2} - \tfrac{1}{2}\sqrt{\Delta_L}.$$

Then it follows that $|\gamma| > \frac{1}{2} + \frac{1}{2}\sqrt{\Delta_L} > \frac{1}{2} + |\gamma| + \frac{1}{2} = |\gamma| + 1$. Hence we must chose the $+$ sign of the square root. From L_- one sees that if $\gamma > 0$ we must chose the $+$ sign of the square root. Also we must have $\frac{1}{2}\sqrt{\Delta_L} > |\gamma| + \frac{1}{2}$. From L_+ one sees that if $\gamma > 0$ and if we chose the $-$ sign of the square root we have $L_+ = |\gamma| - \frac{1}{2} - \frac{1}{2}\sqrt{\Delta_L}$.

Thus from the positivity of L_- and L_+ (or l_- and l_+) one must chose always the $+$ sign of the square root.

We further observe that the expressions (18) and (19) are the analogs of the scalar case with the γ terms appearing as additions due to the third constraint. For $\gamma = 0$ the spinless case is recovered. Thus we see that there is no difference between the spin 0-$\frac{1}{2}$ case with no third constraint and the spinless case. The effect of the third constraint is also visible from l_-, l_+ and L_-, L_+ if we form the differences

$$l_+ - l_- = 2\gamma, \qquad L_+ - L_- = 2\gamma.$$

When $\gamma = 0$, the boundaries of a_{l+} and a_{l-} merge.

We must now evaluate all four quantities t, h, u, and r in terms of the values of a_{l+} and a_{l-} in their respective regions. As in Ref. 1, we find those expressions by integrals. The integrals are over three regions (see Fig. 1). For $\gamma > 0$ the first region is from l_- to l_+ where $a_{l+} = 1$ and a_{l-} is given by Eq. (19). In the second region a_{l+} and a_{l-} are given by Eqs. (18) and (19) respectively. This region extends from l_+ to L_-. Finally, in the third region from L_- to L_+, a_{l+} is given by Eq. (18) and a_{l-} is zero. For $\gamma < 0$, $+$'s and $-$'s are interchanged. The results for the first three quantities are exact because there are no cross terms of a_{l+} and a_{l-} in them:

$$\begin{aligned}
\frac{dA}{dt}\Big|_{t=0} = {} & \tfrac{1}{2}k^{-2}(\sqrt{s}/k)^{\frac{1}{2}}\{\tfrac{1}{12}(3l_+^4 + 8l_+^3 + 6l_+^2) + \tfrac{1}{12}(\alpha + 4\beta\gamma^2)[3(L_+^4 - l_+^4) + 8(L_+^3 - l_+^3) + 6(L_+^2 - l_+^2)] \\
& - \tfrac{1}{60}\beta\,[10(L_+^6 - l_+^6) + 24(L_+^5 - l_+^5) + 15(L_+^4 - l_+^4)] \\
& - \tfrac{1}{30}\beta(1 - 2\gamma)[6(L_+^5 - l_+^5) + 15(L_+^4 - l_+^4) + 10(L_+^3 - l_+^3)] + \tfrac{1}{12}(3l_-^4 + 4l_-^3) \\
& + \tfrac{1}{12}(\alpha + 4\beta\gamma^2 - 2\beta\gamma)[3(L_-^4 - l_-^4) + 4(L_-^3 - l_-^3)] - \tfrac{1}{30}\beta\,[5(L_-^6 - l_-^6) + 6(L_-^5 - l_-^5)] \\
& + \tfrac{1}{20}\beta(1 + 2\gamma)[4(L_-^5 - l_-^5) + 5(L_-^4 - l_-^4)]\}.
\end{aligned} \tag{22}$$

In terms of the Lagrange parameters this reduces, after lengthy calculations, to

$$\frac{dA}{dt}\Big|_{t=0} = \frac{1}{2k^2}\frac{\sqrt{s}}{k}\frac{1}{12\beta^2}[3(\alpha + 4\beta\gamma^2)^2 + 1 + 24\beta\gamma^2(\alpha + 4\beta\gamma^2) - 24\beta\gamma^2(1 - \beta\gamma^2) - 16\gamma^3\beta^2 - 3\alpha]. \tag{23}$$

Similarly, we find for the total cross section

$$\begin{aligned}
\sigma^{\mathrm{T}} = {} & 4\pi/k^2[\tfrac{1}{2}(l_+^2 + 2l_+) + \tfrac{1}{2}(\alpha + 4\beta\gamma^2)(L_+^2 + 2L_+ - l_+^2 - 2l_+) \\
& - \tfrac{1}{12}\beta(3L_+^4 + 4L_+^3 - 3l_+^4 - 4l_+^3) - \tfrac{1}{6}\beta(1 - 2\gamma)(2L_+^3 + 3L_+^2 - 2l_+^3 - 3l_+^2) \\
& + \tfrac{1}{2}l_-^2 + \frac{(\alpha + 4\beta\gamma^2 - 2\beta\gamma)}{2}(L_-^2 - l_-^2) - \tfrac{1}{4}\beta(L_-^4 - l_-^4) - \tfrac{1}{3}\beta(1 + 2\gamma)(L_-^3 - l_-^3)\Big].
\end{aligned} \tag{24}$$

This becomes in terms of the Lagrange parameters

$$\sigma^{\mathrm{T}} = 4\pi/k^2(1/2\beta)[2(\alpha + 4\beta\gamma^2) - 1 + 16\gamma^2 + 8\gamma]. \tag{25}$$

Finally, $\sigma_{\text{el,im}}$ is given by

$$\begin{aligned}\sigma_{\text{el,im}} &= 4\pi/k^2\{\tfrac{1}{2}(l_+^2+2l_+)+\tfrac{1}{2}(\alpha+4\beta\gamma^2)^2(L_+^2+2L_+-l_+^2-2l_+)-\tfrac{1}{3}\beta(1-2\gamma)(\alpha+4\beta\gamma^2)(2L_+^3+3L_+^2-2l_+^3-3l_+^2)\\&+\tfrac{1}{12}\{\beta^2(1-2\gamma)^2-2\beta(\alpha+4\beta\gamma^2)\}(3L_+^4+4L_+^3-3l_+^4-4l_+^3)\\&+\tfrac{1}{10}\beta^2(1-2\gamma)(4L_+^5+5L_+^4-4l_+^5-5l_+^4)+\tfrac{1}{30}\beta^2(5L_+^6+6L_+^5-5l_+^6-6l_+^5)+\tfrac{1}{2}l_-^2\\&+\tfrac{1}{2}(\alpha+4\beta\gamma^2-2\beta\gamma)^2(L_-^2-l_-^2)-\tfrac{2}{3}\beta(1+2\gamma)(\alpha+4\beta\gamma^2-2\beta\gamma)(L_-^3-l_-^3)\\&+\tfrac{1}{4}[\beta^2(1+2\gamma)^2-2\beta(\alpha+4\beta\gamma^2-2\beta\gamma)](L_-^4-l_-^4)+\tfrac{2}{5}\beta^2(1+2\gamma)(L_-^5-l_-^5)+\tfrac{1}{6}\beta^2(L_-^6-l_-^6)\}.\end{aligned} \tag{26}$$

In terms of the Lagrange multipliers this becomes

$$\sigma_{\text{el,im}} = (4\pi/k^2)(1/3\beta)[3(\alpha+4\beta\gamma^2)-2+6\beta\gamma^2]. \tag{27}$$

The evaluation of our fourth quantity r poses some difficulties. The expression for r is given by Eq. (15). It contains products of a_l's with unequal indices. Before we insert the imaginary parts of the partial waves a_{l+} and a_{l-} from Eqs. (18) and (19) into the Eq. (15), we bring r into the following form:

$$r=\sum_{l=1}^{\infty}2(a_{l+}-a_{l-})\frac{l(l+1)}{2l+1}\left[\frac{1}{2l-1}[la_{(l-1)+}+(l-1)a_{(l-1)-}]-\frac{1}{2l+3}[(l+2)a_{(l+1)+}+(l+1)a_{(l+1)-}]\right]. \tag{28}$$

Now we can form from Eqs. (18) and (19) the following combinations:

$$a_{l+}-a_{l-}=2\beta\gamma(2l+1), \tag{29}$$

$$(l+1)a_{l+}+la_{l-}=(2l+1)[(\alpha+4\beta\gamma^2)-\beta l(l+1)]. \tag{30}$$

Inserting these two relations into the Eq. (28), we find for the l th term of the series

$$8\beta^2\gamma l(l+1)(2l+1). \tag{31}$$

However, the l th term has this form only in the region where a_{l+} and a_{l-} are given by the Eqs. (18) and (19). (For $\gamma>0$ between l_+ and L_-). In the transition regions l_- to l_+ and L_- to L_+, both of which have a width of 2γ, only one of the a_l's is given by the Eqs. (18) or (19), whereas the other is either 1 or 0 (Fig. 1). We now go back to Eqs. (16) and (17) and form their difference,

$$a_{l+}-a_{l-}=\gamma\left(-\frac{(l+2)}{(2l+3)}a_{(l+1)}-\frac{(l+1)}{(2l+3)}a_{(l+1)-}\right.$$

$$\left.+\frac{l}{(2l-1)}a_{(l-1)+}+\frac{(l-1)}{(2l-1)}a_{(l-1)-}\right) \tag{32}$$

Inserting this into the Eq. (28), we find for r,

$$r=\sum_{l=1}^{\infty}\frac{2}{\gamma}(a_{l+}-a_{l-})^2\frac{l(l+1)}{2l+1}. \tag{33}$$

This much simpler form has in the first transition region terms like $(2/\gamma)(1-a_{l-})^2l(l+1)/(2l+1)$, and in the second transition region terms like $(2/\gamma)a_{l+}^2\,l(l+1)/(2l+1)$, where a_{l-} and a_{l+} are given by the Eqs. (18) and (19). Putting everything together, we have for r

$$r=\int_{l-}^{l+}\frac{2}{\gamma}(1-a_{l-})^2\frac{l(l+1)}{2l+1}dl+8\beta\gamma^2\int_{l+}^{L-}l(l+1)(2l+1)\,dl+\int_{L-}^{L+}\frac{2}{\gamma}a_{l+}^2\frac{l(l+1)}{2l+1}dl. \tag{34}$$

For large energies we replaced the sums by integrals as is done in Ref. 1. When series are used a careful treatment of transition regions is needed for a strictly correct result, in order to have $l_\pm$ and $L_\pm$.

Let us call these integrals r_1, r_2, and r_3. r_2 is easily evaluated and is

$$r_2=4\beta^2\gamma(L_-^4+2L_-^3+L_-^2-l_+^4-2l_+^3-l_+^2). \tag{35}$$

r_1 and r_3 are not difficult to evaluate but are very long. We therefore will not give the explicit forms of the contributions coming from these regions. If one can look at the spin case as a perturbed spinless case so that those regions with widths of 2γ can be ignored, one can take r_2 as an approximation for r even though the exact formula is known. When expressed in terms of the Lagrange multipliers r_2 has the form

$$r_2=4\beta^2\gamma\{[(\tfrac{1}{2}\sqrt{\Delta_L}-\gamma)^2-\tfrac{1}{4}]^2-[(\tfrac{1}{2}\sqrt{\Delta_l}+\gamma)^2-\tfrac{1}{4}]^2\}. \tag{36}$$

We also give the expression for $d\ln A/dt|_{t=0}=(1/A)dA/dt|_{t=0}$ which is obtained from Eq. (23) and the "optical theorem":

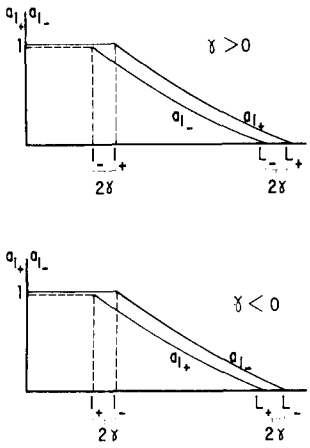


FIG. 1. Three different regions of the imaginary parts a_{l+} and a_{l-} of the partial waves in which they are equal to 1, equal to the expressions given by Eqs. (18) and (19), or zero.

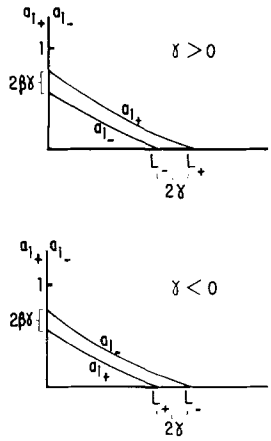


FIG. 2. Case where a_{l+} and a_{l-} start with values less than 1.

$$A(t=0) = [\sqrt{s}k)/4\pi]\sigma^T,$$

$$\frac{d\ln A}{dt}\bigg|_{t=0} = \frac{1}{2k^2}\frac{1}{6\beta} \times \frac{3[(\alpha+4\beta\gamma^2)+4\beta\gamma^2]^2 - 24\beta\gamma^2(1+\beta\gamma^2) - 16\gamma^3\beta^2 - 3\alpha + 1}{2(\alpha+4\beta\gamma^2)+8\gamma(1+2\gamma)-1} \tag{37}$$

Equations (23), (25), (27), (34), (36), and (37) are our results. We observe the following features. For $\gamma = 0$, that is when the third constraint r does not exist, our expressions reduce to

$$\sigma^{\mathrm{T}} = \frac{4\pi}{k^2}\frac{2\alpha-1}{\beta},\ \sigma_{\mathrm{el,im}} = \frac{4\pi}{k^2}\frac{3\alpha-2}{3\beta},\ \text{and}\ \frac{d\ln A}{dt}\bigg|_{t=0} = \frac{1}{2k^2}\frac{1}{6\beta}\frac{3\alpha^2-3\alpha+1}{2\alpha-1}. \tag{38}$$

These equations correspond to spin 0-$\frac{1}{2}$ case with only two constraints, and they are exactly the same relations obtained for the spinless case and lead to the bound

$$\frac{d\ln A}{dt}\bigg|_{t=0} > \tfrac{1}{8}\frac{\sigma^{\mathrm{T}}}{4\pi}\left[1 + 3\left(1 - \frac{\sigma_{\mathrm{el,im}}}{\sigma^{\mathrm{T}}}\right)^2\right]. \tag{39}$$

Thus it is seen that there is no difference in the result for the spinless and spin case when we have only the constraints σ^{T} and $\sigma_{\mathrm{el,im}}$. When we have the third constraint r the elimination of the parameters α, β, and γ is not so easy and calculations have to be made numerically. But in principle one has enough relations to eliminate the Lagrange multipliers and express $d\ln A/dt|_{t=0}$ in terms of σ^{T}, $\sigma_{\mathrm{el,im}}$, and r.

The formulas for the case in which both partial waves start with values less than 1 at $l = 0$ (essentially corresponding to $\alpha + \beta\gamma^2 < 1$) are obtained without difficulty from the basic formulas. The integrals start at $l = 0$ with a_{l+} and a_{l-} given for the entire integration region by the Eqs. (18), (19) or 0 (Fig. 2). In this case, because cancellations of γ terms between the upper and lower limit points do not occur we obtain several terms with γ factors. The equations reduce to the spinless case for $\gamma = 0$. We give below the relevant first terms of the these relations which are followed by terms multiplied with powers of γ.

$$\frac{dA}{dt}\bigg|_{t=0} = \frac{\sqrt{s}}{4k^3}\left(\frac{1}{6}\frac{(\alpha+4\beta\gamma^2)^3}{\beta^2} + \gamma\ \text{terms}\right),$$

$$\sigma^{\mathrm{T}} = \frac{4\pi}{k^2}\left(\tfrac{1}{2}\frac{(\alpha+4\beta\gamma^2)^2}{\beta} + \gamma\ \text{terms}\right),$$

$$\sigma_{\mathrm{el,im}} = \frac{4\pi}{k^2}\left(\frac{1}{3}\frac{\alpha+4\beta\gamma^2)^3}{\beta} + \gamma\ \text{terms}\right), \tag{40}$$

$$r_2 = 4\beta^2\gamma[(\tfrac{1}{2}\sqrt{\Delta_L} - \gamma)^2 - \tfrac{1}{4}]^2.$$

r_2 does not have γ terms because to begin with there were no γ terms coming from the upper and lower limits to cancel.

IV. CONCLUSION

We have considered the variational problem of extremizing $d\ln A/dt|_{t=0}$ with a third constraint in addition to σ^{T} and $\sigma_{\mathrm{el,im}}$. Even though good polarization experiments exist, an angle independent quantity which is given by the imaginary parts of the partial waves only, cannot be obtained from these because the polarization contains cross products of the real and imaginary parts of the partial waves. The quantity defined by a certain integral of the rotation parameter [Eq. (15)] represents a constraint which makes it possible to study the complications introduced both by a third constraint and also by the spin, in the spin 0-$\frac{1}{2}$ case. The existence of two types of partial waves leads to coupled equations for their determination. Unlike the spinless case the imposition of the unitarity defines different regions for the partial waves which greatly complicates the derivations. However, the lower (l_-, l_+) and the upper (L_-, L_+) limits of the partial waves appear always in the right combinations in all quantities except the third constraint, so that they can be replaced by simple combinations of the Lagrange parameters. These parameters can in principle be eliminated from the constraint equations and $d\ln A/dt|_{t=0}$ expressed in terms of the constraints. For $\gamma = 0$ all formulas reduce to the results of the spinless case with two constraints only.

ACKNOWLEDGMENTS

I gratefully acknowledge the hospitality and the support extended to me by the "Universite de Nice" and the "Laboratoire de Physique Théorique" and the director of the Laboratoire, Professor François Rocca. In particular, I would like to thank Professor Le Bellac for making it possible for me to spend my sabbatical leave at his institute.

[1] S. W. MacDowell and A. Martin, Phys. Rev. **135**, B960 (1964).
[2] S. M. Roy and V. Singh, Phys. Rev. Lett. **24**, 28 (1970); Phys. Rev. D **1**, 2638 (1970).
[3] M. B. Einhorn and R. Blankenbecler, Ann. Phys. **67**, 480 (1971).
[4] R. Savit, R. Blankenbecler, and M. B. Einhorn, J. Math. Phys. **12**, 2092 (1971).
[5] G. Auberson, A. Martin, and G. Menessier, Phys. Lett. B **67**, 75 (1977).
[6] G. Auberson and S. M. Roy, Lett. Nuovo Cimento **19**, 97 (1977).
[7] A. D. Gangal and A. S. Vengurlekar, Phys. Rev. D **19**, 1465 (1979).
[8] G. Menessier, S. M. Roy, and V. Singh, Nuovo Cimento A **50**, 443 (1979).

A discussion of the relation "closer than" within simple regions of curved space-time

Gareth Williams
Department of Mathematics, Stetson University, DeLand, Florida 32720

(Received 17 December 1980; accepted for publication 6 March 1981)

A relation "closer than" is introduced for simple regions of curved space-time. This is a total relation on events within a light cone and a partial relation for events on a cone. The group of bijective automorphisms of the relation is the group of bijective homothetic mappings.

PACS numbers: 04.20.Cv, 02.40.Sf

INTRODUCTION

In 1964 Zeeman[1] published a paper showing that the group of automorphisms of Minkowski space that preserved the causal structure was the group generated by the inhomogeneous Lorentz group and dilations. In 1973 Williams[2] related a concept of "closer than" on Minkowski space to Zeeman's causal relations and proved that the group of automorphisms that preserved the relation closer than was that generated by the Poincare group and dilations. In this paper we extend the closer than relation to a simple region of curved space-time of general relativity and examine its group of automorphisms.

DEFINITIONS

Space-time is taken to be a connected, Hausdorff, paracompact, C^∞ real four-dimensional manifold M without boundary, with a C^∞ Lorentz metric and associated pseudo-Riemannian connection. M is taken to be time orientable. For $x \in M$ let us write $I^+(x)$ for the *chronological future* of x, that is, the set of all points in M which can be reached from x by a future directed smooth timelike curve of finite extent. If $y \in I^+(x)$ we write $x < y$. Let $J^+(x)$ denote the *causal future* of x, that is the set of all points in M which can be reached from x by a future directed smooth causal curve (i.e., nonspacelike curve) in M. If $y \in J^+(x)$ we write $x \leqslant y$. The *future horismos* $E^+(x)$ is defined by $E^+(x) = J^+(x) \setminus I^+(x)$. If $y \in E^+(x)$ we write $x \rightarrow y$. The relations $<, \leqslant$, and $\rightarrow$ are respectively called *chronological*, *causal*, and *horismos*. These definitions have their duals with future replaced by past and $+$ by $-$. Let
$E(x) = E^+(x) \cup E^-(x)$ denote the *light cone at* x and
$I(x) = I^+(x) \cup I^-(x)$ denote the interior of the light cone at x.

A subset U of space-time M is called a *simple region* if U has the following properties:

(a) U is an open subset of M (with respect to the manifold topology).

(b) If $x, y \in U$ there is one and only one x and y connecting geodesic curve which we denote $\overline{xy}$. The geodesic $\overline{xy}$ lies entirely within U.

(c) Each geodesic curve $\overline{xy}$ defined by (b) which belongs to U depends continuously on x and y.

(d) The boundary ∂U of U and all closed subsets of U are compact. Let $d(\overline{xy})$ denote the proper length of $\overline{xy}$.

It was shown by Penrose[3] that any space-time can be covered by a locally finite system of simple regions.

Let $x \in U$. All elements of U that lie outside the causal future and past of x are said to be in the *present* of x, denoted $P(x)$.

THE RELATION "CLOSER THAN"

Let us examine the relative closeness to $x \in M$ of various categories of elements of M that lie within a simple region U of M that is a neighborhood of x. We initially consider two categories, $U \cap I(x)$ and $U \cap E(x)$. For convenience we henceforth represent these regions by $I(x)$ and $E(x)$.

Let $a, b \in I(x)$. Generalizing Williams'[2] definition for Minkowski space let us say that "a is closer than b to x if $d(\overline{xa}) < d(\overline{xb})$. We call this relation T and write $x\, ab\, (T)$. Write $x(ab)(T)$ if the distances are equal. This definition is based on the fact that the physical time recorded by a freely-falling observer between the events x and a will be smaller than that recorded by such an observer beween x and b.

Consider $p, q \in E(x)$. We cannot now use proper distance to define comparative closeness to x. However if p and q both lie on the same future (or same past) null geodesic from x we are aware that a concept of closeness to x does exist. The photon from x might reach p before its gets to q and we would then say that "p is closer than q to x". Let us say that p is closer than q to x if p and q lie on the *same* future null geodesic from x and light reaches p first. Similarly, p is closer than q to x if they both lie on the *same* past null geodesic from x and light reached p after q. Write $x\, pq(N)$. Note that under this definition we cannot compare the closeness to x of two events on the same null geodesic if one is in the past of x and the other in the future of x. Neither can we compare the closeness to x of events on distinct null geodescis. There seems to be no physical foundation for comparing the closeness to x of such events. In fact it was seen[2] that even in Minkowski space-time the relation N was a partial relation on the null cone.

AUTOMORPHISMS OF THE RELATIONS

A bijective mapping f of U onto itself is said to be a T-automorphism if and only if for, $a, b \in I(x)$ with $x\, ab\, (T)$, f and f^{-1} preserve T. That is

$$x\, ab\, (T) \leftrightarrow f(x)\, f(a) f(b)\, (T).$$

These are the bijective mappings that preserve the timelike closer than relation. A bijective mapping f of U onto itself is said to be an N-automorphism if for events p and q on a

 0022-2488/81/071443-02$1.00

common null geodesic through x with $x\ pq(N)$

$$x\ pq(N) \leftrightarrow f(x)f(p)f(q)(N).$$

These are the bijective mappings that preserve the null closer than relation.

Lemma 1: A T-automorphism either preserves or reverses chronology.

Proof: Let f be a T-automorphism of U, $x \in U$ and $a,b \in I^+(x)$ with $x\ ab\ (T)$. Thus $f(x)f(a)f(b)(T)$. Suppose f does not preserve or reverse chronology. Let $f(a) < f(x)$ and $f(x) < f(b)$. By transitivity of $<$, $f(a) < f(b)$ and thus $f(b) \in I^+(f(a))$. Since f is a T-automorphism $b \in I(a)$ and further since $x\ ab\ (T)$, $b \in I^+(a)$. However we now arrive at the contradiction that $b\ ax(T)$ but $f(b)\ f(x)f(a)(T)$. Thus f must preserve chronology.

Lemma 2: A T-automorphism preserves timelike geodesics.

Proof: Let $x,b \in U$ with $x < b$. Let f be a T-automorphism. Suppose $f(\overline{xb})$ is not a geodesic; that is, suppose that $f(\overline{xb}) \neq \overline{f(x)f(b)}$. Then there exists $c \in f(\overline{xb})$ such that proper distance between $f(x)$ and c along $f(\overline{xb})$ is equal to $d(\overline{f(x)f(b)})$. We have that $f^{-1}(c) \in \overline{xb}$. But this implies that $xf^{-1}(c)b\ (T)$ while $f(x)\ (cf(b))(T)$, a contradiction. Thus $f(\overline{xb})$ is a geodesic.

Lemma 3: A T-automorphism is a homothetic mapping.

Proof: Let f be a T-automorphism. Consider the *path topology*[4] p on M. This is the finest topology that induces the Euclidean topology on arbitrary timelike curves. Let U_p denote U under the relative p topology. Then since f either preserves or reverses chronology, f is a homeomorphism of U_p.[5] Thus f is a conformal mapping.[4] Furthermore, f preserves timelike geodesics. Therefore f is a homothetic mapping.[6]

Lemma 4: A bijective homothetic mapping of U onto itself is both a T-automorphism and an N-automorphism.

Proof: Homothetic mappings are isometric mappings "up to a scaling factor." These mapping preserve geodesics and geodetic lengths up to a scaling factor. Thus bijective homothetic mappings are T-automorphisms. Furthermore, bijective homothetic mappings, being homeomorphisms of the path topology preserve or reverse causal relations.[4] Consequently such mappings are N-automorphisms.

Combining the results of Lemmas 3 and 4 we now get the following theorem.

Theorem: The group of bijective automorphisms of the relation closer than (i.e., both T and N) on a simple region of curved space-time is the group of bijective homothetic mappings.

FINAL REMARKS

We have in this paper resisted the temptation to define a closer-than-to-x relationship for elements in the present of x. Let $m, n \in P(x)$. In Minkowski space a spacelike closer than relation $x\ mn(\delta)$ has physical meaning in that m and n would be in the physical three-dimensional present of inertial observers at x.[2] However, even though an analogous spacelike relation defined in terms of arc lengths from x along spacelike geodesics would be preserved under homothetic bijective mappings, we do not feel that this relationship has any real physical significance in the curved space-times of general relativity.

[1] E. C. Zeeman, J. Math. Phys. **5**, 490 (1964).
[2] G. Williams, Am. J. Phys. **41**, 871 (1973).
[3] R. Penrose, *Techniques of Differential Topology in Relativity* (SIAM, Philadelphia, 1972), proposition 1.13, p. 6.
[4] S. W. Hawking, A. R. King, and P. J. McCarthy, J. Math Phys. **17**, 174 (1976).
[5] D. B. Malament, J. Math. Phys. **18**, 1399 (1977) (Theorem 2).
[6] R. Göbel, Commun. Math. Phys. **46**, 289 (1976), Prop. 5.8, p. 304.

On the analysis of the gravitational field by reduction of the Weyl tensor to canonical form: Algebraically general case

R. L. Agacy
Department of Mathematics, James Cook University of North Queensland, Queensland, 4811 Australia
L. M. Clendenning
Department of Physics, Humboldt State University, Arcata, California 95521

(Received 27 August 1979; accepted for publication 24 October 1980)

From the Weyl tensor, which can be split up into an "electric" and "magnetic" part of the gravitational field, one can construct a "flux of superenergy." This is analogous to the electromagnetic case where the flux of energy is described by the Poynting vector. In the electromagnetic case, a suitable boost can be made to make the energy flux vanish in a certain frame. By analogy, in this paper we exhibit a variety of methods by which we find a frame in which the flux of superenergy vanishes. These methods are linked together involving spin, Lorentz and complex rotation matrices which are inherently related to a study of the quartic. The algebraically general case only is considered.

PACS numbers: 04.20.Cv,04.20. − q

I. INTRODUCTION

The Weyl tensor with ten real independent components may be represented by two 3×3 real, symmetric, traceless matrices $\mathscr{E}$ and $\mathscr{B}$, the "electric" and "magnetic" parts of the gravitational field.[1,2] The condition that these matrices commute is the necessary and sufficient condition that they can be simultaneously diagonalized[3] which is equivalent physically to being in a frame in which the flux of superenergy, as can be seen from the appropriate components of the Bel–Robinson tensor, vanishes. The question posed in Ref. 1 is that given a frame in which the superenergy flux does not vanish, what method or methods may be employed to find a frame in which it does. This is in analogy to the electromagnetic case where a suitable boost may be made to make the electric and magnetic vectors **E** and **B** parallel in that frame. The problem may also be viewed as reducing the Weyl tensor or Weyl spinor to canonical form. The Weyl spinor being linked to a quartic expression, the equivalent problem is a reduction of the quartic to canonical form. The algebraically general case is considered here.

The contents of Secs. II and III are familiar although we use the complex electromagnetic vector and by analogy a "complex gravitational field" $\mathscr{F}$ comprising the "electric" and "magnetic" parts of the Weyl tensor. Section IV is a straightforward case of matrix diagonalization of $\mathscr{F}$.[4] A study of aspects of the quartic is undertaken in Sec. V. The null rotation method in Sec. VI successively reduces the Weyl spinor or associated quartic to canonical form leading to an equation which is one of the concomitants of the quartic. A complex Euler angle formalism is set up in Sec. VII enabling the actual boosts and spatial rotations necessary for diagonalization of $\mathscr{F}$ (the finding of a frame in which the superenergy flux vanishes) to be determined .

II. MAXWELL SPINOR

We use the notations, letters, and range of indices as in Ref. 5. The metric $g_{\mu\nu}=\text{diag}\,(1,-1,-1,-1)$, tensor indices run from 0, − , − ,3; spinor indices take on values 0,1. The Maxwell tensor is taken to be[6]

$$F_{\mu\nu}=\begin{bmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{bmatrix}.$$

We select the set of Pauli matrices

$$\sigma^0{}_{AB'}=\frac{1}{\sqrt{2}}\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \sigma^1{}_{AB'}=\frac{1}{\sqrt{2}}\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

$$\sigma^2{}_{AB'}=\frac{1}{\sqrt{2}}\begin{pmatrix}0&i\\-i&0\end{pmatrix},\quad \sigma^3{}_{AB'}=\frac{1}{\sqrt{2}}\begin{pmatrix}1&0\\0&-1\end{pmatrix},$$

where the first subscript labels rows. The spinor equivalent $F_{AB'CD'}$ of the Maxwell tensor is given by

$$F_{AB'CD'}=\sigma^{\mu}_{AB'}\sigma^{\nu}_{CD'}F_{\mu\nu}.$$

The symmetric Maxwell spinor ϕ_{AB} can then be obtained from

$$F_{AB'CD'}=\epsilon_{AC}\bar{\phi}_{B'D'}+\phi_{AC}\epsilon_{B'D'} \tag{II.1}$$

and we find

$$\phi_{AB}=\begin{pmatrix}\phi_{00}&\phi_{01}\\\phi_{10}&\phi_{11}\end{pmatrix}=\tfrac{1}{2}\begin{pmatrix}F_x+iF_y&-F_z\\-F_z&-F_x+iF_y\end{pmatrix}$$
$$=\begin{pmatrix}\phi_0&\phi_1\\\phi_1&\phi_2\end{pmatrix}, \tag{II.2}$$

where $F_x=E_x+iB_x, F_y=E_y+iB_y$, $F_z=E_z+iB_z$ is the complex electromagnetic vector.

We may note that the eigenvalues of ϕ_{AB} are $\pm(F_x^2+F_y^2+F_z^2)^{1/2}$ and when this vanishes we have the special or null case (equivalently the condition $\mathbf{E}^2-\mathbf{B}^2=0$ and $\mathbf{E}\cdot\mathbf{B}=0$).

If one considers a unimodular spin transformation of an orthonormal dyad $\{\ell_A, n_A\}$, $\ell_A n^A=1$

$$\begin{pmatrix}l'_A\\n'_A\end{pmatrix}=\begin{pmatrix}a&b\\c&d\end{pmatrix}\begin{pmatrix}l_A\\n_A\end{pmatrix},\quad \text{where } ad-bc=1, \tag{II.3}$$

the spinor equivalent of the Maxwell tensor in a new

 0022-2488/81/071445-07$1.00

"primed" frame is related to its components in the original frame by

$$F'_{AB'CD'} = S_A{}^E \overline{S}_{B'}{}^{F'} S_C{}^G \overline{S}_{D'}{}^{H'} F_{EF'GH'}, \tag{II.4}$$

where $S_A{}^E = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

is the spin transformation, the lower subscript indicating rows.

Similar relations to (II.1) and (II.2) hold in the new frame of reference (with primes on the variables) so that by combining these with (II.2) in (II.4) we can relate the components of the complex electromagnetic vector in the transformed, primed frame to its components in the original frame We find

$$\begin{pmatrix} F'_x \\ F'_y \\ F'_z \end{pmatrix} = \begin{bmatrix} \tfrac{1}{2}(a^2 - b^2 - c^2 + d^2) & \frac{i}{2}(a^2 + b^2 - c^2 - d^2) & -ab + cd \\ -\frac{i}{2}(a^2 - b^2 + c^2 - d^2) & \tfrac{1}{2}(a^2 + b^2 + c^2 + d^2) & i(ab + cd) \\ -ac + bd, & -i(ac + bd) & ad + bc \end{bmatrix} \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} \tag{II.5}$$

or briefly $F' = PF$ where the transformation matrix P is a *complex orthogonal* matrix with unit determinant.

An induced basis in a complex E_3

$$\eta_{0AB} = (i/\sqrt{2})(o_A o_B - \iota_A \iota_B)$$

$$\eta_{1AB} = -(1/\sqrt{2})(\iota_A \iota_B + o_A o_B)$$

$$\eta_{2AB} = (i/\sqrt{2})(\iota_A o_B + o_A \iota_B)$$

which is orthonormal, $\eta_{mAB}\eta_n^{AB} = \delta_{mn}$, allows the Maxwell spinor to be expressed as

$$\phi_{AB} = -(i/\sqrt{2})(F_x \eta_{0AB} + F_y \eta_{1AB} + F_z \eta_{2AB}),$$

where we have chosen $\iota_A = (0,1)$, $o_A = (-1,0)$.

III. WEYL SPINOR

The Weyl tensor in source-free space may be written in the form[7]

$$C^{\alpha\beta}{}_{\lambda\delta} = \begin{pmatrix} -\mathscr{E} & \mathscr{B} \\ -\mathscr{B} & -\mathscr{E} \end{pmatrix},$$

the rows being labelled by $\alpha\beta$, the columns by $\lambda\delta$ with numbering 01, 02, 03, 23, 31, 12 successively down and across,

$$\mathscr{E} = \begin{pmatrix} \mathscr{E}_{xx} & \mathscr{E}_{xy} & \mathscr{E}_{xz} \\ \mathscr{E}_{yx} & \mathscr{E}_{yy} & \mathscr{E}_{yz} \\ \mathscr{E}_{zx} & \mathscr{E}_{zy} & \mathscr{E}_{zz} \end{pmatrix} = \mathscr{E}^T, \quad \mathscr{E}_{xx} + \mathscr{E}_{yy} + \mathscr{E}_{zz} = 0$$

with a similar matrix for $\mathscr{B}$.

We interpret $\mathscr{E}$ and $\mathscr{B}$ as the "electric" and "magnetic" parts of the gravitational field.

The spinor equivalent of the Weyl tensor

$$C_{AB'CD'EF'GH'} = \sigma^{\alpha}{}_{AB'}\sigma^{\beta}{}_{CD'}\sigma^{\gamma}{}_{EF'}\sigma^{\delta}{}_{GH'} C_{\alpha\beta\gamma\delta}$$

can then be calculated. We also have, to determine the totally symmetric Weyl spinor Ψ_{ABCD}, the relation

$$C_{AB'CD'EF'GH'} = -(\epsilon_{AC}\epsilon_{EG}\overline{\Psi}_{B'D'F'H'} + \Psi_{ACEG}\epsilon_{B'D'}\epsilon_{F'H'}).$$

Putting $\Psi_{ABCD} = \psi_{A+B+C+D}$, with Weyl spinor components now denoted $\psi_0, \psi_1, \cdots, \psi_4$, we can form the 3×3 trace-free symmetric matrix

$$\mathscr{F} = \begin{pmatrix} \mathscr{F}_{xx} & \mathscr{F}_{xy} & \mathscr{F}_{xz} \\ \mathscr{F}_{yx} & \mathscr{F}_{yy} & \mathscr{F}_{yz} \\ \mathscr{F}_{zx} & \mathscr{F}_{zy} & \mathscr{F}_{zz} \end{pmatrix} = \Psi$$

$$= \begin{pmatrix} \tfrac{1}{2}(2\psi_2 - \psi_0 - \psi_4) & \tfrac{1}{2}i(\psi_0 - \psi_4) & (\psi_1 - \psi_3) \\ \tfrac{1}{2}i(\psi_0 - \psi_4) & \tfrac{1}{2}(2\psi_2 + \psi_0 + \psi_4) & -i(\psi_1 + \psi_3) \\ (\psi_1 - \psi_3) & -i(\psi_1 + \psi_3) & -2\psi_2 \end{pmatrix}$$

$$= \begin{pmatrix} \psi_{00} & \psi_{01} & \psi_{02} \\ \psi_{10} & \psi_{11} & \psi_{12} \\ \psi_{20} & \psi_{21} & \psi_{22} \end{pmatrix}, \quad \psi_{mn} = \psi_{nm}, \tag{III.1}$$

where $\mathscr{F}_{xx} = \mathscr{E}_{xx} + i\mathscr{B}_{xx}$ etc., and $\mathrm{Tr}\mathscr{F} = 0$.

In terms of the η basis in Sec. II, the Weyl spinor may also be expressed as

$$\Psi_{ABCD} = \sum_{m,n=0}^{4} \Psi_{mn}\eta_{mAB}\eta_{nCD}$$

which has allowed us to form the matrix Ψ_{mn}. We may refer to $\mathscr{F}$ (in analogy with the electromagnetic case) as the "complex gravitational field."

The eigenvalue equation for the Weyl spinor can be expressed as

$$\Psi_{ABCD}\phi^{CD} = \lambda\phi_{AB}$$

or more strikingly in matrix terms as

$$\begin{pmatrix} \mathscr{F}_{xx} & \mathscr{F}_{xy} & \mathscr{F}_{xz} \\ \mathscr{F}_{yx} & \mathscr{F}_{yy} & \mathscr{F}_{yz} \\ \mathscr{F}_{zx} & \mathscr{F}_{zy} & \mathscr{F}_{zz} \end{pmatrix} \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \lambda \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix}, \tag{III.2}$$

where we might wish to interpret in a "formal algebraic sense", ϕ and F as in Sec. II.

Upon performing the spin transformation $S_A{}^E$ of (II.3), new bases η'_{mAB} are induced in E_3 with new, primed, Weyl spinor components $\psi'_0, \ldots \psi'_4$, and new primed matrices $\mathscr{F}' = \Psi'$.

The law of transformation to the primed system is

$$\mathscr{F}' = P\mathscr{F}P^T, (P^T = P^{-1}),$$

where P is the complex orthogonal matrix of (II.5).

IV. DIAGONALIZATION: MATRIX METHOD

The problem posed in Ref. 1 was the "simultaneous diagonalization" of the $\mathscr{E}$ and $\mathscr{B}$ parts of the Weyl tensor. This can be achieved if the matrix $\mathscr{F} = \mathscr{E} + i\mathscr{B}$ is diagona-

lized. That is to say it is the determination of the orthogonal matrix P such that $\mathscr{F}' = P\mathscr{F}P^T$ is diagonal.

In the algebraically general case the eigenvalues of $\mathscr{F} = \mathscr{E} + i\mathscr{B}$ are *distinct* so that $\mathscr{F}$ can be diagonalized to a new matrix $\mathscr{F}'$.

The characteristic equation of $\mathscr{F}(=\Psi)$ is a cubic giving the three distinct eigenvalues and we can then calculate the corresponding eigenvectors which form the rows of the matrix P. We see from the matrix P of (II.5) that we can then obtain the spin transformation $S_A{}^E = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to within a sign. Conversely, if we know the spin transformation, we can obtain the matrix P. A corresponding Lorentz transformation relating the Weyl tensor to its transform in the new, primed, frame can then be calculated (from Ref. 5, p. 53).

V. THE QUARTIC

The Weyl spinor components $(\psi_0, \psi_1, \psi_2, \psi_3, \psi_4)$ may be associated with the quartic

$$u = \psi_0 z^4 + 4\psi_1 z^3 + 6\psi_2 z^2 + 4\psi_3 z + \psi_4 = 0, \qquad \text{(V.1)}$$

whose roots give the principal null directions. There are only two independent invariants of the quartic[8,9]:

$$I = \psi_0\psi_4 - 4\psi_1\psi_3 + 3\psi_2{}^2 = \tfrac{1}{2}\Psi_{ABCD}\Psi^{ABCD}$$

and

$$J = \psi_0\psi_2\psi_4 + 2\psi_1\psi_2\psi_3 - \psi_0\psi_3{}^2 - \psi_1{}^2\psi_4 - \psi_2{}^3 = \tfrac{1}{6}\Psi^{AB}{}_{CD}\Psi^{CD}{}_{EF}\Psi^{EF}{}_{AB}.$$

There are two covariants of the quartic, the Hessian H and a sextic covariant G, these concomitants being connected by the syzygy

$$Iu^2H - 4H^3 - Ju^3 = G^2.$$

If $\alpha, \beta, \gamma, \delta$ are the roots of the quartic, the discriminant Δ of the quartic is given by

$$\psi_0{}^6(\beta-\gamma)^2(\gamma-\alpha)^2(\alpha-\beta)^2(\alpha-\delta)^2(\beta-\delta)^2(\gamma-\delta)^2 = 256\,\Delta,$$

where $\Delta = I^3 - 27J^2$.

Clearly then the vanishing of the discriminant is the necessary and sufficient condition that two roots be equal i.e., two principal directions coincide. Thus when $I^3 = 27J^2$ algebraically special cases arise as noted in Ref. 9.

The characteristic equation $|\Psi - \lambda\,1| = 0$ (where Ψ is the middle matrix of (III.1) and 1 is the unit matrix) when expanded out is

$$\lambda^3 - \lambda I - 2J = 0, \qquad \text{(V.2)}$$

where I and J are the two invariants of the quartic above. Thus we see the link between the eigenvalue or matrix method and the consideration of the quartic. In fact the cubic is the "reducing cubic" (or cubic resolvent) of the quartic! This fact does not appear to have been noticed in the literature! Knowing the roots of the cubic (V.2) i.e., by the eigenvalue method, we can obtain the roots of the quartic and vice versa.

We have the following expressions for the invariants I and J[10]: $I = -\Sigma\, 2\times 2$ Principal minors of

$$\Psi = \tfrac{1}{2}\mathrm{Tr}\mathscr{F}^2 = \tfrac{1}{2}\mathrm{Tr}[\mathscr{E}^2 - \mathscr{B}^2 + 2i\,\mathscr{E}\mathscr{B}],$$

$$J = \tfrac{1}{2}\det\mathscr{F} = \tfrac{1}{6}\mathrm{Tr}\mathscr{F}^3 = \tfrac{1}{6}\mathrm{Tr}[\mathscr{E}^3 - 3\mathscr{E}\mathscr{B}^2 - i(\mathscr{B}^3 - 3\mathscr{B}\mathscr{E}^2)].$$

Setting $\lambda = -2\psi_0\Theta$ the cubic (V.2) is[11]

$$4\psi_0{}^3\Theta^3 - I\psi_0\Theta + J = 0.$$

If the roots of this reducing cubic are $\Theta_1, \Theta_2, \Theta_3$, they are related to the roots of the quartic (Ref. 11, p. 122) by

$$\begin{aligned} 4(\Theta_2 - \Theta_3) &= -(\beta-\gamma)(\alpha-\delta), \\ 4(\Theta_3 - \Theta_1) &= -(\gamma-\alpha)(\beta-\delta), \\ 4(\Theta_1 - \Theta_2) &= -(\alpha-\beta)(\gamma-\delta), \end{aligned} \qquad \text{(V.3)}$$

with $\Theta_1 + \Theta_2 + \Theta_3 = 0$—this condition being equivalent to $\mathrm{Tr}\Psi = \mathrm{Tr}\mathscr{F} = 0$. With $\lambda_i = -2\psi_0\Theta_i$, $i = 1, 2, 3$, we then have for the eignevalues of λ_i in terms of the roots of the quartic (Ref. 11, p. 122)

$$\begin{aligned} \lambda_1 &= \tfrac{1}{6}[(\alpha-\beta)(\gamma-\delta) - (\gamma-\alpha)(\beta-\delta)]\psi_0, \\ \lambda_2 &= \tfrac{1}{6}[(\beta-\gamma)(\alpha-\delta) - (\alpha-\beta)(\gamma-\delta)]\psi_0, \\ \lambda_3 &= \tfrac{1}{6}[(\gamma-\alpha)(\beta-\delta) - (\beta-\gamma)(\alpha-\delta)]\psi_0. \end{aligned} \qquad \text{(V.4)}$$

Let[12]

$$\mu = \frac{(\alpha-\beta)(\gamma-\delta)}{(\alpha-\delta)(\gamma-\beta)} = \frac{(\alpha_A\beta^A)(\gamma_B\delta^B)}{(\alpha_C\delta^C)(\gamma_D\beta^D)}$$

(setting $\alpha = \alpha_0/\alpha_1$ etc.)

be the cross ratio corresponding to the roots $\alpha, \beta, \gamma, \delta$ of the quartic (V.1) and symbolically represented by $\{\alpha, \beta, \gamma, \delta\}$.

Then

$$\{\alpha,\beta,\gamma,\delta\} = \{\beta,\alpha,\delta,\gamma\} = \{\gamma,\delta,\alpha,\beta\} = \{\delta,\gamma,\beta,\alpha\},$$

and it can be seen that the eigenvalues are unchanged under any of the above permutations.

There are six (out of 24) distinct cross ratios in general and these correspond to the six permutations or ways of ordering the eigenvalues or six ways of reducing the quartic to canonical form. If we fix the root α of the quartic in any chosen manner then there is a 1–1 correspondence between the remaining roots (principal null directions) and the eigenvalues.

With an association $\beta, \gamma, \delta \to \lambda_1, \lambda_2, \lambda_3$ we may set up the following table of correspondences using (V.3):

$$1\,2\,3 \to \{\alpha,\beta,\gamma,\delta\} \to \mu = \frac{(\alpha-\beta)(\gamma-\delta)}{(\alpha-\delta)(\gamma-\beta)} = \frac{\lambda_1-\lambda_2}{\lambda_3-\lambda_2}$$
$$z_1{}^{(1)} \quad S_A{}^{E(1)} \quad P^{(1)} \quad L^{(1)},$$

$$2\,3\,1 \to \{\alpha,\gamma,\delta,\beta\} \to \frac{\mu-1}{\mu} = \frac{(\alpha-\gamma)(\delta-\beta)}{(\alpha-\beta)(\delta-\gamma)} = \frac{\lambda_3-\lambda_1}{\lambda_2-\lambda_1}$$
$$z_1{}^{(2)} \quad S_A{}^{E(2)} \quad P^{(2)} \quad L^{(2)},$$

$$3\,1\,2 \to \{\alpha,\delta,\beta,\gamma\} \to \frac{1}{1-\mu} = \frac{(\alpha-\delta)(\beta-\gamma)}{(\alpha-\gamma)(\beta-\delta)} = \frac{\lambda_2-\lambda_3}{\lambda_1-\lambda_3}$$
$$z_1{}^{(3)} \quad S_A{}^{E(3)} \quad P^{(3)} \quad L^{(3)},$$

$$2\,1\,3 \to \{\alpha,\gamma,\beta,\delta\} \to 1-\mu = \frac{(\alpha-\gamma)(\beta-\delta)}{(\alpha-\delta)(\beta-\gamma)} = \frac{\lambda_1-\lambda_3}{\lambda_2-\lambda_3}$$
$$z_1{}^{(4)} \quad S_A{}^{E(4)} \quad P^{(4)} \quad L^{(4)}, \qquad \text{(V.5)}$$

$$1\,3\,2 \rightarrow \{\alpha,\beta,\delta,\gamma\} \rightarrow \frac{\mu}{\mu-1} = \frac{(\alpha-\beta)(\delta-\gamma)}{(\alpha-\gamma)(\delta-\beta)} = \frac{\lambda_2-\lambda_1}{\lambda_3-\lambda_1}$$
$$z_1^{(5)} \quad S_A{}^{E(5)} \quad P^{(5)} \; L^{(5)},$$

$$3\,2\,1 \rightarrow \{\alpha,\delta,\gamma,\beta\} \rightarrow \frac{1}{\mu}\frac{(\alpha-\delta)(\gamma-\beta)}{(\alpha-\beta)(\gamma-\delta)} = \frac{\lambda_3-\lambda_2}{\lambda_1-\lambda_2}$$
$$z_1^{(6)} \quad S_A{}^{E(6)} \quad P^{(6)} \quad L^{(6)}.$$

It may be remarked that for distinct roots of the quartic no complex cross ratio has the value 0, 1 or ∞ which will obviously be the case in the algebraically special cases.

We may solve the quartic (V.1) and the reducing cubic (V.2) i.e., the matrix eigenvalue problem and confirm the relations (V.4) and (V.5).

VI. DIAGONALIZATION: NULL ROTATIONS

A canonical form for the Weyl spinor will have components $(\psi_0',0,\psi_2',0,\psi_0')$, where it can be seen from (III.1) that in a primed frame, the matrix $\Psi' = \mathcal{F}'$ becomes diagonal.

The idea, then, for diagonalization is to transform to a primed frame such that in this frame

$$\psi_1' = \psi_3' = 0 \text{ and } \psi_0' = \psi_4'.$$

The unimodular spin transformation $S_A{}^E = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ can be considered as a product of three spin transformations

$$g_1(z_1) = \begin{pmatrix} 1 & 0 \\ z_1 & 1 \end{pmatrix}, \; g_2(z_2) = \begin{pmatrix} z_2 & 0 \\ 0 & z_2^{-1} \end{pmatrix}, \; g_3(z_3) = \begin{pmatrix} 1 & z_3 \\ 0 & 1 \end{pmatrix}$$

the first and third representing null rotations around ℓ_A and n_A respectively (see II.3); and the second a duality rotation.

Making a null rotation around ℓ_A, we have (see Ref. 5, p. 183)

$$(\psi_0,\ldots,\psi_4) \rightarrow ('\psi_0,\ldots,'\psi_4),$$

where $'\psi_{n-1} = \frac{1}{n}\frac{d}{dz_1}'\psi_n$,

$$'\psi_4 = z_1^4\psi_0 + 4z_1^3\psi_1 + 6z_1^2\psi_2 + 4z_1\psi_3 + \psi_4, \; n = 1,\ldots,4. \tag{VI.1}$$

Making a null rotation around $'n_A$, we have

$$('\psi_0,\ldots,'\psi_4) \rightarrow (''\psi_0,\ldots,''\psi_4),$$

where

$$''\psi_{n+1} = \frac{1}{4-n}\frac{d}{dz_3}''\psi_n,$$
$$''\psi_0 = '\psi_0 + 4z_3{}'\psi_1 + 6z_3{}^{2'}\psi_2 + 4z_3{}^{3'}\psi_3 + z_3{}^{4''}\psi_4, \; n = 0.,,,.3. \tag{VI.2}$$

Making a duality rotation, we have $(''\psi_0,\ldots,''\psi_4,) \rightarrow (\psi_0',\ldots,\psi_4')$ (final primed components), where

$$\psi_n' = z_2{}^{4-2n}\,''\psi_n, \; n = 0,\ldots,4.$$

For $\psi_0' = \psi_4'$, this leads to

$$z_2 = (''\psi_4/''\psi_0)^{1/8}$$

giving z_2 when $''\psi_0$, $''\psi_4$ are known.

For $\psi_3' = 0$ which implies $''\psi_3 = 0$, we have from (VI.2) when $n = 2$

$$z_3 = -\,'\psi_3/'\psi_4$$

determining z_3 when $'\psi_3$, $'\psi_4$ are known.

Using this value of z_3 in (VI.2) when $n = 0$ and equating to zero ($''\psi_1 = 0$ is implied by $\psi_1' = 0$) leads to

$$'\psi_1{}'\psi_4{}^2 - 3'\psi_2{}'\psi_3{}'\psi_4 + 2'\psi_3{}^3 = 0.$$

Each of the $'\psi_n$ are expressible in terms of the original Weyl spinor components and the variable z_1 by (VI.1) so that making the substitutions leads to a sextic equation to determine z_1, and this sextic is precisely the sextic covariant G of the quartic (see Ref. 11, p. 372)

$$G = A_0z_1{}^6 + A_1z_1{}^5 + A_2z_1{}^4 + A_3z_1{}^3 + A_4z_1{}^2 + A_5z_1 + A_6 = 0,$$

where

$$\begin{aligned} A_0 &= \psi_0{}^2\psi_3 - 3\psi_0\psi_1\psi_2 + 2\psi_1{}^3, \\ A_1 &= \psi_0{}^2\psi_4 + 2\psi_0\psi_1\psi_3 - 9\psi_0\psi_2{}^2 + 6\psi_1{}^2\psi_2, \\ A_2 &= 5(\psi_0\psi_1\psi_4 - 3\psi_0\psi_2\psi_3 + 2\psi_1{}^2\psi_3), \\ A_3 &= -10(\psi_0\psi_3{}^2 - \psi_1{}^2\psi_4), \\ A_4 &= -5(\psi_0\psi_3\psi_4 - 3\psi_1\psi_2\psi_4 + 2\psi_1\psi_3{}^2), \\ A_5 &= -\psi_0\psi_4{}^2 - 2\psi_1\psi_3\psi_4 + 9\psi_2{}^2\psi_4 - 6\psi_2\psi_3{}^2, \\ A_6 &= -\psi_1\psi_4{}^2 + 3\psi_2\psi_3\psi_4 - 2\psi_3{}^3. \end{aligned} \tag{VI.3}$$

Solve this sextic and obtain the six roots. For a root z_1 of the sextic and knowing the orginal Weyl spinor components $(\psi_0, \psi_1, \psi_2, \psi_3, \psi_4)$ we can calculate in turn $('\psi_0,\ldots,'\psi_4)$, z_3, $(''\psi_0,\ldots,''\psi_4)$, z_2, and finally $(\psi_0',\ldots,\psi_4')$ with $(\psi_0' = \psi_4'$, $\psi_1' = \psi_3' = 0)$ and so we are assured that $\mathcal{F}' = \Psi'$ is diagonal.

Knowing then z_1, z_2, z_3, we can also calculate the spin transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = g_2g_3g_1 = \begin{pmatrix} z_2(1+z_3z_1) & z_2z_3 \\ z_1/z_2 & 1/z_2 \end{pmatrix}.$$

With the (determined) spin transformation matrices $g_1(z_1)$, $g_2(z_2)$, $g_3(z_3)$ we may construct corresponding complex orthogonal matrices $P_1(z_1)$, $P_2(z_2)$, $P_3(z_3)$ via (II.5):

$$P_1(z_1) = \begin{pmatrix} 1 - z_1{}^2/2 & -iz_1{}^2/2 & z_1 \\ -iz_1{}^2/2 & 1 + z_1{}^2/2 & iz_1 \\ -z_1 & -iz_1 & 1 \end{pmatrix},$$

$$P_2(z_2) = \begin{pmatrix} \frac{1}{2}(z_2{}^2 + z_2{}^{-2}) & \frac{1}{2}i(z_2{}^2 - z_2{}^{-2}) & 0 \\ -\frac{1}{2}i(z_2{}^2 - z_2{}^{-2}) & \frac{1}{2}(z_2{}^2 + z_2{}^{-2}) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$P_3(z_3) = \begin{pmatrix} 1 - z_3{}^2/2 & iz_3{}^2/2 & -z_3 \\ iz_3{}^2/2 & 1 + z_3{}^2/2 & iz_3 \\ z_3 & -iz_3 & 1 \end{pmatrix}.$$

The diagonalization may be viewed as a succession of orthogonal transformations on the original matrix $\mathcal{F}$,

$$'\mathcal{F} = P_1 \; \mathcal{F} \; P_1{}^T,$$
$$''\mathcal{F} = P_3 \; '\mathcal{F} \; P_3{}^T,$$
$$\mathcal{F}' = P_2 \; ''\mathcal{F} \; P_2{}^T,$$

or $\mathcal{F}' = P\mathcal{F}P^T$, where $P = P_2P_3P_1$.

In the diagonalization of a 3×3 matrix with distinct eigenvalues there are six ways of ordering these eigenvalues on the diagonal. These correspond to the six roots of the sextic which lead to six spin transformation matrices of the type g_i (and hence their product) six complex orthogonal matrices of the type P_i, six Lorentz transformations $L^{(i)}$ [Table (V.5)] and is further linked with the six essentially different cross ratios which are formed from the four roots of the quartic.

VII. DIAGONALIZATION: COMPLEX EULER ANGLES

Another method of reduction of the Weyl spinor is the successive determination of complex Euler angles to reduce the Weyl spinor to canonical form.

With respect to the $0xyz$ axes we consider a "boost in a given direction *together with* a rotation about that axis." These operations are commutative so that the order is immaterial. We refer to both operations as a "complex rotation with regard to an axis." Using the Euler angle convention of Ref. 13 we consider (i) a rotation with parameter ϕ_1 and boost with parameter ϕ_2 with respect to the z axis (a complex rotation with regard to the z axis), (ii) a rotation with parameter θ_1 and boost with parameter θ_2 with respect to the new (rotated) x axis, and (iii) a rotation with parameter ψ_1 and boost with parameter ψ_2 with respect to the (final) z axis.

