A class of dissipative evolutions with applications in thermodynamics of fermion systems

M. Fannes[a),b)] and F. Rocca
Physique Théorique, Université de Nice, [c)] France

(Received 17 April 1979; accepted for publication 3 October 1979)

We construct explicitly a special class of semigroups of completely positive maps on the CAR algebra and give an application on a model of thermal contact.

I. INTRODUCTION

Recently, semigroups of completely positive maps on C^*-algebras have been introduced in mathematical physics. The main physical motivation to study such objects comes from non equilibrium statistical mechanics. Groups of automorphisms have been used with success to describe dynamics of reversible processes. They give also a good description of return to equilibrium for weakly interacting systems close to equilibrium.

In opposite, one cannot hope to use groups of automorphisms to describe irreversible processes arising from systems far from equilibrium. The use of semigroups of completely positive maps is a simple theoretical alternative in this case, which conserves the deterministic properties of groups in only one time direction and involves nonhomomorphic transformations of the algebras.

The general structure of such semigroups is only known under some strong continuity conditions.[1] On the other hand several constructions are available in particular cases.[2-5]

This paper is organized as follows. In Sec. II, we introduce semigroups of quasi-free completely positive maps on Clifford algebra and describe their structure. In Sec. III, we consider more specifically a one-dimensional continuous system of fermions and study a special class of quasi-free semigroups, namely those which transform in itself the class of quasi-free states which are stable and invariant under the free evolution. We give explicit constructions in Proposition III. 5.

Finally, Sec. IV is devoted to an application on a model of thermal contact. We exhibit a quasi-free semigroup which asymptotically transforms an initial state describing two subsystems at different temperatures to the expected equilibrium state for the total system.

II. COMPLETELY POSITIVE QUASI-FREE SEMIGROUPS ON CLIFFORD ALGEBRA

Let $(H, s(.,.))$ be a real Hilbert space; its associated Clifford algebra $\mathfrak{A}(H,s)$ is generated by $\mathbb{1}$ and the self-adjoint, real linear elements $B(\psi)$, $\psi \in H$, satisfying

$$\{B(\psi,)B(\varphi)\} = 2s(\psi,\varphi)\mathbb{1}.$$

A linear map $\Phi : \mathfrak{A}(H,s) \to \mathfrak{A}(H,s)$ is called completely positive (CP) if, for all $n \in \mathbb{N}$, $\Phi \otimes 1_{M_n}$ is positive on $\mathfrak{A}(H,s) \otimes M_n$, where M_n is the algebra of $n \times n$ complex matrices.

In the following, we shall deal with CP maps of a special structure, namely the quasi-free completely positive maps.

A linear mapping ρ from the algebra of finite, complex, linear combinations of $\mathbb{1}$ and the monomials $B(\psi_1)...B(\psi_n)$ with $\psi_i \in H$, $n = 1,2,...$, into $\mathbb{C}$ is called a quasi-free functional on $\mathfrak{A}(H,s)$ if it satisfies:

(i) $\rho(B(\psi_1)\cdots B(\psi_{2n+1})) = 0$,

(ii) $\rho(B(\psi_1)\cdots B(\psi_{2n}))$

$$= \sum_P (-1)^{\chi(P)} \rho(B(\psi_{i_1})B(\psi_{i_2}))\cdots\rho(B(\psi_{i_{2n-1}})B(\psi_{i_{2n}}))$$

where the sum is taken over all partitions P of $\{1,2,...,2n\}$ into sets $\{i_{2k-1}, i_{2k}\}$, $k = 1,...,n$, such that $i_{2k-1} < i_{2k}$ and $i_1 < i_3 < \cdots < i_{2n-1}$ and where $\chi(P)$ is the parity of the permutation $\{1,2,...,2n\} \to \{i_1, i_2, ..., i_{2n}\}$.

*Theorem II.*1: Let $T \in \mathfrak{B}(H)$, ρ be a quasi-free functional on $\mathfrak{A}(H,s)$ and Φ be defined by:

$$\Phi(B(\psi_1)\cdots B(\psi_n)) = \sum_P (-1)^{\chi(P)} B(T\psi_{i_1})\cdots B(T\psi_{i_k}) \times \rho(B(\psi_{i_{k-1}})\cdots B(\psi_{i_n})), \quad (1)$$

where the summation is taken over all the partitions P of $\{1,...,n\}$ into two sets $\{i_1,...,i_k\}$, $\{i_{k+1},...,i_n\}$ satisfying $i_1 < i_2 < \cdots < i_k$, $i_{k+1} < \cdots < i_n$ (a monomial of order zero is taken to be $\mathbb{1}$) and where $\chi(P)$ is the parity of the permutation $\{1,...,n\} \to \{i_1,...,i_n\}$.

Then Φ extends to a completely positive unity preserving mapping on $\mathfrak{A}(H,s)$ if and only if

(i) T is a contraction on (H,s);

(ii) ρ extends to a quasi-free state on $\mathfrak{A}(H,s_T)$, where $s_T(.,.) = s(.,.) - s(T.,T.)$.

Proof: The sufficiency of the conditions (i) and (ii) follows from a straightforward generalization of Ref. 4.

We now prove the necessity. As ρ is quasi-free one has for $\psi \in H$

$$\begin{aligned} s(\psi,\psi)^2\rho(\mathbb{1}) &= \rho(B(\psi)B(\psi)B(\psi)B(\psi)) \\ &= \rho(B(\psi)B(\psi))\rho(B(\psi)B(\psi)) \\ &= s(\psi,\psi)^2\rho(\mathbb{1})^2, \end{aligned}$$

and so

$$\rho(\mathbb{1}) = 1.$$

As $\forall \psi \in H$,

$$\Phi(B(\psi)) = B(T\psi)\rho(\mathbb{1}) = B(T\psi),$$

[a)] Permanent Address: Instituut voor Theoretische Fysica, Universiteit Leuven, Celestijnenlaan 200 D, B-3030 Leuven, Belgium.
[b)] Aangesteld Navorser N.F.W.O., Belgium.
[c)] Equipe de Recherche Associée au C.N.R.S. Postal Address: Physique Théorique, Parc Valrose, 06034 Nice Cedex, France.

 0022-2488/80/020221-06$1.00

and as Φ is certainly 2-positive, one has:

$$\begin{aligned} s(T\psi,T\psi)\mathbb{1} &= B(T\psi)B(T\psi) \\ &= \Phi(B(\psi))\Phi(B(\psi)) \\ &\leqslant \Phi(B(\psi)B(\psi)) \\ &= s(\psi,\psi)\Phi(\mathbb{1}) \\ &= s(\psi,\psi)\mathbb{1}. \end{aligned}$$

Therefore T is a contraction.

Finally for $\psi_1, \psi_2 \in H$ and $\lambda \in \mathbb{C}$ one finds, using:

$$\begin{aligned} &\Phi[B(\psi_1)+\lambda B(\psi_2)]\Phi[B(\psi_1)+\lambda B(\psi_2)]^* \\ &\quad \leqslant \Phi[(B(\psi_1)+\lambda B(\psi_2))(B(\psi_1)+\bar{\lambda}B(\psi_2))], \end{aligned}$$

that

$$\begin{aligned} &|\lambda|^2 s_T(\psi_2,\psi_2) + 2\,\mathrm{Re}\{\bar{\lambda}\rho(B(\psi_1)B(\psi_2))\} \\ &\quad + s_T(\psi_1,\psi_1) \geqslant 0. \end{aligned}$$

As T is a contraction on (H,s) this implies that

$$|\rho(B(\psi_1)B(\psi_2))|^2 \leqslant s_T(\psi_2,\psi_2)s_T(\psi_1,\psi_1). \tag{2}$$

Using the self-adjointness of ρ and (2) one now immediately obtains

$$\rho(B(\psi_1)B(\psi_2)) = s_T(\psi_1,\psi_2) + is_T(A\psi_1,\psi_2),$$

where $A \in \mathfrak{B}(H,s_T)$ satisfies

$$s_T(A\psi_1,\psi_2) = -s_T(\psi_1,A\psi_2)$$

and

$$\|A\|_{s_T} \leqslant 1.$$

It follows that ρ extends to a quasi-free state on $\mathfrak{A}(H,s_T)$. ■

A completely positive unity preserving map on $\mathfrak{A}(H,s)$ that satisfies the conditions of Theorem II.1 will be called quasi-free and we use the short-hand notation CPQF to designate it. In order to define a CPQF map it is sufficient to give its action on monomials $B(\psi_1)B(\psi_2)$ of second degree.

Usually, one starts from a complex Hilbert space $(\mathscr{H},\langle\cdot|\cdot\rangle)$. Its underlying real Hilbert space (H,s), with $s(\cdot,\cdot) = \mathrm{Re}\langle\cdot|\cdot\rangle$, coincides with $\mathscr{H}$ as a real vector space. In this case, creation and annihilation operators are defined by

$$\begin{aligned} a^*(\psi) &= \tfrac{1}{2}[B(\psi) - iB(i\psi)], \\ a(\psi) &= \tfrac{1}{2}[B(\psi) + iB(i\psi)], \end{aligned} \qquad \psi \in \mathscr{H},$$

and satisfy the usual CAR. The gauge automorphisms γ_θ are then given by

$$\gamma_\theta a^*(\psi) = e^{i\theta}a^*(\psi), \quad \theta \in [0,2\pi].$$

We are especially interested in strongly continuous semigroups of CPQF maps which commute with the gauge automorphisms.

One immediately checks that such a semigroup Φ_t is defined by

$$\Phi_t : a^*(\psi)a(\varphi) \to a^*(T_t\psi)a(T_t\varphi) + \langle Q_t\varphi\,|\psi\rangle, \tag{3}$$

where $\{T_t | t \in \mathbb{R}^+\}$ is a strongly continuous contraction semigroup on $\mathscr{H}$ and $\{Q_t | t \in \mathbb{R}^+\}$ is a weakly continuous family of bounded operators on $\mathscr{H}$ satisfying

$$0 \leqslant Q_t \leqslant 1 - T_t^*T_t, \quad t \geqslant 0$$

and

$$Q_{t_1+t_2} = Q_{t_1} + T_{t_1}^*Q_{t_2}T_{t_1}, \quad t_1, t_2 \geqslant 0.$$

Let ω_A be a gauge invariant quasi-free state defined by[6]

$$\omega_A(a^*(\psi)a(\varphi)) = \langle\varphi\,|A\psi\rangle,$$

and $\{\Phi_t | t \in \mathbb{R}^+\}$ as in Eq. (3), then $\omega_A \circ \Phi_t$ is again a gauge invariant quasi-free state and

$$\omega_A \circ \phi_t = \omega_{T_t^*AT_t + Q_t}. \tag{4}$$

III. CPQF SEMIGROUPS ASSOCIATED TO FREE EVOLUTIONS

In this section we explicitly construct gauge invariant CPQF semigroups. To keep the situation as simple as possible, we will treat a one dimensional continuous system of free fermions. From now on, $\mathfrak{A}$ is the Clifford algebra built on $\mathscr{L}^2(\mathbb{R},dx) = \mathscr{H}$.

The free evolution on $\mathfrak{A}$ is described by a group $\{\alpha_t | t \in \mathbb{R}\}$ of automorphisms given by

$$\alpha_t a^*(\psi) = a^*(e^{-itA}\psi), \quad \psi \in \mathscr{H}.$$

We pass immediately to the momentum representation. The free evolution is then given by

$$a^*(\psi) \to a^*(M_{\exp(itk^2)}\psi),$$

where M_F denotes the multiplication operator on $\mathscr{H}$ with the function $k \to F(k)$.

A gauge invariant quasi-free state which is invariant and stable for the free evolution is determined by a multiplication operator $M_{a(k)}$ on $\mathscr{H}$ with $0 \leqslant a(k) \leqslant 1$ and $a(k) = a(-k)$.

We now write $\mathscr{H} = \mathscr{H}^- \oplus \mathscr{H}^+$ with

$$\mathscr{H}^\pm = \{\psi \in \mathscr{H} \mid \mathrm{Supp}\psi \subset \mathbb{R}^\pm\},$$

and define the inversion operator

$$(p\psi)(k) = \psi(-k). \tag{5}$$

So, the gauge invariant quasi-free states which are invariant and stable for the free evolution are in one-to-one correspondence with the multiplication operators on $\mathscr{H}^+$, nonnegative and bounded by 1.

We wish now to study the subclass of gauge invariant CPQF semigroups which are essentially characterized by the requirement that they conserve invariance and stability properties of states with respect to the free evolution. As long as only gauge invariant quasi-free states are concerned the previous remarks allow us to restrict our study to semigroups on $\mathfrak{A}(\mathscr{H}^+)$. Then, using in the momentum representation the same notations as in Eq. 3, it is sufficient, considering Eq. 4, to impose that $\tilde{\Phi}_t(\mathscr{M}) \subset \mathscr{M}$, where

$$\tilde{\Phi}_t(A) = T_t^*AT_t + Q_t, \quad A \in \mathfrak{B}(\mathscr{H}^+),$$

and $\mathscr{M}$ is the Von Neuman algebra of multiplication operators on $\mathscr{H}^+$.

For clarity, we summarize the previous notions in the following definition:

Definition III.1: A free semigroup of CPQF maps on $\mathfrak{A}(\mathscr{H}^+)$ is a semigroup of maps Φ_t, $t \in \mathbb{R}^+$, defined by

$$\Phi_t : a^*(\varphi)a(\psi) \to a^*(T_t\varphi)a(T_t\psi) + \langle Q_t\psi\,|\varphi\rangle$$

with the following properties:

(i) $t \in \mathbb{R}^+ \to T_t$ is a strongly continuous semigroup of

contractions on $\mathscr{H}^+$,

(ii) $t\in\mathbb{R}^+\to Q_t$ is weakly continuous and $0\leqslant Q_t \leqslant 1 - T_t^*T_t$,

(iii) $Q_{t_1+t_2} = Q_{t_1} + T_{t_1}^*Q_{t_2}T_{t_1}$, $t_1, t_2 \geqslant 0$,

(iv) $\tilde{\Phi}_t(\mathscr{M}) \subset \mathscr{M}$.

It follows immediately from (iii) that $Q_t\in\mathscr{M}$ and $T_t^*\mathscr{M}T_t \subset \mathscr{M}$. As T_t is a contraction semigroup $T_t^*T_t$ is decreasing and bounded by 1 from above, s-$\lim_{t\to\infty} T_t^*T_t = R_\infty$ exists and belongs to $\mathscr{M}$. It follows also from (ii) and (iii) that Q_t is increasing and bounded by 1, therefore also Q_∞ = s-$\lim_{t\to\infty} Q_t$ exists and belongs to $\mathscr{M}$. Two essentially different cases can be distinguished:

(1) $R_\infty = 0$, then T_t converges strongly to zero and one sees immediately from (1) that for any state ω, $\omega\circ\Phi_t$ converges in the w^*-topology as $t\to\infty$ to one and the same gauge invariant quasi-free state defined by Q_∞. Explicit examples of this type of situation can be found in Ref. 2.

(2) $R_\infty \neq 0$. In this case there exists at least a one parameter family of gauge invariant quasi-free states, invariant under Φ_t; namely those defined by the operators $Q_\infty + \lambda R_\infty$, $0\leqslant\lambda\leqslant 1$. We now investigate this case in more detail.

Lemma III.2: With the same notations as above: Let $\{T_t | t\in\mathbb{R}^+\}$ be a strongly continuous semigroup of contractions on $\mathscr{H}^+$ such that

(i) $T_t^*\mathscr{M}T_t \subset \mathscr{M}$, $t\geqslant 0$,

(ii) $T_t^*M_{F_0}T_t = M_{F_0}$, where $0\neq F_0\in\mathscr{L}^\infty_+(\mathbb{R}^+, dk)$, and where M_{F_0} is the multiplication operator by F_0 on $\mathscr{H}^+$.

There exists then a strongly continuous semigroup $\{V_t | t\in\mathbb{R}^+\}$ of isometries on $\mathscr{H}^+$ such that for $t\geqslant 0$

(i) $V_t^*\mathscr{M}V_t \subset \mathscr{M}$,

(ii) $M_{F_0(1/2)}T_t = V_tM_{F_0(1/2)}$.

Proof: As $T_t^*M_{F_0}T_t = M_{F_0}$ one has for the polar decomposition of $M_{F_0(1/2)}T_t$:

$$M_{F_0(1/2)}T_t = V_tM_{F_0(1/2)},$$

where V_t maps $\{M_{F_0(1/2)}\mathscr{H}^+\}$ into itself. Extending V_t by the identity operator on $\{M_{F_0(1/2)}\mathscr{H}^+\}^\perp$ one obtains a collection of partial isometries, still denoted by V_t, that satisfy:

$$M_{F_0(1/2)}T_t = V_tM_{F_0(1/2)},$$
$$V_t\varphi = \varphi, \quad \varphi\in\{M_{F_0(1/2)}\mathscr{H}^+\}^\perp. \tag{6}$$

Writing

$$M_{F_0} = T_t^*M_{F_0}T_t = M_{F_0(1/2)}V_t^*V_tM_{F_0(1/2)},$$

and using Eq. (6) one obtains $V_t^*V_t = 1$, $t\geqslant 0$. So V_t is an isometry. As

$$V_{t_1+t_2}M_{F_0(1/2)} = M_{F_0(1/2)}T_{t_1+t_2} = M_{F_0(1/2)}T_{t_1}T_{t_2} = V_{t_1}M_{F_0(1/2)}T_{t_2} = V_{t_1}V_{t_2}M_{F_0(1/2)}, \quad t_1, t_2\geqslant 0,$$

one finds, using Eq. (6) again, that $\{V_t | t\in\mathbb{R}^+\}$ is a semigroup. In order to show that it is strongly continuous it is sufficient to observe that

$$\|(V_t - V_{t_1})M_{F_0(1/2)}\varphi\| = \|M_{F_0(1/2)}(T_t - T_{t_1})\varphi\| \leqslant \|(T_t - T_{t_1})\varphi\|, \quad t, t_1\geqslant 0, \ \varphi\in\mathscr{H}^+$$

and to use the strong continuity of $\{T_t | t\in\mathbb{R}^+\}$.

Finally we show that $V_t^*\mathscr{M}V_t \subset \mathscr{M}$. As $\mathscr{M}$ is maximal Abelian, $M_{F_0}\in\mathscr{M}$ and using the construction of V_t it is enough to show that for $F_1, F_2\in\mathscr{L}^\infty(\mathbb{R}^+, dk)$ and $t\geqslant 0$,

$$[V_t^*M_{F_1}V_t, M_{F_2}]\{M_{F_0(1/2)}\mathscr{H}^+\} = 0.$$

Let $\psi\in\mathscr{H}^+$, then

$$\begin{aligned}\langle M_{F_0(1/2)}\psi \,|\, [V_t^*M_{F_1}V_t, M_{F_2}]M_{F_0(1/2)}\psi\rangle &= \langle\psi | M_{F_0(1/2)}V_t^*M_{F_1}V_tM_{F_2}M_{F_0(1/2)}\psi\rangle - \langle\psi | M_{F_0(1/2)}M_{F_2}V_t^*M_{F_1}V_tM_{F_0(1/2)}\psi\rangle \\ &= \langle\psi | T_t^*M_{F_0}M_{F_1}T_tM_{F_2}\psi\rangle - \langle\psi | M_{F_2}T_t^*M_{F_0}M_{F_1}T_t\psi\rangle = 0.\end{aligned}$$

■

Our aim was to construct explicitly the free semigroups of CPQF maps as characterized in the Definition III. 1. However, as we cannot solve the equations in Definition III.1 in their full generality, we will supplement them with a regularity condition introduced in the following lemma.

Lemma III.3: Let $\{V_t | t\in\mathbb{R}^+\}$ be a strongly continuous semigroup of isometries on $\mathscr{H}^+$ which satisfy $V_t^*\mathscr{M}V_t \subset \mathscr{M}$. The following conditions are equivalent:

(i) $\exists t>0$ such that $M_FV_t = V_tV_t^*M_FV_t$, $M_F\in\mathscr{M}$,

(ii) $\exists t>0$ such that $V_tV_t^*\in\mathscr{M}$. If one of these conditions is satisfied then both hold for all $t\geqslant 0$, and $M_F \to V_t^*M_FV_t$ is a semigroup of $*$-homomorphisms of $\mathscr{M}$.

Proof: (i)$\to$(ii) Consider M_F in $\mathscr{M}$. Then

$$M_FV_tV_t^* = V_tV_t^*M_FV_tV_t^* = V_t\{V_tV_t^*M_F^*V_t\}^* = V_t\{M_F^*V_t\}^* = V_tV_t^*M_F.$$

As $\mathscr{M}$ is maximal Abelian, $V_tV_t^*\in\mathscr{M}$.

(ii)$\to$(i)

$$V_tV_t^*M_FV_t = M_FV_tV_t^*V_t = M_FV_t.$$

The last statements of the lemma follow immediately. ■

Proposition III.4: Let $\{V_t | t\in\mathbb{R}^+\}$ satisfy the conditions of Lemma III.3, then there exists a family $\{u_t | t\in\mathbb{R}^+\}$ of absolutely continuous functions $k\in\mathbb{R}^+\to u_t(k)\in\mathbb{R}^+$, and a family $\{\theta_t | t\in\mathbb{R}^+\}$ of real-valued measurable functions $k\in\mathbb{R}^+\to\theta_t(k)\in\mathbb{R}$ such that:

(i) $u_0(k) = k$ and $t\to u_t(k)$ continuous k-a.e.,

(ii) $\theta_0(k) = 0$ and $t\to\theta_t(k)$ continuous k-a.e.,

(iii) $u_{t_1+t_2}(k) = u_{t_2}(u_{t_1}(k))$, $t_1, t_2\geqslant 0$,

(iv) $\theta_{t_1+t_2}(k) = \theta_{t_2}(k) + \theta_{t_1}(u_{t_2}(k))$, $t_1, t_2\geqslant 0$,

(v) $(V_t^*\varphi)(k) = e^{i\theta_t(k)}|u_t'(k)|^{1/2}\varphi(u_t(k))$, $\varphi\in\mathscr{H}^+$.

Proof: Consider the strongly continuous group of unitaries $\{M_{e^{isk}} | s\in\mathbb{R}\}$ on $\mathscr{H}^+$. By Lemma III.3, $V_t^*M_{e^{isk}}V_t$ is unitary. Hence $\{V_t^*M_{e^{isk}}V_t \mid s\in\mathbb{R}\}$ is also a strongly continuous group of unitaries in $\mathscr{M}$. By Stone's theorem

$$V_t^*M_{e^{isk}}V_t = M_{\exp[isu_t(k)]},$$

where $k\to u_t(k)$ is a measurable function from $\mathbb{R}^+$ into $\mathbb{R}^+$. It follows that:

$$V_t^*M_FV_t = M_{F\circ u_t}, \quad F\in\mathscr{L}^\infty(\mathbb{R}^+, dk).$$

By Lemma III.3, $V_tV_t^*\in\mathscr{M}$. We note Δ_t the support of $V_tV_t^*$. Consider now a $\psi\in\mathscr{H}^+$ such that Supp$V_t\psi = \Delta_t$; then, as $V_t^*V_t = 1$, Supp$\psi = \mathbb{R}^+$. One has for $F\in\mathscr{L}^\infty(\mathbb{R}^+, dk)$

$$V_t^*M_{|F|^2}V_t\psi = M_{|F\circ u_t|^2}\psi,$$

so

$$\int |F(k)|^2 |(V_t\psi)(k)|^2\, dk = \int |F(u_t(k))|^2 |\psi(k)|^2\, dk.$$

It follows that $k \to u_t(k)$ is absolutely continuous.

We take now $\psi \in \mathscr{H}^+$. Since

$$V_t^* \varphi = V_t^* V_t V_t^* \varphi,$$

one can suppose without restriction that $\mathrm{Supp}\varphi \subset \Delta_t$, in order to prove (v). We have

$$V_t^* M_{|F|^2} V_t V_t^* \varphi = M_{|F\circ u_t|^2} V_t^* \varphi = V_t^* M_{F|^2} \varphi.$$

Taking the scalar product with $V_t^*\varphi$ one finds:

$$\int_{\Delta_t} |F(k)|^2 |\varphi(k)|^2\, dk = \int |F(u_t(k))|^2 |(V_t^*\varphi)(k)|^2\, dk.$$

Hence, if

$$\int |F(u_t(k))|^2 |(V_t^*\varphi)(k)|^2\, dk = 0, \quad \forall F \in \mathscr{L}^\infty(\mathbb{R}^+, dk),$$

$$\varphi(k) = 0 \text{ on } \Delta_t,$$

and

$$V_t^*\varphi = 0.$$

Consequently

$$\int |F(u_t(k))|^2 |(V_t^*\varphi)(k)|^2\, dk$$

$$= \int |F(u_t(k))|^2 |u_t'(k)|\, |\varphi(u_t(k))|^2\, dk,$$

leads to

$$|(V_t^*\varphi)(k)| = |u_t'(k)|^{1/2} |\varphi(u_t(k))|,$$

and (v) follows.

The other statements of the proposition now follow easily:

(iii) and (iv) are consequence of semigroup properties of $\{V_t | t \in \mathbb{R}^+\}$

(i) and (ii) express the initial conditions and the consequences of strong continuity of $t \to V_t$. ■

We are now in position to construct explicitly a class of free semigroup of CPQF maps.

Let $v : \mathbb{R}^+ \to \mathbb{R}$ be a differentiable, strictly monotone increasing function such that $v(\mathbb{R}^+) + \mathbb{R}^+ \subset v(\mathbb{R}^+)$. Putting in Proposition III.4 $u_t(k) = v^{-1}(v(k) + t)$ and $\theta_t(k) = 0$ one finds by (v) of Proposition III.4 a strongly continuous semigroup of isometries $\{V_t^* | t \in \mathbb{R}^+\}$ of $\mathscr{H}^+$. Let now $h : \mathbb{R} \to \mathbb{C}$ be a measurable function such that $\mathrm{Re}h \geqslant 0$ and such that

$$F_0(k) = \exp\left(-2\int_{v(0)}^{v(k)} ds\, \mathrm{Re}h(s)\right) \in \mathscr{L}^\infty(\mathbb{R}^+, dk).$$

The equation $T_t^* M_{F_0(1/2)} = M_{F_0(1/2)} V_t^*$ of Lemma III.2 then admits as unique solution a strongly continuous semigroup $\{T_t^* | t \in \mathbb{R}^+\}$ of contractions on $\mathscr{H}^+$ satisfying the conditions of Lemma III.2. Finally let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be a bounded measurable function such that $g \leqslant 2\mathrm{Re}(h \circ v)$, and put

$$Q_t = \int_0^t ds\, T_s^* M_g T_s,$$

then $\{T_t | t \in \mathbb{R}^+\}$ and $\{Q_t | t \in \mathbb{R}^+\}$ satisfy Definition III.1. This situation can be generalized as follows:

Proposition III.5: Suppose that

(a) $v : \mathbb{R}^+ \to \mathbb{R}$ is differentiable, strictly monotone increasing and satisfies $v(\mathbb{R}^+) + \mathbb{R}^+ \subset v(\mathbb{R}^+)$;

(b) $h : \mathbb{R} \to \mathbb{C}$ is measurable and $\mathrm{Re}h \geqslant 0$;

(c) $g : \mathbb{R}^+ \to \mathbb{R}^+$ is measurable and $g \leqslant 2\,\mathrm{Re}(h \circ v)$.

Then

(i) $\{T_t | t \in \mathbb{R}^+\}$ and $\{Q_t | t \in \mathbb{R}^+\}$ satisfy Definition III.1, where

$$(T_t^*\psi)(k) = \left|\frac{dv^{-1}}{dk}(v(k) + t)\right|^{1/2} \left|\frac{dv}{dk}(k)\right|^{1/2} \times \exp\left[-\int_0^t \tilde{h}(v(k) + s)\, ds\right] \psi(v^{-1}(v(k) + t)),$$

$$Q_t = M_{K_t},$$

with

$$K_t(k) = \int_0^t ds \exp\left[-2\int_0^s \mathrm{Re}h(v(k) + r)\, dr\right] \times g(v^{-1}(v(k) + s)), \quad k \in \mathbb{R}^+.$$

(ii) The semigroup $\{T_t | t \in \mathbb{R}^+\}$ tends strongly to zero iff

$$\int_{v(0)}^{\infty} \mathrm{Re}h(s)\, ds = +\infty.$$

(iii) M_F, with F absolutely continuous, is invariant under $\{\tilde{\phi}_t | t \in \mathbb{R}^+\}$ iff it satisfies the equation:

$$F' = v'[2F\,\mathrm{Re}(h \circ v) - g].$$

(iv) If

$$\lim_{k \to \infty} F(k) = 0, \quad F \in \mathscr{L}^\infty(\mathbb{R}^+, dk),$$

then

$$\text{s-lim}_{t \to \infty} T_t^* M_F T_t = 0.$$

Proof: (i) is straightforward to check. Note that $K_t \in \mathscr{L}^\infty(\mathbb{R}^+, dk)$ as:

$$0 \leqslant K_t(k) \leqslant \int_0^t ds\, 2\,\mathrm{Re}h(v(k) + s) \times \exp\left[-2\int_0^s \mathrm{Re}h(v(k) + r)\, dr\right] \leqslant 1 - \exp\left[-2\int_0^t \mathrm{Re}h(v(k) + r)\, dr\right].$$

This inequality yields also $Q_t \leqslant 1 - T_t^* T_t$

(ii) Define a unitary operator

$$V : \mathscr{L}^2(\mathbb{R}^+, ds) \to \mathscr{L}^2([v(0), \infty], ds),$$

by

$$(V\psi)(s) = \left|\frac{dv^{-1}}{ds}(s)\right|^{1/2} \psi(v^{-1}(s)), \quad s \in [v(0), \infty], \quad \psi \in \mathscr{H}^+,$$

and the strongly continuous semigroup $\{S_t | t \in \mathbb{R}^+\}$ of (right) shift operators on $\mathscr{L}^2([v(0), \infty], ds)$,

$$(S_t\psi)(s) = 0, \quad s < v(0) + t,$$
$$= \psi(s - t), \quad s \geqslant v(0) + t.$$

One has then

$$T_t = V^* M_{H_t} S_t V, \tag{7}$$

where M_{H_t} is the multiplication operator on $\mathscr{L}^2([v(0), \infty], ds)$ by

$$H_t(s) = \exp\left[- \int_0^t h(s - t + r)\, dr \right].$$

Therefore the semigroup $\{T_t | t \in \mathbb{R}^+\}$ converges strongly to zero on $\mathscr{H}^+$ iff $\{M_{H_t} S_t | t \in \mathbb{R}^+\}$ converges strongly to zero on $\mathscr{L}^2([v(0), \infty], ds)$ and this is the case iff:

$$\int_{v(0)}^{\infty} \mathrm{Re} h(s)\, ds = + \infty,$$

(iii) is immediate;

(iv) with the same notations as in part (ii) of the proof, $VM_F V^*$ is the multiplication operator on $\mathscr{L}^2([v(0), \infty], ds)$ by $s \to F(v^{-1}(s))$. As $\lim_{s \to \infty} F(v^{-1}(s)) = 0$ one finds

$$\text{s-lim}_{t \to \infty} S_t^* M_{H_t}^* M_{F \circ v^{-1}} M_{H_t} S_t = 0,$$

one $\mathscr{L}^2([v(0), \infty], ds)$, and so, using Eq. 7,

$$\text{s-lim}_{t \to \infty} T_t^* M_F T_t = 0, \quad \text{on } \mathscr{H}^+. \qquad \blacksquare$$

To put in evidence the physical content of Proposition III.5, we give immediately a possibility of application in a simple but typical situation.

Consider a system which is prepared in a translation invariant quasi-free state determined by a function $k \to F_0(k)$ which vanishes at infinity. In order to let it evolve towards the translation invariant quasi-free state determined by $k \to F(k)$, where $F(k)$ also vanishes at infinity, one can try to use a semigroup $\{\Phi_t | t \in \mathbb{R}^+\}$ of CPQF maps. One gets the desired result if it is possible to construct functions v, h and g satisfying (a), (b), and (c) of Proposition III.5 and such that

$$F' = v'(2F\,\mathrm{Re}(h \circ v) - g) \quad \text{[Prop. III.5 (iii)]}.$$

Then, by invariance of M_F under $\tilde{\Phi}_t$ and by Proposition III.5 (iv), one gets

$$M_F = \text{s-lim}_{t \to \infty} \tilde{\Phi}_t(M_F) = \text{s-lim}_{t \to \infty} (T_t^* M_F T_t + Q_t) = Q_\infty,$$

and so

$$\text{s-lim}_{t \to \infty} \tilde{\Phi}_t(M_{F_0}) = \text{s-lim}_{t \to \infty} (T_t^* M_{F_0} T_t + Q_t)$$
$$= \text{s-lim}_{t \to \infty} Q_t$$
$$= Q_\infty = M_F.$$

IV. APPLICATION TO A MODEL OF THERMAL CONTACT

We explicitly construct in this section a free semigroup of CPQF maps which describes the physical effect of temperature equalization between two systems of fermions.

We start from a one-dimensional free fermion system which is initially far from the thermodynamical equilibrium: At the left of the origin, it is in equilibrium at the inverse temperature and chemical potential (β_1, μ_1) and at the right at the inverse temperature and chemical potential (β_2, μ_2). We look for a dissipative evolution under which this initial state tends asymptotically to the equilibrium state at the inverse temperature and chemical potential (β_3, μ_3), determined from (β_1, μ_1) and (β_2, μ_2) using standard thermodynamical considerations.

We introduce some notation: $\mathfrak{A}^{\pm}$ is the Clifford algebra built on $\mathscr{L}^2(\mathbb{R}^{\pm}, dx)$; if $k \in \mathbb{R} \to F(k) \in [0,1]$ is a measurable function, ω_F will denote the gauge and translation invariant quasi-free state on $\mathfrak{A} = \mathfrak{A}(\mathscr{L}^2(\mathbb{R}, dx))$ determined by the 2-point function:

$$\omega_F(a^*(\varphi) a(\psi)) = \int_{\mathbb{R}} dk\, \overline{\hat{\psi}(k)}\, F(k) \hat{\varphi}(k).$$

If ω is a state on $\mathfrak{A}$, $\omega^{\pm}$ denotes the restriction of ω to $\mathfrak{A}^{\pm}$. We note $F_{\beta,\mu}$ the Fermi–Dirac distribution function at inverse temperature β and chemical potential μ:

$$F_{\beta,\mu}(k) = \frac{1}{e^{\beta(k^2 - \mu)} + 1}, \quad k \in \mathbb{R}.$$

The initial state of the system is taken as the product state à la Powers[7]:

$$\omega^-_{F_{\beta_1,\mu_1}} \otimes \omega^+_{F_{\beta_2,\mu_2}},$$

and our aim is to calculate

$$\lim_{t \to \infty} \left[(\omega^-_{F_{\beta_1,\mu_1}} \otimes \omega^+_{F_{\beta_2,\mu_2}}) \circ \Phi_t \right](a^*(\varphi) a(\psi)), \quad \varphi, \psi \in \mathscr{H},$$

for the as simple as possible semigroup $\{\Phi_t | t \in \mathbb{R}^+\}$ which leaves $\omega_{F_{\beta_3,\mu_3}}$ invariant and which is constructed with the help of Proposition III.5.

Thus, we consider in momentum representation the action of the semigroup $\{\tilde{\Phi}_t | t \in \mathbb{R}^+\}$. We first describe it on $\mathfrak{B}(\mathscr{L}^2(\mathbb{R}^+, dk))$ and then extend it by inversion with the help of the operator P defined in Eq. (5).

We propose:

$$v(k) = k, \quad k \in \mathbb{R}^+,$$

$$h(k) = -\tfrac{1}{2} \frac{F'_{\beta_3,\mu_3}(k)}{1 - F_{\beta_3,\mu_3}(k)}, \quad k \in \mathbb{R}^+,$$

$$g(k) = -\frac{F'_{\beta_3,\mu_3}(k)}{1 - F_{\beta_3,\mu_3}(k)} = 2h(k), \quad k \in \mathbb{R}^+.$$

One easily verifies that the conditions (a), (b), and (c) of Proposition III.5 are satisfied, and that $M_{F_{\beta_3,\mu_3}}$ is invariant under $\{\tilde{\Phi}_t | t \in \mathbb{R}^+\}$ [(part iii) of the proposition].

The operators T_t and Q_t on $\mathscr{L}^2(\mathbb{R}^+, dk)$ are then given by:

$$(T_t \hat{\varphi}^+)(k) = \begin{cases} 0, & k \leqslant t \\ \left(\dfrac{1 - F_{\beta_3,\mu_3}(k - t)}{1 - F_{\beta_3,\mu_3}(k)} \right)^{1/2} \hat{\varphi}^+(k - t), & k \geqslant t \end{cases},$$

$$(Q_t \hat{\varphi}^+)(k) = \frac{F_{\beta_3,\mu_3}(k) - F_{\beta_3,\mu_3}(k + t)}{1 - F_{\beta_3,\mu_3}(k + t)} \hat{\varphi}^+(k). \tag{8}$$

*Proposition IV.*1: With the same notations as above

$$\text{w}^*\text{-}\lim_{t\to\infty}(\omega_{F_{\beta_1,\mu_1}}^{-}\otimes\omega_{F_{\beta_2,\mu_2}}^{+})\circ\Phi_t=\omega_{F_{\beta_3,\mu_3}}.$$

Proof: As the free semigroups of CPQF maps transform gauge invariant quasi-free states in gauge invariant quasi-free states, it is enough to compute, for all $\varphi\in\mathscr{L}^2(\mathbb{R},dx)$

$$\lim_{t\to\infty}(\omega_{F_{\beta_1,\mu_1}}^{-}\otimes\omega_{F_{\beta_2,\mu_2}}^{+})(\Phi_t(a^*(\varphi)a(\varphi))).$$

This 2-point function can be written as a sum of two terms; the first one contains the action of the operators T_t and the second one pertains to the operators Q_t:

$$(\omega_{F_{\beta_1,\mu_1}}^{-}\otimes\omega_{F_{\beta_2,\mu_2}}^{+})(\Phi_t(a^*(\varphi)a(\varphi)))=I_1(t)+I_2(t),$$

$$I_1(t)=\frac{1}{4\pi^2}\int_{-\infty}^{+\infty}dp\left[F_{\beta_1,\mu_1}(p)\left|\int_{-\infty}^{0}dx\,e^{-ipx}\times\int_{-\infty}^{+\infty}dk\,e^{ikx}(T_t\hat{\varphi})(k)\right|^2+F_{\beta_2,\mu_2}(p)\times\left|\int_0^{+\infty}dx\,e^{-ipx}\int_{-\infty}^{+\infty}dk\,e^{ikx}(T_t\hat{\varphi})(k)\right|^2\right],$$

$$I_2(t)=\int_{-\infty}^{+\infty}dk\,\overline{\hat{\varphi}(k)}\,(Q_t\hat{\varphi})(k).$$

In these formulas T_t and Q_t stand for the operators on $\mathscr{L}^2(\mathbb{R},dx)$ extended from Eq. (8) by inversion. Due to the form of Q_t one obtains immediately

$$\lim_{t\to\infty}I_2(t)=\int_{-\infty}^{+\infty}dk\,\overline{\hat{\varphi}(k)}\,F_{\beta_3,\mu_3}(k)\,\hat{\varphi}(k)=\omega_{F_{\beta_3,\mu_3}}(a^*(\varphi)a(\varphi)).$$

We prove now that

$$\lim_{t\to\infty}I_1(t)=0.$$

$I_1(t)$ is the sum of two terms with the same behavior as $t\to\infty$. We only treat the second one. We can write:

$$\int dp\,F_{\beta_2,\mu_2}(p)\left|\int_0^{\infty}dx\,e^{-ipx}\int_{-\infty}^{+\infty}dk\,e^{ikx}(T_t\hat{\varphi})(k)\right|^2$$

$$\leqslant 2\int dp\,F_{\beta_2,\mu_2}(p)\times\left[\left|\int_0^{\infty}dx\,e^{ipx}\int_{-\infty}^{+\infty}dk\,e^{ikx}(T_t\hat{\varphi}^{+})(k)\right|^2+\left|\int_0^{\infty}dx\,e^{-ipx}\int_{-\infty}^{+\infty}dk\,e^{ikx}(T_t\hat{\varphi}^{-})(k)\right|^2\right],$$

where $\hat{\varphi}^{\pm}$ is the restriction of $\hat{\varphi}$ to $\mathbb{R}^{\pm}$.

We can again limit ourselves to the first term in the right part of the inequality which becomes by introducing Eq. (8) for $(T_t\hat{\varphi}^{+})(k)$:

$$\int_{-\infty}^{+\infty}dp\,F_{\beta_2,\mu_2}(p)\left|\int_0^{\infty}dx\,e^{-ipx}\int_t^{\infty}dk\,e^{ikx}\left(\frac{1-F_{\beta_3,\mu_3}(k-t)}{1-F_{\beta_3,\mu_3}(k)}\right)^{1/2}\hat{\varphi}^{+}(k-t)\right|^2$$

$$\leqslant 2\int_{-\infty}^{+\infty}dp\,F_{\beta_2,\mu_2}(p+t)\left|\int_0^{\infty}dx\,e^{-ipx}\int_0^{\infty}dk\,e^{ikx}[1-F_{\beta_3,\mu_3}(k)]^{1/2}\hat{\varphi}^{+}(k)\right|^2+2\int_{-\infty}^{+\infty}dp\,F_{\beta_2,\mu_2}(p+t)\times\left|\int_0^{\infty}dx\,e^{-ipx}\int_0^{\infty}dk\,e^{ikx}\left(1-\frac{1}{[1-F_{\beta_3,\mu_3}(k+t)]^{1/2}}\right)[1-F_{\beta_3,\mu_3}(k)]^{1/2}\hat{\varphi}^{+}(k)\right|^2.$$

The first term tends to zero as $t\to\infty$ because

$$\lim_{k\to\infty}F_{\beta_2,\mu_2}(k)=0.$$

The second one can be estimated by

$$8\pi^2\|F_{\beta_2,\mu_2}\|_{\infty}\left(1-\frac{1}{[1-F_{\beta_3,\mu_3}]^{1/2}}\right)^2(1-F_{\beta_3,\mu_3}(0))\|\hat{\varphi}^{+}\|^2,$$

and tends to zero because

$$\lim_{k\to\infty}F_{\beta_3,\mu_3}(k)=0.$$

In conclusion, we have explicitly exhibited a semigroup which asymptotically performs the temperature equalization. Remark that $\{T_t\,|\,t\in\mathbb{R}^{+}\}$ does not strongly converge to zero, as follows immediately from Proposition III.5 (ii).

On the other hand, it is not surprising that the final state is to a large extent independent of the initial conditions. Indeed the model is purely descriptive with respect to the thermalization and in particular does not take in account the notions of conservation of particle number and energy which permit to induce theoretically the value of (β_3,μ_3) from the values of (β_1,μ_1) and (β_2,μ_2).

[1]G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).
[2]E.B. Davies, Commun. Math. Phys. **55**, 231 (1977).
[3]D.E. Evans and J.T. Lewis, J. Funct. Anal. **26**, 369 (1977).
[4]D.E. Evans "Completely Positive Quasi-Free Maps on the CAR Algebra", Preprint, University of Copenhagen (1978).
[5]B. Demoen, P. Vanheuverswijn, and A. Verbeure, "Completely Positive Quasi-Free maps of the CCR Algebra", to appear in Rep. Math. Phys.
[6]E. Balslev, J. Manuceau and A. Verbeure, Comm. Math. Phys. **8**, 315 (1968).
[7]R.T. Powers, Princeton thesis (1967).

Generalized back coupling rules for the Racah algebra of Gln

John J. Sullivan
Physics Department, University of New Orleans, New Orleans, Louisiana 70122

(Received 27 February 1979; accepted for publication 17 May 1979)

Two subgroup relations for double coset matrix elements (DCME) of the symmetric group S_N are derived by considering the processes of subduction and induction. The duality of outer product coupling in S_N and inner product coupling in Gln identifies these as generalized back coupling rules for the Racah algebra of Gln. An iterative procedure for evaluating the DCME, once a consistent phase convention is established, is given. As an example the Racah sum formula for the Clebsch–Gordan coefficients of SU2 is derived from consideration of S_N coupling only.

I. DUALITY OF S_N AND Gln

In a series of papers[1] we have been developing the consequences of duality between coupling in the symmetric group S_N and coupling of the integral representations in the general linear group Gln and its unitary Un and unitary unimodular SUn subgroups. Outer product coupling in S_N corresponds to inner product coupling in Gln and/or outer product coupling in $\mathrm{Gl}(n_1+n_2)/\mathrm{Gl}n_1 \otimes \mathrm{Gl}n_2$. This has allowed the identification of Racah recoupling matrix elements in Gln with double coset matrix elements (DCME) in S_N and isoscalar factors in Gln with weighted double coset matrix element (WDCME) in S_N. As a consequence, results which follow from the representation theory of S_N (orthogonality of the matrix representations, adjointness with respect to the alternating group A_N, character theory, and the group orthogonality completeness condition) hold also for the Racah algebra of Gln. Duality shows the Racah factors of Gln are independent of the particular dimension n and requires symmetries (e.g., Regge's symmetries) that appear "hidden" when considered with respect to Gln only. Conversely, the DCME's and WDCME's of S_N are independent of the rank N to the extent equivalence relations in SUn may be invoked, and symmetries which follow from SUn appear hidden when considered with respect to S_N only. The basic combinatorial structure of the general Racah algebra is evidence by the universal occurrence of square roots of rational numbers for the numerical values of all recoupling or isoscalar factors. While developing the structural significance of the duality between S_N and Gln has been our main objective, completion of the work requires developing an evaluative scheme for the DCME and determining a phase convention showing which phases are fixed by consistency with the algebras of S_N and Gln and which phases can be set arbitrarily.

In this paper we develop two recursion relations expressing the DCME of a group in terms of equivalent elements of a subgroup. The first of these relations when used in S_N is the generalization of the Biedenharn, Elliot, and Racah sum rules of angular momentum theory.[2] When used in conjunction with the second relation it allows an interative scheme to be developed for evaluating a general DCME of S_N. The two relations are inverse to one another in the same sense that subduction is inverse to induction. We give some further examples of the nontrivial nature of these two relations including a derivation of the Racah formula[3] for evaluating the Clebsch–Gordan coefficients of SU2.

The simultaneous consideration of several groups and their irrep's requires an initial section on notation. A resume of some results from previous work is also given. The next section derives the recursion relations for a general group. This is particularized to S_N in the next section and the iterative evaluation of a general DCME is demonstrated. Other examples of these relations and some comments on their consequences for a phase convention are also given in this section. A complete phase convention remains to be established.

II. NOTATION AND DOUBLE COSET RESULTS

Any group G can be factored into disjoint double cosets (DC) with respect to two subgroups ${}_iG \backslash G / G_j$ (perhaps identical) as $G = \cup_q {}_iGqG_j$, where q are conveniently chosen (but fixed) double coset representatives (DCR). In bases symmetry adapted to the subgroup sequences (the symbol $\sim$ indicates an isomorphism)

$$G \begin{cases} G_j - q^{-1}\,{}_iGq \cap G_j \\ {}_iG - {}_iG \cap qG_jq^{-1} \end{cases};$$

on the left (bottom) and on the right (top) the double coset matrix element (DCME) can be represented as

$$\begin{bmatrix} \lambda & q \\ {}_i\lambda_i\lambda_j m & \lambda_j\ {}_i\lambda'_j m' \end{bmatrix} = \delta^{{}_i\lambda_j {}_i\lambda'_j}\delta^{mm'} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}_q, \tag{2.1}$$

a result which follows from the isomorphism between the subgroups

$${}_iG \cap qG_jq^{-1} = q(q^{-1}\,{}_iGq \cap G_j)q^{-1} \approx {}_iG_{jq}$$

and an application of Schur's lemma. Here λ with appropriate left or right subscripts labels irrep's of the respective groups. Because the DCR q is implied in the symbol

$$\begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix},$$

it will usually be omitted as a subscript except when a sum over the DC is to be taken. Because of the assumed unitarity of the matrix representation, the DCME are unitary on the indices $({}_i\lambda, \lambda_j)$ with the indices $(\lambda, {}_i\lambda_j)$ fixed. This is the

 0022-2488/80/020227-07$1.00

unitary transformation between the equivalent bases symmetry adapted to the two different subgroup sequences. Let $|\lambda|$ indicate the dimension of the irrep. and $|G|$ the order of the group. A weighting factor occurs frequently in what follows so we introduce the symbol

$$|\lambda\ {}_i\lambda\ {}_j\lambda\ \lambda_j| \equiv \frac{|\lambda|\,|{}_iG|\,|G_j|\,|{}_i\lambda_j|}{|G|\,|{}_i\lambda|\,|\lambda_j|\,|{}_iG_j|}. \tag{2.2}$$

Because of the group orthogonality-completeness condition, the weighted double coset matrix element (WDCME), defined as

$$|\lambda\ {}_i\lambda\ {}_i\lambda_j\ \lambda_j|^{1/2}\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{bmatrix} \equiv \begin{Bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{Bmatrix}, \tag{2.3}$$

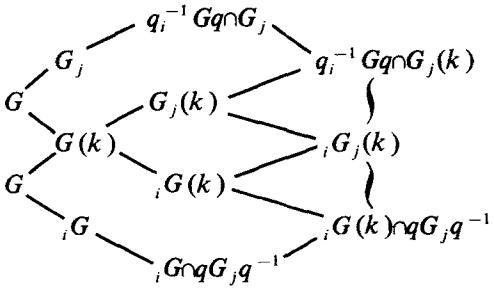
also possess unitary properties on the indices $(\lambda, {}_i\lambda_j q)$ with the indices $({}_i\lambda, \lambda_j)$ fixed. The WDCME is the unitary transformation between two induced representation known to be equivalent by Mackey's subgroup theorem.

To develop the relations of the next section a third subgroup $G(k)$ is introduced with the notation ${}_iG(k)$ $\equiv {}_iG \cap G(k)$ and $G_j(k) \equiv G(k) \cap G_j$ and a similar labeling of the irrep's. Fixed irrep. labels will be designated by λ and summed irrep. labels by ϵ. Limits are always implicit in the DCME's occuring in the sums.

III. EQUIVALENCE RELATIONS FOR DCME

A. By basis transformations in the irrep

Let $G(k)$ be a subgroup containing a given DCR q for which the matrix representation is block diagonal when symmetry adapted to $G(k) \backslash G / G(k)$. If the matrix is known in this basis, equivalence transformations allow the matrix to be evaluated in any other bases. The subgroup sequences of concern for the equivalence transformations are



The isomorphism between the final groups in the sequence follows from $q \in G(k)$. Since the corresponding DCME's accomplished the transformations among the equivalent bases, one has

$$\begin{aligned}\delta^{{}_i\lambda_j,{}_i\lambda_j}&\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{bmatrix}\\ &= \sum\begin{bmatrix}{}_i\lambda & {}_i\epsilon(k)\\ {}_i\lambda_j & {}_i\lambda_j(k)\end{bmatrix}\begin{bmatrix}\lambda & \epsilon(k)\\ {}_i\lambda & {}_i\epsilon(k)\end{bmatrix}\\ &\quad\times\begin{bmatrix}\epsilon(k) & \epsilon_j(k)\\ {}_i\epsilon(k) & {}_i\lambda_j(k)\end{bmatrix}\begin{bmatrix}\lambda & \lambda_j\\ \epsilon(k) & \epsilon_j(k)\end{bmatrix}\\ &\quad\times\begin{bmatrix}\lambda & {}_i\lambda_j'\\ \epsilon_j(k) & {}_i\lambda_j(k)\end{bmatrix}.\end{aligned} \tag{3.1}$$

We prefer the form with no explicit delta factors

$$\begin{aligned}&\begin{bmatrix}\lambda_j & \lambda_j(k)\\ {}_i\lambda_j & {}_i\lambda_j(k)\end{bmatrix}\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{bmatrix}\\ &\quad= \sum\begin{bmatrix}{}_i\lambda & {}_i\epsilon(k)\\ {}_i\lambda_j & {}_i\lambda_j(k)\end{bmatrix}\begin{bmatrix}\lambda & \epsilon(k)\\ {}_i\lambda & {}_i\epsilon(k)\end{bmatrix}\\ &\qquad\times\begin{bmatrix}\epsilon(k) & \lambda_j(k)\\ {}_i\epsilon(k) & {}_i\lambda_j(k)\end{bmatrix}\begin{bmatrix}\lambda & \lambda_j\\ \epsilon(k) & \lambda_j(k)\end{bmatrix}.\end{aligned} \tag{I}$$

This is the first equivalence relation which by itself does not provide an iterative procedure for evaluating a DCME due to the occurrence of similar terms in the sum which except for special choices are presumed unknown.

B. By projection operators

Matrix basis projectors defined by

$$e^{\lambda}_{mm'} \equiv \frac{|\lambda|}{G}\sum_{g\in G}\begin{bmatrix}\lambda & g^{-1}\\ m' & m\end{bmatrix}g \tag{3.2}$$

satisfy the orthogonality condition

$$e^{\lambda}_{mm'}\ e^{\lambda'}_{n'n} = \delta^{\lambda\lambda'}\delta^{m'n'}\ e^{\lambda}_{mn} \tag{3.3}$$

and, on DC decomposition, factor as

$$\begin{aligned}e^{\lambda}_{{}_i\lambda m,\,\lambda_j n} = &\sum_{\epsilon_j pq} e^{{}_i\lambda}_{m,\,{}_i\lambda_j p}\,q\,e^{\lambda_j}_{{}_i\lambda_j p,n}\\ &\times|\lambda\ {}_i\lambda\ {}_i\lambda_j\ \lambda_j|\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\epsilon_j\end{bmatrix}_q(|\,{}_i\epsilon_j\,|)^{-1}.\end{aligned} \tag{3.4}$$

By induction a matrix basis projector of a subgroup becomes a linear combination of matrix basis projectors in the higher group summed over all irrep.'s appearing in the induced representation, i.e.,

$$e^{{}_i\lambda}_{m,m'} = \sum e^{\epsilon}_{{}_i\lambda m,\,{}_i\lambda m'}. \tag{3.5}$$

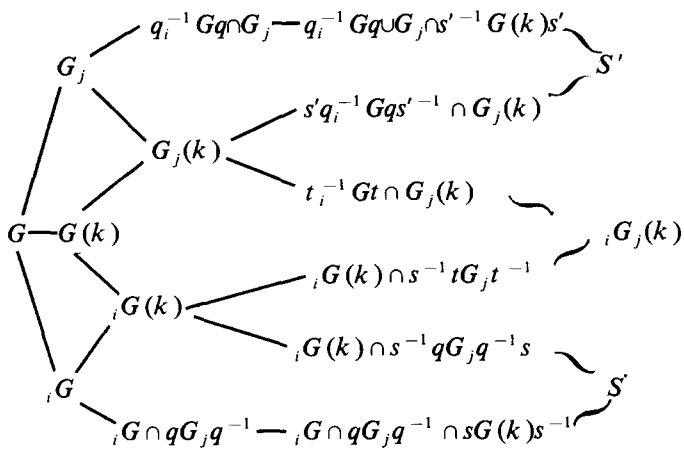
Consider the effect of a sequence (taken here as three) of projections in various subgroups

$$\begin{aligned}&e^{{}_i\lambda}_{{}_i\lambda_j m,\,\lambda(k)p_1}e^{\lambda(k)}_{\lambda(k)p_1,\lambda_j(k)p_2}e^{\lambda_j}_{\lambda_j(k)p_2,\,{}_i\lambda_j m'}\\ &\quad= \sum\begin{bmatrix}\epsilon & {}_i\lambda\\ \lambda(k) & {}_i\lambda(k)\end{bmatrix}e^{\epsilon}_{{}_i\lambda\,{}_i\lambda_j m,\lambda_j\,{}_i\lambda_j m'}\begin{bmatrix}\epsilon & \lambda(k)\\ \lambda_j & \lambda_j(k)\end{bmatrix},\end{aligned} \tag{3.6}$$

where DCME transformations have been used to symmetry adapt the induced middle projector to the subgroup sequences of the initial and final projectors. The intermediate subgroups and irrep.'s indicated by ${}_i\lambda(k)$ and $\lambda_j(k)$ need not be unique and are chosen so there will be some nonvanishing intersection as indicated by the DCME's on the rhs. One can identify the coefficients of identical elements on both sides of this expression (the matrix basis projectors are by construction orthogonal). In particular, the coefficient of the DCR q on the rhs is

$$\delta^{{}_i\lambda_j,{}_i\lambda_j}\delta^{mm'}\sum\frac{|\epsilon|}{|G|}\begin{bmatrix}\epsilon & {}_i\lambda\\ \lambda(k) & {}_i\lambda(k)\end{bmatrix}\begin{bmatrix}\epsilon & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{bmatrix}\begin{bmatrix}\epsilon & \lambda(k)\\ \lambda_j & \lambda_j(k)\end{bmatrix}.$$

Evaluation of the coefficient of q on the lhs requires care and details are presented in the Appendix. Here we note the subgroup sequences involved are



When consideration is restricted to DCR's for which $q = sts'$, the second and third subgroup sequences become identical as do the fourth and fifth subgroup sequences. Moreover, the final subgroups are all isomorphic $S \approx {}_iG_j(k) \approx S'$. the final result demonstrates an equivalence relation among induced representations, i.e.,

$$\sum \begin{Bmatrix} \epsilon & {}_i\lambda \\ \lambda(k) & {}_i\lambda(k) \end{Bmatrix} \begin{bmatrix} \epsilon & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{Bmatrix} \epsilon & \lambda(k) \\ \lambda_j & \lambda_j(k) \end{Bmatrix}$$
$$= \sum_{sts'=q} \begin{Bmatrix} {}_i\lambda & {}_i\lambda(k) \\ {}_i\lambda_j & {}_i\epsilon_j(k) \end{Bmatrix}_s \begin{bmatrix} \lambda(k) & \lambda_j(k) \\ {}_i\lambda(k) & {}_i\epsilon_j(k) \end{bmatrix}_t \times \begin{Bmatrix} \lambda_j & {}_i\lambda_j \\ \lambda_j(k) & {}_i\epsilon_j(k) \end{Bmatrix}_{s'}. \tag{3.7}$$

The DCR subscript is specifically indicated on the right because the sum is over ${}_i\epsilon_j(k)$ and all DCR for which $sts' = q$. A form more comparable to Eq. (I) is obtained by using the orthogonality of the WDCME to obtain

$$\begin{bmatrix} \lambda & \lambda(k) \\ \lambda_j & \lambda_j(k) \end{bmatrix} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}$$
$$= \sum |\lambda_j\, \lambda_j(k)\, {}_i\epsilon_j(k)\, {}_i\lambda_j| \begin{bmatrix} {}_i\lambda & {}_i\epsilon(k) \\ {}_i\lambda_j & {}_i\epsilon_j(k) \end{bmatrix}_s \times \begin{bmatrix} \lambda & \lambda(k) \\ {}_i\lambda & {}_i\epsilon(k) \end{bmatrix}_r \begin{bmatrix} \lambda(k) & \lambda_j(k) \\ {}_i\epsilon(k) & {}_i\epsilon_j(k) \end{bmatrix}_t \times \begin{bmatrix} \lambda_j & {}_i\lambda_j \\ \lambda_j(k) & {}_i\epsilon_j(k) \end{bmatrix}_{s'}. \tag{II}$$

The sum is over all intermediate irrep.'s and compatible DCR. Relations (I) and (II) are inverse to each other in the same sense subduction and induction are inverse processes. Substitution of either relation into the other leads to an identity.

IV. APPLICATION TO THE SYMMETRIC GROUP S_N

For DC decomposition of the symmetric group $\otimes S_{{}_iN} \backslash S_N / \otimes S_{N_j}$ we consider cases where the range of all subscripts is 2. The DC are in one to one correspondence with DC symbols

$$\begin{bmatrix} N & N_j \\ {}_iN & {}_iN_j \end{bmatrix}$$

such that all entries are nonegative integers. For a given DC symbol the DCR may be chosen as the $({}_1N_2)$th power of the cyclic permutation $C_{({}_2N_1 + {}_iN_2)}$ of ordered sets $\{{}_2N_1\}$ and $\{{}_1N_2\}$, i.e., $q = (C_{{}_2N_1 + {}_iN_2})^{{}_iN_2}$ causes the permutation

$$\begin{pmatrix} {}_1N_1 & {}_2N_2 & {}_1N_2 & {}_2N_2 \\ {}_1N_1 & {}_1N_2 & {}_2N_1 & {}_2N_2 \end{pmatrix}$$

all subsets maintaining a standard order. For the DC symbol

$$\begin{bmatrix} 7 & 4 & 3 \\ 5 & 3 & 2 \\ 2 & 1 & 1 \end{bmatrix}$$

with $S_{N_1} \otimes S_{N_2} = S_{1234} \otimes S_{567}$ and $S_{{}_iN} \otimes S_{{}_2N} = S_{12345} \otimes S_{67}$, one may choose the DCR $q = (456)^2 = (654)$. The intertwining subgroup $S_3 \otimes S_1 \otimes S_2 \otimes S_1$ of $\otimes S_{N_j}$ is $S_{123} \otimes S_4 \otimes S_{56} \otimes S_7$, which is isomorphic under conjungation by q to the subgroup $S_{123} \otimes S_6 \otimes S_{45} \otimes S_7$ of $\otimes S_{{}_iN}$. All matrix representations of S_N can be taken as orthogonal, which seems to be the reasonable (but not necessary) choice which we assume here. Since two equivalent orthogonal representations are related by an orthogonal transformation, all double coset matrices of S_N can be considered orthogonal. Orthogonality requires

$$\begin{bmatrix} \lambda & \lambda_i \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} = \begin{bmatrix} \lambda & {}_i\lambda \\ \lambda_j & {}_i\lambda_j \end{bmatrix}.$$

The irrep. label stands for a partition $\lambda \vdash N$ with order parts (rows) $l_i \geqslant l_{i+1}$. The rows or columns of a DCME must couple by the Littlewood–Richardson outer product rules.[4] In the particular case of a one part (node) removal we use the notation λ^i to signify the subduced representation $l^i_j = l_j - \delta_{ij}$. The orthogonality relations alone allow the evaluation [1(e)] of DCME of the form

$$\begin{bmatrix} \lambda & \lambda^i & 1 \\ \lambda^i & \lambda^{ij} & 1 \\ 1 & 1 & 0 \end{bmatrix} = \frac{1}{l_i - l^i_j + j - i} \tag{4.1}$$

or

$$\begin{bmatrix} \lambda & \lambda^j & 1 \\ \lambda^i & \lambda^{ij} & 1 \\ 1 & 1 & 0 \end{bmatrix} = \pm \left\{ \frac{(l_i - l^i_j + j - i + 1)(l_i - l^i_j + j - i - 1)}{(l_i - l^i_j + j - i)^2} \right\}^{1/2} = \pm |\lambda\, \lambda^i \lambda^{ij} \lambda^j|^{-1/2}. \tag{4.2}$$

The invariance of these values under the transformation $l_i \Rightarrow l_i + m, m$ a fixed integer, is an example of a hidden symmetry in S_N which becomes apparent under the identification of the DCME with a Racah recoupling transformation and the use of equivalence in SUn.

The subduction relation (I) is simply the generalization of the closure condition on transformation matrices that has led to a variety of sum rules in angular momentum theory. The induction relation (II) has no such counterpart. DCME's are identical to $(9-j)$ recoupling transformations, and WDCME's for irrep.'s with parts not exceeding the respective dimensions are identical to isoscalar factors [1(d),(e)] for $U(n_1 + n_2)/Un_1 \otimes Un_2$. One may express a recoupling transformation of $U(n_1 + n_2)$ in terms of recoupling trans-

formations of Un_1 and Un_2 and the appropriate isoscalar factors. This leads to a particular case of the induction relation (II). The induction relation is more general since there is no dimensional restriction on the irrep.'s.

An iterative procedure for the evaluation of a general DCME by sequential one node additions can be established. For the indicated choices, Eq. (II) becomes

$$\begin{bmatrix} \lambda & \lambda^{i} & 1 \\ \lambda_1 & \lambda_1^{j} & 1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}^2 = \sum_{k,m} \frac{|\lambda_1 \lambda_1^{j} \lambda_1^{jm} \lambda_1^{j}|}{(l_i - l_k^{i} + k - i)(l_{1j} - l_{1m}^{j} + m - j)} \times \begin{bmatrix} \lambda^{i} & \lambda^{ik} & 1 \\ \lambda_1^{j} & \lambda_1^{jm} & 1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}^2 + \frac{|\lambda_1|}{|\lambda_1^{j}| N_1}, \tag{4.3}$$

which allows an iterative buildup of this form from

$$\begin{bmatrix} \tau & \lambda_2 & 1 \\ 1 & 0 & 1 \\ \lambda_2 & \lambda_2 & 0 \end{bmatrix}^2 = 1. \tag{4.4}$$

(The unit value of this initial form is justified below. Thus

$$\begin{bmatrix} \lambda & \lambda^{i} & 1 \\ \lambda_1 & 1 & 1 \\ \lambda^{ij} & \lambda^{ij} & 0 \end{bmatrix}^2 = \frac{1}{2}\left[\frac{\phi(\lambda_1)}{l_i - l_j^{i} + j - i} + 1\right], \tag{4.5}$$

where $\phi(\lambda_1) = \pm 1$ for $\lambda_2 = [2]$ or $[1^2]$. Presuming DCME's of this form known, relation (I) with the indicated choices becomes

$$\begin{bmatrix} \lambda_1 & \lambda_1^{j} & 1 \\ {}_1\lambda_1 & {}_1\lambda_1^{k} & 1 \\ {}_2\lambda_1 & {}_2\lambda_1 & 0 \end{bmatrix}\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \end{bmatrix} = \sum \begin{bmatrix} {}_1\lambda & {}_1\epsilon & 1 \\ {}_1\lambda_1 & {}_1\lambda_1^{k} & 1 \\ {}_1\lambda_2 & {}_1\lambda_2 & 0 \end{bmatrix}\begin{bmatrix} \lambda & \epsilon & 1 \\ {}_1\lambda & {}_1\epsilon & 1 \\ {}_2\lambda & {}_2\lambda & 0 \end{bmatrix}\begin{bmatrix} \epsilon & \lambda_1^{j} & \lambda_2 \\ {}_1\epsilon & {}_1\lambda_1^{k} & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \end{bmatrix}\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \epsilon & \lambda_1^{j} & \lambda_2 \\ 1 & 1 & 0 \end{bmatrix}, \tag{4.6}$$

which, once a phase convention is established, allows iteration to evaluate any general DCME.

V. OTHER EXAMPLES

Relations (I) and (II) have an important bearing on establishing a consistent phase convention. For the indicated choices, relation (I) becomes

$$\begin{bmatrix} \lambda_1 & {}_1\lambda_1 & {}_2\lambda_1 \\ {}_1\lambda_1 & {}_1\lambda_1 & 0 \\ {}_2\lambda_1 & 0 & {}_2\lambda_1 \end{bmatrix}\begin{bmatrix} \lambda_2 & {}_2\lambda_2 & {}_1\lambda_2 \\ {}_1\lambda_2 & 0 & {}_1\lambda_2 \\ {}_2\lambda_2 & {}_2\lambda_2 & 0 \end{bmatrix}\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \end{bmatrix} = \sum \begin{bmatrix} \epsilon(1) & {}_1\lambda_1 & {}_2\lambda_2 \\ {}_1\lambda_1 & {}_1\lambda_1 & 0 \\ {}_2\lambda_2 & 0 & {}_2\lambda_2 \end{bmatrix}\begin{bmatrix} \epsilon(2) & {}_2\lambda_1 & {}_1\lambda_2 \\ {}_1\lambda_2 & 0 & {}_1\lambda_2 \\ {}_2\lambda_1 & {}_2\lambda_1 & 0 \end{bmatrix}$$
$$\times \begin{bmatrix} {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_1\lambda_1 & {}_1\lambda_1 & 0 \\ {}_1\lambda_2 & 0 & {}_1\lambda_2 \end{bmatrix}\begin{bmatrix} {}_2\lambda & {}_2\lambda_2 & {}_2\lambda_1 \\ {}_2\lambda_1 & 0 & {}_2\lambda_1 \\ {}_2\lambda_2 & {}_2\lambda_2 & 0 \end{bmatrix}\begin{bmatrix} \lambda & \epsilon(1) & \epsilon(2) \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_2 & {}_2\lambda_1 \end{bmatrix}\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \epsilon(1) & {}_1\lambda_1 & {}_2\lambda_2 \\ \epsilon(2) & {}_2\lambda_1 & {}_1\lambda_2 \end{bmatrix}. \tag{4.7}$$

Matrix elements of the identity in a given basis of the form

$$\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \lambda_1 & \lambda_1 & 0 \\ \lambda_2 & 0 & \lambda_2 \end{bmatrix}$$

have unit value. Matrix elements of the form

$$\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 \\ \lambda_1 & \lambda_1 & 0 \end{bmatrix}$$

are identical to a 3-j symbol and must have value $\phi(\lambda - \lambda_1 - \lambda_2) = \pm 1$ because of orthogonality. The above expression thus becomes

$$\begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \end{bmatrix} = \sum \phi({}_2\lambda - {}_2\lambda_1 - {}_2\lambda_2)\phi(\epsilon(2) - {}_1\lambda_2 - {}_2\lambda_1)\phi(\lambda_2 - {}_1\lambda_2 - {}_2\lambda_2) \begin{bmatrix} \lambda & \epsilon(1) & \epsilon(2) \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_2 & {}_2\lambda_1 \end{bmatrix}$$
$$\times \begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ \epsilon(1) & {}_1\lambda_1 & {}_2\lambda_2 \\ \epsilon(2) & {}_2\lambda_1 & {}_1\lambda_2 \end{bmatrix}. \tag{4.8}$$

The product of phases may be regarded as the matrix element of the DCR in a basis giving block diagonal form, while the other DCME's bring about the transformation to the alternate bases. A similar application of Eq. (I) gives the phase upon exchanging the second and third columns (similarly for the rows) as

$$\begin{bmatrix} \lambda & \lambda_2 & \lambda_1 \\ {}_1\lambda & {}_1\lambda_2 & {}_1\lambda_1 \\ {}_2\lambda & {}_2\lambda_2 & {}_2\lambda_1 \end{bmatrix} = \phi(\lambda - \lambda_2 - \lambda_1)\phi({}_1\lambda - {}_1\lambda_2 - {}_1\lambda_1)\phi({}_2\lambda - {}_2\lambda_2 - {}_2\lambda_1) \begin{bmatrix} \lambda & \lambda_1 & \lambda_2 \\ {}_1\lambda & {}_1\lambda_1 & {}_1\lambda_2 \\ {}_2\lambda & {}_2\lambda_1 & {}_2\lambda_2 \end{bmatrix}. \tag{4.9}$$

An obvious constraint on any phase convention is that it be consistent with the characteristic roots of the DCR in S_N. The phase $\phi(\lambda - \lambda_1 - \lambda_2)$ is in general not separable into a product of phases associated with each irrep. as is the case when consideration is restricted to bipartition (SU2) irrep.'s only. The first counter example is $\lambda_1 = [2,1] = \lambda_2$ with $\lambda = [3,2,1]$ occuring in the symmetrized square and the antisymmetrized square. In such cases it is assumed a phase convention $\phi(\lambda - \lambda^{[2]}_{[1]}) = +1$ and $\phi(\lambda - \lambda^{[1^2]}) = -1$ is possible.

As an example of another application of Eq. II, consider $\lambda = [N-M,M]$ a bipartition irrep. of S_N and let the other irrep.'s be the one dimensional symmetric irrep.'s of their respective groups. We wish to calculate the DCME

$$\begin{bmatrix} [N-M,M] & N_1 & N_2 \\ {}_1N & {}_1N_1 & {}_1N_2 \\ {}_2N & {}_2N_1 & {}_2N_2 \end{bmatrix}.$$

Choose $\lambda(k) = [m^2]\cdot[N-M-m,M-m]$, m arbitrary, for which ${}_i\lambda \uparrow \cap \lambda(k)\uparrow = \lambda = \lambda(k)\uparrow \cap \lambda_j\uparrow$. Equation (II) yields

$$\begin{bmatrix} [N-M,M] & N_1 & N_2 \\ {}_1N & {}_1N_1 & {}_1N_2 \\ {}_2N & {}_2N_1 & {}_2N_2 \end{bmatrix} = \sum_L \begin{bmatrix} [m^2] & m & m \\ m & [m-L] & L \\ m & L & [m-L] \end{bmatrix} \times \begin{bmatrix} [N-M-m,M-m] & [N_1-m] & [N_2-m] \\ [{}_1N-m] & [{}_1N_1-m+L] & [{}_1N_2-L] \\ [{}_2N-m] & [{}_2N_1-L] & [{}_2N_2-m+L] \end{bmatrix} \times \binom{{}_1N_1}{m-L}\binom{{}_1N_2}{L}\binom{{}_2N_1}{L}\binom{{}_2N_2}{m-L}\left[\binom{{}_1N}{m}\binom{{}_2N}{m}\binom{N_1}{m}\binom{N_2}{m}\right]^{-1/2}. \tag{4.10}$$

By choosing $m = M$, the problem is reduced to evaluating the DCME

$$\begin{bmatrix} [M^2] & M & M \\ M & [M-K] & K \\ M & K & [M-K] \end{bmatrix},$$

which by an application of the same result is given by

$$\begin{bmatrix} [M^2] & M & M \\ M & [M-K] & K \\ M & K & [M-K] \end{bmatrix} = \sum_L \begin{bmatrix} [m^2] & m & m \\ m & [m-L] & L \\ m & L & [m-L] \end{bmatrix} \times \begin{bmatrix} [(M-m)^2] & [M-m] & [M-m] \\ [M-m] & [M-m-K+L] & [K-L] \\ [M-m] & [K-L] & [M-m-K+L] \end{bmatrix} \times \left[\binom{M-K}{m-L}\binom{K}{L}\binom{M}{m}\right]^2. \tag{4.11}$$

By induction and using

$$\begin{bmatrix} [1^2] & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = -1$$

one obtains

$$\begin{bmatrix} [M^2] & M & M \\ M & [M-K] & K \\ M & K & [M-K] \end{bmatrix} = (-1)^K \binom{M}{K}^{-1} \tag{4.12}$$

and therefore

$$\begin{bmatrix} [N-M,M] & N_1 & N_2 \\ {}_1N & {}_1N_1 & {}_1N_2 \\ {}_2N & {}_2N_1 & {}_2N_2 \end{bmatrix} = \sum_L (-1)^L \binom{M}{L}^{-1} \binom{{}_1N_1}{M-L} \binom{{}_1N_2}{L} \binom{{}_2N_1}{L} \binom{{}_2N_2}{M-L}$$
$$\times \left[\binom{{}_1N}{M} \binom{{}_2N}{M} \binom{N_1}{M} \binom{N_2}{M} \right]^{-1/2}. \tag{4.13}$$

The WDCME of this element is an isoscalar factor identical to a Clebsch–Gordan coefficient of SU2, i.e.,

$$\left(\begin{matrix} {}_1N/2 & {}_2N/2 \\ \frac{({}_1N_1 - {}_1N_2)}{2} & \frac{({}_2N_1 - {}_2N_2)}{2} \end{matrix} \middle| \begin{matrix} \left(\frac{N}{2} - M\right) \\ \frac{(N_1 - N_2)}{2} \end{matrix} \right)_2 = \left[\frac{(N-2M+1)\,{}_1N!\,{}_2N!N_1!N_2!}{(N-M+1)!M!\,{}_1N_1!\,{}_1N_2!\,{}_2N_1!\,{}_2N_2!} \right]^{1/2}$$
$$\times \begin{bmatrix} [N-M,M] & N_1 & N_2 \\ {}_1N & {}_1N_1 & {}_1N_2 \\ {}_2N & {}_2N_1 & {}_2N_2 \end{bmatrix}$$

which agrees even in sign with previously developed sums for evaluating the Clebsch–Gordan coefficients of SU2.[3]

IV. CONCLUSION

We have established two sum relations for DCME or recoupling coefficients. They are sufficient, once a phase convention is established, to allow an iterative evaluation of any DCME. However, an iterative one node addition procedure may be very inefficient for evaluating a specific DCME. One may take advantage of all symmetries of the DCME's and WDCME's to reduce the number of steps relating a desired coefficient to other known coefficients. Association of the irrep.'s of S_N with respect to the alternating group A_N as a subgroup of index two and relations based on the character theory of S_N appear as hidden symmetries from the point of view of SUn. Equivalence of irrep.'s in SUn and complex conjugation appear as hidden symmetries from the point of view of S_N and lead one to consider interchanges of all rows and columns in a DCME. In this regard irrep.'s with rectangular Young patterns seem to have special properties.

There exists historic precedence for a phase convention in SU_2 (and a history of fixing arbitrary phases in different ways). A phase convention for general SUn must be certainly nonsimple.[5] Proposing a phase convention consistent with both S_N and SUn requires further study.

Note added in proof: All symmetries and a consistent phase convention are considered in a paper being prepared for publication.

APPENDIX: Evaluation of the coefficient of q on the lhs of Eq. (3.6)

Under DC decomposition the lhs of Eq. (3.6) expands as

$$\sum e^{{}_i\lambda_j}_{m,{}_i\epsilon_j(k)\,p_1} s\, e^{{}_i\lambda(k)}_{{}_i\epsilon_j(k)\,p_1,\,{}_i\epsilon_j(k)'\,p_2}\, t\, e^{\lambda_j(k)}_{{}_i\epsilon_j(k)'p_2,\,{}_i\epsilon_j(k)''\,p_3}\, s'$$
$$\times e^{{}_i\lambda_j}_{{}_i\epsilon_j(k)''\,p_3,m'} \left| {}_i\lambda_i\,\lambda_j\,{}_i\epsilon_j(k)\,{}_i\lambda(k) \right| \begin{bmatrix} {}_i\lambda & {}_i\lambda(k) \\ {}_i\lambda_j & {}_i\epsilon_j(k) \end{bmatrix}_s$$
$$\times \left| \lambda(k)\,{}_i\lambda(k)\,{}_i\epsilon_j(k)'\,\lambda_j(k) \right| \begin{bmatrix} \lambda(k) & \lambda_j(k) \\ {}_i\lambda(k) & {}_i\epsilon_j(k) \end{bmatrix}_t$$
$$\times \left| \lambda_j\,\lambda_j(k)\,{}_i\epsilon_j(k)''\,{}_i\lambda'_j \right| \begin{bmatrix} \lambda_j & {}_i\lambda'_j \\ \lambda_j(k) & {}_i\epsilon_j(k)'' \end{bmatrix}_{s'}$$
$$\times \left[\left| {}_i\epsilon_j(k) \right| \left| {}_i\epsilon_j(k)' \right| \left| {}_i\epsilon_j(k)'' \right| \right]^{-1}, \tag{A1}$$

where sums are to be taken over all ϵ, p, and DCR for which $sts' = q$. DC decompose the inner two projectors and restrict consideration to the DC with the identity as DCR. In general, such a restriction proceeds as follows:

$$e^{\lambda}_{\tau p,\tau' p'}\Big|_e = \sum e^{\tau}_{p,\epsilon p_1}\, e^{\tau'}_{\epsilon p_1, p'}\, \frac{|\lambda\tau\epsilon\tau'|}{|\epsilon|} \begin{bmatrix} \lambda & \tau' \\ \tau & \epsilon \end{bmatrix}$$
$$= \delta^{\tau\,\tau'} e^{\tau}_{p,p'}\, \frac{|\lambda|\,|T|}{|G|\,|\tau|}, \tag{A2}$$

where the last equality holds iff the subgroups $T \backslash G / T$ are identical as in the present consideration. Carrying the DCME's and weighting factors implicity and just writing down the operator portion, we now have

$$\sum e^{\lambda_j}_{m,{}_i\epsilon_j(k)\,p}\, sts'\, e^{\lambda'_j}_{{}_i\epsilon_j(k)\,p_1 m'}. \tag{A3}$$

Finally, since $sts' = q$, we may commute the remaining projectors with sts' (because of the isomorphism), sum on p, and restrict to the identity element to obtain the factor

$$\delta^{{}_i\lambda_j\,{}_i\lambda'_j} \delta^{mm'}\, \frac{|{}_i\lambda_j|\,|{}_i\epsilon_j(k)|}{|{}_iG_j|}.$$

Collecting all factors generated above and identifying the coefficient of q on the two sides of Eq. (3.6), we have

$$\sum \frac{|\epsilon|}{|G|} \begin{bmatrix} \epsilon & {}_i\lambda \\ \lambda(k) & {}_i\lambda(k) \end{bmatrix} \begin{bmatrix} \epsilon & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{bmatrix} \epsilon & \lambda(k) \\ \lambda_j & \lambda_j(k) \end{bmatrix}$$
$$= \sum \frac{|\lambda|\,|\lambda(k)\,{}_i\lambda(k)\,{}_i\epsilon_j(k)\,\lambda_j(k)|}{|G|\,|\lambda\,{}_i\lambda\,{}_i\lambda_j\,\lambda_j|}$$
$$\times \begin{bmatrix} {}_i\lambda & {}_i\lambda(k) \\ {}_i\lambda_j & {}_i\epsilon_j(k) \end{bmatrix}_s \begin{bmatrix} \lambda(k) & \lambda_j(k) \\ {}_i\lambda(k) & {}_i\epsilon_j(k) \end{bmatrix}_t$$
$$\times \begin{bmatrix} \lambda_j & {}_i\lambda_j \\ \lambda_j(k) & {}_i\epsilon_j(k) \end{bmatrix}_{s'}. \tag{A4}$$

Use of the unitarity on the WDCME removes the sum of the lhs and gives the expression (II).

[1]J.J. Sullivan, J. Math. Phys. (a) **14**, 387 (1973); (b) **16**, 756 (1975); (c) **16**, 1707 (1975); (d) **19**, 1674, 1681 (1978); and (e) Proceedings of the International Symposium on Mathematical Physics, Mexico City, 5–8, January 1976, Vol. I, p. 253.
[2](a) A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Princeton), 2nd edition, pp. 96, 103; (b) the original literature is collected in L.C. Biedenharn and H. Van Dam, *Quantum Theory of Angular Momentum* (Academic, New York, 1965).
[3]G. Racah, Phys. Rev. **62**, 438 (1942) contained in Ref. 2(b); also Eq. (3.6.11) of Ref. 2(a).
[4]D.E. Littlewood, *The Theory of Group Characters* (Clarendon, Oxford, 1950).
[5]P.H. Butler, Philos. Trans. R. Soc. London Ser. A **277**, 545 (1975).

Finite subgroups of the generalized Lorentz groups O(p,q)[a)]

J. Patera
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Quebec, Canada

Y. Saint-Aubin and H. Zassenhaus
Department of Mathematics, Ohio State University, Columbus, Ohio 43210

(Received 14 August 1979; accepted for publication 18 September 1979)

An algorithm is developed which permits us to construct all the finite subgroups of the generalized Lorentz groups O(p,q) up to an O(p,q) conjugation. The application of this algorithm to the Lorentz group O(3,1) is outlined, and the full list of finite subgroups compiled.

I. INTRODUCTION

Properties of the Lorentz group O(3,1) and its representations have been intensively studied for many years in view of the important role of the group in physics.[1] Continuous subgroups of the Lorentz group have been classified[2] a long time ago and have often been used since.

A classification of the finite subgroups of the Lorentz group is undertaken for the first time here, although, as we explain below, the problem has been almost solved on several occasions in a different context.

The purpose of this article is to bring together known general results concerning finite subgroups of the pseudo-orthogonal groups $O(p,q)$ and to formulate a computational procedure which provides the desired list of representatives of $O(p,q)$-conjugacy classes of finite subgroups of $O(p,q)$. Naturally, such an algorithm requires that the finite subgroups of $O(p)$ and $O(q)$ are known. As an illustration we solve the problem completely for the Lorentz group O(3,1).

Technically, a classification of finite subgroups of O(3,1) could proceed in several ways. A possible approach would be to recognize that our problem practically coincides with a problem of Shubnikov and Koptsik[3] and to complete their list of 122 antisymmetry point groups which are extentions of the crystallographic groups to the full list of the antisymmetry point groups acting in R^3. According to Lemma 1 below; every finite subgroup of O(3,1) is an O(3,1)-conjugate of a subgroup of O(4). Hence one could start from the known[4] subgroups of O(4). Still another way would be to use a continuous epimorphism of the group $SL(2,C)$ on the orthochronous Lorentz group $O_1(3,1)$.

In Sec. II two lemmas are formulated which lead to an algorithm for finding all the finite subgroups of the generalized Lorentz group $O(p,q)$ up to $O(p,q)$-conjugation starting from known finite subgroups of $O(p)$ and $O(q)$. In Sec. III the algorithm is applied to the Lorentz group. The last section contains general comments: the relation between the finite subgroups of O(3,1) and the Shubnikov point groups; a criterion which easily distinguishes to which of the five groups DO(3,1), SO(3,1), $O_1(3,1)$, $O_2(3,1)$, and O(3,1), that are locally isomorphic to O(3,1) any given finite subgroup G of O(3,1) minimally belongs; the relationship between the finite subgroups of $\langle SL(2,C), \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)\rangle$ and those of O(3,1); th generating functions and the corresponding integrity base for invariants and general covariants of all irreducible repre sentations of the finite subgroups.

II. THE GENERAL ALGORITHM

A. A first lemma

The apparent simplicity of the classification problem resides in the following misleading argument: Suppose that finite subgroup of O(3,1) contains a Lorentz-boost B, i.e., a O(3,1)-transformation which mixes the space and time coordinates. Since the order of the group must be finite, B^n must be the identity for some finite n. But this is impossibl since one knows physically that the repeated application o the same Lorentz-boost is equivalent to a Lorentz-boost wit a parameter $\gamma = v/c$ approaching 1. Hence a finite subgrou of O(3,1) cannot contain a Lorentz-boost and all its element will be of the form of a direct sum of an O(3)-rotation with possible inversion of the time ($t \to \pm t$).

The error in this reasoning lies in the fact that one has i mind a pure Lorentz-boost which mixes only one of the space coordinates with the time. But if the transformation makes simultaneously a Lorentz-boost and a rotation of th space, it is not impossible that B^n will be the identity for some finite n. In fact, the following O(3,1)-matrix:

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \alpha & \beta \\ 0 & \alpha & \beta^2 & \alpha\beta \\ 0 & -\beta & -\alpha\beta & -\alpha^2 \end{pmatrix},$$

$$\forall\, \alpha, \beta \in R, \text{ such that } \alpha^2 - \beta^2 = 1, \qquad (1)$$

is a Lorentz-boost of order 2, i.e., $B^2 = I$.

With this fact, how can one hope to find all the finite subgroups of O(3,1) with only those of O(3) and the time reflection group $\langle -I_1 \rangle$? (The brackets "$\langle \cdots \rangle$" mean "th group generated by the element(s)···".) The answer to this s given by the following lemma.

Lemma 1: *Let G be a finite subgroup of $O(p,q)$. Then there exists a regular $(p+q)\times(p+q)$ matrix $Y \in O(p,q)$ such that, for all the elements X of G, $Y^{-1}XY$ is of the form*

$$\begin{pmatrix} X'_1 & 0 \\ 0 & X'_2 \end{pmatrix},$$

[a)]Work supported in part by the National Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

 0022-2488/80/020234-06$1.00

where $X'_1 \in O(p)$ and $X'_2 \in O(q)$.

In other words, any finite subgroup of $O(p,q)$ is $O(p,q)$-conjugate to a finite subgroup of $O(p) \oplus O(q)$.

Proof: The group $O(p,q)$ is formed by the elements of the set

$$\{X \mid X \in R^{(p+q)\times(p+q)} \text{ such that } X^T K_{p,q} X = K_{p,q}\}, \tag{2}$$

where $K_{p,q} = I_p \oplus (-I_q)$.

Firstly form the following set $S(G)$:

$$S(G) = \{X \mid X = X^T \in R^{(p+q)\times(p+q)} \text{ such that } g^T X g = X, \forall g \in G\}. \tag{3}$$

By definition (2), $K_{p,q} \in S(G)$. The elements of this set are in one-to-one correspondence with the real quadratic forms $A(x) = x^T X x$ $(x \in R^{(p+q)\times 1})$, that are invariant under the action of G.

Note that

$$P_0 = \sum_{g\in G} g^T g \in S(G), \tag{4}$$

which is a consequence of the finite order of G. Moreover, P_0 is a positive definite form,

$$x^T P_0 x = \sum_{g\in G} x^T g^T I g x = \sum_{g\in G} (gx)^T I (gx) \geqslant 0, \quad \forall x \in R^{(p+q)\times 1}. \tag{5}$$

By a theorem of linear algebra, the two symmetric matrices $K_{p,q}$ and P_0, of which one is positive definite, can be simultaneously diagonalized by means of an equivalent transformation,

$$Z^T K_{p,q} Z \quad \text{and} \quad Z^T P_0 Z, \tag{6}$$

with a nonsingular matrix Z. Since the diagonal matrix $Z^T K_{p,q} Z$ has the same signature as $K_{p,q}$, it is possible upon replacement of Z by $Y = ZM$, where M is a suitable monomial matrix, to have $Y^T K_{p,q} Y = K_{p,q}$. Note that a matrix of degree $p+q$ is said to be monomial if there is precisely one nonzero coefficient in each row and column. And then

$$K_{p,q} = Y^T K_{p,q} Y \quad \text{and} \quad P'_0 = Y^T P_0 Y, \tag{7}$$

and $Y \in O(p,q)$. We claim that the $O(p,q)$-conjugate subgroup H obtained from G by calculating $Y^{-1} g Y$, $\forall g \in G$, is the subgroup we are looking for.

It is easy to show that $P'_0 \in S(H)$; that is a direct consequence of $P_0 \in S(G)$. Define the q following matrices P_{p+i}, $i = 1,\ldots,q$:

$$P_{p+i} = K_{p,q} + \frac{1}{\alpha_{p+i}} P'_0, \tag{8}$$

where α_{p+i} is the $(p+i)$th diagonal entry of P'_0. The P_{p+i} obviously belong to $S(H)$. These diagonal matrices have their first p diagonal entries positive and at least one zero diagonal element in the last q entries.

Since $P_{p+i} \in S(H)$, its null space,

$$\mathrm{Ker} P_{p+i} = \{x \mid x \in R^{(p+q)\times 1} \text{ such that } P_{p+i} x = 0\}, \tag{9}$$

is H-invariant. Further, since $\mathrm{Ker} P_{p+i}$ does not contain isotropic vectors ($x \neq 0$ but $x^T K_{p,q} x = 0$), the decomposition of $R^{(p+q)\times 1}$ in $\mathrm{Ker} P_{p+i}$ and in its $K_{p,q}$-orthogonal complement holds:

$$R^{(p+q)\times 1} = (\mathrm{Ker}\, P_{p+i}) \oplus (\mathrm{Ker}\, P_{p+i})^{\perp}, \tag{10}$$

where

$$(\mathrm{Ker}\, P_{p+i})^{\perp} = \{y \mid y \in R^{(p+q)\times 1} \text{ and } y^T K_{p,q} x = 0, \ \forall x \in \mathrm{Ker}\, P_{p+i}\}. \tag{11}$$

$(\mathrm{Ker}\, P_{p+i})^{\perp}$ is also invariant under the action of H.

Since the intersection of invariant subspaces is invariant, $\cap_{i=1}^{q} \{(\mathrm{Ker}\, P_{p+i})^{\perp}\}$ is invariant. However, the elements of this intersection are just the vectors whose q last coordinates are zero:

$$\bigcap_{i=1}^{q} \{(\mathrm{Ker}\, P_{p+i})^{\perp}\} = \{x \mid x \in R^{(p)\times 1} \oplus 0^{(q)\times 1}\}, \tag{12}$$

where $0^{(q)\times 1}$ means the null vector in $R^{(q)\times 1}$.

One could show in the same manner that

$$\{x \mid x \in 0^{(p)\times 1} \oplus R^{(q)\times 1}\} \tag{13}$$

is also H-invariant.

The invariance of the two subspaces is tantamount to saying that all the elements of H have the desired form:

$$h = \begin{pmatrix} X'_1 & 0 \\ 0 & X'_2 \end{pmatrix}, \text{ where } X'_1 \in O(p) \text{ and } X'_2 \in O(q), \quad \forall h \in H. \tag{14}$$

Thus Lemma 1 is proved. □

B. A second lemma

Although the preceding lemma provides a simplification in the search for finite subgroups of the $O(p,q)$ groups, one still needs an easy test to recognize whether or not two finite subgroups are $O(p,q)$-conjugate. This is achieved by the following lemma.

Lemma 2: Two subgroups G,H of $O(p) \oplus O(q)$ are $O(p,q)$-conjugate if and only if there is an isomorphism:

$$\alpha: G \Rightarrow H$$

of G on H for which:

$$g = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}, \quad a(g) = h = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix},$$

$g_1, h_1 \in O(p)$, *and* $g_2, h_2 \in O(q)$, *and*

$$\mathrm{Tr}(g_1) = \mathrm{Tr}(h_1) \quad \textit{and} \quad \mathrm{Tr}(g_2) = \mathrm{Tr}(h_2).$$

Proof: If there is an isomorphism of G on H of the type described above, then the traces of the upper block g_1 and $h_1 = [\alpha(g)]_1$ are equal, so that these blocks form equivalent representations of the abstract group $\mathcal{G}$. There exists A in $O(p)$ such that

$$h_1 = [\alpha(g)]_1 = A^{-1} g_1 A, \quad \forall g \in G. \tag{15}$$

For the same reason, there exists B in $O(q)$ such that

$$h_2 = [\alpha(g)]_2 = B^{-1} g_2 B, \quad \forall g \in G. \tag{16}$$

Hence

$$h = \alpha(g) = (A \oplus B)^{-1} g (A \oplus B) \tag{17}$$

and $A \oplus B \in O(p,q)$.

Conversely, let G and H be $O(p,q)$-conjugate subgroups of $O(p) \oplus O(q)$ such that $H = A^{-1}GA$ for some $A \in O(p,q)$. Then there is a finite set of nonequivalent unitary irreducible representations $\varphi_1, \varphi_2, \ldots, \varphi_\nu$ of the abstract group $\mathcal{G}$ such that:

$$G \ni g = \begin{pmatrix} G_1(g) & 0 \\ 0 & G_2(g) \end{pmatrix}, \quad G_i \sim \overset{\nu}{\underset{k=1}{\oplus}} a_{ik}\varphi_k ,$$

$$H \ni A^{-1}gA = \begin{pmatrix} H_1(g) & 0 \\ 0 & H_2(g) \end{pmatrix}, \quad H_i \sim \overset{\nu}{\underset{k=1}{\oplus}} b_{ik}\varphi_k , \tag{18}$$

for $i = 1,2$. Here the multiplicities a_{ik}, b_{ik} are nonnegative integers for $i = 1,2$, $k = 1,2,\ldots,\nu$.

It is our aim to show that G_i and H_i are equivalent for $i = 1,2$, i.e., $a_{ik} = b_{ik}$ for $i = 1,2,\ldots,\nu$. Without loss of generality, we may assume that the following equalities already hold:

$$G_i = \overset{\nu}{\underset{k=1}{\oplus}} a_{ik}\varphi_k, \quad H_i = \overset{\nu}{\underset{k=1}{\oplus}} b_{ik}\varphi_k (i = 1,2) . \tag{19}$$

We want to show that $G_i = H_i$ for $i = 1,2$.

Now let V_{φ_k} be the R-linear subspace of the $(p+q)$-column space $R^{(p+q)\times 1}$ which is invariant under the action of $G_1 \oplus G_2$ such that it is maximal among the $G_1 \oplus G_2$-invariant supspaces of the $(p+q)$-column space with the property that every irreducible component is equivalent to φ_k. It is known from general representation theory that V_{φ_k} is uniquely determined by φ_k and it gives rise to a representation of G that is equivalent to $(a_{1k} + a_{2k})\varphi_k$. In fact, it is formed by all $(p+q)$-columns with the property that all jth coefficients are zero unless:

$$\sum_{h<k} a_{1h}[\varphi_h] < j \leqslant \sum_{h\leqslant k} a_{1h}[\varphi_h] ,$$

or

$$\sum_{h<k} a_{2k}[\varphi_h] < j - p \leqslant \sum_{h\leqslant k} a_{2h}[\varphi_h] , \tag{20}$$

where $[\varphi_h]$ denotes the degree of the representation φ_h. The signature of the restriction of $K_{p,q}$ to V_{φ_k} is equal to $(a_{1k}[\varphi_k], a_{2k}[\varphi_k])$. Similarly we define the corresponding subspace V'_{φ_k} for $H_1 \oplus H_2$ and find the signature of the restriction of $K_{p,q}$ to V'_{φ_k} to be equal to $(b_{1k}[\varphi_k], b_{2k}[\varphi_k])$. But because of the $O(p,q)$-conjugacy of $G_1 \oplus G_2$, $H_1 \oplus H_2$ that was assumed above, it follows that $V'_{\varphi_k} = AV_{\varphi_k}$ so that

$$a_{ik} = b_{ik} \quad (i = 1,2) , \tag{21}$$

for $k = 1,2,\ldots,\nu$.

These equalities are tantamount to the equality of the traces of the upper blocks g_1 and $h_1 = [\alpha(g)]_1$ and of the lower blocks g_2 and $h_2 = [\alpha(g)]_2$:

$$\mathrm{Tr}(g_1) = \mathrm{Tr}(h_1) \text{ and } \mathrm{Tr}(g_2) = \mathrm{Tr}(h_2) ,$$
$$\forall g \in G \quad \text{and} \quad h = \alpha(g) \in H.$$

□

C. An algorithm

The two lemmas lead to a natural algorithm for finding the finite subgroups of $O(p,q)$ [up to $O(p,q)$-conjugation] starting from the known finite subgroups of $O(p)$ and $O(q)$.

Let FSG[O(m)] be a list of representatives of conjugacy classes [under O(m)] of the finite subgroups of the group O(m) of orthogonal $m \times m$ matrices. A list FSG[O(p_1, p_2)] of the representatives of the conjugacy classes [under O(p_1, p_2)] of the finite subgroups of the group O(p_1, p_2) of pseudo-orthogonal $(p+q)\times(p+q)$ matrices will be given by the following procedure:

1. Pick out of the FSG[O(p_i)], $i = 1,2$, two members X_i (one for each p_i), construct the "natural" representation ψ_i,

$$\psi_i : X_i \to O(p_i) , \tag{22}$$

and the subgroup of the group $\mathrm{Aut}(X_i)$ of automorphisms of X_i which preserve the character of the representation ψ_i,

$$\mathrm{Aut}_{\psi_i}(X_i) = \{\alpha | \alpha \in \mathrm{Aut}(X_i) \text{ and } \forall x \in X_i : \mathrm{Tr}(x) = \mathrm{Tr}(\alpha x)\} . \tag{23}$$

2. Form a list of the representatives of the conjugacy classes under $\mathrm{Aut}_{\psi_1}(X_1) \times \mathrm{Aut}_{\psi_2}(X_2)$ of the subgroups

$$\hat{G} = \{\Delta_1(g) \times \Delta_2(g) = g | \Delta_i(g) \in X_i, i = 1,2\} , \tag{24}$$

of the direct product $X_1 \times X_2$ such that $\hat{G}$ satisfies the semi-direct product condition,

$$\Delta_i(\hat{G}) = X_i \quad i = 1,2 . \tag{25}$$

3. Form the (finite) list SG(X_1, X_2) of the subgroups of O(p_1, p_2) obtained from $\hat{G}$,

$$G = \{\Delta_1(g) \oplus \Delta_2(g) | g \in \hat{G}\} . \tag{26}$$

4. The union of the lists SG(X_1, X_2) for all pairs $X_1 \in$ FSG[O(p_1)] and $X_2 \in$ FSG[O(p_2)] gives the desired list FSG[O(p_1, p_2)].

III. AN EXAMPLE: FSG[O(3,1)]

The application of this algorithm to the Lorentz group O(3,1) supposes that one already has the lists FSG[O(1)] $\equiv$ FSG[$\langle -I_1 \rangle$] and FSG[O(3)]. In fact, FSG[$\langle -I_1 \rangle$] contains only two elements, $\{(1)\}$ and $\langle -I_1 \rangle = \{(1), (-1)\}$, and FSG[O(3)] is well known. First let us explain how to obtain FSG[O(p)] for p odd, from the known FSG[SO(p)] because the procedure is exactly parallel to the problem of finding FSG[O(3,1)] from the known FSG[O(3)].

A. The list FSG[O(3)]

Any subgroup G of O(p) for p odd is an element of one of the following disjoint sets: (i) the set of subgroups of SO(p), (ii) the set of subgroups of O(p) containing $-I_p$, and (iii) the set of the remaining subgroups of O(p). In the second case (ii), there is the decomposition $G = [G \cap \mathrm{SO}(p)] \times \langle -I_p \rangle$ of G into the direct product of its intersection with SO(p) and the center of O(p). In the third case (iii), we can obtain from $\tilde{G}$ a subgroup $G \subset \mathrm{SO}(p)$, i.e., a subgroup belonging to the set (i), by a Goursat twist[4]:

$$G = \{g | g \in \tilde{G}\,\mathrm{SO}(p) \text{ or } g \in -I_p[\tilde{G} \setminus (\tilde{G} \cap \mathrm{SO}(p))]\} . \tag{27}$$

The intersection $\tilde{G} \cap SO(p)$ is an invariant subgroup of $\tilde{G}$ of index 2. Conversely, from any subgroup G of $SO(p)$ which has a subgroup N of index 2, we can construct a Goursat twist $\tilde{G}$ in the following way:

$$\tilde{G} = \{g | g \in N \text{ or } g \in -I_p \cdot (G \setminus N)\}. \tag{28}$$

If p is odd, it is easy to show that $O(p)$-conjugacy is equivalent to $SO(p)$-conjugacy. As a corollary, two subgroups of the set (ii) [or (iii)] will be $O(p)$-conjugate if and only if their intersection with $SO(p)$ is $SO(p)$-conjugate.

Consequently a representative list FSG[O(p)] (for p odd) is derived from a representative list FSG[SO(p)] upon extending FSG[SO(p)] with the list of the direct products $G \times CO(p)$ [$CO(p)$ is the center of $O(p)$, i.e., $\langle -I_p \rangle$, and G ranges over FSG[SO(p)]] and with the list of the Goursat twist subgroups of the form (28), where $G \in FSG$ [SO(p)] contains a subgroup N of index 2 and N ranges over a representative set of the conjugacy classes of the subgroups of G of index 2 under the normalizer $\mathrm{Nor}_{SO(p)} G$ of G in $SO(p)$.

Explicitly, the list FSG[SO(3)] is known to be formed by:

(a) The cyclic rotation groups $C_n = \langle R_n \rangle$ for $n \in N^+$ of order n, where

$$R_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\frac{2\pi}{n} & \sin\frac{2\pi}{n} \\ 0 & -\sin\frac{2\pi}{n} & \cos\frac{2\pi}{n} \end{pmatrix},$$

which satisfies the relation $R_n^n = I_3$.

(b) The dihedral rotation groups $D_n = \langle R_n, D \rangle$ for $2 \leqslant n \in N^+$ of order $2n$, where

$$D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

The generators satisfy the relations

$$D^2 = (DR_n)^2 = I_3;$$

(c) The tetrahedral rotation group $T = \langle R_2, D, R'_3 \rangle$ of order 12, where

$$R'_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

which satisfies $(DR'_3)^3 = R'^3_3 = I_3$;

(d) The octahedral rotation group $O = \langle R_2, D, R'_3, R'_2 \rangle = \langle R'_3, R'_2 \rangle$ of order 24, where

$$R'_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

which satisfies $(R'_2)^2 = (R'_2 R'_3)^4 = I_3$.

(e) The icosahedral rotation group $I = \langle R_2, R''_3 \rangle$ of order 60, where

$$R''_3 = \begin{pmatrix} \alpha & \beta & 0 \\ -\beta & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} R'_3 \begin{pmatrix} \alpha & -\beta & 0 \\ \beta & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with

$$\alpha = \frac{1}{2}\left(\sqrt{2 + \sqrt{2\sqrt{5} - 2}}\right)$$

and

$$\beta = -\sqrt{1 - \alpha^2},$$

subject to $(R''_3)^3 = (R_2 R''_3)^5 = I_3$.

A list FSG[O(3)] is obtained, as mentioned above, by extending FSG[SO(3)] by the direct products $G \times CO(3)$, $G \in$FSG[SO(3)] and the Goursat twists. These Goursat twist groups are

$\langle -R_n \rangle$ for n even $\geqslant 2$,	obtained from the cyclic groups,
$\langle R_n, -D \rangle$ for $n \geqslant 2$,	obtained from the
$\langle -R_n, D \rangle$ for n even $\geqslant 4$,	dihedral groups,
$\langle -R'_2, R'_3 \rangle$	obtained from the octahedral group.

B. The list FSG[O(3,1)]

To follow the algorithm presented in the preceding section, one must pick $X_1 \in$FSG[O(3)] and $X_2 \in$FSG[$\langle -I_1 \rangle$]. Since FSG[$\langle -I_1 \rangle$] has only two elements, the elements of FSG[O(3,1)] will be groups either with no elements reversing time [$X_2 = \{(1)\}$], or with some elements reversing it ($X_2 = \langle -I_1 \rangle$). The second step of the algorithm consists in constructing the subdirect products of X_1, X_2. Thus the groups with some time reversing elements can be distinguished as the groups which are the direct products of X_1 and $X_2 = \langle -I_1 \rangle$ and the groups which are their subdirect products but not their direct products. In consequence of this procedure, the list FSG[O(3,1)] is partitioned as follows:

(i) $X_1 \oplus 1$, where $X_1 \in$FSG[O(3)] and 1 is the identity acting on the time coordinate,

(ii) $X_1 \oplus \langle (-1) \rangle$, the direct product $X_1 \times \langle -I_1 \rangle$, and

(iii) $[(X_1 \setminus N) \oplus (-1)] \cup [N \oplus (1)]$, where N is a representative of a $\mathrm{Nor}_{O(3)} X_1$-conjugacy classes of subgroups of X_1 of index 2 (see below): The subgroups (iii) are the subdirect products and correspond to the Goursat twists.

Let us give an example of the procedure. Let $X_1 = \langle R_n, D, -I_3 \rangle$, the group obtained by the direct product of D_n with the space inversion. If one takes $X_2 = \{(1)\}$, one obtains $\langle R_n, D, -K_{3,1} \rangle$. [When the $A \in O(3)$ is used to describe a generator of O(3,1), it must be understood as $A \oplus 1$, $-A$, and $-I_4 A$ denote $(-I_3 A) \oplus (+1)$ and $-I_4$ $(A \oplus (+1))$ respectively.] In the same way, if $X_2 = \langle -I_1 \rangle$ and one keeps all the elements of the direct product, one gets $\langle R_n, D, -K_{3,1} \rangle \times \langle -I_4 \rangle$.

The subgroup $X_1 = \langle R_n, D, -I_3 \rangle$ leads to many twists. For any n, one has at least three subgroups of index 2: $\langle R_n, D \rangle$, $\langle R_n, -I_3 \rangle$, and $\langle R_n, -D \rangle$. When n is odd, those three are the only ones. For n even, R_n may not appear as a generator and four other subgroups are possible: $\langle R_{n/2}, D, -I_3 \rangle$, $\langle -R_n, D \rangle$, $\langle R_{n/2}, DR_n, -I_3 \rangle$, and $\langle -R_n, -D \rangle$. (It can be seen that those seven subgroups exhaust the possibilities for n even; in fact, each of them corresponds to a way of removing at least one of the three

generators R_n, D, and $-I_3$ of the elements of the initial group. For example, none of the three generators appears in the last example given above.)

Since $\mathrm{Aut}_{\psi_2}(X_2)$ [see relation (23)] contains only the identity, $\mathrm{Aut}_{\psi_1}(X_1)\times\mathrm{Aut}_{\psi_2}(X_2)$ is isomorphic to $\mathrm{Aut}_{\psi_1}(X_1)$, which is isomorphic to a subgroup of $\mathrm{Nor}_{SO(3)}X_1$. The second step of the algrorithm is then equivalent to finding which of the seven aforementioned subgroups belong to distinct conjugacy classes under $\mathrm{Nor}_{SO(3)}X_1$. One can find $\mathrm{Nor}_{SO(3)}\langle R_n, D, -I_3\rangle$ as follows: The centralizer of the dihedral groups D_n ($n\geqslant 4$; $n=2$ must be studied seperately) is $\langle -I_3\rangle$ in O(3) and simply $\langle I_3\rangle$ in SO(3). But the centralizer of a subgroup is an invariant subgroup of its normalizer and the factor group is isomorphic to a group of automorphisms of the subgroup. Since the D_n are finite, all groups of automorphisms of D_n are also finite. Then $\mathrm{Nor}_{SO(3)}D_n$ must be finite and equivalent to a member of FSG[SO(3)]. Morever it contains D_n. If $R_m\in\mathrm{Nor}_{SO(3)}\langle R_n, D\rangle$, the relation $R_m^{m-1}DR_m = DR_m^2 = DR_n^i$ must hold for a given i. The greatest value of m is then $m=2n$ and $\mathrm{Nor}_{SO(3)}\langle R_n, D\rangle = \langle R_{2n}, D\rangle$. Hence $\mathrm{Nor}_{SO(3)}\langle R_n, D, -I_3\rangle = \langle R_{2n}, D\rangle$. The first three subgroups are not conjugate and they lead to the three twists: $\langle R_n, D, -I_4\rangle$, $\langle R_n, -K_{3,1}, -I_4 D\rangle$, and $\langle R_n, -D, -I_4\rangle$. The sixth and seventh ones are respectively conjugates of the fourth and fifth ones by the matrix R_{2n}. Hence $X_2 = \langle R_n, D, -I_3\rangle$ leads to two other subgroups of the type (iii) for n even $\geqslant$: $\langle R_n \oplus (-1), D, -K_{3,1}\rangle$, $\langle -R_n, D, -I_4\rangle$. The procedure is then finished for this pair (X_1, X_2).

Table I gives the list of representatives of FSG[O(3,1)] in terms of their generators. Each column represents one type [(i), (ii), or (iii)].

IV. CONCLUDING REMARKS

The first observation is that the representative list FSG[O(3,1)] becomes finite if we restrict the representative list FSG[O(3)] to the crystallographic groups. In fact, we must omit the icosahedral groups and all the cyclic and dihedral groups with $n\neq 2,3,4$, and 6. In this case our list will present the 32 crystallographic groups (first column), 32 direct products of the 32 crystallographic groups with the group $\langle I_3 \oplus (-1)\rangle$ (time inversion) (second column), and 58 Goursat twists (third column), i.e., 122 crystallographic groups where the time is allowed to be reversed. These 122 groups are in fact the two-color crystallographic groups or the Shubnikov point groups. This knowledge enables us to represent the elements of the list FSG[O(3,1)] by means of the Shubnikov two-color diagrams.[3]

TABLE I: The Representative List FSG[O(3,1)]. The items in curly brackets are present only when n is even.

	(i) $G\subset O(3)\oplus 1$		(ii) $G\times\langle I_3\oplus(-1)\rangle$		(iii) Goursat twist	
Cyclic Groups	$\langle R_n\rangle$	$n\geqslant 1$	$\langle R_n\rangle\times\langle K_{3,1}\rangle$	$n\geqslant 1$	$\{\langle R_n\oplus(-1)\rangle\}$	$n\geqslant 2$
	$\langle R_n\rangle\times\langle -K_{3,1}\rangle$	$n\geqslant 1$	$\langle R_n\rangle\times\langle -I_3\rangle\times\langle K_{3,1}\rangle$		$\langle R_n\rangle\times\langle -I_4\rangle$	$n\geqslant 1$
					$\{\langle R_n\oplus(-1)\rangle\times\langle -K_{3,1}\rangle\}$	$n\geqslant 2$
					$\{\langle -R_n\rangle\times\langle -I_4\rangle\}$	$n\geqslant 2$
	$\{\langle -R_n\rangle\}$	$n\geqslant 2$	$\{\langle -R_n\rangle\times\langle K_{3,1}\rangle\}$	$n\geqslant 2$	$\{\langle -I_4 R_n\rangle\}$	$n\geqslant 2$
Dihedral Groups	$\langle R_n, D\rangle$	$n\geqslant 2$	$\langle R_n, D\rangle\times\langle K_{3,1}\rangle$	$n\geqslant 2$	$\langle R_n, D\oplus(-1)\rangle$	$n\geqslant 2$
					$\{\langle R_n\oplus(-1), D\rangle\}$	$n\geqslant 4$
	$\langle R_n, D\rangle\times\langle -K_{3,1}\rangle$	$n\geqslant 2$	$\langle R_n, D\rangle\times\langle K_{3,1}\rangle\times\langle -I_4\rangle$	$n\geqslant 2$	$\langle R_n, D\rangle\times\langle -I_4\rangle$	$n\geqslant 2$
					$\langle R_n, -I_4 D\rangle\times\langle -K_{3,1}\rangle$	$n\geqslant 2$
					$\langle R_n, -D\rangle\times\langle -I_4\rangle$	$n\geqslant 2$
					$\{\langle R_n\oplus(-1), D\rangle\times\langle -K_{3,1}\rangle\}$	$n\geqslant 4$
					$\{\langle -R_n, D\rangle\times\langle -I_4\rangle\}$	$n\geqslant 4$
	$\langle R_n, -D\rangle$	$n\geqslant 2$	$\langle R_n, -D\rangle\times\langle K_{3,1}\rangle$	$n\geqslant 2$	$\langle R_n, -I_4 D\rangle$	$n\geqslant 2$
					$\{\langle R_n\oplus(-1), -D\rangle\}$	$n\geqslant 2$
	$\{\langle -R_n, D\rangle\}$	$n\geqslant 4$	$\{\langle -R_n, D\rangle\times\langle K_{3,1}\rangle\}$	$n\geqslant 4$	$\{\langle -R_n, D\oplus(-1)\rangle\}$	$n\geqslant 4$
					$\{\langle D, -I_4 R_n\rangle\}$	$n\geqslant 4$
					$\{\langle D\oplus(-1), -I_4 R_n\rangle\}$	$n\geqslant 4$
Tetrahedral Groups	$\langle R_2, D, R'_3\rangle$		$\langle R_2, D, R'_3\rangle\times\langle K_{3,1}\rangle$			
	$\langle R_2, D, R'_3\rangle\times\langle -K_{3,1}\rangle$		$\langle R_2, D, R'_3\rangle\times\langle K_{3,1}\rangle\times\langle -I_4\rangle$		$\langle R_2, D, R'_3\rangle\times\langle -I_4\rangle$	
Octahedral Groups	$\langle R'_2, R'_3\rangle$		$\langle R'_2, R'_3\rangle\times\langle K_{3,1}\rangle$		$\langle R'_2\oplus(-1), R'_3\rangle$	
	$\langle R'_2, R'_3\rangle\times\langle -K_{3,1}\rangle$		$\langle R'_2, R'_3\rangle\times\langle K_{3,1}\rangle\times\langle -I_4\rangle$		$\langle R'_2, R'_3\rangle\times\langle -I_4\rangle$	
					$\langle -R'_2, R'_3\rangle\times\langle -I_4\rangle$	
					$\langle R'_2\oplus(-1), R'_3\rangle\times\langle -K_{3,1}\rangle$	
	$\langle -R'_2, R'_3\rangle$		$\langle -R'_2, R'_3\rangle\times\langle K_{3,1}\rangle$		$\langle -I_4 R'_2, R'_3\rangle$	
Icosahedral Groups	$\langle R_2, R''_3\rangle$		$\langle R_2, R''_3\rangle\times\langle K_{3,1}\rangle$			
	$\langle R_2, R''_3\rangle\times\langle -K_{3,1}\rangle$		$\langle R'_2, R''_3\rangle\times\langle K_{3,1}\rangle\times\langle -I_4\rangle$		$\langle R_2, R''_3\rangle\times\langle -I_4\rangle$	

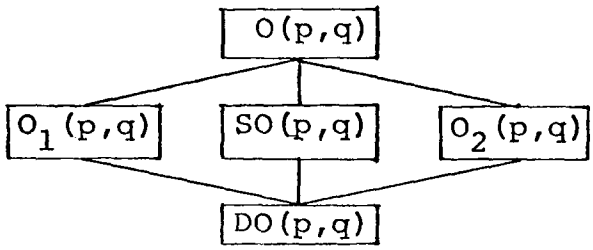


FIG. 1. Hasse diagram for O(p,q). DO(p,q), SO(p,q), $O_1(p,q)$, $O_2(p,q)$, and O(p,q) are respectively the identity component or derived group, the group of $(p+q)\times(p+q)$ real matrices with determinant det $= +1$, the group of transformations with the spinor norm[5] spn $= +1$, the group with det spn $= +1$, and the full pseudo-orthogonal group.

As a second observation, we want to classify the finite subgroups of O(3,1) with respect to which of the five groups DO(3,1), SO(3,1), $O_1(3,1)$, $O_2(3,1)$, O(3,1) they minimally belong. The commutator group of any generalized orthogonal group coincides with the identity component. The factor group of any generalized Lorentz group over its identity component is a Klein 4 group with the Hasse diagram of the intermediary subgroups (see Fig. 1).

Each subgroup G of O(p,q) generates together with DO(p,q) one of the five groups diagrammed above. In particular for each member of FSG[O(3,1)], we find that:

(i) $G \subset$ DO(3,1), if $G = H \oplus 1$ where $H \in$ FSG[SO(3)],

(ii) $\langle G, \mathrm{DO}(3,1)\rangle = \mathrm{O}_1(3,1)$, if $G = H \oplus 1$ where $H \notin$ FSG[SO(3)] but $\in$ FSG[O(3)],

(iii) $\langle G, \mathrm{DO}(3,1)\rangle = \mathrm{SO}(3,1)$ if not all the generators of G are expressible in the form $X \oplus 1$ [$X \in$ O(3)] but all these are of determinant $+1$,

(iv) $\langle G, \mathrm{DO}(3,1)\rangle = \mathrm{O}_2(3,1)$, if not all the generators of G are of the form $X \oplus 1$ but all are either of the form $X \oplus 1$ with $X \in$ SO(3) or $X \oplus (-1)$ with $X \in -I_3$ SO(3),

(v) $\langle G, \mathrm{DO}(3,1)\rangle = \mathrm{O}(3,1)$, otherwise.

Thirdly, we want to point out that our list FSG[O(3,1)] can be used to determine subgroups of

$$\left\langle \mathrm{SL}(2,C), \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\rangle .$$

Observe first that the following decomposition of the Lorentz group into the direct product of the subgroup $O_1(3,1)$ and of the center holds

$$\mathrm{O}(3,1) = \mathrm{O}_1(3,1) \times \langle -I_4 \rangle .$$

Moreover, there is a continuous epimorphism (a surjective homomorphism),

$$\epsilon : \left\langle \mathrm{SL}(2,C), \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\rangle \mapsto \mathrm{O}_1(3,1) ,$$

of the group formed by the complex matrices of dimension 2 and determinant ± 1 on the subgroup $O_1(3,1)$ of the Lorentz subgroup. The epimorphism ϵ is a local isomorphism. Its kernel is the center of SL(2,C) which is generated by $-I_2$. The subgroup SL(2,C) is mapped by ϵ on the identity component of the Lorentz group. Making use of ϵ, our classification also leads to a classification of the finite subgroup conjugacy classes of the factor group of

$$\left\langle \mathrm{SL}(2,C), \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\rangle .$$

over the center $\langle -I_2 \rangle$.

Finally, it should be noted that the generating functions and the integrity bases for all the irreducible representations of the elements of FSG[O(3,1)] are already calculated. When the finite subgroup G lies in the first set (i), it is isomorphic to the subgroup $G' \in$ FSG[O(3)] from which it has been obtained. In this case, Patera, Sharp, and Winternitz[6] have calculated all the generating functions and the integrity bases for all the irreducible representations. When G lies in the second set (ii), the generating functions are obtained from those of $G' \in$ FSG[O(3)] ($G = G' \times \langle -I_4 \rangle$) using the relations (34)–(37) of Ref. 6. Then, in the case (iii), G is isomorphic to the group from which it has been obtained by a Goursat twist. Hence, its generating functions and integrity bases are the same as those of this group.

ACKNOWLEDGMENTS

We are grateful to J. McKay for a number of useful suggestions and for reading the manuscript. A scholarship from the Natural Science and Engineering Research Council of Canada is gratefully acknowledged by one of us (Y.S.-A.).

[1] A compilation of existing literature up to 1971 on the Lorentz group is found in P. Winternitz, *Expansions of Relativistic Scattering Amplitudes and Harmonic Analysis on the Lorentz Group*, Part III (University of Arizona, Arizona, 1971).

[2] D. Finkelstein, Phys. Rev. **100**, 924 (1955); P. Winternitz and I. Friš, Yad Fiz. **1**, 889 (1965) [Sov. J. Nucl. Phys. **1**, 636 (1965)].

[3] A.V. Shubnikov and V.A. Koptsik, *Symmetry in Science and Art* (Plenum, New York, 1974), Table 15; V.A. Koptsik, *Shubnikov's Groups* (Russian) (Publication of the Moscow University, Moscow, 1966).

[4] E.M. Goursat, Ann. Sci. de l'Ecole Normale Sup. **6**, 2 (1883); G. Bagnera, Rend. Circ. Matem. Palermo **15**, 21 (1901).

[5] M. Perroud, J. Math. Phys. **19**, 1265 (1978).

[6] J. Patera, R.T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978); P. Desmier and R.T. Sharp, J. Math. Phys. **20**, 74 (1979).

All solutions to a nonlinear system of complex potential equations

C.B. Collins
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

(Received 19 September 1978; accepted for publication 18 September 1979)

This work employs the powerful geometric methods previously developed in order to determine all solutions of the nonlinear system $(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma)$, $(\nabla^2\gamma)^2 - (1/2)(N+2)f'(\gamma)\nabla^2\gamma + (1/4)(N+1)[f'(\gamma)]^2 - (1/2)Nf(\gamma)f''(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma)$, where $f(\gamma)$ is an arbitrarily assigned function and N is an arbitrary constant. The circumstances are determined under which compatible solutions exist, not only when γ is real, but also when γ is complex, and all of the corresponding solutions are found. This is done by referring the system of equations to a set of coordinates based on the (real or complex) equipotential surfaces of constant γ. The relationship between the solutions and the geometry of the equipotential surfaces is examined, and a close association is discovered between the set of allowable equipotential surfaces and the class of surfaces of constant total (Gaussian) curvature. These results are analogous to those in Collins, Math. Proc. Camb. Phil. Soc. **80**, 165–87 (1976), where the system under study was associated with equipotentials of constant radii of principal normal curvature. The geometrical method used throughout is remarkable in that it yields a knowledge of *all* solutions of the given system. It thereby offers the possibility of application to a wide variety of fields in physics where similar systems of equations are encountered.

1. INTRODUCTION

In a previous article[1], a system of nonlinear partial differential equations, viz.,

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma) \tag{1.1a}$$

and

$$\nabla^2\gamma = g(\gamma), \tag{1.1b}$$

was studied, and an application was later given.[2] The quantity γ was real or complex, the functions $f(\gamma)$ and $g(\gamma)$ were arbitrarily assigned, and the operators ∇ and ∇^2 referred to (complexified) three − dimensional Euclidean space. In effect, all possible functions $f(\gamma)$ and $g(\gamma)$ were determined for which the system (1.1) admits a solution, and, in each case, all compatible solutions for γ were found. This was done by developing a geometric technique due to Friedlander,[3] who solved a certain system of nonlinear partial differential equations by considering the geometry of the equipotential surfaces for one of the dependent variables. In the case of system (1.1), where γ may be complex, equipotential surfaces may be investigated in *complexified* three-dimensional Euclidean space. This engenders a richer class of solutions, since the geometry of such surfaces is more diverse than in the real case. This is particularly exemplified in the case where $f(\gamma)\equiv 0$ in Eqs. (1.1). The equipotentials are then "null" surfaces, having the property that any vector normal to the surface at any point is also tangent to the surface at that point. On the other hand, if $f(\gamma) \neq 0$, then the rate of distortion of the normal congruence is described by means of a symmetric shear tensor, which, since it is in general complex, does not necessarily have three distinct eigenvectors[1]. Such diversity is directly reflected in the wider class of solutions that emerges from the system (1.1) in the case where γ is complex. In Ref. 1 (cf. Ref. 4), the geometry of complex surfaces in complexified three-dimensional Euclidean space was developed, and some familiarity with that discussion will now be assumed. For future reference we shall re-state the following theorem.

Theorem 2.5 (Ref. 1): The set $\mathscr{N}$ of all null surfaces consists of the set $\mathscr{P}$ of all null planes, together with a set $\mathscr{S}$ of surfaces described (parametrically) by the equations

$$x = i\beta\sin\alpha + \mu(\alpha),$$

$$y = i\beta\cos\alpha - \int^{\alpha} \mu_{\alpha}(a)\tan a\, da + \alpha_0,$$

and

$$z = \beta,$$

where $\mu(\alpha)$ is an arbitrary function, and α_0 is an arbitrary constant. No member of $\mathscr{S}$ is a null plane. Symbolically,

$$\mathscr{N} = \mathscr{P} \cup \mathscr{S} \text{ with } \mathscr{P} \cap \mathscr{S} = \phi.$$

In the present article, we shall discuss the following system of nonlinear partial differential equations:

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma) \tag{1.2a}$$

and

$$(\nabla^2\gamma)^2 - \tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma + \tfrac{1}{4}(N+1)\,[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f''(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma), \tag{1.2b}$$

where γ is real or complex, $f(\gamma)$ is an arbitrarily assigned function, and N is an arbitrary constant. This system is superficially much more complicated than the system (1.1). It will be shown, however, that system (1.2) is closely related to system (1.1), and that compatible solutions to Eqs. (1.2) exist only in rather special circumstances. These circumstances are summarized in Table I. An application of the results will be given in a subsequent article.

In the case where the equipotentials of γ are not null

 0022-2488/80/020240-09$1.00

TABLE I. A summary of compatible solutions to the system $(\nabla\gamma)^2 = f(\gamma)$;$(\nabla^2\gamma)^2 - \frac{1}{2}(N+2)f'(\gamma)\nabla^2\gamma + \frac{1}{4}(N+1)[f'(\gamma)]^2 - \frac{1}{2}Nf(\gamma)f'(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma)$. The quantity r is finite and nonzero. The entry $\not\equiv 0 \to 1$ means that $f(\gamma)$ is not identically zero and that it has been transformed to unity by a change of variable. The quantities θ and σ are respectively the area of expansion and the shear scalar of the normal geodesic congruence of the equipotentials $\{\gamma = \text{constant}\}$.

$(\nabla\gamma)^2 = f(\gamma)$	N (constant)	Solution γ	(Generalized) Principal radii of normal curvature of equipotentials	Total (Gaussian) curvature	$\nabla^2\gamma$	θ	σ	Conditions
		$\gamma = \gamma_0$	...	...		...	...	γ_0 constant
	Arbitrary				0			ψ,l,m,n arbitrary; l,m,n not all zero and $l^2+m^2+n^2=0$.
0		$\psi(\gamma) = l(\gamma)x + m(\gamma)y + n(\gamma)z$	∞,∞			0		
				0			0	
	1	(α,β) eliminant of $x = i\beta\sin\alpha + \mu(\alpha,\gamma)$						$\mu(\alpha,\gamma)$, $\alpha_0(\gamma)$ arbitrary functions
		$y = i\beta\cos\alpha \times \int_\alpha \mu_\alpha(a,\gamma)\tan a\,da + \alpha_0(\gamma)$	∞, r		$\neq 0$	$\neq 0$		
		$z = \beta$						
$\not\equiv 0$		$\gamma = \gamma_0$	...	...	0	...	...	γ_0 constant $f(\gamma_0) = (N+1)[f'(\gamma_0)]^2 - 2Nf(\gamma_0)f''(\gamma_0) = 0$
	Arbitrary	$lx + my + nz + A$	∞,∞	0	0	0	0	l,m,n,A constant; $l^2+m^2+n^2=1$ ψ arbitrary; $\psi'' \not\equiv 0$; μ_1,μ_2,μ_3,
		$\mu_1x + \mu_2y + \mu_3z + \psi(\lambda_1x + \lambda_2y + \lambda_3z)$						$\lambda_1,\lambda_2,\lambda_3$ constant; $\mu_1^2+\mu_2^2+\mu_3^2=1$; $\lambda_1,\lambda_2,\lambda_3$ not all zero and $\lambda_1^2+\lambda_2^2+\lambda_3^2=0$; $\lambda_1\mu_1+\lambda_2\mu_2+\lambda_3\mu_3=0$
$\not\equiv 0 \to 1$	1	$l(a)x + m(a)y + n(a)z + A(a)$	∞, r				$\neq 0$	$l^2+m^2+n^2=1$; at most one of $l'(a)$, $m'(a)$, and $n'(a)$ is identically zero; $a = a(x,y,z)$ given by $l'(a)x + m'(a)y + n'(a)z + A'(a) = 0$
		$\pm[(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2]^{1/2} + A$	r, r	$\frac{1}{r^2}$	$\neq 0$	$\neq 0$	0	x_1, y_1, z_1, and constant
	2	$\frac{\partial F}{\partial\beta} + A(\beta)$						$F(\beta,x,y,z) \equiv \phi(\beta) + \frac{1}{2}[(1-\beta^2)x + i(1+\beta^2)y - 2\beta z] = 0$; $A(\beta)$ and $\phi(\beta)$ arbitrary except that $A'(\beta)$ and $\phi'''(\beta)$ are not simultaneously zero

$[f(\gamma) \not\equiv 0]$, we can assume, without loss of generality, that $f(\gamma) = 1$, by a change of dependent variable. Compatible solutions will then exist only if $\nabla^2\gamma = 0$, or if $\nabla^2\gamma \neq 0$, and $N = 1$ or 2. In the real case, the equipotentials of γ are planes, nonplanar developables, and spheres, respectively. In the complex case, the equipotentials are complex generalizations of these surfaces, together with a more complicated set of surfaces which is not admitted for real γ.

If the equipotentials of γ are null $[f(\gamma) \equiv 0]$, compatible solutions exist only if either $\nabla^2\gamma = 0$ or $N = 1$ and $\nabla^2\gamma \neq 0$. The equipotentials (necessarily complex) are, respectively, null planes and arbitrary elements of the set $\mathcal{S}$ of nonplanar null surfaces. In other words, if $(\nabla\gamma)^2 = 0$, then necessarily $(\nabla^2\gamma)^2 = -\nabla\gamma\cdot\nabla(\nabla^2\gamma)$.

The above discussion exhausts all possibilities for the equipotentials of solutions to the system (1.2), and the only other cases admitted are those for which there are no equipotentials at all, viz., γ is identically constant.

The system (1.1) has been investigated,[1] first in the real case, resulting in Theorem 3.1 of Ref. 1 (which gave the conditions for compatible real solutions, and determined all such solutions), and then in the complex case, resulting in Theorem 4.1 of Ref. 1 (which gave conditions for compatible complex solutions, and determined all such solutions). Finally, it was shown in Theorem 4.2 of Ref. 1 that the equipotential surfaces for (nonconstant) solutions of system (1.1) could be characterized by their constant (possibly infinite) principal radii of curvature. In the present article, we present three analogous theorems for the system (1.2). In Sec. 2, we determine all real solutions, in Sec. 3 all complex solutions are found, and in Sec. 4 we characterize the equipotential surfaces for (nonconstant) solutions of system (1.2) by their

constant (possibly infinite) total (i.e., Gaussian) curvature. In particular, we observe that the system (1.2) is a generalization of system (1.1), in the sense that all solutions of Eqs. (1.1) are also solutions of Eqs. (1.2).

2. REAL SOLUTIONS OF THE SYSTEM (1.2)

In this section we examine the system of equations

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma), \tag{1.2a}$$

$$(\nabla^2\gamma)^2 - \tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma + \tfrac{1}{4}(N+1)\,[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f''(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma), \tag{1.2b}$$

determining the functions $f(\gamma)$ and the values of N for which real solutions exist, and determining the solutions themselves in each case. The method used here will be the same as in Ref. 1; this method adapts the technique of Friedlander[3] to the system under discussion. Our general result is now given.

Theorem 2.1: Real solutions to the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma),$$

$$(\nabla^2\gamma)^2 - \tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma + \tfrac{1}{4}(N+1)\,[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f''(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma),$$

(where N is constant) in three-dimensional Euclidean space can exist only when one of the following holds:

(i) $f(\gamma)\equiv 0$; then $\gamma = \gamma_0$, where γ_0 is a constant, or

(ii) $f(\gamma)\not\equiv 0, f(\gamma)\geqslant 0$; then $\gamma = \gamma_0$ is a solution if and only if γ_0 is a common root of $f(\gamma)$ and $(N+1)[f'(\gamma)]^2 - 2Nf(\gamma)f''(\gamma)$, or

(iii) $f(\gamma)\not\equiv 0, f(\gamma)\geqslant 0$, and, if ν is any solution of the equation $\nu' = [f(\gamma)]^{-1/2}$, where a prime (′) denotes differentiation with respect to γ, then

$$(\nabla\nu)^2 \equiv \nabla\nu\cdot\nabla\nu = 1 \tag{1.2c}$$

and

$$(\nabla^2\nu)^2 = -N\nabla\nu\cdot\nabla(\nabla^2\nu). \tag{1.2d}$$

Three possibilities occur: Either (a) $\nabla^2\nu = 0$ and N is an arbitrary constant: The equipotentials are planes, and $\nu = lx + my + nz + A$, where l, m, n, and A are constants, and $l^2 + m^2 + n^2 = 1$; or

(b) $\nabla^2\nu \neq 0$ and $N = 1$: the equipotentials are the nonplanar developable surfaces, and ν is given by the a eliminant of

$$\nu = l(a)x + m(a)y + n(a)z + A(a)$$

and

$$0 = l'(a)x + m'(a)y + n'(a)z + A'(a),$$

where $l^2(a) + m^2(a) + n^2(a) = 1$, and at most one of $l'(a)$, $m'(a)$ and $n'(a)$ is identically zero: or

(c) $\nabla^2\nu \neq 0$ and $N = 2$: The equipotentials are spheres and $\nu = \pm R + A$, where $R \equiv [(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2]^{1/2}$ and x_1, y_1, z_1, and A constants. The constants $l, m, n, x_1, y_1, z_1, \gamma_0$, and A are all real.

Proof: If $f(\gamma)\equiv 0$, then the only real solution is clearly $\gamma = \gamma_0$, where γ_0 is an arbitrary constant. In this case, N is arbitrary.

If $f(\gamma)\not\equiv 0$, then $\gamma = \gamma_0$ is a solution if and only if $f(\gamma)$ and $(N+1)[f''(\gamma)]^2 - 2Nf(\gamma)f''(\gamma)$ possess a common zero γ_0. For example, $\gamma = 0$ is a solution if $f(\gamma) = \gamma^2$.

Real solutions with γ not identically constant will exist only when $f(\gamma)\not\equiv 0$ and $f(\gamma)\geqslant 0$, and we shall now focus attention on this remaining case. Defining a new variable ν by $d\nu/d\gamma \equiv \nu' = [f(\gamma)]^{-1/2}$, it follows that $(\nabla\nu)^2 = 1$. This means[1,3] that the equipotential surfaces are orthogonal to a congruence of straight lines ("rays") and that a Gaussian coordinate system (α,β,ν) can be introduced by specifying any point p (sufficiently locally) by means of its normal distance ν to an initial surface $\nu(x,y,z) = 0$, and the position (α,β) of the point of intersection q of the ray through p with the initial surface. It is most convenient to choose the coordinates (α,β) in the initial surface by using orthogonal curvature line parameters, making the coordinate system (α,β,ν) orthogonal. Using Rodrigues' formulas

$$\mathbf{a}_\alpha = -r\mathbf{n}_\alpha \text{ and } \mathbf{a}_\beta = -s\mathbf{n}_\beta, \tag{2.1}$$

the metric becomes

$$d\mathbf{x}^2 = H^2 d\alpha^2 + K^2 d\beta^2 + d\nu^2, \tag{2.2}$$

where $\mathbf{a}(\alpha,\beta)$ is a vector drawn from the origin to the point q on the initial surface, $\mathbf{n}$ is a unit normal to the initial surface at q, $r = r(\alpha,\beta)$ and $s = s(\alpha,\beta)$ denote the principal radii of curvature (which are nonzero, but may be infinite) of the initial surface, $H \equiv (\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2}(1 - \nu/r)$, $K \equiv (\mathbf{a}_\beta\cdot\mathbf{a}_\beta)^{1/2} \times (1 - \nu/s)$, and suffices denote partial derivatives. Further details are given in Ref. 1. In this coordinate system, Eq. (1.2a), or equivalently Eq. (1.2c), is seen to follow automatically. The equations

$$\nabla^2\nu = \frac{1}{HK}\frac{\partial}{\partial\nu}(HK) = \frac{1}{\nu - r} + \frac{1}{\nu - s}$$

and

$$\nabla\nu\cdot\nabla(\nabla^2\nu) = \frac{\partial}{\partial\nu}\left(\frac{1}{HK}\frac{\partial}{\partial\nu}(HK)\right) = -\frac{1}{(\nu - r)^2} - \frac{1}{(\nu - s)^2}$$

show that Eq. (1.2b), or equivalently Eq. (1.2d), becomes

$$(\nabla^2\nu)^2 = -N\nabla\nu\cdot\nabla(\nabla^2\nu) \Longleftrightarrow \left(\frac{1}{\nu - r} + \frac{1}{\nu - s}\right)^2 = N\left(\frac{1}{(\nu - r)^2} + \frac{1}{(\nu - s)^2}\right). \tag{2.3}$$

Three possibilities now arise. One [case (a)] is that both r and s are finite, and then $r = s$, $N = 2$, and $\nabla^2\nu \neq 0$. A second possibility [case (b)] is that exactly one of r and s is infinite (without loss of generality, $s = \infty$), and then $N = 1$ and $\nabla^2\nu \neq 0$. Finally [case (c)], both r and s can be infinite, in which case Eq. (2.3) represents no restriction on N, and Eq. (1.2b) is identically satisfied, since $\nabla^2\nu = 0$. We shall consider these three cases separately.

Case (a): $r = s = \infty$, N is an arbitrary constant: Here $\nabla^2\nu = 0$ and our system is identical to that treated in Case (3.7a) of Ref. 1. The equipotentials are therefore planes, and the general solution is $\nu = lx + my + nz + A$, where $l^2 + m^2 + n^2 = 1$, and l, m, n, and A are (real) constants.

Case (b): $s = \infty$, r finite, $N = 1$, $\nabla^2\nu = 0$: By Eq. (2.1) we have $\mathbf{a}_\alpha = -r\mathbf{n}_\alpha$ and $\mathbf{n}_\beta = \mathbf{0}$. Thus, $\mathbf{n}_{\alpha\beta} = \mathbf{0}$, which implies $\mathbf{a}_{\alpha\beta} = (r_\beta/r)\mathbf{a}_\alpha$, and so $\mathbf{a}_\beta\cdot\mathbf{a}_{\alpha\beta} = 0$. It follows that

$K = (\mathbf{a}_\beta \cdot \mathbf{a}_\beta)^{1/2}$ depends on β only, and so by redefining β we can arrange for $K \equiv 1$. Then $\mathbf{a}_{\beta\beta}\cdot\mathbf{n} = (\mathbf{a}_\beta\cdot\mathbf{n})_\beta - \mathbf{a}_\beta\cdot\mathbf{n}_\beta = 0$, $\mathbf{a}_{\beta\beta}\cdot\mathbf{a}_\alpha = (\mathbf{a}_\beta\cdot\mathbf{a}_\alpha)_\beta - \mathbf{a}_\beta\cdot\mathbf{a}_{\alpha\beta} = 0$, and $\mathbf{a}_{\beta\beta}\cdot\mathbf{a}_\beta = \frac{1}{2}(\mathbf{a}_\beta\cdot\mathbf{a}_\beta)_\beta = 0$, i.e., $\mathbf{a}_{\beta\beta} = \mathbf{0}$. *A fortiori*, $\mathbf{a}_{\beta\beta\alpha} = \mathbf{a}_{\alpha\beta\beta} = \mathbf{0}$, and so $r_{\beta\beta}\mathbf{n}_\alpha = 0$. Now $\mathbf{n}_\alpha \neq 0$, for otherwise, as shown in the proof of Theorem (3.1) of Ref. 1, $\mathbf{n}_\alpha = \mathbf{n}_\beta = \mathbf{0}$ implies that the equipotentials are planes and $r = \infty$, a contradiction. Consequently, $r_{\beta\beta} = 0$. Moreover, $(\partial/\partial\beta)(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2} = (\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{-1/2}(\mathbf{a}_\alpha\cdot\mathbf{a}_{\alpha\beta}) = (r_\beta/r)(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2}$, and so

$$H_\beta = \frac{r_\beta}{r}(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2}\left(1 - \frac{\nu}{r}\right) + \frac{\nu r_\beta}{r^2}(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2} = \frac{r_\beta}{r}(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2},$$

which implies

$$H_{\beta\beta} = (r_{\beta\beta}/r)(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2} - (r_\beta^2/r^2)(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2} + (r_3^2/r^2)(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)^{1/2} = 0.$$

This is the necessary and sufficient condition that the surfaces $\{\nu = \text{constant}\}$, in the metric induced by Eq. (2.2), should be flat, i.e., of zero total curvature. Thus, the equipotentials are "developable surfaces" and so $\nu(x,y,z)$ satisfy

$$\begin{vmatrix} \nu_{xx} & \nu_{xy} & \nu_{xz} & \nu_x \\ \nu_{xy} & \nu_{yy} & \nu_{yz} & \nu_y \\ \nu_{zz} & \nu_{yz} & \nu_{zz} & \nu_z \\ \nu_x & \nu_y & \nu_z & 0 \end{vmatrix} = 0. \tag{2.4}$$

Furthermore, since $(\nabla\nu)^2 = 1$, it follows that

$$\left.\begin{aligned} \nu_x\nu_{xx} + \nu_y\nu_{xy} + \nu_z\nu_{xz} &= 0, \\ \nu_x\nu_{xy} + \nu_y\nu_{yy} + \nu_z\nu_{yz} &= 0, \\ \nu_x\nu_{xz} + \nu_y\nu_{yz} + \nu_z\nu_{zz} &= 0. \end{aligned}\right| \tag{2.5}$$

Using Eqs. (2.5), elementary row operations performed on the determinant in Eq. (2.4) show that

$$\begin{vmatrix} \nu_{xy} & \nu_{yy} & \nu_{yz} \\ \nu_{xz} & \nu_{yz} & \nu_{zz} \\ \nu_x & \nu_y & \nu_z \end{vmatrix} = \begin{vmatrix} \nu_{xx} & \nu_{xy} & \nu_{xz} \\ \nu_{xz} & \nu_{yz} & \nu_{zz} \\ \nu_x & \nu_y & \nu_z \end{vmatrix} = \begin{vmatrix} \nu_{xx} & \nu_{xy} & \nu_{xz} \\ \nu_{xy} & \nu_{yy} & \nu_{yz} \\ \nu_x & \nu_y & \nu_z \end{vmatrix} = 0. \tag{2.6}$$

Using Eqs. (2.5), elementary column operations performed on the determinants in Eq. (2.6) show that

$$\begin{vmatrix} \nu_{yy} & \nu_{yz} \\ \nu_{yz} & \nu_{zz} \end{vmatrix} = \begin{vmatrix} \nu_{xy} & \nu_{yz} \\ \nu_{xz} & \nu_{zz} \end{vmatrix} = \begin{vmatrix} \nu_{xy} & \nu_{yy} \\ \nu_{xz} & \nu_{yz} \end{vmatrix} = \begin{vmatrix} \nu_{xx} & \nu_{xz} \\ \nu_{xz} & \nu_{zz} \end{vmatrix} = \begin{vmatrix} \nu_{xx} & \nu_{xy} \\ \nu_{xz} & \nu_{yz} \end{vmatrix} = \begin{vmatrix} \nu_{xx} & \nu_{xy} \\ \nu_{xy} & \nu_{yy} \end{vmatrix} = 0. \tag{2.7}$$

The interpretation of Eq. (2.7) in terms of Jacobians shows that ν_x, ν_y, and ν_z are functionally dependent in pairs. In general ($\nabla\nu_x \neq \mathbf{0}$), this gives two simultaneous differential equations:

$$\nu_y = f(\nu_x), \tag{2.8a}$$

$$\nu_z = g(\nu_x), \tag{2.8b}$$

where the functions f and g satisfy $(\nabla\nu)^2 = \nu_x^2 + f^2(\nu_x) + g^2(\nu_x) = 1$. In the special case where $\nabla\nu_x = \mathbf{0}$, then

$$(\nabla\nu)^2 = \nu_y^2 + \nu_z^2 = 1 - \nu_x^2 \tag{2.8c}$$

(there is no special subcase where two of $\nabla\nu_x$, $\nabla\nu_y$, and $\nabla\nu_z$ vanish, since then ν_x, ν_y, and ν_z would then all be constants, and the equipotentials would be planes, for which $r = \infty$).

Equations (2.8) are most readily solved by considering the general case [Eqs. (2.8a) and (2.8b)] and special case (2.8c) separately, employing Jacobi's method[5] in each case. The complete integral is

$$\nu = lx + my + nz + A, \tag{2.9}$$

where l,m,n, and A are constants and $l^2 + m^2 + n^2 = 1$. However, such a solution corresponds to planar equipotentials, where $r = \infty$ and so is inadmissible. Singular solutions do not exist, since $\partial\nu/\partial A = 1$, and so cannot be equated to zero. The general solution is found by regarding l,m,n, and A in Eq. (2.9) not as constants, but as functions of some parameter a, viz.,

$$\nu = l(a)x + m(a)y + n(a)z + A(a), \tag{2.10a}$$

and then eliminating a between Eq. (2.10a) and

$$0 = l'(a)x + m'(a)y + n'(a)z + A'(a). \tag{2.10b}$$

If two of l', m', and n' are identically zero, then so also is the third, and then $A(a)$ is constant, and $r = \infty$, a contradiction. Hence, at most one of l', m', and n' is identically zero.

Case (c): $r = s$(finite), $N = 2$, $\nabla^2\nu \neq 0$: We shall first show that r is constant. By Rodrigues' formula (2.1), we find that $\mathbf{a}_{\alpha\beta} = \mathbf{a}_{\beta\alpha}$ requires $r_\beta\mathbf{a}_\alpha = r_\alpha\mathbf{a}_\beta$. Since $\mathbf{a}_\alpha\cdot\mathbf{a}_\beta = 0$, it follows that $r_\alpha = r_\beta = 0$. Hence, $\nabla^2\nu = 2/(\nu - r)$, with r constant. The problem is now identical to Case (3.7c) of Ref. 1. The equipotentials are spheres, and the general solution is $\nu = r \pm [(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2]^{1/2}$, where x_1, y_1, and z_1 are constants. □

Corollary: Any real nontrivial solution of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma), \tag{1.1a}$$

$$\nabla^2\gamma = g(\gamma) \tag{1.1b}$$

is also a solution of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma), \tag{1.2a}$$

$$(\nabla^2\gamma)^2 - \tfrac{1}{2}(N+2)f'(\gamma)\nabla^2\gamma + \tfrac{1}{4}(N+1)[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f''(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma). \tag{1.2b}$$

Proof: The proof follows immediately, using previous results[1] in conjunction with the proof of Theorem 2.1. □

Remark: A particular example of a solution in case (b) is provided by the case where $l(a) = a$, $m(a) = (1 - a^2 - k^2)^{1/2}$, $n(a) = k$, and $A(a) = 0$, where k is a constant satisfying $-1 < k < 1$. The a eliminant of Eqs. (2.10a) and (2.10b) is then simply

$$\nu = \pm\sqrt{(1 - k^2)(x^2 + y^2)} + kz.$$

If $k = 0$, the equipotentials are concentric right-circular cylinders whose axis is the z axis. If $k \neq 0$, the equipotentials are parallel right-circular cones whose common axis is the z axis.

The equipotentials in cases (a) and (c) are identical with those of Cases (3.7a) and (3.7b) of Ref. 1 for the system (1.1). However, the set of equipotentials in case (b) (i.e., the set of

developable surfaces) is much larger than the set in Case (3.7b) of Ref. 1 (i.e., the set of cylinders). By an appropriate specialization of Eqs. (2.10a) and (2.10b), and bearing in mind the Corollary, it is possible to recover Case (3.7b) of Ref. 1, since the necessary and sufficient condition for the equipotentials to be concentric cylinders is that, in Eqs. (2.10a) and (2.10b), $A(a)$ is constant and $\mathbf{l}\times\mathbf{l}'\equiv/|\mathbf{l}'|$ is a constant unit vector, where $l\equiv(l,m,n)$ and $\mathbf{l}'\equiv d\mathbf{l}/da$. This follows because, if the equipotentials are concentric cylinders, we can, without loss of generality, take the common axis to be the z axis. Then $\nu=\pm(x^2+y^2)^{1/2}+B$, where B is a constant. Now from Eqs. (2.10a) and (2.10b), $\nabla\nu=(l,m,0)=(x,y,0)/(x^2+y^2)^{1/2}$. Since $\mathbf{l}\cdot\mathbf{l}=1$, $\mathbf{l}\cdot\mathbf{l}'=0$, where $\mathbf{l}'\neq\mathbf{0}$, and so $\mathbf{l}'\propto(y,-x,0)$ and $(\mathbf{l}\times\mathbf{l}')/|\mathbf{l}'|=\pm(0,0,1)$. Moreover, $\nu=lx+my+B\Rightarrow A(a)=B$ is constant. Conversely, if $A(a)$ is constant and $(\mathbf{l}\times\mathbf{l}')/|\mathbf{l}'|$ is a constant unit vector [without loss of generality (0,0,1)], then $mn'-nm'=ln'-nl'=0$ and $lm'-l'm=|\mathbf{l}'|$. If $n\neq0$, then $\mathbf{l}=S(a)\mathbf{l}_1$, where $\mathbf{l}_1$ is a constant vector, and $lm'-l'm=0$, a contradiction. Then $n\equiv0$, and, since $n'\equiv0$, $l'm\neq0$. Writing $a=a(l)$ in Eqs. (2.10a) and (2.10b), we obtain two equations from which we can determine the l eliminant, namely, $\nu=\pm(x^2+y^2)^{1/2}+B$, where B is a constant, and the equipotentials are concentric cylinders.

3. COMPLEX SOLUTIONS OF THE SYSTEM (1.2)

In this section we extend our results to the case of complex solutions of the system (1.2). The main result is now given.

Theorem 3.1: Complex solutions to the system

$$(\nabla\gamma)^2\equiv\nabla\gamma\cdot\nabla\gamma=f(\gamma), \tag{1.2a}$$

$$(\nabla^2\gamma)^2-\tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma+\tfrac{1}{4}(N+1)\,[f'(\gamma)]^2-\tfrac{1}{2}Nf(\gamma)f''(\gamma)=-N\nabla\gamma\cdot\nabla(\nabla^2\gamma), \tag{1.2b}$$

(where N is constant) in complexified three-dimensional Euclidean space can exist only when one of the following holds: (i) $f(\gamma)\equiv0$: Three possibilities arise: either (a) $\gamma\equiv\gamma_0$, where γ_0 is constant; or (b) γ satisfies an equation of form $\psi(\gamma)=l(\gamma)x+m(\gamma)y+n(\gamma)z$, where $\psi(\gamma)$, $l(\gamma)$, $m(\gamma)$, and $n(\gamma)$ are arbitrary, except that $l^2+m^2+n^2\equiv0$ and not all of l,m,n are zero; the equipotentials are null planes; $\nabla^2\gamma=0$ and N is an arbitrary constant; or (c) γ is given by eliminating α and β from the equations

$$x=i\beta\sin\alpha+\mu(\alpha,\gamma)$$

$$y=i\beta\cos\alpha-\int^{\alpha}\mu_{\alpha}(a,\gamma)\tan a\,da+\alpha_0(\gamma)$$

and

$$z=\beta,$$

where $\mu(\alpha,\gamma)$ and $a_0(\gamma)$ are arbitrary functions. The equipotentials are members of the set $\mathcal{S}$ of nonplanar null surfaces (see [1]Theorem 2.5); $\nabla^2\gamma\neq0$ and $N=1$; or

(ii) $f(\gamma)\not\equiv0$: Then $\gamma=\gamma_0$ is a solution if and only if γ_0 is a common root of $f(\gamma)$ and $(N+1)\,[f'(\gamma)]^2-2Nf(\gamma)f''(\gamma)$; or

(iii) $f(\gamma)\not\equiv0$, and if ν is any solution of the equation $\nu'=[f(\gamma)]^{-1/2}$, where a prime ($'$) denotes differentiation with respect to γ, then

$$(\nabla\nu)^2\equiv\nabla\nu\cdot\nabla\nu=1 \tag{1.2c}$$

and

$$(\nabla^2\nu)^2=-N\nabla\nu\cdot\nabla(\nabla^2\nu). \tag{1.2d}$$

Five possibilities occur: either (a) $\nabla^2\nu=0$ and N is an arbitrary constant: The equipotentials are planes, and $\nu=lx+my+nz+A$, where l,m,n, and A are constants, and $l^2+m^2+n^2=1$; or (b) $\nabla^2\nu\neq0$ and $N=1$: The equipotentials are the nonplanar developable surfaces, and ν is given by the a eliminant of

$$\nu=l(a)x+m(a)y+n(a)z+A(a)$$

and

$$0=l'(a)x+m'(a)y+n'(a)z+A'(a),$$

where $l^2(a)+m^2(a)+n^2(a)=1$, and at most one of $l'(a)$, $m'(a)$, and $n'(a)$ is identically zero: or (c) $\nabla^2\nu\neq0$ and $N=2$: The equipotentials are spheres and $\nu=\pm R+A$, where $R\equiv[(x-x_1)^2+(y-y_1)^2+(z-z_1)^2]^{1/2}$, and x_1,y_1,z_1, and A are constants: or (d) $\nabla^2\nu=0$ and N is an arbitrary constant: Now $\nu=\mu_1x+\mu_2y+\mu_3z+\psi(\lambda_1x+\lambda_2y+\lambda_3z)$, where ψ is an arbitrary function satisfying $\psi''\not\equiv0$, and $\mu_1,\mu_2,\mu_3,\lambda_1,\lambda_2$, and λ_3 are arbitrary constants satisfying $\mu_1^2+\mu_2^2+\mu_3^2=1$, $\lambda_1^2+\lambda_2^2+\lambda_3^2=0$, and $\mu_1\lambda_1+\mu_2\lambda_2+\mu_3\lambda_3=0$, and not all of λ_1,λ_2, and λ_3 are zero: or (e) $\nabla^2\gamma\neq0$ and $N=2$. Now $\nu=r(\beta)\pm(\partial F/\partial\beta)$, where F and β are given by

$$F(\beta,x,y,z)\equiv\phi(\beta)+\tfrac{1}{2}[(1-\beta^2)x+i(1+\beta^2)y-2\beta z]=0,$$

and the functions $r(\beta)$ and $\phi(\beta)$ are arbitrary, except that $r'(\beta)$ and $\phi'''(\beta)$ are not simultaneously identically zero [i.e., either $r'(\beta)\not\equiv0$, or $\phi'''(\beta)\not\equiv0$, or both].