The spin, the Lorentz, and complex rotation matrices P corresponding to (i), (ii), (iii) are respectively given by

$$D_\phi = \pm\begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2}\end{pmatrix}$$

$$\to\begin{pmatrix} \cosh\phi_2 & 0 & 0 & \sinh\phi_2 \\ 0 & \cos\phi_1 & \sin\phi_1 & 0 \\ 0 & -\sin\phi_1 & \cos\phi_1 & 0 \\ \sinh\phi_2 & 0 & 0 & \cosh\phi_2\end{pmatrix}$$

$$\to P_\phi = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1\end{pmatrix}, \quad \text{where } \phi = \phi_1 + i\phi_2,$$

$$C_\theta = \pm\begin{pmatrix} \cos\theta/2 & -i\sin\theta/2 \\ -i\sin\theta/2 & \cos\theta/2\end{pmatrix}$$

$$\to\begin{pmatrix} \cosh\theta_2 & \sinh\theta_2 & 0 & 0 \\ \sinh\theta_2 & \cosh\theta_2 & 0 & 0 \\ 0 & 0 & \cos\theta_1 & \sin\theta_1 \\ 0 & 0 & -\sin\theta_1 & \cos\theta_1\end{pmatrix}$$

$$\to P_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta\end{pmatrix}, \quad \text{where } \theta = \theta_1 + i\theta_2,$$

$$B_\psi = \pm\begin{pmatrix} e^{-i\psi/2} & 0 \\ 0 & e^{i\psi/2}\end{pmatrix}$$

$$\to\begin{pmatrix} \cosh\psi_2 & 0 & 0 & \sinh\psi_2 \\ 0 & \cos\psi_1 & \sin\psi_1 & 0 \\ 0 & -\sin\psi_1 & \cos\psi_1 & 0 \\ \sinh\psi_2 & 0 & 0 & \cosh\psi_2\end{pmatrix}$$

$$\to P_\psi = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1\end{pmatrix}, \quad \text{where } \psi = \psi_1 + i\psi_2.$$

The product of the spin transformation matrices B_ψ, C_θ, D_ϕ is

$$Q = B_\psi C_\theta D_\phi = \pm\begin{pmatrix} e^{-i(\psi+\phi)/2}\cos\theta/2 & -ie^{-i(\psi-\phi)/2}\sin\theta/2 \\ -ie^{i(\psi-\phi)/2}\sin\theta/2 & e^{i(\psi+\phi)/2}\cos\theta/2\end{pmatrix} = \begin{pmatrix} a & b \\ c & d\end{pmatrix}$$

and the product of the complex rotation matrices P_ψ, P_θ, P_ϕ is

$$P = P_\psi P_\theta P_\phi = \begin{pmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta\end{pmatrix}. \tag{VII.1}$$

It is emphasized that the "angles" ϕ, θ, ψ are complex and we may assume the convention $0\leqslant\phi_1\leqslant 2\pi$, $0\leqslant\theta_1\leqslant\pi$, $0\leqslant\psi_1\leqslant 2\pi$. It will be observed that if the spin transformation is determined by any method then the "complex Euler angles" are determined and vice versa.

We now set out the successive reduction of the Weyl spinor to canonical form by determining the complex Euler angles—this gives the boosts and rotations.

From Eq. (8.69) of Ref. 5 the original components $(\psi_0, \psi_1,...,\psi_4)$ change under the spin transformation D_ϕ to first intermediate components $('\psi_0, '\psi_1,...,'\psi_4)$ given by

$$'\psi_n = e^{i(n-2)\phi}\psi_n \quad (n = 0, 1, ...,4). \tag{VII.2}$$

Under the spin transformation C_θ these first intermediate components change to second intermediate components $(''\psi_0, ''\psi_1,...,''\psi_4)$ given by

$$\begin{pmatrix} {}''\psi_0 \\ {}''\psi_1 \\ {}''\psi_2 \\ {}''\psi_3 \\ {}''\psi_4 \end{pmatrix} = \begin{pmatrix} \tfrac{1}{4}(1+\cos\theta)^2 & -i\sin\theta(1+\cos\theta) & -\tfrac{3}{2}\sin^2\theta & i\sin\theta(1-\cos\theta) & \tfrac{1}{4}(1-\cos\theta)^2 \\ -\tfrac{1}{4}i\sin\theta(1+\cos\theta) & -\tfrac{1}{2}(2\cos^2\theta+\cos\theta-1) & -\tfrac{3}{2}i\sin\theta\cos\theta & \tfrac{1}{2}(2\cos^2\theta-\cos\theta-1) & \tfrac{1}{4}i\sin\theta(1-\cos\theta) \\ -\tfrac{1}{4}\sin^2\theta & -i\sin\theta\cos\theta & -\tfrac{1}{2}+\tfrac{3}{2}\cos^2\theta & -i\sin\theta\cos\theta & -\tfrac{1}{4}\sin^2\theta \\ +\tfrac{1}{4}i\sin\theta(1-\cos\theta) & \tfrac{1}{2}(2\cos^2\theta-\cos\theta-1) & -\tfrac{3}{2}i\sin\theta\cos\theta & \tfrac{1}{2}(2\cos^2\theta+\cos\theta-1) & \tfrac{1}{4}i\sin\theta(1+\cos\theta) \\ \tfrac{1}{4}(1-\cos\theta)^2 & i\sin\theta(1-\cos\theta) & -\tfrac{3}{2}\sin^2\theta & -i\sin\theta(1+\cos\theta) & \tfrac{1}{4}(1+\cos\theta)^2 \end{pmatrix} \times \begin{pmatrix} {}'\psi_0 \\ {}'\psi_1 \\ {}'\psi_2 \\ {}'\psi_3 \\ {}'\psi_4 \end{pmatrix} \tag{VII.3}$$

Under the spin transformation B_ψ these second intermediate components change to the final primed components $(\psi_0', \psi_1', ..., \psi_4')$ given by

$$\psi_n' = e^{i(n-2)\psi}\,{}''\psi_n \quad (n = 0, 1, ..., 4). \tag{VII.4}$$

In a similar manner as used in Sec. VI we want

$$\psi_1' = \psi_3' = 0 \text{ and } \psi_0' = \psi_4'.$$

Working backwards we find from (VII.4) that $\psi_0' = \psi_4'$ leads to

$$e^{4i\psi} = {}''\psi_4/{}''\psi_0$$

determining the complex angle ψ when ${}''\psi_0$, ${}''\psi_4$ are known. The conditions $\psi_1' = \psi_3' = 0$ imply ${}''\psi_1 = {}''\psi_3 = 0$ from (VII.4). Adding and substracting the two equations obtained by putting ${}''\psi_1 = {}''\psi_3 = 0$ in (VII.3) leads to

$$\tan\theta = 2i({}'\psi_3 - {}'\psi_1)/({}'\psi_0 - {}'\psi_4)$$

and

$$\tan 2\theta = -4i({}'\psi_1 + {}'\psi_3)/({}'\psi_0 + 6{}'\psi_2 + {}'\psi_4).$$

Either of these equations determines θ when $({}'\psi_0, {}'\psi_1, ..., {}'\psi_4)$ are known. Eliminating θ gives the following relation:

$$5{}'A_0 - {}'A_2 + {}'A_4 - 5{}'A_6 = 0, \tag{VII.5}$$

where ${}'A_0$, etc. are the same expressions as in (VI.3) but with first intermediate components (single prime on left) used.

Finally using (VII.2) in (VII.5) and putting $\chi = e^{i\phi}$ we obtain a sextic in χ (or a cubic in χ^2) to determine ϕ, viz.,

$$5A_0\chi^6 - A_2\chi^4 + A_4\chi^2 - 5A_6 = 0,$$

where A_0, etc. are the same quantities as in (VI.3).

There are six roots but for any root, or value of ϕ obtained, we can calculate in turn $({}'\psi_0, {}'\psi_1, ..., {}'\psi_4)$, θ, $({}''\psi_0, {}''\psi_1, ..., {}''\psi_4)$, ψ and finally $(\psi_0', \psi_1', ..., \psi_4')$, with

$$\psi_0' = \psi_4' \text{ and } \psi_1' = \psi_3' = 0.$$

The angles ϕ, θ, ψ having been determined, the boosts are obtained from the imaginary parts of these angles and the spatial rotations from the real parts. The spin matrices B_ϕ, C_θ, D_ψ and hence the spin transformation $Q = B_\phi C_\theta D_\psi$ is determined. Then also the complex rotation matrix P of (VII.1) which makes

$$P\mathcal{F}P^T = \mathcal{F}' \text{ (diag } \mathcal{F}') \text{ is determined.}$$

Alternatively one can determine the matrices P_ϕ, P_θ, P_ψ from

$${}'\mathcal{F} = P_\phi\mathcal{F}P_\phi{}^T,$$
$${}''\mathcal{F} = P_\theta{}'\mathcal{F}P_\theta{}^T \quad \text{make } {}''\mathcal{F}_{xz} = {}''\mathcal{F}_{yz} = 0,$$
$$\mathcal{F}' = P_\psi{}''\mathcal{F}P_\psi{}^T \quad \text{make } \mathcal{F}'_{xy} = 0,$$

and working backwards determines the complex Euler angles.

A geometrical interpretation of diagonalizing $\mathcal{F}$ is that of determining the complex rotation necessary to obtain the principal axes of the complex 3-D quadric $\mathcal{F}$ or equivalently the principal axes of the Weyl tensor.

A few points of further interest may be suggested.

Since boosts and rotations are involved it may be worth considering a "Thomas Precession" of this complex gravitational field.

It may also be of interest to investigate the effects on the formalism when boost velocities tend toward the speed of light.

In order to make our results more useful to those interested in further detail we have prepared a somewhat more expanded version in a paper entitled "Reduction of Weyl tensor to canonical form: Algebraically general case," available from Physics Auxiliary Publication Service, A.I.P. See AIP document no. PAPS JMAPA-22-1445-38 for 38 pages of the expanded version of the above document. Order by PAPS number and journal reference from American Institute of Physics, Physics Auxiliary Publication Service, 335 East 45th Street, New York, N.Y. 10017. The price is $1.50 for each microfiche (98 pages), or $5 for photocopies of up to 30 pages with $0.15 for each additional page over 30 pages. Airmail additional. Make checks payable to the American Institute of Physics).

ACKNOWLEDGMENTS

The authors would like to thank Professor John A. Wheeler for suggesting this problem and for his encouragement, and also express gratitude to the University of Texas, Centers for Relativity and Theoretical Physics for their hospitality.

[1] J. A. Wheeler, Phys. Rev. D **16**, 3384 (1977).
[2] A. Matte, Canad. J. Math. **5**, 1 (1953).
[3] L. Bel, Cah. Phys. **16**, 59 (1959).
[4] A. Peres, Phys. Rev. D **18**, 608 (1978).
[5] M. Carmeli, *Group Theory and General Relativity* (McGraw-Hill, New

York, 1977), Chap. 8.
[6]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), p. 550.
[7]The signs of $\mathscr{E}$ and $\mathscr{B}$ are changed from Ref. 1 owing to change of signature.
[8]E. B. Elliot, *Introduction to the Algebra of Quantics* (Chelsea, New York, 1964), 2nd ed. reprinted, p. 271.
[9]R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960); the values of I, J Eq. (4.4) in this reference are respectively twice and six times the values in this paper.
[10]Compare Ref. 3 for the real and imaginary parts of these invariants.
[11]W. S. Burnside and A. W. Panton, *The Theory of Equations* (Longmans, Green, London, 1886), 2nd ed., p. 120.
[12]Same as Ref. 9, p. 183.
[13]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, Mass., 1950); the interpretation of the Euler angles is the same as here but the choice of Pauli matrices is different.

Vacuum counterexamples to the cosmic censorship hypothesis

Bonnie D. Miller
Department of Mathematics and Department of Astronomy and Astrophysics, Michigan State University, East Lansing, Michigan 48824

(Received 17 September 1980; accepted for publication 30 January 1981)

In cylindrically symmetric vacuum spacetimes it is possible to specify nonsingular initial conditions such that timelike singularities will (necessarily) evolve from these conditions. Examples are given; the spacetimes are somewhat analogous to one of the spherically symmetric counterexamples to the cosmic censorship hypothesis.

PACS numbers: 04.20.Cv, 04.20.Jb, 98.80.Bp

I. INTRODUCTION

A number of spacetimes are known which contain timelike singularities, and thus violate the most common formal version of the cosmic censorship hypothesis,[1] the hypothesis that spacetime is globally hyperbolic, or, equivalently, that it has a Cauchy surface. Particularly of interest among these spacetimes are those in which the singularity develops from nonsingular initial conditions.[2] These solutions to Einstein's equations—which we shall call "naked-singularity solutions"—formally represent the possibility that timelike singularities could result from the collapse of initially well-behaved physical systems. (They constrast, for instance, to solutions such as the negative-mass Schwarzschild solution in which the timelike singularity is infinitely old, or to solutions in which the timelike singularity is part of the big bang in a finitely old universe.[3])

As it is not ultimately the above-mentioned formal statement of the cosmic censorship hypothesis (or any of its variants[1]) which is of interest of physics, but rather the hypothesis that naked singularities do not form in physically realistic situations, a primary question to be asked of any counterexample to the former is whether it would also be a counterexample to the latter. As the latter has not yet been precisely formulated, however,[1,4] one asks simply whether the given counterexample is physically realistic. All of the known naked-singularity solutions contain matter; a large part of the last question, then, is just how sensitive the formation of the singularity is, to the detailed properties of the matter comprising the collapsing system. A vacuum analog to any of these counterexamples—or, for that matter, any naked-singularity solution in vacuum—might indicate that there are some types of naked singularities whose formation does not depend critically on the properties of matter, or, more generally, on the presence of matter. Because the Birkhoff theorem implies that such analogs cannot exist in spherical symmetry, and because of the extreme difficulty of solving the time-dependent vacuum field equations in axisymmetric asymptotically-flat spacetimes, the natural first choice of where to look for these vacuum analogs is in cylindrically symmetric spacetimes. This paper gives a simple class of cylindrically symmetric vacuum naked-singularity solutions. It is in several respects analogous to one of the spherically symmetric classes; perhaps the largest difference is that in spherical symmetry, avoiding the formation of a horizon is a major problem in the construction of naked singularities, whereas in cylindrical symmetry there is no possibility of a horizon forming.[5]

Cylindrically symmetric spacetimes are, of course, extremely unrealistic physically. Our point, however, is simply that in the simplest case where the formation of naked singularities in vacuum is not clearly prohibited, it is allowed. That is, roughly speaking, as soon as one gives the field equations enough freedom so that gravitational radiation may be present, that radiation is capable of creating naked singularities.

The spherically symmetric spacetimes,[2] to which our cylindrically symmetric spacetimes are somewhat analogous, may be described as follows: at the center of a collapsing, radiating star, the density becomes infinite as time $T \to 0$. The radiation rate throughout the star is large enough so that the mass function $m(R,T)$ remains less than $R/2$ everywhere. No horizon forms, then, and the singularity at $(R,T) = (0,0)$ is timelike. This singularity, whose mass is zero, may begin to emit photons, in which case it becomes a negative-mass singularity.

Our cylindrically symmetric vacuum spacetimes (which are given in Einstein–Rosen coordinates $[r,t]$) may be described rather similarly: on the symmetry axis of a collapsing system of gravitational radiation, curvature invariants become infinite as $t \to 0$. The (timelike) singularity which thus forms at $(r,t) = (0,0)$ may begin to emit gravitational radiation, at such a rate that the singularity becomes "negative-mass-like"—i.e., repulsive.[6]

Section II discusses the construction of the solutions. Section IIA is a brief discussion of the field equations and their solution in the Einstein–Rosen metric; the solution involves two arbitrary functions. Section IIB specifies a class of choices of these two functions, which produces spacetimes with the following properties:

(i) $r = 0$, $t \geqslant 0$ is singular in all but one case, which is singular at $r = 0$, $t = 0$. In all cases, as $(r,t) \to (0,0)$ from any direction in the past of (0,0), curvature invariants become infinite;

(ii) at all nonsingular points the metric is C^∞, except on the two null hypersurfaces $r \pm t = 0$, on which it is C^2. (That the metric is not C^3 on any of the $t =$ constant < 0 spaceslices is of some significance: if it were, then theorems on

 0022-2488/81/071452-05$1.00

global behavior of solutions to the wave equation[7] would imply the existence of a nonsingular future for that slice. By contrast, the solutions given here do not allow a nonsingular future for such a slice.)

Section II B also discusses a certain function of time, $m(t)$, whose sign determines whether the singularity, at a given time, is attractive or repulsive. As the sign of $m(t)$ may vary with time, the singularity may be alternately attractive and repulsive.

II. CONSTRUCTION OF THE SOLUTIONS

A. The Einstein–Rosen metric

We assume here the Einstein–Rosen form of the metric in cylindrical symmetry:

$$ds^2 = -e^{2(\gamma-\psi)}(dt^2 - dr^2) + e^{2\psi}dz^2 + r^2e^{-2\psi}d\theta^2, \quad (1)$$

where ψ and γ are functions of r and t. The magnitudes of the two Killing vector fields are $|\xi_z| = e^{\psi}$ and $|\xi_\theta| = re^{-\psi} = 1/2\pi$ (circumference of the cylinder $[r,t] = \text{const}$); note that r is the product of $|\xi_z|$ and $|\xi_\theta|$. The function γ specifies Thorne's energy-like "C-energy scalar".[8] At present, none of the functions which have been defined in cylindrically symmetric spacetimes appears to be a fully satisfactory analog to the mass function which occurs in spherically symmetric space times.

The field equations in vacuum are

$$\frac{\partial^2\psi}{\partial t^2} - \frac{\partial^2\psi}{\partial r^2} - \frac{1}{r}\frac{\partial\psi}{\partial r} = 0, \quad (2a)$$

$$\frac{\partial\gamma}{\partial t} = 2r\frac{\partial\psi}{\partial r}\frac{\partial\psi}{\partial t}, \quad (2b)$$

and

$$\frac{\partial\gamma}{\partial r} = r\left[\left(\frac{\partial\psi}{\partial r}\right)^2 + \left(\frac{\partial\psi}{\partial t}\right)^2\right]. \quad (2c)$$

The solution to the linear equation (2a) is the sum of an ingoing wave solution, an outgoing wave solution, and a static solution. The static solutions to Eqs. (2a)–(2c)—i.e., the Levi-Civita solutions—are described in detail by Thorne[8]; they are given by $\psi = -2k\ln(r/r_0)$ and $\gamma = 4k^2\ln(r/r_0)$, where k and r_0 are constants. We will assume that there is no Levi-Civita component to the solution, as we are interested only in solutions which are nonsingular at early times.

Thus, we consider only the time-dependent solution to (2a), the sum of outgoing and ingoing waves,

$$\psi(r,t) = \psi_{\text{out}} + \psi_{\text{in}}. \quad (3)$$

While basically ψ_{out} and ψ_{in} are integrals over the z-axis, we will here write these integrals in the common forms

$$\psi_{\text{out}}(r,t) = \int_{-\infty}^{t-r} \frac{f(\eta)\,d\eta}{[(t-\eta)^2 - r^2]^{1/2}} \quad (4)$$

and

$$\psi_{\text{in}}(r,t) = \int_{t+r}^{\infty} \frac{g(\beta)\,d\beta}{[(\beta-t)^2 - r^2]^{1/2}}. \quad (5)$$

The functions $f(t)$ and $g(t)$—called the "source function" and "sink function", respectively—are arbitrary functions of t, subject only to whatever boundary conditions or differentiability requirements may be present, and to the requirement (easily met) that f and g approach zero fast enough, as $t\to-\infty$ and $t\to\infty$, respectively, that the distant (i.e., $z\to\infty$) contributions to the fields are finite. One would normally assume that f is negative-definite and g positive-definite; we will make these assumptions, and also assume that f and g are bounded for all t and smooth for all $t \neq 0$.

Special cases of the integrals (4) and (5) which will be of interest here are those which occur when f and g, respectively, are identically zero for $t < 0$. When f is zero for $t < 0$, the lower limit of integration in (4) may be taken to be zero, and the integral is identically zero if $u \equiv t - r < 0$. When g is zero for $t < 0$, then if $v \equiv t + r < 0$, the lower limit of integration in (5) may be taken to be zero; that the integral (5) is not identically zero under these circumstances is the time-reverse of the well-known phenomenon that an outgoing cylindrical wave has a "tail." Here, we will refer to this time-reverse as the "nose" of the ingoing wave.

Alternate representations of ψ_{out} and ψ_{in}, useful for calculating the derivatives of ψ with respect to r and t, are

$$\psi_{\text{out}}(r,t) = \int_0^{\infty} f(t - r\cosh\zeta)\,d\zeta \quad (6)$$

and

$$\psi_{\text{in}}(r,t) = \int_0^{\infty} g(t + r\cosh\zeta)\,d\zeta. \quad (7)$$

In general, $\psi(r,t)$ and its derivatives will not be well-behaved at all (r,t), and the resulting spacetimes have singularities as well as nonsingular hypersurfaces on which some derivatives of the metric are discontinuous. Clearly, however, the arbitrariness in the choice of f and g gives one much control over the differentiability (as well as other properties) of constructed solutions.

B. The choice of the source and sink functions

Here, we are interested in choosing the source and sink functions $f(t)$ and $g(t)$ so that the resulting spacetime—a superposition of ingoing and outgoing gravitational waves—allows the following description:

For sufficiently early t—say, for $t < 0$—the space-slices $t = \text{constant}$ are nonsingular. However, gravitational radiation which is present on those slices is collapsing towards the symmetry axis $r = 0$ in such a way that the axis is becoming singular as $t\to0$. The singularity which forms at $t = 0$ may (unpredictably) emit radiation as the ingoing radiation continues its collapse, and the net effect, when both collapsing and emitted radiation are considered, may be that the singularity is repulsive.

To allow the preceding rough interpretation, the spacetime should meet the following conditions:

(1) On the symmetry axis, invariant measures of curvature become infinite as $t\to0$, and the axis is singular for $t \geq 0$;

(2) At all points other than $r = 0$, $t \geq 0$, the metric is suitably differentiable; in particular we will ask that it is C^2; and (optionally)

(3) For $t > 0$, the singularity is repulsive.

In order that the second condition be satisfied, f and g must be chosen so that the integrals for ψ_{out} and ψ_{in} and their first and second derivatives are continuous functions except

at $r = 0$, $t \geqslant 0$. It follows from Eqs. (2b) and (2c) that the $\gamma(r,t)$ then determined (up to a constant) by $\psi = \psi_{\text{out}} + \psi_{\text{in}}$ will be as well-behaved as ψ.

To satisfy conditions (1) and (2), we now impose the following constraints on $f(t)$ and $g(t)$:

(i) $f(t)$ and $g(t)$ are both identically zero for $t < 0$ (i.e., the axis $r = 0$, in both its "source" and its "sink" capacity, is inactive prior to $t = 0$).

(ii) As $t \to 0_+$, the function $g(t)$ is proportional to t^{ν}, where ν is any number in the range $3/2 < \nu < 2$, and the function $f(t)$ is proportional to t^{μ}, where μ is any number larger than $3/2$.

The constraints (i) and (ii) produce the following effects on the spacetime, which is divided now into two regions, $u < 0$ and $u > 0$, which are separated by the wavefront $u = 0$ of the outgoing radiation. (See Fig. 1.)

In the region $u < 0$, $\psi_{\text{out}} \equiv 0$, so that the geometry at a point is determined (up to the additive constant in γ) by $g(t)$. Equation (3) for $\psi(r,t)$ reduces to Eq. (5) for ψ_{in}:

$$\psi(r,t) = \psi_{\text{in}}(r,t) = \int_{t+r}^{\infty} \frac{g(\beta)\, d\beta}{[(\beta - t)^2 - r^2]^{1/2}} \quad (u < 0). \tag{8}$$

There is a further simplification of $\psi(r,t)$ when $v \leqslant 0$:

$$\psi(r,t) = \int_{0}^{\infty} \frac{g(\beta)\, d\beta}{[(\beta - t)^2 - r^2]^{1/2}} \quad (v \leqslant 0). \tag{9}$$

The field $\psi(r,t)$ described by Eq. (9)—the nose of the ingoing cylindrical wave—clearly reduces to a smooth function of time on the axis $r = 0$. More generally, it follows from Eq. (9) and the constraints on g that at any point in the region $v < 0$, ψ and all of its derivatives are continuous functions. However, as $(r,t) \to (0,0)$ from within this region (e.g., from along the line $r = 0$, $t < 0$), $\partial^2\psi/\partial t^2$ and $\partial^2\psi/\partial r^2$ become infinite, so that curvature invariants ($R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$, for instance) become infinite; the condition that these quantities become infinite is the inequality $\nu < 2$. The divergence of $\partial^2\psi/\partial t^2$ and $\partial^2\psi/\partial r^2$ is due to nearby (i.e., $z \to 0$) rather than distant contributions to the relevant integrals, and in this sense, the singularity at $(r,t) = (0,0)$ is caused "locally" rather than by the infinite extent of the symmetry axis.

Examination of the derivatives of ψ_{in} shows that ψ_{in} is C^2 on $v = 0$ (except at $r = 0$); for all $v > 0$ (but $r \neq 0$) ψ_{in} is again C^{∞}; the condition that ψ_{in} be at least C^2 on $v = 0$ is the inequality $3/2 < \nu$.

In the region $u \geqslant 0$, both outgoing and ingoing waves are present; $\psi(r,t)$ is then

$$\psi(r,t) = \int_{0}^{t-r} \frac{f(\eta)\, d\eta}{[(t - \eta)^2 - r^2]^{1/2}} + \int_{t+r}^{\infty} \frac{g(\beta)\, d\beta}{[(\beta - t)^2 - r^2]^{1/2}} \quad (u \geqslant 0). \tag{10}$$

When $u > 0$, ψ_{out} is C^{∞} for $r \neq 0$. As $u \to 0_+$, ψ_{out} and its first and second derivatives—functions which are all identically zero for $u < 0$—all approach zero. Thus ψ_{out} is C^2 on $u = 0$ (except at $r = 0$); the condition that ψ_{out} be at least C^2 is the inequality $3/2 < \mu$.

The locus $r = 0$, $t \geqslant 0$ is a timelike singularity of the spacetime: ψ_{out} and ψ_{in} become infinite—again, for "local" reasons—as one approaches this locus. Near the singularity,[9] $\psi_{\text{out}}(r,t) \approx -f(t)\ln r$ and $\psi_{\text{in}}(r,t) \approx -g(t)\ln r$. As $r \to 0$, then, $\psi(r,t)$ is given approximately by

$$\psi \approx -[f(t) + g(t)]\ln r \quad (r \to 0,\ t \geqslant 0). \tag{11}$$

[We note that, if $f(t) + g(t)$ passes through 0 at a discrete value of t, then $r = 0$ will still be singular at that t, because of the derivatives of f and g. The special case $f(t) + g(t) \equiv 0$ will be discussed separately below.] The corresponding approximation for $\gamma(r,t)$ is

$$\gamma \approx [f(t) + g(t)]^2 \ln r \quad (r \to 0, t \geqslant 0), \tag{12}$$

so that

$$\gamma - \psi \approx [(f + g) + (f + g)^2]\ln r \quad (r \to 0,\ t \geqslant 0). \tag{13}$$

Thus it is the sum of the source and sink functions $f(t) + g(t)$ which determines the behavior of the metric near $r = 0$. In particular, in the three cases $0 < f + g$, $-1 < f + g < 0$, and $f + g < -1$, the geometry as $r \to 0$ on the space-slice $t =$ constant is that of a Levi-Civita solution in which $2k > 0$, $-1 < 2k < 0$, or $2k < -1$, respectively. For instance, if $f + g > 0$, then $|\xi_z| \to \infty$ and $|\xi_\theta| \to 0$ as $r \to 0$; if $-1 < f + g < 0$, then $|\xi_z| \to 0$ and $|\xi_\theta| \to 0$ as $r \to 0$; and if $f + g < -1$, then $|\xi_z| \to 0$ and $|\xi_\theta| \to \infty$ as $r \to 0$. As t varies, then, the geometry near $r = 0$ may undergo enormous fluctuations.

The wide choice of both f and g which one has in writing down Einstein–Rosen solutions does not, of course, carry over to the construction of physical systems (even ignoring the difficulties in setting up a physical system with cylindrical symmetry); while one might construct a system whose initial data—say, on a $t =$ constant < 0 slice—are those determined by one of the sink functions $g(t)$ specified above, the source function $f(t)$ would not be predictable. In that sense, then, no particular source function corresponds to a given sink function. Nonetheless, one may single out as being of particular interest the source function $f(t) \equiv -g(t)$, which—

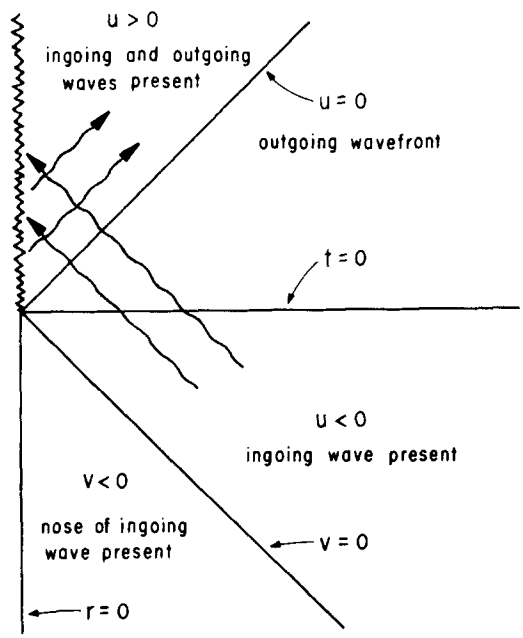


FIG. 1. A cylindrically symmetric vacuum spacetime which is nonsingular for $t < 0$ and singular at $r = 0$ for $t \geqslant 0$.

to the extent possible—cancels the effects of the sink function at $r = 0$. When $f(t) + g(t) \equiv 0$, $r = 0$ will be singular only for $t = 0$ rather than for all $t \geqslant 0$. [And if the constraint (ii) is relaxed—in particular, if the number ν is taken to be large enough so that the metric is C^3 for $t < 0$—then a nowhere singular spacetime will result if $f(t) \equiv -g(t)$.]

A second source function of some interest is the one $f(t) \equiv 0$, which results in a spacetime consisting solely of ingoing radiation. That a singularity forms at $r = 0$ in this case (regardless of how slowly the sink function turns on at $t = 0$), shows that ingoing gravitational waves do not, by themselves, "bounce" or "pass through themselves" when they arrive at the symmetry axis.

We summarize now the differentiability of solutions to the field equations which are subject to the constraints (i) and (ii): they are singular at $r = 0$ when $t \geqslant 0$ (with the above mentioned exception, which is singular only for $t = 0$), C^2 on $u = 0$ $(r \neq 0)$ and on $v = 0$ $(r \neq 0)$, and C^∞ at all other points.

Having satisfied conditions (1) and (2), we turn to condition (3). Repulsiveness of the singularity is achieved by the imposition on f and g of one more constraint, which follows immediately from seeing that the singularity is repulsive or attractive, as the sign of the function

$$m(t) \equiv \lim_{r \to 0} \frac{1}{2} \frac{\partial(\gamma - \psi)}{\partial \ln r} = \frac{1}{2}[(f + g) + (f + g)^2] \quad (14)$$

is negative or positive, respectively. {The definition of $m(t)$ includes the case that there is a static component of the field; when this component is present, $m(t) = \frac{1}{2}[(1 + 4k)(f + g) + (f + g)^2 + 4k^2 + 2k]$, where k is the Levi–Civita constant.} We note that $m(t)$ is confined to the range

$$-\tfrac{1}{8} \leqslant m(t) < \infty, \quad (15)$$

and that it changes sign when $f + g$ passes through 0 or -1. [If $k \neq 0$, the inequality (15) still holds, but $m(t)$ changes sign when $f + g + 2k$ passes through 0 or -1.]

To see that the sign of m determines the repulsive or attractive character of the singularity, it suffices to look at the r-component of the geodesic equations. This equation is

$$\frac{d^2 r}{ds^2} = \frac{1}{r} \frac{\partial(\psi - \gamma)}{\partial \ln r} \left[\left(\frac{dr}{ds}\right)^2 + \left(\frac{dt}{ds}\right)^2\right] + 2 \frac{\partial(\psi - \gamma)}{\partial t} \frac{dr}{ds} \frac{dt}{ds} + e^{4\psi - 2\gamma} r^{-2} \left(\frac{1}{r} - \frac{\partial \psi}{\partial r}\right) P_\theta^2 + e^{-2\gamma} \frac{\partial \psi}{\partial r} P_z^2, \quad (16)$$

where $P_\theta \equiv r^2 e^{-2\psi}\, d\theta/ds$ and $P_z \equiv e^{2\psi}\, dz/ds$ are the two conserved momenta. Also conserved along the geodesic is $u_\beta u^\beta = -1$, where $\mathbf{u}$ is the tangent vector; thus

$$e^{2(\gamma - \psi)} \left[\left(\frac{dr}{ds}\right)^2 - \left(\frac{dt}{ds}\right)^2\right] + e^{-2\psi} P_z^2 + e^{2\psi} r^{-2} P_\theta^2 = -1. \quad (17)$$

We will assume that P_θ and P_z are zero, and later point out that the conclusions are unaffected by this assumption.

We are interested in the behavior of timelike geodesics only near the singularity $r = 0$. As $r \to 0$, the first term on the right-hand side of Eq. (16) is $-2(m/r)[(dr/ds)^2 + (dt/ds)^2]$, or equivalently, because of Eq. (17), $-2(m/r)[2(dr/ds)^2 + r^{-4m}]$. The second term is negligible compared to the first, and we ignore it. As $r \to 0$, then, the equation for radial timelike geodesics may be written

$$\frac{d^2 r}{ds^2} = -2 \frac{m(t)}{r} \left[\left(\frac{dr}{ds}\right)^2 + \left(\frac{dt}{ds}\right)^2\right] \quad (r \to 0), \quad (18)$$

or alternatively,

$$\frac{d^2 r}{ds^2} = -2 \frac{m(t)}{r} \left[2\left(\frac{dr}{ds}\right)^2 + r^{-4m(t)}\right]. \quad (19)$$

Equation (18) or (19) describes radial timelike geodesics near $r = 0$. It is convenient to rewrite Eq. (19), by defining $v \equiv dr/ds$, so that the left-hand side of the equation is $dv/ds = (dv/dr)v = \frac{1}{2} d(v^2)/dr$, and the equation becomes

$$\frac{d(v^2)}{dr} = -8m(t) v^2 r^{-1} - 4m(t) r^{-[1 + 4m(t)]}. \quad (20)$$

Suppose that $m(t) > 0$. Then $d(v^2)/dr$ is negative. Thus if r is initially decreasing along a radial timelike geodesic sufficiently near $r = 0$, the geodesic will reach $r = 0$. That is, the singularity is attractive when $m(t) > 0$.

Now suppose that $m(t) < 0$. Then $d(v^2)/dr$ is positive. Suppose that there is a timelike geodesic G which reaches the singularity at the time t_0: G goes from $r = \epsilon$ (say) to $r = 0$. Let M be the largest and $M - \delta$ the smallest value of $m(t)$ along G (including, if necessary, $m[t_0]$). Thus, $-\frac{1}{8} \leqslant M - \delta < M < 0$. (We assume $\delta > 0$, and will comment below on the special case $\delta = 0$.) Note that, given G, we can pick new values of ϵ which make δ arbitrarily small.

At each point along G, $d(v^2)/dr$ obeys

$$\left.\frac{d(v^2)}{dr}\right|_G > -8Mv^2 r^{-1} - 4Mr^{-(1 + 4M - 4\delta)}. \quad (21)$$

The geodesic G—which we now picture as a curve in the (r, v^2)-plane—can be compared in that plane to the integral curves of the differential equation

$$\frac{d(v^2)}{dr} = -8Mv^2 r^{-1} - 4Mr^{-(1 + 4M - 4\delta)}. \quad (22)$$

Equation (22) describes $v^2(r)$ along curves closely related to G, but which at each point (r, v^2) crossed by G have a smaller slope than G.

The solution of Eq. (22) is

$$v^2(r) = Kr^{-8M} - [M/(\delta + M)] r^{-4M + 4\delta}, \quad (23)$$

where K is a positive constant. If we take ϵ so that $\delta \approx 0$, then a good approximation to Eq. (23) is

$$v^2(r) \approx Kr^{-8M} - r^{-4M}. \quad (24)$$

From either (23) or (24) one sees that, along the integral curves of the differential equation (22), v^2 reaches zero at a value of r which is greater than zero. That is, the integral curves of (22) do not reach $r = 0$. [And in the special case $\delta = 0$, in which Eq. (22) is the geodesic equation, (23) is the solution to the geodesic equation, which shows that there is no such geodesic as G.]

Now consider any of these integral curves which intersects G; call it C. From (21), the intersection point (r_0, v_0^2) is unique, and the slope of G at this point is greater than the slope of C. Thus for all values of v^2 less than v_0^2, r is greater on G than on C. Since the smallest value of r on C is greater than zero, then, the smallest value of r on G must also be greater

than zero. This contradicts the definition of G. Thus, G does not exist. That is, there are no radial timelike geodesics which reach the singularity when $m(t) < 0$. And it is straightforward to show from Eq. (16) that nonzero values of P_θ or P_z generate positive contributions to $d(v^2)/dr$, so that nonradial timelike geodesics are also incapable of reaching the singularity when $m(t) < 0$. No timelike geodesics reach the singularity, then, when $m(t) < 0$. That is, the singularity is repulsive when $m(t) < 0$.

As noted earlier, $m(t)$ changes sign when $f(t) + g(t)$ passes through either 0 or -1: $m(t)$ is positive when either $f + g > 0$ or $f + g < -1$, and it is negative when $-1 < f + g < 0$. The constraint on f and g which produces a repulsive singularity, then, which may be added to the list (i) and (ii), is

(iii) $-1 < f(t) + g(t) < 0$.

It should be noted that there are many pairs of functions $f(t)$ and $g(t)$ which satisfy the constraints (i)–(iii).

The static, Newtonian limit of Eq. (18) (in which the only surviving component of $m[t]$ is $2k^2 + k$) is the equation for radial motion in the gravitational field of a line source of mass (per unit length) m. This formal analog, together with the information carried by $m(t)$, suggests that $m(t)$ might be called the mass (per unit length) of the singularity at time t. Regardless of whether this designation is appropriate, the inequality (15) shows a rather striking phenomenon [due to the nonlinearity of Eq. (2c)] in Einstein–Rosen spacetimes, that there is a limit to how repulsive a singularity can be. In particular, as a source function decreases from zero, it produces first a "negative-mass-like" singularity, but eventually, if it decreases sufficiently, a "positive-mass-like" singularity.

III. CONCLUDING COMMENTS

If there is to eventually be a precise formulation and proof of a hypothesis to the effect that naked singularities do not form in physically realistic circumstances, then counterexamples to the current precise version (s) of the cosmic censorship hypothesis are all, in some sense or another, physically unrealistic. That sense is abundantly clear, for those cases which are cylindrically symmetric, or which for any other reason cannot be thought of as representing an essentially bounded physical system. For cases which pass the most rudimentary physical reality test, asymptotic flatness, Penrose mentions two more or less distinct reasons why the formation of the naked singularity may be unstable[1]: it may depend critically on the equation of state of the matter present, or it may require a carefully chosen set of initial conditions.

Although the cylindrically symmetric counterexamples given in Sec. II are not even asymptotically flat, there is no apparent reason why cylindrically symmetric vacuum—the simplest vacuum is which naked singularities can form—should also be the only vacuum in which they can form. Rather, it would seem more likely that, in other spacetimes which are sufficiently asymmetric that gravitational radiation is allowed (and hence can play a significant role in the dynamics of spacetime) there are further examples of naked singularities forming in vacuum.

Thus, we take the cylindrically symmetric vacuum naked-singularity solutions as an indication that there are other classes of naked singularities (quite possibly in asymptotically flat spacetimes) which form in vacuum; viewed somewhat more generally, the indication is that there are classes which form in a manner which is insensitive to the properties of the ambient matter. Presumably, then, these (hypothetical) classes are unstable to perturbations of initial conditions: to departures from symmetry. Even so, as gravitational radiation is a phenomenon associated expressly with a lack of symmetry, it does not seem altogether certain that such departures will necessarily prevent a naked singularity from forming; just as vacuum naked singularities can first occur when a modest asymmetry is allowed into spacetime, perhaps it is generally so that their occurrence becomes increasingly common as spacetime becomes increasingly asymmetric.

ACKNOWLEDGMENTS

I thank R. Sachs for comments on the manuscript, and R. Geroch and G. Horowitz for comments on cosmic censorship. Also, I thank the Computer Science Division, E.E.C.S., at the University of California, Berkeley, for hospitality extended during the course of this work.

[1] R. Penrose, in *Theoretical Principles in Astrophysics and Relativity*, edited by N. R. Lebovitz, W. H. Reid, and P. O. Vandervoot (University of Chicago Press, Chicago, 1978). This article discusses a number of different formulations of the cosmic censorship hypothesis.

[2] The particular type that we are interested in here is one first mentioned by B. Steinmuller, A. R. King, and J. P. Lasota, in Phys. Lett. A **51**, 191 (1975), and later by B. D. Miller, in Astrophys. J. **208**, 275 (1976), in a brief discussion of solutions by H. Bondi, Proc. Roy. Soc. London, Ser. A **281**, 39 (1964). Other examples are those by P. Yodzis, H. J. Seifert, and M. Muller zum Hagen, Commun. Math. Phys. **34**, 135 (1973) and **37**, 29 (1974).

[3] B. D. Miller, Astrophys. J. **208**, 275 (1976); K. Tomita, Prog. Theor. Phys. **59**, 1150 (1978); B. D. Miller, J. Math. Phys. **20**, 1356 (1979).

[4] R. Geroch and G. Horowitz, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979). Diffuculties in the formulation, as well as approaches to it, are discussed in this article.

[5] K. S. Thorne, in *Magic Without Magic: John Archibald Wheeler*, edited by J. Klauder (Freeman, San Francisco, 1972).

[6] Precisely, a "repulsive" singularity is one which is timelike, but on which no timelike geodesics terminate.

[7] For instance, see G. B. Folland, *Introduction to Partial Differential Equations* (Princeton U. P., Princeton, N.J., 1976), Chap. 5.

[8] K. S. Thorne, "Geometrodynamics of Cylindrical Systems," unpublished Ph. D. Thesis, Princeton University, 1965.

[9] G. B. Whitman, *Linear and Nonlinear Waves* (Wiley, New York, 1974). Section 7.4 discusses several useful details about cylindrical waves.

Solutions to wave equations on black hole geometries [a)]

W. E. Couch
Department of Mathematics & Statistics, The University of Calgary, Calgary, Alberta, T2N 1N4, Canada

(Received 7 October 1980; accepted for publication 6 February 1981)

Classes of explicit special solutions to the scalar wave equation on black hole spacetime geometries are given. Some analogous solutions also exist for fields of spin one or two. The spin-zero radial equation on the Reissner-Nordström geometry with $m = e$ and on the Kerr geometry with $m = a$ is shown to be related to the Whittaker-Hill equation.

PACS numbers: 04.20.Jb

I. INTRODUCTION

In this paper we study the radial differential equations which govern fields on the Reissner-Nordström (RNG), Kerr, and Schwarzschild (SG) black hole background geometries. The metrics of these geometries are given by

$$ds^2 = (1 - 2m/r + e^2/r^2)dt^2 - (1 - 2m/r + e^2/r^2)^{-1}dr^2 - r^2 d\Omega^2 \quad (1)$$

for RNG in the usual exterior coordinates, by

$$ds^2 = (1 - 2mr/\Sigma)dt^2 + (4\,mar\sin^2\theta/\Sigma)dtd\phi - (\Sigma/\Delta)dr^2 - \Sigma d\theta^2 - \sin^2\theta(r^2 + a^2 + 2ma^2 r\sin^2\theta/\Sigma)d\phi^2 \quad (2)$$

for Kerr in Boyer–Lindquist coordinates where $\Delta \equiv r^2 - 2mr + a^2$ and $\Sigma \equiv r^2 + a^2\cos^2\theta$, and by $e = a = 0$ for SG. The parameters m, e, and a are, respectively, the mass, charge, and angular momentum of the black hole. We will call real values of these parameters satisfying $m \geqslant e \geqslant 0$, $m \geqslant a \geqslant 0$, $m > 0$, physical values.

We denote the special RNG satisfying $m = e$ by CMG. The only Kerr geometry which we consider is the special one with $m = a$, we denote it by AMG.

Our purpose is to present some special explicit solutions and to show that the equations on CMG and AMG can be easily related to a standard equation of mathematical physics. The details of our analysis are presented for the massless scalar (spin $s = 0$) field (governed by the covariant scalar wave equation $\Box\Psi = 0$), but we give some examples for massive fields and for higher spin fields and indicate that the results may be extendable to these cases. In all cases the field may be obtained from a single scalar function,[1] Ψ, and in all cases Ψ is separable,[2] $\Psi = e^{i\omega t}\psi(r)S(\theta)e^{im'\phi}$. For spin 0, Ψ is the field itself.

These fields are directly related to interesting problems in the physics of black holes. They reveal the nature of the scattering of radiation which propagates in the gravitational field produced by a black hole; the spin 0 field's reflection coefficient, $R(\omega)$, for incident radiation determines the spectral distribution of the Hawking emission[3]; the fields are involved in the question of the stability of black hole spacetimes.

Except in the special case[4] $m = a$, $\omega = 2am'/(1 + \sqrt{2})$, workers have not been able[5,6] to relate the ordinary differential equations which arise to the standard equations of mathematical physics and so desired information is not easily obtained. Persides[7] has obtained Frobenius series in r for the spin 0 field on a Schwarzschild geometry. His recursion relations are five-term, but he manages to prove convergence, obtain extensions, and relate various solutions with different asymptotic behavior. Most other authors use various analytic approximations[8] or computer calculations[9] to obtain quantities of interest such as $R(\omega)$. However, Hartle and Wilkins[10] use an integral equation and iteration to establish the analyticity properties of $R(\omega)$ and thereby prove (up to questions of completeness[11]) the stability of the Kerr geometry to perturbations having spin 0, $\frac{1}{2}$, 1, and 2.

In this paper we show that the $s = 0$ radial equations on CMG and AMG (with $m' = 0$ in the AMG case) can be related to the well-known Whittaker–Hill equation. This equation occurs often in physical problems and has been studied in the past; it occurs, for example, when the flat space wave equation is separated in a paraboloidal coordinate system.[12]

By using known properties of the Whittaker–Hill equation and generalizations thereof in the cases of RNG and SG, we obtain several classes of solutions in closed form in terms of elementary functions. Some solutions have nonphysical values of the black hole parameters and for other solutions $S(\theta)$ has singularities, e.g. a half-line singularity. The angular singularities occur because the separation constant l is complex. When the parameter values are physical and $S(\theta)$ is nonsingular (case of RNG) the asymptotic behavior of the solutions is investigated. It is found that all solutions blow up on at least one of the null infinities or horizons. This behavior occurs partly because ω has complex values.

For a $s = 0$ field on CMG or AMG the connection with the Whittaker–Hill equation yields convergent infinite series solutions (with three-term recursion relations) which are available in the literature.[13,14]

It is often advantageous to study differential equations with respect to all complex values of their parameters, as potential scattering theory shows.[15] The solutions presented here, with their complex ω and sometimes complex l, may be useful in studies of fields around black holes, especially studies which employ a potential scattering approach. A special case of the solutions found here are the well-known static ($\omega = 0$) fields and they have been, of course, important[16] though they are singular on the horizon.

[a)]Research supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/81/071457-06$1.00

II. THE WHITTAKER–HILL EQUATION

We present in this section the aspects of the Whittaker–Hill equation which are used or generalized in later sections. The material may be found in the book by Arscott.[14]

The equation

$$d^2\psi/dx^2 + A\psi = 0 \tag{3}$$

is Hill's equation if A is a periodic function[17] of x. The Whittaker–Hill equation is the special case

$$d^2\psi/dx^2 + (\theta_0 + 2\theta_1 \cos 2x + 2\theta_2 \cos 4x)\psi = 0, \tag{4}$$

where θ_0, θ_1, θ_2 are real constants. When $\theta_2 = 0$, (4) is the Mathieu equation.

A related important equation due to Ince results from the transformation

$$\psi = Ve^{-(1/4)\xi \cos 2x}. \tag{5}$$

Ince's equation is

$$d^2V/dx^2 + \xi \sin 2x\,(dV/dx) + (\lambda - p\xi \cos 2x)V = 0. \tag{6}$$

The new parameters ξ, p, and λ, introduced for convenience, are defined by

$$\xi^2 = 16\theta_2, \tag{7a}$$

$$(p+1)\xi = -2\theta_1, \tag{7b}$$

$$\lambda = \theta_0 + 2\theta_2. \tag{7c}$$

Now if certain relations hold among θ_0, θ_1, and θ_2, then (6) has solutions which are *finite* series of sines or cosines whereas (4) does not possess such solutions.[18] For example, (6) has the obvious solution $V = 1$ for $p = \lambda = 0$.

The finite series solutions exist if $\theta_2 > 0$ and are the following four types.

Type (i): The series

$$V = \sum_{k=0}^{n} a_k \cos 2kx \tag{8}$$

is a solution to (6) if $p = 2n$, $n =$ nonnegative integer, and the column vector $\mathbf{a} = (a_0, a_1, \cdots, a_n)^T$ of coefficients is determined (up to an overall factor) by

$$L\mathbf{a} = \lambda\mathbf{a}, \tag{9}$$

where L is the $n+1$ by $n+1$ tridiagonal matrix whose three diagonals are given by

$$L = \begin{pmatrix} - & (n+1)\xi, & (n+2)\xi, & (n+3)\xi, & \cdots & & 2n\xi \\ 0, & 4\cdot 1^2, & 4\cdot 2^2, & 4\cdot 3^2, & \cdots & & 4n^2 \\ 2n\xi, & (n-1)\xi, & (n-2)\xi, & (n-3)\xi, & \cdots & \xi & - \end{pmatrix} \tag{10}$$

The main diagonal is the middle row of (10), the adjacent diagonals are the other two rows. All elements of L not on one of these diagonals are zero.

The matrix L is similar to a real, symmetric (Jacobi) matrix and has $(n+1)$ real, distinct, eigenvalues λ. Hence there are $(n+1)$ conditions on $\theta_0 + 2\theta_2$ each of which gives a finite series solution (8). The condition can always be satisfied since we may consider θ_2 to be given, θ_1 to be defined by (7a) and (7b), and θ_0 to be defined by (7c).

After any particular eigenvalue λ is found, all the coefficients a_k may be very easily generated recursively from (9) starting at the last equation with $a_n = 1$. Then each successive equation gives a single a_k in terms of known quantities.

Similarly we have for Type (ii),

$$V = \sum_{k=0}^{n} a_k \cos(2k+1)x, \tag{11}$$

$$p = 2n + 1,$$

$$L = \begin{pmatrix} - & (n+2)\xi, & (n+3)\xi, & \cdots & & (2n+1)\xi \\ 1 + (n+1)\xi, & 3^2 & 5^2, & \cdots & & (2n+1)^2 \\ n\xi, & (n-1)\xi, & (n-2)\xi, & \cdots & \xi & - \end{pmatrix} \tag{12}$$

For Type (iii),

$$V = \sum_{k=1}^{n} a_k \sin 2kx, \tag{13}$$

$$p = 2n,$$

$$L = \begin{pmatrix} - & (n+2)\xi, & (n+3)\xi, & \cdots & & 2n\xi \\ 4\cdot 1^2, & 4\cdot 2^2, & 4\cdot 3^2, & \cdots & & 4n^2 \\ (n-1)\xi, & (n-2)\xi, & (n-3)\xi, & \cdots & \xi & - \end{pmatrix} \tag{14}$$

For Type (iv),

$$V = \sum_{k=0}^{n} a_k \sin(2k+1)x, \tag{15}$$

$$p = 2n + 1,$$

$$L = \begin{pmatrix} - & (n+2)\xi, & (n+3)\xi, & \cdots & (2n+1)\xi \\ 1-(n+1)\xi, & 3^2, & 5^2, & \cdots & (2n+1)^2 \\ n\xi, & (n-1)\xi, & (n-2)\xi, & \cdots\ \xi & - \end{pmatrix}. \tag{16}$$

In all these letter cases λ is any one of the eigenvalues of L (all real and distinct) and $\mathbf{a}$ is determined by $L\mathbf{a} = \lambda\mathbf{a}$.

We shall see in Sec. III that each of these finite series will provide an explicit solution to the $s = 0$ wave equation on CMG and on AMG.

There are also convergent infinite trigonometric series solutions to (6) when $\theta_2 > 0$. They are discussed by Arscott[14] along with a method for dealing with the three-term recursion relation of the series.

If $\theta_2 < 0$ the results are different but in that case periodic solutions to (4) have been found[13] in terms of convergent infinite series solutions to Ince's equation with pure imaginary ξ.

The modified Whittaker–Hill equation is defined by (4) with the cosines occurring there replaced by hyperbolic cosines. It is clear that the modified equation has finite series solutions for V' [V' defined by the hyperbolic version of (5)] obtained merely by the replacements $x \to ix$ and $(\theta_0,\theta_1,\theta_2) \to -(\theta_0,\theta_1,\theta_2)$.

In order to help us make an analogy to (4) in Sec. IV, note that the coefficient of ψ in (4) is a fourth-degree polynomial in $\cos x$ and the finite series expressions for V could all be written as polynomials in $\cos x$ times 1 or $\sin x$.

III. REISSNER–NORDSTRÖM WITH $m = e$ AND KERR WITH $m = a$

In this section we apply the results of Sec. II to CMG and AMG.

A. CMG

The transformation

$$r = m(1 - 1/x') \tag{17}$$

gives the following radial equation for the $s = 0$ field

$$d^2\psi/dx'^2 + m^2\omega^2(1 - 1/x')^4\psi - (L/x'^2)\psi = 0, \tag{18}$$

where $L = l(l+1)$ is a separation constant, $-\infty < x' < 0$ in the exterior, and $1 < x' < \infty$ in the interior.

The further change of variables $x' = -e^{2z}$, $\psi = |x'|^{1/2}\phi$ in the exterior shows that ϕ satisfies the modified Whittaker–Hill equation. Then the formal change $z = ix$ gives

$$d^2\phi/dx^2 - \{4m^2\omega^2[8\cos 2x + 2\cos 4x] + 24m^2\omega^2 - (2l+1)^2\}\phi = 0. \tag{19}$$

Thus solutions to (18) may be obtained from those to (4) via (19) and the transformations. [It does not matter that x should be considered pure imaginary in (19).]

Comparison of (19) with (4) and use of (7) gives

$$m^2\omega^2 = -(p+1)^2/16, \tag{20a}$$

$$\xi = -2(p+1), \tag{20b}$$

$$\lambda = (2l+1)^2 + 2(p+1)^2. \tag{20c}$$

Each of the solutions given in Sec. II thus generates via (20) a value for $m^2\omega^2$, a value for $(2l+1)^2$, and a solution to (18) given by

$$\psi = (|x'|)^{1/2}e^{-(1/4)(p+1)(x' + 1/x')}V, \tag{21}$$

where V is a polynomial in x' and $1/x'$ [times $(|x'|)^{1/2}$ for types (ii) and (iv)]. To illustrate we give the results for type (i), $n = 1$. We find

$$\lambda = 2 \pm 2\sqrt{37},$$
$$L = \tfrac{1}{4}\lambda - 18, \tag{22}$$
$$m^2\omega^2 = -\tfrac{9}{16},$$

and

$$\psi = \frac{r^2 - 2mr(1 + 6/\lambda) + 2m^2(1 + 6/\lambda)}{|r-m|^{3/2}} \times e^{3r/4m}e^{3m/4(r-m)}. \tag{23}$$

For all types, since p is real we are restricted to pure imaginary values of $m\omega$, and l can only possibly by a nonnegative integer if λ is an integer and if $\lambda \geqslant 1 + 2(p+1)^2$. For types (ii), (iii), and (iv), it can be shown that, for each n, the eigenvalues are all bounded below $1 + 2(p+1)^2$, so l is not real and the angular functions $S(\theta)$ are singular. For type (i) it can be shown that, for each n, at most one eigenvalue is greater than or equal to $1 + 2(p+1)^2$; a computer search shows that no such eigenvalues exist for $n \leqslant 30$.

The infinite series, applicable when $\theta_2 > 0$, mentioned in Sec. II will provide additional solutions to (18) with $m^2\omega^2 < 0$; if we require $m\omega$ to be real the relevant series is that for $\theta_2 < 0$. It would be of interest to determine whether these solutions are of value in investigating the physics of fields around black holes. We do not pursue this question here.

B. AMG

In terms of x' the radial equation for an axially symmetric $s = 0$ field on AMG is

$$d^2\psi/dx'^2 + m^2\omega^2(4 - 8/x' + 7/x'^2 - 4/x'^3 + 1/x'^4)\psi - (L/x'^2)\psi = 0. \tag{24}$$

It is easy to see that the transformation $x' = ce^{2z}$, $\psi = (|x'|)^{1/2}\phi$, $z = ix$ will convert $d^2\psi/dx'^2 + A\psi = 0$, $A = d_0 + d_1/x' + \ldots + d_4/x'^4$, into (4) if c is chosen to satisfy $c^2 = d_3/d_1$ and if $d_4/d_0 = c^4$. Equation (24) meets this requirement if we take $c = -\sqrt{2}/2$. We thus obtain

$$d^2\phi/dx^2 + \{(2l+1)^2 - 28m^2\omega^2 - 32(\sqrt{2})m^2\omega^2\cos 2x - 16m^2\omega^2\cos 4x\}\phi = 0. \tag{25}$$

Comparison of (25) with (4) and use of (7) gives

$$m^2\omega^2 = -(p+1)^2/16, \tag{26a}$$

$$\xi = -2(\sqrt{2})(p+1), \tag{26b}$$

$$\lambda = (2l+1)^2 + \tfrac{11}{4}(p+1)^2. \tag{26c}$$

Then via (26) each solution of Sec. II generates values for

$m^2\omega^2$ and $(2l+1)^2$ and a solution to (24) given by

$$\psi = \sqrt{|x'|}\, e^{-(1/2)(p+1)(x'+1/2x')} V. \tag{27}$$

The values of $m\omega$ are again pure imaginary. Real values of l exist because there are eigenvalues λ greater than or equal to $1+\frac{11}{4}(p+1)^2$ but these do not occur for small values of n; for types (i), (ii), (iii), and (iv) such eigenvalues exist only for n greater than 10, 29, 30, and 9, respectively. For $n \leqslant 32$ it has been found that none of the real l values are positive integers.

Remarks similar to those in Sec. III. A concerning infinite series solutions apply to the AMG case.

IV. REISSNER–NORDSTRÖM AND SCHWARZSCHILD GEOMETRIES

A. Solutions on the Reissner–Nordström geometry

It is possible to construct solutions to the radial equation of a $s=0$ field on RNG and SG by considering the following analogue[17] of the Whittaker–Hill equation in which $\cos x$ is replaced by $\operatorname{ctnh} x$

$$d^2\psi/dx^2 + (A_0 + A_1 y + A_2 y^2 + A_3 y^3 + A_4 y^4)\psi = 0, \tag{28}$$

where $y \equiv \operatorname{ctnh} x$, $A_0,\cdots,A_4$ are constants and $A_4 \neq 0$. This equation has an analogous treatment to that of (4) because it can be shown that if

$$\psi = e^{\gamma y} V \tag{29}$$

and if certain conditions on the coefficients of A are satisfied, there are polynomials,

$$V = \sum_{k=0}^{n} b_k y^k, \tag{30}$$

which yield solutions to (28).

The equations on RNG and SG can be put into the form (28) but the conditions on A implied by (29) and (30) are not obeyed by these geometries.

An *ansatz* which *is* applicable to RNG and SG is to use, instead of (29), the generalization

$$\psi = e^{\alpha x}(\sinh x)^{\beta} e^{\gamma y} V, \tag{31}$$

$$A_4 = -\gamma^2, \tag{32}$$

where α and β are two new parameters to be determined along with the coefficients of V in (30). The equation for V is

$$d^2V/dx^2 + 2(\alpha+\gamma+\beta y - \gamma y^2)dV/dx + \tilde{A}V = 0, \tag{33}$$

where

$$\tilde{A} = A + (\alpha+\gamma)^2 + \beta + 2(\alpha\beta + \gamma\beta - \gamma)y + (\beta^2 - \beta - 2\alpha\gamma - 2\gamma^2)y^2 + 2\gamma(1-\beta)y^3 + \gamma^2 y^4. \tag{34}$$

Note that if y is $\tanh x$ in (28) through (31) and $(\sinh x)^{\beta}$ in (31) is replaced by $(\cosh x)^{\beta}$ then (33) and (34) are again obtained.

There are some redundancies built into (28), (30), and (31). A solution ψ characterized by (α,β,γ,n) is equivalent (equal, aside from a nonzero constant factor) to one characterized by $(\alpha-q,\beta+q,\gamma,n+q)$ for every nonnegative integer q. As we shall see in the case of RNG and SG, equation (33) has polynomial solutions which correspond to fields inequivalent for different values of n and inequivalent to those obtained from (29) and (30).

We now apply (31) to RNG. In terms of the variable x defined by

$$x = \frac{1}{2}\ln\left|\frac{r-m_+}{r-m_-}\right|, \tag{35}$$

$\epsilon \equiv \sqrt{1-(e^2/m^2)}$, $m_{\pm} \equiv m \pm m\epsilon$, the radial equation on RNG is the following special case of (28):

$$d^2\psi/dx^2 + [(m^2\omega^2/\epsilon^2)(1-\epsilon y)^4 + L(1-y^2)]\psi = 0, \tag{36}$$

where $y \equiv \operatorname{ctnh} x$ in the exterior and we have

$$r = m(1-\epsilon y). \tag{37}$$

In the interior, $m_- < r < m_+$, (36) holds with $y \equiv \tanh x$.

In terms of r,ψ given by (31) has the form

$$\psi = \frac{e^{-(\gamma/m\epsilon)r}}{|r-m_+|^{(\beta-\alpha)/2}|r-m_-|^{(\alpha+\beta)/2}} V, \tag{38}$$

where V is a polynomial in r of degree n, and in terms of y (essentially r) (33) becomes

$$(y-1)^2(y+1)^2 d^2V/dy^2 - (y+1)(y-1) \times 2[\alpha+\gamma+(\beta-1)y - \gamma y^2]dV/dy + \tilde{A}V = 0. \tag{39}$$

Thus the ansatz (31) merely amounts to taking out, in addition to the exponential, the usual factors of Frobenius series about regular singular points of the r-differential equation for ψ. But a difference is that in (38) the powers $(\beta-\alpha)/2$ and $(\alpha+\beta)/2$ are not necessarily fixed; they may depend on V, i.e., on n, and this, it turns out, will allow V to be a polynomial.

It is clear from (39) that for

$$V = \sum_{k=0}^{n} (-1)^k a_k (y+1)^k \tag{40}$$

the recursion relation on **a** will have four terms whereas the recursion relation on **b** has six terms. In accord with (40) we define coefficients in $\tilde{A}$ by

$$\tilde{A} = \tilde{A}_0 - \tilde{A}_1(y+1) + \tilde{A}_2(y+1)^2 - \tilde{A}_3(y+1)^3; \tag{41}$$

they are easily determined by (34), (37), and

$$A = \frac{m^2\omega^2}{\epsilon^2}\left(\frac{r}{m}\right)^4 + \frac{L}{\epsilon^2}\left[-\left(\frac{r}{m}\right)^2 + 2\frac{r}{m} + \frac{\epsilon^2-1}{\epsilon^2}\right]. \tag{42}$$

[$\tilde{A}$ is of degree three because of (32).] Because of the form of (38) and the fact that $y+1 = -(r-m_+)/\epsilon m$ we may assume, without loss of generality, that $a_0 a_n \neq 0$. From (32) we have

$$m^2\omega^2\epsilon^2 = -\gamma^2. \tag{43}$$

Then substitution of (40) into (39) yields

$$\tilde{A}_0 a_0 = 0 \Rightarrow (\alpha-\beta)^2 = (\gamma^2/\epsilon^4)(1+\epsilon)^4, \tag{44a}$$

$$(\tilde{A}_3 - 2n\gamma)a_n = 0 \Rightarrow \beta = 2\gamma/\epsilon + n + 1, \tag{44b}$$

$$4(k+1)(k+1+\alpha-\beta)a_{k+1} + [\tilde{A}_1 + 2k(2k+\alpha-3\beta-4\gamma+1)]a_k + [\tilde{A}_2 + (k-1)(k-2\beta-8\gamma)]a_{k-1} + 2\gamma(n+2-k)a_{k-2} = 0, \quad k = 0,1,2,\ldots,n, \tag{44c}$$

$$[\tilde{A}_2 + n(n+1-2\beta-8\gamma)]a_n + 2\gamma a_{n-1} = 0, \quad (44d)$$

where we take $a_s = 0$ if $s<0$ or $s\geqslant n+1$.

We may also write $V = \Sigma_{k=0}^{n} a_k'(y-1)^k$ and substitute into (39). This produces an equation similar to the left equation of (44a), in which a_0' is uncoupled from the other components of **a**, and gives the further information:

$$(\alpha+\beta)^2 = (\gamma^2/\epsilon^4)(1-\epsilon)^4. \quad (45)$$

There are three solutions to (44a), (44b), and (45) for α, β, and γ in terms of n and ϵ. One of them we ignore because it may be obtained from one of the others by the symmetry $x\to -x$, $\epsilon\to -\epsilon$ of (36). The remaining two solutions are

$$\gamma = -\tfrac{1}{4}(n+1)\epsilon, \quad (46a)$$

$$\beta = \tfrac{1}{2}(n+1), \quad (46b)$$

$$\alpha = -\tfrac{1}{4}(n+1)(\epsilon^2+1)/\epsilon, \quad (46c)$$

and

$$\gamma = (n+1)(\epsilon^2/(\epsilon-1)^2), \quad (47a)$$

$$\beta = (n+1)(\epsilon^2+1)/(\epsilon-1)^2, \quad (47b)$$

$$\alpha = -2(n+1)\epsilon/(\epsilon-1)^2. \quad (47c)$$

The quantities $\tilde{A}_1 = -4\gamma^2(\epsilon+1)^3/\epsilon^3 - 2L - 2\alpha\beta + 4\gamma\beta - 4\gamma + 2\beta^2 - 2\beta - 4\alpha\gamma$ and $\tilde{A}_2 = -6\gamma^2(\epsilon+1)^2/\epsilon^2 - L + \beta^2 - \beta - 2\alpha\gamma + 4\gamma^2 + 6\gamma(\beta-1)$ which occur in (44c) and (44d) are known from (46) or (47) in terms of ϵ and L. In order to solve the equations (44c) and (44d) we may consider (44c) as $(n+1)$ linear homogeneous equations for the vector **a** and the necessary condition of their vanishing determinant to determine ϵ in terms of L. Equation (44d) is then either a constraint with determines L (and hence ϵ) or it is linearly dependent on (44c) and thereby leaves L undetermined. For each n, $0\leqslant n\leqslant 10$, we have found the latter possibility to be the case and have carried out the procedure just described to obtain inequivalent solutions to (44c) and (44d). We therefore claim, without proving it, that if (44a), (44b), and (45) hold then (44c) and (44d) have inequivalent solutions labelled by unrestricted values of L and nonnegative integer n. [It should be emphasized that the closed form solutions (38) which have been found here hold, at each chosen value of L, for particular values of the black hole parameters $m\omega$ and e.]

The case $n=0$ gives interesting and representative results. For $n=0$, we have $\tilde{A}_1 = 2\tilde{A}_2$ and hence (44c) and (44d), $\tilde{A}_1 = \tilde{A}_2 = 0$, are dependent and L is unrestricted (except that $L\neq 0$). The complete solution in the case of (47) may then be expressed as $\gamma = \frac{1}{4}L^2$, $\alpha+\beta = 1$, $\beta-\alpha = (L+1)^2$, $m^2\omega^2 = -\frac{1}{16}L^2(L+2)^2$, and $\epsilon = L/(L+2)$. Thus we have an infinite set of solutions with nonsingular $S(\theta)$ and physical values of the black hole parameters since for every integer $l\geqslant 1$ we have $\epsilon<1$. Note that, for all integer $l\geqslant 1$, each of the quantities α, β, γ, and $m^2\omega^2$ is an integer, and $\frac{1}{2}(\alpha+\beta)$ and $\frac{1}{2}(\beta-\alpha)$ are both positive half-integers. In this case (and for all n and physical ϵ) $m\omega$ is pure imaginary.

When $n=0$ in (46) we find $4L = -\epsilon^2 - 3$ and $m^2\omega^2 = -\frac{1}{16}$. In this case then, for every integer $l\geqslant 0$ we have a solution with $e^2/m^2>1$, but for $e^2/m^2<1$ the parameter l is complex and $S(\theta)$ is singular.

In the case $n=1$, it is easy to check that (44d) is dependent on (44c), and **a** is easily found from (44c). The condition of vanishing determinant which relates ϵ to L is, in the case of (46), $(\epsilon^2+L+2)^2 = 2\epsilon^2+1$ and in the case of (47) is $(\epsilon-1)^2L^2 + 2(\epsilon-1)(7\epsilon+1)L + 8\epsilon(4\epsilon+1) = 0$. From these relations it may be seen that the nature of the possible values of the black hole parameters is similar to the case of $n=0$.

For increasing n it becomes tedious to write the polynomial in ϵ and L which results from the determinant condition and to check whether (44d) is a dependent equation. We have therefore analyzed the cases $2\leqslant n\leqslant 10$ on a computer. At each of these values of n we have found (44d) to be linearly dependent on (44c) and have found solutions for any chosen value of L. We conclude that L remains unrestricted and infinite sets of solutions exist at each n in each of the cases (46) and (47).

B. Asymptotic behavior on the Reissner–Nordström geometry

Let ψ be given by (38); α, β, and γ by (47); and $\omega_\pm$ by $\omega_\pm \equiv \pm i(n+1)\epsilon/m(\epsilon-1)^2$. Then a solution to (36), linearly independent of ψ, is given by

$$\psi_1 = \psi\int_{m_-}^{r} |r'-m_+|^{\beta-\alpha-1}|r'-m_-|^{\alpha-\beta-1} \times e^{(2\gamma/m\epsilon)r'}V^{-2}\,dr'. \quad (48)$$

We may take (38) and (48) to be valid in both interior and exterior regions. We thus have four fields, denoted by $e^{i\omega\cdot t}(\psi,\psi_1)P_l^{m'}$, which are defined in each of the regions I, II, and III (shown in Fig. 1) belonging to the maximally extended RNG.[19] (Linear combinations of these fields are, of course, also solutions to the wave equation.) We examine the behavior of $e^{i\omega\cdot t}(\psi,\psi_1)$ in I, II, and III, and we choose to express the results in terms of null coordinates $2u = t - r^*$ and $2v = t + r^*$ in the various regions, where $r^* = r + (m_+^2/2m\epsilon)\ln|r-m_+| - (m_-^2/2m\epsilon)\ln|r-m_-|$. The behavior of the fields with $\omega = \omega_-$ may be obtained by time reversal.

At $I^+, I^-, H^+, H^-, H_{II}^+, H_{II}^{+\prime}, H_{II}^-, H_{III}^+$, and $r=0$ we find, respectively, the following limits of $e^{i\omega\cdot t}(\psi,\psi_1)$: $(0,e^{-2|\omega|u})$, $(0,\infty)$, $(e^{-2|\omega|v},0)$, $(\infty,e^{-2|\omega|u})$, (∞,∞), $e^{-2|\omega|v_{II}}(1,1)$, $(\infty,0)$, $e^{-2|\omega|u_{III}}(1,1)$, and $e^{-2|\omega|v}(1,1)$. We have ignored constant nonzero factors in writing these limits.

Different limits may be obtained for special linear combinations, but in no case are all the limits finite.

A situation with no infinities can be defined by supposing the external RNG to be caused by a static spherical body of radius greater than m_+. Then $e^{i\omega\cdot t}\psi$ provides an infinite class of solutions to the wave equation on the external geometry which are all finite everywhere to the future of any spacelike surface $t=$ const.

C. Solutions on the Schwarzschild geometry

The construction of solutions to the radial equation of a $s=0$ field on SG is obtained by setting $\epsilon = 1$ everywhere before (47) in Sec. IV A. Note in particular that

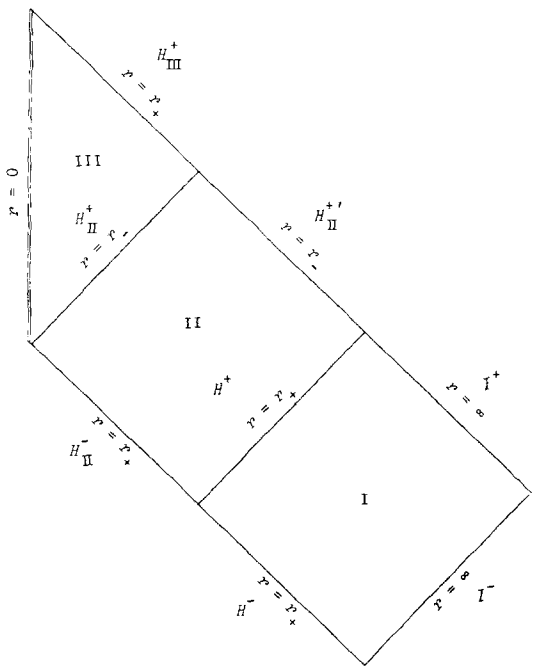


FIG. 1. A portion of the maximally extended Reissner-Nordström geometry showing horizons H and future and past null infinity $I^{\pm}$.

$\alpha + \beta = 0$ in (38). The only possible forms of α, β, and γ are those given by (46) with $\epsilon = 1$, and we claim that with these forms, (44c) and (44d) with $\epsilon = 1$ are satisfied at each n by $n + 1$ values of L. In all the calculated cases we have $L \leqslant -1$ and this is expected to hold for all n.

For the case $n = 0$ and $m' = 0$ we have $L = -1$ and so $l = -\frac{1}{2} \pm i(\sqrt{3})/2$; then $P_l(\cos\theta)$ is a conical function[20] so that the solution is for a half-line singularity at $\theta = \pi$ along the polar axis. The function ψ is simply $e^{r/4m}/|r - 2m|^{1/2}$.

V. OTHER FIELDS

For a massive scalar field of mass μ on RNG or SG the radial equation still has the form of (28) and on CMG or AMG it can be related to (4). Hence the previous methods will generate solutions for the massive field. It appears, however, that the inclusion of μ does not allow new parameter values. For example, for RNG with $n = 0$, it can be shown that it is impossible for all of the following to hold simultaneously: $m^2\mu^2 > 0$, $m^2\omega^2 > 0$, ϵ real, and $l =$ nonnegative integer.

For fields of spin 1 or 2, solutions analogous to those of Sec. III and IV can easily be found for low values of n. (For the Maxwell field on SG, for example, a solution for the radial part of the Newman–Penrose[21] quantity ϕ_2 is $\psi = [\sinh(r/2m - 1)]/(r/2m - 1) - \cosh(r/2m - 1)$ and $\omega = i$, $l = 0$. In this case $\psi/(r - 2m)^{1/2}$ is a modified Bessel function of order 3/2). Furthermore, these solutions have a property in common with those of (39); namely, their particular parameter values reduce the order of certain regular singular points of an appropriate differential equation. This happens in (39) because the parameter conditions (43), (44a), and (45) cause $\tilde{A}$ to have a factor of $(y - 1)(y + 1)$. These facts suggest that results analogous to those of Secs. III and IV may hold for nonzero spin fields.

ACKNOWLEDGMENT

The author wishes to thank R. J. Torrence for several helpful discussions.

[1] J. M. Cohen and L. S. Kegeles, Phys. Lett. A **54**, 5 (1975).
[2] S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).
[3] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[4] S. A. Teukolsky and W. H. Press, Ap. J. **193**, 443 (1974).
[5] S. Persides, J. Math. Phys. **14**, 1017 (1973).
[6] R. F. Arenstorf, J. M. Cohen, and L. S. Kegeles, J. Math. Phys. **19**, 833 (1978).
[7] S. Persides, Commun. Math. Phys. **48**, 165 (1976); **50**, 229 (1976).
[8] See, for example, A. A. Starobinskii and S. M. Churilov, Zh. Eksp. Teor. Fiz. **65**, 3 (1973) [Sov. Phys.-JETP **38**, 1 (1974)].
[9] See, for example, Ref. 4.
[10] J. B. Hartle and D. C. Wilkins, Commun. Math. Phys. **38**, 47 (1974).
[11] J. M. Stewart, Proc. R. Soc. London A **344**, 65 (1975).
[12] I. M. Arscott, Proc. Roy. Soc. Edinburgh A **67**, 265 (1967).
[13] K. M. Urwin and I. M. Arscott, Proc. Roy. Soc. Edinburgh A **69**, 28 (1970).
[14] I. M. Arscott, *Periodic Differential Equations* (MacMillan, New York, 1964).
[15] R. G. Newton, *The Complex j-Plane* (Benjamin, New York, 1964).
[16] For a recent, somewhat novel use of the static fields see R. J. Torrence, "Characteristic Initial Value Problem in Reissner–Nordström Exterior Regions," preprint, 1980.
[17] Much of the interest in Hill's equation has concerned its periodic solutions. We will not be interested in this aspect but rather in how certain of its solutions are constructed. The analogy between (28) and (4) is in the construction of special solutions.
[18] W. Magnus and S. Winkler, *Hill's Equation* (Dover, New York, 1979), p. 110.
[19] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[20] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), pp. 337–9.
[21] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

Linear transport in nonhomogeneous media.I [a)]

Edward W. Larsen
Theoretical Division, University of California, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 30 December 1980; accepted for publication 6 February 1981)

A set of elementary solutions of the linear transport equation is constructed for a class of scattering ratios which vary continuously in space. These solutions are shown to be complete on the half-range $0<\mu\leqslant 1$ for a restricted class of scattering ratios.

PACS numbers: 05.60. + w, 42.68.Db

I. INTRODUCTION

One of the most recent developments in transport theory is the emergence of analytical methods for treating one-dimensional transport problems in media whose cross sections vary continuously with position.[1-9] In this article we continue this line of research for halfspace transport problems of the form

$$\mu\frac{\partial}{\partial x}\Psi(x,\mu)+\Psi(x,\mu)=c\frac{f(x)}{2}\int_{-1}^{1}\Psi(x,\mu')d\mu',\quad 0<x<\infty, \tag{1.1}$$

$$\Psi(0,\mu)=g(\mu),\quad 0<\mu\leqslant 1, \tag{1.2}$$

$$\lim_{x\to\infty}\Psi(x,\mu)=0. \tag{1.3}$$

We require the scattering ratio, written as $cf(x)$, to be non-negative and the half-space $0<x<\infty$ to be subcritical. (Thus the above problem has a unique solution for all values of c between the given value and zero.) The function $f(x)$ has a quite general form described below. However, to introduce notation and motivate the analysis, we first state some known results.

For the well-studied[10] case of a homogeneous medium,

$$f(x)=1,$$

the set of continuum solutions satisfying Eqs. (1.1) and (1.3) can be written

$$\Psi_\nu(x,\mu)=\delta(\nu-\mu)e^{-x/\nu}+c\gamma_\nu(\mu)e^{-x/\nu},\quad 0<\nu<1, \tag{1.4}$$

$$\gamma_\nu(\mu)=\frac{\nu}{2}\left[\frac{1}{\nu-\mu}-\delta(\nu-\mu)\ln\left(\frac{1+\nu}{1-\nu}\right)\right]. \tag{1.5}$$

The distribution γ, which plays a key role in our analysis, satisfies

$$\int_{-1}^{1}\gamma_\nu(\mu)d\mu=0 \tag{1.6}$$

and

$$(1-\mu/\nu)\gamma_\nu(\mu)=\tfrac{1}{2}. \tag{1.7}$$

The set of solutions (1.4) is not complete on the half-range $0<\mu\leqslant 1$; a discrete solution must be appended to the set in order to satisfy Eq. (1.2).

For the case of an exponential medium,

$$f(x)=e^{-x/s},\quad 0<s<\infty,$$

an analogous set of continuum solutions has been constructed by Mullikin and Siewert[2]:

$$\Psi_\nu(x,\mu)=\delta(\nu-\mu)e^{-x/\nu}+c\gamma_\omega(\mu)e^{-x/\omega},\quad 0<\nu<1, \tag{1.8}$$

$$1/\omega=1/s+1/\nu. \tag{1.9}$$

These solutions are complete on the half-range $0<\mu\leqslant 1$ for any finite value of s, provided only that the half-space $0<x<\infty$ is subcritical.[6] Also, these solutions carry over to the previous ones in the limit $s\to\infty$.

In this article, we take $f(x)$ to have the general form

$$f(x)=\int_0^\infty\alpha(t)e^{-x/t}dt, \tag{1.10}$$

and in Sec. II we show that the functions

$$\Psi_\nu(x,\mu)=\delta(\nu-\mu)e^{-x/\nu}+c\int_0^\infty\alpha(t)\gamma_\omega(\mu)e^{-x/\omega}dt$$

$$0<\nu<1, \tag{1.11}$$

$$1/\omega=1/t+1/\nu, \tag{1.12}$$

are a set of elementary solutions of Eqs. (1.1), (1.3). In Sec. III we formulate a simple necessary and sufficient condition on the parameter c for the half space $0<x<\infty$ to be subcritical, and we show that if the half space is subcritical and α satisfies

$$\int_0^\infty|t\alpha(t)|^2\,dt<\infty, \tag{1.13}$$

then the set (1.11), (1.12) is complete on the half-range $0<\mu\leqslant 1$. Finally, in Sec. IV, we discuss and relate our results to the work of Mullikin and Siewert[2] and Kelley and Mullikin,[8] who have considered a different aspect of the problem (1.1)–(1.3), (1.10).

II. CONSTRUCTION OF THE ELEMENTARY SOLUTIONS

For $f(x)=1$ and $f(x)=\exp(-x/s)$, elementary solutions of Eq. (1.1) are given respectively by Eqs. (1.4) and (1.8). The fact that both of these sets of solutions are linear in the parameter c suggests that we seek solutions of Eq. (1.1), with $f(x)$, defined by Eq. (1.10), which are linear in c:

$$\Psi_\nu(x,\mu,c)=g_\nu(x,\mu)+ch_\nu(x,\mu). \tag{2.1}$$

Introducing Eq. (2.1) into Eq. (1.1) gives

$$\mu\frac{\partial}{\partial x}g_\nu(x,\mu)+g_\nu(x,\mu)=0, \tag{2.2}$$

$$\mu\frac{\partial}{\partial x}h_\nu(x,\mu)+h_\nu(x,\mu)=\frac{f(x)}{2}\int_{-1}^{1}g_\nu(x,\mu')d\mu', \tag{2.3}$$

[a)]This work was performed under the auspices of the U. S. Department of Energy.

 0022-2488/81/071463-04$1.00

$$0 = \int_{-1}^{1} h_\nu(x,\mu')d\mu'. \tag{2.4}$$

If we choose

$$g_\nu(x,\mu) = \delta(\nu-\mu)e^{-x/\nu},\ 0<\nu<1,$$

then Eq. (2.2) is satisfied and Eq. (2.3) becomes

$$\mu\frac{\partial}{\partial x}h_\nu(x,\mu) + h_\nu(x,\mu) = \tfrac{1}{2}\int_0^\infty \alpha(t)e^{-x/\omega}dt, \tag{2.5}$$

where

$$1/\omega = 1/t + 1/\nu. \tag{2.6}$$

To solve Eq. (2.5), we take

$$h_\nu(x,\mu) = \int_0^\infty \beta(t,\mu,\nu)e^{-x/\omega}dt,$$

and then Eqs. (2.5) and (2.4) reduce to

$$(1-\mu/\omega)\beta(t,\mu,\nu) = \tfrac{1}{2}\alpha(t),$$

$$0 = \int_{-1}^{1}\beta(t,\mu')d\mu'.$$

The solution of these equations is easily found to be

$$\beta(t,\mu,\nu) = \alpha(t)\gamma_\omega(\mu).$$

Thus, we obtain

$$\Psi_\nu(x,\mu,c) = \delta(\nu-\mu)e^{-x/\nu} + c\int_0^\infty \alpha(t)\gamma_\omega(\mu)e^{-x/\omega}dt,$$

which is just Eq. (1.11).

III. HALF-RANGE COMPLETENESS

The general solution of the transport equation (1.1), which is constructable from the solutions (1.11), and which is bounded at $x = +\infty$, is

$$\Psi(x,\mu) = a(\mu)e^{-x/\mu} + c\int_0^1 a(\nu)\int_0^\infty \alpha(t)\gamma_\omega(\mu)e^{-x/\omega}dt\,d\nu, \tag{3.1}$$

where

$$a(\mu) = 0,\ -1\leqslant\mu<0.$$

To solve the full transport problem (1.1)–(1.3), the boundary condition (1.2) must be satisfied. Using Eq. (3.1) and

$$t = \omega\nu/(\nu-\omega),$$

this condition can be written

$$g(\mu) = a(\mu) + c\int_0^1 a(\nu)\int_0^\nu \alpha(t)\gamma_\omega(\mu)[\nu/(\nu-\omega)]d\omega\,d\nu,$$

$$0<\mu\leqslant 1,$$

or

$$g(\mu) = (I - cL)\,a(\mu), \tag{3.2}$$

where I is the identity operator and L is the integral operator defined by

$$L = L_1 + L_2,$$

$$L_n a(\mu) = \int_0^1 a(\nu)G_n(\nu,\mu)d\nu,\ n = 1,2,$$

$$G_1(\nu,\mu) = \int_0^\nu \alpha(t)\left(\frac{\nu}{\nu-\omega}\right)^2\frac{\omega}{2}\frac{d\omega}{\omega-\mu},$$

and

$$G_2(\nu,\mu) = -\alpha\left(\frac{\mu\nu}{\nu-\mu}\right)\left(\frac{\nu}{\nu-\mu}\right)^2\frac{\mu}{2}\ln\left(\frac{1+\mu}{1-\mu}\right),\ \mu<\nu,$$

$$G_2(\nu,\mu) = 0,\ \nu<\mu.$$

Let us suppose that L_1 and L_2 are completely continuous operators in an suitable function space. Then,[11] either

(i) Eq. (3.2) has a unique solution $a(\mu)$

or

(ii) Eq. (3.2) has a nonzero solution $a(\mu)$ for $g(\mu) = 0$.

If case (ii) were to hold, the half-space $0<x<\infty$ would have a nonzero flux Ψ corresponding to a zero incident flux g, and then the half-space would be critical. However, we have assumed that the half-space is subcritical, and therefore case (ii) is impossible and case (i) must hold. It follows by definition that the set of solutions (1.11) is complete on the half-range $0<\mu\leqslant 1$.

A sufficient condition for the operators L_1 and L_2 to be completely continuous on $L_2(0,1)$ is[12]

$$\int_0^1\int_0^1 |G_n(\nu,\mu)|^2\,d\mu\,d\nu<\infty,\ n = 1,2.$$

Since[13]

$$\int_0^1 |G_1(\nu,\mu)|^2d\mu < \int_{-\infty}^{+\infty}|G_1(\nu,\mu)|^2d\mu$$

$$= \pi^2\int_0^\nu\left|\alpha(t)\left(\frac{\nu}{\nu-\omega}\right)^2\frac{\omega}{2}\right|^2d\omega$$

$$= \frac{\pi^2}{4}\int_0^\nu\left|\alpha(t)\frac{\nu\omega}{\nu-\omega}\right|^2\left(\frac{\nu}{\nu-\omega}\right)^2d\omega$$

$$= \frac{\pi^2}{4}\int_0^\infty |\alpha(t)t|^2dt$$

and

$$\int_0^1 |G_2(\nu,\mu)|^2d\mu$$

$$< \tfrac{1}{4}\int_0^\nu\left|\alpha\left(\frac{\nu\mu}{\nu-\mu}\right)\frac{\mu\nu}{\nu-\mu}\ln\frac{1+\mu}{1-\mu}\right|^2\left(\frac{\nu}{\nu-\mu}\right)^2d\mu$$

$$= \tfrac{1}{4}\int_0^\infty\left|\alpha(s)s\ln\left(\frac{1+s\nu/(s+\nu)}{1-s\nu/(s+\nu)}\right)\right|^2ds$$

$$< \tfrac{1}{4}\ln(1+2\nu)\int_0^\infty |\alpha(s)s|^2ds,$$

then L_1 and L_2 are completely continuous if α satisfies

$$\int_0^\infty |t\alpha(t)|^2dt<\infty. \tag{3.3}$$

To summarize, if the half-space $0<x<\infty$ is subcritical and α satisfies Eq. (1.13), then the set of elementary solutions (1.11) is complete on the half-range $0<\mu\leqslant 1$ in the function space $L_2(0,1)$.

The above remarks imply that if α satisfies Eq. (1.13), then Eq. (3.2) has a unique solution for all values of c such that

$$c\ \mathrm{spr}(L) < 1 \tag{3.4}$$

(where spr denotes spectral radius), and, if

$$c\ \mathrm{spr}(L) = 1, \tag{3.5}$$

then Eq. (3.2) has a nonzero solution for $g = 0$. Therefore, we

can characterize the criticality of the half-space $0<x<\infty$ for functions α satisfying Eq. (1.13) as follows: the half-space is subcritical if Eq. (3.4) holds, is critical if Eq. (3.5) holds, and is supercritical if

$$c\ \mathrm{spr}(L)>1$$

holds.

To conclude this section, we note that if α satifies Eq. (1.13) and $a(\nu)\in L_2(0,1)$, then each term in Eq. (3.1) can be interpreted in $L_2(-1,1)$. This follows from a simple extension of the analysis either in this section or in Ref. 14.

IV. DISCUSSION

The class of scattering ratios which can be represented by Eq. (1.10) is very large. For example, an explicit integration gives

$$\int_0^\infty \frac{e^{-\lambda/t}}{t^2}e^{-x/t}dt=\frac{1}{\lambda+x}. \tag{4.1}$$

(This representation is valid for all $0\leqslant x<\infty$ provided $\mathrm{Re}\lambda>0$, whereas if $\mathrm{Re}\lambda\leqslant 0$, the representation converges only for $-\mathrm{Re}\lambda<x<\infty$. Thus in this discussion we must require $\mathrm{Re}\lambda>0$.) Equation (4.1) can also be integrated and differentiated with respect to λ. For example,

$$\int_0^\infty \frac{e^{-\lambda_0/t}-e^{-\lambda/t}}{t}e^{-x/t}dt=\ln\left(\frac{\lambda+x}{\lambda_0+x}\right) \tag{4.2}$$

and

$$\int_0^\infty \frac{e^{-\lambda/t}}{t^{n+2}}e^{-x/t}dt=\frac{n!}{(\lambda+x)^{n+1}},\quad n=0,1,\cdots. \tag{4.3}$$

If λ is real, then the above functions are all real. However, if λ is complex, then real functions can be generated by adding the complex conjugate or subtracting the complex conjugate and divding by i. By taking linear combinations of such functions, with possibly differing values of λ, one can obtain representations of more complicated functions. For example, any rational function which vanishes at ∞ and has poles only in the left half of the complex plane can be represented in this manner. Moreover, Eqs. (4.1)–(4.3) indicate that the corresponding functions $\alpha(t)$ will satisfy Eq. (1.13), and thus the elementary solutions defined by Eq. (1.11) will be half-range complete.

Other explicit representations can be generated by taking α to be a distribution, but care must be taken to properly interpret the integral term in Eq. (1.11). For example, if

$$\alpha(t)=\delta(t-s),\quad 0<s<\infty,$$

then Eq. (1.10) becomes

$$f(x)=e^{-x/s},$$

and Eq. (1.11) gives

$$\Psi_\nu(x,\mu)=\delta(\nu-\mu)e^{-x/\nu}+c\gamma_\omega(\mu)e^{-x/\omega},$$

where

$$1/\omega=1/s+1/\nu.$$

Since $\alpha(t)$ does not satisfy Eq. (1.13), the analysis of Sec. III does not predict that the above elementary solutions are complete on the half-range $0<\mu\leqslant 1$. However, a more detailed analysis does establish this completeness.[6]

The solutions in the above paragraph were derived by Mullikin and Siewert,[2] who used them not to develop solutions of the transport problem (1.1)–(1.3), but rather to derive an equation which the flux exiting the half-space must satisfy. To make contact with their results, we shall sketch this derivation. One assumes that the transport problem solution can be represented by

$$\Psi(x,\mu)=\int_0^1 a(\nu)\Psi_\nu(x,\mu)d\nu \tag{4.4}$$

and shows that the elementary solutions satisfy

$$\int_{-1}^1 \mu\Psi_{\nu'}(0,-\mu)\Psi_\nu(0,\mu)d\mu=0,\quad 0<\nu,\nu'<1.$$

Then one sets $x=0$ in Eq. (4.4), multiplies by $\mu\,\Psi_{\nu'}(0,-\mu)$, and integrates over μ to get

$$\int_0^1 \mu\Psi_\nu(0,\mu)\Psi(0,-\mu)d\mu$$

$$=\int_0^1 \mu\Psi_\nu(0,-\mu)g(\mu)d\mu,\quad 0<\nu<1,$$

which is a singular integral equation for the exiting flux $\Psi(0,-\mu)$ for $0<\mu\leqslant 1$.

This procedure can be carried out for the elementary solutions defined by Eq. (1.11). The result is

$$S(\nu)+c\int_0^1 S(\mu)\int_0^\infty \alpha(t)\gamma_\omega(\mu)dt\,d\mu$$

$$=c\int_0^1 \mu g(\mu)\int_0^\infty \alpha(t)\gamma_\omega(-\mu)dt\,d\mu,\quad 0<\nu<1, \tag{4.5}$$

where

$$S(\mu)=\mu\Psi(0,-\mu)$$

and

$$1/\omega=1/t+1/\nu. \tag{4.6}$$

Equation (4.5) can be rewritten, after a lengthy amount of algebra, in the form

$$S(\nu)-\frac{c}{2}\int_0^\infty \alpha(t)\omega\int_0^1\left[\frac{S(\omega)-S(\mu)}{\omega-\mu}+\frac{S(\omega)}{\omega+\mu}\right]d\mu\,dt$$

$$=\frac{c}{2}\int_0^1 \mu g(\mu)\int_0^\infty \alpha(t)\frac{\omega}{\omega+\mu}dt\,d\mu,\quad 0<\nu<1. \tag{4.7}$$

Equation (4.7) was also derived by Mullikin and Siewert,[2] although in a completely different manner from that described above because the elementary solutions of Eq. (1.11) were not known to these authors. Kelley and Mullikin[8] have recently proved that Eq. (4.7) has a unique solution for all functions $\alpha(t)$ such that $f(x)\in L_2(0,\infty)$, assuming c to be small enough that the half space $0<x<\infty$ is subcritical. Thus the Kelley–Mullikin analysis shows that Eq. (4.7) has a unique solution for $\alpha(t)=\delta(t-s)$ [which gives $f(x)=e^{-x/s}$, whereas the half-range completeness analysis in Sec. III of this article does not apply since α does not satisfy Eq. (1.13). In addition, Kelley has recently shown[15] that

$$\int_0^\infty f^2(x)\,dx\leqslant\pi\int_0^\infty |t\,\alpha(t)|^2\,dt,$$

and therefore the set of functions f discussed in this article is a strict subset of the set discussed by Kelley and Mullikin.

[1]B. J. Martin, "A New Singular Integral Equation in Radiative Transfer," SIAM J. Appl. Math. **20**, 703 (1971).
[2]T. W. Mullikin and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres," Ann. Nucl. Energy **7**, 205 (1980).
[3]G. C. Pomraning and E. W. Larsen, "Elementary solutions of the linear transport equation for continuously varying spatial media," J. Math Phys. **21**, 1603 (1980).
[4]E. W. Larsen, G. C. Pomraning, and V. C. Badham, "On the singular eigenfunctions for linear transport in an exponential atmosphere," J. Math. Phys. **21**, 2448 (1980).
[5]T. W. Mullikin, "Some Singular Integral Equations in Linear Transport Theory," Sandia National Laboratories Report SAND 80-1069 (1980).
[6]E. W. Larsen and T. W. Mullikin, "Linear transport in an exponential atmosphere," J. Math. Phys. (to appear).
[7]R. D. M. Garcia and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres—Numerical Results," J. Quant. Spectrosc. Radiat. Transfer **25**, 227 (1981).
[8]C. T. Kelley and T. W. Mullikin, "Collocation Methods for Some Singular Integral Equations in Linear Transport Theory," submitted for publication (1980).
[9]G. C. Pomraning, "The Criticality Problem for an Exponential Atmosphere," submitted for publication (1980).
[10]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[11]G. Bachman and L. Narici, *Functional Analysis* (Academic, New York, 1966)
[12]F. Smithies, "A Note on Completely Continuous Transformations," Ann. Math. **38**, 626 (1937).
[13]E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford U. P., London, 1950).
[14]E. W. Larsen, S. Sancaktar, and P. F. Zweifel, "Extension of the Case formulas to L_p. Application to Half and Full Space Problems," J. Math. Phys. **16**, 1117 (1975).
[15]C. T. Kelley (private communication).

Entropy in classical and quantum physics

John F. Cyranski
Theoretical and Physical Chemistry Institute, The National Hellenic Research Foundation, 48, Vassileos Constantinou Avenue, Athens, Greece

(Received 13 September 1978; accepted for publication 28 April 1980)

The enigma of "entropy" is reconsidered from the viewpoint of generalized information theory on a lattice generated from measurements that define the system. A small (incomplete) set of natural axioms for a global information measure is developed sufficiently to deduce as a special case a generalization of Segal's entropy on a W^*-algebra (classical and quantum). A simple relationship between monotonicity of entropy and a semigroup on $[0,\infty]$ (representing composibility of information) is presented. Various extensions of information-theoretic results are incidentally proven, including relations between regular composible informations (on an orthocomplemented complete lattice) and measures (on σ—ideals of the lattice).

PACS numbers: 05.70.Ce, 05.50. + q

I. INTRODUCTION

Although dating from the grimy dawn of the industrial age, entropy has remained one of the most controversial, esoteric, and enigmatic concepts of physics. Recently Wehrl[1] has reviewed the properties of various entropy expressions useful in physics. In addition to the numerous "entropies" he has collected, one stumbles across other formulas which are also dubbed "entropy."[2–5] This proliferation of "entropies" suggests the need for a comprehensive approach to the basic issue, namely: "What *is* entropy?"

Rather than dismiss all expressions that do not conform to "$-\Sigma p_i \ln p_i$," "$-\mathrm{Tr}\hat{\rho}\ln\hat{\rho}$," or even Boltzmann's "$-k\ln W$"—as some reviewers would happily have us do—we prefer to seek a fundamental meaning and axiomatic formulation of a quantity which, when suitably constrained, reduces to such special formulas. Indeed, the need for such an approach is evident from the difficulty of reconciling the second law of thermodynamics with Hamiltonian motion. This problem has led to various approaches to nonequilibrium statistical physics, including different forms of "coarse-graining"[1] (which effectively reject the second law for the *true* system "state," and only work for infinite systems in general[1]), "master equations" (e.g., those based on projections of "relevant" data[6]), semigroups,[7] and modifications of entropy itself.[8]

Of course, all concepts of "entropy" are ultimately related to probabilities (or "states"). This relationship in fact unites the mystery of entropy to the enigma of "probability." Good[9] has distinguished at least five different concepts of "probability." More recently, Fine[10] has compared the various concepts and ultimately concluded with the doubt that "probability" is even necessary for scientific endeavors. Most physicists are familiar only with the so-called "objective"probabilities—relative frequencies over infinite ensembles which, though by definition unmeasurable, are nevertheless considered to represent a physical property ("state"). Rayski[11] has shown how quantum mechanics paradoxes vanish if one relinquishes this interpretation. Elsewhere[12] we have argued how certain arguments against theories of "hidden variables" depend implicitly on the interpretation of probability, and thus cannot be convincing.

An extremely successful step in explaining both enigmas—"entropy" and "probability"—has been made by Jaynes.[13] In essence, Jaynes interpreted "entropy" as the "expected self-information" of a class of mutually exclusive and exhaustive "events," or propositions. Although the association of "entropy" with "information" (negative uncertainty) is neither original with Jaynes[14] nor without opposition,[15] it nevertheless provides an explanation for this hitherto misunderstood quantity. Simultaneously, Jaynes interpreted "probability" to mean the rational "degree of belief"[16] we assign to events (on a zero to one scale) based on available evidence, generally of the form of expectation values of observables. The choice of data used effectively defines the total evidence assumed relevant to the properties being estimated with the degree of belief. This condition is quantified by taking as degree of belief that measure which maximizes the "entropy", i.e., the average uncertainty, subject to the chosen evidence. In a single blow, Jaynes not only explained, simplified (no ensembles!), and generalized statistical thermodynamics, but also provided scientists and engineers with a new, (more or less) consistent technique of *inference* which has been successfully employed in many areas.[17]

Despite these successes, and in part because of them, we again find a need for an axiomatic explication of "entropy." For one thing, Jaynes's program is strictly constrained by the choice of entropy formula[18] most generally, the Gibbs formula classically, and the von Neumann expression in quantum mechanics). Although Shore and Johnson[19] have recently given an axiomatic derivation of the classical entropy in terms of inductive inference, there is no reason to assert that this specialized "entropy" is appropriate or even correct in general. Indeed, as we will elaborate below, most "real-life" situations do not admit the requisite idealizations for Borel algebras, and thus probability-related information measures. It thus may be not only desirable but essential to eliminate probability altogether. Because of this and certain other difficulties,[20] the MEF (maximum entropy formalism) has not been exploited as thoroughly as is possible.

In order to arrive at a fundamental understanding of the significance of "entropy" in physics, we believe it necessary

 0022-2488/81/071467-12$1.00

to reexamine the essence of physics as distilled in its lattice-theoretic formulation. It is convenient to visualize a physical theory as a data-processing algorithm: Empirical data are digested by the theory and predictions result. This is consistent with the view that theories model reality: They are not "explanatory" in the sense of being in one-to-one correspondence with aspects of reality.[21] The model is constructed by first postulating an "operational definition" of the objects under study in terms of the set of all "relevant" empirical relations that can be defined on the objects. This definition is then naturally imbedded in a complete lattice—the "theory," which is the mathematical context for calculations. We review these ideas in Sec. II, extending our earlier work.[22,23] Also, we present certain new results concerning composible informations.

Our approach to "entropy" is based on the idea of "local" informations on a lattice (of "propositions") first studied by Sallantin.[24] In Sec. III we define "entropy" to be a *global* measure of information associated with possibly several "local" informations on the "theory." We do not attempt to characterize our "entropies,"[25] but rather consider certain special cases that ultimately reduce, as desired, to the usual "entropies." In Sec. IV we provide an interpretation of certain qualitatively novel features of our "entropies," which appears to have bearing on irreversible dynamics. Finally, we conclude in Sec. V with a general commentary and some open problems.

II. ABSTRACT THEORIES AND INFORMATIONS

The essence of physics is generally considered to be one or the other of various logico-algebraic models[26] based on abstract lattice theory.[27] While physicists are concerned specifically with deriving quantum and classical physics, we have found a broader viewpoint to be more enlightening. In Refs. 22 and 23 we have proposed a somewhat unusual approach which we describe and extend below. Our approach is reminiscent of the constructive descriptions of logic and language by Watanabe[28] and Sallantin[29]; however, we do not attempt to derive specific theories, but rather describe universal properties. We assume the reader to be familiar with standard terminology of lattice theory.

Our starting point is the obvious one: How do we *define* an object about which we wish to "talk" (i.e., discuss in a literal sense, or scientifically study)? This question is by no means trivial or metaphysical, since practical problems of computer recognition are involved. It is natural to define an object (actually, the class A of all such objects) by a complete list of its properties. Although for real-life objects such a list may not be possible, one can effectively define varying degrees of approximation to "real-life" by a judicious choice of the relevant properties. To account for the intrinsic "fuzziness" of "real-life" descriptions[30,31] we assert that the properties consist of relations $R:A^{n_R}\to Q$, where Q is a quasiordered set of "truth values." (This generalizes slightly the work of Goguen[32] and DeLuca and Termini.[33] Ali and Prugovecki[34] have applied "fuzzy" set theory directly to quantum physics, but not in the "logical" context.) To be operationally consistent we must insist that the order n_R of the relations be finite. The set of all "relevant" relations is denoted by $F:\langle A,Q,F\rangle$ constitutes an empirical relational system. Note that the relations are essentially generalized characteristic functions.

The relations in F are naturally ordered by an "implication," $\Rightarrow$, defined in terms of the quasi-ordering of the truth set Q. This binary (classical) relation induces a quasi-order in F; we denote by P the partially ordered set obtained from F by associating "$\Rightarrow$-equivalent" relations. Thus, P represents a minimal set of relations defining an object class A, with a built-in partial order. See Ref. 23 for details.

At this point we wish to clarify the above "abstract definition" by a further condition on P. Essentially, if R is a relation relevant to an object, it has a quasi-ordered range $\mathscr{R}(R)=\{R(\mathbf{a})\in A^{n_R}\}\subseteq Q$. If we reverse the order of $\mathscr{R}(R)$—which corresponds in the simplest case to interchanging binary (yes-no) responses on an elementary experimental apparatus[35]—we have a new relation, R', whose range is the dual of $\mathscr{R}(R)$. It is natural to allow for each $R\in P$ that $R'\in P$ as well, since the same apparatus or operation that defines R also defines R'. The map $R\to R'$ is an involution by construction:

$$R''=R;\quad R\leqslant S\Rightarrow S'\leqslant R'. \tag{1}$$

Notice, however, that there exists no meaningful relation in P which can in any sense "imply" *both* R and R'. That is

$$\{S\in P:\ S\leqslant R \text{ and } S\leqslant R'\}=\phi. \tag{2}$$

In this axiom, we diverge from the custom of introducing an "absurd" relation $\emptyset\in P$ such that $R\wedge R'=\emptyset$. Thus the "logic" of measurements is not assumed at this level.

So far the objects have been abstractly defined by the poset P with involution; this structure, constructively obtained, corresponds to the rational numbers. It is essential for further progress to introduce idealized elements—analogous to real numbers. We do this in the corresponding manner; we imbed P in a complete lattice $L(P)$ by the "cut procedure."[36] (This method is not alien to quantum mechanics[37]; however, the construction and properties of P used by previous authors are less general than ours.)

Theorem 1: If P is a poset with involution satisfying (2), then $L(P)$ is a complete orthocomplemented lattice.

Proof: Recall[27] that $\overline{E}=\{R\in P:\ R\leqslant S \text{ for all } S\geqslant T \text{ for all } T\in E\}$ defines the closure of $E\subseteq P$, and that $L(P)$ is just the set of closed subsets of P, a complete lattice under set inclusion ($\subseteq$). For $E\in L(P)$ define

$$E^{\perp}=\{R\in P:R\leqslant S' \text{ for all } S\in E\}. \tag{3}$$

Then, $(E^{\perp})^{\perp}=\{R\in P:R\leqslant S' \text{ for all } S'\geqslant T \text{ for all } T\in E\}=E$; $E\subseteq F$ implies $F^{\perp}\subseteq E^{\perp}$ since $R\in F^{\perp}$ implies $R\leqslant S'$ for all $S\in F$ implies $R\leqslant S'$ for all $S\in E$, and thus $R\in E^{\perp}$. Finally, (3) defines an involution since $E^{\perp}\in L(P)$: $\overline{E^{\perp}}=\{R\in P:R\leqslant S \text{ for all } S\geqslant T \text{ for all } T\leqslant W' \text{ for all } W\in E\}=\{R\in P:R\leqslant S \text{ for all } S'\leqslant T', \text{ for all } T'\geqslant W \text{ for all } W\in E\}=\{R\in P:\ R\leqslant S \text{ for all } S'\in\overline{E}=E\}=E^{\perp}$. Now $E\cap E^{\perp}=\{R\in P:R\leqslant S \text{ and } R\leqslant S', \text{ for all } S\in E\}=\phi$ by (2). Thus, $E\to E^{\perp}$ is an orthocomplement.[38] Q.E.D.

The importance of the universality of orthocomplemented complete lattices ($L(P)$) as "theories" lies in the ability we obtain to construct partitions. If $A\in L(P)$, a (finite) *par-*

tition of A is the set

$$\pi(A) = \{E_n \in L(P) : A = \vee_{n=1}^{N} E_n, E_n \perp E_m\} \quad (E \perp F \text{ iff } E \subset F^{\perp}). \tag{4}$$

We denote by Π the set of all (finite) partitions of P, and $\pi(P) = \pi$ for short. The set Π is partially ordered by the relation

$$\pi_1 < \pi_2 \text{ iff for each } E \in \pi_1, \quad \pi(E) \subseteq \pi_2 . \tag{5}$$

If $\pi_1 < \pi_2$, then π_2 is a "finer" partition; note that $\pi_0 = \{\phi, P\}$ is the universal lower bound, and that unless $L(P)$ is finite, Π has no upper bound in general.

We digress briefly to interpret $L(P)$ and Π. Since $R \in P$ is a relation, $\theta(R) = \overline{\{R\}} = \{S \in P : S \leqslant R\}$ (closure of $\{R\}$) is the "proposition" or idealization consisting of all (nonequivalent) relations in P that "imply" R. Since $\overline{E} = \vee_{R \in E} \theta(R)$ (see Ref. 22), $\overline{E}$ is a set of relations plus their antecedents. The elements of $L(P)$ thus effectively summarize sets of properties. This ability is desirable in a theory—i.e., to combine consistently large numbers of experimental features. A partition $\pi \in \Pi$ is a subdivision of $L(P)$ into a finite set of "orthogonal" sets. It is natural to consider such a partition as an experiment whose outcomes are manifestations of the property. Each manifestation is a summary of more "elementary" measurements that "imply" relations comprising the manifestation. Such an interpretation does not disagree with typical physics usage, especially in regard to observables in physics (countable partitions!).[35] One may look upon the elements of P as "outcomes" and the elements of $L(P)$ as "events" in the language of probability theory.

Recall[27] that $\mathscr{T} \subseteq L(P)$ is a σ–ideal if (a) $E \in \mathscr{T}, X \in L(P)$ and $X \subset E$ implies $X \in \mathscr{T}$; and (b) $E_n \in \mathscr{T}$, $n = 1,2,...$, implies $\vee_{n=1}^{\infty} E_n \in \mathscr{T}$. Note that if $E, F \in \mathscr{T}$, $E \cap F^{\perp} \in \mathscr{T}$ as well, which suggests that any σ–ideal is analogous to a σ–ideal in Boolean set theory.[39] We thus are led to define a *generalized measure* as a mapping $\mu : \mathscr{T} \to \overline{\mathbb{R}^+} = [0, \infty]$ such that $\mathscr{T}$ is a σ–ideal, $\mu(\phi) = 0$, $\mu(E) = \mu(F) = 0$ implies $\mu(E \vee F) = 0$ and

$$\mu(\vee_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n), \quad E_n \perp E_m . \tag{6}$$

This definition is a natural extension of Piron's[35] definition 4.38, the distinction being that Piron defines μ on a "tribe": $\overline{\mathscr{T}}$ such that $E \in \overline{\mathscr{T}}$ implies $E^{\perp} \in \overline{\mathscr{T}}$ and $\overline{\mathscr{T}}$ countably closed. In special cases of interest, the σ–ideals reduce to tribes, and the concepts coincide.

Viewing the lattice $L(P)$ as an ordered set of propositions,[40] one naturally discovers certain minimal requirements for a local *information* measure[41] on $L(P)$. First, it is natural to demand that no proposition have negative information:

$$J: \quad L(P) \to \overline{\mathbb{R}^+} = [0, \infty] . \tag{7a}$$

Next, if $E \subset F$, then proposition E is more specific than F in that it contains fewer relations (from P). Thus, as one expects, the more specific proposition have the more informative value:

$$E \subset F \text{ implies } J(E) \geqslant J(F) . \tag{7b}$$

Finally, as P is maximally unspecific, it should have the minimal information value, while ϕ (being absurdly specific) should have undefined value:

$$J(\phi) = \infty; \quad J(P) = 0 . \tag{7c}$$

We have studied properties of such information—defined by (7)—in Ref. 22. In this paper we will restrict attention to the subclass of informations which are (weakly) σ–composible. An information is said to be *composible*[41] (weakly) if there exists a real-valued function F such that

$$J(A \vee B) = F[J(A), J(B)], \quad \text{when } A \perp B . \tag{8}$$

We are motivated largely by the consequences to consider informations which are composed by ROC's:

Definition 1[42]:

A *regular operation of composition* (ROC) is a function F satisfying:

$$F \overline{\mathbb{R}^{+2}} \to \overline{\mathbb{R}^+} \text{ is continuous}, \tag{9a}$$

$$F(x,y) = F(y,x), \tag{9b}$$

$$F[x,F(y,z)] = F[F(x,y),z], \tag{9c}$$

$$F(x,\infty) = x, \tag{9d}$$

$$x < y \text{ implies } F(x,z) \leqslant F(y,z) . \tag{9e}$$

In fact, any composition map must satisfy (9b), (9c), (9d), and (9e) on the set of values defined by

$$\Gamma_n(J) = \{(x_1,x_2,...x_n) : A_1, A_2, ...A_n \in L(P), A_i \perp A_j, x_j = J(A_j)\} \tag{10}$$

for the appropriate $n \geqslant 1$. Condition (9a) extends the domain of F from $\Gamma_2(J)$ to all $\overline{\mathbb{R}^{+2}}$, and furthermore makes F a topological semigroup on $\overline{\mathbb{R}^+}$. Since $J(\vee_{n=1}^{\infty} E_n) = \lim_{N \to \infty} F_N [J(E_1), J(E_2), ..., J(E_N)]$ exists, (where $F_1(x) = x$ and $F_N(x_1,...,x_N) = F[F_{N-1}(x_1,...,x_{N-1}), x_N]$), J is *σ–composible* under F. Incidentally, (9) implies that

$$F(x,y) \leqslant \inf(x,y) . \tag{11}$$

Among the remarkable properties of ROC's is the following characterization theorem[43]:

Theorem 2: Let Λ be a closed subset of $\overline{\mathbb{R}^+}$ so $\overline{\mathbb{R}^+} - \Lambda = \cup_{i \in I} (a_i, b_i)$—$I = \phi$, is finite, or countable. Let $\bar{\mu}_i \in (0, \infty]$, $\theta_i : [0, \bar{\mu}_i] \to [a_i, b_i]$ be strictly decreasing, continuous with $\theta_i(0) = b_i$, $\theta_i(\bar{\mu}_i) = a_i$. Then

$$F(x,y) = \begin{cases} \inf(x,y), & (x,y) \in \overline{\mathbb{R}^{+2}} - \cup_{i \in I} (a_i,b_i)^2 \\ \phi_i[\theta_i^{-1}(x) + \theta_i^{-1}(y)], & (x,y) \in [a_i,b_i]^2 \end{cases} \tag{12}$$

is a ROC with $\Lambda = \Lambda(F) = \{x \in \overline{\mathbb{R}^+} : F(x,x) = x\}$. Here

$$\phi_i(x) = \begin{cases} \theta_i(x), & x \in [0, \bar{\mu}_i] \\ a_i, & x > \bar{\mu}_i \end{cases} . \tag{13}$$

Conversely, any ROC is of form (12) with $\Lambda = \Lambda(F)$.

Perhaps yet more remarkable is the intimate association of ROC–composible informations with certain ideals and measures on ideals. Because $L(P)$ lacks the essential distributivity properties of a Borel algebra, the beautiful results of Ref. 42 apparently do not survive intact. For this reason we extend these results (as best we can) in the following somewhat unaesthetic theorems:

Theorem 3: Let $\Lambda \subseteq \overline{\mathbb{R}^+}$ be closed with $0, \infty \in \Lambda$, $\overline{\mathbb{R}^+} - \Lambda = \cup_{i\in I}(a_i,b_i)$. To each $x\in\Lambda$, associate a σ–ideal $\mathcal{T}_x$ such that for all $x>0$, $\mathcal{T}_x$ is proper, $\mathcal{T}_0 = L(P)$, and $x<y$ implies $\mathcal{T}_x \supseteq \mathcal{T}_y$. On each $\mathcal{T}_{a_i}$, $i\in I$, assume a generalized measure $\mu_i : \mathcal{T}_{a_i} \to [0,\bar{\mu}_i]$ such that $\bar{\mu}_i \in (0,\infty]$ and $\mu_i[\mathcal{T}_{a_i} \cap \mathcal{T}_{b_i}] = 0$. Let $\theta_i : [0,\bar{\mu}_i] \to [a_i,b_i]$ be any continuous, strictly decreasing function with $\theta_i(0) = b_i$ and $\theta_i(\bar{\mu}_i) = a_i$. Let $z_1(A) = \sup\{x\in\Lambda: \ A\in\mathcal{T}_x\}$ and $z_2(A) = \inf\{x\in\Lambda: \ A\notin\mathcal{T}_x\}$.

Then

$$J(A) = \begin{cases} z_1(A), & \text{if } z_1(A) = z_2(A) \\ \phi_i(\mu_i(A)), & \text{if } z_1(A) = a_i < b_i = z_2(A) \end{cases} \tag{14}$$

is a σ–composible information on $L(P)$.

Proof: Note that $J(A)\in[z_1(A),z_2(A)]$ by construction and that $(z_1(A),z_2(A))\cap\Lambda = \phi$ if $z_1(A) < z_2(A)$, since $(a_i,b_i)\cap\Lambda = \phi$. Clearly J is defined on all $A\in L(P)$. Since $\phi\in\mathcal{T}_x$ for all $x\in\Lambda$, $z_1(\phi) = \infty$ so $J(\phi) = \infty$; since $P\in\mathcal{T}_0$, but for all $x>0$, $P\notin\mathcal{T}_x$, then $z_1(P) = z_2(P) = 0$ and $J(P) = 0$, or else $J(P) = \phi_0(\bar{\mu}_0) = a_0 = 0$, if $z_1(P) < z_2(P)$. Now for all $x < z_1(B)$, $B\in\mathcal{T}_x$ and $A\subset B$ implies $A\in\mathcal{T}_x$, so $z_1(B)\leqslant z_1(A)$. If $J(A) < J(B)$, $z_1(A)\leqslant z_1(B)$, so $a_i = z_1(A) = z_1(B)\leqslant J(A) < J(B) \leqslant z_2(B) = z_2(A) = b_i$. But $\mu_i(A)\leqslant\mu_i(B)$ implies $\phi_i(\mu_i(A))\geqslant\phi_i(\mu_i(B))$, so $J(A)\geqslant J(B)$.

Now $A,B\in\mathcal{T}_x$ $(x\in\Lambda)$ iff $A\vee B\in\mathcal{T}_x$, so that $z_1(A\vee B) = \min\{z_1(A),z_1(B)\}\leqslant J(A\vee B)\leqslant\min\{J(A),J(B)\} \leqslant\min\{z_2(A),z_2(B)\}$. For the sake of definiteness, assume $z_1(A)\leqslant z_1(B)$ $(A\perp B)$ so the following cases are possible:

(1) $z_1(A) = z_1(B) = z_2(A) = a_i < z_2(B) = b_i$,

(2) $z_1(A) = z_1(B) = z_2(B) = a_i < z_2(A) = b_i$,

(3) $z_1(A) = z_1(B) = a_i < z_2(A) = z_2(B) = b_i$,

(4) $z_1(A) = a_i < z_1(B) = z_2(A) = z_2(B) = b_i$,

(5) $z_1(A) = z_1(B) = z_2(A) = z_2(B)$,

(6) $z_1(A)\leqslant z_2(A)\leqslant z_1(B)\leqslant z_2(B)$.