Proof: As in the proof of Theorem 2.1, it is clear that γ is identically (complex) constant $\gamma\equiv\gamma_0$ if and only if $f(\gamma_0)=(N+1)\,[f'(\gamma_0)]^2-2Nf(\gamma_0)f''(\gamma_0)=0$, i.e., γ_0 is a common zero of $f(\gamma)$ and $(N+1)[f'(\gamma)]^2-2Nf(\gamma)f''(\gamma)$. This proves case (ii) and case (ia).

If γ is not identically constant, we observe from Eq. (1.2a) that the equipotential surfaces for γ have null normals if and only if $f(\gamma)=0$, and hence if the normal to an equipotential is null at one point, it is null everywhere. Consequently, either $f(\gamma)\equiv0$ [cases (ib) and (ic)], or there are equipotential surfaces whose normals at each point are not null [case (iii)].

Suppose that the normals to all (sufficiently local) equipotential surfaces are null. By Theorem 2.5 of Ref. 1 (see Sec. 1), the equipotentials are either null planes or they lie in the set $\mathcal{S}$, and can be described by the equations

$$x=i\beta\sin\alpha+\mu(\alpha,\gamma),$$

$$y=i\beta\cos\alpha-\int^{\alpha}\mu_{\alpha}(a,\gamma)\tan a\,da+\alpha_0(\gamma),$$

$$z=\beta. \tag{3.1}$$

If the equipotentials are null planes, we must have $\psi(\gamma)=l(\gamma)x+m(\gamma)y+n(\gamma)z$ for some function $\psi(\gamma)$, and for functions $l(\gamma)$, $m(\gamma)$ and $n(\gamma)$ satisfying $l^2(\gamma)+m^2(\gamma)+n^2(\gamma)=0$. One can readily check that $(\nabla\gamma)^2=\nabla^2\gamma=0$

for arbitrary $\psi(\gamma)$. In this case Eq. (1.2b) is satisfied for arbitrary constant N. If, on the other hand, the null equipotentials are not planes, then, in the coordinates of Eqs. (3.1), the metric is of the form[1]

$$ds^2 = AD\alpha^2 + 2B\,d\alpha\,d\gamma + 2C\,d\beta\,d\gamma + D\,d\,\gamma^2,$$

where $A \equiv (i\beta + \mu_\alpha \sec\alpha)^2$ and $C = C(\alpha, \gamma) \not\equiv 0$. It is easily seen that this form guarantees that $(\nabla\gamma)^2 = 0$. Moreover,

$$\nabla^2\gamma = \frac{1}{2C}(\ln A)_\beta = [C(\beta - i\mu_\alpha \sec\alpha)]^{-1},$$

and so

$$\nabla\gamma\cdot\nabla(\nabla^2\gamma) = \frac{1}{C}(\nabla^2\gamma)_\beta = -[C(\beta - i\mu_\alpha\sec\alpha)]^{-2} = -(\nabla^2\gamma)^2.$$

Consequently, Eq. (1.2b) is satisfied if and only if $N = 1$. Note that there are no restrictions on the set of null surfaces in $\mathscr{S}$. This completes case (i).

It means to investigate case (iii). It is clear that whenever the shear tensor of the normal congruence possesses three linearly independent nonnull eigenvectors, we can directly generalize the arguments of the proof of Theorem 2.1, which concerned the real case. The only change that must be made is that the constants l, m, n, x_1, y_1, z_1, and A may now be complex. This completes cases (iiia–c).

In the case where the shear tensor possesses only one nonnull eigenvector, there exists a single null eigenvector $\mathbf{a}_\alpha$ which is orthogonal to $\mathbf{n}$, and only one curvature line parameter[1] α. Since the equipotential surfaces are *not* null, it follows[1] that there are two families of null curves in any surface. One of these families is the set of integral curves of $\mathbf{a}_\alpha$. Let β be a parameter for the second family, i.e., the second family is the set of integral curves of $\mathbf{a}_\beta$, and $\mathbf{a}_\beta\cdot\mathbf{a}_\alpha \neq 0$. We obtain the following generalizations of Rodrigues' formula (2.1):

$$\mathbf{n}_\alpha = -\frac{1}{r}\mathbf{a}_\alpha \tag{3.2}$$

and

$$\mathbf{n}_\beta = -\frac{1}{s}\mathbf{a}_\alpha - \frac{1}{r}\mathbf{a}_\beta, \tag{3.3}$$

where $s = s(\alpha, \beta) \neq 0$ is finite, and $r = r(\alpha, \beta) \neq 0$ is either finite or infinite[1]. Furthermore, the metric can be put in the form

$$ds^2 = 2(\mathbf{a}_\alpha\cdot\mathbf{a}_\beta)\left(1 - \frac{\nu}{r}\right)^2 d\alpha\,d\beta - \frac{2\nu}{s}(\mathbf{a}_\alpha\cdot\mathbf{a}_\beta)\left(1 - \frac{\nu}{r}\right)d\beta^2 + d\nu^2, \tag{3.4}$$

from which it follows that $(\nabla\nu)^2 = 1$ and $\nabla^2\nu = 2/(\nu - r)$ (cf. Ref. 1). Thus, $\nabla\nu\cdot\nabla(\nabla^2\nu) = -2/(\nu - r)^2$, and Eq. (1.2c) and (1.2d) are satisfied.

We consider first the case (iiid) when $r = \infty$. Then $(\nabla\nu)^2 = 1$ and $\nabla^2\nu = 0$, the solution to which[1] is

$$\nu = \mathbf{x}\cdot\boldsymbol{\mu} + \psi(\mathbf{x}\cdot\boldsymbol{\lambda}),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are arbitrary constant vectors satisfying $\boldsymbol{\mu}^2 = 1$, $\boldsymbol{\lambda}^2 = 0$, $\boldsymbol{\lambda}\cdot\boldsymbol{\mu} = 0$, $\boldsymbol{\lambda} \not\equiv \mathbf{0}$, and ψ is an arbitrary function satisfying $\psi'' \not\equiv 0$ [the case $\psi'' \equiv 0$ reduces this solution to that of case (iiia) above]. The constant N is arbitrary. This completes case (iiid).

There now remains only case (iiie), for which $(\nabla\nu)^2 = 1$, $\nabla^2\nu = 2/(\nu - r)$, and $r = r(\alpha, \beta)$ is finite and nonzero. In this case, $N = 2$, as follows from Eqs. (1.2d) and (3.4). Moreover, $r = r(\beta)$, as we now show. We have that $\mathbf{a}_\alpha\cdot\mathbf{a}_{\alpha\alpha} = \frac{1}{2}(\mathbf{a}_\alpha\cdot\mathbf{a}_\alpha)_\alpha = 0$, and so, from Eqs. (3.2) and (3.3),

$$\mathbf{n}_{\alpha\beta} = \mathbf{n}_{\beta\alpha} \Longleftrightarrow \frac{1}{s}\mathbf{a}_{\alpha\alpha} + \left(\frac{r_\beta}{r^2} - \frac{s_\alpha}{s^2}\right)\mathbf{a}_\alpha - \frac{r}{r^2}\mathbf{a}_\beta = \mathbf{0}.$$

The component of this along $\mathbf{a}_\alpha$ yields $r_\alpha = 0$; hence, $r = r(\beta)$. The method of solution to case (iiic) in Ref. 1 applies here, to the point where it is obtained that

$$\nu = r \pm F_\beta, \tag{3.5}$$

where

$$F(\beta, x, y, z) \equiv \phi(\beta) + \tfrac{1}{2}(1 - \beta^2)x + \tfrac{1}{2}i(1 + \beta^2)y - \beta z = 0, \tag{3.6}$$

the function $\phi(\beta)$ being completely arbitrary, and a β suffix denoting partial differentiation with respect to β, holding x, y, and z constant. A direct calculation as in Ref. 1 shows that ν, given by Eq. (3.5), satisfies $(\nabla\nu)^2 = 1$ and $\nabla^2\nu = 2/(\nu - r)$, with $r = r(\beta)$ arbitrary. However, we must still check the validity of Eq. (3.2) and (3.3). Consider the initial surface $\nu = 0$:

$$r(\beta) \pm [\phi'(\beta) - \beta x + i\beta y - z] = 0.$$

The unit normal to this surface is

$$\begin{aligned}\mathbf{n} &= \nabla\nu \\ &= \nabla r \pm \nabla F_\beta \\ &= (r_\beta \pm F_{\beta\beta})\nabla\beta \pm (-\beta, i\beta, -1) \\ &= -\frac{1}{r}\mathbf{G}(\beta) - \frac{\mathbf{a}}{r},\end{aligned}$$

where

$$\begin{aligned}\mathbf{G}(\beta) \equiv\ & \pm r_\beta[-\tfrac{1}{2}(1 - \beta^2), -\tfrac{1}{2}(1 + \beta^2), \beta] \\ & - \{\tfrac{1}{2}\phi''(1 - \beta^2) + \beta\phi' - \phi, \\ & i[\tfrac{1}{2}\phi''(1 + \beta^2) - \beta\phi' + \phi], -\beta\phi'' + \phi'\}.\end{aligned}$$

Thus equation (3.2) is valid, and Eq. (3.3) holds if and only if $-(1/s)\mathbf{a}_\alpha = -(1/r)\mathbf{G}'(\beta) - (r_\beta/r)\mathbf{n}$. A direct calculation shows that this is equivalent to the requirement

$$-\frac{1}{s}\mathbf{a}_\alpha = \frac{1}{r}\{\pm(r_{\beta\beta}F_\beta + r_\beta F_{\beta\beta}) + \phi'''F_\beta + r_\beta^2\}\nabla\beta.$$

Since we require $(1/s) \neq 0$ in Eq. (3.3), it follows that $r(\beta)$ and $\phi(\beta)$ are restricted so that

$$\pm(r_{\beta\beta}F_\beta + r_\beta F_{\beta\beta}) + \phi'''F_\beta + r_\beta^2 \neq 0.$$

We shall show that this condition is violated if and only if both $r_\beta = 0$ and $\phi''' = 0$. For suppose that

$$\pm(r_{\beta\beta}F_\beta + r_\beta F_{\beta\beta}) + \phi'''F_\beta + r_\beta^2 = 0.$$

Then,

$$\begin{aligned}P(\beta, x, y, z) \equiv\ & [(\pm r_{\beta\beta} + \phi''')\phi' + r_\beta^2 \pm r_\beta\phi''] \\ & - [\pm(\beta r_{\beta\beta} + r_\beta) + \beta\phi''']x \\ & + [\pm(\beta r_{\beta\beta} + r_\beta) + \beta\phi''']iy \\ & + [\mp r_{\beta\beta} - \phi''']z = 0.\end{aligned} \tag{3.7}$$

If $P_\beta = 0$, it follows immediately that $r_\beta = \phi''' = 0$. If $P_\beta \neq 0$, the fact that $(\nabla\beta)^2 = 0$ implies $P_x^2 + P_y^2 + P_z^2 = 0$ (where suffices denote differentiation holding β constant), i.e., that $r_{\beta\beta} \pm \phi''' = 0$, and so $r_\beta \pm \phi''$ is constant. Now Eq. (3.7) becomes

$$P(\beta, x, y, z) \equiv r_\beta [(r_\beta \pm \phi'') \mp (x - iy)] = 0 . \quad (3.8)$$

If $r_\beta \neq 0$ in Eq. (3.8), then $r_\beta \pm \phi'' = \pm (x - iy)$, and the fact that $r_\beta \pm \phi''$ is constant yields a contradiction. Consequently, $r_\beta = 0$, and then $\phi''' = 0$. Thus, the only possible case where our solution (3.5) and (3.6) fails is when $r_\beta = 0$ and $\phi''' = 0$. This exceptional case, where r is constant and where the function ϕ is of the form $\phi(\beta) = a\beta^2 + b\beta + c$ for constants a, b, and c, was discussed in Ref. 1. Application of the prescription (3.5) and (3.6) in this particular case yields

$$\nu = r \pm \{[x - (a - c)]^2 + [y - i(a + c)]^2 + (z - b)^2\}^{1/2} ,$$

showing that this case reduces to that of case (iiic) above. We observe[1] that the particular (quadratic) form of the function ϕ is generated from the function $\phi(\beta) \equiv 0$ in Eq. (3.6) by a translation of the origin

$$x \to x - (a - c), \quad y \to y - i(a + c), \quad z \to z - b .$$

This completes case (iiie), and the proof of the theorem. □

Corollary 1: Any complex nontrivial solution of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma \cdot \nabla\gamma = f(\gamma) , \quad (1.1a)$$

$$\nabla^2\gamma = g(\gamma) \quad (1.1b)$$

is also a solution of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma \cdot \nabla\gamma = f(\gamma) , \quad (1.2a)$$

$$(\nabla^2\gamma)^2 - \tfrac{1}{2}(N+2) f'(\gamma) \nabla^2\gamma + \tfrac{1}{4}(N+1)[f'(\gamma)]^2 - \tfrac{1}{2} N f(\gamma) f''(\gamma) = -N \nabla\gamma \cdot \nabla(\nabla^2\gamma) . \quad (1.2b)$$

Proof: The proof follows immediately, using previous results[1] in conjuction with the proof of Theorem 3.1. □

Corollary 2:

$(\nabla\gamma)^2 \equiv \nabla\gamma \cdot \nabla\gamma = 0 \Rightarrow (\nabla^2\gamma)^2 = -\nabla\gamma \cdot \nabla(\nabla^2\gamma)$.

Proof: If γ is identically constant, the result is immediate. If γ is not constant, the result follows from the proof of Theorem 3.1. □

Remark: Case (iiia) can be considered as the special case of (iiid) in which $\psi'' \equiv 0$, and case (iiic) can be considered as the special case of (iiie) with $r_\beta \equiv \phi''' \equiv 0$. It is more convenient to separate these cases in order to compare the results of Theorem 2.1 with those of Theorem 3.1.

It is also possible to recover the complex generalization of Case (3.7b) of Ref. 1 from case (iiib) of Theorem 3.1, in much the same way as was done in the real case following Theorem 2.1. The necessary and sufficient condition for the equipotentials to be concentric cylinders in case (iiib) in Theorem 3.1 is that $A(a)$ is constant, I′ is nonnull, and $(\mathbf{l} \times \mathbf{l}')/|\mathbf{l}'|$ is a constant unit vector. The proof follows precisely as in the real case.

4. CHARACTERIZATION OF THE SOLUTIONS

We shall give two characterizations of the solutions to the system (1.2) in terms of the imbedding of the equipotential surfaces in (complexified) three-dimensional Euclidean space. These two characterizations are analogous to ones that can be obtained for the system (1.1).

Theorem 4.1: A family of surfaces $\{\gamma = \text{constant}\}$ is a set of equipotential surfaces for (nonconstant) solutions of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma \cdot \nabla\gamma = f(\gamma) \quad (1.1a)$$

and

$$\nabla^2\gamma = g(\gamma) \quad (1.1b)$$

if and only if the surfaces are orthogonal to a congruence of straight lines whose area expansion is constant on each equipotential.

Proof: We use a notation in which a semicolon denotes covariant derivative, indices are raised and lowered using the metric tensor, Greek indices denote component indices, δ^μ_λ is the Kronecker delta tensor, and the summation notation for repeated indices is employed. Since $\gamma^{;\alpha}_{;\beta}\gamma^{;\beta} = \gamma_{;\beta}{}^{;\alpha}\gamma^{;\beta} = \tfrac{1}{2}(\gamma_{;\beta}\gamma^{;\beta})^{;\alpha} = \tfrac{1}{2}[(\nabla\gamma)^2]^{;\alpha}$, it follows that the curves of the congruence normal to the equipotential surfaces $\{\gamma = \text{constant}\}$ are geodesics (i.e., straight lines) if and only if $(\nabla\gamma)^2 = f(\gamma)$ for some function $f(\gamma)$, i.e., if and only if Eq. (1.1a) is satisfied. If ν is a solution of the equation $d\nu/d\gamma = [f(\gamma)]^{-1/2}$, then $(\nabla\nu)^2 = 1$ and we can decompose[1] the term $\nu_{;\lambda;\mu}$ as

$$\nu_{;\lambda;\mu} = \theta_{\lambda\mu} = \sigma_{\lambda\mu} + \tfrac{1}{2}\theta h_{\lambda\mu} , \quad (4.1)$$

where $\theta_{\lambda\mu}$, $\sigma_{\lambda\mu}$, and θ are interpreted as the "expansion tensor", the "shear tensor" and the "area expansion" of the normal congruence, respectively, and satisfy $\theta_{\lambda\mu}\nu^{;\mu} = 0$, $\theta_{\lambda\mu} = \theta_{\mu\lambda}$, $\theta^\lambda_\lambda = \theta$, $\sigma_{\lambda\mu}\nu^{;\mu} = 0$, $\sigma_{\lambda\mu} = \sigma_{\mu\lambda}$, and $\sigma^\lambda_\lambda = 0$. The tensor $h_{\lambda\mu}$ is the "projection tensor" into the tangent plane at each point on a surface, i.e., $h_{\lambda\mu} = \delta_{\lambda\mu} - \nu_{;\lambda}\nu_{;\mu}$, $h_{\lambda\mu} = h_{\mu\lambda}$, $h^\lambda_\lambda = 2$, and $h_{\mu\lambda}h^{\mu\tau} = h^\tau_\lambda$. For future reference, we also define the shear scalar σ by $2\sigma^2 = \sigma_{\lambda\mu}\sigma^{\lambda\mu}$, $\sigma \geqslant 0$. Thus, $\nabla^2\gamma = g(\gamma)$ if and only if $\nabla^2\nu = h(\nu)$ for some function $h(\nu)$, which is true if and only if $\theta = \theta(\nu)$, since $\nabla^2\nu = \nu^{;\lambda}_{;\lambda} = \theta$. Thus, Eq. (1.1b) is satisfied if and only if θ is constant on each equipotential $\{\nu = \text{constant}\}$. The above decomposition is analogous to that employed in the pseudo-Riemannian manifolds of general relativistic cosmology (see, for example, Ref. 6).

If $f(\gamma) \equiv 0$, the equipotential surfaces are null. Again, the curves of the congruence normal (and hence tangent) to the equipotential surfaces are geodesics (i.e., straight lines). The quantity $\gamma_{;\lambda;\mu}$ can be decomposed into its trace and tracefree parts as

$$\gamma_{;\lambda;\mu} = \theta_{\lambda\mu} = \sigma_{\lambda\mu} + \theta h_{\lambda\mu} , \quad (4.2)$$

where in this case the interpretations of $\theta_{\lambda\mu}$, $\sigma_{\lambda\mu}$, and θ for the normal congruence are the same as before [for the case $f(\gamma) \not\equiv 0$], but $h_{\lambda\mu}$, while still being a projection tensor orthogonal to $\gamma_{;\mu}$, is now degenerate, since $h_{\lambda\mu}\gamma^{;\mu} = 0$ and $\gamma_{;\mu}\gamma^{;\mu} = 0$. Thus, $h_{\lambda\mu} = h_{\mu\lambda}$, $h^\lambda_\lambda = 1$, and $h_{\mu\lambda}h^{\mu\tau} = h^\tau_\lambda$. The proof of the theorem in this case [$f(\gamma) \equiv 0$] now follows as in

the case where $f(\gamma)\not\equiv 0$. □

Theorem 4.2: A family of surfaces $\{\gamma = \text{constant}\}$ is a set of equipotential surfaces for (nonconstant) solutions of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma) \tag{1.1a}$$

and

$$\nabla^2\gamma = g(\gamma) \tag{1.1b}$$

if and only if the surfaces are orthogonal to a congruence of straight lines and have constant (possibly infinite) principal radii of normal curvature.

Proof: This theorem is merely a restatement of Theorem (4.2) in Ref. 1. □

Two corresponding characterizations for the system (1.2) are given in the next two theorems.

Theorem 4.3: A family of surfaces $\{\gamma = \text{constant}\}$ is a set of equipotential surfaces for (nonconstant) solutions of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma)\,, \tag{1.2a}$$

$$\begin{aligned}(\nabla^2\gamma)^2 &- \tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma \\ &+ \tfrac{1}{4}(N+1)\,[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f'(\gamma) \\ &= -N\nabla\gamma\cdot\nabla(\nabla^2\gamma)\end{aligned} \tag{1.2b}$$

(where N is constant) if and only if the surfaces are orthogonal to a congruence of straight lines whose shear scalar σ and area expansion θ are related by

$$2N(1+\epsilon)\sigma^2 + (N-1-\epsilon)\,\theta^2 = 0\,,$$

where $\epsilon = 0$ for null equipotentials, and $\epsilon = 1$ for nonnull equipotentials.

Proof: As in the proof of Theorem 4.1, it follows that the curves of the congruence normal to the equipotential surfaces are geodesics (i.e., straight lines) if and only if Eq. (1.2a) is satisfied.

If the case $f(\gamma)\not\equiv 0$, we consider the variable $\nu(\gamma)$, being a solution of the equation $d\nu/d\gamma = [f(\gamma)]^{-1/2}$, and satisfying

$$(\nabla\gamma)^2 \equiv \nabla\nu\cdot\nabla\nu = 1 \tag{1.2c}$$

and

$$(\nabla^2\nu)^2 = -N\nabla\nu\cdot\nabla(\nabla^2\nu)\,. \tag{1.2d}$$

Now $\nu_{;\lambda;\mu}\nu^{\lambda;\mu} = \nu_{;\lambda;\mu}\nu^{\mu;\lambda} = (\nu_{;\lambda;\mu}\nu^{\mu})^{;\lambda} - \nu_{;\lambda;\mu}{}^{;\lambda}\nu^{\mu}$ $= -(\nu_{;\lambda}{}^{;\lambda})_{;\mu}\nu^{\mu} = -\nabla\nu\cdot\nabla(\nabla^2\nu)$, recalling that $\nu_{;\lambda;\mu}\nu^{\mu} = 0$. It follows from Eq. (4.1) that $\nabla\nu\cdot\nabla(\nabla^2\nu)$ $= -\frac{1}{2}(4\sigma^2 + \theta^2)$. Thus, Eq. (1.2d), and hence Eq. (1.2b), is satisfied if and only if $4N\sigma^2 + (N-2)\,\theta^2 = 0$.

If $f(\gamma)\equiv 0$, Eq. (1.2b) is satisfied if and only if $2N\sigma^2 + (N-1)\,\theta^2 = 0$, where we have used the decomposition (4.2). □

Corollary: If a set of surfaces is orthogonal to a congruence of straight lines whose area expansion θ is constant on each surface, the shear scalar σ is related to θ by the equation

$$2N(1+\epsilon)\sigma^2 + (N-1-\epsilon)\,\theta^2 = 0\,,$$

where N is a constant, $\epsilon = 0$ for null surfaces, and $\epsilon = 1$ for nonnull surfaces.

Proof: This follows upon application of Theorems 4.1 and 4.3, the Corollary of Theorem 2.1, and Corollary 2 of Theorem 3.1. □

Remark: The expansion θ and shear scalar σ, have been evaluated for each class of nontrivial solution given in Theorems 2.1 and 3.1. Whether or not they vanish identically is indicated in Table I.

Theorem 4.4: A family of surfaces $\{\gamma = \text{constant}\}$ is a set of equipotential surfaces for (nonconstant) solutions of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma)\,, \tag{1.2a}$$

$$\begin{aligned}(\nabla^2\gamma)^2 &- \tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma \\ &+ \tfrac{1}{4}(N+1)\,[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f'(\gamma) \\ &= -N\nabla\gamma\cdot\nabla(\nabla^2\gamma)\end{aligned} \tag{1.2b}$$

if and only if the surfaces are orthogonal to a congruence of straight lines and either have zero total (Gaussian) curvature or have nonzero total (Gaussian) curvature but equal finite principal radii of normal curvature.

Proof: We have already seen in the proof of Theorem 4.1 that the surfaces are orthogonal to a congruence of straight lines if and only if Eq. (1.2a) is satisfied.

Suppose that Eq. (1.2) are satisfied. If the equipotentials are not null and the shear tensor possesses three linearly independent nonnull eigenvectors [case (iii) of Theorem 2.1 and cases (iiia–c) of Theorem 3.1], it follows that the total curvature is zero, unless the principal radii of curvature r and s are finite and equal, in which case they are constant, and hence so is the total curvature. If the equipotentials are not null and the shear tensor possesses only two linearly independent eigenvectors [cases (iiid) and (iiie) of Theorem 3.1], it follows from Eq. (3.2) and (3.3) that the principal curvatures are equal (taking the value $1/r$). If r is infinite, then the total curvature is zero. If r is finite, r is an arbitrary function of the variable β. While the total curvature is variable, this occurs only if the two principal radii of curvature are equal and finite. Finally, if the equipotentials are null, then the total curvature is necessarily zero (principal radii of normal curvature ρ_1 and ρ_2, can be defined for a null surface[1]; the "generalized total curvature" $1/\rho_1\rho_2 = 0$ for all null surfaces).

Conversely, suppose that a set of surfaces $\{\gamma = \text{constant}\}$ is orthogonal to a congruence of straight lines [so Eq. (1.2a) is satisfied], and that the surfaces either have zero total curvature, or have nonzero total curvature with equal finite principal radii of curvature. If the surfaces are not null and if the shear tensor of the normal congruence has three linearly independent eigenvectors, the result follows immediately, by the proofs of Theorems 2.1 and 3.1. If the surfaces are not null and the shear tensor of the normal congruence possesses only two linearly independent eigenvectors, the principal radii of curvature are equal, by Eqs. (3.2) and (3.3). Again the result follows from the proofs of Theorems 2.1 and 3.1. Finally, if the surfaces are null, then $f(\gamma)\equiv 0$, and we see from the proof of cases (ib) and (ic) in Theorem 3.1 that Eq. (1.2d), and hence Eq. (1.2b), is satisfied. □

Corollary: If a set of surfaces is orthogonal to a congruence of straight lines, and if the surfaces have constant (possibly infinite) principal radii of normal curvature, then they

either have zero total (Gaussian) curvature, or have nonzero total (Gaussian) curvature but equal finite principal radii of normal curvature.

Proof: This follows either from previous results[1] or from an application of Theorems 4.2 and 4.4, the Corollary of Theorem 2.1, and Corollary 2 of Theorem 3.1. □

5. CONCLUSION

We have exhibited all compatible real or complex solutions of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma)\,, \tag{1.2a}$$

$$(\nabla^2\gamma)^2 - \tfrac{1}{2}(N+2)f'(\gamma)\,\nabla^2\gamma + \tfrac{1}{4}(N+1)\,[f'(\gamma)]^2 - \tfrac{1}{2}Nf(\gamma)f'(\gamma) = -N\nabla\gamma\cdot\nabla(\nabla^2\gamma)\,. \tag{1.2b}$$

An interesting relationship was discovered in Theorem 4.4 between the nontrivial solutions ($\gamma \neq$ constant), and the set of surfaces whose total curvature is zero, or whose total curvature is nonzero and whose principal radii of curvature are equal (and finite). An alternative characterization was given in Theorem 4.3, involving the shear and expansion scalars of the normal congruence of a set of surfaces. We have also shown that every solution of the system

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = f(\gamma)\,, \tag{1.1a}$$

$$\nabla^2\gamma = g(\gamma) \tag{1.1b}$$

studied previously[1] is also a solution of the system (1.2), but that the converse is false, so our present results represent a generalization of the previous work. Normalizing $f(\gamma)$ so that $f(\gamma) = 1$ in the case where $f(\gamma) \neq 0$, we can conclude by taking real and imaginary parts of Eqs. (1.2) that we have obtained all solutions to the real set of equations

$$(\nabla X)^2 - (\nabla Y)^2 = \epsilon\,,$$

$$\nabla X\cdot\nabla Y = 0\,,$$

$$(\nabla^2 X)^2 - (\nabla^2 Y)^2 = -N\{\nabla X\cdot\nabla(\nabla^2 X) - \nabla Y\cdot\nabla(\nabla^2 Y)\}\,,$$

and

$$2\nabla^2 X\nabla^2 Y = -N\{\nabla X\cdot\nabla(\nabla^2 Y) + \nabla Y\cdot\nabla(\nabla^2 X)\}\,,$$

where $\epsilon = 0$ or 1, and N is a constant. The results are contained in Theorems 2.1 and 3.1, and summarized in Table I of Sec. 1.

The results of the present article have been applied to a discussion of null surfaces in Minkowski space–time and its complexification. This discussion will appear in a subsequent article. Moreover, there are questions in classical physics which do not appear to have been previously examined, and which are readily answered using the present formalism. For instance, in electrostatics, one could ask for what distributions of charge are the charge density and the (magnitude of the) electric field constant on the equipotential surfaces? (The answers to this question are "for planar, cylindrically symmetric, and spherically symmetric distributions only.") Such questions are currently being investigated.

ACKNOWLEDGMENTS

This work was supported in part by Operating Grant #A3978 from the National Research Council of Canada, and by a University of Waterloo Research Grant.

[1] C.B. Collins, Math. Proc. Cambridge Philos. Soc. **80**, 165 (1976).
[2] C.B. Collins, Math. Proc. Cambridge Philos. Soc. **80**, 349 (1976).
[3] F.G. Friedlander, Proc. Cambridge Philos. Soc. **43**, 360 (1947).
[4] L.P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Ginn and Company, Boston, 1909).
[5] H.T.H. Piaggio, *An Elementary Treatise on Differential Equations and their Applications* (Bell, London, 1960).
[6] Relativistic cosmology, p. 104 in *General Relativity and Cosmology*, Proceedings of the International School of Physics "Enrico Fermi" Course XLVII, 1969, ed. R.K. Sachs (Academic, London and New York, 1971).

Complex potential equations, special relativity, and complexified Minkowski space–time

C. B. Collins
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

(Received 19 September 1978; accepted for publication 18 September 1979)

We obtain a number of results on null geodesic congruences in Minkowski space–time M. It is first shown that the only time-invariant hypersurface-orthogonal shear-free null geodesic congruences in M are those that generate either null hyperplanes or null cones. This result is derived in two interesting and quite distinct ways. One method of proof uses Kerr's theorem. The other method proceeds by showing that the equations specifying the class of congruences can be written in the form $(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = 1$, and $(\nabla^2\gamma)^2 = -2\nabla\gamma\cdot\nabla(\nabla^2\gamma)$, where γ is a potential and the operators ∇ and ∇^2 refer to 3-dimensional Euclidean space. The complexified version of this system of equations has been studied previously by the author [preceding paper, J. Math. Phys. **21**, 240 (1980)]. Hypersurface-orthogonal shear-free null geodesic congruences in complexified Minkowski space–time are then investigated, and a close association is discovered between the set of such congruences and the set of (generally, non-hypersurface-orthogonal) shear-free null geodesic congruences in M.

1. INTRODUCTION

In this article, we shall study shearfree null geodesic congruences in real Minkowski space–time M and in its complexification M^*. In Sec. 2, we show that the only time-invariant hypersurface-orthogonal shearfree null geodesic congruences in M are those which (locally) generate either null hyperplanes or null cones, and we obtain, as an elementary consequence, the result that the only null hypersurfaces in M whose null generators are shearfree are either null hyperplanes or null cones. While this last result appears to be well known, there is, as far as the present author is aware, no well-known standard reference to it (cf. the comments in Ref. 1). Furthermore, we shall derive our results in two interesting and quite distinct ways. The first method will be to show that the equations specifying the class of congruences in question can be written in the form

$$(\nabla\gamma)^2 \equiv \nabla\gamma\cdot\nabla\gamma = 1 \tag{1.1a}$$

and

$$(\nabla^2\gamma)^2 = 2\nabla\gamma\cdot\nabla(\nabla^2\gamma), \tag{1.1b}$$

where γ is a (real) potential, and the operators ∇ and ∇^2 refer to (real) three-dimensional Euclidean space. This system is a specialization of that investigated in Theorem 2.1 of Ref. 2, and so the solutions can readily be found. The fact that the potential γ is time independent reflects the time-invariance property of the congruence. Equation (1.1a) can be regarded as a statement that the hypersurface-orthogonal congruence is null, and Eq. (1.1b) is a statement that the shear of the congruence vanishes.

The second method of proving the result of Sec. 2 uses Kerr's theorem in conjunction with a result obtained in the proof of Theorem 3.1 in Ref. 3. We shall consider the metric of Minkowski space–time in Cartesian-like coordinates (t,x,y,z), viz.,

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \tag{1.2}$$

It will also be convenient to express relationships in terms of a set of complex null coordinates $(u,v,\zeta,\bar{\zeta})$, defined by

$$u = \frac{1}{\sqrt{2}}(z+t),\ v = \frac{1}{\sqrt{2}}(z-t),$$

$$\zeta = \frac{1}{\sqrt{2}}(x+iy),\ \bar{\zeta} = \frac{1}{\sqrt{2}}(x-iy), \tag{1.3a}$$

$$\Longleftrightarrow t = \frac{1}{\sqrt{2}}(u-v),\ x = \frac{1}{\sqrt{2}}(\zeta+\bar{\zeta}),$$

$$y = \frac{-i}{\sqrt{2}}(\zeta-\bar{\zeta}),\ z = \frac{1}{\sqrt{2}}(u+v). \tag{1.3b}$$

In this second set of coordinates, the metric (1.2) becomes

$$ds^2 = 2\,du\,dv + 2\,d\zeta\,d\bar{\zeta}. \tag{1.4}$$

Kerr's theorem[4–8] states that in Minkowski space–time the most general analytic shearfree (future pointing, affinely parametrized) null geodesic congruence $\mathbf{k}$ is given by

either

$$\mathbf{k} = k_a dx^a = -(d\mathbf{u} + \bar{Y}d\zeta + Yd\bar{\zeta} - Y\bar{Y}d\mathbf{v}), \tag{1.5a}$$

or

$$\mathbf{k} = k_a dx^a = d\mathbf{v}, \tag{1.5b}$$

where Y is a complex function of the coordinates, defined implicitly by

$$F \equiv F(Y, u + Y\bar{\zeta}, \zeta - Yv) = 0, \tag{1.5c}$$

with F an arbitrary complex analytic function of its three arguments. If we make the further demand that the congruence be time-invariant and hypersurface-orthogonal, we recover the required result.

We can also examine "time"-invariant hypersurface-orthogonal shearfree null geodesic congruences in complexified Minkowski space–time M^* (e.g., see Refs. 7–9) by complexifying the analysis that led to Eq. (1.1). The complex

 0022-2488/80/020249-07$1.00

version of Eq. (1.1) is a specialization of those investigated in Theorem 3.1 of Ref. 1, and the solutions can be readily found. This is done in Sec. 3. In addition to the complex generalizations of the null hyperplanes and the null cones obtained in the real case, there are more complicated sets of hypersurfaces which are not admitted in the real case. Further study shows that when the normals to these hypersurfaces are expanding, they are quite naturally related to a class of (generally, non-hypersurface-orthogonal) time-invariant expanding shearfree null geodesic congruences in *real* Minkowski space–time M.

It should be noted that our discussion is intended to study *local* properties of congruences. It is to be understood that our results may be valid *only* locally, and that separate qualifications to this effect will not be given.

2. NULL CONGRUENCES IN MINKOWSKI SPACE–TIME

If a vector field $\mathbf{k}$ is everywhere tangent to a null geodesic congruence, affinely parametrized so that[10]

$$k_{a;b}k^{b} = 0,$$

then there is a decomposition of $k_{a;b}$ in terms of certain physically relevant quantities (e.g., Ref. 11):

$$k_{a;b} = \theta_{ab} + \omega_{ab},$$

where $\theta_{ab} = \theta_{(ab)} = k_{(a;b)}$ is the "expansion tensor" and $\omega_{ab} = \omega_{[ab]} = k_{[a;b]}$ is the "vorticity tensor." The expansion tensor can be further decomposed into its trace and tracefree parts:

$$\theta_{ab} = \sigma_{ab} + \tfrac{1}{2}\theta h_{ab},$$

where $\theta = \theta^{a}_{\ a} = k^{a}_{\ ;a}$ is the "expansion scalar," $\sigma_{ab} = \sigma_{(ab)} = k_{(a;b)} - \tfrac{1}{2}k^{c}_{\ ;c}h_{ab}$ is the "shear tensor," and h_{ab} is a (degenerate) projection tensor into the space H orthogonal to (and hence containing) $\mathbf{k}$. As in Ref. 11, we will be primarily concerned with projections into the space S, which consists of equivalence classes of vectors in H that differ only by multiples of $\mathbf{k}$. In this case, given a null vector $\mathbf{l}$ satisfying $\mathbf{k}\cdot\mathbf{l} \neq 0$, there is a unique projection tensor h_{ab} with the properties $h_{ab} = h_{(ab)}$, $h_{ab}X^{b} \in S$ for all vectors $\mathbf{X}$, and $h_{ab}X^{b} = X_{a}$ for all $\mathbf{X} \in S$. The null geodesic $\mathbf{k}$ congruence is hypersurface orthogonal if and only if $k_{[a;b}k_{c]} = 0 \Longleftrightarrow \omega = 0$, where $2\omega^{2} \equiv \omega_{ab}\omega^{ab} = k_{[a;b]}k^{a;b} = 0$, and is shearfree if and only if $\sigma = 0$, where σ is the "shear scalar," given by $2\sigma^{2} \equiv \sigma_{ab}\sigma^{ab} = k_{(a;b)}k^{a;b} - \tfrac{1}{2}(k^{a}_{\ ;a})^{2}$, $\sigma \geqslant 0$. We note that the condition $\sigma_{ab} = 0$ is not invariant, since it depends on the choice of h_{ab}, which in turn depends on the choice of the null vector $\mathbf{l}$.

We note in passing that if $\mathbf{k}$ is tangent to a (affinely parametrized) null geodesic congruence, then so also is $f\mathbf{k}$, where f is constant along each curve of the congruence. If, moreover, the $\mathbf{k}$ congruence is shearfree or hypersurface-orthogonal, then so also is the $f\mathbf{k}$ congruence. Furthermore, if the $\mathbf{k}$ congruence is hypersurface-orthogonal, there exists an $f\mathbf{k}$ congruence whose vorticity tensor vanishes.

Given any Killing vector $\boldsymbol{\xi}$, a null geodesic congruence with tangent vector $\mathbf{k}$ is "invariant under the action of $\boldsymbol{\xi}$" if and only if

$$\mathscr{L}_{\xi}\mathbf{k} = \mathbf{0} \Longleftrightarrow [\boldsymbol{\xi},\mathbf{k}] = \mathbf{0} \Longleftrightarrow k_{a;b}\xi^{b} - \xi_{a;b}k^{b} = 0. \tag{2.1}$$

Lemma 2.1: Any null hypersurface in Minkowski space–time is generated by a (unique) null geodesic congruence.

Proof: Let $g(t,x,y,z) = 0$ be the equation of the null hypersurface. Then, defining $\mathbf{k}$ to be the gradient of g, we have $k_{a} = g_{;a}$, where

$$k_{a}k^{a} = 0 \Longleftrightarrow g_{;a}g^{;a}, \text{ and } k_{[a;b]} = 0 \Longleftrightarrow g_{[;a;b]} = 0.$$

Extending the gradient vector field to obtain a null vector field $\mathbf{k}$ defined in a neighborhood of the hypersurface, we see that $k_{a}k^{a} = 0 \Rightarrow k_{a;b}k^{a} = 0$. Thus, on the hypersurface, $k_{b;a}k^{a} = 0$, and so the integral curves of $\mathbf{k}$ form a (hypersurface-orthogonal) null geodesic congruence in the hypersurface.

Since the vector $\mathbf{k}$ is orthogonal to any vector $\mathbf{v}$ tangential to the hypersurface at a point P, it follows that $\mathbf{v}$ is either null or spacelike, and that, if $\mathbf{v}$ is null, it is parallel to $\mathbf{k}$ at P. This property holds at all points P of the hypersurface. Consequently, if a second null congruence $\boldsymbol{\kappa}$ generates the hypersurface, then $\kappa_{a} = fk_{a}$ for some nonzero function f, and $\kappa_{a;b}\kappa^{b} = (f_{;b}k^{b})\kappa_{a}$, i.e., the integral curves of $\boldsymbol{\kappa}$ form a (generally, non-affinely parametrized) null geodesic congruence which coincides with that of the integral curves of $\mathbf{k}$. □

Remark: We will say that a null hypersurface $g(t,x,y,z) = 0$ in M is "shearfree" if an affinely parametrized null geodesic congruence that generates it is shearfree, i.e., $g_{;a;b}g^{;a;b} = \tfrac{1}{2}(g^{;c}_{;c})^{2}$ on the hypersurface. By Lemma 2.1, we see that there is an affinely parametrized null geodesic congruence generating every null hypersurface, and by the proof of Lemma 2.1, and the remarks preceding it, we see that the definition is independent of which particular affine parametrization is used. We will be particularly concerned with (hypersurface-orthogonal shearfree null geodesic) congruences that are invariant under the action of a timelike translational Killing vector (cf. Ref. 4). For brevity, such congruences will be termed "time-invariant." Not every time-invariant shearfree null geodesic congruence in M is necessarily hypersurface orthogonal, as we will show in the proof of Theorem 2.2.

Theorem 2.2: Let $\mathbf{k}$ be a vector field everywhere tangential to a time-invariant hypersurface-orthogonal shearfree null geodesic congruence in Minkowski space–time. Then there is a system of "Cartesian" coordinates (t,x,y,z) such that $\mathbf{k}$ is orthogonal to (and hence generates) the family of null hypersurfaces $\{t - q(x,y,z) = \text{constant}\}$, where either (i) $q = lx + my + nz + A$, with l,m,n, and A constants satisfying $l^{2} + m^{2} + n^{2} = 1$: in this case, the expansion scalar vanishes, and the hypersurfaces are null hyperplanes; or (ii) $q = \pm[(x - x_{1}$ null cones.

Proof: Without loss of generality, the $\mathbf{k}$ congruence is affinely parametrized. Suppose that this congruence is invariant under the action of a timelike translational Killing vector $\boldsymbol{\xi}$. We can choose coordinates (t,x,y,z) in Eqs. (1.2) such that $\boldsymbol{\xi} = \partial/\partial t$. By the proof of Lemma 2.1, there is a real function $g(t,x,y,z) \neq 0$ such that $k_{a} = g_{;a}$. Moreover, by Eq. (2.1), $g_{;a;b}\xi^{b} = 0$. This implies that there is a constant b and a real function $Q(x,y,z)$ such that

$$g(t,x,y,z) = bt + Q(x,y,z). \tag{2.2}$$

The condition that $\mathbf{k}$ is null may be written in the form

$$k_a k^a = 0 \Longleftrightarrow g_{;a} g^{;a} = 0 \Longleftrightarrow (\nabla Q)^2 \equiv \nabla Q \cdot \nabla Q = b^2, \tag{2.3}$$

where ∇ is an operator in the three-dimensional Euclidean space $\{t = \text{constant}\}$. The expansion scalar θ of the $\mathbf{k}$ congruence is

$$\theta = k^a_{\ ;a} = g^{;a}_{\ ;a} = \nabla^2 Q, \tag{2.4}$$

and the shear scalar σ is given by

$$\begin{aligned} \sigma^2 &= \tfrac{1}{2} k_{(a;b)} k^{a;b} - \tfrac{1}{4} (k^a_{\ ;a})^2 \\ &= \tfrac{1}{4} [2 g_{;a;b} g^{;a;b} - (g^{;a}_{\ ;a})^2] \\ &= -\tfrac{1}{4} [(\nabla^2 Q)^2 + 2 \nabla Q \cdot \nabla (\nabla^2 Q)], \end{aligned} \tag{2.5}$$

where the operators ∇ and ∇^2 act in the three-dimensional Euclidean space $\{t = \text{constant}\}$.

Thus, from Eqs. (2.3) and (2.5), any time-invariant hypersurface-orthogonal shearfree null geodesic congruence in Minkowski space–time can be associated with the solution Q to the following system of potential equations:

$$(\nabla Q)^2 \equiv \nabla Q \cdot \nabla Q = b^2 \tag{2.6a}$$

and

$$(\nabla^2 Q)^2 = -2 \nabla Q \cdot \nabla (\nabla^2 Q), \tag{2.6b}$$

where b is a constant. Since Q is real, the case $b = 0$ would require that Q be zero, and then, from Eq. (2.2), we obtain $g \equiv 0$, which is inadmissible. Consequently, $b \neq 0$ in Eq. (2.6a).

The system (2.6) is a particular case of the system of potential equations (1.2) previously investigated.[2] The general solution is therefore obtained by writing $q = Q/b$ and employing Theorem 2.1 of Ref. 2, and the theorem is proved. □

Corollary 1: In Minkowski space–time, any time-invariant hypersurface-orthogonal shearfree null geodesic congruence can be associated with a family of equipotentials in three-dimensional Euclidean space. In case (i) of Theorem 2.2, these equipotentials are planes (the principal curvatures are zero; the normal geodesic congruence satisfies $\theta = \sigma = 0$), and in case (ii) they are spheres (the principal curvatures are equal and nonzero; the normal geodesic congruence satisfies $\theta \neq 0 = \sigma$).

Proof: This follows immediately from Table I of Ref. 2. We note that the quantities θ and σ refer both to the space–time congruence and to the auxiliary congruence of rays orthogonal to the equipotentials in Euclidean space. □

Corollary 2: In Minkowski space–time a null hypersurface H is shearfree if and only if it is either a null hyperplane or a null cone.

Proof: Suppose H is shearfree. We consider Minkowski space–time M with coordinates (t,x,y,z) of metric (1.2), and construct a family of hypersurfaces $H(\nu)$ parallel to H as follows:

$$H(\nu) = \{(t,x,y,z) \in M \,|\, (t - \nu, x, y, z) \in H\},$$

where ν is a parameter that is constant on each surface, and $H(0) = H$. We suppose that H is given by the equation $g(t,x,y,z) \equiv t - f(x,y,z) = 0$. Clearly, for each ν, the null geodesic generators of $H(\nu)$ are parallel to those of H, and are shearfree. We now have a shearfree null geodesic congruence defined throughout M. This congruence is time-invariant, since $\mathcal{L}_\xi \mathbf{k} = \mathbf{0}$, where $\xi = \partial/\partial t$ and $k_a = g_{;a}$, the function g here being $g(t - \nu, x, y, z)$, whose vanishing specifies the hypersurface $H(\nu)$. By Theorem 2.2, this congruence is orthogonal to (and hence generates) a system of hypersurfaces which are either null hyperplanes or null cones, i.e., the family $H(\nu)$ is a system of either null hyperplanes or null cones, and, *a fortiori*, $H = H(0)$ is either a null hyperplane or a null cone.

Conversely, given any null hyperplane or null cone, it is clear by direct calculation that its null geodesic generators are shearfree. □

There is an alternative proof of Theorem 2.2, which uses Kerr's theorem.[4-8] This proof will now be given.

Alternative Proof of Theorem 2.2: Kerr's theorem states the form of the most general analytic (future-pointing affinely parametrized) shearfree null geodesic congruence in M [see Eq. (1.4) and (1.5)].

If Eq. (1.5b) is obeyed, i.e., if $\mathbf{k} = d\mathbf{v}$, then the vector field $\mathbf{k}$ is clearly orthogonal to the hyperplanes $\{t - z = \text{constant}\}$. These intersect the hypersurfaces $\{t = \text{constant}\}$ in the planes $\{z = \text{constant}\}$.

If, on the other hand, Eq. (1.5a) is obeyed, i.e., if $\mathbf{k} = -(d\mathbf{u} + \bar{Y} d\zeta + Y d\bar{\zeta} - Y\bar{Y} d\mathbf{v})$, then $\mathbf{k}$ is hypersurface-orthogonal if and only if $\mathbf{k} \wedge d\mathbf{k} = \mathbf{0}$, which is equivalent to the following set of conditions:

$$\bar{Y} Y_v - Y \bar{Y}_v + \bar{Y} (Y\bar{Y})_{\bar{\zeta}} - Y (Y\bar{Y})_{\bar{\zeta}} = 0, \tag{2.7a}$$

$$Y_v + (Y\bar{Y})_\zeta - Y (Y\bar{Y})_u + Y\bar{Y} Y_u = 0, \tag{2.7b}$$

$$\bar{Y}_v + (Y\bar{Y})_\zeta - \bar{Y} (Y\bar{Y})_u + Y\bar{Y}\, \bar{Y}_u = 0 \tag{2.7c}$$

and

$$\bar{Y}_{\bar{\zeta}} - Y_{\bar{\zeta}} + \bar{Y} Y_u - Y \bar{Y}_u = 0,$$

where a suffix denotes partial differentiation. Forming $\bar{Y} \times$ Eq. (2.7b) $- Y \times$ Eq. (2.7c) and using Eq. (2.7a), we find that

$$Y\bar{Y} (\bar{Y} Y_u - Y \bar{Y}_u) = 0. \tag{2.8a}$$

Consequently, either $Y \equiv 0$, or $Y \not\equiv 0$ and $Y_u / Y = \bar{Y}_u / \bar{Y}$. In either case, Eq. (2.7d) shows that

$$Y_\zeta = \bar{Y}_{\bar{\zeta}}. \tag{2.8b}$$

Recalling that Y is defined implicitly by the relation

$$F(Y, X_1, X_2) = 0, \tag{2.9}$$

where $X_1 \equiv u + Y\bar{\zeta}$ and $X_2 \equiv \zeta - Yv$, we find that $Y_{\bar{\zeta}} = Y Y_u = -Y F_{X_1}/D$ and $Y_v = -Y Y_\zeta = Y F_{X_2}/D$, where $D \equiv F_Y + F_{X_1} \bar{\zeta} - F_{X_2} v \not\equiv 0$. Direct substitution of these relations into Eqs. (2.7) shows that they are identically satisfied if and only if Eq. (2.8) hold, i.e., if and only if either $Y \equiv 0$, or $Y \not\equiv 0$ and both F_{X_1}/YD and F_{X_2}/D are real. Moreover, we find that $\mathbf{k}$ is invariant under the timelike translational Killing vector $\xi = \partial/\partial t$ if and only if $\partial Y/\partial t = 0 \Longleftrightarrow$ nce Y_u generates) the hyperplanes $\{z + t = \text{constant}\}$. These intersect the hypersurfaces $\{t = \text{constant}\}$ in the planes $\{z = \text{constant}\}$.

Now suppose that $Y \not\equiv 0$. If $F_{X_2} \equiv 0$, then $F_{X_1} \equiv 0$ and Eq. (1.5c) requires that Y is identically constant. In this case, $\mathbf{k}$ is orthogonal to (and hence generates) the hyperplanes $\{u + \bar{Y}\zeta + Y\bar{\zeta} - Y\bar{Y} v = \text{constant}\}$. These intersect the hypersurfaces $\{t = \text{constant}\}$ in the planes $\{(Y + \bar{Y})x$

$-i(Y-\bar{Y})y+(1-Y\bar{Y})z=\text{constant}\}$. If $F_{X_2}\neq 0$, then, without loss of generality, we can arrange for F in Eq. (1.5c) to be of the form $F=G(Y,X)-X_2$, where G is an analytic function of Y and X_1; in other words, without loss of generality, $F_{X_2}=-1$. Now $G_{X_1}=Y$, and $G_Y+Y\bar{\zeta}$ is real. Thus, G is of the form $G(Y,X_1)=X_1Y-(\sqrt{2})\phi(Y)$, where $\phi(Y)$ is such that $D=X_1-(\sqrt{2})\phi'(Y)+Y\bar{\zeta}+v$ is real. In this case, using Eqs. (1.3), Eq. (1.5c) can be written as

$$-\frac{1}{\sqrt{2}}F(Y,u+Y\bar{\zeta},\zeta-Yv)\equiv H(Y,x,y,z)$$

$$\equiv\phi(Y)+\tfrac{1}{2}(1-Y^2)x+\tfrac{1}{2}i(1+Y^2)y-Yz=0. \quad (2.10)$$

We require all functions $\phi(Y)$ in Eq. (2.10) with the property that $H_Y=H_Y(x,y,z)=\phi'(Y)-Y(x-iy)-z$ is real. This problem can be readily solved when we observe that Eq. (2.10) is the same as Eq. (4.5) of Ref. 3. The analysis subsequent to Eq. (4.5) of Ref. 3 now applies, and we can therefore deduce that H_Y satisfies

$$(\nabla H_Y)^2=1 \quad (2.11a)$$

and

$$\nabla^2 H_Y=2/H_Y, \quad (2.11b)$$

where H_Y is real. Equations (2.11) are a special case of the system discussed in Theorem 3.1 of Ref. 3, and it follows that the general solution is

$$H_Y=\pm[(x-x_1)^2+(y-y_1)^2+(z-z_1)^2]^{1/2}+A,$$

where x_1, y_1, z_1, and A are constants. Since the normals to the equipotential surfaces $\{H_Y=\text{constant}\}$ possess a shear tensor which has three linearly independent eigenvectors, it follows[2,3] that $\phi'''(Y)=0$, i.e., that $\phi(Y)$ is of the form $\phi(Y)=a+bY+c$, $Y^2$for constants a, b, and c. This function is generated from $\phi(Y)\equiv 0$ in Eq. (2.10) by the shift of origin $x\to x-(a-c)$,$y\to y-i(a+c)$, and $z\to z-b$. In order to determine the hypersurfaces generated by $\mathbf{k}$, it is therefore sufficient to put $\phi(Y)=0$ in Eq. (2.10), which then implies

$$Y=\frac{-z\pm r}{x-iy},\quad r=(x^2+y^2+z^2)^{1/2}.$$

Employing Eq. (1.5a), a direct calculation, given in the Appendix, shows that $k_a=fg_{,a}$, where $g=g\{[(u-v)/\sqrt{2}]\pm[2\zeta\bar{\zeta}+\tfrac{1}{2}(u+v)^2]^{1/2}\}=g(t\pm r)$, and $[1\pm(z/r)]\times fg'(t\pm r)=-\sqrt{2}$ (note that f is constant along the generators $\mathbf{k}$ in keeping with our earlier remarks). Thus, $\mathbf{k}$ generates the hypersurfaces $\{t\pm(x^2+y^2+z^2)^{1/2}=\text{constant}\}$, i.e., a family of null cones whose vertices lie on the t axis. Now invoking the allowed shift of origin, we see that the general family consists of null cones whose vertices lie on a line $\{(x_0,y_0,z_0)=\text{constant}\}$, where $x_0=(a-c)$, $y_0=i(a+c)$, and $z_0=b$ are arbitrary real numbers, i.e., b is an arbitrary real constant and $a=\bar{c}$ is an arbitrary complex constant. These cones intersect the hypersurfaces $\{t=\text{constant}\}$ in spheres centered on (x_0,y_0,z_0). □

Remark: These results are consistent with those of Cox,[4] who has derived expressions for the most general analytic shearfree null geodesic congruence $\mathbf{k}$ in Minkowski space–time that is either (i) invariant under the action of a translational Killing vector $\partial/\partial t$ or (ii) invariant under the action of a spacelike rotational Killing vector $x(\partial/\partial y)-y(\partial/\partial x)$, or (iii) both (i) and (ii). Unfortunately, there are errors in these expressions. In our notation, Cox has shown that in case (i) (Theorem I of Ref. 4), $\mathbf{k}$ is given by $\mathbf{k}=d\mathbf{v}$ or $\mathbf{k}=-(d\mathbf{u}+\bar{Y}d\zeta+Yd\bar{\zeta}-Y\bar{Y}d\mathbf{v})$, where Y is an arbitrary complex function of $\zeta-Yv-Y(u+Y\bar{\zeta})$, in which case it is concluded that there is an analytic function $\phi(Y)$ such that $\phi(Y)+(1/\sqrt{2})(\zeta-Yv)-(1/\sqrt{2})Y(u+Y\bar{\zeta})=0$. However, the case where Y is identically constant appears to have been overlooked. Similarly, in case (iii) (Theorem III of Ref. 4), we have either $\mathbf{k}=d\mathbf{v}$ or $\mathbf{k}=-(d\mathbf{u}+\bar{Y}d\zeta+Yd\bar{\zeta}-Y\bar{Y}d\mathbf{v})$, where Y is a complex function of the coordinates, defined implicitly by $F(Y,u+Y\bar{\zeta},\zeta-Yv)\equiv -iaY+(1/\sqrt{2})(\zeta-Yv)-(1/\sqrt{2})Y(u+Y\bar{\zeta})=0$, where a is real, together with the *additional possibility* that Y could vanish identically. We see from Eq. (2.10) that if *in addition* the congruence is hypersurface orthogonal, then it is given by either $\mathbf{k}=d\mathbf{v}$ or $\mathbf{k}=-d\mathbf{u}$, or $a=0$. It is well known in the theory of Kerr–Schild space–time (e.g., Refs. 6,12, and 13) that, when considered with respect to an auxiliary null geodesic congruence in Minkowski space–time, the case $a\neq 0$ generates the Kerr[14] solution, and the case $a=0$ specializes the solution to that of Schwarzschild. The parameter a is interpreted as the angular momentum/unit mass, and there is thus a close connection between the two hypersurface-orthogonality properties in the curved Kerr–Schild space–time and in the auxiliary flat space–time. In fact, Newman and Winicour[1] have already obtained a result to this effect. They show that the "Kerr congruence" [i.e., the shearfree null geodesic congruence generated in the case when $F(Y,u+Y\zeta,\zeta-Yv)=-iaY+(1/\sqrt{2})(\zeta-Yv)-(1/\sqrt{2})Y(u+Y\bar{\zeta})=0$, where a is real] can be interpreted as a complexified version of the "Schwarzschild congruence" (i.e., the special case when $a=0$). This is also apparent from our discussion above.

3. COMPLEX MINKOWSKI SPACE–TIME

In this section, we examine the complexification of our results, and discover a close relationship between time-invariant shearfree null geodesic congruences in real Minkowski space–time M and "time"-invariant *hypersurface-orthogonal* shearfree null geodesic congruences in *complexified* Minkowski space–time M^*. We allow the coordinates (t,x,y,z) of Eqs. (1.2) to take complex values. In this case, ζ and $\bar{\zeta}$ as defined by Eqs. (1.3) are no longer complex conjugate. Null congruences and null hypersurfaces are defined as direct generalizations of their real counterparts, where the property of being null is with respect to the (complex, non-Hermitian) metric given by Eqs. (1.2). A congruence with tangent vector $\mathbf{k}$ will be called ' "time"-invariant' if there is a nonnull translational Killing vector $\boldsymbol{\xi}$ such that $\mathcal{L}_{\boldsymbol{\xi}}\mathbf{k}=\mathbf{0}$ [cf. Eq. (2.1)]. Since $\boldsymbol{\xi}\cdot\boldsymbol{\xi}$ is a (complex) nonzero constant, $\boldsymbol{\xi}$ may be rescaled so that $\boldsymbol{\xi}\cdot\boldsymbol{\xi}=-1$, and there is then a coordinate system (t,x,y,z) such that $\boldsymbol{\xi}=\partial/\partial t$.

Lemma 2.1 generalizes from M to M^*, to the extent that any null hypersurface in M^* is generated by a null geodesic congruence. However, this congruence is no longer unique.

For instance, the hypersurface $t = x$ is generated by a set of vectors $\delta_0^a + \delta_1^a$, and also by a set of vectors $\delta_2^a + i\delta_3^a$.

Theorem 2.2 generalizes as follows.

Theorem 3.1: Let $\mathbf{k}$ be a vector field everywhere tangential to a "time"-invariant hypersurface-orthogonal shearfree null geodesic congruence in complexified Minkoski space–time. Then there is a system of "Cartesian" coordinates (t,x,y,z) such that $\mathbf{k}$ is orthogonal to (and generates) the family of null hypersurfaces $\{\epsilon t - q(x,y,z) = \text{constant}\}$, where either (i) $\epsilon = 1$ and $q = lx + my + nz + A$, with l,m,n, and A complex constants satisfying $l^2 + m^2 + n^2 = 1$; in this case the expansion scalar vanishes and the hypersurfaces are null hyperplanes; or (ii) $\epsilon = 1$ and $q = \pm[(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2]^{1/2} + A$, with x_1, y_1, z_1, and A complex constants; in this case, the expansion scalar is nonzero and the hypersurfaces are null cones; or (iii) $\epsilon = 1$ and $q = \mu_1 x + \mu_2 y + \mu_3 z + \psi(\lambda_1 x + \lambda_2 y + \lambda_3 z)$, with ψ arbitrary, $\psi'' \not\equiv 0$, and with $\mu_1, \mu_2, \mu_3, \lambda_1, \lambda_2$, and λ_3 complex constants satisfying $\mu_1^2 + \mu_2^2 + \mu_3^2 = 1$, $\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 0$, $\lambda_1\mu_1 + \lambda_2\mu_2 + \lambda_3\mu_3 = 0$, and λ_1, λ_2, and λ_3 are not all zero; in this case the expansion scalar is zero, but the hypersurfaces are not null hyperplanes; or (iv) $\epsilon = 1$ and $q = (\partial F/\partial\beta) + A(\beta)$, where $F(\beta,x,y,z) \equiv \phi(\beta) + \frac{1}{2}[(1 - \beta^2)x + i(1 + \beta^2)y - 2\beta z = 0$; $A(\beta)$ and $\phi(\beta)$ are arbitrary, except that $A'(\beta)$ and $\phi'''(\beta)$ are not simultaneously zero; or (v) $\epsilon = 0$ and $\psi(q) = l(q)x + m(q)y + n(q)z$, where ψ, l, m, and n are arbitrary, with $l^2 + m^2 + n^2 = 0$ and l,m, and n are not all zero; or (vi) $\epsilon = 0$ and q is the (α,β) eliminant of

$$x = i\beta \sin\alpha + \mu(\alpha,\gamma),$$

$$y = i\beta \cos\alpha - \int^{\alpha} \mu_\alpha(a,\gamma) \tan a\, da + \alpha_0(\gamma),$$

and

$$z = \beta,$$

where $\mu(\alpha,\gamma)$ and $a_0(\gamma)$ are arbitrary functions.

Proof: The argument in the proof of Theorem 2.2 generalizes to the complex case to yield the complexification of Eqs. (2.6), viz.,

$$(\nabla Q)^2 \equiv \nabla Q \cdot \nabla Q = b^2 \tag{3.1a}$$

and

$$(\nabla^2 Q)^2 = -2\, \nabla Q \cdot \nabla(\nabla^2 Q), \tag{3.1b}$$

where b is a constant, and $g(t,x,y,z) = bt + Q(x,y,z)$ is such that $k_a = g_{;a}$. Since Q is no longer required to be real, the case $b = 0$ in Eq. (3.1a) is now admissible.

The system (3.1) is a particular case of the system of potential equations (1.2) previously investigated.[2] The general solution is therefore obtained by employing Theorem 3.1 of Ref. 2 (and writing $a = Q/b$ if $b \neq 0$), and the theorem is proved. □

Corollary 1: In complexified Minkowski space–time, any "time-invariant" hypersurface-orthogonal shearfree null geodesic congruence can be associated with a family of equipotentials in complexified three-dimensional Euclidean space. In case (i) of Theorem 3.1, these equipotentials are planes (the principal curvatures are zero; the normal geodesic congruence satisfies $\theta = \sigma = 0$); in case (ii) they are spheres (the principal curvatures are equal and nonzero; the normal geodesic congruence satisfies $\theta \neq 0 = \sigma$); in case (iii) the principal curvatures of the equipotentials are zero and the normal geodesic congruence satisfies $\theta = \sigma = 0$; in case (iv) the principal curvatures of the equipotentials are equal and nonzero and the normal geodesic congruence satisfies $\theta \neq 0 = \sigma$; in case (v) the equipotentials are null planes (i.e., elements of the set[2,3] $\mathscr{S}$) and the normal geodesic congruence satisfies $\theta = \sigma = 0$; and in case (vi) the equipotentials are nonplanar null surfaces (i.e., elements of the set[2,3] $\mathscr{I}$ and the normal geodesic congruence satisfies $\theta \neq 0 = \sigma$.

Proof: This follows immediately from Table I of Ref. 2. We note that the quantities θ and σ refer both to the space–time congruence and to the auxiliary congruence of rays orthogonal to the equipotentials in complex Euclidean space. □

Remark: As observed in Ref. 2, case (ii) is recoverable from case (iv) by the specialization $\phi'''(\beta) = A(\beta) = 0$.

Our main result is as follows:

Theorem 3.2: There is a mapping χ of the set S of expanding (generally, non-hypersurface-orthogonal) shearfree null geodesic $\mathbf{k}$ congruences in real Minkowski space–time that are invariant under a timelike translational Killing vector ξ into the set $\tilde{S}$ in complexified Minkowski space–time M^* of expanding hypersurface-orthogonal shearfree null geodesic $\tilde{\mathbf{k}}$ congruences that are invariant under the action of $\tilde{\xi}$ (the complex analytic extension of ξ), and which satisfy $\tilde{\mathbf{k}} \cdot \tilde{\xi} \neq 0$. Conversely, there is a mapping ψ of the set $\tilde{S}$ of such congruences into the set S of congruences in M.

Furthermore, there is an equivalence relation R on the set $\tilde{S}$ such that the maps $\chi: S \to \tilde{S}/R$ and $\psi: \tilde{S}/R \to S$ are bijective; $\psi \circ \chi =$ identity on S and $\chi \circ \psi =$ identity on $\tilde{S}$.

Proof: (a) We first prove the existence of the maps χ and ψ:

(i) Suppose that we have an expanding shearfree null geodesic congruence in real Minkowski space–time, whose affinely parametrized tangent vector is $\mathbf{k}$. Suppose further that this congruence is invariant under a timelike translational Killing vector ξ. Choose coordinates (t,x,y,z) in Eqs. (1.2) such that $\xi = \partial/\partial t$ (without loss of generality, ξ is a unit vector). Then, by Theorem I of Ref. 4 (in our notation; see also the remark after Theorem 2.2 above), there is a set of coordinates $(u,v,\zeta,\bar{\zeta})$ such that either $k = d\mathbf{v}$ of $\mathbf{k} = -(d\mathbf{u} + \bar{Y} d\zeta + Y d\bar{\zeta} - Y\bar{Y} d\mathbf{v})$, where either Y is constant, or $Y = Y(x,y,z)$ is a complex function, defined implicitly by

$$-\frac{1}{\sqrt{2}} F(Y, u + Y\bar{\zeta}, \zeta - Yv)$$

$$\equiv \phi(Y) + \frac{1}{\sqrt{2}}(\zeta - Yv) - \frac{1}{\sqrt{2}} Y(u + Y\bar{\zeta}) = 0, \tag{3.2}$$

where ϕ is an arbitrary analytic function of Y. The cases where $\mathbf{k} = d\mathbf{v}$ or where Y is constant are inadmissible, since then the congruence is nonexpanding. Define $q = q(x,y,z) \equiv -(1/\sqrt{2})F_Y = \phi'(Y) - (1/\sqrt{2})(u + v + 2Y\bar{\zeta})$. By the remarks following Eq. (2.10), we obtain $(\nabla q)^2 = 1$ and $\nabla^2 q = 2/q$ [which implies $(\nabla^2 q)^2 = -2\nabla q \cdot \nabla(\nabla^2 q)$; cf. Eq. (2.6)]. The hypersurfaces $\{t - q = \text{constant}\}$ are null and, in general, complex [from the remarks following Eq. (2.10), they are real if and only if $\phi(Y) \equiv \bar{c}Y^2 + bY + c$ for some complex constant c and real constant b]. Moreover, the (complex) vector field $\tilde{\mathbf{k}} = d\mathbf{t} - d\mathbf{q}$ is null, hypersurface-orthogonal, invariant under the action of $\tilde{\boldsymbol{\xi}} = \partial/\partial t$, satisfies $\mathbf{k}\cdot\tilde{\boldsymbol{\xi}} \neq 0$, and, from the arguments used in the proofs of Lemma 2.1 and Theorems 2.2 and 3.1, is geodesic, shearfree, and has nonzero expansion scalar. The procedure outlined above serves to define the map χ.

(ii) Suppose conversely that we have an expanding hypersurface-orthogonal shearfree null geodesic congruence in complexified Minkowski space–time that is invariant under the action of $\tilde{\boldsymbol{\xi}} = \partial/\partial t$, whose affinely parametrized tangent vector is $\tilde{\mathbf{k}}$, where $\tilde{\mathbf{k}}\cdot\tilde{\boldsymbol{\xi}} \neq 0$. As in the proof of Theorem 3.1, we may write $\tilde{k}_a = g_{;a}$ for some (complex) function g, where $g(t,x,y,z) = bt + Q(x,y,z)$ and

$$(\nabla Q)^2 \equiv \nabla Q \cdot \nabla Q = b^2 \tag{3.3a}$$

and

$$(\nabla^2 Q)^2 = -2\,\nabla Q \cdot \nabla(\nabla^2 Q), \tag{3.3b}$$

where $b = \tilde{\mathbf{k}}\cdot\tilde{\boldsymbol{\xi}}$ is a nonzero constant.

By Theorem 3.1 and its Corollary, there are two classes of solution to Eqs. (3.3) for which $\theta \neq 0 = \sigma$. These were labeled (ii) and (iv) in Theorem 3.1. As we observed in Ref. 2 and the remark following Theorem 3.1, we can recover solution (ii) from (iv) by the specialization $\phi'''(\beta) = A'(\beta) = 0$. This means that in either case we can construct a function $Y \equiv \beta$ satisfying Eq. (3.2). We now define the real vector $\mathbf{k} = -(d\mathbf{u} + \bar{Y}d\zeta + Yd\bar{\zeta} - Y\bar{Y}d\mathbf{v})$. The integral curves of the $\mathbf{k}$ congruence form an (affinely parametrized) expanding shearfree null geodesic congruence in real Minkowski space–time (this follows from Sec. 2). Moreover, the congruence is invariant under the action of $\boldsymbol{\xi} = \partial/\partial t$, since $Y = Y(x,y,z)$, and the congruence is hypersurface-orthogonal if and only if $\phi(Y) \equiv \bar{c}Y^2 + bY + c$ for some complex constant c and real constant b. The procedure thus described serves to define the map ψ.

(b) We have thus exhibited the existence of the maps χ and ψ and we now examine their properties:

We define a relation R on the set $\tilde{S}$ by the requirement that any two members of $\tilde{S}$ are related by R if and only if, when expressed in terms of solutions to Eqs. (3.3), the functions $\phi(\beta)$ are identical [but the functions $A(\beta)$ may differ]. This relation is clearly an equivalence relation. Moreover, by the above construction of the maps χ and ψ, it is clear that $\psi\circ\chi$ is the identity on S, and that $\chi\circ\psi$ is the identity on $\tilde{S}$, from which it follows that χ and ψ are bijective maps. □

ACKNOWLEDGMENTS

This work was supported in part by Operating Grant #A3978 from the National Research Council of Canada, and by a University of Waterloo Research Grant. I am very grateful to the referee for pointing out important errors in the original manuscript.

APPENDIX. GENERATION OF NULL CONES

We show that if $Y = (-z \pm r)/(x - iy)$, where $r = (x^2 + y^2 + z^2)^{1/2}$, then $k_a = fg_{;a}$, where $g = g\{[(u - v)/\sqrt{2}] \pm [2\zeta\bar{\zeta} + \frac{1}{2}(u + v)^2]^{1/2}\} = g(t \pm r)$, and $[1 \pm (z/r)] \times fg'(t \pm r) = -\sqrt{2}$, in agreement with the claim made in Sec. 2. This is done by employing Eq. (1.5a), which requires that

$$fg_u = -1, \tag{A1}$$

$$fg_v = Y\bar{Y}, \tag{A2}$$

$$fg_\zeta = -\bar{Y}, \tag{A3}$$

and

$$fg_{\bar{\zeta}} = -Y, \tag{A4}$$

where Eqs. (A3) and (A4) imply that $f\zeta g_\zeta = -\zeta\bar{Y} = -\bar{\zeta}Y = f\bar{\zeta}g_{\bar{\zeta}}$, from which we get $g = g(\zeta\bar{\zeta},u,v)$. Let $U = u + v, V = u - v$ and $W = \sqrt{[4\zeta\bar{\zeta} + (u + v)^2]}$. Then Eqs. (A1) + (A2) − (A3) shows that $g_U = 0$. Equations (A1) and (A3) become

$$f\left(g_W \frac{U}{W} + g_V\right) = -1$$

and

$$f\frac{g_W}{W}(W^2 - U^2) = U \mp W.$$

Eliminating f, we obtain

$$Ug_W + Wg_v = (U \pm W)g_W,$$

from which we get

$$g_V = \pm g_W.$$

Thus, g is a function of $V \pm W = u - v \pm \sqrt{[4\zeta\bar{\zeta} + (u + v)^2]}$ only, i.e., $g = g\{[(u - v)/\sqrt{2}] \pm [2\zeta\bar{\zeta} + \frac{1}{2}(u + v)^2]^{1/2}\} = g(t \pm r)$, where $r = (x^2 + y^2 + z^2)^{1/2}$. Direct substitution of this result into Eq. (A1) shows that $[1 + (z/r)]fg'(t \pm r) = -\sqrt{2}$.

[1] E.T. Newman and J. Winicour, J. Math. Phys. **15**, 426 (1974).
[2] C.B. Collins, J. Math. Phys. **21**, 240 (1980).
[3] C.B. Collins, Math. Proc. Cambridge Philos. Soc. **80**, 165 (1976).
[4] D. Cox, J. Math. Phys. **18**, 1188 (1977).
[5] D. Cox and E.J. Flaherty, Jr., Commun. Math. Phys. **47**, 75 (1976).
[6] G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).

[7]R.O. Hansen and E.T. Newman, Gen. Relativity Gravitation **6**, 361 (1975).
[8]R. Penrose, Int. J. Theor. Phys. **1**, 61 (1968).
[9]E.T. Newman, J. Math. Phys. **14**, 774 (1973).
[10]A semicolon denotes convariant derivative, and, as usual, indices are raised and lowered using the metric tensor. The summation notation for repeated indices is employed throughout. Round brackets denote symmetrization and square brackets denote antisymmetrization.
[11]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge University, London and New York, 1973).
[12]C.B. Collins, Math. Proc. Cambridge Philos. Soc. **80**, 349 (1976).
[13]M.M. Schiffer, R.J. Adler, J. Mark, and C. Sheffield, J. Math. Phys. **14**, 52 (1973).
[14]R.P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).

Bäcklund transformations and the symmetries of the Yang equations

Hedley C. Morris

Department of Mathematics,[a] *Oregon State University, Corvallis, Oregon 97331*
and School of Mathematics,[b] *Trinity College Dublin, Dublin, Ireland*

(Received 13 October 1978)

Using the terminology of Jet bundles, we determine generalized symmetries of the Yang equations

$$f(f_{y\bar{y}} + f_{z\bar{z}}) - f_y f_{\bar{y}} - f_z f_{\bar{z}} - e_y g_{\bar{y}} - \epsilon_z g_{\bar{z}} = 0,$$
$$f(e_{y\bar{y}} + e_{z\bar{z}}) - 2 e_y f_{\bar{y}} - 2 e_z f_{\bar{z}} = 0.$$
$$f(g_{y\bar{y}} + g_{z\bar{z}}) - 2 g_{\bar{y}} - 2 g_{\bar{y}} f_y - 2 g_{\bar{z}} f_z = 0,$$

which are equivalent to the self-dual Yang–Mills equations in a particular gauge. The Bäcklund transformations of Corrigan *et al.* are derived and discussed as generalized symmetries.

1. INTRODUCTION

There has been a recent upsurge of interest in the classical notion of a Bäcklund transformation.[1-9] In particular, there is considerable research activity aimed at formulating both an appropriate definition and also an effective computational scheme for the determination of Bäcklund transformations in more than two dimensions. The framework which seems most likely to provide the answer to both aspects is that which surrounds the formalism of jet bundles. We refer to the work of Hermann,[10-12] Pirani,[7-8] and Dodd[13] for details of this theory and wish to concentrate our attention on the computational utility of the language of jet bundles.

By making a special choice of gauge, the R gauge, Yang[14] was able to show that the Yang–Mills gauge potentials in complexified Euclidean space $A^a{}_k$ could be expressed in the form

$$A_k = -\tfrac{1}{2} f \begin{bmatrix} \eta^3{}_{kj}\,\partial_j f, & \eta^{1-i2}_{kj}\,\partial_j e \\ \eta^{1+i2}_{kj}\,\partial_j g, & -\eta^3{}_{kj}\,\partial_j f \end{bmatrix} \tag{1.1}$$

where $A_k = \frac{1}{2} A^a{}_k \sigma^a$ (with Pauli matrices σ^a) and $\eta^a{}_{kj}$ is the 't Hooft[15] tensor given by

$$\eta^a{}_{kj} = \epsilon_{0akj} + \delta_{ak}\,\delta_{0j} - \delta_{aj}\,\delta_{0k}. \tag{1.2}$$

The condition for the field strengths $F_{0k} = \partial_j A_k - \partial_k A_j + i\,[A_j, A_k]$ to be self-dual, ${}^*F_{jk} = \frac{1}{2}\epsilon_{jkie} F_{ie} = F_{jk}$ becomes

$$f(f_{y\bar{y}} + f_{z\bar{z}}) - f_y f_{\bar{y}} - f_z f_{\bar{z}} - e_y g_{\bar{y}} - e_z g_{\bar{z}} = 0, \tag{1.3}$$

$$f(e_{y\bar{y}} + e_{z\bar{z}}) - 2 e_y f_{\bar{y}} - 2 e_z f_{\bar{z}} = 0, \tag{1.4}$$

$$f(g_{y\bar{y}} + g_{z\bar{z}}) - 2 g_y f_y - 2 g_{\bar{z}} f_z = 0, \tag{1.5}$$

where

$$(2)^{1/2} \begin{pmatrix} y & -\bar{z} \\ z & -\bar{y} \end{pmatrix} = x_\mu\,\sigma^\mu$$

and $\bar{y}, \bar{z}$ are complex variables independent of y^* and z^*, the complex conjugates of y and z.

In this paper we will use the language of jet bundle theory to determine symmetries and generalized symmetries[16] of Eqs. (1.3)–(1.5), which we refer to as the Yang equations. In the following section we will determine generalized symmetries of the Yang equations and show that there is essentially only one nontrivial example of the type considered. The map so constructed is the "Bäcklund map" of Corrigan *et al.*[17] Despite the fact that no generally accepted definition of a Bäcklund map in higher dimensions has yet been formulated, it is unlikely that the name will eventually be associated with this example. Generally speaking, some form of integrability condition is involved in a Bäcklund transformation, and, although it is likely that this will need to be weakened, some such notation as involution[13] will probably be retained. Here we show that a generalized symmetry is a correct description of the Corrigan *et al.* mapping.

A Bäcklund transformation from an equation to itself (an auto-Bäcklund) is, in two dimensions, essentially a single discrete symmetry akin to parity or charge conjugation. Acting on a solution ϕ of its appropriate evolution equation, we find that

$$B^2 \phi = \phi. \tag{1.6}$$

With such a single operator a family of operators can be associated by utilizing the symmetries of the relevant evolution equation. Generally, in examples such as the sine–Gordon equation or Korteweg–de Vries equation, there is only a one-parameter family but there is no particular reason for that. If L_α is a symmetry transformation, either space–time or internal, which is paramerized by a set of labels α, we can define a multiparameter Bäcklund map B_α corresponding to a single Bäcklund map B by

$$B_\alpha = L_\alpha{}^{-1}\, B L_\alpha. \tag{1.7}$$

Thus there are two parts to a Bäcklund transformation. We must determine both a discrete mapping and also its interaction with the space–time or internal symmetrys L_α of its evolution equation.

The "Bäcklund map" of Corrigan *et al.*[17] has property (1.6), and in the final two sections of this paper we determine both internal symmetry transformations of the Yang equations which may be compounded with the single map B to give alternative forms of a "Bäcklund map" and also space–

[a] Visiting Associate Professor of Mathematics 1977–78.
[b] Permanent address.

 0022-2488/80/020256-05$1.00

time symmetries which may be combined, as in (1.7), to give a more general multiparameter form. Although most of our results are already known, the overall approach is novel and illustrates clearly the advantage of the jet bundle approach. We answer in the affirmative the question posed by Corrigan *et al.* [17] as to whether or not they have found the most general "Bäcklund map." That we are able to solve the equations that arise is due to the remarkable, and not yet fully explored, structure of the Yang equations.

2: THE "BÄCKLUND TRANSFORMATION" AS A GENERALIZED SYMMETRY

The Yang equations are given by

$$f(f_{,y\bar{y}} + f_{,z\bar{z}}) - f_{,y}f_{,\bar{y}} - f_{,z}f_{,\bar{z}} - e_{,y}g_{,\bar{y}} - e_{,z}g_{,\bar{z}} = 0 \quad (2.1)$$

$$f(e_{,y\bar{y}} + e_{,z\bar{z}}) - 2e_{,y}f_{,\bar{y}} - 2e_{,z}f_{,\bar{z}} = 0, \quad (2.2)$$

$$f(g_{,y\bar{y}} + g_{,z\bar{z}}) - 2g_{,\bar{y}}f_{,y} - 2g_{,\bar{z}}f_{,z} = 0. \quad (2.3)$$

Let us denote the base space with coordinates $z_1 = y$, $\bar{z}_1 = \bar{y}$, $z_2 = z$, $\bar{z}_2 = \bar{z}$ by M. We note that only certain first derivatives occur in the equation, and, to utilize that property, we define the variables ψ^a by

$$(\psi^1, \psi^2, \psi^3\ \psi^4, \psi^5) = (f, e_y, e_z, g_{\bar{y}}, g_{\bar{z}}) \quad (2.4)$$

and denote the space with coordinates ψ^a by N. We denote the jet bundle coordinates corresponding to derivatives $\partial_{z^\alpha}\psi^b$, $\partial_{\bar{z}^\alpha}\psi^b$ by ψ^b_a, $\psi^b_{\bar{a}}$ and similarly for the coordinates corresponding to higher derivatives. Changing to ψ^a eases the computations as one gains first-order expressions for all functions except ψ^1.

Once we add the additional equations $\psi^2_2 = \psi^3_1$ and $\psi^5_1 = \psi^4_2$ the equations in the ψ^a are completely equivalent to the original Yang equations. The resulting equations are functions on $J^2(M, N)$ and we will determine the Corrigan *et al.* "Bäcklund transformation" as a symmetry of the equations expressed in the ψ^a rather than a normal Bäcklund map which is associated with integrability conditions.

In terms of the natural coordinates on the jet bundle $J^2(M, N)$ we have an ideal generated by five functions α_i defined as follows:

$$\psi^1(\psi^1_{1\bar{1}} + \psi^1_{2\bar{2}}) - \psi^1_1\psi^1_{\bar{1}} - \psi^1_2\psi^1_{\bar{2}} - \psi^2\psi^4 - \psi^3\psi^5 = \alpha_1, \quad (2.5)$$

$$\psi^1(\psi^2_{\bar{1}} + \psi^3_{\bar{2}}) - 2\psi^2\psi^1_{\bar{1}} - 2\psi^3\psi^1_{\bar{2}} = \alpha_2, \quad (2.6)$$

$$\psi^1(\psi^4_1 + \psi^5_2) - 2\psi^4\psi^1_1 - 2\psi^5\psi^1_2 = \alpha_3, \quad (2.7)$$

$$\psi^2_2 - \psi^3_1 = \alpha_4, \quad (2.8)$$

$$\psi^5_1 - \psi^4_2 = \alpha_5, \quad (2.9)$$

$\alpha_i: J^2(M, N) \to R$.

If the mapping $f: J^2(M, N) \to R^5$ is defined by

$f: p \to (\alpha_1(p), \alpha_2(p), \ldots, \alpha_5(p))$,

then Yang's equations are expressed in jet bundle language by saying that a function $\psi: M \to N$ is a solution of the Yang equations if

$$f \circ j^2\psi = 0. \quad (2.10)$$

Let us look for a symmetry of the equations. Consider a mapping

$B: N \to N$.

Such a map B defines a map $\tilde{B} = (\mathrm{id}_M \times B): J^0(M, N) \to J^0(M, N)$, which can be uniquely lifted to $\tilde{B}: J^2(M, N) \to J^2(M, N)$. We will call B a symmetry transformation if it has the property that $\tilde{B}$ preserves the ideal generated by the α_i, that is,

$$\tilde{B}_*\alpha_i = \alpha_i \circ \tilde{B} = \gamma_i^k\alpha_k. \quad (2.11)$$

We could allow products of the α_j on the right-hand side, but we do not do so here. The equations

$$\tilde{B}_*\alpha_5 = \gamma_5{}^k\alpha_k$$

and

$$\tilde{B}_*\alpha_4 = \gamma_4{}^k\alpha_k$$

show that if $B: \psi^a \to \psi'^a$, then

$$\psi'^2 = c^1(\psi^1)^{-2}\psi^5 + c^2\psi^2 + c^3, \quad (2.12)$$

$$\psi'^3 = -c^1(\psi^1)^{-2}\psi^4 + c^2\psi^3 + c^4, \quad (2.13)$$

and

$$\psi'^4 = c^5(\psi^1)^{-2}\psi^3 + c^6\psi^4 + c^7, \quad (2.14)$$

$$\psi'^5 = -c^5(\psi^1)^{-2}\psi^2 + c^6\psi^5 + c^8, \quad (2.15)$$

where the c^2 are constants.

The equations

$$\tilde{B}_*\alpha_2 = \gamma_2{}^k\alpha_k$$

and

$$\tilde{B}_*\alpha_3 = \gamma_3{}^k\alpha_k$$

give rise to two solutions.

In the first $c^j = 0, j \neq 1,5$, and $\psi'^1 = c^9/\psi^1$ for constant c^9 and the mapping B takes the form

$$B: (\psi^1, \psi^2, \psi^3, \psi^4, \psi^5) \to \left[\frac{c^9}{\psi^1}, \frac{c^1\psi^5}{(\psi^1)^2}, \frac{-c^1\psi^5}{(\psi^1)^2}, \frac{c^5\psi^3}{(\psi^1)^2}, \frac{-c^5\psi^3}{(\psi^1)^2}\right]. \quad (2.16)$$

The final constraint

$$\tilde{B}_*\alpha_1 = \gamma_1{}^k\alpha_k$$

is automatically satisfied by a B of the above type provided only that

$$c^1c^5 + (c^9)^2 = 0 \quad (2.17)$$

The second solution is the simple scaling symmetry S defined by

$$S: (\psi^1, \psi^2, \psi^3, \psi^4, \psi^5) \to (c^{10}\psi^1, c^2\psi^2, c^2\psi^3, c^6\psi^4, c^6\psi^5), \quad (2.18)$$

where the invariance of α_1 requires that

$$(c^{10})^2 = c^2c^6. \quad (2.19)$$

By using this scaling symmetry with $c_2 = (c^1)^{-1}$, $c_6 = -(c^5)^{-1}$ we can reduce B to the simple form

$$B: (\psi^1\psi^2, \psi^3, \psi^4, \psi^5) \to ((\psi^2)^{-1}, (\psi^1)^{-2}\psi^5, -(\psi^1)^{-2}\psi^4, (\psi^1)^{-2}\psi^3, -(\psi^1)^{-2}\psi^2) \quad (2.20)$$

which is the "Bäcklund transformation" of Corrigan *et al.* [17]

The effect on the ideal generated by the α_i is

$$\tilde{B}_*\alpha_1 = -(\psi^1)^{-4}\alpha_1, \quad (2.21)$$

$$\tilde{B}_*\alpha_2 = +(\psi^1)^{-3}\alpha_5, \quad (2.22)$$

$$\tilde{B}_*\alpha_3 = +(\psi^1)^{-3}\alpha_4, \quad (2.23)$$

$$\tilde{B}_*\alpha_4 = +(\psi^1)^{-3}\alpha_3, \tag{2.24}$$

$$\tilde{B}_*\alpha_5 = +(\psi^1)^{-3}\alpha_2. \tag{2.25}$$

B is a normal symmetry of Eqs. (2.5)–(2.9). Once we introduce the potentials e and g to remove (2.8) and (2.9) the mapping B can be regarded as a generalized symmetry of Eqs. (2.1)–(2.3) in the sense of Ovsjaninkov.[16]

In the following section we return to the Yang equation in the form (2.1)–(2.3) expressed in terms of the potentials f, e, and g. Regarded as equations on $J^2(M,P)$, where P is the space of potentials (f,e,g) they define an ideal of functions spanned by the three functions β_i defined by

$$\beta_1 = f(f_{1\bar{1}} + f_{2\bar{2}}) - f_1 f_{\bar{1}} - f_2 f_{\bar{2}} - e_2 g_{\bar{2}}, \tag{2.26}$$

$$\beta_2 = f(e_{1\bar{1}} + e_{2\bar{2}}) - 2e_1 f_{\bar{1}} - 2e_2 f_{\bar{2}} \tag{2.27}$$

$$\beta_3 = f(g_{1\bar{1}} + g_{2\bar{2}}) - 2g_{\bar{1}} f_1 - 2g_{\bar{2}} f_2 \tag{2.28}$$

3. SYMMETRIES OF THE YANG POTENTIALS

Working now with the Yang equations expressed by the ideal expressed by the β_i, we seek a symmetry transformation L: $P \to P$. Such a map defines a mapping $\bar{L}: J^0(M,P) \to J^0(M,P)$ defined by $\bar{L} = (\mathrm{id}_M \times L)$.

As before, $\bar{L}$ can be uniquely lifted to a mapping $\tilde{L}: J^2(M,P) \to J^2(M,P)$. We require that L leave the ideal spanned by the β_i invariant in the simple linear fashion,

$$\tilde{L}_*\beta_i = \gamma_i^j \beta_j. \tag{3.1}$$

To shorten the account of the calculations, we note that the cancellation in (3.1) of terms involving $e_{\bar{y}}$, $e_{\bar{z}}$, g_y, and g_z may be achieved by taking f', e', g' in the form

$$f' = -e^{\psi}, \tag{3.2}$$

$$e' = -\psi_e, \tag{3.3}$$

$$g' = -\psi_g, \tag{3.4}$$

where $\psi: P \to P$ is a solution of the equation

$$\psi_{ge} = e^{2\psi} \tag{3.5}$$

and we have utilized elementary scaling symmetries to remove unnecessary constants assumed nonzero. By so doing we have eliminated from our solutions those ψ corresponding to such simple scaling symmetries as $(f,e,g) \to (\pm f, e, g)$.

The invariance of β_3 gives the following equations for ψ:

$$(\psi_{gf} e^{-2\psi})_f = -f^{-1}(\psi_{gf} e^{-2\psi}), \tag{3.6}$$

$$(\psi_{gg} e^{-2\psi})_f = -2f^{-1}(\psi_{gg} e^{-2\psi}), \tag{3.7}$$

$$(\psi_{gg} e^{-2\psi})_e = -f^{-1}(\psi_{gf} e^{-2\psi}), \tag{3.8}$$

$$(\psi_{gf} e^{-2\psi})_g = 0 \tag{3.9}$$

$$(\psi_{gf} e^{-2\psi})_e = -2f^{-1}(\psi_{ge} e^{-2\psi}) = -2f^{-1}. \tag{3.10}$$

These equations can easily be reduced to

$$\psi_{gg} e^{-2\psi} = (e^2 + Be + C) f^{-2}, \tag{3.11}$$

$$\psi_{ge} e^{-2\psi} = -(2e + B) f^{-1}, \tag{3.12}$$

where B and C are constants.

Similarly the invariance of β_2 yields the two equations

$$\psi_{ee} e^{-2\psi} = (g^2 + Dg + H) f^{-2}, \tag{3.13}$$

$$\psi_{ef} e^{-2\psi} = -(2g + D) f^{-2}, \tag{3.14}$$

where D and H are constants.

The invariance of β_1 produces the following equations for ψ:

$$f^{-1}\psi_f + \psi_{ff} = e^{-2\psi}\psi_{ef}\psi_{gf}, \tag{3.15}$$

$$(e^{-2\psi})_e = -(2g + D) + f^{-2}(g^2 + Dg + H) \times (2e + B), \tag{3.16}$$

$$(e^{-2\psi})_g = -(2e + B) + f^{-2}(e^2 + Be + C) \times (2g + D), \tag{3.17}$$

$$(e^{-2\psi})_f = 2f - 2f^{-3}(g^2 + Dg + H) \times (e^2 + Be + C). \tag{3.18}$$

Equations (3.16)–(3.18) can be integrated to give

$$e^{-2\psi} = K - (2ge + De + Bg - f^2) + f^{-2}(g^2 + Dg + H)(e^2 + Be + C), \tag{3.19}$$

where K is a constant.

The result must now be substituted back in (3.15) to find the constraints that are implied upon the constants K, B, C, D, H.

If we define

$$\alpha = (g^2 + Dg + H)(e^2 + Be + C), \tag{3.20}$$

$$\beta = -(2ge + De + Bg - K), \tag{3.21}$$

$$\gamma = (2g + D)(2e + B), \tag{3.22}$$

then

$$e^{-2\psi} = f^2 + \alpha f^{-2} + \beta \quad \text{and} \quad \psi_f = \frac{-f + \alpha f^{-3}}{f^2 + \alpha f^{-2} + \beta}. \tag{3.23}$$

This is a solution of Eq. (3.15) which can be written in the form

$$(f\psi_f)_f = \gamma f^{-1} e^{2\psi} \tag{3.24}$$

if

$$\gamma = -2\beta \tag{3.25}$$

and

$$\gamma^2 = 16\alpha \tag{3.26}$$

These conditions require that $4H = D^2$, $4C = B^2$, and

$$K = -DB/2. \tag{3.27}$$

Therefore, our final expression for $e^{-2\psi}$ is

$$e^{-2\psi} = f^2 - 2\tilde{g}\tilde{e} + f^{-2}\tilde{g}^2\tilde{e}^2 = (\tilde{g}\tilde{e} - f^2)^2 f^{-2}, \tag{3.28}$$

where

$$\tilde{g} = g + D/2 \quad \text{and} \quad \tilde{e} = e + B/2. \tag{3.29}$$

An ψ of this form is easily found to satisfy Eqs. (3.4) and (3.11)–(3.14). The potentials $\tilde{g}$ and $\tilde{e}$ are simply constant shifts corresponding to a trivial translational gauge symmetry, and without loss of generality we can take $D = B = 0$.

Finally then we have essentially a single symmetry L which is not of a scaling or translational type and it is given by

$$f' = -e^{+\psi} = -f/(eg - f^2), \tag{3.30}$$

$$e' = -\psi_e = g/(eg - f^2), \tag{3.31}$$

$$g' = -\psi_g = e/(eg - f^2), \tag{3.32}$$

and its action on the ideal generators β_i is

$$\tilde{L}.\beta_1 = [(eg + f^2)\beta_1 - gf\beta_2 - ef\beta_3]/(eg - f^2)^3, \tag{3.33}$$

$$\tilde{L}.\beta_2 = (\beta_3 + g^2\beta_2 - 2g\beta_1)f^2/(eg - f^2)^3, \tag{3.34}$$

$$\tilde{L}.\beta_3 = (\beta_2 + e^2\beta_3 - 2e\beta_1)f^2/(eg - f^2)^3. \tag{3.35}$$

L has the property (1.6) that

$$L^2 = 1. \tag{3.36}$$

This is most easily seen if we define the matrix Q by

$$Q = \begin{pmatrix} e & f \\ f & g \end{pmatrix}; \tag{3.37}$$

then

$L: Q \to Q^{-1}$ and (3.36) is plainly true.

The elementary scaling and translational symmetries corresponding to those in Eqs. (2.15)–(2.19) are $T_{\pm}(\alpha)$ defined by

$$T_{\pm}(\alpha): (f,e,g)(\pm(\alpha_1\alpha_2)^{1/2} f, \alpha_1 e + \alpha_3, \alpha_2 g + \alpha_4). \tag{3.38}$$

This mapping can also be expressed in terms of the matrix Q by

$$T_{\pm}(\alpha): Q \to aQa + b, \tag{3.39}$$

where

$$a = \begin{pmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \pm\sqrt{\alpha_2} \end{pmatrix}$$

and

$$b = \begin{pmatrix} \alpha_3 & 0 \\ 0 & \alpha_4 \end{pmatrix}, \tag{3.40}$$

and in future we will denote $T_{\pm}(\alpha)$ by $T(a,b)$, where a, b are arbitrary diagonal matrices and we can, without loss of anything but trivial solutions, take a as nonsingular. Such $T(a,b)$ form a noncommutative group with composition law

$$T(a,b)\,T(c,d) = T(ac, ada + b) \tag{3.41}$$

and

$$T^{-1}(a,b) = T(a^{-1}, -a^{-1}ba^{-1}). \tag{3.42}$$

The most general transformation constructed from the $T(a,b)$ and L takes the form

$$L_{\pm}(a,b,c,d) = \pm T(a,b)LT(c,d). \tag{3.43}$$

The action of $L_{\pm}(a,b,c,d)$ on Q is

$$L_{\mp}(a,b,c,d): Q \to (AQ + B)(CQ + D)^{-1}, \tag{3.44}$$

where

$$A = \pm ba^{-1}c, \quad B = \pm(ac^{-1} + ba^{-1}dc^{-1}),$$
$$C = a^{-1}c, \quad \text{and} \quad D = a^{-1}dc^{-1}, \tag{3.45}$$

which satisfy the constraint

$$AD - BC = \mp 1. \tag{3.46}$$

This shows that the symmetry group G we have discovered is isomorphic to the set of 4×4 matrices of the form $M(A,B,C,D)$ defined by

$$M(A,B,C,D) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{3.47}$$

where the A,B,C,D are diagonal matrices with the property

$$AD - BC = \mp 1 \tag{3.48}$$

The multiparameter family of discrete maps corresponding to $L_{\mp}$ and defined by

$$L_{\mp}(c,d) = T^{-1}(c,d)L_{\mp}T(c,d) \tag{3.49}$$

corresponds to the subgroup of G consisting of matrices $l_{\mp}(c,d)$ of the form

$$l_{\mp}(c,d) = \begin{pmatrix} \mp d & \pm c^{-2}(1 - d^2) \\ c^2 & d \end{pmatrix}, \tag{3.50}$$

which naturally have the property

$$l_{\mp}^2(c,d) = 1. \tag{3.51}$$

The two components of the symmetry group G_+ and G_- corresponding to $= AD - BC = \pm 1$ are connected by the discrete transformation $T = T(i,0)$

$$TG_{\pm} = G_{\mp}T. \tag{3.52}$$

One cannot combine the maps B and L regarded as symmetries as they act on different spaces. However, if we regard B as a generalized symmetry,[16] we can make sense of the map $\hat{B} = LB$, $\hat{B}: (f,e,g) \to (f',e',g')$, which provides a "Bäcklund transformation" in the form

$$f = -\frac{(e'g' - f'^2)}{f'}, \tag{3.53}$$

$$\partial_{\bar{y}}\left(\frac{e'}{e'g' - f'^2}\right) = -\frac{1}{f^2}\partial_z e, \tag{3.54}$$

$$\partial_z\left(\frac{g'}{e'g' - f'^2}\right) = -\frac{1}{f^2}\partial_{\bar{y}} g, \tag{3.55}$$

$$\partial_{\bar{z}}\left(\frac{e'}{e'g' - f'^2}\right) = \frac{1}{f^2}\partial_y e, \tag{3.56}$$

$$\partial_y\left(\frac{g'}{e'g' - f'^2}\right) = \frac{1}{f^2}\partial_{\bar{z}} g, \tag{3.57}$$

which is the alternative form given by Corrigan *et al.*[17]

4. COORDINATE SYMMETRIES

Let $X: M \to M$ be defined by

$$X: (y, z, \bar{y}, \bar{z}) \to (Y, Z, \bar{Y}, \bar{Z}).$$

If we assume that the ψ^a are scalars, then x may be uniquely lifted to a mapping $\tilde{X}: J^2(M,N) \to J^2(M,N)$, where $\tilde{X}$ is the lift of $\bar{X} = (X \times \mathrm{id}_M)$. Let us seek a symmetry which preserves the ideal generated by the α_i in the sense that

$$X^*\alpha_i = \gamma_i^j \alpha_j', \tag{4.1}$$

where the α_j' are the α_j expressed in the coordinate system $(Y,Z,\bar{Y},\bar{Z})$. The invariance of α_4 and α_5 shows that

$$\begin{aligned} Y &= \alpha y, \quad \bar{Y} = \beta\bar{y}, \\ Z &= \alpha z, \quad \bar{Z} = \beta\bar{z}. \end{aligned} \tag{4.2}$$

This immediately implies the invariance of α_2 and α_3 and the invariance of α_1 requires that $\alpha\beta = 1$. We denote the transformation arrived at by X_α. Thus

$$X_\alpha:(y,z,\bar{y},\bar{z})\to(\alpha y,\alpha z,\alpha^{-1}\bar{y},\alpha^{-1}\bar{z}) \tag{4.3}$$

The action on the ideal generators is given by

$$X^*\alpha_1=\alpha_1',\quad X^*\alpha_2=(1/\alpha)\alpha_2',\quad X^*\alpha_3=\alpha\alpha_3',$$
$$X^*\alpha_4=\alpha\alpha_4',\quad X^*\alpha_5=(1/\alpha)\alpha_5'. \tag{4.4}$$

Clearly as the map B is an internal symmetry, we will not obtain a family of "Bäcklund maps" by combining B and X_α as in (1.7). The nature of the ψ^a as derivative of potentials must be included, and as our next step we must implement the fact that the ψ^a are no longer all scalars when we extend our mapping X to the jet bundle $J^2(M,N)$. Let

$$X:(y,\bar{y},z,\bar{z})\to(Y(y,z),\bar{Y}(\bar{y},\bar{z}),Z(y,z)\bar{Z}(\bar{y},\bar{z})); \tag{4.5}$$

then this defines a mapping $\bar{X}$ on $J^0(M,N)$ given by

$$\bar{X}=(X\times L), \tag{4.6}$$

where $L:N\to N$ is given by

$$L:(\psi^1,\psi^2,\psi^3,\psi^4,\psi^5)\to(\psi'^1,Y_1\psi'^2+Z_1\psi'^3,\psi'^2Y_2+\psi'^3Z_2,$$
$$\psi'^4\bar{Y}_{\bar{1}}+\psi'^5\bar{Z}_{\bar{1}},\ \psi'^4\bar{Y}_{\bar{2}}+\psi'^5\bar{Z}_{\bar{2}}). \tag{4.7}$$

The map $\bar{X}$ can now be lifted uniquely to a map $\tilde{X}$: $J^2(M,N)\to J^2(M,N)$. Such a map $\tilde{X}$ automatically preserves α_5 and α_4 as

$$\tilde{X}^*\alpha_5=(\bar{Y}_{\bar{1}}\bar{Z}_{\bar{2}}-\bar{Z}_{\bar{1}}\bar{Y}_{\bar{2}})\alpha_5', \tag{4.8}$$
$$\tilde{X}^*\alpha_4=(Y_1Z_2-Z_1Y_2)\alpha_4'. \tag{4.9}$$

The invariance of α_1, α_2, and α_3 requires only that

$$Y_1\bar{Y}_{\bar{1}}+Y_2\bar{Y}_{\bar{2}}=Z_1\bar{Z}_{\bar{1}}+Z_2\bar{Z}_{\bar{2}}, \tag{4.10}$$

$$Y_1\bar{Z}_{\bar{1}}+Y_2\bar{Z}_{\bar{2}}=0=Z_1\bar{Y}_{\bar{1}}+Z_2\bar{Y}_{\bar{2}}. \tag{4.11}$$

From these equations one easily shows that the functions Y, $\bar{Y}$,Z, $\bar{Z}$ are linear functions of their respective arguments. We write

$$\begin{pmatrix}Y\\Z\end{pmatrix}=A\begin{pmatrix}y\\z\end{pmatrix}+C \tag{4.12}$$

and

$$\begin{pmatrix}\bar{Y}\\\bar{Z}\end{pmatrix}=B\begin{pmatrix}\bar{y}\\\bar{z}\end{pmatrix}+D \tag{4.13}$$

where A and B are constant matrices and C and D constant vectors. The C and D correspond to translation invariance, and we set them both to zero.

Equations (4.10), (4.11) can be expressed neatly in terms of A and B as

$$AB^T=\mu I, \tag{4.14}$$

where $\mu=\bar{Y}_{\bar{1}}Y_1+Y_2\bar{Y}_{\bar{2}}$ is a constant.

The action on the generations α_i is given by

$$\tilde{X}^*\alpha_i=\mu\alpha_i',\quad i=1,2,3, \tag{4.15}$$
$$\tilde{X}^*\alpha_4=\det A\alpha_4',\quad \tilde{X}^*\alpha_5=\det B\alpha_5'. \tag{4.16}$$

We denote the coordinate symmetry expressed by (4.12)–(4.14) by $X(A,B)$. The mapping $\bar{X}(A,B):J^0(M,N)\to J^0(M,N)$ defined by

$$\bar{X}(A,B):\left[\begin{pmatrix}y\\z\end{pmatrix},\begin{pmatrix}\bar{y}\\\bar{z}\end{pmatrix},\begin{pmatrix}\psi^2\\\psi^3\end{pmatrix},\begin{pmatrix}\psi^4\\\psi^5\end{pmatrix}\right]$$
$$\to\left[A\begin{pmatrix}y\\z\end{pmatrix},B\begin{pmatrix}\bar{y}\\\bar{z}\end{pmatrix},A^T\begin{pmatrix}\psi'^2\\\psi'^3\end{pmatrix},B^T\begin{pmatrix}\psi'^4\\\psi'^5\end{pmatrix}\right] \tag{4.17}$$

can be lifted to a unique transformation $\tilde{X}(A,B):J^2(M,N)\to J^2(M,N)$. As both $\bar{B}$ and $\bar{X}(A,B)$ act on the same space $J^0(M,N)$, we can stay at the $J^0(M,N)$ level and construct a family of "Bäcklund maps" $\bar{B}(A,B)$ from $\bar{B}$ by defining

$$\bar{B}(A,B)=\bar{X}^{-1}(A,B)\,\bar{B}\,\bar{X}(A,B)=(\mathrm{id}_M\times B(A,B)). \tag{4.18}$$

One readily determines that $B(A,B):N\to N$ is given by

$$B(A,B):(\psi^1,\psi^2,\psi^3,\psi^4,\psi^5)\to\Big((\psi^1)^{-1},\frac{\det B}{\mu}\frac{\psi^5}{(\psi^1)^2},$$
$$-\frac{\det B}{\mu}\frac{\psi^4}{(\psi^1)^2},$$
$$\frac{\det A}{\mu}\frac{\psi^3}{(\psi^1)^2},-\frac{\det A}{\mu}\frac{\psi^2}{(\psi^1)^2}\Big). \tag{4.19}$$

From (4.19) we have $\det A\,\det B=\mu^2$, and we see that $B(A,B)$ simply results from B by a scaling symmetry of the type in Eq. (2.18). As we determined all possible B, this had to happen, but it is interesting to see how it has come about.

ACKNOWLEDGMENTS

It should like to thank my colleague Roger Dodd for stimulating discussions and also acknowledge the support of the Department of Mathematics, Oregon State University, where the author was visiting Professor for the academic year 1977–78.

[1]R.M. Miura, Ed. "Bäcklund Transformations, Lecture Notes in Mathematics 515 (Springer–Verlag Berlin, 1976).
[2]H.D. Wahlquist and F.B. Estabrook, Phys. Rev. Lett. **31**, 1386–90 (1973).
[3]H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. **16**, 1–7 (1975).
[4]J.P. Corones, J. Math. Phys. **18**, 163–64 (1977).
[5]H.C. Morris, "Bäcklund transformations and the sine–Gordon equation," in *The* 1976 *Ames Research Center (NASA) Conference on the Geometric Theory of Non-Linear Waves* (Math. Sci. Press., Brookline, Mass. 1977).
[6]H. Chen, Phys. Rev. Lett. **33**, 925–28 (1974).
[7]M. Crampin, F.A.E. Pirani, and D.C. Robinson, Lett. Math. Phys. **2**, 15–19 (1977).
[8]F.A.E. Pirani, D.C. Robinson, and W.F. Shadwick, "Local jet bundle formulation of Bäcklund transformations," preprint, Kings College, London, 1978.
[9]P.F.J. Dhooghe, "Jet bundles and Bäcklund transformations" (in press, 1978).
[10]R. Hermann, *Vector Bundles in Mathematical Physics* Benjamin, New York, 1970), Vol. I.
[11]R. Hermann, *Geometry Physics and Systems* Dekker, New York, 1973).
[12]R. Hermann, *The Geometry of Nonlinear Equations, Bäcklund Transformations and Solitons, Part A*, Interdisciplinary Math., Vol. XII Math. Sci. Press, Brookline, Mass., 1976).
[13]R.K. Dodd, preprint, Trinity College, September 1978.
[14]C.N. Yang, Phys. Rev. Lett. **38**, 1377–79 (1977).
[15]G. 't Hooft, Phys. Rev. Lett. **37**, 8–11 (1976).
[16]L.V. Ovsjannikov, "Properties of Differential Equations," translated by G.W. Bluman (unpublished).
[17]E.F. Corrigan, D.B. Fairlie, R.G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223–40 (1978).

An improvement of Watson's theorem on Borel summability

Alan D. Sokal[a)]
Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544

(Received 27 August 1979; accepted for publication 19 October 1979)

Watson's theorem, which gives sufficient conditions for Borel summability, is not optimal. Watson assumes analyticity and uniform asymptotic expansion in a sector $|\arg z| < \pi/2 + \epsilon$, $|z| < R$, with $\epsilon > 0$; in fact, only the circular region $\mathrm{Re}(1/z) > 1/R$ is required. In particular, one can take $\epsilon = 0$. This improved theorem gives a necessary and sufficient characterization of a large class of Borel-summable functions. I apply it to the perturbation expansion in the ϕ_2^4 quantum field theory.

There has recently been much interest in the Borel summation of perturbation expansions in quantum mechanics and quantum field theory. Proofs of Borel summability[1-8] have universally relied on the classic theorem of Watson,[9,10] which gives a sufficient condition for Borel summability.

The purpose of this note is to point out that Watson's theorem is not optimal, and to call the attention of physicists to an improved version of Watson's theorem. This improved theorem appears in an apparently little-known 1918 paper by F. Nevanlinna,[11,12] and was rediscovered and extended by the present author.[14] This theorem makes clear the "natural" region of analyticity for Borel-summable functions; for as I shall emphasize, the theory of Borel summability is nothing other than the theory of Laplace transforms, written in slightly different variables.

Let us recall that we are concerned with a function $f(z)$ possessing an asymptotic expansion $\Sigma a_n z^n$ as $z \to 0$ in an appropriate region of the complex plane; we wish to determine conditions under which $f(z)$ can be uniquely reconstructed from its perturbation coefficients a_n. One such method of reconstruction is Borel summation.[10,15] We say that the formal power series $\Sigma a_n z^n$ is Borel summable if:

(a) $B(t) = \Sigma a_n t^n/n!$ converges in some circle $|t| < \delta$;

(b) $B(t)$ has an analytic continuation to a neighborhood of the positive real axis;[16] and

(c) $g(z) = (1/z)\int_0^\infty e^{-t/z} B(t)\, dt$ converges (not necessarily absolutely) for some $z \neq 0$.

$B(t)$ is called the Borel transform of the series $\Sigma a_n z^n$, and $g(z)$ is its Borel sum.

Please note that the integral in (c) is a Laplace transform (the conventional variable is $s = 1/z$). It is well known that Laplace transforms converge in right half-planes.[17] Therefore, if the integral in (c) converges for some $z_0 \neq 0$, then it converges for all z with $\mathrm{Re}z^{-1} > \mathrm{Re}z_0^{-1}$, and $g(z)$ is analytic there. This region is a circle tangent to the imaginary axis (Fig. 1).

Watson's theorem[9] gives a sufficient condition for the function $f(z)$ to equal the Borel sum $g(z)$ of its asymptotic Taylor series. Let $f(z)$ be analytic in a sector $|\arg z| < \pi/2 + \epsilon$, $|z| < R$, for some $\epsilon > 0$, and let $f(z)$ have there the asymptotic expansion

$$f(z) = \sum_{k=0}^{N-1} a_k z^k + R_N(z), \tag{1}$$

with

$$|R_N(z)| \leqslant A\sigma^N N! |z|^N \tag{2}$$

uniformly in N and in z in the sector. [The $N!$ bound (2) is crucial.] Under these assumptions, Watson shows that:

(a') $B(t) = \Sigma a_n t^n/n!$ converges in the circle $|t| < 1/\sigma$;

(b') $B(t)$ has an analytic continuation to the sector $|\arg t| < \epsilon$; and

(c') the integral $(1/z)\int_0^\infty e^{-t/z} B(t)\, dt$ is absolutely convergent for $\mathrm{Re}z^{-1} > R^{-1}$ and there equals $f(z)$.[18]

In some applications of Watson's theorem, however, it may be difficult (or impossible) to verify the analyticity and the estimate (2) in a sector with $\epsilon > 0$; it would be desirable if $\epsilon = 0$ could be allowed. Furthermore, it is unnatural for the hypothesis of the theorem to refer to a region of the z plane strictly larger than that in which $f(z)$ is recovered by Borel summation. Both of these disadvantages are remedied in the following theorem, which gives a necessary and sufficient characterization of a large class of Borel-summable functions.

Theorem:[11] Let f be analytic in the circle $C_R = \{z: \mathrm{Re}z^{-1} > R^{-1}\}$ and satisfy there the estimates (1) and (2)

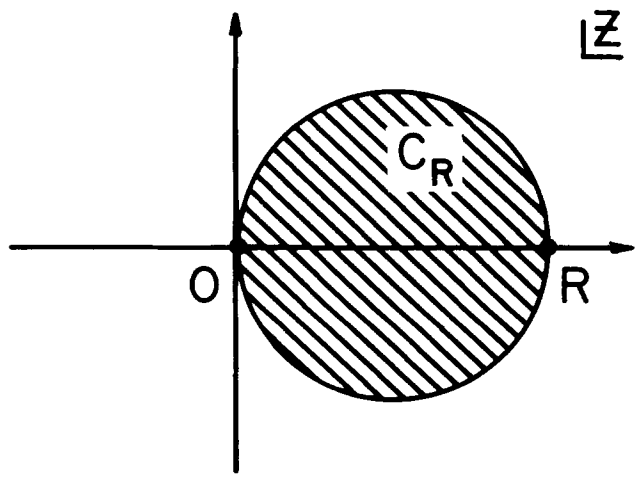


FIG. 1. Minimum region of analyticity of Borel-summable function.

[a)]NSF Predoctoral Fellow (1976–79). Work supported in part by NSF PHY 78-23592.

 0022-2488/80/020261-03$1.00

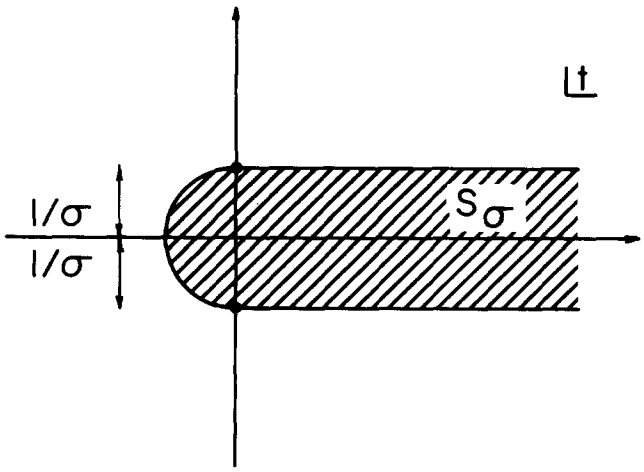


FIG. 2. Region of analyticity of the Borel transform, for a function satisfying the hypotheses of Nevanlinna's theorem.

uniformly in N and in $z \in C_R$. Then $B(t) = \Sigma a_n t^n/n!$ converges for $|t| < 1/\sigma$ and has an analytic continuation to the striplike region $S_\sigma = \{t: \mathrm{dist}(t, \mathbb{R}_+) < 1/\sigma\}$ (Fig. 2), satisfying the bound

$$|B(t)| \leqslant K \exp(|t|/R) \tag{3}$$

uniformly in every $S_{\sigma'}$ with $\sigma' > \sigma$. Furthermore, f can be represented by the absolutely convergent integral

$$f(z) = (1/z) \int_0^\infty e^{-t/z} B(t)\, dt, \tag{4}$$

for any $z \in C_R$.

Conversely, if $B(t)$ is a function analytic in $S_{\sigma''}$ $(\sigma'' < \sigma)$ and there satisfying (3), then the function $f(z)$ defined by (4) is analytic in C_R and satisfies (1) and (2) [with $a_n = B^{(n)}(t)|_{t=0}$] uniformly in every $C_{R'}$ with $R' < R$.

Since the proof of this theorem is very similar to Hardy's proof of Watson's theorem,[10] I will only sketch the method (see also Ref. 11). First one shows, using (1) and (2), that the integrals

$$b_m(t) = a_m + \frac{1}{2\pi i} \oint_{\mathrm{Re}\, z^{-1} = r^{-1}} e^{t/z} z^{-(m+1)} \times \left(f(z) - \sum_{k=0}^{m} a_k z^k \right) dz, \tag{5}$$

are absolutely convergent for $t \geqslant 0$ and independent of r for $0 < r < R$; that $b_0(t)$ is a C^∞ function whose mth derivative is $b_m(t)$; and that

$$|b_m(t)| \leqslant K_1 \sigma^{m+1} (m+1)!\, \exp(t/R), \tag{6}$$

with K_1 independent of t and m. Inserting (1) into (5) and performing a contour integral, one finds

$$b_0(t) = \sum_{k=0}^{N-1} a_k t^k/k! + \frac{1}{2\pi i} \oint_{\mathrm{Re}\, z^{-1} = r^{-1}} e^{t/z} z^{-1} R_N(z)\, dz. \tag{7}$$

Using (2) to bound $R_N(z)$, and choosing $r = t/N$ (with $N > t/R$) to optimize the estimate, the remainder term in (7) is bounded by $K_2 N^{1/2} (\sigma t)^N$. Hence, for $0 \leqslant t < 1/\sigma$, the remainder goes to zero as $N \to \infty$, and $b_0(t)$ there equals $B(t) = \Sigma_{k=0}^\infty a_k t^k/k!$, a series which converges in the circle $|t| < 1/\sigma$. Furthermore, it follows from (6) that each series

$$B_{t_0}(t) = \sum_{m=0}^{\infty} b_m(t_0)(t - t_0)^m/m! \quad (t_0 \geqslant 0),$$

converges in the circle $|t - t_0| < 1/\sigma$ and satisfies there the bound $|B_{t_0}(t)| \leqslant K_1 \exp(t_0/R)(1 - \sigma)|t - t_0)^{-2}$. It is not hard to show that $B_{t_0}(t) = B_{t_1}(t)$ wherever both are defined, and that the union of all these functions therefore defines a single-valued analytic continuation of $B(t)$ into the region S_σ, satisfying the bound (3) uniformly in every $S_{\sigma'}$ with $\sigma' > \sigma$. Finally, inserting (5) (with $m = 0$) into the right-hand side of (4), one obtains

$$a_0 + \frac{1}{z} \int_0^\infty dt\, e^{-t/z} \frac{1}{2\pi i} \oint_{\mathrm{Re}\, z'^{-1} = r^{-1}} dz' \times e^{t/z'} [f(z') - a_0]/z'.$$

Choosing $r^{-1} < \mathrm{Re}\, z^{-1}$ and noting that $[f(z') - a_0]/z'$ is bounded, one sees that the double integral is absolutely convergent; thus one can interchange the order of integration, perform the integral, and verify (4).

Conversely, if $B(t)$ is analytic in $S_{\sigma''}$ and there satisfies (3), one uses Cauchy's integral formula to prove (6) [with σ'' replacing σ] for t real and positive. Integrating by parts N times in (4), one demonstrates (1) and (2). Q.E.D.

In order to understand this theorem, it is instructive to examine how the standard counterexamples $f(z) = \exp(-z^{-\alpha})$, with $\alpha > 0$, fail to satisfy its hypotheses. If $\alpha \geqslant 1$, the asymptotic expansion (1) fails: for any R, there exists a curve lying within C_R along which $f(z)$ does not approach zero as $z \to 0$. If $\alpha < 1$, the expansion (1) holds but the estimate (2) fails: $\sup_{z \in C_R} |z^{-N} R_N(z)|$ is of order $(N!)^{1/\alpha}$.

If $f(z)$ is analytic and satisfies (1) and (2) in a region of opening angle greater than π, then correspondingly stronger conclusions can be drawn. More precisely, let $f(z)$ be analytic and satisfy (1) and (2) in the region $C_{R,\epsilon} = \cup_{|\theta| < \epsilon} e^{i\theta} C_R$; then a trivial modification of the foregoing proof shows that $B(t)$ is analytic in $S_{\sigma,\epsilon} = \cup_{|\theta| < \epsilon} e^{i\theta} S_\sigma$, and (4) holds for any $z \in C_{R,\epsilon}$.

Using Nevanlinna's theorem, one can simplify the proof of Borel summability of the Schwinger functions in the ϕ_2^4 quantum field theory,[3] and also extend it in certain technical aspects. The heart of this proof is to show that the Schwinger functions are analytic and satisfy (1) and (2) uniformly in a semicircle $\mathrm{Re}\, g > 0$, $|g| < R$, where g is the coupling constant. By a scaling argument, one then extends this region to a sector of the form demanded in Watson's theorem, with opening angle strictly greater than π; this argument is valid, however, only if the space–time smearing functions are analytic in a sector containing the positive real axis. Nevanlinna's theorem shows that this scaling argument is unnecessary, and that arbitrary smearing functions can be allowed.

[1] S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[2] B. Simon, Adv. Math. 7, 240 (1971).
[3] J.-P. Eckmann, J. Magnen, and R. Sénéor, Commun. Math. Phys. **39**, 251 (1975); J.-P. Eckmann and H. Epstein, Commun. Math. Phys. **68**, 245 (1979).
[4] J. Magnen and R. Sénéor, Commun. Math. Phys. **56**, 237 (1977).

[5]P. Renouard, Ann. Inst. Henri Poincaré A **27**, 237 (1977); and "Analyticité et Sommabilité de Borel des Fonctions de Schwinger du Modèle de Yukawa en Dimension $d = 2$–II: La Limite Adiabatique," Ecole Polytechnique preprint, July 1979.

[6]J. Avron, I. Herbst, and B. Simon, Phys. Lett. A **62**, 214 (1977); and "Schrödinger Operators in Magnetic Fields III: Atoms and Ions in Constant Fields", in preparation.

[7]S. Graffi and V. Grecchi, Commun. Math. Phys. **62**, 83 (1978).

[8]I. Herbst and B. Simon, Phys. Rev. Lett. **41**, 67 (1978); and "Dilation Analyticity in Constant Electric Field II: The N-Body Problem, Borel Summability," in preparation.

[9]G.N. Watson, Philos. Trans. Soc. London, Ser. A **211**, 279 (1912).

[10]G.H. Hardy, *Divergent Series* (Oxford U.P., London, 1949).

[11]F. Nevanlinna, Ann. Acad. Sci. Fenn. Ser. A**12**, No. 3 (1918–19). A review of this work appears in Jahrb. Fort. Math. **46**, 1463 (1916–18).

[12]Nevanlinna's paper is cited in Hardy's classic book (Ref. 10, p. 199), but Hardy gives no hint that its contents might go beyond Watson's theorem. Few modern mathematical physicists seem to be aware of Nevanlinna's work; one exception is Ref. 13.

[13]W.A. Beyer and L. Heller, J. Math. Phys. **8**, 1004 (1967).

[14]Extensions of Nevanlinna's theorem will be reported in a subsequent article.

[15]E. Borel, *Leçons sur les séries divergentes*, 2ème éd. (Gauthier–Villars, Paris, 1928), pp. 135 ff.

[16]More generally, $B(t)$ can have isolated essential singularities on the positive real axis, provided that $B(t)$ is locally absolutely integrable as one crosses these singularities in the real direction. For example, $B(t) = \exp[-c/(t-t_0)^2]$ with $\mathrm{Re}c \geqslant 0$ is permitted. However, $B(t)$ must not have poles or branch points on the positive real axis: these would require additional prescriptions to resolve the ambiguity in the inverse Borel transform.

[17]D.V. Widder, *The Laplace Transform* (Princeton U.P., Princeton, New Jersey, 1941), Chapter II.

[18]Hardy (Ref. 10) and Borel (Ref. 15) write the inverse Borel transform in the form $\int_0^\infty e^{-t} B(tz)\, dt$, which is equivalent to taking the integral in (c') along the ray $\arg t = \arg z$; as a result, these authors obtain Borel summability only for $|\arg z| < \epsilon$. Watson, however, uses the preferred form (c') and thus obtains the full region of Borel summability (see Ref. 9, pp. 307–310).

The Jost solutions for general Gaussian potentials

L. Trlifaj

Institute of Physics of the Czechoslovak Acad. Sci., Prague, Czechoslovakia
(Received 11 December 1978)

The explicit Born series which represent the regular solutions (of the Schrödinger equation), the Jost solutions, and the Jost functions analytic in some domains of complex wave numbers and angular momenta are found for general potentials of the Gaussian type expressed by the Stiltjes integral. When finding their form we make use of the analogy with the corresponding solutions for general potentials of the Yukawa type, which is based on the special representations of the Bessel and Hankel functions (purely kinematic solutions) in either case. Some other relations are also derived.

1. INTRODUCTION

We define a general Gaussian potential by a Stiltjes integral with respect to a function $\mu(t)$ of bounded variation

$$V(r) = -g\int_\tau^\infty d\mu(t)\exp(-tr^2) \quad (\tau>0). \tag{1}$$

Quantity g is a real constant. We introduce classes M_κ of functions $\mu(t)$ by demanding

$$\nu_p = \int_\tau^\infty d\nu(t)\, t^p < \infty \quad \text{for } p<\kappa \text{ when } \kappa>0 \text{ or } p=\kappa \text{ when } \kappa=0 \text{ or } -1, \tag{2a}$$

where $\nu(t)$ ($d\nu = |d\mu|$) is the total variation of $\mu(t)$. In the class M_κ with $\kappa>0$ there may exist functions which satisfy the inequality (2a) for κ too. This inequality implies that the potential $V(r)$ and its first $[\kappa]$ (or $[\kappa]-1$) derivatives are continuous and finite for $r\geqslant 0$, if $\kappa>0$. The odd derivatives are always zero at $r=0$, the first q even derivatives are also equal to zero there, if

$$\int_\tau^\infty d\mu(t)\, t^p = 0, \quad \text{for integral } p\ 0\leqslant p\leqslant q\leqslant \tfrac{1}{2}\kappa. \tag{2b}$$

We want to show that the regular solutions $\varphi(\lambda,k,r)$ as well as the Jost solutions $f_\pm(\lambda,k,r)$ of the radial Schrödinger equation

$$\left(\frac{d^2}{dr^2} - \frac{\lambda^2-\frac{1}{4}}{r^2} - V(r) + k^2\right)\psi(\lambda,k,r) = 0 \tag{3}$$

with the potential (1) can be found for complex wavenumbers k and angular momenta λ in the form of explicit expressions. The case of the S waves ($\lambda=\frac{1}{2}$) is of a particular interest, since the interval of $r[0,\infty)$ can then be extended to $(-\infty,\infty)$ in Eqs. (1) and (3) and the potential belongs to the symmetric model potentials, which generalize the harmonic oscillator potential.

2. REGULAR SOLUTIONS

A regular solution of Eq. (3), $\varphi(\lambda,k,r)\simeq r^{\lambda+1/2}$, as $r\to 0^+$ is easily found by the technique of the modified Laplace transformation[1]

$$\omega(\lambda,k,s) = \int_0^\infty dr\, r^{\lambda+1/2} e^{-sr^2}\varphi(\lambda,k,r) \quad (\mathrm{Re}s>0, \mathrm{Re}\lambda>0), \tag{4}$$

where Abel's asymptotic behavior is $\omega(\lambda,k,s)\sim\frac{1}{2}\Gamma(\lambda+1)\times s^{-\lambda-1}$ as $s\to\infty$ and $\Gamma(z)$ is the well-known Γ function. Equation (3) is then transformed into a first-order differential equation for $\omega(\lambda,k,s)$

$$4s^2\omega'(\lambda,k,s) + [4(\lambda+1)s - k^2]\,\omega(\lambda,k,s) = g\int_\tau^\infty d\mu(t)\,\omega(\lambda,k,t+s). \tag{5}$$

Respecting the prescribed asymptotic behavior, we integrate it and transform into an integral equation, which can be solved by iterations. The result is

$$\omega(\lambda,k,s) = \tfrac{1}{2}\Gamma(\lambda+1)s^{-\lambda-1}\exp\left(-\frac{k^2}{4s}\right)\sigma(\lambda,k,s) \tag{6a}$$

$$\sigma(\lambda,k,s) = 1 + \sum_{n=1}^\infty \left(-\frac{g}{4}\right)^n \int_0^\infty dz_1 \int_\tau^\infty d\mu(t_1)\cdots\int_0^\infty dz_n \times\int_\tau^\infty d\mu(t_n)\frac{(s+u_1+v_0)^{\lambda-1}}{(s+u_1+v_1)^{\lambda+1}}\cdots\frac{(s+u_n+v_{n-1})^{\lambda-1}}{(s+u_n+v_n)^{\lambda+1}} \times\exp\left[\frac{k^2}{4}\sum_{p=1}^n\left(\frac{1}{s+u_p+v_{p-1}} - \frac{1}{s+u_p+v_p}\right)\right], \tag{6b}$$

where

$$u_p = \sum_{q=1}^p z_q, \quad v_p = \sum_{q=1}^p t_q, \quad v_0 = u_0 = 0. \tag{6c}$$

Since

$$\left|\frac{s+u_p+v_{p-1}}{s+u_p+v_p}\right| < 1 \quad \text{for } \mathrm{Re}s \geqslant -(p-1)\tau,$$

$$\left|\exp\left(\frac{k^2}{4z}\right)\right| \leqslant \exp\left|\frac{k^2}{4z}\right|,$$

$$\int_0^\infty dz_p |s+u_p+v_p|^{-2} < v_p^{-1} \quad \text{for } \mathrm{Re}s\geqslant 0 \tag{7a}$$

and

$$\int_\tau^\infty \frac{d\nu(t_1)}{t_1}\cdots\int_\tau^\infty \frac{d\nu(t_n)}{t_n}\frac{t_1\cdots t_n}{v_1\cdots v_n} = \frac{1}{2}\int_\tau^\infty \frac{d\nu(t_1)}{t_1}\cdots\int_\tau^\infty \frac{d\nu(t_n)}{t_n}\frac{t_3\cdots t_n}{v_3\cdots v_n} = \frac{1}{n!}\left(\int_\tau^\infty \frac{d\nu(t)}{t}\right)^n, \tag{7b}$$

 0022-2488/80/020264-04$1.00

the series converges absolutely and uniformly for $\mathrm{Re}s > 0$ and $\mathrm{Re}\lambda \geqslant 1$ provided $\nu_{-1} = \int_\tau^\infty dv(t)\, t^{-1} < \infty$. If $\nu_0 = \int_\tau^\infty dv(t) < \infty$, the same conclusion can be drawn for $-1 \leqslant \mathrm{Re}\lambda < 1$. Further on we have

$$\left|\frac{s+u_p+v_p}{t_p(s+u_p+v_{p-1})}\right| \leqslant \frac{1}{t_p} + \frac{1}{|s+u_p+v_{p-1}|} \leqslant \tau^{-1} + [|s|^2 + (p-1)^2\tau^2]^{-1/2} \quad \text{for } \mathrm{Re}s \geqslant 0, \tag{7c}$$

so that the series (6b) is absolutely and uniformly convergent for $\mathrm{Re}s > 0$ and $\mathrm{Re}\lambda > -1-\kappa$ provided the function $\mu(t)$ belongs to the class M_κ ($\kappa > 0$). Inequalities analogous to (7a) and (7c) hold also for general complex s except for the points from the interval $[0, -\infty)$, so that the series (6b) represents an analytic function of s in that region for any fixed k.

This enables us to deform the usual integration contour of s for $\mathrm{Re}\lambda > -1$, when we invert the Laplace transformation (4)

$$\varphi(\lambda,k,r) = \frac{1}{2\pi i}\Gamma(\lambda+1)\, r^{-\lambda+1/2} \times \int_C ds\, s^{-\lambda-1} \exp\left(sr^2 - \frac{k^2}{4s}\right)\sigma(\lambda,k,s). \tag{8}$$

The introduced contour C goes from $\infty\exp(-i\pi) - i\epsilon$ to $0 - i\epsilon$, it encircles $s = 0$ in an anticlockwise direction and returns from $0 + i\epsilon$ to $\infty\exp(i\pi) + i\epsilon$ ($\epsilon > 0$) avoiding the points of the negative real axis. Such an integral representation can be continued analytically into the regions of λ in which the function $\sigma(\lambda,k,s)$ is defined. If the function $\mu(t)$ belongs to the class M_κ ($\kappa > 0$), they are determined by $\mathrm{Re}\lambda > -1-\kappa$. In this half-plane the regular solution (8) is an analytic function of λ except for the simple poles at the negative integral values due to the factor $\Gamma(\lambda+1)$. It represents an entire function of $E = k^2$ at the same time. According to Eq. (4) the limit of $\omega(\lambda,k,s)$ tends to the finite value as $s \to 0^+$ for the bound states, which are given by $k = \pm ib$ ($b > 0$) and by the positive integral or half-integral values of λ ($\lambda = \lambda_{\mathrm{phys}}$). The eigenvalue equation consequently reads

$$\sigma(\lambda_{\mathrm{phys}}, \pm ib, 0) = 0 \tag{9}$$

owing to Eq. (6a). It is a general form of the equation derived in Ref. 1 for $\mu(t) = \Theta(t-1)$ [$\tau < 1$, $\Theta(t) = 0$ when $t < 0$ and $\Theta(t) = 1$ when $t > 0$]. As the function $\sigma(\lambda_{\mathrm{phys}},k,0)$ is an analytic function of $E = k^2$ for $\frac{3}{2}\pi > |\arg E| > \pi/2$ and $\lim_{b\to\infty}\sigma(\lambda_{\mathrm{phys}}, \pm ib, 0) = 1$, there is a finite number of the bound states b.

3. THE JOST SOLUTIONS

In the case of the Yukawa potentials there is a relation between the forms of the regular solution and those of the Jost solutions, which can be based on a certain mathematical approach.[2] Since we have not found such an approach in the case investigated, we made use of this formal relationship only and verified that the expressions

$$f_\pm(\lambda,k,r) = \pi^{-1/2}\exp\left[\mp i\frac{\pi}{2}(\lambda - 1/2)\right]\left(\frac{k}{2}\right)^{\lambda+1/2} r^{-\lambda+1/2}\int_{C_\pm} ds\, s^{-\lambda-1}\exp\left(sr^2 - \frac{k^2}{4s}\right) \times \left\{1 + \sum_{n=1}^{\infty}\left(\frac{-g}{4}\right)^n \int_0^\infty dz_1 \int_\tau^\infty d\mu(t_1)\cdots\int_0^\infty dz_n\int_\tau^\infty d\mu(t_n)\exp[-(u_n+v_n)r^2] \times \frac{(s-u_0-v_1)^{\lambda-1}\cdots(s-u_{n-1}-v_n)^{\lambda-1}}{(s-u_1-v_1)^{\lambda+1}\cdots(s-u_n-v_n)^{\lambda+1}}\exp\left[\frac{k^2}{4}\sum_{p=1}^{n}\left(\frac{1}{s-u_{p-1}-v_p} - \frac{1}{s-u_p-v_p}\right)\right]\right\} \tag{10a}$$

where the contours $C_\pm$ of s go from $0\exp[i(\arg k + \gamma)]$ to $\infty\exp[i(\arg k - \gamma \mp \pi)]$ avoiding the points of the line $\mathrm{Im}s = 0$, $\mathrm{Re}s \geqslant \tau$ and $|\gamma| < \pi$, $|\arg k - \gamma| < \pi/2$, represent the Jost solutions of Eq. (3).

Integrating by parts with respect to s or z_n we easily find that the expressions (10a) are indeed solutions of Eq. (3) for either sign of λ. According to Ref. 3 the first terms of the Born series (10a) (corresponding to the choice $g = 0$)

$$h_\pm(\lambda,k,r) = \pi^{-1/2}\exp\left[\mp i\frac{\pi}{2}\left(\lambda - \frac{1}{2}\right)\right]\left(\frac{k}{2}\right)^{\lambda+1/2} r^{-\lambda+1/2}\int_{C_\pm} ds\, s^{-\lambda-1}\exp\left(sr^2 - \frac{k^2}{4s}\right) = h_\pm(-\lambda,k,r) = \exp\left[-i\frac{\pi}{2}\left(\lambda + \frac{1}{2}\right)\right]\left(\frac{\pi \pm kr}{2}\right)^{1/2} H^{(2)}(\pm kr) \quad [-k = \exp(-i\pi)k] \tag{10b}$$

are the Schläfli's integral representations of the Hankel function of the second kind. Using the asymptotic expression for this function we have

$$f_\pm(\lambda,k,r) = h_\pm(\lambda,k,r) = \exp(\mp ikr) \quad \text{as } r\to\infty \quad (-\pi < \arg k < \pi). \tag{11}$$

This is also why the solutions (10a) are the Jost ones. As we are to expect, they are even functions of λ, $f_\pm(\lambda,k,r) = f_\pm(-\lambda,k,r)$ and $f_-(\lambda,k,r) = f_+(\lambda,e^{-i\pi}k,r)$, provided $-\frac{1}{2}\pi < \arg k < \frac{3}{2}\pi$.

In view of the exponential factors the series in Eq. (10a) converges absolutely and uniformly for $r > 0$ and for any finite λ and k provided $\nu_{-1} < \infty$. Thus the expression (10a) represents an analytic function of k with the posssible branch point at $k = 0$ and an analytic function of λ at the regular points k ($r > 0$). In the first quadrant of the λ plane

$$\left|\left[\frac{s-u_{p-1}-v_p}{s-u_p-v_p}\right]^\lambda\right| \leqslant \left[\frac{(u_{p-1}+v_p-x)^2+y^2}{(u_p+v_p-x)^2+y^2}\right]^{(1/2)\mathrm{Re}\lambda}\exp\left[-\frac{z_p|y|}{(u_{p-1}+v_p+\frac{1}{2}z_p+|x|+|y|)^2}\mathrm{Im}\lambda\right] \quad (s = x+iy) \tag{12}$$

for $y = \mathrm{Re}s < 0$. The contours $C_\pm$ always can be chosen in such a way that the bracket in the first factor of the right-hand side of (12) is less than unity. Because of the two possible sign choices of λ in the expression (10a) we have then, asymptotically,

$$f_\pm(\lambda,k,r) \sim h_\pm(\lambda,k,r) \tag{13}$$

as $\mathrm{Re}\lambda \to \pm\infty$ ($\mathrm{Im}\lambda = \text{const.}$). The same relation holds as $\mathrm{Im}\lambda \to \pm\infty$ ($\mathrm{Re}\lambda = \text{const.}$). This is due to the second factor in the right-hand side of (12) as far as the contours $C_\pm$ lie in the lower half-plane $\mathrm{Im}s \leqslant 0$ for $\mathrm{Im}\lambda \to \infty$ and in the upper half-plane $\mathrm{Im}s \geqslant 0$ for $\mathrm{Im}\lambda \to -\infty$ respectively. We can find, in general, regions of k and rays $\lambda = |\lambda|\, e^{i\varphi}$, where the relation (13) holds as $|\lambda| \to \infty$.

The combinations

$$\begin{aligned}\chi(\pm\lambda,k,r) &= \exp\left[i\frac{\pi}{2}\left(\pm\lambda - \frac{1}{2}\right)\right] f_+(\lambda,k,r) - \exp\left[-i\frac{\pi}{2}\left(\pm\lambda - \frac{1}{2}\right)\right] f_-(\lambda,k,r) \\ &= -\pi^{-1/2}\left(\frac{k}{2}\right)^{\pm\lambda+1/2} r^{\mp\lambda+1/2} \int_C ds\, s^{\mp\lambda-1} \exp\left(sr^2 - \frac{k^2}{4s}\right)\left\{1 + \sum_{n=1}^{\infty}\left(\frac{-g}{4}\right)^n \int_0^\infty dz_1 \int_\tau^\infty d\mu(t_1)\cdots\right. \\ &\times \int_0^\infty dz_n \int_\tau^\infty d\mu(t_n) \exp\left(-(u_n+v_n)r^2\right) \frac{(s-u_0-v_1)^{\pm\lambda-1}}{(s-u_1-v_1)^{\pm\lambda+1}} \cdots \frac{(s-u_{n-1}-v_n)^{\pm\lambda-1}}{(s-u_n-v_n)^{\pm\lambda+1}} \\ &\left.\times \exp\left[\frac{k^2}{4}\sum_{p=1}^{n}\left(\frac{1}{(s-u_{p-1}-v_p)} - \frac{1}{s-u_p-v_p}\right)\right]\right\},\end{aligned} \tag{14}$$

where the contour C goes from $\infty \exp[i(\arg k - \gamma - \pi)]$ to $\infty \exp[i(\arg k - \gamma + \pi)]$ encircling the point $s = 0$ in an anticlockwise direction, are, again, solutions of Eq. (3). The first (kinematic, $g = 0$) terms of the series (14) are Schläfli's and Sonin's integral representations of the Bessel function[3] $\chi_0(\pm\lambda,k,r) = -i(2\pi kr)^{1/2} J_{\pm\lambda}(kr)$. The functions $\chi(\pm\lambda,k,r)$ reproduce themselves under the substitution $k \to e^{-i\pi}k$

$$\chi(\pm\lambda, e^{-i\pi}k, r) = \exp[-i\pi(\pm\lambda + 1/2)]\,\chi(\pm\lambda,k,r). \tag{15}$$

The Jost solution $f_\pm(\lambda,k,r)$ can be expressed by the function $x(\pm\lambda,k,r)$ for nonintegral values of λ. The double application of the relation (15) leads then to the known circuit relations of the Jost solutions.[4] We do not give them as many other well-known and general relations,[4-7] as they simply follow from the derived relations or expressions.

4. THE JOST FUNCTIONS

According to Refs. 4–7 the Jost functions are defined by the limits

$$\begin{aligned}f_\pm(\lambda,k) &= \lim_{r\to 0} 2\lambda r^{\lambda-1/2} f_\pm(\lambda,k,r) \\ &= h_\pm(\lambda,k) F_\pm(\lambda,k),\end{aligned} \tag{16}$$

where $F_\pm(\lambda,k)$ are the known normalized Jost functions and the kinematic factor

$$\begin{aligned}h_\pm(\lambda,k) &= \lim_{r\to 0} 2\lambda r^{\lambda-1/2} h_\pm(\lambda,k,r) \\ &= \pi^{-1/2} 2^{\lambda+1/2} \exp[-i(\pi/2)(\lambda - 1/2)] \\ &\quad \times \Gamma(\lambda+1)(\pm k)^{-\lambda+1/2}.\end{aligned}$$

We can substitute the expression (10a) with either sign of λ for $f_\pm(\lambda,k,r)$ in this limit.

The limit corresponding to the plus alternative of λ is easily carried out; we just put $r = 0$ in the integrands of (10a). The convergence of the series in Eq. (10a) changes in the limit; it is absolute and uniform for $\mathrm{Re}\lambda \geqslant 1$ provided $\nu_{-1} < \infty$ and for $-1 \leqslant \mathrm{Re}\lambda < 1$ provided $\nu_0 < \infty$. Because of

$$\begin{aligned}&\lim_{r\to 0} r^{2\lambda} \int_0^\infty dz_n \exp\left(-z_n r^2 - \frac{k^2}{4(s-u_{n-1}-v_n-z_n)}\right) \\ &\quad \times (u_{n-1} + v_n - s + z_n)^{\lambda-1} \\ &\quad = \int_0^\infty dz\, z^{\lambda-1} e^{-z} = \Gamma(\lambda) \quad \text{for } \mathrm{Re}\lambda > 0,\end{aligned} \tag{17}$$

the minus alternative leads to the expressions

$$\begin{aligned}&f_\pm(\lambda,k) - h_\pm(\lambda,k) \\ &= 2\pi^{-1/2} \exp\left[\pm i\frac{\pi}{2}\left(\lambda + \frac{1}{2}\right)\right] \Gamma(\lambda+1)\left(\frac{k}{2}\right)^{1/2-\lambda} \\ &\quad \times \int_{C_-} ds\, s^{\lambda-1} \sum_{n=1}^{\infty}\left(\frac{-g}{4}\right)^n \int_0^\infty dz_1 \int_\tau^\infty d\mu(t_1)\cdots \\ &\quad \times \int_0^\infty dz_{n-1} \int_\tau^\infty d\mu(t_{n-1}) \int_\tau^\infty d\mu(t_n) \exp\left[-\frac{k^2}{4}\right. \\ &\quad \left.\times \sum_{p=1}^{n}\left(\frac{1}{s-u_{p-1}-v_{p-1}} - \frac{1}{s-u_{p-1}-v_p}\right)\right] \\ &\quad \times \frac{(u_1+v_1-s)^{\lambda-1}}{(u_0+v_1-s)^{\lambda+1}} \cdots \frac{(u_{n-1}+v_{n-1}-s)^{\lambda-1}}{(u_{n-1}+v_n-s)^{\lambda+1}} \\ &= -g\pi^{-1/2} \exp\left[\pm i\frac{\pi}{2}\left(\lambda + \frac{1}{2}\right)\right]\left(\frac{k}{2}\right)^{1/2-\lambda} \\ &\quad \times \int_{C_-} ds\, s^{\lambda-1} \exp\left(-\frac{k^2}{4s}\right) \int_\tau^\infty d\mu(t)\, \omega(\lambda,k,t-s) \\ &\quad = \int_0^\infty dr\, V(r) h_\pm(\lambda,k,r)\, \varphi(\lambda,k,r),\end{aligned} \tag{18}$$

according to Eqs. (10a), (6), (4), (1), and (10b). Here, the contours $C_\pm$ can be deformed rather arbitrarily, since there are no limitations as to the choice of the phases $(\arg k - \gamma \mp \pi)$ which is obliged in the formula (10a) due to the exponential factor e^{sr^2}. We see that the properties of the Jost functions $f_\pm(\lambda,k)$ follow from those of the function $\omega(\lambda,k,s)$. Therefore, the Jost functions are analytic functions of k except for the branch point $k = 0$ in general and of λ in

the region $\mathrm{Re}\lambda > -1-\kappa$ except for the simple poles at $\lambda = -1, -2, \cdots$, provided the function $\mu(t)$ belongs to the class M_κ. Their circuit relations are determined by those for the Jost solutions.

The Jost functions $f_\pm(\lambda,k)$ differ from the normalized Jost functions $F_\pm(\lambda,k)$ by the factors $h_\pm(\lambda,k)$. If we choose $\arg k + \gamma = \mp\pi$, i.e., $-\frac{3}{4}\pi < (\arg k,\gamma) < -\pi/4$ and $\pi/4 < (\arg k,\gamma) < \frac{3}{4}\pi$ respectively we can identify the contours $C_\pm$ with the half axes $[0,\infty)\exp(\mp i\pi)$. A substitution $s = \exp(\mp i\pi)x$ shows then, that the normalized Jost functions are identical (in the form) to the function $\sigma(\lambda,k,0)$ of Eq. (6b).

We have demonstrated the direct way of deriving the well-known general relation[4-6] between the Jost functions and the regular solution by the last expression in Eq. (18). Substituting the Jost solutions for the regular solution in this expression, we could derive the other well-known integral relation[4-6] between the normalized Jost functions and the Jost solutions

$$F_\pm(\lambda,k) = 1 - i(\pi/2k)^{1/2}\exp\left[i(\pi/2)\left(\lambda + \frac{1}{2}\right)\right] \times \int_0^\infty dr\, r^{1/2}\, V(r)\, J_\lambda(kr) f_\pm(\lambda,k,r). \tag{19}$$

5. BEHAVIOR OF THE JOST FUNCTIONS AS $|\lambda|\to\infty$ ($\mathrm{Im}k = 0$)

The asymptotic behavior of the Jost functions, which plays an important role in the Sommerfeld–Watson transformation, can be best derived from the second expression in Eqs. (18). As k is real and $\geqslant 0$, we choose the phases $\gamma_\pm$ determining the contours $C_\pm$ such that $\pi/2 > \mp\gamma_\pm > 0$ and deform the contours $C_\pm$ in such a way that, leaving the origin in the $\gamma_\pm$ directions, they stick quickly to the negative (positive) y axis. By the substitution $s = \exp(\mp i\pi/2)w$ they are transformed into the new contours $W_\pm$ which coincide with the x axis except for the neighborhood close to $w = 0$. The expressions for the normalized Jost functions then read

$$\begin{aligned} F_\pm(\lambda,k) - 1 &= \pm i\int_{W_\pm} dw \sum_{n=1}^\infty \left(-\frac{g}{4}\right)^n \int_0^\infty dz_1 \int_\tau^\infty d\mu(t_1)\cdots \\ &\times \int_0^\infty dz_{n-1}\int_\tau^\infty d\mu(t_{n-1})\int_\tau^\infty d\mu(t_n) \\ &\times \frac{(u_0 - v_0 \pm iw)^{\lambda-1}\cdots(u_{n-1}+v_{n-1}\pm iw)^{\lambda-1}}{(u_0+v_1\pm iw)^{\lambda+1}\cdots(u_{n-1}+v_n\pm iw)^{\lambda+1}} \\ &\times \exp\left[\frac{k^2}{4}\sum_{p=1}^n t_p(u_{p-1}+v_{p-1}\pm iw)^{-1}\right. \\ &\left.\times (u_{p-1}+v_p\pm iw)^{-1}\right] \end{aligned} \tag{20}$$

[and $F_+^*(\lambda^*,k) = F_-(\lambda,k)$ as expected]. If the deviation of the contours $W_\pm$ from the x axis is small, which means exactly that $2|y| < \tau$ for any point $w = x + iy$ of $W_\pm$, then either factor in the absolute value

$$\left|\left(\frac{u_{p-1}+v_{p-1}\pm iw}{u_{p-1}+v_p\pm iw}\right)^\lambda\right| = \left[\frac{(u_{p-1}+v_{p-1}\mp y)^2 + x^2}{(u_{p-1}+v_p\mp y)^2 + x^2}\right]^{(1/2)\mathrm{Re}\lambda} \times \exp\{-\mathrm{Im}\lambda\,[\arg(u_{p-1}+v_{p-1}\pm iw) - \arg(u_{p-1}+v_p\pm iw)]\} \quad (p = 1,\ldots,n) \tag{21}$$

is less than unity for $\mathrm{Re}\lambda > 0$ and $\pm\mathrm{Im}\lambda > 0$. But these relations determine the magnitude of the integrands in the expression (20) as for their dependence on λ, so that $F_\pm(\lambda,k)\sim 1$ for $\mathrm{Re}\lambda\to\infty$ ($\mathrm{Im}\lambda = \text{const.}$) and for $\pm\mathrm{Im}\lambda\to\infty$ ($\mathrm{Re}\lambda = \text{const.}$). They diverge asymptotically for $\mp\mathrm{Im}\lambda\to\infty$ ($\mathrm{Re}\lambda = \text{const.}$) as one cannot find any allowed (deformed) contour $W_\pm$, for which the right-hand sides of Eq. (20) would assume a finite value for these transitions.

This conclusion can be corroborated independently by the investigation of the asymptotic behavior of the integral expression (19), in which the relation (13) and the limit (16) prove useful. We do not demonstrate it as it is not so lucid as the approach used.

5. CONCLUSIONS

According to the approach used the main difference between the Jost solutions and the Jost functions, which correspond to the potentials of the general Yukawa and Gaussian type, is the absence of the natural cuts in the complex k plane for the latter type. We have also shown, in accordance with the general theory[4] and unlike the former type, that the normalized Jost functions corresponding to the general Guassian potentials have the unity asymptotic behavior for one of the alternative $\mathrm{Im}\lambda\to\pm\infty$ ($\mathrm{Re}\lambda = \text{const.}$) only and real k. There is also a different asymptotic behavior with respect to the complex variable k in either case.

[1] L. Trlifaj, Aplikace Matematiky **22**, 189 (1977).
[2] L. Trlifaj, "The Jost solutions for Yukawa potentials" (submitted to J. Math. Phys.).
[3] G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., New York, 1945).
[4] R.G. Newton, *The Complex j-Plane* (Benjamin, New York, Amsterdam, 1964).
[5] V. De Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[6] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[7] K. Chadan and P.C. Sabatier, *Inverse Problems in Quantum Scattering* (Springer, New York, 1977).

Partial inner product spaces. III. Compatibility relations revisited

J.-P. Antoine
Institut de Physique Théorique, Université Catholique de Louvain, Belgium [a)]

(Received 22 June 1979; accepted for publication 24 August 1979)

This paper continues our systematic study of partial inner product spaces. We show here that a linear compatibility relation on a vector space V is characterized by special families of vector subspaces of V, called involutive coverings, and vice versa. This result provides the link between partial inner product spaces, defined in an intrinsic way, and various concrete structures, such as rigged or nested Hilbert spaces. Given a linear compatibility, generating sets ("rich subsets") are discussed, and several examples are worked out. Finally, we introduce an order relation among all linear compatibilities on the same vector space.

1. INTRODUCTION

In the first paper of this series (Ref. 1, to be denoted by I in the sequel), we have introduced the concept of partial inner product space (PIP-space) in a rather intrinsic way. The whole structure is entirely determined by the partial inner product and its domain, the so-called compatible pairs of vectors. This procedure is, in a sense, the reverse of the familiar construction of "super-Hilbert" spaces, such as rigged,[2] equipped,[3] extended,[4] or nested Hilbert spaces.[5] Very schematically, there one starts from a family of Hilbert spaces and "ties" them together by duality or compatibility conditions. As was stated in I, all those structures are actually particular cases of PIP-spaces.

The aim of the present paper is to clarify the relationship between the two approaches, and also to work out a number of useful examples. In fact we will stay at the level of compatibility relations, the inner product itself will play no role; in other words, the treatment is purely algebraic, no topology will be needed.

We consider first the "constructive" approach (Sec. 2). The essence of the latter is abstracted in the concept of *involutive covering* of a vector space. This means, a covering family of vector subspaces, stable under finite intersection and equipped with a natural involution, which makes it into a lattice. Then we show, in Sec. 3, the equivalence of this approach with the intrinsic one developed in I. More precisely, an involutive covering $\mathscr{I}$ of V uniquely defines a linear compatibility $\#$ on V, such that the complete involutive lattice of assaying subsets $\mathscr{F}(V,\#)$ is the lattice completion of $\mathscr{I}$. Conversely, given V and $\#$, $\mathscr{F}(V,\#)$ is an involutive covering of V. In addition, we construct explicitly the complete lattice generated by the family of Lebesgue spaces, $\mathscr{I} = \{L^p([0,1];dx), 1 < p < \infty\}$ or, more generally, any reflexive chain (i.e., a totally ordered family with a natural involution) of Banach spaces.