In Case (1), $J(A) = J(A\vee B) = a_i$, $J(B)\geqslant a_i$, so $J(A\vee B) = \inf[J(A),J(B)]$. In Case (2), $J(B) = J(A\vee B) = a_i$, $J(A)\geqslant a_i$, so $J(A\vee B) = \inf[J(A),J(B)]$. In Case (3), $J(A),J(B)$, $J(A\vee B)\in[a_i,b_i]$; $J(A\vee B) = \phi_i[\mu_i(A)+\mu_i(B)]$ and $J(A) = \theta_i(\mu_i(A))$, $J(B) = \theta_i(\mu_i(B))$ (definition of θ_i and μ_i), so $J(A\vee B) = \phi_i[\theta_i^{-1}J(A) + \theta_i^{-1}J(B)]$. In Case (4), $J(A)$, $J(A\vee B)\in[a_i,b_i]$, $J(B) = b_i$, so $J(A\vee B) = \phi_i[\mu_i(A)+0] = J(A) = \inf[J(A),J(B)]$. In Case (5), $x = J(A\vee B) = J(A) = J(B) = \inf[J(A),J(B)]$, where $x\notin(a_i,b_i)$. In Case (6), $J(A),J(A\vee B)\in[z_1(A),z_2(A)]$, $J(B)\geqslant z_2(A)$. Thus, either $z_1(A) = J(A) = J(A\vee B) = z_2(A)$, or $J(A\vee B) = \phi_i[\mu_i(A)+0] = J(A)$ if $a_i = z_1(A) < b_i = z_2(A)\leqslant J(B)$. Thus, $J(A\vee B) = F[J(A),J(B)]$ where F is defined by (12). Since by Theorem 2 F is a ROC, J is σ– composible. Q.E.D.

Comment: If $J(A) < x_0$, $z_1(A) < x_0$ implies $A\notin\mathcal{T}_{x_0}$. Thus, $A\in\mathcal{T}_{x_0}$ implies $J(A)\geqslant x_0$. If $A\in\mathcal{T}_{a_i}$ and $J(A)\in[a_i,b_i]$, by (14) $\mu_i(A) = \theta_i^{-1}J(A)$. Because $\mu_i(A)\leqslant\bar{\mu}_i$, and μ_i is σ–additive, $\Sigma_{J(E_n)\in[a_i,b_i)}\theta_i^{-1}J(E_n)\leqslant\bar{\mu}_i$ for every sequence of mutually orthogonal elements of $\mathcal{T}_{a_i}$.

In the following theorem, if $L(P)$ is orthomodular (15) is satisfied, while if $L(P)$ is distributive (16) is unnecessary.

Theorem 4: Let $L(P)$ be a complete orthocomplemented lattice such that

$A,B\in L(P)$ implies there exists $D\subseteq A$, $E\subseteq B$ such that

$$A\vee B = E\vee D,\ E\perp D. \tag{15}$$

Let $\Gamma_x = \{A\in L(P)$: There exists a countable partition $\pi_\infty(A)$ such that for all $E_n\in\pi_\infty(A)$, $J(E_n) > x\}$, where J is an information on $L(P)$ σ–composible under ROC F. Suppose the ROC is defined by (12), where

$$\bar{\mu}_i = \infty \quad \text{or else } J(A) = a_i \text{ implies } A\notin\Gamma_{a_i}, i\in I. \tag{16}$$

Then there exists a family of σ–ideals $\mathcal{T}_x$, $x\in\Lambda(F)$ such that

$$\mathcal{T}_0 = L(P)\supset\cdots\mathcal{T}_x\supset\cdots\mathcal{T}_y\supset\cdots\supset\mathcal{T}_\infty \quad (x<y) \tag{17}$$

and on each $\mathcal{T}_{a_i}$ there exists a measure $\mu_i : \mathcal{T}_{a_i} \to \overline{\mathbb{R}^+}$. The ideals and measures determine J by (14).

Proof: If $x\in\Lambda(F)$, define $\mathcal{T}_x = \{A\in L(P): J(A)\geqslant x\}$. If $A\in\mathcal{T}_x$, $X\subseteq A$ $(X\in L(P))$, then $J(X)\geqslant J(A)\geqslant x$ implies $X\in\mathcal{T}_x$. By (15) and (9), if $A,B\in\mathcal{T}_x$, $J(A\vee B) = J(E\vee D) = F[J(E),J(D)]\geqslant F(x,x) = x$, so by induction, $E_n\in\mathcal{T}_x$ implie $J(\vee_{n=1}^N E_n)\geqslant x$ for all $N\geqslant 1$, and by σ–continuity, $\vee_{n=1}^\infty E_n\in\mathcal{T}_x$; so $\mathcal{T}_x$, $x\in\Lambda(F)$, are σ–ideals and (17) is immediate by construction.

Define on $\mathcal{T}_{a_i}$ $(i\in I)$

$$\mu_i(A) = \begin{cases} 0, & \text{if } J(A)\geqslant b_i \\ \theta_i^{-1}J(A), & \text{if } A\in\Gamma_{a_i} \text{ and } J(A)\in[a_i,b_i) \\ \infty, & \text{if } A\notin\Gamma_{a_i},\ J(A)\in[a_i,b_i) \end{cases}. \tag{18}$$

By definition of $\mathcal{T}_x$, $x\in\Lambda(F)$, the upper line of (14) holds. In the lower case, $J(A)\in[a_i,b_i)$; if $A\in\Gamma_{a_i}$, by (18) $J(A) = \theta_i[\mu_i(A)] = \phi_i[\mu_i(A)]$, while if $A\notin\Gamma_{a_i}$, $\phi_i[\mu_i(A)] = a_i$. But $A\notin\Gamma_{a_i}$ means that $J(A) = a_i$, for otherwise $(J(A) > a_i)$, $A = A\vee\phi\vee\phi\vee\cdots$ $(\phi\perp A$ for all $A\in L(P))$ implies $A\in\Gamma_{a_i}$. Thus, (14) holds.

Now $\mu_i(\phi) = 0$, since $J(\phi) = \infty\geqslant b_i$, for all $i\in I$. If $E,F\in\mathcal{T}_{a_i}$ and $\mu_i(E) = \mu_i(F) = 0$, then by (18) $J(E),J(F)\geqslant b_i$; (15) implies $J(E\vee F) = J(A\vee B) = F[J(A),J(B)]\geqslant F(b_i,b_i) = b_i$ $(b_i\in\Lambda(F))$, so $\mu_i(E\vee F) = 0$ also. Let $A = \vee_{n=1}^\infty E_n$, $E_n\perp E_m$. In case $A\notin\Gamma_{a_i}$, some E_{n_0} must satisfy $J(E_{n_0}) = a_i$. Thus, $\mu_i(A) = \infty = \Sigma_{n=1}^\infty\mu_i(E_n)$.

If $A\in\Gamma_{a_i}$ and $\bar{\mu}_i = \infty$,

$$J(A) = a_i = \phi_i\left[\Sigma_{J(E_n)\in[a_i,b_i)}\theta_i^{-1}J(E_n)\right],$$

where the sum is $\geqslant\bar{\mu}_i = \infty$. But

$$\sum_{n=1}^\infty \mu_i(E_n) = \sum_{J(E_n)\in[a_i,b_i)}\mu_i(E_n)$$

$$= \sum_{E_n\in\Gamma_{a_i},J(E_n)\in[a_i,b_i)}\theta_i^{-1}J(E_n) + \sum_{E_n\notin\Gamma_{a_i},J(E_n)\in[a_i,b_i)}(\infty) = \infty$$

so that $\mu_i(\vee E_n) = \Sigma\mu_i(E_n) = \infty$. If $A\in\Gamma_{a_i}$ and $\bar{\mu}_i < \infty$, by assumption (16) $J(A) > a_i$. Thus $\Sigma_{J(E_n)\in[a_i,b_i)}\theta_i^{-1}J(E_n) = \Sigma_{n=1}^\infty\mu_i(E_n) < \bar{\mu}_i$, and

$$\phi_i\left[\sum_{J(E_n)\in[a_i,b_i)}\theta_i^{-1}J(E_n)\right] = \theta_i\left[\sum_{n=1}^\infty\mu_i(E_n)\right] = J(A).$$

Thus, $\mu_i(A) = \theta_i^{-1}J(A) = \Sigma_{n=1}^\infty\mu_i(E_n)$. Hence, (18) defines

a generalized measure on $\mathscr{T}_{a_i}$. Q.E.D.

These theorems indicate that there is a very deep relationship between ROC–composible informations and measures. We point out that these results may perhaps be strengthened: However, they suffice for our present purposes and we have not attempted to study further interesting properties (such as uniqueness of the measures).

III. GLOBAL INFORMATIONS

In this section we consider an explanation of "entropy" as the global information associated with $L(P)$ (complete, orthocomplemented lattice) relative to a finite set of "local" informations, each ROC–composible. Although many of the concepts we employ are well known (lattice informations,[41] entropies as functionals of informations on Borel algebras,[44] measure-related entropies on lattices,[29] entropies relative to experiments[5]) our synthesis and generalization of these ideas appear to be unique. In order to maintain universality we seek only minimal (natural) properties of such "entropies" and do not attempt to characterize the special cases we introduce as examples. On the other hand, we sacrifice some generality by considering only ROC–composible informations on $L(P)$. We believe nevertheless that the resultant compromise is of considerable interest.

Let us denote by Z the set of all informations on $L(P)$ which are ROC–composible. We do not specify any ROC, however. We begin by considering properties of the information content of an experiment (partition) $\pi\in\Pi$ relative to a finite vector, $\mathbf{J}=(J_1,J_2,...,J_M)$, of informations, $J_m\in Z$. We denote by $\mathbf{F}$ the vector of corresponding ROC's, $\mathbf{F}=(F_1,F_2,...,F_M)$. Our first axiom is simply that the information content of π be a real-valued functional of the local informations:

Axiom 1:

$$H_\pi^{\mathbf{F}}: Z^M\to\mathbb{R},\quad \pi\in\Pi .$$

Axiom 1 generalizes Ref. 44, which considers only Borel algebras and the case $M=1$. Other axioms from this paper are meaningful in general. Indeed, the trivial proposition, $P\in L(P)$, satisfies $J(P)=0$ for all $J\in Z$, so it is natural that the global information of $\pi_0=\{\phi,P\}$ be zero:

Axiom 2:

$$H_{\pi_0}^{\mathbf{F}}:\quad Z^M\to\{0\} .$$

Our next axiom is not so widely accepted,[45] but we believe that it is in fact most natural to assume that an information on experiments not depend on the labels of the manifestation. Thus we assume

Axiom 3:

$$H_\pi^{\mathbf{F}}(\mathbf{J})=\Phi_{N_\pi}^{\mathbf{F}}\left[\{\mathbf{J}(E_n)\}_{n=1}^{N_\pi}\right];$$

$$\Phi_N^{\mathbf{F}}:\{\{\mathbf{x}_n:n=1,2,...N\}:\mathbf{x}_n\in\overline{\mathbb{R}^+}^M\}\to\mathbb{R},$$

where $\mathbf{J}(E_n)=(J_1(E_n),J_2(E_n),...,J_M(E_n))$, $E_n\in\pi$. That is, the information of $\pi\in\Pi$ relative to $\mathbf{J}$ depends only on the *set* of respective informations of the manifestations of π, and not on their order (index n). This axiom defines "symmetry"[44] of $H_\pi^{\mathbf{F}}$ (under permutations of $n=1,2,...,N$). We assume as well that the function $\Phi_N^{\mathbf{F}}$ has as domain all sets of NM–tuples of reals; we do this primarily for simplicity, but this hypothesis also reflects the fact that the global information is a functional of the $\mathbf{J}\in Z^M$.

Recall that Π is partially ordered under "refinement." It is natural to suppose that the more refined experiment has greater global informative value:

Axiom 4:

$$\pi_1<\pi_2 \text{ implies } H_{\pi_1}^{\mathbf{F}}\leqslant H_{\pi_2}^{\mathbf{F}} .$$

Again, although the order $<$ has varying definitions, the essence of Axiom 4 is well accepted.[5,46]

Now it is not inconceivable for some $\mathbf{J}\in Z^M$ that two partitions be related as follows: $\pi_1=\{E_1,E_2,...,E_N\}$, $\pi_2=\{F_1,F_2,...,F_{N+1}\}$, $\mathbf{J}(E_n)=\mathbf{J}(F_n)$, $n=1,2,...,N$ and $J(F_{N+1})=(\infty,\infty,...,\infty)$. In such a situation, since F_{N+1} is, informationally speaking, impossible, it is natural to demand that π_1 and π_2 possess the same information value. More generally, expansibility[44] takes the form:

Axiom 5:

$$\Phi_{N+1}^{\mathbf{F}}\left[\{\mathbf{x}_n\}_{n=1}^N\cup\{(\infty,\cdots\infty)\}\right]=\Phi_N^{\mathbf{F}}\left[\{\mathbf{x}_n\}_{n=1}^N\right],\ \mathbf{x}_n\in\overline{\mathbb{R}^+}^M.$$

Finally, consider two experiments defined by $\pi_1=\{E_1,E_2,...E_N\}$ and $\pi_2=\{E_1\vee E_2,E_3,...,E_N\}$. Clearly $\pi_2<\pi_1$; in general

$$H_{\pi_1}^{\mathbf{F}}(\mathbf{J})-H_{\pi_2}^{\mathbf{F}}(\mathbf{J})=\Delta_{N_{\pi_1}}^{\mathbf{F}}[\mathbf{J}(E_1),\mathbf{J}(E_2)\mathbf{J}(E_1\vee E_2)] ,$$

but the F–composibility of the J_m's means that $\mathbf{J}(E_1\vee E_2)=\mathbf{F}[\mathbf{J}(E_1),\mathbf{J}(E_2)]$, where $\mathbf{F}(\mathbf{x},\mathbf{y})=(F_1(x_1,y_1),F_2(x_2,y_2),...F_M(x_M,y_M))$—a semigroup operation on $\overline{\mathbb{R}^+}^M$. Then, branching[44] means

Axiom 6:

$$\Phi_N^{\mathbf{F}}\left[\{\mathbf{x}_n\}_{n=1}^N\right]-\Phi_{N-1}^{\mathbf{F}}\left[\{\mathbf{F}(\mathbf{x}_1,\mathbf{x}_2)\}\cup\{\mathbf{x}_n\}_{n=3}^N\right]$$
$$=\Delta_N^{\mathbf{F}}[\{\mathbf{x}_1,\mathbf{x}_2\}] ,$$

where $\Delta_N^{\mathbf{F}}$ is the information gain.[44]

The properties embodied in Axioms 1–6 are certainly minimal, natural, and (excepting Axiom 3) generally accepted. Other axioms involving algebraic independence[46] do not have a clear meaning in our general context, and will be avoided. We define the *global information of* $L(P)$ *relative to* $\mathbf{J}\in Z^M$ by

Axiom 7:

$$H^{\mathbf{F}}=\sup_{\pi\in\Pi}H_\pi^{\mathbf{F}} .$$

Since Π generally has no upper bound, this is the next best thing to taking a "finest partition" of $L(P)$. Even such an axiom has appeared in at least one context.[5]

The following theorem cites sufficient conditions to represent a global information—henceforth "entropy." The conditions are probably necessary as well (in case $M=1$ this is known[44]), but we will not attempt to verify this very tedious result.

Theorem 5: For any $\mathbf{F}=(F_1,F_2,...,F_M)$—$F_m$ a ROC, $M<\infty$—let

$$\varphi^{\mathbf{F}}:\overline{\mathbb{R}^+}^M\to[0,\infty) , \tag{19a}$$

$$\varphi^{\mathbf{F}}(\infty,\infty,...,\infty)=0 , \tag{19b}$$

$$\varphi^{\mathbf{F}}[\mathbf{F}(\mathbf{x},\mathbf{y})]\leqslant\varphi^{\mathbf{F}}(\mathbf{x})+\varphi^{\mathbf{F}}(\mathbf{y}) . \tag{19c}$$

Then, if $\mathbf{J} = (J_1, J_2, \ldots, J_m) \rightarrow J_m$ is F_m–composible on $L(P)$

$$H^{\mathbf{F}}(\mathbf{J}) = -\varphi^{\mathbf{F}}(0,0,\ldots,0) + \sup_{\pi \in \Pi}\left\{\sum_{n=1}^{N_\pi} \varphi^{\mathbf{F}}[\mathbf{J}(E_n)]\right\} \tag{20}$$

is a global information of $L(P)$ relative to $\mathbf{J}$.

Proof: Note that (20) is the π–supremum of

$$H^{\mathbf{F}}_{\pi}(\mathbf{J}) = -\varphi^{\mathbf{F}}(0,\ldots,0) + \sum_{n=1}^{N_\pi} \varphi^{\mathbf{F}}[\mathbf{J}(E_n)]\,. \tag{21}$$

Thus we must show (21) satisfies Axioms 1–6, with $\varphi^{\mathbf{F}}$ defined by (19). By (19a) it is clear the (21) satisfies Axiom 1—$N_\pi < \infty$. Since $\mathbf{J}(P) = (0,\ldots,0)$, by (19b), (21) satisfies Axiom 2. Certainly (21) satisfies Axiom 3, as the sum is commutative. By (19b) Axiom 5 is immediate. Condition 6 is satisfied since

$$\begin{aligned}\Delta^{\mathbf{F}}[\{\mathbf{J}(E_1),\mathbf{J}(E_2)\}] &= \{-\varphi^{\mathbf{F}}(0,\ldots 0) + \sum_{n=1}^{N} \varphi^{\mathbf{F}}(\mathbf{J}(E_n))\} \\ &\quad - \{-\varphi^{\mathbf{F}}(0,\ldots 0) + \varphi^{\mathbf{F}}(\mathbf{J}(E_1 \vee E_2)) + \sum_{n=3}^{N} \varphi^{\mathbf{F}}(\mathbf{J}(E_n))\} \\ &= \varphi^{\mathbf{F}}[\mathbf{J}(E_1)] + \varphi^{\mathbf{F}}[J(E_2)] - \varphi^{\mathbf{F}}[\mathbf{F}(\mathbf{J}(E_1),\mathbf{J}(E_2))]\,. \end{aligned} \tag{22}$$

That is, the difference depends only on $\mathbf{J}(E_1)$ and $\mathbf{J}(E_2)$. By (19c) we see that $\Delta^{\mathbf{F}}(\{\mathbf{J}(E_1),\mathbf{J}(E_2)\}) \geqslant 0$. If $\pi_1 < \pi_2$, there exists a chain $\pi_1 < \pi_a < \pi_b < \cdots < \pi_c < \pi_2$, with $N_{\pi_a} = N_{\pi_1} + 1$, $N_{\pi_b} = N_{\pi_a} + 1$, etc., so that employing (22) and its non-negativity successively yields Axiom 4. Q.E.D.

The next result essentially defines the effect of symmetry on the "entropy." Here $\mathrm{Aut}[L(P)]$ is the group of all automorphisms of $L(P)$—these preserve orthogonality and are complete $(c-)$ morphisms.

Corollary 1: Let $\mathscr{G} = \{g \in \mathrm{Aut}[L(P)]: g(P) = P\}$. Then $\mathbf{H}^{\mathbf{F}}$ is invariant under $\mathscr{G}$.

Proof: From Axiom 7 it is evident that any transformation $\Pi \xrightarrow{\text{onto}} \Pi$ leaves $H^{\mathbf{F}}$ invariant. If $g \in \mathscr{G}$, $g(\vee_{n=1}^{N} E_n) = \vee_{n=1}^{N} g(E_n)$, $g(E) \perp g(F)$ if $E \perp F$ and as $g(P) = P$, then $g(\pi) \in \Pi$ for each $\pi \in \Pi$. In addition, if $\pi \in \Pi$, for each $g \in \mathscr{G}$ there exists g^{-1} and thus $g[g^{-1}(\pi)] = \pi$, showing that g maps Π onto Π. Q.E.D.

Comment: If $\mathscr{T}$ is a subcomplete, orthocomplemented lattice of $L(P)$ with maximum $M_{\mathscr{T}}$, then $\mathscr{G}_{\mathscr{T}} = \{g \in \mathrm{Aut}[\mathscr{T}]: g(M_{\mathscr{T}}) = M_{\mathscr{T}}\}$ keeps invariant that portion of $H^{\mathbf{F}}$ that concerns only $\mathscr{T}$. This will be highly significant, as we show in Sec. IV. Note that transformations other than automorphisms may leave $H^{\mathbf{F}}$ invariant, but these revise the structure of the system and thus are eliminated.

The next result indicates that the semigroup $\mathbf{F}$ on $\overline{\mathbb{R}^+}^M$ induces a potentially useful decomposition of $H^{\mathbf{F}}$.

Corollary 2: Let $\Lambda(\mathbf{F}) = \times_{m=1}^{M} \Lambda(F_m)$, and let $\overline{\mathbb{R}^+}^M - \Lambda(\mathbf{F}) = \cup_{\mathbf{i} \in \mathbf{I}} \overline{[\times_{m=1}^{M}(a_{i_m}, b_{i_m})]}$—, where $\mathbf{i} = (i_1, i_2, \ldots, i_M)$ and $\mathbf{I} = \times_{m=1}^{M} I_m$, $\overline{\mathbb{R}^+} - \Lambda(F_m) = \cup_{i_m \in I_m}(a_{i_m}, b_{i_m})$. Define for $\varphi^{\mathbf{F}}$ satisfying (19)

$$T(\mathbf{i}) = \{E \in L(P): J_m(E) \in (a_{i_m}, b_{i_m}), m = 1,2,\ldots,M\}\,, \tag{23a}$$

$$S = L(P) - \cup_{\mathbf{i} \in I} T(\mathbf{i})\,, \tag{23b}$$

$$G(\pi \cap W) = \sum_{E_n \in \pi \cap W} \varphi^{\mathbf{F}}[\mathbf{J}(E_n)], \quad W \subseteq 2^{L(P)}\,. \tag{23c}$$

Then the global information (20) can be written

$$H^{\mathbf{F}}(\mathbf{J}) = -\varphi^{\mathbf{F}}(0,\ldots,0) + \sum_{\mathbf{i} \in \mathbf{I}} \sup_{\pi \in \Pi} G(\pi \cap T(\mathbf{i})) + \sup_{\pi \in \Pi} G(\pi \cap S) \tag{24}$$

N.B. $\{E \in L(P): J_m(E) \geqslant a_{i_m}, m = 1,2,\ldots,M\} = \cap_{m=1}^{M} \mathscr{T}_{a_{i_m}}$ with $\mathscr{T}_{a_{i_m}} = \{E \in L(P): J_m(E) \geqslant a_{i_m}\}$. These are σ–ideals.

Proof: Clearly (21) can be written as

$$H^{\mathbf{F}}_{\pi}(\mathbf{J}) = -\varphi^{\mathbf{F}}(0,\ldots 0) + \sum_{\mathbf{i} \in \mathbf{I}} G[\pi \cap T(\mathbf{i})] + G(\pi \cap S)\,. \tag{25}$$

Thus, $H^{\mathbf{F}}_{\pi}(\mathbf{J})$ is effectively the sum $\Sigma_{n \in \eta} G(\pi_n)$, where η is at most countable. The problem is therefore to show that

$$\sup_{\pi \in \Pi}\left[\sum_{n \in \eta} G(\pi_n)\right] = \sum_{n \in \eta}\left[\sup_{\pi \in \Pi} G(\pi_n)\right], \tag{26}$$

from which (24) follows. Trivially,

$$\sup_{\pi \in \Pi}\left[\sum_{n \in \eta} G(\pi_n)\right] \leqslant \sum_{n \in \eta}\left[\sup_{\pi \in \Pi} G(\pi_n)\right]. \tag{27}$$

For every $\epsilon > 0$ and $N \geqslant 1$, choose $\pi^n_{N,\epsilon}$ in Π such that

$$\epsilon/N > \sup_{\pi \in \Pi} G(\pi_n) - G(\pi^n_{N,\epsilon})\,. \tag{28}$$

Then, it follows by ordering η naturally that

$$\begin{aligned}\sup_{\pi \in \Pi} \sum_{n=1}^{N} G(\pi_n) + \epsilon &\geqslant \sum_{n=1}^{N} [G(\pi^n_{N,\epsilon}) + \epsilon/N] \\ &\geqslant \sum_{n=1}^{N} \sup_{\pi \in \Pi} G(\pi_n) = V_N, \text{ for all } N \geqslant 1, \epsilon > 0\,. \end{aligned} \tag{29}$$

Clearly V_N and $W_N = \sup \Sigma_{n=1}^{N} G(\pi_n)$ are nondecreasing in N, as $G(\pi_n) \geqslant 0$, so that—remember η is at most countable—

$$\begin{aligned}\operatorname{Lim}_{N \to \|\eta\|} V_N &= \sum_{n \in \eta} \sup_{\pi \in \Pi} G(\pi_n) \leqslant \epsilon + \operatorname{Lim}_{N \to \|\eta\|} W_N \\ &= \epsilon + \sup_{\pi \in \Pi}\left[\sum_{n \in \eta} G(\pi_n)\right]. \end{aligned} \tag{30}$$

As (30) holds for all $\epsilon > 0$, this proves the reverse inequality to (27) and thus (26). Q.E.D.

The next result is helpful in expressing entropy on certain atomistic lattices.[35]

Lemma: Let $L(P)$ be atomistic such that

for each $E \in L(P)$ there exists an at most countable set of mutually $\perp$ atoms A_n such that $E = \vee A_n$. (31)

Then, with Π^* the class of all countable partitions of atoms

$$\pi^* = \{A_n: A_n \perp A_m, A_n \text{ atoms}, P = \vee_{n=1}^{\infty} A_n\}\,,$$

$$H^{\mathbf{F}}(\mathbf{J}) \leqslant \sup_{\pi^* \in \Pi^*} H^{\mathbf{F}}_{\pi^*}(\mathbf{J})\,. \tag{32}$$

If we also assume that

$$\operatorname{Lim}_{N \to \infty} H^{\mathbf{F}}_{\pi_N}(\mathbf{J}) = H^{\mathbf{F}}_{\pi^*}(\mathbf{J}), \pi_N \in \Pi \text{ such that}$$

$$\pi_N = \{A_1, A_2, \ldots, A_{N-1}, \vee_{n \geqslant N} A_n\}\,, \tag{33}$$

then equality holds in (32).

Proof: If $\pi\in\Pi$, there exists $\pi^*\in\Pi^*$ such that $\pi<\pi^*$ by (31) defined by $\cup_{n=1}^{N_\pi}\{A_m:m\in\mathscr{M}_n\}$, $E_n=\vee_{m\in\mathscr{M}_n}A_m$. Thus, by axiom 4, $H_\pi^{\mathbf{F}}(\mathbf{J})\leqslant H_{\pi^*}^{\mathbf{F}}(\mathbf{J})$ and (32) results. For each $\pi^*\in\Pi^*$, consider the sequence $\pi_N=\{A_1,A_2,\ldots A_{N-1}, \vee_{n\geqslant N}A_n:A_n\in\pi^*\}$. Clearly $\pi_N\in\Pi$, $\pi_N<\pi_{N+1}<\pi^*$. By (33) and the fact that $H_{\pi_N}^{\mathbf{F}}(\mathbf{J})\leqslant H^{\mathbf{F}}(\mathbf{J})$,

$$H_{\pi^*}^{\mathbf{F}}(\mathbf{J})\leqslant H^{\mathbf{F}}(\mathbf{J}),\quad \text{for all } \pi^*\in\Pi^* \tag{32'}$$

so that equality in (32) is apparent. Q.E.D.

Let us first illustrate the above results in the case $M=1$. Assume $\varphi^{\mathbf{F}}(\mathbf{x})$ has the form:

$$\varphi^F(x)=\begin{cases}X\psi_i(X),\ \mathrm{X}=\theta_i^{-1}(x), & x\in(a_i,b_i),\\ 0,\ \text{otherwise},\end{cases} \tag{34}$$

where $\psi_i:(0,\bar{\mu}_i]\to[0,\infty)$ is *nonincreasing* and $\psi_i(\bar{\mu}_i)=0$. It is trivial to verify that (34) satisfies (19a) and (19b), while (19c) follows since, because ψ_i is nonincreasing,

$$(X+Y)\psi_i(X+Y)\leqslant X\psi_i(X)+Y\psi_i(Y),\quad X,Y\in(0,\bar{\mu}_i) \text{ with } X+Y\in(0,\bar{\mu}_i). \tag{35}$$

Assume the conditions of Theorem 3, so $J(E)\in(a_i,b_i)$ implies $J(E)=\theta_i[\mu_i(E)]$ with $\mu_i:\mathscr{T}_{a_i}\to[0,\bar{\mu}_i]$ and $\mu_i(E)=0$ for all $E\in\mathscr{T}_{a_i}\cap\mathscr{T}_{b_i}$. We write (24) as

$$H^F(J)=\sum_{i\in I}\sup_{\pi\in\Pi}\left\{\sum_{E_n\in\pi\cap\mathscr{T}_{a_i}}\mu_i(E_n)\psi_i[\mu_i(E_n)]\right\}. \tag{36}$$

Let us further suppose that $M_i=\vee\{E\in\mathscr{T}_{a_i}\}\in\mathscr{T}_{a_i}$ and that the supremum over $\pi\cap\mathscr{T}_{a_i}$ may be replaced by the supremum over $\pi\in\Pi_i$, where Π_i consists of all (finite) partitions of M_i with components in $\mathscr{T}_{a_i}$. In particular, let each $\mathscr{T}_{a_i}$ be an irreducible propositional system, $\mathscr{T}_{a_i}\approx\mathscr{P}[\mathscr{H}^i]$ (lattice isomorphic) the closed submanifolds of a Hilbert space. Then $\mathscr{T}_{a_i}$ is atomistic and (31) holds; moreover, our measure coincides with the CROC-measure of Piron,[35] which reduces to (we choose $\bar{\mu}_i=1, i\in I$)

$$\mu_i(E)=\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_E\quad E\in\mathscr{T}_{a_i},\quad E\leftrightarrow\hat{\Pi}_E\in\mathscr{P}[\mathscr{H}^i], \tag{37}$$

where $\hat{W}_i$ is a density operator on $\mathscr{H}^i$. Hence, (36) becomes

$$H^F(J)=\sum_{i\in I}\sup_{\pi^*\in\Pi_i^*}\left\{\sum_{n=1}^{\infty}\mathrm{Tr}_i[(\hat{W}_i\hat{\Pi}_{E_n})\psi_i(\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{E_n})]\right\}, \tag{38}$$

where $E_n\in\pi^*$ are atoms. Let $p_n^i(\pi^*)=\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{E_n}$. Then, each i-term in (38) is essentially the supremum over all "measurements" of a generalized "Ingarden–Urbanik entropy." To see this, let $\psi_i(z)=-\ln z$. Then clearly $\psi_i:(0,1]\to[0,\infty)$ is nonincreasing with $\psi_i(1)=0$. Moreover, (33) is satisfied since in fact for all $i\in I$:

$$\begin{aligned}\lim_{N\to\infty}\Bigg[&\sum_{n=1}^{N-1}-\mathrm{Tr}_i(\hat{W}_i\hat{\Pi}_{E_n})\ln(\hat{W}_i\hat{\Pi}_{E_n})\\ &+\Big(-\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{\vee_{n\geqslant N}E_n}\ln\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{\vee_{n\geqslant N}E_n}\Big)\Bigg]\\ &=-\sum_{n=1}^{\infty}\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{E_n}\ln\hat{W}_i\hat{\Pi}_{E_n}\end{aligned} \tag{33'}$$

since for large enough N,

$$|-\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{\vee_{n\geqslant N}E_n}\ln(\mathrm{Tr}_i\hat{W}_i\hat{\Pi}_{\vee_{n\geqslant N}E_n})|$$

can be made arbitrarily small. Thus (38) becomes

$$\begin{aligned}H^F(J)&=\sum_{i\in I}\sup_{\pi^*\in\Pi_i^*}\left\{-\sum_{n=1}^{\infty}\mathrm{Tr}_i[\hat{W}_i\hat{\Pi}_{E_n}\ln(\hat{W}_i\hat{\Pi}_{E_n})]\right\}\\ &\geqslant\sum_{i\in I}(-\mathrm{Tr}_i\hat{W}_i\ln\hat{W}_i).\end{aligned} \tag{39}$$

See Ref. 1, p. 255.

Expanding $\hat{W}_i$ in terms of its eigenbasis (π_0^*),

$$\begin{aligned}p_n^i(\pi^*)&=\sum_{m=1}^{\infty}p_m^i(\pi_0^*)T_{mn}(\pi^*);\ T_{mn}(\pi^*)\\ &=\mathrm{Tr}_i\hat{\Pi}_{F_m}\hat{\Pi}_{E_n},\ F_m\in\pi_0^*, E_n\in\pi^*.\end{aligned}$$

Note that $T_{mn}(\pi^*)$ is a non-negative, doubly stochastic matrix. If we assume $\Gamma_i(z)=z\psi_i(z)$ is *convex*, then

$$\begin{aligned}H^F(J)&=\sum_{i\in I}\sup_{\pi^*\in\Pi_i^*}\sum_{n=1}^{\infty}\Gamma_i\Big[\sum_{m=1}^{\infty}p_m^i(\pi_0^*)T_{mn}(\pi^*)\Big]\\ &\leqslant\sum_{i\in I}\sup_{\pi^*\in\Pi_i^*}\Big[\sum_n\sum_m T_{mn}(\pi^*)\Gamma_i(p_m^i(\pi_0^*))\Big]\\ &=\sum_{i\in I}\sum_m\Gamma_i(p_m^i(\pi_0^*)).\end{aligned} \tag{40}$$

But since $p_m^i(\pi_0^*)$ are just the eigenvalues of $\hat{W}_i$, and since $\pi_0^*\in\Pi_0^*$, we have from (40)

$$H^F(J)=\sum_{i\in I}\mathrm{Tr}_i\Gamma_i(\hat{W}_i). \tag{41}$$

For properties of $\mathrm{Tr}\Gamma(\hat{W})$, where Γ is concave, see Ref. 47.

Our next illustration is for the case $M=2$. Again, without any real justification, let us assume that $\varphi^{\mathbf{F}}$ has the form

$$\varphi^{\mathbf{F}}(x,\xi)=\begin{cases}\Psi_{ij}(x,\xi), & (x,\xi)=\mathbf{x}\in\Delta_{ij}=(a_i,b_i)\times(a_j,b_j),(i,j)\in I_1\times I_2\\ 0,\ \text{otherwise}.\end{cases} \tag{42}$$

If we specify that $i<j$ implies $a_i<a_j$, the conditions (19) reduce to

$$\Psi_{ij}(\mathbf{x})\in[0,\infty)\quad \text{for all } \mathbf{x}\in\Delta_{ij},\quad (i,j)\in I_1\times I_2, \tag{43}$$

$$\Psi_{ij}(\mathbf{x})+\Psi_{ij}(\mathbf{y})\geqslant\Psi_{ij}[\mathbf{F}(\mathbf{x},\mathbf{y})]\quad \mathbf{x},\mathbf{y}\in\Delta_{ij},\quad (i,j)\in I_1\times I_2, \tag{44}$$

$$\Psi_{ij}(\mathbf{x})+\Psi_{ik}(\mathbf{y})\geqslant\Psi_{ij}[F_1(x,y),\xi] \quad j<k;\mathbf{x}\in\Delta_{ij},\quad \mathbf{y}\in\Delta_{ik},\quad (F_1(x,y),\xi)\in\Delta_{ij}, \tag{45a}$$

$$\Psi_{ij}(\mathbf{x})+\Psi_{kj}(\mathbf{y})\geqslant\Psi_{ij}[x,F_2(\xi,\eta)] \quad i<k;\mathbf{x}\in\Delta_{ij},\quad \mathbf{y}\in\Delta_{kj},\quad (x,F_2(\xi,\eta))\in\Delta_{ij}. \tag{45b}$$

Note that $F_1(x,y)\in[a_i,b_i)$, if $x,y\in(a_i,b_i)$ in general (likewise for F_2).

We further simplify matters as follows:

$$I_2=\{0,\infty\};\xi=\theta_0^{(2)}(\sigma),\quad \sigma\in(0,\bar{\sigma}),\quad \bar{\sigma}\leqslant\infty, \tag{46}$$

$$\Psi_{i0}(x,\xi)=\begin{cases}\sigma\psi_i(X/\sigma), X=\theta_i^{-1}(x), & x\in(a_i,b_i);\quad \sigma\in(0,\bar{\sigma})\\ 0,\ \text{otherwise}.\end{cases} \tag{47}$$

Note that $X\in(0,\bar{\mu}_i)$ if $x=\theta_i(X)\in(a_i,b_i)$. The constraints (43)–(45) become:

$$\psi_i:(0,\infty)\to[0,\infty),\quad i\in I_1 \tag{48}$$

$$\sigma\psi_i(X/\sigma)+\tau\psi_i(Y/\tau)\geqslant(\sigma+\tau)\psi_i((X+Y)/(\sigma+\tau)), \quad X,Y,X+Y\in(0,\bar{\mu}_i);\ \xi,\tau,\xi+\tau\in(0,\bar{\sigma}), \tag{49}$$

$$\sigma\psi_i(X/\sigma)+\tau\psi_j(Y/\tau)\geqslant(\sigma+\tau)\psi_i(X/(\sigma+\tau)) \text{ if } i<j;\ X\in(0,\bar{\mu}_i),$$

$Y\in(0,\bar{\mu}_j);\quad \sigma,\tau,\sigma+\tau\in(0,\bar{\sigma})$. (50)

A degree of justification for the arbitrariness of choices (42) and (46) can be felt from the following results.

Lemma: $\psi_i(z)$ is a convex, continuous function $\psi_i:(0,\infty)\to[0,\infty)$ and

$$\inf_{z\in(0,\infty)}\psi_j(z)\geqslant\psi_i(0^+)=\lim_{\epsilon\to 0^+}\psi_i(\epsilon),\quad i<j \tag{51}$$

if and only if ψ_i satisfies (48)–(50), for each $i\in I$.

Proof: Clearly (48) means $\psi_i:(0,\infty)\to[0,\infty)$—bounded everywhere. Condition (49) is equivalent to convexity of ψ_i, which because ψ_i is bounded, means continuity as well in $(0,\infty)$. (Ref. 48, p. 91.) First, for $z_1>z_2$, $z_i\in(0,\infty)$, choose $X\in(0,\bar{\mu}_i/2)$ and $\sigma\in(0,\bar{\sigma}/2)$ such that $z_1=X/\sigma$—as X/σ ranges over all $(0,\infty)$ this is always possible. Let $Y=\sigma z_2=(z_2/z_1)X\in(0,\bar{\mu}_i/2)$ and take $\tau=\sigma$ in (49). Thus

$$\psi_i(z_1)+\psi_i(z_2)\geqslant 2\psi_i[(z_1+z_2)/2]\quad\text{for all } z_1,z_2\in[0,\infty) \tag{49'}$$

so that by Ref. 48 (page 70) ψ_i is convex.

Conversely, if ψ_i is convex, then for $z_1,z_2\in(0,\infty),\sigma,\tau\in(0,\infty)$,

$$\frac{\sigma}{\sigma+\tau}\psi_i(z_1)+\frac{\tau}{\sigma+\tau}\psi_i(z_2)\geqslant\psi_i\left[\frac{\sigma}{\sigma+\tau}z_1+\frac{\tau}{\sigma+\tau}z_2\right]$$

so taking $z_1=X/\sigma$ and $z_2=Y/\tau$ we obtain (49) for $X,Y,X+Y\in(0,\bar{\mu}_i)$ and $\sigma,\tau,\sigma+\tau\in(0,\bar{\sigma})$.

We show that given ψ_i are convex and continuous on $(0,\infty)$, (51) is equivalent to (50). In fact, we show that

$$\begin{aligned}\tau\psi_i(0^+)=\sup\{&(\sigma+\tau)\psi_i(X/(\sigma+\tau))-\sigma\psi_i(X/\sigma):\\ &X\in(0,\bar{\mu}_i),\ \sigma\in(0,\bar{\sigma})\text{ such that }\sigma+\tau\in(0,\bar{\sigma})\}\ .\end{aligned} \tag{52}$$

Let $X\in(0,\bar{\mu}_i)$, $Y\in(0,\bar{\mu}_j)$, $\sigma,\tau,\sigma+\tau\in(0,\bar{\sigma})$. Let $X=X'+\delta$ where $X',\delta\in(0,\bar{\mu}_i)$. Then by convexity

$$\begin{aligned}(\sigma+\tau)\psi_i[X/(\sigma+\tau)]-\sigma\psi_i(X/\sigma)&\leqslant\sigma\psi_i(X'/\sigma)\\ &\quad+\tau\psi_i(\delta/\tau)-\sigma\psi_i(X/\sigma)\\ &=\sigma[\psi_i((X-\delta)/\sigma)-\psi_i(X/\sigma)]+\tau\psi_i(\delta/\tau)\ .\end{aligned} \tag{53}$$

Now $|\psi_i((X-\delta)/\sigma)-\psi_i(X/\sigma)|\underset{\delta\to 0}{\longrightarrow}0$ by continuity, so for each $\epsilon>0$, there exists $\delta\in(0,\bar{\mu}_i)$ such that

$$-\epsilon\sigma\leqslant\sigma[\psi_i((X-\delta)/\sigma)-\psi_i(X/\sigma)]\leqslant\epsilon\sigma\ ,$$

$$-\epsilon'\tau\leqslant\tau[\psi_i(0^+)-\psi_i(\delta/\tau)]\leqslant\epsilon'\tau\ .$$

Take $\epsilon=\min(\epsilon,\epsilon')$ for δ.

$$(\sigma+\tau)\psi_i[X/(\sigma+\tau)]-\sigma\psi_i[X/\sigma]\leqslant\epsilon(\sigma+\tau)+\tau\psi_i(0^+)\ ,$$

so

$$(\sigma+\tau)\psi_i(X/(\sigma+\tau))-\sigma\psi_i(X/\sigma)\leqslant\tau\psi_i(0^+)\ . \tag{54}$$

Since for $X=0^+$ equality holds in (54) we obtain (52). Now (50) is equivalent to

$$\begin{aligned}&\tau\psi_j(Y/\tau)\geqslant(\sigma+\tau)\psi_i(X/(\sigma+\tau))-\sigma\psi_i(X/\sigma);\\ &\quad\forall X\in(0,\bar{\mu}_i),Y\in(0,\bar{\mu}_j),\sigma,\tau,\sigma+\tau\in(0,\bar{\sigma})\end{aligned}$$

if and only if

$$\tau\psi_j(Y/\tau)\geqslant\tau\psi_i(0^+),\quad\tau\in(0,\bar{\sigma}),\quad Y\in(0,\bar{\mu}_j)\ ,$$

if and only if

$$\psi_j(Y/\tau)\geqslant\psi_i(0^+),\quad\tau\in(0,\bar{\sigma}),\quad Y\in(0,\bar{\mu}_j)\ ,$$

if and only if

$$\psi_j(z)\geqslant\psi_i(0^+),\quad\forall z\in(0,\infty)\ ,$$

if and only if

$$\inf\{\psi_j(z):z\in(0,\infty)\}\geqslant\psi_i(0^+)\ .$$ Q.E.D.

The following function ψ_i satisfies the Lemma's conditions:

$$\psi_i(z)=z\ln z+(1+i)/e,\quad\text{for all } z\in(0,\infty)\ . \tag{55}$$

Note that $\psi_i(0^+)=(1+i)/e$ ($0\ln 0=0$ convention).

Assume for J_1,J_2 the conditions of Theorem 3, and that $M_i=\vee\{E\in\mathscr{T}_{a_i}\}\in\mathscr{T}_{a_i}$ and let $\sigma_i=\sigma|_{\mathscr{T}_{a_i}}$ (restriction of σ to $\mathscr{T}_{a_i}$). Then, using (24) with Π_i the set of partitions of M_i in $\mathscr{T}_{a_i}$ again replacing $\{\pi\cap\mathscr{T}_{a_i}\}$

$$H^{\mathrm{F}}(J_1,J_2)=\sum_{i\in I_1}h_i\ , \tag{56a}$$

$$h_i=\sup_{\pi\in\Pi_i}\left\{\sum_{\substack{\sigma_i(E_n)\in(0,\bar{\sigma})\\ \mu_i(E_n)\in(0,\bar{\mu}_i)}}\sigma_i(E_n)\psi_i\left[\frac{\mu_i(E_n)}{\sigma_i(En)}\right]\right\}. \tag{56b}$$

In order to discuss classical and quantum physics, we further assume that each $\mathscr{T}_{a_i}$ is a CROC for which only discrete superselection rules are present. We assume also that μ_i is a *state* on $\mathscr{T}_{a_i}$ ($\mu_i(M_i)=\bar{\mu}_i=1$) and $\mu_i\equiv\sigma_i$. Each CROC is associated with a W^*-algebra $\mathscr{A}_i$ of observables on a Hilbert space $\mathscr{H}_i$ and "state" N_i on $\mathscr{A}_i$. Indeed, there exists a homomorphism $\gamma_i:\mathscr{T}_{a_i}\overset{\text{onto}}{\longrightarrow}\mathscr{L}_i=\mathscr{A}_i\cap\mathscr{P}(\mathscr{H}_i)$, where $\mathscr{P}(\mathscr{H}_i)$ is the lattice of projectors on $\mathscr{H}_i$, so $N_i[\gamma_i(E)]=\mu_i(E)$ relates the two types of "states." If γ_i happens to be an isomorphism, then there is a complete algebraic representation of $\mathscr{T}_{a_i}$. We henceforth assume this holds.[49]

We also assume that the measure σ_i corresponds to a faithful, normal, semifinite trace, m_i, on $\mathscr{A}_i$. [See Ref. 1, p. 258, and Ref. 50.] That is, $m_i[\gamma_i(E)]=\sigma_i(E)$ for all $E\in\mathscr{T}_{a_i}$. The main characteristic of such a trace (aside from certain convergence properties) is its unitary invariance: $m_i(\hat{U}^\dagger\hat{I}\hat{I}\hat{U})=m_i(\hat{I}\hat{I})$ for all unitary $\hat{U}\in\mathscr{A}_i$ and all $\hat{I}\hat{I}\in\mathscr{L}_i$. Of the state N_i we demand that there exist $\hat{W}_i>0$ in $\mathscr{A}_i$ such that $N_i(\hat{X})=m_i(\hat{W}_i\hat{X})$ for all $\hat{X}\in\mathscr{A}_i$. We can now describe h_i in terms of partitions of projectors in $\mathscr{L}_i$:

$$h_i=\sup_{\pi_i}\left\{\sum_{m(\hat{P}_n)\neq 0,\infty}m_i(\hat{P}_n)\psi_i[m_i(\hat{W}_i\hat{P}_n)/m_i(\hat{P}_n)]\right\}. \tag{57}$$

[Note that from $\sigma_i\equiv\mu_i$, $\sigma_i(E)=0$ iff $\mu_i(E)=0$, but $\mu_i(E)=\bar{\mu}_i<\infty$ iff $\mu_i(E^\perp)=0$ iff $\sigma_i(E^\perp)=0$ iff $\sigma_i(E)=\bar{\sigma}_i<\infty$. Thus we can eliminate the conditions on μ_i in (56b).] Finally, let us suppose that $\psi_i(\hat{W}_i)$ is "m_i-integrable," i.e., $m_i[\psi_i(\hat{W}_i)]<\infty$. Under these many conditions we have that $h_i=m_i[\psi_i(\hat{W}_i)]$. We prove this in the following Lemma (dropping the index i for convenience.)

Lemma: Let m be a faithful, normal, semifinite trace for a W^*-algebra $\mathscr{A}$ with projection lattice $\mathscr{L}$ (a logic). Let n be a normal state on $\mathscr{A}$ such that there exists self-adjoint $\hat{W}\in\mathscr{A}$, $\hat{W}>0$, such that $n(\hat{X})=m(\hat{W}\hat{X})$ for all $\hat{X}\in\mathscr{A}$. We assume $\psi:(0,\infty)\to(0,\infty)$ is convex and continuous, and $m[\psi(\hat{W})]<\infty$. Then

$$m[\psi(\hat{W})] = \sup_{\pi} \sum_{m(\hat{P}_n)\neq 0,\infty} m(\hat{P}_n)\psi[m(\hat{W}\hat{P}_n)/m(\hat{P}_n)] . \quad (58)$$

Proof: $\hat{W}$ being positive and self-adjoint,

$$\hat{W} = \int_0^\infty d\hat{P}(\omega)\omega \quad (59)$$

is its spectral resolution; we define $\psi(\hat{W})$ by

$$\psi(\hat{W}) = \int_0^\infty d\hat{P}(\omega)\,\psi(\omega) \quad (60)$$

so that with $m_{\hat{W}}(\Delta) = \int_\Delta dm\hat{P}(\omega), \Delta \in B(R^+)$:

$$m[\psi(\hat{W})] = \int_0^\infty dm[\hat{P}(\omega)]\,\psi(\omega) = \int_0^\infty dm_{\hat{W}}\,\psi(\omega) . \quad (61)$$

Since ψ is convex and continuous, its domain decomposes into two disjoint parts: D_1 on which ψ is monotonic increasing, and D_2 on which ψ is monotonic decreasing. Thus, using (61) we obtain

$$\begin{aligned} m[\psi(\hat{W})] &= \sum_{j=1}^{2} \int_{\psi^{-1}(D_j)} dm_{\hat{W}}\,\psi(\omega) \\ &= \sum_{j=1}^{2} \sup_{\{\Delta_n \cap \psi^{-1}(D_j)\}} \Bigg[\sum_{m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)]\neq 0,\infty} m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \\ &\quad \times \inf\{\psi(\omega): \omega \in \Delta_n \cap \psi^{-1}(D_j)\} \Bigg] . \end{aligned} \quad (62)$$

The restriction to elements of the partitions of R^+ with $m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \neq 0,\infty$ follows from the integrability of $\int_{\psi^{-1}(D_j)} dm_{\hat{W}}\,\psi(\omega)$ while the supremum representation of the integral is standard. (See, for example, Ref. 39, p. 115.) On the other hand, for all $\Delta \in B(R^+)$ we have

$$\begin{aligned} m_{\hat{W}}[\Delta \cap \psi^{-1}(D_j)] \inf\{\omega \in \Delta \cap \psi^{-1}(D_j)\} &\leqslant \int_{\Delta \cap \psi^{-1}(D_j)} dm_{\hat{W}}\,\omega \\ &\leqslant m_{\hat{W}}[\Delta \cap \psi^{-1}(D_j)] \sup\{\omega \in \Delta \cap \psi^{-1}(D_j)\} , \end{aligned} \quad (63)$$

where $m_{\hat{W}}[\Delta \cap \psi^{-1}(D_j)] \neq 0,\infty$. Exploiting the monotonicity of ψ on D_j yields

$$\begin{aligned} &\inf\{\psi(\omega): \omega \in \Delta \cap \psi^{-1}(D_j)\} \\ &\quad \leqslant \psi\left[\int_{\Delta \cap \psi^{-1}(D_j)} dm_{\hat{W}}\,\omega / m_{\hat{W}}(\Delta \cap \psi^{-1}(D_j)) \right] . \end{aligned} \quad (64)$$

Thus, by (64) in (62) we get

$$\begin{aligned} m[\psi(\hat{W})] &\leqslant \sum_{j=1}^{2} \sup_{\{\Delta_n \cap \psi^{-1}(D_j)\}} K[\{\Delta_n \cap \psi^{-1}(D_j)\}] \\ &= \sup_{\{\Delta_n\}} \sum_{j=1}^{2} K[\{\Delta_n \cap \psi^{-1}(D_j)\}] , \end{aligned} \quad (65)$$

where

$$\begin{aligned} K[\{\Delta_n \cap \psi^{-1}(D_j)\}] &= \sum_{m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)]\neq 0,\infty} m_{\hat{W}}[\Delta_n \cap \psi^{-1}(D_j)] \\ &\quad \times \psi\left[\int_{\Delta_n \cap \psi^{-1}(D_j)} dm_{\hat{W}}\omega / m_{\hat{W}}(\Delta_n \cap \psi^{-1}(D_j)) \right] \end{aligned}$$

and an argument as in Corollary 2 to Theorem 5 applies. But, $\{\Delta_n \cap \psi^{-1}(D_j)\}$ is a finer partition of R^+ than $\{\Delta_n\}$, and $\{\hat{P}_{\Delta_n \cap \psi^{-1}(D_j)} = \int_{\Delta_n \cap \psi^{-1}(D_j)} d\hat{P}(\omega)\}$ is among *all* projector partitions $\{\hat{P}_n : n = 1,2,\ldots,N_\pi\} \in \Pi$, so that

$$m[\psi(\hat{W})] \leqslant \sup_{\pi} \sum_{m(\hat{P}_n)\neq 0,\infty} m(\hat{P}_n)\psi[m(\hat{P}_n\hat{W})/m(\hat{P}_n)] . \quad (66)$$

On the other hand, for any projector $\hat{P}$ such that $m(\hat{P}) \neq 0,\infty$

$$\begin{aligned} \psi\left[\frac{m(\hat{P}\hat{W})}{m(\hat{P})} \right] &= \psi\left[\int dm\,[\hat{P}\hat{P}(\omega)]\omega / m(\hat{P}) \right] \\ &\leqslant \int dm[\hat{P}\hat{P}(\omega)]\psi(\omega)/m(\hat{P}) = m[\hat{P}\psi(\hat{W})]/m(\hat{P}) , \end{aligned} \quad (67)$$

using Jensen's inequality. Employing (67) with (66) yields

$$\begin{aligned} m[\psi(\hat{W})] &\leqslant \sup_{\pi \in \Pi} \sum_{m(\hat{P}_n)\neq 0,\infty} m[\hat{P}_n]\psi[m(\hat{P}_n\hat{W})/m(\hat{P}_n)] \\ &\leqslant \sup_{\pi} \sum_{m(\hat{P}_n)\neq 0,\infty} m[\hat{P}_n\psi(\hat{W})] \leqslant m[\psi(\hat{W})] \end{aligned} \quad (68)$$

since $\Sigma_{m(\hat{P}_n)\neq 0,\infty} \hat{P}_n \leqslant \hat{I}$. Thus, we have (58). Q.E.D.

The above result is based on the proof of Theorem 1.1 of Ref. 51, a classical version of the Lemma (without the restrictions $\psi \geqslant 0$ and integrable). Our "entropy" is thus given by

$$H^F(J_1,J_2) = \sum_{i\in I_1} m_i[\psi_i(\hat{W}_i)] . \quad (69)$$

In case $\sigma_i(M_i) < \infty$ and $\psi_i(z)$ is given by (55), this becomes

$$\begin{aligned} H^F(J_1,J_2) &= \sum_{i\in I_1} [m_i(\hat{W}_i \ln \hat{W}_i) + (1+i)m_i(\hat{I}_i)/e] \\ &= \sum_{i\in I_1} [-S_i(\hat{W}_i) + (1+i)m_i(\hat{I}_i)/e] , \end{aligned} \quad (70)$$

where $S_i(\hat{W}_i)$ is the "Segal entropy," which includes both the (classical) Gibbs–Boltzmann–Shannon entropy and the (quantum) von Neumann entropies as cases.[50] If, however, $\sigma_i(M_i) = \infty$—our result fails since $m_i[\psi_i(\hat{W}_i)] = \infty \geqslant h_i$. In words, the "global information" is essentially the "neg–entropy" plus the prior measure of the space.

Since our purpose in introducing these illustrations was simply to indicate the relationship of the "global information" concept to typical "entropies," we will not discuss further properties or problems associated with expressions like (70). Rather, we postpone such discussion until the axiom set is completed by an appropriate description of the information of compound systems.

IV. A TENTATIVE INTERPRETATION

In the preceeding sections we have established minimal universal criteria for a "global measure of information on a 'theory' $L(P)$ relative to M (generalized) local information measures"—a quantity which by the examples we considered, apparently relates to a generalization of the neg-entropies of physics. We have purposely avoided questions of characterization, as the universal structure of $L(P)$ is not yet fixed. Nevertheless, a truly surprising result appears already, namely, the introduction of a natural semigroup F (on $\overline{R^{+M}}$) and its natural segmentation of the "entropy." In keeping with our policy of specialization, let us consider the case $M = 1$. (The arguments apply as well to the $M = 2$ case considered in Sec. III.)

With $a_0 = 0 < b_0 \leqslant a_1 < b_1 \leqslant a_2 \ldots$ by the construction of Theorem 3, we may write (24) as

$$H^F(J) = \lim_{t\to\infty} H_t^F(J) , \quad (71)$$

where we assume for convenience that $\varphi^F(0) = 0$ and define for all $t > 0$:

$$H_t^F(J) = \sum_{i\in I, t > a_i} \sup_{\pi\in\Pi} \left[\sum_{J(E_n)\in(a_i,b_i)} \varphi^F[J(E_n)] \right] + \sum_{i\in I, t\geq b_i} \sup_{\pi\in\Pi} \left[\sum_{J(E_n)\in[b_i,\min(t,a_{i+1})]} \varphi^F[J(E_n)] \right]. \quad (72)$$

Note that $H_t^F(J)$ is constant for all $t\in(a_i,b_i)$ and is generally increasing otherwise. The connection between this monotonic nondecreasing "entropy" and dynamics—the "Second Law"—comes about as follows.

Let us model *time* by the positive real line $\overline{R^+}$. This corresponds to the usual "initial value" problem of physics, but can also represent an operational model of a general relativistic cosmos because of the finite age of the universe. A dynamical object is described[35] by

$$P(T) = \bigvee^{\oplus}_{z\in[0,T)} P_z; \quad \bar{P} = \lim_{T\to\infty} P(T), \quad (73)$$

where $\vee^{\oplus}$ is the direct product of posets (the cartesian set product ordered in the obvious way[27]). According to our model, each $P_z, z\in\overline{R^+}$, represents the abstract definition—i.e., the ordered skeleton of defining relations—at the "instant" z. The structure may endure over finite intervals, i.e. $P_z = P$ for all $z\in(a,b)$, whereas it may continuously change over other intervals. Since $L(\vee^{\oplus} P_z) = \vee^{\oplus} L(P_z)$[22] these comments extend directly to the abstract "theory" of the "objects."

Note that the "object" thus described is a "dynamical object." In fact, this description is analogous to the world-line models in relativity. Let us define

$$\mathcal{T}_T = [\vee^{\oplus}_{z\in[0,T)} \phi_z] \vee^{\oplus} [\vee^{\oplus}_{z\geq T} L(P_z)] \subseteq L(\bar{P}). \quad (74)$$

Clearly each $\mathcal{T}_T$ is a complete sublattice of $L(\bar{P})$ (and thus a σ-ideal). Moreover, if $T_1 < T_2$, $\mathcal{T}_{T_1} \supset \mathcal{T}_{T_2}$ is evident. If we impose on $\overline{R^+}$ the semigroup F with idempotents of $\Lambda(F)$, the ideals $\mathcal{T}_z, z\in\Lambda(F)$ correspond to *distinct* components in $L(\bar{P})$, while $\vee^{\oplus}_{z\in(a_i,b_i)} \mathcal{T}_z$ corresponds to $\vee^{\oplus}_{z\in(a_i,b_i)} L(P_z)$, where $L(P_z)$ is the *same* structure for all z in (a_i,b_i). [Compare with the usual physical model of Ref. 35 where the $L(P_z)$ are related to the mutually isomorphic Hilbert spaces.] Thus, intuitively each choice of F (a real-line semigroup!) induces a world-line model that consists of (open) segments with constant entropy (and reversible unitary dynamics) and of (closed) segments of constantly increasing entropy in which dynamics is irreversible in that the physical structure of the object changes. The "entropy" $H_t^F(J)$ defined by (72) is the total information about the system up to "time t"—i.e., the entropy of the "world-line" segment defined by $[0,t]$. As more data about the "world-line" is obtained, the information increases towards the maximum represented by (71).

Each "dynamical object" can therefore be associated with a semigroup $(\overline{R^+}, F)$—*not* the usual physical choice[7]—and $F - \sigma$-composible informations on $L(\bar{P})$ which are related to unitarily evolving measures on $\mathcal{T}_{a_i}$ (that vanish on $\mathcal{T}_{a_i}\cap\mathcal{T}_{b_i}$). In order to see more clearly the significance of this model, consider the example of an atomic system which decays at time T. This system can be described by $L(P_A)$—the lattice of projections of a Hilbert space defining the atomic system— and $L(P_D)$—the lattice of projections of the Hilbert space defining the decay products. Then

$$L(\bar{P}) = [\vee^{\oplus}_{t\in[0,T)} L(P_A)] \vee^{\oplus} [\vee^{\oplus}_{t\geq T} L(P_D)] \quad (75)$$

is the "dynamical object." Note the dependence on T: $\Lambda(F) = \{0, T, \infty\}$ defines the corresponding information. We have $\mathcal{T}_0 = L(\bar{P})$ and $\mathcal{T}_T \approx \vee^{\oplus}_{t\geq T} L(P_D)$. A measure μ_0 on $\mathcal{T}_0$ acts, in fact, solely on $\vee^{\oplus}_{t\in[0,T)} L(P_A)$ since $\mu_0: \mathcal{T}_0\cap\mathcal{T}_T \to 0$. Note that $\mu_0(\bar{\mu}_i = 1)$ is a *joint measure* on $\vee^{\oplus}_{t\in[0,T)} L(P_A)$ whose marginals on $L(P_A)_t$ define the dynamically evolving "states" $\hat{W}_A(t)$ on $\mathcal{H}_A$. The same remarks apply to μ_T on $\mathcal{T}_T$. It should be appreciated that the present description in fact extends the ideas of Piron [see Ref. 35, pp. 117–119] and explains the lattice theoretic model for general irreversible systems in terms of information theory.

The potentiality of the present viewpoint revolves about the interpretation of (71) as an information measure. If one knew precisely J, F, and $L(\bar{P})$, (71) would be fixed and there would be complete knowledge of the "object." In reality, we can only *infer* J, F, and $L(\bar{P})$ from generally inadequate evidence (or hypotheses). Thus, the role of information theory in the inductive process, described in the Introduction, can be exploited in the present context to seek estimates of J and $\Lambda(F)$, e.g., if we fix θ_i and the distinct $L(P_z)$'s. We shall consider these applications elsewhere.

V. CONCLUSION

In this paper we have attempted to provide a universal, intuitively reasonable axiomatic description of "entropy." Our approach takes seriously the information-theoretic interpretation of "entropy," and explores the universal explication of this concept in the context of what might be called "algebraic measurement theory." Although we have not characterized our "entropy" species, we have found that with sufficient effort one can (more or less) recover the more common entropy formulas as special cases of our definition. Besides this encouraging result, we have discovered a surprising relationship between semigroups on $\overline{R^+}$ and Piron's model for irreversibility. This, in conjunction with information-theoretic inference procedures, suggests a new (and relatively simple) way to devise models, etc. for irreversible processes.

While this paper represents a first toddle in hopefully the right direction, it clearly leaves much work to be done. Most urgent, perhaps, is the incorporation of an appropriate "tensor product" in the scheme to describe compound objects whose components preserve a certain identity: The "entropy" of such a compound system will evidently suffer restrictions similar to the usual (sub-)additivity that favors the logarithmic function in a central entropic role. Beyond this, one can establish a "communication theory" between such abstract "languages" which (among interesting social potentials) could perhaps form the basis of a quantitative "metatheory" to compare distinct mathematical theories or physical models. We have already spoken at length of the more and more respectable techniques of information-theoretic

inference which provide the means to use the full "data-processing" potential of a theory to predict based on observations—as well as defining useful criteria for the relevance of parameters within a theory. We anticipate that the "communication theory" envisioned will play a pivotal role in establishing the universal validity of a generalized MEF[17] and in removing the various difficulties it now suffers.

In closing, we note that our axioms already disqualify certain candidates from the entropy zoo. In particular, we have pointed out that the classical and quantal "entropies" correspond to *infinite* global information ($h_i \leqslant \infty$ actually) when the "prior" information corresponds to an infinite measure. In fact, this illustrates the obvious: That in "real-life" priors based on invariance over infinite spaces (Lebesgue measure classically, infinite Hilbert spaces quantally) require an infinite amount of information to specify. In reality, one should account for the finite limits of the *observable* cosmos (operational view) by appropriately delimiting the priors. Otherwise, one must apparently seek other formulas for "entropy."[52]

Noted added in Proof: We have discovered an omission in conditions (43)–(45): Further consequences of Axioms 2 and 4 compel $\psi_i(z)$ in (47) to vanish. Deletion of these axioms permits the representation (20) without $-\phi^{\mathrm{F}}(0,\ldots,0)$, where ϕ^{F} has real range and (19c) is not in force. In this case $\psi_i(z) = z \ln z$ is admissable in (47) and, since non-negativity of ψ_i is not essential in deriving (58) (see Ref. 51), we obtain $H^{\mathrm{F}}(J_1,J_2) = -\Sigma_{i\in I} S(\hat{W}_i)$. If $\hat{W}_i^2 \leqslant \hat{W}_i$, $S(\hat{W}_i) \geqslant 0$ [M. B. Ruskai, Ann. Inst. H. Poincaré **19**, 357 (1973)] so that $H_t^{\mathrm{F}}(J_1,J_2)$ is monotonic (non-increasing) in "time" (Sec. IV). The error noted here does not effect any other result of this paper.

ACKNOWLEDGMENTS

This work has been funded in part by a research grant from the N.H.R.F. and in part by the Theoretical and Physical Chemistry Institute of the N.H.R.F. I wish to acknowledge the hospitality of the members of this Institute, and useful discussions with Dr. I. Komninos and Prof. C. A. Nicolaides.

[1]A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[2]J.-P. Marchand and W. Wyss, J. Stat. Phys. **16**, 349 (1977).
[3]W. Band and J. L. Park, Found. Phys. **6**, 249 (1976).
[4]M. D. Srinivas, J. Math. Phys. **19**, 1952 (1978).
[5]M. Mastrangelo and V. Mastrangelo, Ann. Inst. Henri Poincaré **A 30**, 295 (1979).
[6]B. Robertson, in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus, (MIT Press, Cambridge, 1979), p. 289.
[7]J. Mehra and E. C. G. Sudarshan, N. Cim. **B 11**, 215 (1972).
[8]I. Prigogine, C. George, F. Henin, and L. Rosenfeld, Chem. Scripta **4**, 5 (1973).
[9]I. J. Good, Science **129**, 443 (1959).
[10]T. Fine, *Theories of Probability* (Academic, New York, 1973).
[11]J. Rayski, Found. Phys. **3**, 89 (1973); **7**, 151 (1977); **9**, 217 (1979).
[12]J. F. Cyranski, Found. Phys. **8**, 493 (1978).
[13]E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957).
[14]L. Brillouin, *Science and Information Theory* (Academic, New York, 1956).
[15]J. B. Báron and J. M. Jauch, Helv. Phys. Acta **45**, 220 (1972).
[16]R. Carnap, *Logical Foundations of Probability* (University of Chicago, Chicago, 1950).
[17](a) B. J. Berne, P. Pechukas, and G. D. Harp, J. Chem. Phys. **49**, 3125 (1968). (b) H. Heiner, Phys. Stat. Sol. (b) **91**, 87 (1979). (c) S. Gadre and S. B. Sears, J. Chem. Phys. **71**, 4321 (1979). (d) C. Y. She and H. Heffner, Phys. Rev. **152**, 1103 (1966). (e) B. Robertson, Phys. Rev. **144**, 151 (1966); **160**, 175 (1967). (f) A. Corbet and H. Morowitz, Phys. Rev. **A 6**, 2298 (1972). (g) J. D. Bekenstein, Phys. Rev. **D 12**, 3077 (1975). (h) A. G. Wilson, *Entropy in Urban and Regional Modelling* (Pion, London, 1970). (i) See also articles and Refs. in: *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT, Cambridge, 1979).
[18]W. Ochs, Rep. Math. Phys. **8**, 109 (1975); **9**, 135 (1976); and **9**, 331 (1976).
[19]J. E. Shore and R. W. Johnson, I.E.E.E. Trans. Inf. Theory **IT-26**, 26 (1980).
[20]The MEF suffers from the restrictive form the evidences must take and the problem of choosing the correct "prior" measure. We plan to offer at least a partial solution (classical case) to these problems elsewhere.
[21]J. F. Cyranski, Found. Phys. **8**, 805 (1978).
[22]J. F. Cyranski, Inform. Contr. **41**, 275 (1979).
[23]J. F. Cyranski, Found. Phys. **9**, 641 (1979).
[24]J. Sallantin, C. R. Acad. Sci. Paris **A 274**, 986 (1972).
[25]There exist numerous characterization theorems for various "entropies." See Ref. 1 for examples, and Refs. 18.
[26]For a summary of these models, see *The Logico-Algebraic Approach to Quantum Mechanics*, vols. I and II, edited by C. A. Hooker (Reidel, Dordrecht, Holland, 1975; 1979).
[27]G. Birkhoff, *Lattice Theory*, 3rd ed. (Am. Math. Soc., Providence, 1967).
[28]S. Watanabe, Inform. Contr. **15**, 1 (1969).
[29]J. Sallantin, *Structures de L'Information* No. 8 (C.N.R.S., Paris 1979).
[30]A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*, Vol. I (Academic, New York, 1975).
[31]M. M. Gupta, G. N. Saridis, and B. R. Gaines, editors, *Fuzzy Automata and Decision Processes* (North-Holland, Amsterdam, 1977).
[32]J. A. Goguen, J. Math. Anal. Appl. **18**, 145 (1967).
[33]A. De Luca and S. Termini, J. Math. Anal. Appl. **40**, 373 (1972).
[34]S. T. Ali and E. Prugovecki, J. Math. Phys. **18**, 219 (1977).
[35]C. Piron, *Foundations of Quantum Mechanics* (Benjamin, New York, 1976).
[36]Ref. 27; for other completion alternatives, see. W. R. Tunnicliffe, Proc. Lond. Math. Soc. **3**, 13 (1974).
[37]M. D. MacLaren, Pac. J. Math. **14**, 597 (1964); K. Bugajska and S. Bugajski, Ann. Inst. Henri Poincaré **A 19**, 333 (1973).
[38]This proof extends the relevant part of Lemma (29.11) of F. Maeda and S. Maeda, *Theory of Symmetric Lattices* (Springer, Berlin, 1970).
[39]P. R. Halmos, *Measure Theory* (van Nostrand, Princeton, N.J., 1950).
[40]Strictly speaking, $L(P)$ is not a "logic." However, recent work suggests how an *operation* of implication (conditional) may be introduced in general (at least, orthomodular) lattices: G. M. Hardegree, Z. Naturforsch. **30A**, 1347 (1975). Moreover, quantifiers can be introduced as *mappings* (Boolean context), which suggests a possible avenue to extend quantification to $L(P)$: P. Halmos, *Algebraic Logic* (Chelsea, New York, 1962). If these extensions are accomplished, one could define a theory of "semantic information" on a good deal of everyday language. J. F. Cyranski, in *Advances in Communications*, edited by D. G. Lainiotis and N. S. Tzannes (Reidel, Dordrecht, 1980).
[41]G. Comyn and J. Losfeld, C. R. Acad. Sci. Paris **A 278**, 633 (1974); see also Ref. 24.
[42]J. Kampé de Fériet and P. Benvenuti, C. R. Acad. Sci. Paris **A 274**, 655 (1972).
[43]J. Kampé de Fériet, B. Forte, and P. Benvenuti, C. R. Acad. Sci. Paris **A 269**, 529 (1969).
[44]B. Forte and C. T. Ng, Ann. Mat. Pura. Appl. **101**, 355 (1974).
[45]B. Forte, in *Théorie de L'Information: Développements Récents et Applications*, edited by C. Picard (Edits. du C.N.R.S., Paris, 1978). Forte argues that in statistical physics neither the degrees of freedom nor the phase space (subset of R^{2n}) are known, so that the "entropy" cannot depend solely on the distribution $f_{2n}(x)$ over R^{2n}. Our model has $L(P)$ fixed. See Sec. 4, however.
[46]B. Forte, R.I.R.O. **3**, *R*-2, 63 (1969).
[47]A. Wehrl, Found. Phys. **9**, 939 (1979).
[48]G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, 2nd ed. (Cambridge U. P., London, 1973).
[49]For details see R. Cirelli and F. Gallone, Ann. Inst. H. Poincaré **A 19**, 7

(1973). If one prefers the algebraic method, one can utilize the fact that the projectors of a W^*-algebra form a "quantum logic" and use *this* "logic" as $\mathscr{S}_{a_i}$. See V. S. Varadarajan, *Geometry of Quantum Theory*, Vol. 1 (van Nostrand, Princeton, N.J., 1968).

[50]W. Ochs and H. Spohn, Rep. Math. Phys. **14**, 75 (1978).

[51]S. G. Ghurye, Ann. Math. Stat. **38**, 2056 (1968).

[52]See Ref. 21 and E. T. Jaynes, IEEE Trans. Syst. Sci. Cyber. **SSC-4**, 227 (1968). If $\nu(\Omega) < \infty$ ($M = 2$ context) in general unitary invariance will be lost. Thus, alteration of the "prior" from a "trace" to a weight (Ref. 50, fn. 4) will necessarily modify the form of the "entropy," since $n(\hat{X}) = m(\hat{T}\hat{X}\hat{T})$, $\hat{T} > 0$, is the new "Radon–Nikodym" Theorem. See S. Sakai, Bull. Am. Math. Soc. **71**, 149 (1951).

Yang–Mills classical solutions with a remaining U(1) invariance

Yves Brihaye
University of Mons, Avenue Maistriau, 7000 Mons, Belgium

(Received 17 September 1980; accepted for publication 6 February 1981)

Recently, sufficient conditions have been established, characterizing Yang-Mills fields invariant under a space–time-dependent U(1) subgroup of the full gauge symmetry group SU(2). In this paper we make precise the meaning of the equations and extend the known solutions for larger classes of sources.

PACS numbers: 11.30.Qc, 11.10.Np

I. INTRODUCTION

There have been many attempts to generalize the t'Hooft monopole since its discovery.[1]

The authors of Ref. 2 have studied solutions of a Yang–Mills theory whose gauge potentials are invariant under gauge subgroup of the full gauge group, as is the case for the monopole.

The purpose of this paper is to improve their solutions in such a way that they become regular outside a compact subset of $\mathbb{R}^3$ (space) (see Sec. III) or $\mathbb{R}^4$ (space–time) (see Sec. IV). Section II recalls the ideas of Ref. 2 and the equations to be solved. Some remarks about the meaning of these equations are made.

II. GENERALITIES

Let us consider the Euclidean version of the Yang–Mills theory with a triplet of Higgs scalars described by the Lagrangian density:

$$\begin{aligned} &L = -\tfrac{1}{8}\mathrm{tr}(F^{\mu\nu}F_{\mu\nu}) + \tfrac{1}{4}\mathrm{tr}(D^{\mu}\phi)(D_{\mu}\phi) - V(\phi), \\ &F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} + i[A_{\mu},A_{\nu}], \\ &D_{\mu}\phi = \partial_{\mu}\phi + i[A_{\mu},\phi], \\ &V(\phi) = \lambda\,(\tfrac{1}{2}\mathrm{tr}\phi^2 - a^2)^2 \quad (a = 1 \text{ in what follows}), \end{aligned} \tag{1}$$

$$\begin{aligned} &A_{\mu} = A_{\mu}^{i}\sigma^{i} \quad (\mu = 0,1,2,3), \\ &\phi = \phi^{i}\sigma^{i}, \end{aligned} \tag{2}$$

where σ_i are the Pauli matrices and λ and a are constants. The corresponding equations of motion are

$$\begin{aligned} &D_{\mu}F^{\mu\nu} = \tfrac{1}{2}[\phi,D_{\mu}\phi], \\ &D^{\mu}D_{\mu}\phi = -\lambda\phi\,(\tfrac{1}{2}\mathrm{tr}\phi^2 - 1). \end{aligned} \tag{3}$$

The ansatz of Bacry and Nuyts (see Ref. 2, B–N in what follows) consists in the construction of gauge potentials A_{μ} invariant under a U(1) subgroup [with generator called $G(x)$] of the full gauge group SU(2); i.e.,

$$A'_{\mu} = A_{\mu} = UA_{\mu}U^{-1} - iU\partial_{\mu}U^{-1},$$

where

$$U = U(x) = \exp[i\alpha G(x)]. \tag{4}$$

Taking an infinitesimal U, it is easy to see that A_{μ} must be a function of the matrix G in such a way that

$$D_{\mu}G \equiv \partial_{\mu}G + i[A_{\mu},G] = 0. \tag{4'}$$

Equation (4') requires that $\mathrm{tr}G^2$ be space–time-independent. If we take the generator $G(x)$ equal to the Higgs field $\phi(x)$ [$G(x) \equiv \phi(x)$] and if we normalize it such that it minimizes the Higgs potential [$V(\phi) = 0$], the only equation of motion still to be satisfied is

$$D_{\mu}F^{\mu\nu} = 0. \tag{5}$$

In Ref. 2 the authors construct A_{μ}, the solution of Eq. (4'), in terms of the matrix $\phi(x)$ and of an arbitrary classical vector field $\alpha_{\mu}(x)$:

$$A_{\mu} = i[\partial_{\mu}\phi,\phi] + \alpha_{\mu}\phi. \tag{6}$$

Furthermore, they express both Eq. (5) and the Bianchi identities as a Maxwell-like system:

$$\partial_{\mu}f_{\mu\nu} = 0, \tag{7a}$$

$$\partial_{\mu}\epsilon_{\mu\nu\rho\sigma}f_{\rho\sigma} = 0, \tag{7b}$$

where

$$f_{\mu\nu} = -2i\,\mathrm{tr}(\phi\,[\partial_{\mu}\phi,\partial_{\nu}\phi]) + \partial_{\mu}\alpha_{\nu} - \partial_{\nu}\alpha_{\mu}. \tag{7c}$$

The most general $\phi(x)$ compatible with Eq. (4') may be parametrized in terms of two arbitrary functions $\omega(x)$ and $g(x)$:

$$\phi = \begin{pmatrix} g & (1-g^2)^{1/2}e^{i\omega} \\ (1-g^2)^{1/2}e^{-i\omega} & -g \end{pmatrix}.$$

With this parametrization $f_{\mu\nu}$ takes a simple form [see (8c)].

A completely regular and differentiable field $\phi(x)$ would be gauge-equivalent to a constant one

$$\left[\phi \div \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$$

and one would be topologically trivial. In order to obtain more interesting solutions, such as monopoles, B–N have solved a system like (7) by replacing the null source in Eq. (7a) [resp. 7(b)] by a localized and conserved electric (resp. magnetic) current, i.e.,

$$\partial_{\nu}f_{\mu\nu} = J^{e}_{\mu}, \tag{8a}$$

$$\partial_{\nu}\epsilon_{\mu\nu\rho\sigma}f_{\rho\sigma} = J^{m}_{\mu}, \tag{8b}$$

$$f_{\mu\nu} = (\partial_{\mu}\omega)(\partial_{\nu}g) - (\partial_{\nu}\omega)(\partial_{\mu}g) + \partial_{\mu}\alpha_{\nu} - \partial_{\nu}\alpha_{\mu}. \tag{8c}$$

Let us make some mathematical remarks about this system:

(a) The current components J^{e}_{μ} and J^{m}_{μ} and the field components $f_{\mu\nu}$ have to be considered as distributions over $\mathbb{R}^3$ (space) or $\mathbb{R}^4$ (space–time). Equations (8a) and (8b) together with the condition $f_{\mu\nu}(\infty) = 0$ can be considered as defining $f_{\mu\nu}$, given the currents.

(b) The functions g and ω (and their derivatives) in Eq.

 0022-2488/81/071479-03$1.00

(8c) have to be taken as functions only in a domain D where they are differentiable and continuous, with $D \subset \mathbb{R}^3$ or $\mathbb{R}^4$, depending on the solutions considered.

In Ref. 2 B–N exhibit functions ω and g for various J^m_μ and for $J^e_\omega = 0$ [the arbitrariness on α_μ in Eq. (8a) is sufficient to allow its contribution to kill the electric current J^e_μ]. But some of their solutions have singularities extending to infinity. They don't give an acceptable asymptotic behavior to the field $\phi(x)$.

In the next section we construct solutions with acceptable asymptotic behaviors for the specific forms of the magnetic current considered by B–N.

III. STATIC SOLUTIONS

The static hypothesis is characterized by

$$\partial_0 \omega = \partial_0 g = 0,$$

$$J^m_0 = \rho = p(\nabla) 4\pi \delta^3(\mathbf{x}), \quad J^m_i = 0 \quad (i = 1,2,3), \tag{9}$$

with $p(\mathbf{x})$ a homogeneous polynomial in the space variables. Equation (8) then reduces to

$$\nabla \Lambda \mathbf{f} = 0 \Rightarrow \mathbf{f} = -\nabla V, \tag{10a}$$

$$\nabla \cdot \mathbf{f} = \rho \Rightarrow V = -p(\nabla)(1/R), \tag{10b}$$

$$\mathbf{f} = \nabla \omega \Lambda \nabla g, \tag{10c}$$

with

$$R^2 = x_1^2 + x_2^2 + x_3^2.$$

Given V in the form (10a,b), the first problem is to find two scalar functions with gradient orthogonal to ∇V. Let us note a property that we will use in the following:

If $p(\mathbf{x})$ and $q(\mathbf{x})$ are two homogeneous polynomials of degree d such that

$$V = p(\nabla)(1/R) = q(\mathbf{x})/R^{2d+1}$$

and if $m(\mathbf{x})$ is a function such that

$$\nabla m \cdot \nabla q = 0,$$

then we have for $t = (2d+1)/(d+1)$ and $g(\mathbf{x}) = m(\mathbf{x}/R^t)$

$$\nabla V \cdot \nabla g = 0.$$

This property allows us to construct solutions ω and g of Eq. (10c) for the $\mathbf{f}$ considered by B–N. In each case we give g_1 and g_2, two elementary g functions, defined in $\mathbb{R}^3$ (origin excepted) ($\mathbb{R}^3 \setminus \{0\}$). The true functions ω and g may be obtained as two continuous and differentiable functions of g_1 and g_2 such that

$$\left.\begin{array}{l} g = g(g_1,g_2) \\ \omega = \omega(g_1,g_2) \end{array}\right\} \left| \begin{array}{cc} \dfrac{\partial \omega}{\partial g_1} & \dfrac{\partial \omega}{\partial g_2} \\ \dfrac{\partial g}{\partial g_1} & \dfrac{\partial g}{\partial g_2} \end{array} \right| = 1. \tag{11}$$

The functions ω and g are still regular in $D = \mathbb{R}^3 \setminus \{0\}$; note that a possible choice is $g = g_1$ and $\omega = g_2$ (or $g = g_2$ and $\omega = g_1$).

A. Magnetic dipole

In this case, using the same notation ($d = 1$)

$$p(\mathbf{x}) = \lambda x_3, \quad q(\mathbf{x}) = -\lambda x_3, \quad m_1(\mathbf{x}) = x_1, \quad m_2(\mathbf{x}) = x_2, \tag{12}$$

we choose for g_1 and g_2 (γ_1 and γ_2 are constants)

$$g_1 = \gamma_1 x_1 / R^{3/2}, \quad g_2 = \gamma_2 x_2 / R^{3/2}, \quad \gamma_1 \gamma_2 = 2\lambda. \tag{12'}$$

B. Magnetic quadrupole

As shown in Ref. 2 the most general quadrupole can be obtained from a source ρ [see Eq. (9)] such that

$$p = \alpha x_1^2 + \beta x_2^2, \quad q = 3(\alpha x_1^2 + \beta x_2^2) - (\alpha + \beta) R^2,$$

$$m_1 = x_1 x_3^{(2\alpha - \beta)/(\alpha + \beta)}, \quad m_2 = x_2 x_3^{(2\beta - \alpha)/(\alpha + \beta)}. \tag{13}$$

If $\alpha + \beta \neq 0$, g_1 and g_2 are given by

$$g_1 = \gamma_1 (x_1^{\alpha + \beta} x_3^{2\alpha - \beta} / R^{5\alpha})^{1/(\alpha + \beta)},$$

$$g_2 = \gamma_2 (x_2^{\alpha + \beta} x_3^{2\beta - \alpha} / R^{5\beta})^{1/(\alpha + \beta)},$$

$$\gamma_1 \gamma_2 = 3(\alpha + \beta). \tag{13'}$$

In order for g_1 and g_2 to be defined everywhere asymptotically we see that not all values of α, β are acceptable. In particular, they must satisfy the condition above (see Appendix):

$$\beta/\alpha = (2I - N)/(I + N), \quad 0 \leqslant N \leqslant I, \tag{14}$$

with N an integer and I an odd integer. In the case $\alpha + \beta = 0$, we may choose a basis such that

$$p = \lambda x_1 x_2, \quad q = 3\lambda x_1 x_2, \quad m_1 = x_1^2 - x_2^2, \quad m_2 = x_3, \tag{15}$$

$$g_1 = \gamma_1 (x_1^2 - x_2^2)/R^{10/3}, \quad g_2 = \gamma_2 x_3 / R^{5/3},$$

$$\gamma_1 \gamma_2 = 9\lambda/4. \tag{15'}$$

Furthermore, the hermiticity condition on ϕ [see Eq. (7c)] requires

$$g^2 \leqslant 1. \tag{16}$$

Each function g_i ($i = 1,2$) given above satisfies this last condition outside the surfaces $g_i^2 = 1$ which are in each case included in a compact set of $\mathbb{R}^3$, e.g., the surface

$$g_1^2 = \gamma_1^2 \cdot (x_1^2 / R^3) = 1 \tag{17}$$

is included inside the sphere with radius $R = (\gamma_1)^2$.

Moreover, using Eq. (11), we can define an allowed tranformation $(\omega, g) \rightarrow (\Omega, G)$ such that

$$G = \tanh g, \quad -1 \leqslant G \leqslant +1, \tag{18}$$

$$\Omega = \omega \cdot (\cosh g)^2,$$

and hence the region in which (16) is satisfied is the full $\mathbb{R}^3$ space, origin excepted. This is the crucial point of this paper. To repeat: the transformation (18) guarantees the existence of a Hermitian Higgs field regular in $\mathbb{R}^3 \setminus \{0\}$.

Note: In the dipole case, B–N have taken as ω

$$\omega = \arctan(g_2/g_1) = \arctan(x_2/x_1). \tag{19}$$

This function has a geometrical significance, but unfortunately it introduces singularities of ϕ along the x_3 axis.

IV. NONSTATIC SOLUTIONS

We will consider magnetic current of the form

$$J^m_\mu = p_\mu(\nabla)[2\pi^2 \delta^4(x)], \tag{20a}$$

$$\partial^\mu J^m_\mu = 0, \quad \nabla = \nabla_\nu = (\partial_1, \partial_2, \partial_3, \partial_4), \tag{20b}$$

where the $p_\mu(x)$ are four polynomials of degree d in the variable $x = (x_1, x_2, x_3, x_4)$ constructed in agreement with (20b).

Equation (8a,b) are solved if

$$f_{\mu\nu} = \epsilon_{\mu\beta\rho\sigma}\partial_\rho p_\sigma(\nabla)(1/R^2). \tag{21}$$

Equation (8c) introduces the following requirement which restricts also the possible choices of the polynomials p_μ:

$$\epsilon_{\mu\nu\rho\sigma} f_{\mu\nu} f_{\rho\sigma} = 0. \tag{22}$$

In the following, we present solutions (ω and g) corresponding to the choices of p_μ:

$$p_\mu = \alpha(0,0,-x_4,x_3) \equiv \alpha p_\mu^{\mathrm{BN}}, \tag{23a}$$

$$p_\mu = (\alpha x_1) p_\mu^{\mathrm{BN}}, \tag{23b}$$

$$p_\mu = \gamma(\alpha x_1^2 + \beta x_2^2) p_\mu^{\mathrm{BN}}, \tag{23c}$$

$$p_\mu = (\beta x_2^2 + \gamma x_3^2 + \delta x_4^2, -\beta x_1 x_2, -\gamma x_1 x_3, -\delta x_1 x_4). \tag{23d}$$

These particular forms for the current have been chosen in cannonical space–time positions in such a way that they obey the condition (22). We have not been able to find a general condition for polynomials with arbitrary degree.

Below we write the solutions [up to a transformation like (11)] for the different cases.

(a) Now $R^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$:

$$g_1 = \lambda_1 x_1/R^2, \quad g_2 = \lambda_2 x_2/R^2, \quad \lambda_1\lambda_2 = 4\alpha. \tag{24a}$$

The solution called the dipole ephemeron in Ref. 2 can be obtained in its original form by the transformation (11):

$$\omega = \arctan\left(\frac{\lambda_1}{\lambda_2}\frac{g_2}{g_1}\right), \quad g = 2\alpha\left[\left(\frac{g_1}{\lambda_1}\right)^2 + \left(\frac{g^2}{\lambda_2}\right)^2\right]; \tag{24b}$$

but it is singular if $x_1 = x_2 = 0$ and hence (24a) is a better form.

(b) $$g = 2\alpha x_2/R^2, \quad \omega = (2x_1^2 + 2x_2^2 - x_3^2 - x_4^2)/R^4. \tag{25}$$

In these two cases the solutions are acceptable asymptotically, a transformation like (19) giving a Higgs field Hermitian and regular on all space–time except the origin

(c) If we suppose $\alpha + \beta \neq 0$ (the case $\alpha + \beta = 0$ is completely treated in Ref. 3), the solutions can be written

$$g_1 = \lambda_1 x_1 Y^{n_1}/(R^2)^{m_1}, \quad g_2 = \lambda_2 \frac{x_2 Y^{n_2}}{(R^2)^{m_2}}, \quad \lambda_1\lambda_2 = -8\gamma, \tag{26}$$

with $\alpha + \beta = 1$, we obtain

$$Y = \frac{1-6\alpha}{1+6\alpha}x_1^2 + \frac{6\alpha-5}{7-6\alpha}x_2^2 + x_3^2 + x_4^2, \tag{27}$$

$$n_1 = \frac{6\alpha+1}{4}, \quad n_2 = \frac{5-6\alpha}{4}, \quad m_1 = \frac{6\alpha+1}{2}, \quad m_2 = \frac{7-6\alpha}{2}.$$

For the same reasons as for the magnetic quadrupole [see (14)] these solutions are regular asymptotically if and only if

$$\alpha = (4N+I)/6I, \quad 0 \leqslant N \leqslant I \tag{28}$$

with N an integer and I an odd integer.

(d) $$g = 2x_2^{(\gamma-\delta)} x_3^{(\delta-\beta)} x_4^{(\beta-\delta)},$$

$$\omega = x_2^{1+\delta-\gamma} x_3^{1+\beta-\delta} x_4^{1+\gamma-\beta}/R^6 \tag{29}$$

It is easy to see that the function g is asymptotically singular somewhere because the sum of the exponents is zero. Hence one at least must be negative. Even so this solution is interesing because it is the only known solution which has no cylindrical symmetry in the (x_3,x_4) hyperplane.

Moreover, for specific values of the constants β,γ,δ we have been able to find a transformation like (11) allowing the new functions ω' and g' to be defined everywhere outside the origin point.

Let's choose

$$(\delta-\gamma,\beta-\delta,\gamma-\beta) = [1/(N+N'+N'')](2N-N-N', 2N'-N-N'', 2N''-N-N'), \tag{30}$$

where N, N', N'' are integers. Then

$$g' = g\omega^p, \quad \omega' = \omega^{1-p}/(1-p) \tag{31}$$

are as required above, provided

$$p = 1 - (N+N'+N''/3I), \quad I > N,N',N'', \tag{32}$$

The integer I being odd for the same reasons as before.

V. CONCLUSION

We have obtained in this paper a number of new classical solutions invariant under a U(1) subgroup of an SU(2) gauge theory.

We now list two open questions which we hope to solve in the near future:

(a) generalization of static and nonstatic solutions to all possible pointlike sources;

(b) generalization to larger groups of the basic equations.[4]

ACKNOWLEDGMENTS

I am very grateful to J. Nuyts and A. Martin for reading the manuscript.

APPENDIX

In general for any positive number r, the real function

$$f: x \to x^r \tag{A1}$$

is defined over the full $\mathbb{R}$ provided that r is a rational of the reduced from

$$r = N/I \tag{A2}$$

with N an positive integer, I an odd positive integer. So, in order for the functions g_1 and g_2 [see (13')] to be real asymptotically for positive as well as negative values of x_3, the exponents of x_3 have to be of the form [see (A2)]

$$0 \leqslant (2\alpha-\beta)/(\alpha+\beta) = N/I, \tag{A3}$$

$$0 \leqslant (2\beta-\alpha)/(\alpha+\beta) = 1 - N/I = (I-N)/I. \tag{A4}$$

This implies

$$\beta/\alpha = (2I-N)/(I+N) \tag{A5}$$

and clearly from (A3) and (A4)

$$0 \leqslant N \leqslant I. \tag{A6}$$

This establishes formula (14). The same argument can be used to prove (28) and (32).

[1] G. t'Hooft, Nucl. Phys. B **79**, 276 (1974).
[2] H. Bacry and J. Nuyts, Nucl. Phys. B **151**, 469 (1979).
[3] Y. Brihaye and R. M. Dubois, (to be published).
[4] Y. Brihaye, J. Nuyts, and A. Taormina, (in preparation).

Matrix elements of operators in symmetrical U(6)⊃U(3)⊃U(2)⊃U(1) and U(6)⊃SU(3)⊃SO(3)⊃SO(2) basis

C. Quesne [a)]
Service de Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, B-1050 Brussels, Belgium

(Received 30 September 1980; accepted for publication 12 December 1980)

In this paper we consider a basis for N boson states classified by the chain of groups U(6)⊃U(3)⊃U(2)⊃U(1). We determine analytical expressions for the matrix elements of the boson creation and annihilation operators, as well as for the U(6) generators, with respect to that basis. For such purpose, we use the Wigner–Eckart theorem with respect to U(3) and calculate the reduced matrix elements of the appropriate irreducible tensors from their matrix elements between highest weight states. Then we apply the transformation brackets from the chain U(6)⊃U(3)⊃U(2)⊃U(1) to the physical chain U(6)⊃SU(3)⊃SO(3)⊃SO(2) to obtain the matrix elements of the above-mentioned operators in a basis corresponding to the latter. The matrix elements so determined can be used either in the nuclear interacting boson model of Arima and Iachello or in the microscopic nuclear collective model of Vanagas *et al.*

PACS numbers: 21.60.Fw, 21.60.Cs, 02.20.Nq, 02.20.Qs

1. INTRODUCTION

In nuclear physics authors of recent developments have paid much attention to the basis states of the symmetrical irreducible representation (IR) $[N]$ of the unitary group U(6). The U(6) group has been introduced in various ways with the purpose of giving a unified description of collective states in nuclei. In the work of Dzholos, Dönau, and Janssen,[1] and later in that of Arima and Iachello,[2] the U(6) group is used in connection with active bosons, created by coupled pairs of fermions. In the work of Vanagas and co-workers,[3] it is defined as acting on the six collective variables that can be formed from bilinear products of Jacobi coordinates. Although in this paper we shall adopt the language of Arima and Iachello's interaction boson model and therefore speak of s and d bosons, the results as obtained will be also valid for the other approaches.

Three chains of subgroups of U(6) are considered, namely,

$$\mathrm{U}(6)\supset\mathrm{U}(5)\supset\mathrm{SO}(5)\supset\mathrm{SO}(3)\supset\mathrm{SO}(2), \tag{1.1}$$

$$\mathrm{U}(6)\supset\mathrm{SU}(3)\supset\mathrm{SO}(3)\supset\mathrm{SO}(2), \tag{1.2}$$

and

$$\mathrm{U}(6)\supset\mathrm{SO}(6)\supset\mathrm{SO}(5)\supset\mathrm{SO}(3)\supset\mathrm{SO}(2). \tag{1.3}$$

The chain (1.2), to be considered here, corresponds to the limit of the axial rotor of the Bohr and Mottelson geometrical model.[4]

The knowledge of the matrix elements of various operators with respect to the basis corresponding to the chain (1.2) is required. In the interacting boson model, such operators are for instance the boson creation and annihilation operators, which connect nuclei differing by a pair of nucleons, and the generators of U(6), in whose terms the most general interaction can be expressed.

In a recent work,[5] Castaños, Chacón, Frank, and Moshinsky show that the basis states of the IR $[N]$, classified according to the chain (1.2), can be obtained from those classified according to the chain (1.1) by diagonalizing the matrix of the operator

$$\mathbf{Q}^2=\sum_{m=-2}^{+2}(-1)^m Q_m Q_{-m}. \tag{1.4}$$

Here Q_m ($m=\pm 2,\ \pm 1, 0$) are the components of the quadrupole operator, which, together with the components of the angular momentum $L_\tau(\tau=\pm 1,0)$, form the generators of SU(3) classified by the subgroups SO(3)⊃SO(2). One can then proceed in the basis corresponding to the chain (1.1), which has been extensively studied,[5,6] and calculate the matrix elements of all relevant operators in that basis.

With this paper, we propose an alternative approach, where the matrix elements of the operators under consideration are determined in the canonical basis for U(3), corresponding to the following chain of groups

$$\mathrm{U}(6)\supset\mathrm{U}(3)\supset\mathrm{U}(2)\supset\mathrm{U}(1). \tag{1.5}$$

The matrix elements in the basis corresponding to the chain (1.2) can then be deduced with the help of the well known transformation brackets between the two basis.[7,8] In this way, one can get analytical expressions for the relevant matrix elements as opposed to the numerical values which result from the diagonalization procedure used in Ref. 5. Analytical expressions may be quite useful in some respects. This point will be illustrated in a forthcoming paper, where the formalism developed in this work will be applied to the study of the shape of nuclei in the rotational collective model.[9]

In the next section, we begin by defining our notations for the generators of the groups appearing in the chains (1.2) and (1.5). The following three sections are then concerned with the calculation of matrix elements in the chain (1.5). In Sec. 3, we discuss the construction of the basis states of a symmetrical IR of U(6) in that chain. In Secs. 4 and 5, we obtain the matrix elements of the boson creation and annihilation operators and of the U(6) generators, respectively. Finally, in the concluding section, we indicate how to express the matrix elements in the chain (1.2) in terms of those in the

[a)] Maître de recherches F.N.R.S.

 0022-2488/81/071482-15$1.00

chain (1.5) by using the transformation brackets between the two basis.

2. THE CHAINS U(6)⊃SU(3)⊃SO(3)⊃SO(2) AND U(6)⊃U(3)⊃U(2)⊃U(1)

We shall start by giving the notations required for the generators of the groups contained in the chains (1.2) and (1.5). Let us begin with the physical chain (1.2).[5] The creation operators of s and d bosons are denoted by $\bar{\eta}$ and η_m ($m = \pm 2, \pm 1, 0$), respectively, and the corresponding annihilation operators by $\bar{\xi}$ and ξ^m ($m = \pm 2, \pm 1, 0$). Here m is the projection of the d boson angular momentum on the quantization axis. On occasions, we shall use the single symbols

$$\eta_{lm}(l = 0,2, -l \leqslant m \leqslant l), \quad \eta_{00} = \bar{\eta}, \quad \eta_{2m} = \eta_m, \tag{2.1}$$

and

$$\xi^{lm}(l = 0,2, -l \leqslant m \leqslant l), \quad \xi^{00} = \bar{\xi}, \quad \xi^{2m} = \xi^m, \tag{2.2}$$

to designate the whole set of creation or annihilation operators, respectively.

In terms of those operators, the generators of U(6) are written

$$\mathscr{C}^{l'm'}_{lm} = \eta_{lm}\xi^{l'm'}, \quad l = l' = 0,2. \tag{2.3}$$

The three generators of SO(3) are given by

$$\begin{aligned} L_\tau &= \sqrt{6}\sum_{mm'}\langle 2m1\tau|2m'\rangle\eta_{m'}\xi^m \\ &= \sqrt{5/2}\sum_{mm'}\langle 2m2m'|1\tau\rangle(\eta_m\xi_{m'} - \eta_{m'}\xi_m), \\ \tau &= \pm 1,0, \end{aligned} \tag{2.4}$$

where $\langle|\rangle$ is an ordinary SU(2) Wigner coefficient, and the covariant form of the annihilation operators is defined by

$$\xi_m = (-1)^m\xi^{-m}. \tag{2.5}$$

The operator L_0 generates SO(2).

To get the generators of SU(3), one has to add to the three components $L_\tau(\tau = \pm 1,0)$ of the angular momentum, the five components of the quadrupole operator defined by

$$\begin{aligned} Q_m &= -\sqrt{8\pi/15}\sum_{l'm'l''m''}\langle 2l'm'|r^2Y_{2m}(\theta,\varphi)|2l''m''\rangle \\ &\times\eta_{l'm'}\xi^{l''m''} = \sqrt{7/3}\,[\eta\times\xi]^2_m + 2/\sqrt{3}(\bar{\eta}\xi_m + \eta_m\bar{\xi}), \\ m &= \pm 2, \pm 1,0. \end{aligned} \tag{2.6}$$

Here the ket $|2lm\rangle$ is a two-quantum harmonic oscillator state with the angular momentum and projection indicated, while $[\eta\times\xi]^2_m$ is defined by

$$[\eta\times\xi]^2_m = \sum_{m'm''}\langle 2m'2m''|2m\rangle\eta_{m'}\xi_{m''}. \tag{2.7}$$

Let us pass on to the canonical chain (1.5). As the group theoretical structure of the interacting boson model is similar to that of the many-fermion spectroscopy in the $2s$-$1d$ shell of the three-dimensional harmonic oscillator, which was extensively discussed many years ago,[10,11] the results obtained for the latter case can be applied to the former one, provided fermions be replaced by bosons.

In the corresponding harmonic oscillator problem,[11] the six two-quantum states are obtained by distributing the two quanta over the three one-quantum states $a^+_q|\bar{0}\rangle$, where a^+_q ($q = \pm 1, 0$) denotes the spherical components of the one-quantum creation operators, and $|\bar{0}\rangle$ is the oscillator ground state. The two-quantum states are denoted by $|n_1n_0n_{-1}\rangle$, where n_q ($q = \pm 1, 0$) is the number of quanta of the type a^+_q, and $n_1 + n_0 + n_{-1} = 2$. They form a basis of the IR [200] of the symmetry group U(3) of the harmonic oscillator in the chain U(3)⊃U(2)⊃U(1). This U(3) group is generated by the operators

$$c_q{}^{q'} = a^+_q a^{q'}, \quad q,q' = \pm 1, 0. \tag{2.8}$$

If we choose to enumerate the values of the index q in the order

$$q \to i: \quad +1 \to 1, \quad 0 \to 2, \quad -1 \to 3, \tag{2.9}$$

the states $|n_1n_0n_{-1}\rangle$ can be identified with the Gel'fand states

$$\left|\begin{matrix} 2 & 0 & 0 \\ & n_1 + n_0 & 0 \\ & & n_1 \end{matrix}\right\rangle = |n_1n_0n_{-1}\rangle. \tag{2.10}$$

If moreover we enumerate the states of the IR [200] in the order of decreasing weight, they can be represented by a single index μ, which takes the values $\mu = 1, 2, \cdots, 6$:

$$\begin{aligned} (n_1n_0n_{-1}) \to \mu: \quad & (200)\to 1, \quad (110)\to 2, \quad (101)\to 3, \\ & (020)\to 4, \quad (011)\to 5, \quad (002)\to 6. \end{aligned} \tag{2.11}$$

Finally, the relations between the states $|\mu\rangle = |n_1n_0n_{-1}\rangle$ and the two-quantum states $|2lm\rangle$, classified according to the chain SU(3)⊃SO(3)⊃SO(2), can be deduced from Eq. (6.21a) of Ref. 11 and written:

$$\begin{aligned} &|1\rangle = |222\rangle, \quad |2\rangle = |221\rangle, \\ &|3\rangle = 3^{-1/2}(|220\rangle + \sqrt{2}|200\rangle), \\ &|4\rangle = 3^{-1/2}(\sqrt{2}|220\rangle - |200\rangle), \\ &|5\rangle = |22-1\rangle, \quad |6\rangle = |22-2\rangle. \end{aligned} \tag{2.12}$$

In a similar way, we define boson creation operators ζ_μ, $\mu = 1, \cdots, 6$, classified according to the chain (1.5), by the following relations:

$$\begin{aligned} &\zeta_1 = \eta_2, \quad \zeta_2 = \eta_1, \quad \zeta_3 = 3^{-1/2}(\eta_0 + \sqrt{2}\bar{\eta}), \\ &\zeta_4 = 3^{-1/2}(\sqrt{2}\eta_0 - \bar{\eta}), \quad \zeta_5 = \eta_{-1}, \quad \zeta_6 = \eta_{-2}. \end{aligned} \tag{2.13}$$

The corresponding annihilation operators are denoted by $\zeta^+_\mu, \mu = 1, \cdots, 6$. Let us note that these definitions essentially depend upon the enumeration conventions (2.9) and (2.11), and should be accordingly transformed when different conventions are chosen.

Using the creation and annihilation operators ζ_μ and ζ^+_μ, it is now straightforward to write the generators of the groups appearing in the chain (1.5). The generators of U(6) are given by

$$\mathscr{C}_{\mu}{}^{\mu'} = \xi_{\mu}\xi_{\mu'}^{\dagger}, \quad \mu,\mu' = 1,\cdots,6, \tag{2.14}$$

and are of course different from those defined in Eq. (2.3). To obtain the generators of U(3), it is again useful to consider the corresponding harmonic oscillator problem. In this case, the U(3) subgroup of U(6) is a realization of the symmetry group of the harmonic oscillator, acting in the two-quantum state space. Its generators $\mathbb{C}_{q}{}^{q'}$, $q, q' = \pm 1, 0$, are therefore defined in terms of the generators of U(6) by[11]

$$\mathbb{C}_{q}{}^{q'} = \sum_{\mu,\mu'=1}^{6} \langle \mu | c_{q}{}^{q'} | \mu' \rangle \mathscr{C}_{\mu}{}^{\mu'}. \tag{2.15}$$

Using the enumeration convention (2.9), they can be rewritten as $\mathbb{C}_{i}^{j}$, $i,j = 1, 2, 3$. Explicit expressions for these operators are contained in Eq. (7.17) of Ref. 11. We reproduce them here:

$$\begin{aligned}
\mathbb{C}_{1}{}^{2} &= \sqrt{2}\mathscr{C}_{1}{}^{2} + \sqrt{2}\mathscr{C}_{2}{}^{4} + \mathscr{C}_{3}{}^{5},\\
\mathbb{C}_{1}{}^{3} &= \sqrt{2}\mathscr{C}_{1}{}^{3} + \mathscr{C}_{2}{}^{5} + \sqrt{2}\mathscr{C}_{3}{}^{6},\\
\mathbb{C}_{2}{}^{3} &= \mathscr{C}_{2}{}^{3} + \sqrt{2}\mathscr{C}_{4}{}^{5} + \sqrt{2}\mathscr{C}_{5}{}^{6},\\
\mathbb{C}_{1}{}^{1} &= 2\mathscr{C}_{1}{}^{1} + \mathscr{C}_{2}{}^{2} + \mathscr{C}_{3}{}^{3},\\
\mathbb{C}_{2}{}^{2} &= \mathscr{C}_{2}{}^{2} + 2\mathscr{C}_{4}{}^{4} + \mathscr{C}_{5}{}^{5},\\
\mathbb{C}_{3}{}^{3} &= \mathscr{C}_{3}{}^{3} + \mathscr{C}_{5}{}^{5} + 2\mathscr{C}_{6}{}^{6},\\
\mathbb{C}_{2}{}^{1} &= \sqrt{2}\mathscr{C}_{2}{}^{1} + \sqrt{2}\mathscr{C}_{4}{}^{2} + \mathscr{C}_{5}{}^{3},\\
\mathbb{C}_{3}{}^{1} &= \sqrt{2}\mathscr{C}_{3}{}^{1} + \mathscr{C}_{5}{}^{2} + \sqrt{2}\mathscr{C}_{6}{}^{3},\\
\mathbb{C}_{3}{}^{2} &= \mathscr{C}_{3}{}^{2} + \sqrt{2}\mathscr{C}_{5}{}^{4} + \sqrt{2}\mathscr{C}_{6}{}^{5}.
\end{aligned} \tag{2.16}$$

Equation (2.16) may be applied to the generators of the U(3) group in the case of the interacting boson model as well. The generators of the U(2) and U(1) subgroups are then obtained from those of U(3) by restricting i,j to the values 1, 2, or 1, respectively.

The generators of all the groups appearing in the chain (1.5) have thus been defined, so that we are now in a position to construct N boson states characterized by IR's of the chain of groups (1.5).

3. BASIS STATES OF THE IR [N] OF U(6) IN THE CHAIN U(6)⊃U(3)⊃U(2)⊃U(1)

Since Elliott's pioneering work on the nuclear $2s$-$1d$ shell,[10] it has been well known that the IR's of U(3) contained in the IR[N] of U(6) can be characterized by a partition $[h_1h_2h_3]$, $h_1 \geqslant h_2 \geqslant h_3 \geqslant 0$, where the values of h_1, h_2, and h_3 are restricted by the conditions

$$h_1, h_2, h_3 \quad \text{even}, \tag{3.1a}$$

and

$$h_1 + h_2 + h_3 = 2N. \tag{3.1b}$$

The last one is clearly connected with the fact that the first-order Casimir operator of U(3) is equal to twice that of U(6):

$$\sum_{i=1}^{3} \mathbb{C}_{i}{}^{i} = 2\sum_{\mu=1}^{6} \mathscr{C}_{\mu}{}^{\mu}. \tag{3.2}$$

The IR's of U(3) can be also characterized by three nonnegative integers x, y, and z, satisfying the condition

$$x + 2y + 3z = N. \tag{3.3}$$

The relation between h_1, h_2, h_3 and x, y, z is given by

$$h_1 = 2x + 2y + 2z, \quad h_2 = 2y + 2z, \quad h_3 = 2z. \tag{3.4}$$

N boson states characterized by IR's of U(3), U(2), and U(1) are just the usual Gel'fand states of the U(3) group

$$\left| \begin{matrix} h_{13} & & h_{23} & & h_{33} \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle, \quad h_{ij} \geqslant h_{ij-1} \geqslant h_{i+1,j}, \tag{3.5}$$

with the only restriction that $h_{i3} = h_i$ $(i = 1, 2, 3)$ satisfy the conditions (3.1). To construct the states (3.5), it is sufficient to build the highest weight state (hws) of the IR $[h_1h_2h_3]$,

$$\left| \begin{matrix} h_1 & & h_2 & & h_3 \\ & \text{hws} & & & \end{matrix} \right\rangle = \left| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_1 & & h_2 & \\ & & h_1 & & \end{matrix} \right\rangle, \tag{3.6}$$

because the other states can be obtained from it with the help of lowering operators.[11]

The hws (3.6) is associated with the N th degree polynomial in the ξ_{μ}'s, $P_N(\xi_{\mu})$, which is a simultaneous solution of the equations

$$\begin{aligned}
&\mathbb{C}_{i}{}^{i} P_N(\xi_{\mu}) = h_i P_N(\xi_{\mu}), \quad i = 1,2,3,\\
&\mathbb{C}_{i}{}^{j} P_N(\xi_{\mu}) = 0, \quad 1 \leqslant i < j \leqslant 3,
\end{aligned} \tag{3.7}$$

where the $\mathbb{C}_{i}^{j}$ are interpreted as first-order differential operators. Some years ago, it was shown by Pérez[12] that the solution of Eq. (3.7) leads to the following expression for the hws (3.6):

$$\left| \begin{matrix} h_1 & & h_2 & & h_3 \\ & \text{hws} & & & \end{matrix} \right\rangle = \left| \begin{matrix} 2x+2y+2z & & 2y+2z & & 2z \\ & & & \text{hws} & \end{matrix} \right\rangle = \mathscr{N}(x,y,z)\xi_1{}^{x} Y^{y} Z^{z} |0\rangle, \tag{3.8}$$

in terms of the polynomials ξ_1, Y, and Z associated with the hws of the IR's [200], [220], and [222] of U(3), respectively. Explicit expressions for Y and Z are given by

$$Y = \begin{vmatrix} \xi_1 & \frac{1}{\sqrt{2}}\xi_2 \\ \frac{1}{\sqrt{2}}\xi_2 & \xi_4 \end{vmatrix} \tag{3.9}$$

and

$$Z = \begin{vmatrix} \xi_1 & \frac{1}{\sqrt{2}}\xi_2 & \frac{1}{\sqrt{2}}\xi_3 \\ \frac{1}{\sqrt{2}}\xi_2 & \xi_4 & \frac{1}{\sqrt{2}}\xi_5 \\ \frac{1}{\sqrt{2}}\xi_3 & \frac{1}{\sqrt{2}}\xi_5 & \xi_6 \end{vmatrix}. \tag{3.10}$$

It only remains to determine the normalization coefficient $\mathscr{N}(x,y,z)$ in Eq. (3.8). This can be achieved by using a

method similar to that explained in Ref. 7. (Normalized hws have also been derived independently by Vanagas *et al.*[13]) We only quote hereafter the final result, leaving the complete derivation for Appendix A:

$$N(x,y,z) = 2^{y+z}[x!y!]^{-1} \times [(2x+1)!(2y+1)!(x+y)!(x+y+1)!(y+z)!]^{1/2} \times [z!(2x+2y+1)!(2y+2z+1)! \times (x+y+z+1)!]^{-1/2}. \tag{3.11}$$

The other states of the IR $[h_1h_2h_3]$ can now be written in terms of the hws as follows[11]:

$$\left|\begin{matrix} h_1 & h_2 & h_3 \\ h_{12} & h_{22} & \\ h_{11} & & \end{matrix}\right\rangle = N^{h_{12}h_{22}}_{h_{11}} N^{h_1h_2h_3}_{h_{12}h_{22}} (L_2^{\,1})^{h_{12}-h_{11}} (L_3^{\,1})^{h_1-h_{12}} \times (L_3^{\,2})^{h_2-h_{22}} \left|\begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix}\right\rangle \tag{3.12}$$

In Eq. (3.12), the lowering operators are given by

$$L_2^{\,1} = \mathbb{C}_2^{\,1}, \quad L_3^{\,1} = \mathbb{C}_3^{\,1}(\mathbb{C}_1^{\,1} - \mathbb{C}_2^{\,2} + 1) + \mathbb{C}_2^{\,1}\mathbb{C}_3^{\,2}, \quad L_3^{\,2} = \mathbb{C}_3^{\,2}, \tag{3.13}$$

and the normalization coefficients by

$$N^{h_{12}h_{22}}_{h_{11}} = [(h_{12}-h_{11})!(h_{12}-h_{22})!]^{-1/2}[(h_{11}-h_{22})!]^{1/2},$$

$$N^{h_1h_2h_3}_{h_{12}h_{22}} = [(h_1-h_{12})!(h_1-h_{22}+1)!(h_2-h_{22})!(h_1-h_2)! \times (h_1-h_3+1)!(h_2-h_3)!]^{-1/2}[(h_{12}-h_{22}+1)! \times (h_{12}-h_2)!(h_{12}-h_3+1)!(h_{22}-h_3)!]^{1/2}. \tag{3.14}$$

Let us note that the relative phase of the states (3.12) has been chosen in such a way that the matrix elements of $\mathbb{C}_3^{\,2}$ are nonnegative.

In the following two sections, we will proceed to the calculation of the matrix elements of some operators with respect to the N boson states that have been introduced in this section.

4. MATRIX ELEMENTS OF THE BOSON CREATION AND ANNIHILATION OPERATORS IN A SYMMETRICAL U(6)⊃U(3)⊃U(2)⊃U(1) BASIS

Let us first consider the matrix elements of the creation operators $\zeta_\mu, \mu = 1, \cdots, 6$. These operators transform according to the IR [200] of U(3). We shall denote the components of the corresponding irreducible tensor (IT) T [200] by

$$T\begin{pmatrix} 2 & 0 & 0 \\ r & 0 & \\ t & & \end{pmatrix}, \quad 0 \leqslant r \leqslant 2, \quad 0 \leqslant t \leqslant r. \tag{4.1}$$

The component with $r = t = 2$, corresponding to the hws of the IR, is proportional to ζ_1. We choose to normalize the IT in such a way that

$$T\begin{pmatrix} 2 & 0 & 0 \\ 2 & 0 & \\ 2 & & \end{pmatrix} = \zeta_1. \tag{4.2}$$

The remaining components of the IT are then deduced from Eq. (4.2) by applying an equation similar to Eq. (3.12). The various components of the IT so obtained are listed in Table I.

The matrix elements of the IT T [200] between N boson states of the type (3.12),

TABLE I. Boson creation and annihilation operators as components of the IT's T [200] and T [00 − 2].

r	t	$T\begin{pmatrix} 2 & 0 & 0 \\ r & 0 & \\ t & & \end{pmatrix}$	$T\begin{pmatrix} 0 & 0 & -2 \\ 0 & -r & \\ -t & & \end{pmatrix}$
2	2	ζ_1	ζ_1^{+}
2	1	ζ_2	$-\zeta_2^{+}$
2	0	ζ_4	ζ_4^{+}
1	1	ζ_3	ζ_3^{+}
1	0	ζ_5	$-\zeta_5^{+}$
0	0	ζ_6	ζ_6^{+}

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| T \begin{pmatrix} 2 & 0 & 0 \\ & r & 0 \\ & & t \end{pmatrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle, \tag{4.3}$$

are different from zero only when

$$\begin{aligned} &h'_1 = h_1 + 2, \quad h'_2 = h_2, \quad h'_3 = h_3, \\ &\text{or } h'_1 = h_1, \quad h'_2 = h_2 + 2, \quad h'_3 = h_3, \\ &\text{or } h'_1 = h_1, \quad h'_2 = h_2, \quad h'_3 = h_3 + 2, \end{aligned} \tag{4.4a}$$

and

$$\begin{aligned} &h'_{12} + h'_{22} = h_{12} + h_{22} + r, \\ &h'_{11} = h_{11} + t. \end{aligned} \tag{4.4b}$$

The nonzero matrix elements can be calculated using the Wigner–Eckart theorem with respect to the U(3) group, which is written as

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| T \begin{pmatrix} 2 & 0 & 0 \\ & r & 0 \\ & & t \end{pmatrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle = \langle h'_1 h'_2 h'_3 \| T[200] \| h_1 h_2 h_3 \rangle$$

$$\times \left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| \begin{matrix} & & h'_1 - h_1 & & \\ & h'_1 + h'_2 - h_1 - h_2 & & 0 & \\ 2 & & 0 & & 0 \\ & r & & 0 & \\ & & t & & \end{matrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle. \tag{4.5}$$

In the right-hand side of Eq. (4.5), the first factor is a U(3) reduced matrix element, and the second one is an SU(3) Wigner coefficient. For the latter, we use Biedenharn's notations,[14] where

$$\begin{pmatrix} h'_1 - h_1 & \\ h'_1 + h'_2 - h_1 - h_2 & 0 \end{pmatrix}$$

is an operator pattern, which is entirely determined by the initial and final IR's of U(3), as the Kronecker product $[h_1 h_2 h_3] \times [200]$ is multiplicity free. The SU(3) Wigner coefficient can be factorized into an SU(3) reduced Wigner coefficient and an ordinary SU(2) Wigner coefficient:

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| \begin{matrix} & & h'_1 - h_1 & & \\ & h'_1 + h'_2 - h_1 - h_2 & & 0 & \\ 2 & & 0 & & 0 \\ & r & & 0 & \\ & & t & & \end{matrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle$$

$$= \left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{12} & & \end{matrix} \middle| \begin{bmatrix} & & h'_1 - h_1 & & \\ & h'_1 + h'_2 - h_1 - h_2 & & 0 & \\ 2 & & 0 & & 0 \\ & r & & 0 & \\ & & h'_{12} - h_{12} & & \end{bmatrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{12} & & \end{matrix} \right\rangle$$

$$\times \langle \tfrac{1}{2}(h_{12} - h_{22})\, h_{11} - \tfrac{1}{2}(h_{12} + h_{22})\, \frac{r}{2}\, t - \frac{r}{2} | \tfrac{1}{2}(h'_{12} - h'_{22})\, h'_{11} - \tfrac{1}{2}(h'_{12} + h'_{22}) \rangle. \tag{4.6}$$

The SU(3) reduced Wigner coefficients, contained in Eq. (4.6), were tabulated by Hecht,[15] so that the only quantities left for determination in Eq. (4.5) are the reduced matrix elements of the IT. When using Hecht's tables, one must take into account that they were established with a phase convention for the states of the U(3) IR's which was different from the one adopted in the present work. Hecht indeed chose the matrix elements of $\mathbb{C}_3{}^1$ to be nonnegative, instead of those of $\mathbb{C}_3{}^2$. The states of the IR $[h_1 h_2 h_3]$ therefore contained an additional phase $(-1)^{h_2 - h_{22}}$ with respect to our choice, defined in Eq. (3.12). As a conse-

quence, the values of the SU(3) reduced Wigner coefficients, tabulated by Hecht, have to be multiplied by $(-1)^{h'_2 - h'_{22} + h_2 - h_{22}}$ before they can be used in Eqs. (4.5) and (4.6).