The discussion of Sec. 3 shows that the complete lattice $\mathscr{F}(V,\#)$ of assaying subsets for a given linear compatibility $\#$ can be, in fact, recovered from much smaller families of subspaces. Such generating families, called *rich subsets*, are analyzed in Sec. 4. Several examples are worked out, all of which possess a rich subset consisting entirely of Hilbert spaces: sequence spaces, spaces of locally integrable functions, spaces of functions (or sequences) of "prescribed growth," i.e., whose behavior at infinity is characterized in terms of a fixed family of functions (e.g., functions of polynomial or exponential growth). The existence of such rich subsets in crucial for practical applications; indeed they play the same role as bases of neighborhoods or of open sets do in topology. They will be studied further, together with topological properties, in the next paper of the series.[6]

In the last section we turn to the problem of comparing different compatibilities on the same vector space. *A priori* several order relations may be considered. It turns out, the one that is customarily used in lattice theory is useless in the present context. Another one is introduced, which says that a given compatibility $\#_1$ on V is *coarser* than another one $\#_2$ iff the complete lattice $\mathscr{F}(V,\#_1)$ is a sublattice of $\mathscr{F}(V,\#_2)$, on which the two involutions coincide (*involutive* sublattice). This concept is useful for the construction of PIP-space structures on a given vector space V. Most vector spaces used in mathematical physics carry a natural (partial) inner product, defined on a suitable domain $\Gamma \subseteq V \times V$. With trivial restrictions on Γ (symmetry, bilinearity), the condition: $f \# g$ iff $\{f,g\} \in \Gamma$, actually defines a linear compatibility $\#$ on V. Then *all* linear compatibilities which are admissible for that particular inner product are precisely those that are coarser than $\#$, which in turn are determined by all involutive sublattices of $\mathscr{F}(V,\#)$.

On the other hand, the problem of refining a given compatibility (and then a given PIP-space structure) admits in general no solution, even less a unique maximal one. A counterexample is exhibited (for a PIP-space of analytic functions), in which two different, noncomparable, linear compatibilities have equivalent restrictions to a suitable involutive sublattice. However, partial answers to the refinement problem can be given, but some additional structure is needed, namely topological restrictions on individual assaying subsets (see Ref. 6).

Two appendices conclude the paper. In Appendix A, we rephrase our basic definitions in their natural mathematical framework, namely, Galois connections on partially ordered sets. The lesson is that, in fact, the linearity condition

[a)]Postal address: Institut de Physique Théorique UCL, Chemin du Cyclotron, 2, B-1348 Louvain-La-Neuve, Belgium.

 0022-2488/80/020268-12$1.00

on a compatibility is not essential at all: The whole theory goes through (as already mentioned in I) for a compatibility on an arbitrary set. We comment on this in Appendix B, and present a few examples which offer new insights into familiar subjects (such as topology for instance).

2. INVOLUTIVE COVERING OF A VECTOR SPACE

2.1. *Definitions*: Let A be a partially ordered set, with order $\leqslant$. An (*order reversing*) involution on A is a bijection $r \to \bar{r}$ of A onto itself, such that

(i) $\bar{\bar{r}} = r, \quad \forall\, r \in A$,

(ii) $p \geqslant q$ implies $\bar{p} \leqslant \bar{q}$, $\forall\, p, q \in A$.

In fact, by (i), $p \geqslant q$ iff $\bar{p} \leqslant \bar{q}$.

Let now F be a lattice, with lattice operations $\wedge$ and $\vee$. Then for any involution $r \leftrightarrow \bar{r}$, condition (ii) above is equivalent to

(iii) $\overline{p \vee q} = \bar{p} \wedge \bar{q}$, $\forall\, p, q \in F$.

In the sequel, we will call *involutive lattice* any lattice with an involution. Similarly a *complete involutive lattice* is a complete lattice F with an involution $r \leftrightarrow \bar{r}$ that verifies:

(iii') $\overline{\left(\bigvee_{r_j \in J} r_j\right)} = \bigwedge_{r_j \in J} \bar{r}_j$, for any family $J \subseteq F$.

It should be noticed that condition (iii') holds true for any involution on a lattice, complete or not, whenever the l.u.b. on the l.h.s. exists in the lattice.

If F is a (complete) involutive lattice, a subset $F_1 \subseteq F$ is a (*complete*) *involutive sublattice* of F if it is a sublattice stable under the involution.

2.2. *Remarks*: (a) An involution should not be confused with a complementation: Even if the lattice has a greatest element ω and a least element $\alpha = \bar{\omega}$, one has in general $r \vee \bar{r} \neq \omega$ and $r \wedge \bar{r} \neq \alpha$.

(b) Complete involutive lattices arise naturally in the theory of Galois connections; a summary of the latter approach is given to Appendix A.

Now we particularize those concepts to families of subspaces of a vector space V, that is, subsets of $\mathscr{L}(V)$, the set of all vector subspaces of V, ordered by inclusion. Let $\mathscr{I} = \{V_r\}_{r \in I} \subset \mathscr{L}(V)$ be such a subset.

2.3. *Definition*: The family $\mathscr{I} = \{V_r\}_{r \in I} \subset \mathscr{L}(V)$ is called an *involutive covering* of V if:

(i) $\mathscr{I}$ is a covering of V: $\bigcup_{r \in I} V_r = V$;

(ii) $\mathscr{I}$ is stable under finite intersection;

(iii) $\mathscr{I}$ carries an involution $V_r \leftrightarrow V_{\bar{r}}$.

2.4. *Proposition*: Any involutive covering of V is an involutive lattice, with respect to the given involution and the following lattice operations:

$$V_p \wedge V_q = V_p \cap V_q,$$
$$V_p \vee V_q = V_s \quad \text{where } V_{\bar{s}} = V_{\bar{p}} \cap V_{\bar{q}}.$$

Proof: Simple verification that $\vee$ defines indeed an l.u.b. □

Notice that, in general, one has $V_p \vee V_q \subseteq V_r + V_q$, i.e., $\mathscr{I}$ need not be a sublattice of $\mathscr{L}(V)$: The g.l.b. $\wedge$ is always given by set intersection, but the l.u.b. $\vee$ is determined by the involution. In the sequel we shall use the notation $V_{p \wedge q} \equiv V_p \wedge V_q$, $V_{p \vee q} \equiv V_p \vee V_q$.

2.5. *Proposition*: Let $\mathscr{I}$ be an involutive covering of V. Then every subspace $V_r \in \mathscr{I}$ satisfies the following identity:

$$V_{\bar{r}} = \bigcap_{f \in V_r} \left(\cup\{V_{\bar{q}} \mid V_q \in \mathscr{I}, V_q \ni f\}\right). \tag{1}$$

Proof: (a) For any involutive covering, one has obviously:

$$V_{\bar{r}} \subseteq \bigcap_{f \in V_r} \left(\cup\{V_{\bar{q}} \mid V_q \in \mathscr{I}, V_q \ni f\}\right).$$

(b) To prove the inclusion in the other direction, let g be a vector in the r.h.s. of Eq. (1). That is, given any $h \in V_r$, there exists a $V_q \in \mathscr{I}$, depending on h ($q \equiv q(h)$) such that $h \in V_q$ and $g \in V_{\bar{q}}$. *A fortiori* we have:

$$h \in V_{p(h)} \equiv V_{q(h)} \cap V_r \subseteq V_r,$$
$$g \in V_{\overline{p(h)}} \subseteq V_{\overline{q(h)}},$$

and $V_{p(h)} \in \mathscr{I}$, the lattice being stable under intersection. Hence

$$\bigvee_{h \in V_r} V_{p(h)} \subseteq V_r.$$

On the other hand, every $h \in V_r$ is contained in some $V_{p(h)}$, so:

$$V_r \subseteq \bigcup_{h \in V_r} V_{p(h)} \subseteq \bigvee_{h \in V_r} V_{p(h)}.$$

Thus $\bigvee_{h \in V_r} V_{p(h)}$ exists in $\mathscr{I}$ and equals V_r. This implies that condition (iii') of Def. 2.1. holds true:

$$V_{\bar{r}} = \bigcap_{h \in V_r} V_{\overline{p(h)}}.$$

Since $g \in V_{\overline{p(h)}}$ for every $h \in V_r$, it follows that $g \in V_{\bar{r}}$. □

2.6. *Remark*: Notice that the set $\cup\{V_{\bar{q}} \mid V_q \in \mathscr{I}, V_q \ni f\}$ is in fact a vector subspace of V. For let h, g belong to some $V_{\bar{p}}, V_{\bar{q}}$ in that union. Thus $f \in V_p \cap V_q = V_{p \wedge q}$; hence $V_{\bar{p} \vee \bar{q}}$ belongs to the union also, and h, g and $\lambda h + \mu g (\lambda, \mu \in \mathbb{C})$ all belong to $V_{\bar{p} \vee \bar{q}}$.

3. LINEAR COMPATIBILITY VERSUS INVOLUTIVE COVERING

In the paper I, we have introduced the notion of linear compatibility on a vector space. As we shall see, this concept is equivalent to that of involutive covering of Sec. 2. For convenience, we repeat here the basic definitions.

A *linear compatibility* on a vector space V is a symmetric binary relation $\#$ on V, such that, for every $f \in V$, the set $\{f\}^{\#} = \{g \in V | g \# f\}$ is a vector subspace of V. For any subset $S \subseteq V$, we write $S^{\#} = \cap_{f \in S} \{f\}^{\#}$. The subsets (in fact subspaces) $S \subseteq V$ such that $S^{\#\#} = S$ are called *assaying subsets*; they are precisely the sets of the form $S = Z^{\#}$, for some $Z \subseteq V$. We denote by $\mathscr{F}(V,\#)$ the family of all assaying subsets, ordered by inclusion. Then the basic result[1] reads (see also Appendix A):

3.1. *Theorem*: Let V be a vector space with a linear compatibility $\#$. Then the family $\mathscr{F}(V,\#)$ of all assaying subsets is an involutive covering of V, with the involution $V_r \leftrightarrow V_{\bar{r}} = (V_r)^{\#}$. Moreover, $\mathscr{F}(V,\#)$ is a complete involutive lattice with the following lattice operations:

$$\bigwedge_{j \in J} V_j = \bigcap_{j \in J} V_j, \quad \bigvee_{j \in J} V_j = \left(\sum_{j \in J} V_j \right)^{\#\#}. \qquad (2)\square$$

Conversely, if we start with an involutive covering $\mathscr{S}$ of V, we can associate to it a linear compatibility such that $\mathscr{F}(V,\#)$ is the lattice completion of $\mathscr{S}$.

3.2. *Theorem*: Let V be a vector space with an involutive covering $\mathscr{S}$. Consider in the Cartesian product $V \times V$ the subset $\Delta = \cup_{V_r \in \mathscr{S}} V_r \times V_{\bar{r}}$ and define $f \# g$ to mean $\{f,g\} \in \Delta$. Then:

(i) $\#$ is a linear compatibility relation;

(ii) $\mathscr{S}$ is an involutive sublattice of $\mathscr{F}(V,\#)$;

(iii) Every element of $\mathscr{F}(V,\#)$ is an intersection of elements of the form $\{\cap_{V_r \in \mathscr{S}'} V_r\}^{\#}$ where $\mathscr{S}' \subseteq \mathscr{S}$, i.e., $\mathscr{S}(V,\#)$ is the complete involutive lattice generated by $\mathscr{S}$ through unrestricted lattice operations.

Proof: (i) is easily verified. In order to prove (ii), notice that Eq. (1) can be rewritten as $V_{\bar{r}} = (V_r)^{\#} = \cap_{f \in V_R} \{f\}^{\#}$. This means that all subspaces $V_r \in \mathscr{S}$ are assaying subsets and that the involution $V_r \leftrightarrow V_{\bar{r}}$ coincides with the involution $V_r \leftrightarrow (V_r)^{\#}$. In order to prove (iii), we show first that, for every $h \in V$, one has:

$$\{h\}^{\#\#} = \cap \{V_r | V_r \in \mathscr{S}, V_r \ni h\}. \qquad (3)$$

Indeed, by (i), and Remark 2.4. we have:

$$\{h\}^{\#} = \cup \{V_{\bar{r}} | V_r \in \mathscr{S}, V_r \ni h\}$$
$$= \sum \{V_{\bar{r}} | V_r \in \mathscr{S}, V_r \ni h\}$$

so that

$$\{h\}^{\#\#} = \left(\sum \{V_{\bar{r}} | V_r \in \mathscr{S}, V_r \ni h\} \right)^{\#}$$
$$= \cap \{V_r | V_r \in \mathscr{S}, V_r \ni h\}.$$

To conclude, it suffices to notice that an arbitrary element $V_q \in \mathscr{F}(V,\#)$ can be written as:

$$V_q = \bigcap_{h \in V_q} \{h\}^{\#} = \bigcap_{h \in V_q} (\cap \{V_r | V_r \in \mathscr{S}, V_r \ni h\})^{\#}. \qquad \square$$

3.3. *Remark*: Of course, if the set $\mathscr{S}$ is finite, $\mathscr{F}(V,\#) = \mathscr{S}$. This case, although trivial in the present context, is important for applications. Indeed it covers already such concepts as *rigged*,[2] or *equipped Hilbert spaces*,[3] which are widely used in practice. See, e.g., Ref. 7 for an analysis of some quantum mechanical Hamiltonians in a five-element scale of Hilbert spaces.

Theorem 3.2 makes contact between the abstract formulation of I and the more concrete "constructive" approach developed previously for particular cases such as chains (or scales) of Hilbert or Banach spaces,[8,9] nested Hilbert spaces,[5] or rigged Hilbert spaces.[2] In all of these situations, the starting point is a family of vector spaces $\{V_r\}$ which form an involutive lattice (or a scale), and V is defined as $\cup_r V_r$ (technically, algebraic inductive limit). Theorem 3.2 shows that the two approaches are, in fact, equivalent. In other words, it makes clear that the concept of partial inner product space is a genuine generalization of all these particular structures.

It also follows that all we need is an involutive covering of V, the complete lattice $\mathscr{F}(V,\#)$ is then uniquely determined. In most cases, the lattice completion will remain implicit, for $\mathscr{F}(V,\#)$ is uncomfortably large. (This is just like a topology: One can seldom exhibit explicitly "all" open sets!) But in certain cases, an arbitrary assaying subset may be described, as will be clear from the following examples.

3.A. Example: Sequence spaces

As in I, Example 4(2), consider $V = \omega$, the space of all complex sequences, with the compatibility

$$(x_n) \# (y_n) \Longleftrightarrow \sum_n |x_n y_n| < \infty.$$

An assaying subset is precisely a *perfect* sequence space, as defined by Köthe.[10] This example shows how big and unpractical the complete lattice $\mathscr{F}(V,\#)$ can be: This set $\mathscr{F}(\omega,\#)$ contains almost all possible types of topological vector spaces, many of them with rather awkward properties (it is Köthe's main source of counterexamples!). Thus it is imperative to restrict ourselves to suitable subsets of $\mathscr{F}(\omega,\#)$.

We will consider, in particular, the family $\mathscr{S}$ of all assaying subsets of the form

$$\ell^2(r) = \{(x_n) \in \omega | (x_n r_n^{-1/2}) \in \ell^2\},$$

where $r \equiv (r_n)$ is an arbitrary sequence of positive numbers. Then $\mathscr{S}$ is an involutive covering of ω:

(i) $\mathscr{S}$ is a covering of ω: Given $(x_n) \in \omega$ there is a weight sequence $r = (r_n)$ such that $(x_n) \in \ell^2(r)$; take, for instance, the following weights:

$r_n = |x_n|^2 n^2$ whenever $x_n \neq 0$,

r_n arbitrary whenever $x_n = 0$.

(ii) $\mathscr{S}$ is a lattice for the following operations:

$\ell^2(r) \wedge \ell^2(s) = \ell^2(p)$ where $p_n = \min\{r_n, s_n\}$,

$\ell^2(r) \vee \ell^2(s) = \ell^2(q)$, where $q_n = \max\{r_n, s_n\}$.

(iii) $\mathscr{S}$ is an involutive lattice, with the involution

$r \leftrightarrow \bar{r} \equiv (r_n^{-1})$. Indeed:

$$[\mathscr{P}(p)]^{\#} = \mathscr{P}(\bar{p}) = \mathscr{P}(\bar{r}) \vee \mathscr{P}(\bar{s}),$$

$$[\mathscr{P}(q)]^{\#} = \mathscr{P}(\bar{q}) = \mathscr{P}(\bar{r}) \vee \mathscr{P}(\bar{s}).$$

Actually it turns out that $\mathscr{I}$ is a sublattice of $\mathscr{L}(\omega)$, i.e.:

$$\mathscr{P}(r) \wedge \mathscr{P}(s) = \mathscr{P}(r) \cap \mathscr{P}(s),$$

$$\mathscr{P}(r) \vee \mathscr{P}(s) = \mathscr{P}(r) + \mathscr{P}(s).$$

This is an example of a fairly general situation that will be studied systamatically in the next paper of this series.[6]

According to Theorem 3.2., we can now consider the compatibility $\#_{\mathscr{I}}$ corresponding to the family $\mathscr{I}$. We will see in Sec. 4 that, in fact, $\#_{\mathscr{I}} = \#$, and thus the complete involutive lattice generated by $\mathscr{I} = \{\mathscr{P}(r)\}$ consists of *all* perfect sequence spaces.

3.B. Example: Lebesgue spaces

As a second example, consider the family of Lebesgue spaces over the interval [0,1] with their usual norm topology:

$$\mathscr{I} = \{L^p \equiv L^p([0,1];dx), \quad 1 < p < \infty\}.$$

These spaces form a chain: $p > q$ implies $L^p \subset L^q$, the embedding is continuous and has dense image. With the involution

$$L^p \leftrightarrow (L^p)' = L^{\bar{p}}, \quad p^{-1} + \bar{p}^{-1} = 1,$$

$\mathscr{I}$ is an involutive covering of the space $V = \cup_{1<p<\infty} L^p$. Given the corresponding compatibility $\#$, i.e., $(L^p)^{\#} = L^{\bar{p}}$, we can compute explicitly the complete involutive lattice $\mathscr{F}(V,\#)$, using the results of the extensive study by Davis *et al.*[11]

Remark: Notice that we differ slightly from these authors, in that we do not include the space L^1: By symmetry this would demand inclusion of L^∞ as well, which would invalidate some of the statements made below about duality properties. However, their results apply to our case also, because the first Eq. (4) holds true also with the definition $\hat{S} = (1,t) \cup S$.

First we evaluate "elements of the first generation" of $\mathscr{F}$:

$$L^p(S) \equiv \cap_{p \in S} L^p, \quad L^I(S) \equiv \cap_{p \in S} L^p,$$

where S is an arbitrary subset of $(1,\infty)$. Introducing $r = \inf S$, $t = \sup S$, and defining $\hat{S} = (1,t) \cup S$, $\tilde{S} = (r,\infty) \cup S$, it is shown in Ref. 11 that:

$$L^p(S) = L^p(\hat{S}), \quad L^I(S) = L^I(\tilde{S}). \tag{4}$$

This leaves us with four possible cases:

(i) $t \in S \Rightarrow \hat{S} = (1,t]$ and $L^p(S) = \cap_{1<q\leqslant t} L^q = L^t$;

(ii) $t \notin S \Rightarrow \hat{S} = (1,t)$ and $L^p(S) = \cap_{1<q<t} L^q \equiv L^{t-}$;

(iii) $r \in S \Rightarrow \tilde{S} = [r,\infty)$ and $L^I(S) = \cup_{r\leqslant q<\infty} L^q = L^r$;

(iv) $r \notin S \Rightarrow \tilde{S} = (r,\infty)$ and $L^I(S) = \cup_{r<q<\infty} L^q \equiv L^{r+}$.

Thus we get two new types of spaces, $L^{p\pm}$. Their topological properties are based on the observation that, in the definition of L^{p-}, it is enough [by Eq. (4)] to consider a cofinal countable subset of the V_q's. Then, using the terminology of Floret and Wloka (Ref. 12, §9), L^{p-} is a strict FG-space. Therefore, we get

(i) For $1 < p \leqslant \infty$, $L^{p-} \equiv \cap_{1<q<p} L^q$, with the projective topology, is a non-normable, reflexive Fréchet space, hence barrelled and complete, with dual $(L^{p-})' = (L^{p-})^{\#} = L^{\bar{p}+}$. In particular, $L^{\infty-}$ coincides with the space L^ω of Arens.[13]

(ii) For $1 \leqslant p < \infty$, $L^{p+} \equiv \cup_{p<q<\infty} L^q$, with the inductive topology, is a nonmetrizable (Mackey) complete, barrelled topological vector space, with dual $(L^{p+})' = (L^{p+})^{\#} = L^{\bar{p}-}$.

(iii) Furthermore, the following inclusions are proper:

$$L^{p+} \subsetneq L^p \subsetneq L^{p-} \quad (1 < p < \infty),$$

the embeddings are continuous and have dense image.

3.4. *Proposition*[14]: Let $\mathscr{I}$ be the chain $\{L^p, 1 < p < \infty\}$. Then the complete lattice $\mathscr{F}(V,\#)$ generated by $\mathscr{I}$ is also a chain, obtained by replacing each L^p in $\mathscr{I}$ by the triplet $L^{p+} \subset L^p \subset L^{p-}$, and adding the smallest element $L^{\infty-}$ and the largest element L^{1+}.

Proof: First we evaluate elements of the form $\{f\}^{\#\#}$, then $\{f\}^{\#}$, and finally arbitrary elements of $\mathscr{F}(V,\#)$ through the usual relation:

$$\mathscr{F}(V,\#) \ni V_r = \cap_{f \in V_r} \{f\}^{\#}.$$

Let $f \in V$. Then, by Eq. (3)

$$\{f\}^{\#\#} = \cap_{L^p \ni f} L^p = L^{\bar{q}} \text{ or } L^{\bar{q}-}, \quad \text{for some } \bar{q} \in (1,\infty).$$

Therefore $\{f\}^{\#} = L^q$ or L^{q+} for some $q \in (1,\infty)$. Finally

$$V_r = \cap_{f \in V_r} \{f\}^{\#} = \left(\cap_{q \in S} L^q\right) \cap \left(\cap_{r \in T} L^{r+}\right),$$

where S,T are some subsets of $(1,\infty)$, to be replaced by $\hat{S},\hat{T}$ respectively.

For the first term on the r.h.s., we get

$$\cap_{q \in S} L^q = \cap_{q \in \hat{S}} L^q = L^s \text{ or } L^{s-}, \quad \text{with } s = \sup S.$$

As for the second term, we observe that $r_1 < r_2$ implies $L^{r_2+} \subset L^{r_1+}$ with continuous embedding; the set inclusion is obvious, and the embedding $L^{r_2+} \to L^{r_1+}$ can be factorized continuously through L^r, where r is any real number such that $r_1 < r < r_2$. Therefore all the spaces L^{r+}, $1 \leqslant r < \infty$, form a chain with continuous embeddings. Thus we get

$$\cap_{r \in T} L^{r+} = L^{t+}, \quad \text{where } t = \sup T.$$

Finally, V_r is either of the form $L^s \cap L^{t+}$, or of the form $L^{s-} \cap L^{t+}$. For $s \leqslant t$, we have $L^{s-} \supset L^s \supset L^{t+}$ and $V_r = L^{t+}$. For $s > t$, we have $L^{t+} \supset L^{s-} \supset L^s$, so that $V_r = L^s$ or $V_r = L^{s-}$. This concludes the proof. □

3.C. Generalization: Reflexive chains of Banach spaces

Let $I_0 = (a,b)$ be an open interval of $\mathbb{R}$, and for each $p \in I_0$ let there be a reflexive Banach space V_p. We say that the family $\mathscr{I} = \{V_p\}_{p \in I_0}$ is *a (continuous) reflexive chain of Banach spaces* if the two following conditions hold:

(i) $p < q \Rightarrow V_p \subsetneq V_q$, the inclusion map is injective and continuous;[15]

(ii) I_0 carries an involution $p \leftrightarrow \bar{p}$ such that the dual of V_p is norm-isomorphic to $V_{\bar{p}}$: $(V_p)' \simeq V_{\bar{p}}$.

The following properties follow easily from the definition (see also Ref. 6):

(a) $\mathscr{I}$ is an involutive covering of $V \equiv \cup_{p \in I_0} V_p$, corresponding to the compatibility $(V_p)^\# = V_{\bar{p}}$.

(b) Whenever $p < q$, the inclusion map $V_p \to V_q$ has dense range and each V_p is dense in V (considered as the inductive limit $\lim\overrightarrow{_{p \in I_0}} V_p$).

(c) $V^\# \equiv \cap_{p \in I_0} V_p$ is a dense subspace of every V_p, and of V as well.

Let now $I = [a,b]$ be a closed interval. The analogous definition is obvious, the only difference being that, now $V^\# = V_a, V = V_b$. Within this context, a proposition similar to 3.4. can be formulated. As before we define

$$V_p \equiv \bigcup_{q<p} V_q, \quad V_{p0} \equiv V_p, \quad V_{p+} \equiv \bigcap_{p<q} V_q.$$

Let F_0 be the totally ordered set $F_0 = (a,b) \times \{-, 0, +\}$ with its lexicographic order $(- < 0 < +)$. Then we have:

3.5. *Proposition*: Let $\mathscr{I} = \{V_p\}_{p \in I}$ be a reflexive chain of Banach spaces. Then the complete involutive lattice $\mathscr{F} = \mathscr{F}(V,\#)$ is a chain given explicitly as follows:

(i) If $I = (a,b)$ is an open interval, $\mathscr{F}$ consists of $V^\# \equiv V_{a+}, \mathscr{F}_0, V \equiv V_{b-}$, where $\mathscr{F}_0$ is a chain indexed by F_0.

(ii) If $I = [a,b]$ is a closed interval, $\mathscr{F}$ is the chain $V^\# \equiv V_a, V_{a+}, \mathscr{F}_0, V_{b-}, V \equiv V_b$.

Proof: The argument is exactly the same as the one given in Proposition 3.4 for the case of Lebesgue spaces. First one shows (using the involution $p \leftrightarrow \bar{p}$) that the inclusions $V_{p-} \subset V_{p0} \subset V_{p+}$ are proper. Then, one proves that V_{p+}, with the projective topology, is a reflexive Fréchet space, with dual $V_{\bar{p}-}$, and V_{p-}, with the inductive topology, is a reflexive nonmetrizable, barrelled topological vector space, with dual $V_{\bar{p}+}$. The rest of the argument is then identical and gives (i); as for (ii), it is obvious. □

The situation described above is in fact extremely frequent in applications. The following examples of reflexive chains of Banach spaces are all well known (the first two have the direct order, all the others the inverse order):

(1) the chain of sequence spaces $\{\ell^p, 1 < p < \infty\}$;

(2) the chain of ideals of compact operators in a Hilbert space H: $\{\mathscr{C}^p(H), 1 < p < \infty\}$, which is isomorphic to (1) (see I, Example 4.7);

(3) the chain of Sobolev spaces $\mathscr{H}_s$ $(-\infty < s < \infty)$ defined as follows[16,17]: A tempered distribution $f \in \mathscr{S}'(\mathbb{R}^n)$ belongs to $\mathscr{H}_s$ if its Fourier transform $\hat{f}$ verifies

$$\int d^n\xi \, |\hat{f}(\xi)|^2 (1 + |\xi|^2)^s < \infty;$$

the involution here is $(\mathscr{H}_s)^\# = (\mathscr{H}_s)' = \mathscr{H}_{-s}$;

(4) any chain of Hilbertian spaces, as defined by Palais;[8]

(5) in particular, the chain of Hilbert spaces $\mathscr{F}^\rho$ $(-\infty < \rho < \infty)$ in Bargmann's space $\mathscr{E}'$,[18] defined as follows: $\mathscr{F}^\rho$ consists of all those entire functions $f(z)$, $z \in \mathbb{C}^n$, such that

$$\int_{\mathbb{C}^n} |f(z)|^2 (1 + |z|^2)^\rho e^{-|z|^2} d^n z < \infty$$

and $\mathscr{E}' = \cup_{\rho \in \mathbb{R}} \mathscr{F}^\rho$ [see I, example 4.8 (b)];

(6) a chain of Banach spaces that generalizes the Lebesgue spaces L^p to algebras of unbounded operators has been constructed by Inoue.[19]

3.6. *Remark*: The general considerations given in Secs. 1 and 2 were purely algebraic, no topology was mentioned, except for the explicit examples of Secs. 3.B. and 3.C. Here topological duality was used as a means of defining an adequate involution. We will come back in the next paper[6] to this alternative approach, closer to Refs. 2 and 5.

4. RICH SUBSETS OF $\mathscr{F}(V,\#)$

We have seen in Sec. 3 that an involutive covering on a vector space V uniquely determines a linear compatibility on V. Actually even less is needed: Any family of subsets of V that allows to reconstruct the compatibility will suffice. Considering the fact that the full lattice $\mathscr{F}$ is in general much too large for practical purposes, this result will be most useful for applications. In this section, we will briefly analyze the abstract situation, then exhibit a number of examples.

4.1. *Definition*: A family $\mathscr{I}_0$ of subsets of V is called *rich* for the linear compatibility $\#$ if, given any compatible pair $f, g \in V$, there exists $S \in \mathscr{I}_0$ such that $f \in S^\#$ and $g \in S^{\#\#}$.

Notice that elements of $\mathscr{I}_0$ need not be vector subspaces of V: The involution automatically generates them. But richness is not lost in the process. Indeed, if we define the three following sets:

$$\mathscr{I}_0^{(\#)} = \{S^\# | S \in \mathscr{I}_0\}, \quad \mathscr{I}_0^{(\#\#)} = \{S^{\#\#} | S \in \mathscr{I}_0\},$$

$$\mathscr{I} = \mathscr{I}_0^{(\#)} \cup \mathscr{I}_0^{(\#\#)},$$

we have (the proof is straightforward):

4.2. *Proposition*: Let $\mathscr{J}$ be any one of $\mathscr{I}_0^{(\#)}, \mathscr{I}_0^{(\#\#)}, \mathscr{I}$. Then:

(i) $\mathscr{J}$ is a subset of $\mathscr{F}(V,\#)$:,

(ii) $\mathscr{J}$ is rich and covers V: $\cup_{S \in \mathscr{J}} S = V$,

(iii) $V^\# = \bigcap_{S \in \mathscr{J}} S$. □

The lesson is clear: If $\mathscr{I}_0$ is rich, and $\mathscr{I}_1 \supset \mathscr{I}_0$, then $\mathscr{I}_1$ is also rich. This allows one to enlarge a rich subset by "closing" it with respect to some algebraic operation: Involution (which gives $\mathscr{I}$), lattice operations, or both. In the latter case, we get the involutive sublattice of $\mathscr{F}(V,\#)$ generated by $\mathscr{I}_0$, which is an involutive covering. Anyway there is no loss of generality in assuming a rich subset to contain only assaying subsets (although for some applications it might be more convenient not to do so).

Remark: The above definition may also be phrased as follows. A family $\mathscr{I}$ of assaying subsets is rich iff relation (1) in Proposition 2.5 holds for any assaying subset V_r, or, equivalently, relation (3) of Theorem 3.2 holds for any $h \in V$.

The main reason for introducing rich subsets of $\mathscr{F}(V,\#)$ is topological. Namely, arbitrary assaying subsets, equipped with their Mackey topologies, may be difficult to handle, whereas it is easy in most cases to find *homogeneous* rich subsets, that is rich subsets $\mathscr{I}$ whose elements are all of the same type, such as Hilbert spaces or reflexive Banach spaces. A number of examples are given below. Such rich subsets will be studied in the next paper.[6]

4.A. Example: Hilbert spaces of sequences

We take again Example 3.A, and show that $\mathscr{I} = \{\ell^2(r)\}$ is a rich subset of $\mathscr{F}(\omega,\#)$ consisting of Hilbert spaces only. Let $f,g \in \omega$ with $f \# g$, i.e., $\Sigma_{n=1}^{\infty} |f_n g_n| < \infty$. Define a partition of $n \in \mathbb{N}$ into four disjoint subsets as follows:

$$N_1 = \{n \in \mathbb{N} \mid f_n \neq 0, \quad g_n \neq 0\},$$
$$N_2 = \{n \in \mathbb{N} \mid f_n = 0, \quad g_n \neq 0\},$$
$$N_3 = \{n \in \mathbb{N} \mid f_n \neq 0, \quad g_n = 0\},$$
$$N_4 = \{n \in \mathbb{N} \mid f_n = g_n = 0\}.$$

All four subsets may be finite or infinite. Let $s = (s_n)$ be an arbitrary square integrable sequence, with nonzero elements. Then define a sequence of weights, $r = (r_n)$, as follows:

- for $n \in N_1$, $r_n = |f_n|\,|g_n|^{-1}$,
- for $n \in N_2$, $r_n = |s_n|^2\,|g_n|^{-2}$,
- for $n \in N_3$, $r_n = |f_n|^2\,|s_n|^{-2}$,
- for $n \in N_4$, $r_n > 0$, arbitrary.

Then $f \in \ell^2(r)$, $g \in \ell^2(r^{-1})$; indeed

$$\sum_{n=1}^{\infty} |f_n|^2 r_n^{-1} = \sum_{n \in N_1 \cup N_3} |f_n|^2 r_n^{-1} = \sum_{n \in N_1} |f_n|\,|g_n| + \sum_{n \in N_3} |s_n|^2 < \infty,$$

$$\sum_{n=1}^{\infty} |g_n|^2 r_n = \sum_{n \in N_1 \cup N_2} |g_n|^2 r_n = \sum_{n \in N_1} |f_n|\,|g_n| + \sum_{n \in N_2} |s_n|^2 < \infty.$$

4.B. Example: Locally integrable functions

Let $V = L^1_{\mathrm{loc}}(\mathbb{R}^n, d^n x)$, the space of all functions on $\mathbb{R}^n$ locally integrable with respect to the Lebesgue measure. With the compatibility

$$f \# g \Longleftrightarrow \int |f(x)g(x)|\,dx < \infty$$

one has, as usual, [I, Example 4(5)] $V^{\#} = L^{\infty}_{\mathrm{comp}}(\mathbb{R}^n, d^n x)$, the essentially bounded functions of compact support.

Let $r : \mathbb{R}^n \to \mathbb{R}^+$ be a measurable, a.e. positive function, such that both r and $\bar{r} \equiv r^{-1}$ are locally integrable. Denote by $L^2(r)$ the space of measurable functions $f : \mathbb{R}^n \to \mathbb{C}$ such that $fr^{-1/2}$ is square integrable. Then we claim:

(i) $L^{\infty}_{\mathrm{comp}} \subset L^2(r) \subset L^1_{\mathrm{loc}}$;

(ii) the family of all such $L^2(r)$ is a rich sublattice of $\mathscr{F}(V,\#)$.

Part (i) is a straightforward verification. As for (ii), write $\mathbb{R}^n$ in the form: $\mathbb{R}^n = \cup_{j=1}^{\infty} \Omega_j$ where $\Omega_j = \{x \in \mathbb{R}^n \mid j-1 \leqslant |x| < j\}$. Thus each Ω_j is relatively compact and $\Omega_j \cap \Omega_k = \emptyset$ if $j \neq k$. Let $f,g \in V, f \# g$, and $\{\alpha_j\}, \{\beta_j\}$ two arbitrary sequences of positive numbers such that:

$$\sum_{j=1}^{\infty} \alpha_j \int_{\Omega_j} |f(x)|\,d^n x < \infty, \quad \sum_{j=1}^{\infty} \beta_j \int_{\Omega_j} |g(x)|\,d^n x < \infty.$$

Define

$$r(x) = \sum_{j=1}^{\infty} \chi_j(x) r_j(x),$$

where

$$r_j(x) = \sup(\beta_j, |f(x)|) / \sup(\alpha_j, |g(x)|).$$

χ_j is the characteristic function of Ω_j. (*Remark*: Elements of V are in fact classes of equivalent functions; if in the definition of r_j f or g is replaced by a equivalent function, so is r_j, and nothing changes in the argument.[20]) Thus $r^{-1}(x) = \Sigma_{j=1}^{\infty} \chi_j(x) r_j^{-1}(x)$, and both r and r^{-1} are locally integrable. Furthermore, it is easily shown that $f \in L^2(r)$, $g \in L^2(\bar{r})$. For instance:

$$\int_{\mathbb{R}^n} |f|^2 r^{-1} d^n x = \sum_{j=1}^{\infty} \int_{\Omega_j} |f|^2 \frac{\sup(\alpha_j |g|)}{\sup(\beta_j, |f|)} d^n x$$
$$\leqslant \sum_{j=1}^{\infty} \int_{\Omega_j} |f| \sup(\alpha_j, |g|)\, d^n x$$
$$= \sum_{j=1}^{\infty} \left(\int_{\Omega_j \cap \{x \mid |g(x)| \geqslant \alpha_j\}} |fg|\, d^n x + \int_{\Omega_j \cap \{x \mid |g(x)| < \alpha_j\}} \alpha_j |f|\, d^n x \right)$$
$$\leqslant \sum_{j=1}^{\infty} \left(\int_{\Omega_j} |fg|\, d^n x + \alpha_j \int_{\Omega_j} |f|\, d^n x \right)$$
$$= \int_{\mathbb{R}^n} |fg|\, d^n x + \sum_{j=1}^{\infty} \alpha_j \int_{\Omega_j} |f| d^n x < \infty.$$

Thus $\{L^2(r)\}$ is rich in L^1_{loc}. Moreover it is an involutive sublattice of $\mathscr{F}(V,\#)$, with the following lattice operations:

- $L^2(r) \subseteq L^2(s) \Longleftrightarrow r(x) \leqslant s(x)$ a.e.,
- $L^2(r) \wedge L^2(s) = L^2(p)$, with $p(x) = \inf\{r(x), s(x)\}$,
- $L^2(r) \vee L^2(s) = L^2(q)$, with $q(x) = \sup\{r(x), s(x)\}$,
- $[L^2(r)]^{\#} = L^2(\bar{r})$, with $\bar{r}(x) = r^{-1}(x)$.

4.3. *Remark*: The above result generalizes to the space $L^1_{\mathrm{loc}}(X, d\mu)$, where X is a locally compact and σ-compact space (i.e., $X = \cup_n K_n$, $K_n \subset K_{n+1}$, K_j relatively compact) and μ is a non-negative Radon measure on X. A similar result holds with $V = L^2_{\mathrm{loc}}$ (hence $V^{\#} = L^2_{\mathrm{comp}}$) and $r^{\pm 1} \in L^2_{\mathrm{loc}}$ (instead of L^1_{loc}). Also if X is compact, i.e., $V = L^1(X, d\mu)$, $V^{\#} = L^{\infty}(X, d\mu)$, the statement is simply that the family $\{L^2(r), r^{\pm 1} \in L^1\}$ is rich.

4.C. Example: Functions or sequences of prescribed growth

The result of Example 4.B can be improved if V is restricted to those locally integrable functions which satisfy a growth condition at infinity (such as functions, or sequences, of polynomial or exponential growth), in the sense that the

weight functions $r(x)$ can now be assumed to have the same type of growth.

More precisely, for X and μ as above, let A be a partially ordered set and let $\{F^{(\alpha)}, \alpha \in A\}$ be a family of positive μ-locally integrable functions, indexed by A and monotonically increasing in α:

$$\alpha \leqslant \beta \Longleftrightarrow F^{(\alpha)}(x) \leqslant F^{(\beta)}(x) \quad (\mu\text{- a.e.}) .$$

Assume furthermore that, given $\alpha, \beta \in A$, there exists a positive square integrable function $s_{\alpha\beta}$ which verifies the following inequality for some positive constant $C_{\alpha\beta}$:

$$0 < s^2_{\alpha\beta}(x) \leqslant C_{\alpha\beta} F^{(\alpha)}(x) F^{(\beta)}(x) \quad (\mu\text{- a.e.}) . \tag{5}$$

Define V as the vector space of those functions $f \in L^1_{\text{loc}}(X,d\mu)$ which grow no faster than the functions $F^{(\alpha)}$, i.e., $f \in V$ if there exists $\alpha \in A$ and a constant $c > 0$ such that:

$$|f(x)| \leqslant c F^{(\alpha)}(x) \quad (\mu\text{-a.e.}) .$$

Equip V with the compatibility $\#$ inherited from L^1_{loc}. Then:

4.4. *Proposition*: Let $V, \#$ as above. Consider the family $\mathscr{I} = \{L^2(r)\}$, where each r is a weight function that verifies the following inequalities for some $\alpha, \beta \in A$ and positive constants c', c'':

$$c' [s_{\alpha\beta}(x)/F^{\beta}(x)] \leqslant r^{1/2}(x) \leqslant c'' [F^{(\alpha)}(x)/s_{\alpha\beta}(x)] . \tag{6}$$

Then the family $\mathscr{I}$ is a rich subset of $\mathscr{F}(V,\#)$.

Before proving the statement, let us give a few examples.

(i) *Functions of polynomial growth* [*in* $L^1_{\text{loc}}(\mathbb{R}^n, d^n x)$]:

$$F^{(\alpha)}(x) = (1 + |x|^2)^{\alpha/2} \quad \alpha \in \mathbb{Z} \text{ or } \mathbb{R} ,$$

$$s_{\alpha\beta}(x) = (1 + |x|^2)^{-\gamma/2}$$

with $\gamma > n/2$ and $\gamma \geqslant -\frac{1}{2}(\alpha + \beta)$.

The assaying subsets $L^2(r_{\alpha\beta})$ obtained in this example actually form an involutive sublattice of $\mathscr{F}(V,\#)$ (see Hörmander,[16] Theorem 2.1.1.).

(ii) *Slowly increasing sequences* (*in* ω):

By the same reasoning, we find that the family $\{\ell^2(r)\}$, where $r = (r_n)$ is a sequence of tempered weights, i.e.,

$$c'(1+n)^{-\beta} \leqslant r_n \leqslant c''(1+n)^{\alpha}, \quad c', c'' > 0 ,$$

is rich in the space s' of slowly increasing sequences, equipped with the standard compatibility from ω.

(iii) *Functions of exponential growth* [*in* $L^1_{\text{loc}}(\mathbb{R}^n, d^n x)$]:

$$F^{(\alpha)}(x) = e^{\alpha|x|}, \quad \alpha \in \mathbb{R}$$

$$s_{\alpha\beta}(x) = e^{\gamma|x|}, \quad \text{with } \gamma \in \mathbb{R}, \quad \gamma > \alpha + \beta .$$

(iv) *Entire functions of order* 2 (*in Bargmann's space* $\mathscr{E}'$ [18]):

As in Sec. 3.B, example (5), take $X = \mathbb{C}^n$, with Gaussian measure $d\mu(z) = e^{-|z|^2} dz$; then $\mathscr{E}'$ consists also of those entire functions with growth indexed by the following family:

$$F^{(\alpha)}(z) = e^{1/2|z|^2}(1 + |z|^2)^{-\alpha/2}, \quad \alpha \in \mathbb{R} \text{ or } \mathbb{Z}$$

$$s_{\alpha\beta}(z) = 1, \quad \forall\, \alpha, \beta .$$

Proof of Proposition 4.4: Let $|f(x)| \leqslant c F^{(\alpha)}(x)$, $|g(x)| \leqslant c' F^{(\beta)}(x)$, and $\int_X |f(x)g(x)|\, d\mu < \infty$. We choose a function $s_{\alpha\beta} \in L^2(X,d\mu)$ that verifies Eq. (5). Then we proceed as in Example 4.A, dividing X into four disjoint subsets X_j, $j = 1,\ldots,4$, depending on whether

$$|f(x)| - [s^2_{\alpha\beta}(x)/F^{(\beta)}(x)] \geqslant 0 (X_1 \text{ and } X_3) \text{ or } \leqslant 0 (X_2 \text{ and } X_4) ,$$

$$|g(x)| - [s^2_{\alpha\beta}(x)/F^{(\alpha)}(x)] \geqslant 0 (X_1 \text{ and } X_2) \text{ or } \leqslant 0 (X_3 \text{ and } X_4) .$$

We define a weight function $r_{\alpha\beta}$ as follows:

$$\begin{aligned} r_{\alpha\beta}(x) &= |f(x)|\,|g(x)|^{-1}, \quad x \in X_1 \\ &= s^2_{\alpha\beta}(x)|g(x)|^{-2}, \quad x \in X_2 \\ &= |f(x)|^2 s^{-2}_{\alpha\beta}(x), \quad x \in X_3 \\ &= \text{arbitrary for } x \in X_4 \text{ provided Eq. (6) is verified} . \end{aligned}$$

It is then straightforward to verify that $f \in L^2(r_{\alpha\beta})$, $g \in L^2(\overline{r_{\alpha\beta}})$, and also that $r_{\alpha\beta}$ verifies Eq. (6) on all of X. □

4.D. Two counterexamples

(i) The family $\{\ell^p\}_{1 \leqslant p \leqslant \infty}$ is not rich in ℓ^∞, with the compatibility from ω. Indeed we have:

$f \equiv [(\log n)^{-1}]$ belongs to ℓ^∞, but not to $\ell^{\infty -} \equiv \cup_{p < \infty} \ell^p$,

$g \equiv [(n \log n)^{-1}]$ belongs to $\ell^{1+} \equiv \cap_{1 < p \leqslant \infty} \ell^p$, but not to ℓ^1, and yet

$\Sigma_{n=1}^{\infty} |f_n g_n| = \Sigma_{n=1}^{\infty} n^{-1} (\log n)^{-2} < \infty$ by Ref. 21.

This example carries over immediately to the ideals $\mathscr{C}^p$ $(1 \leqslant p \leqslant \infty)$ of compact operators in a Hilbert space [cf. Sec. 3.B., example (2)].

(ii) The family $\{\ell^2(r^{(k)})\}_{k \in \mathbb{Z}}$ where $r_n^{(k)} = n^k$, is not rich in the space s', for the same compatibility. Take indeed:

$$f_n = n^l \text{ and } g_n = n^{-l-1} (\log n)^{-1-\delta} (\delta > 0) .$$

Then

$$\sum_{n=1}^{\infty} |f_n g_n| = \sum_{n=1}^{\infty} n^{-1} (\log n)^{-1-\delta} < \infty ,$$

but there is no $k \in \mathbb{Z}$ such that $f \in \ell^2(r^{(k)})$ and $g \in \ell^2(\overline{r^{(k)}})$:

$$\sum_{n=1}^{\infty} |f_n|^2 [r_n^{(k)}]^{-1} = \sum_{n=1}^{\infty} n^{2l-k} < \infty, \quad \text{iff } k > 2l+1 ,$$

$$\sum_{n=1}^{\infty} |g_n|^2 r_n^{(k)} = \sum_{n=1}^{\infty} n^{k-2l-2} (\log n)^{-2-2\delta} < \infty, \quad \text{iff } k \leqslant 2l+1 .$$ [21]

It is interesting to notice that further enlarging the family of weights does not improve the situation, as long as it remains totally ordered: For instance, if one uses $\{r^{(k,j)}\}$ with $r_n^{(k,j)} = n^k (\log n)^j$, a similar counterexample can be obtained. The moral is that whenever the compatibility $\#$ is given by absolute convergence of a series or an integral, no totally ordered subset of $\mathscr{F}(V,\#)$ will be rich: Two vectors f, g may be compatible, not because one is "good" and the other is "bad", but because there are cancellations between f_n and g_n. An extreme example is that of two wildly increasing sequences, f and g, such that $f_{2m} = 0$ and $g_{2m+1} = 0$ for all m. Simpler yet, the complete lattice generated by a totally ordered subset will be totally ordered, i.e., a chain, while $\mathscr{F}(V,\#)$ is not (compare Sec. 3.B).

5. COMPARISON OF COMPATIBILITY RELATIONS ON A VECTOR SPACE

One of the main reasons to consider partial inner product spaces is the possibility they offer to control very singular operators. For this purpose the question frequently arises as to whether a given compatibility relation is "fine" enough for a particular operator, or, on the contrary, too fine in that it leads to an unmanageable lattice of assaying subsets. What we need is obviously a way of comparing different compatibility relations on the same vector space.

Let V be a vector space, $\#_1$ and $\#_2$ linear compatibility relations on V. We will say that $\#_1$ is *weakly finer* than $\#_2$, say, if $\#_1$ has more compatible pairs; equivalently, $f \#_2 g$ implies $f \#_1 g$ for any pair $f, g \in V$, or $S^{\#_2} \subseteq S^{\#_1}$ for any subset $S \subseteq V$. This relation is a partial order on the set Comp(V) of all linear compatibility relations on V (see also Appendix A). Actually, Comp(V) is even a complete lattice for this order, since it is stable under arbitrary intersections and has a greatest element $\hat{\#}$. Here, intersection is defined as follows: $S^{\wedge_{i\in I}\#_i} = \cap_{i\in I} S^{\#_i}$ for an arbitrary subset $S \subseteq V$. The greatest element is the trivial compatibility $\hat{\#}$ for which every pair of vectors is a compatible pair:

$$f \hat{\#} g, \ \forall f, g \in V \text{ or } \{f\}^{\hat{\#}} = V, \ \forall f \in V .$$

The lattice $\mathcal{F}(V, \hat{\#})$ has only one element, namely V itself. Thus $\mathcal{F}(V, \#)$ is *not* a subset of $\mathcal{F}(V, \hat{\#})$, for any $\# \in$ Comp(V); more generally, a weakly finer compatibility does not lead to more assaying subsets. As a consequence, this ordering of Comp(V), although standard in the context of Galois connections (see Appendix A), is useless for our purposes.

Before passing to a stronger order, it is instructive to exhibit an example of two compatibilities, which are not weakly comparable.

5.1. *Example*: Let $\mathcal{E}'$ be Bargmann's space of entire holomorphic functions defined on the complex plane. Define the following compatibilities on $\mathcal{E}'$:

$$f \#_1 g \quad \text{iff} \quad \int_{\mathbb{C}} |f(z)\, g(z)|\, e^{-|z|^2}\, dz < \infty ,$$

$$f \#_1 g \quad \text{iff} \quad \sum_{n=0}^{\infty} n! \,|a_n b_n| < \infty ,$$

where $f(z) = \Sigma_{n=0}^{\infty} a_n z^n$ and $g(z) = \Sigma_{n=0}^{\infty} b_n z^n$. Let now $f(z) = \exp(\frac{1}{2} z^2)$ and $g_1(z) = \exp(\frac{i}{2} z^2)$. Then f and g_1 are $\#_1$-compatible, but they are not $\#_2$-compatible. On the other hand, if $g_2(z) = z\exp(\frac{1}{2} z^2)$, then f and g_2 are (trivially) $\#_2$-compatible and they are not $\#_1$-compatible. Hence $\#_1$ and $\#_2$ are not comparable in the weak sense.

As discussed above, a good order on Comp(V) must satisfy the two following conditions: If $\#_1$ is finer than $\#_2$, then $\#_2$-compatible vectors should be $\#_1$-compatible (i.e., finer implies weakly finer) and $\mathcal{F}(V, \#_2)$ should be a subset of $\mathcal{F}(V, \#_1)$. However, these requirements are not sufficient for a comparison of compatibilities. It is essential that the involution in $\mathcal{F}(V, \#_2)$ be the restriction of the involution in $\mathcal{F}(V, \#_1)$. This condition leads to the next definition.

5.2. *Definition*: We shall say that $\#_1$ is *finer* than $\#_2$, or that $\#_2$ is *coarser* than $\#_1$ ($\#_2 \leqslant \#_1$) if $\mathcal{F}(V, \#_2)$ is an involutive sublattice of $\mathcal{F}(V, \#_1)$.

A simple criterion for comparing two compatibilities on V is given by the following result.

5.3. *Proposition*: $\#_1$ is finer than $\#_2$ iff

$$S^{\#_2\#_2} = S^{\#_2\#_1} \tag{7}$$

for every subset $S \subseteq V$.

Proof: We know that $A \subseteq V$ belongs to $\mathcal{F}(V, \#_i)$ if and only if $A = B^{\#_i\#_i}$ for some $B \subseteq V$, or equivalently, $A = C^{\#_i}$ for some $C \subseteq V$. Thus Eq. (7) means that every element $S^{\#_2\#_2}$ of $\mathcal{F}(V, \#_2)$ belongs to $\mathcal{F}(V, \#_1)$. Furthermore, for any $D = S^{\#_2} \in \mathcal{F}(V, \#_2)$, one has $D^{\#_2} = D^{\#_1}$. This proves that $\mathcal{F}(V, \#_2)$ is an involutive sublattice of $\mathcal{F}(V, \#_1)$ if and only if Eq. (7) holds for every $S \subseteq V$. □

5.4. *Corollary*: If $\#_2 \leqslant \#_1$, then $S^{\#_2} \subseteq S^{\#_1}$ and $S^{\#_2\#_2} \supseteq S^{\#_1\#_1}$ for every subset $S \subseteq V$. In particular, $f \#_2 g$ implies $f \#_1 g$ for any $f, g \in V$, i.e., $\#_1$ is weakly finer than $\#_2$.

Proof: From Eq. (7), one gets, for any $S \subseteq V$

$$S^{\#_2} = S^{\#_2\#_2\#_2} = S^{\#_2\#_2\#_1} \subseteq S^{\#_1} ,$$

where the inclusion follows from $S \subseteq S^{\#_2\#_2}$. In particular for $S = \{f\}$, this gives $\{f\}^{\#_2} \subseteq \{f\}^{\#_1}$. Furthermore, using Eq. (7) again, we have

$$S^{\#_2\#_2} = S^{\#_2\#_1} \supseteq S^{\#_1\#_1} . \qquad \square$$

What about Comp(V)? First, the relation $\leqslant$ is a partial order on it, as can be checked immediately. Then Comp(V) is inductively ordered, i.e., every totally ordered subset $\{\#_n\}_{n\in J}$ has an upper bound $\#_\infty$, defined as follows:

$$f \#_\infty g \text{ iff there exists } n \in J \text{ such that } f \#_n g .$$

Equivalently, $S^{\#_\infty} = \cup_{n\in J} S^{\#_n}$, for any $S \subseteq V$. Therefore Zorn's lemma applies: Every element of Comp(V) is majorized by a maximal element. However, there is *no* greatest element. The only possible candidate would be $\hat{\#}$, for which $S^{\hat{\#}} = V$ for every $S \subseteq V$. But Eq. (7) implies that $\hat{\#}$ cannot be finer than any $\# \in$ Comp(V), except itself! In fact, Comp(V) is in general neither directed to the left, nor to the right, *a fortiori* it is not a lattice.

Let again $\#$ be a linear compatibility on V and let $\mathcal{J}$ be a *rich* subset of $\mathcal{F}(V, \#)$. As was remarked after Proposition 4.2, we may always assume that $\mathcal{J}$ is an involutive sublattice of $\mathcal{F}(V, \#)$, i.e., an involutive covering of V. Moreover, by Theorem 3.2, $\mathcal{F}(V, \#)$ is the lattice completion of $\mathcal{J}$.

What happens now if we start from a sublattice $\mathcal{J}$ cofinal to $\mathcal{F}(V, \#)$, but *not* rich? Again by Theorem 3.2, we can associate to $\mathcal{J}$ a new compatibility relation $\#_{\mathcal{J}}$, such that $\mathcal{J}$ generates the complete involutive lattice $\mathcal{F}(V, \#_{\mathcal{J}})$.

Then it is easy to see that $\mathcal{F}(V,\#_{\mathcal{I}})$ is actually a complete involutive sublattice of $\mathcal{F}(V,\#)$; in other words, $\#_{\mathcal{I}} < \#$. This result in fact gives a description of all compatibility relations on V coarser than a given one $\#$, namely:

5.5. *Theorem*: Let $V,\#$ be as usual

(a) Let $\mathcal{I}$ be a cofinal involutive sublattice of $\mathcal{F}(V,\#)$. Then the compatibility $\#_{\mathcal{I}}$ determined by $\mathcal{I}$ is coarser than $\#$: $\#_{\mathcal{I}} \leqslant \#$.

(b) Conversely, if $\#_1 \leqslant \#$, then there exists a sublattice $\mathcal{I} \subseteq \mathcal{F}(V,\#)$ cofinal with $\mathcal{F}$ and stable under the involution, such that $\#_{\mathcal{I}} = \#_1$. □

Thus, given $\#$, a compatibility relation on V, coarser than $\#$, is the same thing as a complete involutive sublattice of $\mathcal{F}(V,\#)$. The set of all of these is stable under intersection and it has a greatest element, namely $\mathcal{F}(V,\#)$, hence this set is itself a complete lattice contrary to Comp(V). It has also a smallest element. $\mathcal{F}(V,\#_0)$ consisting exactly of $V^{\#}$ and V, and corresponding to the trivial compatibility relation $\#_0$ defined as

$f \#_0 g$ iff at least one of them belongs to $V^{\#}$.

In a sense this trivial compatibility relation correponds, in our language, to the standard situation of the theory of distributions: Only two kinds of objects are available, test functions ($V^{\#}$) and distributions (V).

Let us give a few examples of comparable compatibility relations.

5.6. *Examples*: (i) We have already encountered non-rich sublattices in Sec. 4. Take for instance, the functions of polynomial growth of Example 4.C (i): The set $\mathcal{I} = \{L^2(r_\delta), r_\delta(x) = (1 + |x|^2)^{\delta/2}, \delta \in \mathbb{R}\}$ is not rich. The same example in the discrete case was given in 4.C (ii). In each case, $\mathcal{I}$ defines a new compatibility relation, coarser that the original one. Additional examples (for sequence spaces) can be found in Ref. 22.

(ii) As pointed out at the end of Sec. 4, no totally ordered subset will be rich for a compatibility which is defined by absolute convergence of a series or an integral. Bargmann's space $\mathcal{B}'$ illustrates this point beautifully. Three compatibilities arise naturally on $\mathcal{B}'$: $\#_1$ and $\#_2$, as defined in Example 5.1, and $\#_3$ defined by the chain $\{\mathcal{F}^\rho, \rho \in \mathbb{R}\}$ described in Sec. 3.B, example (5). Then it can be checked easily that $\#_3$ is strictly coarser that both $\#_1$ and $\#_2$: Neither the pair f, g_1 nor the pair f, g_2 are $\#_3$-compatible. Indeed, both f and g_1 belong to $\mathcal{F}^\rho$ iff $\rho < 0$, and $g_2 \in \mathcal{F}^\rho$ iff $\rho < -1$.

The main application of Theorem 5.5. is to the construction of partial inner product spaces. Given a vector space V, one can define a *partial inner product* on V directly.

5.7. *Definition*: *A partial inner product* on a vector space V is a Hermitian form defined on a domain $\Gamma \subseteq V \times V$, such that:

(i) Γ is symmetric: $\{f,g\} \in \Gamma$ iff $\{g,f\} \in \Gamma$.

(ii) Γ is "partially linear": For every $f \in V$, the set $\{g | \{f,g\} \in \Gamma\}$ is a vector subspace of V.

Typical examples are all those partial inner products whose domain of definition is defined by the absolute convergence of a series or an integral; such are, for instance, the "natural" inner products on ω, $L^1_{\text{loc}}(X,d\mu)$ or $\mathcal{B}'$ obtained by extension of the inner product of a Hilbert space.

According to this definition, a partial inner product uniquely defines a linear compatibility relation $\#$:

$f \# g$ iff $\{f,g\} \in \Gamma$.

However, quite often the complete lattice $\mathcal{F}(V,\#)$ generated by $\#$ is too large, and thus one is led, for practical purposes, to consider a coarser compatibility relation on V. The point of Theorem 5.5. is that, first, one knows all possible candidates, and second, each of them can be used as a domain for the initial partial inner product. In particular, if the latter is nondegenerate, i.e., $(V^{\#}) = \{0\}$, it will remain so, whichever coarser compatibility $\#_1$ one chooses, including the trivial one, $\#_0$, for which $V^{\#_0} = V^{\#_1} = V^{\#}$.

Theorem 5.5 solves the problem of coarsening a given compatibility relation on V. In practice, the converse problem will often arise, namely, how to refine a given compatibility. Here however, there is no canonical solution.

There is one case where a solution can be found, namely when the compatibility is given in terms of an involutive covering of V and there is an explicit finer covering. Typically, a partial inner product is introduced, which has a bigger "natural" domain. However, even in that case, uniqueness is not guaranteed. Once again, Bargmann's space gives a counterexample. If we start with the compatibility $\#_3$ defined by the involutive covering $\mathcal{F}^\rho, \rho \in \mathbb{R}\}$, and introduce, as in Example 5.1, the usual inner product

$$\langle f | g \rangle = \int \overline{f(z)}\, g(z) e^{-|z|^2/2}\, dz .$$

The latter is obviously defined whenever the integral converges absolutely: This leads to the compatibility $\#_1$, finer than $\#_3$. Now if $f \#_3 g$, then the inner product is also given by the expression

$$\langle f | g \rangle = \sum_{n=1}^{\infty} n!\, \overline{a_n}\, b_n .$$

which is defined whenever the series converges absolutely, leading to $\#_2 \geqslant \#_3$. However, we have seen that $\#_1$ and $\#_2$ are not comparable although they are both finer than $\#_3$.

Apart from that situation, very little can be said about the problem of refinement. As a first step, one might try to increase the number of compatible pairs. Equivalently, one can try to extend the domain of the linear forms $\varphi_f \equiv \langle f | \cdot \rangle$, initially defined on $\{f\}^{\#}$, and continuous in the Mackey topology $\tau(\{f\}^{\#}, \{f\}^{\#\#})$. This however requires topological considerations. For instance, if $\{f\}^{\#}$ is not Mackey complete, one can extend φ_f continuously to its completion. Or if φ_f is continuous for a weaker topology $\tau(V_r, V_{\bar{r}})$, with $\{f\}^{\#} \subseteq V_r$, φ_f can be extended to all of V_r. Also φ_f might be not continuous, but closable for such a weaker topology; then one can replace it by its closure

φ_f. However, explicit examples show that all these procedures will yield compatibilities that are not comparable with the original one, because the involution will be modified. We must conclude that the problem of refinement has in general no solution.

Yet there are cases in which partial results can be obtained, namely one can enlarge the set of *explicit* assaying subsets. For instance, suppose the compatibility on V is given in terms of an involutive family $\mathcal{F} = \{V_r\}$ of subspaces, which is *not* stable under intersection (thus not a chain). Assume in addition that every V_r, with its Mackey topology $\tau(V_r, V_{\bar{r}})$, is a reflexive Banach space (in particular, a Hilbert space). Then it can be shown[6] that every element of the lattice generated by $\mathcal{F}$ is again a reflexive Banach space, and is given explicitly as follows:

$$V_{p\wedge q} = V_p \cap V_q ,$$

$$\text{with norm } \|f\|_{p\wedge q} = \|f\|_p + \|f\|_q ,$$

$$V_{p\vee q} = V_p + V_q ,$$

$$\text{with norm } \|f\|_{p\vee q} = \inf_{f=g+h} (\|g\|_p + \|h\|_q) .$$

Here the infimum is taken over all decompositions $f = g + h$, with $g \in V_p$, $h \in V_q$. In this way, one obtains an explicit enlargement of $\mathcal{F}$ although not a finer compatibility as defined above. Another example is the explicit completion of a reflexive chain that was discussed at the end of Sec. 3.

Notice that, in both cases, essential use is made of topological properties of assaying subsets. These are determined entirely by the partial inner product, which defines the duality between pairs V_r, $V_{\bar{r}}$; the compatibility alone no longer suffices. So the next step in our analysis is to study systematically the topological structure of PIP-spaces: This will be done in the next paper of the series.[6]

ACKNOWLEDGMENTS

It is a pleasure to thank Alex Grossmann for innumerable discussions on the subject of this paper, and also F. Debacker-Mathot and M. Vause for some profitable conversations.

APPENDIX A. GALOIS CONNECTIONS

Let $\mathcal{M}$ be a partially ordered set. A *closure* on $\mathcal{M}$ is a map $X \to \bar{X}$ from $\mathcal{M}$ to $\mathcal{M}$ such that: (i) $X \subseteq \bar{X}$; (ii) $\bar{\bar{X}} = \bar{X}$; (iii) $X \subseteq Y$ implies $\bar{X} \subseteq \bar{Y}$.[23] An element $X \in \mathcal{M}$ is said to be *closed* iff $X = \bar{X}$. Let $\mathcal{C}(\mathcal{M})$ be the set of all closed elements of $\mathcal{M}$, with the induced order. Then *if* $\mathcal{M}$ is a complete lattice, so is $\mathcal{C}(\mathcal{M})$ with respect to the lattice operations ($\mathcal{N} \subseteq \mathcal{C}(\mathcal{M})$):

$$\bigwedge_{X\in\mathcal{N}} X \Big|_{\mathcal{C}(\mathcal{M})} = \bigwedge_{X\in\mathcal{N}} X \Big|_{\mathcal{M}} ,$$

$$\bigvee_{X\in\mathcal{N}} X \Big|_{\mathcal{C}(\mathcal{M})} = \overline{\left(\bigvee_{X\in\mathcal{N}} X \Big|_{\mathcal{M}}\right)} . \tag{A1}$$

Let now $\mathcal{L}$ and $\mathcal{M}$ be two partially ordered sets. A *Galois connection*[23,24] between $\mathcal{L}$ and $\mathcal{M}$ is a pair of maps $\alpha: \mathcal{L} \to \mathcal{M}$ (we write $\alpha(X) \equiv X^\alpha$) and $\beta: \mathcal{M} \to \mathcal{L}$, such that:

(i) both α and β reverse order,

(ii) $S \subseteq S^{\alpha\beta}$ for each $S \in \mathcal{L}$ and $T \subseteq T^{\beta\alpha}$ for each $T \in \mathcal{M}$.

It follows from the definition that $S \to S^{\alpha\beta}$ (resp. $T \to T^{\beta\alpha}$) is a closure on $\mathcal{L}$ (resp. $\mathcal{M}$).

From now on, we will assume that both $\mathcal{L}$ and $\mathcal{M}$ are complete lattices. So are then the two sets of closed elements $\mathcal{C}(\mathcal{L})$ and $\mathcal{C}(\mathcal{M})$. Furthermore, α (resp. β) is a lattice anti-isomorphism of $\mathcal{C}(\mathcal{L})$ onto $\mathcal{C}(\mathcal{M})$ [resp. $\mathcal{C}(\mathcal{M})$ onto $\mathcal{C}(\mathcal{L})$]; for every subset $\mathcal{N} \subseteq \mathcal{C}(\mathcal{L})$ one has

$$\left(\bigvee_{X\in\mathcal{N}} X\right)^\alpha = \bigwedge_{X\in\mathcal{N}} X^\alpha , \tag{A2a}$$

$$\left(\bigwedge_{X\in\mathcal{N}} X\right)^\alpha = \bigvee_{X\in\mathcal{N}} X^\alpha , \tag{A2b}$$

and similarly for β. Actually[25] in the case where both $\mathcal{L}$ and $\mathcal{M}$ are complete lattices, the two maps α and β are not independent: α generates a Galois connection iff it satisfies the single condition (A2a); β is then uniquely determined and given by:

$$T^\beta = \bigvee_{X\in\mathcal{N}} X .$$

Further insight into the structure of Galois connections can be found in a paper by Shmuely.[26] Two points are of interest for us:

(1) Galois correspondences between the complete lattices $\mathcal{L}$ and $\mathcal{M}$ are in 1-1 correspondence with certain subsets of $\mathcal{L} \times \mathcal{M}$, called G-ideals. Since these, with the natural order inherited $\mathcal{L} \times \mathcal{M}$, form a complete lattice, it follows that the set of all Galois maps $\alpha: \mathcal{L} \to \mathcal{M}$ that generate a Galois connection also form a complete lattice.

(2) Every Galois connection between $\mathcal{L}$ and $\mathcal{M}$ can be generated by a *binary relation*, that is a subset $\Gamma \subseteq \mathcal{L} \times \mathcal{M}$; for instance $\Gamma = \{(S,T) | T \leqslant S^\alpha\}$. Conversely, every binary relation $\Gamma \subseteq \mathcal{L} \times \mathcal{M}$ generates a Galois connection, namely the one that corresponds to the minimal G-ideal generated (by lattice operations) by $\Gamma, (0_{\mathcal{L}}, 1_{\mathcal{M}})$ and $(1_{\mathcal{L}}, 0_{\mathcal{M}})$ where 0, resp. 1, denotes the smallest, resp. largest, element of the lattice indicated.

Next we specialize the discussion to the case $\mathcal{L} = \mathcal{M} = \mathcal{P}(S)$, the complete lattice of all subsets of a given set S. If we assume furthermore that $\alpha = \beta$ (such an α is called an *involution*[24]), the resulting self-dual Galois connection on $\mathcal{P}(S)$ is exactly what was called *compatibility on* S in I. Indeed $\alpha = \beta$ is equivalent to the corresponding binary relation Γ being symmetric: $(X,Y) \in \Gamma$ iff $(Y,X) \in \Gamma$, which we can write, as usual, $X \# Y$ (with $\# \equiv \alpha = \beta$). The closed elements of $\mathcal{P}(S)$ are precisely the *assaying subsets*, which constitute the complete lattice $\mathcal{F}(S,\#)$. The map $\#$ of $\mathcal{F}(S,\#)$ onto itself is an involution and a lattice anti-isomorphism. Property (1) above means that the set Comp(S) of all compatibilities on S is in a 1-1 correspondence with the set of all symmetric G-ideals of $\mathcal{P}(S) \times \mathcal{P}(S)$ and the latter is a complete lattice with respect to the order inherited from $\mathcal{P}(S) \times \mathcal{P}(S)$. That order is exactly the notion of the weak comparability ("weakly finer", etc.) introduced in Sec. 5. Property (2) yields the notion of generating subset for a Galois connection: These are our *rich subsets*, discussed in Sec. 4.

Finally, we come back to the linear case. Let V be a vector space and $\#$ a linear compatibility on V. By the very definition, the relation $f \# g$ ($f, g \in V$) is equivalent to $[f] \# [g]$, where $[f]$ is the one-dimensional subspace generated by f. Thus we may take as complete lattice $\mathscr{L}(V)$, the set of all vector subspaces of V: A linear compatibility on V is the same thing as a self-dual (or involutive) Galois map on $\mathscr{L}(V)$. The whole discussion above then goes through.

APPENDIX B. NONLINEAR COMPATIBILITY RELATIONS

As stated in I and in Appendix A, compatibility relations may be defined naturally on arbitrary sets; they are just symmetric binary relations. In this Appendix we will exhibit some interesting examples of such general compatibilities, which illustrate some points of the paper.

Let S be an arbitrary set, and $\#$ a symmetric binary relation on S. The latter extends to an involutive Galois connection on $\mathscr{P}(S)$:

- if $T = \emptyset$, the empty set, we put $T^{\#} = S$,
- if $T = \{f\}$, $T^{\#} = \{g \in S \mid g \# f\}$,
- if T is arbitrary, $T^{\#} = \{g \in S \mid g \# f \; \forall f \in T\}$

$$= \bigcap_{f \in T} \{f\}^{\#}.$$

Thus the set $\mathscr{F}(S, \#)$ of all assaying subsets ($T = T^{\#\#}$), ordered by inclusion, is a complete lattice with respect to the operations $\wedge$ = set intersection and $\vee$ = $\#\#$-closure of set union, and $\#$ is an anti-isomorphism of $\mathscr{F}(S, \#)$.

Example B.1.: Equivalence relation. In general, $\#$ need not be reflexive, nor transitive. If it is, i.e., an equivalence relation, then the above construction gives the following: If $T \subseteq S$ is entirely contained in one equivalence class, then $T^{\#} = T^{\#\#}$ = equivalence class containing T. If T contains at least two nonequivalent elements, then $T^{\#} = \emptyset$ and $T^{\#\#} = S$. So, if S contains at least two inequivalent elements, the assaying subsets of S are $\emptyset$, S, and single equivalence classes. We see in particular that the union of two assaying subsets need not be one (i.e., the sup is not set union).

Example B.2: Polarity. Let S be $\mathbb{R}^n$ with the Euclidean inner product; write $\xi \# \eta$ iff $|(\xi, \eta)| \leqslant 1$. Then for any subset $T \subseteq S$, $T^{\#}$ is the absolute polar of T, and the assaying subsets are precisely all closed, absolutely convex subsets of $\mathbb{R}^n$.[10]

Example B.2a. As a variant of B.2, define a family of compatibilities for every $p > 0$ as follows: $\xi \#_p \eta$ iff $|(\xi, \eta)| \leqslant p$. Then for $p < q$, $\#_q$ has more compatible pairs than $\#_p$ and $A^{\#_p} \subsetneq A^{\#_q}$ for every $A \subseteq \mathbb{R}^n$. On the other hand, the assaying subsets are the same for every $p > 0$, namely all closed, absolutely convex subsets of $\mathbb{R}^n$; hence $\mathscr{F}(\#_p) = \mathscr{F}(\#_q)$ as sets and $A^{\#_p \#_p} = A^{\#_q \#_q}$, the closed, absolutely convex cover of A. In other words, $\#_p$ is weakly coarser, but not coarser than (indeed, not comparable to) $\#_q$ if $p < q$.

Example B.3: Vicinity. Let S be an abelian topological group and $N \subseteq S$ a balanced neighborhood of the identity, i.e., $-N = N$. Define $f \# g$ iff $f - g \in N$. Take for instance $S = \mathbb{R}^n$, N the open unit sphere. Then any open sphere S_R of radius $R \leqslant 1$ with arbitrary center, as well as any closed sphere $\bar{S}_R$ of radius $R < 1$, is assaying, and we have $(S_R)^{\#} = \bar{S}_{1-R}$, $(\bar{S}_R)^{\#} = S_{1-R}$. More generally, if T is an arbitrary subset of diameter $d \geqslant 2$, then $T^{\#} = \emptyset$. Otherwise $T^{\#}$ is a small connected set, $T^{\#} = \cap_{f \in T} S_1[f]$, where $S_1[f]$ is the open unit sphere centered at f.

Example B.3a. In the same setup, take instead $\bar{N}$, the closed unit sphere. Then only spheres $\bar{S}_R$ ($R \leqslant 1$) are assaying and $(\bar{S}_R)^{\#} = \bar{S}_{1-R}$.

Example B.3b. On the other hand, if N is taken to be a subgroup (e.g., $\mathbb{R}^k$, $k < n$) we return to a equivalence relation as discussed in Example B.1 above.

[1] J.-P. Antoine and A. Grossmann, "Partial inner product spaces I. General properties," J. Funct. Anal. **23**, 369–378 (1976).
[2] I.M. Gel'fand and N.Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. IV.
[3] Yu.M. Berezanskii, *Expansions in Eigenfunctions of Self-Adjoint Operators* (Amer. Math. Soc., Providence, RI, 1968).
[4] E. Prugovečki, "The bra and ket formalism in extended Hilbert space," J. Math. Phys. **14**, 1410–1422 (1973).
[5] A. Grossmann, "Elementary properties of nested Hilbert spaces," Commun. Math. Phys. **2**, 1–30 (1966).
[6] J.-P. Antoine, "Partial inner product spaces. IV. Topological considerations" (in preparation).
[7] A. Grossmann, R. Hoegh-Krohn, and M. Mebkhout, "A class of explicitly soluble, local, many center, Hamiltonians for one-particle Quantum Mechanics in two and three dimensions. I," J. Math. Phys. (to be published).
[8] R. Palais, *Seminar on the Atiyah-Singer Index Theorem*, Chap. VII Annals of Math. Studies, #57 (Princeton University, Princeton, N.J., 1965).
[9] S.G. Krein and Ju. I. Petunin, "Scales of Banach spaces," Usp. Mat. Nauk **21**, 89–168 (1966); Engl. Transl.: Russian Math. Surveys **21**, 85–159 (1966). This review paper contains references to earlier work.
[10] G. Köthe, *Topological Vector Spaces I* (Springer–Verlag, Berlin, Heidelberg, New York, 1969).
[11] H.W. Davis, F.J. Murray, and J.K. Weber, "Families of L_p-spaces with inductive and projective topologies," Pacific J. Math. **34** 619–638 (1970); "Inductive and projective limits of L_p spaces," Portug. Math. **31**, 21–29 (1972).
[12] K. Floret and J. Wloka, *Einführung in die Theorie der lokalkonvexen Räume*, Lect. Notes in Math. #56 (Springer–Verlag, Berlin, Heidelberg, New York 1968) (esp. §9).
[13] R. Arens, "The space L^{ω} and convex topological rings," Bull. Amer. Math. Soc. **52**, 931–935 (1946).
[14] Nonstandard analysis could also be used here: V_{p+} is really $V_{p+\epsilon}$, ϵ infinitesimal.
[15] *Warning*: The order here is direct: $p \to V_p$ is monotone *increasing*, whereas it was *decreasing* in the Lebesgue case (inverse order).
[16] L. Hörmander, *Linear Partial Differential Equation* (Springer–Verlag, Berlin, Heidelberg, New York, 1963).
[17] F. Trêves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York and London, 1967).
[18] V. Bargmann, "On a Hilbert space of analytic functions and an associated integral transform. Part II. A family of related functions spaces; application to distribution theory," Commun. Pure Appl. Math. **20**, 1–101 (1967).
[19] A. Inoue, "L^p-spaces and maximal unbounded Hilbert algebras," J. Math. Soc. Japan, **30**, 667–686 (1978).
[20] N. Bourbaki, *Intégration* (Hermann, Paris 1965–1967), Chaps. 3 and 5.
[21] See, e.g., I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York and London, 1965), (cf. 0.224).
[22] F. Debacker-Mathot, "Some operator algebras in nested Hilbert spaces," Commun. Math. Phys. **42**, 183–193 (1975).

[23]G. Birkhoff, *Lattice Theory* (AMS Colloquium Publ., Providence, R.I., 1967), 3rd ed.
[24]O. Ore, "Galois connexions," Trans. Amer. Math. Soc. **55**, 493–515 (1944).
[25]G. Pickert, "Bemerkungen über Galois-Verbindungen," Archiv Math. **3**, 385–389 (1952).
[26]Z. Shmuely, "The structure of Galois connections," Pacific J. Math. **54**, 209–225 (1974).

A modified asymptotic Padé method. Application to multipole approximation for the plasma dispersion function Z

Pablo Martín and Guillermo Donoso
Departamento de Física, Universidad Simón Bolivar, Apartado 80659, Caracas 108, Venezuela

Jorge Zamudio-Cristi
Universidad Nacional Abierta, Apartado 8226, Caracas 101, Venezuela

(Received 18 July 1979; accepted for publication 18 September 1979)

An improved asymptotic Padé method is presented. The approximations expressed as rational functions are developed both in a power series and asymptotic expansions. Identifying these developments with those for the exact function, the approximated rational functions are obtained. Three- and four-pole approximations for the plasma dispersion function have been determined with this method. Our approximations (Z_{lm}) give closer results to the exact function than all the published ones.

I. INTRODUCTION

In a previous work[1] a two-pole approximation for the plasma dispersion function Z (defined as a Hilbert transform of the Gaussian[2,3]) has been obtained. In that work we have modified the Padé method, in order to include, besides the power series, one term of the asymptotic expansion. In this paper we generalize the method in order to include several terms of the asymptotic expansion besides the power series. There are of course several possibilities of approximation depending on how many terms of the power series and asymptotic expansion are taken into account.

The method here is used to approximate the function Z, but the procedure is quite general and it can be used in other problems to improve the Padé method[4] in the cases where the asymptotic expansions are known.

Though the Martín and González two-pole approximation[1] for the Z function is better than that of Fried *et al.*,[3] however for $s \geqslant 2$, the accuracy in the imaginary part of $Z(s)$ and $Z'(s)$ is not good. Since the imaginary part of $Z'(s)$ becomes important to compute the damping or growth of plasma waves, it appears very convenient to have better approximations to the Z function that are still easy to calculate. For this reason we have applied the general procedure to three- and four-pole approximations. As in the two-pole case, a straightforward Padé method gives only good results for very small values of the arguments. The reason is that in the usual Padé approximation only the power series is used.

Our best results are obtained by using five terms of the power series and three terms of the asymptotic expansion. This gives excellent agreement for all values of s in the upper half-plane. We think that this approximation can now be used for most of the computations of $Z(s)$ and $Z'(s)$. Three poles can also be used with a little less accuracy. Similarly to the two-pole case,[1] the multipole approximations still preserve the symmetry property of $Z(s)$ and all the parameters are defined in an exact way, avoiding the need of visual inspection of computed displays as in Ref. 3.

The material is arranged in the following way. The general procedure for obtaining the poles of the approximation is described in Sec. II. The results are presented in a graphical form in Sec. III. An analysis and discussion of the various approximations are also carried out. The last section is devoted to the conclusions.

II. GENERAL PROCEDURE

We look for rational functions to approximate a given function. The approximations are obtained by identifying the terms of the power series and asymptotic expansion of the exact and approximated functions. Considering the plasma dispersion function $Z(s)$, the rational function is written in the form

$$Z(s) \simeq Z_{ap}(s) = \sum_{i=0}^{n-1} p_i s^i \Big/ \left(1 + \sum_{j=1}^{n} q_j s^j\right). \tag{1}$$

Here the polynomial numerator is chosen one degree less than the polynomial denominator in order to get the required asymptotic behavior to $O(1/s)$. Once the parameters of the approximation are found an expansion in partial fractions

$$Z_{ap}(s) = \sum_{k=1}^{n} \frac{b_k}{s - a_k} \tag{2}$$

is performed, showing explicitly that this function is an n-pole approximation.

To obtain the values $p_0, p_1, \ldots, p_{n-1}$ and $q_1, q_2, \ldots, q_n$, the power series and asymptotic expansion for $Z_{ap}(s)$ is written as

$$Z_{ap} = p_0 + (p_1 - p_0 q_1)s + (p_2 - p_0 q_2 - p_1 q_1 + p_0 q_1^2)s^2 + \cdots, \tag{3}$$

$$Z_{ap} = \frac{p_{n-1}}{q_n}\frac{1}{s} + \frac{p_{n-1}}{q_n}\left(\frac{p_{n-2}}{p_{n-1}} - \frac{q_{n-1}}{q_n}\right)\frac{1}{s^2} + \frac{p_{n-1}}{q_n}\left(\frac{p_{n-3}}{p_{n-1}} - \frac{q_{n-2}}{q_n} - \frac{p_{n-2}\,q_{n-1}}{p_{n-1}\,q_n} + \frac{q_{n-1}^2}{q_n^2}\right)\frac{1}{s^3} + \cdots. \tag{4}$$

Now the coefficients of the first l-terms of series (3) are

 0022-2488/80/020280-06$1.00

TABLE I. p_i parameters of the polynomial numerator and q_j parameters of the polynomial denominator of $Z_{lm}(s)$.

	$n=2$	
Z_{31}	$p_0 = i\sqrt{\pi}$	$q_1 = -i\dfrac{\sqrt{\pi}}{\pi-2}$
	$p_1 = \dfrac{4-\pi}{\pi-2}$	$q_2 = -\dfrac{4-\pi}{\pi-2}$
Z_{22}	$p_0 = i\sqrt{\pi}$	$q_1 = -i\sqrt{\pi}$
	$p_1 = \pi-2$	$q_2 = -(\pi-2)$
Z_{13}	$p_0 = i\sqrt{\pi}$	$q_1 = -i\sqrt{\pi}$
	$p_1 = 2$	$q_2 = -2$
	$n=3$	
Z_{51}	$p_0 = i\sqrt{\pi}$	$q_1 = \dfrac{-3\pi+10}{2(5\pi-16)}i\sqrt{\pi}$
	$p_1 = \dfrac{3\pi^2-30\pi+64}{2(5\pi-16)}$	$q_2 = \dfrac{21\pi-64}{6(5\pi-16)}$
	$p_2 = \dfrac{9\pi-28}{6(5\pi-16)}i\sqrt{\pi}$	$q_3 = -\dfrac{9\pi-28}{6(5\pi-16)}i\sqrt{\pi}$
Z_{42}	$p_0 = i\sqrt{\pi}$	$q_1 = -\dfrac{3\pi-8}{3\sqrt{\pi}(\pi-3)}i$
	$p_1 = \dfrac{10-3\pi}{3(\pi-3)}$	$q_2 = -\dfrac{10-3\pi}{3(\pi-3)}$
	$p_2 = -\dfrac{16-5\pi}{3\sqrt{\pi}(\pi-3)}$	$q_3 = \dfrac{(16-5\pi)i}{3\sqrt{\pi}(\pi-3)}$
Z_{33}	$p_0 = i\sqrt{\pi}$	$q_1 = -\dfrac{1}{4-\pi}i\sqrt{\pi}$
	$p_1 = \dfrac{3\pi-8}{4-\pi}$	$q_2 = -\dfrac{3\pi-8}{4-\pi}$
	$p_2 = \dfrac{-2(\pi-3)}{4-\pi}i\sqrt{\pi}$	$q_3 = \dfrac{2(\pi-3)}{4-\pi}i\sqrt{\pi}$
	$n=4$	
	$p_0 = i\sqrt{\pi}$	$q_1 = -\dfrac{9\pi-28}{2(6\pi^2-29\pi+32)}i\sqrt{\pi}$
	$p_1 = \dfrac{-15\pi^2+88\pi-128}{2(6\pi^2-29\pi+32)}$	$q_2 = -\dfrac{-36\pi^2+195\pi-256}{6(6\pi^2-29\pi+32)}$
Z_{53}		
	$p_2 = -\dfrac{-33\pi+104}{6(6\pi^2-29\pi+32)}i\sqrt{\pi}$	$q_3 = \dfrac{-33\pi+104}{6(6\pi^2-29\pi+32)}i\sqrt{\pi}$
	$p_3 = -\dfrac{9\pi^2-69\pi+128}{3(6\pi^2-29\pi+32)}$	$q_4 = \dfrac{9\pi^2-69\pi+128}{3(6\pi^2-29\pi+32)}$

identified with the first l-terms of the exact power series of $Z(s)$:

$$Z(s) = i\sqrt{\pi} - 2s - i\sqrt{\pi}\,s^2 + \tfrac{4}{3}s^3 + \frac{i\sqrt{\pi}}{2}s^4 - \tfrac{8}{15}s^5 + \cdots. \tag{5}$$

Similarly the coefficients of the first m-terms of expansion (4) are identified with the first m-terms of the asymptotic expansion for $Z(s)$:

$$Z(s) = i\sqrt{\pi}\,\sigma e^{-s^2} - \frac{1}{s} - \frac{1}{2s^3} - \frac{3}{4s^5} - \frac{15}{8s^7} - \cdots. \tag{6}$$

Here the exponential terms $i(\pi)^{1/2}\sigma\ \exp(-s^2)$ is of higher order for real s ($\sigma = 1$), and zero in the upper half-plane ($\sigma = 0$). Therefore it is not considered here.

In the lower half-plane ($\sigma = 2$) this term becomes important. However, the multipole approximations cannot be used directly, all the poles being in the lower half-plane. In this case, $Z(s)$ is obtained computing Z^* for s^* and then adding $2i(\pi)^{1/2}\exp(-s^2)$ as in Ref. 3.

In an n-pole approximation we have $2n$ unknowns and we need $2n$ equations. Therefore it will be necessary that

$$2n = l + m. \tag{7}$$

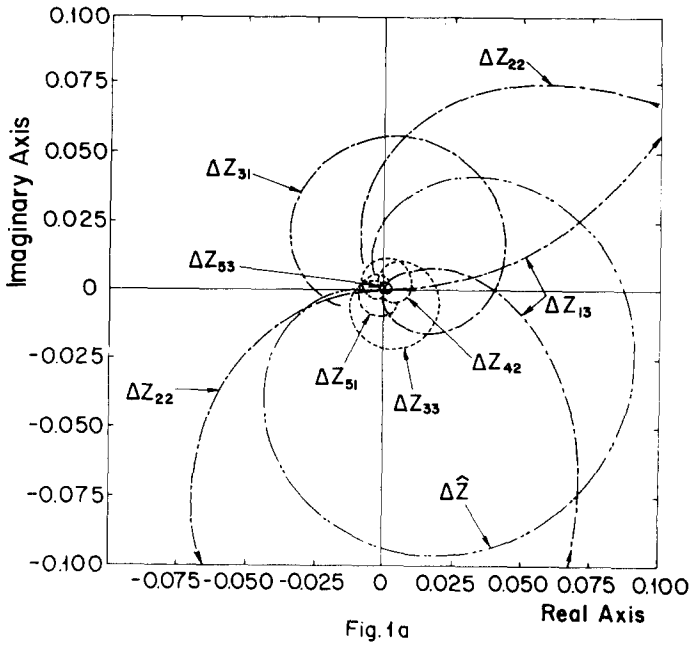
The approximation $Z_{ap}(s)$ for a given l and m will be denoted by $Z_{lm}(s)$. It is clear that there are several possibilities in order to verify Eq. (7) for the same n.

For a given n, if a good approximation is needed for small values of s, l should be large. For large values of s, l should be small. To obtain a good approximation for all values of s, l and m should be of about the same value. The analysis in the next section for the two- and three-pole cases shows that l slightly larger than m give the best results.

The set of equations for the p's and q's are nonlinear but in all cases we have examined all the p's and q's give unique values.

TABLE II. a_k poles and b_k pole-residues of $Z_{lm}(s)$.

$n=2$		
Z_{31}	$a_1 = -0.5138 - 1.0324i$	$b_1 = -0.5 + 1.2891i$
	$a_2 = 0.5138 - 1.0324i$	$b_2 = -0.5 - 1.2891i$
Z_{22}	$a_1 = -0.5228 - 0.7763i$	$b_1 = -0.5 + 0.7425i$
	$a_2 = 0.5228 - 0.7763i$	$b_2 = -0.5 - 0.7425i$
Z_{13}	$a_1 = -0.5510 - 0.4431i$	$b_1 = -0.5 + 0.4021i$
	$a_2 = 0.5510 - 0.4431i$	$b_2 = -0.5 - 0.4021i$
$n=3$		
Z_{51}	$a_1 = -0.8957 - 1.3245i$	$b_1 = 1.2401 + 0.8399i$
	$a_2 = -1.4096i$	$b_2 = -3.4802$
	$a_3 = 0.8957 - 1.3245i$	$b_3 = 1.2401 - 0.8399i$
Z_{42}	$a_1 = -0.9050 - 1.1317i$	$b_1 = 0.5809 + 0.7400i$
	$a_2 = -1.2278i$	$b_2 = -2.1618$
	$a_3 = 0.9050 - 1.1317i$	$b_3 = 0.5809 - 0.7400i$
Z_{33}	$a_1 = -0.9217 - 0.9091i$	$b_1 = 0.1822 + 0.5756i$
	$a_2 = -1.0204i$	$b_2 = -1.3643$
	$a_3 = 0.9217 - 0.9091i$	$b_3 = 0.1822 - 0.5756i$
$n=4$		
Z_{53}	$a_1 = -1.2359 - 1.2150i$	$b_1 = 0.5468 - 0.0372i$
	$a_2 = -0.3786 - 1.3509i$	$b_2 = -1.0468 + 2.1018i$
	$a_3 = 0.3786 - 1.3509i$	$b_3 = -1.0468 - 2.1018i$
	$a_4 = 1.2359 - 1.2150i$	$b_4 = 0.5468 + 0.0372i$



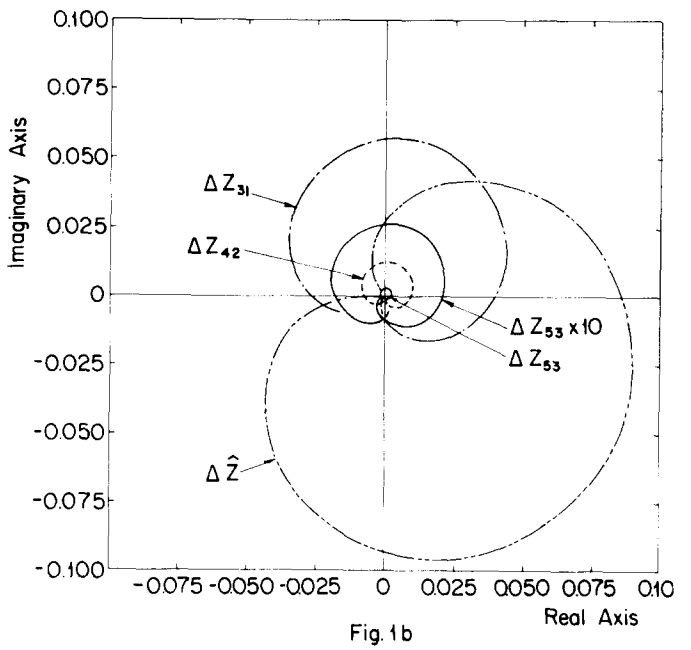


FIG. 1. Map in the complex ($\Delta Z_{lm} = Z_{lm} - Z$) plane of the real s axis, $0 \leqslant s \leqslant 4$ under $s \rightarrow \Delta Z_{lm}(s)$. (a) Curves for ΔZ_{13}, $\Delta \hat{Z}$, ΔZ_{22}, ΔZ_{31}, ΔZ_{33}, ΔZ_{42}, ΔZ_{51}, and ΔZ_{53}. (b) Curves for $\Delta \hat{Z}$, ΔZ_{31}, ΔZ_{42}, ΔZ_{53}, and $10 \times \Delta Z_{53}$.

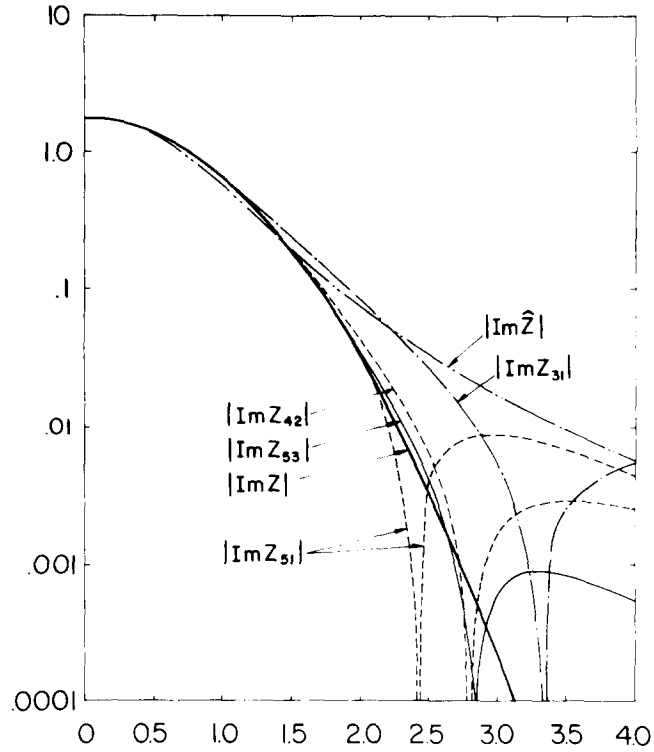


FIG. 2. Imaginary parts of $Z(s)$ in logarithmic scale for $\hat{Z}$, Z_{31}, Z_{42}, Z_{51}, Z_{53}, and Z (exact).

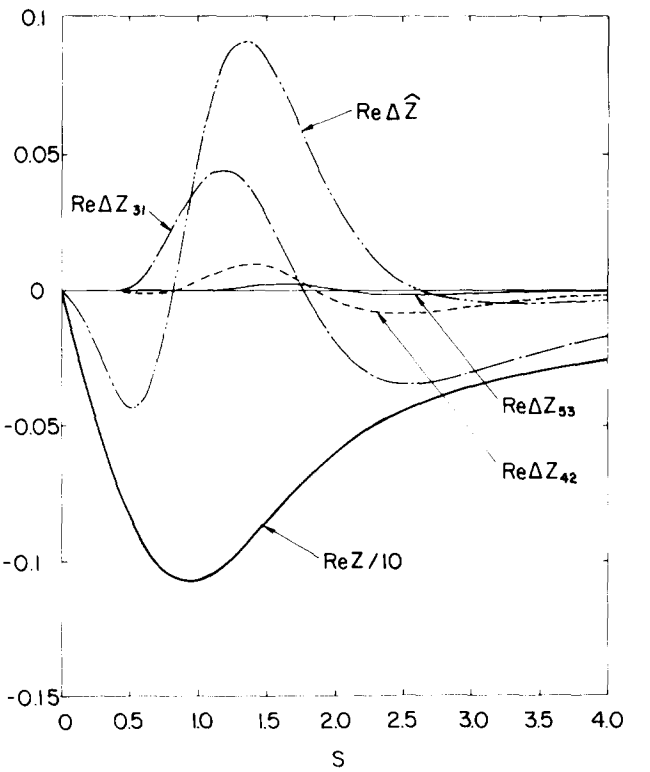


FIG. 3. Real parts of $\Delta Z_{lm}(s)$ for $\Delta \hat{Z}$, ΔZ_{31}, ΔZ_{42}, and ΔZ_{53}. Real part of $Z/10$ is also plotted for reference.

Following the general procedure previously described we get the rational approximations whose coefficients are given in Tables I and II.

In the two-pole case, the first function $Z_{31}(s)$ is the same discussed in a previous paper under the notation $\tilde{Z}(s)$.[1]

We point out that the symmetry property

$$Z_{lm}(s) = -[Z_{lm}(-s^*)]^* \tag{8}$$

is automatically verified by all the approximations here considered. This is because all the p_i's for i even, and all the

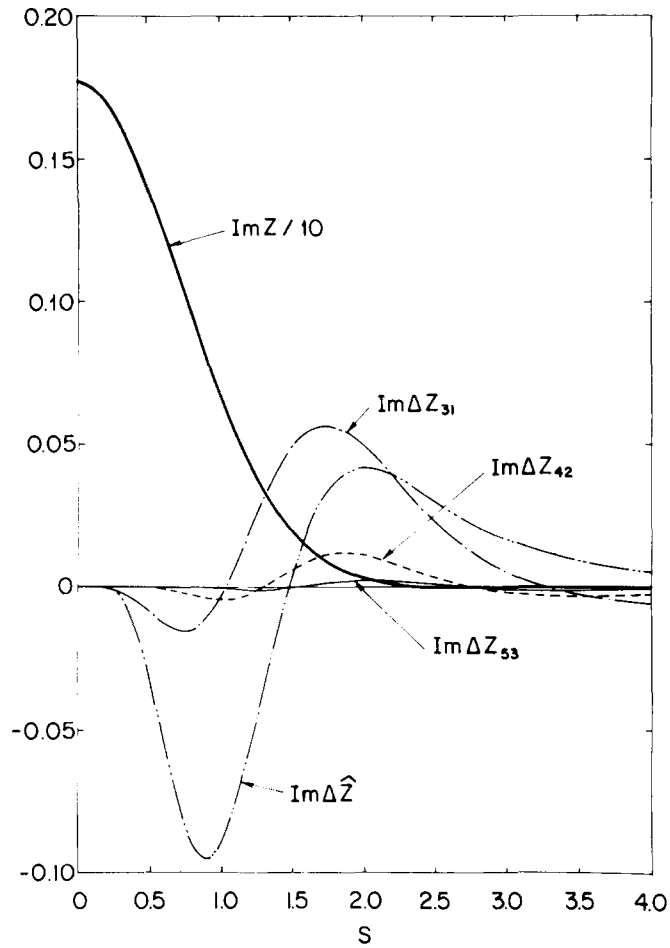


FIG. 4. Imaginary parts of $\Delta Z_{lm}(s)$ for $\Delta\widehat{Z}$, ΔZ_{31}, ΔZ_{42}, and ΔZ_{53}. Imaginary parts of $Z/10$ is also plotted as reference.

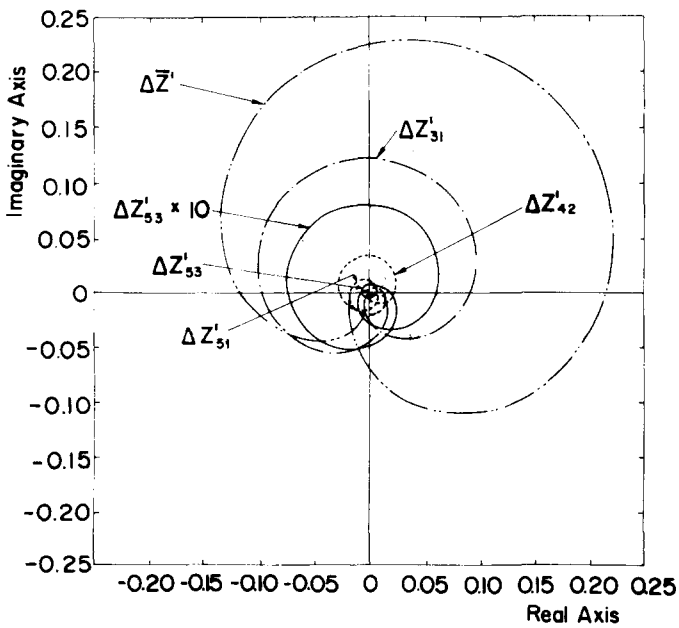


FIG. 5. Map in the complex $\Delta Z'_{lm} = Z'_{lm} - Z'$ plane of the real s axis $0 \leqslant s \leqslant 4$ under $s \to \Delta Z'_{lm}(s)$ for $\Delta\bar{Z}'$, $\Delta Z'_{31}$, $\Delta Z'_{42}$, $\Delta Z'_{51}$, $\Delta Z'_{53}$, and $10 \times \Delta Z'_{53}$.

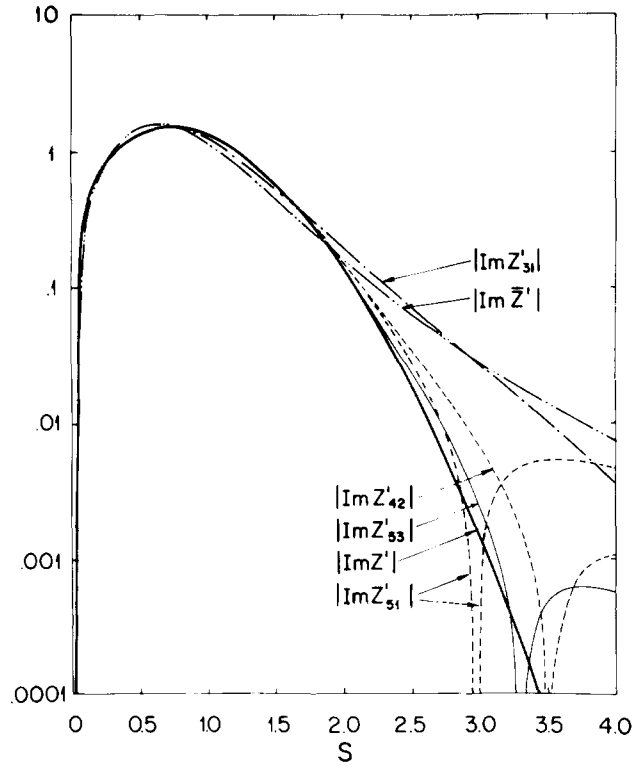


FIG. 6. Imaginary parts of $Z'_{lm}(s)$ in logarithmic scale for $\bar{Z}'$, Z'_{31}, Z'_{42}, Z'_{53}, and Z' (exact).

q_j's for j odd, result in pure imaginary and the others real. Equivalently all the poles a_k appear to be symmetric with respect to the imaginary axis and the coefficients b_k are symmetric with respect to the real axis.

III. ANALYSIS AND DISCUSSION

The map of the real axis for $0 \leqslant s \leqslant 4$ under $[Z_{lm}(s) - Z(s)]$ is shown in Fig. 1. First of all, it is clear from Fig. 1(a) that increasing the number of poles results in a better approximation. For two poles, the best approximation is $Z_{31}(s)$ as it was previously shown.[1] However, the Fried *et al.* calculations $\widehat{Z}(s)$ give better results than Z_{13} and Z_{22}. For three poles, Z_{42} gives similar results to Z_{51} and better results than Z_{33}. However, $\mathrm{Im}Z_{42}$ fits to $\mathrm{Im}Z$ better than $\mathrm{Im}Z_{51}$ (see Fig. 2). The locus $Z'_{51} - Z'$ is closer to the origin than the locus of Z'_{42} (see Fig. 5), but $\mathrm{Im}Z'_{42}$ fits better than $\mathrm{Im}Z'_{51}$ (see Fig. 6). Since small discrepancies in the imaginary part of Z' can become important in the calculation of the damping of the waves, we have considered Z'_{42} as a better approximation than Z_{51}.

In Fig. 1(b) we have replotted the best approximation for a given number of poles. The locus $10(Z_{53} - Z)$ is also drawn, and still lies inside the locus $Z_{31} - Z$.

For all $Z_{lm}(s)$ the agreement becomes better as s moves away from the real axis in the upper half-plane. This results from the fact that all the poles of Z_{lm} are in the lower half-plane, as was the case for the two-pole approximations previously calculated.[1,3]

For this reason all the graphs that follow will be done for real values of s in the range $0 \leqslant s \leqslant 4$.

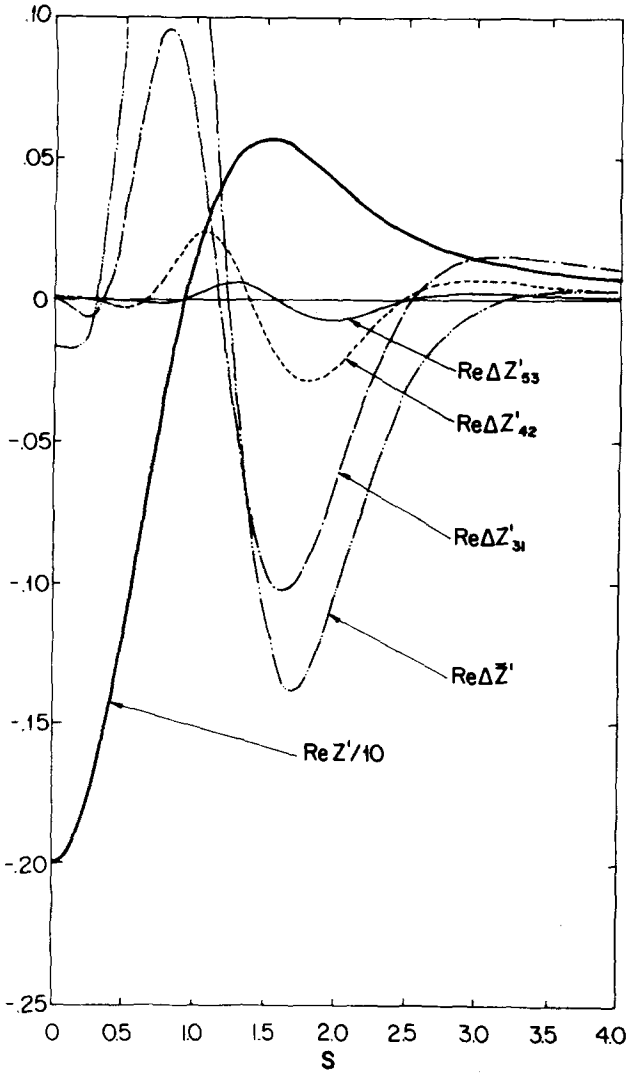


FIG. 7. Real parts of $\Delta Z'_{lm}(s)$ for $\Delta\bar{Z}'$, $\Delta Z'_{42}$, and $\Delta Z'_{53}$. Real part of $Z'/10$ is also plotted for reference.

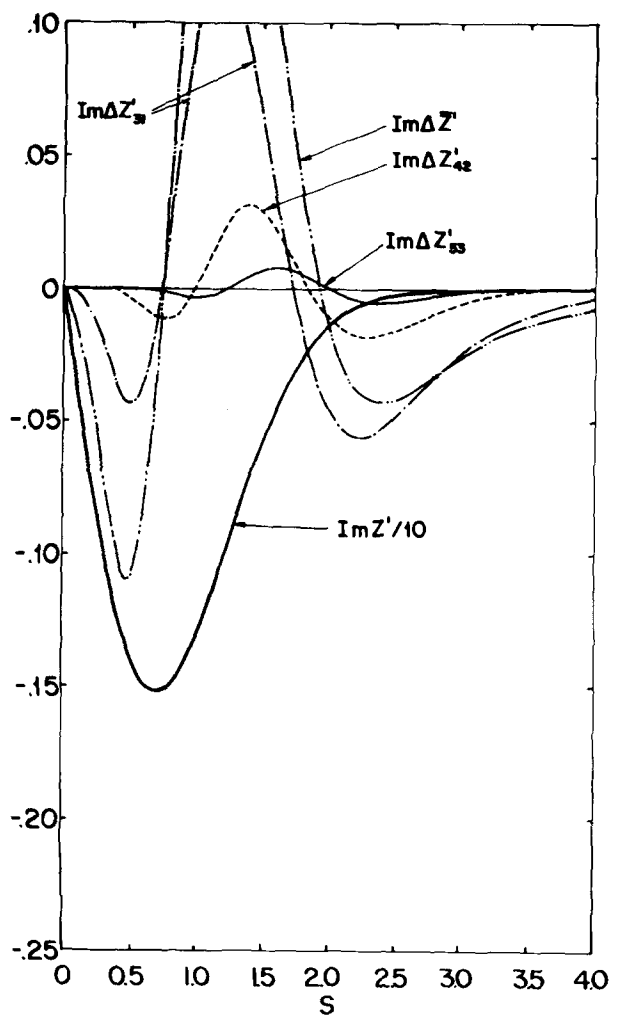


FIG. 8. Imaginary parts of $\Delta Z'_{lm}(s)$ for $\Delta\bar{Z}'$, $\Delta Z'_{31}$, $\Delta Z'_{42}$, and $\Delta Z'_{42}$, and $\Delta Z'_{53}$. Imaginary Part of $Z'/10$ is also plotted for reference.

Figure 2 shows $|\mathrm{Im}Z|$ for $\hat{Z}$ (Ref. 3), Z (exact value), Z_{31}, Z_{42}, Z_{53}, and Z_{51} in a logarithmic scale. Figures 3 and 4 show the real and imaginary parts of the $Z_{lm} - Z$ for each number of poles. Figures 5–8 are the corresponding plots for Z'.

The agreement in the real part of Z_{lm} is quite good, and for Z_{53} the relative error becomes less than 0.5%. For $\mathrm{Im}Z_{lm}$ the agreement is good for small s ($0 < s \lesssim 2.5$) and it is not so good for large s, mainly because $\mathrm{Im}Z$ goes to zero very fast. In all approximations $\mathrm{Im}Z_{lm}$ goes through zero, reaches a small minimum, and goes asymptotically to zero (Figs. 2 and 4). $\mathrm{Im}Z$ never becomes negative in the range where Z_{lm} is negative; but it is very small, and for $\mathrm{Im}Z_{53}$ the absolute error here is less than 0.001 (Fig. 2).

The derivative functions Z'_{lm} do not fit as well as the Z_{lm} functions. This is reflected in the change of scale for Fig. 5 with respect to Fig. 1. The relative error of the real part of Z'_{53} is less than 2%, compared with 0.5% for $\mathrm{Re}Z_{53}$ (Fig. 7). For the imaginary part of Z' the agreement is not so good as for Z (Fig. 8), though for large s, the maximum absolute error is about the same (0.001) (Fig. 6).

IV. CONCLUSION

The usual Padé method has been improved by considering the asymptotic expansion together with the power series in order to obtain better approximations to a given function. Using this method we have obtained multipole approximations for the plasma dispersion function.

In this way we get good approximation to $Z(s)$ not only for small values of s, but also for large values of s.

The approximations agree well for real values of s and in the upper half-plane. All the poles of these approximations are in the lower half-plane. The symmetry condition is also automatically verified. Better agreement is obtained in general when more poles are considered. A good balance between the power and asymptotic series is obtained by using two more terms of the power series than the asymptotic expansion. In this way we obtain the best fit for the whole range of s. If we want a better agreement for a small (large) value of s, more (less) terms of the power series must be used. Our four-pole approximation here denoted by $Z_{53}(s)$ gives a relative error for $\mathrm{Re}Z$ and $\mathrm{Re}Z'$ less than 0.5 and 2%, respective-

ly. The imaginary parts do not agree so well, mainly for $s \gtrsim 2.5$. Nevertheless the absolute errors for $\mathrm{Im}Z$ and $\mathrm{Im}Z'$ are less than 0.001 in this region.

Approximations with three poles are also obtained which fit better than all the published ones, but with less accuracy than Z_{53}.

ACKNOWLEDGMENT

This work was supported in part by the Venezuelan Government Agency CONICIT, Project No. 31.26.51.0606.

[1]P. Martín and M. A. González, Phys. Fluids **22**, 1413 (1979).

[2]B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).

[3]B. D. Fried, C. L. Hedrick, and J. McCune, Phys. Fluids **11**, 249 (1968).

[4]George A. Baker, Jr., *Essentials of Padé Approximants* (Academic, New York, 1975).

Dynamical symmetry breaking in the Lewis–Riesenfeld oscillator[a)]

Neil J. Günther
Department of Physics, The University of Southampton, England, U.K. S09 5NH

(Received 8 May 1978; revised manuscript received 18 October 1978)

Dynamical realization is given to the generator responsible for breaking the usual SU(3) invariance symmetry in the case of the Lewis–Riesenfeld time-dependent oscillator: $H(t) = 1/2\,\Sigma_i^3[p_ip_i + w^2(t)q_iq_i]$. This is achieved by finding a simple expression for the time-dependent Hamiltonian in the interaction picture: $H = H_0 + gH_1$. The breaking interaction is shown to transform as $(\mathbf{6}\oplus\bar{\mathbf{6}})$ of SU(3). Time-dependent dilations relate the broken Hamiltonian to the Lewis–Riesenfeld form. This scaling generates the algebra of the Weyl group and its role in oscillator noninvariance symmetries is considered.

1. INTRODUCTION

Renewed interest in the role of noninvariance symmetries for the harmonic oscillator has been independently generated in a number of recent articles.[1–4] Such groups permit a complete characteriztion of the dynamics in terms of the associated group representations. The noninvariance symmetry includes the usual invariance symmetry of the oscillator Hamiltonian as a subgroup in such a way that the symmetric tensor representations of the larger group contain the corresponding oscillator energy levels[5–7] (finite- and infinite-dimensional). The larger group is not a true symmetry of the total Hamiltonian in the sense that the extra generators do not commute with the Hamiltonian.

In particular our investigations of the symmetries of the Lewis–Riesenfeld time-dependent oscillator[8,9] have shown that the usual SU(3) dynamical symmetry, associated with the time-independent oscillator in three dimensions, constitutes a noninvariance symmetry of the Lewis–Riesenfeld system, while the true invariance symmetry is SO(3). The introduction of an arbitrary time-dependent frequency $\omega(t)$ into the oscillator Hamiltonian has the effect of reducing the invariance symmetry in a very particular way, and it is our purpose in this paper to give a dynamical description of this symmetry breaking.

The form of the operators which are responsible for breaking the SU(3) invariance are perhaps not immediately obvious because of the "sourceless" form of the $3H_{\rm LR}$-Hamiltonian, i.e., the arbitary function of time $\omega(t)$ represents the result of some time-dependent dynamical effects, e.g., contact with a thermal reservoir or electromagnetic interactions. This Hamiltonian is suggestive of so-called "open systems."[10] However, as we shall show, the symmetry of the breaking generator is independent of the arbitrary function of time. This is to be expected since the symmetry algebra is constructed only from canonically conjugate variables.

Symmetry breaking effects are most conveniently expressed in the interaction picture with the total Hamiltonian written in the partitioned form $H = H_0 + gH_1$, where g is an arbitrary constant and in general H, H_0, and H_1 do not commute with each other. If H_0 is exactly invariant under a dynamical (degeneracy) group G_0, which has a subgroup g_0, and H is invariant under g_0, then H_1 is also invariant under g_0 while the noninvariance group $G \supset G_0$ has some of its generators not commuting with H. For the $3H_{\rm LR}(t)$-Hamiltonian we have $g_0 \sim$ SO(3) and $G \sim$ SU(3).

In this paper we write a simple expression for the interaction Hamiltonian on the basis of the arbitrary nature of the time dependence which relates the $3H_{\rm LR}(t)$-Hamiltonian to the usual time-independent $3H$-oscillator generators and discuss the significance of dilations and scaling transformations. First we recall how the noninvariance symmetry is generated.

2. NONINVARIANCE ALGEBRA

The Lewis–Riesenfeld Hamiltonian

$$H(t) = \tfrac{1}{2}\sum_i^3 [p_i\,p_i + \omega^2(t)q_iq_i] \tag{1}$$

is clearly not a constant of the motion, but we have established the existence of a symmetric tensor invariant

$$A_{ij}(q,p,\rho,\dot\rho) = \tfrac{1}{2}[\rho^{-2}q_iq_j + (\rho p - \dot\rho q)_i(\rho p - \dot\rho q)_j], \tag{2}$$

where $\rho(t)$ satisfies the constraint equation ($\lambda = \text{const}$)[11]

$$\rho^3(t)[\ddot\rho(t) + \omega^2(t)\rho(t)] = \lambda^2. \tag{3}$$

This invariant is a constant of the motion even though it does not commute with $H(t)$;

$$[A_{ij}(q,p,\rho,\dot\rho),H(t)] = \frac{\partial A_{ij}}{\partial t} \neq 0 \tag{4}$$

The antisymmetric **3**-tensor

$$J_k = i\epsilon_{ijk}q_ip_j, \quad i,j,k = 1,2,3 \tag{5}$$

together with the traceless symmetric **5**-tensor

$$B_{ij} = A_{ij} - \frac{\delta_{ij}}{3}\sum_k^3 A_{kk} \tag{6}$$

span the regular **8**-representation of the SU(3) algebra

$$[M_\mu,M_\nu] = 2iC_{\mu\nu\kappa}M_k, \quad \mu,\nu,\kappa = 1,2,\ldots,8, \tag{7}$$

[a)]Part of this work was delivered at the Seventh Integrative Conference on Group Theory and Mathematical Physics, The University of Texas at Austin, 1978.

 0022-2488/80/020286-04$1.00

where

$$J_1 = M_7, \quad J_2 = -M_5, \quad J_3 = M_2,$$
$$B_{12} = M_1, \quad B_{23} = M_6, \quad B_{13} = M_4,$$
$$(1/\sqrt{3})(B_{11} + B_{22} - 2B_{33}) = M_8,$$
$$B_{11} - B_{22} = M_3.$$

3. INTERACTION HAMILTONIAN

Following the discussion of Sec. 1, we devise a way of expressing the Lewis–Riesenfeld Hamiltonian in the desired interaction form. In doing so, the following criteria must be satisfied:

(i) The total Hamiltonian H must be equivalent to the Lewis–Riesenfeld Hamiltonian described in Sec. 2.

(ii) H_0 is the usual time-independent oscillator Hamiltonian (in three dimensions) which is exactly SU(3) invariant.

(iii) The time dependence of the righ-hand side is relegated to the interaction term H_1 in such a way that if it is no longer explicitly time-dependent, then H becomes equivalent to the time-independent Hamiltonian.

(iv) H_1 must be exactly invariant under SO(3) transformations. By writing the total Hamiltonian so as to fulfil these conditions, the nature of the symmetry breaking can be discussed along conventional lines.

The approach taken here is somewhat different from previous discussions of broken symmetries in quantum mechanical systems. [12,13] Usually the form of H_1 is given in terms of the appropriate dynamical variables on the basis of mathematical or physical assumptions. In this way the symmetry properties of the total Hamiltonian are determined, e.g., if $H_1 = g/q^2$, then the simplest noninvariance group is $SO(3)\times SO(2,1)$.

The problem being considered here amounts to a reversal of this situation since we already know the form of the total Hamiltonian H and its symmetries. [14] Accordingly there is the question of uniqueness of choice regarding the form H_1. However, we avoid this problem by noting that since $\omega(t)$ in Eq. (1) is an arbitary function, we may simply write $\omega^2(t) = \omega_0^2 + \Omega^2(t)$ to give an interaction expression for the Lewis–Riesenfeld Hamiltonian

$$H(t) = \tfrac{1}{2}(p_i^2 + \omega_0^2 q_i^2) + \tfrac{1}{2}\Omega^2(t) q_i^2, \tag{8}$$

which fulfills each of the above requirements.

4. CANONICAL TRANSFORMATIONS

A. Classical transformations

A similar breakup of the Lewis–Riesenfeld Hamiltonian can be achieved through the use of appropriate canonical transformations. [1] These results will be important for the discussion in Sec. 5.

The canonical transformations

$\{p,q\}$:space	$\{\bar{p},\bar{q}\}$:space
$p = \rho^{-1}\bar{p} + \dot{\rho}q$	$\bar{p} = \rho p - \dot{\rho}q$
$q = \rho\bar{q}$	$\bar{q} = \rho^{-1}q$

are generated by the function

$$F(p,\bar{q},\rho,\dot{\rho}) = -\rho p\bar{q} + \tfrac{1}{2}\rho\dot{\rho}\bar{q}^2 \tag{9}$$

and the $3H(t)$-Hamiltonian $H(q,p,t)$ becomes

$$H(q,p,t) = \rho^{-2}H_0(\bar{q},\bar{p}) - \frac{\partial F}{\partial t}. \tag{10}$$

With this observation we are able to write the desired interaction term as

$$H_1(\bar{q},\bar{p}) = -\frac{\partial F}{\partial t}(\bar{q},\bar{p}) = \rho^{-1}\dot{\rho}\bar{p}\bar{q} + \tfrac{1}{2}(\dot{\rho}^2 - \rho\ddot{\rho})\bar{q}^2. \tag{11}$$

B. Quantum transformations

The corresponding canonical transformations in quantum mechanics connecting the representation $\{p,q\}$ to $\{\bar{p},\bar{q}\}$ are given by

$$\bar{p} = UpU^{-1}, \quad \bar{q} = UqU^{-1},$$

where U is a unitary operator, thus ensuring the existence of the canonical commutation relations

$$[q_i,q_j] = 0, \quad [p_i,p_j] = 0, \quad [q_i,p_j] = i\delta_{ij}$$

in each representation. The form of these transformations is related then to the classical generating function by a theorem due to Dirac. [15]

If $|q\rangle$ and $|\bar{q}\rangle$ are eigenstates of the respective representations, then the transformation function between these eigenstates is defined through $F(\bar{q},q)$ treated as an operator in the mixed representation

$$\langle\bar{q}|q\rangle = \exp[iF(\bar{q},q)]. \tag{12}$$

For a given representation $|r\rangle$ we have the operator equations

$$\begin{aligned}\langle\bar{q}|\bar{p}|r\rangle &= \frac{\partial F(\bar{q},q)}{\partial\bar{q}}\langle\bar{q}|q\rangle,\\ \langle\bar{q}|p|r\rangle &= -\frac{\partial F(\bar{q},q)}{\partial q}\langle\bar{q}|q\rangle,\end{aligned} \tag{13}$$

corresponding to the usual classical equations.

The generator defined in Sec. 3A is written in the coordinate-momentum representation. Using the identity

$$\sum_q \langle\bar{q}|q\rangle\langle q|p\rangle = \langle\bar{q}|p\rangle, \tag{14}$$

we have, e.g.,

$$\langle\bar{q}|\bar{p}|p\rangle = \langle q|p\rangle\frac{\partial F(\bar{q},p)}{\partial\bar{q}}\langle\bar{q}|q\rangle = \frac{\partial F\langle\bar{q}|p\rangle}{\partial\bar{q}} \tag{15}$$

for which the corresponding transformation function is

$$\langle\bar{q}|p\rangle = \exp[iF(\bar{q},p)]. \tag{16}$$

The most general canonical transformations are generated by the classical action $\int dtL(\bar{q},\dot{\bar{q}},t)$ where the $3H_{LR}(t)$-Lagrangian is

$$L_{LR} = \tfrac{1}{2}\rho^2\dot{\bar{q}}^2 - \rho\dot{\rho}\bar{q}\cdot\dot{\bar{q}} - \tfrac{1}{2}(\rho^{-2} - \rho\ddot{\rho})\bar{q}^2, \tag{17}$$

and the Hamiltonian is

$$H_{LR} = \tfrac{1}{2}[(\rho^{-1}\bar{p} + \dot{\rho}\bar{q})^2 + (\rho^{-2} - \rho\ddot{\rho})\bar{q}^2]. \tag{18}$$

The generating function $F(\bar{q},p)$ represents an action related

to the simpler Lagrangian

$$L = -\tfrac{1}{2}\rho\dot{\rho}\dot{\vec{q}}\cdot\vec{q} + \tfrac{1}{2}(\dot{\rho}^2 + \rho\ddot{\rho})\vec{q}^2. \tag{19}$$

The associated Hamiltonian is a special case of the above, viz.

$$H = \tfrac{1}{2}(\rho^{-2}\vec{p}^2 - \dot{\rho}\ddot{\rho}\vec{q}^2). \tag{20}$$

4. BREAKING GENERATOR

The canonical variables in the interaction Hamiltonian evidently do not form a realization of SU(3) infinitesimal generators so that it is not possible to express the total Hamiltonian in terms of the quadratic Casimir invariant [16,17]

$$C_2^{[\alpha,\beta]} = \tfrac{1}{3}(\alpha_2 + \alpha\beta + \beta^2 + 3\alpha + 3\beta). \tag{21}$$

We can, however, determine how H_1 transforms under SU(3) generators. The canonical bilinear matrices

$$(Q_iQ_j),(P_iP_j),(Q_iP_j + P_iQ_j)$$

span the algebra of Sp(6,$\mathbb{R}$) [2,18] and H_1 can be associated with the trace, $\Sigma_i^3 Q_iQ_i$. Using the branching rule method described in Ref. 2, we can find its representation in SU(3). Recalling [2] that

$$\mathrm{SP}(6)\downarrow\mathrm{SU}(3):\ \mathbf{21} \supset \mathbf{8} \oplus \mathbf{6} \oplus \bar{\mathbf{6}} \oplus \mathbf{1}, \tag{22}$$

we see that H_1 must lie in a **6** and $\bar{\mathbf{6}}$. The bilinear $\frac{1}{2}\Sigma_k^3(q_k\,q_k - p_kp_k)$ transforms exactly as a **6** according to the commutation relations among SU(3) generators (See Sec. 2):

$$\begin{aligned}
&\tfrac{1}{2}\sum_k [J_i,(q_kq_k - p_k\,p_k)] = 0,\\
&\tfrac{1}{2}\sum_k [B_{ij},(q_kq_k - p_k\,p_k)]\\
&\quad = -i\left\{(q_i\,p_j + p_i\,q_j) - \frac{\delta_{ij}}{3}(q\cdot p + p\cdot q)\right\},\\
&\tfrac{1}{2}\sum_{ij} [B_{ij},[B_{ij},(q_kq_k - p_k\,p_k)]] = \tfrac{10}{3}\sum_k (q^2 - p^2).
\end{aligned} \tag{23}$$

In this normalization $c_2^{[2,0]} = \frac{10}{3}$ corresponds to **6**. [16]

5. DILATIONS AND THE WEYL GROUP

A. Time-dependent dilations

The transformations used in Sec. 3 are of three related types.

Type (I): $\{p,q\}\leftrightarrow\{\bar{p},\bar{q}\}$:

$$\begin{aligned} p &= p^{-1}\bar{p} + \dot{\rho}\bar{q}, \\ q &= \rho\bar{q} \end{aligned} \quad \text{and} \quad \begin{aligned} \bar{p} &= \rho p - \dot{\rho}q, \\ \bar{q} &= \rho^{-1}q \end{aligned};$$

Type (II): $\{p,q\}\leftrightarrow\{P,Q\}$

$$\begin{aligned} p &= P + (\dot{\rho}/\rho), \\ q &= Q \end{aligned} \quad \text{and} \quad \begin{aligned} P &= p - (\dot{\rho}/\rho) \\ Q &= q \end{aligned};$$

Type (III): $\{\bar{p},\bar{q}\}\leftrightarrow\{P,Q\}$

$$\begin{aligned} \bar{p} &= \rho P, \\ \bar{q} &= \rho^{-1}Q \end{aligned} \quad \text{and} \quad \begin{aligned} P &= \rho^{-1}\bar{p}, \\ Q &= \rho q \end{aligned}.$$

Type (I) transformations were used previously to establish the connections

$$\begin{array}{ccc} \mathrm{Tr}A_{ij}(q,p) & \rightarrow & H_0(\bar{q},\bar{p}) \\ \Big| & & \Big| \\ H(q,p,t) & \rightarrow & \rho^{-2}H_0(\bar{q},\bar{p}) \end{array}$$

Type (III) transformations are simply time-dependent dilations defined [19,20] by

$$\begin{aligned} U^{-1}(\alpha)PU(\alpha) &= e^{\alpha}P, \\ U^{-1}(\alpha)QU(\alpha) &= e^{-\alpha}Q, \end{aligned} \tag{24}$$

where

$$U(\alpha) = \exp[-i\alpha(t)]D. \tag{25}$$

and $\alpha = \log\rho$.

The generator $D = \Sigma_i^3 q_i\partial/\partial q_i$ spans the algebra of the Weyl group $W_G(3)$:

$$[q_i,D] = iq_i, \quad [p_i,D] = -ip_i, \quad [J_i,D] = 0. \tag{26}$$

Applying the scaling transformations of Type (III) to the Hamiltonian,

$$H_0(\bar{q},\bar{p}) = \tfrac{1}{2}(\bar{p}^2 + \omega_0^2\bar{q}^2), \tag{27}$$

furnishes the interaction form in $\{P,Q\}$-space

$$H_0(\bar{q},\bar{p}) = \tfrac{1}{2}[P^2 + \omega^2(t)Q^2 + (\ddot{\rho}/\rho)Q^2], \tag{28}$$

where Eq. (3) has been written

$$\lambda^2 = \omega_0^2 = \rho^3[\ddot{\rho} + \omega^2(t)\rho]. \tag{29}$$

B. Time-independent scale transformations

Consider the case where $\Omega(t) = \text{const}$; then the total Hamiltonian is just the time-independent oscillator. Moreover, if $\rho = \text{const} = \omega_0$, then

$$H_0 = \tfrac{1}{2}\rho^{-2}(p^2 + q^2). \tag{30}$$

In this case the time-independent oscillator scales under dilations, although the generators of $W_G(3)$ do not commute with those of SU(3). It seems plausible that this may be the underlying feature which has led to the recent suggestion that the semidirect product group $\mathrm{SU}(3)\odot W_G(3)$ may be a suitable noninvariance group for the $3H$-oscillator. The irreducible representations may be found using Mackey's theorem. [21] The induced reps U_W of $W_G(n)\uparrow\mathrm{SU}(n)$ which intertwine the projective representations Λ_n of SU(n) are evaluated from the group action. The resulting irreps

$$(U_W\Lambda_n)\uparrow\mathrm{SU}(n) \sim \Lambda_n \quad \left(\equiv \sum_{\mu}^{\infty} \oplus \Lambda_n^{\mu}\right) \tag{31}$$

are, in general, the tensor reps. corresponding to the $(a + I)$th energy eigenvalue [2,6,7] as described in Sec. 1.

6. CONCLUSION

In this paper we have resolved the outstanding problem of determining a dynamical realization for the interaction responsible for the absence of an SU(3) invariance symmetry in the Lewis–Riesenfeld oscillator. To arrive at this result requires the construction of an expression for the $3H_{\mathrm{LR}}(t)$-Hamiltonian in the interaction picture. This was achieved most simply by appealing to a previously defined canonical transformation. By using the branching rules for

$Sp(6)\downarrow SU(3)$, it was shown that the breaking generators transform as a $(\mathbf{6} \oplus \bar{\mathbf{6}})$ under SU(3). Time-dependent dilatations were found to relate the broken Hamiltonian to that of $3H_{LR}(t)$. Finally, it was suggested that time-independent scaling transformations applied to the $3H$-oscillator may provide a physical reason underlying the existence of the semidirect product group $W_G(3) \odot SU(3)$ as a noninvariance symmetry of the time-independent oscillator.

ACKNOWLEDGMENTS

During the corse of this work I have benefited from the continuing critical interest Dr. P. D. Jarvis and seveal helpful suggestions due to Dr. P. H. Dondi. Financial assistance through a University of Southampton Research Grant is also greatfully acknowledged.

[1]N.J. Günther and P.G. Leach, J. Math. Phys. **18**, 572 (1977).
[2]N.J. Günther, "Unitary Declension of Dynamical Symmetries for the Time-dependent Harmonic Oscillator," J. Math. Phys. (to be published).
[3]M.E. Major, J. Math. Phys. **18**, 1938, 1944, 1952 (1977).
[4]G. Burdet, M. Perrin, and M. Perroud, Comm. Math. Phys. **58**, 241 (1978).
[5]A. Barut, Phys. Rev. **135**, B893 (1964).
[6]N. Mukunda, L. O'Raifeartaigh, and E.C.G. Sudarshan, Phys. Rev. Lett. **15**, 1041 (1965).
[7]E.C.G. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, Phys. Rev. Lett. **19**, 322 (1965).
[8](a) H.R. Lewis, J. Math. Phys. **9**, 1976 (1968); (b) H.R. Lewis and W.B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[9]The Lewis-Riesenfeld Hamiltonian is generally denoted here $3H_{LR}(t)$, consistent with the notation developed in Refs. 1 and 2.
[10]E.B. Davies, *Quantum Theory of Open Systems* (Academic, New York, 1976).
[11]C.J. Eliezer and N.J. Günther, "Motion of a Harmonic Oscillator in a Magnetic Field—An Alternative Method of Solution," La Trobe University Report, 1976.
[12]J.P. Elliot, Proc. Roy. Soc. (Lond.) A **245**, 128 (1958).
[13]J. Lánik, Nucl. Phys. B **5**, 523 (1968).
[14]For this reason the problem here more closely resembles the formulation of broken SU(3) model of Hadronic mass spectra. See Ref. 16.
[15]J. Hamilton, *Theory of Elementary Particles* (Clarendon, Oxford, 1961).
[16]P. Carruthers, *Introduction to Unitary Symmetry* (Wiley, Sydney, 1966).
[17]It should be emphasised that the motivation to search for an interaction form of the $3H_{LR}(t)$-Hamiltonian is to determine a realization for the breaking generators. There is no need to consider methods for solving the total Hamiltonian since the exact algebraic solution is known. See Ref. 8b.
[18]K. Helmers, Nucl. Phys. B **23**, 594 (1961); M. Moshinsky, and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[19]P. Carruthers, Phys. Rep. C **I**, 1 (1971).
[20]D.J. Almond, unpublished lecture notes: "Unitary Ray Representations of Lie Groups and Their Applications to Quantum Mechanics," University of Southampton (1973).
[21]A.J. Coleman, "Induced and Subduced Representations," in *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1968), Vol. I.

The physical optics method in electromagnetic scattering

John S. Asvestas
Hughes Aircraft Company, Radar Systems Group, El Segundo, California 90247

(Received 8 August 1979; accepted for publication 2 November 1979)

The purpose of this work is to analyze the physical optics method as applied to electromagnetic scattering theory and to point out its physical and mathematical drawbacks. The main conclusions are (1) that the boundary values assumed by physical optics lead to electromagnetic fields that do not satisfy the finiteness of energy condition and, as a consequence, that integral representations of these fields cannot be obtained via the divergence theorem; (2) that the commonly accepted representations are not solutions of the physical optics problem because they fail to reproduce the assumed discontinuities of the fields on the scatterer. Despite the above conclusions, the present work should not be construed as an attempt to discredit the method but rather as an effort toward a better understanding of it. As it is well known, there have been a number of occasions in which physical optics has yielded quite satisfactory results.

I. INTRODUCTION

One of the most widely used methods for estimating the electromagnetic fields scattered off a three-dimensional, closed, perfectly conducting surface S is that of physical optics.[1,2] Briefly, the physical argument involved in this method is the following: If at each point of the scatterer the principal radii of curvature are large compared to the wavelength of the incident radiation, then for all practical purposes, the incident wave "sees" at that point an infinite, perfectly conducting plane tangent to the scattering surface at the point. For a point directly exposed to the incident rays (in the sense of geometric optics) the total field there can be reasonably approximated to be that for the infinite plane, while for a point not exposed to the incident rays it can be taken to be equal to zero. Once this assumption is made, the scattered fields off the surface are computed via integral representations which involve the assumed field distribution and which will be discussed below in detail. If, for simplicity, S is taken to be convex, then the dark region, S_1, of the scatterer is separated from the illuminated region, S_2, by a simple closed contour C. Denoting by $\{\mathbf{E}^i, \mathbf{H}^i\}$ the incident fields, by $\{\mathbf{E}, \mathbf{H}\}$ the scattered ones, and by $\hat{n}$ the exterior unit normal on S, then according to physical optics

$$\hat{n}\times\mathbf{E} = -\hat{n}\times\mathbf{E}^i, \quad \hat{n}\times\mathbf{H} = -\hat{n}\times\mathbf{H}^i, \tag{1}$$

on the dark side, while

$$\hat{n}\times\mathbf{E} = -\hat{n}\times\mathbf{E}^i, \quad \hat{n}\times\mathbf{H} = \hat{n}\times\mathbf{H}^i \tag{2}$$

on the illuminated side. Maxwell's equations together with (1) and (2) imply that

$$\hat{n}\cdot\mathbf{E} = -\hat{n}\cdot\mathbf{E}^i, \quad \hat{n}\cdot\mathbf{H} = -\hat{n}\cdot\mathbf{H}^i \quad \text{on } S_1, \tag{3}$$

$$\hat{n}\cdot\mathbf{E} = \hat{n}\cdot\mathbf{E}^i, \quad \hat{n}\cdot\mathbf{H} = -\hat{n}\cdot\mathbf{H}^i \quad \text{on } S_2. \tag{4}$$

This approximation to the exact scattering problem, also referred to as the Kirchhoff–Kottler formulation, has been the subject of much discussion over the years. For this reason another paper on the subject demands some justification. Bouwkamp[3] in his 1954 report on diffraction theory notes, "Many authors have recently devoted their attention to the classical Kirchhoff–Kottler formulation of the Huygens principle and its application to practical problems. Apart from the applications, essentially new knowledge has not been obtained. Old results were often re-formulated, re-interpreted, re-derived, re-named, or just simplified (?)." Fifteen years later, during which physical optics was the subject of considerable discussion,[1,4,5] Bowman *et al*[2] remarked "... recent years have seen a growing tendency to credit physical optics with an accuracy which is in no sense justifiable. It is therefore unfortunate that necessary and sufficient conditions for the validity of the method cannot be stated and, indeed, several of the most fruitful applications have been in circumstances where prior justification would be difficult." Indeed, a careful search through the literature referenced in the above works reveals that very little work has been done in understanding the method itself, the physical implications of the assumptions on the field distribution on S, and the structure (and mathematical implications) of the integral representations associated with it. The purpose of this paper is to address these questions and point out the physical and mathematical drawbacks of the method.

Much of the the confusion surrounding physical optics stems not from the assumption on the field distribution on S but from the integral representations used for calculating the scattered fields. The present plan is to accept (1)–(4) as the fundamental assumptions of physical optics and then attempt to derive integral representations. As it is well known, integral representations for electromagnetic scattered fields can be obtained in two ways: either by using the divergence theorem or by first postulating them in terms of surface density functions (as is commonly done in potential theory) and subsequently verifying their validity by checking whether they satisfy the requirements of the problem. Both approaches will be used here. Before doing so, however, a few notes on the accepted representations will be offered together with a more detailed plan of the paper and its main conclusions.

[a]This work was partly supported by the U.S. Air Force Office of Scientific Research Grant 74-2634.

 0022-2488/80/020290-10$1.00

The integral representations commonly associated with the physical optics method have their origin in the work of Kottler[6,7] on the black screen problem. Their derivation assumes that the fields off the surface are due to electric and magnetic charge and current densities on S as well as to electric and magnetic line charges on the closed contour C that separates the illuminated from the dark side of S. In this manner the fields are expressed in terms of surface integrals, known as the Stratton–Chu formulas,[8] and line integrals which bear the name of Kottler.

These modified Stratton–Chu formulas have been used through the years in their original form or in a modified but equivalent one.[9] In 1967, Kottler[10] attempted to derive them for the black screen problem via the divergence theorem but it was subsequently shown[11] that the derivation was in error for the resulting fields failed to satisfy the specified conditions at the edge of the black screen. In 1968, Sancer[5] followed similar lines for deriving the corresponding formulas for the physical optics case. As with Kottler[10] his derivation violated one of the asumptions of the divergence theorem, namely, that the function involved in the theorem must be continuous in the closed region in which the theorem is applied.

In the present work, the question of whether integral representations can be obtained through a legitimate use of the divergence theorem is addressed in Sec. II. Although the answer is in the negative, the process yields the interesting by-product that the fields resulting from the physical optics approximation do not satisfy the finiteness of energy condition; thus, once the surface fields are specified as in (1)–(4), then irrespective of the integral representations to be employed, the scattered fields will violate a physical condition that scattered fields in all situations are supposed to obey. In Sec. III, Hertz vector-potentials are employed for deriving representations (potential method). It is shown subsequently that the fields so obtained fail to reproduce the assumed discontinuities across C and, hence, are not solutions of the physical optics problem. The analysis employed also answers in the negative the long standing question of whether the fields resulting from the physical optics approximation can be used as the first term in an iterative scheme that would converge to the exact solution of the scattering problem. Seciton IV offers some concluding remarks, while detailed computations are left to Appendices A, B, and C.

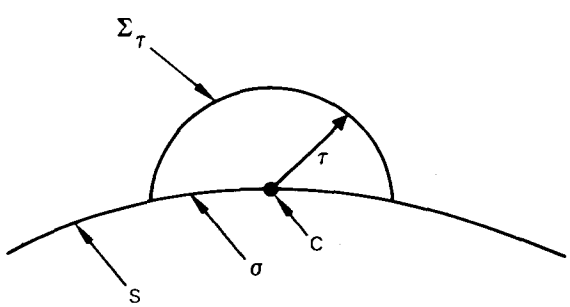


FIG. 1. Cross section of S showing surfaces σ and Σ_τ, and curve C.

II. THE DIVERGENCE THEOREM AND THE ENERGY CONDITION

The scattering surface under consideration is a three-dimensional, closed and bounded surface S. Its mathematical properties will be described in greater detail in the next section but, for the present, it is sufficient to assume that it is smooth enough for the divergence theorem to be applicable.[12,13] The incident electromagnetic wave is assumed to be such that the dark region, S_1, is separated from the illuminated region, S_2, by a simple closed contour C which is positively oriented with respect to the exterior normal on S_1. From (1)–(4), the scattered fields obey the following conditions on C:

$$\hat{n}_2\times\mathbf{E}_2 - \hat{n}_1\times\mathbf{E}_1 = \mathbf{0}\,, \quad \hat{n}_2\times\mathbf{H}_2 - \hat{n}_1\times\mathbf{H}_1 = 2\hat{n}\times\mathbf{H}^i\,, \quad \text{on } C\,, \tag{5}$$

$$\hat{n}_2\cdot\mathbf{E}_2 - \hat{n}_1\cdot\mathbf{E}_1 = 2\hat{n}\cdot\mathbf{E}^i\,, \quad \hat{n}_2\cdot\mathbf{H}_2 - \hat{n}_1\cdot\mathbf{H}_1 = 0\,, \text{ on } C\,, \tag{6}$$

where the subscript 1 (2) denotes the approach to C from S_1 (S_2), and $\hat{n}$ denotes the exterior unit normal on S.

Since the scattered fields are not continuous across C, the divergence theorem is not directly applicable.[12,13] For this reason an auxiliary surface Σ_τ, is constructed by using each point P of C as the center of a circle of radius τ which lies on the plane normal to both S and C at P. For C sufficiently smooth and τ sufficiently small, the resulting structure is a tubelike, doubly connected, closed surface surrounding C. Denoting by σ the part of S intercepted by this tube and Σ_τ the part of the tube exterior to S (Fig. 1), the divergence theorem is applied in the closed region whose boundary consists of the closed surface $(S-\sigma)\,U\Sigma_\tau$ and the surface of a sphere whose radius eventually recedes to infinity.

Under the assumption that the scattered fields are continuous in the closed region above and that they satisfy the radiation conditions at infinity, then the application of the divergence theorem in the form of a vector[8] or a dyadic[14] Green's identity yields

$$\mathbf{E}(\mathbf{r}') = \lim_{\tau\to 0}\int_{(S-\sigma)U\Sigma_\tau}\mathbf{A}\,dS - \frac{i}{\omega\epsilon}\lim_{\tau\to 0}\int_{(S-\sigma)U\Sigma_\tau}\hat{n}\cdot\nabla\times(\mathbf{H}\nabla g)\,dS\,, \quad \mathbf{r}'\notin(S-\sigma)U\Sigma_\tau\,, \tag{7}$$

with

$$\mathbf{A} = \hat{n}\cdot\mathbf{E}\nabla g + (\hat{n}\times\mathbf{E})\times\nabla g + i\omega\mu g\hat{n}\times\mathbf{H}\,, \tag{8}$$

$$g(\mathbf{r},\mathbf{r}') = e^{ik|\mathbf{r}-\mathbf{r}'|}/4\pi|\mathbf{r}-\mathbf{r}'|\,, \tag{9}$$

and a similar expression holding for $\mathbf{H}$. The constants ϵ and μ stand for the permittivity and permeability of the medium while ω and k stand for the angular frequency and the wave number of the incident wave ($e^{-i\omega t}$ time-dependence).

The last integral in (7) vanishes by virtue of Stokes' theorem; moreover, since $S-\sigma\to S$ as $\tau\to 0$, the integral of $\mathbf{A}$ over $S-\sigma$ tends to the integral of $\mathbf{A}$ over S, so that (7) becomes

$$\mathbf{E}(\mathbf{r}') = \int_S \mathbf{A}\,dS + \lim_{\tau\to 0}\int_{\Sigma_\tau}\mathbf{A}\,dS\,. \tag{10}$$

Since the values that $\mathbf{A}$ takes on Σ_τ are not known, the last integral in (10) cannot be evaluated directly. The only way this integral can be handled is by means of the finiteness of energy condition, the only remaining physical condition that can be employed in this problem. This condition requires that the scattered fields are locally square-integrable or, in physical terms, that the energy contained in a finite region of space is finite and tends to zero with the volume of the region. If this condition obtains, it can be shown that the second integral in (10) tends to zero with τ; if not, then no conclusion can be drawn and the divergence theorem fails to produce representations.

The next step then is to determine whether the present problem satisfies the energy condition which, in mathematical form and for the region of interest, reads

$$\int_{V_\tau} |\mathbf{E}|^2 \, dV < \infty \,, \quad \int_{V_\tau} |H|^2 \, dV < \infty \quad , \tag{11}$$

V_τ being the region of space enclosed by $\sigma U \Sigma_\tau$. To this end the following theorem is proved in Appendix A with $\hat{t}$ being the unit vector tangent to C and pointing in the direction in which the curve is traced.

Theorem 1: If conditions (11) obtain, then the line integrals

$$\int_C \hat{t}\cdot(\mathbf{H}_1 - \mathbf{H}_2) g \, ds \quad \text{and} \quad \int_C \hat{t}\cdot(\mathbf{E}_1 - \mathbf{E}_2) g \, ds \,, \quad \mathbf{r}' \notin S \,, \tag{12}$$

are necessarily equal to zero.

An examination of the proof of this theorem reveals that the theorem remains valid if g is replaced by any function f which is defined and has continuous first partial derivatives in the interior of $\sigma U \Sigma_\tau$. This observation is of consequence because of the following well-known theorem.[15]

Theorem 2: If a function F is (Lebesgue) integrable in an open set ω and if, for any function G which is continuous on ω, the equation $\int_\omega FG \, dV = 0$ holds, then F must satisfy the condition $\int_\omega |F| \, dV = 0$, and consequently F is equal to zero almost everywhere on ω.

These two theorems imply the following:

Theorem 3: If the function $\hat{t}\cdot(\mathbf{H}_1 - \mathbf{H}_2)$ and $\hat{t}\cdot(\mathbf{E}_1 - \mathbf{E}_2)$ are (Lebesgue) integrable on C and if (11) holds, then the two functions are necessarily equal to zero almost everywhere on C.

In the physical optics approximation the functions of Theorem 3 are integrable on C since they are given in terms of the incident fields; in fact, they are continuous except at those isolated points at which $\hat{t}$ may not be continuous. Were conditions (11) then to hold, these functions would be identically zero except at a number of isolated points. This is certainly not the case for $\hat{t}\cdot(\mathbf{H}_1 - \mathbf{H}_2)$ as witnessed from the second of (5) and, hence, the finiteness of energy condition is not satisfied. In turn, the second integral in (10) cannot be evaluated and the divergence theorem does not yield representations.

As it is well known,[9] scattered fields are required to satisfy the finiteness of energy condition and, therefore, the failure of the physical optics fields in this respect constitutes a most serious drawback of the method. Since, also, this failure is a direct result of the very initial assumptions, i.e., Eqs. (1)–(4), it is evident that the method is off to a bad start.

As mentioned in the Introduction, Sancer,[5] ignoring the continuity requirement of the divergence theorem and by means of it, obtained a pair of representations for the fields resulting from the physical optics approximation. Using a dyadic form of the theorem he derived (7) with the integration extending over S rather than $(S - \sigma)U\Sigma_\tau$. For the second integral he employed Stokes' theorem over S_1 and S_2 to convert the surface integral into a line integral, his final results being

$$\mathbf{E}(\mathbf{r}') = \int_S [\hat{n}\cdot\mathbf{E}\nabla g + (\hat{n}\times\mathbf{E})\times\nabla g + i\omega\mu g \hat{n}\times\mathbf{H}] \, dS - \frac{i}{\omega\epsilon}\int_C \hat{t}\cdot(\mathbf{H}_1 - \mathbf{H}_2)\nabla g \, ds \,, \tag{13}$$

$$\mathbf{H}(\mathbf{r}') = \int_S \{\hat{n}\cdot\mathbf{H}\nabla g + (\hat{n}\times\mathbf{H})\times\nabla g - i\omega\epsilon g \hat{n}\times\mathbf{E}\} \, dS + \frac{i}{\omega\mu}\int_C \hat{t}\cdot(\mathbf{E}_1 - \mathbf{E}_2)\nabla g \, ds \,. \tag{14}$$

These are the Stratton–Chu[8] modified formulas and, as it will be shown in the next section, they do not constitute integrals representations for the physical optics fields.

Sancer's oversight of the continuity requirement can be seen to lead to inconsistent results. The same integral that led to the line integral in (13) can be written via the divergence theorem in the form

$$\int_S \hat{n}\cdot\nabla\times(\mathbf{H}\nabla g) \, dS = -\lim_{\rho\to 0}\int_{S_\rho} \hat{n}\cdot\nabla\times(\mathbf{H}\nabla g) \, dS - \lim_{R\to\infty}\int_{S_R} \hat{n}\cdot\nabla\times(\mathbf{H}\nabla g) \, dS \,, \tag{15}$$

where S_ρ is a sphere of radius ρ and center at $\mathbf{r}'$, while S_R is a sphere of radius R and center at $\mathbf{r}'$. Both the integrals on the right are well-behaved and vanish before taking the indicated limits by virtue of Stokes' theorem. Naturally, the statement in (15) is questionable since the first integrand is not continuous on S, but Sancer's is not either. It is true that the continuity condition is only a sufficient condition of the divergence theorem; nevertheless, it is a condition under which the theorem has been proven and as such it cannot be ignored. In Appendix B examples are given where the violation of this condition leads to inconsistent results.

III. AN EXAMINATION OF THE EXISTING REPRESENTATIONS

Since the divergence theorem fails to ascertain whether the modified Stratton–Chu formulas (13) and (14) are indeed solutions of the problem resulting from the physical optics approximation, it then remains to examine whether these formulas satisfy the requirements of the problem. As it is mentioned in the Introduction, these formulas were originally obtained using physical arguments. Specifically, the electric and magnetic charge and current density on S were taken into consideration as well as electric and magnetic line charges on C. An alternative way of obtaining them is to consider only the electric and magnetic current densities and form the Hertz vectors

$$\pi(\mathbf{r}') = \frac{i}{\omega\epsilon}\int_S \hat{n}\times\mathbf{h}g\,dS\,, \quad \pi^*(\mathbf{r}') = -\frac{i}{\omega\mu}\int_S \hat{n}\times\mathbf{e}g\,dS\,, \tag{16}$$

where $\hat{n}\times\mathbf{h}$ and $\hat{n}\times\mathbf{e}$ denote the assigned values of the tangential magnetic and electric scattered fields on S as given by (1) and (2). In terms of these vectors the scattered fields at points off the surface are given by[9]

$$\mathbf{E}(\mathbf{r}') = \nabla'\times\nabla'\times\pi(r') + i\omega\mu\nabla'\times\pi^*(\mathbf{r}')\,, \tag{17}$$

$$\mathbf{H}(\mathbf{r}') = \nabla'\times\nabla'\times\pi^*(r') - i\omega\epsilon\nabla'\times\pi(\mathbf{r}')\,. \tag{18}$$

These formulas are often called the Franz[16] representations. The modified Stratton–Chu formulas follow from them by performing the indicated differentiations and using standard vector identities; conversely, (17) and (18) can be obtained from (13) and (14) by reversing the steps. Thus, the two sets of representations are mathematically equivalent.

For (13) and (14) to be solutions of the problem they must satisfy Maxwell's equations at points exterior to S, the radiation conditions at infinity, and also reproduce the assigned values of the fields on the scatterer. It can be readily shown that the first two requirements are satisfied. The third one, however, and, specifically, the conditions (5) and (6) across C do not obtain. This can be shown either by using (13) and (14) and utilizing the results of Müller[13] or by first transforming (13) and (14) to what is known as Kottler's formulas and then employing potential-theoretic techniques. These formulas are[10]

$$\mathbf{E}(\mathbf{r}') = \int_S\left(\mathbf{e}\frac{\partial g}{\partial n} - g\frac{\partial \mathbf{e}}{\partial n}\right)dS - \int_C g(\mathbf{e}_1-\mathbf{e}_2)\times\hat{t}\,ds - \frac{i}{\omega\epsilon}\int_e \hat{t}\cdot(\mathbf{h}_1-\mathbf{h}_2)\nabla g\,ds\,, \tag{19}$$

$$\mathbf{H}(\mathbf{r}') = \int_S\left(\mathbf{h}\frac{\partial g}{\partial n} - g\frac{\partial \mathbf{h}}{\partial n}\right)dS - \int_C g(\mathbf{h}_1-\mathbf{h}_2)\times\hat{t}\,ds + \frac{i}{\omega\mu}\int_C \hat{t}\cdot(\mathbf{e}_1-\mathbf{e}_2)\nabla g\,ds\,, \tag{20}$$

and can be obtained from (13) and (14) through a series of vector operations; conversely, (13) and (14) can be obtained from (19) and (20) by reversing the steps so that the two sets of formulas are mathematically equivalent.