The calculation of the reduced matrix elements of the IT $T[200]$ can be carried out by applying Eq. (4.5) to hws in both bra and ket. From Eq. (4.4b), the values of r and t are then found to be equal to

$$r = h'_1 + h'_2 - h_1 - h_2 \text{ and } t = h'_1 - h_1. \tag{4.7}$$

The matrix element of the left-hand side of Eq. (4.5),

$$\left\langle \begin{matrix} h'_1 & h'_2 & h'_3 \\ & \text{hws} & \end{matrix} \middle| T \begin{pmatrix} 2 & 0 & 0 \\ h'_1 + h'_2 - h_1 - h_2 & 0 & \\ & h'_1 - h_1 & \end{pmatrix} \middle| \begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix} \right\rangle, \tag{4.8}$$

can be computed from the explicit expression of the hws, contained in Eq. (3.8). According as $[h'_1 h'_2 h'_3] = [h_1 + 2h_2 h_3]$, $[h_1 h_2 + 2h_3]$ or $[h_1 h_2 h_3 + 2]$, the component of the IT appearing in Eq. (4.8) is ζ_1, ζ_4, or ζ_6. It can be shown that the matrix elements (4.8) can be expressed in terms of normalization coefficients $\mathcal{N}(x,y,z)$ corresponding to various sets of values of x, y, and z. As an example, the case of the matrix element of ζ_4 is treated in Appendix B. The complete results for the matrix elements (4.8) are given in Table II.

It is now straightforward to obtain the reduced matrix elements of $T[200]$ by dividing the matrix elements (4.8) by the SU(3) Wigner coefficients (4.6) corresponding to hws. The results are contained in Table III.

Once the matrix elements of the boson creation operators have been obtained, those of the annihilation operators are given in terms of them by the following Hermitian conjuguate relation:

$$\left\langle \begin{matrix} h'_1 & h'_2 & h'_3 \\ & h'_{12} & h'_{22} \\ & h'_{11} & \end{matrix} \middle| \zeta_\mu^+ \middle| \begin{matrix} h_1 & h_2 & h_3 \\ & h_{12} & h_{22} \\ & h_{11} & \end{matrix} \right\rangle = \left\langle \begin{matrix} h_1 & h_2 & h_3 \\ & h_{12} & h_{22} \\ & h_{11} & \end{matrix} \middle| \zeta_\mu \middle| \begin{matrix} h'_1 & h'_2 & h'_3 \\ & h'_{12} & h'_{22} \\ & h'_{11} & \end{matrix} \right\rangle. \tag{4.9}$$

Equation (4.9) leads to a relation between the corresponding reduced matrix elements.

To establish this relation, it is first necessary to express the annihilation operators in terms of an IT. It is clear that the annihilation operators transform according to the IR $[00-2]$ of U(3) and that ζ_6^+ is of highest weight. Let us choose the normalization of the IT $T[00-2]$ in such a way that

$$T \begin{pmatrix} 0 & 0 & -2 \\ 0 & 0 & \\ & 0 & \end{pmatrix} = \zeta_6^+. \tag{4.10}$$

Then it is straightforward to see that the remaining components of $T[00-2]$ are as given in Table I. They are related to the components of $T[200]$ by the following equation:

TABLE II. Matrix elements of $T \begin{pmatrix} 2 & 0 & 0 \\ r & 0 & \\ & t & \end{pmatrix}$ between hws.

h'_1	h'_2	h'_3	$\left\langle \begin{matrix} h'_1 & h'_2 & h'_3 \\ & \text{hws} & \end{matrix} \middle\| T \begin{pmatrix} 2 & 0 & 0 \\ r & 0 & \\ & t & \end{pmatrix} \middle\| \begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix} \right\rangle$
$h_1 + 2$	h_2	h_3	$\left[\frac{(x+1)(2x+2y+3)(x+y+z+2)}{(2x+3)(x+y+2)}\right]^{1/2}$
h_1	$h_2 + 2$	h_3	$\left[\frac{2x(y+1)(2y+2z+3)}{(2x+1)(2y+3)}\right]^{1/2}$
h_1	h_2	$h_3 + 2$	$\left[\frac{y(2x+2y+1)(z+1)}{(2y+1)(x+y+1)}\right]^{1/2}$

TABLE III. Reduced matrix elements of $T[200]$.

h'_1	h'_2	h'_3	$\langle h'_1 h'_2 h'_3 \| T[200] \| h_1 h_2 h_3 \rangle$
h_1+2	h_2	h_3	$\left[\frac{(x+1)(2x+2y+3)(x+y+z+2)}{(2x+3)(x+y+2)}\right]^{1/2}$
h_1	h_2+2	h_3	$\left[\frac{2x(y+1)(2y+2z+3)}{(2x-1)(2y+3)}\right]^{1/2}$
h_1	h_2	h_3+2	$\left[\frac{y(2x+2y+1)(z+1)}{(2y-1)(x+y)}\right]^{1/2}$

TABLE IV. Reduced matrix elements of $T[00-2]$.

h'_1	h'_2	h'_3	$\langle h'_1 h'_2 h'_3 \| T[00-2] \| h_1 h_2 h_3 \rangle$
h_1-2	h_2	h_3	$\left[\frac{x(2x+2y+1)(x+y+z+1)}{(2x-1)(x+y)}\right]^{1/2}$
h_1	h_2-2	h_3	$\left[\frac{2(x+1)y(2y+2z+1)}{(2x+3)(2y-1)}\right]^{1/2}$
h_1	h_2	h_3-2	$\left[\frac{(y+1)(2x+2y+3)z}{(2y+3)(x+y+2)}\right]^{1/2}$

$$T\begin{pmatrix} 0 & 0 & -2 \\ 0 & & -r \\ & -t & \end{pmatrix} = (-1)^{r+t}\left[T\begin{pmatrix} 2 & 0 & 0 \\ r & & 0 \\ & t & \end{pmatrix}\right]^{+}. \tag{4.11}$$

Finally, when combining Eq. (4.11) with the following symmetry relation of the SU(3) reduced Wigner coefficients,[15]

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_1 & & h'_2 & \\ & & h'_1 & & \end{matrix} \middle\| \begin{bmatrix} & h'_1-h_1 & \\ h'_1+h'_2-h_1-h_2 & & 0 \\ 2 & 0 & 0 \\ & r & \\ & h'_1-h_1 & \end{bmatrix} \middle\| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_1 & & h_2 & \\ & & h_1 & & \end{matrix} \right\rangle$$

$$= (-1)^{(1/3)(h_1-2h_2+h_3-h'_1+2h'_2-h'_3-2)-h_2+h'_2}\left[\frac{\dim([h_1h_2h_3])(h'_1-h'_2+1)}{\dim([h'_1h'_2h'_3])(h_1-h_2+1)}\right]^{1/2}$$

$$\times\left\langle \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_1 & & h_2 & \\ & & h_1 & & \end{matrix} \middle\| \begin{bmatrix} & h_1-h'_1 & \\ 0 & h_1+h_2-h'_1-h'_2 & \\ 0 & 0 & -2 \\ & 0 & -r \\ & h_1-h'_1 & \end{bmatrix} \middle\| \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_1 & & h'_2 & \\ & & h'_1 & & \end{matrix} \right\rangle, \tag{4.12}$$

where

$$\dim([h_1h_2h_3]) = (2x+1)(2y+1)(x+y+1) \tag{4.13}$$

is the dimension of the IR $[h_1h_2h_3]$, Eq. (4.5), written for hws, leads to the relation between reduced matrix elements of $T[200]$ and $T[00-2]$, that we were looking for:

$$\langle h'_1 h'_2 h'_3 \| T[00-2] \| h_1 h_2 h_3 \rangle = \left[\frac{\dim([h_1h_2h_3])}{\dim([h'_1h'_2h'_3])}\right]^{1/2} \langle h_1 h_2 h_3 \| T[200] \| h'_1 h'_2 h'_3 \rangle. \tag{4.14}$$

In deriving Eq. (4.14), we have used the fact that the phase

$$(-1)^{(1/3)(h_1-2h_2+h_3-h'_1+2h'_2-h'_3-2)}$$

is equal to $+1$ for IR's $[h'_1h'_2h'_3]$ satisfying Eq. (4.4a). The reduced matrix elements of $T[00-2]$, calculated by means of Eq. (4.14), are listed in Table IV.

5. MATRIX ELEMENTS OF THE U(6) GENERATORS IN A SYMMETRICAL U(6)⊃U(3)⊃U(2)⊃U(1) BASIS

The method that is going to be used to calculate the matrix elements of the U(6) generators is essentially the same as that devised in the previous section for the matrix elements of the boson creation and annihilation operators. There are however two additional difficulties with respect to the previous case: first the U(6) generators do not transform irreducibly under U(3), so that one has to separate them into IT's, and secondly the Kronecker products of IR's of U(3), that appear when considering matrix elements of those IT's, are not multiplicity free.

Let us begin by organizing the U(6) generators into IT's with respect to U(3)⊃U(2)⊃U(1). The U(6) generators are the products of a boson creation operator transforming according to the IR [200], and an annihilation operator transforming according to [00 − 2]. Therefore they can be separated into three IT's with classification [000], [10 − 1], and [20 − 2], respectively. The [000] IT is just the first-order Casimir operator, $\Sigma_\mu \mathscr{C}_\mu{}^\mu$, of U(6) [or the first-order Casimir operator, $\Sigma_i C_i{}^i$, of U(3), owing to Eq. (3.2)]. The [10 − 1] IT is made of the eight generators of SU(3), obtained from the generators (2.16) of U(3)

by dropping the trace $\Sigma_i C_i{}^i$. Finally, the [20 − 2] IT is made of the 27 remaining independent linear combinations of the generators. Its components, denoted by

$$T\begin{pmatrix} 2 & 0 & -2 \\ & r & s \\ & & t \end{pmatrix}, \quad -2\leqslant s\leqslant 0\leqslant r\leqslant 2, \quad s\leqslant t\leqslant r, \tag{5.1}$$

are listed in Table V. They have been chosen in accordance with Eq. (3.12) and with the normalization convention

$$T\begin{pmatrix} 2 & 0 & -2 \\ & 2 & 0 \\ & & 2 \end{pmatrix} = \mathscr{C}_1{}^6. \tag{5.2}$$

As the matrix elements of $T[000]$ and $T[10-1]$ are well known,[11] we only need to calculate those of $T[20-2]$, i.e.,

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| T\begin{pmatrix} 2 & 0 & -2 \\ & r & s \\ & & t \end{pmatrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle. \tag{5.3}$$

The latter differ from zero only when either of the following conditions is satisfied,

$$h'_1 = h_1 + 2, \quad h'_2 = h_2, \quad h'_3 = h_3 - 2, \tag{5.4a}$$
$$h'_1 = h_1 + 2, \quad h'_2 = h_2 - 2, \quad h'_3 = h_3, \tag{5.4b}$$
$$h'_1 = h_1, \quad h'_2 = h_2 + 2, \quad h'_3 = h_3 - 2, \tag{5.4c}$$
$$h'_1 = h_1, \quad h'_2 = h_2, \quad h'_3 = h_3, \tag{5.4d}$$
$$h'_1 = h_1, \quad h'_2 = h_2 - 2, \quad h'_3 = h_3 + 2, \tag{5.4e}$$
$$h'_1 = h_1 - 2, \quad h'_2 = h_2 + 2, \quad h'_3 = h_3, \tag{5.4f}$$
$$h'_1 = h_1 - 2, \quad h'_2 = h_2, \quad h'_3 = h_3 + 2, \tag{5.4g}$$

and when moreover

$$h'_{12} + h'_{22} = h_{12} + h_{22} + r + s,$$

and (5.5)

$$h'_{11} = h_{11} + t.$$

In the calculation of the nonzero matrix elements, the multiplicity problem only appears for the diagonal case, corresponding to Eq. (5.4d): the multiplicity of the IR $[h'_1 h'_2 h'_3]$ in the Kronecker product $[h_1 h_2 h_3]\times[20-2]$ is indeed equal to three for $[h'_1 h'_2 h'_3] = [h_1 h_2 h_3]$, and one in the remaining cases.

The Wigner–Eckart theorem for the matrix elements (5.3) is written

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| T\begin{pmatrix} 2 & 0 & -2 \\ & r & s \\ & & t \end{pmatrix} \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle$$

$$= \sum_{(\gamma)} \langle h'_1 h'_2 h'_3 \| T[20-2;(\gamma)] \| h_1 h_2 h_3 \rangle$$

$$\times \left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle\| \begin{matrix} & & (\gamma) & & \\ 2 & & 0 & & -2 \\ & r & & s & \\ & & t & & \end{matrix} \middle\| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle, \tag{5.6}$$

where

$$(\gamma) = \begin{pmatrix} & \gamma_{11} & \\ \gamma_{12} & & \gamma_{22} \end{pmatrix}$$

is an operator pattern.[14] As can be seen in Table VII, it is entirely determined by the initial and final IR's in the nondiagonal cases, whereas in the diagonal case it takes the three following values:

$$(\gamma)=\begin{pmatrix}0&\\0&0\end{pmatrix},\ \begin{pmatrix}0&\\1&-1\end{pmatrix},\ \text{and}\ \begin{pmatrix}0&\\2&-2\end{pmatrix}.$$

The factorization of the SU(3) Wigner coefficient now becomes:

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{11} & & \end{matrix} \middle| \left\langle \begin{matrix} & (\gamma) & \\ 2 & 0 & -2 \\ r & & s \\ & t & \end{matrix} \right\rangle \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix} \right\rangle$$

$$=\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{12} & & \end{matrix} \middle| \left[\begin{matrix} & (\gamma) & \\ 2 & 0 & -2 \\ r & & s \\ & h'_{12}-h_{12} & \end{matrix} \right] \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{12} & & \end{matrix} \right\rangle$$

$$\times\langle \tfrac{1}{2}(h_{12}-h_{22})\,h_{11}-\tfrac{1}{2}(h_{12}+h_{22})\,\tfrac{1}{2}(r-s)\,t-\tfrac{1}{2}(r+s)|\tfrac{1}{2}(h'_{12}-h'_{22})\,h'_{11}-\tfrac{1}{2}(h'_{12}+h'_{22})\rangle. \tag{5.7}$$

As the IR [20 − 2] of U(3) is equivalent to [42] when considered as an IR of SU(3), the SU(3) reduced Wigner coefficients needed in Eq. (5.7) are equal to those where one of the IR's is [42]:

$$\left\langle \begin{matrix} h'_1 & & h'_2 & & h'_3 \\ & h'_{12} & & h'_{22} & \\ & & h'_{12} & & \end{matrix} \middle| \left[\begin{matrix} & \gamma_{11} & \\ \gamma_{12} & & \gamma_{22} \\ 2 & 0 & -2 \\ r & & s \\ h'_{12} & -h_{12} & \end{matrix} \right] \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{12} & & \end{matrix} \right\rangle$$

$$=\left\langle \begin{matrix} h'_1+2 & & h'_2+2 & & h'_3+2 \\ & h'_{12}+2 & & h'_{22}+2 & \\ & & h'_{12}+2 & & \end{matrix} \middle| \left[\begin{matrix} & \gamma_{11}+2 & \\ \gamma_{12}+2 & & \gamma_{22}+2 \\ 4 & 2 & 0 \\ r+2 & & s+2 \\ & h'_{12}-h_{12}+2 & \end{matrix} \right] \middle| \begin{matrix} h_1 & & h_2 & & h_3 \\ & h_{12} & & h_{22} & \\ & & h_{12} & & \end{matrix} \right\rangle. \tag{5.8}$$

The latter were tabulated by Castilho Alcaras, Biedenharn, Hecht, and Neely.[16] Their values may be introduced directly into Eq. (5.7) as the phase convention chosen by these authors for the states of the U(3) IR's coincides with that used in the present paper.

Let us turn now to the calculation of the reduced matrix elements of $T[20-2]$. We shall treat first the nondiagonal cases, leaving for later the diagonal one, where there is a multiplicity problem. We have to actually compute one half of the nondiagonal reduced matrix elements as it is clear that the other half can be deduced from it by using the hermiticity property of the IT:

$$\left[T\begin{pmatrix}2&0&-2\\&r&s\\&&t\end{pmatrix}\right]^{+}=(-1)^{r+s+t}T\begin{pmatrix}2&0&-2\\&-s&-r\\&&-t\end{pmatrix}. \tag{5.9}$$

So let us consider the cases (5.4a,b,c). When taking hws in both bra and ket in Eq. (5.6), it turns out that the values of r, s, and t, compatible with Eqs. (5.1) and (5.5), are entirely fixed by the initial and final U(3) IR's. The evaluation of the matrix element of the left-hand side of Eq. (5.6) then leads to the determination of the unique reduced matrix element of the right-hand side. That evaluation can be carried out straightforwardly by using methods similar to that explained in Appendix B.

The results are contained in Tables VI and VII.

The reduced matrix elements corresponding to the cases (5.4e,f,g) can now be calculated from the previous ones by using the following symmetry relation for the reduced matrix elements:

$$\langle h'_1 h'_2 h'_3 \| T[20-2;(\gamma)] \| h_1 h_2 h_3 \rangle = \left[\frac{\dim([h_1 h_2 h_3])}{\dim([h'_1 h'_2 h'_3])} \right]^{1/2} \langle h_1 h_2 h_3 \| T[20-2;(\tilde{\gamma})] \| h'_1 h'_2 h'_3 \rangle, \tag{5.10}$$

where

$$(\tilde{\gamma}) = \begin{pmatrix} & -\gamma_{11} & \\ -\gamma_{22} & & -\gamma_{12} \end{pmatrix}.$$

This relation can be proved by combining Eq. (5.9) with the following symmetry property of the SU(3) reduced Wigner coefficients[16]:

$$\left\langle \begin{matrix} h_1 & h_2 & h_3 \\ & h_1 & h_2 \\ & & h_1 \end{matrix} \middle| \begin{matrix} & -\gamma_{11} & \\ -\gamma_{22} & & -\gamma_{12} \\ 2 & 0 & -2 \\ -s & & -r \\ & h_1 - h'_1 & \end{matrix} \middle| \begin{matrix} h'_1 & h'_2 & h'_3 \\ & h'_1 & h'_2 \\ & & h'_1 \end{matrix} \right\rangle$$

$$= (-1)^{r+s+h_1-h'_1} \left[\frac{\dim([h_1 h_2 h_3])(h'_1 - h'_2 + 1)}{\dim([h'_1 h'_2 h'_3])(h_1 - h_2 + 1)} \right]^{1/2}$$

$$\times \left\langle \begin{matrix} h'_1 & h'_2 & h'_3 \\ & h'_1 & h'_2 \\ & & h'_1 \end{matrix} \middle| \begin{matrix} & \gamma_{11} & \\ \gamma_{12} & & \gamma_{22} \\ 2 & 0 & -2 \\ r & & s \\ & h'_1 - h_1 & \end{matrix} \middle| \begin{matrix} h_1 & h_2 & h_3 \\ & h_1 & h_2 \\ & & h_1 \end{matrix} \right\rangle. \tag{5.11}$$

It remains to consider the diagonal case for which there are three independent reduced matrix elements for

$$(\gamma) = \begin{pmatrix} & 0 & \\ 0 & & 0 \end{pmatrix}, \begin{pmatrix} & 0 & \\ 1 & & -1 \end{pmatrix}, \text{ and } \begin{pmatrix} & 0 & \\ 2 & & -2 \end{pmatrix}$$

to be determined. When taking hws in both bra and ket in Eq. (5.6), it turns out that for $[h'_1 h'_2 h'_3] = [h_1 h_2 h_3]$ three sets of values of r, s, and t are compatible with Eqs. (5.1) and (5.5), namely

$$\begin{pmatrix} r & & s \\ & t & \end{pmatrix} = \begin{pmatrix} 0 & & 0 \\ & 0 & \end{pmatrix}, \begin{pmatrix} 1 & & -1 \\ & 0 & \end{pmatrix}, \text{ and } \begin{pmatrix} 2 & & -2 \\ & 0 & \end{pmatrix}.$$

Therefore if one considers Eq. (5.6) for each of these sets, one gets a system of three linear equations in the three independent reduced matrix elements. Its solution leads to the results of Table VII, which have been expressed in terms of the boson number N, given by Eq. (3.3), and the eigenvalues of the second-and third-order Casimir operators of SU(3), respectively, equal to

$$g = 2(2x^2 + 2xy + 2y^2 + 3x + 3y), \tag{5.12}$$

and

$$\Gamma = 2(x-y)(4x + 2y + 3)(2x + 4y + 3). \tag{5.13}$$

6. TRANSFORMATION BRACKETS BETWEEN STATES IN THE CHAINS U(6)⊃SU(3)⊃SO(3)⊃SO(2) AND U(6)⊃U(3)⊃U(2)⊃U(1)

The transformation brackets from the canonical chain U(3)⊃U(2)⊃U(1) to the physical chain U(3)⊃SO(3)⊃SO(2) were

determined by Moshinsky some years ago.[7,8] Hereafter we shall quote his results and indicate how they can be adapted to the present problem.

Denoting by

$$\left|\begin{matrix} h_1 h_2 h_3 \\ qLM \end{matrix}\right\rangle$$

the Bargmann–Moshinsky basis states of the IR $[h_1 h_2 h_3]$ of U(3) corresponding to the chain $U(3)\supset SO(3)\supset SO(2)$,[17] the transformation brackets we have to consider are

$$\left\langle \begin{matrix} h_1 & h_2 & h_3 \\ & h_{12} \quad h_{22} & \\ & h_{11} & \end{matrix} \middle| \begin{matrix} h_1 & h_2 & h_3 \\ & qLM & \end{matrix} \right\rangle = \left\langle \begin{matrix} \kappa_1 & \kappa_2 & 0 \\ & \kappa_{12} \quad \kappa_{22} & \\ & \kappa_{11} & \end{matrix} \middle| \begin{matrix} \kappa_1 & \kappa_2 & 0 \\ & qLM & \end{matrix} \right\rangle, \tag{6.1}$$

where

$$\kappa_i = h_i - h_3, \quad \kappa_{i2} = h_{i2} - h_3, \quad i = 1,2,$$
$$\kappa_{11} = h_{11} - h_3. \tag{6.2}$$

In Eq. (6.1), we have used the fact that the U(3) transformation brackets for the IR $[h_1\, h_2\, h_3]$ reduce to SU(3) transformation brackets for the IR $[\kappa_1\, \kappa_2] = [h_1 - h_3, h_2 - h_3]$.

For $M = L$, the transformation bracket (6.1) is given by[7,8]

$$\left\langle \begin{matrix} \kappa_1 & \kappa_2 & 0 \\ & \kappa_{12} \quad \kappa_{22} & \\ & \kappa_{11} & \end{matrix} \middle| \begin{matrix} \kappa_1 & \kappa_2 & 0 \\ & qLM & \end{matrix} \right\rangle$$
$$= \left\{ \frac{[\tfrac{1}{2}(\kappa_{12} - \kappa_{22} + L)]!\kappa_{22}!(\kappa_{12} - \kappa_{22} + 1)!(\kappa_{12} + 1)!(\kappa_1 - \kappa_{22} + 1)!(\kappa_{12} - \kappa_2)!(\kappa_1 - \kappa_{12})!}{[\tfrac{1}{2}(\kappa_{12} - \kappa_{22} - L)]!(\kappa_1 - \kappa_2 + 1)!(\kappa_2 - \kappa_{22})!} \right\}^{1/2}$$
$$\times (-2)^{(1/2)(\kappa_{12} - \kappa_{22} - L)} \sum_{\rho\beta} E^q_\rho \frac{[\tfrac{1}{2}(\kappa_1 - L - \rho)]!\rho!(\kappa_2 - \beta)!}{2^\beta [\tfrac{1}{2}(\kappa_1 - \kappa_{12} - \kappa_{22} - \rho) + \beta]!(\kappa_{22} - \beta)!(\rho - \beta)!\beta!(\kappa_{12} + 1 - \beta)!}, \tag{6.3}$$

where ρ is even (odd) if $\kappa_1 - L$ is even (odd) and

$$E^q_\rho = \binom{q}{\frac{1}{2}\rho} (-1)^{\frac{1}{2}\rho}, \quad \text{if } \rho \text{ is even},$$
$$= \binom{q}{\frac{1}{2}(\rho - 1)} (-1)^{\frac{1}{2}(\rho - 1)}, \quad \text{if } \rho \text{ is odd}. \tag{6.4}$$

For $M \neq L$, the transformation bracket (6.1) can be obtained from the one with $M = L$ and the well known matrix elements of $\mathbb{C}_i{}^j$ with respect to the canonical chain because

$$\left| \begin{matrix} \kappa_1 & \kappa_2 & 0 \\ & qLM & \end{matrix} \right\rangle = \left[\frac{(L + M)!2^{L-M}}{(L - M)!(2L)!} \right]^{1/2} (L_{-1})^{L-M} \left| \begin{matrix} \kappa_1 & \kappa_2 & 0 \\ & qLL & \end{matrix} \right\rangle, \tag{6.5}$$

and L_{-1}, defined by Eq. (2.4), can be expressed in terms of $\mathbb{C}_i{}^j$ as

$$L_{-1} = \mathbb{C}_2{}^1 + \mathbb{C}_3{}^2. \tag{6.6}$$

Equation (6.3) cannot however be applied directly to the present case because, through its derivation in Ref. 7, the values of the index q were assumed to be enumerated in the order

$$q \to i: \quad +1 \to 1, \quad -1 \to 2, \quad 0 \to 3, \tag{6.7}$$

which does not coincide with our convention, defined in Eq. (2.9). To overcome that difficulty, one could use the same procedure as that devised in Ref. 8. It consists in applying the transposition (2,3) to the bra in Eq. (6.3) and using the explicit result of Chacón and Moshinsky for the matrix elements of (2,3) between basis states corresponding to the canonical chain.[18]

There is however a simpler method, where Eq. (6.3) is retained without any modification and the transformation from the enumeration convention (2.9) to (6.7) is performed in the remaining equations. The latter is easily made once it has been noted that the enumeration convention of the index q values does only determine the relations between the ζ_μ's and the η_{lm}'s, in conjunction with the enumeration conven-

TABLE V. Components of the IT $T\begin{pmatrix} 2 & 0 & -2 \\ r & & s \\ & t & \end{pmatrix}$ [20 − 2].

r	s	t	$T\begin{pmatrix} 2 & 0 & -2 \\ r & & s \\ & t & \end{pmatrix}$
2	0	2	$\mathcal{C}_1^6$
2	0	1	$\mathcal{C}_2^6$
2	0	0	$\mathcal{C}_4^6$
2	−1	2	$-\mathcal{C}_1^5$
2	−1	1	$3^{-1/2}[\mathcal{C}_1^3 - \sqrt{2}\mathcal{C}_2^5]$
2	−1	0	$3^{-1/2}[\sqrt{2}\mathcal{C}_2^3 - \mathcal{C}_4^5]$
2	−1	−1	$\mathcal{C}_4^3$
2	−2	2	$\mathcal{C}_1^4$
2	−2	1	$2^{-1/2}[-\mathcal{C}_1^2 + \mathcal{C}_2^4]$
2	−2	0	$6^{-1/2}[\mathcal{C}_1^1 - 2\mathcal{C}_2^2 + \mathcal{C}_4^4]$
2	−2	−1	$2^{-1/2}[\mathcal{C}_2^1 - \mathcal{C}_4^2]$
2	−2	−2	$\mathcal{C}_4^1$
1	0	1	$15^{-1/2}[-2\mathcal{C}_1^3 - \sqrt{2}\mathcal{C}_2^5 + 3\mathcal{C}_3^6]$
1	0	0	$15^{-1/2}[-\sqrt{2}\mathcal{C}_2^3 - 2\mathcal{C}_4^5 + 3\mathcal{C}_5^6]$
1	−1	1	$10^{-1/2}[\mathcal{C}_1^2 + \mathcal{C}_2^4 - 2\sqrt{2}\mathcal{C}_3^5]$
1	−1	0	$10^{-1/2}[-\mathcal{C}_1^1 + 2\mathcal{C}_3^3 + \mathcal{C}_4^4 - 2\mathcal{C}_5^5]$
1	−1	−1	$10^{-1/2}[-\mathcal{C}_2^1 - \mathcal{C}_4^2 + 2\sqrt{2}\mathcal{C}_5^3]$
1	−2	1	$\mathcal{C}_3^4$
1	−2	0	$3^{-1/2}[-\sqrt{2}\mathcal{C}_3^2 + \mathcal{C}_5^4]$
1	−2	−1	$3^{-1/2}[\mathcal{C}_3^1 - \sqrt{2}\mathcal{C}_5^2]$
1	−2	−2	$\mathcal{C}_5^1$
0	0	0	$30^{-1/2}[\mathcal{C}_1^1 + \mathcal{C}_2^2 - 3\mathcal{C}_3^3 + \mathcal{C}_4^4 - 3\mathcal{C}_5^5 + 3\mathcal{C}_6^6]$
0	−1	0	$15^{-1/2}[\sqrt{2}\mathcal{C}_3^2 + 2\mathcal{C}_5^4 - 3\mathcal{C}_6^5]$
0	−1	−1	$15^{-1/2}[-2\mathcal{C}_3^1 - \sqrt{2}\mathcal{C}_5^2 + 3\mathcal{C}_6^3]$
0	−2	0	$\mathcal{C}_6^4$
0	−2	−1	$-\mathcal{C}_6^2$
0	−2	−2	$\mathcal{C}_6^1$

tion of the $(n_1 n_0 n_{-1})$ values. By this way, the conventions (2.9) and (2.11) lead to the relations (2.13). When choosing now the convention (6.7) for the index q values and combining it with the following modified convention for the $(n_1 n_0 n_{-1})$ values,

$$(n_1 n_0 n_{-1}) \to \mu: \quad (200) \to 1, \quad (101) \to 2,$$
$$(110) \to 3, \quad (002) \to 4, \quad (011) \to 5, \quad (020) \to 6, \tag{6.8}$$

the relations between the ζ_μ's and the η_{lm}'s become

$$\zeta_1 = \eta_2, \quad \zeta_2 = 3^{-1/2}(\eta_0 + \sqrt{2}\bar{\eta}), \quad \zeta_3 = \eta_1,$$
$$\zeta_4 = \eta_{-2}, \quad \zeta_5 = \eta_{-1}, \quad \zeta_6 = 3^{-1/2}(\sqrt{2}\eta_0 - \bar{\eta}), \tag{6.9}$$

while the relations (2.16) between the U(3) and U(6) generators remain unchanged, as well as all the results contained in Tables I to VII. Therefore Eq. (6.3) can be used just as it is, provided that the results are interpreted in accordance with Eq. (6.9). Let us note also that in this case Eq. (6.6) has to be replaced by

$$L_{-1} = \mathbb{C}_3^1 + \mathbb{C}_2^3. \tag{6.10}$$

ACKNOWLEDGMENTS

The author is indebted to the following contributors: M. Moshinsky and E. Chacón for their initiating collaboration and subsequent cooperation; L. C. Biedenharn, J. A. Castilho Alcaras, K. T. Hecht, and A. Partensky for their help in updating the tables of Ref. 16.

This work was initiated during a stay at the Instituto de Física of the Universidad Nacional Autónoma de México. The author would like to thank this institution for its hospitality and the F.N.R.S. for its financial support.

TABLE VI. Matrix elements of $T\begin{pmatrix} 2 & 0 & -2 \\ r & & s \\ & t & \end{pmatrix}$ between hws.

h'_1	h'_2	h'_3	$\begin{pmatrix} r & & s \\ & t & \end{pmatrix}$	$\left\langle \begin{matrix} h'_1 & h'_2 & h'_3 \\ & \text{hws} & \end{matrix} \middle\| T\begin{pmatrix} 2 & 0 & -2 \\ r & & s \\ & t & \end{pmatrix} \middle\| \begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix} \right\rangle$
h_1+2	h_2	h_3-2	$\begin{pmatrix} 2 & & 0 \\ & 2 & \end{pmatrix}$	$\left[\frac{(x+1)(y+1)z(2x+2y+3)(2x+2y+5)(x+y+z+2)}{(2x+3)(2y+3)(x+y+2)(x+y+3)}\right]^{1/2}$
h_1+2	h_2-2	h_3	$\begin{pmatrix} 2 & & -2 \\ & 2 & \end{pmatrix}$	$\left[\frac{2(x+1)(x+2)y(2x+2y+3)(2y+2z+1)(x+y+z+2)}{(2x+3)(2x+5)(2y+1)(x+y+2)}\right]^{1/2}$
h_1	h_2+2	h_3-2	$\begin{pmatrix} 2 & & 0 \\ & 0 & \end{pmatrix}$	$\left[\frac{2x(y+1)(y+2)z(2x+2y+3)(2y+2z+3)}{(2x+1)(2y+3)(2y+5)(x+y+2)}\right]^{1/2}$
h_1	h_2	h_3	$\begin{pmatrix} 0 & & 0 \\ & 0 & \end{pmatrix}$	$30^{-1/2}\left[x+2y+\frac{5z(2xy+2y^2+x+3y)}{(2y+3)(x+y+2)}\right]$
h_1	h_2	h_3	$\begin{pmatrix} 1 & & -1 \\ & 0 & \end{pmatrix}$	$-10^{-1/2}x\left[1+\frac{5z}{(2y+3)(x+y+2)}\right]$
h_1	h_2	h_3	$\begin{pmatrix} 2 & & -2 \\ & 0 & \end{pmatrix}$	$6^{-1/2}\frac{x}{2x+3}\left[2x+4y+3+\frac{z(8xy+8y^2+10x+24y+15)}{(2y+3)(x+y+2)}\right]$

TABLE VII. Reduced matrix elements of $T[20-2]$.

h'_1	h'_2	h'_3	(γ)	$\langle h'_1 h'_2 h'_3 \| T[20-2;(\gamma)] \| h_1 h_2 h_3 \rangle$
h_1+2	h_2	h_3-2	$\begin{pmatrix} & 2 & \\ 2 & & 0 \end{pmatrix}$	$\left[\frac{(x+1)(y+1)z(2x+2y+3)(2x+2y+5)(x+y+z+2)}{(2x+3)(2y+3)(x+y+2)(x+y+3)}\right]^{1/2}$
h_1+2	h_2-2	h_3	$\begin{pmatrix} & 2 & \\ 2 & & -2 \end{pmatrix}$	$\left[\frac{2(x+1)(x+2)y(2x+2y+3)(2y+2z+1)(x+y+z+2)}{(2x+3)(2x+5)(2y-1)(x+y+2)}\right]^{1/2}$
h_1	h_2+2	h_3-2	$\begin{pmatrix} & 0 & \\ 2 & & 0 \end{pmatrix}$	$\left[\frac{2x(y+1)(y+2)z(2x+2y+3)(2y+2z+3)}{(2x-1)(2y+3)(2y+5)(x+y+2)}\right]^{1/2}$
h_1	h_2	h_3	$\begin{pmatrix} & 0 & \\ 0 & & 0 \end{pmatrix}$	$\frac{\Gamma+5g(N+3)}{3[30g(g-4)]^{1/2}}$
h_1	h_2	h_3	$\begin{pmatrix} & 0 & \\ 1 & & -1 \end{pmatrix}$	$\frac{1}{2}\left[\frac{-\Gamma^2+g^2(4g+9)}{30g(g-9)}\right]^{1/2}$
h_1	h_2	h_3	$\begin{pmatrix} & 0 & \\ 2 & & -2 \end{pmatrix}$	$\frac{1}{6}\left\{\frac{-\Gamma^2+g^2(4g+9)}{6(g-4)(g-9)[-\Gamma^2+4g(g-9)^2]}\right\}^{1/2}[-\Gamma+4(g-9)(N+3)]$
h_1	h_2-2	h_3+2	$\begin{pmatrix} & 0 & \\ 0 & & -2 \end{pmatrix}$	$\left[\frac{2(x+1)(y-1)y(z+1)(2x+2y+1)(2y+2z+1)}{(2x+3)(2y-3)(2y-1)(x+y)}\right]^{1/2}$
h_1-2	h_2+2	h_3	$\begin{pmatrix} & -2 & \\ 2 & & -2 \end{pmatrix}$	$\left[\frac{2(x-1)x(y+1)(2x+2y+1)(2y+2z+3)(x+y+z+1)}{(2x-3)(2x-1)(2y+3)(x+y)}\right]^{1/2}$
h_1-2	h_2	h_3+2	$\begin{pmatrix} & -2 & \\ 0 & & -2 \end{pmatrix}$	$\left[\frac{xy(z+1)(2x+2y-1)(2x+2y+1)(x+y+z+1)}{(2x-1)(2y-1)(x+y-1)(x+y)}\right]^{1/2}$

APPENDIX A: NORMALIZATION COEFFICIENTS OF HWS IN THE CHAIN U(6)⊃U(3)⊃U(2)⊃U(1)

In this appendix, we determine the value of the normalization coefficient $\mathcal{N}(x,y,z)$ appearing in Eq. (3.8). By definition, this coefficient is given by the following equation:

$$[\mathcal{N}(x,y,z)]^{-2} = \langle 0|(\xi_1^+)^x(Y^+)^y(Z^+)^z\xi_1^x Y^y Z^z|0\rangle. \tag{A1}$$

In the calculation of (A1), we shall use throughout the fact that ξ_μ^+ can be interpreted as a differential operator, i.e.,

$$\xi_\mu^+ = \partial/\partial\xi_\mu. \tag{A2}$$

Let us first apply the operator Z^+ to the state $\xi_1^x Y^y Z^z|0\rangle$. From Eq. (3.10), Z^+ can be expanded as follows:

$$Z^+ = Y^+\xi_6^+ + \frac{1}{\sqrt{2}}\dot{Y}^+\xi_5^+ + \frac{1}{\sqrt{2}}\ddot{Y}^+\xi_3^+, \tag{A3}$$

where $\dot{Y}$ and $\ddot{Y}$ are defined by

$$\dot{Y} = \begin{vmatrix} \frac{1}{\sqrt{2}}\xi_3 & \frac{1}{\sqrt{2}}\xi_5 \\ \xi_1 & \frac{1}{\sqrt{2}}\xi_2 \end{vmatrix},$$

$$\text{and}\quad \ddot{Y} = \begin{vmatrix} \frac{1}{\sqrt{2}}\xi_2 & \xi_4 \\ \frac{1}{\sqrt{2}}\xi_3 & \frac{1}{\sqrt{2}}\xi_5 \end{vmatrix}. \tag{A4}$$

As ξ_6^+, ξ_5^+, and ξ_3^+ act neither on ξ_1 nor on Y, and give the following results when acting on Z:

$$\begin{aligned} \xi_6^+ Z|0\rangle &= Y|0\rangle, \\ \xi_5^+ Z|0\rangle &= \sqrt{2}\dot{Y}|0\rangle, \\ \xi_3^+ Z|0\rangle &= \sqrt{2}\ddot{Y}|0\rangle, \end{aligned} \tag{A5}$$

one obtains

$$\begin{aligned} &Z^+\xi_1^x Y^y Z^z|0\rangle \\ &\quad = z(Y^+Y + \dot{Y}^+\dot{Y} + \ddot{Y}^+\ddot{Y})\xi_1^x Y^y Z^{z-1}|0\rangle. \end{aligned} \tag{A6}$$

Using the commutation relations of the boson creation and annihilation operators, the operator $Y^+Y + \dot{Y}^+\dot{Y} + \ddot{Y}^+\ddot{Y}$ can be rewritten in terms of the generators of U(6) as follows:

$$\begin{aligned} Y^+Y + \dot{Y}^+\dot{Y} + \ddot{Y}^+\ddot{Y} &= \mathscr{C}_1{}^1\mathscr{C}_4{}^4 + \tfrac{1}{2}\mathscr{C}_1{}^1\mathscr{C}_5{}^5 + \tfrac{1}{4}(\mathscr{C}_2{}^2)^2 \\ &\quad + \tfrac{1}{4}\mathscr{C}_2{}^2\mathscr{C}_3{}^3 + \tfrac{1}{4}\mathscr{C}_2{}^2\mathscr{C}_5{}^5 + \tfrac{1}{2}\mathscr{C}_3{}^3\mathscr{C}_4{}^4 + \tfrac{3}{2}\mathscr{C}_1{}^1 + \tfrac{5}{4}\mathscr{C}_2{}^2 \\ &\quad + \tfrac{3}{4}\mathscr{C}_3{}^3 + \tfrac{3}{2}\mathscr{C}_4{}^4 + \tfrac{3}{4}\mathscr{C}_5{}^5 + 3 - \frac{1}{2\sqrt{2}}\mathscr{C}_2{}^1(\sqrt{2}\mathscr{C}_2{}^4 + \mathscr{C}_3{}^5) \\ &\quad - \frac{1}{2\sqrt{2}}\mathscr{C}_4{}^2(\sqrt{2}\mathscr{C}_1{}^2 + \mathscr{C}_3{}^5) \\ &\quad - \tfrac{1}{4}\mathscr{C}_5{}^3(\sqrt{2}\mathscr{C}_1{}^2 + \sqrt{2}\mathscr{C}_2{}^4). \end{aligned} \tag{A7}$$

The last three terms in the right-hand side of Eq. (A7) can be transformed in such a way that they give rise to either terms with a raising generator of U(3) acting on the right, which gives zero when applied to the hws $\xi_1^x Y^y Z^{z-1}|0\rangle$, or terms only containing weight generators of U(6). Let us make the explicit transformation for one of them:

$$\mathscr{C}_2{}^1(\sqrt{2}\mathscr{C}_2{}^4+\mathscr{C}_3{}^5)=\mathscr{C}_2{}^1(\mathbb{C}_1{}^2-\sqrt{2}\mathscr{C}_1{}^2)$$
$$=\mathscr{C}_2{}^1\mathbb{C}_1{}^2-\sqrt{2}\mathscr{C}_2{}^2(\mathscr{C}_1{}^1+1). \quad \text{(A8)}$$

As a result, the operator $Y^+Y+\dot{Y}^+\dot{Y}+\ddot{Y}^+\ddot{Y}$, when applied to the hws $\zeta_1{}^xY^yZ^{z-1}|0\rangle$, is equivalent to a second degree polynomial in the weight generators of U(6). This polynomial can now be expressed in terms of weight generators of U(3) only. The final result writes

$$Z^+\zeta_1{}^xY^yZ^z|0\rangle=z(\tfrac{1}{4}\mathbb{C}_1{}^1+1)(\mathbb{C}_2{}^2+3)\zeta_1{}^xY^yZ^{z-1}|0\rangle$$
$$=\tfrac{1}{2}z(x+y+z+1)(2y+2z+1)\zeta_1{}^xY^yZ^{z-1}|0\rangle. \quad \text{(A9)}$$

In the last step of Eq. (A9), use has been made of Eq. (3.7). Repeating this process for the other $(z-1)$ Z^+ operators, one obtains:

$$(Z^+)^z\zeta_1{}^xY^yZ^z|0\rangle$$
$$=[2^{2z}(2y+1)!(x+y+1)!(y+z)!]^{-1}y!z!$$
$$\times(2y+2z+1)!(x+y+z+1)!\zeta_1{}^xY^y|0\rangle. \quad \text{(A10)}$$

Let us now apply the operator Y^+ to the state (A10). As $\zeta_4{}^+$ and $\zeta_2{}^+$ do not act on ζ_1, and give the following results when acting on Y:

$$\zeta_4{}^+Y|0\rangle=\zeta_1|0\rangle,$$
$$\zeta_2{}^+Y|0\rangle=-\zeta_2|0\rangle, \quad \text{(A11)}$$

one obtains

$$Y^+\zeta_1{}^xY^y|0\rangle=y\left[\zeta_1{}^+\zeta_1+\frac{1}{2}\zeta_2{}^+\zeta_2\right]\zeta_1{}^xY^{y-1}|0\rangle. \quad \text{(A12)}$$

The operator $\zeta_1{}^+\zeta_1+\frac{1}{2}\zeta_2{}^+\zeta_2$ can be transformed into

$$\zeta_1{}^+\zeta_1+\tfrac{1}{2}\zeta_2{}^+\zeta_2=\mathscr{C}_1{}^1+\tfrac{1}{2}\mathscr{C}_2{}^2+\tfrac{3}{2}$$
$$=\tfrac{1}{2}(\mathbb{C}_1{}^1-\mathscr{C}_3{}^3+3), \quad \text{(A13)}$$

where $\mathscr{C}_3{}^3$ disappears when applied to the state $\zeta_1{}^xY^{y-1}|0\rangle$. Equation (A12) thus becomes

$$Y^+\zeta_1{}^xY^y|0\rangle=\tfrac{1}{2}y(\mathbb{C}_1{}^1+3)\zeta_1{}^xY^{y-1}|0\rangle$$
$$=\tfrac{1}{2}y(2x+2y+1)\zeta_1{}^xY^{y-1}|0\rangle. \quad \text{(A14)}$$

Repeating this process y times, one obtains

$$(Y^+)^y\zeta_1{}^xY^y|0\rangle$$
$$=[2^{2y}(2x+1)!(x+y)!]^{-1}x!y!(2x+2y+1)!\zeta_1{}^x|0\rangle. \quad \text{(A15)}$$

It only remains to apply $\zeta_1{}^+$ x times to Eq. (A15) and use the result

$$(\zeta_1{}^+)^x\zeta_1{}^x|0\rangle=x!|0\rangle. \quad \text{(A16)}$$

By combining Eqs. (A1), (A10), (A15), and (A16), one gets Eq. (3.11) for the normalization coefficient $\mathscr{N}(x,y,z)$.

APPENDIX B: MATRIX ELEMENT OF ζ_4 BETWEEN HWS

In this appendix, we determine the matrix element of ζ_4 between the hws corresponding to the IR's $[h_1h_2h_3]$ and $[h_1h_2+2h_3]$.

By definition, it is equal to

$$\left\langle \begin{matrix} h_1 & h_2+2 & h_3 \\ & \text{hws} & \end{matrix}\right|\zeta_4\left|\begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix}\right\rangle$$
$$=\mathscr{N}(x-1,y+1,z)$$
$$\times\left\langle \begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix}\right|\zeta_4{}^+\zeta_1{}^{x-1}Y^{y+1}Z^z|0\rangle, \quad \text{(B1)}$$

where use has been made of its reality. As the action of $\zeta_4{}^+$ on Y and Z is given by

$$\zeta_4{}^+Y|0\rangle=\zeta_1|0\rangle,$$
$$\zeta_4{}^+Z|0\rangle=[\zeta_1\zeta_6-\tfrac{1}{2}\zeta_3{}^2]|0\rangle, \quad \text{(B2)}$$

Eq. (B1) can be transformed into

$$\left\langle \begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix}\right|\zeta_4\left|\begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix}\right\rangle$$
$$=\mathscr{N}(x-1,y+1,z)\{(y+1)[\mathscr{N}(x,y,z)]^{-1}$$
$$+z\mathscr{N}(x,y,z)\langle 0|(\zeta_1{}^+)^x(Y^+)^y(Z^+)^z$$
$$\times\left[\zeta_1\zeta_6-\frac{1}{2}\zeta_3{}^2\right]\zeta_1{}^{x-1}Y^{y+1}Z^{z-1}|0\rangle\Big\}, \quad \text{(B3)}$$

where it remains to calculate the matrix element of the right-hand side.

Using the reality of the matrix elements again, we obtain from Eq. (A9)

$$\langle 0|(\zeta_1{}^+)^{x-1}(Y^+)^{y+1}\left[\zeta_1{}^+\zeta_6{}^+-\frac{1}{2}(\zeta_3{}^+)^2\right]$$
$$\times(Z^+)^{z-1}\zeta_1{}^xY^yZ^z|0\rangle$$
$$=[2^{2z-2}(2y+3)!(x+y+2)!(y+z)!]^{-1}(y+1)!z!$$
$$\times(2y+2z+1)!(x+y+z+1)!$$
$$\times\langle 0|(\zeta_1{}^+)^{x-1}(Y^+)^{y+1}\left[\zeta_1{}^+\zeta_6{}^+-\frac{1}{2}(\zeta_3{}^+)^2\right]$$
$$\times\zeta_1{}^xY^yZ|0\rangle, \quad \text{(B4)}$$

where

$$\langle 0|(\zeta_1{}^+)^{x-1}(Y^+)^{y+1}\left[\zeta_1{}^+\zeta_6{}^+-\frac{1}{2}(\zeta_3{}^+)^2\right]\zeta_1{}^xY^yZ|0\rangle$$
$$=\langle 0|(\zeta_1{}^+)^{x-1}(Y^+)^{y+1}\left[\zeta_1{}^+\zeta_1{}^xY^{y+1}+\frac{1}{2}\zeta_4\zeta_1{}^xY^y\right]|0\rangle$$
$$=x[\mathscr{N}(x-1,y+1,0)]^{-2}+\frac{1}{2}(2y+3)$$
$$\times\langle 0|(\zeta_1{}^+)^{x-1}(Y^+)^{y+1}\zeta_4\zeta_1{}^xY^y|0\rangle. \quad \text{(B5)}$$

The last matrix element then reduces to

$$\langle 0|(\zeta_1{}^+)^x(Y^+)^y\zeta_4{}^+\zeta_1{}^{x-1}Y^{y+1}|0\rangle=(y+1)[\mathscr{N}(x,y,0)]^{-2}. \quad \text{(B6)}$$

Combining the results (B3), (B4), (B5), and (B6), and using Eq. (3.11) for $\mathscr{N}(x,y,z)$, we finally obtain the matrix element of ζ_4 as

$$\left\langle \begin{matrix} h_1 & h_2+2 & h_3 \\ & \text{hws} & \end{matrix}\right|\zeta_4\left|\begin{matrix} h_1 & h_2 & h_3 \\ & \text{hws} & \end{matrix}\right\rangle$$
$$=\left[\frac{2x(y+1)(2y+2z+3)}{(2x+1)(2y+3)}\right]^{1/2}. \quad \text{(B7)}$$

[1] R. V. Dzholos, F. Dönau, and D. Janssen, Yad. Fiz. **22**, 965 (1975) [Sov. J. Nucl. Phys. **22**, 503 (1976)].

[2] A. Arima and F. Iachello, Ann. Phys. (N.Y.) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).

[3]V. Vanagas, E. Nadjakov, and P. Raychev, Preprint IC/75/140, Trieste 1976 and Bulg. J. Phys. **11**, 6 (1975); V. Vanagas, "*The Microscopic Nuclear Theory Within the Framework of the Restricted Dynamics*" in *Lecture Notes in Physics* (University of Toronto, Toronto, 1977).
[4]A. Bohr, Mat. Fys. Medd. Danske Vid. Selsk. **26**, No. 14 (1952); A. Bohr and B. R. Mottelson, Mat. Fys. Medd. Danske Vid. Selsk. **27**, No. 16 (1953).
[5]O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[6]E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 668 (1976); E. Chacón and M. Moshinsky, *ibid.* **18**, 870 (1977).
[7]M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962).
[8]M. Moshinsky and V. Syamala Devi, J. Math. Phys. **10**, 455 (1969).
[9]A. Partensky and C. Quesne (to be published).
[10]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 562 (1958).
[11]M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1967).
[12]R. Pérez, "SU (3) Symmetry and Realistic Interactions," Thesis, U.N.A.M., Mexico (1967).
[13]V. Vanagas, S. Ališauskas, R. Kalinauskas, and E. Nadjakov, Bulg. J. Phys. **7**, 168 (1980).
[14]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).
[15]K. T. Hecht, Nucl. Phys. **62**, 1 (1965).
[16]J. A. Castilho Alcaras, L. C. Biedenharn, K. T. Hecht, and G. Neely, Ann. Phys. (N.Y.) **60**, 85 (1970).
[17]V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).
[18]E. Chacón and M. Moshinsky, Phys. Lett. **23**, 567 (1966); M. Moshinsky and E. Chacón, in *Spectroscopic and Group Theoretical Methods in Physics*, edited by F. Bloch, S. G. Cohen, A. de-Shalit, S. Sambursky, and I. Talmi (North-Holland, Amsterdam, 1968), p. 99.

Paraxial self-trapped beams in nonlinear optics

Robert D. Small
Department of Mathematics, University of New Brunswick, P. O. Box 4400, Fredericton, New Brunswick, E3B 5A3, Canada

(Received 3 September 1980; accepted for publication 23 January 1981)

Self-trapped beam solutions to the equations of nonlinear optics are studied by perturbing the straight ray solutions. In two dimensions, analytic forms for the perturbation solutions are obtained while for the radially symmetric case numerical solutions are produced. In both cases the perturbation amplitude contains an arbitrary function of distance so that a large class of self-trapped beams is produced. The solutions obtained are paraxial, and it is argued that they are the only symmetric paraxial solutions. It is further pointed out that solving the well-known paraxial equations leads to errors unless the solutions obtained are of the form found here since otherwise they violate the paraxial approximation.

PACS numbers: 42.65.Jx

I. INTRODUCTION

Light rays in certain nonlinear media tend to converge due to the nonlinear effects. The property responsible for self-focusing and self-trapping of beams is that the phase velocity of the waves decreases with increasing amplitude. A disturbance that has a distribution of amplitude along an initially flat wavefront will be retarded at its more intense points, and rays will bend inwards there. The convergence of rays increases the amplitude enhancing the nonlinear effect and forcing greater convergence of rays until diffraction effects become felt. Diffraction, which resists sudden changes in amplitude, causes rays to bend outward to disperse the energy of the waves. When the opposing effects of nonlinear focusing and diffraction are present, a great variety of results can take place. It has been found experimentally that beams can focus and emerge from the focus, contract to one or more self-trapped filaments, or merely diffuse away from the initial distribution. Askary'an[1] proposed on theoretical grounds that a beam of light could create a self-induced waveguide. Later, Talanov[2] and Chiao *et al.*[3] calculated beam profiles based on balancing self-focusing and diffraction so that the beam propagates undistorted. This paper is devoted to constructing profiles that generalize these undistorted beams by perturbing them with slowly varying functions.

Beams that vary slowly in the direction of propagation relative to distances of the order of a beam width are said to obey the paraxial approximation since rays remain close to parallel to the beam axis. When the paraxial approximation is satisfied, the governing nonlinear equations are simplified. It is necessary for the validity of the results, however, that solutions that satisfy the simplified equations also satisfy the paraxial approximation. In this paper we produce general paraxial solutions by perturbing beams that are independent of the direction of propagation and then by accepting only slowly varying perturbations rather than solving the paraxial equations and checking the results for validity.

The paraxial approximation was introduced into nonlinear optics as a natural extension of the properties of linear optics. Chiao *et al.*[3] use it without reference, and later Kelley[4] in using the equations of Chiao refers to it as a "familiar" procedure. Apparently independently, Talanov[2] used this approximation in nonlinear optics following the example of Vainstein[5] who used it in resonator and waveguide problems. Vainstein gives as reference the review of diffraction phenomena by Malyuzhinets,[6] who indicated the validity of the approximation for diffraction of nearly plane waves of small wave length by apertures, the waves governed by the linear wave equation. Since then a tremendous volume of research has been generated using the paraxial equations as starting point. The reviews by Marburger[7] and Svelto[8] give substantial lists of references that cover this research up to their time of publication. Some of the papers appearing since then are included in Refs. 9–26. Marburger seems to be the only one to mention the possibility of nonphysical solutions due to neglect of higher space derivatives, and he justifies his own analysis by prescribing symmetry between certain time derivatives and derivatives in the direction of propagation. It is the opinion of this writer that much of the paraxial research on self-focusing does not account for the extraneous solutions that satisfy the paraxial equations but violate the underlying Maxwell equations due to the additional nonlinear effects not present in linear theories of diffraction.

In Sec. II we derive the differential equations for the envelope quantities without invoking the paraxial approximation. A general nonlinearity is utilized, although it is emphasized that the cubic nonlinearity is the dominant one.

In Sec. III we perturb the undistorted beams both in two and three dimensions. The former is treated because, while primarily of academic interest, it possesses analytic solutions, while radially symmetric beams in three dimensions must be expressed in terms of the undistorted beam profiles that are only known numerically. For the latter case we produce graphs of the perturbed profiles and of their transverse wavenumbers. It turns out that the perturbed beams contain an unspecified function of distance along the beam axis. Thus a great variety of beams can occur which include many of the cases found experimentally.

In Sec. IV we derive the paraxial equations and indicate the relation between paraxial solutions and the perturbation solutions of Sec. III. As a comparison with paraxial theory, we investigate the effect of Zakharov's[27] criterion for self-trapping on the perturbation solutions.

 0022-2488/81/071497-07$1.00

II. THE ENVELOPE EQUATIONS

Maxwell's equations describing waves in a nonlinear medium reduce to

$$c^2\nabla^2\mathbf{E} - \mathbf{E}_{tt} = (1/\epsilon_0)\mathbf{P}_{tt} \tag{1}$$

if one makes the assumption

$$\nabla\cdot\mathbf{E} = 0. \tag{2}$$

Equation (2) is exact for the two-dimensional analysis but is only approximate for the three-dimensional case, where it depends on the electric field not being unusually large so that the relation between $\mathbf{E}$ and $\mathbf{P}$ is close to linear. In this circumstance the dominant term of the Maxwell equation

$$\nabla\cdot(\epsilon_0\mathbf{E} + \mathbf{P}) = 0$$

asserts (2). Marburger,[7] in his review, indicates that if this approximation is not made, then "there are no confined wave solutions with purely linear polarization everywhere." He cites the work of Abakarov *et al.*[28] and Pohl,[29] who construct special field configurations in which $\nabla\cdot\mathbf{E}$ is identically zero. Svelto's[8] explanation for accepting (2) is that this approximation retains the nonlinear effect of self-focusing while neglecting the nonlinear effect that causes rotation of the polarization ellipse.