The representations in (19) and (20) are easier to work with in terms of a local analysis in a neighborhood of a point of C, for the surface integrals have been extensively studied in potential theory.[12,17] The class of surfaces to be considered is a fairly common one in both potential[12] and scattering theory,[13] and it can be described as follows. The surface S possesses a tangent plane at each one of its points. Moreover, for any point M of S and a rectangular coordinate system xyz with origin at M and the z axis in the direction of the normal, there exists a real number $\eta>0$ such that for points (x, y) belonging to the disk $x^2+y^2\leqslant\eta^2$, the part S_0 of S intercepted by the sphere $x^2+y^2+z^2\leqslant\eta^2$ can be expressed in the form

$$z = f(x, y)\,, \tag{21}$$

with

$$f(0,0) = f_x(0,0) = f_y(0,0) = 0\,, \tag{22}$$

and f possessing continuous second partial derivatives in $x^2+y^2\leqslant\eta^2$. For M a point of C, the y axis points in the direction in which C is traced while the x axis points toward the illuminated region (S_2). To avoid unnecessary complications of the mathematical arguments the portion C_0 of C intercepted by the sphere of radius η is assumed to lie on the yz plane, i.e.,

$$C_0 = \{(y, z): z = f(0, y)\,,\ |y|\leqslant\eta\}\,. \tag{23}$$

The local analysis is performed on (19) and (20) in the neighborhood of a point M of C. The point $\mathbf{r}'$ is allowed first to approach the points $\mathbf{r}_1 = (-\delta, 0, f(-\delta, 0))$ and $\mathbf{r}_2 = (\delta, 0, f(\delta, 0))$ of S, in turn, and then these points are taken to the origin along the curve Γ which, for simplicity, is defined by

$$\Gamma = \{(x, z): z = f(x, 0)\,,\ |x|\leqslant\delta<\eta\}\,. \tag{24}$$

It is noted first that the line integrals in (19) and (20) are well behaved at $\mathbf{r}_1$ and $\mathbf{r}_2$. Secondly, the function g can be written in the form

$$g(\mathbf{r}, \mathbf{r}') = \frac{\exp(ik|\mathbf{r}-\mathbf{r}'1}{4\pi|\mathbf{r}-\mathbf{r}'|} + \frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|} = G(\mathbf{r}, \mathbf{r}') + \frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}\,, \tag{25}$$

where G and its gradient have expansions of the form

$$G(\mathbf{r}, \mathbf{r}') = \frac{ik}{4\pi}\sum_{n=0}^{\infty}\frac{(ik|\mathbf{r}-\mathbf{r}')^n}{(n+1)!}\,, \tag{26}$$

$$\nabla G(\mathbf{r}, \mathbf{r}') = \frac{k^2}{4\pi}\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\sum_{n=0}^{\infty}\frac{(n+1)(ik|\mathbf{r}-\mathbf{r}'|)^n}{(n+2)!}\,.$$

The surface integral in (19) with g replaced by G is a continuous function of $\mathbf{r}'$ everywhere in space. That this is the case can be directly concluded from Lemmas 73 and 74 of Müller.[13]

The remaining surface integral, namely

$$\frac{1}{4\pi}\int_S\left(\mathbf{e}(\mathbf{r})\frac{\partial}{\partial n}|\mathbf{r}-\mathbf{r}'|^{-1} - |\mathbf{r}-\mathbf{r}'|^{-1}\frac{\partial\mathbf{e}(\mathbf{r})}{\partial n}\right)dS, \tag{27}$$

represents the sum of a double layer and a simple layer potential. Since the density of the simple layer is a bounded and integrable function, then[17] the potential is continuous in the entire space. Similarly,[17] since the density of the double layer is continuous in a neighborhood of $\mathbf{r}_1$, then

$$\lim_{\mathbf{r}'\to\mathbf{r}_1}\frac{1}{4\pi}\int_S \mathbf{e}(\mathbf{r})\frac{\partial}{\partial n}|\mathbf{r}-\mathbf{r}'|^{-1}\,dS = \tfrac{1}{2}\mathbf{e}(\mathbf{r}_1) + \frac{1}{4\pi}\int_S \mathbf{e}(\mathbf{r})\frac{\partial}{\partial n}|\mathbf{r}-\mathbf{r}_1|^{-1}\,dS \tag{28}$$

with the last integral being continuous[17] on S, and with a similar statement holding as $\mathbf{r}\to\mathbf{r}_2$.

With these results it follows from (19) that

$$\mathbf{E}(\mathbf{r}_1) = \tfrac{1}{2}\mathbf{e}(\mathbf{r}_1) + \int_S\left(\mathbf{e}(\mathbf{r})\frac{\partial g(\mathbf{r}, \mathbf{r}_1)}{\partial n} - g(\mathbf{r}, \mathbf{r}_1)\frac{\partial\mathbf{e}(\mathbf{r})}{\partial n}\right)dS - \int_C[\mathbf{e}_1(\mathbf{r})-\mathbf{e}_2(\mathbf{r})]\times\hat{t}g(\mathbf{r}, \mathbf{r}_1)\,ds - \frac{i}{\omega\epsilon}\int_C\hat{t}\cdot[\mathbf{h}_1(\mathbf{r})-\mathbf{h}_2(\mathbf{r}_2)]\nabla g(\mathbf{r}, \mathbf{r}_1)\,ds\,, \tag{29}$$

with the corresponding expression for $\mathbf{E}(\mathbf{r}_2)$ being obtainable

from (29) by replacing $\mathbf{r}_1$ by $\mathbf{r}_2$; moreover, the expression for $\mathbf{H}$ at these points follows directly from (20). Using (29) to form the differences in (5) and (6) results in

$$\begin{aligned}\hat{\nu}\times(\mathbf{E}_2-\mathbf{E}_1) &= \lim_{\delta\to 0}[\hat{n}_2\times\mathbf{E}(\mathbf{r}_2)-\hat{n}\times\mathbf{E}(\mathbf{r}_1)]\\ &= -\tfrac{1}{2}\lim_{\delta\to 0}[\hat{n}_2\times\mathbf{e}(\mathbf{r}_2)-\hat{n}_1\times\mathbf{e}(\mathbf{r}_1)]\\ &\quad -\lim_{\delta\to 0}\int_C\{\hat{n}_2\times[(\mathbf{e}_1-\mathbf{e}_2)\times\hat{t}\,]g(\mathbf{r},\mathbf{r}_2)\\ &\quad -\hat{n}_1\times[(\mathbf{e}_1-\mathbf{e}_2)\times\hat{t}\,]g(\mathbf{r},\mathbf{r}_1)\}\,ds\\ &\quad -\frac{i}{\omega\epsilon}\lim_{\delta\to 0}\int_C\hat{t}\cdot(\mathbf{h}_1-\mathbf{h}_2)\\ &\quad\times[\hat{n}_2\times\nabla g(\mathbf{r},\mathbf{r}_2)-\hat{n}_1\times\nabla g(\mathbf{r},\mathbf{r}_1)]\,ds\,,\end{aligned}\tag{30}$$

where $\hat{\nu}$ is the normal on S at M. The exclusion of the surface integral is due to its aforementioned property of continuity. Expressed in terms of the incident fields (30) becomes

$$\begin{aligned}&\hat{\nu}\times(\mathbf{E}_2-\mathbf{E}_1)\\ &= 2\lim_{\delta\to 0}\int_C\hat{n}\cdot\mathbf{E}^i\{(\hat{n}_2\times\hat{q})g(\mathbf{r},\mathbf{r}_2)-(\hat{n}_1\times\hat{q})g(\mathbf{r},\mathbf{r}_1)\}\,ds\\ &\quad+\frac{2i}{\omega\epsilon}\lim_{\delta\to 0}\int_C\hat{t}\cdot\mathbf{H}^i\{\hat{n}_2\times\nabla g(\mathbf{r},\mathbf{r}_2)-\hat{n}_1\times\nabla g(\mathbf{r},\mathbf{r}_1)\}\,ds\,,\end{aligned}\tag{31}$$

with

$$\hat{q}=\hat{t}\times\hat{n}\,.\tag{32}$$

Similarly,

$$\begin{aligned}&\hat{\nu}\cdot(\mathbf{E}_2-\mathbf{E}_1)\\ &= \hat{\nu}\cdot\mathbf{E}^i-2\lim_{\delta\to 0}\int_C\hat{n}\cdot\mathbf{E}^i\{\hat{n}_2\cdot\hat{q}g(\mathbf{r},\mathbf{r}_2)-\hat{n}_1\cdot\hat{q}g(\mathbf{r},\mathbf{r}_1)\}\,ds\\ &\quad+\frac{2i}{\omega\epsilon}\lim_{\delta\to 0}\int_C\hat{t}\cdot\mathbf{H}^i\{\mathbf{n}_2\cdot\nabla g(\mathbf{r},\mathbf{r}_2)-\hat{n}_1\cdot\nabla g(\mathbf{r},\mathbf{r}_1)\}\,ds\end{aligned}\tag{33}$$

while,

$$\begin{aligned}\hat{\nu}\times(\mathbf{H}_2-\mathbf{H}_1) = \hat{\nu}\cdot\mathbf{H}^i+2\lim_{\delta\to 0}\int_C\hat{q}\cdot\mathbf{H}^i\{(\hat{n}_2\times\hat{n})g(\mathbf{r},\mathbf{r}_2)\\ -(\hat{n}_1\times\hat{n})g(\mathbf{r},\mathbf{r}_1)\}\,ds\,,\end{aligned}\tag{34}$$

$$\hat{\nu}\cdot(\mathbf{H}_2-\mathbf{H}_1)=2\lim_{\delta\to 0}\int_C\hat{q}\cdot\mathbf{H}^i\{\hat{n}_2\cdot\hat{n}g(\mathbf{r},\mathbf{r}_2)-\hat{n}_1\cdot\hat{n}g(\mathbf{r},\mathbf{r}_1)\}\,ds\,.\tag{35}$$

All but one of the limits in (31)–(35) exist and are equal to zero as shown in Appendix C. The one that does not exist is that of the second integral in (31). From (C20), Appendix C,

$$\begin{aligned}&\int_C\hat{t}\cdot\mathbf{H}^i\{\hat{n}_2\times\nabla g(\mathbf{r},\mathbf{r}_2)-\hat{n}_1\times\nabla g(\mathbf{r},\mathbf{r}_1)\}\,ds\\ &\quad=[\hat{n}(M)+\hat{q}(M)]O(1)+\hat{t}(M)O(\delta^{-1})\,,\quad\delta\to 0^+\,;\end{aligned}\tag{36}$$

thus the conditions (5) and (6) across C do not obtain and, hence, (19) and (20) [or, equivalently, (17) and (18), or (13) and (14)] are not solutions of the problem posed by the physical optics approximation.

The above calculations not only show the incompatibility between the physical optics problem and its purported solution but also conclusively answer the question asked by various authors,[2,3,18] namely, whether (19) and (20) could be used in an iteration scheme which might converge to the correct solution of the original problem. Bouwkamp[3] answers the question in the negative and, in support, he cites the work of Schelkunoff.[18] The latter, however, used an argument which, in light of the conclusions here, is in error. His argument was that the fields resulting from (19) and (20), when brought to the surface, would reduce to those assumed by the physical optics approximation and, thus, the same scattered fields would result iteration after iteration. The true reason, of course, is that the limit in (31) does not exist in the direction $\hat{t}$ in which C is traced and, therefore, at least the last integral in (20) will not exist in the second iteration. It is worth noting that, were the limit in (31) to exist and be equal to zero, then successive iterations of (19) and (20) might converge to the exact solution since the discontinuities in (5) and (6) are cut in half after one iteration as it is clear from (33) and (34).

IV. CONCLUSION

The purpose of this work has been twofold: first, to point out that the assumed surface fields of physical optics result in scattered fields which do not obey the finiteness of energy condition. A consequence of this is that the divergence theorem cannot be used in deriving integral representations for these fields. The second is to show that the modified Stratton–Chu formulas are not solutions for the scattered fields of physical optics and that they cannot be iterated to improve the physical behavior of these fields.

In conclusion, it should be mentioned that although the modified Stratton–Chu formulas are, in a mathematical sense at least, arbitrary formulas for the fields of an already approximated problem, they do seem to yield satisfactory results on occasion; thus, the present work, should not be construed as an attempt to discredit the method but rather as an effort toward a better understanding of it.

APPENDIX A

It is shown here that if the fields satisfy (11), then the line integrals in (12) are necessarily equal to zero. The first step toward this end is the parametrization of the curve C and the erection of local coordinates.

With respect to an arbitrary rectangular coordinate system xyz, the curve C can be parametrized in terms of its arc length starting with an arbitrary point P of C; thus

$$C=\{(x,y,z)\colon x=f(s)\,,\ y=g(s)\,,\ z=h(s)\,,\ 0\leqslant s\leqslant L\}\,,\tag{A1}$$

where L is the length of C, and f, g, and h are assumed to have continuous second derivatives with $f'^2+g'^2+h'^2=1$. The unit tangent vector to C is given by

$$\hat{t}=f'(s)\hat{x}+g'(s)\hat{y}+h'(s)\hat{z}\,.\tag{A2}$$

At each point P of C it is assumed that there exists a unit vector $\hat{n}$ normal to S at P which is also normal to $\hat{t}$. Letting

$$\hat{t}'=\hat{n}\times\hat{t}\,,\tag{A3}$$

then the triple $(\hat{n},\hat{t},\hat{t}')$ is a positive triple of orthonormal vectors at each point of C.

The vectors $\hat{n}$ and $\hat{t}'$ lie on a plane perpendicular to both S and C at the point P under consideration. On this plane the polar coordinate system (ρ, ϕ), $\rho \geqslant 0$, $0 \leqslant \phi < 2\pi$, is introduced with the pole at P and the angle measured from $\hat{t}$ to $\hat{n}$. If $\mathbf{r}_0$ is the position vector to a point of C, and if $\boldsymbol{\rho}$ is the position vector on the $\hat{t}'$–$\hat{n}$ plane, then the equation

$$\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z} = \mathbf{r}_0 + \boldsymbol{\rho} \tag{A4}$$

defines a transformation of points (s, ρ, ϕ) to points (x, y, z) with Jacobian

$$J = \frac{\partial \mathbf{r}}{\partial s} \cdot \frac{\partial \mathbf{r}}{\partial \rho} \times \frac{\partial \mathbf{r}}{\partial \phi} = \rho\left(1 - \rho\boldsymbol{\rho} \cdot \frac{d\hat{t}}{ds}\right). \tag{A5}$$

The continuity of the second derivatives of f, g, and h guarantees the boundedness of $|d\hat{t}/ds|$ which in turn guarantees that, for ρ sufficiently small, the Jacobian is positive and the transformation one-to-one.

The next step is to restate the energy conditions (11) in a different but equivalent form. The claim is that if the integrals

$$\int_{\Sigma_\rho} |\mathbf{E}|^2 \, dS \quad \text{and} \quad \int_{\Sigma_\rho} |\mathbf{H}|^2 \, dS$$

exist and are continuous for $\rho > 0$, then conditions (11) are equivalent to

$$\int_{\Sigma_\rho} |\mathbf{E}|^2 \, dS = O(\rho^{-1+\alpha}),$$

$$\int_{\Sigma_\rho} |\mathbf{H}|^2 \, dS = O(\rho^{-1+\alpha}), \quad \rho \to 0, \tag{A6}$$

where α is a positive real number. The surface Σ_ρ is defined in Sec. II and it is here referred to the coordinate system (s, ρ, ϕ).

The proof of this theorem has as follows: If the first of (A6) holds then the limit

$$\lim_{\delta \to 0^+} \int_\delta^\tau d\rho \int_{\Sigma_\rho} |\mathbf{E}|^2 \, dS \tag{A7}$$

exists and, by Fubini's theorem

$$\int_{V_\tau} |\mathbf{E}|^2 \, dV = \lim_{\delta \to 0^+} \int_\delta^\tau d\rho \int_{\Sigma_\rho} |\mathbf{E}|^2 \, dS. \tag{A8}$$

Conversely, if the left-hand side of (A8) exists, the limit in (A8) exists and, since the integrand (surface integral) is nonnegative and continuous for $0 < \rho \leqslant \tau$, the first of (A6) has to hold. The proof for the second of (A6) follows the same lines.

With this background, Theorem 1 of Sec. II can now be proved. Letting $\mathbf{r}'$ be a point strictly off the surface $(S - \sigma) \cup \Sigma_\rho$ and using Maxwell's equations and standard vector identities, it follows that

$$\int_{\Sigma_\rho} \hat{n} \cdot \nabla \times [g(\mathbf{r}, \mathbf{r}')\mathbf{H}(\mathbf{r})] \, dS$$

$$= \int_{\Sigma_\rho} \boldsymbol{\rho} \cdot [\nabla g \times \mathbf{H} + g \nabla \times \mathbf{H}] \, dS$$

$$= -\int_{\Sigma_\rho} [\nabla g \cdot (\hat{\rho} \times \mathbf{H}) + i\omega\epsilon g \hat{\rho} \cdot \mathbf{E}] \, dS, \tag{A9}$$

so that

$$\left| \int_{\Sigma_\rho} \hat{n} \cdot \nabla \times [g(\mathbf{r}, \mathbf{r}')\mathbf{H}(\mathbf{r})] \, dS \right|$$

$$\leqslant \left| \int_{\Sigma_\rho} \nabla g \cdot (\hat{\rho} \times \mathbf{H}) \, dS \right| + \omega\epsilon \left| \int_{\Sigma_\rho} g \hat{\rho} \cdot \mathbf{E} \, dS \right|. \tag{A10}$$

The last term in this can be estimated by means of (A6) and the Cauchy–Schwarz inequality as follows,

$$\left| \int_{\Sigma_\rho} g \hat{\rho} \cdot \mathbf{E} \, dS \right|^2$$

$$\leqslant \left(\int_{\Sigma_\rho} |g|^2 \, dS \right) \left(\int_{\Sigma_\rho} |\hat{\rho} \cdot \mathbf{E}|^2 \, dS \right)$$

$$\leqslant \left[\rho \int_{\Sigma_\rho} |g|^2 \left(1 - \rho\hat{\rho} \cdot \frac{d\hat{t}}{ds}\right) d\phi ds \right] \left(\int_{\Sigma_\rho} |\mathbf{E}|^2 \, dS \right)$$

$$\leqslant \rho d_1(\mathbf{r}') d_2 \rho^{-1+\alpha} = D_1(\mathbf{r}') \rho^\alpha, \tag{A11}$$

where $d_1(\mathbf{r}')$ stands for the first integral; similarly,

$$\left| \int_{\Sigma_\rho} \nabla g \cdot (\hat{\rho} \times \mathbf{H}) dS \right|^2 \leqslant D_2(\mathbf{r}') \rho^\alpha. \tag{A12}$$

Since $\alpha > 0$, both integrals tend to zero with ρ and, consequently, so does the integral in the left-hand side of (A9). But that integral can be transformed by means of Stokes' theorem to the following

$$\int_{\Sigma_\rho} \hat{n} \cdot \nabla \times (g\mathbf{H}) \, dS = \int_{C_2} \hat{t} \cdot \mathbf{H} g \, ds - \int_{C_1} \hat{t} \cdot \mathbf{H} g \, ds$$

$$\underset{\rho \to 0}{\to} \int_C \hat{t} \cdot (\mathbf{H}_2 - \mathbf{H}_1) g \, ds, \tag{A13}$$

where C_1 and C_2 are the curves where S_ρ intersects S. The curve C_1 lies on S_1 and C_2 on S_2. Since the surface integral in (A13) vanishes with ρ, so does the line integral to its right, which completes the proof of Theorem 1.

APPENDIX B

Examples are given here of the violation of the continuity assumption of the divergence theorem. The region of application is the unit cube $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $0 \leqslant z \leqslant 1$. The first example is with respect to the function

$$f(x, y, z) = \begin{cases} x, & 0 \leqslant x < 1, \quad 0 \leqslant y \leqslant 1, \quad 0 \leqslant z \leqslant 1, \\ 0, & x = 1, \quad 0 \leqslant y \leqslant 1, \quad 0 \leqslant z < \frac{1}{2}, \\ \frac{1}{2}, & x = 1, \quad 0 \leqslant y \leqslant 1, \quad \frac{1}{2} < z \leqslant 1. \end{cases} \tag{B1}$$

For the vector $\mathbf{A} = \hat{x} f$,

$$\int_V \nabla \cdot \mathbf{A} \, dV = \int_V \frac{\partial f}{\partial x} dV = 1, \tag{B2}$$

$$\int_S \hat{n} \cdot \mathbf{A} \, dS = -\int_{x=0} \hat{x} \cdot \mathbf{A} \, dS + \int_{x=1} \hat{x} \cdot \mathbf{A} \, dS$$

$$= 0 + \tfrac{1}{2} \int_0^1 \int_{1/2}^1 dz dy = \tfrac{1}{4}. \tag{B3}$$

Equations (B2) and (B3) show that the conclusion of the theorem does not hold; on the other hand, if

$$g(x, y, z) = \begin{cases} x, & 0 \leqslant x < 1, \quad 0 \leqslant y \leqslant 1, \quad 0 \leqslant z \leqslant 1, \\ 0, & x = 1, \quad 0 \leqslant y \leqslant 1, \quad 0 \leqslant z < \frac{1}{2}, \\ 2, & x = 1, \quad 0 \leqslant y \leqslant 1, \quad 1/2 < z \leqslant 1, \end{cases} \tag{B4}$$

then, with $\mathbf{A} = \hat{x} g$,

$$\int_V \nabla\cdot\mathbf{A}\, dV = 1 = \int_S \hat{n}\cdot\mathbf{A}\, dS\,, \tag{B5}$$

which shows that the continuity condition is not a necessary one.

The importance of the continuity condition can be better visualized by considering examples for the one-dimensional analog of the divergence theorem, namely the fundamental theorem of calculus. For example, for the function

$$f(x) = \begin{cases} x, & 0<x<1, \\ 1, & x=0, \\ 2, & x=1, \end{cases}$$

the conclusion of the theorem holds, i.e.,

$$\int_0^1 f'(x)\, dx = f(1) - f(0)\,,$$

but for the functions

$$g(x) = \begin{cases} 0, & x=0 \\ 1, & 0<x\leqslant 1 \end{cases}, \quad h(x) = \begin{cases} x, & 0\leqslant x<1 \\ 0, & x=1 \end{cases}$$

it does not.

APPENDIX C

The limits of (31) and (33)–(35) are computed here. To this end the curve C is split into two parts, the part C_0 of (23) and the rest. The integrals over $C - C_0$ have continuous integrands which in the limit as $\delta\to 0$ tend to zero. Each integral over C_0 is expressed as the sum of two integrals by means of (25) and (26). The integrals involving G are well behaved and tend to zero with δ. What remains are the integrals over C_0 involving the fundamental solution of Laplace's equation and its gradient.

The first limit to be computed is that of the second integral in (31) which is

$$\mathbf{I}_1 = \int_{-\eta}^{\eta} g(y)\left(\frac{\hat{n}_2\times\hat{R}_2}{R_2^2} - \frac{\hat{n}_1\times\hat{R}_1}{R_1^2}\right) dy\,, \tag{C1}$$

where $\hat{n}_1$ and $\hat{n}_2$ are the exterior unit normals on S at the points

$$\mathbf{r}_1 = -\delta\hat{x} + f(-\delta, 0)\hat{z}\,, \quad \mathbf{r}_2 = \delta\hat{x} + f(\delta, 0)\hat{z}\,, \tag{C2}$$

and

$$\begin{gathered} \mathbf{R}_1 = \mathbf{r} - \mathbf{r}_1\,, \quad \mathbf{R}_2 = \mathbf{r} - \mathbf{r}_2\,, \quad R_1 = |\mathbf{r} - \mathbf{r}_1|\,, \\ R_2 = |\mathbf{r} - \mathbf{r}_2|\,, \quad \hat{R}_1 = \mathbf{R}_1/R_1\,, \quad \hat{R}_2 = \mathbf{R}_2/R_2\,, \end{gathered} \tag{C3}$$

with

$$\mathbf{r} = y\hat{y} + f(0, y)\hat{z}\,. \tag{C4}$$

The function g is defined by

$$\begin{aligned} g(y) &= \hat{t}\cdot\mathbf{H}^i[1 + f_y^2(0, y)]^{1/2} \\ &= [\hat{y} + f_y(0, y)\hat{z}]\cdot\mathbf{H}^i = H_y^i(y) + f_y(0, y)H_z^i(y)\,. \end{aligned} \tag{C5}$$

The integral in (C1) can be written in the form

$$\begin{aligned} \mathbf{I}_1 = &\int_{-\eta}^{\eta} [g(y) - g(0)]\left(\frac{\hat{n}_2\times\hat{R}_2}{R_2^2} - \frac{\hat{n}_1\times\hat{R}_1}{R_1^2}\right) dy + g(0) \\ &\times\int_{-\eta}^{\eta}\left(\frac{\hat{n}_2\times\hat{R}_2 - \hat{z}\times\hat{R}_2}{R_2^2} - \frac{\hat{n}_1\times\hat{R}_1 - \hat{z}\times\hat{R}_1}{R_1^2}\right) dy + g(0) \\ &\times\int_{-\eta}^{\eta}\left[\left(\frac{\hat{z}\times\hat{R}_2}{R_2^2} - \frac{\hat{z}\times\hat{\rho}_2}{\rho^2}\right) - \left(\frac{\hat{z}\times\hat{R}_1}{R_2^2} - \frac{\hat{z}\times\hat{\rho}_1}{\rho^2}\right)\right] dy + g(0)\int_{-\eta}^{\eta}\frac{\hat{z}\times(\hat{\rho}_2 - \hat{\rho}_1)}{\rho^2}\, dy\,, \end{aligned} \tag{C6}$$

where

$$\rho = (\delta^2 + y^2)^{1/2}\,, \quad \hat{\rho}_1 = (\delta\hat{y} + y\hat{y})\rho^{-1}\,, \quad \hat{\rho}_2 = (-\delta\hat{x} + y\hat{y})\rho^{-1}\,. \tag{C7}$$

In order to estimate (C6) the following information is needed. With K symbolizing from now on a generic constant,

$$\left|\frac{1}{R_1} - \frac{1}{R_2}\right| = \frac{|R_2^2 - R_1^2|}{R_1R_2(R_1 + R_2)} \leqslant K\frac{\delta^2(\delta^2 + y^2)}{\rho^3} \leqslant K\delta\,, \tag{C8}$$

$$\left|\frac{1}{R_1^3} - \frac{1}{R_2^3}\right| = \left|\frac{1}{R_1} - \frac{1}{R_2}\right|\left|\frac{1}{R_1^2} + \frac{1}{R_1R_2} + \frac{1}{R_2^2}\right| \leqslant K\frac{\delta}{\rho^2}\,. \tag{C9}$$

These inequalities follow from (C2), (C3) and the properties of f as described by (21) and (22). Similarly,[13] if $\hat{n}$ and $\hat{n}'$ are unit normals on the portion of S under consideration, then

$$|\hat{n} - \hat{n}'| \leqslant K|\mathbf{r} - \mathbf{r}'|\,, \tag{C10}$$

where $\mathbf{r}$ and $\mathbf{r}'$ are the position vectors to the normals. Moreover, by the mean value theorem, the z component, n_{1z} and n_{2z}, of $\hat{n}_1$ and $\hat{n}_2$ can be written in the form

$$n_{1z}(-\delta) = 1 + n_{1z}'(-\delta_1)\delta\,, \quad -\delta < -\delta_1 < 0\,, \quad n_{2z}(\delta) = 1 - n_{2z}'(\delta_2)\delta\,, \quad 0<\delta_2<\delta\,, \tag{C11}$$

with the prime denoting differentiation.

By (C5) the first factor of the first integral, $\mathbf{I}_{11}$, of (C6) satisfies the inequality

$$|g(y) - g(0)| \leqslant K|y| \tag{C12}$$

while, from (C8)–(C11), the terms of the second factor are bounded by

$$\frac{|n_{2x}[f(0, y) - f(\delta, 0)]|}{R_2^3} \leqslant \frac{K\delta}{\rho}\,, \quad \frac{|n_{2y}[f(0, y) - f(\delta, 0)]|}{R_2^3} \leqslant \frac{K\delta}{\rho}\,, \quad \frac{|n_{2x}y + n_{2y}\delta|}{R_2^3} \leqslant \frac{K\delta}{\rho}\,, \tag{C13}$$

$$\left|y\left(\frac{n_{2z}}{R_2^3}-\frac{n_{1z}}{R_1^3}\right)\right|\leqslant K\frac{\delta}{\rho^2},\quad \delta\left|\frac{n_{2z}}{R_2^3}+\frac{n_{1z}}{R_1^3}\right|\leqslant K\frac{\delta}{\rho^3}. \tag{C14}$$

The same inequalities hold for the terms involving $\hat{n}_1$ and R_1. This practice of making statements only at the point $\mathbf{r}_2$ (with the understanding that analogous ones hold at $\mathbf{r}_1$) will continue to the end of the Appendix. From the last three statements

$$|\mathbf{I}_{11}|\leqslant K\delta\int_{-\eta}^{\eta}\frac{dy}{\rho^2}=2K\tan^{-1}\left(\frac{\eta}{\delta}\right)\underset{\delta\to 0}{\to}K\pi. \tag{C15}$$

By (C13) and the inequality

$$\left|\frac{y(1-n_{2z})}{R_2^3}\right|=\left|\frac{yn_{2z}'(\delta_2)\delta}{R_2^3}\right|\leqslant K\frac{\delta}{\rho^2} \tag{C16}$$

the second integral in (C6) behaves as the one in (C15). Moreover, since

$$\frac{\hat{z}\times\hat{R}_2}{R_2^2}-\frac{\hat{z}\times\hat{\rho}_2}{\rho^2}=-(y\hat{x}+\delta\hat{y})\left(\frac{1}{R_2^3}-\frac{1}{\rho^3}\right) \tag{C17}$$

and

$$\left|\frac{1}{R_2^3}-\frac{1}{\rho^3}\right|\leqslant\frac{K}{\rho} \tag{C18}$$

the third integral in (C6) also has the same behavior.

The last integral, $\mathbf{I}_{14}$, in (C6) can be computed explicitly,

$$\mathbf{I}_{14}=-\hat{y}2\delta\int_{-\eta}^{\eta}\frac{dy}{(\delta^2+y^2)^{3/2}}=-\hat{y}\frac{4\eta}{\delta(\delta^2+\eta^2)^{1/2}} \tag{C19}$$

so that, together with (C15) and the subsequent discussion,

$$\mathbf{I}_1=(\hat{x}+\hat{z})O(1)+\hat{y}O(\delta^{-1}),\quad \delta\to o^+, \tag{C20}$$

which completes the proof of (36).

The next limit to be computed is the second one in (33) which is

$$\begin{aligned}I_2&=\int_{-\eta}^{\eta}g(y)\left(\frac{\hat{n}_2\cdot\hat{R}_2}{R_2^2}-\frac{\hat{n}_1\cdot\hat{R}_1}{R_1^2}\right)dy\\&=\int_{-\eta}^{\eta}[g(y)-g(0)]\left(\frac{\hat{n}_2\cdot\hat{R}_2}{R_2^2}-\frac{\hat{n}_2\cdot\hat{R}_1}{R_1^2}\right)dy+g(0)\int_{-\eta}^{\eta}\left(\frac{n_{2x}R_{2x}+n_{2y}R_{2y}}{R_2^3}-\frac{n_{1x}R_{1x}+n_{1y}R_{1y}}{R_1^3}\right)dy\\&\quad+g(0)\int_{-\eta}^{\eta}\left(\frac{n_{2z}R_{2z}}{R_2^3}-\frac{n_{1z}R_{1z}}{R_1^3}\right)dy,\end{aligned} \tag{C21}$$

where g is defined in (C5). By writing

$$\begin{aligned}\frac{\hat{n}_2\cdot\hat{R}_2}{R_2^2}-\frac{\hat{n}_1\cdot\hat{R}_1}{R_1^2}&=-\delta\left(\frac{n_{2x}}{R_2^3}+\frac{n_{1x}}{R_1^3}\right)+y\left(\frac{n_{2y}}{R_2^3}-\frac{n_{1y}}{R_1^3}\right)\\&\quad+\frac{n_{2z}[f(0,y)-f(\delta,0)]}{R_2^3}-\frac{n_{1z}[f(0,y)-f(-\delta,0)]}{R_1^3}\end{aligned} \tag{C22}$$

it can be readily shown that each of the first two terms as well as the combination of the last two are bounded, in absolute value, by $K\delta/\rho^2$; hence, the first integral, I_{21}, on the right-hand side of (C21) is bounded by

$$|I_{21}|\leqslant K\delta\int_{-\eta}^{\eta}\frac{dy}{(\delta^2+y^2)^{1/2}}=2K\delta\log\left(\frac{\eta+(\delta^2+\eta^2)^{1/2}}{\delta}\right)\underset{\delta\to 0^+}{\to}0. \tag{C23}$$

The second integral, I_{22}, in (C21) can be written in the form

$$\begin{aligned}I_{22}&=\int_{-\eta}^{\eta}\left[(n_{2x}R_{2x}+n_{2y}R_{2y})\left(\frac{1}{R_2^3}-\frac{1}{\rho^3}\right)-(n_{1x}R_{1x}+n_{1y}R_{1y})\left(\frac{1}{R_1^3}-\frac{1}{\rho^3}\right)\right]dy\\&\quad+\int_{-\eta}^{\eta}\frac{n_{2x}R_{2x}+n_{2y}R_{2y}-n_{1x}R_{1x}-n_{1y}R_{1y}}{\rho^3}dy.\end{aligned} \tag{C24}$$

Since

$$|n_{2x}R_{2x}+n_{2y}R_{2y}|\left|\frac{1}{R_2^3}-\frac{1}{\rho^3}\right|\leqslant K(\delta^2+\delta|y|)\frac{1}{\rho}\leqslant K\delta \tag{C25}$$

the first integral on the right of (C24) is of $O(\delta)$. The second integral, I_{222}, becomes, after substitution of the appropriate components of $\mathbf{R}_1$ and $\mathbf{R}_2$,

$$I_{222} = \int_{-\eta}^{\eta} \frac{-\delta(n_{2x}+n_{1x})+y(n_{2y}-n_{1y})}{\rho^3}\,dy = -\delta(n_{2x}+n_{1x})\int_{-\eta}^{\eta}\frac{dy}{\rho^3}$$
$$= \delta^2\left(\frac{f_{xx}(\delta_2,0)}{[1+f_x^2(\delta,0)+f_y^2(\delta,0)]^{1/2}} - \frac{f_{xx}(-\delta_1,0)}{[1+f_x^2(-\delta,0)+f_y^2(-\delta,0)]^{1/2}}\right)\frac{2\eta}{\delta^2(\eta^2+\delta^2)^{1/2}} \tag{C26}$$

and goes to zero with δ due the continuity property of the second partials of f; thus

$$I_{22} = o(1), \quad \delta\to 0^+. \tag{C27}$$

The last integral, I_{23}, in (C21) can be written in the form

$$I_{23} = \int_{-\eta}^{\eta}\left(\frac{(n_{2z}-1)R_{2z}}{R_2^3} - \frac{(n_{1z}-1)R_{1z}}{R_1^3}\right)dy + \int_{-\eta}^{\eta}\left(\frac{R_{2z}}{R_2^3} - \frac{R_{1z}}{R_1^3}\right)dy \tag{C28}$$

and, since

$$\frac{|(n_{2z}-1)R_{2z}|}{R_2^3} = \frac{\delta|n'_{2z}(\delta_2)[f(0,y)-f(\delta,0)]|}{R_2^3} \leqslant K\frac{\delta}{\rho} \tag{C29}$$

the first integral in (C28) tends to zero with δ as in (C23). Moreover, since

$$\left|\frac{R_{2z}}{R_2^3} - \frac{R_{1z}}{R_1^3}\right| \leqslant K\delta + \left|\frac{f(\delta,0)}{R_2^3} - \frac{f(-\delta,0)}{R_1^3}\right| \tag{C30}$$

and

$$\left|\frac{f(\delta,0)}{R_2^3} - \frac{f(-\delta,0)}{R_1^3}\right| \leqslant K\delta + \frac{|f(\delta,0)-f(-\delta,0)|}{\rho^3} = K\delta + \frac{|f_{xx}(\delta_2,0)-f_{xx}(-\delta_1,0)|}{2\rho^3}, \tag{C31}$$

then the second integral in (C28) behaves as the one in (C26), so that

$$I_{23} = o(1), \quad \delta\to 0^+. \tag{C32}$$

From (C23), (C27), and (C32) it follows that

$$I_2 = o(1), \quad \delta\to 0^+. \tag{C33}$$

The next limit to be computed is the one in (35) which is

$$I_3 = \int_{-\eta}^{\eta} g(y)\left(\frac{\hat{n}_2\cdot\hat{n}}{R_2} - \frac{\hat{n}_1\cdot\hat{n}}{R_1}\right)dy \tag{C34}$$

with

$$g(y) = \hat{q}\cdot\mathbf{H}^i[1+f_y^2(0,y)]^{1/2}. \tag{C35}$$

Rewriting I_3 in the form

$$I_3 = \int_{-\eta}^{\eta} g(y)\left(\frac{n_{2x}n_x+n_{2y}n_y}{R_2} - \frac{n_{1x}n_x+n_{1y}n_y}{R_1}\right)dy + \int_{-\eta}^{\eta} g(y)\left(\frac{n_{2z}n_z}{R_2} - \frac{n_{1z}n_z}{R_1}\right)dy \tag{C36}$$

and noting that

$$\frac{|n_{2x}n_x|}{R_2} \leqslant K\frac{\delta|y|}{\rho} \leqslant K\delta, \quad \frac{|n_{2y}n_y|}{R_2} \leqslant \frac{K\delta|y|}{\rho} \leqslant K\delta, \tag{C37}$$

and

$$\left|\frac{(n_{2z}-1)n_z}{R_2} - \frac{(n_{1z}-1)n_z}{R_1} + n_z\left(\frac{1}{R_2}-\frac{1}{R_1}\right)\right| \leqslant \delta\left|\frac{n'_{2z}(\delta_2)}{R_1} + \frac{n'_{1z}(-\delta_1)}{R_2}\right| + \left|\frac{1}{R_2}-\frac{1}{R_1}\right| \leqslant K\frac{\delta}{\rho}, \tag{C38}$$

it follows that

$$I_3 = o(\delta^{1-\gamma}), \quad \delta\to 0^+, \quad 0<\gamma<1. \tag{C39}$$

The fourth limit to be computed is the first one in (33), which is

$$I_4 = \int_{-\eta}^{\eta} g(y)\left(\frac{\hat{n}_2\cdot\hat{q}}{R_2} - \frac{\hat{n}_1\cdot\hat{q}}{R_1}\right)dy, \tag{C40}$$

where $\hat{q}$ is defined in (32) and

$$g(y) = \hat{n}\cdot\mathbf{E}^i[1+f_y^2(0,y)]^{1/2}. \tag{C41}$$

By writing I_4 in a manner analogous to that in (C36) it readily follows that

$$I_4 = o(\delta^{1-\gamma}), \quad \delta \to 0^+, \quad 0 < \gamma < 1. \tag{C42}$$

The fifth limit is the one in (34) which reads

$$\mathbf{I}_5 = \int_{-\eta}^{\eta} g(y)\left(\frac{\hat{n}_2 \times \hat{n}}{R_2} - \frac{\hat{n}_1 \times \hat{n}}{R_1}\right) dy \tag{C43}$$

with g defined in (C35). By writing this integral in the form

$$\mathbf{I}_5 = \int_{-\eta}^{\eta} g(y)\left\{\frac{\hat{x}n_{2y}n_x - \hat{y}n_{2x}n_z + \hat{z}[n_{2x}n_y - n_{2y}n_x]}{R_2} - \frac{\hat{x}n_{1y}n_z - \hat{y}n_{1x}n_z + \hat{z}[n_{1x}n_y - n_{1y}n_x]}{R_1}\right\} dy$$
$$- \hat{x}\int_{-\eta}^{\eta} g(y)n_y\left(\frac{n_{2z}}{R_2} - \frac{n_{1z}}{R_1}\right) dy + \hat{y}\int_{-\eta}^{\eta} g(y)n_x\left(\frac{n_{2z}}{R_2} - \frac{n_{1z}}{R_1}\right) dy \tag{C44}$$

it can be readily shown that the first integral is of $o(\delta^{1-\gamma})$ while the last two, by steps similar to those of (C38), are of $O(\delta)$; thus,

$$\mathbf{I}_5 = o(\delta^{1-\gamma}), \quad \delta \to 0^+, \quad 0 < \gamma < 1. \tag{C45}$$

The last limit to be computed is the first one in (31) which is

$$\mathbf{I}_6 = \int_{-\eta}^{\eta} g(y)\left(\frac{\hat{n}_2 \times \hat{q}}{R_2} - \frac{\hat{n}_1 \times \hat{q}}{R_1}\right) dy \tag{C46}$$

with g defined in (C41). Since

$$\hat{n}_2 \times \hat{q} = \hat{n}_2 \times (\hat{t} \times \hat{n}) = \hat{t}(\hat{n}_2 \cdot \hat{n}) - \hat{n}(\hat{t} \cdot \hat{n}_2), \tag{C47}$$

then

$$\mathbf{I}_6 = \int_{-\eta}^{\eta} g(y)\hat{t}\left(\frac{\hat{n}_2 \cdot \hat{n}}{R_2} - \frac{\hat{n}_1 \cdot \hat{n}}{R_1}\right) dy - \int_{-\eta}^{\eta} g(y)\hat{n}\left(\frac{\hat{n}_2 \cdot \hat{t}}{R_2} - \frac{\hat{n}_2 \cdot \hat{t}}{R_1}\right) dy. \tag{C48}$$

Both of these integrals can be treated in a manner analogous to that of (C36); thus

$$\mathbf{I}_6 = o(\delta^{1-\gamma}), \quad \delta \to 0^+, \quad 0 < \gamma < 1. \tag{C49}$$

[1]R.G. Kouyoumjian, Proc. IEEE **53**, 864 (1965).
[2]J.J. Bowman, T.B.A. Senior, and P.L.E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North–Holland, Amsterdam, 1969).
[3]C.J. Bouwkamp, Rep. Progr. Phys. **17**, 35 (1954).
[4]S. Silver, J. Opt. Soc. Am. **52**, 131 (1962).
[5]M.I. Sancer, Radio Sci. **3**, 141 (1968).
[6]F. Kottler, Ann. Physik **71**, 457 (1923). For an account of this reference the author relied on Baker and Copson.[7]
[7]B.B. Baker and E.T. Copson, *The Mathematical Theory of Huygens' Principle* (Oxford University, London, 1950), 2nd ed.
[8]J.A. Stratton, *Electromagnetic Theory* (McGraw–Hill, New York, 1941).
[9]D.S. Jones, *The Theory of Electromagnetism* (MacMillan, New York, 1964)
[10]F. Kottler, Progr. Optics **6**, 331 (1967).
[11]J.S. Asvestas, J. Opt. Soc. Am. **65**, 155 (1975).
[12]O.D. Kellogg, *Foundations of Potential Theory* (Dover, New York, 1953).
[13]C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves* (Springer–Verlag, New York, 1969).
[14]J. Van Bladel, *Electromagnetic Fields* (McGraw–Hill, New York, 1964).
[15]S.L. Sobolev, *Partial Differential Equations of Mathematical Physics* (Addison–Wesley, Reading, Mass., 1964).
[16]W. Franz, Z. Naturforschg. **3a**, 500 (1948).
[17]N.M. Günter, *Potential Theory* (Ungar, New York, 1967).
[18]S.A. Schelkunoff, Commun. Pure Appl. Math. **4**, 43 (1951).

The complete symmetry group of the one-dimensional time-dependent harmonic oscillator

P. G. L. Leach
Department of Applied Mathematics, La Trobe University, Bundoora, 3083, Australia

(Received 12 January 1979)

The five invariants for the time-dependent one-dimensional harmonic oscillator Hamiltonian are constructed. Using the linear transformation to the time-dependent oscillator Hamiltonian, the five invariants for the latter are obtained. The differential operators which generate the dynamical symmetry of this Hamiltonian have the same commutator relations as these of the time-independent problem. An additional three operators are obtained using the method of extended Lie groups and have the same properties as those for the time-independent problem. Thus the complete dynamical symmetry of the time-dependent problem is the eight-parameter Lie group SL(3,R).

1. INTRODUCTION

In recent years there has been considerable interest in harmonic oscillator systems, both time independent and time dependent. This interest has expressed itself in several ways, especially in the construction of invariants and the determination of symmetry groups for such systems. Different approaches are found in the literature. To some extent these differences follow from different concepts of what is the basic dynamical expression for the oscillator system—Newton's equation(s) of motion, the Lagrangian, or the Hamiltonian. There are differences in the results obtained. To be more precise consider the one-dimensional time-independent harmonic oscillator with equation of motion, Lagrangian, and Hamiltonian, respectively,

$$\ddot{q} + q = 0, \tag{1.1}$$

$$L = \tfrac{1}{2}\dot{q}^2 - \tfrac{1}{2}q^2, \tag{1.2}$$

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 \tag{1.3}$$

(there is no essential loss of generality in taking the customary ω^2 as unity). The application of the Lie theory of extended groups to Eq. (1.1)[1] showed that the complete symmetry group was the eight-parameter Lie group SL(3,R). Applying Noether's theorem to Eqs. (1.2), Lutzky[2] obtained a five-parameter subgroup of SL(3,R) corresponding to two linear and three quadratic constants of the motion. To obtain the additional three-parameter subgroup, Lutzky used the Langrange's equation of motion. The members of this subgroup do not preserve the invariance of the action integral as does a Noether-derived operator, but they do preserve the invariance of the equation of motion since solutions are transformed into solutions. There does not appear to have been a similar treatment for the Hamiltonian (1.3) although for this problem the extension of the results for the Lagrangian (1.2) is particularly obvious.

In considering the time-dependent one-dimensional harmonic oscillator (we restrict ourselves to one dimension to keep the discussion as simple as possible; the extension to higher dimensions is more a matter of algebraic rather than conceptual difficulty), again there have been different approaches. Defining the problem by

$$\ddot{q} + \omega^2(t)q = 0, \tag{1.4}$$

$$L = \tfrac{1}{2}\dot{q}^2 - \tfrac{1}{2}\omega^2(t)q^2, \tag{1.5}$$

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2(t)q^2, \tag{1.6}$$

Lewis,[3] applying Kruskal's method[4] in closed form, constructed an exact invariant which is equally valid for the Lagrangian (1.5) or the Hamiltonian (1.6). This was

$$I_L = \tfrac{1}{2}\{q^2/\rho^2 - (\rho\dot{q} - \dot{\rho}q)^2\}, \tag{1.7}$$

$$I_H = \tfrac{1}{2}\{q^2/\rho^2 - (\rho p - \dot{\rho}q)^2\}, \tag{1.8}$$

where the suffixes L and H refer to Lagrangian and Hamiltonian formulation, respectively, and $\rho(t)$ is any solution of

$$\ddot{\rho} + \omega^2(t)\rho = \rho^{-3}. \tag{1.9}$$

Leach[5] used the method of time-dependent linear canonical transformations to obtain a form similar to Eq. (1.8) by transforming Eq. (1.6) to (1.3). Lutzky[6] applied Noether's theorem to the Lagrangian (1.5) to obtain Eq. (1.7).

None of the writers mentioned above has provided a discussion of the symmetry group and invariants of the one-dimensional time-dependent harmonic oscillator. In this note we provide such discussion in the context of the Hamiltonian formalism. We start with a simple method for obtaining the five invariants for the Hamiltonian (1.3). Using the linear canonical transformation between Eqs. (1.3) and (1.6) we construct the five invariants of Eq. (1.6) from those of Eq. (1.3). The five corresponding group generators are given. The remaining three operators which leave Newton's equation of motion invariant are also given. It is demonstrated that the operators have the same commutator properties as those for the time-independent problem, hence showing that the time-dependent oscillator also possesses the dynamical symmetry of SL(3,R).

2. THE INVARIANTS OF H(1.3)

All manner of polynomial invariants for H (1.3) may be constructed by postulating a general form for the invariant with undertermined coefficients and then using the requirement that, if I is an invariant of H,

$$\frac{dI}{dt} = [I,H]_{\mathrm{PB}} + \frac{\partial I}{\partial t} = 0. \tag{2.1}$$

 0022-2488/80/020300-05$1.00

Writing H (1.3) as

$$H = \tfrac{1}{2}\mathbf{z}^T\mathbf{z}, \tag{2.2}$$

where

$$\mathbf{z} = \begin{bmatrix} q \\ p \end{bmatrix}, \tag{2.3}$$

a linear invariant, denoted by I_1, has the form

$$I_1 = \mathbf{r}^T\mathbf{z}, \tag{2.4}$$

in which $\mathbf{r}$ is a coordinate free 2-vector.

Since the Poisson bracket of two scalars F and G is given by

$$[F,G]_{PB} = \left(\frac{\partial F}{\partial \mathbf{z}}\right)^T J \left(\frac{\partial G}{\partial \mathbf{z}}\right), \tag{2.5}$$

where J is the 2×2 symplectic matrix

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{2.6}$$

Eq. (2.1) is simply

$$\mathbf{r}^T\mathbf{z} + \dot{\mathbf{r}}^T J\mathbf{z} = 0. \tag{2.7}$$

Making use of the linear independence of the elements of $\mathbf{z}$, Eq. (2.7) reduces to

$$\dot{r}_1 = r_2, \quad \dot{r}_2 = -r_1, \tag{2.8}$$

where r_1 and r_2 are the elements of $\mathbf{r}$. This Hamiltonian system of first order linear differential equations has a singular point at

$$r_1 = 0, \quad r_2 = 0. \tag{2.9}$$

This particular solution is trivial in the present context. The general solution of Eq. (2.8) is

$$r_1 = a\cos t + b\sin t, \tag{2.10a}$$

$$r_2 = -a\sin t + b\cos t, \tag{2.10b}$$

in which a and b are arbitrary constants. The general linear invariant of H (1.3) is

$$I_1 = (a,b)\mathbf{z}\cos t + (b,-a)\mathbf{z}\sin t. \tag{2.11}$$

If we write the row vector (a,b) as $\mathbf{c}^T$, the invariant is

$$I_1 = \mathbf{c}^T(I\cos t - J\sin t)\mathbf{z}, \tag{2.12}$$

i.e., to within a factor of the magnitude of $\mathbf{c}$, I_1 is the magnitude of the projection of the vector

$$\mathbf{I}_1 = (I\cos t - J\sin t)\mathbf{z} \tag{2.13}$$

in the direction of some arbitrary constant vector $\mathbf{c}$. Clearly the basic invariant is $\mathbf{I}_1$, whose two linearly independent elements are

$$I_{11} = q\cos t - p\sin t, \tag{2.14a}$$

$$I_{12} = q\sin t + p\cos t. \tag{2.14b}$$

We observe that the invariant vector $\mathbf{I}_1$ is simply the position vector of the point on the phase plane occupied by the particle at time $t = 0$. The time development of $\mathbf{z}$ is given by

$$\mathbf{z} = (I\cos t + J\sin t)\mathbf{I}_1 \tag{2.15a}$$

$$= (I\cos t + J\sin t)\mathbf{z}(0) \tag{2.15b}$$

which describes a circle of radius $z(0)$ in the phase plane. We note that

$$\frac{\partial I_{12}}{\partial t} = I_{11}, \tag{2.16a}$$

$$\frac{\partial I_{11}}{\partial t} = -I_{12}. \tag{2.16b}$$

This indicates that there are only two symmetry mappings associated with the first order invariants, a result which follows from the discussion given by Katzin and Levine.[7]

We could proceed to construct the quadratic invariants of H (1.3) by postulating the form

$$I_2 = \tfrac{1}{2}\mathbf{z}^T M\mathbf{z}, \tag{2.17}$$

in which M is a 2×2 coordinate free symmetric matrix, and solving the equation corresponding to Eq. (2.17), viz.,

$$\mathbf{z}^T(JM + \tfrac{1}{2}\dot{M})\mathbf{z} = 0. \tag{2.18}$$

For the one-dimensional system being considered here it eventuates that there are only three types of quadratic invariant which are given by the following products of the linear invariants:

$$2I_{21} = I_{11}^2 + I_{12}^2 = p^2 + q^2, \tag{2.19a}$$

$$2I_{22} = 2I_{11}I_{12} = (q^2 - p^2)\sin 2t + 2qp\cos 2t, \tag{2.19b}$$

$$2I_{23} = I_{11}^2 - I_{12}^2 = (q^2 - p^2)\cos 2t - 2qp\sin 2t. \tag{2.19c}$$

The result is not surprising. Any quadratic form in two variables is a linear combination of the three linearly independent expressions given above and no more than three linearly independent quadratic expression can be formed from two linearly independent variables.

The five invariants which have been obtained here correspond to the five invariants derived by Lutzky in Ref. 2 where he used Noether's theorem to obtain the five Lie group operators and hence the invariants. The slight differences are due to the present writer wishing to keep the form of the invariants related to their physical interpretations. Thus, I_{11} and I_{12} are the initial positions (at time $t = 0$) in the phase while I_{21} is the Hamiltonian which in this problem is the conserved energy.

We point out that the method outlined here generalizes easily to multidimensional oscillators and that suitable operators may be obtained for the corresponding quantum mechanical problem.[8] In the latter problem the products are symmetrized, a process which follows naturally if a matrix formulation is used. For the quantum mechanical problem it is convenient to define new operators. They are

$$A^{\pm} = 2^{-1/2}\{I_{11} \pm iI_{12}\}, \tag{2.20a}$$

$$B^{\pm} = I_{23} \pm iI_{22}, \tag{2.20b}$$

$$C = I_{21}. \tag{2.20c}$$

The A's are the time-dependent creation and annihilation operators. We note that in higher dimensional problems there is an additional class of invariants whose elements constitute the angular momentum tensor.

3. TRANSFORMATION TO THE TIME-DEPENDENT OSCILLATOR

A canonical transformation of the Hamiltonian

$$H = \tfrac{1}{2}\mathbf{z}^T A \mathbf{z} \tag{3.1}$$

to the Hamiltonian

$$\bar{H} = \tfrac{1}{2}\bar{\mathbf{z}}^T \bar{A} \bar{\mathbf{z}} \tag{3.2}$$

may be accomplished by the linear transformation

$$\bar{\mathbf{z}} = S\mathbf{z}, \tag{3.3}$$

where (cf. Ref. 5)

$$\dot{S} = J\bar{A}\dot{S} - SJA, \tag{3.4a}$$

$$SJS^T = J. \tag{3.4b}$$

The equation (3.4a) arises from the requirement that the description of the time development of the system be equivalent in either the $\mathbf{z}$ or $\bar{\mathbf{z}}$ coordinate system, i.e.,

$$\dot{\bar{\mathbf{z}}} = J\frac{\partial \bar{H}}{\partial \bar{\mathbf{z}}} = J\bar{A}S\mathbf{z} \tag{3.5a}$$

and

$$\dot{\bar{\mathbf{z}}} = \frac{d(S\mathbf{z})}{dt} = \dot{S}\mathbf{z} + SJA\mathbf{z}. \tag{3.5b}$$

The number of arbitrary constants in the general solution of Eq. (3.4a) is reduced by the requirement (3.4b), which is the condition that the transformation be canonical.

For the particular case of the transformation from H (1.3) to H (1.6) we take H (1.3) as $\bar{H}$. The transformation has been shown to be[5]

$$\mathbf{z} \rightarrow \mathbf{z}:\mathbf{z} = RS\mathbf{z}, \tag{3.6}$$

where

$$R = \begin{bmatrix} C_1 \cos W_1 + C_2 \cos W_2, & -C_1 \sin W_1 - C_2 \sin W_2 \\ C_1 \sin W_1 - C_2 \sin W_2, & C_1 \cos W_1 - C_2 \cos W_2 \end{bmatrix}, \tag{3.7a}$$

$$S = \begin{bmatrix} \rho^{-1} & 0 \\ -\dot{\rho} & \rho \end{bmatrix}, \tag{3.7b}$$

and

$$W_1 = \int_{t_0}^{t} (\rho^{-2} - 1)dt', \quad W_2 = \int_{t_0}^{t} (\rho^{-2} + 1)dt', \tag{3.8a}$$

$$\ddot{\rho} + \omega^2(t)\rho = \rho^{-3}, \tag{3.8b}$$

$$C_1^2 - C_2^2 = 1. \tag{3.8c}$$

4. INVARIANTS FOR THE TIME-DEPENDENT OSCILLATOR HAMILTONIAN

As invariance is independent of the coordinate representation, there are five invariants for H (1.6) which are obtained by expressing the invariant derived in section two in terms of the new coordinates. Using the transformation (3.6) with the specific expressions for R and S given by Eqs. (3.7a) and (3.7b), respectively, we have

$$I_{11} = (\cosh C + \sinh C)\{\rho^{-1} q \cos W - (\rho p - \dot{\rho} q)\sin W\}, \tag{4.1a}$$

$$I_{12} = (\cosh C - \sinh C)\{\rho^{-1} q \sin W + (\rho p - \dot{\rho} q)\cos W\}, \tag{4.1b}$$

$$2I_{21} = \cosh 2C\{\rho^{-2} q^2 + (\rho p - \dot{\rho} q)^2\} + \sinh 2C\{\rho^{-2} q^2 - (\rho p - \dot{\rho} q)^2 \cos 2W - 2\rho^{-1} q(\rho p - \dot{\rho} q)\sin 2W\}, \tag{4.1c}$$

$$2I_{22} = \rho^{-2} q^2 \sin 2W - (\rho p - \dot{\rho} q)^2 \times \sin 2W + 2\rho^{-1} q(\rho p - \dot{\rho} q)\cos 2W, \tag{4.1d}$$

$$2I_{23} = \cosh 2C\{\rho^{-2} q^2 \cos 2W - (\rho p - \dot{\rho} q)^2 \cos 2W - 2\rho^{-1} q(\rho p - \dot{\rho} q)\sin 2C\} + \sinh 2C\{\rho^{-2} q^2 + (\rho p - \dot{\rho} q)^2\}, \tag{4.1e}$$

in which we have written

$$\cosh C = C_1, \quad \sinh C = C_2, \tag{4.2a}$$

$$W = \int_{t_0}^{t} \rho^{-2} dt'. \tag{4.2b}$$

If C_2 is set at zero, I_{21} is the form of the invariant reported by Lewis.[3] The general form given above was reported by Leach.[5b]

We may define a set of invariants independent of C by writing

$$J_{11} = (\cosh C - \sinh C)I_{11} = \rho^{-1} q \cos W - (\rho p - \dot{\rho} q)\sin W, \tag{4.3a}$$

$$J_{12} = (\cosh C + \sinh C)I_{12} = \rho^{-1} q \sin W + (\rho p - \dot{\rho} q)\cos W, \tag{4.3b}$$

$$2J_{21} = 2(I_{21}\cosh 2C - I_{23}\sinh 2C) = \rho^{-2} q^2 (\rho p - \dot{\rho} q)^2, \tag{4.3c}$$

$$2J_{22} = 2I_{22} = \{\rho^{-2} q^2 - (\rho p - \dot{\rho} q)^2\}\sin 2W + 2\rho^{-1} q(\rho p - \dot{\rho} q)\cos 2W, \tag{4.3d}$$

$$2J_{23} = 2(-I_{21}\sinh 2C + I_{23}\cosh 2C) = \{\rho^{-2} q^2 - (\rho p - \dot{\rho} q)^2\}\cos 2W - 2\rho^{-1} q(\rho p - \dot{\rho} q)\sin 2W. \tag{4.3e}$$

We note that the Lewis invariant now occurs without a parameter in Eq. (4.3c).

The physical interpretation of some of these invariants is facilitated if we make use of the intermediate Hamiltonian

$$H' = \tfrac{1}{2}\rho^{-2}(p'^2 + q'^2), \tag{4.4}$$

which is related to H (1.6) by the canonical transformation

$$\begin{bmatrix} q' \\ p' \end{bmatrix} = \begin{bmatrix} \rho^{-1} & 0 \\ -\dot{\rho} & \rho \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix}. \tag{4.5}$$

Then

$$J_{11} = q' \cos W - p' \sin W, \tag{4.6a}$$

$$J_{12} = q' \sin W + p' \cos W, \tag{4.6b}$$

$$J_{21} = \tfrac{1}{2}(q'^2 + p'^2). \tag{4.6c}$$

In the (q', p') phase plane, taking $t_0 = 0$,

$$J_{11} = q'(0), \quad J_{12} = p'(0). \tag{4.7}$$

The motion of the particle along the phase space trajectory is given by

$$\begin{bmatrix} q' \\ p' \end{bmatrix} = \begin{bmatrix} \cos W & \sin W \\ -\sin W & \cos W \end{bmatrix} \mathbf{J}_1, \tag{4.8}$$

where $\mathbf{J}_1$ is defined similarly to $\mathbf{I}_1$. This is a circular motion and J_{21} simply represents the constancy of the radius, being half the square of the radius.

Reverting to Eq. (4.3) we may obtain the formal solution for q by eliminating $\rho p - \dot{\rho} q$ from (4.3a) and (4.3b). This is

$$q = \rho(J_{11}\cos W + J_{12}\sin W), \tag{4.9}$$

which is of the same form as that given in Ref. 6, Eq. (10).

Finally, we note that only two of the five invariants listed are functionally independent. It seems to us to be natural to select J_{11} and J_{12} as the independent quantities. Then

$$2J_{21} = (J_{11})^2 + (J_{12})^2, \tag{4.9a}$$

$$2J_{23} = (J_{11})^2 - (J_{12})^2, \tag{4.9b}$$

$$2J_{22} = 2J_{11}J_{12}. \tag{4.9c}$$

These relations are the same as those which are found in the time-independent case (cf. Ref. 8).

5. OPERATORS OF THE FIVE-PARAMETER SUBGROUP

Now that we have the five invariants for the Hamiltonian of the time-dependent harmonic oscillator it is a simple task to obtain the corresponding differential operators which are the generators of transformations. To facilitate comparison with the time-independent harmonic oscillator as discussed by Lutzky,[2] we adopt the nomenclature used in that paper. To summarize this, a generator G is given by

$$G(q,t) = \xi(q,t)\frac{\partial}{\partial t} + \eta(q,t)\frac{\partial}{\partial q} \tag{5.1}$$

and the corresponding invariant, in the Lagrangian formulation, is

$$\Phi(q,\dot{q},t) = (\xi\dot{q} - \eta)\frac{\partial L}{\partial \dot{q}} - \xi L + f(q,t). \tag{5.2}$$

In this instance

$$L(q,\dot{q},t) = \tfrac{1}{2}\dot{q}^2 - \tfrac{1}{2}\omega^2(t)q^2. \tag{5.3}$$

We may determine ξ, η and f from the invariants given by Eqs. (4.3a)–(4.3c) and we have, replacing p by $\dot{q}$,

$$\Phi_1(q,\dot{q},t) = -2J_{22}(q,\dot{q},t), \tag{5.4a}$$

$$\Phi_2(q,\dot{q},t) = -2J_{23}(q,\dot{q},t), \tag{5.4b}$$

$$\Phi_3(q,\dot{q},t) = -2J_{12}(q,\dot{q},t), \tag{5.4c}$$

$$\Phi_4(q,\dot{q},t) = 2J_{11}(q,\dot{q},t), \tag{5.4d}$$

$$\Phi_5(q,\dot{q},t) = 2J_{21}(q,\dot{q},t). \tag{5.4e}$$

The corresponding operators are

$$G_1 = \sin 2W\frac{\partial}{\partial W} + (\rho\dot{\rho}\sin 2W + \cos 2W)q\frac{\partial}{\partial q}, \tag{5.5a}$$

$$G_2 = \cos 2W\frac{\partial}{\partial W} + (\rho\dot{\rho}\cos 2W - \sin 2W)q\frac{\partial}{\partial q}, \tag{5.5b}$$

$$G_3 = \rho\cos W\frac{\partial}{\partial q}, \tag{5.5c}$$

$$G_4 = \rho\sin W\frac{\partial}{\partial q}, \tag{5.5d}$$

$$G_5 = \frac{\partial}{\partial W} + \rho\dot{\rho}q\frac{\partial}{\partial q}. \tag{5.5e}$$

We note that if ω is constant and equal to one, $\rho^{-2} = \omega = 1$, $W = t$, and the expressions for the generators are the same as those given by Lutzky Ref. 2, Eqs. (24a)–(24e) for the time-independent problem. These operators generate a five-parameter Lie group and have the following commutative relations:

$$[G_1,G_2] = -2G_5, \quad [G_5,G_1] = 2G_2, \quad [G_2,G_5] = 2G_1, \tag{5.6a}$$

$$[G_3,G_4] = 0, \tag{5.6b}$$

$$[G_3,G_1] = [G_2,G_4] = [G_5,G_4] = G_3, \tag{5.6c}$$

$$[G_1,G_4] = [G_3,G_5] = [G_2,G_3] = G_4, \tag{5.6d}$$

which are exactly the same as those for the time-independent problem.

We conclude this section by listing the expressions for ξ, η, and f. The order corresponds to the one to five ordering of the Φ's and G's:

$\xi(q,t)$	$\eta(q,t)$	$f(q,t)$
$\rho^2\sin 2W$	$\frac{\partial}{\partial q}\{\frac{1}{2}q^2(\rho\dot{\rho}\sin 2W + \cos 2W)\}$	$\frac{\partial}{\partial t}\{\frac{1}{2}q^2(\rho\dot{\rho}\sin 2W + \cos 2W)\}$
$\rho^2\cos 2W$	$\frac{\partial}{\partial q}\{\frac{1}{2}q^2(\rho\dot{\rho}\cos 2W - \sin 2W)\}$	$\frac{\partial}{\partial t}\{\frac{1}{2}q^2(\rho\dot{\rho}\cos 2W - \sin 2W)\}$
0	$\frac{\partial}{\partial q}\{q\rho\cos W\}$	$\frac{\partial}{\partial t}\{q\rho\cos W\}$
0	$\frac{\partial}{\partial q}\{q\rho\sin W\}$	$\frac{\partial}{\partial t}\{q\rho\sin W\}$
ρ^2	$\frac{\partial}{\partial q}\{\frac{1}{2}q^2\rho\dot{\rho}\}$	$\frac{\partial}{\partial t}\{\frac{1}{2}q^2\rho\dot{\rho}\}$

(5.7)

It should be noted that the ξ's listed in Eqs. (5.7) are as coefficients of $\partial/\partial t$. In Eq. (5.6), $\partial/\partial W$ is used, the two being related by

$$\frac{\partial}{\partial W} = \rho^2\frac{\partial}{\partial t}. \tag{5.8}$$

The appropriate variable to use in the operators in W and not t because W is now the effective time variable. This is most readily seen from the Hamiltonian H' introduced in Eq. (4.4). Under the time scale transformation from t to W, Hamilton's equations for H' are

$$\frac{dq'}{dW} = p', \quad \frac{dp'}{dW} = -q'. \tag{5.9}$$

6. OPERATORS OF THE THREE-PARAMETER SUBGROUP

The five invariants obtained in Sec. 4 relate to the Hamiltonian. The complete dynamical symmetry of the time-dependent oscillator is made up not only of the five corresponding operators, but also any operators which leave the Newtonian equation of motion invariant. These operators, as well as those given in Sec. 5, may be obtained by using the Lie theory of extended groups. This method is quite adequately described in Wulfman and Wybourne (1.b) and it is not proposed to repeat their working. We shall simply summarize the relevant results.

Suppose that the generator G defined by

$$G = \xi(q,t)\frac{\partial}{\partial t} + \eta(q,t)\frac{\partial}{\partial q} \tag{6.1}$$

is a member of the complete dynamical symmetry group. Its second extension is

$$\begin{aligned} G'' = {} & \xi\frac{\partial}{\partial t} + \eta\frac{\partial}{\partial q} + \{\eta_t + (\eta_q - \xi_t)\dot{q} - \xi_q\dot{q}^2\xi\}\frac{\partial}{\partial \dot{q}} \\ & + \{\eta_{tt} + (2\eta_{qt} - \xi_{tt})\dot{q} \\ & + (\eta_{qq} - 2\xi_{qt})\dot{q}^2 - \xi_{qq}\dot{q}^3 \\ & + (\eta_q - 2\xi_t - 3\dot{q}\xi_q)\ddot{q}\}\frac{\partial}{\partial \ddot{q}}. \end{aligned} \tag{6.2}$$

If the Newtonian equation of motion is

$$N(\ddot{q},\dot{q},q,t) = 0, \tag{6.3}$$

$$G''N = 0 \tag{6.4}$$

since G is a generator of the symmetry group. The requirement that Eq. (6.4) be true whenever Eq. (6.3) is true leads to a set of partial differential equations for ξ and η.

In the case of the time-dependent harmonic oscillator, Newton's equation is

$$\ddot{q} + \omega^2(t)q = 0. \tag{6.5}$$

When G'' acts on this equation and the resulting differential equations are solved, in addition to the five operators already given, we obtain

$$G_6 = q\frac{\partial}{\partial q}, \tag{6.6a}$$

$$G_7 = \rho^{-1}q\sin W\frac{\partial}{\partial W} + (\dot{\rho}\sin W + \rho^{-1}\cos W)q^2\frac{\partial}{\partial q}, \tag{6.6b}$$

$$G_8 = \rho^{-1}q\cos W\frac{\partial}{\partial W} + (\dot{\rho}\cos W - \rho^{-1}\sin W)q^2\frac{\partial}{\partial q}. \tag{6.6c}$$

These three operators form a subgroup with the commutation relations

$$[G_6,G_7] = G_7, \quad [G_6,G_8] = G_8, \quad [G_7,G_8] = 0, \tag{6.7}$$

which are the same as for the time-independent problem.

These three operators also have the same commutator relations with the other five operators as in the time-independent case. They are

$$[G_6,G_1] = [G_6,G_2] = [G_6,G_5] = 0, \tag{6.8a}$$

$$[G_6,G_3] = -G_3, \quad [G_6,G_4] = -G_4, \tag{6.8b}$$

$$[G_7,G_1] = -G_7, \quad [G_7,G_2] = -G_8, \tag{6.8c}$$

$$[G_8,G_1] = G_8, \quad [G_8,G_2] = -G_7, \tag{6.8d}$$

$$[G_7,G_3] = -\tfrac{1}{2}(G_1 + 3G_6),$$

$$[G_8,G_4] = \tfrac{1}{2}(-G_1 + 3G_6), \tag{6.8e}$$

$$[G_7,G_4] = \tfrac{1}{2}(G_2 - G_5),$$

$$[G_8,G_3] = -\tfrac{1}{2}(G_2 + G_5), \tag{6.8f}$$

$$[G_7,G_5] = -G_8 \quad [G_8,G_5] = G_7. \tag{6.8g}$$

Thus we have the result that the complete dynamical symmetry of the time-dependent one-dimensional harmonic oscillator is SL(3,R).

ACKNOWLEDGMENTS

We thank C.J. Eliezer and G. Prince for useful discussions.

[1] R.L. Anderson and S.M. Davison, J. Math. Anal. Appl. **48**, 301–15 (1974); C.E. Wulfman and B.G. Wybourne, J. Phys. A **9**, 507–18 (1976).
[2] M. Lutzky, J. Phys. A **11**, 249–58 (1978).
[3] H.R. Lewis Jr., Phys. Rev. Lett. **18**, 510–12 (1967); Phys. Rev. **172**, 1313–315 (1968); J. Math. Phys. **9**, 1976–986 (1968).
[4] M. Kruskal, J. Math. Phys. **3**, 806–28 (1962).
[5] P.G.L. Leach, J. Math. Phys. **18**, 1608–11 (1977); J. Appl. Math. **34**, 496–503 (1978).
[6] M. Lutzky, Phys. Lett. A **68**, 3–4 (1978).
[7] G.H. Katzin and J. Levine, J. Math. Phys. **18**, 1267–74 (1977).
[8] P.G.L. Leach, "The Invariants of Quadratic Hamiltonians, I: Linear and Quadratic Invariants for the Time-Independent Isotropic Harmonic Oscillator," Preprint, Department of Applied Mathematics, La Trobe University, 1978.

On the infrared problem in nonrelativistic quantum electrodynamics

W. F. Wreszinski
Instituto de Física, Universidade de São Paulo, São Paulo

(Received 22 June 1979; accepted for publication 11 October 1979)

Following a suggestion of Hepp and Lieb, it is shown rigorously that the infrared divergences which occur in the problem of thermodynamic stability of a system of atoms (with a finite number of levels) interacting with an ultraviolet-cutoff quantized radiation field are entirely removed upon inclusion of the term in $\mathbf{A}^2$ in the Hamiltonian, in the special case where the atoms are placed at the points of a regular lattice.

Since the appearance of the work of Hepp and Lieb,[1,2] several papers have studied the system of atoms or molecules with a finite number of levels interacting with the quantized radiation field (with an ultraviolet cutoff) from the point of view of statistical mechanics. (See, for instance, Ref. 3 and references given there.) In particular, the influence of the approximations involved in the various studied caricatures of the above mentioned system on the existence or nonexistence of the superradiant phase transition (first found in Ref. 2) was considered in greatest generality in Ref. 3. There it was found that a necessary condition for a phase transition is the removal of either the dipole approximation, or of the limitation to a finite number of modes of the field, or of both.

It seems thus important to study the system without any approximation and, in particular, to take into account the infinite number of modes of the radiation field. The most important result so far in this direction was the proof of thermodynamic stability by Hepp and Lieb in Ref. 2. In their paper they suggested that the infrared divergences which arose in treating this question (and which they ignored, by extra assumptions) would be mitigated by taking into account the term $\mathbf{A}^2$ in the Hamiltonian, whose influence on the thermodynamics has, incidently, been studied by several authors ever since.[4-7] Their suggestion was taken up by Rzazewski and Wodkiewicz in Ref. 8, but their treatment was heuristic in one essential aspect: They ignored the proportionality of the number of modes of the field to the volume, in the thermodynamic limit.[9] In this note, we prove rigorously that the inclusion of the term $\mathbf{A}^2$ not only mitigates but also eliminates the infrared divergences entirely, in the special case where the atoms or molecules are placed at the points of a (regular) lattice.

Let K and L be integers, Λ be a parallelepiped in the three-dimensional cubic lattice $\mathbb{Z}^3$, given by

$$\Lambda \equiv \{\mathbf{n}\in\mathbb{Z}^3:\ 0\leqslant n_i\leqslant L-1,\quad i=1,2,3\}, \tag{1}$$

and Λ^* be the set

$$\Lambda^* \equiv \{\mathbf{K}: K_j=(2\pi/L)n_j:$$
$$n_j=-(KL/2)+1,\ldots,KL/2\quad (KL \text{ even}),$$
$$\text{or } n_j=-\tfrac{1}{2}(KL-1),\ldots,\tfrac{1}{2}(KL-1)\ (KL \text{ odd}); j=1,2,3\}. \tag{2}$$

If $K=1$, Λ^* is the lattice dual to Λ. K will play the role of an ultraviolet cutoff. We shall need the relation

$$\frac{1}{|\Lambda|}\sum_{\mathbf{n}\in\Lambda} e^{i\mathbf{K}\cdot\mathbf{n}}=\delta_{\mathbf{K},0},\quad \forall K\in\Lambda^*. \tag{3}$$

This is proved in the same way as the analogous relation for $K=1$.

Consider a system of N atoms or molecules (which we take, for simplicity of notation, to be two-level and identical) in a cubical box Λ [for the moment, not necessarily of the form of Eq. (1)] of volume $V=L^3$, interacting with an ultraviolet-cutoff quantized radiation field. The Hamiltonian may be written

$$H_\Lambda(\Lambda')=\widetilde{H}_\Lambda(\Lambda')+U+T, \tag{4}$$

where T is the N-particle kinetic energy operator, and $U=U(\mathbf{x}_1,\ldots,\mathbf{x}_N)$ is the operator of interaction between the atoms (assumed to satisfy conditions such as those mentioned in Ref. 2.) $\widetilde{H}_\Lambda(\Lambda')$ describes the remaining part of the system, with Λ' denoting the set of modes of the field occurring in the formal Fourier expansion of the vector potential. The latter is not defined at $\mathbf{K}=\mathbf{0}$, hence Λ' must exclude this point, and we shall take

$$\Lambda'=\Lambda^*-\{\mathbf{0}\}, \tag{5}$$

where Λ^* is the set Eq. (2). Under this condition, the following facts are a direct consequence of the methods of Ref. 2. $H_\Lambda(\Lambda')$ is self-adjoint and bounded below on the Hilbert space

$$\mathscr{H}_\Lambda=\left(\bigotimes_{i=1}^{N}\mathbb{C}_i^2\right)\otimes\mathscr{F}_{\Lambda'}\otimes\mathscr{G}_\Lambda,$$

where $\mathscr{F}_{\Lambda'}$ denotes Fock space for the (finite) set of modes in Λ', and $\mathscr{G}_\Lambda=L^2(\Lambda^N)$. The partition function

$$Z_\Lambda(\Lambda')\equiv \mathrm{tr}_{\mathscr{H}_\Lambda} e^{-\beta H_\Lambda(\Lambda')}, \tag{6}$$

may be shown to exist and satisfy the bound

$$Z_\Lambda(\Lambda')\leqslant\widetilde{Z}_\Lambda(\Lambda'), \tag{7}$$

where

$$\widetilde{Z}_\Lambda(\Lambda')\equiv 2^N \mathrm{tr}_{\mathscr{F}_\Lambda}\,\mathrm{tr}_{\mathscr{G}_\Lambda}\int d\Omega^N \exp[-\beta H_\Lambda(\Lambda',\Omega^N)]. \tag{8}$$

above Ω^N is the N-fold cartesian product of copies of the unit sphere in $\mathbb{R}^3$, with

$$d\Omega^N=\prod_{\mathbf{n}\in\Lambda} d\Omega^{\mathbf{n}},$$
$$d\Omega^{\mathbf{n}}=\sin\theta^{\mathbf{n}}\,d\theta^{\mathbf{n}}\,d\varphi^{\mathbf{n}},\quad \mathbf{n}\in\Lambda, \tag{9}$$

 0022-2488/80/020305-03$1.00

$$H_\Lambda(\Lambda',\Omega^N) = \tilde{H}_\Lambda(\Lambda',\Omega^N) + U + T,$$

where U and T are the same operators described before, and $\tilde{H}_\Lambda(\Lambda',\Omega^N)$

$$\equiv \sum_{\mathbf{K},\mathbf{K}'\in\Lambda'} \frac{1}{(\nu_\mathbf{K}\nu_{\mathbf{K}'})^{1/2}} \{[\nu_\mathbf{K}\,\delta_{\mathbf{K},\mathbf{K}'} + 2q^*_{\mathbf{K}',\mathbf{K}}] \times a^*_{\mathbf{K}'}a_\mathbf{K} + p_{\mathbf{K},\mathbf{K}'}a_{\mathbf{K}'}a_\mathbf{K} + p^*_{\mathbf{K},\mathbf{K}'}a^*_\mathbf{K}a^*_{\mathbf{K}'}\}$$
$$+ \sum_{\mathbf{K}\in\Lambda'} \frac{1}{(\nu_\mathbf{K})^{1/2}} (r_\mathbf{K}a_\mathbf{K} + r^*_\mathbf{K}a^*_\mathbf{K})$$
$$+ \epsilon \sum_{\mathbf{n}\in\Lambda} s^z_\mathbf{n} + \sum_{\mathbf{K}\in\Lambda'} \frac{1}{\nu_\mathbf{K}} q_{\mathbf{K},\mathbf{K}}\,. \tag{10}$$

In (10),

$$r_\mathbf{K} = r^*_{-\mathbf{K}} = \frac{1}{N^{1/2}} \sum_{\mathbf{n}\in\Lambda} [\lambda_{\mathbf{K},\mathbf{n}} s^+_\mathbf{n} + \mu_{\mathbf{K},\mathbf{n}} s^-_\mathbf{n}], \tag{11}$$

$$\lambda_{\mathbf{K},\mathbf{n}} = \mu^*_{-\mathbf{K},\mathbf{n}} = \frac{\rho^{1/2}e}{2^{1/2}m} \int_{\mathbb{R}^3} \psi^*_u(\mathbf{x}) \times (\boldsymbol{\epsilon}^{(y)}_\mathbf{K}\cdot\mathbf{p})\psi_d(\mathbf{x})\, e^{i\mathbf{K}\cdot\mathbf{x}}\, d^3x. \tag{12}$$

In the above formulas, $\rho = N/V$ denotes the density, $\epsilon > 0$ is the energy difference between the "up" and "down" levels, ψ_u and ψ_d, respectively, are the corresponding wavefunctions, $a_\mathbf{K}$ and $a^*_\mathbf{K}$ are the usual annihilation and creation operators, satisfying $[a_\mathbf{K}, a^*_{\mathbf{K}'}] = \delta_{\mathbf{K},\mathbf{K}'}$, and $\nu_\mathbf{K} = |\mathbf{K}|$ is the photon energy. Further,

$$s^z_\mathbf{n} = \tfrac{1}{2}\cos\theta_\mathbf{n}, \quad s^\pm_\mathbf{n} = \tfrac{1}{2}\sin\theta_\mathbf{n} e^{\pm i\varphi_\mathbf{n}},$$

$$p_{\mathbf{K},\mathbf{K}'} \equiv \frac{e^2\rho}{4m}\frac{1}{N}\sum_{\mathbf{n}\in\Lambda} e^{i(\mathbf{K}+\mathbf{K}')\cdot\mathbf{n}} \boldsymbol{\epsilon}^{(y)}_\mathbf{K}\cdot\boldsymbol{\epsilon}^{(y)}_{\mathbf{K}'}, \tag{13a}$$

$$q_{\mathbf{K},\mathbf{K}'} \equiv \frac{e^2\rho}{4m}\frac{1}{N}\sum_{\mathbf{n}\in\Lambda} e^{i(\mathbf{K}-\mathbf{K}')\cdot\mathbf{n}} \boldsymbol{\epsilon}^{(y)}_\mathbf{K}\cdot\boldsymbol{\epsilon}^{(y)}_{\mathbf{K}'}. \tag{13b}$$

We have taken the vector potential polarized along the y axis, for subsequent simplicity. Equation (10) differs from the Hamiltonian considered in Ref. 2 just by addition of the diamagnetic term, proportional to $\mathbf{A}^2$.

We now observe that, if we place the atoms at the points of a (regular) lattice, that is, take for Λ the set (1), it follows from Eq. (3) that the quantities p and q in Eqs. (13a) and (13b) above simplify to

$$p_{\mathbf{K},\mathbf{K}'} = (e^2/4m)\delta_{\mathbf{K},-\mathbf{K}'}, \tag{14a}$$

$$q_{\mathbf{K},\mathbf{K}'} = (e^2/4m)\delta_{\mathbf{K},\mathbf{K}'}. \tag{14b}$$

It is this special case that we shall treat subsequently. Note that $\rho = 1$ and

$$\boldsymbol{\epsilon}^{(y)}_\mathbf{K}\cdot\boldsymbol{\epsilon}^{(y)}_{+\mathbf{K}} = 1.$$

(other cases involve other factors which, of course, do not alter the final result.)

Hamiltonian (10) reduces to ($N = |\Lambda| = L^3$)

$\tilde{H}_\Lambda(\Lambda',\Omega^{|\Lambda|})$

$$= \sum_{\mathbf{K}\in\Lambda'} \nu_\mathbf{K}\, a^*_\mathbf{K}a_\mathbf{K} + \epsilon\sum_{\mathbf{n}\in\Lambda} s^z_\mathbf{n}$$
$$+ \sum_{\mathbf{K}\in\Lambda'} \frac{1}{(\nu_\mathbf{K})^{1/2}} (r_\mathbf{K}\,a_\mathbf{K} + r^*_\mathbf{K}\,a^*_\mathbf{K})$$
$$+ \eta \sum_{\mathbf{K}\in\Lambda'} \frac{1}{\nu_\mathbf{K}} (a_\mathbf{K}\,a_{-\mathbf{K}} + a^*_\mathbf{K}\,a^*_{-\mathbf{K}} + 2a^*_\mathbf{K}\,a_\mathbf{K}) + \eta \sum_{\mathbf{K}\in\Lambda'} \frac{1}{\nu_\mathbf{K}}, \tag{15}$$

where $\eta \equiv e^2/4m$. The above Hamiltonian may be diagonalized by a unitary Bogoliubov-type transformation:

$$b_\mathbf{K} = \alpha_\mathbf{K}a_\mathbf{K} + \beta_\mathbf{K}\,a^*_{-\mathbf{K}} + \gamma_\mathbf{K}\,. \tag{16}$$

Condition $[b_\mathbf{K}, b^*_{\mathbf{K}'}] = \delta_{\mathbf{K},\mathbf{K}'}$ implies

$$|\alpha_\mathbf{K}|^2 - |\beta_\mathbf{K}|^2 = 1, \quad \forall\mathbf{K}\in\Lambda'. \tag{17}$$

We shall choose the phase factors for each $\mathbf{K}$ such as to render $\alpha_\mathbf{K}, \beta_\mathbf{K}$, and $\gamma_\mathbf{K}$ real.

We now require

$$\tilde{H}_\Lambda(\Lambda',\Omega^{|\Lambda'|}) = \sum_\mathbf{K} (\omega_\mathbf{K} b^*_\mathbf{K} b_\mathbf{K} + \rho_\mathbf{K}), \tag{18}$$

with $\tilde{H}$ given by Eq. (15), and $\omega_\mathbf{K} \geqslant 0, \rho_\mathbf{K}$ real, for all $\mathbf{K}$. From Eqs. (18) and (15), we obtain the following conditions:

$$\alpha_\mathbf{K} = \alpha_{-\mathbf{K}}, \quad \beta_\mathbf{K} = \beta_{-\mathbf{K}}, \quad \gamma_\mathbf{K} = \gamma_{-\mathbf{K}},$$

$$\omega_\mathbf{K}\gamma_\mathbf{K}(\alpha_\mathbf{K} + \beta_\mathbf{K}) = (1/\nu_\mathbf{K}^{1/2})r_\mathbf{K}, \tag{19}$$

$$\omega_\mathbf{K}(\alpha^2_\mathbf{K} + \beta^2_\mathbf{K}) = \nu_\mathbf{K} + (2\eta/\nu_\mathbf{K}), \tag{20}$$

$$\omega_\mathbf{K}\alpha_\mathbf{K}\beta_\mathbf{K} = \eta/\nu_\mathbf{K}, \tag{21}$$

$$\omega_\mathbf{K}(\beta^2_\mathbf{K} + \gamma^2_\mathbf{K}) + \rho_\mathbf{K} = \eta/\nu_\mathbf{K}. \tag{22}$$

From Eqs. (17), (20), and (21) we obtain $\omega_\mathbf{K}$, $\alpha_\mathbf{K}$, and $\beta_\mathbf{K}$:

$$\omega_\mathbf{K} = (\nu^2_\mathbf{K} + 4\eta)^{1/2},$$

$$\alpha^2_\mathbf{K} = \frac{1}{2}\left(\frac{\nu_\mathbf{K} + 2\eta/\nu_\mathbf{K}}{\omega_\mathbf{K}} + 1\right),$$

$$\beta^2_\mathbf{K} = \frac{1}{2}\left(\frac{\nu_\mathbf{K} + 2\eta/\nu_\mathbf{K}}{\omega_\mathbf{K}} - 1\right).$$

From Eq. (19) we obtain

$$\gamma_\mathbf{K} = \frac{r_\mathbf{K}}{(\nu_\mathbf{K})^{1/2}\omega_\mathbf{K}(\alpha_\mathbf{K} + \beta_\mathbf{K})}$$

whence

$$\rho_\mathbf{K} = \frac{\eta}{\nu_\mathbf{K}} - \frac{\omega_\mathbf{K}}{2}\left(\frac{\nu_\mathbf{K} + 2\eta/\nu_\mathbf{K}}{\omega_\mathbf{K}} - 1\right) - \omega_\mathbf{K}\frac{|r_\mathbf{K}|^2}{\omega_\mathbf{K}(\nu^2_\mathbf{K} + 4\eta)}$$
$$= \tfrac{1}{2}(\omega_\mathbf{K} - \nu_\mathbf{K}) - \frac{|r_\mathbf{K}|^2}{\nu^2_\mathbf{K} + 4\eta}. \tag{23}$$

Cancellation of the term $\eta/\nu_\mathbf{K}$ in the last formula is crucial.

The finite volume partition function Z_Λ is defined as the limit (which we shall *assume* to exist)

$$Z_\Lambda \equiv \lim_{\Lambda'\to\Lambda^*} Z_\Lambda(\Lambda'), \tag{24}$$

where Λ^* is the set (2). By Eq. (5), "limit" above means just adjunction of the point $\{\mathbf{0}\}$, and the above existence assumption means that we suppose $Z_\Lambda(\Lambda')$ to be defined for $\Lambda' = \Lambda^*$, although formal expressions occurring in it are not.

Proposition: There exists a constant $C > 0$ such that

$$Z_\Lambda \leqslant \exp[C\,|\Lambda\,|]. \tag{25}$$

Proof: Let $E_\Lambda(\Lambda')$ denote the ground state energy of $\widetilde{H}_\Lambda(\Lambda', \Omega^{|\Lambda|})$. By Eqs. (23) and (18),

$$E_\Lambda(\Lambda') = \tfrac{1}{2}\sum_{\mathbf{K}\in\Lambda'}(\omega_{\mathbf{K}} - \nu_{\mathbf{K}}) - \sum_{\mathbf{K}\in\Lambda'}\frac{|r_{\mathbf{K}}|^2}{\nu_{\mathbf{K}}^2 + 4\eta}. \tag{26}$$

It follows from Eqs. (7), (24), and a simple argument. [Actually $\widetilde{H}_\Lambda$ should be as in Eq. (15) but with a factor of $\frac{1}{2}$ multiplying the photon energy term in order that the argument be valid, but this could have been done without affecting our conclusions (Ref. 2, p. 2521).], that (25) would result from the following lower bound:

$$\lim_{\Lambda'\to\Lambda^*} E_\Lambda(\Lambda') \geqslant -D\,|\Lambda\,|, \quad \text{for some } D \geqslant 0. \tag{27}$$

Now, by Eq. (26)

$$\lim_{\Lambda'\to\Lambda^*} E_\Lambda(\Lambda') = \tfrac{1}{2}\sum_{\mathbf{K}\in\Lambda^*}(\omega_{\mathbf{K}} - \nu_{\mathbf{K}}) - \sum_{\mathbf{K}\in\Lambda^*}\frac{|r_{\mathbf{K}}|^2}{\nu_{\mathbf{K}}^2 + 4\eta}. \tag{28}$$

The first term at the rhs of Eq. (28) satisfies Eq. (27) with $D = 0$. The second term may be treated exactly as in Ref. 2, leading to a proof of Eq. (27). The arguments of Ref. 2 are somewhat simplified here, because the analogs of the functions h_i in Ref. 2, p. 2520, decay in the present case faster than any power of $|\mathbf{K}|$ in $\mathbf{K}$-space. ■

We note, finally, that the transformation from $a_{\mathbf{K}}$ to $b_{\mathbf{K}}$ [given by Eq. (16)] is implemented by an operator which is unitary on $\mathcal{F}_{\Lambda'}$, but which ceases to exist on the full Fock space, because of vacuum polarization.[10] For an analysis of the infrared problem in nonrelativistic quantum electrodynamics from the point of view of quantum field theory, see Ref. 11.

[1] K. Hepp and E.H. Lieb, Ann. Phys. **76**, 360 (1973).
[2] K. Hepp and E.H. Lieb, Phys. Rev. A **8**, 2517 (1973).
[3] I. Bialynicki-Birula and K. Rzazewski, "No-go theorem concerning the superradiant phase transition in atomic systems," Pittsburgh preprint 188.
[4] K. Rzazewski, K. Wodkiewicz, and W. Zakowicz, Phys. Rev. Lett. **35**, 432 (1975).
[5] K. Rzazewski, K. Wodkiewicz, and W. Zakowicz, Phys. Lett. A **58**, 211 (1976).
[6] R. Gilmore and C. Bowden, J. Math. Phys. **17**, 1617 (1976).
[7] A.A.S. Brito and W.F. Wreszinski, Helv. Phys. Acta **50**, 373 (1977).
[8] K. Rzazewski and K. Wodkiewicz, Phys. Rev. A **13**, 1967 (1976).
[9] I should like to thank Dr. K. Rzazewski for an instructive and pleasant talk in Zurich about the subject.
[10] See, e.g., A.S. Wightman, "Introduction to some aspects of the relativistic dynamics of quantized fields," in *High Energy Electromagnetic Interactions and Field Theory*, edited by M. Levy (Gordon and Breach, New York, 1967).
[11] B. Schroer, Fortschr. Phys. **11**, 1 (1963).

Two sided estimates of the scattering amplitude for low energy

A. G. Ramm
Institute for Theoretical Physics, University of Vienna
and Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109 [a]

(Received 18 July 1979; accepted for publication 26 October 1979)

Estimates and iterative processes for calculating the scattering amplitude for $k = 0$ are given.

1. INTRODUCTION

Consider the problem

$$(H - k^2)\Psi = [-\nabla^2 + V(x) - k^2]\Psi = 0,$$

$$x \in \mathbb{R}^3, \quad \Psi = \exp[ik(n,x)] + v, \tag{1}$$

$$v \sim [\exp(ik|x|)/|x|] f(n,\nu,k), \quad \text{as } |x| \to \infty, \quad \nu = x/|x|. \tag{2}$$

Our main assumptions are ($\int \equiv \int_{\mathbb{R}^3}$):

$$V(x) = 0 \text{ for } |x| > a, \quad \int |V(x)|\,dx < \infty, \tag{3}$$

$$ka \ll 1, \quad H > 0, \quad V(x) \in L^2_{\text{loc}}.$$

The problem consists in estimating the scattering amplitude $f(n,\nu,k)$. The contents of this paper can be summarized as follows:

(1) Two sided estimates for f are obtained;

(2) Iterative process to calculate f is given; the process converges as a geometrical progression:

(3)Hard-core potential is considered;

(4) Some qualitative properties of f are described;

(5) Explicit formula for one-dimensional scattering is given.

2. PRELIMINARIES

We start with the equation

$$\Psi(x) = \exp[ik(n,x)] - \int \frac{\exp(ik|x-y|)}{4\pi|x-y|} V(y)\Psi(y)\,dy. \tag{4}$$

If $ka \ll 1$ we can write this equation as

$$\Psi(x) = 1 - \int \frac{V(y)\Psi(y)}{4\pi|x-y|}\,dy, \tag{5}$$

the error being $O(ka)$. With the same accuracy we get

$$f(n,\nu,k) \equiv f = -(1/4\pi)\int \exp[-ik(\nu,y)]V(y)\Psi(y)\,dy = -(1/4\pi)\int V\Psi\,dy. \tag{6}$$

Equation (5) is equivalent to the problem

$$H\Psi = -\nabla^2\Psi + V(x)\Psi = 0 \quad \text{in} \quad \mathbb{R}^3, \quad \Psi(\infty) = 1. \tag{7}$$

Let $\Psi = 1 + \varphi$. Then

$$H\varphi = -\nabla^2\varphi + V\varphi = -V, \quad \varphi(\infty) = 0, \tag{8}$$

$$f = -\frac{1}{4\pi}\int V\,dx - \frac{1}{4\pi}\int V\varphi\,dy. \tag{9}$$

In what follows we make use of the following theorem which was formulated and proved in Ref. 1.

Theorem 1: Let A be a linear self-adjoint operator on a Hilbert space H, $D(A) = \text{dom } A$, $R(A) = \text{range } A$, $f \in R(A)$, $A\varphi = f$. Then the representation

$$(A\varphi,\varphi) = \max_{g \in D(A)} \frac{|(g,f)|^2}{(Ag,g)}, \tag{10}$$

holds if and only if $A \geqslant 0$, i.e., $(Ag,g) \geqslant 0$, $\forall g \in H$. For $(Ag,g) = 0$ we define the expression under the sign max as zero.

Remark 1: The "if" part is known, the "only if" part is less obvious. The theorem gives a necessary and sufficient condition for the Schwinger stationary principle to be an extremal principle.

3. POSITIVE POTENTIAL

If $V \geqslant 0$ we derive from equations (5) the following equation

$$(I + B)h = V(x)^{1/2},$$

$$Bh = \int \frac{V(x)^{1/2}V(y)^{1/2}}{4\pi|x-y|} h(y)\,dy, \tag{11}$$

where $B \geqslant 0$ in $H = L^2(\mathbb{R}^3)$. From (11) and Theorem 1 it follows that

$$-4\pi f = \int V^{1/2}(x)h(x)\,dx = \max_{g \in H} \frac{(V^{1/2},g)^2}{(g + Bg,g)}, \tag{12}$$

where (h,g) denotes the inner product in H. Hence

$$f \leqslant -\frac{1}{4\pi}\frac{|(V^{1/2},g)|^2}{(g + Bg,g)}, \quad \forall g \in H. \tag{13}$$

To obtain a lower bound for f we apply (10) to Eq. (8) and use inequality $H > 0$. As a result we get

$$(-V,\varphi) = \max_{g \in D(H)} \frac{|(g,V)|^2}{(Hg,g)}. \tag{14}$$

From (14) it follows that

$$f \geqslant -\frac{1}{4\pi}\int V\,dx + \frac{|(g,V)|^2}{4\pi(Hg,g)}, \quad \forall g \in D(H), \quad H = -\nabla^2 + V(x). \tag{15}$$

[a] Address after June 1, 1979: Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109.

 0022-2488/80/020308-03$1.00

Remark 2: Bound (15) was obtained without assumption $V(x)\geqslant 0$. We used only assumption $H>0$.

Another lower bound can be obtained so: Consider the functional

$$E(g)=\int (|\nabla g|^2+Vg^2+2Vg)\,dx, \tag{16}$$

defined on $W_2^1(\mathbb{R}^3)$. Equation (8) is the necessary condition for the minimum of functional (16). Since functional (16) is quadratic, Eq. (8) is also the sufficient condition for minimum of this functional. Hence

$$E(g)\geqslant E(\varphi)=\int \varphi V\,dx=-4\pi f-\int V\,dx. \tag{17}$$

Here we took into account formula (g) and the identity

$$\int (|\nabla\varphi|^2+V\varphi^2+V\varphi)\,dx=0, \quad \text{[see (8)]}. \tag{18}$$

From (17) we get

$$f\geqslant -\frac{1}{4\pi}\int V\,dx-\frac{1}{4\pi}\int (|\nabla g|^2+Vg^2+2Vg)\,dx, \quad \forall g\in W_2^1(\mathbb{R}^3). \tag{19}$$

This inequality is similar to the estimation of electrical capacity.[2] In Ref. 3 an inequality similar to (13) can be found.

4. ITERATIVE PROCESS TO CALCULATE f

Theorem 2: Equation (11) can be solved by iterative process

$$h_{n+1}=\gamma h_n-qBh_n+qV^{1/2}(x), \quad h_0=qV^{1/2}(x), \tag{20}$$

where

$$\gamma=\|B\|/(2+\|B\|), \quad q=2/(2+\|B\|), \tag{21}$$

$\|B\|$ is the norm of the operator B: $L^2(D)\to L^2(D)$, $D=\operatorname{supp} V(x)$. If $h=\lim_{n\to\infty} h_n$, then $\|h-h_n\|=O(\gamma^n)$.

Corollary: We can calculate f from the formula

$$f=-\frac{1}{4\pi}\lim_{n\to\infty}\int V^{1/2}(x)h_n(x)\,dx. \tag{22}$$

Remark 3: It is clear that

$$\|B\|^2\leqslant\int\int [V(x)V(y)\,dx\,dy]/(16\pi^2|x-y|^2), \quad \text{so } \|B\|\leqslant(\|V\|_R)/(4\pi),$$

where $\|V\|_R$ is the Rollnik norm of $V(x)$.

To prove Theorem 2 we denote $I+B=A$, $V^{1/2}(x)=\omega(x)$, so that $Ah=\omega$, $h=(I-qA)h+q\omega$. If we take q as in (21) and set $\gamma=1-q$, we can see that $I-qA\leqslant\max_{1<\lambda<1+\|B\|}(1-q\lambda)\leqslant\|B\|/(2+\|B\|)=\gamma$. Hence process (20) converges as a geometrical progression with the denominator γ. Such an iterative process can be found, for example, in Ref. 4.

5. ONE-DIMENSIONAL SCATTERING PROBLEM

If $x\in\mathbb{R}^1$, we get the following formulas instead of (5) and (6):

$$\Psi=1+\int\frac{V(y)\Psi(y)\,dy}{2ik}, \quad f=\frac{1}{2ik}\int V\Psi\,dy. \tag{23}$$

Multiplying by V and integrating we find

$$F=\frac{\int V\,dx}{2ik-\int V\,dx}, \quad x\in\mathbb{R}^1. \tag{24}$$

6. HARD-CORE POTENTIAL

If $V(x)=\infty$ in a bounded domain $D\in\mathbb{R}^3$ with a smooth boundary Γ, $a=\operatorname{diam} D$, $ka\ll 1$, then the solution to problems (1) and (2) can be found in the form

$$v=\int_\Gamma\frac{\exp(ik|x-s|)}{4\pi|x-s|}\sigma(s)\,ds, \tag{25}$$

$$v|_\Gamma=-\exp[ik(n,s)]|_{ka\ll 1}=-1,$$

$$f=\frac{1}{4\pi}\int_\Gamma\exp[-ik(\nu,s)]\sigma(s)\,ds\Big|_{ka\ll 1}=\frac{1}{4\pi}\int_\Gamma\sigma(s)\,ds. \tag{26}$$

From (25) we get

$$\int_\Gamma\frac{\sigma(s)\,ds}{4\pi|t-s|}=-1. \tag{27}$$

Hence $\int_\Gamma\sigma(s)=-C$, where C is the electrical capacity of the conductor D. Therefore $f=-C/4\pi$. In Refs 1 and 2, two sided estimates for C are given. In Ref. 5 an iterative process and approximate analytical formulas to calculate C with the prescribed accuracy were given. In particular, $C\geqslant 4\pi S^2J^{-1}$, where $S=\operatorname{meas}\Gamma$, $J=\int_\Gamma\int_\Gamma|s-t|^{-1}\,ds\,dt$. In many cases the formula $C\simeq 4\pi S^2J^{-1}$ gives an error $\sim 3\%$ (e.g., in the calculation of the capacity of a parallelepiped of arbitrary shape, of a circle cylinder of arbitrary length). If R and r are the radii of the minimal ball containing D and maximal ball which is contained in D, then $r<C<R$. So

$$-R/4\pi\leqslant f\leqslant -r/4\pi. \tag{28}$$

The Neumann boundary condition and low-frequency electromagnetic wave scattering was considered in Ref. 6.

7. MONOTONICITY OF Ψ

If $0\leqslant V_1\leqslant V_2$, then from (5) and the maximum modulus principle it follows that $\Psi_1\geqslant\Psi_2\geqslant 0$, where Ψ_j corresponds to V_j, $j=1,2$. If $V_1\leqslant V\leqslant V_2$, $V_j=\text{const}$, then $\Psi_1\geqslant\Psi\geqslant\Psi_2\geqslant 0$, where for Ψ_j we can give an explicit formulas since $V_j=\text{const}$.

8. ALTERNATING POTENTIAL

Because of Remark 2 we must obtain only an upper bound for f. We are not going to get the best estimates, but rather to describe a simple method of getting the upper bound and give a simple example. First we note that if $V(x)=V_+(x)-V_-(x)$, where $V_+(x)=V(x)$ if $V(x)\geqslant 0$, $V_+(x)=0$ if $V(x)<0$, $V_-(x)=|V(x)|$ if $V(x)\leqslant 0$, $V_-(x)=0$ if $V(x)>0$, and $H\equiv-\nabla^2-V_-(x)$, then $(Hg,g)\geqslant(H_-g,g)$. To simplify the study we make additional assumption

$$H_->0. \tag{29}$$

Then from (9) and (14) it follows that

$$f \leqslant -\frac{1}{4\pi}\int V(x)\,dx + \max_{g\in D(H_-)}\frac{|(g,V)|^2}{4\pi(H_- g,g)}. \tag{30}$$

To obtain a simple upper bound for f we must estimate $(H_- g,g)$ from below. In Ref. 7 the following inequality was proved ($\|\ \| = \|\ \|_{H=L^2(\mathbb{R}^3)}$):

$$(H_- g,g) \geqslant \|\nabla g\|^2(1-\mu_p^{-1}N_p), \tag{31}$$

where

$$H_- = H_0 - V_-, \quad N_p^p = \int |y-x|^{2p-3}V_-^p(x)\,dx,$$

$$p' = \frac{p}{p-1}, \quad p>1,$$

$$\mu_p \equiv \frac{p}{p-1}\left(\frac{4\pi(p-1)\Gamma^2(p)}{\Gamma(2p)}\right)^{1/p}$$

$\Gamma(p)$ is Gamma function, and $y\in\mathbb{R}^3$ is an arbitrary point. Our assumption concerning V_- can now be formulated so: There exist a point $y\in\mathbb{R}^3$ and a number $p>1$ such that

$$N_p < \mu_p. \tag{32}$$

We note that inequality (32) implies (29).

Theorem 3: If inequality (32) holds, then

$$f \leqslant -\frac{1}{4\pi}\int V(x)\,dx + \frac{1}{4\pi}\int |V(x)|\,dx\,\frac{\mu_p}{\mu_p - N_p}. \tag{33}$$

Proof: We have

$$\max\frac{|(g,V)|^2}{(Hg,g)} \leqslant \frac{\int|V|\,dx\int|V(x)||g|^2\,dx}{\|\nabla g\|^2(1-\mu_p^{-1}N_p)+(V_+ g,g)}. \tag{34}$$

Here we used the inequality (31). Since $H_0 - V_- \geqslant 0$ we have

$$\|\nabla g\|^2 \geqslant \int V_-|g|^2\,dx. \tag{35}$$

Hence the right side of (34) is less than

$$\int |V|\,dx\,\frac{A+B}{cA+B}, \quad A\equiv(V_- g,g),$$
$$B\equiv(V_+ g,g), \quad c\equiv 1-\mu_p^{-1}N_p. \tag{36}$$

From (32) we conclude that $0<c<1$. Therefore

$$\frac{A+B}{cA+B} \leqslant \frac{1}{c}. \tag{37}$$

From (37), (36), and (34), we get (33).

Remark 4: We could get the upper bound using different ways. Here are two examples. If for some $\sigma(x)>0$, $\int|V|^2\sigma^{-1}(x)\,dx \leqslant \infty$ and $(Hg,g)\geqslant\int\sigma|g|^2\,dx$, then $\max(|(g,V)|^2)/(Hg,g) \leqslant \int|V|^2\sigma\,dx$. This is the first way to get an upper bound for f. In Ref. 8 inequality $(H_0 g,g) \geqslant \mu_q\|r^{(q-3)/2q}g\|^2_{L^{2q}(\mathbb{R}^3)}$ was used. Connections with Padé approximation were indicated and a very good upper bound for f was obtained. In general any norm $|||g|||$ such that $|(V,g)| \leqslant C_1(V)|||g|||$ and $(Hg,g)\geqslant C_2(V)|||g|||^2$ can be used for obtaining an upper bound for f:

$$f \leqslant -\frac{1}{4\pi}\int V\,dx + \frac{C_1^2(V)}{4\pi C_2(V)}. \tag{38}$$

Remark 5: If inequality

$$\|H_0^{-1}V\|_{C(\mathbb{R}^3)} \leqslant b < 1 \tag{39}$$

holds, then Eq. (5) can be solved by means of iterative process, $\Psi = \Sigma_{j=0}^{\infty}(-1)^j(H_0^{-1}V)^j 1$. In this case we can also get a simple upper bound for f:

$$f = -\frac{1}{4\pi}\int V\,dx + \frac{1}{(4\pi)^2}\int\int\frac{V(x)V(y)}{|x-y|}\,dx\,dy + \cdots$$
$$\leqslant -\frac{1}{4\pi}\int V\,dx + \frac{1}{16\pi^2}\int\int\frac{V(x)V(y)}{|x-y|}\,dx\,dy$$
$$+ \frac{1}{4\pi}\int|V|\,dx\,\frac{b^2}{1-b}. \tag{40}$$

ACKNOWLEDGMENTS

I thank Professor W. Thirring and the University of Vienna for hospitality, and Dr. H. Grosse and Dr. H. Narnhoffer for stimulating discussions.

[1] A.G. Ramm, Ukr. Phys. J. **5**, 534 (1975).
[2] G. Polya and G. Szego, *Isoperimetric Inequalitites in Mathematical Physics* (Princeton University, Princeton, New Jersey, 1951).
[3] R. Blankenbecler and R. Sugar, Phys. Rev. **136**, 472 (1964).
[4] M.A. Krasnoselskij *et al.*, *Approximate Solution of Operator Equations* (Noordhoff Groningen, The Netherlands, 1972).
[5] A.G. Ramm, Sov. Phys. Dokl. **14**, 425 (1969); **15**, 1108 (1971); Radiofizika **14**, 613 (1971); **17**, 1062 (1974).
[6] A.G. Ramm, "Electromagnetic wave scattering by small bodies of arbitrary shape," in Proc. Int. Conference on Scattering Theory, Columbus, Ohio, June 1979 (Pergamon, New York, 1979).
[7] V. Glaser, A. Martin, H. Grosse, and W. Thirring, in *Studies in Mathematical Physics*, edited by E. Lieb, B. Simon, and A. Wightman (Princeton University, Princeton, New Jersey, 1976), p. 169.
[8] H. Grosse, Acta. Phys. Austriaca. **48**, 215 (1978).

A study of the completeness properties of resonant states

W. J. Romo[a)]
Department of Physics, Carleton University, Ottawa, Ontario K1S 5B6

The completeness properties of the discrete set of bound state, virtual states, and resonant states of a Hamiltonian H is investigated, where H describes a system in which a single nonrelativistic-spinless particle moves in a central cutoff potential. A limited form of completeness is obtained. It is shown that the convergence of the resulting "completeness series" is very sensitive to the detailed mathematical structure of the potential.

I. INTRODUCTION

This is the final article of a set of articles[1-3] in which we examine the completeness properties of the set of all bound states, virtual states and resonant states of a simple quantum mechanical system. Resonant states were first introduced by Gamow[4] in his study of alpha decay. Later they were employed by Siegert in a derivation of a dispersion formula for nuclear reactions[5] and by Humblet and Rosenfeld in the formulation of a nuclear reaction theory.[6] Many of the subsequent papers dealt with the development of a scalar product for resonant and virtual states,[7-14] with their inclusion in perturbation calculations,[11-13] with their role in the shell model approach to nuclear reactions[11,15,16] and with their employment in calculations of cross sections for direct reactions to unbound final states.[17,18] The completeness properties of these states has received much less attention. Perhaps, the most practical form of a completeness relationship that involves the resonant states is the one given by Berggren.[10] Starting from the usual completeness relationship consisting of a sum over bound states and an integral over continuum states, Berggren derived his relationship by analytically continuing the integrand of the continuum integral and deforming the path of integration past a finite number of resonance poles. He then separated the resonance and background contributions, obtaining a completeness relationship that involved a sum over the bound states plus a finite set of resonance terms and an integral over the deformed path. By inserting this expansion of the unit operator between the bra and ket of a scalar product, he then[19] obtained a modified sum rule that includes resonant state contributions. Work of a similar nature has been reported by Berrondo and García-Calderón.[20]

In this article we shall determine the conditions under which matrix elements of the form $\langle\psi_1|P_a|\psi_2\rangle$ can be expanded in terms of the infinite but discrete set of bound, virtual and resonant states of a Hamiltonian H describing a system in which a spinless nonrelativistic particle of mass m moves in a central potential $V(r)$ that vanishes for $r>R$. The operator P_a in the matrix element is the projection operator defined by

$$P_a\psi(\mathbf{r}) = \begin{cases} \psi(\mathbf{r}), & \text{if } r\leqslant a, \\ 0, & \text{if } r>a. \end{cases} \tag{1.1}$$

In Ref. 1 we established the existence of such an expansion for the particular case in which $V(r)=\lambda\delta(r-R)$, $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ were any pair of s-wave eigenfunctions of an infinite square well potential of radius R, and the radius parameter a is less than or equal to R. Here we shall extend our consideration to include all partial waves and all potentials that vanish beyond the fixed radius R and are reasonably smooth functions of r for $r\leqslant R$. (A precise definition of the class of potentials will be given in Sec. 2.) To derive the expansion of $\langle\psi_1|P_a|\psi_2\rangle$ we follow the same procedure employed in Ref. 1. One first determines the conditions under which the Mittag–Leffler expansion of the function

$$I_{12}(k,a) = \langle\psi_1|P_a(E+i\epsilon-H)^{-1}P_a|\psi_2\rangle,$$

is convergent. One then obtains the desired expansion of $\langle\psi_1|P_a|\psi_2\rangle$ from a careful examination of $\lim_{E\to\infty}(EI_{12})$. The first step, that of establishing the Mittag–Leffler expansion of I_{12} was carried out in Ref. 2, however, when an attempt was made to complete the study it became clear that the asymptotic forms of the radial wave functions and Jost functions employed in Ref. 2 were inadequate to complete the project, so a detailed calculation of these asymptotic expressions was carried out and the results reported in Ref. 3. To simplify the derivation of the asymptotic formulas, the behavior of the potential near the origin, $r=0$, and near the cutoff radius, $r=R$, was restricted somewhat more than it has been in Ref. 2. On the other hand, some restrictions placed on the behavior of $V(r)$ in the region $0<r<R$ were removed. Consequently, the class of potentials considered in Ref. 3, which we shall again employ in this work, is different from that considered in Ref. 2, and therefore the derivation of the Mittag–Leffler expansion of I_{12} must be repeated.

In Sec. 2 specification of the properties of the potential $V(r)$ and the wavefunction $\psi_i(\mathbf{r})$ are given and the conditions for the convergence of the Mittag–Leffler expansion of $I_{12}(k,a)$ are determined. In Sec. 3 the asymptotic form of the terms of the Mittag–Leffler series is examined and convergence is examined in more detail. In Sec. 4 the limiting procedure alluded to above is carried out and the desired "completeness relationship" for the set of resonant, bound and vertical states is obtained. Finally, in Sec. 5, the article is concluded with a discussion of the results.

2. MITTAG-LEFFLER EXPANSION

In this section we shall determine the convergence conditions for the Mittag–Leffler expansion of the matrix element

$$I_{12}(k,a) = \langle\psi_1|P_a(k^2+i\epsilon-H)^{-1}P_a|\psi_2\rangle, \tag{2.1}$$

[a)]Supported in part by a research grant from the National Research Council of Canada.

 0022-2488/80/020311-16$1.00

where ψ_1 and ψ_2 are wavefunctions of the form

$$\psi_i(\mathbf{r}) = r^{-1}\chi_i(r)Y_{lm}(\hat{\mathbf{r}}), \tag{2.2}$$

P_a is the projection operator defined by Eq. (1.1) and H is the single particle Hamiltonian

$$H = -\nabla^2 + V(r). \tag{2.3}$$

(We shall take $\hbar^2/2m = 1$ throughout this paper.) It shall be assumed that $V(r)$ is a real central potential that satisfies the following list of properties:

(a) $V(r) = 0$ for $r > R$, where R is a fixed positive radius.

(b) Immediately below the cutoff radius, i.e., for $R(1-\delta) \leqslant r \leqslant R$ with $0 < \delta \ll 1$, $V(r) = [1-(r/R)]^{\sigma} \times P_0(1-r/R)$, where $\sigma \geqslant 0$ and $P_0(t) \equiv \Sigma_{j=0}^{\infty} v_j t^j$ is a power series that converges for all t in $[0, \delta]$.

(c) Near the origin, $V(r) = r^{-\gamma}Q_0(r)$, where $\gamma < 2$ and $Q_0(r) = \Sigma_{j=1}^{\infty} \tau_j r^j$ is a power series that converges for all r in the closed interval $[0, \epsilon R]$ with $0 < \epsilon \ll 1$.

(d) $V(r)$ is piecewise continuous and bounded, and all of its derivatives up to order $L \geqslant \sigma + 2$ are also piecewise continuous and bounded in the open interval $(0, R)$.

(e) $d^L V(r)/dr^L$ has n points of discontinuity in the open interval $(0, R)$. One defines R_i, with $\epsilon R < R_1 < R_2 < \cdots < R_n < R(1-\delta)$, to be such a point, and further define m_i to be the smallest interger for which $d^{m_i}V(r)/dr^{m_i}$ is discontinuous at R_i.

(f) Finite right-and left-hand limits of $d^L V(r)/dr^L$ are assumed to exist at each of the points R_i.

Since the states ψ_1 and ψ_2 are eigenstates of $\mathbf{L}^2$ and L_Z, and $V(r)$ is a central potential, I_{12} can also be written as

$$I_{12}(k, a) = \int_0^a dr \int_0^a dr'\, \chi_1^*(r) G_l^{(+)}(k; r, r')\chi_2(r'), \tag{2.4}$$

where $G_l^{(+)}(k;r,r')$ is the l th partial wave component of the full Green's function $\langle \mathbf{r}|(k^2 + i\epsilon - H)^{-1}|\mathbf{r}'\rangle$. The Green's function $G_l^{(+)}(k;r,r')$ can be written[21]

$$G_l^{(+)}(k; r, r') = (-1)^{l+1} k^l \phi_l(k, r_<) f_l(-k, r_>)/f_l(-k), \tag{2.5}$$

where $r_<$ and $r_>$ are the lesser and greater, respectively of r and r', while $\phi_l(k,r)$ and $f_l(-k,r)$ are regular and irregular solutions, respectively, of the radial Schrödinger equation

$$-\frac{d^2}{dr^2}\psi_l + V(r)\psi_l + \frac{l(l+1)}{r^2}\psi_l = k^2\psi_l. \tag{2.6}$$

The regular solution also satisfies the boundary condition as $r \to 0$ given by

$$\lim_{r \to 0} (2l+1)!! r^{-l-1}\phi_l(k, r) = 1, \tag{2.7}$$

while the irregular solution satisfies

$$\lim_{r \to \infty} e^{-ikr} f_l(-k, r) = i^l. \tag{2.8}$$

Finally, $f_l(-k)$ is the Jost function defined by

$$f_l(-k) = \lim_{r \to 0} (-kr)^l f_l(-k, r)/(2l-1)!! \tag{2.9}$$

To complete the specification of $I_{12}(k,a)$ it shall be assumed that the radial functions $\chi_i(r)$, with $i = 1$ and 2, and all of its derivatives of order $\leqslant L$ are bounded and piecewise continuous for all $0 < r < R$, and that near the origin

$$\chi_i(r) = r^{\xi_i} \sum_{n=0}^{\infty} x_{in} r^n, \tag{2.10}$$

where the power series converges for all r in the closed interval $[0, \epsilon R]$ with $0 < \epsilon \ll 1$, and $\xi_i > -1/2$ for $i = 1,2$. Once again, as in the case of $V(r)$, the properties of $\chi_i(r)$ are more carefully spelled out in this paper than they were in Ref. 2.

Although the class of functions to which $\chi_i(r)$ and $V(r)$ belong are somewhat different in this paper than the corresponding classes of functions in Ref. 2, the arguments used to obtain the general analytic properties of $I_{12}(k,a)$ given in that reference were sufficiently general so that they apply in the present situation as well. Hence, one finds that $I_{12}(k,a)$ is a meromorphic function of k with an infintie number of poles that lie at the zeros of $f_l(-k)$. The zeros of $f_l(-k)$ in the upper half-plane are restricted to lie at points iK_n on the imaginary axis, where $-K_n^2$ is a bound-state energy. These zeros will be of unit order.

The zeros of $f_l(-k)$ in the lower half-plane will be infinite in number, of finite order and symmetrically distributed about the imaginary axis. There may be at most a finite number of zeros along the imaginary axis. The only zero that $f_l(-k)$ can have on the real axis will be found at $k = 0$. For simplicity, and without appreciable loss of generality, as was explained in Ref. 2, it shall be assumed that all zeros are simple zeros.

The poles of $I_{12}(k,a)$ or equivalently the zeros of $f_l(-k)$, shall be ordered as follows: The poles $k_1, k_2, \cdots k_N$ will correspond to the N poles which lie along the imaginary k axis, where N is a finite number.[21] The poles k_n with $n > N$ will correspond to poles in the fourth quadrant of the k plane, so ordered that if $n > m > N$ then either $|k_n| > |k_m|$, or $|k_n| = |k_m|$ and $|\mathrm{Im}k_n| > |\mathrm{Im}k_m|$. The poles in the third quadrant will be labeled by k_{-n} with $n = N+1, N+2, \cdots$. They are related to the fourth quadrant poles by

$$k_{-n} = -k_n^*. \tag{2.11}$$

Since the location of the poles vary from partial wave to partial wave, they should be labeled by an additional subscript l to identify the partial wave, however, for simplicity this label has been supressed.

We are now in a position to consider the Mittag–Leffler expansion of $I_{12}(k,a)$. The approach shall be the same as the one used in Ref. 2. According to Cauchy's residue theorem.[22]

$$\frac{1}{2\pi i}\oint_{C_m} \frac{I_{12}(k', a)dk'}{k' - k} = I_{12}(k, a) + \sum_{n=1}^{N} \frac{\mathscr{R}_n(a)}{k_n - k} + \sum_{n=N+1}^{\infty} \sum_{\alpha=-1}^{+1} \frac{\mathscr{R}_{\alpha n}(a)}{k_{\alpha n} - k}, \tag{2.12}$$

where $m \geqslant N_0 \gg N$, C_m is a circular path[23] centered at $k' = 0$ that encloses all poles $k_{\pm n}$ with $|n| \leqslant m$, but excludes all other poles, $\mathscr{R}_{\pm n}$ is the residue of $I_{12}(k', a)$ at $k' = k_{\pm n}$, and k is a point inside the contour C_m at which $I_{12}(k', a)$ is analytic. N_c is an integer that is sufficiently large so that k_m

assumes its asymptotic form, to be given later, and $|k_n| > |k_j|$ for all $|n| > N_c$ and $|j| \leqslant N$. The Mittag–Leffler expansion follows from Eq. (2.12) if one can show that the contour integral on the left-hand side of Eq. (2.12) vanishes as m tends to infinity. For sufficiently large m, $|k'| \geqslant 2|k|$ for every point k' on C_m, and thus

$$\left| \oint_{C_m} \frac{I_{12}(k', a)dk'}{k' - k} \right| \leqslant 2 \oint_{C_m} \left| \frac{I_{12}(k', a)}{k'} \right| |dk'| . \tag{2.13}$$

From the properties of the radial wavefunctions $\chi_i(r)$ with $i = 1$ and 2, it can be seen that $|\chi_i(r)|$ is integrable over the interval $0 \leqslant r \leqslant R$. Using this fact and the bounds on $G_l^{(+)}(k;r,r')$ given by Newton,[21] one can easily show that $|I_{12}(k,a)/k| \leqslant \text{const} \times |k|^{-2}$ for all k on C_m with Im $k \geqslant 0$. Hence, the contribution to the contour integral arising from that portion of C_m for which Im $k' \geqslant 0$ will vanish as $m \to \infty$, for any $a \leqslant R$.

Although the changes in the classes of functions to which $\chi_i(r)$ and $V(r)$ belong, do not produce any important changes in the asymptotic form of $I_{12}(k,a)$ for points in the upper half k plane, the same is not true for points in the lower half k plane. To determine the asymptotic form of $I_{12}(k,a)$ as $k \to \infty$ with Im $k < 0$, one must first determine the asymptotic forms of $\phi_l(k,r)$, $f_l(-k,r)$ and $f_l(-k)$, as can be seen from Eqs. (2.4) and (2.5). The asymptotic forms of $\phi_l(k,r)$ and $f_l(-k)$ for the class of potentials being considered were reported in Ref. 3. The techniques employed in that reference to determine the asymptotic form of $\phi_l(k,r)$ can also be employed to determine the asymptotic form of $f_l(-k,r)$. To illustrate the technique and to define some important functions we shall give a brief derivation of the asymptotic form of $f_l(\pm k,r)$ as $k \to \infty$ in $\mathscr{D}$, where $k \in \mathscr{D}$ if $|k| \geqslant \kappa \gg 1$ and Im $k < 0$.

One first constructs a pair of linearly independent solutions to Eq. (2.6) for which the asymptotic form is known in the radial interval $R_j \leqslant r \leqslant R_{j+1}$. This construction is repeated until a pair of solutions is obtained for each of the intervals $[R_j, R_{j+1}]$ with j ranging from 0 to $n + 2$. The radius R_j with $j = 1$ to n was defined in the specification of $V(r)$, while the additional radii R_{n+1} and R_{n+2} are defined by

$$R_{n+1} = R - |k|^{-1} \text{ and } R_{n+2} = R , \tag{2.14}$$

and R_0 will be a radius depending on $|k|$ defined so that $R_0 \to 0$ but $|k| R_0 \gg 1$ as $|k| \to \infty$, for example one might choose $R_0 = \text{const} \div \ln(|k| R)$ as was done in Ref. 3 or possibly $R_0 = 100(l + 1/2)^2 |k|^{-1}$, as in Ref. 2. The general techniques for the construction of the linearly independent auxiliary solutions to Eq. (2.6) can be found in the text by Erdélyi.[24] The required specialization of those techniques with a few necessary modifications was discussed in Ref. 3. Defining the two independent solutions associated with the interval $\Sigma_j \equiv [R_j, R_{j+1}]$ to be $y_{\alpha,j}(k,r)$ with $\alpha = \pm 1$ (note that the angular momentum label has been suppressed), one finds that

$$y_{\alpha,j}(k, r) = Y_j(\alpha k, r)\{1 + O[(kR_j)^{-x_j}]\} , \tag{2.15a}$$

$$\frac{dy_{\alpha,j}(k, r)}{dr} = \frac{dY_j(\alpha k, r)}{dr} \{1 + O[(kR_j)^{-x_j}]\} , \tag{2.15b}$$

uniformly for all r in Σ_j and k in $\mathscr{D}$, where

$$Y_j(k, r) = e^{ikr} \sum_{m=0}^{M_j} a_m^j(r) k^{-m} , \tag{2.15c}$$

with

$$a_0^j(r) = 1 \text{ for every } j, \text{ and} \tag{2.15d}$$

$$\frac{da_m^j(r)}{dr} = (2i)^{-1}[\{l(l+1)r^{-2} + V(r)\} \times a_{m-1}^j(r) - d^2 a_{m-1}^j(r)/dr^2] , \tag{2.15e}$$

for all j and $m \geqslant 1$. For the power X_j appearing in Eqs. (2.15a) and (2.15b) one has $X_j = L$ for $j \leqslant n - 1$ and $X_n = X_{n+1} = \sigma + 2$. For the upper limit of the sum in Eq. (2.15c) one has $M_j = L$ for $0 \leqslant j \leqslant n$ and $M_{n+1} = [\sigma] + 1$, where $[\sigma]$ is the greatest integer $\leqslant \sigma$. The arbitrary constants of integration that arise when one integrates Eq. (2.15e), to obtain $a_m^j(r)$ were chosen so that

$$a_m^{n+1}(r) = \int_\infty^r dr' \frac{da_m^{n+1}(r')}{dr'} , \tag{2.16a}$$

and $r \geqslant R_{n+1}$ and $1 \leqslant m \leqslant [\sigma] + 1$,

$$a_m^n(r) = \int_{R_n}^r dr' \frac{da_m^n(r')}{dr'} , \tag{2.16b}$$

for r in Σ_n and $[\sigma] + 1 < m \leqslant L$, and

$$a_m^j(R_{j+1}) = a_m^{j+1}(R_{j+1}) , \tag{2.16c}$$

for every $m \leqslant L$ if $0 \leqslant j \leqslant n - 1$, and also for $1 \leqslant m \leqslant [\sigma] + 1$ if $j = n$.

A number of important properties of the $a_n^j(r)$ can be determined from Eq. (2.15e) and the properties of $V(r)$ given earlier.[3] For $0 < j < n$, one finds that $a_n^j(r)$ and all of its derivatives of order $\leqslant L + 2 - n$ are continuous throughout the open interval (R_j, R_{j+1}) and are bounded throughout the closed interval $[R_j, R_{j+1}]$. The coefficients $a_n^0(r)$ and all of its derivatives of order $\leqslant L + 2 - n$ are continuous for all r for which $R_0 \leqslant r \leqslant R_1$, and satisfy

$$\left| \frac{d^m a_n^0(r)}{dr^m} \right| \leqslant \text{const} r^{-n-m} , \tag{2.17a}$$

for $R_0 \leqslant r \leqslant R_1$. For $j = n$ and $n + 1$, $a_n^j(r)$ and all of its derivatives of order $\leqslant M_j + 2 - n$ are continuous in the open interval (R_j, R_{j+1}) and satisfy

$$\left| \frac{d^m a_n^j(r)}{dr^m} \right| \leqslant \text{const}[1 + (1 - r/R)^{\sigma+2-m-n}] \tag{2.17b}$$

for $R_j \leqslant r \leqslant R_{j+1}$.

Since the two solutions $y_{\pm 1,j}(k,r)$ linearly independent one can write $f_l(\pm k,r)$ as a linear combination of the two:

$$f_l(-\beta k, r) = \sum_{\alpha=1}^{+1} A_{\alpha,\beta}^j(k) y_{\alpha,j}(k, r) , \quad \beta = \pm 1 , \tag{2.18a}$$

with

$$A_{\alpha,\beta}^j(k) = \frac{W[y_{-\alpha,j}(k, r), f_l(-\beta k, r)]}{W[y_{-\alpha,j}(k, r), y_{\alpha,j}(k, r)]} , \tag{2.18b}$$

where $W[f(r), g(r)]$ is the Wronskian defined by

$$W[f(r), g(r)] = f(r) \frac{dg(r)}{dr} - g(r) \frac{df(r)}{dr} .$$

Since the Wronskians are independent of r, they may be evaluated at any convenient point. Clearly the Wronskians involving $y_{\alpha,j}(k,r)$ will be evaluated with r in Σ_j, where the asymptotic expansions given by Eqs. (2.15)–(2.17) are valid. To begin the calculation of the coefficients $A^{j}_{\alpha,\beta}(k,r)$, one observes that[21]

$$f_l(k,r) = w_l(kr) \tag{2.19}$$

for all $r \geqslant R$, where $V(r) = 0$, and where $w_l(kr)$ is a Riccati–Hankel function. Using the known analytic form for $w_l(kr)$ and its derivative,[25] one finds that

$$\begin{aligned} A^{n+1}_{\alpha,\beta}(k) &= W[y_{-\alpha,n+1}(k,r), w_l(-\beta kr)] \\ &\quad \times \{W[y_{-\alpha,n+1}(k,r), y_{\alpha,n+1}(k,r)]\}^{-1}|_{r=R} \\ &= i^l\{\delta_{\alpha,+1}[1 + O(k^{-N-2})] \\ &\quad + \delta_{\alpha,-1} O(k^{-\sigma-2} e^{2i\beta kR})\}, \end{aligned} \tag{2.20}$$

with $N = [\sigma]$. Using Eq. (2.18a) with $j = n+1$, together with its derivative with respect to r and the asymptotic forms of $y_{\alpha,n+1}(k,r)$, $dy_{\alpha,n+1}(k,r)/dr$ and $A^{n+1}_{\alpha,n+1}(k)$ given by Eqs. (2.15a), (2.15b), and (2.20), respectively, gives us the asymptotic form of $f_l(-\beta k,r)$ and its derivative with respect to r for the range $r = R_{n+1}$ to R. This enables us to compute the asymptotic form of $A^{n}_{\alpha,\beta}(k)$ by means of Eq. (2.18) with $r = R_{n+1}$. Continuing this procedure, one finds the asymptotic form of each of the $A^{j}_{\alpha,\beta}(k)$ with $0 \leqslant j \leqslant n$, namely

$$\begin{aligned} A^{j}_{\alpha,\beta}(k) = i^l \Bigg\{ & \delta_{\alpha,+1}\Bigg[1 + \sum_{p=j+1}^{n+1} O(k^{-\mu_p-2}) \\ & + \sum_{j+1\leqslant p<q\leqslant n+1} O(k^{-\mu_p-\mu_q-4}\exp[2i\beta k(R_q - R_p)])\Bigg] \\ & + \delta_{\alpha,-1} \sum_{j+1\leqslant p\leqslant n+1} O\left[k^{-\mu_p-2}\exp(2i\beta kR_p)\right]\Bigg\}, \end{aligned} \tag{2.21}$$

where $\mu_{n+1} = [\sigma]$ and $\mu_j = \min[m_j, L-2]$ for $j \leqslant n$. Eqs. (2.18a), (2.21), and (2.15) yield the asymptotic forms for $f_l(\pm k,r)$ and $df_l(\pm k,r)/dr$ for all points from R_0 to R. Using the same expansion functions, one finds[3]

$$\phi_l(k,r) = \sum_{\alpha=-1}^{+1} B_{\alpha j}(k) y_{\alpha j}(k,r) \tag{2.22}$$

$$\begin{aligned} B_{\alpha j}(k) = {} & \tfrac{1}{2}(i\alpha k)^{-l-1}\{1 + K_l(\alpha k) + O(k^{-1}\ln k) \\ & + \sum_{q=0}^{j} O(k^{-\mu_q-2}\exp[-2i\alpha k R_q])\}, \end{aligned} \tag{2.23a}$$

where

$$K_l(k) = \begin{cases} \sum_{j=1}^{m} (2ik)^{j(\gamma-2)} b(l,j) & \text{for } 1<\gamma<2, \\ 0 & \text{for } \gamma \leqslant 1, \end{cases} \tag{2.23b}$$

and $m = [(2-\gamma)^{-1}]$. The sum over q is absent in the $j = 0$ case. In Ref. 3 it was also shown that

$$\begin{aligned} f_l(\mp k) = {} & 1 + K_l(\mp k) + 0(k^{-1}\ln k) + [1 + K_l(\pm k) \\ & + O(k^{-1}\ln k)]\sum_{j=1}^{n+1} A_j(\pm iz)^{-\alpha_j}\exp(\pm i\beta_j z), \end{aligned} \tag{2.24a}$$

as $k \to \infty$ in $\mathcal{D}$, where $z = 2kR$,

$$\beta_j = \frac{R_j}{R}, \alpha_j = m_j + 2, A_j = (-1)^{l+1}(-R)^{\alpha_j}\Delta_j, \tag{2.24b}$$

for $j = 1, n$,

$$\beta_{n+1} = 1, \alpha_{n+1} = \sigma + 2, A_{n+1} = (-1)^l v_0 R^2 \Gamma(\sigma+1), \tag{2.24c}$$

and

$$\Delta_j = \lim_{\epsilon \to 0+} \frac{d^{m_j} V(r)}{dr^{m_j}}\Bigg|_{R_j-\epsilon}^{R_j+\epsilon}. \tag{2.24d}$$

Since we now have the asymptotic forms of $\phi_l(k,r)$, $f_l(-k,r)$, and $f_l(-k)$ as $k \to \infty$ with $\mathrm{Im}k < 0$, we are in a position to determine a bound on the portion of the integral on the left-hand side of Eq. (2.13) that arises from the segment of C_m for which $\mathrm{Im}k' < 0$. The approach we shall take is identical to that employed in Appendix C of Ref. 2. One first defines $F_j(k)$ with $j = 1,4$ by

$$\begin{aligned} F_1(k) &= \int_{R_n}^{a} dr \int_{R_n}^{a} dr' \mathcal{F}(r,r'), \\ F_2(k) &= \int_{0}^{R_n} dr \int_{R_n}^{a} dr' \mathcal{F}(r,r'), \\ F_3(k) &= \int_{R_n}^{a} dr \int_{0}^{R_n} dr' \mathcal{F}(r,r'), \\ F_4(k) &= \int_{0}^{R_n} dr \int_{0}^{R_n} dr' \mathcal{F}(r,r'), \end{aligned} \tag{2.25}$$

with

$$\begin{aligned} \mathcal{F}(r,r') = {} & (-1)^{l+1} k^l \chi_1^*(r)\phi_l(k,r_<) f_l(-k,r_>) \\ & \times \chi_2(r')/f_l(-k) \end{aligned} \tag{2.26}$$

and $R_0 = 100(l+1/2)^2/|k|$.

One then has

$$I_{12}(k,a) = \sum_{i=1}^{4} F_i(k).$$

To obtain bounds on the interior integrals in which r and, or r' run from 0 to R_0, one replaces the functions $f_l(-k,r)$ and $\phi_l(k,r)$ that appear in the integrand by the bounds on their absolute magnitude given by Newton,[21] and also replaces $\chi_1^*(r)$ an $\chi_2(r')$ by the following bounds on their absolute magnitudes

$$|\chi_i(r)| < \mathrm{const} r^{-\xi} \quad \text{for } i = 1 \text{ and } 2, \tag{2.27a}$$

where

$$\xi = \max[\xi_1, \xi_2]. \tag{2.27b}$$

This bound follows immediately from the properties of $\chi_i(r)$ given earlier. To obtain a bound on the magnitude of the exterior integrals one first divides these integrals into subintegrals over the subsegments Σ_j, then replaces $\phi_l(k,r)$ and $f_l(-k,r)$ by their asymptotic forms that are valid in Σ_j. The integrand will then be a sum of terms, each proportional to a function of the form $e^{i\beta kr}F(r)$, where $\beta = 0, \pm 1$, or ± 2, and $F(r)$ is a bounded function of r which is independent of k. If $\beta \neq 0$, one integrates by parts at least one time. If $\beta = 0$ one has a suitable expression as it stands. One next sums over j to

obtain the asymptotic form of the exterior integrals and then uses the properties of the functions $\chi_i(r)$ and $a_n^j(r)$ to determine a bound on the absolute magnitudes of the external integrals. After a considerable amount of algebra one finds that there exist positive constants C and $D_p, p = 1$ to $n+1$, such that for sufficiently large $|k|$

$$|I_{12}(k)||f_l(-k)| \leqslant C|k|^{-1}[e^{2|\nu|a}|kR|^{-2}+1 + \sum_{p=1}^{n} D_p e^{2|\nu|R}|kR|^{-\mu_p-2} + D_{n+1}e^{2|\nu|R}|kR|^{-\sigma-2}], \quad (2.28)$$

where $\nu = \mathrm{Im}k < 0$. In Ref. 3 the location of those zeros of $f_l(-k)$ for which $|k| \to \infty$ were determined. Such zeros lie in either the third or fourth quadrant of the k plane, and are given by

$$k_M^j = (2R)^{-1}[\zeta_j M - i\lambda_j \ln|M| + O(1)], \quad (2.29a)$$

$$\zeta_j = 2\pi(\beta_{N(j)} - \beta_{N(j-1)})^{-1}, \quad (2.29b)$$

where the index M takes on all integral values satisfying $|M| \geqslant N_{c,j}$, where $N_{c,j}$ is some large positive integer, and the index j runs from 1 to $\tau \leqslant n+1$. As before, n is the total number of discontinuities of $d^L V(r)/dr^L$. The constants λ_j are defined by

$$\lambda_j = \min_{m|m>N(j-1)|}\left(\frac{\alpha_m - \alpha_{N(j-1)}}{\beta_m - \beta_{N(j-1)}}\right) = \frac{\alpha_{N(j)} - \alpha_{N(j-1)}}{\beta_{N(j)} - \beta_{N(j-1)}}, \quad (2.30)$$

for $j = 1$ to τ, where $N(0) \equiv 0$ and $N(j)$ equals the value of m for which the minimum occurs. Note that Eq. (2.30) implies that $N(j)$ increases with j and that $N(\tau) = n+1$. It shall be assumed that Eq. (2.30) has unique solutions for each j. (Any degeneracies can be removed by infinitesimal changes in the parameters β_m.) It then follows that $\lambda_j < \lambda_{j+1}$ for each j from 1 to $\tau - 1$. The two parameters α_0 and β_0, which are required in Eq. (2.30) when $j = 1$, are defined by

$$\alpha_0 = \beta_0 = 0. \quad (2.31)$$

Equation (2.29) is an asymptotic expansion in terms of the index M. The remainder term $O(1)$ is of zero order in M. It should be noted that Eq. (2.29) involves a double index label for the zeros of $f_l(-k)$, while the ordering of zeros previously introduced involves only one index. Clearly the transformation from a single label to a double label corresponds to a reordering of the zeros. It will prove convenient to retain the single label form when establishing the conditions for convergence of the Mittag–Leffler expansion of $I_{12}(k,a)$. Once the convergence has been established, further information will be obtained by reordering the resulting infinite sum. The validity of this procedure will be established when it is required. From Eq. (2.29) one sees that a contour plot of the function $|f_l(-k)|^{-1}$ in the right-half k plane would look like a set of τ mountain ranges, with each range labelled by a fixed value of j, with roughly equal spacing between mountains in a given range. Since a zero of $f_l(-k)$ corresponds to an infinity of $|f_l(-k)|^{-1}$, the altitudes of these "mountains" are infinity high. However, the contour C_m passes between the poles and it can be deformed so that for each values of Im k the real value is chosen so that $|f_l(-k)|^{-1}$ is minimized. If this is done it follows that

$$|f_l(-k)|^{-1} \leqslant C|k|^{\alpha N(j)}\exp(-2\beta_{N(j)}|\nu|R), \quad (2.32)$$

for k on C_m^j, where C_m^j is that portion of C_m for which Re $k > 0$ and

$$-\lambda_{j+1}\ln(2|k|R) \leqslant 2R\,\mathrm{Im}k \leqslant -\lambda_j \ln(2|k|R),$$

with $j = 0$ to $\tau - 1$. For the segment C_m^τ defined to be that part of C_m for which Re $k > 0$ and $\mathrm{Im}k < -(2R)^{-1}\lambda_\tau \times \ln(2|k|R)$, one has

$$|f_l(-k)|^{-1} \leqslant C|k|^{2+\sigma}\exp(-2|\nu|R). \quad (2.33)$$

Using the symmetries of $f_l(-k)$, $f_l(-k,r)$, and $\phi_l(-k,r)$ given in Ref. 21, one finds that

$$|I_{12}(k,a)/k| = |I_{12}(-k^*,a)/(-k^*)|,$$

which implies that the contribution to the integral on the right-hand side of Eq. (2.13) made by the segment of C_m lying in the third quadrant exactly equals the contribution from the segment in the fourth quadrant. Hence, one only needs to evaluate the latter contribution. Combining Eqs. (2.13), (2.28), (2.32), and (2.33), and carrying out the resulting integrals, one finds that

$$\left|\oint_{C_m}\frac{I_{12}(k',a)dk'}{k'-k}\right| \leqslant C\rho_m^{-1} + \sum_{j=0}^{\tau-1}\left.\frac{C_j\rho_m^{\alpha_{N(j)}+\lambda(a/R-\beta_{N(j)})+4}}{a-R_{N(j)}}\right|_{\lambda=\lambda_j}^{\lambda=\lambda_{j+1}} + C'\rho_m^{\sigma-\lambda_\tau(1-a/R)-2}, \quad (2.34)$$

provided that $a < R$, where $\lambda_0 \equiv 0$ and ρ_m is the radius of the circular path C_m. The first term on the right-hand side of Eq. (2.34) is a bound on the contribution of the integral over the portion of C_m that lies in the upper half k' plane, the sum on j is a bound on the integral over the segments for which Imk satisfies

$$-(2R)^{-1}\lambda_\tau \ln\rho_m \leqslant \mathrm{Im}k' < 0,$$

and the last term is a bound on the integral over the line segment for which $\mathrm{Im}k' < -(2R)^{-1}\lambda_\tau \ln\rho_m$. Since $\rho_m \to \infty$ as $m \to \infty$ the contour integral will vanish as $m \to \infty$ provided that

$$a/R < 1, \quad (2.35a)$$

$$a/R < \beta_\tau^c = 1 - (\sigma - 2)/\lambda_\tau, \quad (2.35b)$$

$$a/R < \beta_j^c = \beta_{N(j)} - (\alpha_{N(j)} - 4)/\lambda_j, \quad \text{if } \alpha_{N(j)} \geqslant 4, \quad (2.35c)$$

$$a/R < \beta_j^c = \beta_{N(j)} + (4 - \alpha_{N(j)})/\lambda_{j+1}, \quad \text{if } \alpha_{N(j)} \leqslant 4. \quad (2.35d)$$

Employing the definition of λ_j given by Eq. (2.30) and the fact that $\lambda_j, \beta_{N(j)}$, and $\alpha_{N(j)}$ are all positive quantities that increase with j, one finds that

$$\beta_j^c \geqslant \beta_J^c$$

for all j, where J is the value of j for which $|\alpha_{N(j)} - 4|$ is minimized as j runs from 0–τ, i.e.,

$$|\alpha_{N(J)} - 4| = \min_{0 \leqslant j \leqslant \tau}|\alpha_{N(j)} - 4|. \quad (2.36)$$

Hence the Mittag–Leffler series $I_{12}(k,a)$ converges if Eq.

(2.35a) is satisfied and also

$$a/R < \beta_J^c \tag{2.37}$$

with J defined by Eq. (2.36) and β_j^c defined by Eqs. (2.35b)–(2.35d).

To help clarify the situation we shall give the convergence conditions for three special cases:

(i) If $\sigma \leqslant 2$, then $\alpha_{N(j)} < 4$ for all $j \leqslant \tau$ and $J = \tau$. Equation (2.35b) then yields $\beta_J^c \geqslant 1$ and the Mittag–Leffler series converges for all a for which Eq. (2.35a) is satisfied.

(ii) If $\alpha_{N(1)} > 4$, then either $J = 0$ or 1. In either case $\beta_J^c = 4/\lambda_1$ and the series converges for all a for which $a < 4R_{N(1)}/\alpha_{N(1)}$.

(iii) If there exists a $j = J$ for which $\alpha_{N(J)} = 4$, which implies that $V(r)$ has a discontinuity in its second derivative at $R_{N(J)}$, then $\beta_J^c = \beta_{N(J)}$ and the series converges for all a for which $a < R_{N(J)}$.

If all of the derivatives of $V(r)$ up to order L are continuous for $0 < r < R$, i.e., the $n = 0$ case, then the potential will be a member of the same class of functions examined in Ref. 2. For $\sigma \leqslant 2$, case (i) above applies, which agrees with the result obtained in Ref. 2. For $\sigma > 2$, since $n = 0$ then $\tau = 1$ and $\alpha_{N(\tau)} = \alpha_{N(1)} = \sigma + 2$. Thus (ii) holds and the series converges if $a < 4R\,(2+\sigma)^{-1}$. This also agrees with the result of Ref. 2 in those for which both $\chi_1(a)$ and $\chi_2(a)$ are nonvanishing. The analysis of Ref. 2 shows that the radius of convergence is extended if either one or both χ_1 and χ_2 vanish at a. This refinement has been ignored in the present derivation. The additional convergence condition $a < (3+\sigma_j)2R/(2+\sigma)$ for $j = 1,2$ that appeared in Ref. 2 has been eliminated in the present paper by improving the bound on $|I_{12}(k,a)|$.

In concluding this section, we would like to point out that although we have derived the convergence condition for a particular orbital angular momentum l, the covnergence conditions given by Eqs. (2.35a)–(2.35d) or equivalently by Eqs. (2.35a), (2.36), and (2.37) are the same for every l.

3. DETAILED STUDY OF THE CONVERGENCE OF THE MITTAG-LEFFLER SERIES

When Eqs. (2.35a) and (2.37) are satisfied the Mittag–Leffler expansion for $I_{12}(k,a)$ converges, so that from Eq. (2.12) one has

$$I_{12}(k,a) = \sum_{n=1}^{N} \frac{\mathscr{R}_n(a)}{k-k_n} + \sum_{n=N+1}^{\infty} \left(\frac{\mathscr{R}_n(a)}{k-k_n} + \frac{\mathscr{R}_{-n}(a)}{k-k_{-n}} \right). \tag{3.1}$$

From the definition of $I_{12}(k,a)$, Eq. (2.4), and the relationship of $\phi_l(k,r)$ and $f_l(-k,r)$ at a zero of $f_l(-k)$, one finds that[2]

$$\begin{aligned} \mathscr{R}_n(a) &= \mathrm{Res}[I_{12}(k,a),\ k = k_n] \\ &= \frac{\langle \chi_1 | P_a | \Phi_n \rangle \langle \Phi_{-n} | P_a | \chi_2 \rangle}{2k_n}, \end{aligned} \tag{3.2a}$$

where

$$\langle \chi_i | P_a | \Phi_n \rangle = \langle \Phi_n | P_a | \chi_i \rangle^* = \int_0^R dr \chi_i^*(r) \Phi_n(r), \tag{3.2b}$$

$$\Phi_n(r) = 2k_n^{l+1} [i f_l(k_n) d f_l(-k)/dk|_{k=k_n}]^{-1/2} \times \phi_l(k_n, r), \tag{3.2c}$$

and

$$\Phi_{-n}(r) = \Phi_n^*(r). \tag{3.2d}$$

Equations (3.2a)–(3.2c) hold for both positive and negative values of n. The wavefunction $\Phi_n(r)$ is the usual normalized resonant state wavefunction[2,10,11] corresponding to the pole at k_n.

To obtain further insight into the convergence of the Mittag–Leffler expansion and to gain information that will be essential when we examine the completeness properties of the resonance states, we shall determine the asymptotic form of the residues $\mathscr{R}_{\pm n}(a)$ as $n \to \infty$. To determine these asymptotic forms one must first construct the asymptotic form of the integral

$$\mathscr{I}(k,a) = \int_0^a dr \chi(r) \phi_l(k,r) \tag{3.3}$$

in the limit that $|k| \to \infty$ with $\mathrm{Im}k < 0$, where $\chi(r)$ can be any one of the functions $\chi_1(r)$, $\chi_2(r)$, $\chi_1^*(r)$, and $\chi_2^*(r)$. Employing Eqs. (3.2d), (2.22), (2.15a), and (2.15c), one finds that with r in $\Sigma_j \equiv [R_j, R_{j+1}]$, the integrand of Eq. (3.3) becomes[26]

$$\begin{aligned} \chi(r)\phi_l(k,r) = \sum_{\alpha=-1}^{+1} B_{\alpha,j}(k) \chi(r) e^{i\alpha k r} \\ \times \left\{ \sum_{m=0}^{M_j} a_m^j(r) (\alpha k)^{-m} + O\left[(kR_j)^{-X_j}\right] \right\}. \end{aligned} \tag{3.4}$$

From the assumed properties of $\chi_1(r)$ and $\chi_2(r)$ and the definition of $\chi(r)$ it follows that $d^L\chi(r)/dr^L$ is piecewise continuous and bounded for r in $(0,R)$. Let us subdivide the segments Σ_j into subsegments $[R_{j,\beta}, R_{j,\beta+1}]$ with β ranging from 0 to $\beta(j)$, where $R_{j,0} \equiv R_j$, $R_{j,\beta(j)+1} \equiv R_{j+1}$ and the $R_{j,\beta}$ with $0 < \beta \leqslant \beta(j)$ are the complete set of discontinuities of $d^L\chi(r)/dr^L$ in Σ_j. From the properties of the expansion coefficients $a_m^j(r)$ given in Sec. 2 and in Ref. 3, it then follows that the function $F_{jm} = \chi(r) a_m^j(r)$ is d_m times differentiable throughout $[R_{j,\beta}, R_{j,\beta+1}]$, where

$$d_m = \min[L, L+2-m], \tag{3.5}$$

and that

$$|F_{jm}^{(\nu)}(r)| \leqslant \text{const} \quad \text{for } 1 \leqslant j \leqslant n, \tag{3.6}$$

where

$$|F_{0m}^{(\nu)}(r)| \leqslant \text{const} \times r^{-\xi-m-\nu}, \tag{3.7}$$

with ξ given by Eq. (2.27b). Hence the mth term in the integral can be evaluated by repeated integration by parts. One obtains

$$\begin{aligned} \int_{R_{j,\beta}}^{R_{j,\beta+1}} dr \chi^*(r) \phi_l(k,r) \\ = \sum_{\alpha=-1}^{+1} B_{\alpha,j}(k) \left\{ \sum_{m=0}^{M_j} (i\alpha k)^{-m} \right. \\ \left. \times \left(- \sum_{\nu=0}^{d_m-1} (-i\alpha k)^{-\nu-1} e^{i\alpha k r} F_{jm}^{(\nu)}(r) \Big|_{R_{j,\beta}}^{R_{j,\beta+1}} \right. \right. \end{aligned}$$

$$+ (i\alpha k)^{-d_m} \int_{R_{j,\beta}}^{R_{j,\beta+1}} dr\, e^{i\alpha k r} F_{jm}^{(d_m)}(r)\Big) + O\left[(kR_j)^{-X_j}\left|\int_{R_{j,\beta}}^{R_{j,\beta+1}} dr\, e^{i\alpha k r}\chi(r)\right|\right]\Big\}. \tag{3.8}$$

Bounds on the two integrals on the right-hand side of Eq. (3.8) can be constructed by using Eqs. (2.27a), (3.6), and (3.7). For the integral from $r=0$ to R_0, $\chi(r)$ is replaced by a Born series expansion as was done in evaluating the corresponding "internal" integrals in Refs. 2 and 3. The first term of the resulting series expansion can easily be evaluated and it is seen that the leading R_0 dependent terms cancel against the leading R_0 dependent terms of the external integral. The higher order terms are not as easily evaluated but using the techniques of Ref. 3 one finds that

$$\int_{R_n}^{0} dr\chi(r)\phi_l(k,r) = k^{-l-2-\xi}\left[\sum_{j=0}^{m} C_j k^{j(\gamma-2)} + O\,|k|^{-1}\ln|k|)\right] + (R_0 \text{ dependent terms}), \tag{3.9}$$

where m is the greatest integer $\leqslant (2-\gamma)^{-1}$.

If one defines J and b as the values of j and β for which $R_{j,\beta}\leqslant a\leqslant R_{j,\beta+1}$, then the external integral consists of a sum over all those subintegrals given by the left-hand side of Eq. (3.8) for which j and β satisfy the constraints $R_{j,\beta}\leqslant R_{J,b}$, plus an additional integral of the same form, but with $R_{J,b}$ replacing $R_{j,\beta}$ and a replacing $R_{j,\beta+1}$. Adding the interior integral to the exterior integral, using properties of the amplitudes $B_{\alpha,j}(k)$ given in Ref. 3, gives

$$\int_0^a dr\,\phi_l(k,r)\chi(k) = \sum_{\alpha=-1}^{+1}\Big(B_{\alpha,0}(k)\Big\{e^{i\alpha k a}\chi(a)(i\alpha k)^{-1} - \lim_{\epsilon\to 0}(i\alpha k)^{-1}{\sum_{j,\beta}}' \exp(i\alpha k R_{j,\beta}) \times[\chi(R_{j,\beta}+\epsilon)-\chi(R_{j,\beta}-\epsilon)]\Big\} + (i\alpha k)^{\xi-1}\tilde{B}_{\alpha,0}(k) \times[1+O\,|k|^{-1})]\Big), \tag{3.10}$$

where $B_{\alpha,0}(k)$ was defined in Eq. (2.23) and $\tilde{B}_{\alpha,0}(k)$ also satisfies an equation of the form (2.23), but with $b(l,j)$ replaced by a new set of constants $\tilde{b}(l,j)$. The prime on the sum over j and β indicates that the sum is restricted to those values of j and β for which $R_{j,\beta}\leqslant R_{J,b}$. We note that the R_0 dependence of the interior integral cancels against the R_0 dependence of the exterior integral, as it clearly must since R_0 is an arbitrary matching point and both $V(r)$ and $\chi(r)$ are assumed to have derivatives of all order at R_0.

To complete the evaluation of the residues one needs the asymptotic forms of $f_l(k)$ and $df_l(-k)/dk$ and also the asymptotic pole positions. The asymptotic form of $f_l(k)$ is given by Eq. (2.24a), the asymptotic form of $df_l(-k)/dk$ evaluated at a zero of $f_l(-k)$ is derived in Appendix A and given by Eq. (A6), and finally, the asymptotic form for the pole positions is given by Eq. (2.29). Substituting these asymptotic expressions into the equation for the residue of $I_{12}(k,a)$ at $k=k^j_m$ and collecting terms yields

$$\mathscr{R}^j_m(a)(k-k^j_m)^{-1} = \sum_\kappa C^j_{a,\kappa}(a)(i\zeta_j M)^{\pi^j_\kappa(a)} \times\exp(i\phi^j_\kappa(a)M) \times[1+O(M^{-1}\ln|M|)] \tag{3.11}$$

for every fixed value of k as $|M|\to\infty$, where the sum on κ has a finite number of terms and the amplitudes $C^j_{\alpha,\kappa}(a)$, the powers $\pi^j_\kappa(a)$ and the phase factors $\phi^j_\kappa(a)$ are each independent of M. The indices $\alpha=+1$ and $\alpha=-1$ correspond to $M>0$ and $M<0$, respectively. The derivation also reveals that $|C^j{}_{+1,\kappa}(a)| = |C^j{}_{-1,\kappa}(a)|$ for every κ, and that $C^j{}_{+1,\kappa}(a) = C^j{}_{-1,\kappa}(a)$ for every κ for which both l_k and $m_{i,\kappa}$ vanish. A more detailed expression for these coefficients is given in Appendix B.

Before we examine Eq. (3.11) in more detail, we address ourselves to a complication touched upon in the previous section, the terms of the Mittag–Leffler series were labelled by a single index n, which we associated with the pole k_n, while our asymptotic expression for the pole positions, and hence for the residue poles, has two labels j, which labels the family, and M, which labels the various members of the family. It would be extremely difficult and of doubtful value to identify each pole k^j_m with a pole k_n. So we shall use a different approach to investigate the convergence properties of the Mittag–Leffler series. Suppose, for the sake of argument, that the pair of infinite sums

$$S^j{}_{+1} \equiv \sum_{M>0} \mathscr{R}^j{}_M(k-k^j{}_M)^{-1} \quad \text{and}$$

$$S^j{}_{-1} = \sum_{M>0} \mathscr{R}^j{}_{-M}(k-k^j{}_{-M})^{-1}, \tag{3.12}$$

converge for each and every j family. Then by the Cauchy condition,[27] for every $\epsilon>0$ there exists an N_ϵ such that

$$\left|\sum_{M=k}^{N} \frac{\mathscr{R}^j{}_{\pm M}}{k-k^j{}_{\pm M}}\right| < \frac{\epsilon}{2\tau},$$

for every j as long as $N>K>N_\epsilon$. What does this imply about the convergence of the Mittag–Leffler series? To answer this question we consider a partial sum for the Mittag–Leffler series and note that it can be rewritten as

$$\sum_{m=K}^{N}\left(\frac{\mathscr{R}_m}{k-k_m}+\frac{\mathscr{R}_{-m}}{k-k_{-m}}\right) = \sum_{j=1}^{\tau}\left(\sum_{m=K^j}^{N^j}\frac{\mathscr{R}^j_m}{k-k^j_m}+\sum_{m=K^j}^{N^j}\frac{\mathscr{R}^j{}_{-m}}{k-k^j{}_{-m}}\right).$$

The reordering of terms is valid since we are dealing with finite sums. By the nature of the mapping from k_m to k^j_m it follows that if one takes K sufficiently large, say $K>M_\epsilon$, then $K^j>N_\epsilon$ for every j, thus to each $\epsilon>0$ there exists an M_ϵ such that if $N>K>N_\epsilon$ then

$$\left|\sum_{m=K}^{N}\left(\frac{\mathscr{R}_m}{k-k_m}+\frac{\mathscr{R}_{-m}}{k-k_{-m}}\right)\right| \leqslant \sum_{j=1}^{\tau}\sum_{\alpha=-1}^{+1}\left|\sum_{M=K^j}^{N^j}\frac{\mathscr{R}^j_{\alpha M}}{k-k^j_{\alpha M}}\right| < \epsilon.$$

Hence, the Mittag–Leffler expansion converges if both $S^j{}_{+1}$ and $S^j{}_{-1}$ converge for every j. One can go one step further. If one defines infinite sums

$$S^j_{\alpha,\kappa}(a) = \sum_{M>0} C^j_{\alpha,\kappa}(a)\,(i\alpha\zeta M)^{\pi^j_\kappa(a)}\exp[i\alpha\phi^j_\kappa(a)M]\,, \quad (3.13)$$

if follows that S^j_α will converge if $S^j_{\alpha,\kappa}$ converges for each value κ in the sum on the right-hand side of Eq. (3.11). Thus, if $S^j_{+1,\kappa}$ and $S^j_{-1,\kappa}$ converge for every j and κ, the Mittag–Leffler expansion will converge. The converse, however, is not true, that is convergence of the Mittag–Leffler series does not imply convergence of $S^j_{\alpha,\kappa}$ for every j and κ. We shall see examples of this later on, but for the moment, let us determine the convergence conditions for the series $S^j_{\pm 1,\kappa}(a)$. We begin by examining Eq. (3.11) in greater detail.

Each of the phases $\phi^j_\kappa(a)$ and each of the powers $\pi^j_\kappa(a)$ has the general form

$$\phi^j_\kappa(a) = \zeta_j X^j_\kappa \quad (3.14a)$$

and

$$\pi^j_\kappa(a) = \lambda_j X^j_\kappa + \alpha_{N(j)} - {\sum_{i=1}^{n+1}}'' n^j_{\kappa,i}(\alpha_1 - \alpha_{N(j)}) - l^j_\kappa\,(2-\gamma) + \sum_{i=1}^{2} m^j_{\kappa,i}\xi_i - 4\,, \quad (3.14b)$$

with

$$X^j_\kappa = \beta^\kappa - \beta_{N(j)} + {\sum_{i=1}^{n+1}}'' n^j_{\kappa,i}(\beta_i - \beta_{N(j)})\,, \quad (3.14c)$$

where the double prime on the sums means that $i = N(j)$ and $i = N(j-1)$ are excluded from the sums, the coefficient $n^i_{\kappa,i}$ are nonnegative integers, l^j_κ is an integer ranging from 0 to m, with m the greatest integer $\leqslant (2-\gamma)^{-1}$, and $m^j_{\kappa,i} = 0$ or 1. The quantities $\lambda_j, \zeta_j, \beta_i, \beta_{N(j)}, \alpha_i, \alpha_{N(j)}, \gamma$, and ξ_i where defined in Sec. 2.

The various terms of the sum over κ in Eq. (3.11) fall into three basic classes which we list below:

(a) In this class $m^j_{\kappa,1} = m^j_{\kappa,2} = 0$ and β^κ times R can be any one of the following; a, $a \pm r^j_p$, $r^1_p \pm r^2_q$, $-a$, $-a \pm r^j_p$, $-r^1_p \pm r^2_p$, or 0, where r^j_p is the position of the pth discontinuity of $\chi_i(r)$. The i can be either 1 or 2 in r^j_p, but p and q are restricted to values for which $r^j_p < a$ and $r^2_q < a$.

(b) In this class either $m^j_{\kappa,1} = 1$ and $m^j_{\kappa,2} = 0$ or $m^j_{\kappa,1} = 0$ and $m^j_{\kappa,2} = 1$, while $\beta^\kappa R$ may be either $\pm a/2$ or $\pm r^j_p/2$, where again i may be either 1 or 2, but p is restricted to values for which $r^j_p < a$. If $\beta^\kappa R = \pm r^j_p/2$ then $m^j_{\kappa,1} = 0$.

(c) In this class $m^j_{\kappa,1} = m^j_{\kappa,2} = 1$ and $\beta^\kappa = 0$. If one defines $P(i,j,\lambda)$ by

$$P(i,j,\lambda) = \lambda(\beta_i - \beta_{N(j)}) - \alpha_i + \alpha_{N(j)}\,, \quad (3.15)$$

as was done in Ref. 3, then

$$\pi^j_\kappa(a) = \lambda_j\beta^\kappa + P(0,j,\lambda_j) + {\sum_{i=1}^{n+1}}'' n^j_{\kappa,i}P(i,j,\lambda_j) - l^j_\kappa(2-\gamma) + \sum_{i=1}^{2} m^j_{\kappa,i}\xi_i - 4\,. \quad (3.16)$$

From the theorem established in Sec. 3 of Ref. 3 it follows that $P(i,j,\lambda_j) < 0$ for every i in the sum while $P(0,j,\lambda_j) \leqslant 0$ for every j (the quality only holds for $j = 1$). Since $l_\kappa(2-\gamma) \geqslant 0$ and $\xi_i < \frac{1}{2}$, one then finds that

$$\pi^j_\kappa(a) - \lambda_j\beta^\kappa < \tfrac{1}{2}\sum_{i=1}^{2} m_{\kappa,i} - 4 \leqslant -3\,, \quad (3.17)$$

for every κ. If in addition $\beta^\kappa \leqslant 0$, then the infinite series $S^j_{\alpha,\kappa}(a)$ will be absolutely convergent. Therefore to complete our study of constraints on a which guarantee the convergence of $S^j_\alpha(a)$, we need only examine those series $S^j_{\alpha,\kappa}(a)$ associated with values of $\beta^\kappa > 0$. This eliminates from further consideration all values of κ associated with terms in class (c) and many values of κ associated with terms in classes (a) and (b) as well.

For those series $S^j_{\alpha,\kappa}(a)$ that survive after those for which $\beta^\kappa \leqslant 0$ have been eliminated, a necessary condition for convergence[28] is $\pi^j_\kappa(a) < 0$. From Eq. (3.15) and the properties of the $P(i,j,\lambda_j)$, it follows that the maximum value of $\pi^j_\kappa(a)$ for κ in class (a), occurs when $l_k = n_{\kappa,i} = 0$ for every i in the sum and $\beta^\kappa = a/R$. Calling this term the $\kappa = 1$ term, one has

$$\pi^j_1(a) = \lambda_j(a/R - \beta_{N(j)}) + \alpha_{N(j)} - 4\,. \quad (3.18)$$

Now $\pi^j_1(a) < 0$ implies that a must satisfy

$$a/R < \beta_{N(j)} - (\alpha_{N(j)} - 4)/\lambda_j\,, \quad (3.19)$$

or else $S^j_{\alpha,1}(a)$ will diverge. If Eq. (3.19) is satisfied and the phase $\phi^j_\kappa(a)$ is not an integral multiple of 2π, the series will converge by the Dirichlet test.[29] If, for the moment, we assume that none of the phases $\phi^j_\kappa(a)$ is an integral multiple of 2π, then $S^j_{\alpha,\kappa}(a)$ will converge for every κ belonging to class (a), since $\pi^j_\kappa(a) < \pi^j_1(a) < 0$ for every $\kappa \neq 1$ in class (a).

For κ in class (b), one sees that the maximum power is

$$\pi^j_\kappa(a) = \lambda_j[a/(2R) - \beta_{N(j)}] + \alpha_{N(j)} + \xi - 4\,, \quad (3.20)$$

with ξ as defined in Eq. (2.27b). If a satisfies Eq. (3.19), then by Eqs. (2.31), (3.15), and (3.18), Eq. (3.20) becomes

$$\pi^j_\kappa(a) = [\pi^j_1(a) + P(0,j,\lambda_j)]/2 + \xi - 2 < -\tfrac{3}{2}\,,$$

so every one of the series $S^j_{\alpha,\kappa}(a)$ with κ in class (b) is absolutely convergent. Thus, if a satisfies Eq. (3.19) and $\phi^j_\kappa(a) \neq 0 \pmod{2\pi}$ [i.e., $\phi^j_\kappa(a)$ does not equal an integral multiple of 2π] for every κ in class (a) for which $-1 \leqslant \pi^j_\kappa(a) < 0$, then $S^j_\alpha(a)$ will also converge. It may seem that we have neglected terms associated with the factor $O(M^{-1}\ln|M|)$ on the right-hand side of Eq. (3.11), but these terms are of order $O(M^{-1-\pi^j_\kappa(a)}\ln|M|) = o(M^{-1})$, so they too give rise to an absolutely convergent sum when Eq. (3.19) is satisfied.

If $a/R < \beta^c_j$, with β^c_j as defined in Sec. 2, then not only is Eq. (3.19) satisfied for every j, but the Mittag–Leffler series converges. Hence, one gets convergence even when one or more of the phases $\phi^j(a)$ equals 0 (mod 2π). In Appendix B it is shown that if $\phi^j_\kappa(a) = 0 \pmod{2\pi}$ and $-1 \leqslant \pi^j_\kappa(a) < 0$, then either the coefficient $C^j_{\alpha,\kappa}(a)$ vanishes or there exists a term in a neighboring family $j' = j \pm 1$ whose phase also equals 0 (mod 2π) and whose power equals $\pi^j_\kappa(a)$. Taking the ratio of the two terms one obtains

$$C^j_{\alpha,\kappa}(a)/C^{j'}_{\alpha,\kappa'}(a) = -\zeta_j/\zeta_{j'}\,. \quad (3.21)$$

But the right-hand side of Eq. (3.21) is, apart from sign, just the ratio of the spacing between poles of the j family to the spacing between poles of the j' family. Therefore, if one combines the terms of the series $S^j_{\alpha,\kappa}$ and $S^{j'}_{\alpha,\kappa'}$ to form a new series whose terms are ordered according to the magnitude

of the poles, as is done in ordering terms of the Mittag–Leffler series, then the new series converges.

Now that we have examined the Mittag–Leffler series in some detail we are in a position to consider the limiting process that leads to the "completeness" series.

4. COMPLETENESS

We shall begin this section by determining the asymptotic form of $I_{12}(k,a)$ as $k\to\infty$ in $\mathscr{E}$, whre $\mathscr{E}$ is a domain in the complex k plane, containing all points for which $\Theta_0 \leqslant \arg k \leqslant \pi - \Theta_0$ with $0 < \Theta_0 < \pi/2$, and $|k| \geqslant X$, where X is sufficiently large so that all bound state poles have $|k_i| < X$. For the class of potentials being considered, X is finite. From Eqs. (2.4), (2.5), and (2.26) it can be seen that

$$I_{12}(k,a) = \sum_{i=1}^{4} F_i(k)\,, \tag{4.1}$$

where the $F_i(k)$ are defined by Eq. (2.25). Taking $R_0 = [\ln(|k|R)]^{-1}$ in Eq. (2.25) and employing the asymptotic form of $G_l^{(+)}(k;r,r)$ given by Newton,[21] i.e.,

$$\begin{aligned} G_l^{(+)}(k;r,r') &= (-1)^{l+1}k^l\phi_l(k,r_<) \\ &\quad\times f_l(-k,r_>)/f_l(-k) \\ &= (2ik)^{-1}[e^{ik(r_>-r_<)} - (-1)^l e^{ik(r+r')} \\ &\quad + o(e^{-|\nu|(r_>-r_<)})]\,, \end{aligned} \tag{4.2}$$

for all r and $r' \geqslant R_0$, where $\nu = \mathrm{Im}k$, one can evaluate the asymptotic form of $F_1(k)$. Using the bounds on $|G_l^{(+)}(k';r;r')|$ given by Newton and the bounds on $|\chi_i(r)|$ given by Eq. (2.27a), one can place bounds on the contributions of $F_2(k)$, $F_3(k)$, and $F_4(k)$ to $I_{12}(k,a)$. Combining these results, one finds that

$$I_{12}(k,a) = k^{-2}\int_0^a dr\,\chi_1^*(r)\chi_2(r)\,(1+o(1))\,, \tag{4.3}$$

as $k\to\infty$ in $\mathscr{E}$. Therefore,

$$\lim_{k\to\infty} k^m I_{12}(k,a) = \delta_{m,2}\langle\chi_1|P_a|\chi_2\rangle\,, \tag{4.4a}$$

for $m = 1$ and 2 and k in $\mathscr{E}$, where

$$\langle\chi_1|P_a|\chi_2\rangle \equiv \int_0^a dr\,\chi_1^*(r)\chi_2(r)\,. \tag{4.4b}$$

Next, we apply the same limiting process to the pole expansion of $I_{12}(k,a)$, i.e., we consider

$$\lim_{\substack{k\to\infty \\ (k\text{ in }\mathscr{E})}} \left[\sum_{n=1}^{N}\frac{\mathscr{R}_n(a)}{k-k_n} + \sum_{n=N+1}^{\infty}\left(\frac{\mathscr{R}_n(a)}{k-k_n} + \frac{\mathscr{R}_{-n}(a)}{k-k_{-n}}\right)\right]. \tag{4.5}$$

Suppose, for the sake of argument, that one can interchange the order of the limit and the sum. For the $m = 1$ case one would obtain

$$\sum_{n=0}^{N}\mathscr{R}_n(a) + \sum_{n=N+1}^{\infty}[\mathscr{R}_n(a) + \mathscr{R}_{-n}(a)] = 0\,. \tag{4.6}$$

Continuing on with the supposition, if one multiplies Eq. (4.6) by k and Eq. (3.1) by k^2, adds the resulting equations and combines terms to form a single infinite sum, he will obtain

$$\begin{aligned} k^2 I_{12}(k,a) &= \sum_{n=1}^{N}\frac{k_n\mathscr{R}_n(a)}{1-k^{-1}k_n} \\ &\quad + \sum_{n=N+1}^{\infty}\left(\frac{k_n\mathscr{R}_n(a)}{1-k^{-1}k_n} + \frac{k_{-n}\mathscr{R}_{-n}(a)}{1-k^{-1}k_{-n}}\right). \end{aligned} \tag{4.7}$$

Assuming further, that the ordering of the limit and the sum can once again be interchanged, Eq. (4.4) then gives us

$$\begin{aligned} \langle\chi_1|P_a|\chi_2\rangle &= \sum_{n=1}^{N} k_n\mathscr{R}_n(a) \\ &\quad + \sum_{n=N+1}^{\infty}[k_n\mathscr{R}_n(a) + k_{-n}\mathscr{R}_{-n}(a)]\,. \end{aligned} \tag{4.8}$$

Substituting Eqs. (3.2a) and (3.2b) into Eq. (4.8) gives us the desired completeness relationship for the bound and resonance states.

It is now our task to determine the range of values of a for which the assumed interchanges are valid. Since the derivation will involve a number of steps, we begin with a brief outline of the procedure.

(i) They asymptotic form of the terms of the series is established.

(ii) The range of a for which the series is uniformly convergent in the domain $\mathscr{E}$ is determined.

(iii) The behavior of the terms of the series as $k\to\infty$ in $\mathscr{E}$ is examined.

(iv) Finally, by combining the results of steps (ii) and (iii) it is shown that one can interchange the order of the sum and limit for any a for which the series is uniformly convergent in $\mathscr{E}$.

Step (i) is easily accomplished. The two series of concern are

$$\begin{aligned} S^{(m)}(k,a) &\equiv \sum_{n=0}^{N}\frac{(k_n)^{m-1}\mathscr{R}_n(a)}{1-k^{-1}k_n} \\ &\quad + \sum_{n=N+1}^{\infty}\left(\frac{(k_n)^{m-1}\mathscr{R}_n(a)}{1-k^{-1}k_n} + \frac{(k_{-n})^{m-1}\mathscr{R}_{-n}(a)}{1-k^{-1}k_{-n}}\right) \end{aligned} \tag{4.9}$$

with $m = 1$ and 2. $S^{(1)}(k,a)$ corresponds to the series of Eq. (4.5), and $S^{(2)}(k,a)$ corresponds to the series of Eq. (4.7). From Eqs. (3.11) and (2.29) it is seen that the asymptotic form of the terms of $S^{(m)}(k,a)$ expressed in the (jM) representation are

$$\begin{aligned} &\frac{(k^j{}_m)^{m-1}\mathscr{R}^j{}_M(a)}{1-k^{-1}k^j{}_M} + \frac{(k^j{}_{-m})^{m-1}\mathscr{R}^j{}_{-M}(a)}{1-k^{-1}k^j{}_{-M}} \\ &\quad = \sum_{\alpha=-1}^{+1}\sum_{\kappa}\mathscr{A}^{m,j}_{\alpha,\kappa,M}(k,a)\,, \end{aligned} \tag{4.10a}$$

with

$$\begin{aligned} \mathscr{A}^{m,j}_{\alpha,\kappa,M}(k,a) &= D^{m,j}_{\alpha,\kappa}(a)\,(i\alpha\zeta_j M)^{\pi^j_\kappa(a)+m}\exp(i\alpha\phi^j_\kappa(a)M) \\ &\quad\times(1-k^{-1}k^j_{\alpha m})^{-1} \\ &\quad\times[1+O(M^{-1}\ln M)]\,, \end{aligned} \tag{4.10b}$$

$$D^{m,j}_{\alpha,\kappa}(a) = -i^{-m}C^j_{\alpha,\kappa}(a)\,. \tag{4.10c}$$

As a preliminary to step (ii) we shall derive bounds on $|1-k^{-1}k^j_m|^{-1}$ and $|(1-k^{-1}k^j_M)^{-1}$

$-(1-k^{-1}\zeta_j M)^{-1}|$ for all j, all k in $\mathscr{E}$ and every M for which $|M|\geqslant M^c\gg 1$. Let $\psi=\arg(k/k^j_m)$, then

$$|1-k^{-1}k^j_M|^2 = |e^{i\psi}-|k^j_M/k||^2 = (\cos\psi-|k^j_m/k|)^2+\sin^2\psi\geqslant\sin^2\theta_0\,,$$

therefore

$$|1-k^{-1}k^j_M|^{-1}\leqslant(\sin\Theta_0)^{-1}\,, \tag{4.11}$$

for every j, every k in $\mathscr{E}$, and every $|M|\geqslant M^c$. To obtain the second the second inequality, we first note from Eqs. (2.29) and (4.10d) that

$$(1-k^{-1}k^j_M)^{-1}-(1-k^{-1}\zeta_j M)^{-1} = -i(1-k^{-1}k^j_m)^{-1}[\lambda_j\ln|M|+O(1)]\times(k-\zeta_j M)^{-1}$$

while

$$|k-\zeta_j M|^2 = ||k|e^{i\theta}-\zeta_j M|^2 = (|k|-\zeta_j M\cos\theta)^2+(\zeta_j M\sin\theta)^2 \geqslant(\zeta_j M\sin\theta_0)^2\,,$$

therefore

$$|(1-k^{-1}k^j_M)^{-1}-(1-k^{-1}\zeta_j M)^{-1}| \leqslant\text{const}\times|M|^{-1}\ln|M|\,, \tag{4.12}$$

for every $j\leqslant\tau$, k in $\mathscr{E}$ and $|M|\geqslant M^c\gg 1$.

The approach we shall take in determining the condition for uniform convergence of the series $S^{(m)}(k,a)$, will be similar to that used in determining the condition for convergence of the Mittag–Leffler series in Sec. 3, we shall reorder the terms of the series, writing $S^{(m)}(k,a)$ as a sum of a finite number of subseries, each subseries consisting of all terms of $S^{(m)}(k,a)$ that have the same value of j,κ, and α. A straightforward generalization of the argument presented in Sec. 3, then shows that $S^{(m)}(k,a)$ will be uniformly convergent in $\mathscr{E}$ if each of the subseries is uniformly convergent in $\mathscr{E}$. From Eqs. (4.10) and (4.11) it is seen that a necessary condition for the convergence of the subseries $\Sigma_M(\mathscr{A}^{m,j}_{\alpha,\kappa,M})$ is that $\pi^j_\kappa(a)+m<0$. This inequality will be satisfied for every value of j,κ, and α and for every k in $\mathscr{E}$, if

$$\frac{a}{R}<\beta^c(m)\equiv\min_{j(j\geqslant 1)}\{\beta_{N(j)}+\lambda_j^{-1}[4-\alpha_{N(j)}-m]\}\,. \tag{4.13}$$

Since $\alpha_j\geqslant 2$ for every $j\geqslant 1$, it follows that the minimum is achieved at $j=1$ if $m\geqslant 2$, i.e.,

$$\beta^c(m)=(4-m)\beta_{N(j)}/\alpha_{N(j)}\,,\quad\text{for } m=2 \text{ and } 3\,. \tag{4.14}$$

Lt us now asume that $a/R\leqslant\beta^c(m)-\epsilon_0$, where ϵ_0 is an arbitrary small positive number, and let us define

$$\mathscr{B}^{m,j}_{\alpha,\kappa,M}(k,a) = D^{m,j}_{\alpha,\kappa}(a)\,(i\alpha\zeta_j M)^{\pi^j_\kappa(a)+m}\exp[i\alpha\phi^j_\kappa(a)M]\times(1-k^{-1}\zeta_j\alpha M)^{-1}\,, \tag{4.15a}$$

$$\mathscr{C}^{m,j}_{\alpha,\kappa,M}(k,a)=\mathscr{A}^{m,j}_{\alpha,\kappa,M}(k,a)-\mathscr{B}^{m,j}_{\alpha,\kappa,M}(k,a)\,. \tag{4.15b}$$

It then follows from Eqs. (4.10), (4.12), and (4.15) that

$$|\mathscr{C}^{m,j}_{\alpha,\kappa,M}(k,a)|<CM^{-1-2\epsilon_0}\ln M<CM^{-1-\epsilon_0}\,, \tag{4.16}$$

for every j,κ,α, and for every k in $\mathscr{E}$. (In deriving Eq. (4.16) we have also used the fact that $\lambda_j\geqslant 2$ for every $j\geqslant 1$.) According to Eq. (4.16) each of the series $\Sigma_M(\mathscr{C}^{m,j}_{\alpha,\kappa,M})$ is uniformly and absolutely convergent in $\mathscr{E}$. Hence, uniform convergence of $S^{(m)}(k,a)$ will follow if one can show that each of the subseries $\Sigma_M(\mathscr{B}^{m,j}_{\alpha,\kappa,M})$ is uniformly convergent in $\mathscr{E}$. For conveneince we shall refer to these sereis as $\mathscr{B}$ series.

The key for the analysis of the $\mathscr{B}$ series is provided by the theorem of Appendix C. To apply the theorem one takes the quantities a_M, z, and p defined in the appendix to be

$$a_M = D^{m,j}_{\alpha,\kappa}(a)\,(i\alpha\zeta_j M)^{\pi^j_\kappa(a)+M}\exp[i\alpha\phi^j_\kappa(a)M]\,, \tag{4.17a}$$

$$|z|=\zeta_j/|k|\,, \tag{4.17b}$$

$$\arg z=-\arg k+(\alpha-1)\pi/2\,, \tag{4.17c}$$

and $p=1$. We again assume that $a/R\leqslant\beta^c(m)-\epsilon_0$. If we further assume that $\phi^j_\kappa(a)\neq 0(\text{mod }2\pi)$, then the series $\Sigma_M a_M$ converges by the Dirichlet test,[29] and then the theorem of Appendix C tells us that the series $\Sigma_M[\mathscr{B}^{m,j}_{\alpha,\kappa,M}(k,a)]$ is uniformly convergent for all k in $\mathscr{S}_\alpha$, where $\mathscr{S}_{+1}$ consists of all points of the k plane for which $\theta_0\leqslant\arg k\leqslant 2\pi-\theta_0$, and $\mathscr{S}_{-1}$ consists of all points for which $\theta_0-\pi\leqslant\arg k\leqslant\pi-\theta_0$. Since the domain $\mathscr{E}$ is contained in the intersect of $\mathscr{S}_{+1}$ and $\mathscr{S}_{-1}$ we conclude that both of the series, i.e., the one with $\alpha=+1$ and the one with $\alpha=-1$, are uniformly convergent in $\mathscr{E}$.

If there are no (j,κ) for which is $\phi^j_\kappa(a)=0(\text{mod}2\pi)$ and $\pi^j_\kappa(a)\geqslant -1-m$, then each of the $\mathscr{B}$ series is uniformly convergent in $\mathscr{E}$, and it follows from Eq. (4.15b) and the uniform and absolute convergence of the $\mathscr{C}$ series that each of the $\mathscr{A}$ series is also uniformly convergent in $\mathscr{E}$. Finally, as mentioned earlier, the uniform convergence of each of the $\mathscr{A}$ series implies the uniform convergence of the entire series $S^{(m)}(k,a)$ for k in $\mathscr{E}$.

The case in which one of the phases $\phi^j_\kappa(a)=0(\text{mod }2\pi)$ and $-1\leqslant\pi^j_\kappa(a)+m<0$, is treated in Appendix B, where it is shown that the divergent series that results can be combined with a second divergent $\mathscr{B}$ series to form a resultant series, made up of terms from the two original series ordered according to $|k_n|$, that is uniformly convergent in $\mathscr{E}$. Thus, the series $S^{(m)}(k,a)$ will always be uniformly convergent in $\mathscr{E}$ as long as $a/R\leqslant\beta^c(m)-\epsilon_0$.

To facilitate the discussion of step (iii), we define a partial sum $S^{(m)}_I(k,a)$ by

$$S^{(m)}_I(k,a)\equiv\sum_n^N\frac{(k_n)^{m-1}\mathscr{R}_n(a)}{1-k^{-1}k_n}+\sum_{n=N+1}^{I}\sum_{\alpha=-1}^{+1}\frac{(k_{\alpha n})^{m-1}\mathscr{R}_{\alpha n}(a)}{1-k^{-1}k_{\alpha n}}\,, \tag{4.18}$$

where $I>N$, but is otherwise arbitrary. Since $S^{(m)}_I(k,a)$ is a finite sum one can immediately pass to the limit $k\to\infty$ in $\mathscr{E}$, obtaining

$$S^{(m)}_I(\infty,a)=\sum_{n=1}^N(k_n)^{m-1}\mathscr{R}_n(a)+\sum_{n=N+1}^{I}\sum_{\alpha=-1}^{+1}(k_{\alpha n})^{m-1}\mathscr{R}_{\alpha n}(a)\,. \tag{4.19}$$

If $a/R<\beta^c(m)-\epsilon_0$; then there exists a constant C such that

$$|(k_n)^{m-1}\mathscr{R}_n(a)| \leqslant C\,, \tag{4.20}$$

for all n, both positive and negative. Therefore

$$|S_I^{(m)}(k,a) - S_I^{(m)}(\infty,a)|$$

$$= \left| \sum_{n=1}^{N} \frac{(k_n)^m \mathscr{R}_n(a)}{k-k_n} + \sum_{n=N+1}^{\infty} \sum_{\alpha=-1}^{+1} \frac{(k_{\alpha n})^m \mathscr{R}_{\alpha n}(a)}{k-k_{\alpha n}} \right|$$

$$< (2CI) \times \max_n \left| k_n (k-k_n)^{-1} \right|\,, \tag{4.21}$$

where the maximum is taken over all n for which $|k_n| \leqslant k_I$. Let us now define a neighborhood of infinity in $\mathscr{E}$, call it $\mathscr{N}_{I,\epsilon}$, to be the set of all points in $\mathscr{E}$ for which

$$|k| > (1 + 6CI/\epsilon)|k_I|\,, \tag{4.22}$$

where ϵ is an arbitrary positive number, then

$$|S_I^{(m)}(k,a) - S_I^{(m)}(\infty,a)| < \epsilon/3\,, \tag{4.23}$$

for every k in $\mathscr{N}_{I,\epsilon}$ as long as $a/R < \beta^c(m) - \epsilon_0$.

We can now proceed to step (iv). The interchange of limit and sum we wish to establish can be expressed as follows:

$$\lim_{k\to\infty (k \text{ in } \mathscr{N})} S^{(m)}(k,a) = \lim_{I\to\infty} S_I^{(m)}(\infty,a) = S^{(m)}(\infty,a)\,. \tag{4.24}$$

The last equality defines $S^{(m)}(\infty,a)$. From Eqs. (4.6), (4.8), and (4.19), one sees that $S^{(1)}(\infty,a)$ is just the series on the left-hand side of Eq. (4.6) and $S^{(2)}(\infty,a)$ is the series on the right-hand side of Eq. (4.8). Since the series $S^{(m)}(k,a)$ is uniformly convergent for every k in $\mathscr{E}$, to every $\epsilon > 0$, there exists an integer I such that

$$|S_J^{(m)}(k,a) - S^{(m)}(k,a)| < \epsilon/3\,,$$

for every k in $\mathscr{E}$, $J \geqslant I$, and $a/R < \beta^c(m) - \epsilon_0$. Hence

$$|S^{(m)}(k,a) - S^{(m)}(\infty,a)| \leqslant |S^{(m)}(k,a) - S_I^{(m)}(k,a)|$$
$$+ |S_I^{(m)}(k,a) - S_I^{(m)}(\infty,a)|$$
$$+ |S_I^{(m)}(\infty,a) - S^{(m)}(\infty,a)| < \epsilon\,,$$

for every k in $\mathscr{N}_{I,\epsilon}$ and $a \leqslant \beta^c(m) - \epsilon_0$. Therefore, Eq. (4.24) is valid whenever $a \leqslant \beta^c(m) - \epsilon_0$.

Since $S^{(1)}(k,a) = kI_{12}(k,a)$, one sees that Eqs. (4.6) and (4.7) are valid for every $a < \beta^c(1) - \epsilon_0$. It then follows that $S^{(2)}(k,a) = k^2 I_{12}(k,a)$ for every every $a < \beta^c(1) - \epsilon_0$, and that Eq. (4.8) holds for every $a < \beta^c(2) - \epsilon_0$. Since these are the relations we wished to establish we now write them out in a more complete form.

$$\sum_{n=1}^{N} \langle \chi_1 | P_a | \Phi_n \rangle \langle \Phi_n | P_a | \chi_2 \rangle (2k_n)^{-1}$$
$$+ \sum_{n=N+1}^{\infty} \left[\frac{\langle \chi_1 | P_a | \Phi_n \rangle \langle \Phi_{-n} | P_a | \chi_2 \rangle}{2k_n} \right.$$
$$\left. + \frac{\langle \chi_1 | P_a | \Phi_{-n} \rangle \langle \Phi_n | P_a | \chi_2 \rangle}{2k_{-n}} \right]$$
$$= 0\,, \tag{4.25}$$

for $a/R < \beta^c(1)$, and

$$\tfrac{1}{2}\left\{ \sum_{n=1}^{N} \langle \chi_1 | P_a | \Phi_n \rangle \langle \Phi_n | P_a | \chi_2 \rangle \right.$$
$$\left. + \sum_{n=N+1}^{\infty} [\langle \chi_1 | P_a | \Phi_n \rangle \langle \Phi_{-n} | P_a | \chi_2 \rangle \right.$$
$$\left. + \langle \chi_1 | P_a | \Phi_{-n} \rangle \langle \Phi_n | P_a | \chi_2 \rangle] \right\} = \langle \chi_1 | P_a | \chi_2 \rangle\,, \tag{4.26}$$

for every $a < 2R_{N(1)}/\alpha_{N(1)}$.

The class of functions to which $\chi_1(r)$ and $\chi_2(r)$ belong has been defined so broadly that any number of complete sets of functions, such as the radial wavefunctions for a harmonic oscillator potential, will belong to the class. Hence, Eqs. (4.25) and (4.26) will hold for any two square integrable functions, and one concludes that[30]

$$\tfrac{1}{2}\left\{ \sum_{j=1}^{N} P_a |\Phi_j\rangle \langle \Phi_j | P_a \right.$$
$$\left. + \sum_{j=N+1}^{n} P_a [|\Phi_j\rangle \langle \Phi_{-j}| + |\Phi_{-j}\rangle \langle \Phi_j|] P_a \right\} \to P_a\,, \tag{4.27}$$

for $a/R \leqslant \beta^c(2) - \epsilon_0$, and

$$\sum_{j=1}^{N} \frac{P_a |\Phi_j\rangle \langle \Phi_j | P_a}{2k_j}$$
$$+ \sum_{j=N+1}^{n} P_a \left[\frac{|\Phi_j\rangle \langle \Phi_{-j}|}{2k_j} + \frac{|\Phi_{-j}\rangle \langle \Phi_j|}{2k_{-j}} \right] P_a \to 0\,, \tag{4.28}$$

for $a/R \leqslant \beta^c(1) - \epsilon_0$. Where the arrows mean that the operators on the left-hand side of the arrow weakly converge to the operator on the right-hand side as $n \to \infty$.

A resonant state expansion of the wavefunction $\chi_i(r)$ can be constructed by the same procedure used to construct the series for $\langle \chi_1 | P_a | \chi_2 \rangle$. One first defines a function $J_i(k,r)$ by

$$J_i(k,r) = \int_0^a dr'\, G_l^{(+)}(k;r,r')\chi_i(r')\,, \tag{4.29}$$

with $r < a$. He then multiplies $J_i(k,r)$ by k^2 and takes the limit $k \to \infty$ in $\mathscr{E}$, obtaining

$$\lim_{k\to\infty \text{ in } \mathscr{E}} [k^2 J_i(k,r)] = \chi_i(r)\,. \tag{4.30}$$

One next constructs the Mittag–Leffler series for $J_i(k,r)$, multiplies the series by k^2 and again lets $k \to \infty$ in $\mathscr{E}$. Upon equating the two limiting expressions one obtains

$$\chi_i(r) = \tfrac{1}{2}\left\{ \sum_{j=1}^{N} \phi_j(r) \langle \Phi_j | P_a | \chi_i \rangle \right.$$
$$\left. + \sum_{j=N+1}^{\infty} [\Phi_j(r) \langle \Phi_{-j} | P_a | \chi_i \rangle + \Phi_{-j}(r) \langle \Phi_j | P_a | \chi_i \rangle] \right\}\,, \tag{4.31}$$

which is valid for $r + a < 2R_{N(1)}/\alpha_{N(1)}$, or equivalently for $r < R_{N(1)}/\alpha_{N(1)}$. Note that the radius of convergence of the series for $\chi_i(r)$, Eq. (4.31), is less than half the radius of convergence of the series for $\langle \chi_i | P_a | \chi_2 \rangle$, Eq. (4.26).

5. COMMENTS AND CONCLUSIONS

Eqs. (4.26), (4.27), and (4.31) represent the "completeness" properties of the set of bound, virtual, and resonant states of the simple quantum mechanical system under consideration. Eq. (4.26), or equivalently, Eq. (4.27) is an ex-

pression of how completeness applies to matrix elements, and Eq. (4.31) is an expression of how completeness applies to wavefunctions. The completeness properties of these states differs from true completeness in several ways. First, resonant state expansions only hold within a finite volume of configuration space, this is a property shared with an expansion in terms of R-matrix states.[31] Perhaps this fact is not too surprising since, as Seigert pointed out,[5] resonant state wavefunctions bear some resemblance to Kapur–Peierls wavefunctions,[32] and the latter are just a particular choice of R-matrix wavefunctions. However, unlike the R-matrix case, the volume in which the expansion is valid is restricted by the physical properties of the system rather than being completely arbitrary. Furthermore, the bound-virtual-, and resonant-state wavefunctions are neither normalized or orthogonal over the finite region $0 \leqslant r \leqslant a$. Finally, the convergence of the series expansion of the operator P_a is only weakly convergent in contrast to strong convergence in the R-matrix case.

In addition to the completeness relations for the set of bound, virtual, and resonant states of the system under study, we have determined the convergence conditions of the Mittag–Leffler series for $I_{12}(k,a)$, i.e., for certain matrix elements of the Green's function. The convergence of the Mittag–Leffler series can be improved by one or more subtractions, as was pointed out by Gracía-Calderón[33] and Bang et al.[34] A simple analysis based upon the work of Sec. 3, in particular Eq. (3.14b), shows that the Mittag–Leffler series for $I_{12}(k,a)$ with p subtractions will be absolutely convergent for all $a \leqslant R$, if $p > \sigma - 1$, and that each additional subtraction will increase the rate of convergence.

One rather interesting result was obtained in Ref. 33, was further dealt with in Ref. 34, and bears mentioning here. Since the physical solution of Eq. (2.6) is related to the Green's function by

$$\psi_l^{(+)}(k, r) = (-1)^{l+1} k G_l^{(+)}(k; r, R)/w_l(-k, R) \tag{5.1}$$

for all $r < R$, then a resonant state expansion for $\psi_l^{(+)}(k, r)$ can immediately be obtained for $r < R$ by substituting the Mittag–Leffler series for $G_l^{(+)}(k; r, r')$ into the right-hand side of Eq. (5.1). Once again the convergence can be improved by including more and more subtractions.

For potentials with an exponential tail such as Yukawa or Woods–Saxon potentials, the Green's function will have cuts in the complex k plane in addition to the bound-, virtual-, and resonant-state poles. Thus, one cannot obtain a simple Mittag–Leffler series for $I_{12}(k,a)$ when such a potential is substituted for the cutoff potentials we have been considering. Perhaps, a generalized expansion could be obtained, but in all likelihood, it would contain an integral term, arising from the cuts, as well as the discrete sum over pole states. However, it has been argued that if one truncates these potentials at some large radius, the physical consequences will be negligible, even though the analytic properties may be drastically changed. Such a truncation would yield a potential falling in the class we have considered, in fact one with $\sigma = 0$. To the extent that the physical consequences of truncating a potential with an exponential tail are small, one can also truncate a cutoff potential for which $\sigma > 0$ at a radius $R - \epsilon$, with ϵ an arbitrarily small positive number, converting the potential into a physically equivalent potential with $\sigma = 0$.For potentials in this restricted but important class, one has the following results:

(i) The Mittag–Leffler series for $I_{12}(k,a)$ is absolutely convergent for each fixed real value of k and $a \leqslant R$.

(ii) The resonant-state expansion for $\langle \chi_i | P_a | \chi_2 \rangle$ converges for all $a < R$, and it is absolutely convergent if $a < R/2$.

(iii) The resonant state expansion for P_a covnerges weakly for all $a < R$.

(iv) The resonant state expansion for $\chi_i(r)$ converges for all $a < R/2$. The series is absolutely convergent if $a < R/4$. For a Woods–Saxon potential $V(r) = V_0/\{1 + \exp[(r - r_0)/a_0]\}$ that is truncated at $r = R$ with $R \gg r_0 \gg a_0$, the expansions are absolutely convergent up to a radius that is well out into the tail of the potential.

Finally, if one applies the results of Secs. 2 and 3 to study the convergence of the pole expansions of the partial wave components of the S matrix and fully-off-shell T matrix given by Eqs. (2.45) and (2.43), respectively, of Ref. 2., he finds that the series for $S_l(k)$ is uniformly and absolutely convergent for all real k, and that the series for $a_l(k;p,q)$ is uniformly and absolutely convergent for all real k, p, and q. In both cases these results hold for all potentials that fall in the class defined in Sec. 2.

ACKNOWLEDGMENTS

The author would like to thank Dr. P.R. Beesack, Dr. F. Fiala, and Dr. J.E. Hardy for discussions.

APPENDIX A

The derivation of the asymptotic form of $df_l(-k)/dk$ is complicated by the fact that one cannot in general differentiate an asymptotic expansion.[35] Fortunately, there is a way of avoiding this difficulty. One begins with the integral representation for $f_l(-k)$[21]

$$f_l(-k) = 1 + (-1)^l k^l \int_0^R dr\, \phi_l(k, r) V(r) w_l(-kr), \tag{A1}$$

where $w_l(x)$ is a Riccati–Hankel function. Applying d/dk to this equation gives

$$\begin{aligned} \frac{df_{l'}(-k)}{dk} &= \frac{l(f_{l'}(-k) - 1)}{k} \\ &\quad - (-1)^{l'} k^{l'} \int_0^R dr\, \phi_{l'}(k, r) r V(r) w_{l'}'(-kr) \\ &\quad + (-1)^{l'} k^{l'} \int_0^R dr\, w_{l'}(-kr) V(r) \frac{d\phi_{l'}(k, r)}{dk}, \end{aligned} \tag{A2}$$

where $w_l'(x) = dw_l(x)/dx$. Since we have already constructed the asymptotic expansion of $f_l(-k)$ and $\phi_l(k,r)$, and we also know the asymptotic forms of $w_l(-kr)$ and $w_l'(-kr)$, one can immediately evaluate the asymptotic form of the first pair of terms on the right-hand side of Eq. (A2). Before

the final integral can be evaluated one must construct the asymptotic expansion of $d\phi_l(k,r)/dk$. If one differentiates the radial Schrödinger equation for $\phi_l(k,r)$ with respect to k and converts the resulting differential equation into an integral equation he obtains

$$\frac{d\phi_l(k,r)}{dk} = -2k\int_0^r g_l(k;r,r')\phi_l(k,r')\,dr', \tag{A3a}$$

where

$$g_l(k;r,r') = (-1)^l(2ik)^{-1} \times [f_l(-k,r)f_l(k,r') - f_l(k,r)f_l(-k,r')]. \tag{A3b}$$

Upon replacing $f_l(k,r)$ and $f_l(-k,r)$ by their asymptotic expansions in Eq. (A3b) and substituting the resulting asymptotic expansions for $g_l(k;r,r')$ along with the asymptotic expansion of $\phi_l(k,r)$ into the integral on the right-hand side of Eq. (A3a) and integrating, one obtains the desired expansion for $d\phi_l(k,r)/dk$. As usual the asymptotic expansion changes form in each of the segments Σ_j. Finally, if the asymptotic expansion of $d\phi_l(k,r)/dk$ and $w_l(-kr)$ are inserted into the final integral of Eq. (A2) and the integral is evaluated one finds the asymptotic expansion of $df_l(-k)/dk$,

$$\frac{df_l(-k)}{dk} = 2iR\,[1+K_l(k)+O(k^{-1}\ln k)]\sum_{q=1}^{n+1} A_q\beta_q(iZ)^{-\alpha_q} \times\exp(i\beta_q Z)[1+O(k^{-1})], \tag{A4}$$

where $Z = 2kR$.

If k^j_M is a zero of $f_l(-k)$ belonging to the jth family, then

$$\sum_{q=0}^{n+1} A_q(iZ^j_M)^{-\alpha_q}\exp(i\beta_q Z^j_M) = 0, \tag{A5}$$

where $Z^j_M = 2k^j_M R$, $\alpha_0 = \beta_0 = 0$ and

$$A_O \equiv [1+K_l(k)+0(k\ln k)]/[1+K_l(-k) + O(k\ln k)].$$

Evaluating Eq. (A4) at $k = k^j_M$ and using Eq. (A5) to eliminate the $q = N(j-1)$ term gives

$$\left.\frac{df_l(-k)}{dk}\right|_{k=k^j_M} = 2i[R_{N(j)} - R_{N(j-1)}][1+K_l(k^j_M)+O(M\ln M)] \times(iZ^j_M)^{-\alpha_{N(j)}}A_{N(j)}\exp(i\beta_{N(j)}Z^j_M) \times\left(1+\sum_{\substack{q=0\\ q\neq N(j),N(j+1)}}^{n+1}(A_q/A_{N(j)})\frac{\beta_q-\beta_{N(j-1)}}{\beta_{N(j)}-\beta_{N(j-1)}} \times(iZ^j_M)^{\alpha_{N(j)}-\alpha_q}\exp[iZ^j_M(\beta_q-\beta_{N(j)})]\right). \tag{A6}$$

From a theorem which was established in Ref. 3, it follows that each of the terms in the sum over q on the right-hand side of Eq. (A6) tends to zero as $|M|\to\infty$, while the factor $(iZ^j_M)^{-\alpha_{N(j)}}\exp(i\beta_{N(j)}Z^j_M)$ tends to unity as $|M|\to\infty$ if $j=1$, or it tends to infinity as $|M|\to\infty$ if $j>1$.

APPENDIX B

In this Appendix we shall investigate the convergence properties of the Mittag–Leffler series and of the series $S^{(m)}(k,a)$ of Sec. 4 when one or more of the phases $\phi^j_\kappa(a) = 2\pi I$ for integral I. We know in the case of the Mittag–Leffler series for $I_{12}(k,a)$ that the series converges as long as $a/R < \beta^c_j$, no matter what the individual phases may be. Thus, if the Mittag–Leffler series converges, but one of the subseries $S^{(j)}_{\alpha,\kappa}(a)$, cf. Eq. (3.13), diverges because its phases equals $2\pi I$, then it is reasonable to assume that there is another divergent subseries in the set $\{S^{(j')}_{\alpha,\kappa'}(a)\}$ whose terms can be combined with the terms of the divergent series $S^{(j)}_{\alpha,\kappa}(a)$ to form a convergent series if the terms of the new series have been ordered according to the same rule that was used in ordering the terms of the Mittag–Leffler series.

To pursue this possibility let us assume that $\phi^j_\kappa(a) = 2\pi I$, where $\phi^{(j)}_{\alpha,\kappa}(a)$ is the phase factor for the terms of the series $S^{(j)}_{\alpha,\kappa}(a)$. We next examine the two series $S^{(j-1)}_{\alpha,\kappa'}(a)$ and $S^{(j+1)}_{\alpha,\kappa''}(a)$ generated by the two neighboring families of poles k^{j-1}_M and k^{j+1}_M, respectively, where the parameters defining these series are given by

$$\beta^{\kappa'} = \beta^\kappa,\quad m^{j-1}_{\kappa',i} = m^j_{\kappa,i},\quad l^{j-1}_{\kappa'} = l^j_\kappa,\quad n^{j-1}_{\kappa',N(j-2)} = 0,$$
$$n^{j-1}_{\kappa',N(j)} = -(I+N^j_\kappa),\ \text{and}\ n^{j-1}_{\kappa',i} = n^j_{\kappa,i}\ \text{for all other}\ i, \tag{B1}$$

for the parameters of $S^{(j-1)}_{\alpha,\kappa'}$, where

$$N^j_\kappa = 1 + \sum_i'' n^j_{\kappa,i}, \tag{B2}$$

and

$$\beta^{\kappa''} = \beta^\kappa,\quad m^{j+1}_{\kappa'',i} = m^j_{\kappa,i},\quad l^{j+1}_{\kappa''} = l^j_\kappa,\quad n^{j+1}_{\kappa'',N(j+1)} = 0,$$
$$n^{j+1}_{\kappa'',N(j-1)} = I\ \text{and}\ n^{j+1}_{\kappa'',i} = n^j_{\kappa,i}\ \text{for all other}\ i, \tag{B3}$$

for the parameters of $S^{(j+1)}_{\alpha,\kappa''}(a)$. With these parameters and the definitions of the phase and power functions given by Eqs. (3.14a) and (3.14b), one readily finds that

$$\pi^{j-1}_{\kappa'}(a) = \pi^{j+1}_{\kappa''}(a) = \pi^j_\kappa(a),$$
$$\phi^{j-1}_{\kappa'}(a) = 2\pi n^j_{\kappa,N(j-2)},\quad \phi^{j+1}_{\kappa''}(a) = 2\pi(I+N^j_\kappa). \tag{B4}$$

Hence, the two neighboring series have the same power and phases (mod 2π) as the series $S^{(j)}_{\alpha,\kappa}(a)$, so they can potentially be the ones that can be combined with it to produce a single convergent series. However, the constrains that the parameters $n^{j-1}_{\kappa',N(j)}$ and $n^{j+1}_{\kappa'',N(j-1)}$ must be nonnegative integers, cf. Sec. 3, implies that the parametrization of $S^{(j-1)}_{\alpha,\kappa'}(a)$ is only valid if $I \leqslant -N^j_\kappa$, and that the parametrization of $S^{(j+1)}_{\alpha,\kappa''}(a)$ will only be allowed if $I \geqslant 0$.

Our effort now divides into three separate tasks. First we must show that when $I \leqslant -N^j_\kappa$ the series $S^{(j)}_{\alpha,\kappa}(a)$ and $S^{(j-1)}_{\alpha,\kappa'}(a)$ combine to form a convergent series, then we must show that if $I \geqslant 0$ the series $S^{(j)}_{\alpha,\kappa}(a)$ and $S^{(j+1)}_{\alpha,\kappa''}(a)$ can be so combined, and, finally, we must explain what happens when $1 - N^j_\kappa \leqslant I \leqslant -1$.

The key to carrying out all three tasks is provided by a closer examination of the amplitude function $C^j_{\alpha,\kappa}(a)$, that appears in Eq. (3.11). The detailed derivation of the asymp-

totic form of $\mathscr{R}^j_\kappa(a)$ reveals that

$$C^j_{\alpha,\kappa}(a) = \frac{F_a(\beta^\kappa)\zeta_j \Gamma(x+N^j_\kappa)}{[A_{N(j+1)}]^{N^j_\kappa}\Gamma(x+1)} \times \prod_{i=0}^{n+1}\left(\frac{(A_i)^{n^j_{\kappa,i}}}{(n^j_{\kappa,i})!}\right)\left(\frac{A_{N(j-1)}}{A_{N(j)}}\right)^{x+N^j_\kappa}, \tag{B5}$$

where $x = \zeta_j X^j_\kappa/2\pi$, $A_0 = 1$, and X^j_κ, ζ_j and A_i with $i = 1, n-1$, are defined by Eqs. (3.14c), (2.29), and (2.24), respectively. The amplitude $F_\alpha(\beta^\kappa)$ is the same for all families, i.e., it is independentof j, and it depends upon the wave functions χ_1 and χ_2, for example, if $\beta^\kappa = a/R$ then $F_\alpha(\beta^\kappa) = 2R^3(-1)^l\chi_1^*(a)\chi_2(a)/\pi$. (The remaining possibilities will not be listed since $F_\alpha(\beta^\kappa)$ will play no role in our arguments.)

If $\phi^j_\kappa(a) = 2\pi I$, then $x = I$, and one finds that

$$C^j_{\alpha,\kappa}(a) \propto \frac{(I+N^j_\kappa-1)!}{I!} = (I+1)(I+2)\cdots(I+N^j_\kappa-1). \tag{B6}$$

Clearly $C^j_{\alpha,\kappa}(a)$ will vanish if I is an integer satisfying $1-N^j_\kappa \leqslant I \leqslant -1$. This result completes one of our tasks. Although there is no compensating series for $S^{(j)}_{\alpha,\kappa}(a)$, when I is an integer in the range $1-N^j_\kappa \leqslant I \leqslant -1$, none is needed for then $S^{(j)}_{\alpha,\kappa}(a) = 0$.

Next we form the ratios of the two coefficients of interest, obtaining

$$\frac{C^i_{\alpha,\nu}(a)}{C^j_{\alpha,\kappa}(a)} = -\frac{\zeta_i}{\zeta_j}, \tag{B7}$$

where $i = j-1$ and $\nu = \kappa'$ if $I \leqslant -N^j_\kappa$, and $i = j+1$ and $\nu = \kappa''$ if $I \geqslant 0$.

Before we combine the two series $S^{(j)}_{\alpha,\kappa}(a)$ and $S^{(i)}_{\alpha,\nu}(a)$ to form a new series $\bar{S}^{(j)}_{\alpha,\kappa}(a)$ and show that the latter series converges, we shall consider the series $T^{m,j}_{\alpha,\kappa}(k,a)$ defined by

$$T^{m,j}_{\alpha,\kappa}(k,a) \equiv \sum_{M>0} \mathscr{B}^{m,j}_{\alpha,\kappa,M}(k,a), \tag{B8}$$

which was discussed in Sec. 4, where it was imply referred to as the $\mathscr{B}$ series. From Eq. (4.10b), one sees that the phase and power functions, $\phi^j_\kappa(a)$ and $\pi^j_\kappa(a)$, respectively, of the series $T^{m,j}_{\alpha,\kappa}(k,a)$ are identical to those of $S^{(j)}_{\alpha,\kappa}(a)$, therefore, if we once again assume that $\phi^j_\kappa(a) = 2\pi I$, it follows from Eqs. (4.10c), (B5), and (B7) that $D^{m,j}_{\alpha,\kappa}(a)$, which is the amplitude function for $\mathscr{B}^{m,j}_{\alpha,\kappa,M}$, vanishes for $1-N^j_\kappa \leqslant I \leqslant -1$, and that

$$\frac{D^{m,i}_{\alpha,\nu}(a)}{D^{m,j}_{\alpha,\kappa}(a)} = -\frac{\zeta_i}{\zeta_j}, \tag{B9}$$

where again $i = j-1$ and $\nu = \kappa'$ if $I \leqslant -N^j_\kappa$, and $i = j+1$ and $\nu = \kappa''$ if $I \geqslant 0$. As was done in the case of the $S^{(j)}_{\alpha,\kappa}(a)$ series, the terms of the $T^{m,j}_{\alpha,\kappa}(k,a)$ and $T^{m,i}_{\alpha,\nu}(k,a)$ series shall be combined to form a convergent series $\bar{T}^{m,j}_{\alpha,\kappa}(k,a)$.

To construct the convergent sum $\bar{S}^{(j)}_{\alpha,\kappa}(a)$[or $\bar{T}^{m,j}_{\alpha,\kappa}(k,a)$], one combines the terms of $S^{(j)}_{\alpha,\kappa}(k,a)$ and $S^{(i)}_{\alpha,\nu}(a)$[or $T^{m,j}_{\alpha,\kappa}(k,a)$ and $T^{m,i}_{\alpha,\nu}(k,a)$] into a single series with the terms ordered according to $|k_n|$, or in the case of degeneracies in $|k_n|$, they are ordered according to $|\mathrm{Im}k_n|$, where k_n is the pole of $I_{12}(k,a)$ from which the term arose. According to the Cauchy condition[27] a series Σa_n converges if the partial sum $\Sigma_{n=M,N} a_n$ tends to zero for every $N > M$ as M tends to infinity. Suppose we form such a partial sum from the series $\bar{S}^{(j)}_{\alpha,\kappa}(a)$ [or $\bar{T}^{m,j}_{\alpha,\kappa}(k,a)$]. Since the partial sum is a finite sum, it can be written as the sum of two numbers with one being the sum over all terms in the partial sum that come from $S^{(j)}_{\alpha,\kappa}(a)$ [or $T^{m,j}_{\alpha,\kappa}(k,a)$] and the other being the sum over all terms from $S^{(i)}_{\alpha,\nu}(a)$ [or $T^{m,i}_{\alpha,\nu}(k,a)$]. Each of these numbers can be evaluated by means of the Euler–Maclaurin summation formula.[25] The major contribution to each number will come from the integral term in the Euler–Maclaurin formula. But by virtue of Eq. (B7) [or (B9)] the integrand of the integral associated with the (j,κ) sum equals minus one times the integrand associated with the (i,ν) sum, and the limits of integration are nearly equal. Thus, the sum of the two integrals nearly vanishes and one finds that the magnitude of the partial sum has the same order of magnitude as that of its first term. Hence, the magnitude of the partial sum for the series $\bar{S}^{(j)}_{\alpha,\kappa}(a)$ is $O(M^{-\pi^j_\kappa})$, and the magnitude of the series $\bar{T}^{m,j}_{\alpha,\kappa}(k,a)$ is $O(M^{-\pi^j_\kappa+m}|1-k^{-1}\zeta_j M|^{-1})$. Therefore, if $a/R < \beta^c_j$ so that $\pi^j_\kappa > 0$, the partial sum for $\bar{S}^{(j)}_{\alpha,\kappa}(a)$ will tend to zero as its lower index tends to infinity. This proves the convergence of $\bar{S}^{(j)}_{\alpha,\kappa}(a)$. From Eq. (4.11) we see that $|1-k^{-1}\zeta_j M|^{-1}$ is uniformly bounded in $\mathscr{E}$ for every value of j. Therefore, if $a/R < \beta^c(m)$, which implies that $m-\pi^j_\kappa < 0$, the magnitude of the partial sum of the series $\bar{T}^{m,j}_{\alpha,\kappa}(k,a)$ will tend to zero uniformly in $\mathscr{E}$ as the lower index of the partial sum tends to infinity. This proves that $\bar{T}^{m,j}_{\alpha,\kappa}(k,a)$ is uniformly convergent in $\mathscr{E}$.

APPENDIX C

Theorem: Given an arbitrary convergent series $A = \Sigma^\infty_{n=1} a_n$, then the series $B_p(z) = \Sigma^n_\infty{}^{=1} a_n(1+nz)^{-p}$, with $p = 0,1,2,\cdots$, converges uniformly in the region $T = \{z = re^{i\theta} | r \geqslant 0, \ -\theta_0 \leqslant \theta \leqslant \theta_0\}$ with $\pi/2 \leqslant \theta_0 < \pi$.

Proof: The theorem is obviously true for $p = 0$. Let us assume that it also holds for some integer $p > 0$, and define a remainder series R_N by

$$R_N \equiv \sum_{n=N}^{\infty} a_n(1+nz)^{-p} \tag{C1}$$

Then

$$R_n - R_{n+1} = a_n(1+nz)^{-p}, \tag{C2}$$

and

$$\begin{aligned}\sum_{n=N}^{M}\frac{a_n}{(1+nz)^{p+1}} &= \sum_{n=N}^{M}\frac{R_n-R_{n+1}}{1+nz} \\ &= \sum_{n=N-1}^{M-1}\frac{R_{n+1}}{1+(n+1)z} - \sum_{n=N}^{M}\frac{R_{n+1}}{1+nz} \\ &= \frac{R_N}{1+Nz} - \frac{R_{M+1}}{1+Mz} + \sum_{n=N}^{M-1}\left(\frac{1}{1+(n+1)z} - \frac{1}{1+nz}\right)R_{n+1} \\ &= \frac{R_N}{1+Nz} - \frac{R_{M+1}}{1+Mz} + \sum_{n=N}^{M-1}\left(\frac{1}{|1+(n+1)z|^2} - \frac{1}{|1+nz|^2}\right)R_{n+1}\end{aligned}$$

$$+z^*\sum_{n=N}^{M-1}\left(\frac{n+1}{|1+(n+1)z|^2}-\frac{n}{|1+nz|^2}\right)R_{n+1},$$

so that

$$\left|\sum_{n=N}^{M}\frac{a_n}{(1+nz)^{p+1}}\right|\leqslant\frac{|R_N|}{|1+Nz|}+\frac{|R_{M+1}|}{|1+Mz|}+\sum_{n=N}^{M-1}\left|\frac{1}{|1+(n+1)z|^2}-\frac{1}{|1+nz|^2}\right|\times|R_{n+1}|+|z|\sum_{n=N}^{M-1}\left|\frac{n+1}{|1+(n+1)z|^2}-\frac{n}{|1+nz|^2}\right||R_{n+1}|.$$

Since the series $B_p(z)$ converges by our induction hypothesis, then for any $\epsilon_0>0$ there exists an N_0 such that

$$|R_n|<\epsilon_0,\ \text{for every } n>N_0 \text{ and } z\in T. \tag{C3}$$

It then follows for every $M>N>N_0$, that

$$\left|\sum_{n=N}^{M}a_n(1+nz)^{-p-1}\right|\leqslant\epsilon_0\left[|1+Nz|^{-1}+|1+Mz|^{-1}+S_1+rS_2\right], \tag{C4}$$

where

$$S_1=\sum_{n=N}^{M-1}\left||1+(n+1)z|^2-|1+nz|^{-2}\right| \tag{C5}$$

and

$$S_2=\sum_{n=N}^{M-1}\left|(n+1)|1+(n+1)z|^{-2}-n|1+nz|^{-2}\right|. \tag{C6}$$

Let us examine the term S_1. First we note that the terms within the outermost absolute value signs on the right-hand side of (C5) can be written

$$\frac{1}{|1+(n+1)z|^2}-\frac{1}{|1+nz|^2}=\frac{-2r\cos\theta-(2n+1)r^2}{|1+(n+1)z|^2|1+nz|^2}. \tag{C7}$$

Thus, it is apparent that these terms are positive (or zero) up to a certain $n=n_1-1\leqslant-\frac{1}{2}-(\cos\theta)/r<n_1$, and negative thereafter. Therefore, if $M\leqslant n_1$ or $N\geqslant n_1$

$$S_1=\left|\sum_{n=N}^{M-1}(|1+(n+1)z|^{-2}-|1+nz|^{-2})\right|=\left||1+Mz|^{-2}-|1+Nz|^{-2}\right|,$$

and if $N<n_1<M$

$$S_1=\sum_{n=N}^{n-1}(|1+(n+1)z|^{-2}-|1+nz|^{-2})+\sum_{n=n_1}^{M-1}(|1+nz|^{-2}-|1+(n+1)z|^{-2})=2|1+n_1z|^{-2}-|1+Nz|^{-2}-|1+Mz|^{-2}.$$

In all cases one has

$$S_1\leqslant\frac{2}{|1+mz|^2},\quad\text{where } m=\begin{cases}M, & \text{for } n_1\geqslant M,\\ n_1, & \text{for } N<n_1<M,\\ N, & \text{for } n_1\leqslant N.\end{cases} \tag{C8}$$

Next we consider S_2. Since

$$\frac{n+1}{|1+(n+1)z|^2}-\frac{n}{|1+nz|^2}=\frac{1-n(n+1)r^2}{|1+(n+1)z|^2|1+nz|^2},$$

one sees that these terms are positive or zero for $n\leqslant n_2-1\leqslant-1/2+1(4+r^2)^{-1/2}/2R<n_2$, and negative thereafter. Thus,

$$S_2=\left|\frac{M}{|1+Mz|^2}-\frac{N}{|1+Nz|^2}\right|,\quad\text{for}N\geqslant n_2 \text{ or } M\leqslant n_2,$$

$$S_2=\frac{2n_2}{|1+n_2z|^2}-\frac{M}{|1+Mz|^2}-\frac{N}{|1+Nz|^2},$$

and thus for $N<n_2<M$,

$$S_2\leqslant\frac{2m}{|1+Mz|^2},\quad\text{where}m=\begin{cases}M, & \text{for } n_2\geqslant M,\\ n_2, & \text{for } N<n_2<M,\\ N, & \text{for } n_2\leqslant N.\end{cases} \tag{C9}$$

To complete the proof one needs upper bounds on $|1+nz|^{-2}$ and $nr|1+nz|^{-2}$ that hold for all $n\geqslant1$ and $z\in T$. For the first one has

$$|1+nz|^2=1+2nr\cos\theta+n^2r^2\geqslant1+2nr\cos\theta_0+n^2r^2=\sin^2\theta_0+(\cos\theta_0+nr)^2,$$

so that

$$|1+nz|^{-2}\leqslant(\sin\theta_0)^{-2},\quad\text{for all } n\geqslant1 \text{ and } z\in T. \tag{C10}$$

To obtain an upper bound on the second we observe that $(1-nr)^{-2}\geqslant0$ and thus

$$2nr(1+\cos\theta)\leqslant1+2nr\cos\theta+n^2r^2=|1+nz|^2,$$

so that

$$nr|1+nz|^{-2}\leqslant\tfrac{1}{2}(1+\cos\theta)^{-1}\leqslant\tfrac{1}{2}(1+\cos\theta_0)^{-1}. \tag{C11}$$

From (C8) and (C10) we see that

$$S_1\leqslant2(\sin\theta_0)^{-2}, \tag{C12}$$

while from (C9) and (C11) we find that

$$rS_2\leqslant(1+\cos\theta_0)^{-1}. \tag{C13}$$

Combining (C4), (C10), (C12) and (C13) yields

$$\left|\sum_{n=N}^{M-1}\frac{a_n}{(1+nz)^{p+1}}\right|\leqslant\epsilon_0\left(\frac{2}{\sin\theta_0}+\frac{2}{(\sin\theta_0)^2}+\frac{1}{1+\cos\theta_0}\right)\equiv\epsilon_0h(\theta_0). \tag{C14}$$

Since $h(\theta_0)$ is positive and finite over the range of values of θ_0 being considered, it can be seen that for any $\epsilon>0$, if one chooses $\epsilon_0=\epsilon/h(\theta_0)$ and selects N_0 according to the induction hypothesis (A3), then

$$\left|\sum_{n=N}^{M-1}\frac{a_n}{(1+nz)^{p+1}}\right|<\epsilon,$$

for every choice of $M>N>N_0$, and every $z\in T_0$. Hence, $B_{p+1}(z)$ satisfies the Cauchy condition for uniform convergence of the series in T, and the induction proof is completed.

[1]W.J. Romo, Nucl. Phys. A **237**, 275 (1975).
[2]W.J. Romo, Nucl. Phys. A **302**, 61 (1978).
[3]W.J. Romo, J. Math. Phys. **20**, 1210 (1979).
[4]G. Gamow, Zs. f. Phys. **51**, 204 (1928).
[5]A.J.F. Siegert, Phys. Rev. **56**, 750 (1939).
[6]J. Humblet and L. Rosenfeld, Nucl. Phys. **26**, 529 (1961).
[7]Ya. B. Zel'dovich, JETP (Sov. Phys.) **12**, 542 (1961).
[8]J.S. Bell, Nuovo Cimento. **24**, 452 (1962).
[9]N. Hokkyo, Prog. Theor. Phys. **33**, 1116 (1965).
[10]T. Berggren, Nucl. Phys. A **109**, 265 (1968); Phys. Lett. B **37**, 240 (1971); Phys. Lett. **B38**, 61 (1972).
[11]W.J. Romo, Ph.D. Thesis, University of Wisconsin (1967); Nucl. Phys. A **116**, 618 (1968); Nucl. Phys. A **191**, 65 (1972); Nucl. Phys. A **209**, 396 (1973).
[12]B. Gyarmati and T. Vertse, Nucl. Phys. A **106**, 523 (1971).
[13]R. Moore and E. Gerjuoy, Phys. Rev. A **7**, 1288 (1973).
[14]G. García-Calderón and R. Peierls, Nucl. Phys. A **265**, 443 (1976).
[15]K.W. McVoy and W.J. Romo, Nucl. Phys. A **126**, 161 (1969); Ann. Phys. (N.Y.) **57**, 496 (1970); Ann. Phys. (N.Y.) **62**, 544 (1971).
[16]C. Mahaux and H.A. Weidenmuller, *Shell-Model Approach to Nuclear Reactions* (North-Holland, Amsterdam, 1969), pp. 83 and 84.
[17]J. Bang and J. Zimanyi, Nucl. Phys. A **139**, 534 (1969).
[18]T. Berggren, Nucl. Phys. A **169**, 353 (1971).
[19]T. Berggren, Phys. Lett. B **44**, 23 (1973).
[20]M. Berrondo and G. García-Calderón, Lett. Nuovo Cimento **20**, 34 (1977).
[21]R. Newton, J. Math. Phys. **1**, 319 (1960); *Scattering Theory of Waves and Particles* (McGraw–Hill, New York, 1966).
[22]W. Kaplan, *Advanced Calculus* (Addison–Wesley, Reading, Massachusetts, 1957), p. 525.
[23]When $|k_{m+1}| = |k_m|$ the "circular path" C_m must be slightly deformed so that the poles k_m and k_{-m} are enclosed and the poles k_{m+1} and k_{-m-1} are excluded from the interior of C_m.
[24]A. Erdélyi, *Asymptotic Expansions* (Dover, New York, 1956), pp. 78–85.
[25]*Handbook of Mathematical Functions*, edited by M. Abromowitz and A. Stegun (National Bureau of Standards, Washington, D.C., 1965).
[26]In this section we shall take the upper limit of Σ_n to be $R_{n+1} = R - \epsilon$ with ϵ a small fixed positive number, rather than $R - |k|^{-1}$ as in Sec. 2, since we are only interested in values of r for which $r \leqslant a < R$. As a result of the new choice of R_{n+1}, it follows that $X_n = L$ in Eq. (3.4), and Eq. (3.6) holds when $j = n$.
[27]T.M. Apostol, *Mathematical Analysis* (Addison–Wesley, Reading, Massachusetts, 1957), p. 356.
[28]K. Rektorys, *Survey of Applicable Mathematics* (Iliffe, London, 1969), p. 382.
[29]Ref. 28, p. 388.
[30]R. Newton, *Scattering Theory of Waves and Particles* (McGraw–Hill, New York, 1966), p. 169.
[31]Ref. 16, pp. 138–76.
[32]P.L. Kapur and R.E. Peierls, Proc. R. Soc. A **166**, 277 (1938).
[33]G. García-Calderón, Nucl. Phys. A **261**, 130 (1976).
[34]J. Bang, F.A. Gareev, M.H. Gizzatkulov, and S.A. Goncharov, Nucl. Phys. A **309**, 381 (1978).
[35]Ref. 24, p. 7.

Inverse scattering problems in higher dimensions: Yang–Mills fields and the supersymmetric sine-Gordon equation

H. C. Morris

Department of Mathematics, [a] *Oregon State University, Corvallis, Oregon 97331*
and School of Mathematics, [b] *Trinity College, Dublin, Republic of Ireland*
(Received 11 October 1978)

It is shown that the notion of a prolongation structure can be extended to higher dimensions and used to determine inverse scattering problems. The relationship to generalized Lax representations is also considered. The method is illustrated on the self-dual Yang–Mills equations. A generalization to include Grassman algebra valued variables is shown to provide a scattering problem for the supersymmetric sine-Gordon equation.

1. INTRODUCTION

In previous work we have attempted to extend the prolongation structure method[1–5] in two distinct directions. We have tried to generalize and develop the technique in higher dimensions[6–9] and also to extend it by the inclusion of Grassman algebra valued variables.[10] In this paper we further explore the development of this approach for the determination of inverse scattering problems in three and four dimensions and illustrate its utility by the construction of an inverse scattering problem for the self dual Yang–Mills equations.[11,12] We then go on to combine the two approaches in order to determine an inverse scattering problem for the supersymmetric sine-Gordon equation.[13–15]

In Sec. 2 we examine the equations that may be related to a particular type of linear prolongation form in three dimensions, derive in a slightly different way some of the results of Ref. 6, and show the connection to the Lax representation approach of Zakharov and Shabat.[16]

In Sec. 3 we develop the analogous equations for four dimensions and emphasize in particular the relationship to a generalized Lax representation.[17] We wish to stress the point that the prolongation structure method can allow the direct determination of such a representation from the equations if it exists. The process is not trivially implementable however, and considerable experience and insight is required to implement the method in its current state of development.

To illustrate the direct approach we tackle in Sec. 4 equations of the self-dual Yang–Mills equations in their simplest form due to Yang[11] and Zakharov and Belavin.[12] It is not necessary to use this particular complex form, but it does reduce the algebra to a more compact and manageable form which is easier to follow.

In Sec. 5 we link this work to that of Ref. 10 by considering a theory involving superfields. This example is the first of its kind in that it involves not only four dimensions, two of which are fermionic, but also leads to a coordinate dependent prolongation form related to a generalized prolongation structure which is an incomplete graded Lie algebra.

We do not attempt to determine solutions using the scattering problems defined. The purpose of this paper is to show the all-embracing nature of the prolongation structure idea, as applied to partial differential equations, to determine inverse scattering problems. However, a detailed analysis of the new results of this paper will be presented elsewhere in the context of specific equations.

2. LINEAR PROLONGATION FORMS IN THREE DIMENSIONS

Consider linear prolongation forms Ω of the following structure,

$$\Omega = \omega_a dx^a \wedge d\xi + (F^c \sigma_c)\xi, \tag{2.1}$$

where $\sigma^c = \frac{1}{2}\epsilon^{abc} dx^a \wedge dx^b$ and the ω_a are a set of constant matrices which commute with one another,

$$[\omega_a, \omega_b] = 0, \quad \forall a,b. \tag{2.2}$$

In general some coordinate dependence is possible in the ω_a but we will not consider that possibility in this paper in any generality. We will present an example of coordinate dependent ω_a in our construction of an inverse scattering problem for the supersymmetric sine-Gordon equation in Sec. 5.

Taking the exterior derivative of (2.1) gives

$$d\Omega = (dF_c \wedge \sigma^c)\xi + (F_c \sigma^c) \wedge d\xi \tag{2.3}$$

If η is a 1-form of the type

$$\eta = \eta_a dx^a, \tag{2.4}$$

then

$$d\Omega - \eta \wedge \Omega = (dF_c \wedge \sigma_c - \eta_c F_c \sigma) + (F_c \sigma_c - \eta_a \omega_b \epsilon^{abc} \sigma_c) \wedge d\xi, \tag{2.5}$$

where $\sigma = dx^1 \wedge dx^2 \wedge dx^3$.

We see from (2.5) that Ω is a prolongation form for the ideal spanned by the elements α_{ij} of the matrix valued 3-form α defined by

$$\alpha = dF_c \wedge \sigma_c - \eta_c F_c \sigma \tag{2.6}$$

where η_a and F_c are related by

$$F_c = \eta_a \omega_b \epsilon^{abc}. \tag{2.7}$$

 0022-2488/80/020327-07$1.00

As $[\omega_a,\omega_b]=0$ it follows that we have a constraint on the F_a given by

$$F_c\omega_c=0. \tag{2.8}$$

If we now suppose that ω_3 is invertible, then by redefining ξ it may be taken to be unity. To determine the content of Eq. (2.6), which we would like to express in terms of the F_a, we must simplify the quantity η_cF_c which occurs. We note that

$$\eta_cF_c=\eta_1F_1+\eta_2F_2+\eta_3(-F_1\omega_1-F_2\omega_2) \tag{2.9}$$

$$=(\eta_1-\eta_3\omega_1)F_1+(\eta_2-\eta_3\omega_2)F_2$$

$$-\eta_3([F_1,\omega_1]+[F_2,\omega_2]), \tag{2.10}$$

and so by virtue of Eq. (2.7) we have

$$\eta_cF_c=[F_1,F_2]-\eta_3([F_1,\omega_1]+[F_2,\omega_2]), \tag{2.11}$$

Therefore, if we choose ω_1 and ω_2 so that

$$[F_1,\omega_1]+[F_2,\omega_2]=0, \tag{2.12}$$

our basic 3-form takes the form

$$\alpha=dF_c\wedge\sigma_c-[F_1,F_2]\sigma, \tag{2.13}$$

which is closed and consequently equivalent to a set of partial differential equations. We should also recall that we have the condition imposed upon ω_1,ω_2 that,

$$[\omega_1,\omega_2]=0. \tag{2.14}$$

The existence of the nontrivial solution for ω_1 and ω_2 to (2.12) and (2.14) means that the three-dimensional equations (2.13) have an inverse scattering problem obtained by restricting Ω to a solution manifold of that system. In that case we obtain the equations

$$\epsilon_{abc}\omega_a\xi_{,b}=-F_c\xi, \tag{2.15}$$

which for $c=1,2$ are

$$\xi_{,2}-\omega_2\xi_{,3}=F_1\xi \tag{2.16}$$

and

$$\xi_{,1}-\omega_1\xi_{,3}=-F_2\xi, \tag{2.17}$$

corresponding to the equation

$$F_{c,c}=[F_1,F_2], \tag{2.18}$$

which may be reduced to

$$\partial_1F_1+\partial_2F_2-\partial_3(\omega_1F_1+\omega_2F_2)=[F_1,F_2] \tag{2.19}$$

by use of (2.8).

Clearly we have shown that these equations have the generalized Lax form

$$[L_1,L_2]=0, \tag{2.20}$$

where

$$L_1=\partial_2-\omega_2\partial_3-F_1, \tag{2.21}$$

$$L_2=\partial_1-\omega_1\partial_3+F_2, \tag{2.22}$$

and ω_1 and ω_2 are any pair of constant commuting matrices having the additional property that

$$[\omega_1,F_1]+[\omega_2,F_2]=0. \tag{2.23}$$

The principal interest lies in the case where the ω_a are either matrices and/or parameter dependent. For example, if we take

$$\omega_1=\lambda \text{ and } \omega_2=\lambda^2 \tag{2.24}$$

and

$$F_1=A_2-\lambda^2A_3, \tag{2.25}$$

$$F_2=A_3\lambda-A_1, \tag{2.26}$$

$$F_3=A_1\lambda^2-A_2\lambda, \tag{2.27}$$

which conforms to the constraint (2.8) which in this case is expressed in the form

$$F_3+\lambda F_1+\lambda^2F_2=0. \tag{2.28}$$

Equations (2.19) then become

$$F_{ab}=A_{a,b}-A_{b,a}+[A_a,A_b]=0 \quad (a,b=1,2,3), \tag{2.29}$$

the vanishing of the Yang–Mills fields in three dimensions. It should be recalled that this gives only special solutions because only in the case of self dual fields in four dimensions is $F_{ab}=0$ equivalent to the full field equations. However, we have obtained an interesting inverse scattering problem for the SU(2) Yang–Mills equation $F_{ab}=0$ in threee dimensions which is given by

$$\xi_{,2}-\lambda^2\xi_{,3}=(A_2-\lambda^2A_3)\xi, \tag{2.30}$$

$$\xi_{,1}-\lambda\xi_{,3}=(A_1-\lambda A_3)\xi. \tag{2.31}$$

3. LINEAR PROLONGATION FORMS IN FOUR DIMENSIONS

Consider a linear prolongation form Ω of the type

$$\Omega=\tfrac{1}{2}(\omega_{ab}dx^a\wedge dx^b)\wedge d\xi+F_c\sigma^c\xi \tag{3.1}$$

where $\sigma^d=(1/6)\epsilon^{abcd}dx^a\wedge dx^b\wedge dx^c$ and ω_{ab} are constants. The general case of matrix valued ω_{ab} is more difficult and will only be considered in a special case. Taking the exterior derivative of (3.1) we obtain,

$$d\Omega=dF_c\wedge\sigma^c\xi+F_c\sigma^c\wedge d\xi. \tag{3.2}$$

If η is a 1-form of the type

$$\eta=\eta_adx^a, \tag{3.3}$$

then

$$d\Omega-\eta\wedge\Omega=(-F_c\sigma^c-\tfrac{1}{2}\epsilon_{abcd}\eta_c\omega_{ab}\sigma_d)\wedge d\xi$$

$$+(dF_c\wedge\sigma_c-\eta_cF_c\sigma)\xi, \tag{3.4}$$

and so we see that Ω is a prolongation form for the ideal spanned by the elements α_{ij} of the matrix valued 4-form α defined by

$$\alpha=dF_c\wedge\sigma_c-\eta_cF_c\sigma, \quad \text{where } \sigma=dx^1\wedge dx^2\wedge dx^3\wedge dx^4, \tag{3.5}$$

and η_c and F_a are connected by the relationship

$$F_d=-\tfrac{1}{2}\epsilon_{abcd}\eta_c\omega_{ab}=\tilde{\omega}_{dc}\eta_c, \tag{3.6}$$

where

$$\tilde{\omega}_{dc}=\tfrac{1}{2}\epsilon_{abdc}\omega_{ab}=\tfrac{1}{2}\epsilon_{dcab}\omega_{ab}. \tag{3.7}$$

The most interesting cases arise when the matrix $\tilde{\omega}_{dc}$ is singu-

lar and we restrict our attention in this paper to the case when $\tilde{\omega}_{dc}$ is of rank 2. The explicit form of $\tilde{\omega}$ is given by

$$\tilde{\omega} = \begin{pmatrix} 0 & \omega_{34} & \omega_{42} & \omega_{23} \\ -\omega_{34} & 0 & \omega_{14} & \omega_{31} \\ -\omega_{42} & -\omega_{14} & 0 & \omega_{12} \\ -\omega_{23} & -\omega_{31} & -\omega_{12} & 0 \end{pmatrix}. \tag{3.8}$$

If we take the first two rows as independent, then there exist r_i such that

$$r_1(0,\omega_{34},\omega_{42},\omega_{23}) + r_2(\omega_{34},0,\omega_{14},\omega_{31}) = (-\omega_{42}, -\omega_{14},0,\omega_{12}) \tag{3.9}$$

and

$$r_3(0,\omega_{34},\omega_{42},\omega_{23}) + r_4(-\omega_{34},0,\omega_{14},\omega_{31}) = (-\omega_{23}, -\omega_{31}, -\omega_{12},0); \tag{3.10}$$

we easily find that

$$r_2 = \frac{\omega_{42}}{\omega_{34}}, \quad r_1 = -\frac{\omega_{14}}{\omega_{34}}, \tag{3.11}$$

$$r_4 = \frac{\omega_{23}}{\omega_{34}}, \quad r_3 = -\frac{\omega_{31}}{\omega_{34}}, \tag{3.12}$$

and

$$\omega_{31}\omega_{42} - \omega_{14}\omega_{23} = \omega_{12}\omega_{34}. \tag{3.13}$$

From (3.6) we then obtain

$$F_3 = r_1F_1 + r_2F_2, \tag{3.14}$$

$$F_4 = r_3F_1 + r_4F_2. \tag{3.15}$$

As in Sec. 2 we need to analyze the term $(\eta_c F_c)$ in α and we see that

$$\begin{aligned} \eta_c F_c &= \eta_1F_1 + \eta_2F_2 + \eta_3(r_1F_1 + r_2F_2) + \eta_4(r_3F_1 + r_4F_2) \\ &= (\eta_1 + \eta_3r_1 + \eta_4r_3)F_1 + (\eta_2 + r_2\eta_3 + r_4\eta_4)F_2 \\ &= \frac{1}{\omega_{34}}\{[+\omega_{34}\eta_1 - \omega_{14}\eta_3 - \omega_{31}\eta_4]F_1 \\ &\quad + [\omega_{34}\eta_2 + \omega_{42}\eta_3 + \omega_{23}\eta_4]F_2\}. \end{aligned} \tag{3.16}$$

From (3.6) we have

$$F_1 = \omega_{34}\eta_2 + \omega_{42}\eta_3 + \omega_{23}\eta_4, \tag{3.17}$$

$$F_2 = -\omega_{34}\eta_1 + \omega_{14}\eta_3 + \omega_{31}\eta_4. \tag{3.18}$$

Therefore, we have

$$\eta_c F_c = \frac{1}{\omega_{34}}[F_1,F_2]. \tag{3.19}$$

Without loss of generality we can choose $\omega_{34} = 1$, and our equation for α becomes

$$\alpha = dF_c \wedge \sigma_c - [F_1,F_2]\sigma, \tag{3.20}$$

and the elements α_{ij} generate a closed ideal. The corresponding field equations are therefore

$$F_{c,c} = [F_1,F_2] \tag{3.21}$$

or

$$F_{1,1} + F_{2,2} - (r_1F_1 + r_2F_2)_{,3} - (r_3F_1 + r_4F_2)_{,4} = [F_1,F_2],$$

$$(F_{1,1} - r_1F_{1,3} - r_3F_{1,4}) + (F_{2,2} - r_2F_{2,3} - r_4F_{2,4}) = [F_1,F_2]. \tag{3.22}$$

Again the equations have a generalized Lax representation

$$[L_1,L_2] = 0, \tag{3.23}$$

with

$$L_1 = (\partial_2 - r_2\partial_3 - r_4\partial_4 - F_1) \tag{3.24}$$

and

$$L_2 = (\partial_1 - r_1\partial_3 - r_3\partial_4 + F_2) \tag{3.25}$$

and so if we introduce the definitions of the r_i we obtain

$$L_1 = (\partial_2 - \omega_{42}\partial_3 - \omega_{23}\partial_4 - F_1), \tag{3.26}$$

$$L_2 = (\partial_1 + \omega_{14}\partial_3 + \omega_{31}\partial_4 + F_2), \tag{3.27}$$

where condition (3.13) simply defines ω_{12} to be given by

$$\omega_{12} = \omega_{31}\omega_{42} - \omega_{14}\omega_{23}, \tag{3.28}$$

which places no constraint on the coefficients $\omega_{42},\omega_{23},\omega_{14},\omega_{31}$ which appear in L_1 and L_2.

We can easily allow the ω_{ab} in this special case to be matrix valued if we add in the additional constraints that

$$[\omega_{ab},\omega_{cd}] = 0, \tag{3.29}$$

and taking $\omega_{34} = I$ the identity matrix the analogue of (2.23) is

$$[\omega_{14},F_1] = [\omega_{42},F_2], \tag{3.30}$$

and

$$[\omega_{31},F_1] = [\omega_{23},F_2], \tag{3.31}$$

together with the definition of ω_{12} as

$$\omega_{12} = (\omega_{42}\omega_{31} - \omega_{14}\omega_{23}). \tag{3.32}$$

These equations greatly generalize some of those presented in Ref. 9.

We could go on to determine that the analog of the free Yang–Mills fields that arise in our study of three-dimensional examples in Sec. 2 are the self-dual Yang–Mills equations. Rather than adopt the same approach we attempt in the following section to show how one may, by using the prolongation method ideas, derive an inverse scattering problem for the self-dual Yang–Mills equations directly from the equations themselves.

4. THE SELF-DUAL YANG–MILLS EQUATIONS

The self-dual Yang–Mills fields expressed in the complex coordinates introduced by Yang[11] and Zakharov and Belavin[12] take the form

$$\partial_{z_1}B_2 - \partial_{z_2}B_1 + [B_1,B_2] = 0, \tag{4.1}$$

$$\partial_{\bar{z}_1}B_2^{+} - \partial_{\bar{z}_2}B_1^{+} + [B_2^{+},B_1^{+}] = 0, \tag{4.2}$$

$$\partial_{z_1}B_1^{+} + \partial_{z_2}B_2^{+} + \partial_{\bar{z}_1}B_1 + \partial_{\bar{z}_2}B_2 + [B_1,B_1^{+}] + [B_2,B_2^{+}] = 0, \tag{4.3}$$

and may be expressed by the closed ideal of 4-forms spanned by the forms α_i defined by

$$\alpha_1 = (dB_1 \wedge dz_1 + dB_2 \wedge dz_2 + [B_1,B_2]dz_1 \wedge dz_2)d\bar{z}_1 \wedge d\bar{z}_2, \tag{4.4}$$

$$\alpha_2 = (dB_{\bar{1}}^{+} \wedge d\bar{z}_1 + dB_{\bar{2}}^{+} \wedge d\bar{z}_2 + [B_{\bar{2}}^{+},B_{\bar{1}}^{+}]d\bar{z}_1 \wedge d\bar{z}_2)dz_1 \wedge dz_2, \tag{4.5}$$

$$\alpha_3 = (dB_{\bar{1}}^{+} \wedge dz_2 - dB_{\bar{2}}^{+} \wedge dz_1) \wedge d\bar{z}_1 \wedge d\bar{z}_2 + (dB_1 \wedge d\bar{z}_2 - dB_2 \wedge d\bar{z}_1) \wedge dz_1 \wedge dz_2 + ([B_1,B_{\bar{1}}^{+}] + [B_2,B_{\bar{2}}^{+}])\sigma, \tag{4.6}$$

where $\sigma = dz_1 \wedge dz_2 \wedge d\bar{z}_1 \wedge d\bar{z}_2$.

Consider a linear prolongation form Ω having the structure

$$\Omega = d\xi \wedge (\omega_{12}dz_1 \wedge dz_2 + \omega_{1\bar{1}}dz_1 \wedge d\bar{z}_1 + \omega_{1\bar{2}}dz_1 \wedge d\bar{z}_2 + \omega_{2\bar{1}}dz_2 \wedge d\bar{z}_1 + \omega_{2\bar{2}}dz_2 \wedge d\bar{z}_2 + \omega_{1\bar{2}}dz_1 \wedge d\bar{z}_2) + (F_1dz_2 \wedge d\bar{z}_1 \wedge d\bar{z}_2 + F_2dz_1 \wedge d\bar{z}_1 \wedge d\bar{z}_2 + F_{\bar{1}}dz_1 \wedge dz_2 \wedge d\bar{z}_2 + F_{\bar{2}}dz_1 \wedge dz_2 \wedge d\bar{z}_1)\xi, \tag{4.7}$$

where the ω_{ab} are scalars. The requirement that Ω prolong the ideal $\langle\alpha_1,\alpha_2,\alpha_3\rangle$ spanned by α_1, α_2 and α_3 is

$$d\Omega = f^i\alpha_i + (\eta_1dz_1 + \eta_2dz_2 + \eta_{\bar{1}}d\bar{z}_1 + \eta_{\bar{2}}d\bar{z}_2) \wedge \Omega, \tag{4.8}$$

and this shows us that we must have

$$F_1 = F_1(B_2,B_{\bar{1}}^{+}), \tag{4.9}$$

$$F_2 = F_2(B_1,B_{\bar{2}}^{+}), \tag{4.10}$$

$$F_{\bar{1}} = F_{\bar{1}}(B_1,B_{\bar{2}}^{+}), \tag{4.11}$$

$$F_{\bar{2}} = F_{\bar{2}}(B_2,B_{\bar{1}}^{+}), \tag{4.12}$$

$$F_{1'B_2} = F_{2'B_1}, \tag{4.13}$$

$$F_{\bar{1}'B_{\bar{2}}} = F_{\bar{2}'B_{\bar{1}}}, \tag{4.14}$$

$$F_{1'B_{\bar{1}}} = F_{\bar{1}'B_1} = -F_{\bar{2}'B_2} = -F_{2'B_{\bar{2}}}, \tag{4.15}$$

the relationships

$$F_1 = -\omega_{1\bar{2}}\eta_2 + \omega_{2\bar{2}}\eta_{\bar{1}} - \omega_{2\bar{1}}\eta_{\bar{2}}, \tag{4.16}$$

$$F_1 = -\omega_{1\bar{2}}\eta_1 + \omega_{1\bar{2}}\eta_{\bar{1}} - \omega_{1\bar{1}}\eta_{\bar{2}}, \tag{4.17}$$

$$F_{\bar{1}} = -\omega_{2\bar{2}}\eta_1 + \omega_{1\bar{2}}\eta_2 - \omega_{12}\eta_{\bar{2}}, \tag{4.18}$$

$$F_{\bar{2}} = -\omega_{21}\eta_1 + \omega_{1\bar{1}}\eta_2 - \omega_{12}\eta_{\bar{1}}, \tag{4.19}$$

and the central equation

$$[B_1,B_2]F_{1'B_{\bar{2}}} + [B_{\bar{2}}^{+},B_{\bar{1}}^{+}]F_{\bar{1}'B_{\bar{2}}} + F_{1'B_{\bar{1}}}\{[B_1,B_{\bar{1}}^{+}] + [B_2,B_{\bar{2}}^{+}]\} = (\eta_1F_1 - \eta_2F_2 + \eta_{\bar{1}}F_{\bar{1}} - \eta_{\bar{2}}F_{\bar{2}}). \tag{4.20}$$

From (4.9)–(4.15) we see that possible forms for F_1, F_2, $F_{\bar{1}}$ and $F_{\bar{2}}$ are

$$F_1 = (x_1B_2 + x_2B_{\bar{1}}^{+}), \tag{4.21}$$

$$F_2 = (x_1B_1 - x_2B_{\bar{2}}^{+}), \tag{4.22}$$

$$F_{\bar{1}} = (x_2B_1 + x_3B_{\bar{2}}^{+}), \tag{4.23}$$

$$F_{\bar{2}} = (-x_2B_2 + x_3B_{\bar{1}}^{+}), \tag{4.24}$$

where the x_i are scalars. In order to reduce the algebra of the general case we make the additional ansatz consistent with (4.21)–(4.24), that

$$F_{\bar{1}} = \lambda F_2, \tag{4.25}$$

$$F_{\bar{2}} = -\lambda F_1, \tag{4.26}$$

which is equivalent to

$$x_2 = +\lambda x_1 \quad \text{and} \quad x_3 = -\lambda x_2. \tag{4.27}$$

This in not necessary but greatly reduces the required algebra and leads to useful, if not most general, results.

The term $(\eta_1F_1 - \eta_2F_2 + \eta_{\bar{1}}F_{\bar{1}} - \eta_{\bar{2}}F_{\bar{2}})$ must be converted into commutation relations between the F_α if the normal prolongation ideas are to be used. Using our ansatz (4.25)–(4.26) this term becomes

$$(\eta_1 + \lambda\eta_{\bar{2}})F_1 + (\lambda\eta_{\bar{1}} - \eta_2)F_2,$$

and we need to express the combinations $(\eta_1 + \lambda\eta_{\bar{2}})$ and $(\lambda\eta_{\bar{1}} - \eta_2)$ in terms of the F_α if we are to obtain a Lie algebra-like structure. Equations (4.16)–(4.19) give

$$\eta_1(\omega_{2\bar{2}} - \lambda\omega_{1\bar{2}}) - \eta_2\omega_{1\bar{2}} + \eta_{\bar{2}}(\omega_{12} - \lambda\omega_{1\bar{1}}) = 0, \tag{4.28}$$

$$\eta_1\omega_{21} + \eta_2(\lambda\omega_{1\bar{2}} - \omega_{1\bar{1}}) + \eta_{\bar{1}}(\omega_{12} - \lambda\omega_{2\bar{2}}) = 0. \tag{4.29}$$

These are solved by the relations

$$\omega_{2\bar{2}} = \lambda\omega_{1\bar{2}}, \tag{4.30}$$

$$\omega_{1\bar{2}} = 0 = \omega_{21}, \tag{4.31}$$

$$\omega_{12} = \lambda\omega_{1\bar{1}} = \lambda\omega_{2\bar{2}}, \tag{4.32}$$

in which case (4.16) and (4.17) become

$$F_1 = -\omega_{1\bar{2}}(\eta_2 - \lambda\eta_{\bar{1}}), \tag{4.33}$$

$$F_2 = -\omega_{1\bar{2}}(\eta_1 + \lambda\eta_{\bar{2}}), \tag{4.34}$$

and so we are able to express the quantity $(\eta_1F_1 - \eta_2F_2 + \eta_{\bar{1}}F_{\bar{1}} - \eta_{\bar{2}}F_{\bar{2}})$ as

$$-(\omega_{1\bar{2}})^{-1}[F_2,F_1].$$

Clearly we can take $\omega_{1\bar{2}} = 1$ without loss of generality and so the essential equation (4.20) becomes

$$[B_1,B_2]F_{1'B_{\bar{2}}} + [B_{\bar{2}}^{+},B_{\bar{1}}^{+}]F_{1'B_{\bar{2}}} + F_{1'B_{\bar{1}}}\{[B_1,B_{\bar{1}}^{+}] + [B_2,B_{\bar{2}}^{+}]\} + [F_2,F_1] = 0. \tag{4.35}$$

From the forms (4.21)–(4.24) with condition (4.27) we get

$$F_{1,B_{\bar{2}}} = x_1, \quad F_{1,B_{\bar{1}}} = x_2 = +\lambda x_1, \quad F_{\bar{1},B_{\bar{2}}} = x_3 = -\lambda^2x_1$$

and (4.35) becomes

$$(x_1 + x_1^2)\{[B_1,B_2] + \lambda\{[B_1,B_{\bar{1}}^{+}] + [B_2,B_{\bar{2}}^{+}]\} + \lambda^2[B_{\bar{1}}^{+},B_{\bar{2}}^{+}]\} = 0, \tag{4.36}$$

which is satisfied for arbitrary λ if $x_1 + x_1^2 = 0$ and $x_1 = -1$. We have therefore obtained a prolongation form which looks as follows,

$$\Omega = d\xi \wedge (d\bar{z}_1 \wedge d\bar{z}_2 + \lambda(dz_1 \wedge d\bar{z}_1 + dz_2 \wedge d\bar{z}_2) + \lambda^2dz_1 \wedge dz_2) - dz_2 \wedge d\bar{z}_1 \wedge (d\bar{z}_2 - \lambda dz_1)(B_2 + \lambda B_{\bar{1}}^{+})\xi + dz_1 \wedge d\bar{z}_2 \wedge (d\bar{z}_1 + \lambda dz_2)(B_1 - \lambda B_{\bar{2}}^{+})\xi. \tag{4.37}$$

If we section this onto a solution manifold of the self-dual Yang–Mills equations we obtain the equations

$$\xi_{z_1} + \lambda\xi_{\bar{z}_2} = (B_1 - \lambda B_2^{+})\xi, \tag{4.38}$$

$$\xi_{z_2} + \lambda\xi_{\bar{z}_1} = (B_2 + \lambda B_1^{+})\xi, \tag{4.39}$$

If we denote by D_1 and D_2 the differential operators $D_1 = \partial_{z_1} + \lambda\partial_{\bar{z}_2}$ and $D_2 = \partial_{z_2} - \lambda\partial_{\bar{z}_1}$ we have

$$D_1\xi = (B_1 - \lambda B_2^{+})\xi, \tag{4.40}$$

$$D_2\xi = (B_2 + \lambda B_1^{+})\xi, \tag{4.41}$$

and the integrability conditions $D_1D_2 = D_2D_1$ give

$$D_2(B_1 - \lambda B_2^{+}) - D_1(B_2 + \lambda B_1^{+}) - [(B_1 - \lambda B_2^{+}),(B_2 + \lambda B_1^{+})] = 0, \tag{4.42}$$

which is equivalent to

$$\{B_{1,2} - B_{2,1} - [B_1,B_2]\} + \lambda\{-B^{+}_{2,2} - B_{1,\bar{1}} - B^{+}_{1,1} - B_{2,\bar{2}} - [B_1,B_1^{+}] - [B_2,B_2^{+}]\} + \lambda^2\{B^{+}_{2,\bar{1}} - B^{+}_{1,\bar{2}} + [B_2^{+},B_1^{+}]\} = 0, \tag{4.43}$$

and satisfied for all λ if B_α is a solution of the self dual Yang–Mills equation . Belavin and Zakharov have recently shown[12] how this inverse scattering problem may be used to construct instanton solutions for this set of equations. Drawing on the experience gained from this example we will now show how a ω_{ab} constructed from super fields may be used to determine an inverse scattering problem for the super-symmetric sine-Gordon equation.

5. THE SUPERSYMMETRIC SINE-GORDON EQUATION

A natural way of incorporating fermions into a model is via a supersymmetric extension. The model which results from the completely integrable sine-Gordon equation was considered by Di Vecchia and Ferrara[14] and Hruby[15] and shown to have the interesting property that instanton and soliton solutions remain along with other attractive properties.

The model is defined by the action in superspace given by

$$\int d^2x d^2\theta(-\tfrac{1}{2}i\Phi D_2D_1\Phi + m\cos\Phi), \tag{5.10}$$

where D_1 and D_2 are defined below and which yields the superfield equation of motion

$$D_2D_1\Phi = im\sin\Phi. \tag{5.2}$$

Φ is the superfield with components ϕ, ψ, and F defined by

$$\Phi(x,\theta) = \phi(x) + i\bar{\theta}\psi(x) + \tfrac{1}{2}i\bar{\theta}\theta F(x) \tag{5.3}$$

in terms of which the field equation (5.2) can be expressed as

$$F = -m\sin\phi, \quad \partial\psi = -m\psi\cos\phi, \tag{5.4}$$

$$\Box\phi = -m(F\cos\phi + \tfrac{1}{2}i\bar{\psi}\psi\sin\phi) = \tfrac{1}{2}(m^2\sin2\phi - im\bar{\psi}\psi\sin\phi). \tag{5.5}$$

We will now show how an inverse scattering problem can be constructed for such an equation. Let us take as defining forms

$$\alpha_1 = (d\theta_1\wedge dx\wedge dT + id\theta_1\wedge d\theta_2\wedge dT\theta_2)\wedge dP + im\sin\phi\sigma, \tag{5.6}$$

$$\alpha_2 = (d\theta_2\wedge dx\wedge dT + id\theta_1\wedge d\theta_2\wedge dx\theta_1)\wedge d\phi - \sigma P, \tag{5.7}$$

where $\sigma = d\theta_1\wedge d\theta_2\wedge dx\wedge dT$; where the fermionic fields are concerned we are using left derivatives, so for example,

$$dP = (d\theta_1P_{\theta_1} + d\theta_2P_{\theta_2} + dxP_x + dTP_T), \tag{5.8}$$

where P_{θ_i} is the left derivative of P defined uniquely by

$$P = \theta_1P_{\theta_1} + \text{terms independent of } \theta_1. \tag{5.9}$$

Sectioning these forms yields

$$\tilde{\alpha}_1 = \sigma\{(P_{\theta_2} - i\theta_2P_x) - im\sin\phi\} = 0, \quad \text{or} \quad D_2 = im\sin\phi, \tag{5.10}$$

where

$$D_2\psi = -\psi_{\theta_2} + i\theta_2\psi_{,x}, \tag{5.11}$$

$$\tilde{\alpha}_2 = \sigma\{(\phi_{\theta_1} + i\theta_1\phi_T) - P\} = 0, \text{ or } P = D_1\phi,$$

where

$$D_1\psi = \psi_{\theta_1} + i\theta_1\psi_{,T}. \tag{5.12}$$

Thus the closed ideal $\langle\alpha_1,\alpha_2\rangle$ spanned by α_1 and α_2 is completely equivalent to the supersymmetric sine-Gordon equation

$$D_2D_1\phi = im\sin\phi, \tag{5.13}$$

and we note that the operators D_1 and D_2 anticommute $\{D_1,D_2\} = 0$. Consider a prolongation term Ω having the structure

$$\Omega = (dx\wedge dT - d\theta_1\wedge dx\beta d\theta_2\wedge dT + \gamma d\theta_1\wedge d\theta_2)\wedge d\xi + [d\theta_1\wedge dx\wedge dT + d\theta_1\wedge d\theta_2\wedge dT\beta]F\xi + [d\theta_2\wedge dx\wedge dT + d\theta_1\wedge d\theta_2\wedge dx\alpha]G\xi, \tag{5.14}$$

where $F = x_1 + Px_2$, $G = \sin\phi x_3 + \cos\phi x_4$ and $\alpha = i\theta_1$, $\beta = i\theta_2$, and $\gamma = -\alpha\beta = +\theta_1\theta_2$. Clearly x_1 and x_2 may be functions of θ_1, and x_3 and x_4 functions of θ_2 without affecting the calculation. We then easily determine that

$$\begin{aligned} -d\Omega = {} & (d\theta_1\wedge dx\wedge dT + d\theta_1\wedge d\theta_2\wedge dT\beta)\\ & \wedge(\epsilon x_1 - P\epsilon x_2)d\xi + (d\theta_2\wedge dx\wedge dT\\ & + d\theta_1\wedge d\theta_2\wedge dx\alpha)\wedge(\sin\phi\epsilon x_3 + \cos\phi\epsilon x_4)d\xi\\ & + (d\theta_1\wedge dx\wedge dT + d\theta_1\wedge d\theta_2\wedge dT\beta)\wedge dPx_2\xi\\ & + (d\theta_2\wedge dx\wedge dT + d\theta_1\wedge d\theta_2\wedge dx\alpha)\wedge d\phi\\ & \times(\cos\phi x_3 - \sin\phi x_4)\xi, \end{aligned} \tag{5.15}$$

where for a matrix M which may be split into boson and fermion parts M_B, and M_F, so that $M = M_B + M_F$ we define ϵM to be given by

$$\epsilon M \overset{\text{def}}{=} M_B - M_F. \tag{5.16}$$

In terms of the forms α_1, α_2 we have

$$\begin{aligned} -d\Omega = {} & \alpha_1x_2\xi + \alpha_2(\cos\phi x_3 - \sin\phi x_4)\xi\\ & - im\sin\phi\sigma x_2\xi + \sigma P(\cos\phi x_3 - \sin\phi x_4)\xi\\ & + (d\theta_1\wedge dx\wedge dT + d\theta_1\wedge d\theta_2\wedge dT\beta) \end{aligned}$$

$$\times(\epsilon x_1 - P\epsilon x_2)d\xi + (d\theta_2 \wedge dx \wedge dT$$
$$+ d\theta_1 \wedge d\theta_2 \wedge dx\alpha) \wedge (\sin\phi\epsilon x_3 + \cos\phi\epsilon x_4)\delta\xi \tag{5.17}$$

If η is a 1-form given as

$$\eta = n_1 d\theta_1 + \eta_2 d\theta_2 + \eta_3 dx + \eta_4 dT, \tag{5.18}$$

then

$$\eta \wedge \Omega = (\eta_1 - \eta_4\alpha)(d\theta_1 \wedge dx \wedge dT + \beta d\theta_1 \wedge d\theta_2 \wedge dT) \wedge d\xi$$
$$+ (\eta_2 + \eta_3\beta)(d\theta_2 \wedge dx \wedge dT + \alpha d\theta_1 \wedge d\theta_2 \wedge dx)$$
$$\wedge d\xi + (\eta_1 - \eta_4\alpha)\sigma G\xi + (\eta_2 + \eta_3\beta)\sigma F\xi. \tag{5.19}$$

Therefore, we will have

$$d\Omega = -\eta \wedge \Omega \mathrm{mod}\langle\alpha_1,\alpha_2\rangle, \tag{5.20}$$

provided that

$$\eta_1 - \eta_4\alpha = (\epsilon X_1 - P\epsilon X_2) \tag{5.21}$$

and

$$\eta_2 + \eta_3\beta = (\sin\phi\epsilon X_3 + \cos\phi\epsilon X_4), \tag{5.22}$$

together with the relationship

$$-im\sin\phi X_2 + P(\cos\phi X_3 - \sin\phi X_4)$$
$$= (\sin\phi\epsilon X_3 + \cos\phi\epsilon X_4)(X_1 + PX_2)$$
$$+ (\epsilon X_1 - P\epsilon X_2)(\sin\phi X_3 + \cos\phi X_4), \tag{5.23}$$

which may be decomposed into the equations

$$-imX_2 = (\epsilon X_3)X_1 + (\epsilon X_1)X_3, \tag{5.24}$$
$$X_3 = X_4X_2 - (\epsilon X_2)X_4, \tag{5.25}$$
$$-X_4 = X_3X_2 - (\epsilon X_2)X_3, \tag{5.26}$$
$$0 = (\epsilon X_4)X_1 + (\epsilon X_1)X_4, \tag{5.27}$$

These equations can be reduced to the form of an incomplete graded Lie algebra by splitting each of the X_i into a bosonic and fermionic part. If we write $X_i = X_i^B + X_i^{AF}$ then Eqs. (5.24)–(5.27) become

$$\{X_1^B,X_3^B\} - \{X_3^F,X_1^F\} = -imX_2^B, \tag{5.28}$$
$$[X_3^B,X_1^F] + [X_1^B,X_3^F] = -imX_2^F, \tag{5.29}$$
$$\{X_2^F,X_4^F\} + [X_4^B,X_2^B] = X_3^B, \tag{5.30}$$
$$[X_4^F,X_2^B] + \{X_4^B,X_2^F\} = X_3^F, \tag{5.31}$$
$$\{X_2^F,X_3^F\} + [X_3^B,X_2^B] = -X_4^B, \tag{5.32}$$
$$[X_3^F,X_2^B] + \{X_3^B,X_2^F\} = -X_4^F, \tag{5.33}$$
$$\{X_4^B,X_1^B\} - \{X_4^F,X_1^F\} = 0, \tag{5.34}$$
$$[X_4^B,X_1^F] - [X_4^F,X_1^B] = 0. \tag{5.35}$$

These equations are a generalized prolongation structure for the supersymmetric sine-Gordon equation.

A very much reduced algebra may be obtained if X_2, X_3 and X_4 are taken to be purely bosonic. The algebra then reduces to

$$[X_3^B,X_1^F] = 0, \tag{5.36}$$
$$[X_4^B,X_2^B] = X_3^B, \tag{5.37}$$
$$[X_3^B,X_2^B] = -X_4^B, \tag{5.38}$$
$$[X_4^B,X_1^F] = 0, \tag{5.39}$$
$$\{X_1^B,X_3^B\} = -imX_2^B. \tag{5.40}$$

We could also take X_1^F to be zero, but more generally it has the form

$$X_1^F = \theta_1 Y, \tag{5.41}$$

where Y is a bosonic matrix and satisfies the bracket relations

$$[X_3^B,Y] = 0 = [X_4^B,Y], \tag{5.42}$$

and any multiple $-\lambda I$ of the identity matrix would suffice. To obtain a representation of the matrices $X_1^B, X_2^B, X_3^B, X_4^B$ we will first attempt to close the X_2^B, X_3^B, X_4^B algebra and then impose the anticommutation result (5.40) as a constraint.

We can close the X_2^B, X_3^B, X_4^B algebra if we introduce the additional bracket relation

$$[X_3^B,X_4^B] = -\sigma^2 X_2. \tag{5.43}$$

If we put

$$X_4 = \sigma Y_1,\ X_3 = +\sigma Y_2 \quad \text{and} \quad X_2 = +Y_3, \tag{5.44}$$

we obtain the algebra

$$[Y_1,Y_2] = Y_3, \tag{5.45}$$
$$[Y_1,Y_3] = Y_2, \tag{5.46}$$
$$[Y_2,Y_3] = -Y_1, \tag{5.47}$$

and the constraint to be imposed is that

$$\{X_1^B,Y_2\} = -\frac{im}{\sigma}Y_3. \tag{5.48}$$

A three-dimensional representation of the algebra, (5.45)–(5.47), which is essentially the adjoint representation, is given by

$$Y_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix},\quad Y_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix},$$
$$Y_3 = \begin{pmatrix} 0 & -1 & 0 \\ +1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.49}$$

If we substitute Y_2 and Y_3 into (5.48) we find that any matrix of the form

$$X_1^B = \begin{pmatrix} a & 0 & c \\ 0 & e & \frac{im}{\sigma} \\ -c & \frac{-im}{\sigma} & -a \end{pmatrix} \tag{5.50}$$

will do. We choose to put $a = e = c = 0$ and take X_1^B in the form

$$X_1^B = \frac{im}{\sigma}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{5.51}$$

Sectioning Ω onto a solution manifold of the supersymmetric sine-Gordon equation we obtain the inverse scattering

problem

$$\xi_{,\theta_1} + i\theta_1\xi_{,T} = -(X_1 + PX_2)\xi, \tag{5.52}$$

$$\xi_{,\theta_2} - i\theta_2\xi_{,X} = -(X_3\sin\phi + X_4\cos\phi)\xi. \tag{5.53}$$

In terms of the representation just constructed we have the problem

$$D_1\xi = \begin{pmatrix} \lambda\theta_1 & P & 0 \\ -P & \lambda\theta_1 & \dfrac{-im}{\sigma} \\ 0 & \dfrac{im}{\sigma} & \lambda\theta_1 \end{pmatrix}\xi, \tag{5.54}$$

$$D_2\xi = \sigma\begin{pmatrix} 0 & 0 & -\sin\phi \\ 0 & 0 & -\cos\phi \\ -\sin\phi & -\cos\phi & 0 \end{pmatrix}\xi. \tag{5.55}$$

Both σ and λ are arbitrary constants in these equations.

A connection may be obtained if we try to relate this inverse scattering problem to the inverse scattering equations of the classical sine-Gordon. The operator D_1 has the property that

$$D_1D_1 = +i\partial_T, \tag{5.56}$$

and applying D_1 to each side of (5.54) gives

$$\partial_T\xi = \begin{pmatrix} -i\lambda & \partial_T\phi & \dfrac{Pm}{\sigma} \\ -\partial_T\phi & -i\left(\lambda + \dfrac{m^2}{\sigma^2}\right) & 0 \\ -\dfrac{Pm}{\sigma} & 0 & -i\left(\lambda + \dfrac{m^2}{\sigma^2}\right) \end{pmatrix}\xi. \tag{5.57}$$

If we select $m^2/\sigma^2 = -2\lambda$, i.e., $\sigma = im/\sqrt{2\lambda}$, we obtain

$$\partial_T\xi = \begin{pmatrix} i\lambda & \partial_T\phi & -i\sqrt{2\lambda}\,P \\ -\partial_T\phi & -i\lambda & 0 \\ i\sqrt{2\lambda}\,P & 0 & -i\lambda \end{pmatrix}\xi, \tag{5.58}$$

which can be thought of as an isospectral $\partial_x\lambda = 0$ eigenvalue problem. We note the similarity of the problem to the analogous classical equation to which it reduces in the absence of fermionic operators. Thus, our final inverse scattering problem is given by

$$D_1\xi = \begin{pmatrix} \lambda\theta_1 & P & 0 \\ -P & \lambda\theta_1 & -\sqrt{2\lambda} \\ 0 & \sqrt{2\lambda} & \lambda\theta_1 \end{pmatrix}\xi, \tag{5.59}$$

$$D_2\xi = \frac{im}{\sqrt{2\lambda}}\begin{pmatrix} 0 & 0 & -\sin\phi \\ 0 & 0 & -\cos\phi \\ -\sin\phi & -\cos\phi & 0 \end{pmatrix}\xi, \tag{5.60}$$

and is the inverse scattering formulation recently announced by Girardello and Sciuto.[18]

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of the Department of Mathematics, Oregon State University, Corvallis, Oregon, 97331, where he was visiting Associate Professor of Applied Mathematice for the academic year 1977–1978, during which time this paper was conceived.

[1]F.B. Estabrook and H.D. Wahlquist, J. Math. Phys. **16**, 1 (1975).
[2]F.B. Estabrook and H.D. Wahlquist, J. Math. Phys. **17**, 1993 (1976).
[3]F.B. Estabrook, H.D. Wahlquist, and Robert Hermann, Differential-Geometric Prolongations and Bäcklund Transformations, The 1976 Ames Research Center (NASA) Conference on the Geometric Theory of Non-Linear Waves (Math. Sci. Press, 1977).
[4]Robert Hermann, Interdisciplinary Mathematics Vol. XII, *The Geometry of Non-Linear Differential Equations, Bäcklund Transformations, and Solitons*, Part A (Math. Sci. Press, 1976).
[5]R.K. Dodd and D. Gibbons, Proc. R. Soc. London, Ser. A **359**, 411–433 (1978).
[6]H.C. Morris, J. Math. Phys. **17**, 1870 (1976).
[7]H.C. Morris, J. Math. Phys. **18**, 285 (1977).
[8]H.C. Morris, "Prolongation structures and nonlinear evolution equations in two spatial dimensions: A general class of equations," J. Phys. A (to appear).
[9]H.C. Morris, "Prolongation structures of nonlinear evolution equations in two and three spatial dimensions," The 1976 Ames Research Center (NASA) Conference on the Geometric Theory of Non-linear Waves. (Math. Sci. Press, 1977).
[10]H.C. Morris, J. Math. Phys. **19**, 85 (1978).
[11]C.N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[12]V. Zakharov and A. Belavin, Phys. Lett. B **73**, 53 (1978).
[13]S. Ferrara, L. Girardello, and S. Scuito, Phys. Lett. B **76**, 303 (1978).
[14]P. Di Vecchia and S. Ferrara, Nucl. Phys. B **130**, 93, (1977).
[15]J. Hruby, Nucl. Phys. B **131**, 275 (1977).
[16]V. Zakharov and A. Shabat, J. Funct. Anal. Appl. **8**, 226 (1974).
[17]V. Zakharov, "The Inverse Scattering Method," in *Solitons*, edited by R. Bullough (Springer, New York, to appear).
[18]L. Girardello and S. Scuito: CERN preprint TH. 2496, May 1978.

Numerical computation of off-shell two-body Coulomb amplitude values

Stanley J. Sramek
Physical Dynamics, Inc., P.O. Box 556, La Jolla, California 92038
and University of California, Department of Physics, La Jolla, California 92093

(Received 18 July 1979; accepted for publication 3 October 1979)

Practical methods of computing numerical values for off-shell two-body Coulomb scattering amplitudes are developed. The methods are based on well-known analytic representations of the amplitudes. The results of a comparative computational study of the different methods are presented. It is found that the methods produce numerical values sufficiently precise for use in treating many-body problems formulated in terms of two-body amplitudes.

I. INTRODUCTION

Many physical problems involving a system of three or more interacting bodies can be conveniently formulated in terms of the off-shell two-body amplitudes for scattering between every pair of bodies in the system. Therefore, an ability to compute accurate numerical values for these amplitudes is useful in solving the more complex problems. Examples of such problems involving simple Coulomb interactions are the binding or scattering of an electron or positron from a hydrogen atom, treated by the Faddeev equations approach.

For problems involving Coulomb interactions, several analytic representations are known for the off-shell two-body amplitude. These representations can be used as the starting points for developing practical procedures for computing accurate numerical values for the amplitudes. This author has conducted a comparative numerical study of three such procedures, each based on a different analytic representation, and each restricted to use for negative values of the two-body energy. Two of the procedures were developed by this author. For a many-body problem involving negative values of the total energy, the negative energy values of the two-body amplitudes are the only values entering the problem's formulation.

The off-shell scattering amplitude for any two-body system satisfies the well-know Lippmann–Schwinger integral equation. In operator form, this equation is

$$T(E) = V + V G_0(E)\, T(E),$$

where

$$G_0(E) = (E - H_0)^{-1},$$

E is an energy treated as an independent complex variable, H_0 is the relative two-body kinetic energy operator, and V is the scattering potential operator. In the plane-wave representation, this equation becomes

$$T(\mathbf{k}_1, \mathbf{k}_2; E) = V(\mathbf{k}_1, \mathbf{k}_2) + \int d\mathbf{k}\, \frac{V(\mathbf{k}_1, \mathbf{k})}{E - (\hbar^2 k^2/2\mu)}\, T(\mathbf{k}, \mathbf{k}_2; E), \quad (1)$$

where

$$V(\mathbf{k}_1, \mathbf{k}_2) = (2\pi)^{-3} \int d\mathbf{r} \exp[i(\mathbf{k}_2 - \mathbf{k}_1)\cdot\mathbf{r}]\, V(\mathbf{r}),$$

and μ is the two-body reduced mass. The off-shell amplitude and the potential can both be resolved into partial wave components:

$$t_l(k_1, k_2; E) = \pi^2 \int d(\hat{k}_1 \cdot \hat{k}_2)\, P_l(\hat{k}_1 \cdot \hat{k}_2)\, T(\mathbf{k}_1, \mathbf{k}_2; E),$$

$$v_l(k_1, k_2) = \pi^2 \int d(\hat{k}_1 \cdot \hat{k}_2)\, P_l(\hat{k}_1 \cdot \hat{k}_2)\, V(\mathbf{k}_1, \mathbf{k}_2),$$

where P_l is the Legendre polynomial. Note that k_1, k_2, and E mathematically are all independent variables. The l th partial wave amplitude satisfies the integral equation

$$t_l(k_1, k_2; E) = v_l(k_1, k_2) + \frac{2}{\pi} \int_0^\infty dk\, \frac{k^2 v_l(k_1, k)}{E - (\hbar^2 k^2/2\mu)}\, t_l(k, k_2; E). \quad (2)$$

For use in most many-body problems, numerical values of the partial wave amplitudes rather than of the full amplitude are needed.

For two-body Coulomb interactions, analytic expressions are known for the full potential and for its partial wave components:

$$V(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{2\pi^2} \cdot \frac{Ze^2}{|\mathbf{k}_1 - \mathbf{k}_2|^2},$$

$$v_l(k_1, k_2) = \frac{Ze^2}{2k_1 k_2}\, Q_l\left(\frac{k_1^2 + k_2^2}{2k_1 k_2}\right), \quad (3)$$

where Ze^2 is the product of the charges of the two bodies, and Q_l is the Legendre function of the second kind.

II. METHODS OF COMPUTATION

For all of the computational methods developed below, we rewrite the integral equations (1) and (2) as

$$T(\mathbf{k}_1, \mathbf{k}_2; E) = V(\mathbf{k}_1, \mathbf{k}_2) + W(\mathbf{k}_1, \mathbf{k}_2; E),$$

$$t_l(k_1, k_2; E) = v_l(k_1, k_2) + w_l(k_1, k_2; E).$$

The functions w_l and W, of course, satisfy the angular projection relation

$$w_l(k_1, k_2; E) = \pi^2 \int d(\hat{k}_1 \cdot \hat{k}_2)\, P_l(\hat{k}_1 \cdot \hat{k}_2) W(\mathbf{k}_1, \mathbf{k}_2 E). \quad (4)$$

[a]Present address: Department of Physics, University of Nebraska, Lincoln, Nebraska.

 0022-2488/80/020334-09$1.00

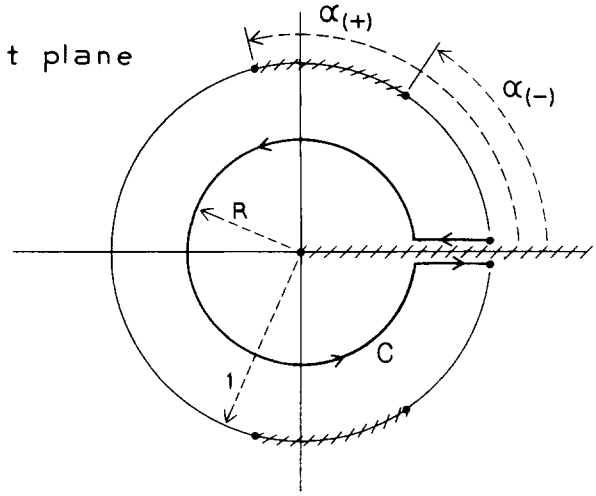


FIG. 1. The contour C along which the integrals (6) and (7) are defined. The cross-hatched segments along the positive real axis and the unit circle are the branch cuts in the integrand (7).

In every method numerical values for t_l are obtained by adding values of v_l and w_l together. The v_l values are obtained using Eq. (3), for which the values of the Q_l function are obtained using the simple recursion formula

$$Q_0(z) = \tfrac{1}{2}\log[(z+1)/(z-1)],$$
$$Q_1(z) = zQ_0(z) - 1,$$
$$Q_l(z) = (1/l)[(2l-1)zQ_{l-1}(z) - (l-1)Q_{l-2}(z)], \quad l > 1. \tag{5}$$

The methods differ only in the procedure used to obtain w_l values.

The Q_l functions exhibit branch point singularities in the complex z plane at $z = 1$ and $z = -1$. In computing values of the Q_l functions for complex arguments, the cut in the complex logarithmic function must be selected so that the Q_l functions exhibit a cut along the real axis between these branch point, but are continuous along all other parts of the real axis.

A. Method of contour integration

The development of the method of contour integration begins with the integral representation of the full amplitude given in a review article by Chen and Chen[1]:

$$W(\mathbf{k}_1,\mathbf{k}_2;E) = \frac{2\sigma Ze^2}{\pi^2(e^{i2\pi\sigma}-1)} \times \int_C dt \frac{t^{\sigma}}{S(1-t)^2 - 4t\,|\mathbf{k}_1 - \mathbf{k}_2|^2}, \tag{6}$$

where, for negative real energies,

$$\sigma = (Ze^2/\hbar)\,|\mu/2E\,|^{1/2},$$
$$S = (2\mu E - \hbar^2 k_1^2)\,(2\mu E - \hbar^2 k_2^2)/(2\mu E\hbar^2),$$

and C is any contour in the complex t plane of the form shown in Fig. 1. We obtain a contour integral representation for w_l by substituting Eq. (6) into the angular projection Eq. (4), interchanging the order of the integrations, and evaluating the angular integral analytically. This evaluation is simple, because the angular integral is of the tabulated form[2]

$$\int_{-1}^{1} dx \frac{P_l(x)}{a+bx} = -\frac{2}{b} Q_l\left(-\frac{a}{b}\right),$$

with

$$x = (\hat{k}_1 \cdot \hat{k}_2),$$
$$a = S(1-t)^2 - 4t(k_1^2 + k_2^2),$$
$$b = 8tk_1k_2.$$

We thus obtain

$$w_l(k_1,k_2;E) = -\frac{\sigma Ze^2}{2k_1k_2(e^{i2\pi\sigma}-1)} \times \int_C dt\, t^{\sigma-1} Q_l\left(\frac{4t(k_1^2+k_2^2) - S(1-t)^2}{8tk_1k_2}\right). \tag{7}$$

The above integrand exhibits several branch point singularities in the complex t plane, as shown in Fig. 1. The $t^{\sigma-1}$ factor gives rise to a branch point at $t = 0$, with the cut lying along the positive real axis. The Q_l factor gives rise to four branch points lying on the unit circle, two in the upper half and two in conjugate positions in the lower half plane. The angular positions of these branch points are

$$\cos\alpha_{(\pm)} = 1 + (2/S)(k_1 \pm k_2)^2,$$

where $\alpha_{(\pm)}$ is the angle between the branch point's complex radius vector and the positive real axis. The members of each pair of branch points are connected by a cut, also lying on the unit circle. In the limit $k_2 \to k_1$, $\alpha_{(-)} \to 0$; that is, the two branch points nearer the positive real axis move arbitrarily close to the endpoints of the C contour.

We now reduce the contour integral (7) to a form suitable for numerical evaluation. We consider the integrals along the circular segment and along the two straight segments of C separately, and make the following changes of variables:

$t = x$ (straight segment in u.h.p.),
$t = Re^{i\alpha}$ (circular segment),
$t = xe^{i2\pi}$ (straight segment in l.h.p.).

The new variables x and α are considered pure real. R is the radius of the circular segment of C, and satisfies the inequalities $0 < R < 1$. The phase factor $e^{i2\pi}$ assures that the phase difference between the integrals along the two straight segments of C, caused by the cut between them, is properly taken into account.

The integral along the straight segment in the l.h.p. is equivalent to that in the u.h.p., differing only in the direction of integration and in the phase factor due to the cut. The α integral, with limits $[0,2\pi]$, reduces to an integral with limits $[0,\pi]$ if the periodicity of the elementary trigonometric functions and the relationship $Q_l(z^*) = Q_l^*(z)$ are used. Using these simplifications, the integral expression (7) straightforwardly reduces to

$$w_l(k_1,k_2;E) = -(\sigma Ze^2/2k_1k_2) \times \{I_1 + R^{\sigma}[\mathrm{Re}(I_2)\cot(\pi\sigma) + \mathrm{Im}(I_2)]\},$$

where

$$I_1 = \int_R^1 dx\, x^{\sigma-1} Q_l\left(A - B\frac{(1-x)^2}{x}\right),$$

$$I_2 = \int_0^{\pi} d\alpha\, e^{i\sigma\alpha}\, Q_l[A - Bf(\alpha,R)],$$

$$A = (k_1^2 + k_2^2)/2k_1k_2, \qquad B = S/8k_1k_2,$$

$$f(\alpha,R) = (R + 1/R)\cos\alpha + i(R - 1/R)\sin\alpha - 2.$$

The integrals I_1 and I_2 can be computed easily by standard numerical methods. Values for the Q_l functions can be obtained using the recursion relation (5). The arguments of Q_l are always real in computing I_1 but are complex in computing I_2. The R parameter should be chosen so that the circular segment of the C contour is not too close to either the singularity at $t = 0$ or to the singularities on the unit circle. If k_1 is nearly equal to k_2, then two of the singularities on the unit circle are near the C contour's endpoints, and a dense grid of points must be used near $x = 1$ in computing I_1.

B. Method of angular projection

The development of the method of angular projection begins with the hypergeometric function representation of the full amplitude given by Ford.[3] For negative energies this representation can be written

$$W(\mathbf{k}_1,\mathbf{k}_2;E) = (2Ze^2/S\pi^2\sin\alpha)\ \mathrm{Im}[F(1,\sigma;\ 1+\sigma;e^{i\alpha})], \quad (8)$$

where F is the well-known hypergeometric function and α is defined by

$$\cos\alpha = 1 + (2/S)\,|\mathbf{k}_1 - \mathbf{k}_2|^2 \qquad 0 \leqslant \alpha \leqslant \pi.$$

In the method of angular projection, we use this hypergeometric function representation to compute numerical values of W, and then obtain values of w_l by performing the angular projection integration (4) numerically.

We must reduce Eq. (8) to a form suitable for numerical computation. By substituting the hypergeometric power series into the equation, we obtain

$$W(\mathbf{k}_1,\mathbf{k}_2;E) = \frac{2\sigma Ze^2}{S\pi^2\sin\alpha}\sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda+\sigma}. \quad (9)$$

This equation is not suitable for numerical computation. The magnitude of the series terms decreases as λ^{-1} for large λ, so that the series converges far too slowly for accurate term-by-term summation. Even slight inaccuracies in the W values can cause serious errors in the higher-l w_l values when the integration (4) is performed.

We can convert the series in Eq. (9) to a series in which the magnitude of the terms decreases as λ^{-4} for large λ, thus obtaining much more rapid numerical convergence. We write

$$\frac{1}{\lambda+\sigma} = \frac{1}{\lambda} - \frac{\sigma}{\lambda^2} + \frac{\sigma^2}{\lambda^3} - \frac{\sigma^3}{\lambda^3(\lambda+\sigma)}. \quad (10)$$

Equation (9) then becomes

$$W(\mathbf{k}_1,\mathbf{k}_2;E)$$

$$= (2\sigma Ze^2/S\pi^2\sin\alpha)[S_1^0 - \sigma S_2^0 + \sigma^2 S_3^0 - \sigma^3 S_4^\sigma],$$

where

$$S_1^0 = \sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda}, \qquad S_2^0 = \sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda^2},$$

$$S_3^0 = \sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda^3}, \qquad S_4^\sigma = \sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda^3(\lambda+\sigma)}.$$

Only S_4^σ must be summed numerically, because S_1^0, S_2^0, and S_3^0 can be summed analytically using the tabulated series[4]

$$\sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda} = \tfrac{1}{2}(\pi - \alpha), \qquad 0 < \alpha < 2\pi,$$

$$\sum_{\lambda=1}^{\infty}\frac{\cos\lambda t}{\lambda} = \tfrac{1}{2}\log[2(1-\cos t)], \qquad 0 < t < 2\pi,$$

$$\sum_{\lambda=1}^{\infty}\frac{1}{\lambda^2} = \tfrac{1}{6}\pi^2.$$

Clearly,

$$S_1^0 = \tfrac{1}{2}(\pi - \alpha),$$

$$S_2^0 = \int_0^{\alpha} dt\left(\sum_{\lambda=1}^{\infty}\frac{\cos\lambda t}{\lambda}\right)$$

$$= -\tfrac{1}{2}\int_0^{\alpha} dt\,\log[2(1-\cos t)],$$

$$S_3^0 = \int_0^{\alpha}\left(\sum_{\lambda=1}^{\infty}\frac{\cos\lambda t}{\lambda}\right)$$

$$= \int_0^{\alpha} dt\left[\sum_{\lambda=1}^{\infty}\frac{1}{\lambda^2} - \int_0^{t} du\left(\sum_{\lambda=1}^{\infty}\frac{\sin\lambda u}{\lambda}\right)\right]$$

$$= \int_0^{\alpha} dt\left(\tfrac{1}{6}\pi^2 - \tfrac{1}{2}\int_0^{t} du\,(\pi - u)\right)$$

$$= \tfrac{1}{12}\alpha(\pi - a)(2\pi - \alpha).$$

We thus obtain

$$W(\mathbf{k}_1,\mathbf{k}_2;E) = \frac{\sigma Ze^2}{S\pi^2\sin\alpha}\Big((\pi-\alpha) + \sigma[\alpha\log 2 + \mathrm{Sr}(\alpha)] + \tfrac{1}{6}\sigma^2\alpha(\pi-\alpha)(2\pi-\alpha) - 2\sigma^3 \times \sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda^3(\lambda+\sigma)}\Big), \quad (11)$$

where

$$\mathrm{Sr}(\alpha) \equiv \int_0^{\alpha} dt\,\log(1-\cos t). \quad (12)$$

Equation (11) is suitable for use in computing numerical values of W, which can then be used in performing the angular projection integration (4). A large set of values of the special function $\mathrm{Sr}(\alpha)$ can be tabulated by performing a single numerical integration; values to be used in Eq. (11) can be obtained from the tabulated values using some appropriate interpolation scheme. These values can be obtained with a high degree of precision with little difficulty, if both the numerical integration and the interpolation scheme take proper account of the logarithmic singularity at $t = 0$ of the integrand (12) (see Appendix A). The series in Eq. (11) must be summed numerically term by term. Such a numerical summation can include only a finite number of terms; the contribution of those terms $\lambda > \lambda_c$, where λ_c is the cutoff index, should be approximated by adding a single correction term to the finite sum (see Appendix B):

$$\sum_{\lambda=1}^{\infty}\frac{\sin\lambda\alpha}{\lambda^3(\lambda+\sigma)} \approx \sum_{\lambda=1}^{\lambda_c}\frac{\sin\lambda\alpha}{\lambda^3(\lambda+\sigma)} + \frac{\cos((\lambda_c+\tfrac{1}{2})\alpha)}{2\lambda_c^3(\lambda_c+\sigma)\sin\tfrac{1}{2}\alpha}. \quad (13)$$

The expansion (10), which makes possible the reduc-

tion of the series (9) to a more rapidly converging series, can be extended to terms of higher order in σ, thus resulting in a still more rapidly converging sum over λ. However, the resulting formula for W involves special functions defined by successive integrations of the special function $\mathrm{Sr}(\alpha)$. Each such successive integration magnifies the small errors present in the original set of $\mathrm{Sr}(\alpha)$ values, so that no advantage is gained.

C. Method of separable series summation

The development of the method of separable series summation begins with the well-known Sturmian functions expansion of the partial wave amplitude[5]:

$$w_l(k_1,k_2;E) = \sum_{n=l+1}^{\infty} \frac{\phi_{nl}(k_1,E)\phi_{nl}(k_2,E)}{\gamma_n(E)[1-\gamma_n(E)]}, \tag{14}$$

where

$$\gamma_n(E) = -\frac{n}{\sigma},$$

$$\phi_{nl}(k,E) = N_{nl}(E)\frac{(\hbar k)^l}{(\hbar^2k^2/2\mu - E)^{l+1}} \times C_{n-l-1}^{l+1}\left(\frac{\hbar^2k^2/2\mu + E}{\hbar^2k^2/2\mu - E}\right), \tag{15}$$

$$N_{nl} = \hbar^{3/2} l!\left(\frac{2^{4l+3}n(n-l-1!)}{(n+l)!}\right)^{1/2}\left(-\frac{E}{2\mu}\right)^{(2l+3)/4},$$

and C_{n-l-1}^{l+1} is a Gegenbauer polynomial. The w_l values are computed simply by summing the series (14) to a large number of terms. A limited numerical study of this method has previously been reported by Chen and Ishihara.[6]

This author wishes to point out that, although the above expressions for the Sturmian functions ϕ_{nl} appear somewhat cumbersome, numerical values of these functions to be used in summing the series can actually be obtained very easily. The Gegenbauer polynomials satisfy the recursion relation[7]

$$C_0^\lambda(x) = 1; \quad C_1^\lambda(x) = 2\lambda x;$$
$$C_n^\lambda(x) = (1/n)\,[2(n+\lambda-1)\,x\;C_{n-1}^\lambda(x) - (n+2\lambda-2)\;C_{n-2}^\lambda(x)], \quad n>1.$$

This recursion relation implies the following recursion relation for the Sturmian functions:

$$\phi_{l+1,l}(k,E) = N_{l+1,l}(E)\frac{(\hbar k)^l}{(\hbar^2k^2/2\mu - E)^{l+1}},$$
$$\phi_{l+2,l}(k,E) = (2l+4)^{1/2}\,x\phi_{l+1,l}(k,E),$$
$$\phi_{nl}(k,E) = (1/A_{nl})\,[2x\phi_{n-1,l}(k,E) - A_{n-1,l}\,\phi_{n-2,l}(k,E)] \quad n>l+2;$$

where

$$x = \left(\frac{\hbar^2k^2/2\mu + E}{\hbar^2k^2/2\mu - E}\right), \quad A_{nl} = \left(\frac{(n+l)(n-l-1)}{n(n-1)}\right)^{1/2}.$$

Once the Sturmian functions values for the first two terms of the series (14) are obtained, the values of all succeeding terms can be rapidly obtained using this recursion formula. This recursion formula has in fact been used by previous workers.[6]

D. Systems of units and normalizations

Previous papers on two-body Coulomb amplitudes and their application to many-body problems differ considerably among themselves in both the units systems to which their equations are specialized, and in the normalization conventions adopted in defining the partial wave amplitudes. Before performing any calculations combining the present work with any of this previous work, care must be taken to assure that all formulas involved use consistent units systems and normalizations.

All equations in this paper have been written without specialization to any system of units. To specialize them to atomic units, simply set $\hbar$ and e^2 equal to unity in all of them, and assume that the reduced mass is expressed in units of electron masses.

Some papers use a system of units involving mass-scaled momenta, defined by $p^2 = \hbar^2k^2/2\mu$. The equations in this paper can be converted to such a system by replacing each wave number k by $(2\mu)^{1/2}p/\hbar$. Papers using mass-scaled momenta sometimes further specialize their equations to atomic units, and sometimes specialize them instead to Rydberg units.

Some papers define the partial wave components v_l and t_l of the potential and of the amplitude in a manner differing by a factor of 2 from the definitions used in this paper. This difference or any similar difference in normalization conventions causes the multiplicative factors on all angular projection relations, and on the integral terms of all integral equations, to differ from the multiplicative factors appearing in this paper. This differing of multiplicative factors can be used to detect an underlying difference in normalization conventions.

III. COMPARATIVE NUMERICAL STUDY OF THE METHODS

The author has conducted a computational study of the three methods described above. This study was conducted in order to determine whether the numerical values of $t_l(k_1,k_2;E)$, produced by the methods, become stable for reasonably small values of each method's "convergence variables"; and to determine whether the methods' relative usefulness differs significantly in different regions of the space defined by the variables $(k_1,k_2;E)$. The convergence variables of each method are: the numbers of intervals N_1 and N_2 used in numerical evaluation of the integrals I_1 and I_2 (method of contour integration); the number of terms λ_c included in summing the modified hypergeometric series and the number of intervals N_a used in numerical evaluation of the angular projection integral (method of angular projection); and the number of terms N_t included in summing the Sturmian functions expansion (method of separable series summation). The radius R of the circular contour segment is an additional computation parameter involved in the method of contour integration.

A. Test procedure

A grid of test points in the space $(k_1,k_2;E)$ was selected. At each of these points, the values of t_l, for $l = 0$ through

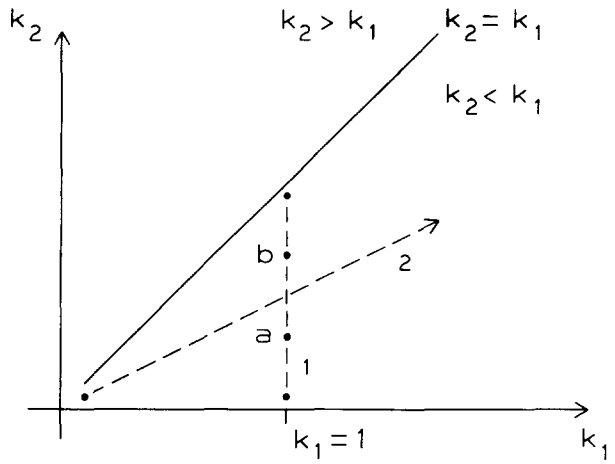


FIG. 2. The (k_1,k_2) plane. Lines 1 and 2 contain the pairs (k_1,k_2) at which the principal numerical test were conducted. The line $k_2 = k_1$ marks a logarithmic singularity in all $t_l(k_1,k_2;E)$. The coordinates of points a and are (1.0, 0.32) and (1.0, 0.68), respectively.

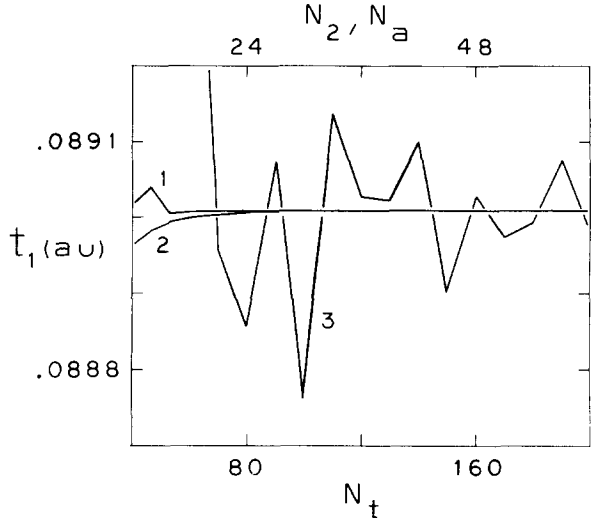


FIG. 3. The convergence of $t_1(k_1,k_2;E)$, for $Z = 1, \mu = 1.0, E = -0.375$, and $(k_1,k_2) = (1.0, 0.32)$. Curves 1, 2, and 3, respectively, represent the methods of contour integration, of angular projection, and of separable series summation. For each curve the abscissa represents N_2, N_a, and N_t, respectively. For curves 1 and 2, the values of R, N_1, and λ_c are held constant at 0.5, 60, and 60, respectively.

$l = 4$, were computed by each of the three methods as functions of that method's convergence variables. The maximum values used for the convergence variables were 60 for N_1, N_2, λ_c, and N_a, and 200 for N_t. The three methods consume comparable computer times using these maximum convergence variables for this entire range of l values. The values used for R were 0.2, 0.5, and 0.8.

Atomic units were used in all test computations. Most were performed for $\mu = 1.0$, $E = -0.375$, and $Z = 1$ and -1 both. Two sets of pairs (k_1,k_2) were used. In the first set, k_1 was fixed at 1.0 and k_2 took the values 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.50, 0.68, 0.84, 0.92, 0.96, 0.98, and 0.99. In the second set, the ratio k_2/k_1 was fixed at 0.5 and k_1 took the values 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 1.0, 3.0, 6.0, 12.0, 25.0, 50.0, and 100.0. A few additional computations were performed, some for $\mu = 0.5$, some for $E = -0.625$ and -0.187, and some for a few pairs (k_1,k_2) lying off the two principal sets. These additional computations were used to check the general validity of conclusions drawn from the basic tests.

Figure 2 depicts the plane (k_1, k_1). Because $t_l(k_2,k_1;E) = t_l(k_1,k_2;E)$, all tests may be restricted to the region $k_1 > k_2$. The two principle sets of test points lie along the two dashed lines. The first set (line 1) is used to investigate the limiting behavior of the three methods as $k_2 \to 0$ and $k_2 \to k_1$ for fixed k_1. The second set (line 2) is used to investigate the limiting behavior as k_1 and k_2 both $\to 0$ and both $\to \infty$. The line $k_2 = k_1$ marks a logarithmic singularity in t_l.

At each of the principal test points (k_1,k_2), the t_l values produced by the three methods were compared, to determine which method's results converged most rapidly or least rapidly to stable numerical values as functions of the convergence variables. From these comparisons, general conclusions were drawn regarding the method's relative usefulness in different parts of the (k_1,k_2) plane.

B. Test results

Figures 3 and 4 illustrates the test results, for $l = 1$ and $Z = 1$, at the two points $(k_1,k_2) = (1.0, 0.32)$ and $(1.0, 0.68)$; these points are labeled a and b in Fig. 2. Each curve represents the t_1 value, computed by one of the three methods, as a function of one of that method's convergence variables. For the method of contour integration (curve 1) the abscissa represents N_2; for the method of angular projection (curve 2) the abscissa represents N_a; and for the method of separable series summation (curve 3) the abscissa represents N_t. For curve 1, the convergence variable N_1 is constant at its maximum value 60, and the parameter R is constant at 0.5. For curve 2, the convergence variable λ_c is constant at its maximum value 60. As mentioned previously, the computations by the three methods for the maximum values of the convergence variables consumed comparable computer times; it is

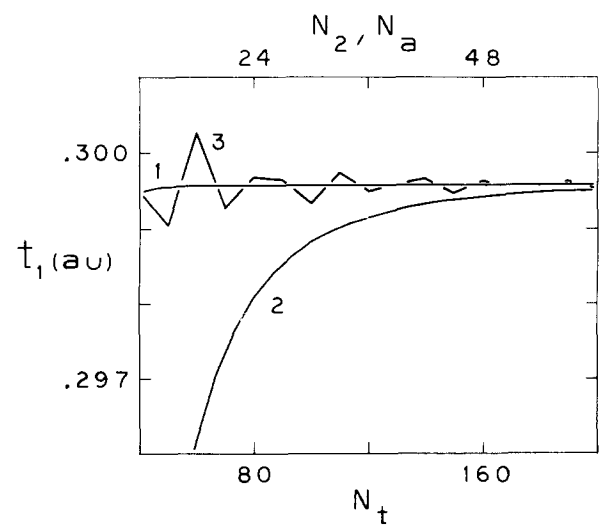


FIG. 4. The same as Fig. 3, but with $(k_1,k_2) = (1.0, 0.68)$.

TABLE I. The convergence test results along line 1 (Fig. 2) for the method of contour integration; for $E = -0.375$, $Z = 1$, and $\mu = 1.0$. Each entry is the number of significant figures, in the computed values of $t_l(k_1,k_2;E)$, judged stable as N_2 varies upward to 60, for a fixed $N_1 = 60$ and $R = 0.5$.

k_2/k_1	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$
0.01	6	6	3	1	0
0.02	5	6	1	0	0
0.04	6	6	5	2	0
0.08	6	7	7	5	2
0.16	6	6	6	6	5
0.32	5	6	6	5	6
0.50	6	6	5	5	6
0.68	6	6	6	5	5
0.84	4	5	6	5	5
0.92	5	5	5	5	6
0.96	6	5	6	5	6
0.98	6	7	7	6	7
0.99	6	6	7	7	7

TABLE III. The same as Table I, but for the method of separable series summation. Each entry is the number of significant figures in t_l judged stable as N_t varies upward to 200.

k_2/k_1	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$
0.01	2	0	0	0	0
0.02	2	2	0	0	0
0.04	4	1	0	0	0
0.08	3	2	1	0	0
0.16	5	2	2	0	0
0.32	4	1	3	1	1
0.50	4	3	2	3	2
0.68	3	3	3	4	3
0.84	4	3	3	4	3
0.92	2	3	3	3	2
0.96	3	4	2	3	3
0.98	3	3	4	3	3
0.99	4	4	4	4	4

therefore reasonable to plot three curves on one graph for comparison purposes even though the three curves' abscissae have different mathematical meanings.

In Fig. 3, we see that curves 1 and 2 very quickly converge to the same stable value of approximately 0.08901, whereas curve 3 oscillates about this value and converges very slowly, if at all. In Fig. 4, we see that curve 1 again very quickly converges to a stable value, and that curve 3 again oscillates about this value of approximately 0.2996. But curve 3 now clearly converges to this same value. Curve 2 also appears to converge to this same value, but only very slowly. Thus, for these two examples of the test results, we see the following:

(1) The method of contour integration produces rapidly converging results at both point a, near the $k_2 \to 0$ limiting region, and at point b, near the $k_2 \to k_1$ logarithmic singularity.

(2) The method of angular projection produces rapidly converging results at point a but slowly converging results at point b.

TABLE II. The same as Table I, but for the method of angular projection. Each entry is the number of significant figures in t_l judged stable as N_a varies upward to 60, for a fixed $\lambda_c = 60$.

k_2/k_1	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$
0.01	9	8	6	0	0
0.02	8	6	4	0	0
0.04	9	7	5	1	0
0.08	7	7	5	3	0
0.16	7	7	4	3	1
0.32	5	5	4	3	2
0.50	5	4	3	3	3
0.68	3	3	2	2	1
0.84	2	2	1	1	1
0.92	1	1	1	1	1
0.96	1	0	0	0	0
0.98	1	0	0	0	0
0.99	0	0	0	0	0

(3) The method of separable series summation produces very poorly converging results at point a but fairly rapidly converging results at point b.

These two examples illustrate a general conclusion drawn from an analysis of all of the test results, which we now describe.

The test results were analyzed by first condensing the numerical data generated to 36 tables, six examples of which are shown here (Tables I–VI). Table I, for example, represents the test results for the method of contour integration, for all l values tested and for all (k_1,k_2) pairs lying along line 1 in Fig. 2, for $Z = 1$, $R = 0.5$, and for the convergence variable N_1 held fixed at its maximum value of 60 while N_2 is varied. Each entry of Table I is the number of significant figures, in the numerical value of a particular $t_l(k_1,k_2;E)$, judged as having become stable when N_2 is varied up to its maximum value of 60. Table VI, for example, also represents the test results for the method of separable series summation, for all l values tested and for all (k_1,k_2) pairs laying along line 2 in Fig. 2, for $Z = 1$. Each entry is the number of significant figures judged stable when N_t is varied up to its maximum value of 200. By studying the 36 tables, general conclusions can be drawn.

TABLE IV. The same as Table I, but along line 2 (Fig. 2).

k_1	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$
0.01	8	7	4	1	0
0.02	7	7	4	0	0
0.04	6	6	5	1	0
0.08	6	6	8	4	1
0.16	6	6	6	4	3
0.32	6	6	6	5	4
1.0	6	6	5	5	6
3.0	8	8	8	8	9
6.0	8	8	8	8	7
12.0	8	8	7	8	7
25.0	8	6	6	6	6
50.0	7	7	6	7	6
100.0	6	4	5	4	5

TABLE V. The same as Table II, but along line 2 (Fig. 2).

k_1	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$
0.01	7	5	6	5	4
0.02	6	5	5	5	4
0.04	6	6	4	4	4
0.08	6	4	4	4	3
0.16	6	5	4	4	2
0.32	5	5	4	3	3
1.0	5	4	3	3	3
3.0	5	5	3	3	3
6.0	5	4	4	3	2
12.0	5	5	4	3	3
25.0	6	5	5	4	3
50.0	6	6	5	4	3
100.0	7	6	5	4	4

The results obtained by the method of contour integration are of comparable precision for $R = 0.2$ and $R = 0.5$, and of poorer quality for $R = 0.8$. This result is not surprising, because for $R = 0.8$ the circular part of the contour (Fig. 1) passes close to the logarithmic singularities on the unit circle. The results converge more rapidly as N_1 increases than as N_2 increases; therefore, in comparing the method of contour integration with the other methods, we consider primarily its convergence precision as N_2 increases.

The results obtained by the method of angular projection converge more rapidly as λ_c increases than as N_a increases; therefore, in comparing this method with the other methods, we consider primarily its convergence precision as N_a increases.

Consider now Tables II and III. Table II shows that, along line 1 (Fig. 2), the method of angular projection in general gives good result for $k_2/k_1 \leqslant 0.5$. In the limit $k_2 \to 0$ we see that, for $l = 0$, 1, and 2, the method continues to give results accurate to approximately four to eight significant figures. For $l = 3$ and 4 the method deteriorates as $k_2 \to 0$; however, approximate values of t_l for very small k_2 values can be obtained by extrapolation from the values obtained for somewhat larger k_2, assuming a k_2^l dependence. For $k_2/k_1 > 0.5$, the angular projection results deteriorate rapidly. As $k_2 \to k_1$, that is, as we near the logarithmic singularity in t_l, this method becomes useless. On the other hand, Table III shows that the method of separable series summation in general gives poor results for $k_2/k_1 < 0.5$ and good results for $k_2/k_1 \geqslant 0.5$. As $k_2 \to 0$, this method's results deteriorate rapidly for all l except $l = 0$. As $k_2 \to k_1$, this method continues to give results accurate to approximately three or four significant figures, for all l. Thus, the methods of angular projection and of separable series summation can be regarded as complementary: In the region $k_2/k_1 < 0.5$ the former produces accurate results, and in the region $k_2/k_1 > 0.5$ the latter produces accurate results. Figures 3 and 4, previously discussed, illustrate this general conclusion.

Consider now Table I. This table shows that the method of contour integration produces good results almost everywhere along line 1, showing only some deterioration for higher l values as $k_2 \to 0$. This deterioration appears to be comparable in seriousness to that shown by the method of angular projection as $k_2 \to 0$. For $l = 0$ and 1, this method continues to produce accurate results as $k_2 \to 0$, although this accuracy is not quite as great as that attained by angular projection. As $k_2 \to k_1$, this method continues to give results accurate to approximately six or seven significant figures, for all l. Thus, the method of contour integration produces results superior to those produced by the method of separable series summation along that part of line 1 where the latter method produces its best results; there is no portion of line 1 where the method of separable series summation is not inferior to at least one of the other two methods. In the region $k_2/k_1 \leqslant 0.5$, the methods of contour integration and angular projection appear equally useful.

We now consider the tests conducted along line 2 (Fig. 2). Table IV indicates that the method of contour integration remains accurate to four or five significant figures for all l as k_1 and k_2 both $\to \infty$, but that serious deterioration occurs for higher l as k_1 and k_2 both $\to 0$. Table V indicates that the method of angular projection remains accurate to four to six significant figures for all l both as k_1 and $k_2 \to \infty$ and as k_1 and $k_2 \to 0$. Table VI indicates that the method of separable series summation deteriorates substantially both as k_1 and $k_2 \to \infty$ and as k_1 and $k_2 \to 0$, thus confirming the conclusion that it is always inferior to at least one of the other two methods.

We thus reach the following final conclusions, which are also supported by an analysis of the tabulated test results not presented here:

(1) In the entire region $k_2/k_1 < 0.5$, the methods of contour integration and of angular projection are approximately equally useful; the method of angular projection is probably preferable because its results show no deterioration along line 2. The method of separable series summation is useless in this region.

(2) In the entire region $k_2/k_1 > 0.5$, the method of contour integration is clearly superior to both of the other methods.

Figures 3 and 4, previously discussed, illustrate these general conclusions.

C. Mathematical significance of test results

The failures of the methods of angular projection and of

TABLE VI. The same as Table III, but along line 2 (Fig. 2)

k_1	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$
0.01	3	2	2	2	3
0.02	3	2	2	1	1
0.04	3	3	2	3	1
0.08	3	2	2	1	1
0.16	4	4	2	3	2
0.32	3	3	3	2	2
1.0	4	3	2	3	2
3.0	3	3	4	3	3
6.0	4	3	3	2	2
12.0	3	3	2	2	1
25.0	3	2	2	1	1
50.0	2	3	1	2	2
100.0	3	2	2	2	1

separable series expansion along certain portions of line 1 can be understood in terms of the methods' basic mathmatical natures.

The method of angular projection fails for higher l values as $k_2 \to 0$, and for all l values as $k_2 \to k_1$. To understand these failures, we consider the qualitative behavior of the full amplitude $W(\mathbf{k}_1,\mathbf{k}_2;E)$, as a function of $\hat{k}_1 \cdot \hat{k}_2$, in these two limits.