For an isotropic medium $\mathbf{P}$ is parallel to $\mathbf{E}$, and, when wave patterns are stationary, the frequency is fixed so that $|\mathbf{P}|$ depends on $|\mathbf{E}|$ algebraically. We retain the differential equation describing an inertial medium, however, as the calculations are simplified if the amplitude of $\mathbf{P}$ is determined first and then eliminated. This relation is

$$\mathbf{P}_{tt} + \omega_0{}^2\mathbf{P} - F(|\mathbf{P}|^2)\mathbf{P} = \epsilon_0\omega_p{}^2\mathbf{E}, \tag{3}$$

and it may be regarded naively as Newton's law for the motion of electrons in the medium. They vibrate with a natural frequency ω_0, which is the resonant frequency of the medium, and are driven by the external electric field. The nonlinearity of the medium is expressed by $F(|\mathbf{P}|^2)$ with any linear part absorbed into the $\omega_0{}^2\mathbf{P}$ term. Since $F(|\mathbf{P}|^2)$ must be small compared to $\omega_0{}^2$, its Taylor series will be dominated by the term quadratic in $|\mathbf{P}|$ except for very unusual media. Both Akhmanov[30] and Zakharov[27] take quadratic and quartic terms in F, where the quartic one dominates near focal points as the medium saturates. Many of the physical mechanisms that lead to the nonlinear term are described in the review by Svelto.[8]

To obtain equations governing the stationary wave envelopes, we make the substitution

$$\begin{aligned}\mathbf{E} &= (a(x,y,z)\cos[\theta(x,y,z) - \omega t],0,0),\\ \mathbf{P} &= (b(x,y,z)\cos[\theta(x,y,z) - \omega t],0,0)\end{aligned} \tag{4}$$

for linearly polarized waves, or

$$\begin{aligned}\mathbf{E} &= (a(x,y,z)\cos[\theta(x,y,z) - \omega t],\\ &\quad a(x,y,z)\sin[\theta(x,y,z) - \omega t],0),\\ \mathbf{P} &= (b(x,y,z)\cos[\theta(x,y,z) - \omega t],\\ &\quad b(x,y,z)\sin[\theta(x,y,z) - \omega t],0)\end{aligned} \tag{5}$$

for circularly polarized waves, where the wavenumber vector $\mathbf{k}$ is defined by

$$\mathbf{k}(x,y,z) = \nabla\theta. \tag{6}$$

Since the resulting equations have no preferred direction, substitution of waves (4) and (5) have the same effect as expressions for waves propagating in arbitrary directions. The equations obtained by substituting (4) or (5) into (1) and (3), collecting terms in $\cos(\theta - \omega t)$ and $\sin(\theta - \omega t)$ and eliminating b to one order of smallness in F are

$$\nabla^2 a - (k^2 - k_0{}^2)a + f(a^2)a = 0, \tag{7}$$

$$\nabla\cdot(\mathbf{k}a^2) = 0, \tag{8}$$

$$\nabla\times\mathbf{k} = 0, \tag{9}$$

where

$$k_0{}^2 = \frac{\omega^2}{c^2}\left(1 + \frac{\omega_p{}^2}{\omega_0{}^2 - \omega^2}\right)$$

and

$$f(a^2) = \frac{\omega^2\omega_p{}^2}{2\pi c^2(\omega_0{}^2 - \omega^2)^2} \times\int_0^{2\pi} F\left(\frac{\epsilon_0{}^2\omega_p{}^4 a^2\cos^2\theta}{(\omega_0{}^2 - \omega^2)^2}\right)\cos^2\theta\, d\theta$$

for the case of linear polarization and

$$f(a^2) = \frac{\omega^2\omega_p{}^2}{c^2(\omega_0{}^2 - \omega^2)^2} F\left(\frac{\epsilon_0{}^2\omega_p{}^4 a^2}{(\omega_0{}^2 - \omega^2)^2}\right)$$

for circular polarization. For either polarization the example

$$F(|\mathbf{P}|^2) = \alpha|\mathbf{P}|^{2n}$$

leads to the expression

$$f(a^2) = \beta a^{2n},$$

but with different values of the constant β. Similarly the saturating medium cited previously,

$$F(|\mathbf{P}|^2) = \alpha_1 a^2 + \alpha_2 a^4,$$

leads to the form

$$f(a^2) = \beta_1 a^2 + \beta_2 a^4.$$

We can make some observations immediately by solving (7) for the phase velocioty ω/k, obtaining

$$\frac{\omega}{k} = \left[\frac{\omega_0{}^2 + \omega_p{}^2 - \omega^2}{c^2(\omega_0{}^2 - \omega^2)} + \frac{\nabla^2 a}{\omega^2 a} + \frac{f(a^2)}{\omega^2 a}\right]^{-1/2}.$$

Neglecting the second derivatives and the term f, we have the phase velocity of a uniform plane wave in a linear medium. If $\omega_p = 0$, the medium becomes uncoupled from the electromagnetic wave and the wave passes at speed c. With coupling due to finite ω_p, its speed is less than c and depends on the wave frequency ω. This is called a dispersive medium because the group velocity $\partial\omega/\partial\kappa$ depends on ω so that wavepackets of different frequencies separate. When $\omega = \omega_0$, the wave resonates with the medium and cannot pass through it. Also there is a frequency band $\omega_0 \leqslant \omega \leqslant (\omega_0{}^2 + \omega_p{}^2)^{1/2}$, where the wave reflects backwards from the medium, the amplitude falling exponentially as the wave penetrates.

When the second derivatives of amplitude become large relative to the amplitude, there is an additional influence on the phase velocity. The derivatives are called higher disper-

sion terms since they modify the previously discussed dispersive effects and they are also responsible for diffraction. Near the peak of a local maximum in amplitude we have $\nabla^2 a < 0$. This negative term increases the phase velocity relative to other portions of the wave front giving a defocussing effect. Finally the term containing f gives the nonlinear influence of the medium on the wave. For $f > 0$ this term causes the phase velocity to fall as amplitude rises so that a local maximum in amplitude tends to focus. When diffraction and nonlinearity are in perfect balance, we obtain the self-trapped beam of Chiao *et al.*,[3] where rays are straight.

III. SOLUTIONS OF THE ENVELOPE EQUATIONS

In this section we perturb exact straight ray solutions to (7), (8), and (9) to obtain slowly varying amplitude profiles. The perturbed solutions obey the paraxial approximation and can be compared to the work of other authors. Analytic expressions for the straight ray solutions are available in two dimensions with cubic nonlinearity (F is quadratic), and hence this analysis is presented to prepare for the treatment of three-dimensional waves. In the latter case, with radial symmetry and arbitrary nonlinearity, the straight ray solutions must be determined numerically and the perturbed solutions are then expressed in terms of the computed functions.

A. Two-dimensional waves with cubic nonlinearity

In two dimensions with $F(|\mathbf{P}|) = \alpha a^2$, (7), (8), and (9) take the form

$$a_{xx} + a_{yy} - (k^2 - k_0^2)a + \beta a^3 = 0, \tag{10}$$

$$(k_1 a^2)_x + (k_2 a^2)_y = 0, \tag{11}$$

$$k_{1y} = k_{2x}. \tag{12}$$

We first find a solution that represents a uniform self-trapped beam by seeking solutions independent of x with the underlying waves propagating in the x direction. Such a beam is

$$a = \bar{a}(y) = [2(K^2 - k_0^2)/\beta]^{1/2} \operatorname{sech}[(K^2 - k_0^2)^{1/2} y], \tag{13}$$

$$k_1 = K, \quad k_2 = 0.$$

This solution decays for large $|y|$ while all others oscillate in y without decay. It has been discussed by Chiao,[3] Akhmanov,[30] and many others.

To upset the balance that maintains straight rays, we perturb this solution by functions that vary slowly in x. We set

$$\bar{x} = \epsilon x, \tag{14}$$

$$\bar{y} = (1 + \epsilon^2 \delta_2 + \epsilon^3 \delta_3 + \cdots) y,$$

where ϵ is a small positive constant and seek solutions of (10), (11), and (12) that are power series in ϵ of the form

$$a = \bar{a}(\bar{y}) + \epsilon G_1(\bar{x}, \bar{y}) + \cdots,$$

$$k_1 = K + \epsilon U_1(\bar{x}, \bar{y}) + \cdots, \tag{15}$$

$$k_2 = \epsilon V_1(\bar{x}, \bar{y}) + \epsilon^2 V_2(\bar{x}, \bar{y}) + \cdots.$$

The hierarchy of equations is found by equating coefficients of each power of ϵ. The δ_i are determined by suppressing secular terms in the G_i at each stage of solution. To the order of accuracy that we are solving we are able to replace $\bar{y}$ by y. We seek beams that are symmetric about the $\bar{x}$ axis, and this requires that G_1 and U_1 be even in y and that V_1 and V_2 be odd in y.

Placing (15) into (10), (11), and (12), we obtain the following leading order equations, the first two orders being taken from (11):

$$G_{1yy} - (K^2 - k_0^2)G_1 - 2K\bar{a}U_1 + 3\beta\bar{a}^2 G_1 = 0, \tag{16}$$

$$(\bar{a}^2 V_1)_y = 0, \tag{17}$$

$$\bar{a}^2 U_{1\bar{x}} + 2K\bar{a}G_{1\bar{x}} + (2\bar{a}V_1 G_1)_y + (\bar{a}^2 V_2)_y = 0, \tag{18}$$

$$U_{1y} = 0. \tag{19}$$

From (17) and (13) we obtain

$$V_1 = C(\bar{x}) \cosh^2[(K^2 - k_0^2)^{1/2} y].$$

Since V_1 must be odd, we take $C(\bar{x}) \equiv 0$, and this necessitates finding V_2 to obtain the y component of $\mathbf{k}$. From (19) we have

$$U_1(\bar{x}, y) = u(\bar{x}),$$

where $u(\bar{x})$ is an arbitrary function of $\bar{x}$. Succeeding terms in expansion (15) will not be small unless $u(\bar{x})$ and all of its derivatives are slowly varying; hence we must take $u(\bar{x})$ to be analytic in $\bar{x}$ with all derivatives no greater than $O(1)$. When U_1 is placed into (16) the $\bar{x}$ dependence factors out. Setting

$$G_1 = u(\bar{x})g(y),$$

(16) becomes

$$g_{yy} - (K^2 - k_0^2)\{1 - 6 \operatorname{sech}^2[(K^2 - k_0^2)^{1/2} y]\}g = 2^{3/2} K [(K^2 - k_0^2)/\beta]^{1/2} \operatorname{sech}[(K^2 - k_0^2)^{1/2} y]. \tag{20}$$

It may be noticed that $\partial\bar{a}/\partial K$ satisfies (20). This is the bounded symmetric solution. The general solution contains in addition an odd term proportional to $\partial\bar{a}/\partial y$ and an unbounded term which may be found by variation of parameters. Both of these latter terms are discarded, leaving

$$g(y) = (2/\beta)^{1/2} K \{(K^2 - k_0^2)^{-1/2} \operatorname{sech}[(K^2 - k_0^2)^{1/2} y] - y \operatorname{sech}[(K^2 - k_0^2)^{1/2} y] \tanh[(K^2 - k_0^2)^{1/2} y]\}.$$

Finally (18) gives

$$V_2 = -u'(\bar{x})\left\{\frac{K^2 y}{K^2 - k_0^2} + \frac{2K^2 - k_0^2}{2(K^2 - k_0^2)^{3/2}} \sinh[2(K^2 - k_0^2)^{1/2} y]\right\}.$$

The perturbed solution is now fixed when the arbitrary function $u(\bar{x})$ is specified. We can obtain a great variety of beams by the choice of $u(\bar{x})$. For a beam with periodically converging and diverging rays we could choose for example $u(\bar{x}) = \cos\bar{x}$. If $u(\bar{x})$ is chosen to have a singularity, then this solution can possibly describe a focusing beam. It will be an accurate description of the behavior only in its approach to focus, the various approximations breaking down both due to fast variations in $\bar{x}$ and due to high amplitudes.

B. Three-dimensional waves with radial symmetry

We now seek radially symmetric solutions to (7), (8), and (9). For a general nonlinearity these equations become

$$a_{xx} + a_{rr} + (1/r)a_r - (k^2 - k_0^2)a + f(a^2)a = 0,$$
$$(k_1 a^2)_x + (1/r)(r k_2 a^2)_r = 0, \tag{21}$$
$$k_{1r} = k_{2x},$$

where k_1 and k_2 are the components of $\mathbf{k}$ in the x and r directions respectively. A uniform beam is sought in the form

$$a = \bar{a}(r), \quad k_1 = K, \quad k_2 = 0. \tag{22}$$

Substitution into (21) gives

$$\bar{a}_{rr} + (1/r)\bar{a}_r - (K^2 - k_0^2)a + f(\bar{a}^2)\bar{a} = 0,$$
$$\bar{a}_r(0) = 0, \quad \bar{a}(\infty) = 0. \tag{23}$$

Under suitable restrictions on $f(a^2)$, (23) possesses an infinite set of solutions that have any integral number of zeros. The following demonstration of the existence of this set resembles that given by Finkelstein *et al.*,[31] who credit it to Bohnenblust. They dealt with an equation describing probability amplitudes for spinor fields.

If the term $(1/r)\bar{a}_r$ is neglected from (23), there is an integral

$$B = \bar{a}_r^2 - (K^2 - k_0^2)\bar{a}^2 + \hat{f}(\bar{a}^2), \tag{24}$$

where $\hat{f}(\xi) = \int_0^\xi f(\xi')\,d\xi'$. The constant B parametrizes non-intersecting level curves in the phase plane plot, $\bar{a}_r$ vs $\bar{a}$. The singular points are at $\bar{a}_r = 0, \bar{a} = 0$ and at $\bar{a}_r = 0, \bar{a}$ a root a_0 of

$$f(a_0^2) = K^2 - k_0^2. \tag{25}$$

If there is only one positive root for a_0^2 in (25), then there are three singular points, as shown in Fig. 1. The origin is a saddle point, and the remaining two are centers. The level curves emanating from the origin correspond to $B = 0$ and the regions containing curves for $B > 0$ and $B < 0$ are separated by this curve. Any curve interior to another curve has a lower value of B than the latter.

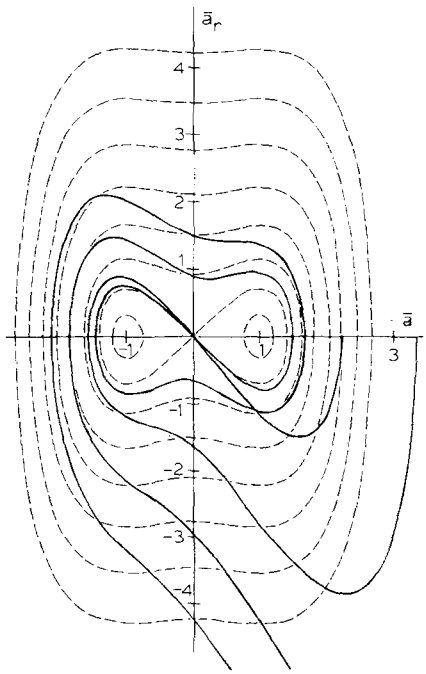


FIG. 1. Phase plane trajectories for radially symmetric beam amplitudes. The broken lines represent integral curves for the quantity B defined in the text. The solid lines represent the trajectories which continually cross integral curves in the direction of decreasing B. The scale is normalized by setting $a_0 = 1$.

We now regard (24) as the definition of a quantity B for the exact equation (23). Differentiation of (24) and the use of (23) gives

$$\frac{dB}{dr} = -\frac{2}{r}a_r^2. \tag{26}$$

Thus, along a trajectory given by (23), as r increases B decreases, and the trajectory in the phase plane proceeds across curves of Fig. 1, always toward more interior regions. The trajectories that begin on the $\bar{a}$ axis and end at the origin are the ones that satisfy the boundary conditions. They separate dense sets of trajectories that are trapped by the singular points $\bar{a} = \pm a_0$. Figure 1 is drawn for the case $f(a^2) = a^2$ and the profiles for the desired trajectories appear in Fig. 2. These trajectories were computed first by Chiao *et al.*,[3] who did the lowest order one and Haus,[32] who computed the next four on an analog computer. Zakharov[27] gives the starting amplitudes for the first 15 trajectories.

We now assume that $f(\xi)$ is monotonic increasing so that there is a unique positive root to (25) and we consider (22) as a definite solution to (23). Perturbing this solution by a slowly varying perturbation, we set

$$a = \bar{a}(\bar{r}) + \epsilon G_1(\bar{x},\bar{r}) + \cdots,$$
$$k_1 = K + \epsilon U_1(\bar{x},\bar{r}) + \cdots, \tag{27}$$
$$k_2 = \epsilon V_1(\bar{x},\bar{r}) + \epsilon^2 V_2(\bar{x},\bar{r}) + \cdots,$$

where ϵ is a small positive quantity and

$$\bar{x} = \epsilon x, \quad \bar{r} = (1 + \delta_2\epsilon^2 + \cdots)r.$$

As in the two-dimensional case, equating coefficients of like powers of ϵ gives a hierarchy of equations and the δ_i are determined by suppressing secular terms. To the order of accuracy that we compute we can take $\bar{r} = r$.

The equations obtained by substituting (27) into (21) are

$$G_{1rr} + (1/r)G_{1r} - [K^2 - k_0^2 - f(\bar{a}^2) - 2\bar{a}^2 f'(\bar{a}^2)]G_1 = 2K\bar{a}U_1, \tag{28}$$
$$(1/r)(r\bar{a}^2 V_1)_r = 0, \tag{29}$$
$$\bar{a}^2 U_{1\bar{x}} + 2K\bar{a}G_{1\bar{x}} + (1/r)(r\bar{a}^2 V_2 + 2r\bar{a}G_1 V_1)_r = 0, \tag{30}$$
$$U_{1r} = 0. \tag{31}$$

From (29) we obtain

$$V_1 = [C(\bar{x})/r]\bar{a}^{-2}.$$

Since this is singular at $r = 0$, we must take $C(\bar{x}) \equiv 0$. Equation (31) indicates that $U_1(\bar{x},r) \equiv u(\bar{x})$, where $u(\bar{x})$ is an analytic function of $\bar{x}$ with derivatives no larger than $O(1)$. Then the $\bar{x}$ dependence of (28) factors out and the substitution

$$G_1(\bar{x},r) = u(\bar{x})g(r)$$

gives

$$g_{rr} + (1/r)g_r - [K^2 - k_0^2 - f(\bar{a}^2) - 2\bar{a}^2 f'(\bar{a}^2)]g = 2K\bar{a}. \tag{32}$$

It may be verified that $g = \partial\bar{a}/\partial K$ is a solution to (32), and it is the only solution such that $\partial g(0)/\partial r = 0, g(\infty) = 0$. Finally

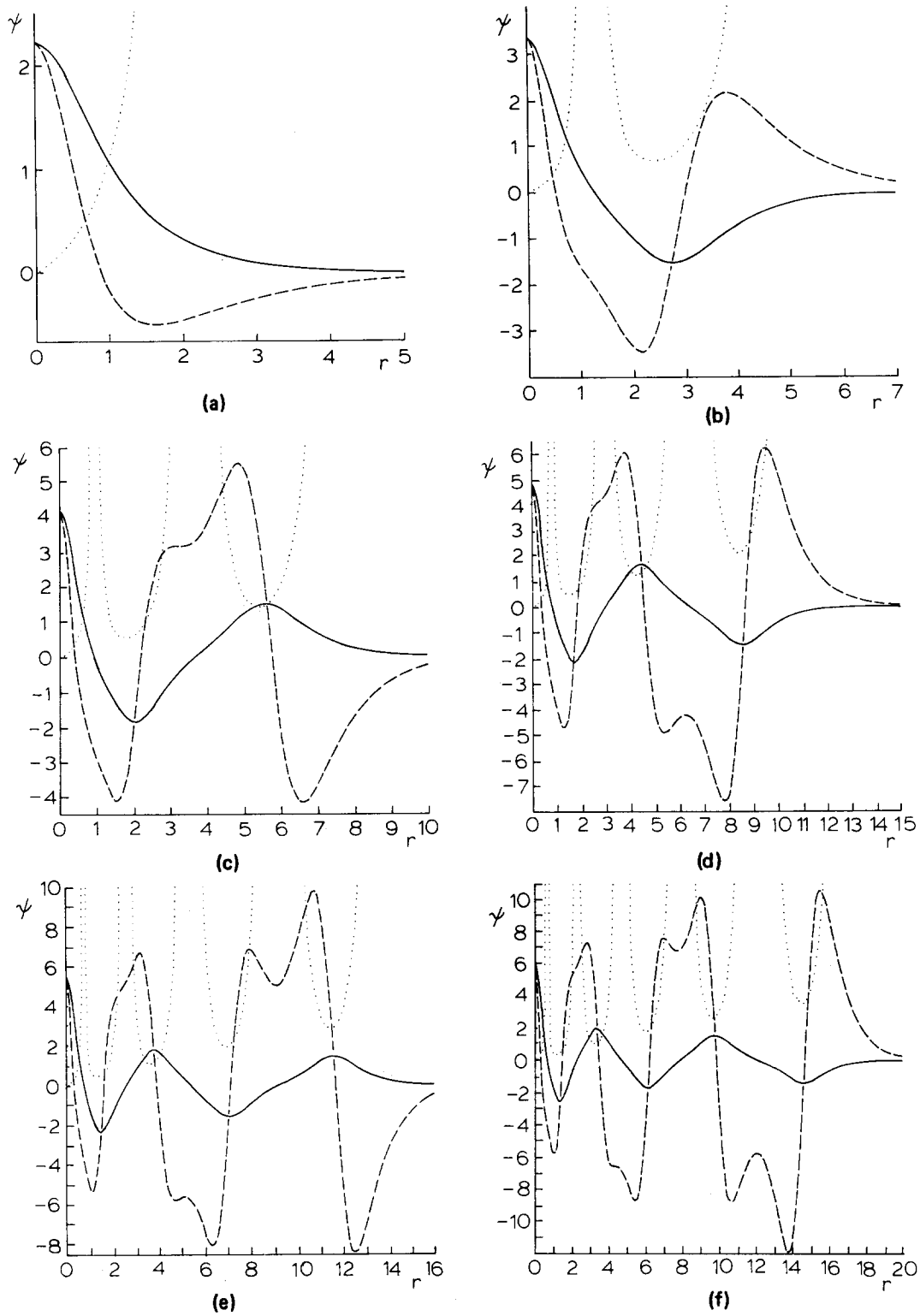


FIG. 2. Profiles of undistorted beam amplitude, perturbation amplitude, and transverse wavenumber. The solid lines represent the first six eigensolutions ψ of Eq. (36). The broken lines represent the corresponding perturbation amplitudes μ, and the dotted lines represent one term ν of the perturbation transverse wavenumber.

(30) gives

$$V_2 = -\frac{u'(\bar{x})}{r\bar{a}^2}\int_0^r \left(\bar{a}^2 + 2K\bar{a}\frac{\partial \bar{a}}{\partial K}\right) r\,dr. \tag{33}$$

This form of the final solution can be evaluated numerically only with great difficulty in the general case. For the special case

$$f(\bar{a}^2) = \beta \bar{a}^{2n}, \tag{34}$$

where $\beta > 0$, in order to have a solution to (23), differentiation of (23) with respect to both K and r and some manipulation gives

$$\frac{\partial \bar{a}}{\partial K} = \frac{K}{K^2 - k_0^2}\left(\frac{\bar{a}}{n} + r\bar{a}_r\right) \tag{35}$$

so that the perturbed solution is determined in a simple way by the original straight ray solution and its derivative. We now obtain from (33) the result

$$V_2 = -u'(\bar{x})\left\{\frac{K^2 r}{K^2 - k_0^2} + \left[1 + \frac{2K^2}{K^2 - k_0^2}\left(\frac{1}{n} - 1\right)\right] \times \frac{1}{r\bar{a}^2}\int_0^r \bar{a}^2 r\, dr\right\}.$$

As mentioned previously, the cubic term is the dominant nonlinearity for other than very special media. Taking $n = 1$, the scaling

$$\bar{a} = [(K^2 - k_0^2)/\beta]^{1/2}\psi[(K^2 - k_0^2)^{1/2} r]$$

reduces (23) to

$$\psi_{\xi\xi} + (1/\xi)\psi_\xi - \psi + \psi^3 = 0, \qquad \psi_\xi = 0, \quad \psi(\infty) = 0. \tag{36}$$

The perturbation amplitude becomes

$$G_1 = \{u(\bar{x})K/[\beta(K^2 - k_0^2)]\}^{1/2}\,\mu(\xi),$$

where $\mu(\xi) = \psi + \xi\psi_\xi$. The transverse perturbation wavenumber becomes

$$V_2 = -u'(\bar{x})\left\{\frac{K^2\xi}{(K^2 - k_0^2)^{3/2}} + \frac{\nu(\xi)}{(K^2 - k_0^2)^{1/2}}\right\}$$

where $\nu(\xi) = (\int_0^\xi \psi^2 \xi\, d\xi)/\psi^2\xi$.

Figures 2(a)–(f) display the first six numerically computed eigensolutions to (36) and the corresponding plots of $\mu(\xi)$ and $\nu(\xi)$. It may be noted that k_2 is singular at zeros of ψ, but these are precisely the places where the wavenumber need not be defined.

As in the two-dimensional case, if $u(\bar{x})$ is chosen to have a singularity, then the solution can possibly describe the approach to focus of a focusing beam.

IV. THE PARAXIAL EQUATIONS AND CONCLUSIONS

In this section we present the derivation of the paraxial equations and an argument that the perturbed solutions of Sec. III form the general symmetric solutions of these equations that also satisfy the paraxial condition. Since a great volume of research has been generated from the paraxial equations, we do not attempt a general comparison of results but choose instead the asymptotic analysis of Zakharov *et al.*[27] as a single example. These workers have produced criteria that determine asymptotic properties of solutions of the paraxial equations based on certain integrals of their initial values.

The paraxial equations can be obtained from (7), (8), and (9) by setting

$$\theta = Kx - \phi(\bar{x}, y, z),$$

where

$$\bar{x} = \epsilon x \quad \text{and} \quad \epsilon \ll K.$$

If the terms in ϵ^2 are discarded and then x is reintroduced for $\bar{x}/\epsilon$, the results are

$$a_{yy} + a_{zz} - [K^2 - k_0^2 - f(a^2) - 2K\phi_x + \phi_y^2 + \phi_z^2]a = 0, \tag{37}$$

$$K(a^2)_x = (a^2\phi_y)_y + (a^2\phi_z)_z,$$

which are the paraxial equations in common use. If we write

$$u = a(x, y, z)e^{i\phi(x, y, z)},$$

then (37) is equivalent to the complex form

$$2iKu_x = u_{yy} + u_{zz} - [K^2 - k_0^2 - f(|u|^2]u, \tag{38}$$

which is the nonlinear Schrödinger equation that stands as the starting point for many analyses.

The following argument now relates the perturbation solutions to the legitimate solutions of (37) and (38). We consider any solution of (37) that is slowly varying in x so that the paraxial condition applies. It may be expanded by Taylor series as

$$a(\bar{x}, y, z) = a(0, y, z) + \epsilon x\frac{\partial a}{\partial \bar{x}}(0, y, z) + O(\epsilon^2),$$

$$\mathbf{k}(\bar{x}, y, z) = \mathbf{k}(0, y, z) + \epsilon x\frac{\partial \mathbf{k}}{\partial \bar{x}}(0, y, z) + O(\epsilon^2),$$

where the derivatives with respect to $\bar{x}$ cannot be large. Thus in any domain in x of $O(1/\epsilon)$ the paraxial solution must be a perturbation of a solution that is independent of x. These are precisely the solutions of Sec. III, which appear to be linear in x in any domain of $O(1/\epsilon)$ but which can be valid for much larger domains.

Two examples are now presented in which working with (37) and (38) leads to false conclusions because the underlying paraxial approximation is violated. The first is that a number of treatments of the paraxial equations assume an initial Gaussian profile and proceed either analytically or numerically to solve for the waveform. This necessarily violates the paraxial approximation since there are no Gaussian solutions to (23).

As a second example we examine the asymptotic analysis of Zakharov *et al.*[27] It turns out that what appears to be a reasonable criterion for determining from the initial data whether self-trapping occurs becomes so weak a statement as to be ineffective. Zakharov first defines two integrals of Eqs. (37), namely

$$I_1 = \int\int a^2\, dy\, dz$$

and

$$I_2 = \int\int [a_y^2 + a_z^2 + (\phi_y^2 + \phi_z^2)a^2 - \hat{f}(a^2)]\, dy\, dz.$$

The domain of integration is an initial plane at any fixed value of x, and $\hat{f}(\xi)$ has the same definition as in (24). These quantities are independent of x for all solutions of (36) but produce quantities of $O(\epsilon^2)$ that depend on x for solutions of (7), (8), (9) that satisfy the paraxial approximation. The criterion for a beam to be self-trapped is that it initially obey

$$I_1(a) > I_1(\bar{a}) \quad \text{and} \quad I_2(a) < 0,$$

where $\bar{a}$ is the solution to (23) of lowest energy and hence has no zeros for finite r. The beam is then self-trapped in the sense that the maximum amplitude of the beam is bounded below by the positive constant $-2I_2 I_1^{-1}$.

Testing this theory on the perturbation solutions of Sec. III, we first use nonlinearity (34) and the resulting solution (35), obtaining

$$I_1 = \left[1 + \frac{2\epsilon u(\bar{x})K}{K^2 - k_0^2}\left(\frac{1}{n} - 1\right)\right]\int_0^\infty \bar{a}^2 r\,dr,$$

$$I_2 = \left[1 + \frac{2\epsilon u(\bar{x})K}{n(K^2 - k_0^2)}\right]\beta\frac{n-1}{n+1}\int_0^\infty \bar{a}^{2n+2} r\,dr.$$

For all positive integers n, the negativity of I_2 is violated, and hence the criterion does not hold for this large class of beams. In the case of a nearly cubic medium,

$$f(a^2) = a^2 - \epsilon_1 a^4$$

where $\epsilon_1 \ll 1$, the object of Zakharov's study, we obtain

$$I_1 = \int_0^\infty \bar{a}^2 r\,dr + O(\epsilon,\epsilon_1),$$

$$I_2 = -\frac{\epsilon}{3}\int_0^\infty \bar{a}^6 r\,dr + \frac{\epsilon u(\bar{x})K}{K^2 - k_0^2}\int_0^\infty \bar{a}^4 r\,dr.$$

Now the criterion states that $\max a \geqslant O(\epsilon)$. Zakharov indicates that the beam is now self-trapped, but surely a beam that falls in amplitude from $O(1)$ to $O(\epsilon)$ in a distance of $O(1/\epsilon)$ is one that is dispersing! From these examples one might expect that strong statements about solutions to the paraxial equations may refer only to those that also violate the paraxial approximation.

In conclusion we have presented a perturbation expansion of a symmetric beam with straight rays both in two and three dimensions. We have criticized other researchers for solving the paraxial equations without verifying that the results obey the paraxial approximation. The perturbation solutions have been shown to be paraxial and thus to form the subclass of solutions of the paraxial equations that satisfy the paraxial approximation.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada Grant A8785 and by the United States Army under Contract No. DAAG29-75-C-0024.

[1]G. A. Askary'an, Zh. Eksp. Teor. Fiz. **42**, 1567 (1962) [Sov. Phys. JETP **15**, 1088 (1962)].
[2]V. I. Talanov, Zh. Eksp. Teor. Fiz. Pis. Red. **2**, 223 (1965) [JETP Lett. **2**, 138 (1965)].
[3]R. Y. Chiao, E. Garmire, and C. H. Townes, Phys. Rev. Lett. **13**, 479 (1964).
[4]P. L. Kelley, Phys. Rev. Lett. **15**, 1005 (1965).
[5]L. A. Vainstein, Zh. Tekh. Fiz. **34**, 193 (1964) [Sov. Phys. Tech. Phys. **9**, 157 (1964)].
[6]G. D. Malyuzhinets, Usp. Fiz. Nauk **69**, 321 (1959) [Sov. Phys. Usp. **2**, 749 (1959)].
[7]J. M. Marburger, Prog. Quantum Electron. **4**, 35 (1975).
[8]O. Svelto, Prog. Opt. **11**, 1 (1974).
[9]A. L. Dyshko, V. N. Lugovoi, and A. M. Prokhorov, Zh. Eskp. Teor. Fiz. **65**, 1367 (1973) [Sov. Phys. JETP **38**, 680 (1974)].
[10]A. L. Dyshko and V. N. Lugovoi, Opt. Acta **23**, 483 (1976).
[11]J. A. Fleck, J. Comput. Phys. **16**, 324 (1974).
[12]A. A. Kanashov and A. M. Rubenshik, Opt. Commun. **24**, 121 (1978).
[13]A. A. Kolokolov and A. I. Sukov, Opt. Spectrosc. **37**, 308 (1974).
[14]J. F. Lam, B. Lippman, and F. Tappert, Opt. Commun. **15**, 419 (1976).
[15]V. N. Lugovoi, Zh. Eksp. Teor. Fiz. **65**, 886 (1973) [Sov. Phys. JETP **38**, 439 (1974)].
[16]S. V. Manakov, Zh. Eksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1974)].
[17]F. P. Mattar, Appl. Phys. **17**, 53 (1978).
[18]A. F. Mastryukov and V. S. Synakh, Kvant. Elektron. (Moscow) **3**, 2473 (1976) [Sov. J. Quantum Elect. **6**, 1353 (1977)].
[19]N. G. Phu-Xuan and G. Rivoire, Apt. Acta **25**, 233 (1978).
[20]A. S. Reztsov and N. A. Tikhonov, Dokl. Akad. Nauk SSSR **231**, 324 (1976) [Sov. Phys. Dokl. **21**, 655 (1977)].
[21]N. N. Rozanov and V. A. Smirnov, Zh. Eksp. Teor. Fiz. **70**, 2060 (1976) [Sov. Phys. JETP **43**, 1075 (1976)].
[22]Y. R. Shen, Prog. Quantum Electron. **4**, 1 (1975).
[23]Y. R. Shen, Rev. Mod. Phys. **48**, 1 (1976).
[24]A. B. Shvartsburg, Zh. Eksp. Teor. Fiz. **70**, 1640 (1976) [Sov. Phys. JETP **43**, 855 (1976)].
[25]N. A. Rikhonov, Dokl. Akad. Nauk SSSR **231**, 592 (1976) [Sov. Phys. Dokl. **21**, 663 (1977)].
[26]V. E. Zakharov, Zh. Eksp. Teor. Fiz. **68**, 940 (1975) [Sov. Phys. JETP **41**, 465 (1975)].
[27]V. E. Zakharov, V. V. Sobolev, and V. C. Synakh, Zh. Eksp. Teor. Fiz. **60**, 136 (1971) [Sov. Phys. JETP **33**, 77 (1971)].
[28]D. I. Abakarov, A. A. Akopyan, and S. I. Pekar, Zh. Eksp. Teor. Fiz. **52**, 463 (1967) [Sov. Phys. JETP **25**, 303 (1967)].
[29]D. Pohl, Opt. Commun. **2**, 305 (1970).
[30]S. A. Akhmanov, A. P. Sukhorukov, and R. V. Khokhlov, Zh. Eksp. Teor. Fiz. **50**, 1537 (1966) [Sov. Phys. JETP **23**, 1025 (1966)].
[31]R. J. Finkelstein, R. LeLevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
[32]H. A. Haus, Appl. Phys. Lett. **8**, 128 (1966).

1- and 2-topology of reaction networks

Oktay Sinanoğlu
Departments of Chemistry and of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

(Received 2 December 1980; accepted for publication 13 February 1981)

The kinds of reaction networks introduced earlier by this writer are capable of diverse applications in chemical physics, biochemistry, chemical engineering, economics, ecological, and other dynamics. All possible mechanisms or pathways as a function of the numbers of reaction steps ρ or species σ are generated with them. They also give the rate laws, multiplicity of steady states, the nature of their dynamic instabilities, oscillations, etc. These properties are related to a large extent on the 1- and 2-topology of the networks, $\{\mathcal{N}\}$. The $\{\mathcal{N}\}$ are graphs of two kinds of lines and two kinds of vertices. They can be planar or nonplanar. The *genus g* and *thickness t* of any $\mathcal{N}$ are related to ρ, σ and the numbers of catalytic and autocatalytic cycles. The *Betti numbers* $\mathbb{B}^i(p)$ of 1- and 2-complexes constituted by $\mathcal{N}$ and other topological invariants of the networks under two kinds of homeomorphisms are given. A number of theorems are stated and proved. The above reaction networks are interesting mathematical objects in that they help classify coupled nonlinear differential equations.

PACS numbers: 82.20.Wt

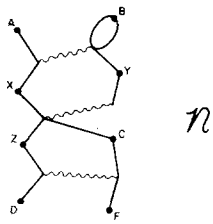
I. INTRODUCTION, MOTIVATION, AND DEFINITIONS

The kinds of networks we shall define below and study topologically should be useful in a number of fields from chemistry to economics, but we shall concentrate below on the case of chemical reaction networks. These are capable of diverse applications in chemistry, biochemistry, chemical engineering, and chemical physics. They also make the presentation more concrete.

A *mechanism* $\mathcal{M}$ or a synthetic *pathway* $\mathcal{P}$ is a set of coupled chemical reaction steps. Each $\mathcal{M}$ or $\mathcal{P}$ we represent[1] by a chemical reaction network $\mathcal{N}$; e.g.,

$$\left.\begin{array}{rcl} A+X & \rightarrow & Y+2B \\ Y & \rightarrow & Z+X+C \\ Z+D & \rightarrow & F+C \end{array}\right\} \mathcal{M}\,(\text{or } \mathcal{P})$$

$$\text{OVR:}\quad A+D \rightleftharpoons 2B+2C+F, \tag{1a}$$



$\mathcal{N}$ (1b)

Each step is stoichiometric; therefore so is the *overall reaction* OVR.

In a mechanism, the steps are *elementary reaction steps* implying *molecularity*. In most $\mathcal{M}$, therefore one usually has reaction steps which are uni- or bimolecular (uni–uni, uni–bi, bi–bi combinations), more than two-body collisions being much less probable. If all the steps of an $\mathcal{M}$ are only $\in$ {uni–uni, uni$\rightleftarrows$bi, bi$\rightarrow$bi}, we shall call this $\mathcal{M}$ a "*strict mechanism*," $\mathcal{M}^*$. Otherwise, it will be referred to with $\mathcal{M}_>$ (more than bimolecular steps included). In $\mathcal{P}$, the steps being procedural, there is no such restriction as bi or uni, etc., on reaction vertices.

For rate processes[1,2] $\mathcal{M}$ is more fundamental. For thermodynamic cycles[3] or for *a priori* synthesis design and for chemical yields, $\mathcal{P}$ can be used.[4]

A reaction network $\mathcal{N}$ is a graph of two kinds of lines and two kinds of vertices:

Lines of $\mathcal{N}$ are: (i) (~~~) = *rx-line* = reaction step line denoting any one of ($\rightarrow$) or ($\rightleftarrows$) or (=) in each step of $\mathcal{M}$ (or $\mathcal{P}$), Eq. (1a).

(ii) (—) = species-mole line = *sm-line*. *Each mole of each species* in $\mathcal{M}$ is one distinct line.

Vertices of $\mathcal{N}$ are: (a) (x) = (~~<) = reaction, *rx-vertex*, and (b) (●) = (<) = species, *s-vertex*. All the sm-lines in an $\mathcal{N}$ corresponding to different moles of the same species are joined to that species' s-vertex as in Eq. (1b).

An arbitrary $\mathcal{N}$ may be *planar* or *nonplanar*, but any $\mathcal{N}$ can be drawn in the three-dimensional Euclidean space $\mathscr{E}_3$ without crossings.

In this paper, we shall study the topological properties of arbitrary $\mathcal{N}$. (1) We shall find their topological invariants which also classify $\{\mathcal{N}\}$; (2) we shall obtain relations among the number of chemical species σ, the number of rx-steps ρ, the number of catalysts, intermediates, and autocatalysts or inhibitors in $\{\mathcal{M}\}$ (or reagents, products, by-products, etc., in $\mathcal{P}$) that can be accommodated in any general rx-network $\mathcal{N}$; (3) we shall see (e.g., in terms of ρ and σ) when $\{\mathcal{N}\}$ are planar and when nonplanar and the extent of their nonplanarity in terms of the *thickness t* (cf. below) of $\mathcal{N}$ and their covering graphs and the *genus g* of the closed surfaces in $\mathscr{E}_3$ on which an $\mathcal{N}$ can be drawn without crossings; (4) we shall relate the physical quantities like ρ and σ to the *Betti numbers* $\{\mathbb{B}^i_{(p)}\}$ of $p = 1$ *and 2-complexes*[5] defined on $\mathcal{N}$.

Our purpose in studying $\{\mathcal{N}\}$ is severalfold: (1) In the past there was no systematic way to generate $\{\mathcal{M}\}$ for various numbers σ of species (or of ρ) in a reacting mixture. Our methods[1] allowed such a systematics which could also be used for specified observed overall reaction (OVR) types.[2] Then also one can see which and how many $\{\mathcal{M}\}$ would be

 0022-2488/81/071504-09$1.00

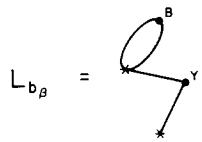
invoked to match an overall observed rate law. (2) For each $\mathcal{N}$, the rate equation can be written down. These are in general a set of coupled nonlinear differential equations. (3) For each $\mathcal{N}$, the existence and solutions for multiple steady states can be deduced [1a] using also other graphs (below) obtained from $\mathcal{N}$. (4) The nature of the steady states, the singular points of the system, the kinds of instabilities,[6] the existence of chemical oscillations,[7,8] and possibly chaotic behavior[9] can be deduced [1a] to a large extent from $\{\mathcal{N}\}$ and 1- and 2-topology. This program was already stated and carried out to a considerable extent for (1)–(3) and part of (4) by 1973–74.[1,10] At that time, however, we had found it convenient to separate the $\{\mathcal{N}\}$ into what we called "laminar" and "turbulent" networks[2] and to deal with these separately. The present set of papers by the writer, treat arbitrary, general $\{\mathcal{N}\}$ without any restrictions, with more efficient methodology, and with many new results. (5) Finally, the kinds of $\{\mathcal{N}\}$ defined above, which have two kinds of lines and two kinds of vertices, and their topology would be useful in other fields like economics, ecology, biophysics, etc. These $\{\mathcal{N}\}$ are mathematically interesting objects. They help classify coupled nonlinear differential equations.

II. GENERAL NETWORK: ITS LINEBLOCKS, ITS SKELETON, AND TOPOLOGICAL INVARIANCE

All the sm-lines of a general $\mathcal{N}$, such as Eq. (1b), that are directly connected to each other constitute a *lineblock* L_{b_i} of $\mathcal{N}$.

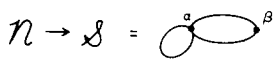
Removal of the rx-lines from an $\mathcal{N}$ produces several L_{b_i}.

In Fig. 1(b) there are two $L_{bi}, i\in \alpha,\beta$, e.g.,



(2)

The *skeleton*[1,2,10] $\mathcal{S}$ of $\mathcal{N}$ is obtained by compressing each L_{b_i} of $\mathcal{N}$ into a dot-point. Denoting the (~~~) rx-lines with thick lines (▬▬), we get from Fig. 1(b),

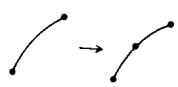


(3)

The skeleton is a graph of one kind of lines (ρ of them) and one kind of vertices (γ of them). It can be a *simple graph*, a *multigraph*, or a *pseudograph* as in Eq. (3).

We shall investigate first the 1-topology of $\{\mathcal{S}\}$ and $\{\mathcal{N}\}$ under line *subdivisions*.

On $\mathcal{S}$, an elementary subdivision is simply



(4a)

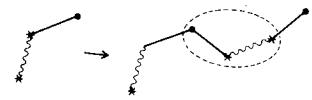
which makes

$$\rho\to\rho+1 \quad \text{and} \quad \gamma\to\gamma+1, \tag{4b}$$

Such subdivisions on $\mathcal{S}$ are homeomorphisms. Properties of $\{\mathcal{S}\}$ thereby unchanged are topological 1-invariants of $\mathcal{S}$.

In $\mathcal{N}$, subdivisions of sm-lines (—) and of (~~~~) rx-lines can be made. However, the line subdivisions which preserve the definition of $\mathcal{N}$ (and its physical relevance) are the following.

(A) *sm-line subdivision on* $\mathcal{N}$



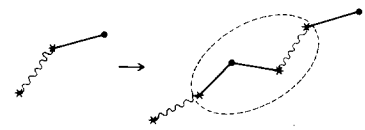
(5a)

which makes

$$\sigma\to\sigma+1, \quad n\to n+2, \quad \rho\to\rho+1, \tag{5b}$$

$\sigma\equiv$no. of s-vertices, $n\equiv$no. of sm-lines, $\rho\equiv$no. of rx-vertices, in $\mathcal{N}$.

(B) *rx-line subdivision of* $\mathcal{N}$



(6a)

$$\sigma\to\sigma+1, \quad n\to n+2, \quad \rho\to\rho+1. \tag{6b}$$

(A) and (B) are 1-homeomorphisms on $\mathcal{N}$. We consider the topological 1-invariants of $\{\mathcal{N}\}$ with respect to (A) and (B).

Comment: There are no subdivision homeomorphisms of an individual lineblock which keep it a lineblock.

III. THE 1-TOPOLOGY OF SKELETONS

Arbitrary $\mathcal{S}$ (ρ lines, γ vertices) is "three-dimensional," i.e., it can be drawn in $\mathcal{E}_3$ without spurious line crossings.

For *any* $\mathcal{S}$,

$$\gamma=\rho-r+1 \tag{7}$$

with $r\equiv$no. of "empty rings"[1] of $\mathcal{S}$. (Our own proof for any $\mathcal{S}$ is in Ref. 1.) More precisely, r is the maximum number of linearly independent rings [mod 2], hence the *rank* of he *ring-incidence* matrix $\mathbf{n}_r^1$ (rows labelled by the rings, columns labelled by the lines of $\mathcal{S}$).

Viewing $\mathcal{S}$ as a "1-*complex*"[10] (made of 1-simplexes and 0-simplexes)

$$r=\mathbb{B}^1(1) \tag{8}$$

the $i=1, p=1$-*Betti number* of $\mathcal{S}$ (cf. Sec. VIII).

Physically r is closely related to the number of generalized catalysts possible in any $\mathcal{M}$ of ρ-steps.[1]

Theorem S1: For a given number ρ of rx-steps, all possible $\{\mathcal{S}\}_\rho$ have

$$r\in\{0,1,2,\ldots,\rho\}. \tag{9}$$

Proof follows from $\gamma\geqslant 1$.

Theorem S2: The r of $\mathcal{S}$ is subdivision-invariant [proof from Eq. (4b) put in Eq. (7)].

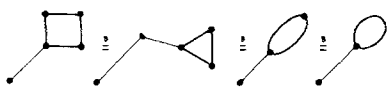
Note that subdivision homeomorphism ($\overset{s}{=}$) is an equivalence relation. Thus:

Theorem S3: (a) Any $\mathscr{S}$ (ρ any integer) belongs to one and only one equivalence class C^s_μ under $\overset{s}{=}$),

$$\{\mathscr{S}\} = \underset{\mu}{\cup} C^s_\mu \tag{10}$$

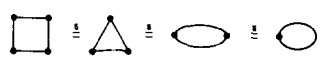
(b) If $\mathscr{S}_1$ and $\mathscr{S}_2 \in C^s_\nu$, then $r_1 = r_2$. (c) But if $r_1 = r_2$, $\mathscr{S}_1$ and $\mathscr{S}_2$ may not be subdivision homeomorphs.

Example:



(11a)

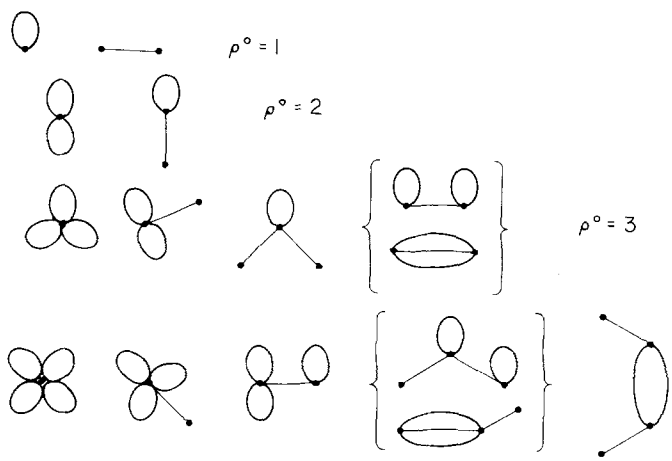
and



(11b)

but

(11c)



and so on for each larger ρ°'s.

Each of the $\mathscr{S}_0$ above is a "smallest" (ρ minimum) $\mathscr{S}$, since no further subdivision vertices can be removed from such $\mathscr{S}_0$. In the integer partitions of 2°_ρ, $k^\circ_1 = 2$'s are missing (except $\rho^\circ = 1$) as a 2 would be either a subdivision or a disconnected loop (◯).

The $\{k^\circ_1\}$ (and the $r = \hat{r}$) are invariant under subdivisions as is easily seen (and Eq. (7)).

Comment: Note that all the degrees $\{k^\circ_1\}$ and/or $\rho^\circ, r, \gamma^\circ$ still do not uniquely specify the "smallest" (i.e., $\rho_0 \leqslant \rho$) member of each C^s_μ class. In Eq. (14), note, e.g., the $(3+3)$, $r = 2$, $\rho^\circ = 3$ degeneracy, also the $(3+4+1)$, $r = 2, \rho^\circ = 4$ degeneracy.

Theorem S4B: A more complete set of labels for the s-equivalence classes $\{\mathscr{S}_0 \overset{h}{\rightarrow} \{\mathscr{S}\}\}$ is the set which consists of (a) the degrees $\{k^*_i\}$ in the *simple graph* $\hat{\mathscr{S}}_0$ (≡ the "*basic simple graph*" $\hat{\mathscr{S}}$ for each $\mathscr{S}_0$) obtained from $\mathscr{S}_0$ by removing all the loops and "extra" multilines from $\mathscr{S}_0$, (b) the numbers $\{r^*_i\}$ of *loops* (◯) at each corresponding vertex of $\mathscr{S}_0$, and (c) the numbers $\{e^*_{ij}\}$ of extra-multilines between the ith and jth (necessarily adjacent) vertices of $\mathscr{S}_0$.

Thus the value of r is not suffcient to label an s-equivalence class. Additional s-invariants are needed to specify C^s_μ uniquely.

Further (though still incomplete) s-invariants are provided by:

Theorem S4A: (a) The set of degrees ("k-star values"[1]) $\{k^\circ_1, k^\circ_2, \cdots\}$ such that

$$k^\circ_1 + k^\circ_2 + \cdots = 2^\circ_\rho, \quad \rho^\circ \in \{1,2,3,\cdots\}, \tag{12}$$

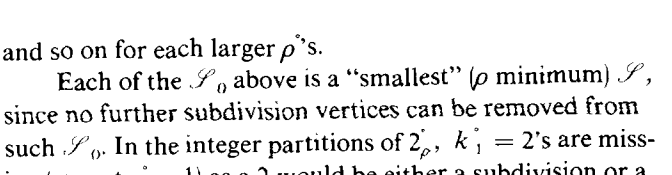
and such that $\{k^\circ_i\}$ are the vertex degrees of a "smallest" ($\rho \geqslant \rho^\circ$) connected s-homeomorph, is a set of s-invariants. (Higher homeomorphs of $\mathscr{S}_0$ contain vertices of these degrees as well as some new ones of degree 2.)

Proof: The possible $\{k^\circ_1\}$ sets for the above are

$2\rho^\circ$	$(k^\circ_1 + k^\circ_2 \cdots) = 2\rho^\circ$
2	$(2+0)$, $(1+1)$
4	$(4+0)$, $(3+1)$
6	$(6+0)$, $(5+1)$, $(4+1+1)$, $(3+3)$
8	$(8+0)$, $(7+1)$, $(5+3)$, $(4+3+1)$, $(3+3+1+1)$
⋮	⋯

(13)

with the corresponding $\{\mathscr{S}_0\}$:



(14)

Proof: Each $\rho^0 = \rho_{\min}$, $\mathscr{S}_0 \overset{h}{\leftarrow}$ "smallest" s-homeomorph [e.g., in Eq. (14)] is a simple graph $\hat{\mathscr{S}}_0$ to which some (one-line) loops have been added at its vertices $\{i\}$ and to which some occasional extra multilines have been added between two adjacent vertices i and j. (Otherwise there is a

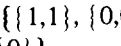
removable vertex of degree 2 and the minimum ρ° and γ° have not been reached). E.g., Eqs. (14) give only the *basic simple graphs* $\{\hat{\mathscr{S}}_0\}$,

(14')

The degree k_i^* of the ith vertex of the basic simple graph plus 2 for each loop at i and plus 1 for each extra multiline at i gives the k_i° of $\mathscr{S}_0$:

$$k_i^{\circ} = k_i^* + 2r_i^* + \sum_j e_{ij}^*$$

(Q.E.D.)

Remark 1: The $\{\{k_i^*\}, \{r_i^*\}, \{e_{ij}^*\}\}$ set distinguishes, e.g., between the "smallest protopyte" homeomorphs $\{\rho^{\circ}=3,\ r^{\circ}=2\}$:  . For the spindle, the numbers are $\{\{1,1\}, \{0,0\}, \{2\}\}$, while for the other $\mathscr{S}_0$, $\{\{1,1\}, \{1,1\}, \{0\}\}$.

Remark 2: (a) Let $\mathscr{S}_0$ have a total number of r_0 independent rings. This r_0 which is one of the s-invariants of $\{\mathscr{S}_0 \overset{k^{\circ}}{\longrightarrow} \mathscr{S}\}$ equals

$$r_0 = \hat{r}_0 + \sum_i r_i^* + \sum_i e_{i,i+1}.$$

The remaining number of rings, i.e., $\hat{r}_0$ is the ring number s-invariant of $\hat{\mathscr{S}}_0$. It is the number of independent rings which are not shrinkable to a loop (), nor to a spindle (). (b) Reinterpreting the loop as a ring attached to $\hat{\mathscr{S}}_0$ at a single vertex (of degree k_i^*) and a spindle as a ring attched to $\hat{\mathscr{S}}_0$ at two vertices (of k_i^* and k_j^*), we see that the rings of $\{\mathscr{S}\}$ are classified into three and more kinds of rings. The number of each type of ring in $\mathscr{S}$ is an s-invariant.

IV. WHEN IS A SKELETON NONPLANAR?

Lemma: If an $\mathscr{S}$ with ρ and r, has l loops and c multiple extra lines we can remove them without affecting the planarity. The resulting *simple graph* $\hat{S}$ has

$$\hat{r} = r - l - c, \quad \hat{\rho} = \rho - l - c, \quad \hat{\gamma} = \gamma. \tag{15}$$

Proof: Loop can be drawn as small as needed; multilines can be drawn as closely as needed (Q.E.D.). For Eq. (15), cf. (7). Each multiline adds a ring, as does a loop.

Theorem S5: (a) As long as rx-steps in an $\mathscr{M}$ are $\rho < 9$, the skeletons are planar (b) If $\rho = \hat{\rho} + l + c \geqslant 9 + l + c$ and if $\hat{\mathscr{S}}$ has at least six vertices $\{i\}$ of k-star value (degree) $k_i \geqslant 3$, then some $\mathscr{S}$ can be nonplanar.

Proof: By Kuratowski's theorem $\hat{\mathscr{S}}$ is nonplanar if it contains an s-homeomorph of K_5 or $K_{3,3}$. But $K_{3,3}$ has $\rho = 9$ and six vertices with $k_i = 3$. The K_5 has $\rho = 10$ and five vertices with $k_i = 4$.

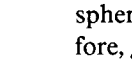
V. MAPPING A SKELETON ON A g-TORUS

Mapping $\mathscr{S}$ on a plane or on a sphere ($g = 0$) are equivalent.[11] If $\mathscr{S}$ is nonplanar, it can be mapped on the surface of a g-torus (sphere with g handles, $g \geqslant 1$). For a given $\mathscr{S}$ (ρ,r,l,c), how do we predict the minimum g, the *genus* of $\mathscr{S}$? (The g we shall need, e.g., in predicting the dynamic stability of $\mathscr{M}$.)

The simplest nonplanar $\hat{\mathscr{S}}$ is the $K_{3,3}$:

$K_{3,3}$ (16)

All except the last dotted line () can be drawn on a sphere (or plane). The remaining line goes on a handle; therefore, $g = 1$.

An equality giving $g(\rho,\gamma)$ is not known even for *simple graphs* (no loops ( ) or multilines (), ···) except[12] for K_γ. However, several, theorems (Heawood, Ringel and Youngs, Euler, ···)[13] combined may be used to estimate the $\bar{g}$ of a connected *simple graph* $\bar{G}(\bar{\rho}\bar{\gamma},\bar{r})$ if $\bar{\gamma} \geqslant 3$,

$$\text{int}_> \left\{\frac{\bar{\rho}}{6} - \frac{(\bar{\gamma}-2)}{2}\right\} \leqslant \bar{g} \leqslant \text{int}_> \left\{\frac{(\bar{\gamma}-3)(\bar{\gamma}-4)}{12}\right\} \tag{17}$$

(if ρ and $\bar{r}$ are specified, use $\bar{\gamma} = \bar{\rho} - \bar{r} + 1$) $\text{int}_>\{\ \}$ means larger integer.

Further,[13] if $\bar{\gamma} \geqslant 3$ and

$$\bar{\rho} \geqslant \bar{\gamma}(\bar{\gamma} - 1)/2 - 5, \tag{18}$$

then Eq. (17) yields either the correct $\bar{g}$ or $\bar{g} + 1$.

However, our skeletons $\{\mathscr{S}(\rho,r)\}$ *are not simple graphs*; therefore, Eqs. (17) and (18) must be extended to graphs $\mathscr{S}$ with loops and multilines. This is done below (Theorem S6).

Were g known, $\mathscr{S}$ could be drawn on a g-torus and its *faces* (f of them) (2-simplexes) defined in terms of the *Jordan curve theorem*.[5] Then the second Euler equation holds:

$$\gamma = \rho - f + (2 - 2g). \tag{19}$$

(This holds for *nonsimple* $\mathscr{S}$ *as well*, since every loop and every multiline can be drawn so as to give a face. Thus $f_l = r_l$, and $f_c = r_c$). One should also recall that Eq. (19) holds only for $\mathscr{S}$ on the surface of a g-torus (planar $\mathscr{S}$ for $g = 0$), while Eq. (7) is for any three-dimensional $\mathscr{S}$ of even unknown g.