From Eq. (8), we see that $W(\mathbf{k}_1,\mathbf{k}_2;E)$ is a continuous function of $\mathbf{k}_2$ at $\mathbf{k}_2 = 0$, so that, for k_2 small but nonzero, its dependence on the direction $\hat{k}_2$ must be very weak. The integrand of the angular projection integral (4) is therefore a nearly constant function of $\hat{k}_1 \cdot \hat{k}_2$; its slight deviations from a constant value determine the integral's nonzero value for higher l values. Even small errors in computing the numerical values of W will *not* be small in comparison to these slight deviations, and will therefore substantially affect the integral's value for higher l. The method thus fails for k_2 near zero, for higher l.

From Eq. (8), we see that, for $k_2 = k_1$, $W(\mathbf{k}_1,\mathbf{k}_2;E)$ is a singular function of $\hat{k}_1 \cdot \hat{k}_2$ in the limit as the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$ becomes zero. Therefore, for k_2 nearly equal to k_1, the integrand of the angular projection integral (4) is nearly singular at $\hat{k}_1 \cdot \hat{k}_2 = 1$, so that the integral cannot be accurately computed. The method thus fails for k_2 near k_1.

The method of separable series summation fails as $k_2 \to 0$ for $l \neq 0$. To understand this failure, we consider the series (14), and investigate the $n \to \infty$ behavior of the terms for small k_2.

From Eq. (15), we see that, for $k_2 \to 0$, the argument of the Gegenbauer polynomial C^{l+1}_{n-l-1} becomes -1. The values of the Gegenbauer polynomials for arguments of unit magnitude are known[8]:

$$\begin{aligned} C^{l+1}_{n-l-1}(-1) &= (-1)^{n-l-1} C^{l+1}_{n-l-1}(1) \\ &= [(-1)^{n-l-1}/(2l+1)!\,] \\ &\quad \times [(n+l!)/(n-l-1)!\,]; \end{aligned}$$

furthermore,

$$(n+l)!/(n-l-1)! \sim n^{2l+1} \quad \text{as } n \to \infty.$$

The factors involving $\gamma_n(E)$ then introduce an n^{-2} dependence for large n, so that as $n \to \infty$ the magnitude of the terms of the Sturmian functions expansion behaves as n^{2l-1}. Thus, for $l = 0$, the series will converge numerically for small k_2, although rather slowly. But for $l \neq 0$ the series will fail to converge numerically.

ACKNOWLEDGMENTS

I wish to thank Professor Joseph C.Y. Chen for suggesting this investigation to me, and for helpful discussion during the investigation. This research was supported in part by the Physical Dynamics, Inc., Independent Research and Development Fund; and in part by a grant from the Committee on Research, University of California at San Diego.

APPENDIX A: NUMERICAL VALUES OF Sr(α)

To tabulate accurate values of $\mathrm{Sr}(\alpha)$, write the integrand (12) as a sum of two terms:

$$\log(1-\cos t) = 2\log t + \log[(1-\cos t)/t^2].$$

The first term, containing the logarithmic singularity, can be integrated analytically. The result is

$$\mathrm{Sr}(\alpha) = 2\alpha(\log\alpha - 1) + \int_0^\alpha dt \log\left(\frac{1-\cos t}{t^2}\right). \tag{16}$$

The remaining term must be integrated numerically. Because this integrand has no singularity at $t = 0$ the integration can be accurately performed using some standard method such as Simpson's rule.

The numerical value of $\mathrm{Sr}(\alpha)$ for some arbitrary α can be obtained from the tabulated values by applying an appropriate interpolation scheme to those values tabulated at points lying near the α of interest. The author's test computations showed that, for $\alpha > \pi/20$, a simple polynomial interpolation scheme applied directly to the tabulated Sr values produces satisfactory results. The author's interpolation scheme consisted of fitting a third order polynomial to the Sr values tabulated at the four points nearest the α of interest. The author tabulated the values of Sr at the endpoints of 400 intervals of length $\pi/400$, covering the interval $[0,\pi]$.

For $\alpha < \pi/20$, such an interpolation scheme produces unsatisfactory results. The function $\mathrm{Sr}(\alpha)$ has an infinite derivative at $\alpha = 0$, and no polynomial fit applied directly to the tabulated Sr values for small α can adequately mimic this infinite slope feature of the function. To obtain $\mathrm{Sr}(\alpha)$ for small α, the two portions (16) of the function Sr must be treated separately. The first portion is known analytically; and the second, because it does not exhibit the infinite slope feature, can be accurately treated with a simple polynomial interpolation scheme.

APPENDIX B: NUMERICAL SUMMATION OF CERTAIN SERIES

Consider two series, of the forms

$$C^\sigma_n = \sum_{\lambda=1}^{\infty} \frac{\cos\lambda\alpha}{\lambda^{n-1}(\lambda+\sigma)},$$

$$S^\sigma_n = \sum_{\lambda=1}^{\infty} \frac{\sin\lambda\alpha}{\lambda^{n-1}(\lambda+\sigma)}.$$

A series of the latter form, with $n = 4$, must be summed mathematically in computing Coulomb amplitude values by the method of angular projection. To study the numerical convergence properties of such series, and to justify use of the correction term in Eq. (13), we first note that the two series are the real and imaginary parts of a series

$$E^\sigma_n \equiv \sum_{\lambda=1}^{\infty} \frac{e^{i\lambda\sigma}}{\lambda^{n-1}(\lambda+\sigma)} \equiv \sum_{\lambda=1}^{\infty} z(\lambda).$$

Let λ_c denote a cutoff index. We denote the series truncated after λ_c terms as

$$\epsilon(\lambda_c) \equiv \sum_{\lambda=1}^{\lambda_c} z(\lambda).$$

We wish to determine a complex correction terms $c(\lambda_c)$ such that $\epsilon(\lambda_c) + c(\lambda_c) \to E^\sigma_n$ much more rapidly as $\lambda_c \to \infty$ than does $\epsilon(\lambda_c)$.

We regard E^σ_n and all of the z, c, and ϵ as vectors lying in the complex plane, and study their geometric properties.

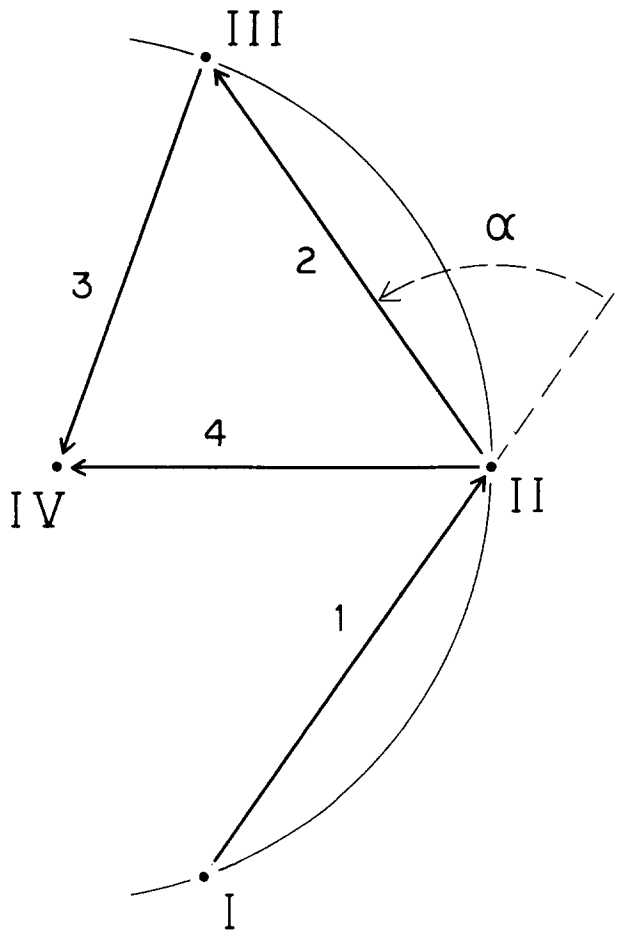


FIG. 5. The geometric construction of the complex correction term $c(\lambda_c)$. The labeled points are: I, $\epsilon(\lambda_c - 1)$; II, $\epsilon(\lambda_c)$; III, $\epsilon(\lambda_c + 1)$; and IV, E_n^σ. The labeled vectors are: 1, $z(\lambda_c)$; 2, $z(\lambda_c + 1)$; 3, $c(\lambda_c + 1)$; and 4, $c(\lambda_c)$. The curve represents the spiral path along which the points ϵ lie. The segment shown is considered approximately circular, with the point IV lying at its center.

The most important properties are: For all λ, the angle between vectors $z(\lambda)$ and $z(\lambda + 1)$ is exactly equal to α, and, for sufficiently large λ, the ratio of these vectors' magnitudes becomes arbitrarily close to unity. Since each $\epsilon(\lambda_c)$ is the vector sum of the first λ_c vectors $z(\lambda)$, these properties imply: For a sufficiently large minimum λ_c', the chain of points $\epsilon(\lambda_c)$ for $\lambda_c > \lambda_c'$ lies on a spiral path, which very slowly spirals inward around the point E_n^σ. For any $\lambda_c > \lambda_c'$, the points ϵ lying near $\epsilon(\lambda_c)$ lie approximately on a circle, the radius of which very slowly diminishes to zero as $\lambda_c \to \infty$. (This circle's radius is initially much smaller for $n = 4$ than for $n = 1$, but in either case this radius diminishes to zero very slowly as $\lambda_c \to \infty$.) The point E_n^σ lies very near the center of this circle. We therefore choose $c(\lambda_c)$ as a vector which, when added to $\epsilon(\lambda_c)$, reaches the point lying at that center.

Consider Fig. 5. The points $\epsilon(\lambda_c - 1)$, $\epsilon(\lambda_c)$, $\epsilon(\lambda_c + 1)$, and E_n^σ; and the vectors $z(\lambda_c)$, $z(\lambda_c + 1)$, $c(\lambda_c)$, and $c(\lambda_c + 1)$ are shown. Also shown is a part of the spiral on which the points ϵ lie. By vector addition,

$$c(\lambda_c) = z(\lambda_c + 1) + c(\lambda_c + 1).$$

Because of the approximate circular geometry,

$$z(\lambda_c + 1) \approx z(\lambda_c)\, e^{i\alpha},$$

$$c(\lambda_c + 1) \approx c(\lambda_c)\, e^{i\alpha}.$$

Therefore,

$$c(\lambda_c) \approx [z(\lambda_c) + c(\lambda_c)]\, e^{i\alpha}, \qquad c(\lambda_c) \approx z(\lambda_c)/(e^{-i\alpha} - 1).$$

By taking the real and imaginary parts of this $c(\lambda_c)$ we finally obtain correction terms to be used in truncating the series C_n^σ and S_n^σ:

$$C_n^\sigma \approx \sum_{\lambda=1}^{\lambda_c} \frac{\cos\lambda\alpha}{\lambda^{n-1}(\lambda + \sigma)} - \frac{\sin((\lambda_c + \frac{1}{2})\alpha)}{2\lambda_c^{n-1}(\lambda_c + \sigma)\sin\frac{1}{2}\alpha},$$

$$S_n^\sigma \approx \sum_{\lambda=1}^{\lambda_c} \frac{\sin\lambda\alpha}{\lambda^{n-1}(\lambda + \sigma)} + \frac{\cos((\lambda_c + \frac{1}{2})\alpha)}{2\lambda_c^{n-1}(\lambda_c + \sigma)\sin\frac{1}{2}\alpha}.$$

[1] J.C.Y. Chen and A.C. Chen, Adv. At. Mol. Phys. **8**, 88 (1972).
[2] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series, and Products (Academic, New York, 1965), 4th ed., p. 821.
[3] W.F. Ford, J. Math. Phys. 7, 626 (1966).
[4] Ref. 2, pp. 38, 7, and 1079.
[5] Ref. 1, p. 92.
[6] J.C.Y. Chen and T. Ishihara, Phys. Rev. **186**, 25 (1969).
[7] Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, p. 178.
[8] Ref. 7, p. 176.

Remarks on characteristic classes of four-dimensional Einstein manifolds

Y. L. Xin

Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794 and Department of Mathematics, Fudan University, Shanghai, China

(Received 28 June 1979; accepted for publication 5 September 1979)

Using the decomposition of the curvature tensor by means of the dual * operator, several useful expressions for characteristic classes in some kinds of Einstein manifolds are obtained. Consequently, some topological restrictions of the underlying manifolds are considered.

1. INTRODUCTION

The most intersting Riemannian manifolds in physics have dimension 4. In this case the duality * operator plays an important role. The Riemannian curvature tensor can be decomposed by means of the * operator. Using such a decomposition, it has been shown that there exist several inequalities between the Euler number, Pontrjagin number, and the Yang–Mills functional on the manifolds.[1] This paper discusses some important special cases. For Einstein–Kähler manifolds and half-conformally flat Einstein manifolds, which are interesting to the gauge theory,[2] we obtain several useful expressions for the Euler class and Pontrjagin class, from which some topological restrictions of the manifolds are obtained. In particular, H. Donnelly's inequality[3] for the Einstein–Kähler manifolds is a consequence of these expressions.

2. DECOMPOSITION OF THE CURVATURE TENSOR AND CHARACTERISTIC CLASSES

Let M be an oriented Riemannian manifold of dimension 4, and let R and Λ^p be the curvature tensor and the bundle of exterior p-forms of the manifold M, respectively. The Riemannian curvature tensor defines a symmetric transformation in Λ^2 given by

$$R(e_h \Lambda e_i) = \frac{1}{2}\sum_{j,k} R_{hijk}(e_j \Lambda e_k),$$

where $\{e_i\}$ is a local orthonormal basis of 1-forms. At any point $x \in M$ such linear transformations form a vector space with inner product given by $\langle R \cdot R' \rangle = \mathrm{tr} R R'$.

The duality operator $*: \Lambda^2 \to \Lambda^2$ is defined by

$$\alpha \Lambda *\beta = (\alpha, \beta)\Omega,$$

where $\alpha, \beta \in \Lambda^2$, (α, β) is the inner product on 2-forms, and Ω is the volume form.

As is well known, the bundle Λ^2 splits a direct sum

$$\Lambda^2 = \Lambda^2_+ \oplus \Lambda^2_-,$$

where $\Lambda^2_\mp$ are the ∓ 1 eigenspaces of the * operator.

For the sake of the simplicity we denote

$$*R(e_h \Lambda e_i) = \frac{1}{2}\sum_{j,k} R_{hijk} *(e_j \Lambda e_k) = \frac{1}{2}\sum_{j,k} R^{*}_{hijk}(e_j \Lambda e_k).$$

Setting

$$R^{*}_{hijk} = R^{*}_{jkhi}, \tag{2.1}$$

we define

$$R^{*}(e_h \Lambda e_i) = \tfrac{1}{2}\sum_{j,k} R^{*}_{hijk}(G_j \Lambda e_k),$$

$$*R^{*}(e_h \Lambda e_i) = \tfrac{1}{2}\sum_{j,k} R^{**}_{hijk}(e_j \Lambda e_k).$$

Moreover, we define

$$\begin{aligned} R^{++} &= \tfrac{1}{4}(R + R^{*\cdot} + R^{\cdot *} + R^{**}), \\ R^{--} &= \tfrac{1}{4}(R - R^{*\cdot} - R^{\cdot *} + R^{**}), \\ R^{+-} &= \tfrac{1}{4}(R + R^{*\cdot} - R^{\cdot *} - R^{**}), \\ R^{-+} &= \tfrac{1}{4}(R - R^{*\cdot} + R^{\cdot *} - R^{**}). \end{aligned} \tag{2.2}$$

It is well known that $R^{+-} = 0$ is equivalent to $R^{-+} = 0$ and M is a Einstein manifold iff $R^{+-} = 0$[1] which is the case of our consideration through this paper. On the other hand, the curvature R can be written as a block matrix relative to the decomposition $\Lambda^2_+ \oplus \Lambda^2_-$:

$$\begin{pmatrix} A^{+} & B \\ B^{T} & A^{-} \end{pmatrix},$$

where $|A^{+}|^2 = |R^{++}|^2$, $|A^{-}|^2 = |R^{--}|^2$, and $|B|^2 = |B^T|^2 = |R^{+-}|^2$. Moreover, we have $\mathrm{tr} A^{+} = \mathrm{tr} A^{-} = \frac{1}{4}s$, where s is the scalar curvature, which is a constant in an Einstein manifold, and

$$\begin{aligned} W^{+} &= A^{+} - \frac{s}{12} id, \\ W^{-} &= A^{-} - \frac{s}{12} id, \\ W &= W^{+} + W^{-}, \end{aligned} \tag{2.3}$$

where id denotes the identity transformation and W is the Weyl conformal curvature tensor.

We define the manifolds with $W^{+} = 0$ or $W^{-} = 0$ as the half-conformally flat manifolds.

From the above relations the Euler characteristic class[4,1] of M can be expressed as

$$\frac{1}{32\pi^2}\sum \epsilon_{i_1 \cdots i_4} R(e_{i_1} \Lambda e_{i_2}) \Lambda R(e_{i_3} \Lambda e_{i_4}) = \frac{1}{8\pi^2}(|R^{++}|^2 + |R^{--}|^2 - 2|R^{+-}|^2)\Omega \tag{2.4}$$

and

$$\frac{1}{8\pi^2}\left(|W^{+}|^2 + |W^{-}|^2 + \frac{s^2}{24} - 2|R^{+-}|\right)\Omega. \tag{2.5}$$

Similarly, the first Pontrjagin class[4,1] is

 0022-2488/80/020343-04$1.00

$$\frac{1}{8\pi^2}\sum \delta^{j_1 j_2}_{i_1 i_2} R(e_{i_1}\Lambda e_{j_1})\Lambda R(e_{i_2}\Lambda e_{j_2})$$

$$= \frac{1}{4\pi^2}(|R^{++}|^2 - |R^{--}|^2)\Omega$$

$$\frac{1}{4\pi^2}(|W^+|^2 - |W^-|^2)\Omega\,. \tag{2.6}$$

As a consequence, Eqs. (2.4) and (2.6) immediately give Hitchin's inequality about Einstein manifolds $|p|/2 \leqslant \chi$.[5]

As is known the Yang–Mills functional of M is

$$YM(M) = \int_M |R|^2\Omega\,. \tag{2.7}$$

Though this paper, if we do not assume a manifold to be compact without boundary, but assume that the Yang–Mills functional is finite, then χ and p lose original topological meanings, but integral inequalities in the paper still hold. They can be considered as constraints on the Riemannian metrics.

3. EINSTEIN–KÄHLER CASE

Let M be a Kähler manifold.

Lemma 3.1: For the Kähler manifolds

$$W^+ = \begin{pmatrix} \frac{s}{6} & 0 & 0 \\ 0 & -\frac{s}{12} & 0 \\ 0 & 0 & -\frac{s}{12} \end{pmatrix}. \tag{3.1}$$

Proof: We take a Hermitian basis e_1, Je_1, e_2, Je_2,[4] where J is the canonical almost complex structure. From this Hermitian basis $\{e_1\Lambda Je_1 + e_2\Lambda Je_2, e_1\Lambda e_2 + Je_2\Lambda Je_1, e_1\Lambda Je_2 + Je_1\Lambda e_2\}$ form a basis for Λ^2_+. Using the properties about Kähler manifold,[4] we have

$$R_{e_1\Lambda e_2 + Je_2\Lambda Je_1} = 0\,, \quad R_{e_1\Lambda Je_2 + Je_1\Lambda e_2} = 0\,,$$

and

$$R_{e_1\Lambda Je_1 + e_2\Lambda Je_2} = Ric\cdot J \tag{3.2}$$

by direct computation. Thus, Eq. (3.2) gives

$$\langle R_{e_1\Lambda Je_1 + e_2\Lambda Je_2}, e_1\Lambda Je_1 + e_2\Lambda Je_2\rangle$$
$$= \langle Ric\cdot J, e_1\Lambda Je_1 + e_2\Lambda Je_2\rangle$$
$$= \langle RicJ(e_1), Je_1\rangle + \langle RicJ(e_2), Je_2\rangle = \frac{s}{2}\,,$$

$$\langle R_{e_1\Lambda Je_1 + e_2\Lambda Je_2}, e_1\Lambda e_2 + Je_2\Lambda Je_1\rangle$$
$$= \langle RicJ, e_1\Lambda Je_2 + Je_2\Lambda Je_1\rangle$$
$$= \langle RicJ(e_1), e_2\rangle + \langle RicJJ(e_2), Je_1\rangle = 0\,.$$

Similarly,

$$\langle R_{e_1\Lambda Je_1 + e_2\Lambda Je_2}, e_1\Lambda Je_2 + Je_1\Lambda e_2\rangle = 0\,.$$

So we have

$$A^+ = \begin{pmatrix} s/4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.3}$$

From Eq. (3.3) we obtain Eq. (3.1). Equations (2.5), (2.6) and (3.1) give the following:

Theorem 3.2: Let M be a four-dimensional Einstein–Kähler manifold. Then its caracteristic numbers have the following relations:

(i) $\chi(M) = \frac{1}{8\pi^2}\int_M \left(|W^-|^2 + \frac{s^2}{12}\right)\Omega\,,$

(ii) $p(M) = \frac{1}{4\pi^2}\int_M \left(\frac{s^2}{24} - |W^-|^2\right)\Omega\,,$

(iii) $\chi(M) + \frac{1}{2}p(M) = \frac{s^2 \mathrm{Vol}(M)}{2^6\pi^2}\,,$

(iv) $\chi - p = \frac{3}{8\pi^2}\int_M |W^-|^2\Omega\,,$

(v) $-2\chi \leqslant p \leqslant \chi\,,$

where $-2x = p$ holds iff $s = 0$ which is equivalent to $W^+ = 0$ or $R^{++} = 0$, and $p = x$ holds iff $W^- = 0$. So both are half-conformally flat manifolds. It is well known that the first one is the K_3 surface if it is simply connected, and the second one has constant holomorphic section curvature which we shall show later.

Remark 1: if $s \neq 0$, $\mathrm{Vol}(M) < \infty$. If $s = 0$, $\chi(M) + \frac{1}{2}p(M) = 0$ holds also

Remark 2: (V) have been obtained by H. Donnelly in a different approach[3].

If we employ the normal form of the curvature tensor for the four-dimensional Einstein manifold[6], then we can express characteristic numbers in other forms. This theorem states that there exists an orthonormal basis such that relative to the corresponding basis $\{e_1\Lambda e_2, e_1\Lambda e_3, e_1\Lambda e_4, e_3\Lambda e_4, e_4\Lambda e_2, e_2\Lambda e_3\}$ of Λ^2, the matrix of curvature components has the form

$$[R] = \begin{pmatrix} A & B \\ B & A \end{pmatrix},$$

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \text{ and } B = \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix},$$

where any 2-planes, which form the basis of Λ^2, are critical points for sectional curvature function. The relative critical values are λ_1, λ_2, and λ_3 with $\lambda_1 + \lambda_2 + \lambda_3 = s/4$. The first Bianchi identity implies $\mu_1 + \mu_2 + \mu_3 = 0$. From now on we call such an orthonormal basis of Λ^2 to be a normal basis, from which we take the basis $\{e_1\Lambda e_2 + e_3\Lambda e_4, e_1\Lambda e_3 + e_4\Lambda e_2, e_1\Lambda e_4 + e_2\Lambda e_3, e_1\Lambda e_2 - e_3\Lambda e_4, e_1\Lambda e_3 - e_4\Lambda e_2, e_1\Lambda e_4 - e_2\Lambda e_3\}$ relative to the decomposition (2.3). It has a more explicit form

$$W^+ = \begin{pmatrix} \lambda_1 + \mu_1 - \frac{s}{12} & 0 & 0 \\ 0 & \lambda_2 + \mu_2 - \frac{s}{12} & 0 \\ 0 & 0 & \lambda_3 + \mu_3 - \frac{s}{12} \end{pmatrix},$$

$$W^- = \begin{pmatrix} \lambda_1 - \mu_1 - \frac{s}{12} & 0 & 0 \\ 0 & \lambda_2 - \mu_2 - \frac{s}{12} & 0 \\ 0 & 0 & \lambda_3 - \mu_3 - \frac{s}{12} \end{pmatrix}, \tag{3.4}$$

and obviously $B = 0$.

Lemma 3.3: There exists a Hermitian basis $\{e_2, Je_1, e_2, Je_2\}$ in an Einstein–Kähler manifold such that the curvature tensor has the normal form in the following basis: $\{e_1 \wedge Je_1, e_1 \wedge e_2, e_1 \wedge Je_2, e_2 \wedge Je_2, Je_2 \wedge Je_1, Je_1 \wedge e_2\}$,

Proof: According to the fact that the 2-plane P is a critical point of R iff $RP = \lambda P + \mu P^{\perp}$,[6] if P is a critical 2 planes, and so is JP. Therefore, there is either $P = \pm JP$ or $P^{\perp} = \pm JP$, which correspond to $P \cap JP \neq \phi$ or $P \cap JP = \phi$, respectively, where $\pm$ means the same or opposite orientation of the planes. For any normal basis $P_1 = e_1 \wedge e_2$, $P_2 = e_1 \wedge e_3$, $P_3 = e_1 \wedge e_4$, $P_1^{\perp} = e_3 \wedge e_4$, $P_2^{\perp} = e_4 \wedge e_2$, $P_3^{\perp} = e_2 \wedge e_3$, it is impossible that $JP_i = \pm P_i^{\perp}$, $i = 1, 2, 3$. Otherwise, $J(P_1 \cap P_2) \subset \pm P_1^{\perp} \cap \pm P_2^{\perp} = \{e_4\}$, namely, $Je_1 = \pm e_4$. Similarly, we have $Je_1 = \pm e_2$, which is contradictory. Therefore, there exists a P_i such that $JP_i = \pm P_i$. It means P_i and $P_i^{\perp}$ are holomorphic 2-planes. (Maybe there is a need to change the orientation of the planes.) Without loss of generality, we assume them to be P_1 and $P_1^{\perp}$. We take that e_1 and Je_1 span P_1, e_2 and Je_2 span $P_1^{\perp}$. So we obtain such a Hermitian basis that $e_1 \wedge Je_1$, $e_1 \wedge e_2$, $e_2 \wedge Je_2$, $Je_2 \wedge Je_1$ are critical 2-planes. By the symmetry of the Riemmanian curvature tensor $e_1 \wedge Je_2$ and $Je_1 \wedge e_2$ are critical 2-planes as well.

As a result, in any Einstein–Kähler manifold one can choose a Hermitian basis $\{e_1, Je_1, e_2, Je_2\}$ with critical 2-planes of the section curvature function $e_1 \wedge Je_1$, $e_2 \wedge Je_2$, ... Equations (3.1) and (3.4) give

$$|W^-|^2 = \left(2\lambda_1 - \frac{s}{3}\right)^2 + \left(2\lambda_2 - \frac{s}{12}\right)^2 + \left(2\lambda_3 - \frac{s}{12}\right)^2. \tag{3.5}$$

So we obtain the result as follows:

Theorem 3.4: Let M be a four-dimensional Einstein–Kähler manifold, then we have

$$\chi(M) = \frac{1}{8\pi^2}\int_M \left\{\left(2\lambda_1 - \frac{s}{3}\right)^2 + \left(2\lambda_2 - \frac{s}{12}\right)^2 + \left(2\lambda_3 - \frac{s}{12}\right)^2 + \frac{s^2}{12}\right)\Omega,$$

$$p(M) = \frac{1}{4\pi^2}\int_M \left\{\frac{s^2}{24} - \left(2\lambda_1 - \frac{s}{3}\right)^2 + \left(2\lambda_2 - \frac{s}{12}\right)^2 + \left(2\lambda_3 - \frac{s}{12}\right)\right\}\Omega, \tag{3.6}$$

where s is the scalar curvature in M; Ω is the volume form of M; $\lambda_1, \lambda_2, \lambda_3$ are critical values of section curvature functions and λ_1 is the holomorphic sections curvature also.

Remark 1: Using the relations between section curvatures and holomorphic section curvatures, we can evaluate the upper bound of the Euler number and the lower bound of the Pontrjagin number, if we give bounds to holomorphic section curvatures.

Remark 2: $W^- = 0$ iff $\lambda_1 = s/6$, $\lambda_2 = s/24$, $\lambda_3 = s/24$.

If we notice Eqs. (3.1) and (3.4) again, we have $\mu_1 = s/12$, $\mu_2 = -s/24$, $\mu_3 = -s/24$. Consequently, its holomorphic sections curvatures equal a constant $s/6$, which are already obtained by Gu Chaohao and Hu Hesheng.[7]

4. HALF-CONFORMALLY FLAT EINSTEIN CASE

Recall that conditions for half-conformally flat Einstein are $W^- = 0$ (or $W^+ = 0$) and $R^{+-} = 0$. We have the following results:

Theorem 4.1: Let M be a four-dimensional half-conformally flat Einstein manifold. Its characteristic numbers have relations

(i) $$\chi(M) = \frac{1}{8\pi^2}\int_M \left(|W^+|^2 + \frac{s^2}{24}\right)\Omega \times \left(\text{or } \chi(M) = \frac{1}{8\pi^2}\int_M \left(|W^-|^2 + \frac{s^2}{24}\right)\Omega\right)$$

(ii) $$p(M) = \frac{1}{4\pi^2}\int_M |W^+|^2\Omega \left(\text{or } p(M) = -\frac{1}{4\pi^2}\int_M |W^-|^2\Omega\right)$$

(iii) $$\chi(M) = \frac{|p(M)|}{2} + \frac{s^2 \text{Vol}(M)}{2^6 \cdot 3\pi^2}.$$

Proof: (i) and (ii) can be obtained from Eqs. (2.5), (2.6), and the conditions for half-conformally flat Einstein manifolds. (iii) follows from (i) and (ii).

If we name the half-conformally flat Einstein manifolds with $S \neq 0$ and $W \neq 0$ as nontrivial half-conformally flat Einstein manifolds, we have the following corollary.

Corollary 4.2: There are no nontrivial half-conformally flat Einstein metrics with the global topology of S^4, $S^2 \times S^2$, and K_3.

Proof: According to the Theorem 4.1, $S \neq 0$ is equivalent to $X > |p|/2$ and $W \neq 0$ is equivalent to $p(M) \neq 0$ from the facts that $p(s^4) = p(s^2 \times s^2) = 0$, $p(K_3) = -48$ and $X(K_3) = 24$; hence, we obtain the conclusion.

If we employ the normal form of curvature tensor for a four-dimensional Einstein manifold, we have another expression for characteristic numbers. $W^- = 0$ gives $\lambda_i - \mu_i = s/12$, so we have

$$|W^+|^2 = \left(2\lambda_1 - \frac{s}{6}\right)^2 + \left(2\lambda_2 - \frac{s}{6}\right)^2 + \left(2\lambda_3 - \frac{s}{6}\right)^2. \tag{4.1}$$

Substituting Eq. (4.1) into expressions for the Euler number and the Pontrjagin number, we get

$$\chi(M) = \frac{1}{8\pi^2}\int_M \left[\left(2\lambda_1 - \frac{s}{6}\right)^2 + \left(2\lambda_2 - \frac{s}{6}\right)^2 + \left(2\lambda_3 - \frac{s}{6}\right)^2 + \frac{s^2}{24}\right]\Omega,$$

$$p(M) = \frac{1}{4\pi^2}\int_M \left[\left(2\lambda_1 - \frac{s}{6}\right)^2 + \left(2\lambda_2 - \frac{s}{6}\right)^2 + \left(2\lambda_3 - \frac{s}{6}\right)^2\right]\Omega. \tag{4.2}$$

If we assume that $\lambda_i \lambda_j \geqslant 0$, $ij = 1, 2, 3$, we obtain then

$$0 \leqslant |W^+|^2 \leqslant \frac{s^2}{6}. \tag{4.3}$$

Theorem 4.1 and inequality (4.3) give

$$\frac{s^3 \text{vol}(M)}{2^6 \cdot 3\pi^2} \leqslant \chi(M) \leqslant \frac{5s^2 \text{vol}(M)}{2^6 \cdot 3\pi^2} \tag{4.4}$$

and

$$0 \leqslant p(M) \leqslant \frac{s^2 \mathrm{vol}(M)}{2^3 \cdot 3\pi^2} \tag{4.5}$$

so we have

$$0 \leqslant \frac{p(M)}{2^3} \leqslant \frac{s^2 \mathrm{vol}(M)}{2^6 \cdot 3\pi^2} \leqslant \chi(M) \leqslant \frac{5 \cdot s^2 \mathrm{vol}(M)}{2^6 \cdot 3\pi^2}.$$

For $W^+ = 0$ we can obtain similar results. We summarize the results as follows:

$$0 \leqslant \frac{|p(M)|}{2^3} \leqslant \frac{s^2 \mathrm{vol}(M)}{2^6 \cdot 3\pi^2} \leqslant \chi(M) \leqslant \frac{5 \cdot s^2 \mathrm{vol}(M)}{2^6 \cdot 3\pi^2}. \tag{4.6}$$

Theorem 4.3: Let M be a four-dimensional half-conformally flat Einstein manifold with nonnegative (or nonpositive) section curvatures. Then its characteristic numbers, scalar curvature, and volume satisfy inequality (4.6).

For CP_2 with the Fubini-study metric, which is a nontrivial half-conformally flat Einstein manifold, we have $p = 3$, $x = 3$, $s = 24$, $\mathrm{Vol}(cp_2) = \pi^2/2$, which obviously satisfy inequality (4.6). It is not known whether there exist any other nontrivial half-conformally flat Einstein manifolds except the Einstein–Kähler manifolds with constants holomorphic sectional curvatures.

ACKNOWLEDGMENTS

The author would like to thank Professor C.N. Yang for his encouragement and guidance and also to thank Professor Gu Chauhao for his advise and help. He is grateful to the State University of New York at Stony Brook for its hospitality.

[1] C.H. Gu, H.S. Hu, D.Q. Li, C.L. Shen, Y.L. Xin, and C.N. Yang, Sci. Sin. **21** 475 (1978).
[2] C.N. Yang (private communication).
[3] H. Donnelly, J. Diff. Geom. **11**, 259 (1976).
[4] S. Kobayashi, and K. Nomizu, Found. Diff. Geom. **2**, (1969).
[5] N. Hitchin, J. Diff. Geom. **9**, 435 (1974).
[6] I.M. Singer, and J.A. Thorpe, The curvature of 4-dimensional Einstein spaces, *Global Analysis*, Paper in honor of K. Kodaira (Princeton University, Princeton, 1969), pp. 355-365.
[7] C.H. Gu, and H.S. Hu (private communication).

Equilibrium of slowly rotating relativistic fluids[a)]

H. P. Künzle
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

J.R. Savage
Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2J1

(Received 1 August 1978; accepted for publication 8 June 1979)

The equilibrium conditions of a relativistic fluid with nonzero viscosity and heat conduction coefficients are known to reduce to Einstein's equations for a barotropic perfect fluid in rigid motion. We consider here the linearization of these equations on a static spherically symmetric background and show that the solution space is three-dimensional (parametrized by the angular velocity vector, for example), provided the exterior vacuum region is asymptotically Euclidean and the equation of state $\rho = \rho(p)$ (satisfying $\rho \geq p \geq 0$ and $dp/d\rho > 0$) is fixed, as well as the central values of the pressure and the gravitational potential. In the exterior region this solution agrees with the Kerr solution, linearized on the Schwarzschild background. This result is the first step towards proving a certain uniqueness of the possible equilibrium configurations of slowly rotating relativistic fluids. It is obtained using invariantly defined global conditions, without assuming the existence of particular coordinate systems.

1. INTRODUCTION

We begin in this paper a mathematical analysis of the set of possible equilibrium configurations of a slowly rotating (simple, thermodynamic) relativistic fluid body surrounded by vacuum.

Slowly rotating relativistic stars have been extensively discussed during the last decade, but mainly from an astrophysical point of view.[1] Our objective is to extend the classical mathematical theory of the equilibrium configurations of nonrelativistic self-gravitating fluids[2] to Einstein's theory of general relativity. The main tool in the Newtonian case is the classical potential theory which does not generalize easily to Riemannian manifolds. On the other hand, for the stationary vacuum equations in general relativity, together with the boundary conditions of asymptotical flatness and regularity at a horizon, one has the uniqueness theorems for black holes of Israel, Carter, Hawking, and Robinson.[3] The situation is considerably more complicated if sources are present. Nevertheless, if the equation of state is given and fixed and the field equations are supplemented by the conditions for rigid motions, one obtains a system of equations whose linearization may be expected to be elliptic, once the coordinate freedom is factored out. The set of solutions of these linearized equations together with appropriate boundary conditions will then be finite-dimensional.

Of course, this reasoning is very heuristic and only a very general guideline. Proving that there exists a solution to the linearized system does not guarantee the existence of a solution to the nonlinear equations (it could be linearization instable[4]). Nor does uniqueness of the solution to the linearized system imply uniqueness for the nonlinear case (there could be bifurcation points). However, the general relativistic case is less degenerate than the Newtonian one, so there is more hope of proving that there exists a finite-dimensional family of slowly rotating configurations with a given equation of state. (In the Newtonian case elliptic configurations with constant density have received by far the greatest attention,[5] but both these special assumptions do not make much sense in general relativity.) For static (nonrotating) cases the result is intuitively obvious: The equilibrium configuration must be spherically symmetric. But while a mathematical proof exists for the Newtonian case,[6] it is not yet complete for the general relativistic case.[7-9]

In this paper our aim is quite modest. We consider the linearized equations on the static spherically symmetric background, where they simplify considerably. This allows us to show that for a fixed equation of state and fixed central values of the gravitational potential and pressure the space of solutions is three-dimensional and can be thought of as parametrized by the angular velocity vector. This result is obtained by means of elementary estimates very similar to those used in Ref. 8.

In the rest of this section we will make our assumptions more precise and outline the method used. We consider a simple relativistic fluid with viscosity and heat conduction, i.e., we assume that there exists a timelike vector field u^α and a matter (baryon number) density n satisfying[10]

$$\nabla_\mu (n u^\mu) = 0, \tag{1.1}$$

as well as a specific thermal energy ϵ, specific entropy s, thermodynamic pressure p, and temperature $T(>0)$ related by

$$d\epsilon = T ds + p n^{-2} dn. \tag{1.2}$$

The stress–energy tensor is assumed to be of the form

$$T^{\alpha\beta} = \rho u^\alpha u^\beta + 2u^{(\alpha} q^{\beta)} + (p - \zeta\theta)(g^{\alpha\beta} + u^\alpha u^\beta) - \pi^{\alpha\beta}, \tag{1.3}$$

where $\rho := n(1+\epsilon)$ and $\theta := \nabla_\mu u^\mu$ (expansion rate), and where the heat flow vector q^α and the shear stress tensor $\pi^{\alpha\beta}$ are orthogonal to u^α and of the form

a)Partially supported by the National Sciences and Engineering Research Council, Grant No. A8059.

 0022-2488/80/020347-08$1.00

$$\pi_{\alpha\beta} = 2\eta\sigma_{\alpha\beta}, \tag{1.4}$$

$$q^{\alpha} = -\kappa(g^{\alpha\mu} + u^{\alpha}u^{\mu})(\partial_{\mu}T + T\dot{u}_{\mu})(\cdot := u^{\mu}\nabla_{\mu}). \tag{1.5}$$

Here $\sigma_{\alpha\beta} := \nabla_{(\alpha}u_{\beta)} - \frac{1}{3}\theta(g_{\alpha\beta} + u_{\alpha}u_{\beta}) + \dot{u}_{(\alpha}u_{\beta)}$ is the shear deformation rate. Thermal equilibrium is then defined by the condition that

$$\nabla_{\mu}s^{\mu} = 0 \tag{1.6}$$

for the entropy flux vector $s^{\mu} := nsu^{\mu} + T^{-1}q^{\mu}$.

Now Eqs. (1.1)–(1.6), together with the conservation law

$$\nabla_{\mu}T^{\mu}_{\alpha} = 0, \tag{1.7}$$

and the assumptions that the shear and bulk viscosities η and ζ as well as the heat conductivity κ are all strictly positive, are known[11] to imply that

(i) the Lie derivative with respect to $\theta^{\alpha} := T^{-1}u^{\alpha}$ of the metric $g_{\alpha\beta}$ and all thermodynamical quantities vanish, in particular the motion of the fluid is rigid ($\sigma_{\alpha\beta} = 0 = \theta$);

(ii) the heat flux q^{α} vanishes;

(iii) the equation of hydrodynamical equilibrium (momentum balance) is equivalent to

$$Tdp = (\rho + p)dT. \tag{1.8}$$

From (1.2) and (1.8) it follows, moreover, that the fluid is barotropic [with an equation of state of the form $\rho = \rho(p)$, say]. Properties (i) and (ii) of a fluid in equilibrium can also be derived from relativistic kinetic theory.[12]

By (i) the space–time region occupied by the fluid is stationary. It would seem reasonable that for an isolated equilibrium fluid in empty space the space–time region outside the material must also be stationary. But this is not easy to prove rigorously and the result would depend crucially on the asymptotic conditions imposed. On the other hand, if one assumes the space–time to be stationary *a priori* (with timelike Killing vector field ξ^{α}, say) then the matter region is either *static* (θ^{α} is hypersurface orthogonal and proportional to ξ^{α}) or θ^{α} and ξ^{α} are non parallel commuting Killing vector fields, which means that this region of space–time is *axisymmetric*. Lindblom[13] has shown that one can extend θ^{α} to a Killing vector field in the vacuum region near the boundary of the fluid and can thus argue that, with reasonable asymptotic conditions, the whole space–time must be axisymmetric. We will not assume global axisymmetry, however, since it does not lead to any major simplification in our approach.

We therefore consider an equation of state $\rho = \rho(p)$ given and propose to study the set $\mathcal{S}$ of all globally stationary and asymptotically Euclidean space–times supporting a perfect fluid, $T_{\alpha\beta} = (\rho + p)u_{\alpha}u_{\beta} + pg_{\alpha\beta}$, in a spatially compact region B such that the scalar T, defined on B by (1.8) (up to a constant) makes $\theta^{\alpha} = T^{-1}u^{\alpha}$ into a Killing vector field. The conjecture is that this set can be described by a finite number of parameters (at least locally, i.e., in a neighborhood of the spherically symmetric static solutions).

The paper is organized as follows. In Sec. 2 we define globally stationary space–times and then write the field equations and equilibrium conditions in terms of the three-dimensional quotient geometry consisting of a Riemannian metric γ_{ij}, a vector field h^{i} (related to the angular momentum), and a scalar field U (the gravitational potential). This formalism goes back to Lichnerowicz[14] and our version is closest to the one of Ehlers.[15] The vector field h^{i} is divergence free and its curl is a kind of mass current so that it behaves in most respects like a magnetic field. Our derivation of these three-dimensional equations uses the global theory of connections,[16] but this argument need not be followed in detail for an understanding of the remainder of the paper.

In the static case (characterized by $h^{i} \equiv 0 \Longleftrightarrow \theta^{i} \equiv 0$) it follows that all thermodynamical quantities are functions of U only, in particular the pressure, so that the surface of the fluid coincides with an equipotential surface that will be topologically spherical. For configurations close to spherically symmetric ones it is then very convenient to study the family of two-dimensional geometries of these equipotential surfaces since one can use mathematical results that hold only for compact manifolds. We choose to use this approach also in the stationary case, although it is much less convenient since p is no longer a function of U only. Using another global assumption (namely that there exists a θ-invariant 1-form α such that $\partial_{[i}\alpha_{j]} = \epsilon_{ijk}h^{k}$) one can construct a modified potential U_{eff} on whose level surfaces p will be constant. But this effective potential satisfies much more complicated equations and can only be extended into the vacuum region if global axisymmetry is assumed *a priori*. Moreover, its level surfaces near infinity degenerate into cylinders, rather than remaining topological spheres.

In Sec. 3 we translate the field equations and equilibrium conditions into a set of equations for quantities defined on a family of 2-manifolds, parametrized by U. (We assume that U has only one critical point at the center so that all equipotential surfaces are diffeomorphic to 2-spheres.[7,8])

These equations are linearized on the spherical background in Sec. 4, where we also establish what constitutes a physically nontrivial deformation, namely one not due to coordinate transformations.

In Sec. 5 we show that the solution space of this linear system is three-dimensional for an arbitrary equation of state $\rho = \rho(p)$ with $\rho \geqslant p \geqslant 0$ (Hawking's dominant energy condition[17]) and $dp/d\rho > 0$ everywhere. This is achieved by exactly the same method as in Ref. 8 since on the spherical background the stationary case reduces to a large extent to the static one. We find, not surprisingly, that the only quantity that does not vanish to first order on the spherical background is h^{i}. Moreover, δh^{i} in the vacuum region is exactly of the same form that one obtains by linearizing the Kerr solution on the Schwarzschild background.

The equations for δh^{i} can be solved numerically quite easily for any explicitly known static and spherically symmetric background solution. As an example we give in Sec. 6 the results for the case of the interior Schwarzschild and one of the Tolman[18] solutions.

2. GLOBALLY STATIONARY SPACE–TIMES

In the entire paper we assume for simplicity that all manifolds and tensor fields are C^{∞} except at the (C^{∞}) boundary of the region occupied by matter. We shall call such tensor fields regular. Our other basic assumption is that

the space–time (M,g) is *globally stationary*, i.e., that there exists a one-dimensional isometry group whose orbits are timelike submanifolds diffeomorphic to $\mathbb{R}$ and whose generator is therefore an everywhere timelike vector field ξ tangent to these orbits. Let

$$e^{2U} = -g(\xi,\xi) \tag{2.1}$$

so that $\xi(U) = 0$ and U is a regular function on M.

We assume that the set of all orbits Σ has a manifold structure such that the canonical projection $\pi: M \to \Sigma$ is C^∞. This is the case iff for all $x_0 \in M$ there exists a neighborhood $V \ni x_0$ that is intersected by each orbit in at most one connected segment.[19] Thus M is a principal fibre bundle over Σ with structure group $(\mathbb{R}, +)$.

We orthogonally decompose $T_x M$ into a one-dimensional vertical subspace V_x parallel to ξ and a complementary horizontal subspace H_x by writing $\eta \in T_x M$ as $\eta = \eta^0 \xi + \tilde{\eta}$ with $\eta^0 = e^{-2U} g(\eta,\xi)$. This decomposition defines a connection since it is invariant under the group action of $(\mathbb{R}, +)$.[16] It can be equivalently characterized by a connection form, i.e., in this case a real valued 1-form α that satisfies

$$\text{(i) } \mathcal{L}_\xi \alpha = 0, \quad \text{(ii) } \alpha(X_x) = 0 \Longleftrightarrow X_x \in H_x, \quad \text{(iii) } \alpha(\xi) = 1. \tag{2.2}$$

These equations define α uniquely if the horizontal subspace is given. The curvature form H associated to the connection form α is then $H = d\alpha + \alpha \wedge \alpha = d\alpha$ and $H = \pi^* \tilde{H}$ for a unique 2-form $\tilde{H}$ on Σ. Similarly we have $U = \pi^* \tilde{U}$.

We define a Riemannian metric $\tilde{\gamma}$ on Σ by

$$\tilde{\gamma}(\tilde{\eta},\tilde{\xi})|_{\tilde{x}} = e^{2\tilde{U}} g(\eta,\xi)|_x \tag{2.3}$$

for any $x \in \pi^{-1}(\tilde{x})$, where $\eta,\xi \in T_x M$ are the unique horizontal lifts of $\tilde{\eta},\tilde{\xi} \in T_{\tilde{x}} \Sigma$, respectively. Then we find

$$g = -\pi^*(e^{2\tilde{U}})\alpha \otimes \alpha + \pi^*(e^{-2\tilde{U}}\tilde{\gamma}). \tag{2.4}$$

We now drop the tilde and consider only the three–geometries (Σ,γ,α,U) assumed globally defined and C^∞, except at the star boundary, with γ a positive definite Riemannian metric, α a smooth connection form on an $\mathbb{R}$-principal bundle over Σ, and $U \in C^\infty(\Sigma,\mathbb{R})$. Note that there is no reason for the closed 2-form H to be exact.

On a local neighborhood of $x \in M$ there exist coordinates $(t,x^i) = x^\alpha$ such that $\xi = \partial_t$, i.e., $\mathcal{L}_\xi g = 0 \Longleftrightarrow \partial_t g_{\alpha\beta} = 0$. Letting

$$g_{00} = -e^{2U}, \quad g_{0k} = -e^{2u} a_k, \quad g_{kl} = e^{-2U}\gamma_{kl} - e^{2U} a_k a_l, \tag{2.5}$$

we have that

$$\alpha = dt + a_k dx^k, \quad \gamma = \gamma_{kl} dx^k \otimes dx^l, \quad H_{ij} := 2\partial_{[i} a_{j]}. \tag{2.6}$$

We now use such local adapted charts to derive the three-geometry formulas. First we observe that for any vector field η such that $\mathcal{L}_\xi \eta = 0$, $\eta^0 = g(\xi,\eta)$ and $\eta^i \partial_i$ are a scalar and vector field, respectively, on Σ. By writing all the tensorial equations with the 0 index lowered and the Latin indices raised, the a_k does not appear except through H_{ij}. We use γ_{ij} to lower and raise all Latin indices.

Einstein's equations for a perfect fluid,

$${}^4R_{\alpha\beta} - (\tfrac{1}{2})\,{}^4R g_{\alpha\beta} = T_{\alpha\beta} = (\rho + p)T^2\theta_\alpha\theta_\beta + p g_{\alpha\beta}, \tag{2.7}$$

then become in the three-dimensional formulation

$$\Delta U := \gamma^{ij}\nabla_i \partial_j U = M - \tfrac{1}{2}e^{4U}h^2 + (\rho + p)e^{-4U}T^2\theta^2, \tag{2.8}$$

$$\epsilon^{ijk}\nabla_j(e^{4U}h_k) = 2(\rho + p)e^{-2U} v T \theta^i, \tag{2.9}$$

$${}^3R^{ij} = 2\gamma^{ir}\partial_r U \gamma^{js}\partial_s U + \tfrac{1}{2}e^{4U}h^i h^j + (\rho + p)e^{-4U}T^2\theta^i\theta^j - [2pe^{-2U} + (\rho + p)e^{-4U}T^2\theta^2]\gamma^{ij}, \tag{2.10}$$

where $h^i = \frac{1}{2}\epsilon^{ijk}H_{jk}$, $h^2 = \gamma_{ij}h^i h^j$, $\theta^2 = \gamma_{ij}\theta^i\theta^j$, $M = \frac{1}{2}(\rho + 3p)e^{-2U}$, and $v = -u_0 > 0$ so that $v^2 = e^{2U} + T^2\theta^2$. On the other hand θ^α is a symmetry of the space–time metric and all thermodynamic quantities iff (for $\theta = \theta^i\partial_i$)

$$\mathcal{L}_\theta U = 0, \quad \mathcal{L}_\theta \gamma = 0, \quad \mathcal{L}_\theta p = \mathcal{L}_\theta \rho = \mathcal{L}_\theta T = 0, \tag{2.11}$$

and

$$\theta \lrcorner H + d(vT^{-1}e^{-2U}) = 0, \tag{2.12}$$

which implies that

$$\mathcal{L}_\theta H = d(\theta \lrcorner H) + \theta \lrcorner dH = 0 \quad \text{whence } \mathcal{L}_\theta h = 0. \tag{2.13}$$

Since H is closed we also have

$$\nabla_r h^r = 0. \tag{2.14}$$

It can be shown that (M,g) is static iff $H \equiv 0$. Then (2.9) implies $\theta^i = 0$ and the field equations reduce to those in Ref. 8.

We assumed M to be C^∞ and the metric to be regular, i.e., C^∞ except at the star boundary where we now require it to be C^1 and piecewise C^3 (i.e., second and third derivatives have finite limits on both sides of the hypersurface.)[14] We also assume ξ to be regular and C^0 at the boundary. Together with the Killing equations this implies it is C^1 at the boundary. Hence U and γ_{ij} are C^1 and piecewise C^3; h^i and $\partial_i U$ are C^0 and piecewise C^2 while ${}^3R_{ij}$ may have a jump singularity. As in Ref. 8 we assume the space–time to be *asymptotically Euclidean* in the sense of Lichnerowicz.[14] In terms of the 3-geometry this means

(i) there exists a compact $K \subset \Sigma$ and a diffeomorphism $\phi: \Sigma \setminus K \to \mathbb{R}^3 \setminus B$, where B is a closed ball centered at the origin;

(ii) with respect to the standard coordinate system in $\mathbb{R}^3$

$$\gamma_{ij} = \delta_{ij} + O(|x|^{-1}), \quad \partial_k \gamma_{ij} = O(|x|^{-2}),$$
$$U = O(|x|^{-1}), \quad \partial_k U = O(|x|^{-2}), \tag{2.15}$$
$$h^i = O(|x|^{-2}), \quad \text{where } |x|^2 = \sum_{i=1}^{3} (x^i)^2.$$

3. (2 + 1)-DIMENSIONAL FORMALISM

In this section we rewrite our equations in terms of the two-dimensional geometry of the equipotential surfaces (which are all Riemannian 2-spheres since we assume that the space–time is nearly spherically symmetric). Let S be an abstract 2-sphere and $i_c: S \to \Sigma$ be the imbedding map of S into Σ such that $i_c(S) = U^{-1}(c) =: S_c$ for any $c \in (U_{\min},0)$. Then any Σ 1-form is characterized by $\bar{\omega} := i^*\omega$ and

$$\omega^0 := i^*(\nabla U \lrcorner \omega). \tag{3.1}$$

This decomposition extends to any Σ-tensor field since Σ is Riemannian. An intrinsically defined normal derivative to the S_c surfaces is given by

$$D\bar{\omega} = i^*(|\nabla U|^{-2}\nabla_{\nabla U}\omega) \tag{3.2}$$

while the second fundamental form of S_c is[16]

$$\Omega = i^*(\nabla(|\nabla U|^{-1}\nabla U)). \tag{3.3}$$

To find the local coordinate expression let $(\bar{x}^A, A = 1,2)$ be a chart of S and $i_c : \bar{x}^A \to (U = c, x^A = \bar{x}^A(c))$ [using $(U,\bar{x}^A)$ as a chart of Σ]. Then for a 1-form

$$\bar{\omega}_A = \frac{\partial U}{\partial \bar{x}^A}\omega_1 + \frac{\partial x^B}{\partial \bar{x}^A}\omega_B = \omega_A,$$
$$\omega^0 = i^*(\nabla U \lrcorner \omega) = i^*(\omega_i \gamma^{ij}\partial_j U) = \omega^1. \tag{3.4}$$

In particular

$$\gamma_{AB} = \bar{\gamma}_{AB}, \quad \gamma^{00} = \gamma^{ij}\partial_i U \partial_j U =: W^2,$$
$$\gamma^0{}_A = \gamma^{ir}\partial_r U \gamma_{iA} = \partial_A U = 0, \tag{3.5}$$

and $\bar{\gamma}_{AB}$ is used to raise and lower all indices of S-tensors. Since U has only one (nondegenerate) critical point at the center, ∇U is a vector field that vanishes nowhere else and is orthogonal to the surfaces $U^{-1}(c)$. It can thus be used to construct a global polar type coordinate system $(U,\bar{x}^A)$ on Σ such that

$$(\gamma_{ij}) = \begin{pmatrix} W^{-2} & 0 \\ 0 & \bar{\gamma}_{AB} \end{pmatrix}. \tag{3.6}$$

Regarding an S-tensor $\bar{K}$ as a function of U we then see that

$$D\bar{K} = \frac{\partial \bar{K}}{\partial U} \tag{3.7}$$

and, in particular,

$$D\bar{\gamma}_{AB} = 2W^{-1}\Omega_{AB}, \tag{3.8}$$

where Ω_{AB} are the components of the second fundamental form Ω.

After some work it is found that the Einstein equations take the form

$$DW = -\Omega + MW^{-1} - \tfrac{1}{2}e^{4U}(W^{-3}h^{0^2} + W^{-1}h^{-2}) + (\rho + p)W^{-1}e^{-4U}T^2\theta^2, \tag{3.9}$$

$$\partial_{[A}\bar{h}_{B]} = 0, \tag{3.10}$$

$$D\bar{h}_A = \partial_A(h^0 W^{-2}) - 4\bar{h}_A + 2(\rho + p)W^{-1} \times e^{-6U}vT\bar{\epsilon}_{AB}\bar{\theta}^B, \tag{3.11}$$

$$D\Omega_{AB} = 2W^{-1}\Omega_{AC}\Omega^C_B - W^{-1}\Omega\Omega_{AB} - 2W^{-3}\partial_A W \partial_B W + W^{-2}\bar{\nabla}_A\partial_B W + \tfrac{1}{2}W^{-1}\bar{R}\bar{\gamma}_{AB} - \tfrac{1}{2}W^{-1}e^{4U}\bar{h}_A\bar{h}_B - (\rho + p)W^{-1}e^{-4U}T^2(\bar{\theta}_A\bar{\theta}_B - \bar{\theta}^2\bar{\gamma}_{AB}) + 2pW^{-1}e^{-2U}\bar{\gamma}_{AB}, \tag{3.12}$$

$$\bar{\nabla}_B\Omega^B_A - \partial_A\Omega = \tfrac{1}{2}W^{-1}e^{4U}h^0\bar{h}_A, \tag{3.13}$$

$$\bar{R} - \Omega^2 + \Omega_{AB}\Omega^{AB} + 2W^2 - \tfrac{1}{4}e^{4U}(\bar{h}^2 - W^{-2}h^{0^2}) = -2pe^{-2U}, \tag{3.14}$$

where $\bar{h}^2 := \bar{\gamma}^{AB}\bar{h}_A\bar{h}_B$, $\theta^2 := \bar{\gamma}^{AB}\bar{\theta}_A\bar{\theta}_B$, $\Omega := \bar{\gamma}^{AB}\Omega_{AB}$, $\bar{\epsilon}_{AB} := W\epsilon_{1AB} = \bar{\gamma}^{1/2}\delta^{12}_{AB}$, $\bar{R}_{AB} = \tfrac{1}{2}\bar{R}\bar{\gamma}_{AB}$, and $\bar{\nabla}$ is the covariant derivative with respect to the connection defined by $\bar{\gamma}_{AB}$. Using $\theta^0 = \theta^i\partial_i U = 0$, Eq. (2.11) becomes

$$\mathcal{L}_{\bar{\theta}}W = 0, \tag{3.15}$$

$$D\bar{\theta}^A = 0, \quad \bar{\nabla}_{(A}\bar{\theta}_{B)} = 0, \tag{3.16}$$

while Eqs. (2.14) and (2.12) yield, respectively,

$$Dh^0 - W^{-1}h^0 DW - W^{-1}\bar{h}^A\partial_A W + \bar{\nabla}_A\bar{h}^A + W^{-1}\Omega h^0 = 0 \tag{3.17}$$

and

$$D(vT^{-1}e^{-2U}) = W^{-1}\bar{\epsilon}_{AB}\bar{\theta}^A\bar{h}^B, \quad \partial_A(vT^{-1}e^{-2U}) = W^{-1}h^0\bar{\epsilon}_{AB}\theta^B \tag{3.18}$$

and Eq. (2.13) translates into

$$\mathcal{L}_{\bar{\theta}}h^0 = 0, \quad \mathcal{L}_{\bar{\theta}}\bar{h}_A = 0, \tag{3.19}$$

Note that, in fact, the Lie derivative with respect to $\bar{\theta}^A$ of any S tensor vanishes. We can now regard a stationary, rigid, perfect fluid space–time with a fixed equation of state and a fixed central potential $U_0 = \min(U|x\in\Sigma)$ as characterized by the set $\mathcal{G} = \{\bar{\gamma},\Omega,W,\bar{h},h^0,p,\bar{\theta},T\}$ where all these tensors are functions of $U\in(U_0,0)$ and where we drop the bars from now on when there is no danger of confusion.

Since U is only a C^1 function on Σ the junction conditions must be derived with some care in this formalism. Let z be a coordinate such that $z^{-1}(0)$ defines the surface of discontinuity, the star boundary, and use z and x^A as C^∞ coordinates. Let

$$[K] = \lim_{z\to 0^+} K - \lim_{z\to 0^-} K \tag{3.20}$$

for any Σ tensor or S tensor K. The first discontinuity in derivatives of U is $[U_{zz}] =: J$ (while $[U_{zA}] = 0$). Translating the junction conditions of Sec. 2 to the coordinates $\{U = U(z,x^A), \bar{x}^A = x^A\}$, using $\partial_z\phi = (\partial\phi/\partial U)U_z =: \phi' U_z$, $\partial_A\phi = \partial_{\bar{A}}\phi + \phi' U_A$, where ϕ is any scalar, and noting that $U_A = -(\partial z/\partial\bar{x}^A)U_z$, we obtain

$$[DW] = JW(Dz)^2 \tag{3.21}$$

$$[\Omega_{AB}] = JU_z^{-2}W^{-1}U_A U_B = JW^{-1}\left(\frac{\partial z}{\partial \bar{x}^A}\right)\left(\frac{\partial z}{\partial \bar{x}^B}\right). \tag{3.22}$$

4. LINEARIZATION ON THE SPHERICALLY SYMMETRIC BACKGROUND

Consider now a one-parameter family $\mathcal{G}(\lambda)$ where $\mathcal{G}(0)$ is the spherically symmetric solution and let $\delta f = [df(\lambda)/d\lambda]|_{\lambda=0}$ be the variation of any $f\in\mathcal{G}$ on the background. Specifically, we write $\delta\gamma_{AB} = C_{AB}$, $\delta W = w$, $\delta h^0 = K$, $\delta h_A = k_A$, $\delta\theta^A = \theta^A$, $\delta T = \tau$. Now, on the compact Riemannian manifold S there is a decomposition

$$C_{AB} = \Phi_{AB} + \mathcal{L}_\xi\gamma_{AB} \tag{4.1}$$

such that

$$\nabla_B\Phi^B_A = 0, \tag{4.2}$$

where Φ is unique and ξ is unique up to a Killing vector field of γ.[21] Here the Φ_{AB} measures a physical change of the geometry while ξ represents an infinitesimal coordinate change. Therefore, if we select any such ξ and define $\hat{\delta}f = df - \mathcal{L}_\xi f$ for any $f\in\mathcal{G}$, then the change of $\mathcal{G}$ is due to a mere coordinate transformation iff $\hat{\delta}f = \mathcal{L}_\eta f$ where η satisfies $\mathcal{L}_\eta\gamma = 0$. Since the Lie derivatives of all the equations for $\mathcal{G}$ simply give the linearized equations for variations of the form $\mathcal{L}_\xi f$ we now drop the caret and compute the variations of (3.9)–(3.19) assuming that

$$\delta\gamma_{AB} = \Phi_{AB}, \quad \nabla_B \Phi^B_A = 0. \tag{4.3}$$

A Riemannian 2-sphere S has no closed two-dimensional subgroup of its isometry group I, and is isometric to the Euclidean 2-sphere if $\dim I = 3$.[22] Therefore, in the case $\dim I < 3$, we have inside matter $\eta^A = c\theta^A$ for some constant $c \in \mathbb{R}$. But all the quantities are invariant under θ^A so all variations vanish under a coordinate transformation. Outside matter θ^A and τ are no longer defined, but we expect axisymmetry,[13] so it is likely that all variations will vanish under a coordinate transformation. If S is isometric to the Euclidean 2-sphere (for all U), all quantities become functions of U only, $\theta^A = h^A = 0$, and all variations will vanish under a coordinate transformation. On a spherical background we have $\theta^A = h^A = h^0 = 0$ and as shown in Ref. 8, $\Omega_{AB} = \frac{1}{2}\Omega\gamma_{AB}$. Linearization of (3.18) gives $D\tau + \tau = 0 = \partial_A \tau$ whence $\tau = \tau_0 e^{U_0 - U}$ which vanishes since $\delta p = (\rho + p)T^{-1}\tau$ and we keep the central pressure fixed under the variation.[23] Using these facts we now find for the linearized Einstein equations

$$(MW^{-1} + \Omega)w + WDw + \tfrac{1}{2}W^2 D\Phi = 0, \tag{4.4}$$

$$\Omega W^{-2}\partial_A w + \partial_A D\Phi = 0, \tag{4.5}$$

$$(\bar{\Delta} + \tfrac{1}{2}\bar{R})\Phi + \tfrac{1}{2}W\Omega D\Phi = -2W^{-1}(\bar{R} + 2pe^{-2U})w, \tag{4.6}$$

$$\bar{\Delta}(W^{-1}w) - W\Omega D(W^{-1}w) - 4pW^{-1}e^{-2U}w - \tfrac{1}{2}W^2 DD\Phi - \tfrac{1}{2}MD\Phi = 0, \tag{4.7}$$

$$Dk_A + 4k_A - W^{-2}\partial_A k - 2W^{-1}(\rho + p)e^{-6U}T_0 e^{U_0}\epsilon_{AB}\theta^B = 0, \tag{4.8}$$

$$\partial_{[A}k_{B]} = 0, \tag{4.9}$$

where $\Phi = \gamma^{AB}\Phi_{AB}$. Linearization of the remaining equilibrium conditions yields

$$\nabla_{(A}\theta_{B)} = 0 \quad \text{and} \quad D\theta^A = 0, \tag{4.10}$$

when $\rho \neq 0$, and similarly, linearization of ${}^3\nabla_r h^r$ gives

$$Dk + (2\Omega W^{-1} - MW^{-2})k + \nabla_A k^A = 0. \tag{4.11}$$

5. SOLUTION OF THE LINEARIZED EQUATIONS

Referring to Ref. 8 we can show that (4.4) – (4.7) are equivalent[24] to the linearized field equations in the static, spherically symmetric background, and the argument[25] in that paper gives us the result that w and Φ_{AB} vanish. Since the 1-form $k_A dx^A$ is closed on S_c by (4.9), and since $p_1(S^2) = 0$, $k_A dx^A$ is exact so that there exists a function $K(U, x^A)$ on S_c determined up to an additive function of U only such that $k_A = \partial_A K$. Equations (4.8) and (4.11) now become

$$\partial_A(DK + 4K - W^{-2}k) = 2T_0 e^{U_0 - 6U}W^{-1} \times (\rho + p)\epsilon_{AB}\theta^B, \tag{5.1}$$

$$\bar{\Delta}K = -Dk - W^{-1}(2\Omega - MW^{-1})k. \tag{5.2}$$

To solve for K and k we let $\gamma_{AB}dx^A dx^B = r^2(U)(d\theta^2 + \sin^2\theta d\phi^2)$. Equation (4.10) implies that θ^A is a U independent Killing vector field on the Euclidean 2-sphere. Choosing the direction of the rotation axis to be the z-axis we let $\theta^A\partial_A = a\partial_\phi$ and expand into spherical harmonics,

$$K = \sum_{l,m} K_{lm}(r)Y_{lm}(\theta,\phi), \quad k = \sum_{l,m} k_{lm}(r)Y_{lm}(\theta,\phi), \tag{5.3}$$

where

$$\int\int Y_{lm}Y_{l'm'}\sin\theta\, d\theta\, d\phi = \delta_{ll'}\delta_{mm'}.$$

As in Ref. 8, $\Omega = 2W\mathscr{R}'$ where $\mathscr{R}' = \partial_U(\ln r)$. Equation (5.2) then becomes $0 = r^2 Dk_0 + r^2 W^{-1}(2\Omega - MW^{-1})k_0$ for $l = 0$ and

$$K_{lm} = r^2[l(l+1)]^{-1}[Dk_{lm} + (4\mathscr{R}' - MW^{-2})k_{lm}] \quad (l \neq 0) \tag{5.4}$$

while (5.1) yields

$$\sum_{l,m}(DK_{lm} + 4K_{lm} - W^{-2}k_{lm})\partial_\theta Y_{lm} = 2aT_0 e^{U_0 - 6U}W^{-1}(\rho + p)r^2\sin\theta$$

and

$$\sum_{l,m}(DK_{lm} + 4K_{lm} - W^{-2}k_{lm})\partial_\phi Y_{lm} = 0$$

so that

$$DK_{lm} + 4K_{lm} - W^{-2}k_{lm} = 0 \quad \text{for} \quad m \neq 0 \text{ or } l \neq 0,1 \tag{5.5}$$

and

$$DK_1 + 4K_1 - W^{-2}k_1 = -2a(4\pi/3)^{1/2}T_0 e^{U_0 - 6U} \times W^{-1}(\rho + p)r^2. \tag{5.6}$$

Note that K_0 is an arbitrary function of U and need not be known to determine $k_A dx^A$ uniquely.

With the use of Eqs. (3.9) and (3.14) on the spherical background we have, from the equation for k_0, $k_0 = cW \times r^{-2}$, for $c =$ constant. By doing an expansion in normal coordinates (y^i) at the center we find that $W = (\frac{1}{3})M_0|y| + O(|y|^3)$ and $r = |y| + O(|y|^3)$ so that in order for k_0 to vanish at the center (since $h^0 = h^i\partial_i U = 0$ at the center), $c = 0$, whence $k_0 \equiv 0$. Equation (5.5), with the use of Eq. (5.4) to eliminate K_{lm} and Eqs. (3.9) and (3.14) on the spherical background, becomes

$$W^2 DD\hat{k}_{lm} + MD\hat{k}_{lm} - F_l(U)\hat{k}_{lm} = 0 \quad \text{for} \quad m \neq 0 \text{ or } l \neq 0,1, \tag{5.7}$$

where $\hat{k}_{lm} = rW^{-1}e^{2U}k_{lm}$ and

$$F_l(U) = W^2(5 - 4\mathscr{R}') + l(l+1)r^{-2} + 2M - pe^{-2U}. \tag{5.8}$$

Since the three-dimensional Laplacian Δ in the spherically symmetric case has the form $\Delta f = W^2 DDf + MDf + \bar{\Delta}f$, Equation (5.7) is equivalent to

$$\Delta\hat{k}_{lm} = F_l(U)\hat{k}_{lm} \quad \text{for} \quad m \neq 0 \text{ or } l \neq 0,1. \tag{5.9}$$

We now use an argument similar to that for ω in Ref. 8. Let $u = Wr$, $v = pr^2 e^{-2U}$, and $x = Mr^2/(3u)$ so that $u, v, x \geq 0$ in the physical domain, and $ur =$ constant and $v = x = 0$ in vacuo. We can show that $F_l(U) \geq 0$ is equivalent to

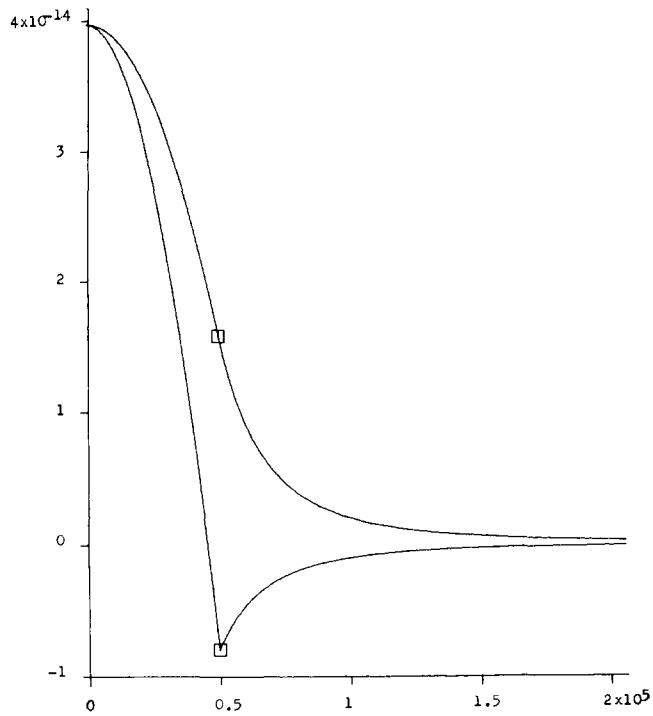


FIG. 1. Radial (upper curve) and tangential components of the vector field k as a function of the radial distance from the center for the Tolman solution IV with total gravitational mass $m = 1$ and $\beta = p_0/\rho_0 = 10^{-5}$ (qualitatively typical for the Newtonian limit). The symbols □ indicate surface values.

$$l^2(l+1)^2 + 9u^4 + 36u^2x^2 + v^2 + 2[5l(l+1) - 8]u^2 + 12l(l+1)ux - 2l(l+1)v + 60u^3x - 26u^2v - 12uxv \geqslant 0.$$

Since $p, \rho \geqslant 0$, we have that $0 \leqslant v \leqslant 2xu$ so that the above inequality is implied by

$$l^2(l+1)^2 + 9u^4 + 12u^2x^2 + 2[5l(l+1) - 8]u^2 + 8u^3x + v^2 \geqslant 0.$$

Since $5l(l+1) - 8 \geqslant 2$ for $l \geqslant 1$, this shows that $F_l(U) > 0$ for $l \geqslant 1$. Asymptotic flatness conditions with Einstein's vacuum equations give us that $h^0 = O(r^{-4})$, $U = mr^{-1} + O(r^{-2})$, and $W = mr^{-2} + O(r^{-3})$ in the asymptotic region so that $\hat{k}_{lm} = O(r^{-1})$ in this region. Similarly we find that $\hat{k}_{lm} = 0$ at the center. These boundary conditions together with (5.9) and $F_l(U) \geqslant 0$ for $l \geqslant 1$ yield $\hat{k}_{lm} = 0$ for $m \neq 0$ or $l \neq 0,1$.

Thus the only nonvanishing component is $\hat{k}_1 = \hat{k}_{10}$ which, by (5.6) and (5.4), satisfies

$$W^2 D D\hat{k}_1 + MD\hat{k}_1 - F_1(U)\hat{k}_1 = -c_0 re^{-4U}(\rho + p), \tag{5.10}$$

where $c_0 = 4a(4\pi)^{1/2}3^{-1/2}T_0 e^{U_0}$. In vacuo one can easily show, using the exterior Schwarzschild solution, that $W = m^{-1}\sinh^2 U$, $r = -m\sinh^{-1}U$ and $\mathscr{R}' = -\coth U$. The only solution of (5.10) satisfying the asymptotic conditions is then found to be

$$\hat{k} = 4Ae^U \sinh^2 U. \tag{5.11}$$

It agrees with what one obtains by linearizing the Kerr solution on a spherical background, in this formalism.

To determine the solutions in general we introduce a coordinate z defined by $e^U = e^{U_0} + \alpha^2 z^2$ where $\alpha > 0$ is chosen such that z (surface of star) $= 1$. Note that z behaves like a radial polar coordinate near the center since U has a positive definite critical point. Equation (5.10) can then be written as

$$ad^2\hat{k}/dz^2 + bd\hat{k}/dz + c\hat{k} + c_0 q = 0, \tag{5.12}$$

where

$$a = e^{2U}W^2, \quad b = 2\alpha^2 z e^{2U}M - z^{-1}(e^{2U_0} - \alpha^4 z^4)W^2,$$
$$c = -4\alpha^4 z^2[2M - pe^{-2U} - 4W^2\mathscr{R}' + 5W^2 + 2r^{-2}],$$

and

$$q = 8\alpha^4 z^2 e^{-2U} r(M - pe^{-2U}).$$

From the junction conditions (Sec. 3) we see that a is C^0 while b, c, and q may have jump discontinuities at $z = 1$. Also, since k is C^0, $\hat{k}$ must be C^0 at the boundary and K is also clearly C^0 at the boundary. We assume the equation of state is analytic in order to do the following expansions near the center in terms of z. After a lengthy calculation (and dividing by a common factor) we obtain

$$a = z^2 + a_4 z^4 + O(z^6), \quad b = 2z + b_3 z^3 + O(z^5),$$
$$c = -2 + O(z^2), \quad q = q_3 z^3 + q_5 z^5 + O(z^7), \tag{5.13}$$

where the coefficients depend on α, $U_0, p_0, \rho_0, dp/d\rho$ $(z = 0)$, etc. Considering the homogeneous equation to (5.12), we see that $z = 0$ is the only singularity in [0,1] and it is a regular singular point, so if we make suitable power series expansions we find that the general solution of (5.12) satisfying the regularity conditions at the center is

$$\hat{k} = \mu\hat{k}_h + \lambda\hat{k}_i, \tag{5.14}$$

where $\hat{k}_h$ is the solution of the homogeneous equation with $\hat{k}_h(0) = 0$, $dk_h/dz(0) = 1$ and $\hat{k}_i$ the solution of (5.12) with $\hat{k}_i(0) = 0$, $dk_i/dz(0) = 0$. To determine λ and μ we use the fact that $d\hat{k}/dU$ is continuous at the star boundary by (5.4) and the junction conditions to find that the boundary conditions are $\hat{k}(1) = Ae^{-U_s}(1 - e^{2U_s})^2 = :S$, $d\hat{k}/dz(1) = -2\alpha^2 Ae^{-2U_s}(1 - e^{2U_s})(1 + 3e^{2U_s}) = :V$ where $e^{U_s} = e^{U_0} + \alpha^2$. Thus λ and μ are determined by $\mu\hat{k}_h(1) + \lambda\hat{k}_i(1) = S$ and $\mu d\hat{k}_h/dz(1) + \lambda d\hat{k}_i/dz(1) = V$. These equations will uniquely determine μ and λ provided $\mathscr{W}(1) \neq 0$ where $\mathscr{W}(z) = \hat{k}_h d\hat{k}_i/dz - \hat{k}_i d\hat{k}_h/dz$. Now, $\mathscr{W}(z)$ obeys, by (5.12),

$$a\frac{d\mathscr{W}}{dz} + b\mathscr{W} + q\hat{k}_1 = 0 \tag{5.15}$$

with $\mathscr{W}(0) = 0$.

A power series expansion gives a unique solution for (5.15), $\mathscr{W} = -z^3/5 + O(z^5)$ where the factor q_3 has been absorbed into c_0 which is just a fixed constant. Thus for small $z > 0$, $\mathscr{W} < 0$ and $d\mathscr{W}/dz < 0$. Now, $a, q > 0$ in (0,1] [see (5.13)] and $\hat{k}_h(0) > 0$ in (0,1] since $\hat{k}_h(0) = 0$, $d\hat{k}_h/dz(0) > 0$ and if $z_1 \in (0,1]$ is the smallest z such that $d\hat{k}_h/dz(z_1) = 0$, then $d^2\hat{k}_h/dz^2(z_1) = -(c/a)\hat{k}_h(z_1) \geqslant 0$ since $c = -4\alpha^4 z^2 F_1(U) \leqslant 0$. Thus $\hat{k}_h$ has no maximum in (0,1].

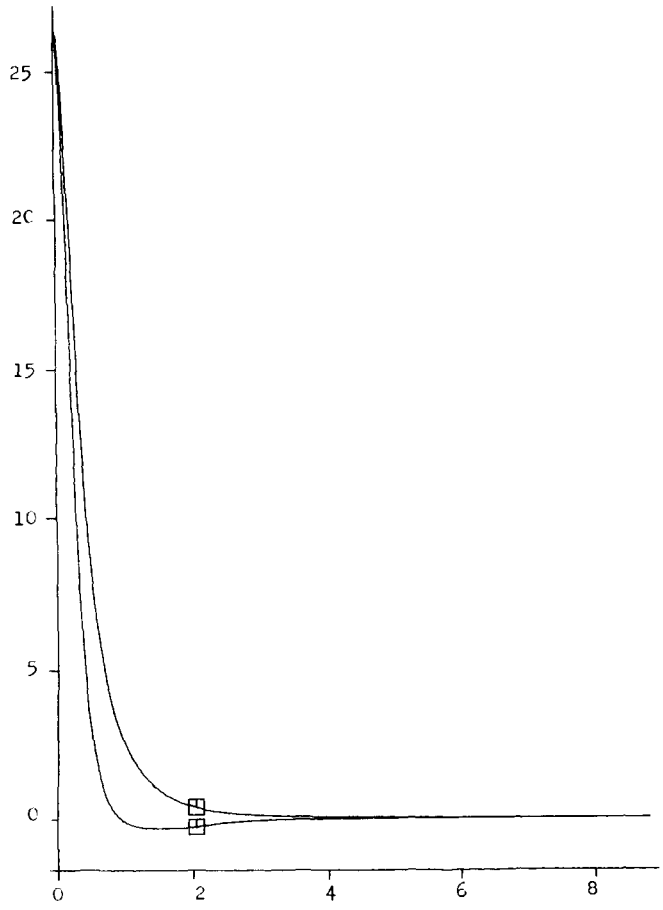


FIG. 2. As Fig. 1, but $\beta = 0.28$. For $\beta > 1/3$ the condition that $\rho \geqslant p$ everywhere is violated and the magnitude of **k** at the center becomes infinite.

Now suppose z_2 is the smallest $z \in (0,1]$ such that $\mathscr{W}(z_2) = 0$. Then $d\mathscr{W}/dz(z_2) = -q(z_2)k(z_2)a^{-1}(z_2) < 0$ so that there exists an $\epsilon > 0$ such that $\mathscr{W}(z) > 0$ for $z \in (z_2 - \epsilon, z_2)$ which contradicts $\mathscr{W} < 0$, $d\mathscr{W}/dz < 0$ for small z. Therefore, $\mathscr{W}(1) \neq 0$ and there exist unique λ and μ solving the equations giving the boundary conditions of $\hat{k}$. We summarize these results in the following theorem.

Theorem: Let $\mathscr{G}$ denote the set of (isometry classes of) globally stationary, asymptotically Euclidean space–times consisting of

(i) a spatially bounded barotropic perfect fluid in rigid motion with a fixed analytic equation of state $\rho = \rho(p)$, satisfying $\rho \geqslant p \geqslant 0$, $dp/d\rho > 0$, and fixed central values p_0 of the pressure and U_0 of the gravitational potential;

(ii) an exterior vacuum region.

Then the space of infinitesimal deformations in $\mathscr{G}$ of the unique static spherically symmetric $g_0 \in \mathscr{G}$ is three-dimensional (corresponding to the space of infinitesimal isometries $\theta^i \partial_i$ of $\gamma_{ij} dx^i dx^j$).

6. EXAMPLES

By (2.9) and (2.14) the vector field $k^i = \delta h^i$ satisfies on the spherical background the equations

$$\epsilon^{irs} \partial_r k_s = 2T_0 e^{U_0}(\rho + p) e^{-6U} \theta^i \tag{6.1}$$

and

$$\nabla_s k^s = 0, \tag{6.2}$$

where $\theta^i = \delta\Theta^i$ is a Killing vector field of the 3-metric γ_{ij}. The field $\mathbf{k} = (k^i)$ therefore behaves qualitatively just like the magnetic field of the earth, for example. It does not seem to change qualitatively under a change of the equation of state. It vanishes in the nonrelativistic limit, while in the ultrarelativistic limit the peak of the magnitude of **k** at the center becomes sharper.

In Figs. 1 and 2 we plotted the radial component k_r and the tangential component k_θ of **k** against r for a weakly and a highly relativistic case (characterized by the value of the quotient $\beta = p_0/\rho_0$ of central pressure and density being small and big, respectively). The plots are for Tolman's solution IV.[18] We also plotted the curves for the Schwarzschild interior, but there is no qualitative difference.

ACKNOWLEDGMENT

We thank Werner Israel for several interesting discussions and Walter Aiello for help with the numerical calculations.

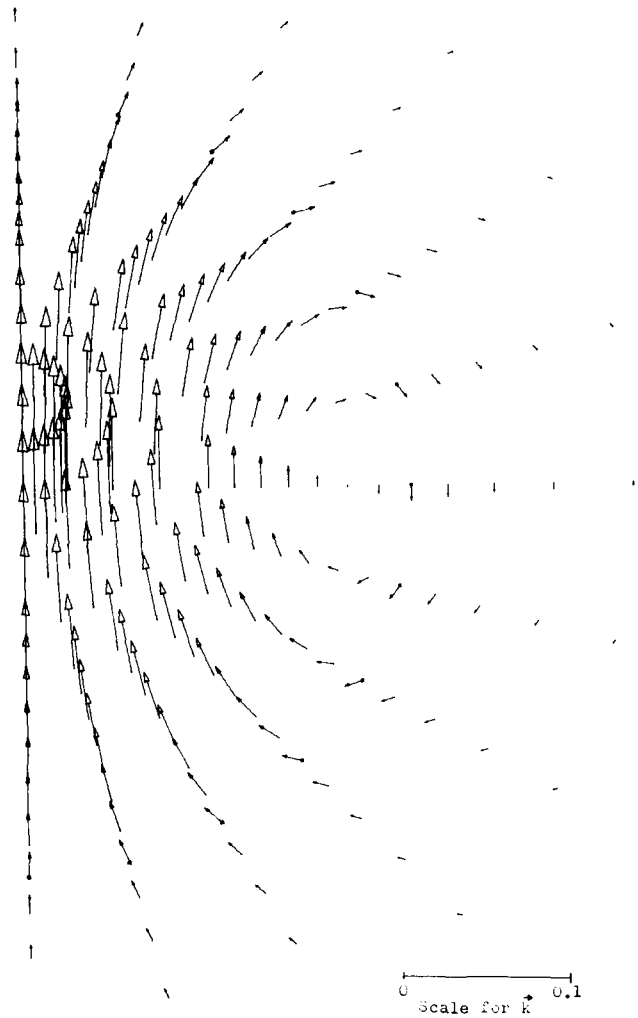


FIG. 3. The vector field **k** plotted in a half plane $r \geqslant 0$, $0 \leqslant \theta \leqslant \pi$, $\varphi = \text{const}$ (for $m = 1$ and $\beta = 0.1$). The symbols □ indicates the surface ($r = 5.408$).

[1]See, mainly, J.B. Hartle, Astrophys. J. **150**, 1005–1029 (1967) and several subsequent papers in this series, and, for example, S. Chandrasekhar and J.L. Friedman, Astrophys. J. **175**, 379–405 (1972) (stability criteria).
[2]L. Lichtenstein, *Gleichgewichtsfiguren rotierender Flüssigkeiten* (Springer, Berlin, 1933).
[3]W. Israel, Phys. Rev. **164**, 1776–779 (1967); B. Carter, Phys. Rev. Lett. **26**, 331–33 (1971); S.W. Hawking, Commun. Math. Phys. **25**, 152–66 (1972); D.C. Robinson, Phys. Rev. Lett. **34**, 905 (1975); Gen. Rel. Grav. **8**, 695–98 (1977).
[4]A.E. Fischer and J.E. Marsden, Bull. Am. Math. Soc. **79**, 997–1003 (1973).
[5]See S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale Univ., New Haven, 1969). Nonelliptic and heterogeneous configurations were already treated in Ref. 2. For more recent references see, e.g., B.L. Smith, Astrophys. Space Sci. **25**, 195–205 (1973).
[6]T. Carleman, Math. Zeitschr. **3**, 1–7 (1918).
[7]A. Avez, Ann. Inst. Henri Poincaré **1**, 291–300 (1964).
[8]H.P. Kunzle, Commun. Math. Phys. **20**, 85–100 (1971).
[9]D.W. Marks, Astrophys. J. **211**, 266–69 (1977), claims to prove spherical symmetry for static fluids. However, he assumes, without justification, that the curvature of equipotential surfaces is constant.
[10]Our notations and conventions are those of C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973); except for the constant in Einstein's equations which we write $G_{\alpha\beta} = T_{\alpha\beta}$.
[11]See, for example, Ref. 13.
[12]See, for example, J.M. Stewart, *Nonequilibrium Relativistic Kinetic Theory* (Springer, New York, 1971).
[13]L. Lindblom, Astrophys. J. **208**, 873–80 (1976).
[14]A. Lichnerowicz, *Théories relativistes de la gravitation et de l'éléctromagnetisme*, (Masson, Paris, 1955).
[15]J. Ehlers, Proc. London Conf. relativistic theories of gravitation (GR 4) (London, 1965).
[16]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* I, II (Interscience, New York, 1963, 1969). See also S. Sternberg, *Lectures on Differential Geometry* (Prentice Hall, Englewood Cliffs, 1964).
[17]S.W. Hawking and G.F.R. Ellis *The Large Scale Structure of Space-Time* (Cambridge Univ., Cambridge, 1973).
[18]R.C. Tolman, Phys. Rev. **55**, 364–73 (1939).
[19]R.S. Palais, *A Global Formulation of the Lie Theory of Transformation Groups* (A.M.S. Memoirs, Providence, 1957).
[20]We follow here the notation of Ref. 19, e.g., if π is a differentiable map, then π_* denotes the tangent map and π^* the pullback of differential forms.
[21]M. Berger and D. Ebin, J. Diff. Geom. **3**, 379–92 (1969).
[22]S. Kobayashi, *Transformation Groups in Differential Geometry* (Springer, New York, 1972), p. 47.
[23]For static spherically symmetric solutions with a given equation of state $\rho = \rho(p)$ there remains only one arbitrary parameter, which we have already chosen to be U_0. This will determine the central pressure, central density, total gravitational mass, radius of the star, etc. For the slowly rotating case one would expect only one more parameter, for example, the angular velocity ω. But since the central values of the density, pressure and potential do not depend on the direction of the rotation they will depend at least quadratically on ω so that, when the equations are linearized on the spherical background, a degeneracy occurs. This is why we must fix p_{center} independently in this case.
[24]In Ref. 8 it was not realized that the components γ_{1A} can be made to vanish globally. Therefore, the linearized equations there contain an additional quantity Q_A that can, in fact, be put equal to zero.
[25]A lemma is stated and used in this argument, but the proof given is not complete as it ignores problems of differentiability. For a discussion and a proof of a revised (weaker) lemma sufficient for our argument see H.P. Künzle and J.R. Savage, C.R. Acad. Sci. Paris, Ser. A **287**, 249–50 (1978). Also note that the first term of (4.14) in Ref. 8 should have a negative sign. The argument in the appendix needed to show that $\omega = 0$ can be improved to hold also in the case where $0 \leqslant p \leqslant \rho$.

Equilibrium states for mean field models

M. Fannes,[a] H. Spohn,[b] and A. Verbeure
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 20 March 1979; accepted for publication 15 October 1979)

We rigorously characterize the KMS and the limiting Gibbs states for mean field models. As an application we prove the convergence of the Gibbs states for the Dicke Maser model in the infinite volume limit.

I. INTRODUCTION

Let $\mathscr{H}$ be a Hilbert space and let Λ be a finite ordered subset $\{i_1,...,i_n\}$ of $\mathbb{N}$ let $\mathscr{B}_\Lambda$ denote the C^*-algebra

$$1 \otimes \cdots \otimes \underset{\uparrow}{\mathscr{B}}_{i_1} \otimes \cdots \otimes \underset{\uparrow}{\mathscr{B}}_{i_2} \cdots \otimes \underset{\uparrow}{\mathscr{B}}_{i_n},$$

where $\mathscr{B}$ is the C^*-algebra of all bounded operators on $\mathscr{H}$. For $\Lambda \subset \Lambda'$ there is a canonical isotony of $\mathscr{B}_\Lambda$ in $\mathscr{B}_{\Lambda'}$, allowing us to define the C^*-algebra $\mathscr{B}^\infty$ as the inductive limit of $\{\mathscr{B}_\Lambda | \Lambda \subseteq \mathbb{N}\}$.

A state ω of $\mathscr{B}^\infty$ is called *locally normal* if for every finite $\Lambda \subset \mathbb{N}$, the restriction ω_Λ of ω to $\mathscr{B}_\Lambda$ is a normal state of $\mathscr{B}_\Lambda$.

A state ω of $\mathscr{B}^\infty$ is called *symmetric* if for all elements of the type $X_1 \otimes \cdots \otimes X_n$, $X_i \in \mathscr{B}$, $n \in \mathbb{N}$

$$\omega(X_1 \otimes \cdots \otimes X_n) = \omega(X_{i_1} \otimes \cdots \otimes X_{i_n}),$$

where $i = (i_1,...,i_n)$ is a permutation of $(1,2,...,n)$.

A state ω of $\mathscr{B}^\infty$ is called a *product state* if for all Λ_1 and Λ_2 finite subsets of $\mathbb{N}$ such that $\Lambda_1 \cap \Lambda_2 = \phi$ and $X_i \in \mathscr{B}_{\Lambda_i}$, $i = 1,2$, holds: $\omega(X_1 X_2) = \omega(X_1)\,\omega(X_2)$.

We denote by ω_ρ the symmetric locally normal product state defined by the density matrix ρ on $\mathscr{H}$ such that

$$\omega_\rho(X_1 \otimes \cdots \otimes X_n) = \prod_{i=1}^{n} \operatorname{Tr} \rho X_i .$$

Then the following result holds:[1]

Theorem I.1: Let ω be a locally normal symmetric state of $\mathscr{B}^\infty$. Then there exists a unique probability measure μ on the set $\mathscr{B}^1$ of density matrices on $\mathscr{H}$ such that $\omega = \int d\mu(\rho)\omega_\rho$. We start from this theorem to study general properties of the equilibrium statistical mechanics of all models of the mean-field type.

Our main result consists in an abstract proof of the existence of the so-called "reduced problem" per lattice site. More precisely: Any KMS state is described by a measure μ with support on those locally normal product states ω_ρ, for which ρ is a solution of the gap-equation. Furthermore for limiting Gibbs states the support of the measure μ is concentrated those locally normal product states ω_ρ satisfying the variational principle. As a by-product the convergence of the free energy is shown. Finally the results are applied to the Dicke maser model.

[a] Aangesteld Navorser NFWO, Belgium.
[b] On leave of absence from Fachbereich Physik, Universität München.

II. EQUILIBRIUM STATES FOR MEAN-FIELD MODELS

For purely technical convenience, in this section we restrict ourselves to the case $\dim\mathscr{H} < \infty$. At the end of the paper we treat a model where $\dim\mathscr{H} = \infty$. By mean-field models we mean systems with a local Hamiltonian of the following type: for each finite subset Λ of $\mathbb{N}$:

$$H_\Lambda = \sum_{i \in \Lambda} A_i + \frac{1}{2N(\Lambda)} \sum_{\substack{i \neq j \\ i,j \in \Lambda}} B_{ij}, \tag{1}$$

where $A_i \in \mathscr{B}_{\{i\}}$ and all $A_i i \in \Lambda$ are copies of $A = A^* \in \mathscr{B}$ and where $B_{ij} \in \mathscr{B}_{\{i,j\}}$ and all B_{ij} $(i,j \in \Lambda)$ are copies of a self-adjoint B in $\mathscr{B} \otimes \mathscr{B}$, which is invariant under the symmetry $(ij) \to (ji)$. $N(\Lambda)$ is the number of elements in Λ. Remark that our results extend trivially to many body interactions of mean-field type. For infinite $\Lambda \subset \mathbb{N}$, let $\omega_{\beta,\Lambda}$ be the canonical Gibbs state of $\mathscr{B}_\Lambda$ at inverse temperature $\beta > 0$

$$\omega_{\beta,\Lambda}(X) = \frac{\operatorname{Tr} e^{-\beta H_\Lambda} X}{\operatorname{Tr} e^{-\beta H_\Lambda}}; \quad X \in \mathscr{B}_\Lambda . \tag{2}$$

An alternative way to characterize the canonical Gibbs state $\omega_{\beta,\Lambda}$ is to impose the condition that for all $X \in \mathscr{B}_\Lambda$:[2]

$$\beta\omega_{\beta,\Lambda}(X^*[H_\Lambda, X]) \geq \omega_{\beta,\Lambda}(X^*X) \ln \frac{\omega_{\beta,\Lambda}(X^*X)}{\omega_{\beta,\Lambda}(XX^*)}. \tag{3}$$

Notice that $\mathscr{B}_\Lambda$ is isomorphic with $\mathscr{B}_{\{1,...,N(\Lambda)\}}$ and because of the mean-field character of the Hamiltonian this isomorphism maps the Gibbs state $\omega_{\beta,\Lambda}$ into $\omega_{\beta,\{1,...,N(\Lambda)\}}$. Therefore for the thermodynamic limit we can restrict ourselves to increasing intervals $\Lambda_n = \{1,...,n\}$. We denote $\omega_{\beta,\Lambda_n} = \omega_{\beta,n}$. Consider a w^*-limit point ω_β of the set $\{\omega_{\beta,n} | n \in \mathbb{N}\}$. Then due to the symmetry of the H_Λ, ω_β is a symmetric state, and Theorem I.1 yields

$$\omega_\beta = \int d\mu(\rho)\omega_\rho , \tag{4}$$

for some probability measure μ.

Lemma II.1: Let X be any element of $\mathscr{B}_\Lambda$, then with ω_β as above

$$\beta \int d\mu(\rho)\omega_\rho(X^*[H_\rho^\Lambda, X]) \geq \omega_\beta(X^*X) \ln \frac{\omega_\beta(X^*X)}{\omega_\beta(XX^*)},$$

where $H_\rho^\Lambda = \Sigma_{i \in \Lambda} H_{\rho,i}$ and $H_{\rho,i} \in \mathscr{B}_{\{i\}}$ is a copy of $H_\rho = A + B_\rho$ at the ith site with $B_\rho = \operatorname{Tr}_2 (1 \otimes \rho)B$. Tr_2 is the partial trace over the second Hilbert space.

Proof: Because of the assumption and (3) it is sufficient to compute:

 0022-2488/80/020355-04$1.00

$$\lim_{n\to\infty}\omega_{\beta,n}(X^*[H_{\Lambda_n},X]).$$

Clearly

$$\lim_{n\to\infty}\omega_{\beta,n}\left(X^*\left[\sum_{i\in\Lambda_n}A_i,X\right]\right)=\omega_\beta\left(X^*\left[\sum_{i\in\Lambda}A_i,X\right]\right),$$

and

$$\lim_{n\to\infty}\omega_{\beta,n}\left(X^*\left[\frac{1}{2n}\sum_{\substack{i\neq j\\ i,j\in\Lambda_n}}B_{ij},X\right]\right)$$

$$=\lim_{n\to\infty}\omega_{\beta,n}\left(X^*\frac{1}{2n}\left(\sum_{\substack{i\neq j\\ i,j\in\Lambda}}+\sum_{\substack{i\in\Lambda\\ j\in\Lambda_n\setminus\Lambda}}+\sum_{\substack{j\in\Lambda\\ i\in\Lambda_n\setminus\Lambda}}\right)[B_{ij},X]\right)$$

$$=\lim_{n\to\infty}\frac{n-N(\Lambda)}{n}\omega_{\beta,n}\left(X^*\sum_{i\in\Lambda}[B_{ij_0},X]\right)\quad(j_0\notin\Lambda),$$

by symmetry. Remark that $X^*\Sigma_{i\in\Lambda}[B_{ij_0},X]$ is a local element independent of Λ_n. Therefore

$$\omega_\beta\left(X^*\sum_{i\in\Lambda}[B_{ij_0},X]\right)=\int d\mu(\rho)\omega_\rho\left(X^*\sum_{i\in\Lambda}[B_{\rho,i},X]\right)\ \blacksquare$$

Lemma II.1 proves that any limit point of local equilibrium states satisfies the inequality: For any $X\in\mathscr{B}_\Lambda$

$$\beta\int d\mu(\rho)\,\omega_\rho(X^*[H_\rho^\Lambda,X])\geqslant\omega(X^*X)\ln\frac{\omega(X^*X)}{\omega(XX^*)},\qquad(5)$$

with $\omega=\int d\mu(\rho)\omega_\rho$. We remark that for any density matrix ρ on $\mathscr{H}$, the state ω_ρ of $\mathscr{B}^\infty$ is a factor state. If $\rho\neq\rho'$, then the representations of ω_ρ and of $\omega_{\rho'}$ are quasi-inequivalent, which implies that the representations are disjoint.[3,4]

Lemma II.2: If a symmetric state ω of $\mathscr{B}^\infty$ satisfies the inequality (5) then for μ almost all ρ one has

$$[\rho,H_\rho]=0.\qquad(6)$$

Proof: For any X in some $\mathscr{B}_{\{1,2,\ldots,n\}}$ consider

$$Y_N=\frac{1}{N}\left(X\otimes 1\otimes 1\cdots+1\otimes X\otimes 1\cdots+\overset{N-1}{\otimes}1\in X\right),$$

then

$$\lim_{N\to\infty}\omega(Y_N^*Y_N)=\lim_{N\to\infty}\omega(Y_NY_N^*)=\omega(X\otimes X^*),$$

and inequality (5) becomes

$$\int d\mu(\rho)\omega_\rho(X^*)\omega_\rho([H_\rho^{\{1,\ldots,n\}},X])\geqslant 0.$$

By taking $X=X^*$ remark that the left-hand side is purely imaginary, hence

$$\int d\mu(\rho)\omega_\rho(X)\omega_\rho([H_\rho^{\{1,\ldots,n\}},X])=0.$$

By a straightforward algebraic manipulation (polarization)

$$\int d\mu(\rho)\omega_\rho(Y)\omega_\rho([H_\rho^\Lambda,X])=0,\quad\text{for all } X\in\mathscr{B}_\Lambda,$$

and all $Y\in\mathscr{B}_\Lambda$. For any measurable set Δ in $\mathscr{B}_1$ such that $\mu(\Delta)>0$ consider the decomposition of ω

$$\omega=\mu(\Delta)\omega_\Delta+\mu(\mathscr{B}_1\setminus\Delta)\omega_{\mathscr{B}_1\setminus\Delta},$$

into the disjoint states

$$\omega_\Delta=\frac{1}{\mu(\Delta)}\int_\Delta d\mu(\rho)\omega_\rho$$

and

$$\omega_{\mathscr{B}_1\setminus\Delta}=\frac{1}{\mu(\mathscr{B}_1\setminus\Delta)}\int_{\mathscr{B}_1\setminus\Delta}d\mu(\rho)\omega_\rho.$$

Now there exists a net $\{Y_\alpha\}_\alpha$ of positive elements $Y_\alpha\in\mathscr{B}_{\Lambda(\alpha)}$ such that $\|Y_\alpha\|\leqslant 1$ and $\lim_\alpha\omega^\Delta(Y_\alpha)=1$, $\lim_\alpha\omega^{\mathscr{B}_1\setminus\Delta}(Y_\alpha)=0$.[5] Since $\|\omega_\rho([H_\rho^\Lambda,X])\|\leqslant 2N(\Lambda)\|X\|(\|A\|+\|B\|)$,

$$\frac{1}{\mu(\Delta)}\int_\Delta d\mu(\rho)\omega_\rho([H_\rho^\Lambda,X])=0.$$

Lemma II.3: If ω is a symmetric state of $\mathscr{B}^\infty$ satisfying the inequality (5), then for μ almost all ρ

$$\beta\omega_\rho(X^*[H_\rho^\Lambda,X])\geqslant\omega_\rho(X^*X)\ln\frac{\omega_\rho(X^*X)}{\omega_\rho(XX^*)},\qquad(7)$$

for all $\Lambda\subset N$ and all $X\in\mathscr{B}_\Lambda$.

Proof: Again for any measurable set Δ in $\mathscr{B}_1$ such that $\mu(\Delta)>0$ we consider the decomposition $\omega=\mu(\Delta)\omega_\Delta+\mu(\mathscr{B}_1\setminus\Delta)\omega_{\mathscr{B}_1\setminus\Delta}$ into disjoint states and the net $\{Y_\alpha\}_\alpha$ as in the proof of Lemma II.2. Consider now the net $\{Z_\alpha\}_\alpha$ where

$$Z_\alpha=\frac{1}{n_\alpha}\left\{X\otimes Y_\alpha\otimes 1\otimes\cdots+X\otimes\overset{n_\alpha-1}{\otimes}1\otimes Y_\alpha\right\},$$

$X\in\mathscr{B}_\Lambda$ for some $\Lambda\subset\mathbf{N}$, and where $\{n_\alpha\}_\alpha$ is a net of natural numbers to be specified later on.

One checks:

$$\lim_\alpha\omega_\Delta(Z_\alpha^*Z_\alpha)=\omega_\Delta(X^*X),$$

$$\lim_\alpha\omega_\Delta(Z_\alpha Z_\alpha^*)=\omega_\Delta(XX^*),$$

$$\lim_\alpha\omega_{\mathscr{B}_1\setminus\Delta}(Z_\alpha^*Z_\alpha)=\lim_\alpha\omega_{\mathscr{B}_1\setminus\Delta}(Z_\alpha Z_\alpha^*)=0.$$

Since $Z_\alpha\in\mathscr{B}_{\Lambda'(\alpha)}$ for some finite $\Lambda'(\alpha)$, for each ω_ρ

$$\omega_\rho(Z_\alpha^*[H_\rho^{\Lambda'(\alpha)},Z_\alpha])$$

$$=\frac{1}{n_\alpha}\omega_\rho(X^*\otimes Y_\alpha[H_\rho^{\Lambda'(\alpha)},X\otimes Y_\alpha])$$

$$+\frac{n_\alpha-1}{n_\alpha}\omega_\rho(X^*\otimes Y_\alpha\otimes 1\times[H_\rho^{\Lambda'(\alpha)},X\otimes 1\otimes Y_\alpha])$$

$$=\frac{1}{n_\alpha}\omega_\rho(X^*[H_\rho^\Lambda,X])\omega_\rho(Y_\alpha^2)$$

$$+\frac{1}{n_\alpha}\omega_\rho(X^*X)\omega_\rho(Y_\alpha[H_\rho^{\Lambda(\alpha)},Y_\alpha])$$

$$+\frac{n_\alpha-1}{n_\alpha}\omega_\rho(X^*[H_\rho^\Lambda,X])\omega^2(Y_\alpha),$$

where Lemma II.2 is used. The net $\{n_\alpha\}_\alpha$ is specified such that

$$\lim_\alpha\frac{1}{n_\alpha}\omega_\rho(Y_\alpha[H_\rho^{\Lambda(\alpha)},Y_\alpha])=0,$$

for all $\rho \otimes \mathscr{B}_1$. Then clearly

$$\lim_\alpha \omega_\rho(Z_\alpha^*[H_\rho^{\Lambda'(\alpha)}, Z_\alpha]) = \begin{cases} \omega_\rho(X^*[H_\rho^\Lambda, X]), & \text{if } \rho \in \Delta, \\ 0, & \text{if } \rho \in \mathscr{B}_1 \setminus \Delta. \end{cases}$$

Hence by taking $X = Z_\alpha$ in the inequality (4) and taking the α-limit one gets the results. ■

Now we formulate our main result:

Theorem II.4: Let ω be any symmetric state of $\mathscr{B}^\infty$, ω satisfies inequality (5) if and only if

$$\rho = e^{-\beta H_\rho}/\mathrm{Tr}\, e^{-\beta H_\rho}, \quad \mu \text{ a.e.} \tag{8}$$

Proof: Using Lemma II.3 with $X \in \mathscr{B}$, one gets

$$\beta \,\mathrm{Tr}\rho X^*[H_\rho, X]) \geqslant \mathrm{Tr}\rho X^* X \ln \frac{\mathrm{Tr}\rho X^* X}{\mathrm{Tr}\rho X X^*}, \quad \mu \text{ a.e.}$$

Using the techniques of Ref. 2 the result is obtained. ■

From the theorem it follows that the problem of mean-field systems is reduced to solving the one-site nonlinear Eq. (8). For high temperatures we have:

Proposition II.5: For $\beta < 1/2\|B\|$, Eq. (8) has a unique solution.

Proof: Let F be the map $\rho \in \mathscr{B}_1 \to F(\rho) = e^{-\beta H_\rho}/Z(\rho) \in \mathscr{B}_1$

where $Z(\rho) = \mathrm{Tr} e^{-\beta H_\rho}$. We show that F is a strict contraction if $\beta < 1/2\|B\|$. Denote by $\|\cdot\|_1$ the trace norm, then for $\rho_1, \rho_2 \in \mathscr{B}_1$:

$$\begin{aligned} &\|F(\rho_1) - F(\rho_2)\|_1 \\ &\quad = \|\exp[-\beta H_{\rho_1} - \log Z(\rho_1)] \\ &\qquad - \exp[-\beta H_{\rho_2} - \log Z(\rho_2)]\|_1 \\ &\quad = \Big\| \int_0^1 dt \exp[-(1-t)(\beta H_{\rho_1} \\ &\qquad + \log Z(\rho_1)]\{\beta(B_{\rho_1} - B_{\rho_2}) \\ &\qquad + \log Z(\rho_1)/Z(\rho_2)\} \exp\{-t[\beta H_{\rho_2} \\ &\qquad + \log Z(\rho_2)]\} \Big\|_1 \\ &\quad \leqslant \beta\|B\|\|\rho_1 - \rho_2\|_1 + |\log Z(\rho_1 - \log Z(\rho_2)| \\ &\quad \leqslant 2\beta\|B\|\|\rho_1 - \rho_2\|_1 . \end{aligned}$$

The result follows from the contraction mapping principle.[6]

This result together with Lemma II.1 implies that the mean-field models we are considering have a unique limiting Gibbs state for high enough temperatures.

For low temperatures one cannot expect to have a unique solution of Eq. (8). In fact for the BCS-model[7] one shows that below the critical temperature Eq. (8) has the solutions ρ_{norm} corresponding to the normal phase, and $\rho(\alpha)$ ($\alpha \in [0, 2\pi]$) corresponding to the superconducting phase. Furthermore the limiting Gibbs state is given by the state

$$\omega_\beta = \frac{1}{2\pi}\int_0^{2\pi} d\alpha\, \omega_{\rho(\alpha)} .$$

This shows that the support of the measure $\mu(\rho)$ defining a limiting Gibbs state (4), can be strictly smaller than the set of solutions of (8). We expect this to be a general feature of mean-field theories. The support of the measure is reduced by the following proposition.

Define the pressure

$$p = \sup_\omega p(\omega),$$

where the sup is taken over all symmetric states of $\mathscr{B}^\infty$, $p(\omega) = s(\omega) - \beta e(\omega)$, where $s(\omega)$ is the entropy density and where $e(\omega)$ is the energy density of the state ω.[8]

Proposition II.6: If ω_β is a limiting Gibbs state, then

$$p(\omega^\rho) = p, \quad \mu \text{ a.e.}$$

Proof: We denote as before by $\omega_{\beta,n}$ the Gibbs state for the volume $\Lambda_n = \{1,\ldots,n\}$ and $\omega_{\beta,n}^m$ the restriction of $\omega_{\beta,n}$ to the subalgebra $\mathscr{B}_{\Lambda_m}$ $(m < n)$. Let $\{\omega_{\beta,n}\}_n$ be a subsequence of Gibbs states converging to ω_β. Let us divide up the interval $[1,n]$ into equal intervals of length m plus a remainder interval of length less than m. Then by the subadditivity and monotonicity of the total entropy S:

$$\frac{1}{n} S(\omega_{\beta,n}) \leqslant \frac{1}{n}\left(\left[\frac{n}{m}\right] + 1\right) S(\omega_{\beta,n}^m).$$

Since by assumption $\omega_{\beta,n}^m \underset{n}{\to} \omega_\beta^m$ as $n \to \infty$ and by continuity of entropy

$$\limsup_n (1/n) S(\omega_{\beta,n}) \leqslant (1/m) S(\omega_\beta^m),$$

this implies

$$\limsup_n (1/n) S(\omega_{\beta,n}) \leqslant s(\omega_\beta).$$

From the convergence of the energy density

$$\limsup_n (1/n)[S(\omega_{\beta,n} - \beta\omega_{\beta,n}(H_{\Lambda n})] \leqslant p(\omega_\beta).$$

On the other hand

$$\liminf_n (1/n)[S(\omega_{\beta,n}) - \beta\omega_{\beta,n}(H_{\Lambda n})] \geqslant p \geqslant p(\omega_\beta),$$

as a consequence of Klein's inequality. Hence $p = p(\omega_\beta)$. Since $\dim \mathscr{H} < \infty$, we have[8]

$$p(\omega_\beta) = \int d\mu(\rho) p(\omega_\rho),$$

and the proposition follows. ■

From the proof of this proposition we have

Corollary II.7: The thermodynamic limit of the pressure for the mean-field models considered exists and is given by

$$p = \lim_{n\to\infty} \frac{1}{n} \ln \mathrm{Tr}\, e^{-\beta H_{\Lambda_n}}.$$

III. APPLICATION–THE DICKE MASER MODEL

As an application we study the Dicke Maser model.[9] Hepp and Lieb[10] gave a rigorous treatment of the thermodynamics and of the convergence of the intensive observables for the one-mode version of the model. Here we discuss the infinite mode version with the Hamiltonian[11]

$$H_n = \sum_{k=1}^n a_k^+ a_k + \epsilon \sum_{k=1}^n \sigma_k^+ \sigma_k^-$$

$$+\frac{\lambda}{n}\left(\sum_{k,l=1}^{n} a_k^+\sigma_l^- + \text{h.c.}\right),$$

with $0<\epsilon\leqslant\lambda^2$.

a_k^+, a_k are copies of the creation and annihilation operators (a^+,a) of the harmonic oscillator describing the photon mode, acting on a Hilbert space $\mathscr{K}$ and satisfying the commutation relation $[a_k,a_a^*]=\mathbb{1}$. $\sigma_k^{\pm}$ are copies of the Pauli matrices σ^+ and σ^- acting on a two-dimensional Hilbert space $\mathbb{C}^2$. They satisfy the relations

$$\{\sigma^+,\sigma^+\}=2(\sigma^+)^2=\{\sigma^-,\sigma^-\}=0,$$
$$\{\sigma^+,\sigma^-\}=\mathbb{1}.$$

In Ref. 11 the variational principle of statistical mechanics for this model was solved partly in the sense that one found the infimum of the free energy by variation only over a class of states which factorize with respect to observables in the photon and atomic algebras. Also the thermodynamic limit of the free energy was proved.

Clearly, the model is a mean-field model with permutation symmetry, such that the analysis of the previous sections is applicable. The main results of this section will be the extension of the variational principle to all locally normal symmetric states, and the proof of the convergence of the Gibbs state.

For this model the algebra $\mathscr{B}$ becomes the set of all bounded operators on the Hilbert space $\mathscr{H}=\mathscr{K}\otimes\mathbb{C}^2$ which is an infinite dimensional Hilbert space. This means that as far as the applicability of Theroem I.1 is concerned not every state is locally normal. Moreover as the local Hamiltonians are unbounded operators, the correlation inequality in Sec. II should be applied with some care. The problem consists in extending the correlation inequality (3) to unbounded operators. For this model the domain questions are easily settled by observing that for finite volumes the total particle number operator is a conserved quantity.[12] The problem of local normality is solved in:

Proposition III.1: A limiting Gibbs state ω_β for the Dicke Maser model is locally normal.

Proof: We have to prove that for any finite volume Λ, the restriction $\omega_{\beta,\Lambda}$ of ω_β to $\mathscr{B}_\Lambda$ is a regular state; because of the symmetry it is sufficient to prove the regularity on $\mathscr{B}_{|1|}$.

This follows from von Newmann's uniqueness theorem, if the map

$$z\in\mathbb{C}\to\omega_\beta[\exp i(za_1^+ + \bar{z}a_1)],$$

is continuous at $z=0$. Let $\{\omega_{\beta,n}\}_n$ be a subnet of Gibbs states w^*-converging to ω_β, then

$$\begin{aligned}&|\omega_{\beta,n}[\exp i(za_1^+ + \bar{z}a_1)-\mathbb{1}]|^2\\ &\leqslant 2\{1-\omega_{\beta,n}[\cos(za_1^+ + \bar{z}a_1)]\}\\ &\leqslant 4\omega_{\beta,n}\left[\sin^2\left(\frac{za_1^+ + \bar{z}a_1}{2}\right)\right]\\ &\leqslant \omega_{\beta,n}[(za_1^+ + \bar{z}a_1)^2]\\ &= |z|^2\omega_{\beta,n}(\mathbb{1}+2a_1^+a_1).\end{aligned}\tag{9}$$

For an upper bound of the right-hand side, we apply inequality (3) with $X=a_1$. then

$$-\beta\omega_{\beta,n}(a_1^+a_1)+\lambda\beta\omega_{\beta,n}(a_1^+\sigma_1^-)$$
$$\geqslant\omega_{\beta,n}(a_1^+a_1)\ln\frac{\omega_{\beta,n}(a_1^+a_1)}{\omega_{\beta,n}(a_1a_1^+)}.$$

With $X=\sigma_1^+\sigma_1^-$, one obtains

$$-\lambda\beta\omega_{\beta,n}(a_1^+\sigma_1^-)\geqslant 0,$$

hence

$$-\beta\omega_{\beta,n}(a_1^+a_1)\geqslant\omega_{\beta,n}(a_1^+a_1)\ln\frac{\omega_{\beta,n}(a_1^+a_1)}{\omega_{\beta,n}(a_1a_1^+)}$$

or

$$\omega_{\beta,n}(a_1^+a_1)\leqslant 1/(e^\beta-1).\tag{10}$$

Therefore (9) becomes

$$\omega_{\beta,n}[\exp i(za_1^+ + \bar{z}a_1)-1]\leqslant[(e^\beta+1)/(e^\beta-1)]|z|^2.$$

On taking the limit $n\to\infty$, the proposition follows. ■

We are now in a position to apply Theorems I.1 and II.4. The problem is reduced to solving the so-called gap Eq. (8). But its solutions are well known: They exhibit a phase transition at a critical temperature T_c $(0<T_c<\infty)$. Above T_c equation (8) has a unique solution ρ_{norm} corresponding to the normal phase. Below T_c, in addition to the normal phase ρ_{norm} one obtains also a one-parameter family of solutions $\omega_{\rho(\alpha)}$, labeled by the spontaneously broken gauge symmetry parameter $\alpha\in[0,2\pi]$. $\omega_{\rho(\alpha)}$ has the properties that, if τ_α is the gauge autormorphism of $\mathscr{B}^\infty$ defined by

$$\tau_\alpha a_k^+ = e^{i\alpha}a_k^+,\quad \tau_\alpha\sigma_k^+ = e^{i\alpha}\sigma_k^+\quad (k=1,2,\ldots),$$

then

$$\omega_{\rho(\alpha)}\cdot\tau_{\alpha'}=\omega_{\rho[(\alpha+\alpha')\bmod 2\pi]},\tag{11}$$

and that[11]

$$p=p(\omega_{\rho(\alpha)})>p(\omega_{\rho_{\text{norm}}}).\tag{12}$$

Proposition II.6 holds true also for the Dicke Maser model, since by Proposition III.1 the local normality of the limiting Gibbs states is guaranteed and since by bound (10), energy and entropy densities are finite. For $T\geqslant T_c$, clearly $\omega_\beta=\omega_{\rho_{\text{norm}}}$. For $T<T_c$, Proposition II.6 and (12) imply

$$\omega_\beta=\int d\mu(\alpha)\,\omega_{\rho(\alpha)}.$$

Because of gauge invariance of ω_β and (11) necessarily $d\mu(\alpha)=(1/2\pi)d\alpha$. Therefore the Gibbs states $\omega_{\beta,n}$ converge to $\omega_\beta=(1/2\pi)\int_0^{2\pi}d\alpha\,\omega_{\rho(\alpha)}$ as $n\to\infty$.

[1] R.L. Hudson and G.R. Moody, Z. Wahrscheinlichkeitstheorie verw. Gebiete **33**, 343 (1976).
[2] M. Fannes and A. Verbeure, Commun. Math. Phys. **55**, 125 (1977); **57**, 165 (1977).
[3] P.T. Powers, Ann. Math. (N.Y.) **86**, 138 (1967).
[4] E. Størmer, J. Funct. Anal. **3**, 48 (1969).
[5] J. Dixmier, *Les C*-algèbres et leurs représentations* (Gauthier–Villars, Paris, 1964), Propositions 8.2.4 and 5.2.8.
[6] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1973).
[7] W. Thirring, Commun. Math. Phys. **7**, 181 (1968).
[8] D.W. Robinson, *The Thermodynamic Pressure in Quantum Statistical Mechanics*, Lecture Notes in Physics, No. 9 (1971).
[9] R.H. Dicke, Phys. Rev. **93**, 99 (1954).
[10] K. Hepp and E.H. Lieb, Ann. Phys. (N.Y.) **76**, 360 (1973).
[11] M. Fannes, P.N.M. Sisson, A. Verbeure, and J.C. Wolfe, Ann. Phys. **98**, 38 (1976).
[12] P.N.M. Sisson, A. Verbeure, and J.C. Wolfe, Ann. Soc. Sc. Brux. **89**, 552 (1975).

Approximate equilibrium states for two models of an interacting boson gas

Robert Critchley

Dublin Institute for Advanced Studies, Dublin, Ireland
and National Institute for Higher Education, Plassey House, Limerick, Ireland
(Received 7 July 1978; revised manuscript received 27 November 1978)

Two models for an infinitely extended boson gas are studied. The aim is to determine approximate equilibrium states for them by minimizing the free energy density over a class of trial states. For one of the models, in which the interaction is proportional to the square of the number of particles, the equilibrium state derived from the usual thermodynamic limiting argument (the Gibbs state) is also obtained. It is seen that the set of approximate equilibrium states consists of this Gibbs state and states with the same thermodynamic properties. Using the other model, in which the particles are assumed to interact according to a delta function interaction potential, some of the problems of approximate solutions are discussed.

1. INTRODUCTION

In this note we discuss two models of infinitely extended interacting boson systems in quantum statistical mechanics. The first one (Sec. 3) we are able to solve exactly; the second (Sec. 4) we solve approximately and are able to illustrate some of the problems of approximate solutions.

The method used is to apply the variational principle described by Robinson[1] and Švarc,[2] amongst others. This states that the equilibrium states of a quantum mechanical many-body system are the states that minimize the free energy density of that system expressed as a functional defined on a suitable set of states. "Suitable" means that we look only at the physically reasonable states. For example, Robinson[1] restricts attention to locally normal translation invariant states and proves, subject to certain restrictions on the models considered, the coincidence of the free energy density defined using the variational principle and that defined through the more usual limiting process.

The difficulty in following this program is that it requires an explicit expression for the free energy density functional in terms of the parameters of the state. This is not available in general, but for a restricted class of states it is both available and usable. We describe these states below, and then by minimizing the free energy density over these states we obtain an approximate equilibrium state.

2. QUASIFREE APPROXIMATION

A state of an infinitely extended boson system may be regarded as a functional $\mu: M \to C$ which is defined on a space $M \subseteq L^2(\mathbb{R}^3)$ of test functions and which satisfies conditions obtained by Araki[3] and Segal.[4] These authors also describe the connection between this definition of a state and the more usual algebraic one. Briefly, μ is related to the conjugate fields $\psi(x)$, $\psi^*(x)$ which satisfy the canonical commutation relations $[\psi(x), \psi^*(y)] = \delta(x-y)$, in that $\mu(h)$ is regarded as the expected value in a thermodynamic (or algebraic) state of $\exp[i(\psi(h)+\psi^*(h))/\sqrt{2}]$, i.e.,

$$\mu(h) = \langle \exp[i(\psi(h)+\psi^*(h))/\sqrt{2}] \rangle, \tag{2.1}$$

where

$$\psi(h) = \int \psi(x)\overline{h(x)}d^3x, \quad \psi^*(h) = \int \psi^*(x)h(x)d^3x, \tag{2.2}$$

and $\langle\ \rangle$ denotes the expectation value in the thermodynamic state. This formal connection motivates the definitions given below for the physical interpretation of μ.

Quasifree states were introduced by Robinson[5]; they are states determined by particularly simple functionals. It has been shown by various authors (Araki and Woods,[6] Lanford and Robinson,[7] Lewis and Pulè,[8] Critchley and Lewis[9]) that they emerge naturally in the study of boson systems. The quasifree states considered here are given by

$$\mu(h) = e^{-\|h\|^2/4}\, e^{(i/\sqrt{2})(\alpha\overline{\hat{h}(0)} + \bar{\alpha}\hat{h}(0))}\, e^{-\langle h, Ah\rangle/2}, \tag{2.3}$$

where $\alpha \in \mathbb{C}$, $\hat{}$ denotes Fourier transform, and A is a positive linear operator on M. We assume that A is a multiplication operator in the Fourier transformed space:

$$\langle h_1, Ah_2\rangle = \int \overline{\hat{h}_1(p)}\hat{a}(p)\hat{h}_2(p)\frac{d^3p}{(2\pi)^3}. \tag{2.4}$$

The state (2.3) is not invariant under the gauge transformation $h \to e^{i\theta}h$, but a gauge invariant state is obtained by averaging (2.3) over the gauge group:

$$\int \mu(e^{i\theta}h)\frac{d\theta}{(2\pi)} = e^{-\|h\|^2/4} J_0(\sqrt{2}|\alpha||\hat{h}(0)|) \times \exp(-\tfrac{1}{2}\langle h, Ah\rangle), \tag{2.5}$$

where J_0 is a Bessel function. Lewis and Pulè[8] showed how this state arises in the study of the free boson gas. It has the same thermodynamic properties as the state (2.3), but different algebraic ones. Both are interpreted as describing a boson gas undergoing Bose–Einstein condensation. The $e^{i\sqrt{2}(\cdot)}$

 0022-2488/80/020359-05$1.00

term in (2.3), and the $J_0(.)$ term in (2.5) describe the condensate of density $|\alpha|^2$, and the $e^{-\langle\cdot\rangle/2}$ term describes the normal component of the gas with density $\int \hat{a}(p) d^3p/(2\pi)^3$ (see below). The fact that the state (2.5) can be decomposed as a gauge average of (2.3) demonstrates the breaking of gauge symmetry expected of a condensing boson gas (see for example Johnston[10]).

The thermodynamic quantities that we require for the variational problem are defined as follows:

$$\bar{k}(\mu) = \int \frac{p^2}{2} \hat{a}(p) \frac{d^3p}{(2\pi)^3} \quad \text{(kinetic energy density)}, \tag{2.6}$$

$$\bar{n}(\mu) = \int \hat{a}(p) \frac{d^3p}{(2\pi)^3} \quad \text{(particle density)}, \tag{2.7}$$

$$\bar{s}(\mu) = \int [(1+\hat{a}(p))\log(1+\hat{a}(p)) - \hat{a}(p)\log\hat{a}(p)] \frac{d^3p}{(2\pi)^3}, \quad \text{(entropy density)}. \tag{2.8}$$

To see how these arise we look by way of example at $\bar{n}(\mu)$. If $\langle\ \rangle$ denotes the expectation in the thermodynamic state corresponding to μ, then $\langle\psi^*(x)\psi(x)\rangle$ integrated over a finite volume V is interpreted as the number of particles in that volume. So if $\langle\psi^*(x)\psi(x)\rangle$ is independent of x (as we find is the case here) then it may be interpreted as the density of particles in the state. Using the formal identifications of (2.1) and (2.2) we find that for the state (2.3) [or (2.5)]

$$\langle\psi^*(x)\psi(y)\rangle = |\alpha|^2 + a(x-y),$$

so that

$$\bar{n}(\mu) = \langle\psi^*(x)\psi(x)\rangle = |\alpha|^2 + a(0) = |\alpha|^2 + \int \hat{a}(p) \frac{d^3p}{(2\pi)^3}.$$

The expression for the kinetic energy density $\bar{k}(\mu)$ arises similarly. That for the entropy density $\bar{s}(\mu)$ is to be expected as it represents the entropy density of a linear superposition of harmonic oscillators. The problem in passing to the thermodynamic limit, however, is that in the weak *-topology the entropy density is only upper semicontinuous as a function of the states (Robinson[1]). The validity of the expression (2.8) has been established under various assumptions by Lanford and Robinson[11] and Fannes.[12] Critchley and Lewis[13] have also considered the problem, and Fannes[14] has derived the analogous expression for Fermi systems.

A model of an interacting statistical mechanical system amounts to the assumption of a potential energy density functional $\bar{u}(\mu)$. For example if the interaction is mediated by an Euclidean invariant two-body function $U(x-y)$, then in the state (2.3) [or (2.5)]

$$\bar{u}(\mu) = 2|\alpha|^2 \int \hat{U}(p)\hat{a}(p) \frac{dp}{(2\pi)^3} + \int \hat{a}(p)\hat{U}(p-q)\hat{a}(q) \frac{d^3p\, d^3q}{(2\pi)^6} + \hat{U}(0)[\bar{n}(\mu)]^2. \tag{2.9}$$

This arises from the potential energy term,

$$\int \psi^*(x)\psi^*(y)U(x-y)\psi(y)\psi(x) d^3x\, d^3y,$$

in the many-body Hamiltonian, so that the potential energy density in an Euclidean invariant state [in which $\langle\psi^*(x)\psi^*(y)\psi(y)\psi(x)\rangle$ depends only on $|x-y|$] is

$$\int U(x)\langle\psi^*(0)\psi^*(x)\psi(x)\psi(0)\rangle d^3x,$$

which can be evaluated using (2.1) and (2.2).

The free energy density functional at temperature T is then given by

$$\bar{f}(\mu) = \bar{k}(\mu) + \bar{u}(\mu) - T\bar{s}(\mu), \tag{2.10}$$

and we may now minimize this over the trial states (2.3) [or (2.5)] subject to the condition that the particle density $\bar{n}(\mu)$ should take a fixed value denoted by ρ. To take account of this we introduce a Lagrange multiplier γ and minimize instead

$$\bar{F}(\mu) = \bar{f}(\mu) - \gamma(\bar{n}(\mu) - \rho). \tag{2.11}$$

We first investigate whether there is a local minimum of $\bar{F}(\mu)$ by differentiating it with respect to the parameters determining the state. This gives the following equations:

$$\frac{\partial\bar{F}}{\partial\hat{a}(p)} : T\log\left(1 + \frac{1}{\hat{a}(p)}\right) = \frac{p^2}{2} - \gamma + \frac{\partial\bar{u}}{\partial\hat{a}(p)}, \tag{2.12}$$

$$\frac{\partial\bar{F}}{\partial|\alpha|} : 2|\alpha|\gamma = \frac{\partial\bar{u}}{\partial|\alpha|}, \tag{2.13}$$

$$\frac{\partial\bar{F}}{\partial\gamma} : \rho = |\alpha|^2 + \int \hat{a}(p) \frac{d^3p}{(2\pi)^3}. \tag{2.14}$$

The argument of the complex number α does not appear explicitly in $\bar{F}(\mu)$ [assuming it does not appear in $\bar{u}(\mu)$], so it is undetermined. Thus, if $|\alpha|$ is nonzero in the minimizing state, then there will be a family of states (each determined by a different value of the argument of α) minimizing the free energy. This nonuniqueness of the equilibrium stated illustrates the breaking of gauge symmetry mentioned earlier.

3. EXAMPLE 1: $\bar{u}(\mu) = 2c[\bar{n}(\mu)]^2$

Davies[15] has considered a class of interacting systems in which the interaction is a function of the particle density. Here we look at the particular case of this in which the finite volume Hamiltonian is

$$H = K + 2cN^2/|V| \tag{3.1}$$

(K = kinetic energy term, N = number operator, c is a positive constant and $|V|$ is the volume). This gives a potential energy density functional

$$\bar{u}(\mu) = 2c\bar{n}(\mu)^2. \tag{3.2}$$

Lebowitz and Penrose[16] and Švarc[2] have shown how this represents a repulsive interaction in the limit of extreme long range and extreme weakness.

For this model, Eq. (2.12)–(2.14) are

$$T\log\left(1 + \frac{1}{\hat{a}(p)}\right) = \frac{p^2}{2} - \gamma + 4c, \tag{3.3}$$

$$2|\alpha|(\gamma - 4c\rho) = 0, \tag{3.4}$$

$$|\alpha|^2 + \int \hat{a}(p) \frac{d^3p}{(2\pi)^3} = \rho. \tag{3.5}$$

To solve and interpret these equations let

$$\beta = 1/T \quad \text{and} \quad z_0 = \exp[\beta(\gamma - 4c\rho)]. \tag{3.6}$$

Then from (3.3), $\hat{a}(p) = z_0(e^{\beta p^2/2} - z_0)^{-1}$, (3.7)

from (3.4), $|\alpha| \log z_0 = 0$, (3.8)

from (3.5), $\rho = |\alpha|^2 + (2\pi\beta)^{-3/2} g_{3/2}(z_0)$, (3.9)

where

$$g_{3/2}(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^{3/2}}$$

$$= (2\pi\beta)^{3/2} \int \frac{z}{e^{\beta p^2/2} - z} \frac{d^3p}{(2\pi)^3}, \quad z \leqslant 1. \tag{3.10}$$

From (3.8) either $z_0 = 1$ or $|\alpha| = 0$. Let $\rho_c = (2\pi\beta)^{-3/2} g_{3/2}(1)$. If $\rho < \rho_c$, then from (3.9) $z_0 < 1$, and so we must have $|\alpha| = 0$. Therefore, z_0 is the unique solution of $\rho = (2\pi\beta)^{-3/2} g_{3/2}(z_0)$. If $\rho > \rho_c$, then from (3.9) we must have $|\alpha| > 0$, and so, from (3.8) $z_0 = 1$. Hence $|\alpha|^2 = \rho - \rho_c$. Thus, the gauge invariant state (2.5) that minimizes the free energy density is

$$\text{(i) } \rho < \rho_c: \mu(h) = e^{-\|h\|^2/4} \exp[-\tfrac{1}{2}\langle h, A(z_0)h\rangle], \tag{3.11}$$

$$\text{(ii) } \rho > \rho_c: \mu(h) = e^{-\|h\|^2/4} J_0(\sqrt{2}(\rho - \rho_c)^{1/2} |\hat{h}(0)|) \times \exp[-\tfrac{1}{2}\langle h, A(1)h\rangle], \tag{3.12}$$

where $\langle h_1, A(z)h_2\rangle = \int \frac{\overline{\hat{h}_1(p)} z \hat{h}_2(p)}{(e^{\beta p^2/2} - z)} \frac{d^3p}{(2\pi)^3}, \quad z \leqslant 1.$ (3.13)

This may be interpreted as showing that at density ρ_c the system undergoes a phase transition. For densities less than this there is a unique minimizing state and no Bose–Einstein condensation; at higher densities there is a continuum of minimizing states of the form (2.3) each determined by a different argument of the complex number α, and such that (3.12) is their average. Further, since $|\alpha|$ is nonzero the gas exhibits a Bose–Einstein condensation. Note that the fugacity $z' = e^{\beta\gamma}$ is related to the parameter z_0 according to $z_0 = z' e^{-4c\rho\beta}$.

To see that (3.11), (3.12) is the exact solution to this problem (not just an approximation) we compute the state using the Hamiltonian (3.1). If we do this in the canonical ensemble (in which the number of particles is fixed) then the interaction term $cN^2/|V|$ is just a constant. So this term cancels out in the density matrix expression for the state $(e^{-\beta H}/\text{trace} e^{-\beta H})$ and the system is equivalent, from a mathematical point of view, to the canonical ensemble of the free boson gas. But Cannon[17] has shown that the thermodynamic limit of this is the state (3.11), (3.12); therefore, this is also the limit of the equilibrium state determined by (3.1). Furthermore, Davies[15] has shown that, in the limit, the grand canonical and the canonical equilibrium states determined by (3.1) are the same. Therefore, (3.11), (3.12) is also the limit of the grand canonical state determined by (3.1). Hence, it may be regarded unequivocally as the solution of the model.

4. EXAMPLE 2: IMPENETRABLE POINT PARTICLES

The second example has an interaction given by the two-body potential

$$U(x - y) = c\delta(x - y) \tag{4.1}$$

or, more precisely, $\hat{U}(p) = c$ for all $p \in \mathbb{R}^3$. The interaction of (3.1) is sometimes considered as an approximation to this at small densities. For larger densities Huang[18] suggests that $2c(N^2 - \frac{1}{2}n_0^2)/|V|$ is more accurate, where n_0 is the number of particles in the ground state (the condensate).

To determine the potential energy density we use the expression (2.9). This gives

$$\bar{u}(\mu) = 2|\alpha|^2 c(\bar{n}(\mu) - |\alpha|^2) + c(\bar{n}(\mu) - |\alpha|^2)^2$$

$$+ c[\bar{n}(\mu)]^2 = 2c([\bar{n}(\mu)]^2 - \tfrac{1}{2}|\alpha|^4). \tag{4.2}$$

As $|\alpha|^2$ is interpreted as the density of the condensate, this is identical to Haung's suggestion.

With $\bar{u}(\mu)$ given by (4.2), and with $z_1 = \exp\beta(\gamma - 4c\rho)$ Eq. (2.12)–(2.14) can now be investigated. Firstly, we note that $\partial\bar{u}/\partial\hat{a}(p)$ is the same in this example as it was in Example 1, and so (2.12) becomes, as in (3.7)

$$\hat{a}(p) = z_1(e^{\beta p^2/2} - z_1)^{-1}. \tag{4.3}$$

The extra term in the potential (4.2) changes [(3.8) or (2.13)] to

$$|\alpha| \log z_1 = |\alpha|(-2c\beta|\alpha|^2); \tag{4.4}$$

(3.9) [or (2.14)] remains as

$$\rho = |\alpha|^2 + (2\pi\beta)^{-3/2} g_{3/2}(z_1). \tag{4.5}$$

Retaining the notation of Sec. 3, we see that for $\rho < \rho_c$, a solution of these is

$$|\alpha| = 0, \quad \rho = (2\pi\beta)^{-3/2} g_{3/2}(z_1). \tag{4.6}$$

But solutions with $|\alpha| \neq 0$ are also possible for some values of ρ, as we now demonstrate. From (4.4) if $|\alpha| \neq 0$, then $z_1 = \exp(-2c\beta|\alpha|^2)$. So from (4.5) $|\alpha|$ must satisfy

$$\rho - |\alpha|^2 = (2\pi\beta)^{-3/2} g_{3/2}(e^{-2c\beta|\alpha|^2}). \tag{4.7}$$

Graphical considerations (Fig. 1) show that for values of ρ near to but less than ρ_c, this equation has two solutions for $|\alpha|^2$. For $\rho > \rho_c$, (4.5) shows that we must have $|\alpha| > 0$, and so from (4.4) $z_1 = \exp(-2c\beta|\alpha|^2)$. Substituting back into (4.5) shows that $|\alpha|$ is determined by (4.7). From Fig. 1 this has a unique solution, and hence there is a unique solution to (4.3)–(4.5) for these values of ρ. A further immediate deduction from Fig. 1 is that when $\rho > \rho_c$, the solution $|\alpha|^2$ of (4.7) is greater than $\rho - \rho_c$; i.e., the density of the condensate is greater than $\rho - \rho_c$. This is in marked contrast to the free boson gas and to the model of Sec. 3.

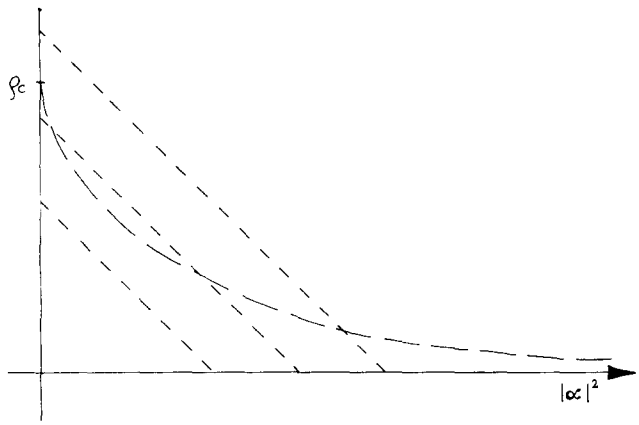


FIG. 1. $\rho - |\alpha|^2$ for three values of ρ(– –), plotted against $(2\pi\beta)^{-3/2} g_{3/2}(e^{-2c\beta|\alpha|^2})$(— —).

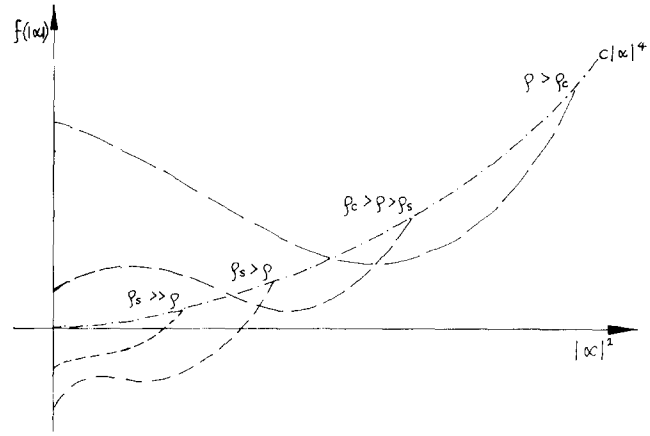


FIG. 2. The free energy density as a function of $|\alpha|^2$ for four values of ρ.

We now determine which of the solutions (4.6) or (4.7) is an absolute minimum of $\bar{f}(\mu)$. The expression (4.3) is substituted into $\bar{f}(\mu)$ (2.10) and the integrations performed. This results in a function f of the one variable $|\alpha|$:

$$f(|\alpha|) = c(2\rho^2 - |\alpha|^4) + \beta^{-1}\log z_1(\rho - |\alpha|^2) - (2\pi\beta)^{-3/2}\beta^{-1} g_{5/2}(z_1), \tag{4.8}$$

where z_1 depends on $|\alpha|$ through

$$\rho = |\alpha|^2 + (2\pi\beta)^{-3/2} g_{3/2}(z_1),$$

ρ is fixed, and $g_r(z) = \sum_{n=1}^{\infty} z^n/n^r$. Differentiating (4.8):

$$\frac{df}{d|\alpha|} = -4c|\alpha|^3 - \frac{2|\alpha|}{\beta}\log z_1,$$

$$\frac{d^2 f}{d|\alpha|^2} = -12c|\alpha|^2 - \frac{2}{\beta}\log z_1 - \frac{2|\alpha|}{\beta z_1}\frac{dz_1}{d|\alpha|}.$$

Thus the solution with $|\alpha| = 0$ is a local minimum, since

$$\left.\frac{d^2 f}{d|\alpha|^2}\right|_{\alpha=0} > 0.$$

Now assume that ρ is less than ρ_c but such that there are two nonzero solutions $|\alpha_1|, |\alpha_2|$ of (4.7). Assume further that $|\alpha_1| < |\alpha_2|$. We wish to compare $f(|\alpha_1|)$ and $f(|\alpha_2|)$. Substituting for z_1 in (4.8) gives

$$f(|\alpha_i|) = 2c\rho^2 + c|\alpha_i|^4 - 2c\rho|\alpha_i|^2 - (2\pi\beta)^{-3/2}\beta^{-1} \times g_{5/2}[\exp(-2\beta c|\alpha_i|^2)].$$

Let

$$g(x) = 2c\rho^2 + cx^4 - 2c\rho x^2 - (2\pi\beta)^{-3/2}\beta^{-1} g_{5/2}(e^{-2\beta c x^2}),$$

then $g(|\alpha_i|) = f(|\alpha_i|)$ and

$$\frac{dg}{dx} = -4cx[\rho - x^2 - (2\pi\beta)^{-3/2} g_{3/2}(e^{-2c\beta x^2})].$$

But is is clear from Fig. 1 that if $|\alpha_1| < x < |\alpha_2|$, then $(\rho - x^2) > (2\pi\beta)^{-3/2} g_{3/2}(e^{-2c\beta x^2})$. So for x in this range, $dg/dx < 0$. Thus,

$$f(|\alpha_1|) = g(|\alpha_1|) > g(|\alpha_2|) = f(|\alpha_2|).$$

Hence $|\alpha_1|$ is a local maximum of f, and $|\alpha_2|$ is a local minimum (as $|\alpha|^2 \to \rho$ so $z_1 \to 0$ and $df/d|\alpha| \to \infty$). Thus the minimizing solution of (4.3)–(4.5) is either the one with $\alpha = 0$, or the one with $|\alpha| = |\alpha_2|$ depending on whether $f(0)$ or $f(|\alpha_2|)$ is the smaller.

Let ρ_s be the value of ρ at which $f(0) = f(|\alpha_2|)$. Such a value exists since at the threshold [i.e., the smallest value of ρ for which (4.7) has a solution], we have $|\alpha_1| = |\alpha_2|$ and so $f(0) \leqslant f(|\alpha_1|) = f(|\alpha_2|)$, whereas at $\rho = \rho_c, \alpha_1 = 0$ and so $f(0) = f(|\alpha_1|) \geqslant f(|\alpha_2|)$. It can be shown that ρ_s is unique, so that for all $\rho < \rho_s$ the minimum is $f(0)$, and for all $\rho > \rho_s$ the minimum is $f(|\alpha_2|)$. (See Fig. 2.)

We now summarize this. There is a critical density ρ_s such that for $\rho < \rho_s$ the state is given by

$$\mu(h) = e^{-\|h\|^2/4} e^{-\langle \cdots \rangle/2} e^{\langle h, A(z_1)h\rangle/2},$$

where z_1 is the unique solution of $\rho = (2\pi\beta)^{-3/2} g_{3/2}(z_1)$. For $\rho > \rho_s$ the (gauge invariant) minimizing state is

$$\mu(h) = e^{-\|h\|^2/4} J_0(\sqrt{2}|\alpha||\hat{h}(0)|) e^{-\langle h, A(z_1)h\rangle/2},$$

where $z_1 = e^{-2c\beta|\alpha|^2}$ and $|\alpha|$ is the larger solution of (4.7).

If ρ_s is to be interpreted as a critical density at which a phase transition occurs then it is evident that this transition manifests itself in a rather dramatic way. The amount of condensate changes discontinuously from 0 to $|\alpha_2|^2$. Other interpretations using the two local minima of the free energy density, and the two critical densities ρ_c and ρ_s are also possible, but this uncertainty suggests that caution should be exercised in the interpretation of approximate states.

ACKNOWLEDGMENTS

The author wishes to acknowledge many helpful discussions with Professor J.T. Lewis, and suggestions from Dr. E.B. Davies.

[1]D.W. Robinson, *Thermodynamic Pressure in Quantum Statistical Mechanics*, Springer Lecture Notes in Physics, Vol. 9 (Springer-Verlag, Berlin, 1971).

[2]A.S. Ŝvarc, Trans. Mosc. Math. Soc. **22**, (1970) [A.M.S. translation] (1972)].
[3]H. Araki, J. Math. Phys. **1**, 492 (1960).
[4]I.E. Segal, Can. J. Math. **13** (1961).
[5]D.W. Robinson, Commun. Math. Phys. **1**, 89 (1965).
[6]H. Araki and E.J. Woods, J. Math. Phys. **4**, 637 (1963).
[7]O.E. Lanford and D.W. Robinson, Commun. Math. Phys. **24**, 193 (1972).
[8]J.T. Lewis and J.V. Pulè, Commun. Math. Phys. **36**, 1 (1974).
[9]R.H. Critchley and J.T. Lewis, Commun. Math. Phys. **44**, 107 (1975).
[10]J.R. Johnston, Am. J. Phys. **38**, 516 (1970).
[11]O. Lanford and D.W. Robinson, unpublished work (1971).
[12]M. Fannes, preprint (Leuven, 1976).
[13]R.H. Critchley and J.T. Lewis, "The Entropy Density of Quasi-Free States", D.I.A.S. preprint (1975)
[14]M. Fannes Commun. Math. Phys. **31**, 279 (1973).
[15]E.B. Davies, Commun. Math. Phys. **28**, 69 (1972).
[16]J.L. Lebowitz and O. Penrose, J. Math. Phys. **7**, 98 (1966).
[17]J.T. Cannon, Commun. Math. Phys. **29**, 89 (1973).
[18]K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).

Geometrical gauge conditions in Yang–Mills theory: Some nonexistence results

Alan Chodos[a] and Vincent Moncrief[b]

Department of Physics, Yale University, New Haven, Connecticut 06520

We investigate the possibility of defining an "orthogonal" gauge for non-Abelian Yang–Mills theory. Such a gauge would generalize, in a geometrical way, the orthogonality features of the Coulomb gauge in electrodynamics. We show however that such a gauge does not exist (even locally) in the non-Abelian case. Specifically we prove that the tangent spaces defined at every point by orthogonality to the gauge group orbits admit no integral submanifolds. We also study the question of existence of global gauge conditions in phase space. We show that such global gauges, should they exist, would induce globally defined gauges in configuration space contradicting Singer's result. We thus conclude (modulo certain technical points which might render Singer's argument inapplicable to our function spaces) that global gauge conditions in phase space do not exist.

I. INTRODUCTION

Coulomb gauge plays a special role in electrodynamics. The gauge condition $\nabla\cdot\tilde{a} = 0$ serves to eliminate precisely the gauge variant part of the vector potential $\tilde{a}$, leaving only the physical transverse degrees of freedom in the theory. While Coulomb gauge may not always be the most convenient for calculation, it tends to be the most useful for the purpose of laying bare the structure of the theory.

To understand why Coulomb gauge is thus distinguished, let us consider the configuration space of gauge potentials $\tilde{a}$. At any point of this manifold, the directions tangent to the gauge orbits are given by the gradients $\nabla\alpha$ of functions α obeying suitable asymptotic conditions. Now a gauge condition determines a surface in the configuration manifold, and a geometrical way of defining such a surface and thus isolating the gauge variant part of the potential is to require orthogonality (in a suitable inner product) of the surface to the gauge orbits which intersect it. That is, if $\tilde{t}$ is tangent to the surface at the point $\tilde{a}$, we require

$$\langle \nabla\alpha, \tilde{t} \rangle = \sum_i \int_{\mathbf{R}^3} d^3x \, \nabla_i \alpha(\mathbf{x}) t^i(\mathbf{x}) = 0 \qquad \text{(I.1)}$$

for all suitably behaved α. This clearly implies that $\nabla\cdot\tilde{t} = 0$, i.e., that the tangent space to the surface at any point consists of the transverse subspace of the full tangent space. Any of the family of surfaces defined by

$$\nabla\cdot\tilde{a} - f = 0, \qquad \text{(I.2)}$$

where f is a fixed scalar field, has this property. We can specialize to the usual Coulomb gauge by requiring in addition that the surface pass through $\tilde{a} = 0$.

The purpose of this paper is to see whether this line of reasoning generalizes to the non-Abelian case. Our objective is to define, at least in some local region of configuration space, a submanifold (of the appropriate dimensions) whose tangent space at any point is orthogonal to the gauge group orbit through that point. Such a surface would generalize, in a more geometrical way, the Coulomb gauge condition to the non-Abelian case. One knows from the work of Gribov[1] and Singer[2] that a global, continuous gauge condition does not exist. However, as Singer suggests, one might be able to patch together several local gauge conditions to define a (not everywhere continuous) global one. Our proposed generalization of the Coulomb gauge might then seem to be a natural candidate for the needed local gauge conditions.

We shall show however that no such gauge condition exists (even locally) anywhere in configuration space. Specifically we shall show that the subspaces of the tangent spaces which are orthogonal to the gauge group orbits at every point do not admit any integral submanifolds. The geometrical basis for this argument (in a suitable function space setting) is discussed in Sec. II. The nonexistence argument is given in Sec. III. Section IV discusses the possibility of global gauge conditions in phase space and shows that their existence would imply the existence of global gauge conditions in configuration space, contradicting Singer's proof. We are led to conclude (modulo certain technical assumptions which might render Singer's argument inapplicable to our function spaces) that global gauge conditions in phase space do not exist.

While writing this paper, we became aware of work by Creutz, Muzinich, and Tudron[3] in which, among other things, they discuss a special case of the problem considered here. They show that there is no local orthogonal gauge condition through $\tilde{a} = 0$. Our results show that the point $\tilde{a} = 0$ is not exceptional and further avoids using the assumption of a "function" defining the hypothetical gauge-orthogonal surface. An earlier study of this same question (pointed out to us by Michael Creutz) was made by Treat.[4] His work made use of certain elliptic operators which, as Gribov subsequently showed, fail to have globally defined inverses. We avoid using such assumptions and thus extend this earlier work.

[a]Research supported in part (Yale Report #C00-3075-225) by the U.S. Department of Energy under Contract No. EY-76-C-02-3075.
[b]Research supported in part by an NSF Grant to Yale, PHY76-82353.

 0022-2488/80/020364-08$1.00

II. GEOMETRY OF CONFIGURATION SPACE

Let G be a g-dimensional, compact, semisimple Lie group and let $\{\theta_a\}$, with $a = 1,2,...,g$, be the Hermitian generators of a representation of G for which

$$[\theta_a, \theta_b] = if^{abc}\theta_c \tag{II.1}$$

and

$$\mathrm{tr}\theta_a\theta_b = k\delta_{ab} \tag{II.2}$$

for some constant k. The structure constants f^{abc} may be assumed completely antisymmetric.

A suitable configuration space $\mathscr{A}$ for our purposes would be a Hilbert or Banach space of vector potentials $\tilde{a} = a_i^{(a)}(\mathbf{x})\,\theta_a\,dx^i$ obeying

$$\mathrm{tr}\int_{\mathbf{R}^3} dx^3 \tilde{a}_i\,\tilde{a}_i = k\int_{\mathbf{R}^3} dx^3\, a_i^{(a)}\,a_i^{(a)} < \infty. \tag{II.3}$$

Some specific choices for $\mathscr{A}$ will be discussed below.

The main reason we want square integrable potentials is that we wish to define the Riemannian metric $\mathfrak{G}(\ ,\ )$ on $\mathscr{A}$ by setting, for any pair of tangent vectors $\tilde{t}$, and $\tilde{t}'$ at $\tilde{a}$,

$$\mathfrak{G}(\tilde{t},\tilde{t}') = \mathrm{tr}\int_{\mathbf{R}^3} dx^3\,\tilde{t}_i\,\tilde{t}_i' \equiv \langle\tilde{t},\tilde{t}'\rangle_{\tilde{a}}. \tag{II.4}$$

Here $\tilde{t}$ and $\tilde{t}' \in T_{\tilde{a}}\mathscr{A} \approx \mathscr{A}$ so we require $\mathscr{A} \subset L_2$ for $\mathfrak{G}(\ ,\ )$ to be well defined.

The metric $\mathfrak{G}(\ ,\ )$ is the natural metric defined by the kinetic energy term in the Yang–Mills Lagrangian. It has the important property that every guage transformation is an isometry (or symmetry transformation) of it. To see this, we let $\tilde{a}(\lambda)$ be an arbitrary smooth curve of vector potentials and $\tilde{a}_U(\lambda)$ be the transformed curve under a fixed gauge transformation U:

$$\tilde{a}_U(\lambda) = U\tilde{a}(\lambda)U^{-1} + iU(dU^{-1}), \tag{II.5}$$

where

$$(dU^{-1}) = (\partial_i U^{-1}dx^i). \tag{II.6}$$

Differentiating with respect to λ, we get the transformation law for tangent vectors

$$\tilde{t}_U = \left(\frac{d}{d\lambda}\tilde{a}_U(\lambda)\right)\Big|_{\lambda=0} = U\left(\frac{d\tilde{a}(\lambda)}{d\lambda}\right)\Big|_{\lambda=0}U^{-1} = U\tilde{t}U^{-1}$$

or

$$\tilde{t}_U = U\tilde{t}U^{-1}. \tag{II.7}$$

It follows that

$$\begin{aligned}\langle\tilde{t}_U,\tilde{t}'_U\rangle_{\tilde{a}_U} &= \mathrm{tr}\int_{\mathbf{R}^3} d^3x\,(\tilde{t}_U)_i(\tilde{t}'_U)_i \\ &= \mathrm{tr}\int_{\mathbf{R}^3} d^3x\,(U\tilde{t}_iU^{-1}U\tilde{t}_i'U^{-1}) \\ &= \mathrm{tr}\int_{\mathbf{R}^3} d^3x\,\tilde{t}_i\tilde{t}_i' \\ &= \langle\tilde{t},\tilde{t}'\rangle_{\tilde{a}}.\end{aligned} \tag{II.8}$$

Thus for arbitrary $\tilde{a}$ and arbitrary gauge transformation U acting on $\mathscr{A}$ we have, for an $\tilde{t}$ and $\tilde{t}' \in T_{\tilde{a}}\mathscr{A}$,

$$\langle\tilde{t}_U,\tilde{t}'_U\rangle_{\tilde{a}_U} = \langle\tilde{t},\tilde{t}'\rangle_{\tilde{a}}, \tag{II.9}$$

which is the isometry property of $\mathfrak{G}(\ ,\ )$. Many other metrics can be constructed which also have this symmetry but $\mathfrak{G}(\ ,\ )$ is singled out by the Yang–Mills Lagrangian.

Suppose Σ is a submanifold of $\mathscr{A}$ with the property that its tangent space $T_{\tilde{a}}\Sigma$ at any point $\tilde{a}\in\Sigma$ is orthogonal to the gauge group orbit $\mathscr{O}_{\tilde{a}}$ through $\tilde{a}$. We define such a submanifold to be *gauge-orthogonal*. Then we claim that, for any $U\in\mathscr{G}$ = the gauge group acting on $\mathscr{A}$, the set Σ_U defined by

$$\Sigma_U = \{\tilde{a}_U\in\mathscr{A} \mid \tilde{a}_U = U\tilde{a}U^{-1} + iUdU^{-1} \text{ for some } \tilde{a}\in\Sigma\} \tag{II.10}$$

is also gauge-orthogonal. Here the orbit $\mathscr{O}_{\tilde{a}}$ is defined by

$$\mathscr{O}_{\tilde{a}} = \{\tilde{a}'\in\mathscr{A} \mid \tilde{a}' = U\tilde{a}U^{-1} + iUdU^{-1} \text{ for some } U\in\mathscr{G}\}, \tag{II.11}$$

and orthogonality means that for any $\tilde{t}_\Sigma\in T_{\tilde{a}}\Sigma$ and any $\tilde{t}_\omega\in T_{\tilde{a}}\mathscr{O}_{\tilde{a}}$ we have

$$\mathfrak{G}(\tilde{t}_\Sigma,\tilde{t}_\omega) = 0. \tag{II.12}$$

The tangent space to the orbit $T_{\tilde{a}}\mathscr{O}_{\tilde{a}}$ is given explicitly by the infinitesimal gauge transformations of $\tilde{a}$

$$T_{\tilde{a}}\mathscr{O}_{\tilde{a}} = \{\tilde{t}_\omega\in T_{\tilde{a}}\mathscr{A} \mid \tilde{t}_\omega = d\tilde{\omega} + i[\tilde{\omega},\tilde{a}] \text{ for some } \tilde{\omega}\in\Omega\}. \tag{II.13}$$

Here Ω is a space of Lie algebra valued functions $\tilde{\omega} = \omega^{(a)}(\mathbf{x})\theta_a$ with suitable asymptotic conditions (discussed below). The orthogonality condition is equivalent to

$$\begin{aligned}0 &= \int_{\mathbf{R}^3} d^3x\, t_i^{(a)}(\omega_{,i}^{(a)} - f^{abc}\omega^b a_i^{(c)}) \\ &= \int_{\mathbf{R}^3} d^3x\,[\partial_i(t_i^{(a)}\omega^{(a)}) - \omega^{(a)}(\partial_i t_i^{(a)} - f^{abc}t_i^{(b)}a_i^{(c)})]\end{aligned} \tag{II.14}$$

for arbitrary $\tilde{\omega}\in\Omega$. For the asymptotic conditions we shall use, the divergence term drops out, giving

$$\partial_i t_i^{(a)} - f^{abc}t_i^{(b)}a_i^{(c)} = 0. \tag{II.15}$$

Thus $\tilde{t}\in T_{\tilde{a}}\mathscr{A}$ is orthogonal to $\mathscr{O}_{\tilde{a}}$ iff it has vanishing covariant divergence

$$\nabla_{\tilde{a}}\cdot\tilde{t} \equiv \delta\cdot\tilde{t} + i[\tilde{t}\cdot,\tilde{a}] = 0 \tag{II.16}$$

(where $\delta\cdot\tilde{t} \equiv \partial_i\tilde{t}_i$ and $[\tilde{t}\cdot,\tilde{a}] \equiv [\tilde{t}_i,\tilde{a}_i]$).

We can demonstrate explicitly that Σ_U is gauge-orthogonal, as claimed above. Let $\tilde{t}_U$ be a tangent vector to Σ_U at some point $\tilde{a}_U\in\Sigma_U$ and let $\tilde{h}_U(\tilde{\omega})$ be an element of $T_{\tilde{a}_U}\mathscr{O}_{\tilde{a}_U}$. Then $\tilde{h}_U(\tilde{\omega})$ has the form

$$\tilde{h}_U(\tilde{\omega}) = d\tilde{\omega} + i[\tilde{\omega},\tilde{a}_U] \tag{II.17}$$

for some $\tilde{\omega}\in\Omega$. We claim that

$$\langle\tilde{h}_U(\tilde{\omega}),\tilde{t}_U\rangle_{\tilde{a}_U} = 0 \tag{II.18}$$

for any such $\tilde{h}_U(\tilde{\omega})$, $\tilde{t}_U$, and

$$\tilde{a}_U = U\tilde{a}U^{-1} + iUdU^{-1}. \tag{II.19}$$

This follows from noting that

$$\begin{aligned}\tilde{h}_U(\tilde{\omega}) &= U\{d(U^{-1}\tilde{\omega}U) + i[(U^{-1}\tilde{\omega}U),\tilde{a}]\}U^{-1} \\ &= U(\tilde{h}(U^{-1}\tilde{\omega}U))U^{-1},\end{aligned} \tag{II.20}$$

i.e., that $\tilde{h}_U(\tilde{\omega})$ is the transform of an element $\tilde{h}(U^{-1}\tilde{\omega}U)\in T_{\tilde{a}}\mathscr{O}_{\tilde{a}}$. Furthermore, $\tilde{t}_U$ is the transform of some $\tilde{t}\in T_{\tilde{a}}\Sigma$. To see this, let $\tilde{a}_U(\lambda)$ be any curve in Σ_U which

induces

$$\tilde{t}_U = \left(\frac{d}{d\lambda} \tilde{a}_U(\lambda) \right)\Big|_{\lambda=0} \quad \text{at} \quad \tilde{a}_U = \tilde{a}_U(0). \tag{II.21}$$

Then, by the definition of Σ_U, it follows that

$$\tilde{a}_U(\lambda) = U\tilde{a}(\lambda)U^{-1} + iU(dU^{-1}), \tag{II.22}$$

where $\tilde{a}(\lambda)$ is a curve in Σ with $\tilde{a}(0) = \tilde{a}$. Differentiating with respect to λ gives

$$\tilde{t}_U = U\tilde{t}U^{-1}, \tag{II.23}$$

where $\tilde{t} \in T_{\tilde{a}}\Sigma$. Since $T_{\tilde{a}}\Sigma$ is orthogonal to $T_{\tilde{a}}\mathcal{O}_{\tilde{a}}$ for any $\tilde{a} \in \Sigma$ by assumption, we have

$$\langle \tilde{t}, \tilde{h}(U^{-1}\tilde{\omega}U) \rangle_{\tilde{a}} = 0. \tag{II.24}$$

From the transformation properties of $\tilde{t}$ and $\tilde{h}$ derived above and the isometry property of $\mathfrak{G}(\ ,\ )$ under gauge transformations it follows that

$$\langle \tilde{t}_U, \tilde{h}_U(\tilde{\omega}) \rangle_{\tilde{a}_t} = \langle \tilde{t}, \tilde{h}(U^{-1}\tilde{\omega}U) \rangle_{\tilde{a}} = 0, \tag{II.25}$$

proving the claim made in Eq. (II.18). Intuitively Σ_U is geometrically equivalent to Σ since it is just the image of Σ under an isometry transformation of $\mathfrak{G}(\ ,\ )$.

So far we have assumed nothing about the dimensionality of Σ. It is easy to show that at least one-dimensional examples of gauge-orthogonal submanifolds exist. One can simply take integral curves of any of the invariant vector fields which we consider below. However, we are interested here only in the possible existence of *maximal* gauge-orthogonal submanifolds, i.e., submanifolds Σ whose tangent space $T_{\tilde{a}}\Sigma$ at any point coincides with the subspace of $T_{\tilde{a}}\mathcal{A}$, which is orthogonal to the orbit through $\tilde{a}$. Recalling Eq. (II.16), we see that such a submanifold would have

$$T_{\tilde{a}}\Sigma = \{ \tilde{t} \in T_{\tilde{a}}\mathcal{A} \mid \nabla_{\tilde{a}} \cdot \tilde{t} = 0 \}$$

at each $\tilde{a} \in \Sigma$. Our main result will be to show that such submanifolds do not exist.

Suppose that a maximal gauge-orthogonal submanifold did exist (we shall henceforth let Σ represent such a hypothetical submanifold). Then Σ would define, at least locally, a choice of gauge on $\mathcal{A}$. That is, for any $\tilde{a} \in \Sigma$ there would be a neighborhood $\mathcal{N}(\tilde{a})$ (in $\mathcal{A}$) of $\tilde{a}$ in which each potential is gauge equivalent to a unique potential in $\mathcal{N}(\tilde{a}) \cap \Sigma$. In geometrical terms Σ would be a "slice" for the group action of $\mathcal{G}$ on $\mathcal{A}$ or, if we view $\mathcal{A}$ as a principal fiber bundle over the orbit space $\mathcal{A}/\mathcal{G}$, then Σ would be a local cross section of $\mathcal{A}$. (We remark that the asymptotic conditions discussed below ensure that $\mathcal{G}$ acts freely on $\mathcal{A}$ and thus that $\mathcal{A}/\mathcal{G}$ would in fact be a manifold.)

Let $\mathcal{N}(\tilde{a})$ be as above and $\tilde{a}'$ be any element of $\mathcal{N}(a)$. Then $\tilde{a}'$ is gauge equivalent to some unique $\tilde{a} \in \Sigma$ so we can write

$$\tilde{a}' = \tilde{a}_U = U\tilde{a}U^{-1} + iU(dU^{-1}) \tag{II.26}$$

for some $U \in \mathcal{G}$. Now define Σ_U as before and conclude that there is a uniquely determined gauge-orthogonal slice through each point of $\mathcal{N}(\tilde{a})$. Thus the existence of a guage-orthogonal slice through $\tilde{a}$ would imply the existence of a geometrically equivalent slice through each point in some neighborhood of $\tilde{a}$.

The main idea of our nonexistence argument is now easy to outline. We shall construct a set of vector fields over $\mathcal{A}$, which are everywhere orthogonal to the orbits of $\mathcal{G}$. On any neighborhood $\mathcal{N}(\tilde{a})$ with the property discussed above each of these vector fields would be tangent to the gauge-orthogonal slices defined through the points of $\mathcal{N}(\tilde{a})$. If follows that the Lie bracket of any pair of such vector fields would have to be tangent to the (hypothetical) slice defined through any point of $\mathcal{N}(\tilde{a})$ and thus orthogonal to the orbit of $\mathcal{G}$ through that point. We shall, however, construct a set of vector fields which are everywhere gauge-orthogonal but whose Lie brackets are not all gauge-orthogonal on any open set of $\mathcal{A}$. This will show that no slices orthogonal to $\mathcal{G}$ exist anywhere in $\mathcal{A}$.

The basic property of Lie brackets alluded to above is discussed in the context of Banach manifolds by Lang.[5] The existence of a neighborhood $\mathcal{N}(\tilde{a})$ foliated by gauge-orthogonal slices Σ_U would define, in the sense of Lang, a subbundle of the tangent bundle of $\mathcal{N}(\tilde{a})$. Any pair of vector fields lying in that subbundle (i.e., tangent to the slice Σ_U through any given point) would have a Lie bracket lying in that subbundle (and thus tangent to the slice Σ_U through the given point). Since our hypothetical slices are orthogonal to $\mathcal{G}$, the Lie brackets of vector fields tangent to these slices would have to be orthogonal to $\mathcal{G}$. We shall see in the next section that this is impossible.

For our purposes the most convenient representation of a vector field $\tilde{X}$ on $\mathcal{A}$ is that of a linear, first-order functional differential operator acting on the space of real-valued functions $\mathcal{F}$ on $\mathcal{A}$

$$\mathcal{F} \to \tilde{X}\mathcal{F}.$$

Thus $\tilde{X}$ can be written in component form as

$$\tilde{X} = \int_{\mathbb{R}^3} d^3x \left[X_i^{(a)}(\tilde{a}, \mathbf{x}) \frac{\delta}{\delta a_i^{(a)}(\mathbf{x})} \right]. \tag{II.27}$$

The Lie bracket $[\ ,\ ]$ of any pair of such vector fields $\tilde{X}$, $\tilde{Y}$ is then simply the commutator of the two differential operators. (See Lang[5] for a coordinate free discussion of differential geometry on Banach manifolds.)

Given a smooth function $\mathcal{F}$, we can define the gradient (or differential) $d\mathcal{F}$ to have the "components" $\delta\mathcal{F}/\delta a_i^{(a)}(\mathbf{x})$. Using $\mathfrak{G}(\ ,\ )$, which in this formal notation has the components $\delta_{ab}\,\delta^{ij}\delta(\mathbf{x},\mathbf{x}')$, we have the associated vector field

$$\tilde{X}_{d\mathcal{F}} = \int_{\mathbb{R}^3} d^3x \left[\frac{\delta\mathcal{F}}{\delta a_i^{(a)}(\mathbf{x})} \frac{\delta}{\delta a_i^{(a)}(\mathbf{x})} \right]. \tag{II.28}$$

$\tilde{X}_{d\mathcal{F}}$ will be everywhere orthogonal to the orbits of $\mathcal{G}$ iff $\delta\mathcal{F}/\delta a_i^{(a)}$ obeys suitable differential and asymptotic conditions (so that it defines an element of $T_{\tilde{a}}\mathcal{A}$ at each $\tilde{a}$), and it obeys the orthogonality condition

$$\partial_i \left(\frac{\delta\mathcal{F}}{\delta a_i^{(a)}} \right) - f^{abc} \frac{\delta\mathcal{F}}{\delta a_i^{(b)}} a_i^{(c)} = 0. \tag{II.29}$$

Contracting this equation with an arbitrary $\tilde{\omega} \in \Omega$ and integrating over $\mathbb{R}^3$, we get

$$0 = \int_{\mathbb{R}^3} d^3x \left[\partial_i \left(\omega^{(a)} \frac{\delta\mathcal{F}}{\delta a_i^{(a)}} \right) - (\omega_{,i}^{(a)} - f^{abc}\omega^{(b)} a_i^{(c)}) \frac{\delta\mathcal{F}}{\delta a_i^{(a)}} \right]. \tag{II.30}$$

With appropriate asymptotic conditions (see below) the divergence term drops out so that the transversality condition reduces to the condition that $\mathscr{F}$ be invariant under arbitrary infinitesimal gauge transformations (induced by elements $\tilde{\omega}\in\Omega$). Conversely, the gradient of an invariant function gives rise to a vector field which obeys Eq. (II.29).

It is straightforward to construct a large class of gauge invariant function on $\mathscr{A}$, to derive their associated gradient vector fields, and to compute the Lie brackets of pairs of such vector fields. We shall do this in the next section for functions of the type

$$\mathscr{F}_s = \mathrm{tr}\int_{\mathbb{R}^3} d^3x\, s_{ij}(\mathbf{x})\mathscr{B}^i\mathscr{B}^j, \tag{II.31}$$

where $s_{ij}(\mathbf{x})$ is any smooth symmetric tensor field over $\mathbb{R}^3$ with suitable asymptotic behavior (it suffices to consider functions of compact support) and where $\widetilde{\mathscr{B}}$ is the magnetic field of $\tilde{a}$

$$\mathscr{B}^i = \tfrac{1}{2}\epsilon^{ijk}\widetilde{F}_{jk}, \tag{II.32}$$

$$\widetilde{F}_{jk} = \partial_j\tilde{a}_k - \partial_k\tilde{a}_j - i[\tilde{a}_j,\tilde{a}_k]. \tag{II.33}$$

Since $s_{ij}(\mathbf{x})$ is independent of $\tilde{a}$ by assumption, it is straightforward to show that $\mathscr{F}_s$ is gauge invariant.

In the computations to follow it will be convenient to have the following lemma on gauge transformations. We claim that for any $\tilde{a}\in\mathscr{A}$ and any $\mathbf{x}_0\in\mathbb{R}^3$ there is always a gauge transformation U such that

$$\tilde{a}_U(\mathbf{x}_0) = 0$$

i.e., one can always transform the gauge potential to zero at a fixed point in space. In fact this can be accomplished with a U of the form $U = \exp(i\omega^{(a)}\theta_a)$, where $(\omega^a)(\mathbf{x}_0) = 0$. To see this, we write out $\tilde{a}_U(\mathbf{x}_0)$ explicitly [putting $U(\mathbf{x}_0) = 1$]:

$$\begin{aligned}\tilde{a}_U(\mathbf{x}_0) &= U\tilde{a}U^{-1}(\mathbf{x}_0) + iU(dU^{-1})(\mathbf{x}_0)\\ &= \tilde{a}(\mathbf{x}_0) + d\tilde{\omega}(\mathbf{x}_0)\\ &= [a_i^{(a)}(\mathbf{x}_0) + \partial_i\omega^{(a)}(\mathbf{x}_0)]\theta_a\,dx^i\end{aligned} \tag{II.34}$$

and note that one can always choose $\tilde{\omega}$ such that $\partial_i\omega^{(a)}(\mathbf{x}_0) = -a_i^{(a)}(\mathbf{x}_0)$.

In a recent paper[6] one of us applied Cantor's[7,8] weighted Sobolev spaces of functions to the study of Gribov degeneracies. Such spaces have a number of convenient analytical properties and were used previously by Cantor[7,8] and by Fischer, Marsden, and Choquet-Bruhat[9] to solve several outstanding problems in general relativity. We shall adopt them here for our function spaces $\mathscr{A}$ and Ω.

Following Cantor we define the Banach space of functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ ($m = 3$ in our case) with the norm

$$\|f\|_{p,s,\delta} = \sum_{0\le\alpha\le s}\|\sigma^{\alpha+\delta}\,D^\alpha f\|_{L_p}, \tag{II.35}$$

where $1<p<\infty$, $\delta\in\mathbb{R}$, s is a nonnegative integer, and $\sigma(\mathbf{x}) = (1+|\mathbf{x}|^2)^{1/2}$ and $\|\ \|_{Lp}$ is the usual L_p norm on $\mathbb{R}^3$:

$$\|f\|_{L_p} = \left(\int_{\mathbb{R}^3} d^3x|f|^p\right)^{1/p}. \tag{II.36}$$

Following Cantor, we shall designate these spaces as $\mathscr{M}^p_{s,\delta}(\mathbb{R}^3,\mathbb{R}^n)$ or simply as $\mathscr{M}^p_{s,\delta}$. With a suitable choice of n, one can make $\mathscr{M}^p_{s,\delta}$ spaces of vector potentials $\mathscr{A}$ or Lie-algebra-valued functions Ω. The choice of p, s, δ assures certain differentiability and asymptotic properties. In particular the space $\mathscr{M}^p_{s,\delta}$ includes functions which behave asymptotically as

$$f\sim\frac{1}{|\mathbf{x}|^{\delta+\epsilon+(3/p)}},\quad Df\sim\frac{1}{|\mathbf{x}|^{\delta+1+\epsilon+(3/p)}},$$

$$\ldots, D^sf\sim\frac{1}{|\mathbf{x}|^{\delta+s+\epsilon+(3/p)}}, \tag{II.37}$$

where ϵ is an arbitrary number greater than zero and D^α designates the partial derivatives of order α.

The choices found most convenient in Ref. 6 were

$$\Omega = \mathscr{M}^p_{s,\delta},\quad \mathscr{A} = \mathscr{M}^p_{s-1,\delta+1}, \tag{II.38}$$

where

$$p>3,\quad \delta+3/p\ge\tfrac{1}{2},$$

$$s\ge3,\quad 0\le\delta<1-3/p.$$

These ensured the invertibility of certain elliptic operators such as the Laplacian and guaranteed the square integrability of the potentials. They also ensure that the integrals

$$\int_{\mathbb{R}^3} d^3x\,[\partial_i(\omega^{(a)}t_i^{(a)})]$$

vanish identically for $\tilde{\omega}\in\Omega = \mathscr{M}^p_{s,\delta}$ and $\tilde{t}\in T_{\tilde{a}}\mathscr{A}\approx\mathscr{M}^p_{s-1,\delta+1}$ since $(\omega^{(a)}t_i^{(a)})\sim 1/|\mathbf{x}|^{2+\epsilon'}$ for some $\epsilon'>0$. Though we do not require the inverse Laplacian in the present context, we shall retain [Eq. (II.38)] for convenience.

There is a technical difficulty associated with regarding the gradients of the invariant functions $\mathscr{F}_s$ and their Lie brackets as vector fields over $\mathscr{A}$. The variational derivatives $\delta\mathscr{F}_s/\delta\tilde{a}$ involve the second spatial derivatives of $\tilde{a}$ and so, for a generic $\tilde{a}\in\mathscr{A}$, are less differentiable than necessary to lie in $T_{\tilde{a}}\mathscr{A}\approx\mathscr{M}^p_{s-1,\delta+1}$. However, we can restrict the definition of these gradients to the dense subspaces ($k = 0,1,2,\cdots$)

$$\mathscr{D}_k = \mathscr{M}^p_{s+k,\delta+1}\subset\mathscr{M}^p_{s-1,\delta+1}. \tag{II.39}$$

On $\mathscr{D}_1$ the gradient fields $\widetilde{X}_{d\mathscr{F}_s}$ are densely defined vector fields, i.e., $\widetilde{X}_{d\mathscr{F}_s}(\tilde{a})\in T_{\tilde{a}}\mathscr{A}$ for all $\tilde{a}\in\mathscr{D}_1$. Similarly the Lie brackets

$$[\widetilde{X}_{d\mathscr{F}_s},\widetilde{X}_{d\mathscr{F}_{s'}}]$$

are defined on the dense domain $\mathscr{D}_3\subset\mathscr{A}$.

This necessity of restricting the domain of definition of "vector fields" to subspaces of the original configuration (or phase) space is a common problem in the Hamiltonian dynamics of infinite dimensional systems. It is discussed extensively by Chernoff and Marsden in Ref. 10. It arises there for the same reason that it occurs here; one "loses" derivatives in taking the gradient (or symplectic gradient) of some relevant function (e.g., the Hamiltonian).

A natural way of circumventing this complication (proposed to us by J. Marsden[11]) is to require that our hypothetical submanifolds Σ restrict to submanifolds of $\mathscr{D}_1$ and $\mathscr{D}_3$ as well, i.e., that $\Sigma\cap\mathscr{D}_1$ and $\Sigma\cap\mathscr{D}_3$ are submanifolds of $\mathscr{D}_1$ and $\mathscr{D}_3$, respectively. In this case one can show that the Lie brackets of the gradient vector fields we construct would have to be tangent to Σ at every point of $\Sigma\cap\mathscr{D}_3$, the dense domain on which these brackets are defined. By showing this

to be impossible we rule out the existence of such gauge-orthogonal slices.

The asymptotic behavior of the potentials required in Eq. (II.38) allows potentials which decay as $\tilde{a} \sim 1/|\mathbf{x}|^{3/2+\epsilon}$. This is a minimal requirement for the potentials to be square integrable. If we want $\mathscr{A}$ to be a vector space (rather than say an affine space obtained by "shifting" $\mathscr{M}^{p}_{s-1,\delta+1}$ by some fixed potential $\tilde{a}^*$ with different asymptotic behavior) on which the metric $\mathfrak{G}(\ ,\ )$ is well defined, we need this square integrability. Note, however, that this requirement excludes configurations with nonvanishing magnetic charge for which $\tilde{a} \sim 1/|\mathbf{x}|$ behavior would be needed. The reason such monopoles are not excluded *a priori* in Yang–Mills theory (by, say, the requirement of finite energy) is that the metric $\mathfrak{G}$ enters the Yang–Mills Hamiltonian only through the contraction of the electric fields. This contraction reduces to an inner product of velocities $\partial\tilde{a}/\partial t$ (i.e., tangent vectors in $\mathscr{A}$) only in a gauge with $\tilde{a}_0 = 0$. This $\mathfrak{G}(\ ,\ )$ is not, strictly speaking, a natural metric to define on $\mathscr{A}$ unless one demands that $\tilde{a}_0 = 0$ always be an allowed gauge condition.

III. THE NONEXISTENCE THEOREM

In this section we shall use the family of gauge invariant functionals

$$\mathscr{F}_s[\tilde{a}] = \sum_{a,i,j} \int_{\mathbb{R}^3} d^3x\, s_{ij}(\mathbf{x})\, \mathscr{B}_i^{(a)}(\mathbf{x})\, \mathscr{B}_j^{(a)}(\mathbf{x}), \tag{III.1}$$

where

$$\mathscr{B}_i^{(a)} \equiv \tfrac{1}{2}\epsilon^{ikl}(\partial_k\, a_l^{(a)} - \partial_l\, a_k^{(a)} + f^{abc}\, a_k^{(b)}\, a_l^{(c)}) \tag{III.2}$$

is the magnetic field and $s_{ij}(\mathbf{x})$ is an arbitrary smooth symmetric tensor field with compact support on $\mathbb{R}^3$. The vector field

$$\tilde{X}_{d\mathscr{F}_s} \equiv \tilde{X}_s \tag{III.3}$$

associated with each such functional is defined as in Eq. (II.28). We shall compute the Lie brackets

$$[\tilde{X}_s, \tilde{X}_{s'}] \equiv \tilde{X}_{ss'} \tag{III.4}$$

of pairs of such vector fields and show that they cannot be orthogonal to the orbits of $\mathscr{G}$ for all possible $s_{ij}(\mathbf{x})$ and $s'_{ij}(\mathbf{x})$ on any open set in functional space. This will show that no $\mathscr{G}$-orthogonal slices exist.

The Lie bracket of $\tilde{X}_s$ and $\tilde{X}_{s'}$ is given by

$$\begin{aligned}[\tilde{X}_s, \tilde{X}_{s'}] &= \int_{\mathbb{R}^3} d^3y \int_{\mathbb{R}^3} d^3x \left[\frac{\delta \mathscr{F}_s}{\delta a_i^{(a)}(\mathbf{x})} \frac{\delta^2 \mathscr{F}_{s'}}{\delta a_i^{(a)}(\mathbf{x})\, \delta a_j^{(b)}(\mathbf{y})} \right. \\ &\quad \left. - \frac{\delta \mathscr{F}_{s'}}{\delta a_i^{(a)}(\mathbf{x})} \frac{\delta^2 \mathscr{F}_s}{\delta a_i^{(a)}(\mathbf{x})\, \delta a_j^{(b)}(\mathbf{y})} \right] \frac{\delta}{\delta\, a_j^{(b)}(\mathbf{y})} \\ &= \int_{\mathbb{R}^3} d^3y\, (X_{ss'})_j^{(b)}(\mathbf{y}) \frac{\delta}{\delta\, a_j^{(b)}(\mathbf{y})}. \end{aligned} \tag{III.5}$$

Now we wish to test whether

$$0 = \partial_j\, (X_{ss'})_j^{(a)}(\mathbf{x}) + f^{abc}\, a_j^{(b)}(\mathbf{x})(X_{ss'})_j^{(c)}. \tag{III.6}$$

We know, from the gauge invariance of $\mathscr{F}_s$ and $\mathscr{F}_{s'}$, that

$$\partial_j \left(\frac{\delta \mathscr{F}_s}{\delta a_j^{(a)}(\mathbf{x})} \right) + f^{abc}\, a_j^{(b)}(\mathbf{x}) \frac{\delta \mathscr{F}_s}{\delta a_j^{(c)}(\mathbf{x})} = 0 \tag{III.7}$$

and similarly for $\mathscr{F}_{s'}$. Thus we have

$$\begin{aligned}&[\partial_j\, \delta^{ac} + f^{abc}\, a_j^{(b)}](\tilde{X}_{ss'})_j^{(c)}(\mathbf{x}) \\ &= \int_{\mathbb{R}^3} d^3y \frac{\delta \mathscr{F}_s}{\delta a_i^{(a)}(\mathbf{y})} \left[(\partial_j\, \delta^{ac} + f^{abc}\, a_j^{(b)}) \frac{\delta \mathscr{F}_{s'}}{\delta a_j^{(c)}(\mathbf{x})} \right] \\ &\quad - f^{abc} \frac{\delta \mathscr{F}_s}{\delta a_j^{(b)}(\mathbf{x})} \frac{\delta \mathscr{F}_{s'}}{\delta a_j^{(c)}(\mathbf{x})} - (\mathscr{F}_s \leftrightarrow \mathscr{F}_{s'}) \\ &= -2\, f^{abc} \frac{\delta \mathscr{F}_s}{\delta a_j^{(b)}(\mathbf{x})} \frac{\delta \mathscr{F}_{s'}}{\delta a_j^{(c)}(\mathbf{x})}. \end{aligned} \tag{III.8}$$

So, in order for a $\mathscr{G}$-orthogonal slice to exist through $\tilde{a}_0$, we must have

$$Q^{(a)}(\tilde{a}, \mathbf{x}) \equiv f^{abc} \frac{\delta \mathscr{F}_s}{\delta a_j^{(b)}(\mathbf{x})} \frac{\delta \mathscr{F}_{s'}}{\delta a_j^{(c)}(\mathbf{x})} = 0 \tag{III.9}$$

for all $\tilde{a}$ in some neighborhood of $\tilde{a}_0$ and for all choices of $\mathscr{F}_s$ and $\mathscr{F}_{s'}$. Note that we shall encounter no trouble in the Abelian case for which $f^{abc} = 0$. This accords with our knowledge that the surfaces defined by $\nabla \cdot \tilde{a} = f(\mathbf{x})$ are $\mathscr{G}$-orthogonal slices in that case.

Returning to the non-Abelian case, we evaluate

$$\frac{\delta \mathscr{F}_s}{\delta a_j^{(b)}(\mathbf{x})} = 2 \int_{\mathbb{R}^3} d^3y \left[s_{ik}(\mathbf{y}) \mathscr{B}_i^{(a)}(\mathbf{y}) \frac{\delta \mathscr{B}_k^{(a)}\mathbf{y}}{\delta a_j^{(b)}(\mathbf{x})} \right] \tag{III.10}$$

and find, with the definition [Eq. (III.2)] of $\tilde{\mathscr{B}}$, that

$$\frac{\delta \mathscr{F}_s}{\delta a_j^{(b)}(\mathbf{x})} = +\epsilon^{jmk} [\partial_m\, C_k^{(b)} + f^{bde}\, a_m^{(d)}\, C_k^{(e)}], \tag{III.11}$$

where

$$C_k^{(b)}(\mathbf{x}) \equiv s_{ki}(\mathbf{x}) \mathscr{B}_i^{(b)}(\mathbf{x}). \tag{III.12}$$

In order to simplify the succeeding steps, let us imagine that we are seeking to verify the equation

$$Q^{(a)}(\tilde{a}, \mathbf{x}) = 0 \tag{III.13}$$

at some point $\mathbf{x}$ for a particular choice of $\tilde{a}$. Since this equation must hold for all possible choices of $\tilde{s}$ and $\tilde{s}'$, we can choose them both to vanish at $\mathbf{x}$ and regard their derivatives $\partial_m\, s_{ij}(\mathbf{x})$ and $\partial_m\, s'_{ij}(\mathbf{x})$ as arbitrary parameters (symmetric in i and j). Then $Q^{(a)}(\tilde{a}, \mathbf{x}) = 0$ implies

$$f^{abc}\, \epsilon^{jmk}\, \epsilon^{jnl} (\partial_m s_{ki})(\partial_m\, s'_{lp})\, \mathscr{B}_i^{(b)}(\mathbf{x}) \mathscr{B}_p^{(c)}(\mathbf{x}) = 0. \tag{III.14}$$

Symmetrizing in the index pairs (ki) and (lp), we derive the requirement

$$\begin{aligned} 0 &= f^{abc}(\epsilon^{jmk}\, \epsilon^{jnl}\, \mathscr{B}_i^{(b)}\, \mathscr{B}_p^{(c)} + \epsilon^{jmk}\, \epsilon^{jnp}\, \mathscr{B}_i^{(b)}\, \mathscr{B}_l^{(c)} \\ &\quad + \epsilon^{jmi}\, \epsilon^{jnl}\, \mathscr{B}_k^{(b)}\, \mathscr{B}_p^{(c)} + \epsilon^{jmi}\, \epsilon^{jnp}\, \mathscr{B}_k^{(b)}\, \mathscr{B}_l^{(c)}), \\ &= f^{abc}[(\delta^{mn}\delta^{kl} - \delta^{ml}\delta^{kn}) \mathscr{B}_i^{(b)}\, \mathscr{B}_p^{(c)} \\ &\quad + (\delta^{mn}\delta^{kp} - \delta^{mp}\delta^{kn}) \mathscr{B}_i^{(b)}\, \mathscr{B}_l^{(c)} \\ &\quad + (\delta^{mn}\delta^{il} - \delta^{ml}\delta^{in}) \mathscr{B}_k^{(b)}\, \mathscr{B}_p^{(c)} \\ &\quad + (\delta^{mn}\delta^{ip} - \delta^{mp}\delta^{in}) \mathscr{B}_k^{(b)}\, \mathscr{B}_l^{(c)}]. \end{aligned} \tag{III.15}$$

Now let us choose, for example, $m = n = 1$, $k = l = 2$. we have

$$\begin{aligned} 0 &= f^{abc}[\mathscr{B}_i^{(b)}\, \mathscr{B}_p^{(c)} + \delta^{2p}\, \mathscr{B}_i^{(b)}\, \mathscr{B}_2^{(c)} + \delta^{i2}\, \mathscr{B}_2^{(b)}\, \mathscr{B}_p^{(c)} \\ &\quad + (\delta^{ip} - \delta^{1p}\delta^{1i}) \mathscr{B}_2^{(b)}\, \mathscr{B}_2^{(c)}]. \end{aligned} \tag{III.16}$$

The last term vanishes by the antisymmetry of f^{abc}. Since i and p are still at our disposal, we deduce

$$0 = f^{abc}\,\mathscr{B}_i^{(b)}\,\mathscr{B}_p^{(c)}, \quad i,p = 1,2,3. \tag{III.17}$$

So it remains to show that this condition cannot be satisfied everywhere throughout any neighborhood of the given point $\tilde{a}$. To do this, we shall show that if the condition holds at $\tilde{a}_0$, one can always find arbitrarily small perturbations of $\tilde{a}_0$ which violate the condition.

For this purpose we need the following

Lemma: It is always possible, by an appropriate perturbation $\delta a_i^{(a)}(\mathbf{x})$ of $a_i^{(a)}(\mathbf{x})$, to induce an arbitrary perturbation $\delta\mathscr{B}_i^{(a)}(\mathbf{x}_0)$ in $\mathscr{B}_i^{(a)}(\mathbf{x}_0)$ at some fixed point $\mathbf{x}_0$.

Proof: If we perform a gauge transformation with some fixed $U(\mathbf{x})$, then

$$\mathscr{B}'_i(\mathbf{x}) = U(\mathbf{x})\mathscr{B}_i(\mathbf{x})U^{-1}(\mathbf{x}) \tag{III.18}$$

and similarly

$$\delta\mathscr{B}'_i(\mathbf{x}) = U(\mathbf{x})\delta\mathscr{B}_i(\mathbf{x})U^{-1}(\mathbf{x}). \tag{III.19}$$

By the lemma at the end of Sec. II we can always choose U so that $a_i^{(a)}(\mathbf{x}_0) = 0$. In this gauge let us choose, near $\mathbf{x} = \mathbf{x}_0$

$$\delta a_i'^{(a)}(\mathbf{x}) = C_{ij}^{a}(x^j - x_0^j) \tag{III.20}$$

with $C_{ij}^{a} = -C_{ij}^{a}$. Then

$$\begin{aligned}\delta\mathscr{B}'^{(a)}_k(\mathbf{x}_0) &= \tfrac{1}{2}\epsilon_{kij}(\partial_i\delta a_j'^{(a)} - \partial_j\delta a_i'^{(a)})(\mathbf{x}_0)\\ &= -\epsilon_{kij}\,C_{ij}^{a}.\end{aligned} \tag{III.21}$$

Defining

$$C_{ij} \equiv C_{ij}^{a}\,\theta_a, \tag{III.22}$$

we have

$$\begin{aligned}C_{ij} &= -\tfrac{1}{2}\epsilon_{ijk}\delta\mathscr{B}'_k(\mathbf{x}_0)\\ &= -\tfrac{1}{2}\epsilon_{ijk}U(\mathbf{x}_0)\delta\mathscr{B}_k(\mathbf{x}_0)U^{-1}(\mathbf{x}_0).\end{aligned}$$

Since the constants C_{ij} can be chosen arbitrarily any desired value for $\delta\mathscr{B}_k(\mathbf{x}_0)$ can thus be achieved.

Now suppose that Eq. (III.13) is satisfied at some point $\tilde{a}_0$. We shall show that it cannot be satisfied throughout any neighborhood of $\tilde{a}_0$ by appropriately selecting $\delta\widetilde{\mathscr{B}}$.

There are two cases to be distinguished.

Case I: $f^{abc}\,\mathscr{B}_p^{(c)}(\mathbf{x})$ is not identically zero for all p and all $\mathbf{x}$. Suppose in particular that $f^{abc}\,\mathscr{B}_{p_0}^{(c)}(\mathbf{x}_0) \neq 0$. Then we have

$$0 = f^{abc}\,\mathscr{B}_i^{(b)}(\mathbf{x}_0)\mathscr{B}_{p_0}^{(c)}(\mathbf{x}_0) \tag{III.23}$$

and we let $\mathscr{B}_j^{(a)}(\mathbf{x}_0) \to \mathscr{B}_j^{(a)}(\mathbf{x}_0) + \delta\mathscr{B}_j^{(a)}(\mathbf{x}_0)$ where, by the lemma above, we can regard $\delta\mathscr{B}_j^{(a)}(\mathbf{x}_0)$ as arbitrary. If condition (III.13) is to be satisfied every where in a neighborhood of $\tilde{a}_0$, then we must have

$$0 = f^{abc}\,[\delta\mathscr{B}_i^{(b)}(\mathbf{x}_0)\mathscr{B}_{p_0}^{(c)}(\mathbf{x}_0) + \mathscr{B}_i^{(b)}(\mathbf{x}_0)\delta\mathscr{B}_{p_0}^{(c)}(\mathbf{x}_0)]. \tag{III.24}$$

Let us choose $\delta\mathscr{B}_{p_0}^{(c)}(\mathbf{x}_0) = 0$. Then $\delta\mathscr{B}_i^{(b)}(\mathbf{x}_0)$ is still arbitrary for $i \neq p_0$. Then the only way Eq. (III.24) can hold is to have

$$0 = f^{abc}\,\mathscr{B}_{p_0}^{(c)}(\mathbf{x}_0), \tag{III.25}$$

which contradicts our assumption.

Case II: $f^{abc}\,\mathscr{B}_p^{(c)}(\mathbf{x}) \equiv 0$. In this case the first order variations of Eq. (III.17) vanish so we must consider the second order variations which are

$$f^{abc}\,\delta\mathscr{B}_i^{(b)}(\mathbf{x}_0)\delta\mathscr{B}_p^{(c)}(\mathbf{x}_0).$$

For this to vanish for arbitrary $\delta\widetilde{\mathscr{B}}$, we must require $f^{abc} = 0$, which contradicts our assumption that the gauge group is non-Abelian.

This completes the proof that, for non-Abelian gauge theories, no $\mathscr{G}$-orthogonal slices exist even locally in function space. Roughly speaking, the orbits are sufficiently "twisted" in function space that no local slices can be orthogonal to every orbit which intersects it.

It is important to remember that the question of existence of orthogonal slices to a particular group action depends crucially upon the metric used to define orthogonality. Consider, for example, Euclidean 3-space in cylindrical coordinates for which

$$ds^2 = d\rho^2 + dz^2 + \rho^2 d\varphi^2. \tag{III.26}$$

The vector field

$$\widetilde{G} = \frac{1}{l}\frac{\partial}{\partial\varphi} + \frac{\partial}{\partial z}, \tag{III.27}$$

where l is a nonzero constant, defines a group action on this space whose orbits are the helical curves

$$\rho(\lambda) = \rho_0, \tag{III.28a}$$

$$z(\lambda) = z_0 + \lambda, \tag{III.28b}$$

$$\varphi(\lambda) = \varphi_0 + \lambda/l, \tag{III.28c}$$

where ρ_0, z_0, φ_0 are constants.

Since the metric is simultaneously invariant under both translations along, and rotations about, the z axis, it is invariant under the helical motion defined by their combination (i.e., the motion generated by $\widetilde{G}$). Any single orthogonal slice through a point p could therefore be dragged along the orbits of G to yield a neighborhood of p filled with orthogonal slices. We can prove, however, that such slices do not exist by displaying a set of vector fields orthogonal to G whose Lie brackets are not orthogonal to G throughout any open set in the space. For example, we can define the gradient fields of the invariant functions

$$h \equiv \tfrac{1}{2}\rho^2, \quad k_c \equiv \tfrac{1}{2}\rho^2\sin(\varphi - z/l + c), \tag{III.29}$$

where c is an adjustable constant. Computing these vector fields and their Lie brackets, one finds

$$[\widetilde{X}_{dh}, \widetilde{X}_{dk_c}]\cdot\widetilde{G} = -(\rho^2/l)\cos(\varphi - z/l + c). \tag{III.30}$$

This vanishes for all values of c only on the axis $\rho = 0$ and so not on any open set.

However, if we take the alternative metric

$$(ds')^2 = d\rho^2 + dz^2 + \rho^2(d\varphi - dz/l)^2, \tag{III.31}$$

then the coordinate transformation

$$\rho' = \rho, \quad z' = z, \quad \varphi' = \varphi - z/l \tag{III.32}$$

reduces $(ds')^2$ to Euclidean form and the "helical" curves to vertical lines. In this case the existence of orthogonal surfaces $z = \text{const}$ is obvious.

IV. CONCLUDING REMARKS

In this paper we have formulated a criterion, gauge-

orthogonality, by means of which we tried to find a Coulomb-like local gauge condition in Yang–Mills theory. We have shown that in configuration space, gauge orthogonal surfaces do not exist. In addition, we know from the work of Singer[2] that in configuration space no global gauges exist. The question arises as to whether if we work in phase space instead of configuration space, either or both of these restrictions can be circumvented.

It turns out that the question of gauge-orthogonality is clouded by the fact that no obviously natural metric exists on phase space, and, as we pointed out above, the choice of metric is crucial for orthogonality. Instead of pursuing this question we shall devote the remainder of this section to showing that since, as Singer has proved, there are no global sections on configuration space, then there cannot be any on phase space either.

The phase space for Yang–Mills theory may be regarded as a space of pairs $(\tilde{a},\tilde{e})$, where $\tilde{a}$ and $\tilde{e}$ are the potential and electric field respectively. We shall abbreviate this space as $T^*\mathscr{A}$ = the cotangent bundle of $\mathscr{A}$. Some specific models for $T^*\mathscr{A}$ using $\mathscr{M}^p_{s,\delta}$ spaces are discussed in Ref. 6.

The gauge group $\mathscr{G}$ acts on $T^*\mathscr{A}$ in a well-known way:

$$\tilde{a}\to U\tilde{a}U^{-1} + iUdU^{-1}, \tag{IV.1a}$$

$$\tilde{e} = U\tilde{e}U^{-1}. \tag{IV.1b}$$

If we consider only gauge transformations which reduce to the identity as $|\mathbf{x}|\to\infty$, we find, as in Ref. 6, that $\mathscr{G}$ acts freely on $T^*\mathscr{A}$ (i.e., no point $(\tilde{a},\tilde{e})$ remains fixed under any infinitesimal gauge transformation other than the trivial (identity) transformation). In this case we may regard $T^*\mathscr{A}$ as a principal fiber bundle over the quotient space $T^*\mathscr{A}/\mathscr{G}$, and it is natural to ask whether $T^*\mathscr{A}$ has any global cross sections.

Let us formulate this question more explicitly. Does $T^*\mathscr{A}$ admit a submanifold which intersects each of the orbits of $\mathscr{G}$ at one and only one point? The orbits of $\mathscr{G}$ in $T^*\mathscr{A}$ are, of course, just the equivalence classes of gauge equivalent pairs $(\tilde{a},\tilde{e})$ and the existence of a global section would be equivalent to a continuous global choice of gauge (to get the strongest possible result Singer assumes only continuous sections; for simplicity we shall look for differentiable ones).

The main idea for reducing the question of existence of global sections of $T^*\mathscr{A}$ to that of global sections of $\mathscr{A}$ is to note that the submanifold defined by $\tilde{e} = 0$ may be naturally identified with $\mathscr{A}$ itself and consists entirely of orbits of $\mathscr{G}$ acting in $T^*\mathscr{A}$ (i.e., all gauge transformations preserve $\tilde{e} = 0$). If Σ^* were a global section of $T^*\mathscr{A}$, then the intersection of Σ^* with $\mathscr{A}$ (regarded here as the submanifold $\tilde{e} = 0$ in $T^*\mathscr{A}$) would intersect each of the orbits of $\mathscr{A}$ at one and only one point. Therefore, we need only to show that $\Sigma^*\cap\mathscr{A}$ is a submanifold of $\mathscr{A}$ to conclude that it defines a global section of $\mathscr{A}$.

This last step follows from noting that the intersection of the two submanifolds Σ^* and $\mathscr{A}$ is transversal,[12] i.e., that at every point $(\tilde{a},0)$ of the intersection the tangent spaces $T_{(\tilde{a},0)}\Sigma^*$ and $T_{(\tilde{a},0)}\mathscr{A}$ together span $T_{(\tilde{a},0)}(T^*\mathscr{A})$. Formally at least this result follows from the assumption that Σ^* is a global cross section of $T^*\mathscr{A}$ and the observation that the base space $\mathscr{A}$ consists entirely of orbits of $\mathscr{G}$. Near any point $(\tilde{a},0)\in\Sigma^*\cap\mathscr{A}$, Σ^* provides a slice for the group action on $T^*\mathscr{A}$. Therefore, $T_{(\tilde{a},0)}(T^*\mathscr{A})$ is spanned by $T_{(\tilde{a},0)}\Sigma^*$ and the tangent space to the orbit through $(\tilde{a},0)$. However, $T_{(\tilde{a},0)}\mathscr{O}$, the tangent space of this orbit, is a subspace of $T_{(\tilde{a},0)}\mathscr{A}$ since $\mathscr{A}$ consists entirely of orbits of $\mathscr{G}$. Hence for every $(\tilde{a},0)\in\Sigma^*\cap\mathscr{A}$ we have

$$T_{(\tilde{a},0)}\Sigma^* + T_{(\tilde{a},0)}\mathscr{A} = T_{(\tilde{a},0)}(T^*\mathscr{A}),$$

and the intersection is transversal and therefore a manifold. Since the existence of such a manifold would contradict Singer's result, we must conclude that no global section of $T^*\mathscr{A}$ exists.

The above formal argument does not adequately treat the question of appropriate function spaces and asymptotic conditions. Singer considered C^∞ fields which can be compactified to S^3 while we have taken $\mathscr{M}^p_{s,\delta}$ spaces with asymptotic conditions on $\mathbb{R}^3$. Subtleties in the different treatment of asymptotic conditions might prohibit the application of Singer's theorem to our configuration space. Nonetheless, the above argument shows that the question of existence of a global section in $T^*\mathscr{A}$ is reducible to that of the existence of a global section in $\mathscr{A}$.

A modified form of the above question is to ask whether global cross sections of the constraint submanifold[6] of phase space exist. This question makes sense since the constraint subset is a submanifold of $T^*\mathscr{A}$ (see Ref. 6), which consists entirely of orbits of $\mathscr{G}$ (i.e., the constraints are preserved by gauge transformations). However, the constraint manifold contains the $\tilde{e} = 0$ space as a submanifold which may again be identified with the configuration space $\mathscr{A}$. The remainder of the argument is precisely as before, and we conclude that global sections are excluded since none exist for $\mathscr{A}$.

Another recent paper (pointed out to us by the referee) which discusses, for the SU(2) gauge theory, the nonexistence of global gauge conditions in phase space is that of Narasimhan and Ramadas.[13] They also derive some interesting geometrical results concerning the *connection* on the bundle $\mathscr{A}\to\mathscr{A}/\mathscr{G}$ defined by the constraint equations restricted to the $\tilde{a}_0^{(a)} = 0$ gauge. The horizontal subspace defined by this connection at any point of $\mathscr{A}$ coincides with what we have called the orthogonal subspace at the point.

ACKNOWLEDGMENT

We are grateful to Professor Jerrold Marsden for helpful discussions.

[1]V.N. Gribov, Nucl. Phys. B **139**, 1 (1978).
[2]I.M. Singer, Commun. Math. Phys. **60**, 7 (1978).
[3]M. Creutz, I.J. Muzinich, and T.N. Tudron, Phys. Rev. D **19**, 531 (1979).
[4]R. Treat, J. Math. Phys. **13**, 1704 (1972).
[5]S. Lang, *Differential Manifolds* (Addison-Wesley, Reading, Mass., 1972).
[6]V. Moncrief, J. Math. Phys. **20**, 579 (1979).
[7]M. Cantor, Ind. Univ. Math. J. **24**, 897 (1975).
[8]M. Cantor, "The existence of non-trivial asymptotically flat initial data for vacuum spacetimes," preprint, Univ. of Texas (1977).
[9]A. Fischer, J. Marsden, and Y. Choquet-Bruhat, "Maximal hypersurfaces and positivity of mass," preprint, Univ. of Paris and Univ. of California (1977).

[10]P. Chernoff and J. Marsden, *Properties of Infinite Dimensional Hamiltonian Systems*, Springer Lecture Notes #425 (Springer-Verlag, Berlin, 1974).
[11]J. Marsden (private communication).
[12]V. Guillemin and A. Pollack, *Differential Topology* (Prentice-Hall, Englewood Cliffs, N.J., 1974). See also Ref. 5, Chap. 2, for a more advanced treatment covering the infinite-dimensional case.
[13]M.S. Narasimhan and T.R. Ramados, "Geometry of SU(2) gauge fields," preprint, Tata Institute (1978).

The dominant partition method[a),b)]

Robert M. Dixon
Department of Physics, Morgan State University, Baltimore, Maryland 21239

Edward F. Redish[c)]
Laboratory for Astronomy and Solar Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20770 and Department of Physics and Astronomy, [d)] University of Maryland, College Park, Maryland 20742

(Received 17 April 1979; accepted for publication 20 September 1979)

Employing the L'Huillier, Redish, and Tandy (LRT) wave function formalism we develop a partially connected method for obtaining few-body reductions of the many-body problem in the LRT and Bencze, Redish, and Sloan (BRS) formalisms. This method for systematically constructing fewer body models for the N-body LRT and BRS equations is termed the dominant partition method (DPM). The DPM maps the many-body problem to a fewer-body one using the criterion that the truncated formalism must be such that consistency with the full Schrödinger equation is preserved. The DPM is based on a class of new forms for the irreducible cluster potential, introduced in the LRT formalism. Connectivity is maintained with respect to all partitions containing a given partition which is referred to as the dominant partition. Degrees of freedom corresponding to the breakup of one or more of the clusters of the dominant partition are treated in a disconnected manner. This approach for simplifying the complicated BRS equations is appropriate for physical problems where a few-body reaction mechanism prevails. We also show that the dominant-partition-truncated form of the BRS equations may be obtained by distributing the residual interaction in the exit channel in a manner consistent with the dominant partition truncations of the irreducible cluster potential.

I. INTRODUCTION

Connected kernel equations (CKE's) have enjoyed considerable prominence in the recent history of reaction theory. The equations due to Alt, Grassberger, and Sandhas[1] (AGS), Bencze, Redish, and Sloan[2] (BRS), and Kouri, Levin, and Tobocman[3] (KLT) are representative examples. In the formalisms of AGS, BRS, and KLT the many-body scattering problem is formulated in terms of a set of coupled integral equations for the transition operators. These CKE's are often viewed as extensions of the three-body formalism of Faddeev.[4]

Although the CKE's provide mathematically correct formulations of the N-body scattering problem, these equations have not inspired extensive usage in direct reaction analysis. The distorted wave Born approximation continues to be the primary method employed in the analyses of direct reactions. Significantly, there exists in the community an understanding that many-body effects should be included in reaction analysis;[5] however, the CKE's are not generally regarded as offering a viable approach for such inclusions. Even in the three-body case, the Faddeev equations are often regarded as useful for mathematical proofs but not as feasible for calculations.[6]

The complicated nature of CKE's as well as an uncertainty about how the dynamics is distributed in these equations have been important factors in limiting the role of CKE's in reaction analysis. Necessarily any use of CKE's must involve truncations. This compelling necessity for methods of truncating CKE's probably has contributed to their limited use.

The application of the CKE's to nuclear and atomic systems is rendered difficult because of the number of equations involved and the fact that many channels are treated theoretically on an equal footing. We therefore consider what simplification can be achieved by reducing the number of equations and/or channels. In practice, since certain channels may be ignored or treated phenomenologically, it may not be necessary to preserve connectedness in them. We call the formalism derived from relaxation of full connectedness in a set of CKE's a *partially connected formalism*. A partially connected approach has been advocated by Hahn and Watson as a means of circumventing the difficulty imposed by the coupling of all rearrangement channels in the three-body problem.[7]

The choice of criteria by which one truncates a CKE is an open question. A possible approach to simplifying these equations for some problems is to map the many-body space into that of a fewer-body problem. This approach will be useful in the case that the physics seems to be dominated by a few-body mechanism. One example is the deuteron-alpha scattering at energies below the threshold for breakup of the alpha. Notably, such a mapping does not destroy all of the many-body information which would be lost in the arbitrary imposition of a few-body model on a given many-body system.

Actually, when one considers such a truncation it becomes clear that many of the CKE's do not lend themselves to such a reduction method. The AGS and KLT equations

[a)] Work based on material submitted in partial fulfillment of requirement for Ph.D., University of Maryland, 1977.
[b)] Supported in part by U.S. Department of Energy.
[c)] N.R.C.-N.A.S. Senior Resident Research Associate.
[d)] Permanent address.

 0022-2488/80/020372-06$1.00

are notable examples. The explicit dependence on the number of particles as exhibited by the AGS equations and the dependence on the number of channels as exhibited by the KLT equations tend to make the structures of these equations rather rigid. The AGS equations make explicit the number of particles through kernels which contain all subsystem transition operators. The KLT equations are written for a fixed number of channels which structurally excludes the possibility of later dropping one of the channels.

In this article we develop a method of truncating the many-body BRS equations to a fewer-body problem. The method is developed in the LRT[8] connected-kernel wavefunction formalism, and is similar in spirit to the Hahn–Watson reduction method.[7] This truncation is termed the *Dominant Partition Method* (DPM). It maps the given many-body problem to a fewer-body problem whose solutions satisfy the full Schrödinger equation. The equations obtained constitute a partially connected set, the disconnectedness appearing in those channels which are not considered explicitly.

In Section II the LRT wavefunction formalism and the related irreducible cluster potential are reviewed. In Section III the dominant partition theorem is presented and in Section IV this reduction method is applied to the BRS equations. In Section IV it is also shown that the reduced set of BRS equations may be obtained via a distribution method. The summary and conclusion are presented in Section V.

II. THE L'HUILLIER, REDISH, TANDY WAVEFUNCTION FORMALISM

In this section a set of coupled connected kernel equations for the wave function describing the scattering between many-body ($N \geqslant 4$) clusters is obtained. These equations are derived by using the BRS equations and the Green function for the system. The system under consideration has N distinguishable particles which interact via two-body potentials. (The case of three- and many-body forces is not considered here.) A division of the N particles into n clusters is termed an a_n-partition. The Greek alphabet is used to label two-cluster partitions and the N-cluster partition is labelled 0. The partition Hamiltonians H_a, residual interactions V^a and associated Green functions are defined by

$$H_a = H_0 + V_a \,, \tag{1}$$

$$V^a = H - H_a \,, \tag{2}$$

$$G_a = (Z - H_a)^{-1} \,, \tag{3}$$

where V_a is the sum of two-body interactions internal to the a-partition and Z is the complex energy parameter ($Z = E + i\epsilon$). The full and free Green functions are given by

$$G = (Z - H)^{-1} \,, \tag{4}$$

$$G_0 = (Z - H_0)^{-1} \,, \tag{5}$$

where H is the full N-particle Hamiltonian and H_0 is the total kinetic energy operator.

Consider the N-body scattering problem initiated by incoming bound states of the two clusters comprising the partition β. The full wave functions is[9]

$$\Psi_\beta = \lim_{\epsilon \to 0} i\epsilon G \Phi_\beta \,, \tag{6}$$

where Φ_β describes a relative motion plane wave times the internal bound-state wavefunctions for the two clusters. The Green function is expressible in terms of G_β,

$$G = G_\beta + G V^\beta G_\beta \,. \tag{7}$$

Using that

$$G_\beta \Phi_\beta = (i\epsilon)^{-1} \Phi_\beta \,,$$

we obtain

$$\Psi_\beta = (1 + G V^\beta) \Phi_\beta \,. \tag{8}$$

Using (7) we have

$$\Psi_\beta = G G_\beta^{-1} \Phi_\beta \,. \tag{9}$$

Noting that

$$V^a \Psi_\beta = T^{a\beta} \Phi_\beta \,, \tag{10}$$

where $T^{a\beta}$ is a transition operator[10] we write (9) as

$$\Psi_\beta = G_0 G_\beta^{-1} \Phi_\beta + G_0 T^{0\beta} \Phi_\beta \,. \tag{11}$$

Employing the BRS equation[2]

$$T^{a\beta} = V_\beta^a + \sum_\sigma K_\sigma^{a0} G_0 T^{\sigma\beta} \,, \tag{12}$$

we obtain

$$\Psi_\beta = G_0 (G_\beta^{-1} + V_\beta) \Phi_\beta + \sum_\sigma G_0 K_\sigma^{00} G_0 T^{\sigma\beta} \Phi_\beta \,. \tag{13}$$

(The operator satisfying the BRS equation differs from the standard transition operators by an off-shell transformation. See Ref. 2.) In (12) V_β^a is the sum of two-body interactions internal to β and external to a. The kernel K_σ^{a0} is the sum of all Weinberg graphs[11] of connectivity[2] σ which begin with any interaction and do not end with an interaction in a. Defining $K_\sigma^{00} \equiv K_\sigma$, using (3) and (10) we have

$$\Psi_\beta = \Phi_\beta + \sum_\sigma G_0 K_\sigma G_0 V^\sigma \Psi_\beta \,. \tag{14}$$

This integral equation has a completely connected kernel. The operator K_σ is the sum of all σ-connected Weinberg graphs.

Decomposition of the wave function into parts associated with the two-cluster partitions of the N-body problem is achieved by writing

$$\Psi_\beta = \sum_\gamma \Psi_\beta^{(\gamma)} \,, \tag{15}$$

where

$$\Psi_\beta^{(\gamma)} = \Phi_\beta \delta_{\gamma\beta} + G_0 K_\gamma G_0 V^\gamma \Psi_\beta \,. \tag{16}$$

The wave function $\Psi_\beta^{(\gamma)}$ only has outgoing waves of the γ type, that is, bound clusters of the γ partition or direct breakup from the γ partition. Equation (16) may be written in a convenient form by considering again (9) and writing G in terms of G_γ. We find

$$\Psi_\beta = G_\gamma G_\beta^{-1} \Phi_\beta + G_\gamma V^\gamma G G_\beta^{-1} \Phi_\beta \,. \tag{17}$$

Using (9) and (10) we have

$$\Psi_\beta = G_\gamma G_\beta^{-1} \Phi_\beta + G_\gamma T^{\gamma\beta} \Phi_\beta \,. \tag{18}$$

Multiplying from the left by $G_0 G_\gamma^{-1}$ we have

$$G_0 G_\gamma^{-1} \Psi_\beta = G_0 G_\beta^{-1} \Phi_\beta + G_0 T^{\gamma\beta} \Phi_\beta \,. \tag{19}$$

Using (19) in (16) we get

$$\Psi_\beta^{(\gamma)} = (\delta_{\gamma\beta} - G_0 K_\gamma G_0 G_\beta^{-1}) \Phi_\beta + G_0 K_\gamma G_0 G_\gamma^{-1} \Psi_\beta \,. \tag{20}$$

We introduce the operator $\mathscr{V}_\gamma$ which is defined by

$$K_\gamma G_0 \equiv \mathscr{V}_\gamma G_\gamma \,. \tag{21}$$

The operator $\mathscr{V}_\gamma$ is termed the *irreducible γ connected potential*. It is the sum of γ connected graphs which become less than γ connected if the rightmost interaction is removed.[8] We have

$$\Psi_\beta^{(\gamma)} = (\delta_{\gamma\beta} - G_0 \mathscr{V}_\gamma G_\gamma G_\beta^{-1}) \Phi_\beta + G_0 \mathscr{V}_\gamma \Psi_\beta \,, \tag{22}$$

and $G_0 \mathscr{V}_\gamma G_\gamma G_\beta^{-1}$ on shell is the same as $G_0 \mathscr{V}_\beta \delta_{\beta\gamma}$. This yields

$$\Psi_\beta^{(\gamma)} = \delta_{\gamma\beta}(1 - G_0 \mathscr{V}_\beta) \Phi_\beta + G_0 \mathscr{V}_\gamma \Psi_\beta \,. \tag{23}$$

Noting that

$$(1 - G_0 \mathscr{V}_\beta) \Phi_\beta = 0 \,, \tag{24}$$

which is most easily seen on examining the anticluster expansion for $\mathscr{V}_\beta$ (Section III), we obtain

$$\Psi_\beta^{(\gamma)} = G_0 \mathscr{V}_\gamma \Psi_\beta \,. \tag{25}$$

We write this in differential form as

$$(E - H_0 - \mathscr{V}_\gamma) \Psi_\beta^{(\gamma)} = \mathscr{V}_\gamma \sum_{\sigma(\neq\gamma)} \Psi_\beta^{(\sigma)} \,, \tag{26}$$

which is reminiscent of the Faddeev three-body result. These are the LRT[8] equations. The relation of these equations to other N-body CKE's is discussed in detail in Ref. 12.

III. THE DOMINANT PARTITION THEOREM

The anticluster expansion for the kernel in the BRS equation has been previously shown to be[13]

$$K_\sigma^{\beta 0} G_0 = \sum_{m=2}^{N-1} \sum_{(\sigma \supset) a_m} N(\sigma, a_m) V_{a_m}^\beta G_{a_m} \,. \tag{27}$$

The N's in the above equation are termed counting coefficients.[13] They depend on both σ and a_m. From (21) we have the anticluster expansion for the irreducible cluster potential:

$$\mathscr{V}_\sigma = \sum_{m=2}^{N-1} \sum_{(\sigma \supset) a_m} N(\sigma, a_m) V_{a_m} G_{a_m} G_\sigma^{-1} \,. \tag{28}$$

Summing the components of equation (26) we note the interesting result that

$$(E - H_0 - \sum_\sigma \mathscr{V}_\sigma) \Psi_\beta = 0 \,. \tag{29}$$

This property provides the underpinning for the truncation of the BRS equations that will be presented in this section. Insight into the method is afforded by the following theorem.

Theorem I. For arbitrary N, if $\mathscr{V}_\sigma$ is given exactly then

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = V \Psi_\beta$$

where V is the full potential and Ψ_β solves the Schrödinger equation for the N-body system.

It is clear that this must be the case from a comparison between (29) and the Schrödinger equation

$$(E - H_0 - V) \psi_\beta = 0 \,.$$

Nonetheless, we give a more complicated proof whose structure generalizes to the approximate case discussed below.

Proof: From (28) we have

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = \sum_\sigma \sum_{m=2}^{N-1} \sum_{(\sigma \supset) a_m} N(\sigma, a_m) V_{a_m} G_{a_m} G_\sigma^{-1} \Psi_\beta \,. \tag{30}$$

Interchanging sums and using

$$G_\sigma^{-1} = G_{a_m}^{-1} - V_\sigma^{a_m}, \quad \sigma \supset a_m \,, \tag{31}$$

we have

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = \sum_{m=2}^{N-1} \sum_{a_m} \sum_{\sigma(\supset a_m)} N(\sigma, a_m) (V_{a_m} - V_{a_m} G_{a_m} V_\sigma^{a_m}) \Psi_\beta \,. \tag{32}$$

Noting that

$$V_\sigma^{a_m} = V_\sigma - V_{a_m} = V^{a_m} - V^\sigma, \quad \sigma \supset a_m \,, \tag{33}$$

we obtain

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = \sum_{m=2}^{N-1} \sum_{a_m} \sum_{\sigma(\supset a_m)} N(\sigma, a_m) V_{a_m} \times (1 - G_{a_m} V^{a_m} + G_{a_m} V^\sigma) \Psi_\beta \,. \tag{34}$$

We use the result,[14]

$$\Psi_\beta = \delta_{\beta a_m} \Phi_\beta + G_{a_m} V^{a_m} \Psi_\beta \tag{35}$$

to obtain

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = V_\beta \Phi_\beta + \sum_{m=2}^{N-1} \sum_{a_m} \sum_{(\supset a_m)} N(\sigma, a_m) V_{a_m} G_{a_m} V^\sigma \Psi_\beta \,. \tag{36}$$

It is known[15] that

$$\sum_{\sigma(\supset a_m)} N(\sigma, a_m) V^\sigma = C_m V^{a_m} \,, \tag{37}$$

where $C_m = (-1)^m (m-1)!$. Using this in (36) and again employing (35) we obtain

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = \sum_{m=2}^{N-1} C_m \sum_{a_m} V_{a_m} \Psi_\beta \,. \tag{38}$$

Now we employ the lemmas:

Lemma 1:[16] $\sum_{c_k} V_{c_k}^{a_m} = S_{N-1}^{(k)} V^{a_m}$, and (39)

Lemma 2:[17] $\sum_{n=2}^{N-1} C_n S_{N-1}^{(n)} = 1, \quad N \geqslant 3 \,.$ (40)

The $S_{N-1}^{(k)}$ are Stirling numbers of the second kind. $S_{N-1}^{(k)}$ is the number of distinct ways of making k clusters out of $(N-1)$ objects. From (38) we now obtain

$$\sum_\sigma \mathscr{V}_\sigma \Psi_\beta = V \Psi_\beta \,. \quad \text{Q.E.D.} \tag{41}$$

This means that (29) with the σ sum taken over the two-cluster partitions is the Schrödinger equation. This result motivates the consideration of truncations of Σ_σ and/or $\mathscr{V}_\sigma$

such that the corresponding summed version of (29) remains the full Schrödinger equation. We present a method which satisfies this condition through the following results.

Remark: For a given N, if $\mathscr{V}_\sigma$ is given exactly the sum $(\Sigma \mathscr{V}_\sigma \Psi_\beta)$ over an arbitrary subset of the two-cluster partitions does not result in $(E - H_0 - \Sigma_\sigma \mathscr{V}_\sigma)\Psi_\beta = 0$ becoming the Schrödinger equation.

Example: Consider the case $N = 3$ for which $\mathscr{V}_\sigma = V_\sigma$. We label the possible σ's as $\sigma_1 = (1)\,(23)$, $\sigma_2 = (2)\,(13)$, and $\sigma_3 = (3)\,(12)$. Note that

$$\sum_{\sigma_1, \sigma_2} \mathscr{V}_\sigma \Psi_\beta = \sum_{\sigma_1, \sigma_2} V_\sigma \Psi_\beta \,, \tag{42}$$

$$\sum_{\sigma_1, \sigma_2} \mathscr{V}_\sigma \Psi_\beta \neq V \Psi_\beta \,. \tag{43}$$

Remark: For a given N and $\mathscr{V}_\sigma$ truncated arbitrarily then the sum over all σ does not result in $(E - H_0 - \Sigma_{\text{all }\sigma} \mathscr{V}_\sigma^T \Psi_\beta = 0)$ becoming the Schrödinger equation, where $\mathscr{V}_\sigma^T$ is a truncated version of $\mathscr{V}_\sigma$.

Example: Consider $N = 4$ and suppose we truncate $\mathscr{V}_\sigma$ by taking $\mathscr{V}_\sigma \sim \mathscr{V}_\sigma^T = V_\sigma$, so

$$\sum_{\text{all }\sigma} \mathscr{V}_\sigma^T \Psi_\beta = \sum_{\text{all }\sigma} V_\sigma \Psi_\beta \,. \tag{44}$$

From Lemma 1, $\Sigma_\sigma V_\sigma = S_{N-1}^{(2)} V$ and we have

$$\sum_{\text{all }\sigma} \mathscr{V}_\sigma^T \Psi_\beta = S_3^{(2)} V \Psi_\beta \,. \tag{45}$$

We now introduce notation to represent a particular class of truncations of the irreducible cluster potential. The operator $\mathscr{V}_\sigma^{a_m}$ is defined as the truncation of the anticluster expansion (28) of $\mathscr{V}_\sigma$ which includes only those terms corresponding to the partitions that can be formed by joining the clusters of a single "dominant" partition a_m, with $3 \leqslant m \leqslant N-1$. Explicitly this is

$$\mathscr{V}_\sigma^{a_m} = \sum_{n=2}^{m} \sum_{(\sigma \supset) a_n (\supseteq a_m)} N(\sigma, a_n) V_{a_n} G_{a_n} G_\sigma^{-1} \,. \tag{46}$$

This class of truncated operators allows us to introduce the dominant partition theorem (DPT). The term dominant partition derives from the role played by a fixed partition a_m in the truncation of $\mathscr{V}_\sigma$ and in limiting the sum on two-cluster partitions.

Theorem II: (dominant partition theorem): For arbitrary N, and an arbitrary fixed partition a_m

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = V \Psi_\beta \,,$$

where Ψ_β solves the Schrödinger equation for the N-body system $(3 \leqslant m \leqslant N-1)$.

Proof: Using the definition (46) gives

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = \sum_{\sigma(\supset a_m)} \sum_{n=2}^{m} \sum_{(\sigma) \supset a_n (\supseteq a_m)} N(\sigma, a_n) \times V_{a_n} G_{a_n} G_\sigma^{-1} \Psi_\beta \,. \tag{47}$$

We write (46) in a more convenient form by picking off the a_m term

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = \sum_{\sigma(\supset a_m)} \sum_{n=2}^{m-1} \sum_{(\sigma \supset) a_n (\supseteq a_m)} N(\sigma, a_n) V_{a_n} G_{a_n} G_\sigma^{-1} \Psi_\beta + \sum_{\sigma(\supset a_m)} N(\sigma, a_m) V_{a_m} G_{a_m} G_\sigma^{-1} \Psi_\beta \,. \tag{48}$$

Again employing (31) we obtain

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = \sum_{\sigma(\supset a_m)} \left[\sum_{n=2}^{m-1} \sum_{(\sigma \supset) a_n (\supseteq a_m)} N(\sigma, a_n) V_{a_n} \times (1 - G_{a_n} V_\sigma^{a_n}) + N(\sigma, a_m) V_{a_m} \times (1 - G_{a_m} V_\sigma^{a_m}) \right] \Psi_\beta \,. \tag{49}$$

Using (33) and (35) we have

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = V_\beta \Phi_\beta + \sum_{\sigma(\supset a_m)} \sum_{n=2}^{m-1} \sum_{(\sigma \supset) a_n (\supseteq a_m)} N(\sigma, a_n) \times V_{a_n} G_{a_n} V^\sigma \Psi_\beta + \sum_{\sigma(\supset a_m)} N(\sigma, a_m) \times V_{a_m} G_{a_m} V^\sigma \Psi_\beta \,. \tag{50}$$

Interchanging sums we have

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = V_\beta \Phi_\beta + \sum_{n=2}^{m-1} \sum_{a_n (\supseteq a_m)} \sum_{\sigma(\supset a_m)} N(\sigma, a_n) \times V_{a_n} G_{a_n} V^\sigma \Psi_\beta + \sum_{\sigma(\supset a_m)} N(\sigma, a_m) \times V_{a_m} G_{a_m} V^\sigma \Psi_\beta \,. \tag{51}$$

Use of (37) and (35) yields

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = \left(\sum_{n=2}^{m-1} \sum_{a_n (\supseteq a_m)} C_n V_{a_n} + C_m V_{a_m} \right) \Psi_\beta \,. \tag{52}$$

We now use[16]

Lemma 3:

$$\sum_{a_n (\supset a_m)} V_{a_n}^{a_m} = S_{m-1}^{(n)} V^{a_m} \,. \tag{53}$$

This yields

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = \left[\sum_{n=2}^{m-1} C_n (S_m^{(n)} V_{a_m} + S_{m-1}^{(n)} V^{a_m}) + C_m V_{a_m} \right] \times \Psi_\beta \,, \tag{54}$$

where we have used (33) (with σ replaced by a_n) to express V_{a_n} in (52) as $V_{a_m} + V_{a_n}^{a_m}$.

Noting the results

$$\sum_{n=2}^{m-1} C_n S_m^{(n)} = 1 - C_m \tag{55}$$

and

$$\sum_{n=2}^{m-1} C_n S_{m-1}^{(n)} = 1 \,, \tag{56}$$

which follow from Lemma 2 we have

$$\sum_{\sigma(\supset a_m)} \mathscr{V}_\sigma^{a_m} \Psi_\beta = V \Psi_\beta \,. \quad \text{Q.E.D.} \tag{57}$$

Theorem II provides the basis for the DPM. It shows that we may truncate $\mathscr{V}_\sigma$ through the anticluster expansion

by retaining only those partitions that contain a given dominant partition. Note that the full partition Green functions G_{a_n} are retained in (46). They are *not* projected on the Hilbert space corresponding to bound states of the dominant partition a_m. The description of the breakup of these clusters is contained in these Green functions.

The reduced problem is then solved by solutions to the original Schrödinger equation. This theorem provides a consistent means of reducing the many-body problem in the LRT wavefunction formalism to a few-body problem.

Nuclear reactions are commonly analyzed in terms of a few-body picture. For a given N there are $S_N^{(2)} = 2^{N-1} - 1$ two-cluster channels. Any realistic attempt to solve the many-body problem cannot treat all of these channels on an equal footing. Moreover, it is reasonable to expect that in direct reactions the processes involved are not so extensive that all possible rearrangement and inelastic processes must be included. In many cases a realistic approach to many-body reaction theory will be afforded by systematically building few-body models.

IV. THE DOMINANT PARTITION AND THE BRS EQUATION

We now obtain dominant partition truncations of the BRS equations. This is accomplished by restricting the sum on two-cluster partitions to the class defined by $\sigma(\supset a_m)$, where a_m is taken to be the dominant partition. Correspondingly we introduce the appropriate truncation of the BRS kernel by using the anticluster expansion (27). Restricting the sum on two-cluster partitions to those that contain a particular a_m terminates the anticluster expansion with that term explicitly involving a_m. We write the truncated kernel as $\kappa_{\beta\sigma}^{a_m}$, that is

$$\kappa_{\beta\sigma}^{a_m} = \sum_{n=2}^{m} \sum_{(\sigma\supset)a_n(\supseteq a_m)} N(\sigma, a_n) V_{a_n}^{\beta} G_{a_n} . \tag{58}$$

With $\beta,\alpha(\supset a_m)$ we write the truncated BRS equations as

$$T^{\beta\alpha} = V_{\alpha}^{\beta} + \sum_{\sigma(\supset a_m)} \kappa_{\beta\sigma}^{a_m} T^{\sigma\alpha} . \tag{59}$$

It will be recalled that in the derivation of the BRS equations that a crucial step was the democratic distribution of the residual interaction over all partitions.[16] If we restrict the distribution to those partitions containing a_m and proceed with the derivation, are the truncated equations obtained the same as those we have termed the dominant partition truncated BRS equations? The answer is yes and provides the next theorem.

Theorem III: The dominant-partition-truncated BRS equations (59) are obtained by distributing the residual interaction V^{β} over the subset of all possible partitions containing a_m and proceeding as in the derivation of the BRS equations (Ref. 16).

This theorem provides a satisfying degree of consistency in the reduction of the BRS equations. These considerations are displayed in Fig. 1.

Proof: From (39) and (53) we have

$$\sum_{d_j} V_{d_j}^{\beta} = S_{N-1}^{(j)} V^{\beta} \tag{60}$$

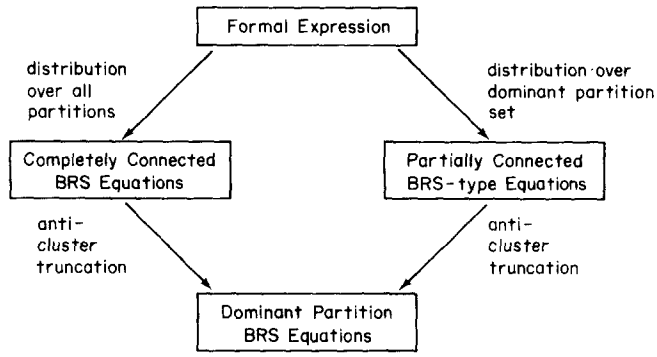


FIG. 1.

and

$$\sum_{d_j(\supseteq a_m)} V_{d_j}^{a_m} = S_{m-1}^{(j)} V^{a_m} . \tag{61}$$

Also note that

$$\sum_{d_j(\supset a_m)} V_{d_j}^{\beta} = S_{m-1}^{(j)} V^{\beta}, \quad \beta\supset a_m . \tag{62}$$

Using (56) we obtain

$$V^{\beta} = \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j V_{d_j}^{\beta}, \quad \beta\supset a_m . \tag{63}$$

We use this in the definition of the transition operator[2]

$$T_{+}^{\beta\alpha} = V^{\beta} G G_{\alpha}^{-1} . \tag{64}$$

We obtain

$$T_{+}^{\beta\alpha} = \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j V_{d_j}^{\beta} G_{d_j} G_{\alpha}^{-1} + \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j V_{d_j}^{\beta} G_{d_j} T_{+}^{d_j\alpha}, \quad \beta\supset a_m . \tag{65}$$

Using the Lippmann identity[16,18] to transform the Born term gives the following equation for a new set of operators $T^{\beta\alpha}$ which are equal to $T_{+}^{\beta\alpha}$ on the half-shell

$$T^{\beta\alpha} = \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j V_{d_j}^{\beta} \delta_{d_j\alpha} + \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j V_{d_j}^{\beta} G_{d_j} T^{d_j\alpha}, \quad \beta\supset a_m . \tag{66}$$

This yields

$$T^{\beta\alpha} = V_{\alpha}^{\beta} + \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j V_{d_j}^{\beta} G_{d_j} T^{d_j\alpha}, \quad \alpha\supset a_m, \ \beta\supset a_m . \tag{67}$$

We use the Yakubovskii cluster expansion[19] to decompose the transition operator internal to partition d_j into pieces of different connectivities. We write

$$T_{+d_j}^{\gamma\alpha} = \sum_{n=j}^{m-1} \sum_{(d_j\supset)d_n} K_{d_n}^{\gamma\alpha} , \tag{68}$$

and

$$T_{+d_j}^{\beta 0} G_0 = V_{d_j}^{\beta} G_{d_j} = \sum_{n=j}^{m-1} \sum_{(d_j\supset)d_n} K_{d_n}^{\beta 0} G_0 . \tag{69}$$

The anticluster truncation procedure is to replace $K_{d_n}^{\beta 0} G_0$ by

$\kappa^{a_m}_{\beta d_n}$ for $n > m$ and to drop all other terms. We then get

$$T^{\beta\alpha} = V^{\beta}_{\alpha} + \sum_{j=2}^{m-1} \sum_{d_j(\supset a_m)} C_j \sum_{n=j}^{m-1} \sum_{(d_j \supset) d_n} \kappa^{a_m}_{\beta d_n} T^{d_j \alpha} . \quad (70)$$

We interchange the n and j sums realizing that m is the largest number of clusters that we can have in the limited space. We obtain

$$T^{\beta\alpha} = V^{\beta}_{\alpha} + \sum_{n=2}^{m-1} \sum_{d_n(\supset a_m)} \kappa^{a_m}_{\beta d_n} \sum_{j=2}^{n} \sum_{d_j(\supset d_n)} C_j V^{d_j} G G^{-1}_{\alpha} , \quad (71)$$

where (64) has been used. Employing

Lemma 4:[16]

$$\sum_{j=2}^{n} \sum_{d_j(\supset d_n)} C_j V^{d_j} = \delta_{2n} V^{d_n} , \quad (72)$$

yields

$$T^{\beta\alpha} = V^{\beta}_{\alpha} + \sum_{n=2}^{m-1} \sum_{d_n(\supset a_m)} \kappa^{a_m}_{\beta d_n} \delta_{2n} T^{d_n \alpha} \quad (73)$$

and

$$T^{\beta\alpha} = V^{\beta}_{\alpha} + \sum_{\sigma(\supset a_m)} \kappa^{a_m}_{\beta\sigma} T^{\sigma\alpha} , \quad \beta, \alpha \supset a_m . \quad \text{Q.E.D.} \quad (74)$$

V. SUMMARY AND CONCLUSIONS

We have developed a generalization of the Hahn and Watson's[7] "partially connected" strategy appropriate for constructing n-cluster models for N-body problems where it is intended that $n \ll N$. The cases treated are those in which the only channels treated explicitly are obtained by combining the clusters of an n-cluster "dominant" partition, a_n. We obtain n-body equations of the BRS type for transition operators and of the LRT type for wave functions. The resulting equations are connected in the degrees of freedom corresponding to the relative motion of the clusters of a_n, but not in those internal to a single cluster of a_n. Following Hahn and Watson we assume that these degrees of freedom are to be handled in some manner different from operator integral equations (e.g., by statistical or phenomenological methods).

Our main result is that the BRS and LRT equations for the small number of clusters is in fact exact if the subsystem Green functions are put in from some other source. This means that no incoming waves associated with channels breaking the cluster of a_n are to be admitted. Furthermore, one can obtain the partially connected equations by either truncating the anticluster expansion for the kernel or by distributing the residual potential only over the appropriate, limited set of partitions. The same equations are obtained by both procedures.

A specific example where a procedure such as described here may be relevant is in the six-nucleon problem where the initial channel is a low energy ($E < 20$ MeV) deuteron incident on a ^{4}He nucleus. As is well known[20], this is well described as a three-body problem. Here, our dominant partition would be $a_3 = (n)\,(p)\,(nnpp)$ where the effects of exchange are ignored. The three-body equation would fall out immediately upon approximating the Green function G_{a_3} by its part having the ^{4}He pole. This approximation would yield real effective nucleon-^{4}He interactions.

More general results, including the appearance of complete effective interactions, can be obtained in a number of ways, the simplest of which is the introduction of projection operators at the Green function G_{a_3}. The part corresponding to everything but the ^{4}He pole is then solved formally a la Feshback.[21] This leads to the appearance of generalized optical potentials as effective interactions plus the well-known[22] effective three-body force.

[1] E.O. Alt, P. Grassberger, and W. Sandhas, Nucl. Phys. **B2**, 167 (1967); JINR Report No. 4-6688 (1972); W. Sandhas, in *Few-Body Dynamics*, A. Mitra, I. Slaus, V. Bhasin, and V. Gupta, Eds. (North-Holland, New York, 1976), pp. 540–57.

[2] G. Bencze, Nucl. Phys. A **210**, 568 (1973); E.F. Redish, Nucl. Phys. A **225**, 16 (1974); I.H. Sloan, Phys. Rev. C **6**, 1945 (1972).

[3] D.J. Kouri and F.S. Levin, Nucl. Phys. A **253**, 395 (1975); W. Tobocman, Phys. Rev. C **9**, 2466 (1974).

[4] L.D. Faddeev, Sov. Phys. JETP **12**, 1014 (1961).

[5] E.F. Redish, in *Modern Three Hadron Physics*, A.W. Thomas, Ed. (Springer, New York, 1977), pp. 181–239.

[6] S.R. Cotanch and C.M. Vincent, Phys. Rev. C **14**, 1739 (1976).

[7] Y. Hahn and K.M. Watson, Phys. Rev. A **5**, 1718 (1972).

[8] M. L'Huillier, E.F. Redish, and P.C. Tandy, Bull. Am. Phys. Soc. **21**, 49 (1976); "The Structure of the Scattering Wave Function in the N-body Problem," Int. Symp. on Nuclear Reaction Models, Balatonfüred, Hungary, June 27–July 1, 1977; J. Math. Phys. **19**, 1276 (1978).

[9] M. Goldberger and K.M. Watson, *Collision Theory* (Wiley, New York, 1964).

[10] R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[11] S. Weinberg, Phys. Rev. **133**, B232 (1964).

[12] V. Vanzani, in *Few-Body Nuclear Physics* (I.A.E.A., Vienna, 1978), p. 57.

[13] E.F. Redish, Nucl. Phys. A **235**, 82 (1974).

[14] W. Glöckle, Nucl. Phys. **141**, 620 (1970).

[15] K.L. Kowalski, Phys. Rev. C**16**, 7 (1977).

[16] P. Benoist-Gueutal, M. L'Huillier, E.F. Redish, and P.C. Tandy, Phys. Rev. C **17**, 1924 (1978).

[17] *Handbook of Mathematical Functions*, M. Abramowitz and I.A. Stegun, Eds. (National Bureau of Standards, 1964), p. 825.

[18] B.A. Lippmann, Phys. Rev. **102**, 264 (1956).

[19] O.A. Yakubovskii, Sov. J. Nucl. Phys. (U.S.A.) **5**, 937 (1967).

[20] G. Plattner, in *Few-Body Systems and Nuclear Forces* II, Proc. VIII Int. Conf., Graz, Austria, Aug. 24–30, 1978, H. Zingle, M. Haftel, and H. Zankel, Eds. (Springer-Verlag, Berlin, 1978), p. 376.

[21] H. Feshbach, Ann. Phys. (N.Y.) **5**, 357 (1958); **19**, 287 (1962); Gy. Bencze, "Effective Three-Body Problems in Nuclear Reactions," U. of Helsinki preprint no. 2-73 (1973), unpublished.

[22] N. Austern and K.C. Richards, Ann. Phys. (N.Y.) **49**, 309 (1968).

Initial value problem for two oscillators interacting with electromagnetic field

K. Rzążewski

Institute of Theoretical Physics, University of Warsaw, 00—681 Warsaw, Hoża 69, Poland

W. Żakowicz

Institute of Physics of Polish Academy of Sciences, 02-668 Warsaw, Al.Lotników 32, Poland

(Received 29 September 1978)

We solve exactly equations of motion for two charged harmonic oscillators interacting with electromagnetic field. Detailed analyses of the system's behavior for small times is presented. We go beyond the one pole approximation. A simple model for the so-called soft mode instability is given. The scattering cross section is derived from time dependent solutions.

I. INTRODUCTION

The interaction of two atoms with radiation has been studied for many years.[1-8] This is the simplest system for which cooperative behavior can be studied and due to its simplicity a dynamical description can be given.

The long time behavior has been studied in the frame of the so-called single pole approximation. Stephen[1] was probably the first two found the correct distance dependent expressions for the width and shifts of the emission line. To some extent the expression has been qualitatively verified in the experiments of Drexhage *et al.*,[9] on the scattering of light by the atom placed in front of the mirror.

The interesting problem was the demonstration of causal behavior of the system due to the finite velocity of light propagation. If one of the atoms is excited in the initial moment, one expects proper delay before the other one will be excited. Milloni and Knight[6] have shown how this step type behavior matches the long time description given by single pole methods.

The purpose of this paper is to discuss the problem of two sources by means of two harmonic oscillators interacting with electromagnetic field. This problem can be exactly solved in the frame of the dipole approximation with a cutoff removing the corresponding divergence for high photon frequencies $\omega \to \infty$.

A similar problem of two harmonic oscillators but interacting with the scalar field has been recently discussed by Aichelburg and Grosse.[7,8] They were able to renormalize their model and to remove the cutoff parameter from solution.

In our model the cutoff frequency Ω remains in the solution via the single oscillator life time and frequency shift. In the limit $\Omega \to \infty$ the single oscillator life time remains finite while the frequency shift is linearly divergent. One could remove this divergence by a similar renormalization procedure as in Refs. 7, 8, but the resulting oscillator dynamics contains so-called "run away" solutions which do not have any physical meaning.

Two different methods can be used to express solutions. One is convenient for short time behavior and study of retardation properties, the other for long time evolution.

We go also beyond the single pole approximation. We find all subsidiary poles of the resolvent functions in the problem and point out their role played for short time behavior, especially discussing their contribution in matching the initial data.

The model predicts instability for very short distance between sources. This instability is entirely due to the electrostatic dipole–dipole interaction. An instability of this kind is responsible for the spontaneous generation of the static dipole moment in the system of two-level atoms, phenomenon recently discussed by several authors.[10,11]

The paper contains a section containing a derivation and discussion of the scattering cross section of radiation on two sources. The problem was investigated by Lyuboshits[2] in a quite different framework.

II. MODEL AND ITS SOLUTION

We consider a system composed of two harmonic oscillators located at fixed points $\bar{r}_1$ and $\bar{r}_2$. Each oscillator is composed of two opposite charges e, $-e$, with Coulomb interaction represented by the elastic force. One of these charges is spread out uniformly in a sphere and has infinite mass. We look for the motion of the second one (electron) having the mass m and moving inside this sphere.

The Hamiltonian of the system, written in the Coulomb gauge, is

$$H = \frac{1}{2m}\sum_{i=1}^{2}\left(\bar{p}_i - \frac{e}{c}\bar{A}(\bar{r}_i + \bar{x}_i)\right)^2 + \tfrac{1}{2}m\omega_0^2\sum_{i=1}^{2}\bar{x}_i^2 + \frac{e^2}{r_{12}^3}\left[\bar{x}_1\cdot\bar{x}_2 - 3(\bar{x}_1\cdot\bar{n})(\bar{x}_2\cdot\bar{n})\right]$$

 0022-2488/80/020378-11$1.00

$$+\frac{1}{8\pi}\int d_3 r[\bar{E}_T^2(\bar{r})+\bar{B}^2(\bar{r})], \tag{2.1}$$

where x_1, x_2 denote displacements of electrons from their equilibrium positions r_1 and r_2, p_1 and p_2 are cannonical moments, $r_{12}=|\bar{r}_1-\bar{r}_2|$, and $\bar{n}=\bar{r}_{12}/r_{12}$.

For the vector potential $\bar{A}(\bar{r})$ we have the following plane wave decomposition:

$$\bar{A}(\bar{r})=\frac{(\hbar c)^{1/2}}{2\pi}\sum_{\mu=1}^{2}\int d_3 k\,\bar{e}_{\bar{k}\mu}k^{-1/2}(a_{\bar{k}\mu}e^{i\bar{k}\cdot\bar{r}}+a_{k\mu}^{+}e^{-i\bar{k}\cdot\bar{r}}), \tag{2.2}$$

where $a_{k\mu}^{+}$ and $a_{\bar{k}\mu}$ are creation and annihilation operators for photons with the wave vector $\bar{k}$ and linear polarization $\bar{e}_{\bar{k}\mu}$ normalized according to the following commutation relations:

$$[a_{\bar{k}\mu},a_{\bar{k}'\mu'}^{+}]=\delta_{\mu\mu'}\,\delta_3(\bar{k}-\bar{k}'),$$
$$[a_{\bar{k}\mu},a_{\bar{k}'\mu'}]=0=[a_{\bar{k}\mu}^{+},a_{\bar{k}'\mu'}^{+}].$$

The model we will consider corresponds to the Hamiltonian (2.1) taken in the dipole approximation. In the dipole approximation we replace $\bar{A}(\bar{r}_i+\bar{x}_i)$ by $\bar{A}(\bar{r}_i)$. Making such a replacement, however, we are losing a tempering oscillatory factor $e^{i\bar{k}\cdot\bar{x}_i}$ and the model becomes divergent at high photon frequencies. To stimulate the converging effect of $e^{i\bar{k}\cdot\bar{x}}$ and to make our dipole approximation finite, we introduce in the coupling terms the form factor which cut off high frequency photons.

With these approximations our model is defined by the following Hamiltonian:

$$H=\frac{1}{2m}\sum_{i=1}^{2}\left(\bar{p}_i-\frac{e}{c}\bar{A}_g(\bar{r}_i)\right)^2+\tfrac{1}{2}m\omega_0^2\sum_{i=1}^{2}\bar{x}_i^2$$
$$+\frac{e^2}{r_{12}^3}[\bar{x}_1\cdot\bar{x}_2-3(\bar{x}_1\cdot\bar{n})(\bar{x}_2\cdot\bar{n})]+\frac{1}{8\pi}\int d_3 r$$
$$\times[\bar{E}_T^2(\bar{r})+\bar{B}^2(\bar{r})], \tag{2.3}$$

where $\bar{A}_g(\bar{r})$ corresponds to Eq. (2.2) with $k^{-1/2}$ being replaced by the form factor $g(k)$. We take the form factor in the form

$$g(k)=\frac{1}{k^{1/2}}\frac{\Omega}{(\Omega^2+k^2c^2)^{1/2}} \tag{2.4}$$

with the cutoff frequency Ω being of the order of $2\pi c/d$, where d is the radius of oscillator.

We are going to solve the Heisenberg equations of motion for the dynamical variables of the system. Instead of the canonical moments $\bar{p}_i$ appearing in the Hamiltonian (2.1) we will use velocites (kinetic momenta). This choice is dictated by causality requirements as was discussed by us in Ref. 12.

The equations of motion for charges are

$$\frac{d\bar{x}_1}{dt}=\bar{v}_1, \tag{2.5a}$$

$$\frac{d\bar{v}_1}{dt}=-\omega_0^2\bar{x}_1+\frac{e^2}{mr_{12}^3}[\bar{x}_2-3\,\bar{n}(\bar{n}\cdot\bar{x}_2)]$$
$$+i\frac{e}{m}\frac{(\hbar c)^{1/2}}{2\pi}\sum_{\mu}\int d^3k\,kg(k)\bar{e}_{\bar{k}\mu}$$
$$\times(a_{\bar{k}\mu}e^{i\bar{k}\cdot\bar{r}_1}-a_{\bar{k}\mu}^{+}e^{-i\bar{k}\cdot\bar{r}_1}), \tag{2.5b}$$

plus the similar equation for the second oscillator (1↔2).

The field equations are ($\omega=|\bar{k}|\cdot c$)

$$\frac{da_{\bar{k}\mu}}{dt}=-i\omega a_{\bar{k}\mu}+i\frac{e}{2\pi(\hbar c)^{1/2}}g(k)\bar{e}_{\bar{k}\mu}$$
$$\cdot\sum_{i=1}^{2}\bar{v}_i\cdot e^{-i\bar{k}\bar{r}_i} \tag{2.6}$$

plus conjugate equation for $a_{\bar{k}\mu}^{+}$.

To solve the set of equations (2.5), (2.6), we will use the Laplace transform method, $\tilde{f}(z)=\int_0^\infty dt\,e^{-zt}f(t)$. Using transformed Eq. (2.6),

$$\tilde{a}_{\bar{k}\mu}(z)=\frac{a_{\bar{k}\mu}(0)}{z+i\omega}+i\frac{e}{2\pi(\hbar c)^{1/2}}\frac{g(k)}{z+i\omega}\bar{e}_{\bar{k}\mu}$$
$$\cdot\sum_{j=1}^{2}e^{-i\bar{k}\cdot\bar{r}_j}\bar{v}_j(z), \tag{2.7}$$

one can eliminate dynamical field variables from the equation of motion of oscillators, getting the following set of equations:

$$\mathscr{H}(z)\tilde{\bar{v}}_1(z)+\frac{e^2}{mc^2}e^{-zr_{12}/c}\left[(\hat{I}-\bar{n}\otimes\bar{n})\frac{z^2}{r_{12}}\right.$$
$$\left.+(\hat{I}-3\,\bar{n}\otimes\bar{n})\left(\frac{zc}{r_{12}^2}+\frac{c^2}{r_{12}^3}\right)\right]\cdot\tilde{\bar{v}}_2(z)$$
$$=\bar{K}_1=-\omega_0^2\bar{x}_1(0)+z\bar{v}_1(0)$$
$$-\frac{e^2}{mr_{12}^3}(\hat{I}-3\,\bar{n}\otimes\bar{n})\cdot\bar{x}_2(0)$$
$$+ize\frac{(\hbar c)^{1/2}}{2\pi m}\sum_{\mu}\int d^3k\,k^{1/2}\,\bar{e}_{\bar{k}\mu}\left(\frac{a_{\bar{k}\mu}(0)}{z+i\omega}e^{i\bar{k}\cdot\bar{r}_1}\right.$$
$$\left.-\frac{a_{\bar{k}\mu}^{+}(0)}{z-i\omega}e^{-i\bar{k}\cdot\bar{r}_1}\right) \tag{2.8}$$

plus the second equation for $\tilde{\bar{v}}_2(z)$. Here the function $\mathscr{H}(z)$ defines resolvent function for single oscillator.[12] It is given by

$$\mathscr{H}(z)=\omega_0^2+z^2\left(1+\frac{4e^2}{3\pi mc}\int_0^\infty dk\,\frac{k^3g^2(k)}{z^2+\omega^2}\right), \tag{2.9}$$

and with our form factor (2.4) for that branch of H which is relevent for the future evolution we get

$$\mathscr{H}(z)=\omega_0^2+z^2\left(1+\frac{2}{3}\frac{e^2}{mc^3}\frac{\Omega^2}{z+\Omega}\right). \tag{2.10}$$

We point out that the function $\mathscr{H}(z)$ defined by (2.9) is a multivalued function of z. With the form factor (2.4) it is double-valued. In passing to (2.10) we have chosen that branch which has no zeros for $\mathrm{Re}z>0$. The same Riemanian sheet was used when evaluating the contribution to the motion of one oscillator coming from the second one. This term appears with the factor $\exp(-zr_{12}/c)$ responsible for retardation effects in the interaction of both oscillators via electromagnetic field.

Notice that, although the dynamical variables of the field has been eliminated from the equations of motion, the initial data for the field are still necessary. This means that we are still dealing with the coupled problem for the motion of charges and field. A similar system of equations for N oscillators occupying a small spherical volume was analyzed in Ref. 17.

The components of velocities parallel to the vector $\bar{n}$, $\bar{v}_{\|,i}=(\bar{v}_i\cdot\bar{n})\bar{n}$ and transverse to $\bar{n}$, $\bar{v}_{\perp i}=\bar{n}\times\bar{v}_i$ satisfy separated equations

$$\mathscr{H}\tilde{\bar{v}}_{\parallel 1} - \mathscr{U}_{\parallel}\tilde{\bar{v}}_{\parallel 2} = \bar{K}_{\parallel 1}, \tag{2.11a}$$

$$-\mathscr{U}_{\parallel}\tilde{\bar{v}}_{\parallel 1} + \mathscr{H}\tilde{\bar{v}}_{\parallel 2} = \bar{K}_{\parallel 2}, \tag{2.11b}$$

and

$$\mathscr{H}\tilde{\bar{v}}_{\perp 1} + \mathscr{U}_{\perp}\tilde{\bar{v}}_{\perp 2} = \bar{K}_{\perp 1}, \tag{2.11c}$$

$$\mathscr{U}_{\perp}\tilde{\bar{v}}_{\perp 1} + \mathscr{H}\tilde{\bar{v}}_{\perp 2} = \bar{K}_{\perp 2}, \tag{2.11d}$$

where

$$\mathscr{U}_{\parallel} = 2\gamma e^{-z\rho}\left(\frac{z}{\rho^2} + \frac{1}{\rho^3}\right), \tag{2.12a}$$

$$\mathscr{U}_{\perp} = \gamma e^{-z\rho}\left(\frac{z^2}{\rho} + \frac{z}{\rho^2} + \frac{1}{\rho^3}\right) \tag{2.12b}$$

with $\gamma = e^2/mc^3$ and $\rho = r_{12}/c$.

These equations separate further into symmetric $\bar{v}_s = \bar{v}_1 + \bar{v}_2$ and antisymmetric vibrations $\bar{v}_a = \bar{v}_1 - \bar{v}_2$. The solution of Eqs. (2.11) reads

$$\tilde{\bar{v}}_{\parallel s} = \frac{1}{\mathscr{H} - \mathscr{U}_{\parallel}}\bar{K}_{\parallel s}, \quad \tilde{\bar{v}}_{\parallel a} = \frac{1}{\mathscr{H} + \mathscr{U}_{\parallel}}\bar{K}_{\parallel a}, \tag{2.13}$$

$$\tilde{\bar{v}}_{\perp s} = \frac{1}{\mathscr{H} + \mathscr{U}_{\perp}}\bar{K}_{\perp s}, \quad \tilde{\bar{v}}_{\perp a} = \frac{1}{\mathscr{H} - \mathscr{U}_{\perp}}\bar{K}_{\perp a}.$$

The functions $(\mathscr{H} \pm \mathscr{U}_{\parallel})^{-1}$ and $(\mathscr{H} \pm \mathscr{U}_{\perp})^{-1}$ we will call resolvent functions of the two oscillator problem.

Using the solution (2.13), we get for the velocity of the first electron

$$\tilde{\bar{v}}_1(z) = \alpha(z)\bar{n}(\bar{n}\cdot\bar{K}_1) + \beta(z)\bar{K}_1 + \gamma(z)\bar{n}(\bar{n}\cdot\bar{K}_2) + \delta(z)\bar{K}_2, \tag{2.14}$$

where

$$\alpha(z) = \frac{\mathscr{H}}{\mathscr{H}^2 - \mathscr{U}_{\parallel}^2} - \frac{\mathscr{H}}{\mathscr{H}^2 - \mathscr{U}_{\perp}^2}, \tag{2.15a}$$

$$\beta(z) = \frac{\mathscr{H}}{\mathscr{H}^2 - \mathscr{U}_{\perp}^2}, \tag{2.15b}$$

$$\gamma(z) = \frac{\mathscr{U}_{\parallel}}{\mathscr{H}^2 - \mathscr{U}_{\parallel}^2} + \frac{\mathscr{U}_{\perp}}{\mathscr{H}^2 - \mathscr{U}_{\perp}^2}, \tag{2.15c}$$

$$\delta(z) = \frac{\mathscr{U}_{\perp}}{\mathscr{H}^2 - \mathscr{U}_{\perp}^2}. \tag{2.15d}$$

Before we discuss the time dependence of these solutions, we find the emitted electromagnetic field. We get the solution for the creation and annihilation operators $a_{k\mu}^{+}(z)$ and $a_{\bar{k}\mu}(z)$ substituting (2.14) into (2.7). Using these operators, one can perform the reconstruction of electromagnetic field in space. Our procedure and discussion is similar to the one in Ref. 12.

For the transverse part of the electric field

$$\tilde{\bar{E}}_T(\bar{r},z) = i\frac{(\hbar c)^{1/2}}{2\pi}\sum_{\mu}\int d^3k\, k^{1/2}\bar{e}_{\bar{k}\mu} \times (a_{\bar{k}\mu}(z)e^{i\bar{k}\cdot\bar{r}} - a_{k\mu}^{+}(z)e^{-\bar{k}\cdot\bar{r}}) \tag{2.16}$$

we get after k integration and polarization summation

$$\tilde{\bar{E}}_T(\bar{r},z) = \tilde{\bar{E}}^{\text{free}}(\bar{r},z) - e\sum_{i=1}^{2}\left\{e^{-z|\bar{r}-\bar{r}_i|/c}\left[(\hat{I} - \bar{n}_i\otimes\bar{n}_i)\frac{z}{|\bar{r}-\bar{r}_i|c^2} + (\hat{I} - 3\,\bar{n}_i\otimes\bar{n}_i)\left(\frac{1}{|\bar{r}-\bar{r}_i|^2 c} + \frac{1}{z|\bar{r}-\bar{r}_i|^3}\right)\right] - \frac{1}{z|\bar{r}-\bar{r}_i|^3}(\hat{I} - 3\,\bar{n}_i\otimes\bar{n}_i)\right\}\cdot\tilde{\bar{v}}_i(z) \tag{2.17}$$

here $\bar{n}_i = (\bar{r} - \bar{r}_i)/|\bar{r} - \bar{r}_i|$ and $\tilde{\bar{E}}^{\text{free}}$ represents the freely evolving part of the electric field.

The longitudinal part of the field is given by

$$\tilde{\bar{E}}_L(\bar{r},z) = -e\sum_{i=1}^{2}(\hat{I} - 3\bar{n}_i\otimes\bar{n}_i)\cdot\frac{\tilde{\bar{x}}_i(z)}{|\bar{r}-\bar{r}_i|^3}. \tag{2.18}$$

Adding both parts, we get

$$\tilde{\bar{E}}(\bar{r},z) = \tilde{\bar{E}}^{\text{free}}(\bar{r},z) - e\sum_{i=1}^{2}(\hat{I} - 3\bar{n}_i\otimes\bar{n}_i)\frac{\bar{x}_i(0)}{z|\bar{r}-\bar{r}_i|^3} - e\sum_{i=1}^{2}e^{-z|\bar{r}-\bar{r}_i|/c}\left[(\hat{I} - \bar{n}_i\otimes\bar{n}_i)\frac{z}{c^2|\bar{r}-\bar{r}_i|} + (\hat{I} - 3\bar{n}_i\otimes\bar{n}_i)\left(\frac{1}{|\bar{r}-\bar{r}_i|^2} + \frac{1}{z|\bar{r}-\bar{r}_i|^3}\right)\right]\cdot\tilde{\bar{v}}_i(z). \tag{2.19}$$

The first term of the right-hand side represents the static dipole field initially present due to the initial value of the dipole moments. The second term gives the contribution to the field due to the motion of charges. Due to the factor $\exp(-z|\bar{r}-\bar{r}_i|/c)$ the field at the point $\bar{r}$ is properly retarded. Notice, that this was not the case either for the transverse electric field or for the longitudinal one.

The dependence of $\bar{v}_i$ on $\bar{K}_j$ and therefore on the initial data [the initial state of oscillators $\bar{x}_i(0)$ and $\bar{v}_i(0)$ and of the field $a_{\bar{k}\mu}(0)$ and $a_{k\mu}^{+}(0)$, [Eq. (2.8)] leads to natural decomposition of the emitted field into spontaneous radiation and scattered radiation. Because of analytical properties of the resolvent functions, the initial state of oscillators cannot contribute to the spontaneous radiation at the point $\bar{r}$ before time $ct = \min(|\bar{r}-\bar{r}_1|, |\bar{r}-\bar{r}_2|)$. For times

$$\min(|\bar{r}-\bar{r}_1|, |\bar{r}-\bar{r}_2|) < tc < \max(|\bar{r}-\bar{r}_1|, |\bar{r}-\bar{r}_2|)$$

only one oscillator contributes. For later times both oscillator contribute, and the radiation is influenced by the interference effect.

To demonstrate causal properties of the scattered radiation, it is more convenient to express the initial data for field by the initial field distributions in space, i.e., $\bar{E}_T(\bar{r},0)$ and $\bar{B}(\bar{r},0)$ instead of $a_{\bar{k}\mu}(0)$ and $a_{k\mu}^{+}(0)$.

That part of $\bar{K}_i$ which depends on the initial data for the field may be written as

$$\bar{K}_j^f = ize\frac{(\hbar c)^{1/2}}{2\pi m}\sum_{\mu}\int d^3k\, k^{1/2}\cdot\bar{e}_{\bar{k}\mu}\left(\frac{a_{\bar{k}\mu}(0)}{z + i\omega}e^{i\bar{k}\cdot\bar{r}_j} - \frac{a_{k\mu}^{+}(0)}{z - i\omega}e^{-i\bar{k}\cdot\bar{r}_j}\right) = \frac{e}{4\pi mc}\int d^3r'\frac{e^{-z|\bar{r}_j - \bar{r}'|/c}}{|\bar{r}_j - \bar{r}'|}\times\left(\frac{z^2}{c}\bar{E}_T(\bar{r}',0) + z\,\text{curl}\bar{B}(\bar{r}',0)\right) \tag{2.20}$$

where we have used the equation

$$\sum_{\nu} \bar{e}_{\bar{k}\mu} a_{\bar{k}\mu} = -i \frac{\pi}{(\hbar\omega)^{1/2}} \int \frac{d^3r}{(2\pi)^3} e^{-i\bar{k}\cdot\bar{r}} \times \left[\bar{E}_T(\bar{r}) - \frac{\bar{k}}{k} \otimes \bar{B}(\bar{r}) \right] \tag{2.21}$$

plus the conjugate equation for $\Sigma_\mu \bar{e}_{\bar{k}\mu}\, a^+_{\bar{k}\mu}$. Again, the scattered field at the point $\bar{r}$ is influenced by the initial field at the point $\bar{r}\,'$ only for times t for which the electromagnetic signal can travel from $\bar{r}\,'$ to $\bar{r}$ via at least one oscillator.

It is worth adding at this point a remark about an alternative approach to the electric dipole interaction which is usually refered to as the $\bar{d}\cdot\bar{E}$ Hamiltonian.[13,14,15] One can write down an alternative Hamiltonian to (2.1) in the form

$$H = \frac{1}{2m}(\bar{\pi}_1^2 + \bar{\pi}_2^2) + \tfrac{1}{2}m\omega_0^2(\bar{x}_1^2 + \bar{x}_2^2) + \frac{1}{8\pi}\int d^3r(\bar{E}^2 + \bar{B}^2) - e(\bar{x}_1\cdot\bar{E}(\bar{r}_1) + \bar{x}_2\cdot\bar{E}(\bar{r}_2)) + 2\pi \int d^3r[(\bar{P}_1^{\perp})^2 + (\bar{P}_2^{\perp})^2], \tag{2.22}$$

which is unitarily equivalent to (2.1) but has an advantage of containing no direct interaction between oscillators (explicitly!). In this description kinetic momenta $\bar{\pi}_1$ and $\bar{\pi}_2$ *are* canonical; they commute with photon creation and annihilation operators.

The whole difference between (2.1) and (2.22) is in the different definition of the radiation field.

Studying the Maxwell equation

$$\mathrm{div}\bar{E} = 4\pi\rho, \tag{2.23}$$

we can write its general solution as a sum of the general solution of homogeneous equation and special solution of inhomogeneous equation. The latter is not unique. Consequently, the notion of the radiation field and the notion of the photon is not unique. The $\bar{p}\cdot\bar{A}$ interaction supplemented by the Coulomb gauge, used throught this paper, can be characterized by the special solution of (2.23) which is given by the Poisson integral:

$$\bar{E}(\bar{r}) = -\mathrm{grad}\left(\int \frac{\rho(\bar{r}\,')}{|\bar{r} - \bar{r}\,'|} d^3r' \right). \tag{2.24}$$

On the other hand, in the $\bar{d}\cdot\bar{E}$ interaction, the special solution of the inhomogeneous equation is given by the following distribution:

$$\bar{E}(\bar{r}) = 4\pi e \sum_{i=1}^{2} \bar{x}_i \delta(\bar{r} - \bar{r}_i) = 4\pi\bar{P}(\bar{r}); \tag{2.25}$$

that is, it is proportional to the polarization. Of course, all physical situations described in terms of the initial distribution of electric and magnetic field may be equivalently described either way, but the answers are ususally a little different if the initial state is described in terms of its photon's contents. In particular, photon vacua in the presence of the excited system are different in both descriptions.

The vacuum of the $\bar{p}\cdot\bar{A}$ Hamiltonian analyzed in terms of photons entering the $\bar{d}\cdot\bar{E}$ Hamiltonian contains photons necessary to build the Coulomb dipole field in the whole space. On the other hand, the vacuum of the $\bar{d}\cdot\bar{E}$ Hamiltonian contains $\bar{p}\cdot\bar{A}$ type photons. Therefore, the notion of spontaneous emission (emission to the vacuum) is slightly different in both cases. It manifests itself both for very short and very long times. In particular, t powers of the nonexponential tail of the decay are different.[14]

It is worth stressing that the scattering amplitude is the same in both descriptions. In the remote future and remote past $\bar{x}_i$ excitation amplitudes tend to zero, and Eq. (2.23) becomes homogeneous and no nonuniqueness arises. Formal proof of the above statement can be found in Ref. 14.

III. TIME EVOLUTION OF THE SYSTEM

To get time dependence of the solutions, we have to perform the inverse Laplace transformation of $\tilde{\bar{v}}_i(z)$ and $\tilde{\bar{E}}(\bar{r},z)$. This inverse transformation is given by

$$\bar{v}(t) = \int_\Gamma \frac{dz}{2\pi i} e^{zt}\tilde{\bar{v}}(z). \tag{3.1}$$

The form of the solutions shows that the analytical properties of the resolvent functions $(\mathscr{H} \pm \mathscr{U}_{\parallel})^{-1}$ and $(\mathscr{H} \pm \mathscr{U}_{\perp})^{-1}$ are crucial for these inverse transformations. As is discussed in the Appendix, these functions are analytic functions of z in the right half of the z plane, $\mathrm{Re}z > 0$, if separation of oscillators exceeds certain critical distance. The situation when the distance between oscillators is smaller than the critical one will be further discussed in connection with the static phase transition in the Dicke model (Sec. IV). For typical situations the contour of integration Γ may be chosen parallel and to the right of the imaginary axis.

To perform these inverse Laplace integrals, we can use the theorem of residues. There are two different methods for evaluation of these integrals. The first one uses the poles of the integrand functions. Beyond some imaginary poles which are connected with field initial data terms, the others are poles of the resolvent functions.

The second approach uses the resolvent functions expanded into the power series [e.g., $(\mathscr{H} - \mathscr{U}_{\parallel})^{-1} = \Sigma_{j=1}^{\infty} \mathscr{U}_{\parallel}^{j} \mathscr{H}^{-j-1}$] and the inverse integrals are expressed in terms of single oscillator poles. The order of these poles is, however, increasing for higher terms in the expansion. Within this approach the solution explicitly demonstrates retardation effects in the coupling of both oscillators. It is particularly convenient for description of the short time behavior, for times of the order of few transit times.

On the other hand, the first method is particularly suitable for long time behavior. Then, one can take into account only poles which are nearest to the imaginary axis. In fact each resolvent function has two, mutually conjugate poles, which are close to the imaginary axis. These principal poles can be easily found by perturbation of the free oscillator poles. $\pm i\omega_0$, and they are known for a long time[1-6] Beyond these poles, the resolvent functions, which are associated with exponential polynomials, have an infinite number of poles.[16] All these poles are much further to the left from the imaginary axis; therefore their contribution is important only at the initial stage of the system's evolution.

We start from the first approach. The principal roots of the equations $\mathscr{H} \pm \mathscr{U}_{\parallel} = 0$ and $\mathscr{H} \pm \mathscr{U}_{\perp} = 0$ may be found using the small value of the coupling parameter γ and the expansion

$$\zeta = \zeta_0 + \gamma\zeta_1 + \gamma^2\zeta_2 + \cdots. \tag{3.2}$$

As the first term ζ_0 we choose the principal root of the single oscillator function H. It is assymptotically given by

$$\zeta_0 = i\tilde{\omega} - \Gamma \simeq \frac{i\omega_0}{(1+\frac{2}{3}\gamma\Omega)^{1/2}} - \frac{1}{3}\frac{\gamma\omega_0^2}{(1+\frac{2}{3}\gamma\Omega)^2} + \cdots. \tag{3.3}$$

In the neighborhood of ζ_0 it is convenient to represent the function $\mathscr{H}$ in the form

$$\mathscr{H} = A(z-\zeta_0)(z-\zeta_0^*), \tag{3.4}$$

where $$A = \frac{z-\zeta_3}{z+\Omega} \simeq 1 + \frac{2}{3}\gamma\frac{\Omega^2}{\Omega+i\omega_0}. \tag{3.5}$$

[$\zeta_3 \simeq -\Omega(1+\frac{2}{3}\gamma\Omega)$ is the third root of $\mathscr{H}$.] For $\mathscr{H} \mp \mathscr{U}_{\parallel}$ we get

$$\zeta^{\parallel}_{1\{a\}^s} = \pm\frac{1}{i\tilde{\omega}A}e^{-\zeta_0\rho}\left(\frac{\zeta_0}{\rho^2}+\frac{1}{\rho^3}\right), \tag{3.6a}$$

$$\zeta^{\parallel}_{2\{a\}^s} = \frac{1}{2\tilde{\omega}^3}e^{-2\zeta_0\rho}\left(\frac{\zeta_0}{\rho^2}+\frac{1}{\rho^3}\right) \times\left[\frac{2\tilde{\omega}\zeta_0}{\rho} - i\left(\frac{\zeta_0}{\rho^2}+\frac{1}{\rho^3}\right)\right], \tag{3.6b}$$

and for zeros of $\mathscr{H} \pm \mathscr{U}_{\perp}$

$$\zeta^{\perp}_{1\{a\}^s} = \pm\frac{i}{2\tilde{\omega}A}e^{-\zeta_0\rho}\left(\frac{\zeta_0^2}{\rho}+\frac{\zeta_0}{\rho^2}+\frac{1}{\rho^3}\right),$$

$$\zeta^{\perp}_{2\{a\}^s} = -\frac{e^{-2\zeta_0\rho}}{4\tilde{\omega}^2}\left(\frac{\zeta_0^2}{\rho}+\frac{\zeta_0}{\rho^2}+\frac{1}{\rho^3}\right) \times\left[\frac{i}{2\tilde{\omega}}\left(\frac{\zeta_0^2}{\rho}+\frac{\zeta_0}{\rho^2}+\frac{1}{\rho^3}\right)+\frac{\zeta_0}{\rho}-\zeta_0^2\right]. \tag{3.6c}$$

Obviously the complex conjugate values are also roots of these equations.

The imaginary and real parts of these roots describe the frequency of oscillations and decay constants for different cooperative modes of the system. Their dependence on the separation of oscillators was discussed, e.g., in Ref. 5.

Applying these principal poles, one can find approximate solutions for the motion of oscillators and radiation field. For the moment we will consider those parts of Eq. (2.13) which depend on the initial state of oscillators only. The contribution of the initial excitation of the electromagnetic field will be discussed in the next section, where we deal with the scattering.

The assymptotic motion for all four modes of the system is given by ($\eta = \omega_0\rho$)

$$\bar{v}_{\parallel\{a\}^s}(t) \simeq \bar{x}_{\parallel\{a\}^s}(0)\left(\omega_0 \mp \frac{2\gamma}{\rho^3\omega_0}\right) \times\mathrm{Re}\left\{ie^{\zeta^{\parallel}_{\{a\}^s}t}\left[1-\tfrac{1}{3}\gamma\Omega+\tfrac{2}{3}i\gamma\omega_0 \pm \gamma\omega_0 e^{-i\eta} \times\left(\frac{1}{\eta^3}+\frac{i}{\eta^2}-\frac{1}{\eta}\right)\right]\right\} + \bar{v}_{\parallel\{a\}^s}(0) \times\mathrm{Re}\left\{e^{\zeta^{\parallel}_{\{a\}^s}t}\left(1-\tfrac{2}{3}\gamma\Omega+i\gamma\omega_0 \mp \gamma\omega_0\frac{e^{-i\eta}}{\eta}\right)\right\}$$

$$+ \text{field contribution} \tag{3.7a}$$

$$\bar{v}_{\perp\{a\}^s} \simeq \bar{x}_{\perp\{a\}^s}(0)\left(\omega_0 \pm \frac{\gamma}{\rho^3\omega_0}\right) \times\mathrm{Re}\left\{ie^{\zeta^{\perp}_{\{a\}^s}t}\left[1-\tfrac{1}{3}\gamma\Omega+\tfrac{2}{3}i\gamma\omega_0 \pm \tfrac{1}{2}\gamma\omega_0 e^{-i\eta} \times\left(i-\frac{i}{\eta^2}-\frac{1}{\eta^3}\right)\right]\right\} + \bar{v}_{\perp\{a\}^s}(0) \times\mathrm{Re}\left\{e^{\zeta^{\perp}_{\{a\}^s}t}\left[1-\tfrac{2}{3}\gamma\Omega+i\gamma\omega_0 \pm \tfrac{1}{2}\gamma\omega_0 e^{-i\eta}\left(i-\frac{1}{\eta}\right)\right]\right\}$$

$$+ \text{field contribution.} \tag{3.7b}$$

The amplitudes of oscillations, given by the residua of $(\mathscr{H} \mp \mathscr{U}_{\parallel,\perp})^{-1}$ and $z(\mathscr{H} \mp \mathscr{U}_{\parallel,\perp})^{-1}$ include first order term in γ while the second order terms are taken into account in the oscillation frequency.

Velocities of charges (and hence positions) exhibit damped oscillations. Each normal component has its own life time and frequency. Therefore, the spectrum of radiation may be quite complicated. In the general case it is composed of four peaks (they may overlap) of different widths and hights.

For a small system $r_{12} \ll \lambda$ or $\eta \ll 2\pi$ the decay constants are equal to

$$\Gamma_{\parallel s} = \tfrac{2}{3}\gamma\omega_0^2(1-\tfrac{1}{20}\eta^2), \quad \Gamma_{\perp s} = \tfrac{2}{3}\gamma\omega_0^2(1-\tfrac{1}{10}\eta^2),$$
$$\Gamma_{\parallel a} = \tfrac{1}{30}\gamma\omega_0^2\eta^2, \quad \Gamma_{\perp a} = \tfrac{1}{15}\gamma\omega_0^2\eta^2. \tag{3.8}$$

The decay times for symmetric excitations are almost two times shorter than for a single isolated oscillator. That is caused by a cooperative emission. On the other hand, the antisymmetric vibrations have very long life time. Their decay is due to the higher multipole radiation, which is very weak.

The frequency shifts of different modes of the system measured with respect to the single oscillator frequency $\tilde{\omega}$, are equal to:

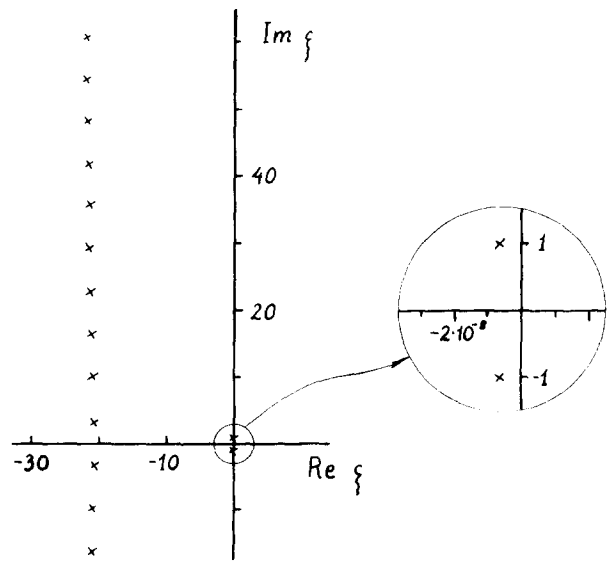


FIG. 1. Position of poles for the resolvent function of the longitudinal symmetric mode.

TABLE I.

		Position term $\operatorname{Re}\operatorname{Res}\frac{1}{\mathscr{H}-\mathscr{U}_{\parallel}}$	Velocity term $\operatorname{Re}\operatorname{Res}\frac{z}{\mathscr{H}-\mathscr{U}_{\parallel}}$
1	Principal pair of poles	1.2068×10^{-8}	$1-667.2\times10^{-8}$
2	First pair of subsidiary poles	$446\,007.3846\times10^{-8}$	-9.72×10^{-2}
3	Sum of N pairs of subsidiary poles		
	$N=10$	$444\,522.1\times10^{-8}$	-0.376
	$N=5000$	$1\,011.9534\times10^{-8}$	$-\frac{1}{2}+29112.8\times10^{-8}$
3'	$N=40000$	125.4436×10^{-8}	$-\frac{1}{2}+4486.4\times10^{-8}$
4	asymptotic expressions	-126.6514×10^{-8}	-3820.0×10^{-8}
	$1+3'+4$	-0.0008×10^{-8}	$\frac{1}{2}-0.8\times10^{-8}$

$$\Delta\omega_{\parallel\left\{\begin{smallmatrix}s\\a\end{smallmatrix}\right\}}=\mp\gamma\omega_0^2\left(\frac{1}{\eta^3}+\frac{1}{2\eta}-\frac{1}{8}\eta\right),$$
$$\Delta\omega_{\perp\left\{\begin{smallmatrix}s\\a\end{smallmatrix}\right\}}=\pm\frac{\gamma\omega_0^2}{2}\left(\frac{1}{\eta^3}-\frac{1}{2\eta}+\frac{3}{8}\eta\right). \tag{3.9}$$

The shifts of the symmetric parallel and transverse modes are of opposite sign. For the transverse symmetric mode the dipole Coulomb field of one oscillator increases the force acting on the second one, unlike the case for the parallel mode when this force is decreased.

It is evident that the solution based on principal poles is not valid for short times. Then, one has to take into account also contributions from all subsidiary poles. Similarly to the exponential polynomials, the resolvent functions have infinite number of poles (Fig. 1). We illustrate their role showing their contribution to match initial data. Without the subsidiary poles, the formulas (3.7) taken at $t=0$ are in a disagreement with the initial data. Although the disagreements are small, they are of the same order as the relevent physical quantities like the frequency shifts and decay constants, and it is worthwhile to comment about possible improvements.

Let us consider, as an example, the parallel symmetric mode connected with function $\mathscr{H}-\mathscr{U}_{\parallel}$. The initial data conditions require that

$$\lim_{t\to0^+}\left(g(t)=\int_\Gamma\frac{dz}{2\pi i}\frac{e^{zt}}{\mathscr{H}-\mathscr{U}_{\parallel}}\right)=0,$$
$$\lim_{t\to0^+}\left(f(t)=\int_\Gamma\frac{dz}{2\pi i}\frac{ze^{zt}}{\mathscr{H}-\mathscr{U}_{\parallel}}\right)=1. \tag{3.10}$$

Using standard method,[16] one can find the values of the asymptotic poles of the resolvent function valid for large integer k

$$z_k=-\frac{1}{\rho}\left(\ln\frac{\pi\rho k}{\gamma}-\frac{1}{4k}\right)-i\left[\frac{\pi}{\rho}(2k-\tfrac{1}{2})+\frac{\ln(\pi k\rho/\gamma)+1}{2\pi\rho k}\right]+O(k^{-2}) \tag{3.11}$$

For small k this formula is not valid, but may be used as a starting point for a numerical search of poles.

Fitting the initial data requires high accuracy and computer analysis. The computer gives the position of principle and subsidiary poles, calculated residua of the integrand functions, and sums them. The high accuracy is required because of a slow convergence of the series. The computer summed up contributions from 40 000 poles. The remaining terms were taken into account analytically using asymptotic formulas. Table I illustrates the procedure and final result. We have assumed $\rho\omega_0=1$, $\gamma\omega_0=10^{-8}$, $\Omega/\omega_0=10^3$. The improvement of the position term is evident. Notice that the first subsidiary poles give correction which exceeds the initial discrepency by five orders of magnitude. A slow convergence of the series of residua results in the fact that with 40 000 of them the deviation is still two orders of magnitude worse than the initial error. Then, however, we can use the asymptotic formula, matching finally the initial data three orders of magnitude better than the principal poles contribution did.

The properties of the velocity term are not so evident. The initial discrepancy from 1 seems to be equal to 7×10^{-6}. Adding all contributions, we get the value very close to $\frac{1}{2}$, namely $\frac{1}{2}-0.8\times10^{-8}$. This is because, summing all residua, we get $f(0)$ not $f(0^+)$, which fits the initial data. As the function $f(t)$ is discontinuous and, as $f(t<0)=0$, we have $f(0^+)=2f(0)$ and with the help of residua we get $f(0^+)=1-1.6\times10^{-8}$, which is quite good even in comparison with the erroneous treatment of the "principal poles" values. It remains, however, a puzzle as to why the value obtained with principal poles only is so clase to 1 which might suggest that the assymptotic function is also good for short times.

The above test shows how crucial all subsidiary poles are for short time behavior. The subsidiary poles are separated from the imaginary axis by a distance $\sigma\simeq(1/\rho)\ln(\pi\rho/\gamma)$. Therefore, their contribution dies in time with the decay time $T_s=\sigma^{-1}$, which is a fraction of the transit time for a light passing between oscillators. This does not mean that the asymptotic solutions based on the principal poles are valid for such short times. Large values of residua at the principal poles results in the fact that their contribution survives for many decay times T_s. The second method, which will be discussed now, shows that the asymptotic solutions given by Eqs. (3.7) are valid after several transit times.

The above discussion shows that it is not very convenient to describe a short time behavior using the first method. In that case the second, based on the single oscillator poles, is much more convenient. It is illustrated again for the $\bar{v}_{\|s}$ mode.

The resolvent function entering the integrands can be expanded as follows:

$$\frac{1}{\mathscr{H}-\mathscr{U}_{\|}}=\sum_{n=0}^{\infty}\frac{\mathscr{U}_{\|}^{n}}{\mathscr{H}^{n+1}}. \tag{3.12}$$

Applying this expansion, we can write

$$\begin{aligned}\bar{v}_{\|s}(t)=&\sum_{n=0}^{\infty}\int\frac{dz}{2\pi i}(2\gamma)^{n}\frac{e^{z(t-n\rho)}(z/\rho^{2}+1/\rho^{3})}{[\mathscr{H}(z)]^{n+1}}\\&\times\left[-\left(\omega_{0}^{2}-\frac{2\gamma}{\rho^{3}}\right)\bar{x}_{\|s}(0)+z\bar{v}_{\|s}(0)\right]\\&+\text{field contribution}\\=&-\left(\omega_{0}^{2}-\frac{2\gamma}{\rho^{3}}\right)\bar{x}_{\|s}(0)g(t)+f(t)\bar{v}_{s}(0)\\&+\text{field contribution}\end{aligned} \tag{3.13}$$

In spite of the fact that the upper limit of the sum is infinite, for finite times there are only a finite number of terms, namely $N=\mathrm{int}(t/\rho)$, which contribute to the sum. Indeed, for $n>N$ the contour Γ can be shifted to the right to $+\infty$ and all integrals vanish.

The poles which determine the values of relevant terms coincide with the poles of one oscillator resolvent function $\mathscr{H}^{-1}$. The order of these poles in subsequent terms grows. It is therefore difficult to use this method for exact treatment of the long time behavior. On the other hand, one has no difficulties for short times. For example, it is very easy to check the initial data condition. Then, there is only one term relevant (with $n=0$), and because the point at infinity has no essential singularity, the value of $\bar{v}(0^{+})$ is given by the limit

$$\bar{v}_{\|s}(0^{+})=\lim_{z\to\infty}z\tilde{\bar{v}}_{\|s}(z)=\bar{v}_{\|s}(0). \tag{3.14}$$

Before the light signal passes the distance between oscillators, the motion of each oscillator is not modified by the presence of the second one. At $t=\rho$ we have to add a new term to $\bar{v}(t)$. It expresses the mutual influence of oscillators on their evolution, which is due to radiation emitted by one oscillator and absorbed by the second oscillator. The third term, which we must include for $t\geqslant 2\rho$, includes also the radiation emitted by one oscillator scattered by the second one and reabsorbed by the first one. This effect interferes with another one caused by spontaneous radiation emitted at time $\rho<t<2\rho$ by the second oscillator. As time grows, the number of possible processes dramatically increases.

For long time behavior not all the processes are of equal importance. Due to the small value of the coupling parameter ($\gamma\omega_{0}\approx 10^{-8}$ for atomic parameters) the subsequent terms give smaller contributions. For large t, one can sum the leading terms finding a proper asymptotic behavior found within the first method. The procedure below, extends the method presented in Ref. 6.

Substituting in (3.13) the approximate expression (3.4) for $\mathscr{H}$ and calculating the residua, we get (we skip indices $\|,s$)

$$\begin{aligned}g(t)=&2\mathrm{Re}\sum_{n=0}^{\mathrm{int}(t/\rho)}\frac{1}{A}\left(\frac{2\gamma}{A}\right)^{n}\frac{1}{n!}\frac{d^{n}}{dt^{n}}\\&\times\frac{e^{z(t-n\rho)}(z/\rho^{2}+1/\rho^{3})^{n}}{(z-\zeta_{0}^{*})^{n+1}}\Bigg|_{z=\zeta_{0}}\\\simeq&\mathrm{Re}\frac{e^{\zeta_{0}t}}{i\tilde{\omega}A}\sum_{n=0}^{\mathrm{int}(t/\rho)}\frac{\alpha^{n}}{n!}[t^{n}+t^{n-1}(an^{2}+bn)+\cdots],\end{aligned}$$

$$\alpha=\frac{\gamma}{i\tilde{\omega}A}e^{-\zeta_{0}\rho}\left(\frac{\zeta_{0}}{\rho^{2}}+\frac{1}{\rho^{3}}\right)=\gamma\zeta_{1},$$

$$a=-\frac{\zeta_{0}}{\rho(\zeta_{0}/\rho^{2}+1/\rho^{3})}-\frac{1}{2i\tilde{\omega}},\quad b=-\frac{1}{2i\tilde{\omega}}. \tag{3.15}$$

For times which are of the order of the decay time and which are of a physical interest, only a certain number of initial terms, much less than $N=\mathrm{int}(t/\rho)$, determine the value of the function $g(t)$. As the remaining terms are very small, we will make a very small error extending the summation to infinity. Then we get

$$\begin{aligned}g(t)=&\mathrm{Re}\frac{e^{(\zeta_{0}+\gamma\zeta_{1})t}}{i\omega_{0}}\Big[1-\tfrac{1}{3}\gamma\Omega+\tfrac{2}{3}i\omega_{0}\gamma\\&+\gamma\tilde{\omega}e^{-i\eta}\left(\frac{1}{\eta^{3}}+\frac{i}{\eta^{2}}-\frac{1}{\eta}\right)+\gamma^{2}\zeta_{2}t\Big],\end{aligned} \tag{3.16}$$

This expression is in agreement with formulas (3.7a) found using principal poles. Only the second order correction to the position of the pole, $\gamma^{2}\zeta_{2}$, appears here as the time dependent amplitude correction.

In the same way one can find the agreement for the velocity factor $f(t)$. Milloni and Knight[6] kept only t^{n} terms, getting the proper first order correction to frequency shift and life time but not the amplitude.

IV. INSTABILITY REGION

Up to now we have considered normal situations when the separation of oscillators exceeds a certain critical distance and the resolvent functions have no poles in the right half-plane of complex z plane. If the separation is smaller, there are poles to the right from the imaginary axis, which means instability in the model. Similar instability was also found in the scalar field model in Ref. 8.

Consider the function $\mathscr{H}(z)-\mathscr{U}_{\|}(z)$, which describes the symetric vibration of longitudinal mode. This function has one real and positive zero, if

$$\omega_{0}^{2}<2e^{2}/mr^{3}. \tag{4.1}$$

Indeed, then $\mathscr{H}(0)-\mathscr{U}_{\|}(0)<0$ but $\lim_{z\to+\infty}[\mathscr{H}(z)-\mathscr{U}_{\|}(z)]=+\infty$. Due to this zero and instability, the total dipole moment of oscillators is exponentially growing in time.

This instability is exactly the same as the recently discussed[10,11] appearance of the macroscopic polarization in the system of two-level atoms interacting with the electromagnetic field, with Coulomb dipole–dipole interaction taken into account. This polarization develops if the density of two-level atoms exceeds a certain critical value. A similar instability occurs also in a spherical system of many oscillators.[17]

One can easily understand that linear character of oscil-

lators produces unbounded effects once a distance between oscillators is smaller than the critical one.

Since the "catastrophe" we are discussing now is at zero frequency (for real z), it can be easily understood from purely electrostatic considerations.

Two oscillators interacting through dipole–dipole force obey the following Newton equations:

$$m\ddot{\bar{x}}_{1,2} = -m\omega_0^2\bar{x}_{1,2} - e^2\frac{\hat{I}-3\,\bar{n}\otimes\bar{n}}{r^3}\bar{x}_{2,1}. \tag{4.2}$$

For the longitudinal mode $\bar{x}_{\|s} = \bar{n}[\bar{n}\cdot(\bar{x}_1+\bar{x}_2)]$ we get

$$m\ddot{\bar{x}}_{\|s} = (2e^2/r^3 - m\omega_0^2)\bar{x}_{\|s}, \tag{4.3}$$

from which we get immediately the instability condition (4.1).

There are two remarks to be added:

(i) One can easily get instability with finite displacement of both oscillators (spontaneous generation of finite total dipole moment) by taking into account Coulomb interaction between different charges constituting oscillators exactly and not in dipole–dipole form.

(ii) If elastic force binding oscillators is of electrostatic origin, ω_0 is not a free parameter, but $\omega_0^2 = e^2/md^3$. Therefore, the instability condition requires such two extended atoms to overlap. In this case none of this results really holds. For this reason there is probably also very little physics in the static phase transition for two-level atoms. The density required would cause electronic shells of atoms to overlap.

V. SCATTERING PROBLEM

The scattering problem was considered by Lyuboshitz,[2] who applied equations describing the scattering by many centers. Now, we shall present the derivation of the scattering cross section starting from our general solution of the initial value problem. In this solution, we have distinguished terms which describe the scattering of electromagnetic radiation. These terms contain initial field operators $a_{\bar{p}\mu}(0)$ and $a^+_{\bar{p}\mu}(0)$ [or equivalently $\bar{E}_T(\bar{r},0)$ and $\bar{B}(\bar{r},0)$]. This scattered part of the photon creation operator $a^+_{\bar{k}\mu}(z)$ is given by

$$a^{+\,sc}_{\bar{k}\mu}(z) = -\frac{e^2}{4\pi^2 m}\frac{z}{(z-i\omega)\omega^{1/2}}\sum_\nu\int d^3p\,p^{1/2} \times\left(\frac{a^+_{\bar{p}\nu}(0)}{z-i\omega_p}F_{\mu\nu}(\bar{k},\bar{p};z) + \frac{a_{\bar{p}\nu}(0)}{z+i\omega_p}F^*_{\mu\nu}(\bar{k},\bar{p};z)\right),$$

where

$$F_{\mu\nu}(\bar{k},\bar{p};z) = (e^{i(\bar{k}-\bar{p})\bar{r}_1} + e^{i(\bar{k}-\bar{p})\bar{r}_2})[\alpha(z)(\bar{e}_{\bar{p}\nu}\cdot\bar{n})(\bar{e}_{\bar{k}\mu}\cdot\bar{n}) + \beta(z)(\bar{e}_{\bar{p}\nu}\cdot\bar{e}_{\bar{k}\mu})] + (e^{i(\bar{k}\bar{r}_1-\bar{p}\bar{r}_2)} + e^{i(\bar{k}\bar{r}_2-\bar{p}\bar{r}_1)}) \times (\bar{e}_{\bar{p}\nu}\cdot\bar{n})(\bar{e}_{\bar{k}\mu}\cdot\bar{n}) + \delta(z)(\bar{e}_{\bar{p}\nu}\cdot\bar{e}_{\bar{k}\mu})]$$

and the functions $\alpha(z)$, $\beta(z)$, $\gamma(z)$, and $\delta(z)$ are given by Eqs. (2.15).

We have previously used this expression for reconstruction of the scattered field in space and in the discussion of its causal properties. However, in a typical scattering experiment such a detailed information about the scattered field is not controlled. What is usually measured is a scattering cross section.

To extract the scattering cross section, we have to specify the initial data for the field to be a monochromatic plane wave. Assuming that the field is in a coherent state, it is enough to know mean values for the field strengths:

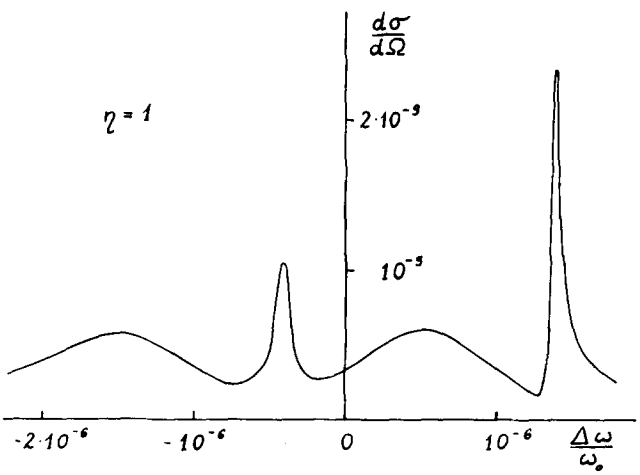
$$\langle\bar{E}_T(\bar{r},t=0)\rangle = \bar{\epsilon}_0\mathscr{E}_0\sin(\bar{p}\cdot\bar{r}), \tag{5.1a}$$

$$\langle\bar{B}(\bar{r},t=0)\rangle = (\hat{p}\times\bar{\epsilon}_0)\mathscr{E}_0\sin(\bar{p}\cdot\bar{r}). \tag{5.1b}$$

Using relation (2.21), we find that the expectation values for $a_{\bar{q}\nu}(0)$ and $a^+_{\bar{q}\nu}(0)$ are

$$\sum_\nu\bar{e}_{\bar{q}\nu}\langle a_{\bar{q}\nu}(0)\rangle = -\frac{2\pi}{(\omega_p\hbar)^{1/2}}\mathscr{E}_0\bar{\epsilon}_0\delta^{(3)}(\bar{q}-\bar{p}) = \sum_\nu\bar{e}_{\bar{q}\nu}\langle a^+_{\bar{q}\nu}(0)\rangle. \tag{5.2}$$

With the help of these expressions we can compute expecta-



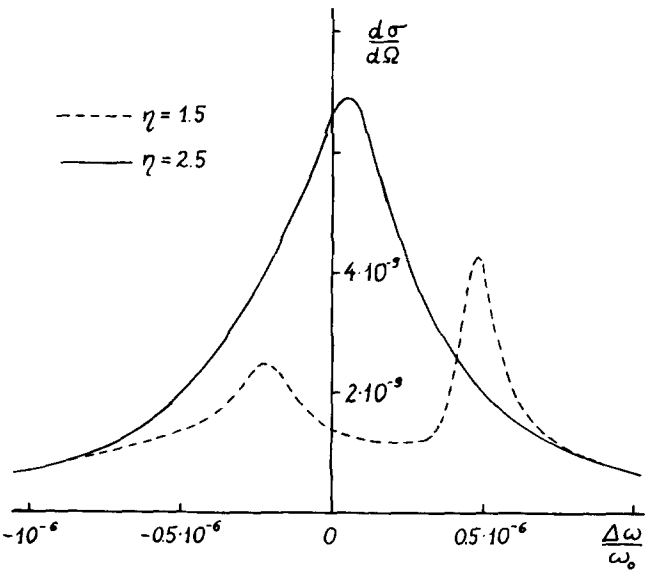


FIG. 2. Differential cross section as a function of frequency, $\Delta\omega = \omega - \omega_0(1 - \frac{1}{3}\gamma\Omega)$ for different separation of oscillators. The scattering parameters are $\mathbf{n} = (1/2^{1/2})(1,1,0)$, $\mathbf{k}_i = (0,1,0)$, $\mathbf{k}_f = (1/2^{1/2})(0,1,1)$, $\mathbf{e}_i = \mathbf{e}_f = (1,0,0)$.

tion values of the scattered part of the field. When doing so, we get terms of different type. For $t\to\infty$ we can neglect terms resulting from poles of the resolvent functions $(\mathscr{H}\pm\mathscr{U}_{\parallel})^{-1}$ and $(\mathscr{H}\pm\mathscr{U}_{\perp})^{-1}$ (they are damped) and counter-rotating terms because of their nonresonant and purely oscillatory character.

Remaining is the expression of the following type:

$$\langle a^{+\,\mathrm{sc}}_{\bar{k}\mu}(t)\rangle \underset{t\to\infty}{\to} +i\frac{e^2}{4\pi^2 m}\omega^{1/2}\frac{2\pi}{\hbar^{1/2}}\mathscr{E}_0 F_{\mu 0}(\bar{k},\bar{q};i\omega) \times\frac{e^{i\omega t}-e^{i\omega_q t}}{i(\omega-\omega_q)}. \tag{5.3}$$

The density of the energy flux for scattered light in the direction $\hat{k}$ and polarization μ is given by

$$\bar{S}_{\bar{k}\mu}(\omega,t)=\hat{k}\hbar\frac{\omega^3}{c^3}a^{+\,\mathrm{sc}}_{\bar{k}\nu}(t)a^{\mathrm{sc}}_{\bar{k}\nu}(t), \tag{5.4}$$

where the normalization is such that $\Sigma_\mu\int d\omega\int d\Omega\,\bar{S}_{\bar{k}\mu}$ gives the total energy flux of the scattered field. The expectation value of the operator $\bar{S}_{\bar{k}\mu}(\omega,t)$ can be calculated with the help of the mean values $\langle a^{\mathrm{sc}}_{\bar{k}\mu}(t)\rangle$ and $\langle a^{+\,\mathrm{sc}}_{\bar{k}\mu}(t)\rangle$. The other terms are negligible when $t\to\infty$.

The differential cross section is defined by the limit

$$\frac{d\sigma_\mu(\bar{k})}{d\Omega}=\lim_{t\to\infty}\frac{\langle\bar{S}_{\bar{k}\mu}(\omega,t)\rangle}{t\,|\bar{S}|}, \tag{5.5}$$

where $|\bar{S}|=(c/2\pi)\mathscr{E}_0^2$ is the mean density of the energy flux for the initial field. The time t appearing in the denominator gives the density of energy flux of scattered light per unit time.

The linear growth of the numerator is recovered from the familiar formula

$$\lim_{t\to\infty}\{[1-\cos(\omega-\omega_P)t\,]/(\omega-\omega_P)^2\}\to\pi t\delta(\omega-\omega_p). \tag{5.6}$$

The presence of the function $\delta(\omega-\omega_p)$ means that we have only the elastic scattering. For the scattering differential cross section we get

$$\frac{d\sigma_\mu(\omega)}{d\Omega_{\hat{k}}}=\frac{e^2\omega^4}{m^2c^4}|F_{\mu\nu}(\bar{k},\bar{p};i\omega)|^2, \tag{5.7}$$

and, writing $|F_{\mu\nu}|^2$ explicitly,

$$\begin{aligned}\frac{d\sigma_\mu(\omega)}{d\Omega}=2\frac{e^2\omega^4}{m^2c^4}&(\{1+\cos[\bar{r}_{12}\cdot(\bar{k}-\bar{p})]\}\{|\alpha(i\omega)|^2A^2_{\mu 0}\\&+|\beta(i\omega)|^2B^2_{\mu 0}+2A_{\mu 0}B_{\mu 0}\,\mathrm{Re}[\alpha(i\omega)\beta^*(i\omega)]\}\\&+\{1+\cos[\bar{r}_{12}\cdot(\bar{k}+\bar{p})]\}\{|\gamma(i\omega)|^2A^2_{\mu 0}\\&+|\delta(i\omega)|^2B^2_{\mu 0}+2A_{\mu 0}B_{\mu 0}\,\mathrm{Re}[\gamma(i\omega)\delta^*(i\omega)]\}\\&+2(\cos\bar{r}_{12}\cdot\bar{p}+\cos\bar{r}_{12}\cdot\bar{k})\,\mathrm{Re}\{A^2_{\mu 0}\alpha(i\omega)\gamma^*(i\omega)\\&+B^2_{\mu 0}\beta(i\omega)\delta^*(i\omega)+A_{\mu 0}B_{\mu 0}[\alpha(i\omega)\delta^*(i\omega)\\&+\beta(i\omega)\gamma^*(i\omega)]\}),\end{aligned} \tag{5.8}$$

where

$$A_{\mu 0}=(\bar{e}_{\bar{k}\mu}\cdot\bar{n})(\bar{n}\cdot\bar{\epsilon}_0),\quad B_{\mu 0}=\bar{e}_{\bar{k}\mu}\cdot\bar{\epsilon}_0.$$

This expression exhibits resonant and interference effects in the scattering. The spectrum is composed of four peaks resulting from principal poles of the functions α, β, γ, and δ. For separation of oscillators $r_{12}>\lambda$ these peaks overlap, producing single resonance. For $r\ll\lambda$ all peaks are separated. In that case, in the vicinity of all resonances we can get approximate expressions for the cross section. Near the resonance of the transverse symmetric mode we have

$$\delta=\beta=-\alpha=-\gamma=1/2(\mathscr{H}+\mathscr{U}_\perp),$$

and for the cross section we get

$$\frac{d\sigma_\mu}{d\Omega_{\hat{k}}}=4\frac{e^4\omega^4}{m^2c^3}[(\bar{e}_{\bar{k}\mu}\cdot\bar{n})(\bar{\epsilon}_0\cdot\bar{n})-(\bar{e}_{\bar{k}\mu}\cdot\bar{\epsilon}_{0)}]^2 \times\frac{1-\frac{1}{4}(\bar{k}\bar{r})^2-\frac{1}{4}(\bar{p}\bar{r})^2+\cdots}{(\omega^2-\omega_{s\perp}^2)^2+(\frac{4}{3}e^2\omega_0^3/mc^3)^2}. \tag{5.9}$$

For the resonance of the transverse antisymmetric mode the corresponding expressions are

$$\beta=\gamma=-\alpha=-\delta=\tfrac{1}{2}\frac{1}{(\mathscr{H}-U_\perp)},$$

and

$$\frac{d\sigma_\mu}{d\Omega_{\hat{k}}}=\frac{1}{4}\frac{e^4\omega^4}{m^2c^3}[(\bar{e}_{\bar{k}\mu}\cdot\bar{n})(\bar{\epsilon}_0\cdot\bar{n})-(\bar{e}_{\bar{k}\mu}\cdot\bar{\epsilon}_{0)}]^2 \times\frac{(\bar{r}\cdot\bar{k})^2(\bar{r}\cdot\bar{p})^2}{(\omega^2-\omega_{a\perp}^2)^2+[\frac{2}{15}(e^2\omega_0^3/mc^3)\eta^2]^2}. \tag{5.10}$$

For the longitudinal mode resonances the polarization term is replaced by $[(\bar{\epsilon}_0\cdot\bar{n})(\bar{e}_{\bar{k}\mu}\cdot\bar{n})]^2$ and in the denominator we have the functions $|\mathscr{H}-\mathscr{U}_{\parallel}|^2$ and $|\mathscr{H}+\mathscr{U}_{\parallel}|^2$ for symmetric and antisymmetric modes respectively.

Notice that at the exact resonances the magnitudes of the cross section for the symmetric and antisymmetric modes are of the same order. Only the widths of the antisymmetric modes are much smaller than width of the symmetric ones, making the symmetric resonances much stronger and easier to detect.

Some examples of the geometrical and spectral properties of the scattering are displayed in Figs. 2a, 2b and 3.

We point out Fig. 3, which shows the situation when the back scattering exceeds 2.5 times the forward scattering.

Summing over the final polarizations and integrating over directions of the scattered radiation, one can find the total cross section. Below, we give the value of the total cross section for unpolarized incoming radiation when both oscillators, separated by a distance r, are randomly oriented:

$$\begin{aligned}\sigma(\omega,r)=\frac{16\pi e^4\omega^4}{m^2c^4}\Bigg(&[|\alpha|^2+|\gamma|^2+2\,\mathrm{Re}(\alpha\beta^*+\gamma\delta^*)]\\&\times\left[\frac{1}{9}+\left(\frac{\sin\eta}{\eta^3}-\frac{\cos\eta}{\eta^2}\right)^2\right]+(|\beta|^2+|\delta|^2)\\&\times\left\{\frac{1}{3}+\frac{1}{2}\left[\frac{\sin\eta}{\eta}\left(\frac{\sin\eta}{\eta}+\frac{\cos\eta}{\eta^2}-\frac{\sin\eta}{\eta^3}\right)\right.\right.\\&\left.\left.-\left(\frac{\sin\eta}{\eta^3}-\frac{\cos\eta}{\eta^2}\right)\left(\frac{\sin\eta}{\eta}+3\frac{\cos\eta}{\eta^2}-3\frac{\sin\eta}{\eta^3}\right)\right]\right\}\\&+\frac{4}{3}\,\mathrm{Re}(\alpha\gamma^*+\alpha\delta^*+\beta\gamma^*)\left(\frac{\sin\eta}{\eta^3}-\frac{\cos\eta}{\eta^2}\right)\\&+\frac{4}{3}\frac{\sin\eta}{\eta}\,\mathrm{Re}(\beta\delta^*)\Bigg).\end{aligned} \tag{5.11}$$

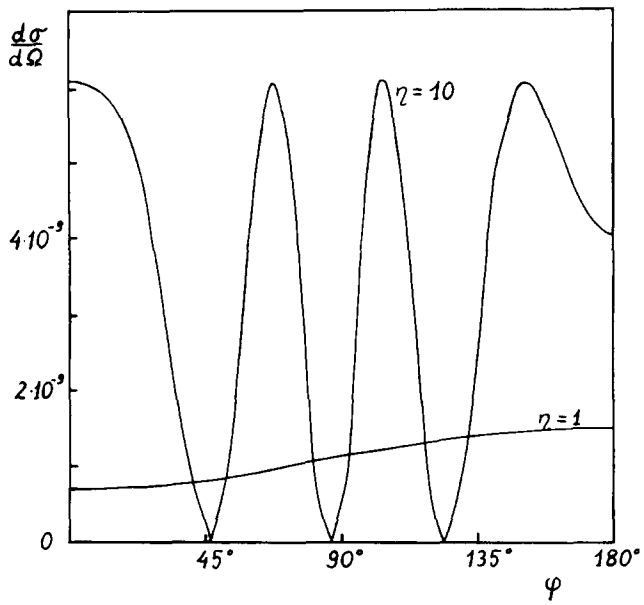


FIG. 3. Angular dependence of the cross section. The scattering geometry is defined by $\mathbf{n} = \mathbf{k}_i = (0,0,1)$ $\mathbf{e}_i = \mathbf{e}_f = (0,1,0)$, $\mathbf{k}_f = (\sin\varphi, 0, \cos\varphi)$, and $\omega = \omega_0(1 - \frac{1}{3}\gamma\Omega)$.

The total cross section for different oscillator separations are shown in Fig. 4.

When the separation of oscillators $r \to \infty$, $\alpha = \gamma = \delta = 0, \beta = 1/\mathscr{H}$, and we get

$$\sigma(\omega) = 16\pi e^2\omega^4/3m^2c^4 |\mathscr{H}(i\omega)|^2, \tag{5.12}$$

which is two times bigger than for one oscillator.

In the opposite case, when separation of oscillators tends to zero, we get

$$\sigma(\omega) = 16\pi e^2\omega^4/9m^2c^4 |\mathscr{H} + U_\perp|^2 \tag{5.13}$$

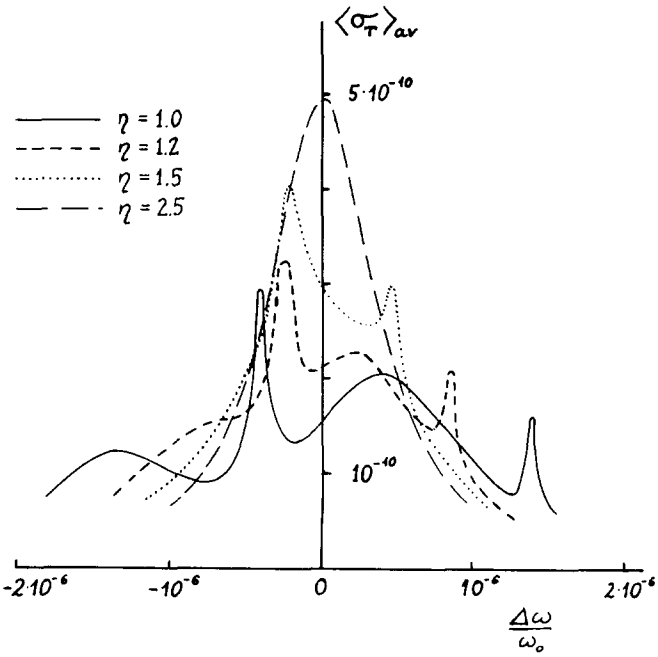


FIG. 4. Averaged total cross section as function of frequency for different separation of oscillators.