From each nonsimple $\mathscr{S}$ (ρ,r,g) of l loops and c extra multilines, we can get two simple graphs $\hat{\mathscr{S}}$ $(\hat{\rho},\hat{r},\hat{g})$ and $\hat{\hat{\mathscr{S}}}(\hat{\hat{\rho}},\hat{\hat{r}},\hat{\hat{g}})$ such that

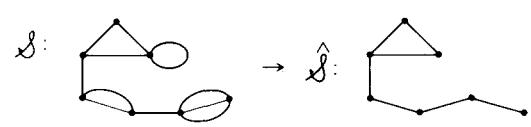
$$\hat{g} = g = \hat{\hat{g}} \tag{20}$$

Definition:

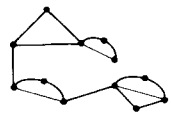
$$\hat{\mathscr{S}} \equiv \mathscr{S} - \{\text{loops,multilines}\}, \tag{21}$$

$$\hat{\hat{\mathscr{S}}} \equiv \mathscr{S} + \{\text{vertices and lines from subdivisions of all loops and multilines}\}. \tag{22}$$

Example:



(21x)

$\mathscr{S} \rightarrow \hat{\hat{\mathscr{S}}}$: 

(22x)

Lemma: $\mathscr{S}$, $\hat{\mathscr{S}}$, and $\hat{\hat{\mathscr{S}}}$ have the same genus.

$\mathscr{S} \rightarrow \hat{\mathscr{S}}$

$$\begin{cases} \gamma \rightarrow \hat{\gamma}, \\ \rho \rightarrow \hat{\rho} = \rho - l - c, \\ r \rightarrow \hat{r} = r - l - c, \\ f \rightarrow \hat{f} = f - l - c, \end{cases} \tag{23}$$

and

$\mathscr{S} \rightarrow \hat{\hat{\mathscr{S}}}$,

$$\begin{cases} \gamma \rightarrow \hat{\hat{\gamma}} = \gamma + 2l + c, \\ \rho \rightarrow \hat{\hat{\rho}} = \rho + 2l + c, \\ r \rightarrow \hat{\hat{r}}, \\ f \rightarrow \hat{\hat{f}}. \end{cases} \tag{24}$$

Proof follows from Eqs. (18), (23), and (24), or from a useful relation we obtain by combining Eqs. (7) and (18),

$$r - f + 1 = 2g. \tag{25}$$

Comment: Note that $\mathscr{S}$ and $\hat{\hat{\mathscr{S}}}$ are s-homeomorphs,

$\mathscr{S} \overset{s}{\sim} \hat{\hat{\mathscr{S}}}$, but $\mathscr{S}$ and $\hat{\mathscr{S}}$ *are not*,

$$\mathscr{S} \overset{s}{\neq} \hat{\mathscr{S}}. \tag{26}$$

Yet Eq. (20) still holds, as easily seen from Eqs. (25) and (23).

Theorem S6: The genus g of any $\mathscr{S}$ (nonsimple graph, etc.) can be estimated using Eqs. (17) and (18) for $\hat{\mathscr{S}}$ and/or $\hat{\hat{\mathscr{S}}}$ with Eqs. (23) and (24).

Proof follows from $\hat{\mathscr{S}}$ and $\hat{\hat{\mathscr{S}}}$ being simple graphs and from $g = \hat{g} = \hat{\hat{g}}$.

Comment: The bounds obtained from $\hat{\hat{\mathscr{S}}}$ are closer to g than the bounds from $\hat{\mathscr{S}}$.

Theorem S7: The *thickness* t (cf., e.g., Ref. 13) of any $\mathscr{S}$ (including *nonsimple* ones) with $\gamma \geqslant 3$ satisfies

$$t(\mathscr{S}) = t(\hat{\mathscr{S}}) = t(\hat{\hat{\mathscr{S}}}). \tag{27}$$

Thus $t(\mathscr{S})$, the minimum number of independent planes superimposed on which $\mathscr{S}$ can be drawn without crossings, can be estimated from

$$t(\mathscr{S}) \geqslant \text{int}_{>} \{(\rho - l - c)/(3\gamma - 6)\} \tag{28}$$

or

$$t(\mathscr{S}) \geqslant \text{int}_{>} \{(\rho + 2l + c)/[3(\rho + 2l + c) - 6]\}. \tag{29}$$

Proof follows from the inequality for simple graphs[13] and Eqs. (23), (24), (27).

[Note also that it appears that often $t = g + 1$, but this is in general not true. For $g \geqslant 1$, more generally $2 \leqslant t \leqslant (g + 1)$ (e.g., if one plane contains lines going on different handles of a sphere with g handles).]

VI. TOPOLOGY OF A GENERAL LINEBLOCK

Lemma L 1: Any lineblock L_{b_i} of a general reaction network $\mathscr{N}$ is a bipartite graph by itself.

Proof: Each sm-line of an L_{b_i} has as its boundary one species s-vertex, and one rx-vertex. Also a single L_{b_i} is connected.

Note: L_{b_i} can have multilines, but no loops.

Lemma L 2: If the rx-lines of an $\mathscr{N}$ are removed, there results a disconnected graph of γ bipartite components.

Proof: $\#\{L_{b_i}\}$ of $\mathscr{N} = \#\{$vertices of its $\mathscr{S}\} = \gamma$.

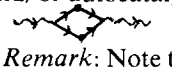
[*Example*: see, e.g., Eq. (1b), then (3).]

Theorem L1: If a vertex i (dot-point)[1] of $\mathscr{S}$ of an $\mathscr{N}$ has double weight[14] (ω_i, n_i), and k-star value[1] (degree) k_i, the corresponding L_{b_i} of $\mathscr{N}$ displays ω_i s-vertices, n_i sm-lines, k_i rx-vertices, and R_i sm-rings[14] such that

$$\boxed{\omega_i + k_i = n_i - R_i + 1.} \tag{30}$$

Further $0 \leqslant R_i \leqslant (n_i - 1)$.

Proof follows from Lemma L1 and the analog of Eq. (7), noting that the total number of vertices of L_{b_i} of either kind is $(\omega_i + k_i)$. Also $(\omega_i + k_i) \geqslant 2$, so that $R_i \leqslant (n_i - 1)$.

[In the chemical context,[14] R_i is the number of homolytic and/or autocatalytic and/or "heterolytic" (e.g., ) rings in L_{b_i}.]

Remark: Note that R_i is *not* invariant under sm-subdivisions defined by Eq. (5a). Neither are there any other kinds of line subdivisions which preserve the connectedness[15] and the *bipartite nature* of an individual L_{b_i}.

Theorem L2: (a) Any lineblock L_{b_i} of $\mathscr{N}$, by itself, is always planar ($g_i = 0$) if $\mathscr{N}$ corresponds to a "*strict mechanism*" $\mathscr{M}^*$ [recall[1] that a "*strict mechanism*" $\mathscr{M}^*$ has only rx-steps of the uni–uni, uni–bi, or bi–bi molecular types (Sec. I)]. (b) If $\mathscr{N}$ is that of a "synthetic pathway" $\mathscr{P}$, or a nonstrict mechanism $\mathscr{M}$, then an L_{b_i} in that $\mathscr{N}$ can be nonplanar only if $n_i \geqslant 9$, $k_i \geqslant 3$, $\omega_i \geqslant 3$, *and* each rx-vertex in L_{b_i} is of sm-degree $\geqslant 3$, *and* each s-vertex in L_{b_i} is of degree $\geqslant 3$.

Proof: (a) If L_{b_i} is that of "strict mechanism" $\mathscr{M}^*$ ($\sim \mathscr{N}$), then, by the definitions in Sec. I (and in Ref. 1) which distinguish an $\mathscr{M}^*$ from a $\mathscr{P}$, each rx-vertex of L_{b_i} is of sm-degree $\in \{1,2\}$ only. Hence L_{b_i} cannot contain $K_{3,3}$. Therefore, L_{b_i} cannot be nonplanar. (b) L_{b_i} is bipartite; so is the nonplanarity criterion graph $K_{3,3}$. But $K_{3,3}$ has three vertices of each of the two kinds, each vertex of degree 3, and it has nine lines (Q.E.D.).

Note: Even if each L_{b_i} of an $\mathscr{N}$ ($\sim \mathscr{M}^*$) were planar, the full $\mathscr{N}$ ($\sim \mathscr{M}^*$) could be nonplanar, through the nonplanarity of its $\mathscr{S}$ according to Theorem S5.

Genus of a pathway lineblock: The genus g_i of an L_{b_i} contained in an $\mathscr{N}$ ($\sim \mathscr{P}$) can be estimated from the analog of Eqs. (17) with the substitutions:

$$\left\{\begin{matrix} \bar{\rho} \rightarrow (n_i - l_i^b) \\ \bar{\gamma} \rightarrow (\omega_i + k_i) \end{matrix}\right\}. \tag{31}$$

l_i^b = no. of multilines in L_{b_i} (one has $0 \leqslant l_i^b \leqslant R_i$).

[The g_i of an $L_{b_i} \subset \mathscr{N}(\sim \mathscr{M}^*) = 0$ as stated in Theorem L2a.]

Thickness of a pathway lineblock: A lower bound to the t_i of $L_{b_i} \subset \mathscr{N}$ ($\sim \mathscr{P}$) is obtained from the analog of Eq. (28) (for $L_{b_i} \rightarrow \hat{L}_{b_i}$):

$$t(L_{b_i}) \geqslant \text{int}_{>} \{(n_i - l_i^b)/3(\omega_i + k_i) - 6\} \tag{32}$$

[The $t_i = 1$ of an $L_{b_i} \subset \mathcal{N}$ ($\sim \mathcal{M}^*$) since these L_{b_i} are planar.]

VII. TOPOLOGY OF THE FULL NETWORK

When the rx-vertices of the $\{L_{b_i}\}$ are connected by ρ rx-lines $\{\sim\sim\sim\}$ according to the $\mathcal{S}$, we get the full network $\mathcal{N}$.

Theorem N1: Any network $\mathcal{N}$ representing some $\mathcal{M}$ or $\mathcal{P}$ with ρ rx-steps, σ chemical species, n moles of all species (n = absolute sum of all stoichiometric coefficients appearing in all the steps of $\mathcal{M}$ or $\mathcal{P}$; thus $n > \sigma$) satisfies the relation

$$\boxed{\rho + \sigma = n - (r + R) + 1.} \tag{33}$$

$R = \Sigma_{i\geqslant 1}^{\gamma} R_i$, the sum of the *sm-ring numbers* of all the $L_{b_i} \subset \mathcal{N}$. The r is the ring number (may be called also the "*rx-ring number*") of the $\mathcal{S}$ of $\mathcal{N}$ as before.

Proof: Sum Eq. (30) over all the γ lineblocks $\{L_{b_i}\} \subset \mathcal{N}$. We have

$$\sigma = \sum_{i\geqslant 1}^{\gamma} \omega_i, \quad 2\rho = \sum_{i\geqslant 1}^{\gamma} k_i, \quad \sum_{i\geqslant 1}^{\gamma} R_i = R,$$

$$\sum_{i}^{\gamma} n_i = n, \quad \sum_{i}^{\gamma} 1 = \gamma, \quad \text{and} \quad \gamma = \rho - r + 1. \quad \text{(Q.E.D.)}$$

Corollary 1/*N* 1: If the values n', ρ', σ' are specified (e.g., physically), any such $\mathcal{N}(n', \rho', \sigma')$ will have an R value in the range

$$\boxed{1 + (n' - \sigma') - 2\rho' \leqslant R \leqslant 1 + (n' - \sigma') - \rho'.} \tag{34}$$

Proof from Eq. (33) and for any $\mathcal{S}$; hence its $\mathcal{N}$, $0 \leqslant r \leqslant \rho$.

Corollary 2/*N* 1: If the $\mathcal{S}$ (ρ'',r'') is specified, any $\mathcal{N}$ having this particular $\mathcal{S}$ will have "*excess moles*" [$\equiv (n - \sigma)$] such that

$$\boxed{(n - \sigma) \geqslant \rho'' + r'' - 1.} \tag{35}$$

Proof from Eq. (33) and $R \geqslant 0$; also in any $\mathcal{N}$, $n > \sigma$ because each L_{b_i} is connected.

The $(r + R) \equiv$ composite ring number of $\mathcal{N}$ is the $i = 1$ *Betti number* ($p = 1$) of $\mathcal{N}$ disregarding the distinction between sm-lines and rx-lines:

$$(r + R) = \mathbb{B}^1 (p = 1). \tag{36}$$

(cf. also Sec. VIII).

A. Topological 1-invariants of $\mathcal{N}$

Theorem N2: Under sm-line subdivisions and/or rx-line subdivisions defined by Eqs. (5) and (6, the sum $(r + R)$ remains invariant. [We say $(r + R)$ is an "(*sm and rx*)-*invariant*".]

Proof: Subsitution of Eq. (5b) and/or (6b) into Eq. (33) yields the same $(r + R)$.

Remark: (a) Under sm-subdivisions on $\mathcal{N}$, r, and R may each change, but their sum does not. (b) Under rx-subdivisions on $\mathcal{N}$, not only $(r + R)$, but also r and R individually are invariant, because rx-subdivisions alone, are also s-subdivisions of the $\mathcal{S}$ of $\mathcal{N}$ preserving r. [Examples of (a) and (b) occurred in Ref. 14.]

Theorem N3: The degrees of both the species vertices $\{\bullet\}$ and the rx-vertices $\{\times\}$ of $\mathcal{N}$ are *sm- and rx-invariants*.

(For proof note that initial vertices are not affected by sm- and rx-subdivisions although new vertices of degree two arise.)

Corollary 1/*N* 3: We classified general networks before[14] with respect to the types of species vertices ("source", "sink", "internal," etc.)[14] the $\{\mathcal{N}\}$ contain. These network types[14] [(1) "*strictly laminar*," (2) "*laminar*," (3) "*internally turbulent*," (4*a*) "*autocatalytic*," (4*b*) "*self-inhibitory*"] are *sm- and rx-invariant* (since the vertex types depend on their degree).

Corollary 2/*N* 3: Likewise, if $\mathcal{N}$ represents a *strict mechanism*, $\mathcal{M} = \mathcal{M}^*$ (i.e., all of the rx-vertices of $\mathcal{M}^*$ are of the types: $\sim\!\!<$ and $\sim\!\!-$ only), or if $\mathcal{N}$ represents a pathway $\mathcal{P}$, or more general mechanism $\mathcal{M}$, (i.e., rx-vertices like $\sim\!\!<$ and more are allowed), then this nature of $\mathcal{N}$ (i.e., whether $\mathcal{N} \sim \mathcal{M}^*$ or $\mathcal{N} \sim \mathcal{P}$ or $\mathcal{M}$) is sm-/rx-invariant. This means successively larger $\mathcal{N}$'s generated will remain of the same type of $\mathcal{M}$ (or $\mathcal{P}$) (this is used in the mechanism generation methods of our Ref. 14).

B. Nonplanarity of $\mathcal{N}$

Theorem N4: The genus $g_{\mathcal{N}}$ of any general network is given by

$$\sigma + \rho - n + f_{\mathcal{N}} = 2 - 2g_{\mathcal{N}}. \tag{37}$$

Proof: Disregard the distinction between the two kinds of lines, and the two kinds of vertices in $\mathcal{N}$. Total no. of vertices of any kind $= \sigma + 2\rho$; total no. of lines of any kind $= \rho + n$(*Q.E.D.*).

Note: The "faces" of $\mathcal{N}$, total no. $f_{\mathcal{N}}$, are bounded by either all sm-lines ($\equiv$"*sm-faces*") or by some sm- and some rx-lines ($\equiv$"*sm/rx faces*").

$$f_{\mathcal{N}} = f_{\mathcal{N}}^{(\mathrm{sm})} + f_{\mathcal{N}}^{(\mathrm{sm/rx})}. \tag{38}$$

We now investigate the relation between the genus $g_{\mathcal{N}}$ of the full $\mathcal{N}$ (strictly of $\langle\mathcal{N}\rangle$ discussed in Sec. VIII) and the $\{g_i\}$ of individual lineblocks $\{L_{b_i}\}$ and the $g_{\mathcal{S}}$ of the skeleton of $\mathcal{N}$. The g_i of each L_{b_i} is defined by the mapping of that L_{b_i} on a separate closed surface (g_i-torus). The $\mathcal{S}$ is also mapped on a separate closed surface, a $g_{\mathcal{S}}$-torus in defining $g_{\mathcal{S}}$.

Lemma N 1: In combining the $\{L_{b_i}\}$ according to $\mathcal{S}$ into an $\mathcal{N}$, we first map all of the $\{L_{b_i}\}$ which are as yet disconnected by rx-lines onto a single closed surface. The genus g_L of this surface is

$$g_L = \max\{g_1, g_2, \cdots, g_\gamma\} \equiv g_i^{\max}. \tag{39}$$

Proof: Any $L_{b_i}(g_i)$ can be mapped on any surface with $g \geqslant g_i$. Hence all L_{b_i} can be mapped onto the surface with largest g_i.

Theorem N4: The number of faces and genus of $\mathcal{N}$ is related to those of $\mathcal{S}$ and of $\cup_L \{L_{b_i}\}$ by

$$\boxed{(f_{\mathcal{N}} - 2g_{\mathcal{N}}) = -[(f_L + f_{\mathcal{S}} - 1) - 2(g_L + g_{\mathcal{S}})].} \tag{40}$$

Proof: For the union of disconnected $\{L_{b_i}\}$ on the surface L, we have [c.f. Eq. (25) which holds for a disconnected graph on a single surface as well as for a connected one, even though Eq. (7) is modified to $\gamma = \rho - r + m$ for m disconnected components]

$$f_L = R + 1 - 2g_L; \tag{41a}$$

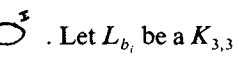
similarly for $f_{\mathscr{S}}$ with r, and for $\mathscr{N}$

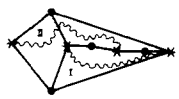
$$f_{\mathscr{N}} = (r + R) + 1 - 2g_{\mathscr{N}}. \tag{41b}$$

Hence,

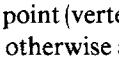
$$\boxed{f_{\mathscr{N}} - (f_L + f_{\mathscr{S}} - 1) = -2[g_{\mathscr{N}} - (g_L + g_{\mathscr{S}})].} \tag{40'}$$

Lemma N 2: Even if all of its L_{b_i} are planar and its $\mathscr{S}$ is planar, the $\mathscr{N}$ may not be planar.

Proof: Take, e.g., and $\mathscr{S} =$  . Let L_{b_i} be a $K_{3,3}$ sm-graph with its bridge line missing. The first rx-line added [Eq. (42)] keeps the graph planar, but the second makes it nonplanar (Q.E.D.)

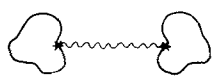


(42)

Theorem N5: If each L_{b_i} of $\mathscr{N}$ has at most one sm-ring, $\{R_i \leqslant 1\}$, and each dot point (vertex) of the $\mathscr{S}$ has at most one loop (), but otherwise any number r of rings, then

$$\boxed{\begin{array}{l} g_{\mathscr{N}} = g_{\mathscr{S}}, \\ (\{R_i \leqslant 1\}, \quad \mathscr{S}:\{l_i \leqslant 1 \text{ loop per vertex}\}). \end{array}} \tag{43}$$

Proof: Any $R_i \leqslant 1$, L_{b_i} is planar; also any of its rx-vertices is accessible by an rx-line without crossing any closed-Jordan sm-curves [Eq. (44a)] or closed Jordan sm/rx-curves [Eq. (44b)].



(44a)

(44b)

Theorem N6: For a general $\mathscr{N}$ with its $\mathscr{S}$ and $\{L_{b_i}\}$,

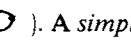
$$\boxed{g_{\mathscr{N}} \geqslant \max\{g_L, g_{\mathscr{S}}\}} \tag{45}$$

(the larger of g_L or $g_{\mathscr{S}}$).

Proof: Equality occurs when rx-lines added to $\cup_L \{L_{b_i}\}$ do not cross any sm or sm/rx rings of any L_{b_i}. Otherwise, $g_{\mathscr{N}}$ may increase.

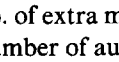
Theorem N7:

$$g_{\mathscr{N}} \leqslant \text{int}_> \{(\sigma + 2\rho - 3)(\sigma + 2\rho - 4)/12\}. \tag{46}$$

Proof: $\mathscr{N}$ viewed as a graph of $\sigma + 2\rho$ vertices of either kind (s-vertices and rx-vertices), and sm- and rx-lines distinction ignored, has no 1-line loops, only some sm-multilines (). A *simple graph* $[\hat{\mathscr{N}}]_m$ is obtained removing such multilines, with the same genus and number of vertices. Then apply the upper bound equation [as in Eq. (17)]. (Q.E.D.)

Remark: The upper bound $= \{g_{\mathscr{N}}$ or $g_{\mathscr{N}} + 1\}$ if

$$n + \rho - c_{\text{sm}} \geqslant (\sigma + 2\rho)(\sigma + 2\rho - 1)/2 - 5. \tag{47}$$

c_{sm} = no. of extra multi-sm-lines in $\mathscr{N}$ () (related to the number of autocatalytic or homolytic steps).

Removal of c_{sm} gives $\mathscr{N} \to [\hat{\mathscr{N}}]$, and so Eq. (18) can be applied.

VIII. EXTENSION TO 2-COMPLEXES AND THE BETTI NUMBERS IN P = 1 AND 2

The network $\mathscr{N}$ and its skeleton $\mathscr{S}$ treated as graphs are 1-complexes. Then the Betti numbers are $\mathbb{B}^0$ and $\mathbb{B}^1$.

Every graph can be mapped onto a closed orientable 2-surface embedded in $\mathscr{E}_3$. This surface is a g-torus, $g \in \{0,1,2,3,\cdots\}$.

On the surface, the union of the graph G with its faces is the 2-complex $\langle G \rangle$ on G. The Betti numbers, in $p = 1$, are

$$\mathbb{B}^0, \mathbb{B}^1 \quad \text{of} \quad G \tag{48}$$

and those of $\langle G \rangle$ are for $p = 2$. The latter

$$\overline{\mathbb{B}}^0, \overline{\mathbb{B}}^1, \overline{\mathbb{B}}^2 \quad \text{of} \quad \langle G \rangle \tag{49}$$

are such that in general $\overline{\mathbb{B}}^i \neq \mathbb{B}^i$ for the same i.

$\{\mathbb{B}^0, \mathbb{B}^1\}$ are invariants ("s-") in dim = 1, i.e., under line subdivisions on G.

$\{\overline{\mathbb{B}}^0, \overline{\mathbb{B}}^1, \overline{\mathbb{B}}^2\}$ are invariants ("t-") dim = 2, under face subdivisions (hence under "triangulations") on $\langle G \rangle$ as well as under line subdivisions. The $\mathbb{B}^0, \mathbb{B}^1$ are not t-invariants, while Eqs. (49) are.

$\{\overline{\mathbb{B}}^i\}$ are related to the *incidence matrices* $\{\overline{\mathbf{I}}_i\}$ for the 2-complex by

$$\boxed{\begin{array}{l} \overline{\mathbb{B}}^0 = \overline{\alpha}_0 - \text{rank } \overline{\mathbf{I}}_0, \\ \overline{\mathbb{B}}^1 = \overline{\alpha}_1 - \text{rank } \overline{\mathbf{I}}_1 - \text{rank } \overline{\mathbf{I}}_0, \\ \overline{\mathbb{B}}^2 = \overline{\alpha}_2 - \text{rank } \overline{\mathbf{I}}_2 - \text{rank } \overline{\mathbf{I}}_1, \end{array}} \tag{50}$$

where $\overline{\alpha}_i = \#\{i\text{-simplexes}\}$.

$$\overline{\mathbf{I}}_i = (i\text{-simplex to } (i+1)\text{-simplex incidence matrix}). \tag{50'}$$

For a p-complex, rank $\overline{\mathbf{I}}_p = 0$. Note also that Eqs. (5) satisfy the Euler–Poincaré relation

$$\overline{\mathbb{B}}^0 - \overline{\mathbb{B}}^1 + \overline{\mathbb{B}}^2 = \overline{\alpha}_0 - \overline{\alpha}_1 + \overline{\alpha}_2. \tag{51}$$

A. The 2-complex $\langle \mathscr{S} \rangle$ on the skeleton $\mathscr{S}$

First for $\mathscr{S}$ itself as 1-complex

$$\mathscr{S}: \quad \mathbb{B}^0 = \gamma - \text{rank } \mathbf{I}_0, \quad \mathbb{B}^1 = \rho - \text{rank } \mathbf{I}_0. \tag{52}$$

As $\mathbf{I}_0$ is the usual vertex→line incidence matrix ($\mathcal{S}$ is connected, so $m = 1$)

$$\text{rank } \mathbf{I}_0 = \gamma - 1 \tag{53}$$

(or rank $\mathbf{I}_0 = \rho - r$).

Thus

$$\mathbb{B}^0 = 1 \text{ (for } \mathcal{S},\ m = 1) \text{ and } \mathbb{B}^1 = r \tag{54}$$

which verifies the s-invariance of r. (Further for disconnected G of m-pieces, $\mathbb{B}^0 = m$.)

Next, including the faces of $\mathcal{S}$ on a g-torus

$$\mathcal{S} \rightarrow \langle \mathcal{S} \rangle. \tag{55}$$

As the g-torus is an orientable surface, Eqs. (50), etc. are for mod. 2 or for any directed $\mathcal{S}$ obtained from $\mathcal{S}$ by arbitrary assignments of arrows to each line.

Then,

$\langle \mathcal{S} \rangle$:

$$\bar{\mathbb{B}}^0 = 1 \ (m = 1), \tag{56a}$$

$$\bar{\mathbb{B}}^1 = r - \text{rank } \bar{\mathbf{I}}_1, \tag{56b}$$

$$\mathbb{B}^2 = f - \text{rank } \bar{\mathbf{I}}_1. \tag{56c}$$

Note the $\bar{\mathbb{B}}^1 - \mathbb{B}^1 = -\text{rank } \bar{\mathbf{I}}_1 \neq 0$.

$\bar{\mathbf{I}}_1$ is the edge→face (1→2) incidence matrix.

Lemma: For $\langle \mathcal{S} \rangle$ (of $\mathcal{S}$ with $m = 1$)

$$\text{rank } \bar{\mathbf{I}}_1 = f - 1 = \rho - r_f, \tag{57}$$

where r_f = maximum no. of linearly independent 2-rings in $\langle \mathcal{S} \rangle$ [a ring made of faces ("umbrellas") going around a vertex of $\langle \mathcal{S} \rangle$].

Proof: Draw the "face-graph" $\mathcal{S}_f$ of $\langle \mathcal{S} \rangle$ by

$$\begin{cases} \text{face of } \langle \mathcal{S} \rangle \rightarrow \text{vertex of } \mathcal{S}_f, \\ \text{adjacent faces in } \langle \mathcal{S} \rangle \rightarrow \text{connected vertices in } \mathcal{S}_f. \end{cases}$$

Then one has $f = \gamma_f$; $\rho = \rho_f$ and can use Eq. (53) (Q.E.D.).

Theorem: For the 2-complex $\langle \mathcal{S} \rangle$ on any $\mathcal{S}$ (which is connected; $m = 1$),

$$\boxed{\begin{aligned} \bar{\mathbb{B}}^0 &= 1 \quad (m = 1), \\ \bar{\mathbb{B}}^1 &= r - f + 1, \\ \bar{\mathbb{B}}^2 &= 1. \end{aligned}} \tag{58}$$

Proof: Eqs. (56) and (57).

Corollary 1: For $\langle \mathcal{S} \rangle$ of $\mathcal{S}$

$$\boxed{\bar{\mathbb{B}}^1 = 2g.} \tag{58'}$$

Proof: Eqs. (58) and (25).

Theorem: Given two skeletons $\mathcal{S}$ $(\rho, r, \cdots)$ and $\mathcal{S}'$ $(\rho', r', \cdots)$, if their extension 2-complexes $\langle \mathcal{S} \rangle_{\rho,r,g,\ldots}$ and $\langle \mathcal{S} \rangle_{\rho',r',g',\ldots}$ are homeomorphic, then $g = g'$ and $r = r'$.

Proof: For 2-complexes homeomorphism can be studied by fs-subdivisions (i.e., face-subdivisions and line-subdivisions, or by triangulations). As can be seen from Eqs. (58) under any such subdivision, g remains invariant. But in addition to $\{\bar{\mathbb{B}}^i,\ i = 0,1,2\}$ remaining invariant, so must the $\{\mathbb{B}^i\}$ in the lower dimension under s-subdivisions only. Thus $r = r'$ (Q.E.D.).

Remark 1: If $\langle \mathcal{S} \rangle$ and $\langle \mathbb{B}' \rangle$ are 2-homeomorphic, then $\mathcal{S}$ and $\mathcal{S}'$ are 1-homeomorphic (proof omitted). Note that this is *not* a corollary of the above theorem, because:

Remark 2: The reverse of the above theorem is not always true; i.e., if $g = g'$ and $r = r'$ 2-homeomorphism is not guaranteed. (Same with $\{\mathcal{S}\}$ and $r = r'$ only.)

"Facial rings" ("2-rings") and the genus of $\mathcal{S}$: The maximum number of linearly independent (mod 2) "*umbrellas*" of $\langle \mathcal{S} \rangle$ ("*facial rings*," or "2-*rings*") is r_f, in Eq. (57). From Eqs. (57), (56b), and (58'), we have

$$\boxed{r_f = \rho - f + 1} \tag{59}$$

or

$$\boxed{r_f = 2g + \gamma - 1.} \tag{60}$$

Equation (60) is a very useful, practical relation, in calculating the genus or the facial ring rank, r_f. Thus, e.g., for planar graphs ($g = 0$), $r_f = \gamma - 1$. Those $\{\langle \mathcal{S} \rangle\}$ with larger genus will have the greater number r_f of independent 2-rings, given the same number of vertices.

Another relation useful in giving insight into the genus is, from Eqs. (60) and (7),

$$g_{\mathcal{S}} \equiv g_{\langle \mathcal{S} \rangle} = \boxed{g = \tfrac{1}{2}[(r + r_f) - \rho].} \tag{61}$$

B. The 2-complex $\langle \mathcal{N} \rangle$ on the network $\mathcal{N}$

If we do not distinguish between the two kinds of lines (*sm*-: ——× and *rx*: ~~~×) and the two kinds of vertices (s-: ≺ and rx-: ~~~≺) in $\mathcal{N}$ and $\langle \mathcal{N} \rangle$, we get the Betti numbers (and $r_f(\langle \mathcal{N} \rangle)$, $g_{\mathcal{N}}$ etc., directly (i.e., from the "*graph of* $\mathcal{N}$" $\equiv {}^{G}\mathcal{N}$ thereby defined, and its $\langle {}^{G}\mathcal{N} \rangle$).

For *the* 1-*complex* ($\mathcal{N} \sim {}^{G}\mathcal{N}$):

$$\begin{aligned} \mathbb{B}^0_{\mathcal{N}} &= 1 \quad (m = 1), \\ \mathbb{B}^1_{\mathcal{N}} &= (r + R) \end{aligned} \tag{62}$$

[R defined in Eq. (33)].

For *the* 2-*complex* ($\langle \mathcal{N} \rangle \sim \langle {}^{G}\mathcal{N} \rangle$):

$$\begin{aligned} \bar{\mathbb{B}}^0_{\mathcal{N}} &= 1 \quad (m = 1), \\ \bar{\mathbb{B}}^1_{\mathcal{N}} &= (r + R) - f_{\mathcal{N}} + 1, \\ \bar{\mathbb{B}}^2_{\mathcal{N}} &= 1 \end{aligned} \tag{63}$$

($f_{\mathcal{N}} \equiv f_{G_{\mathcal{N}}}$ as in previous sections). Also

$$\bar{\mathbb{B}}^1_{\mathcal{N}} = 2g_{\mathcal{N}} \tag{63'}$$

(other methods of calculating or estimating $g_{\mathcal{N}}$ were given in Sec. VII).

The independent "facial rings" of $\langle \mathcal{N} \rangle$, now can be *sm-faces* (bounded by sm-lines only) or (sm/rx)-faces (bounded by sm- and rx-lines). The total number is $f_{\mathcal{N}}$. The maximal independent 2-rings (regardless of vertex tip type (s- or rx-) of the "umbrella") number

$$r_f^{\,l} = 2g_{\,l} + (\sigma + 2\rho) - 1 \tag{64}$$

or

$$r_f^{\,l} = (n + \rho) - f_{\,l} + 1. \tag{64'}$$

Also,

$$2g_{\,l} = (r + R + r_f^{\,l}) - (\rho + n). \tag{65}$$

C. The 1- and 2-homeomorphisms on $\langle \mathcal{N} \rangle$

Going back from ${}^{G}\mathcal{N} \to \mathcal{N}$, the only thing to note is that not all 1- and 2-subdivisions preserve the definition (and physical meaning) of $\mathcal{N}$, although they are homeomorphisms on ${}^{G}\mathcal{N}$.

Only the invariance, in corresponding dimensions, of the Betti numbers, Eqs. (62) and (63), under a subgroup of 1- and 2-homeomorphisms on ${}^{G}\mathcal{N}$ which preserve the definition of $\mathcal{N}$ (the bipartiteness of the lineblocks, etc.), is needed (and relevant).

The elements of this subgroup are:

(i) "*sm-line subdivisions*," Eq. (5), on $\mathcal{N}$,

(ii) "*rx-line subdivisions*," Eq. (6), on $\mathcal{N}$,

(iii) *sm-face subdivisions by sm-lines* on $\langle \mathcal{N} \rangle$ (and $\langle \mathcal{L} \rangle$),

(iv) *sm-face subdivisions by sm/rx-composite lines* (•—∿∿—•) on $\langle \mathcal{N} \rangle$,

(v) *(sm/rx)-face subdivisions by sm-lines* on $\langle \mathcal{N} \rangle$,

(vi) *(sm/rx)-face subdivisions by sm/rx-composite lines* on $\langle \mathcal{N} \rangle$.

Calling $\{(i), (ii)\} \equiv$ "$\mathcal{N}$*-subdivisions*" and $\{iii, \cdots vi\} \equiv \langle \mathcal{N} \rangle$*-subdivisions*, we have the following:

Theorem: The definition (and physical meaning) of $\mathcal{N}$ and Eqs. (62) are preserved under $\mathcal{N}$-subdivisions. The definition of $\mathcal{N}$ (and of $\langle \mathcal{N} \rangle$) and Eqs. (63′) are preserved under $\langle \mathcal{N} \rangle$-subdivisions.

Proof: These constitute subgroups of the arbitrary 1- and 2-homeomorphism groups of ${}^{G}\mathcal{N}$ and $\langle {}^{G}\mathcal{N} \rangle$. Further any of the $\mathbb{B}$ in Eqs. (62), (63) $= {}^{G}\mathbb{B}$ (Q.E.D.)

[1]O. Sinanoğlu, (a) in *C.A. Coulson Memorial Session, Fifth Canadian Symp. Theo. Chem.* (June 1974); also Boğaziçi Evrenkenti Lectures (1973); (b) J. Am. Chem. Soc. **97**, 2309 (1975); (c) Nippon Kagaku Kaishi **9**, 1358 (1976).

[2]O. Sinanoğlu and L. S. Lee, Theor. Chim. Acta, (a) **48**, 287 (1978); (b) **51**, 1 (1979).

[3]O. Sinanoğlu, "A Priori Design of Thermochemical Cycles...," *Third International Congress of Quant. Chem. Kyoto (Japan) Abstracts* (Nov. 1979). 1979).

[4]Each step i of a $\mathscr{P}$ could be opened up into a mechanism $\mathscr{M}_i$.

[5]See, e.g., W. S. Massey, *Algebraic Topology: An Introduction* (Harcourt, Brace World, New York, 1967).

[6]For a review and other references see: H. Haken, *Synergetics—An Introduction; Nonequilibrium Phase Transitions and Self-Organization in Chemistry, Physics, and Biology* (Springer, New York, 1977).

[7]G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley-Interscience, New York, 1977) and many references therein.

[8]A. T. Winfree, see, e.g., *Physics Today* **28**, 34 (1975) and his other references. (I thank Prof. Arthur Winfree for stimulating me into the study of nonplanar networks further.)

[9]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[10]The deduction of steady state solutions from graphs obtained from I', and the kinds of instabilities they would have, were carried out by graph methods for linear regions around steady states by this writer (O.S.) and L. S. Lee and reported in the Ph.D. thesis of L. S. Lee with O.S. (Chemistry Department, Yale University, 1973–74), but the last two parts are not published otherwise, although discussed briefly in Refs. 1a,b.

[11]S. Lefschetz, *Introduction to Topology* (Princeton U. P., Princeton, N.J., 1949).

[12]G. Ringel and J. W. T. Youngs, Proc. Natl. Acad. Sci. (USA) **60**, 438 (1968).

[13]See, e.g., R. J. Trudeau, *Dots and Lines* (The Kent State Univ. Press, Kent, Ohio, 1976), p. 156.

[14]O. Sinanoğlu, "Chemical Mechanisms, Pathways, Cycles and Their Networks—Systematic and Topological Generation of Autocatalytic and other General Networks" (in press).

[15]For an excellent summary of some additional aspects of graph theory, see J. W. Essam and M. E. Fisher, Rev. Mod. Phys. **42**, 271 (1970).

Inverse scattering—exact solution of the Gel'fand-Levitan equation

Kay R. Pechenick and Jeffrey M. Cohen
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 14 January 1981; accepted for publication 20 February 1981)

One formulation of inverse scattering theory involves the Gel'fand–Levitan equation. We present a procedure for finding exact solutions to this equation; this procedure can be applied whenever the reflection coefficient is a rational function of the wave number k, with an arbitrary number of poles. We present graphs of the potential as a function of distance, for several cases with 3, 4, 5, and 6 poles. Prior to this paper, no 4-, 5-, or 6-pole case had ever been treated successfully.

PACS numbers: 94.20. − y, 94.20.Tt, 03.80. + r

I. INTRODUCTION

In the last 40 years, most work on ionospheric structure determination has used an approximate method based on the Abel integral equation. From scattering data it is possible to compute the ionospheric electron density using this method. Using the same data, it is also possible to use a full-wave method to obtain the ionospheric electron density with greatly reduced errors. Since the data used is identical for the approximate and full wave theories, there is no need to modify experimental equipment. The difference in treatment is essentially computational, since the same data gives the position-dependent electron density using the approximate and full wave theories.

In principle, the full wave inverse scattering method is exact. However, in practice, approximate analytic or numerical methods are normally employed to solve the Gel'fand-Levitan equation. To circumvent the possibility of roundoff errors, numerical instabilities, etc., in solving the Gel'fand-Levitan equation numerically, we have solved the equation exactly, using a generalization of Kay's[1] procedure for rational fraction reflection coefficients. Previous attempts along these lines have given usable results when the number of poles is not too large (3 poles—Ahn and Jordan[2]; 1 pole—Moses[3]; 3 poles—Moses[4]; 1, 2, and 3 poles—Pechenick and Cohen[5]). As the number of poles increases, the complexity of the analytic expressions increases rapidly. For more than 5 poles, quintic and higher-order polynomial equations must be solved, which is difficult in view of Galois' proof that the general 5th and higher-order equations are insoluble by radicals. Using our procedure, these difficulties are circumvented because we evaluate our analytic solutions by computer, thereby avoiding numerical instabilities which may arise in numerical solutions of integral and differential equations, and avoiding errors which often arise when analytic expressions are evaluated by hand. The higher-order polynomial equations are solved numerically, thus evading Galois' theorem. In this way we can treat an arbitrarily large number of poles. To check our procedure, we have repeated calculations carried out analytically by others[2–4] and have obtained agreement. We have also checked our computer codes against analytic solutions which we have obtained.[5] After our computer codes were shown to be operating properly, we treated a large number of cases. A number of these are described in this communication.

II. SOLUTION OF EQUATION

The determination of the ionospheric electron density can be reduced[6,7] to the problem of solving the Gel'fand-Levitan[8] equation, which is

$$R(x+t)+K(x,t)+\int_{-\infty}^{x} K(x,z)R(z+t)\,dz=0. \quad (1)$$

Now, if $R(x)$ is zero for $x<0$, we can write

$$R(x)=R_1(x)\theta(x) \quad (2)$$

for some function $R_1(x)$, where the step function $\theta(x)=0$ for $x<0$, and 1 otherwise. Substituting (2) into (1), we find that $K(x,t)$ has the form

$$K(x,t)=K_1(x,t)\theta(x+t), \quad (3)$$

where $K_1(x,t)$ satisfies the equation

$$R_1(x+t)+K_1(x,t)+\int_{-t}^{x} K_1(x,z)R_1(z+t)\,dz=0, \quad (4)$$

for $x+t\geqslant 0$. [We shall actually solve (4) for *all* x and t.]

Now let

$$R(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}r(k)\,dk, \quad (5)$$

and let $r(k)$ be a rational function[1,9] of k

$$r(k)=N(k)/D(k), \quad (6)$$

where $N(k)$ and $D(k)$ are polynomials in k, and $D(k)$ has the form

$$D(k)=\prod_{i=1}^{n}(k-k_i), \quad (7)$$

where the k_i are distinct complex numbers in the lower half-plane. Then we find that

$$R(x)=-i\sum_{j=1}^{n} e^{-ik_jx}C_j\theta(x), \quad (8)$$

where

$$C_j=N(k_j)\prod_{\substack{i=1\\ i\neq j}}^{n}\frac{1}{k_j-k_i}. \quad (9)$$

C_j is the residue of $r(k)$ at $k=k_j$; it is a function of the k_i.

To solve the Gel'fand-Levitan equation, we assume that

$$K_1(x,t)=\sum_{\alpha} f_\alpha(x)e^{a_\alpha t}, \quad (10)$$

 0022-2488/81/071513-04$1.00

where the summation is over a finite but as yet unspecified number of values of α and the $f_\alpha(x)$ are unknown functions of x. We also have

$$R_1(x) = -i\sum_{j=1}^{n} e^{-ik_j x} C_j, \tag{11}$$

from (2) and (8). Substituting (10) and (11) into (4), we obtain

$$-i\sum_{j=1}^{n} e^{ik_j(x+t)} C_j + \sum_\alpha f_\alpha(x) e^{a_\alpha t} - i\int_{-t}^{x} \left[\sum_\alpha f_\alpha(x) e^{a_\alpha z} \sum_{j=1}^{n} e^{-ik_j(z+t)} C_j\right] dz = 0. \tag{12}$$

The integral in (12) is equal to

$$\sum_\alpha \sum_{j=1}^{n} C_j f_\alpha(x) \frac{1}{a_\alpha - ik_j} \left[e^{a_\alpha x - ik_j(x+t)} - e^{-a_\alpha t}\right]. \tag{13}$$

Substituting the expression (13) for the integral into (12), we obtain an equation which can be written

$$A + B = 0, \tag{14}$$

where

$$A \equiv \sum_\alpha f_\alpha(x) e^{a_\alpha t} + i\sum_\alpha \sum_{j=1}^{n} C_j f_\alpha(x) \frac{1}{a_\alpha - ik_j} e^{-a_\alpha t}, \tag{15}$$

and

$$B \equiv -i\left[\sum_{j=1}^{n} e^{-ik_j(x+t)} C_j + \sum_\alpha \sum_{j=1}^{n} C_j f_\alpha(x) \frac{1}{a_\alpha - ik_j} e^{a_\alpha x - ik_j(x+t)}\right]. \tag{16}$$

We shall show that it is possible to set both $A = 0$ and $B = 0$ simultaneously, and thus find a solution to (12).

If we let

$$a_{-\alpha} = -a_\alpha \tag{17}$$

for all α, then the equation $A = 0$ may be rewritten

$$\sum_\alpha \left[f_\alpha(x) - i\sum_{j=1}^{n} C_j f_{-\alpha}(x) \frac{1}{a_\alpha + ik_j}\right] e^{a_\alpha t} = 0. \tag{18}$$

To solve (18), we let

$$f_\alpha(x) - i\sum_{j=1}^{n} C_j f_{-\alpha}(x) \frac{1}{a_\alpha + ik_j} = 0, \tag{19}$$

for each α. Replacing α by $-\alpha$ in (19) and using (17), we obtain

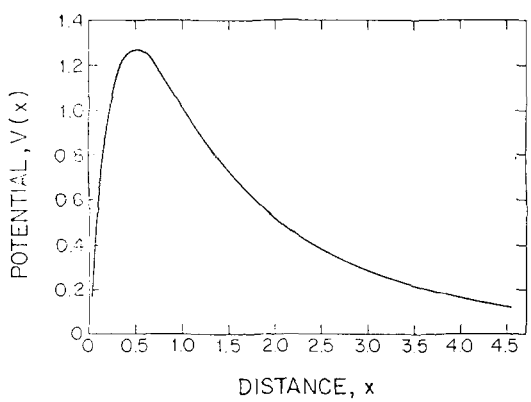


FIG. 1. Potential $V(x)$ vs distance x for 3 poles: $k_{1,2} = \pm 0.8 - 0.499i$, $k_3 = -i$.

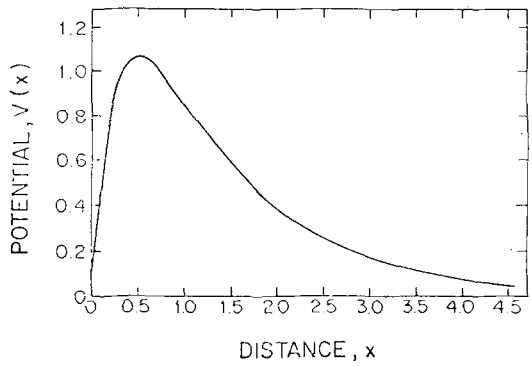


FIG. 2. Potential $V(x)$ vs distance x for 3 poles: $k_{1,2} = \pm 0.7 - 0.4999i$, $k_3 = -i$.

$$f_{-\alpha}(x) + i\sum_{j=1}^{n} C_j f_\alpha(x) \frac{1}{a_\alpha - ik_j} = 0. \tag{20}$$

To find a nontrivial solution to the homogeneous set of equations (20), we express it in the matrix form

$$\begin{pmatrix} 1 & -i\sum_{j=1}^{n} C_j \frac{1}{a_\alpha + ik_j} \\ i\sum_{j=1}^{n} C_j \frac{1}{a_\alpha - ik_j} & 1 \end{pmatrix} \begin{pmatrix} f_\alpha(x) \\ f_{-\alpha}(x) \end{pmatrix} = 0, \tag{21}$$

and set the determinant of the matrix equal to zero. This condition may be written as

$$\left(\sum_{j=1}^{n} \frac{C_j}{a_\alpha - ik_j}\right)\left(\sum_{l=1}^{n} \frac{C_l}{a_\alpha + ik_l}\right) = 1. \tag{22}$$

Equation (22) determines the possible values of a_α. Because of (17), our procedure will work only if the negative of every possible value of a_α is also a possible value of a_α.

By using some elementary results of complex variable theory, we can rewrite Eq. (22). C_j is the residue of $r(k)$ at $k = k_j$, and $C_j/(a_\alpha - ik_j)$ is the residue of $r(k)/(a_\alpha - ik)$ at $k = k_j$, where k is considered to be a complex variable. Therefore

$$\sum_{j=1}^{n} \frac{C_j}{a_\alpha - ik_j},$$

is the sum of all residues of $r(k)/(a_\alpha - ik)$, except for the one at $k = -ia_\alpha$. Similarly,

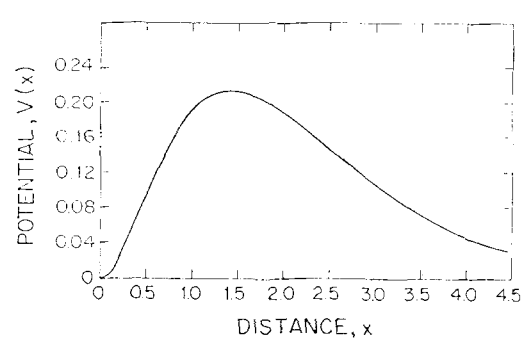


FIG. 3. Potential $V(x)$ vs distance x for 4 poles: $k_{1,2} = \pm 0.2403 - 0.3666i$, $k_{3,4} = \pm 0.3571 - 0.6019i$.

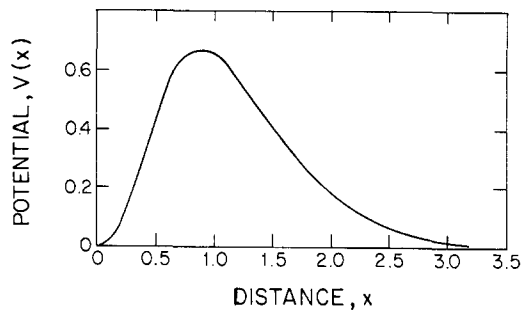


FIG. 4. Potential $V(x)$ vs distance x for 5 poles: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$, $k_4 = -1.4i$, $k_5 = -1.5i$.

$$\sum_{l=1}^{n} \frac{C_l}{a_\alpha + ik_l},$$

is the sum of all residues of $r(k)/(a_\alpha + ik)$, except for the one at $k = ia_\alpha$. But in each case, the sum of *all* residues is zero, since the integral of $r(k)/(a_\alpha \pm ik)$ around a closed contour at infinity is zero. Thus,

$$\sum_{j=1}^{n} \frac{C_j}{a_\alpha - ik_j},$$

is minus the residue of $r(k)/(a_\alpha - ik)$ at $-ia_\alpha$, and

$$\sum_{l=1}^{n} \frac{C_l}{a_\alpha + ik_l},$$

is minus the residue of $r(k)/(a_\alpha + ik)$ at ia_α. Thus,

$$\sum_{j=1}^{n} \frac{C_j}{a_\alpha - ik_j} = -ir(-ia_\alpha), \tag{23}$$

and

$$\sum_{l=1}^{n} \frac{C_l}{a_\alpha + ik_l} = ir(ia_\alpha), \tag{24}$$

so Eq. (22) becomes

$$r(ia_\alpha)r(-ia_\alpha) = 1. \tag{25}$$

From (25) it is clear that the negative of every possible value of a_α is also a possible value of a_α.

Now, $r(k)$ is required to have the properties[10,11] that

$$r(0) = -1, \tag{26}$$

and

$$r(-k) = r(k)^*. \tag{27}$$

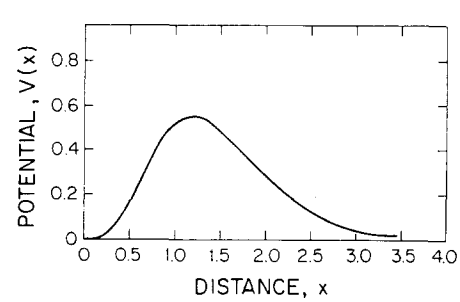


FIG. 5. Potential $V(x)$ vs distance x for 6 poles: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$, $k_4 = -1.4i$, $k_5 = -1.5i$, $k_6 = -1.6i$.

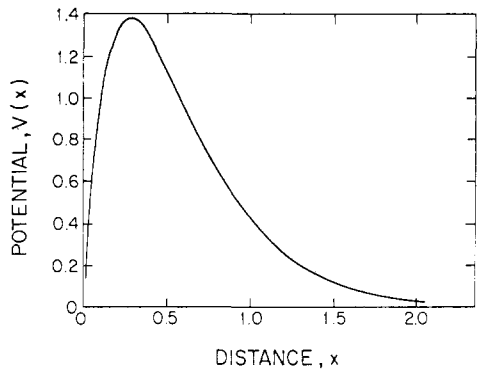


FIG. 6. Potential $V(x)$ vs distance x for 3 poles: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$.

Because of (26), $a_\alpha = 0$ is a solution to (25). It can also be shown that (25) is equivalent to a polynomial equation of degree n for ${a_\alpha}^2$. Since one root is zero, we only need to be able to solve an equation of degree $n-1$, for the case in which $r(k)$ has n poles. We solve this equation numerically on the computer.

Following our procedure, we let the a_α be the roots of Eq. (25); we check that $r(k)$ is such that the n possible values of ${a_\alpha}^2$ are all distinct. We let α take on the integer values from 1 to n and from -1 to $-n$, with $a_{-\alpha} = -a_\alpha$. For all other values of α, $f_\alpha(x)$ must be zero, so we may neglect all other values of α. Thus, there are $2n$ nonzero functions $f_\alpha(x)$. According to Eq. (20), the $f_\alpha(x)$ with positive α are related to those with negative α by the equation

$$f_{-\alpha}(x) = \left(-i\sum_{j=1}^{n} \frac{C_j}{a_\alpha - ik_j}\right) f_\alpha(x), \tag{28}$$

which is valid for both positive and negative values of α.

We now turn to the equation $B = 0$, which reduces to n equations, one for each value of j: $j = 1,2,\ldots,n$; and in which the summation over α includes both positive and negative values of α. The equations are

$$1 + \sum_{\alpha} \frac{1}{a_\alpha - ik_j} e^{a_\alpha x} f_\alpha(x) = 0. \tag{29}$$

By substituting (28) into (29), we obtain

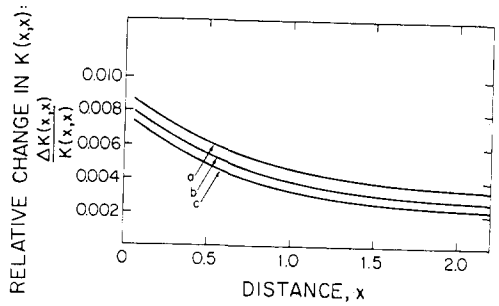


FIG. 7. Relative change in $K(x,x)$, vs distance x for three perturbations of a 3-pole case: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$. Perturbation (a): $k_1 \to -1.11i$, k_2 and k_3 unchanged. Perturbation (b): $k_2 \to -1.21i$, k_1 and k_3 unchanged. Perturbation (c): $k_3 \to -1.31i$, k_1 and k_2 unchanged.

$$1 + \sum_{\alpha=1}^{n} \left(F_{j\alpha} e^{a_\alpha x} + G_{j\alpha} e^{-a_\alpha x} \right) f_\alpha(x) = 0, \tag{30}$$

$j = 1,2,...,n$, where

$$F_{j\alpha} \equiv \frac{1}{a_\alpha - ik_j}, \tag{31}$$

$$G_{j\alpha} \equiv \frac{i}{a_\alpha + ik_j} \sum_{l=1}^{n} \frac{C_l}{a_\alpha - ik_l} = \frac{r(-ia_\alpha)}{a_\alpha + ik_j}. \tag{32}$$

Equation (30) is a set of n simultaneous linear equations for the $f_\alpha(x)$; we solve these equations numerically for a set of values of x.

Now,

$$K(x,t) = \sum_\alpha f_\alpha(x) e^{a_\alpha t} \theta(x+t), \tag{33}$$

from (3) and (10), so

$$K(x,x) = \sum_\alpha f_\alpha(x) e^{a_\alpha x} \theta(x), \tag{34}$$

where in (33) and (34) we are summing over positive and negative α. Substituting (28) into (34) we obtain

$$K(x,x) = \left\{ \sum_{\alpha=1}^{n} \left[e^{a_\alpha x} - r(-ia_\alpha) e^{-a_\alpha x} \right] f_\alpha(x) \right\} \theta(x). \tag{35}$$

Following our procedure, we substitute into (35) the numerical results for the $f_\alpha(x)$ obtained by solving (30).

We have applied our procedure to several rational function reflection coefficients $r(k)$. In all of them, the numerator $N(k)$ in (6) is a constant. Because of Eq. (26), $N(k)$ must be the constant $-(-k_1)(-k_2)\cdots(-k_n)$, so that $r(k)$ is completely determined by its poles, $k_i, i = 1,2,...,n$. Because of Eq. (27), the negative of the complex conjugate of every pole is also a pole. Thus, for each value of i, either k_i is pure imaginary or there is some j such that $k_j = -k_i^*$. There is another condition which $r(k)$ must satisfy: $|r(k)| \leqslant 1$ for all real k. We have checked that all of our $r(k)$ satisfy this condition. We have plotted graphs of the potential $V = (d/dx)K(x,x)$ as a function of x.

III. DISCUSSION

In this communication we have presented an exact class of inverse scattering solutions. Although in principle 5 or fewer poles can be treated analytically by radicals, no previous authors have treated more than 3 poles. This is because of the necessary complexity of the calculations. Here we present results, not only for 3 poles, but also for 4, 5, and 6 poles.

Figure 1 shows V versus x for the 3-pole case in which $k_1 = 0.8 - 0.499i$, $k_2 = -0.8 - 0.499i$, $k_3 = -i$, while Fig. 2 depicts the 3-pole case with $k_1 = 0.7 - 0.4999i$, $k_2 = -0.7 - 0.4999i$, $k_3 = -i$. These are the two cases which were studied by Ahn and Jordan,[2] and our results agree with theirs.

Figure 3 shows V versus x for a 4-pole case: $k_{1,2} = \pm 0.2403 - 0.3666i$, $k_{3,4} = \pm 0.3571 - 0.6019i$. We see that the curve rises more gently near $x = 0$ for 4 poles than for 3 poles. Fig. 4 depicts a 5-pole case: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$, $k_4 = -1.4i$, $k_5 = -1.5i$, and the curve rises even more slowly near $x = 0$ in this 5-pole case than in the 4-pole case.

Figure 5 depicts a 6-pole case: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$, $k_4 = -1.4i$, $k_5 = -1.5i$, $k_6 = -1.6i$. We see that this curve rises even more gradually near the origin. Note that a closed-form solution to the 6-pole case would require a solution to a 5th-order polynomial equation, which cannot be solved by radicals in general.

Figure 6 shows V versus x for another 3-pole case: $k_1 = -1.1i$, $k_2 = -1.2i$, $k_3 = -1.3i$. To test the stability of the Gel'fand–Levitan procedure, we also considered three perturbations of this 3-pole case:

(a) $k_1 = -1.11i$, k_2 and k_3 unchanged;

(b) $k_2 = -1.21i$, k_1 and k_3 unchanged;

(b) $k_3 = -1.31i$, k_1 and k_2 unchanged.

These perturbations of the poles, which are of the order of one percent, cause small changes in $K(x,x)$, which are of the order of one percent. Hence the procedure is stable in this case. In Fig. 7, we plotted the change in $K(x,x)$ (perturbed value minus unperturbed value) divided by the unperturbed value of $K(x,x)$, for cases (a), (b), and (c).

Thus our procedure gives useful results even in cases where other methods fail. In future communications we intend to extend our method still further.

ACKNOWLEDGMENTS

This work was supported in part by the Air Force Office of Scientific Research and the National Science Foundation. One of us (J.M.C.) is indebted to Dr. Angelo J. Skalafuris for his hospitality at the Mathematics Research Center, where this work was begun.

[1] I. Kay, Commun. Pure Appl. Math. **13**, 371 (1960).
[2] S. Ahn and A. K. Jordan, IEEE Trans. Antennas Propag. **24**, 879 (1976).
[3] H. E. Moses, Stud. in Appl. Math. **60**, 177 (1979).
[4] H. E. Moses (unpublished).
[5] K. R. Pechenick and J. M. Cohen, University of Pennsylvania Report (1980).
[6] H. H. Szu, C. E. Carroll, C. C. Yang, and S. Ahn, J. Math. Phys. **17**, 1236 (1976).
[7] J. M. Cohen and M. W. Kearney, University of Pennsylvania Report (1980).
[8] I. M. Gel'fand and B. M. Levitan, Izv. Akad. Nauk SSSR, Ser. Math. **15**, 309 (1951), Amer. Math. Soc. Transl. **1**, 253 (1955).
[9] I. Kay, NYU Res. Report No. EM-74 (1955).
[10] I. Kay and H. E. Moses, Nuovo Cimento **3**, 276 (1956).
[11] I. Kay and H. E. Moses, J. Appl. Phys. **27**, 1503 (1956).