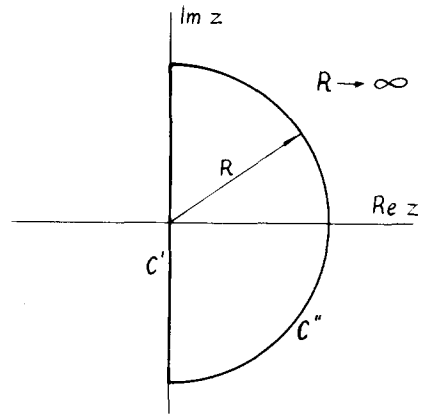


FIG. 5. The contour used for the principle of argument.

near the resonance of the transverse symmetric mode and

$$\sigma(\omega) = 32\pi e^2\omega^4/9m^2c^4 |\mathscr{H} - U_\parallel|^2 \tag{5.14}$$

near the resonance of the parallel symmetric mode. The last expressions maintain the coherent character of the scattering by two oscillators.

APPENDIX

We are going to consider the possibility of zeros of the functions $\mathscr{H} \pm \mathscr{U}_\parallel$ lying in the right-half plane (RHP) of complex z plane. These functions written explicitly are ($\xi = z/\omega_0$)

$$F_\pm(\xi) = 1 + \xi^2\left(1 + \frac{2}{3}\gamma\frac{\Omega^2}{\Omega + \xi}\right) \pm 2\gamma e^{-\xi\eta}\left(\frac{\xi}{\eta^2} + \frac{1}{\eta^3}\right). \tag{A1}$$

Since they have no poles in the RHP, the number of zeros is equal to the increment of argument when the point is enclosing the RHP (see Fig. 5).

The change of argument on the semicircle C'' (with $R \to \infty$)is

$$\Delta \arg(F_\pm)_{C''} = \Delta \arg R^2 e^{2i\varphi}[1 + O(R^{-2})]\,|_{-\pi/2}^{\pi/2} = 2\pi. \tag{A2}$$

On the imaginary axis $\xi = iy$

$$\mathrm{Re}[F_\pm(iy)] = 1 - y^2\left(1 + \frac{2}{3}\gamma\frac{\Omega^3}{\Omega^2 + y^2}\right) \pm 2\gamma\left(\frac{\cos y\eta}{\eta^3} + \frac{y\sin y\eta}{\eta^2}\right), \tag{A3}$$

$$\mathrm{Im}[F_\pm(iy)] = \frac{2}{3}\gamma\left[\frac{\Omega^2 y^3}{\Omega^2 + y^2} \pm 3\left(\frac{y\cos y\eta}{\eta^2} + \frac{\sin y\eta}{\eta^3}\right)\right]. \tag{A4}$$

When $\Omega\eta > 4$, one can numerically check that

$$\begin{aligned} &\mathrm{Im}[F_\pm(iy)] > 0 \quad \text{for} \quad \infty > y > 0, \\ &\mathrm{Im}[F_\pm(iy)] < 0 \quad \text{for} \quad 0 > y > -\infty, \end{aligned} \tag{A5}$$

i.e., the contour $(\mathrm{Re}F_\pm, \mathrm{Im}F_\pm)$ crosses the real $F_\pm$ axis only for $y = 0$. Always for F_+ and for F_- when $2\gamma < \eta^3$, these crossing points lie on the positive part of $\mathrm{Re}F_\pm$ axis. Therefore, the contours $F_\pm$ encircle the center of coordinate system and, when y varies from $+\infty$ to $-\infty$, the change of $\arg F_\pm$ is equal to -2π. The total change of the $\arg F_\pm$ is

equal to 0, which proves that there are no zeros of $F_{\pm}$ in the RHP of the z plane.

When $2\gamma > \eta^3$, the crossing points for F_{-} lie on the negative part of the ReF_{-} axis. There is no change of argF_{-} when ξ moves along the imaginary axis. The total change of the argument of F_{-} remains equal to 2π which shows that there is one root of F_{-} lying in the RHP of z plane. This root is real and is responsible for the instability in the model discussed in Sec. IV.

The solutions of the model we have been discussing are valid only for stable situations, and then the condition is always satisfied.

A similar discussion can be given for transverse modes and functions $H \pm U_{\perp}$.

[1]J. Stephen, J. Chem. Phys. **40**, 669 (1964).
[2]V.L. Lyuboshits, Zh. Eksp. Teor. Fiz. **52**, 926 (1967) [Sov. Phys. JETP **25**, 612 (1967)].
[3]H. Morawitz, Phys. Rev. **187**, 1792 (1969).
[4]H. Steudel, Ann. Physik **7**, 57 (1971).
[5]R.H. Lehmberg, Phys. Rev. A **1**, 883 (1970).
[6]P.W. Milloni and P.L. Knight, Phys. Rev. A **10**, 1096 (1974).
[7]P.C. Aichelburg, Phys. Rev. Lett. **38**, 451 (1977).
[8]P.C. Aichelburg and H. Grosse, Phys. Rev. D **16**, 1900 (1977).
[9]K. Drexhage, H. Kuhn, and F. Schaefer, Ber. Bunsenges. Phys. **72**, 329 (1968).
[10]V.I. Emelianov and Yu.L. Klimontovich, Phys. Lett. A **59**, 366 (1976).
[11]M. Yamanoi and M. Takatsuji, in *Coherence and Quantum Optics IV*, edited by L. Mandel and E. Wolf (Plenum, New York, 1978).
[12]K. Rzażewski and W. Żakowicz, J. Phys. A **9**, 1959 (1976).
[13]E.A. Power and S. Zienau, Trans. R. Phil. Soc. London A **251**, 427 (1959).
[14]L. Davidovich, PhD thesis, University of Rochester, 1975 (unpublished).
[15]K. Rzażewski in Ref. 11.
[16]R. Bellman and K.L. Cooke, *Differential-Difference Equations* (Academic, New York, 1963).
[17]W. Żakowicz, Phys. Rev. A **17**, 343 (1978).

Electromagnetic scattering by an infinite inhomogeneous dielectric cylinder: New Green's function and integral equations

R. F. Alvarez-Estrada
Departamento de Física Teórica, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

M. L. Calvo
Departamento de Optica y Estructura de la Materia, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

(Received 7 August 1979; accepted for publication 20 September 1979)

The scattering of electromagnetic waves by an infinite dielectric cylinder with variable dielectric permeability presents, in general, certain mathematical difficulties regarding the construction of rigorous solutions. We give a new divergenceless tensor Green's function, specially appropriate for cylindrical symmetry, and, in terms of it, present new scattering integral equations. We prove that the series for med by all successive iterations of those scattering integral equations converge under certain conditions. Suitable transformations lead to new integral equations with Hilbert–Schmidt kernels, which imply further rigorous results.

I. INTRODUCTION

The scattering of electromagnetic (e.m.) waves by physical systems which can be assimilated to infinite cylinders constitutes an important branch of e.m. scattering theory. Moreover, it also provides a useful tool for investigating the dielectric properties of those systems, as it often happens in the fields of radio- and microwaves, and even of optics (plasma columns, meteor trails consisting of long columns of ionized gas, optical fibers,···). Compared to e.m. scattering by spherical systems, the corresponding study for infinite cylinders appears as less developed. Explicit solutions are known for homogeneous dielectric cylinders, while the more difficult case of inhomogeneous ones, of considerable physical and mathematical interest as well (our previous examples are, frequently, inhomogeneous dielectric cylinders), has been less explored. On the other hand, most research has concentrated on the case where the incoming wave vector is orthogonal to the axis of the cylinder, although the situation corresponding to oblique incidence has also received attention. As a suitable set of references on e.m. scattering by cylinders and related subjects.[1-9]

A general study of e.m. scattering by infinite dielectric cylinders faces, among others, the following interrelated difficulties: (a) the vector and transverse (divergenceless) nature of e.m. fields in situations showing cylindrical symmetry, (b) the case of oblique incidence, (c) for inhomogeneous cylinders, the wave equations and, then, the associated scattering integral equations contain additional derivatives of fields, which come from the very structure of Maxwell's equations and of the matter equations.

In this paper, we shall present a general and rigorous approach to the infinite cylinder scattering problem which solves the above difficulties (a), (b), and (c). Our main results are: (i) we give a new divergenceless tensor Green's function, G^{CT}, specially appropriate for cylindrical symmetry, in order to cope both with difficulties (a) and (b) (Sec. 3), (ii) we present new scattering integral equations in terms of G^{CT} and, through suitable transformations, we prove that the series formed by the iterations of those integral equations converge under certain conditions, thereby solving the above difficulty (c) as well (Sec. 4), (iii) we reduce the scattering integral equations to a new set with Hilbert–Schmidt kernels, which implies further convergence results (Sec. 5). Some approximate estimates of the convergence conditions are also presented (Secs. 4 and 5).

II. TIME-INDEPENDENT FORMULATION OF SCATTERING PROBLEM

We consider a macroscopically homogeneous and isotropic infinite medium (say, the air), with constant dielectric permeability $\epsilon_0 = 1$, vanishing conductivity and magnetic permeability equal to that of the vacuum. Inside it, there is a dielectric cylinder of infinite length, whose cross section Ω has an arbitrary shape (circular, elliptic,···). The coordinate system is chosen so that the x_3-axis is parallel to the axis of the cylinder and $\bar{u}_3$ will denote a unit vector along it. The position of a point in space is represented by $(\boldsymbol{\rho}, x_3)$, where $\boldsymbol{\rho} = (x_1, x_2)$ is a two-dimensional vector orthogonal to $\bar{u}_3$.

The dielectric cylinder has variable dielectric permeability $\epsilon(\boldsymbol{\rho})$ ($\boldsymbol{\rho}$ varying inside Ω) and the same conductivity and magnetic permeability as the surrounding infinite medium.

A classical monochromatic (e.m.) wave propagates in the remote past at infinite distance from the cylinder with wave vector $\bar{k} = (\mathbf{k}, k_3)$. The two-dimensional wave vector $\mathbf{k} = (k_1, k_2)$ is orthogonal to $\bar{u}_3$ and $k = (\mathbf{k}\cdot\mathbf{k})^{1/2} > 0$. We represent the e.m. wave by the magnetic field $\bar{H}_0(\bar{k},\lambda) \times \exp i(\mathbf{k}\boldsymbol{\rho} + k_3 x_3)$ (time-dependent terms being factored out in this paper). Here, $\bar{H}_0(\bar{k},\lambda)$ is a complex polarization vector such that $\bar{k}\cdot\bar{H}_0(\bar{k},\lambda) = 0$ (the polarization index λ takes only two values), and

$$\sum_{\lambda} H_0(\bar{k},\lambda)_\alpha H_0^*(\bar{k},\lambda)_\beta = \delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2 + k_3^2}, \quad \alpha,\beta = 1,2,3. \tag{2.1}$$

Let $\bar{H}(\boldsymbol{\rho})\cdot\exp ik_3 x_3$ be the total complex magnetic field describing the propagation of the e.m. wave and its scattering by the cylinder. The reason for using $\bar{H}$ instead $\bar{E}$ or $\bar{D}$ is due to the simplification for the mathematical developments.

 0022-2488/80/020389-06$1.00

The fact that $\epsilon(\rho)$ be x_3-independent leads naturally to factor out $\exp ik_3 x_3$. Maxwell's equations[10,11] and the above factorization lead to the following basic equations for $\bar{H} = (H_\alpha)$, $\alpha = 1,2,3$:

$$(\Delta_\rho + k^2)\bar{H} = \bar{j}, \quad \sum_{\alpha=1}^{2} \frac{\partial}{\partial x_\alpha} H_\alpha + ik_3 H_3 = 0, \tag{2.2}$$

$$\bar{j} = \begin{cases} 0, & \rho \text{ outside } \Omega \\ -(\epsilon - 1)(k^2 + k_3^2)\bar{H} - (\bar{\nabla} \ln\epsilon) \\ \times [(\bar{\nabla} \times \bar{H}) + ik_3 \bar{u}_3 \times \bar{H}], & \rho \text{ inside } \Omega, \end{cases} \tag{2.3}$$

where

$$\Delta_\rho = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}.$$

We shall assume that ϵ, $\partial\epsilon/\partial x_1$, and $\partial\epsilon/\partial x_2$ are finite and continuous both inside and at the boundary of the cylinder, without any discontinuity. Regarding $\partial^2\epsilon/\partial x_\alpha \partial x_\beta$, $\alpha,\beta = 1,2$, they are assumed to be finite and continuous almost everywhere inside and at the boundary of Ω, but they are allowed to have a finite number of finite discontinuities. We remark that our treatment will be valid, without modifications, for e.m. scattering by a finite number of dielectric cylinders (each of which fulfills the above conditions) whose axes be parallel to one another.

III. THE CYLINDRICAL TRANSVERSE TENSOR GREEN'S FUNCTION

We shall start by introducing the following tensor Green's function $G^{CT}(\rho) = [G^{CT}_{\alpha\beta}(\rho)]$, $\alpha,\beta = 1,2,3$ ("CT" standing for "cylindrical transverse," for reasons which will become obvious soon), in order to satisfy automatically the second Eq. (2.2), and whose interest will be shortly appreciated:

$$G^{CT}_{\substack{\alpha\beta \\ \alpha,\beta=1,2,3}}(\rho) = \frac{1}{(2\pi)^2} \int d^2\mathbf{l} \frac{\exp i\mathbf{l}\cdot\rho}{k^2 + k_3^2} \left\{ \frac{[l_\alpha l_\beta]^C - (k^2 + k_3^2)\delta_{\alpha\beta}}{l^2 - (k^2 + i\varepsilon)} - \frac{[l_\alpha l_\beta]^C}{l^2 + k_3^2} \right\}, \quad \varepsilon \to 0^+, \tag{3.1}$$

$$[l_\alpha l_\beta]^C = \begin{cases} l_\alpha l_\beta, & \alpha,\beta = 1,2, \\ l_\alpha k_3, & \alpha = 1,2, \ \beta = 3, \\ k_3 l_\beta, & \alpha = 3, \ \beta = 1,2, \\ k_3^2, & \alpha = \beta = 3 \end{cases} \tag{3.2}$$

Notice that G^{CT} is a symmetric tensor.

Two important properties of G^{CT} are:

$$(1) \sum_{\alpha=1}^{2} \frac{\partial}{\partial x_\alpha} G^{CT}_{\alpha\beta}(\rho) + ik_3 G^{CT}_{3\beta}(\rho) = 0, \quad \beta = 1,2,3, \tag{3.3}$$

$$(2) (\Delta_\rho + k^2) G^{CT}_{\alpha\beta}(\rho - \rho') = \delta^{CT}_{\alpha\beta}(\rho - \rho'), \quad \alpha,\beta = 1,2,3. \tag{3.4}$$

where we have also introduced what may well be called the cylindrical transverse Dirac's δ-function: $\delta^{CT}(\rho) = [\delta^{CT}_{\alpha\beta}(\rho)]$,

$$\delta^{CT}_{\substack{\alpha\beta \\ \alpha,\beta=1,2,3}}(\rho) = \frac{1}{(2\pi)^2} \int d^2\mathbf{l} \exp i\mathbf{l}\cdot\rho \left\{ \delta_{\alpha\beta} - \frac{[l_\alpha l_\beta]^C}{l^2 + k_3^2} \right\}. \tag{3.5}$$

The latter also fulfills

$$\sum_{\alpha=1}^{2} \frac{\partial}{\partial x_\alpha} \delta^{CT}_{\alpha\beta}(\rho) + ik_3 \delta^{CT}_{3\beta}(\rho) = 0, \quad \beta = 1,2,3.$$

Further formal properties which help to clarify the previous developments and statements are the following. Let $\bar{A}(\rho) = [A_\alpha(\rho)]$, $\alpha = 1,2,3$ be an arbitrary vector and let us introduce the new vectors:

$$\bar{A}^{CT}_G(\rho) = \int d^2\rho' G^{CT}(\rho - \rho')\bar{A}(\rho'),$$

$$\bar{A}^{CT}_\delta(\rho) = \int d^2\rho' \delta^{CT}(\rho - \rho')\bar{A}(\rho').$$

Then: (i) both $\bar{A}^{CT}_G(\rho)$ and $\bar{A}^{CT}_\delta(\rho)$ satisfy the second Eq. (2.2) and $(\Delta_\rho + k^2)\bar{A}^{CT}_G(\rho) = \bar{A}^{CT}_\delta(\rho)$, (ii) if $\bar{A}(\rho)$ fulfills the second Eq. (2.2), one has $\bar{A}^{CT}_\delta(\rho) = \bar{A}(\rho)$ and $(\Delta_\rho + k^2)\bar{A}^{CT}_G(\rho) = \bar{A}(\rho)$.

The formal proofs of the above properties proceed through elementary Fourier transformations, and, for brevity, we omit them.

Let

$$H_0^{(1)}(k|\rho|) = \frac{1}{i\pi^2} \int \frac{d^2\mathbf{l} \exp i\mathbf{l}\cdot\rho}{l^2 - (k^2 + i\varepsilon)}, \quad \varepsilon \to 0^+ \tag{3.6}$$

be Hankel's function of first kind and order zero with outgoing-wave behavior.[12] Its short distance behavior is (γ being Euler's constant):

$$H_0^{(1)}(z) \underset{z \to 0}{\to} 1 - \frac{z^2}{4} + \frac{2i}{\pi} \left\{ \left[\ln\left(\frac{z}{2}\right) + \gamma \right] \times \left(1 - \frac{z^2}{4}\right) + \frac{z^2}{4} \right\} + O(z^{4-\varepsilon}), \tag{3.7}$$

where $O(z^{4-\varepsilon})$ vanishes as $z^{4-\varepsilon}$ for any small strictly positive ε, for $z \to 0$. Upon starting from the standard large-distance behavior of $H_0^{(1)}$,[13] one proves easily that if $|\rho| \to \infty$, for fixed $\rho/|\rho|$, ρ' ($\mathbf{k}' = (k/|\rho|),\rho$):

$$H_0^{(1)}(k|\rho - \rho'|) \to [2/\pi k|\rho|]^{1/2} \cdot \exp i(k|\rho| - \pi/4) \cdot \exp(-i\mathbf{k}'\rho'). \tag{3.8}$$

Upon comparing Eqs. (3.1) and (3.6) and differentiating, one finds the following representation for G^{CT} in terms of Hankel's function $H_0^{(1)}$:

$$G^{CT}_{\substack{\alpha\beta \\ \alpha,\beta=1,2}} = \frac{i}{4(k^2 + k_3^2)} \left\{ -(k^2 + k_3^2)\delta_{\alpha\beta} H_0^{(1)}(k|\rho|) - \frac{x_\alpha}{|\rho|} \cdot \frac{x_\beta}{|\rho|} \left[k^2 \left(\frac{d^2 H_0^{(1)}(z)}{dz^2} \right)_{z=k|\rho|} + k_3^2 \left(\frac{d^2 H_0^{(1)}(z)}{dz^2} \right)_{z=+i|k_3||\rho|} \right] + \frac{x_\alpha x_\beta - |\rho|^2 \delta_{\alpha\beta}}{|\rho|^3} \left[k \left(\frac{dH_0^{(1)}(z)}{dz} \right)_{z=k|\rho|} - i|k_3| \left(\frac{dH_0^{(1)}(z)}{dz} \right)_{z=+i|k_3||\rho|} \right] \right\}, \tag{3.9}$$

$$G_{\substack{\alpha 3 \\ \alpha=1,2}}(\boldsymbol{\rho}) = \frac{k_3}{4(k^2+k_3^2)} \cdot \frac{x_\alpha}{|\boldsymbol{\rho}|} \left[k \left(\frac{dH_0^{(1)}(z)}{dz} \right)_{z=k|\boldsymbol{\rho}|} - i|k_3| \left(\frac{dH_0^{(1)}(z)}{dz} \right)_{z=+i|k_3||\boldsymbol{\rho}|} \right], \tag{3.10}$$

$$G_{33}^{\mathrm{CT}}(\boldsymbol{\rho}) = -\frac{i}{4(k^2+k_3^2)} [k^2 H_0^{(1)}(k|\boldsymbol{\rho}|) + k_3^2 H_0^{(1)}(+i|k_3||\boldsymbol{\rho}|)]. \tag{3.11}$$

Another important property is: all $G_{\alpha\beta}^{\mathrm{CT}}(\boldsymbol{\rho})$, for $\alpha,\beta = 1,2,3$, diverge only as $\ln|\boldsymbol{\rho}|$, if $|\boldsymbol{\rho}| \to 0$. This follows easily by combining Eqs. (3.9)–(3.11) with Eq. (3.7) and noticing that the $1/|\boldsymbol{\rho}|^2$ and $1/|\boldsymbol{\rho}|$ singularities cancel exactly.

Thus, the tensor Green's function defined in Eq.(3.1) and (3.2), so that it automatically fulfills the second Eq. (2.2) (say, so that it is divergenceless or "transverse"), is as singular as $H_0^{(1)}(k|\boldsymbol{\rho}|)$ as $|\boldsymbol{\rho}| \to 0$ and no more. Actually, $G^{\mathrm{CT}}(\boldsymbol{\rho})$ is the direct generalization to cylindrical geometry of a certain divergenceless Green's function, $G^{T}(\boldsymbol{\rho},x_3)$ introduced in a sketchy way by Morse and Feshbach[8] and studied in detail in previous works, and which turned out to be particularly useful in treating e.m. scattering by dielectric finite-volume defects and diffraction by periodic slabs.[14]

Specifically, one has $[\bar{l} = (l_\alpha) = (\mathbf{l}, l_3),\ \alpha = 1,2,3]$

$$G_{\alpha\beta}^{T}(\boldsymbol{\rho},x_3) = \frac{1}{(2\pi)^3(k^2+k_3^2)} \int d^3\bar{l} \left\{ \frac{l_\alpha l_\beta - (k^2+k_3^2)\delta_{\alpha\beta}}{\bar{l}^2 - (k^2+k_3^2+i\varepsilon)} - \frac{l_\alpha l_\beta}{\bar{l}^2} \right\} \exp i(\mathbf{l}\boldsymbol{\rho} + l_3 x_3), \quad \alpha,\beta = 1,2,3, \tag{3.12}$$

and the formal connection between G^{CT} and G^{T} can be easily seen to be:

$$G_{\alpha\beta}^{\mathrm{CT}}(\boldsymbol{\rho}) = \int_{-\infty}^{+\infty} dx_3 \exp(-ik_3 x_3) \cdot G_{\alpha\beta}^{T}(\boldsymbol{\rho} x_3), \quad \alpha,\beta = 1,2,3. \tag{3.13}$$

The direct generalization to cylindrical symmetry of the commonly used Green's function[15]

$$\Gamma_{\alpha\beta}(\boldsymbol{\rho},x_3) = -\left(\delta_{\alpha\beta} + \frac{1}{k^2+k_3^2} \cdot \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \right) \times \left(\frac{\exp i(k^2+k_3^2)^{1/2}(\boldsymbol{\rho}^2+x_3^2)^{1/2}}{4\pi(\boldsymbol{\rho}^2+x_3^2)^{1/2}} \right), \quad \alpha,\beta = 1,2,3, \tag{3.14}$$

namely

$$\Gamma_{\alpha\beta}^{C}(\boldsymbol{\rho}) = \int_{-\infty}^{+\infty} dx_3 \exp(-ik_3 x_3) \Gamma_{\alpha\beta}(\boldsymbol{\rho},x_3),$$

constitutes another cylindrical tensor Green's function (see also Uzunoglu and Holt[9]). However, $\Gamma_{\alpha\beta}^{C}(\boldsymbol{\rho})$ has the following disadvantages:

(i) it does not satisfy the Eq. (3.3),

(ii) it is more singular than $G_{\alpha\beta}^{\mathrm{CT}}(\boldsymbol{\rho})$ as $|\boldsymbol{\rho}| \to 0$ [$\Gamma_{\alpha\beta}^{C}(\boldsymbol{\rho})$ diverges as $1/|\boldsymbol{\rho}|^2$].

For these reasons, and except for some further remarks about $\Gamma_{\alpha\beta}^{C}$ in Sec. 4, we shall work exclusively with $G_{\alpha\beta}^{\mathrm{CT}}$ throughout this paper.

Upon combining Eqs. (3.9)–(3.11) with (3.8), one finds the asymptotic behavior of $G_{\alpha\beta}^{\mathrm{CT}}(\boldsymbol{\rho}-\boldsymbol{\rho}')$ as $|\boldsymbol{\rho}| \to \infty$ for fixed $\boldsymbol{\rho}'$ and $\mathbf{k}' = (k/|\boldsymbol{\rho}|)\boldsymbol{\rho}$

$$G_{\alpha\beta}^{\mathrm{CT}}(\boldsymbol{\rho}-\boldsymbol{\rho}') \to -\frac{i}{4}\left(\frac{2}{\pi k|\boldsymbol{\rho}|}\right)^{1/2} \cdot \exp i(k|\boldsymbol{\rho}| - \pi/4) \exp(-i\mathbf{k}'\boldsymbol{\rho}') \times \sum_{\lambda} H_0(\bar{k}',\lambda')_\alpha\, H_0(\bar{k}',\lambda')_\beta^*, \quad \alpha,\beta = 1,2,3. \tag{3.15}$$

Here, $\bar{k}' = (\mathbf{k}', k_3)$, the polarization index λ' takes only two values and the complex polarization vectors $\bar{H}_0(\bar{k}',\lambda')$ satisfy:

$$\sum_{\lambda'} H_0(\bar{k}'\lambda')_\alpha\, H_0(\bar{k}'\lambda')_\beta^* = \delta_{\alpha\beta} - \frac{k'_\alpha k'_\beta}{k^2+k_3^2}, \quad \alpha,\beta = 1,2,3,$$

$$\bar{k}' \cdot \bar{H}_0(\bar{k}'\lambda') = 0.$$

IV. SCATTERING INTEGRAL EQUATIONS

The total magnetic field $\bar{H}(\boldsymbol{\rho})$ which satisfies both Eqs. (2.2) and which, in a time-dependent (wave packet) formulation, would coincide with the incoming magnetic field at infinite distance from the dielectric cylinder in the remote past, satisfies the general scattering integral equation

$$\bar{H}(\boldsymbol{\rho}) = \bar{H}_0(\bar{k},\lambda) \exp i\mathbf{k}\boldsymbol{\rho} + \int_\Omega d^2\boldsymbol{\rho}'\, G^{\mathrm{CT}}(\boldsymbol{\rho}-\boldsymbol{\rho}')\bar{j}(\boldsymbol{\rho}'). \tag{4.1}$$

In fact, the rhs of the equation fulfills both Eqs. (2.2), by virtue of properties (3.3) and (3.4). (Recall also the formal properties studied after Eq. (3.5).] On the other hand, the fulfillment of the incoming-wave condition is warranted, according to the general prescriptions of scattering theory,[16] by the recipe used to integrate over the singularity at $\mathbf{l}^2 = k^2$ in Eq. (3.1) [namely, the instruction $\mathbf{l}^2 - k^2 \to \mathbf{l}^2 - (k^2 + i\varepsilon)$].

By letting $|\boldsymbol{\rho}| \to +\infty$ for fixed $\boldsymbol{\rho}/|\boldsymbol{\rho}|$ and recalling (3.15), Eq. (4.1) becomes

$$H_\alpha(\boldsymbol{\rho}) \to H_0(\bar{k},\lambda)_\alpha \exp i\mathbf{k}\boldsymbol{\rho} + \frac{\exp ik|\boldsymbol{\rho}|}{|\boldsymbol{\rho}|^{1/2}} \times \sum_{\lambda'} H_0(\bar{k}'\lambda')_\alpha \cdot T(\bar{k}'\lambda',\bar{k}\lambda), \quad \alpha = 1,2,3, \tag{4.2}$$

where we have introduced the scattering amplitude

$$T(\bar{k}'\lambda',\bar{k}\lambda) = -\frac{i}{4}\left(\frac{2}{\pi k}\right)^{1/2} \exp(-i\pi/4) \times \int_\Omega d^2\boldsymbol{\rho}'\, \bar{H}_0^*(\bar{k}'\lambda') \exp(-i\mathbf{k}'\boldsymbol{\rho}') \cdot \bar{j}(\boldsymbol{\rho}'). \tag{4.3}$$

The presence of derivatives of $\bar{H}(\boldsymbol{\rho})$ inside the integral in Eq. (4.1) prevents a rigorous study of the convergence of its iterations, as it stands. Fortunately, the following transformations will lead to new and mathematically more tractable integral equations: (i) by applying a standard vector identity to $-G^{\mathrm{CT}}(\boldsymbol{\rho}-\boldsymbol{\rho}')\{[\bar{\nabla}\ln\epsilon(\boldsymbol{\rho}')] \times [\nabla \times \bar{H}(\boldsymbol{\rho}')]\}$ [which

contributes to $G^{CT}\bar{j}$ in Eq. (4.1)], we replace derivatives of $\bar{H}(\rho)$ by derivatives of $G^{CT}(\rho-\rho')$ and $\bar{\nabla}\ln\epsilon(\rho')$ plus the divergence of certain products of all of them, (ii) by using Green's integral theorem, we transform the contribution from the above divergence into a line integral over the boundary of Ω, which vanishes since $\bar{\nabla}\epsilon=0$ on the latter. After some lengthy algebra, one gets the announced integral equation:

$$\bar{H}(\rho)=\bar{H}_0(\bar{k}\lambda)\exp i\mathbf{k}\rho+\int_\Omega d^2\rho'\Big[G^{CT}(\rho-\rho')L^{(0)}(\rho')+\sum_{i=1}^{2}\frac{\partial G^{CT}(\rho-\rho')}{\partial(x_i-x_i')}L^{(i)}(\rho')\Big]\bar{H}(\rho'), \tag{4.4}$$

$$\frac{\partial}{\partial x_i}G^{CT}(\rho)=\left(\frac{\partial}{\partial x_i}G^{CT}_{\alpha\beta}(\rho)\right),\quad \alpha,\beta=1,2,3, \tag{4.5}$$

$$L^{(h)}=(L^{(h)}_{\alpha\beta}),\quad \alpha,\beta=1,2,3,\quad h=0,1,2,$$

$$L^{(0)}_{\alpha\beta}=-\left[(\epsilon-1)(k^2+k_3^2)\delta_{\alpha\beta}+ik_3\frac{\partial\ln\epsilon}{\partial x'_\beta}\delta_{\alpha3}(1-\delta_{3\beta})+\sum_{\sigma,\mu,\nu=1}^{3}\varepsilon_{\alpha\sigma\mu}\varepsilon_{\nu\mu\beta}\frac{\partial}{\partial x'_\nu}\left(\frac{\partial(\ln\epsilon)}{\partial x'_\sigma}\right)\right], \tag{4.6}$$

$$L^{(1)}=\begin{pmatrix}0 & -\sum_{\sigma=1}^{3}\varepsilon_{1\sigma3}\frac{\partial\ln\epsilon}{\partial x'_\sigma} & \sum_{\sigma=1}^{3}\varepsilon_{1\sigma2}\frac{\partial\ln\epsilon}{\partial x'_\sigma}\\ 0 & -\sum_{\sigma=1}^{3}\varepsilon_{2\sigma3}\frac{\partial\ln\epsilon}{\partial x'_\sigma} & \sum_{\sigma=1}^{3}\varepsilon_{2\sigma2}\frac{\partial\ln\epsilon}{\partial x'_\sigma}\\ 0 & -\sum_{\sigma=1}^{3}\varepsilon_{3\sigma3}\frac{\partial\ln\epsilon}{\partial x'_\sigma} & \sum_{\sigma=1}^{3}\varepsilon_{3\sigma2}\frac{\partial\ln\epsilon}{\partial x'_\sigma}\end{pmatrix}, \tag{4.7}$$

$$L^{(2)}=\begin{pmatrix}\sum_{\sigma=1}^{3}\varepsilon_{1\sigma3}\frac{\partial\ln\epsilon}{\partial x'_\sigma} & 0 & -\sum_{\sigma=1}^{3}\varepsilon_{1\sigma1}\frac{\partial\ln\epsilon}{\partial x'_\sigma}\\ \sum_{\sigma=1}^{3}\varepsilon_{2\sigma3}\frac{\partial\ln\epsilon}{\partial x'_\sigma} & 0 & -\sum_{\sigma=1}^{3}\varepsilon_{2\sigma1}\frac{\partial\ln\epsilon}{\partial x'_\sigma}\\ \sum_{\sigma=1}^{3}\varepsilon_{3\sigma3}\frac{\partial\ln\epsilon}{\partial x'_\sigma} & 0 & -\sum_{\sigma=1}^{3}\varepsilon_{3\sigma1}\frac{\partial\ln\epsilon}{\partial x'_\sigma}\end{pmatrix}. \tag{4.8}$$

$\varepsilon_{\alpha\beta\gamma}$ is the totally antisymmetric tensor ($\varepsilon_{123}=+1$).

Equation (4.4) does allow one to construct a rigorous iterative solution for $\bar{H}(\rho)$. Let us introduce

$$\bar{H}^{(0)}(\rho)=\bar{H}_0(\bar{k}\lambda)\exp i\mathbf{k}\rho, \tag{4.9}$$

$$\alpha=\operatorname*{Max}_{\substack{\rho\\ \nu=1,2,3}}\int_\Omega d^2\rho'\Big(\sum_{\beta,\gamma=1}^{3}|G^{CT}_{\nu\beta}(\rho-\rho')|\cdot|L^{(0)}_{\beta\gamma}(\rho')|+\sum_{i=1}^{2}\sum_{\beta,\gamma=1}^{3}\left|\frac{\partial G^{CT}_{\nu\beta}(\rho-\rho')}{\partial(x_i-x_i')}\right|\cdot|L^{(i)}_{\beta\gamma}(\rho')|\Big). \tag{4.10}$$

Then, the following series, obtained by successive iterations of Eq. (4.4):

$$\bar{H}(\rho)=\sum_{n=0}^{+\infty}\bar{H}^{(n)}(\rho), \tag{4.11}$$

$$\bar{H}^{(n)}(\rho)=\int_\Omega d^2\rho'\Big(G^{CT}(\rho-\rho')L^{(0)}(\rho')+\sum_{i=1}^{2}\frac{\partial G^{CT}(\rho-\rho')}{\partial(x_i-x_i')}L^{(i)}(\rho')\Big)\bar{H}^{(n-1)}(\rho'),\quad n=1,2,3,\ldots, \tag{4.12}$$

converges when $\alpha<1$, since

$$\operatorname*{Max}_{\substack{\rho\\ \nu=1,2,3}}|H_\nu(\rho)|\leqslant\frac{\operatorname{Max}_{\nu=1,2,3}|H_0(\bar{k},\lambda)_\nu|}{1-\alpha}.$$

The proof proceeds by the direct majoration $\operatorname{Max}_{\substack{\rho\\ \nu=1,2,3}}|H^{(n)}_\nu(\rho)|\leqslant\alpha\operatorname{Max}_{\substack{\rho\\ \nu=1,2,3}}|H^{(n-1)}_\nu(\rho)|$ and the summation of the resulting geometric series. Notice that $G^{CT}_{\alpha\beta}(\rho)$ and $\partial G^{CT}_{\alpha\beta}(\rho)/\partial x_i$ diverge as $\ln|\rho|$ and $|\rho|^{-1}$ for $|\rho|\to0$, respectively. Since, for finite ρ, $\int_{|\rho'|<\eta}d^2\rho'\,|\ln|\rho-\rho'||<+\infty$ and $\int_{|\rho'|<\eta}d^2\rho'\,(|\rho-\rho'|)^{-1}<+\infty$ (η being a small strictly positive fixed number), one proves easily that $\alpha<+\infty$, under the assumed conditions on $\epsilon(\rho)$.

The convergence condition $\alpha<1$ holds when $\epsilon(\rho)$ is close to 1 and varies slowly for any ρ inside Ω.

Having established this convergence result, a comment on the practical usefulness of Eq. (4.1) is in order. When $\alpha<1$, the series generated by all the successive iterations of Eq. (4.1) provides another convergent representation for $\bar{H}(\rho)$, which is, moreover, simpler than the series obtained by iterating (4.4). In fact, the nth iterate of Eq. (4.1) becomes, by applying to it the same transformations which led from Eq. (4.1) to Eq. (4.4), the nth term in the series obtained by iterating (4.4). Since the latter series converges, so does that for (4.1).

Let us consider the analogue of Eq. (4.1) using Γ^C instead of G^{CT}, namely

$$\bar{H}(\rho)=\bar{H}^{(0)}(\rho)+\int_\Omega d^2\rho'\Gamma^C(\rho-\rho')\bar{j}(\rho'). \tag{4.13}$$

Then: (a) One can prove formally that the nth iterate of Eq. (4.13) coincides with $\bar{H}^{(n)}$ [Eq. (4.12)]. The proof proceeds by noticing that since $\bar{H}^{(0)}$ [Eq. (4.9)] satisfies the second Eq. (2.2), so does the first iterate of (4.13) [even if Γ^C fails to satisfy Eq. (3.3)], and generalizing inductively, for higher iterates of (4.13).

(b) Equation (4.13) can be easily shown to lead to the asymptotic behavior (4.2), with the same scattering amplitude (4.3).

In spite of the formal equivalence between (4.1) and (4.13), a convergence proof for the iterations of (4.13) similar to that for (4.11) breaks down. In fact, the analog of α [Eq. (4.10)], with G^{CT} replaced by Γ^C, diverges due to the more singular behavior of the latter as $|\rho-\rho'|\to0$ ($\operatorname{Max}_\rho\int_{|\rho'|<n}d^2\rho'\,(|\rho-\rho'|^n)^{-1}=\infty$ for $n\geqslant2$).

We shall study the convergence condition $\alpha<1$ under the following assumptions:

(a) $\epsilon(\rho)$ varies slowly throughout Ω (although it is not constant) so that the contributions from $L^{(1)}_{\alpha\beta}$ and $L^{(2)}_{\alpha\beta}$ in Eqs. (4.4) and (4.10) are negligible (although they do not vanish exactly) compared to $L^{(0)}_{\alpha\beta}$, which, in turn, reduces approximately to $-(\epsilon-1)(k^2+k_3^2)\cdot\delta_{\alpha\beta}$ (ϵ being interpreted as an average value of $\epsilon(\rho)$ in Ω). Notice that if $\epsilon(\rho)$ was a constant throughout Ω and was discontinuous at the boundary of Ω, we would violate the assumptions made in

Sec. 2, and our previous developments could be expected to lose their validity, in general.

(b) If $a(\Omega)$ is a length characterizing the size of Ω [if Ω is circular, $a(\Omega)$ is the radius], the conditions $ka(\Omega) \ll 1$ and $k_3 a(\Omega) \ll 1$ hold, that is, the wavelength is larger than $a(\Omega)$.

Then, by replacing $G^{CT}_{\alpha\beta}(\rho - \rho')$, $\alpha,\beta = 1,2,3$ by the most singular contributions for small $|\rho - \rho'|$ (which are logarithmically divergent) and performing some approximate estimates, the convergence condition becomes $\alpha < 1$, where α is the largest of the following two numbers

$$\frac{|\epsilon - 1|}{2}[[ka(\Omega)]^2 \, |\ln ka(\Omega)| + [k_3 a(\Omega)]^2 \times |\ln k_3 a(\Omega)|],$$

$$\frac{|\epsilon - 1|}{4}\{[(ka(\Omega))^2 + 2(k_3 a(\Omega))^2] \, |\ln ka(\Omega)| + (k_3 a(\Omega))^2 \, |\ln k_3 a(\Omega)|\}.$$

For other rigorous approaches to long-wavelength e.m. scattering, see Ref. 17.

V. REDUCTION TO INTEGRAL EQUATIONS WITH HILBERT–SCHMIDT KERNELS

The structure of the kernels in Eq. (4.4) and the fact that $\partial G^{CT}(\rho - \rho')/\partial(x_i - x_i')$ diverges as $1/|\rho - \rho'|$ for $|\rho - \rho'| \to 0$, prevents those kernels from being Hilbert–Schmidt. In order to derive further rigorous results, we shall transform Eq. (4.4) into a new system whose kernels are Hilbert–Schmidt.

We shall have to add the following new assumption: At each ρ in Ω there exist three 3×3 matrices $[L^{(h)}(\rho)]^{1/2}$ which are square roots of $L^{(h)}(\rho)$, respectively $\{[L^{(h)}(\rho)]^{1/2}\}^2 = L^{(h)}(\rho)$, $h = 0,1,2$. For later convenience, we shall introduce, for any ρ,ρ' inside and at the boundary of Ω:

$$\bar{Q}^{(h)}(\rho) = [L^{(h)}(\rho)]^{1/2}\bar{H}(\rho), \quad h = 0,1,2, \tag{5.1}$$

$$\bar{Q}_0^{(h)}(\rho) = [L^{(h)}(\rho)]^{1/2}\{\bar{H}_0(\bar{k}\lambda)\exp i\mathbf{k}\rho + \int_\Omega d^2\rho' \left[G^{CT}(\rho - \rho')L^{(0)}(\rho') + \sum_{i=1}^{2}\frac{\partial G^{CT}(\rho - \rho')}{\partial(x_i - x_i')}L^{(i)}(\rho')\right] \times \bar{H}_0(\bar{k},\lambda)\exp i\mathbf{k}\rho'\Big\}, \quad h = 0,1,2, \tag{5.2}$$

$$M^{(h,0)}(\rho,\rho') = [L^{(h)}(\rho)]^{1/2}\int_\Omega d^2\rho'' \left[G^{CT}(\rho - \rho'')L^{(0)}(\rho'') + \sum_{i=1}^{2}\frac{\partial G^{CT}(\rho - \rho'')}{\partial(x_i - x_i'')}L^{(i)}(\rho'')\right] \times G^{CT}(\rho'' - \rho')[L^{(0)}(\rho')]^{1/2}, \quad h = 0,1,2, \tag{5.3}$$

$$M^{(h,j)}(\rho,\rho') = [L^{(h)}(\rho)]^{1/2}\int_\Omega d^2\rho'' \left[G^{CT}(\rho - \rho'')L^{(0)}(\rho'') + \sum_{j=1}^{2}\frac{\partial G^{CT}(\rho - \rho'')}{\partial(x_j - x_j'')}L^{(j)}(\rho'')\right] \times \frac{\partial G^{CT}(\rho'' - \rho')}{\partial(x_i'' - x_i')}[L^{(i)}(\rho')]^{1/2}, \quad i = 1,2, \quad h = 0,1,2. \tag{5.4}$$

Then, iterating Eq. (4.4) once, multiplying the once iterated equation by $[L^{(h)}(\rho)]^{1/2}$ and using Eqs. (5.1–4), one gets the announced equations:

$$\underset{h=0,1,2}{\bar{Q}^{(h)}}(\rho) = \bar{Q}_0^{(h)}(\rho) + \sum_{h'=0}^{2}\int_\Omega d^2\rho' M^{(h,h')}(\rho,\rho')\cdot\bar{Q}^{(h')}(\rho'). \tag{5.5}$$

The reasons for having iterated Eq. (4.4) and having introduced Q's and M's will become clear upon showing that the kernels of (5.5) are Hilbert–Schmidt. We shall introduce L^2-norms for the set of three vectors $Q = (\bar{Q}^{(0)},\bar{Q}^{(1)},\bar{Q}^{(2)})$, and for the set of nine 3×3 matrices $M = [M^{(h,h')}]$ respectively as

$$||Q||_2 = \left[\sum_{h=0}^{2}\sum_{\alpha=1}^{3}\int_\Omega d^2\rho\,|Q_\alpha^{(h)}(\rho)|^2\right]^{1/2},$$

$$||M||_2 = \left[\sum_{h,h'=0}^{2}\sum_{\alpha,\beta=1}^{3}\int_\Omega d^2\rho\, d^2\rho' \, |M_{\alpha\beta}^{(h,h')}(\rho,\rho')|^2\right]^{1/2},$$

and so on for $||Q_0||_2$, with $Q_0 = (\bar{Q}_0^{(0)},\bar{Q}_0^{(1)},\bar{Q}_0^{(2)})$.

Then, one has the following results:

(a) The set of nine 3×3 matrix integral kernels M is Hilbert–Schmidt: $||M||_2 < +\infty$. In fact, as commented before, $\partial G^{CT}(\rho - \rho')/\partial(x_i - x_i')$ diverges as $(|\rho - \rho'|)^{-1}$ for $|\rho - \rho'| \to 0$, but the iteration involved in the transformation leading from Eq. (4.4) to Eq. (5.5) or, equivalently, the integration over ρ'' appearing in Eqs. (5.3) and (5.4) are enough to smooth out the $(|\rho - \rho'|)^{-1}$ singularity. A detailed inspection shows that the singularities in $M^{(h,h')}(\rho,\rho')$ as $|\rho - \rho'| \to 0$ are of the following types:

$$J_1 = \int_{|\rho''|<\eta} d^2\rho'' \, \frac{1}{|\rho - \rho''|}\cdot\frac{1}{|\rho'' - \rho'|},$$

$$J_2 = \int_{|\rho''|<\eta} d^2\rho'' \ln(k\,|\rho' - \rho''|)\cdot\frac{1}{|\rho'' - \rho|},$$

$$J_3 = \int_{|\rho''|<\eta} d^2\rho'' \ln(k\,|\rho - \rho''|)\cdot\ln(k\,|\rho'' - \rho'|),$$

η being a small strictly positive fixed number. By setting $\rho' = 0$, choosing ρ along the x_1 axis, introducing polar coordinates for the ρ''-integration and integrating over $|\rho''|$ first, one finds

$$J_1 = \int_0^{2\pi} d\varphi \ln\Big\{[\eta - |\rho|\cos\varphi + (\eta^2 + |\rho|^2 - 2|\rho|\eta\cos\varphi)^{1/2}]\,(|\rho|[1 - \cos\varphi])^{-1}$$

which diverges as $\ln|\rho|$ for $|\rho| \to 0$. By looking at the expressions for J_2, J_3, one realizes that they are less singular than J_1 for $|\rho - \rho'| \to 0$. On the other hand, the factors $[L^{(h)}(\rho)]^{1/2}$ (which are assumed to exist) at right and left in Eqs. (5.1)–(5.4) and the subsequent reduction of M's to ρ,ρ' varying inside Ω $\{[L^{(h)}(\rho)]^{1/2} = 0$ for ρ outside $\Omega!\}$ eliminate any possible divergence in the ρ- and ρ'-integrations as $|\rho| \to \infty$, $|\rho'| \to \infty$. All these facts imply $||M||_2 < +\infty$

The factors $[L^{(h)}(\rho)]^{1/2}$ are the natural generalization to the actual vector case of a simpler factor, which proves useful in studying the scattering integral equation for a nonrelativistic scalar quantum particle, which interacts with an external potential $V(\vec{x})$. In fact, in the latter (quantum) case, the multiplication by $V^{1/2}$ transforms the scattering integral equation into another one whose kernel is Hilbert–Schmidt.[18]

(b) One has $\|Q_0\|_2 < +\infty$. This property comes from the factor $[L^{(h)}(\rho)]^{1/2}$ in Eq. (5.2) and the fact that $\int_{|\rho'|<\eta} d^2\rho' |\ln|\rho - \rho'|| < +\infty$ and $\int_{|\rho'|<\eta} d^2\rho' (|\rho - \rho'|)^{-1} < +\infty$ for finite ρ.

(c) Assume $\|M\|_2 < 1$. Then, the series for $\bar{Q}^{(h)}(\rho)$, $h = 0,1,2$ obtained by successive iterations of Eq. (5.5) converge. In fact, by iterating Eq. (5.5), majorizing term by term in L^2-norm and summing the resulting geometric series, one gets $\|Q\|_2 \leqslant (1 - \|M\|_2)^{-1} \cdot \|Q_0\|_2$.

(d) Moreover, since $\|M\|_2 < +\infty$ and $\|Q_0\|_2 < +\infty$, the modified Fredholm theory gives the solution of Eq. (5.5) as a ratio of two series, provided that the modified Fredholm determinant does not vanish. Both series always converge for any cross section of the cylinder and any magnitude and variation of $\epsilon(\rho)$ (which be compatible with the assumptions made upon it). We shall omit the detailed expressions for the Fredholm series.[19]

We shall study briefly the convergence condition $\|M\|_2 < 1$ under the same simplifying assumptions (a) [$\epsilon(\rho) \simeq \epsilon$, $L^{(1)}$ and $L^{(2)}$ are neglected, etc.], and (b) [the wavelength is larger than a characteristic length $a(\Omega)$ of Ω] as in Sec. 4. Again, the contributions from $\partial\epsilon/\partial x_i$, $i = 1,2$ are taken as negligible. Upon performing the same approximation on $G^{CT}_{\alpha\beta}$ as there, the condition $\|M\|_2 < 1$ becomes

$$\frac{|\epsilon - 1|}{2} \{\{[ka(\Omega)]^2 \, |\ln ka(\Omega)| + [k_3 a(\Omega)]^2 \times |\ln k_3 a(\Omega)|\}^2 + \tfrac{3}{4}\{[(ka(\Omega))^2 + 2(k_3 a(\Omega))^2] \times |\ln ka(\Omega)| - (k_3 a(\Omega))^2 \, |\ln k_3 a(\Omega)|\}^2\}^{1/2} < 1.$$

This condition, as well as the one obtained at the end of Sec. 4, ensure the convergence of the iterations of Eqs. (4.1) [or (4.4)] and (5.5), respectively, for suitably small values of $ka(\Omega)$ and $k_3 a(\Omega)$. The latter may well correspond, physically, to the domain of radio-and, even, microwaves. In the optical domain, at least $ka(\Omega)$ is, frequently, much larger than 1, so that our short-distance approximation for G^{CT} fails. If $ka(\Omega) \gg 1$ and $k_3 a(\Omega) \gg 1$, and under the same conditions upon ϵ, the leading contributions to both α and $\|M\|_2$ can be estimated by using Eq. (3.15). One finds:

$$\alpha \simeq 2 \left[\frac{2\pi}{ka(\Omega)}\right]^{1/2} \cdot |\epsilon - 1| \, \{[ka(\Omega)]^2 + [k_3 a(\Omega)]^2\},$$

$$\|M\|_2 \simeq \left[\frac{\pi}{2ka(\Omega)}\right]^{1/2} \cdot |\epsilon - 1| \times \{[ka(\Omega)]^2 + [k_3 a(\Omega)]^2\}.$$

Then, the series of iterations for Eqs. (4.4) and (5.5) may still converge provided that $|\epsilon - 1|$ be sufficiently small.

ACKNOWLEDGMENTS

We are grateful to Prof. A. Duran for his interest in this line of research. One of us (R.F.A.–E.) is also indebted to Instituto de Estudios Nucleares, Junta de Energia Nuclear, Madrid, for partial financial support.

[1] M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation* (Academic, New York, 1969), Chapter 6, and references therein.
[2] J.R. Mentzner, *Scattering and Diffraction of Radio Waves* (Pergamon, Oxford, 1955).
[3] R.W.P. King and T.T. Wu, *The Scatter and Diffraction of Waves* (Harvard University, Cambridge, 1959).
[4] J.A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), Chapter 6.
[5] J.R. Wait, *Electromagnetic Radiation from Cylindrical Structures* (Pergamon, Oxford, 1959).
[6] V. Twersky, *Electromagnetic Waves*, edited by R.E. Langer (University of Wisconsin Press, Madison, Wisconsin, 1962, pp. 361–389; see also V. Twersky, "Multiple Scattering of Waves by Periodic and by Random Distributions" and references therein, in *Electromagnetic Scattering*, edited by Piergiorgio L.E. Uslenghi (Academic, New York, 1978).
[7] R.F. Harrington, *Field Computation by Moment Methods*, edited by R.F. Harrington (Cazenovia, New York 1968), Chapter 3.
[8] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. II, Chapter 13.
[9] N.K. Uzunoglu and A.R. Holt, J. Phys. **A 10**, 413 (1977); D.L. Jain and R.P. Kanwal, Ann. Phys. (N.Y.) **91**, 1 (1975).
[10] M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1965), Sec. 1.2.
[11] L. Landau and E. Lifshitz, *Electrodynamique des Milieux Continus* (Editions Mir, Moscow 1969), Chapter 10, Section 68.
[12] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chapter 5, Section 3.
[13] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (Dover, New York, 1965), Chapter 9.
[14] M.L. Calvo and A. Duran, Nuovo Cimento **B 29**, 277 (1975), and **45**, 68 (1978); R.F. Alvarez–Estrada and M.L. Calvo, J. Phys. **A 11**, 1855 (1978).
[15] V. Twersky, J. Math. Phys. **3**, 83 (1962); **19**, 215 (1978). See also, Ref. 8, as well as R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw–Hill, New York, 1965), Chapter 3.
[16] M.L. Goldberger and K.M. Watson, *Collision Theory* (Wiley, New York, 1964), Chapters 3 and 5.
[17] R.E. Kleinman, "Low Frequency Electromagnetic Scattering" in *Electromagnetic Scattering*, edited by Piergiorgio L.E. Uslenghi (Academic, New York, 1978).
[18] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms*, Princeton Series in Physics (Princeton University, Princeton, New Jersey, 1971), Chapters 1, 3, and 5, and references therein; E. Prugovecki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971), Chapter 5, Section 6.
[19] F. Smithies, *Integral Equations* (Cambridge Unviersity, London, 1965).

Reciprocal-type transformations in magnetogasdynamics

C. Rogers
University of Waterloo, Waterloo, Ontario, Canada NZL 3G1

J. G. Kingston
University of Nottingham, Nottingham, England

W. F. Shadwick
University of North Carolina, Chapel Hill, North Carolina

(Received 27 August 1979; accepted for publication 15 October 1979)

The equations of plane, steady nondissipative magnetogasdynamics are formulated in terms of an exterior differential system. Reciprocal-type invariant transformations are derived as symmetries of this system.

1. INTRODUCTION

It was Haar[1] in 1928 who first presented explicitly a class of transformations which leave invariant, up to the equation of state, the governing equations of plane potential gasdynamics. However, it may be shown that the pressure-density approximation to the adiabatic gas law introduced as early as 1904 by Chaplygin[2] in his now classical work on gas jets may be set in the context of a class of Bäcklund transformations of the hodograph equations of gasdynamics (Loewner[3]).

In 1938, Bateman[4] constructed a further class of transformations which leave invariant the gasdynamic equations. These have been termed the "reciprocal relations" and a specialization of these were used by Tsien[5] in connection with the approximation of certain subsonic adiabatic gas flows. Bateman[6] subsequently observed that both the Haar transformations and the reciprocal relations are of the Bäcklund type. Since that time reciprocal and other invariant transformations have been the subject of extensive enquiry.[7-23]

In the present paper, a new class of transformations are introduced which leave invariant the equations of two-dimensional, steady, nondissipative magnetogasdynamics. The basic equations may be written in the form of four conservation laws[22]

$$\partial_1 A^i_{\ 2} - \partial_2 A^i_{\ 1} = 0, \quad i = 1,...,4, \tag{1.1}$$

together with the relation

$$\mu(uH_2 - vH_1) = \alpha, \tag{1.2}$$

where $[A^i_{\ j}]$ is the matrix

$$[A^i_{\ j}] = \begin{bmatrix} \rho uv - \mu^2 H_1 H_2 & -\{\mathscr{P} + \rho u^2 - \mu^2 H_1^2\} \\ -\{\mathscr{P} + \rho v^2 - \mu^2 H_2^2\} & \rho uv - \mu^2 H_1 H_2 \\ \rho v & -\rho u \\ \mu H_2 & -\mu H_1 \end{bmatrix}, \tag{1.3}$$

while $\mathscr{P}$ is the total magnetic pressure, ρ is the gas density, u and v are the velocity components, H_i $(i = 1,2)$ are the components of the magnetic field, μ is the magnetic permeability (assumed constant), and α is a constant.

Equations in the form of conservation laws in two independent variables may be conveniently expressed in terms of differential 1-forms. Thus, if we coordinatize $\mathbb{R}^8$ by $x^1, x^2, u, v, \mathscr{P}, \rho, H_1$, and H_2 we may introduce coordinates w^i $(i = 1,...,4)$ on $\mathbb{R}^4$ and an exterior differential system Σ on $\mathbb{R}^8 \times \mathbb{R}^4$ whose solutions are the solutions of the system (1.1)–(1.3), namely, the exterior ideal generated by the 0-form

$$h := \mu(uH_2 - vH_1) - \alpha, \tag{1.4}$$

and the 1-forms

$$\beta^i := dw^i - A^i_{\ a}\, dx^a. \tag{1.5}$$

A *solution* of Σ will be defined as a two-dimensional submanifold (ϕ, N), $\phi : N \to \mathbb{R}^8 \times \mathbb{R}^4$ which is transversal to the coordinates x^a, and satisfies

$$\phi^* \Sigma = 0.$$

A solution of Σ is therefore given locally by a map of the form

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \to \begin{pmatrix} x^1 \\ x^2 \\ u(x^1, x^2) \\ v(x^1, x^2) \\ \mathscr{P}(x^1, x^2) \\ \rho(x^1, x^2) \\ H_1(x^1, x^2) \\ H_2(x^1, x^2) \\ w^1(x^1, x^2) \\ w^2(x^1, x^2) \\ w^3(x^1, x^2) \\ w^4(x^1, x^2) \end{pmatrix},$$

such that $u, v, \mathscr{P}, \rho, H_1$, and H_2 satisfy the system (1.1)–(1.3).

A diffeomorphism ψ of $\mathbb{R}^8 \times \mathbb{R}^4$ which satisfies $\psi^* \Sigma \subset \Sigma$ will be called a *symmetry* of Σ. Thus, if ψ is a symmetry of Σ, and (ϕ, N) is a solution of Σ, then so is $(\psi \circ \phi, N)$ [provided $\psi \circ \phi$ is transversal to the coordinates (x^a)]. We shall demonstrate in Sec. 2 that reciprocal type symmetries of Σ exist.

 0022-2488/80/020395-03$1.00

2. RECIPROCAL TYPE SYMMETRIES OF Σ

Consider a diffeomorphism ψ of $\mathbb{R}^8 \times \mathbb{R}^4$ which is linear in the coordinates x^a, ω^i, namely

$$x'^a := \psi^* x^a = B^a_j \omega^j + C^a_k x^k \tag{2.1}$$

and

$$\omega'^i := \psi^* \omega^i = b^i_j \omega^j + c^i_k x^k, \tag{2.2}$$

where B^a_j, C^a_k, b^i_j, and c^i_k are constants. The condition

$$\psi^* \Sigma \subset \Sigma$$

may now be used to determine the remaining quantities

$$u' := \psi^* u, \quad v' := \psi^* v, \quad \mathscr{P}' := \psi^* \mathscr{P}, \quad \rho' := \psi^* \rho$$
$$H'_i := \psi^* H_1, \quad i = 1,2,$$

as follows.

From Eqs. (2.1) and (2.2)

$$\psi^* \beta^i = \psi^* dw^i - \psi^* A^i_a \cdot \psi^* dx^a$$
$$= b^i_j dw^j - c^i_a dx^a - \psi^* A^i_k (B^k_j dw^j + C^k_a dx^a),$$

and from (1.4) and (1.5)

$$dw^j = A^j_a dx^a, \quad \text{mod } \Sigma,$$

so that

$$\psi^* \beta^i = [b^i_j A^j_a + c^i_a - A'^i_k (B^k_j A^j_a + C^k_a)] dx^a, \quad \text{mod } \Sigma, \tag{2.3}$$

where

$$A'^i_k := \psi^* A^i_k = A^i_k \circ \psi.$$

Thus, $\psi^* \beta^i \in \Sigma$ iff

$$b^i_j A^j_a + c^i_a - A'^i_k (B^k_j A^j_a + C^k_a) = 0. \tag{2.4}$$

The condition (2.4) provides eight equations for the six unknowns in terms of u, v, $\mathscr{P}$, ρ, and the H_i ($i = 1,2$), together with the constants b^i_j, c^i_a, B^k_j, and C^k_a. The requirement

$$\psi^* h = lh, \tag{2.5}$$

where l is a constant, provides a ninth equation. Thus, in general, it is to be expected that it will be necessary to impose three additional constraints in order to obtain symmetries of the present type. Such overconstrained systems are characteristic of the algebra associated with Bäcklund transformations. It is possible however to reduce the number of constraints to two as is evidenced in the following illustration of the method.

3. AN INVARIANT TRANSFORMATION

The matrices $[b^i_j]$, $[c^i_a]$, $[B^k_j]$, and $[C^k_a]$ are chosen to be of the forms

$$[b^i_j] = \begin{bmatrix} 0_2^2 & 0_2^2 \\ 0_2^2 & \Lambda(a_1, a_2) \end{bmatrix}, \quad [c^i_a] = \begin{bmatrix} c & c \\ -c & c \\ & 0_2^2 \end{bmatrix},$$

$$[B^k_j] = \begin{bmatrix} b & -b \\ & 0_2^2 \\ b & b \end{bmatrix}, \quad [C^k_a] = [0_2^2], \tag{3.1}$$

and $[A^i_j]$ is written as

$$[A^i_j] = \begin{bmatrix} S & -U \\ -V & S \\ \rho v & -\rho u \\ \mu H_2 & -\mu H_1 \end{bmatrix}, \tag{3.2}$$

where

$$S := \rho u v - \mu^2 H_1 H_2, \quad U := \mathscr{P} + \rho u^2 - \mu^2 H_1^2,$$
$$V := \mathscr{P} + \rho v^2 - \mu^2 H_2^2.$$

In the above, a_i $i = 1, 2$ and b, c are real constants. It follows that

$$B^k_j A^j_a + C^k_a = b \begin{bmatrix} S+V & -(S+U) \\ S-V & S-U \end{bmatrix}, \tag{3.3}$$

so that

$$\det(B^k_j A^j_k + C^k_a) = 2b^2(S^2 - UV) = -2b^2 J, \tag{3.4}$$

where

$$J = \mathscr{P}^2 + \mathscr{P}(\rho q^2 - \mu^2 H^2) - \rho \alpha^2. \tag{3.5}$$

Thus,

$$[B^k_j A^j_a + C^k_a]^{-1} = -\frac{1}{2bJ} \begin{bmatrix} S-U & S+U \\ -(S-V) & S+V \end{bmatrix}, \tag{3.6}$$

while

$$[b^i_j A^j_a + c^i_a] = \begin{bmatrix} c & c \\ -c & c \\ a_1 \rho v & -a_1 \rho u \\ a_2 \mu H_2 & -a_2 \mu H_1 \end{bmatrix}, \tag{3.7}$$

so that

$$[A'^i_j] = -\frac{1}{2bJ} \begin{bmatrix} c(V-U) & c(2S+U+V) \\ -c(2S-U-V) & c(V-U) \\ -a_1 \rho[\mathscr{P}(u+v) + \mu\alpha(H_1 - H_2)] & -a_1 \rho[\mathscr{P}(u-v) - \mu\alpha(H_1 + H_2)] \\ -a_2[\mu\mathscr{P}(H_1 + H_2) + \rho\alpha(u-v)] & -a_2[\mu\mathscr{P}(H_1 - H_2) - \rho\alpha(u+v)] \end{bmatrix}. \tag{3.8}$$

In this case, the nature of the matrices B^k_j and c^i_a has guaranteed that $A'^1_1 = A'^2_2$ so that the number of equations for $\mathscr{P}'$, ρ', u', v', and $\mu H'_i$ ($i = 1,2,$) has been reduced to seven, obtained by comparison of (3.8) and the primed counterpart of (1.3), augmented by the constraint (2.5). Thus, in general, $\mathscr{P}'$, ρ', u', v', and $\mu H'_i$ may be determined in terms of $\mathscr{P}$, ρ, u, v, and μH_i subject to two constraints on the latter six quantities.

Now, comparison of the (4,1) and (4,2) terms of (3.8) and

$$[A'^i_j] = \begin{bmatrix} \rho' u' v' - \mu^2 H'_1 H'_2 & -\{\mathscr{P}' + \rho' u'^2 - \mu^2 H'^2_1\} \\ -(\mathscr{P}' + \rho' v'^2 - \mu^2 H'^2_2) & \rho' u' v' - \mu^2 H'_1 H'_2 \\ \rho' v' & -\rho' u' \\ \mu H'_2 & -\mu H'_1 \end{bmatrix}, \tag{3.9}$$

produces the required expressions for the new components H_i' of the magnetic field in terms of the original magnetogasdynamic variables, namely

$$\mu H_1' = -a_2[\mathscr{P}\mu(H_1 - H_2) - \rho\alpha(u+v)]/2Jb$$

and (3.10)

$$\mu H_2' = a_2[\mathscr{P}\mu(H_1 + H_2) + \rho\alpha(u-v)]/2Jb$$

or,

$$\mu H' = -\frac{a_2}{2Jb}[\mathscr{P}\mu(1-i)\bar{H} - \rho\alpha(1+i)\bar{q}], \quad (3.11)$$

where $H := H_1 + iH_2$, $q := u + iv$, and similarly for the primed quantities. On the other hand, comparison of the (3.1) and (3.2) entries of (3.8) and (3.9), yields

$$\rho' u' = -a_1\rho[\mathscr{P}(u-v) - \mu\alpha(H_1 + H_2)]/2Jb$$

and (3.12)

$$\rho' v' = a_1\rho[\mathscr{P}(u+v) + \mu\alpha(H_1 - H_2)]/2Jb,$$

or,

$$\rho' q' = -\frac{a_1\rho}{2Jb}[\mathscr{P}(1-i)\bar{q} - \mu\alpha(1+i)\bar{H}]. \quad (3.13)$$

Now, (2.5) implies that

$$2i\rho'\alpha' = -\mu\{\rho' q'\bar{H}' - \rho'\bar{q}'H\}, \quad (3.14)$$

where $\alpha' = l\alpha$, so that, from (3.11) and (3.13) we obtain the expressions for the new gas density and the new gas velocity components in the forms

$$\rho' = -a_1 a_2 \alpha\rho/2b^2\alpha' J \quad (3.15)$$

and

$$q' = (b\alpha'/a_2\alpha)[\mathscr{P}(1-i)\bar{q} - \mu\alpha(1+i)\bar{H}], \quad (3.16)$$

respectively.

Further, comparison of the (1,2) or (2,1) entries of (3.8) and (3.9) gives the new total magnetic pressure, namely

$$\mathscr{P}' = \frac{a_1\alpha'}{2a_2\alpha J}[\mathscr{P}(u+v) + \mu\alpha(H_1 - H_2)]^2 + \frac{c}{2bJ}[2\mathscr{P} + \rho(u-v)^2 - \mu^2(H_1 - H_2)^2] - \frac{a_2^2}{4J^2b^2}[\mathscr{P}\mu(H_1 + H_2) + \rho\alpha(u-v)]^2. \quad (3.17)$$

Finally, the (1,1) terms of (3.8) and (3.9) yield

$$\rho' u' v' - \mu^2 H_1' H_2' = (c/2bJ)[\rho(u^2 - v^2) - \mu^2(H_1^2 - H_2^2)], \quad (3.18)$$

while the (1,2) and (2,1) terms give, for $\mathscr{P}'$ to be defined consistently,

$$\rho'(u'^2 - v'^2) - \mu^2(H_1'^2 - H_2'^2) = (2c/bJ)(\rho uv - \mu^2 H_1 H_2). \quad (3.19)$$

The conditions (3.18) and (3.19) may be conveniently combined to give the single relation

$$\rho' q'^2 - \mu^2 H'^2 = (c/bJ)(\rho q^2 - \mu^2 H^2). \quad (3.20)$$

Thus, to summarize, it has been shown that the magnetogasdynamic Eqs. (1.1)–(1.3) are invariant under the transformations defined by (2.1)–(2.2) subject to the requirement (3.20). Moreover, the new magnetogasdynamic variables are given explicitly by (3.11) and (3.15)–(3.17).

ACKNOWLEDGMENTS

One of the authors (C.R) would like to express his gratitude to the Department of Applied Mathematics and Theoretical Physics, University of Cambridge where part of this work was undertaken. Support under National Research Council of Canada Grant No. A0879 is also acknowledged.

[1] A. Haar, Math. Ann. **100**, 481 (1928).
[2] S.A. Chaplygin, Sci. Mem. Moscow Univ. Math. Phys. Sec. **21**, 1 (1904).
[3] C. Loewner, NACA Technical Note 2065 (1950).
[4] H. Bateman, Proc. Nat. Acad. Sci. USA **24**, 246 (1938).
[5] H.S. Tsien, J. Aero. Sci. **6**, 399 (1939).
[6] H. Bateman, Quart. Appl. Math. **1**, 281 (1944).
[7] A.A. Nikol'skii, PMM **27**, 740 (1963).
[8] E.D. Tomilov, PMM **29**, 959 (1965).
[9] V.A. Rykov, PMM **29**, 178 (1965).
[10] L.A. Movsesian, PMM **31**, 137 (1967).
[11] M.D. Ustinov, Izv. AN. SSSR Mekh. Zhid. I. Gaza **3**, 68 (1966).
[12] M.D. Ustinov, Izv. AN. SSSR Mehk. Zhid. I. Gaza **4**, 88 (1967).
[13] R.C. Prim, J. Appl. Phys. **20**, 448 (1949).
[4] G. Power and P. Smith, Appl. Sci. Res. A **8**, 386 (1959).
[5] G. Power and P. Smith, J. Math. Mech. **10**, 349 (1961).
[6] P. Smith. Arch. Rat. Mech. Anal. **15**, 147 (1964).
[7] C. Rogers, Zeit. Angew. Math. Phys. **19**, 58 (1968).
[8] C. Rogers, Zeit. Angew. Math. Phys. **20**, 370 (1969).
[9] G. Power and C. Rogers, Appl. Sci. Res. **21**, 176 (1969).
[10] C. Rogers, Arch. Rat. Mech. Anal. **47**, 36 (1972).
[11] S.P. Castell and C. Rogers, Quart. Appl. Math. **32**, 241 (1974).
[12] C. Rogers, S.P. Castell, and J.G. Kingston, J. de Mécanique **13**, 343 (1974).
[13] C. Rogers and J.G. Kinston, J. de Mécanique **15**, 185 (1976).

Nonperiodic oscillations of Langmuir waves

P. K. C. Wang
School of Engineering and Applied Science, University of California, Los Angeles, California 90024

(Received 21 December 1978; accepted for publication 15 October 1979)

It is shown that the single-mode equations derived from Zakharov's model for Langmuir turbulence in a plasma in the presence of an external spatially homogeneous electric field oscillating at the electron plasma frequency has nonperiodic chaotic solutions whose power spectra have turbulence-like features. Bounds for these chaotic solutions are derived. Typical numerical results are presented for the one-dimensional case.

1. INTRODUCTION

It has been observed that certain nonlinear ordinary differential equations have chaotic or turbulence-like solutions.[1-4] A simple example is the Lorenz model for thermal convection in a fluid layer.[1] Recently, Ruelle and Takens proposed that fluid turbulence can be mathematically characterized by this class of solutions whose trajectories in the state space are attracted to a nonempty set ("strange attractor") which is neither an equilibrium set nor a periodic orbit.[5] On this set, the trajectories exhibit chaotic oscillations. Moreover, they are sensitive to variations in the initial conditions. Here, we shall demonstrate that the single-mode equations derived from the Zakharov's model for Langmuir turbulence in a plasma have nonperiodic chaotic solutions.

We begin with the following dimensionless form of Zakharov's equations describing the nonlinear interaction of high-frequency electron oscillations with an ion fluid in the presence of an external spatially homogeneous electric field oscillating at the electron plasma frequency ω_p:[6-8]

$$\nabla\cdot\left[i\frac{\partial \mathbf{E}}{\partial t}+\nabla^2\mathbf{E}-n(\mathbf{E}+\mathbf{E}_0)\right]=0\,, \tag{1}$$

$$\frac{\partial^2 n}{\partial t^2}-\nabla^2 n=\nabla^2\left[|\mathbf{E}|^2+\mathbf{E}_0\cdot(\mathbf{E}+\mathbf{E}^*)\right], \tag{2}$$

where $i=\sqrt{-1}$; $(\cdot)^*$ denotes complex conjugation, and $\mathbf{a}\cdot\mathbf{b}$ the usual scalar product of two real or complex vectors $\mathbf{a}$ and $\mathbf{b}$. $\mathbf{E}=(E_1,\ldots,E_N)$ is the complex amplitude of the high-frequency electric field $\mathscr{E}$ given by

$$\mathscr{E}(t,\mathbf{x})=\mathrm{Re}[\mathbf{E}(t,\mathbf{x})\exp(-i\omega_p t)]; \tag{3}$$

and n is a real quantity corresponding to the low-frequency perturbation in the ion density from its constant equilibrium value n_0. The units of time t, spatial coordinates $\mathbf{x}=(x_1,\ldots,x_N)$, electric fields $\mathbf{E}_0$ and $\mathbf{E}$, and ion density perturbation are respectively, $3/(2\alpha\omega_p)$, $(3/2)\alpha^{-1/2}\lambda_D$, $[(64/3)\pi n_0 m_e c_s^2]^{1/2}$, and $(4/3)\alpha n_0$, where α is the electron-ion mass ratio m_e/m_i, λ_D the Debye length, and c_s is the ion acoustic speed. Here, $\mathbf{E}_0=(E_{01},\ldots,E_{0N})$ is a real constant vector corresponding to the normalized amplitude of the external electric field. It is of interest to determine the behavior of the solutions of (1) and (2) (with appropriate damping terms) as a function of the parameter $\mathbf{E}_0$, in particular, the existence of turbulence-like solutions for some $\mathbf{E}_0$, and the onset of such solutions as $\mathbf{E}_0$ tends to some threshold values.

2. SIMPLIFIED MODEL

Let the spatial domain Ω be a bounded open subset of the N-dimensional Euclidean space $\mathbb{R}^N$, and $L^2(\Omega)$ denote the Hilbert space of real square-integrable functions defined on Ω with inner product $\langle u,v\rangle=\int_\Omega u(\mathbf{x})v(\mathbf{x})\,d\Omega$. Let $\{\phi_k\}$ be a countable orthonormal basis for $L^2(\Omega)$. We seek solutions to (1) and (2) in the form:

$$\mathbf{E}(t,\mathbf{x})=\sum_k \mathbf{E}_k(t)\phi_k(\mathbf{x}),\quad n(t,\mathbf{x})=\sum_k n_k(t)\phi_k(\mathbf{x})\,. \tag{4}$$

If the boundary of Ω is sufficiently smooth, then the Laplacian with suitable homogeneous boundary conditions is a negative operator with a countable point spectrum. We may take ϕ_k to be the orthonormalized eigenfunction of ∇^2 corresponding to the eigenvalue $\lambda_k=-\mu_k^2$. In this case, we may substitute (4) into (1) and (2), multiply both sides of the equations by $\phi_m(\mathbf{x})$, and integrate over Ω to give a countably infinite system of ordinary differential equations for $\mathbf{E}_m$ and n_m:

$$i\frac{d\mathbf{E}_m}{dt}-\mu_m^2\mathbf{E}_m=n_m\mathbf{E}_0+\sum_k\sum_{k'}\alpha_{mkk'}n_k\mathbf{E}_{k'}, \tag{5}$$

$$\frac{d^2 n_m}{dt^2}+\mu_m^2 n_m=-\mu_m^2\mathbf{E}_0\cdot(\mathbf{E}_m+\mathbf{E}_m^*)+\sum_k\sum_{k'}\beta_{mkk'}\mathbf{E}_k\cdot\mathbf{E}_{k'}\,, \tag{6}$$

where $\mathbf{E}_k\cdot\mathbf{E}_{k'}=\Sigma_j E_{kj}E^*_{k'j}$ and

$$\alpha_{mkk'}=\int_\Omega \phi_k(\mathbf{x})\phi_{k'}(\mathbf{x})\phi_m(\mathbf{x})\,d\Omega\,. \tag{7}$$

$$\beta_{mkk'}=\int_\Omega \nabla^2[\phi_k(\mathbf{x})\phi_{k'}(\mathbf{x})]\phi_m(\mathbf{x})\,d\Omega\,. \tag{8}$$

By retaining only the terms involving $\mathbf{E}_m$ and n_m in (5) and (6), we obtain the following simplified equations for a single mode m:

$$i\frac{d\mathbf{E}_m}{dt}-(\mu_m^2-i\gamma_m)\mathbf{E}_m=n_m(\mathbf{E}_0+\alpha_m\mathbf{E}_m)\,, \tag{9}$$

$$\frac{d^2 n_m}{dt^2}+2\Gamma_m\frac{dn_m}{dt}+\mu_m^2 n_m=-\mu_m^2\mathbf{E}_0\cdot(\mathbf{E}_m+\mathbf{E}_m^*)+\beta_m|\mathbf{E}_m|^2\,, \tag{10}$$

where

 0022-2488/80/020398-10$1.00

$$\alpha_m = \int_\Omega \phi^3_m(\mathbf{x})\,d\Omega, \quad \beta_m = \int_\Omega \nabla^2[\phi^2_m(\mathbf{x})]\phi_m(\mathbf{x})\,d\Omega\,. \tag{11}$$

Also, we have added the phenomenological damping coefficients γ_m and Γ_m. They may represent Landau damping of the high- and low-frequency waves. In what follows, we shall analyze the behavior of the solutions of (9) and (10) as $\mathbf{E}_0$ varies. For brevity, the subscript m in (9) and (10) will be omitted in places where ambiguity does not arise.

A starting point for searching the strange attractor or chaotic solutions is to study the nature of the equilibrium points as $\mathbf{E}_0$ varies. It is known that chaotic solutions could arise after finite number of Hopf bifurcations,[5] therefore we shall establish the existence of Hopf bifurcation points.

3. EQUILIBRIUM POINTS

Consider the following equations for determining the equilibrium points of (9) and (10) for any given $\mathbf{E}_0$:

$$(i\gamma - \mu^2)\mathbf{E} = n(\mathbf{E}_0 + \alpha\mathbf{E})\,, \tag{12}$$

$$\mu^2 n = -\mu^2\mathbf{E}_0\cdot(\mathbf{E}+\mathbf{E}^*) + \beta\,|\mathbf{E}|^2\,. \tag{13}$$

Using (13) to eliminate n in (12), we obtain an equation for $\mathbf{E}$:

$$(i\gamma - \mu^2)\mathbf{E} = [-\mathbf{E}_0\cdot(\mathbf{E}+\mathbf{E}^*) + \beta\mu^{-2}\,|\mathbf{E}|^2](\mathbf{E}_0 + \alpha\mathbf{E})\,. \tag{14}$$

Evidently, since $\mathbf{E}_0$ is a real N-dimensional vector and n is real, a solution of (14) must be a complex scalar multiple of $\mathbf{E}_0$ (i.e., $\mathbf{E} = \xi\,\mathbf{E}_0$ for some $\xi = \xi_R + i\xi_I$). Thus, the solution of (14) reduces to finding ξ. Note that $\mathbf{E} = \mathbf{0}$ is a solution of (14) for any $\mathbf{E}_0$. Substituting $\mathbf{E} = \xi\,\mathbf{E}_0$ into (14) leads to the following equations for ξ_R and ξ_I:

$$-(\mu^2\xi_R + \gamma\xi_I) = \{\beta\mu^{-2}(\xi_R^2 + \xi_I^2) - 2\xi_R\}(1+\alpha\xi_R)\|\mathbf{E}_0\|^2, \tag{15}$$

$$\gamma\xi_R - \mu^2\xi_I = \{\beta\mu^{-2}(\xi_R^2 + \xi_I^2) - 2\xi_R\}\alpha\xi_I\|\mathbf{E}_0\|^2, \tag{16}$$

and

$$n = \{\beta\mu^{-2}(\xi_R^2 + \xi_I^2) - 2\xi_R\}\|\mathbf{E}_0\|^2, \tag{17}$$

where $\|\mathbf{E}_0\|^2 = \mathbf{E}_0\cdot\mathbf{E}_0$.

Dividing (15) by (16) gives

$$\xi_R^2 + \xi_I^2 = (\alpha\gamma)^{-1}(\mu^2\xi_I - \gamma\xi_R)\,, \tag{18}$$

which implies that a solution must lie on the circle:

$$\left(\xi_R + \frac{1}{2\alpha}\right)^2 + \left(\xi_I - \frac{\mu^2}{2\alpha\gamma}\right)^2 = \frac{(\mu^4+\gamma^2)}{(2\alpha\gamma)^2}\,. \tag{19}$$

Now, we substitute (18) into (16) and solve for ξ_R in terms of ξ_I:

$$\xi_R = \xi_I(\mu^2 + \beta\gamma^{-1}\|\mathbf{E}_0\|^2\xi_I) \times [\gamma + (2\alpha + \beta\mu^{-2})\|\mathbf{E}_0\|^2\xi_I]^{-1}\,. \tag{20}$$

Finally, using (20) to eliminate ξ_R in (18) leads to the following quadratic equation for ξ_I:

$$A\xi_I^2 + B\xi_I + C = 0\,, \tag{21}$$

where

$$A = [(\beta/\gamma)^2 + \delta^2]\|\mathbf{E}_0\|^4\,, \tag{22}$$

$$B = 2(\mu^2\beta\gamma^{-1} + \gamma\delta)\|\mathbf{E}_0\|^2 - 2\mu^2\delta\gamma^{-1}\|\mathbf{E}_0\|^4\,, \tag{23}$$

$$C = \mu^4 + \gamma^2 - 2\mu^2\|\mathbf{E}_0\|^2 \tag{24}$$

$$\delta = (2\alpha + \beta\mu^{-2})\,. \tag{25}$$

If $B^2 - AC \geqslant 0$ or

$$\|\mathbf{E}_0\|^4 - 4\alpha\beta\delta^{-2}\|\mathbf{E}_0\|^2 - (2\alpha\gamma/\delta)^2 \geqslant 0\,, \tag{26}$$

then (21) has real roots given explicitly by

$$\xi_I^{\pm} = \frac{\delta\gamma\mu^2}{(\beta^2+\delta^2\gamma^2)\|\mathbf{E}_0\|^2}\left\{\|\mathbf{E}_0\|^2 - \frac{(\mu^2\beta+\gamma^2\delta)}{\delta\mu^2} \pm \left[\|\mathbf{E}_0\|^4 - 4\alpha\beta\delta^{-2}\|\mathbf{E}_0\|^2 - \left(\frac{2\alpha\gamma}{\delta}\right)^2\right]^{1/2}\right\} \tag{27}$$

For $\|\mathbf{E}_0\|^2 \geqslant 0$, condition (26) is satisfied if and only if

$$\|\mathbf{E}_0\|^2 \geqslant E_{oc}^2 \triangleq 2\alpha\delta^{-2}[\beta + (\beta^2+\delta^2\gamma^2)^{1/2}]\,. \tag{28}$$

Thus, we conclude that for $0 \leqslant \|\mathbf{E}_0\| < E_{oc}$, the origin $(\mathbf{E},n,\dot{n}) = (\mathbf{0}+j\mathbf{0},0,0)$ is the only equilibrium state of system (9) and (10), where $\dot{n}$ denotes dn/dt. When $\|\mathbf{E}_0\| = E_{oc}$, a new equilibrium state $(\mathbf{E},n,\dot{n}) = (\mathrm{Re}(\mathbf{E}) + i\mathrm{Im}(\mathbf{E}),n,0)$ emerges, where $\mathrm{Re}(\mathbf{E}) = \xi_R\,\mathbf{E}_0$, $\mathrm{Im}(\mathbf{E}) = \xi_R\,\mathbf{E}_0$ with ξ_I given by

$$\xi_I = \frac{\delta\gamma\mu^2}{(\beta^2+\delta^2\gamma^2)E_{oc}^2}\left(E_{oc}^2 - \frac{(\mu^2\beta+\gamma^2\delta)}{\delta\mu^2}\right), \tag{29}$$

and ξ_R, n are given by (20) and (17) respectively. As $\|\mathbf{E}_0\|$ increases from E_{oc}, the foregoing nonzero equilibrium state bifurcates into two distinct equilibrium states $(\xi_R^+\mathbf{E}_0 + i\xi_I^+\mathbf{E}_0, n^+, 0)$ and $(\xi_R^-\mathbf{E}_0 + i\xi_I^-\mathbf{E}_0, n^-, 0)$ where $\xi_I^{\pm}$ are given by (27), whose corresponding $\xi_R^{\pm}$ and $n^{\pm}$ are determined respectively by (20) and (17). We note that ξ_R and ξ_I depend on $\|\mathbf{E}_0\|^2$. Also, the coefficient C defined by (24) vanishes when

$$\|\mathbf{E}_0\|^2 = \tilde{E}_{oc}^2 \triangleq (\mu^4+\gamma^2)/(2\mu^2)\,. \tag{30}$$

Consequently, ξ_I^- also vanishes. Thus, in this case, the equilibrium set consists of the origin and the point $[(\xi_R^+ + i\xi_I^+)\mathbf{E}_0, n^+, 0]$. When $\|\mathbf{E}_0\|$ increases from $\tilde{E}_{oc}$, we have again three distinct equilibrium points.

4. STABILITY OF EQUILIBRIUM

Let $\mathbf{E}_R = \mathrm{Re}(\mathbf{E})$, $\mathbf{E}_I = \mathrm{Im}(\mathbf{E})$, and $\mathbf{z}$ denote the $2(N+1)$-dimensional real vector $(n,\dot{n},\mathbf{E}_R,\mathbf{E}_I)^T$, where $(\cdot)^T$ denotes transposition. We rewrite (9) and (10) in the form:

$$\frac{d\mathbf{z}}{dt} = \mathbf{f}(\mathbf{z};\mathbf{E}_0) \triangleq \begin{bmatrix} \dot{n} \\ -\mu^2 n - 2\Gamma\dot{n} - 2\mu^2\,\mathbf{E}_0\cdot\mathbf{E}_R + \beta\,(\|\mathbf{E}_R\|^2 + \|\mathbf{E}_I\|^2) \\ -\gamma\mathbf{E}_R + (\mu^2 + \alpha n)\mathbf{E}_I \\ -\gamma\mathbf{E}_I - (\mu^2 + \alpha n)\mathbf{E}_R - n\mathbf{E}_0 \end{bmatrix}. \tag{31}$$

Let $\mathbf{z}_e = (n^e, 0, \mathbf{E}_R^e, \mathbf{E}_I^e)$ be an equilibrium point of (31) as given in Sec. 3, and $\delta\mathbf{z}(t) \triangleq \mathbf{z}(t) - \mathbf{z}_e$. We consider the following linearized system of (31) about $\mathbf{z}_e$:

$$\frac{d\delta\mathbf{z}}{dt} = \mathbf{J}_f(z_e;\mathbf{E}_0)\delta\mathbf{z}, \tag{32}$$

where $\mathbf{J}_f(\mathbf{z}_e;\mathbf{E}_0)$ is the Jacobian matrix of $\mathbf{f}$ at $\mathbf{z}_e$ given by

$$\mathbf{J}_f(\mathbf{z}_e;\mathbf{E}_0) = \begin{bmatrix} 0 & 1 & \mathbf{O}_N^T & \mathbf{O}_N^T \\ -\mu^2 & -2\Gamma & 2(\beta\mathbf{E}_R^e - \mu^2\mathbf{E}_0)^T & 2\beta(\mathbf{E}_I^e)^T \\ \alpha\mathbf{E}_I^e & 0 & -\gamma\mathbf{I}_N & (\mu^2 + \alpha n^e)\mathbf{I}_N \\ -(\mathbf{E}_0 + \alpha\mathbf{E}_R^e) & 0 & -(\mu^2 + \alpha n^e)\mathbf{I}_N & -\gamma\mathbf{I}_N \end{bmatrix}, \tag{33}$$

where $\mathbf{O}_N$ and $\mathbf{I}_N$ are the N-dimensional zero vector and $N \times N$ identity matrix respectively. It can be shown (see Appendix) that the characteristic polynomial of $\mathbf{J}_f(\mathbf{z}_e;\mathbf{E}_0)$ is given by

$$\begin{aligned} &\det[\mathbf{J}_f(\mathbf{z}_e;\mathbf{E}_0) - \lambda\,\mathbf{I}_{2(N+1)}] \\ &\quad = [(\gamma+\lambda)^2 + (\mu^2 + \alpha n^e)^2]^N(\lambda^2 + 2\Gamma\lambda + \mu^2) + 2[(\gamma+\lambda)^2 + (\mu^2 + \alpha n^e)^2]^{N-1} \\ &\qquad \times [(\mu^2 + \alpha n^e)\{(\beta\mathbf{E}_R^e - \mu^2\mathbf{E}_0)\cdot(\mathbf{E}_0 + \mathbf{E}_R^e) + \alpha\beta\|\mathbf{E}_I^e\|^2\} + (\gamma+\lambda)(\alpha\mu^2 + \beta)\mathbf{E}_0\cdot\mathbf{E}_I^e]. \end{aligned} \tag{34}$$

For the case where $\mathbf{z}_e = \mathbf{0}$, the above expression reduces to

$$\det[\mathbf{J}_f(\mathbf{0};\mathbf{E}_0) - \lambda\,\mathbf{I}_{2(N+1)}] = [(\gamma+\lambda)^2 + \mu^4]^{N-1}[(\lambda^2 + 2\Gamma\lambda + \mu^4)\{(\gamma+\lambda)^2 + \mu^4\} - 2\mu^4\|\mathbf{E}_0\|^2] \tag{35}$$

Evidently, when $\mathbf{E}_0 = \mathbf{0}$, the spectrum of $\mathbf{J}_f(\mathbf{0};\mathbf{0})$ is given by $[-\gamma \pm i\mu^2$ (multiplicity N), $-\Gamma \pm (\Gamma^2 - \mu^2)^{1/2}]$, which implies the asymptotic stability of the origin for $\gamma, \Gamma > 0$. For $\mathbf{E}_0 \neq 0$, the eigenvalues $\lambda = -\gamma \pm i\mu^2$ [multiplicity $(N-1)$] remain invariant, while the remaining eigenvalues are roots of the quartic equation:

$$\lambda^4 + 2(\gamma+\Gamma)\lambda^3 + (\mu^4 + \mu^2 + 4\gamma\Gamma + \gamma^2)\lambda^2 + 2[\Gamma(\gamma^2 + \mu^4) + \gamma\mu^2]\lambda + \mu^2[(\gamma^2 + \mu^4) - 2\mu^2\|\mathbf{E}_0\|^2] = 0. \tag{36}$$

Obviously, $\mathbf{J}_f(\mathbf{0};\mathbf{E}_0)$ has a zero eigenvalue when $\|\mathbf{E}_0\|^2 = \tilde{E}_{oc}^2 \triangleq (\mu^4 + \gamma^2)/(2\mu^2)$. This coincides with condition (30) for which one of the equilibrium points returns to the origin. It can be readily shown by using Routh's criterion[9] that the origin becomes unstable when $\|\mathbf{E}_0\| > \tilde{E}_{oc}$. In fact, (36) has only one unstable root and it is real and positive. Thus, the origin has a saddle point structure in a two-dimensional manifold. So we conclude that Hopf bifurcation cannot occur at the origin for any value of $\|\mathbf{E}_0\|$.

For the case where $\|\mathbf{E}_0\| \geqslant \mathrm{E}_{oc}$, there exist nonzero equilibrium states $\mathbf{z}_e$ which depend on $\|\mathbf{E}_0\|$. We observe from (34) that $\lambda = -\gamma \pm \mathrm{i}(\mu^2 + \alpha n^e)$ are stable eigenvalues of $\mathbf{J}_f(\mathbf{z}_e;\mathbf{E}_0)$ with a multiplicity of $(N-1)$. The remaining eigenvalues are given by the roots of the quartic equation:

$$\lambda^4 + a_3\lambda^3 + a_2(\|\mathbf{E}_0\|)\lambda^2 + a_1(\|\mathbf{E}_0\|)\lambda + a_0(\|\mathbf{E}_0\|) = 0. \tag{37}$$

where

$$\begin{aligned} &a_3 = 2(\gamma+\Gamma), \quad a_2(\|\mathbf{E}_0\|) = \gamma^2 + \mu^2 + 4\gamma\Gamma + (\mu^2 + \alpha n^e)^2, \\ &a_1(\|\mathbf{E}_0\|) = 2\{\Gamma[\gamma^2 + (\mu^2 + \alpha n^e)^2] + \gamma\mu^2 + (\alpha\mu^2 + \beta)\mathbf{E}_0\cdot\mathbf{E}_I^e\}, \\ &a_0(\|\mathbf{E}_0\|) = \mu^2[\gamma^2 + (\mu^2 + \alpha n^e)^2] + 2\{(\mu^2 + \alpha n^e)[(\beta\mathbf{E}_R^e - \mu^2\mathbf{E}_0)\cdot(\mathbf{E}_0 + \alpha\mathbf{E}_R^e) \\ &\qquad + \alpha\beta\|\mathbf{E}_I^e\|^2] + \gamma(\alpha\mu^2 + \beta)\mathbf{E}_0\cdot\mathbf{E}_I^e\}, \end{aligned} \tag{38}$$

where $\mathbf{E}_R^e$, $\mathbf{E}_I^e$, and n^e (given in Sec. 3) depend on $\|\mathbf{E}_0\|$. To determine the value of $\|\mathbf{E}_0\|$ for which Hopf bifurcation occurs, it is necessary to determine the existence of purely imaginary roots of (37) for some value of $\|\mathbf{E}_0\|$. From Routh's criterion, we can deduce that if

$$a_3 a_2(\|\mathbf{E}_0\|) > a_1(\|\mathbf{E}_0\|) \tag{39}$$

and

$$a_1(\|\mathbf{E}_0\|)[a_3 a_2(\|\mathbf{E}_0\|) - a_1(\|\mathbf{E}_0\|)] = a_3^2 a_0(\|\mathbf{E}_0\|), \tag{40}$$

then (38) has a pair of purely imaginary roots given by $\lambda = \pm\{a_3 a_0[a_3 a_2(\|\mathbf{E}_0\|) - a_1(\|\mathbf{E}_0\|)]\}^{1/2} i$.

Let E_{0H} be the value of $\|\mathbf{E}_0\|$ such that both (39) and (40) are satisfied; and $\tilde{\lambda}_\pm(\|\mathbf{E}_0\|) = \tilde{\lambda}_R(\|\mathbf{E}_0\|) \pm i\tilde{\lambda}_I(\|\mathbf{E}_0\|)$ be the roots of (37) such that $\tilde{\lambda}_R(E_{0H}) = 0$. By a lengthy but straightforward computation, it can be shown that $\tilde{\lambda}'_R$, the derivative of $\tilde{\lambda}_R$ with respect to the parameter $\|\mathbf{E}_0\|$, is given by

$$\begin{aligned} \tilde{\lambda}'_R(\|\mathbf{E}_0\|) = &[a_3^2 a_0'(\|\mathbf{E}_0\|) + 2a_1(\|\mathbf{E}_0\|)a_1'(\|\mathbf{E}_0\|) - a_3 a_2'(\|\mathbf{E}_0\|)a_1(\|\mathbf{E}_0\|) \\ &- a_3 a_2(\|\mathbf{E}_0\|)a_1'(\|\mathbf{E}_0\|)]\{2a_3[a_3 a_1(\|\mathbf{E}_0\|) + a_2^2(\|\mathbf{E}_0\|) - 4a_0(\|\mathbf{E}_0\|)]\}^{-1}, \end{aligned} \tag{41}$$

where a_j' denotes the derivative of a_j with respect to $\|\mathbf{E}_0\|$. For Hopf bifurcation,[10] $\tilde{\lambda}'_R(E_{0H}) > 0$. Due to the complicated dependence of a_2, a_1, and a_0 on $\|\mathbf{E}_0\|$, it is difficult to determine the threshold values of $\|\mathbf{E}_0\|$ for Hopf bifurcation. We shall resort to numerical computation at this point.

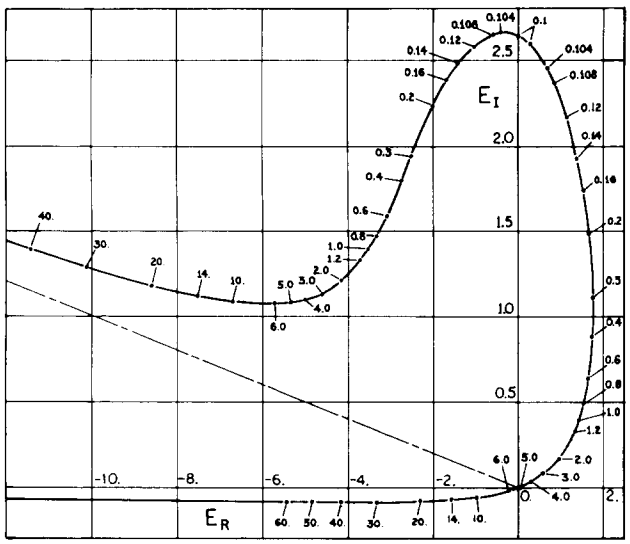


FIG. 1. Locus of equilibrium electric field $E^e = E_R^e + iE_I^e$ with E_0^2 as a parameter. ($E_e^+ = \xi^+ E_0$ and $E_e^- = \xi^- E_0$ are denoted by solid dots and circles respectively.

We note here that the Hopf bifurcation problem for (9) and (10) with spatial dimension N or $\dim(\mathbf{E}) > 1$ can be completely studied by considering only (37) which is the characteristic equation for the case with $N = 1$. Since for $\gamma > 0$ and $\Gamma > 0$, the additional eigenvalues $\lambda = -\lambda \pm i(\mu^2 + \alpha n^e)$ for $N > 1$ are stable, and $a_2, a_3(\|\mathbf{E}_0\|) > 0$ for all $\|\mathbf{E}_0\|$, the dimension of the unstable manifold associated with a nonzero equilibrium state is at most three.

5. BOUNDS FOR CHAOTIC OSCILLATIONS

The existence of chaotic oscillations depends on the manner in which the stable and unstable manifolds associated with the equilibrium points intersect with each other. At present, there are no readily verifiable analytical sufficient conditions for the existence of chaotic solutions for finite dimensional systems of ordinary differential equations. Here, we assume the existence of chaotic oscillations and proceed to derive bounds for their amplitudes, thus providing estimates for the size of the invariant manifold generated by the chaotic oscillations.

First, we shall make use of a function V of the form:

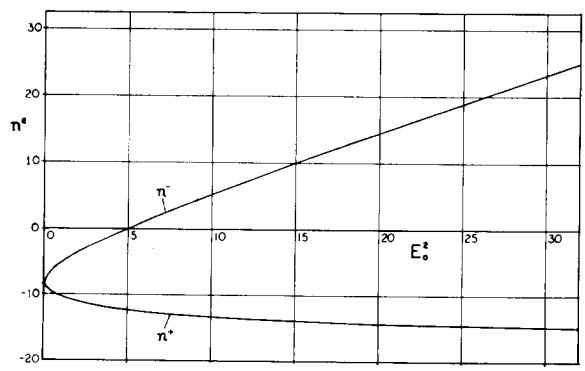


FIG. 2. Locus of equilibrium ion densitities n^+ and n^- with E_0^2 as a parameter.

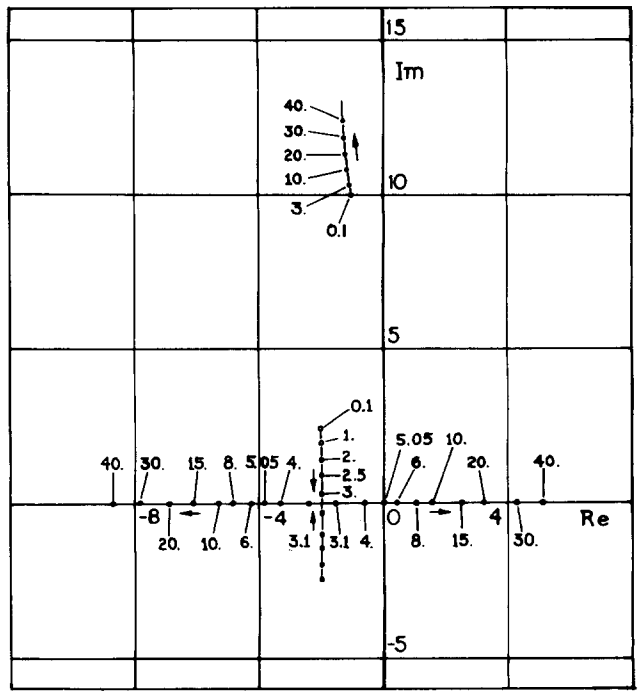


FIG. 3. Locus of the eigenvalues of $\mathbf{J}_f(0;E_0)$ with E_0^2 as a parameter.

$$V(\mathbf{E}) = |\mathbf{E} - \tilde{\mathbf{E}}|^2/2\,, \tag{42}$$

to estimate the magnitude of chaotic oscillations of the electric field where $\tilde{\mathbf{E}} = \mathrm{Re}\tilde{\mathbf{E}} + i\mathrm{Im}\tilde{\mathbf{E}}$ is to be determined. By direct computation:

$$\begin{aligned}\frac{dV}{dt} = &-\gamma\{\|\mathbf{E}_R + (\tfrac{1}{2})[\gamma^{-1}(\mu^2 + \alpha n)\mathrm{Im}\tilde{\mathbf{E}} - \mathrm{Re}\tilde{\mathbf{E}}]\|^2 \\ &+ \|\mathbf{E}_I + (\tfrac{1}{2})[\gamma^{-1}(\mu^2 + \alpha n)\mathrm{Re}\tilde{\mathbf{E}} - \mathrm{Im}\tilde{\mathbf{E}} + \gamma^{-1} n\mathbf{E}_0]\|^2 \\ &- \tfrac{1}{4}[\|\gamma^{-1}(\mu^2 + \alpha n)\mathrm{Im}\tilde{\mathbf{E}} - \mathrm{Re}\tilde{\mathbf{E}}\|^2 \\ &+ \|\gamma^{-1}(\mu^2 + \alpha n)\mathrm{Re}\tilde{\mathbf{E}} - \mathrm{Im}\tilde{\mathbf{E}} + \gamma^{-1} n\mathbf{E}_0\|^2]\}. \end{aligned} \tag{43}$$

If we set $\mathrm{Re}\tilde{\mathbf{E}} = -\alpha^{-1}\mathbf{E}_0$ and $\mathrm{Im}\tilde{\mathbf{E}} = \mathbf{0}$, then (43) reduces to

$$\frac{dV}{dt} = -\gamma\{\|\mathbf{E}_R + (2\alpha)^{-1}\mathbf{E}_0\|^2 + \|\mathbf{E}_I - \mu^2\mathbf{E}_0/(2\alpha\gamma)\|^2$$

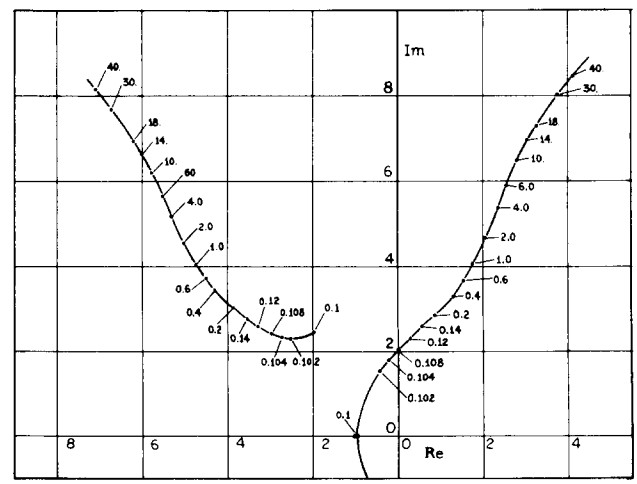


FIG. 4. Locus of the eigenvalues of $\mathbf{J}_f(\mathbf{z}_e^+;E_0)$ with E_0^2 as a parameter.

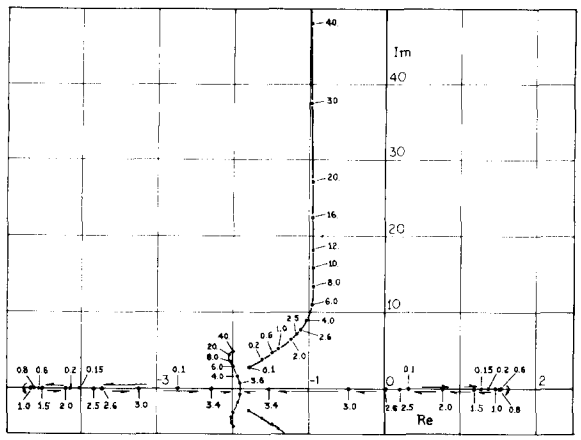


FIG. 5. Locus of the eigenvalues of $\mathbf{J}_f(\mathbf{z}_s;E_0)$ with E_0^2 as a parameter.

$$-(2\alpha)^{-2}(1+\mu^4\gamma^{-2})\|\mathbf{E}_0\|^2\}$$
$$=-\gamma[|\mathbf{E}-\mathbf{E}_s|^2-(2\alpha)^{-2}(1+\mu^4\gamma^{-2})\|\mathbf{E}_0\|^2], \quad (44)$$

where $\mathbf{E}_s=-(2\alpha)^{-1}\mathbf{E}_0+i\mu^2\mathbf{E}_0/(2\alpha\gamma)$. We note that with the foregoing choice of $\tilde{\mathbf{E}}$, n does not appear in (44). Moreover, for $\gamma>0$, $dV/dt<0$ at any point $\mathbf{E}$ exterior to the set $\Xi=\{\mathbf{E}:|\mathbf{E}-\mathbf{E}_s|\leqslant(2\alpha)^{-1}(1+\mu^4\gamma^{-2})^{1/2}\|\mathbf{E}_0\|\}$. Let $\Xi_\delta=\{E:V(E)=\frac{1}{2}|E+\alpha^{-1}E_0|^2\leqslant\delta^2\}$. Since for any $\delta>0$, Ξ_δ is a ball in $\mathbf{E}$-space centered about the point $-\alpha^{-1}\mathbf{E}_0$ with radius $\sqrt{2}\delta$, it is possible to select a δ such that $\tilde{\Xi}\subset\Xi_\delta$. In fact, elementary geometric considerations show that the smallest δ having the foregoing inclusion property is given by $\check{\delta}=(\sqrt{2}\alpha)^{-1}(1+\mu^4\gamma^{-2})^{1/2}\|\mathbf{E}_0\|$. Evidently, $dV/dt<0$ at any point $\mathbf{E}$ exterior to $\Xi_{\check{\delta}}$. This implies that for a solution of (31) initiated from any point $\mathbf{z}(0)=(n(0),\dot{n}(0),\mathbf{E}_R(0),\mathbf{E}_I(0))$ at $t=0$ with $\mathbf{E}(0)=\mathbf{E}_R(0)+i\mathbf{E}_I(0)$ exterior to $\Xi_{\check{\delta}}$, its corresponding $\mathbf{E}(t)$, $t>0$ either eventually enters $\Xi_{\check{\delta}}$ at some finite time $t_1>0$ and remains in $\Xi_{\check{\delta}}$ for all $t>t_1$ or tends to $\Xi_{\check{\delta}}$ as $t\to\infty$. Clearly, $\Xi_{\check{\delta}}$ contains all the points $\mathbf{E}(t)$ along any chaotic solution of (31) when it exists.

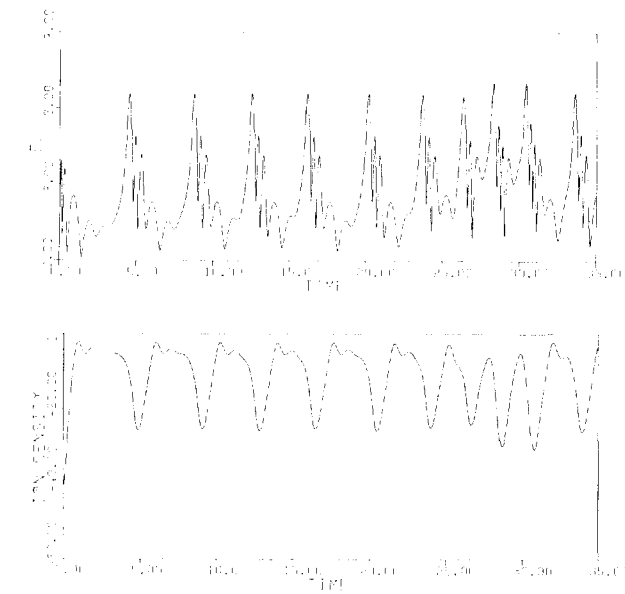


FIG. 6. $|E(t)|$ and $n(t)$ vs time t corresponding to the solution of (53) and (54) with $m=1, L=\pi/\sqrt{10}, \Gamma_1=2.0, \gamma_1=1.0, E_0^2=1.625$; initial data: $E(0)=1.426+0.5071i$, $n(0)=-40.75$ and $\dot{n}(0)=34.58$.

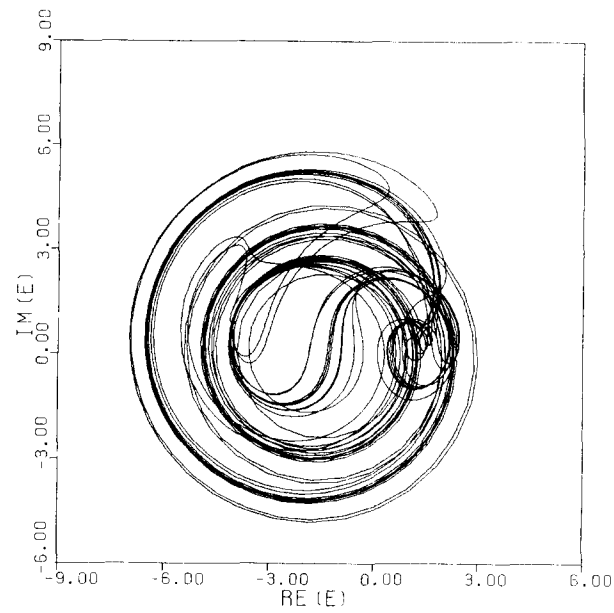


FIG. 7. Projection of the trajectory of (53) and (54) (with parameters as given in Fig. 6) onto the (E_R,E_I)-plane.

Next, we derive a bound for the magnitude of ion density oscillations. Let $\tilde{n}=n+\mu^2\beta^{-1}\|\mathbf{E}_0\|^2$. We can rewrite the first two equations in (31) as

$$\frac{d\mathcal{N}}{dt}=\mathcal{A}\mathcal{N}+[0,\beta|\mathbf{E}-\mu^2\beta^{-1}\mathbf{E}_0|^2]^T, \quad (45)$$

where $\mathcal{N}=(\tilde{n},\dot{\tilde{n}})^T$ and

$$\mathcal{A}=\begin{bmatrix}0 & 1\\ -\mu^2 & -2\Gamma\end{bmatrix}. \quad (46)$$

Given $\mathcal{N}(0)$, the initial data for $\mathcal{N}$ at $t=0$, (45) is equivalent to the integral equation:

$$\mathcal{N}(t)=[\exp\mathcal{A}t\,]\mathcal{N}(0)+\int_0^t\exp[\mathcal{A}(t-\tau)]\times[0,\beta|\mathbf{E}(\tau)-\mu^2\beta^{-1}\mathbf{E}_0|^2]^T d\tau. \quad (47)$$

Thus,

$$\|\mathcal{N}(t)\|\leqslant\|\exp\mathcal{A}t\,\|\|\mathcal{N}(0)\|+\int_0^t|\beta|\,\|\exp\mathcal{A}(t-\tau)\|\times|\mathbf{E}(\tau)-\mu^2\beta^{-1}\mathbf{E}_0|^2 d\tau. \quad (48)$$

For $\mu^2\geqslant\Gamma>0$, we can find a constant $\mathscr{C}>0$ such that $\|\exp(\mathcal{A}t)\|\leqslant\mathscr{C}\exp(-\Gamma t)$. Also, we have already established that along any chaotic solution $\mathbf{E}(t)\subset\Xi_{\check{\delta}}$ or $|\mathbf{E}(t)+\alpha^{-1}\mathbf{E}_0|\leqslant\check{\delta}=(\sqrt{2}\alpha)^{-1}(1+\mu^4\gamma^{-2})^{1/2}\|\mathbf{E}_0\|$ for all t. Hence,

$$\begin{aligned}&|\mathbf{E}(\tau)-\mu^2\beta^{-1}\mathbf{E}_0|\\ &\quad=|\mathbf{E}(\tau)+\alpha^{-1}\mathbf{E}_0-(\alpha^{-1}+\mu^2\beta^{-1})\mathbf{E}_0|\\ &\quad\leqslant|\mathbf{E}(\tau)+\alpha^{-1}\mathbf{E}_0|+|\alpha^{-1}+\mu^2\beta^{-1}|\|\mathbf{E}_0\|\\ &\quad\leqslant\check{\delta}+|\alpha^{-1}+\mu^2\beta^{-1}|\|\mathbf{E}_0\|\leqslant\Psi\|\mathbf{E}_0\|, \end{aligned} \quad (49)$$

where $\Psi=(\sqrt{2}\alpha)^{-1}(1+\mu^4\gamma^{-2})^{1/2}+|\alpha^{-1}+\mu^2\beta^{-1}|$. It follows from (48) that

$$\|\mathcal{N}(t)\|\leqslant\mathscr{C}\{\Psi|\beta|\|\mathbf{E}_0\|^2\Gamma^{-1}+[\|\mathbf{z}(0)\|-\Psi|\beta|\|\mathbf{E}_0\|^2\Gamma^{-1}]\exp(-\Gamma t)\}$$

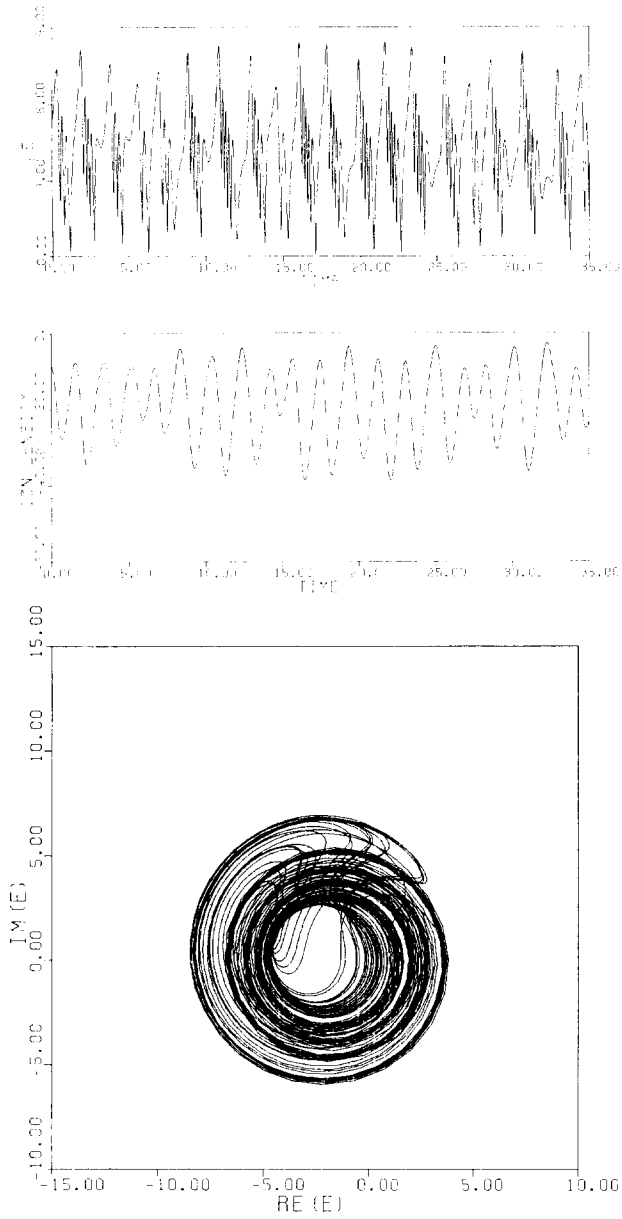


FIG. 8. Solution of (53) and (54) with $E_0^2 = 2.669$, $E(0) = -3.07656 + 4.22328i$, $n(0) = -8.38056$, $\dot{n}(0) = -8.25344$, and other parameters as given in Fig. 6.

$$\leqslant \mathscr{C} \max\{\|\mathbf{z}(0)\|, \Psi \mid \beta \mid \|\mathbf{E}_0\|^2 \Gamma^{-1}\}, \tag{50}$$

for all $t \geqslant 0$ and $\mathbf{E}(0) \in \Xi_\delta$. When $\|\mathbf{z}(0)\| \leqslant \Psi \mid \beta \mid \|\mathbf{E}_0\|^2 \Gamma^{-1}$, we have $\|\mathbf{z}(t)\| \leqslant \mathscr{C} \Psi \mid \beta \mid \|\mathbf{E}_0\|^2 \Gamma^{-1}$ for all $t \geqslant 0$, a bound which is independent of $\mathbf{z}(0)$.

6. ONE-DIMENSIONAL EXAMPLE

Consider the case where $N = 1$ with a bounded spatial domain $\Omega = [0, L]$. Assuming that both E and n vanish at the boundary points $x = 0$ and $x = L$, we can take $\phi_k(x) = (2/L)^{1/2} \sin(k\pi x/L)$, $\mu_k = k\pi/L$, $k = 1,2\cdots$. For this case, the coefficients α_m and β_m defined in (11) become

$$\alpha_m = \int_0^L (2/L)^{3/2} \sin^3(m\pi x/L)\, dx = \begin{cases} 0 & \text{for } m \text{ even}, \\ (2^7/L)^{1/2}/(3m\pi) & \text{for } m \text{ odd}, \end{cases} \tag{51}$$

$$\beta_m = \int_0^L \left(\frac{2}{L}\right)^{3/2} \frac{d^2}{dx^2} [\sin^2(m\pi x/L)] \sin(m\pi x/L)\, dx = \begin{cases} 0 & \text{for } m \text{ even}, \\ -8m\pi (2/L^5)^{1/2}/3 & \text{for } m \text{ odd}. \end{cases} \tag{52}$$

Thus, for an odd integer m, Eqs. (9) and (10) have the form:

$$i\frac{dE_m}{dt} - \left(\frac{m\pi}{L}\right)^2 E_m + i\gamma_m E_m = n_m \left[E_0 + \left(\frac{2^7}{L}\right)^{1/2} /(3m\pi) E_m\right], \tag{53}$$

$$\frac{d^2 n_m}{dt^2} + 2\Gamma_m \frac{dn_m}{dt} + \left(\frac{m\pi}{L}\right)^2 n_m = -\left(\frac{m\pi}{L}\right)^2 E_0 (E_m + E_m^*) - \left[8m\pi \frac{(2/L^5)^{1/2}}{3}\right] |E_m|^2, \tag{54}$$

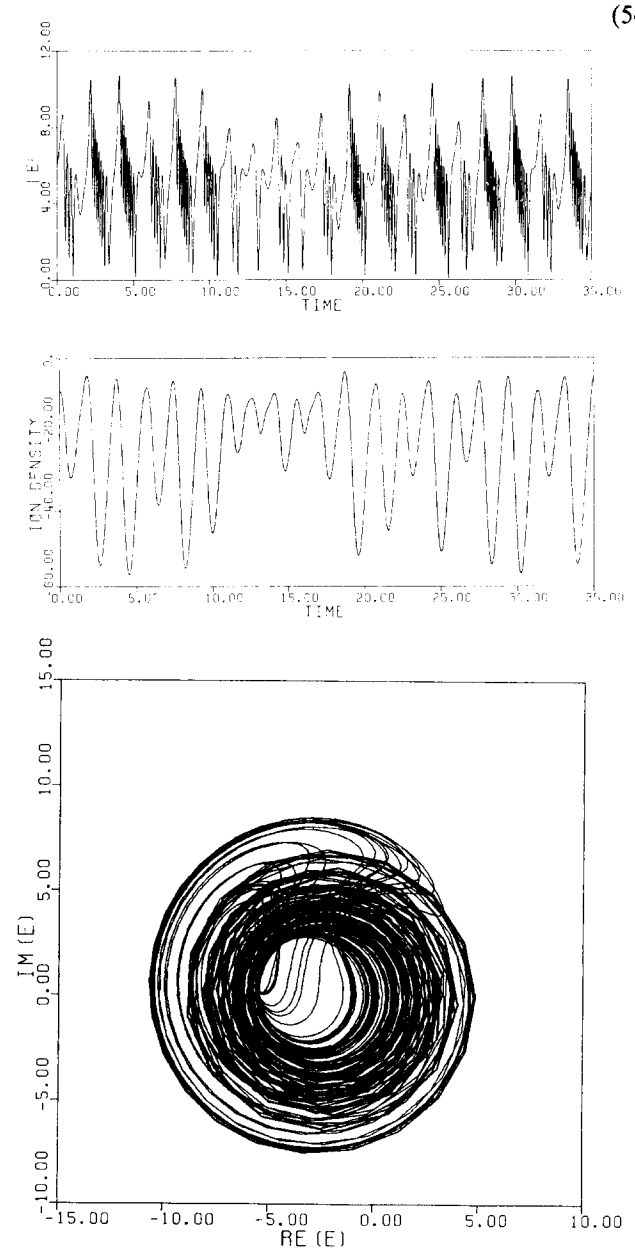


FIG. 9. Solution of (53) and (54) with $E_0^2 = 5.05$, $E(0) = -5.0069 + 3.26742i$, $n(0) = -8.63114$, $\dot{n}(0) = 1.86353$, and other parameters as given in Fig. 6.

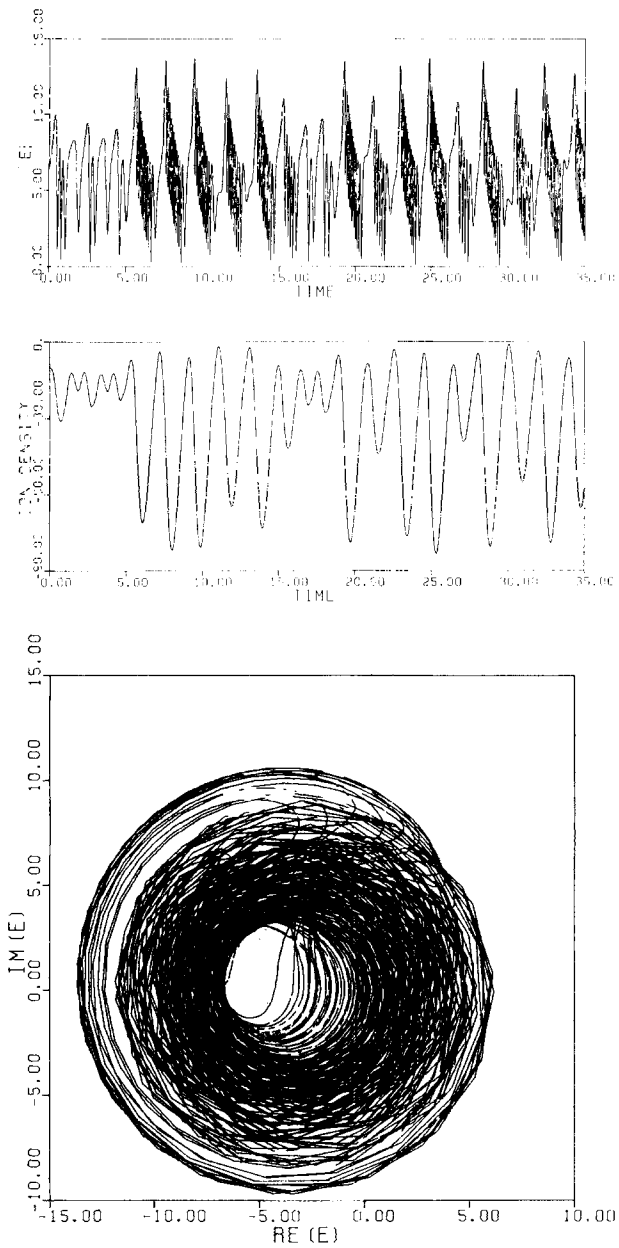


FIG. 10. Solution of (53) and (54) with $E_0^2 = 10.0$, $E(0) = -6.0 + 2.0i$, $n(0) = -10$, $\dot{n}(0) = 0$, and other parameters as given in Fig. 6.

where E_0 is a real nonnegative parameter.

To illustrate the qualitative features of the solutions of the above equations, numerical results are obtained for the case where $m = 1$, $L = \pi/\sqrt{10}$, $\Gamma_1 = 2.0$, and $\gamma_1 = 1.0$. These values are chosen to simplify the numerical computation. They may not correspond to any particular physical situation. Numerical results for specific physical situations will be presented elsewhere.

First, we compute the locus of the equilibrium electric field E^e as a function of E_0^2 using (27) and (20). Figure 1 shows the locus in the (E_R, E_I)-plane. It can be seen that bifurcation occurs at $E_0^2 = E_{0c}^2 = 0.09974$. At $E_0^2 = \tilde{E}_{0c}^2 = 5.05$, one of the equilibrium points E^e returns to the origin. As $E_0^2 \to \infty$, the locus of one of the equilibrium points is asymptotic to the line $E_I = \gamma\beta^{-1}(2\alpha_1 + \beta_1\mu_1^{-1})E_R = -0.1\,E_R$, while the other one tends to $(E_R, E_I) = (-\infty, 0)$. Figure 2 shows the equilibrium ion densities n^e as a function of E_0^2.

Next, we examine the nature of each equilibrium state $\mathbf{z}_e$ for various values of E_0^2 by determining the roots of (37) or the eigenvalues of $\mathbf{J}_f(\mathbf{z}_e; E_0)$. Figure 3 shows the locus of the eigenvalues of $\mathbf{J}_f(0; E_0)$ as E_0^2 varies. As established in Sec. 4, when $E_0^2 < \tilde{E}_{0c}^2 = 5.05$, all the eigenvalues have negative real parts implying that the origin is asymptotically stable. When E_0^2 exceeds $\tilde{E}_{0c}^2$, one of the real eigenvalues crosses the imaginary axis. Consequently, the origin becomes unstable. Figure 4 shows the eigenvalue locus of $\mathbf{J}_f(\mathbf{z}_e^+; E_0)$ for $0.1 \leqslant E_0^2 \leqslant 40.0$, where the components of $\mathbf{z}_e^+ = (n^+, 0, \xi_R^+ E_0, \xi_I^+ E_0)$ are given by (17), (20), and (27). We note that for $0.1 \leqslant E_0^2 \leqslant 0.108$, all the eignevalues have negative real parts, and at $E_0^2 \approx 0.108$, a complex conjugate pair of eigenvalues cross the imaginary axis into the right-half plane. It can be verified that Hopf bifurcation takes place at this point. The locus of the eigenvalues of $\mathbf{J}_f(\mathbf{z}_e^-; E_0)$ is shown in Fig. 5. Here, for $0.1 \leqslant E_0^2 \leqslant 2.669$, $\mathbf{J}_f(\mathbf{z}_e^-; E_0)$ has a positive real eigenvalue. When $E_0^2 \geqslant 2.67$, all the eigenvalues of $\mathbf{J}_f(\mathbf{z}_e^-; E_0)$ are in the left-half plane.

An inspection of the eigenvalue loci given by Figs. 3–5 suggests that one might search for the existence of chaotic solutions in the neighborhood of $\mathbf{z}_e^+$ for $E_0^2 > 0.108$ (Hopf bifurcation point). Numerical integration of (53) and (54) with various initial conditions was performed for progressively larger values of E_0^2. The results suggest that the periodic solutions in the neighborhood of $\mathbf{z}_e^+$ (whose existence is ensured by the Hopf bifurcation theorem) are unstable and the bifurcation is subcritical. Figure 6 shows the time-domain buildup of a nearly periodic solution which evolves into chaotic oscillations. The projection of the trajectory onto the (E_R, E_I)-plane is shown in Fig. 7. Figures 8–10 show the chaotic solutions for various values of E_0^2. It was found that

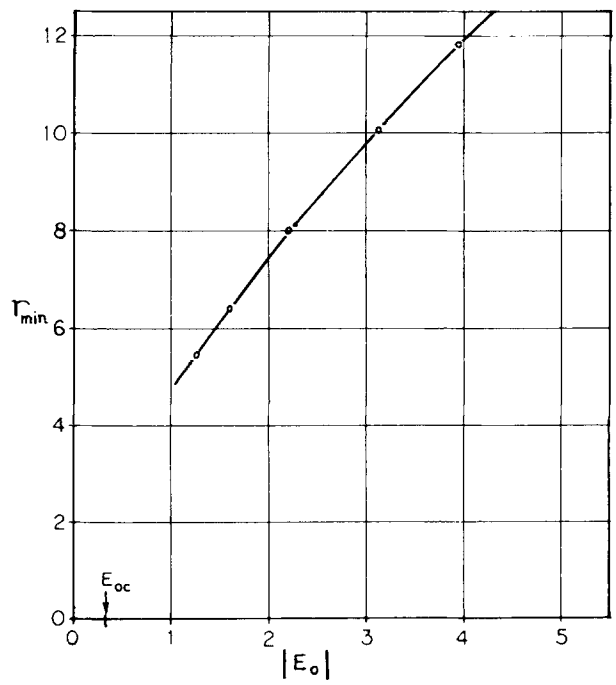


FIG. 11. Variation of r_{min} as a function of $|E_0|$.

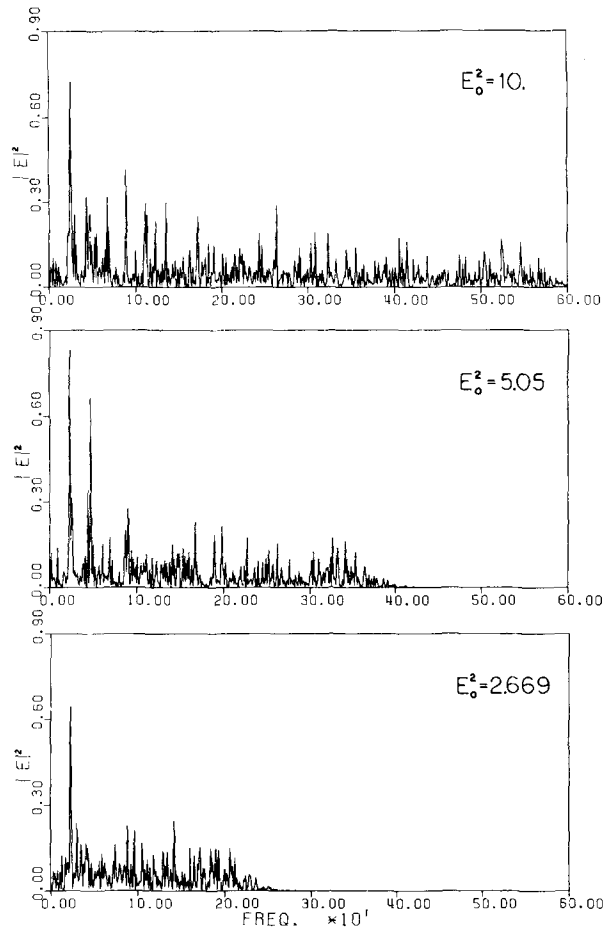


FIG. 12. Power spectra of the electric field corresponding to the solutions of (53) and (54) as shown in Figs. 8–10; frequency scale: 1/40.96 normalized unit.

these solutions are highly sensitive to initial conditions. Also, not all trajectories in the z space tend to the chaotic solutions as $t \to \infty$. This is apparent from the fact that for $0.108 \leqslant E_0^2 < \tilde{E}_{0c}^2 = 5.05$, the origin $\mathbf{z} = \mathbf{0}$ is a stable equilibrium point, and for any $E_0^2 > \tilde{E}_{0c}^2$, $\mathbf{z}_e^-$ is always a stable equilibrium point. We note from Figures 8b–10b that in each case, there exists a circle with minimum radius $r_{\min}$ which encloses the projection of the trajectories onto the (E_R, E_I)-plane. Figure 11 shows the variation of $r_{\min}$ as a function $|E_0|$ as obtained from the numerical solutions. Evidently, $r_{\min}$ can be bounded by a linear fundtion of $|E_0|$. This is consistent with the estimate δ given in Sec. 5. It can be readily verified that in each case, the projection of the trajectories of the chaotic oscillations onto the (E_R, E_I)-plane is completely contained in $\Xi_\delta = \{E: |E + \alpha^{-1} E_0| \leqslant \sqrt{2\delta} = 8.3445 |E_0|\}$, where $\alpha = 1.204367$. Also, we observe from these solutions that the maximum depth of the ion density troughs increases with E_0^2, and the electric field oscillates more rapidly during the ion density dips. This can be roughly explained by considering the following equations for E_R and E_I derived from (31):

$$\frac{d^2 E_R}{dt^2} + \left(2\gamma - \frac{\alpha \dot{n}}{(\mu^2 + \alpha n)}\right)\frac{dE_R}{dt} + \left(\gamma^2 + (\mu^2 + \alpha n)^2 - \frac{\alpha\gamma\dot{n}}{(\mu^2 + \alpha n)}\right)E_R = -n(\mu^2 + \alpha n)E_0, \tag{55}$$

$$\frac{d^2 E_I}{dt^2} + \left(2\gamma - \frac{\alpha \dot{n}}{(\mu^2 + \alpha n)}\right)\frac{dE_I}{dt} + \left(\gamma^2 + (\mu^2 + \alpha n)^2 - \frac{\alpha\gamma\dot{n}}{(\mu^2 + \alpha n)}\right)E_I = -\left(\dot{n} + \left(\gamma - \frac{\alpha \dot{n}}{(\mu^2 + \alpha n)}\right)n\right)E_0. \tag{56}$$

Considering n as a slowly time-varying parameter, the frequency of electric-field oscillations is roughly equal to $\omega = [\gamma^2(\mu^2 + \alpha n)^2 - \alpha\gamma\dot{n}(\mu^2 + \alpha n)^{-1}]^{1/2}$, and the effective damping coefficient is $2\gamma - \alpha\dot{n}(\mu^2 + \alpha n)^{-1}$. Let T denote the time interval corresponding to an ion density dip, and t^* is the minimum point of n over T where $\dot{n}(t^*) = 0$. Then $\omega(t^*) > \omega(t)$ for all t in T such that $\dot{n}(t)$ and $\mu^2 + \alpha n(t)$ have the same sign. This condition is satisfied for the solutions shown here.

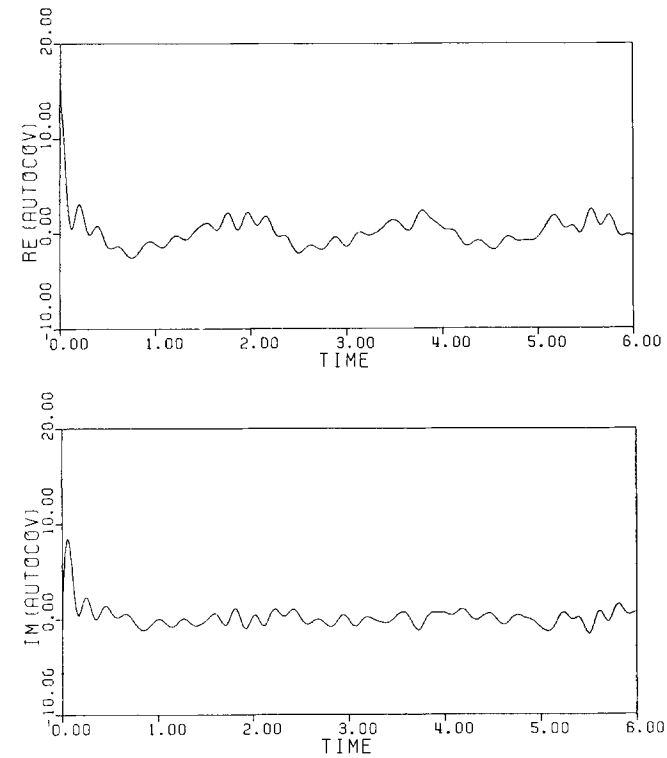
Figure 12 shows the power spectra of the electric field computed by means of the fast Fourier transform method. The results resemble those corresponding to turbulence. Also, the spectral bandwidth increases with E_0^2 as expected from physical considerations. Finally, the truncated discrete version of the autocovariance function of E given by

$$\rho(j\Delta) = \frac{1}{(N - j + 1)} \sum_{i=1}^{N-j+1} [E(i\Delta) - \bar{E}] \times \{E[(i + j - 1)\Delta] - \bar{E}\}^* \tag{57}$$

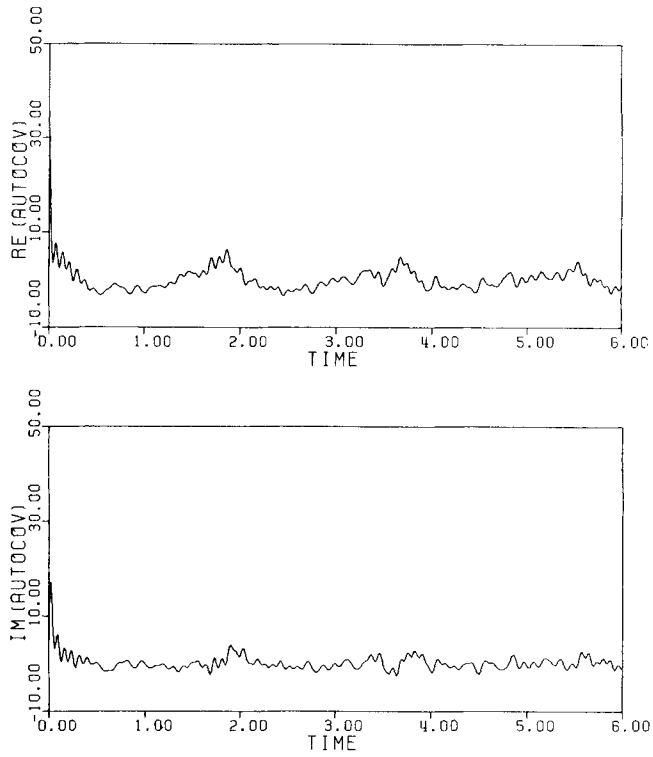
is computed, where $\bar{E}$ denotes the mean-value of E, and Δ is the time-step size. Figures 13a–13c show the real and imaginary parts of $\rho(j\Delta)$ for $E_0^2 = 1.669$, 5.05 and 10.0. It can be seen that both $\mathrm{Re}\,\rho(j\Delta)$ and $\mathrm{Im}\,\rho(j\Delta)$ decay from their maximum values and then fluctuate about zero. But we cannot deduce that the autocovariance function actually tends to zero as the time delay $\tau \to \infty$ as in the case of solutions on a strange attractor.

7. CONCLUDING REMARKS

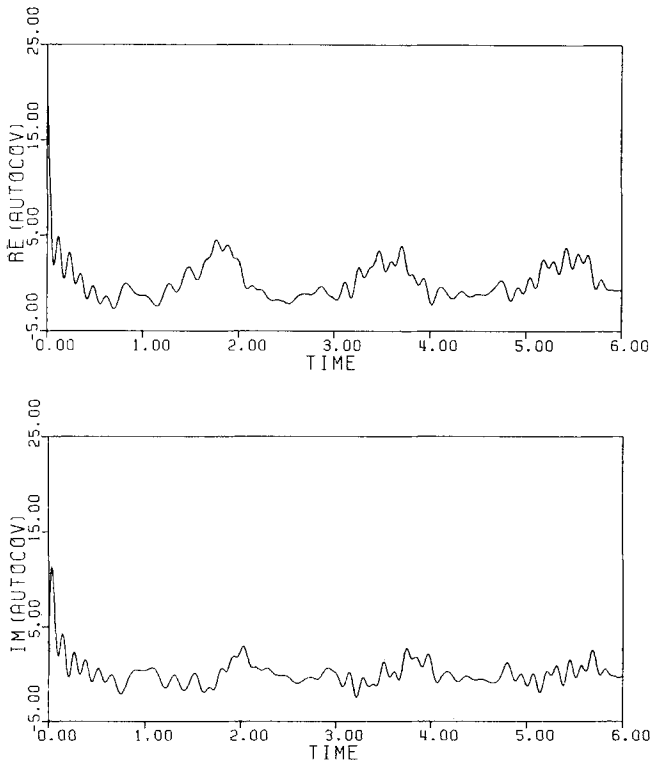
It was found that the single-mode equations derived from the Zakharov's model for Langmuir turbulence in a plasma with phenomenological damping exhibit chaotic solutions whose power spectra have turbulence-like features. In the case of multiple modes, if all the mode coupling terms are omitted, then we obtain sets of uncoupled equations of the form (9) and (10). Each set is capable of producing chaotic solutions when E_0^2 exceeds a certain threshold value (generally different for each mode). The total power spectrum of the electric field is simply the sum of the single-mode power spectra. This seems to imply that energy transfer between various modes is not necessary in producing turbulence which is contrary to the cascade theory of turbulence. There are a number of computer studies[7] of Langmuir turbulence induced by interacting collapsing solitary waves based on Zakharov's model with phenomenological damping. Perhaps these computer results actually correspond to some form of chaotic solutions which are inherent in the model.



(a)



(c)



(b)

FIG. 13. Real and imaginary parts of the autocovariance function of E corresponding to the solutions shown in Figs. 8–10; (a) $E_0^2 = 2.669$, (b) $E_0^2 = 5.05$, and (c) $E_0^2 = 10$.

In this work, we have sought solutions in terms of the eigenfunctions of the Laplacian operator over a bounded spatial domain. Of course, we may expand the solutions in terms of any suitable countable basis for $L^2(\Omega)$ and arrive at a countably infinite system of ordinary differential equations similar to that given in (5) and (6). One may also consider directly the Hopf bifurcation problem for Zakharov's model (1) and (2) without resorting to modal expansions. Some results in this direction have been obtained recently. They will be reported elsewhere.

Finally, we note that the presence of the phenomenological damping coefficients γ_m and Γ_m in the simplified equations for each mode m is essential for the existence of chaotic solutions. But there does not exist a clearcut way of introducing the damping terms into the Zakharov's model based on physical considerations. Also, a detailed study of the structure of the stable and unstable manifolds associated with the equilibrium states is necessary for revealing the nature of the chaotic oscillations described here. Unfortunately, this task is complicated by the system's dimensionality.

ACKNOWLEDGMENT

The author wishes to thank the referee for his helpful comments and suggestions. This work was supported by AFOSR grants No. 74-2662 and No. 79-0050.

APPENDIX

From (33), it is evident that

$$\det[\mathbf{J}_f(\mathbf{z}^e;E_0)-\lambda\,\mathbf{I}_{2(N+1)}]=\lambda(2\Gamma+\lambda)\det\mathbf{P}-\det\mathbf{Q}, \tag{A1}$$

where

$$\mathbf{P}\begin{bmatrix}-(\gamma+\lambda)\mathbf{I}_N & (\mu^2+\alpha n^e)\mathbf{I}_N\\ -(\mu^2+\alpha n^e)\mathbf{I}_N & -(\gamma+\lambda)\mathbf{I}_N\end{bmatrix},$$

$$Q=\left[\begin{array}{c|cc} -\mu^2 & 2(\beta\mathbf{E}_R^e-\mu^2E_0)^T & 2\beta(\mathbf{E}_I^e)^T\\ \alpha\mathbf{E}_I^e & & \\ \hline & \multicolumn{2}{c}{\mathbf{P}}\\ -(E_0+\alpha\mathbf{E}_R^e) & & \end{array}\right]. \tag{A2}$$

Since $\det\mathbf{P}=[(\gamma+\lambda)^2+(\mu^2+\alpha n^e)^2]^N>0$, $\mathbf{P}^{-1}$ exists and is given by

$$\mathbf{P}^{-1}=[(\gamma+\lambda)^2+(\mu^2+\alpha n^e)^2]^{-1}\times\begin{bmatrix}-(\gamma+\lambda)\mathbf{I}_N & -(\mu^2+\alpha n^e)\mathbf{I}_N\\ (\mu^2+\alpha n^e)\mathbf{I}_N & -(\gamma+\lambda)\mathbf{I}_N\end{bmatrix}. \tag{A3}$$

Now, $\det\mathbf{Q}$ can be computed by considering the matrix

$$\mathbf{S}=\mathbf{Q}\begin{bmatrix}1 & \mathbf{O}_{2N}^T\\ \mathbf{O}_{2n} & \mathbf{P}^{-1}\end{bmatrix}=\left[\begin{array}{c|cc} -\mu^2 & 2[(\beta\mathbf{E}_R^e-\mu^2E_0)^T & \beta(\mathbf{E}_I^e)^T]\mathbf{A}^{-1}\\ \alpha\mathbf{E}_I^e & & \\ \hline & \multicolumn{2}{c}{\mathbf{I}_{2N}}\\ -(E_0+\alpha\mathbf{E}_R^e) & & \end{array}\right]. \tag{A4}$$

Since $\det\mathbf{S}=(\det\mathbf{Q})(\det\mathbf{P}^{-1})$, $\det\mathbf{P}^{-1}=[(\gamma+\lambda)^2+(\mu^2+\alpha n^e)^2]^{-N}$ and

$$\det\mathbf{S}=-\mu^2-2[(\beta\mathbf{E}_R^e-\mu^2E_0)^T]\mathbf{A}^{-1}\begin{bmatrix}\alpha\mathbf{E}_I^e\\ -(E_0+\alpha\mathbf{E}_R^e)\end{bmatrix}, \tag{A5}$$

we have $\det\mathbf{Q}=\det\mathbf{S}/\det\mathbf{P}^{-1}$. The expression (34) is obtained directly from (A1) and (A5).

[1]E.N. Lorenz, J. Atmos. Sci. **20**, 130 (1972).
[2]J.B. McLaughlin and P.C. Martin, Phys. Rev. A **12**, 186 (1975).
[3]O.E. Rössler, in *Synergetics, A Workshop*, edited by H. Haken (Springer-Verlag, New York, 1977), p. 184.
[4]Y. Kuramoto and T. Yamada, Prog. Theor. Phys. **55**, 679 (1976).
[5]D. Ruelle and F. Takens, Commun. Math. Phys. **20**, 167 (1971).
[6]V.E. Zakharov, Zh. Eksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].
[7]N.R. Pereira, R.N. Sudan, and J. Denavit, Phys. Fluids **20**, 271 (1977).
[8]T.A. Davydova and K.P. Shamrai, Fiz. Plazmy **3**, 591 (1977) [Sov. J. Plasma Phys. **3**, 333 (1978)].
[9]E.J. Routh, *Advanced Dynamics of a System of Rigid Bodies* (Dover, New York, 1955).
[10]J.E. Marsden and M. McCracken, *The Hopf Bifurcation and its Applications* (Springer-Verlag, New York, 1976).

A comparison of iteration schemes for Chandrasekhar H-equations in multigroup neutron transport

C. T. Kelley

Department of Mathematics, North Carolina State University, Raleigh, North Carolina
(Received 18 October 1978)

An iteration scheme for the Chandrasekhar H-equations in multigroup neutron transport is shown to converge to the solution of physical interest. Moreover, the convergence is more rapid than that provided by direct iteration.

I. INTRODUCTION

Matrix-valued analogs of the Chandrasekhar H-function [1] that arise in multigroup neutron transport theory [2] satisfy a coupled system of nonlinear integral equations. We write these equations in matrix form as

$$\begin{pmatrix} H_l(\mu,\omega) & 0 \\ 0 & \tilde{H}_r(\mu,\omega) \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + \mu\omega \begin{pmatrix} H_l(\mu,\omega) & 0 \\ 0 & \tilde{H}_r(\mu,\omega) \end{pmatrix} \times \int_0^1 \begin{pmatrix} \Psi(\nu) & 0 \\ 0 & \tilde{\Psi}(\nu) \end{pmatrix} \begin{pmatrix} H_r(\nu,\omega) & 0 \\ 0 & \tilde{H}_l(\nu,\omega) \end{pmatrix} \frac{d\nu}{\mu+\nu} \tag{1}$$

In Eq. (1), I is the $n \times n$ identity matrix, $\sim$ denotes transpose, Ψ is a matrix-valued function of ν with integrable, nonnegative entries, and ω is a complex parameter. $n \times n$ matrix-valued functions, H_l and H_r, are sought.

If A_r and A_l are $n \times n$ matrices and A is the $2n \times 2n$ matrix given by

$$A = \begin{pmatrix} A_l & 0 \\ 0 & \tilde{A}_r \end{pmatrix}, \tag{2}$$

we define A^* by

$$A^* = \begin{pmatrix} A_r & 0 \\ 0 & \tilde{A}_l \end{pmatrix}. \tag{3}$$

If we let E denote the $2n \times 2n$ identity matrix and, as in Refs. 3–6, define $2n \times 2n$ matrices H and D by

$$H = \begin{pmatrix} H_l & 0 \\ 0 & \tilde{H}_r \end{pmatrix}, \quad D = \begin{pmatrix} \Psi & 0 \\ 0 & \tilde{\Psi} \end{pmatrix}, \tag{4}$$

we may write Eq. (1) in the more compact form,

$$H(\mu,\omega) = E + \mu\omega H(\mu,\omega) \int_0^1 D(\nu) H^*(\nu,\omega) \frac{d\nu}{\mu+\nu} \tag{5}$$

If we let L denote the linear integral operator in Eq. (5), we may write

$$H(\omega) = E + \omega H(\omega) L H^*(\omega). \tag{6}$$

Equation (6) has multiple solutions, only one of which is of physical interest. We denote this solution by $H(\mu,\omega)$. H is the only solution of Eq. (6) that is analytic in ω near $\omega = 0$.

If we make the normalization

$$\| \int_0^1 \Psi(\nu) d\nu \|_{sp} = \tfrac{1}{2}, \tag{7}$$

where $\|\cdot\|_{sp}$ denotes spectral radius, then $H(\mu,\omega)$ is analytic in ω for $|\omega| < 1$ and continuous in μ and ω for $0 \leqslant \mu \leqslant 1$ and $|\omega| \leqslant 1$.

Of interest here is solvability of Eq. (6) by iteration. We consider two interation schemes. The first is direct iteration:

$$H_0 = E, \quad H_n = E + \omega H_{n-1} L H^*_{n-1}, \quad n \geqslant 1. \tag{8}$$

The iteration scheme given by Eq. (8) has been studied extensively. [5–11] The main result [6] is that $H_n(\mu,\omega)$ converges to $H(\mu,\omega)$ uniformly in μ and ω for $0 \leqslant \mu \leqslant 1$ and $|\omega| \leqslant 1$.

The second scheme is direct iteration of the following equation, which is equivalent to Eq. (6):

$$H(\omega) = [E - \omega L H^*(\omega)]^{-1}. \tag{9}$$

The iteration scheme is

$$K_0 = E, \quad K_n = (E - \omega L K^*_{n-1})^{-1} \tag{10}$$

In Ref. 7 the author showed that in the scalar case K_n coverges to H. Bowden [12] has also considered this question. In this paper we show not only that K_n coverges to H but also that the convergence of K_n to H is more rapid than that of H_n to H. Before stating our main result, we define a norm as follows. If A is a $2n \times 2n$ matrix-valued function of μ for $0 \leqslant \mu \leqslant 1$, define

$$\|A\| = \max_{\substack{1 \leqslant i,j \leqslant 2n \\ 0 \leqslant \mu \leqslant 1}} |A_{ij}(\mu)|. \tag{11}$$

We prove the following theorem:

Theorem 1: $K_n(\mu,\omega)$ converges to $H(\mu,\omega)$ uniformly in μ and ω for $0 \leqslant \mu \leqslant 1$ and $|\omega| \leqslant 1$. Moreover, for all c, $0 \leqslant c \leqslant 1$, and $n \geqslant 0$,

$$\max_{|\omega| \leqslant c} \|H(\omega) - K_n(\omega)\| \leqslant \max_{|\omega| \leqslant c} \|H(\omega) - H_n(\omega)\| \tag{12}$$

As will become clear, the proof of Theorem 1 generalizes directly to the more general case of operator-valued H-functions as described in Ref. 3–5 and 13.

II. PROOF OF THE THEOREM

For $|\omega| \leqslant 1$ we may write

$$H(\mu,\omega) = \sum_{m=0}^{\infty} \omega^m P_m(\mu), \tag{13}$$

 0022-2488/80/020408-02$1.00

$$H_n(\mu,\omega) = \sum_{m=0}^{\infty} \omega^m Q_{m,n}(\mu). \tag{14}$$

In Eq. (14), $Q_{m,n} = 0$ if $m > 2^n$, and $Q_{m,n} = P_m$ if $0 \leqslant m \leqslant n$. For matrices A and B, we say $A \geqslant B$ if $A_{ij} \geqslant B_{ij}$ for all i and j. We have, as in Ref. 5, that $0 \leqslant Q_{m,n} \leqslant Q_{m,n+1} \leqslant P_m$ for all m and n.

Now for $|\omega|$ sufficiently small, we may write

$$K_n(\mu,\omega) = \sum_{m=0}^{\infty} \omega^m R_{m,n}(\omega). \tag{15}$$

We require the following lemma:

Lemma 1: For all $m,n \geqslant 0$ we have

$$0 \leqslant Q_{m,n} \leqslant R_{m,n} \leqslant P_m. \tag{16}$$

Assuming the lemma for the present, we may now prove Theorem 1. Note that an immediate consequence of Lemma 1 is that K_n exists for $|\omega| \leqslant 1$.

Lemma 1 implies that the power series cofficients of H-K_n and H-H_n are matrix-valued functions of μ with nonnegative entries. Hence, for $0 \leqslant c \leqslant 1$

$$\max_{|\omega| \leqslant c} \|H(\omega) - H_n(\omega)\| = \|H(c) - H_n(c)\| \tag{17}$$

and

$$\max_{|\omega| \leqslant c} \|H(\omega) - K_n(\omega)\| = \|H(c) - K_n(c)\|. \tag{18}$$

As $Q_{m,n} \leqslant R_{m,n}$ we must have, for $0 \leqslant c \leqslant 1$,

$$0 \leqslant H(c) - K_n(c) \leqslant H(c) - H_n(c). \tag{19}$$

Hence $\|H(c) - K_n(c)\| \leqslant \|H(c) - H_n(c)\|$. This fact, together with Eqs. (17) and (18) implies inequality (12). As $H_n(\mu,\omega)$ converges [6] to $H(\mu,\omega)$ uniformly in μ and ω for $0 \leqslant \mu \leqslant 1$ and $|\omega| \leqslant 1$, the proof of Theorem 1 is complete.

It remains only to prove Lemma 1. Using Eqs. (6), (8), and (10), we derive the relations

$$P_m = \sum_{k+l=m-1} P_k L P^*_l, \quad m \geqslant 1, \tag{20}$$

$$Q_{m,n} = \sum_{k+l=m-1} Q_{k,n-1} L Q^*_{l,n-1}, \quad m \geqslant 1, \tag{21}$$

$$R_{m,n} = \sum_{k+l=m-1} R_{k,n} L R^*_{l,n-1}, \quad m \geqslant 1. \tag{22}$$

We proceed by induction on m and n. Note that, for all n, $P_0 = Q_{0,n} = R_{0,n} = E$. Also, for all m, $P_m \geqslant Q_{m,0} = R_{m,0} \geqslant 0$. Hence inequality (16) holds for all m if $n = 0$ and for $m = 0$ for any value of n. Now assume that inequality (16) holds for all m if $n \leqslant N$ and for $m \leqslant M$ if $n = N+1$. We will be done if we show that this implies that inequality (16) holds for $n = N+1$ and $m = M+1$.

Equation (21), the induction hypothesis, and the fact that, for all m and n, $Q_{m,n} \leqslant Q_{m,n+1}$ imply that

$$\begin{aligned} 0 \leqslant Q_{M+1,N+1} &= \sum_{k+l=M} Q_{k,N} L Q^*_{l,N} \\ &\leqslant \sum_{k+l=M} Q_{k,N+1} L Q^*_{l,N} \\ &\leqslant \sum_{k+l=M} R_{k,N+1} L R^*_{l,N}. \end{aligned} \tag{23}$$

From Eqs. (20), (22) and the induction hypothesis we obtain

$$\begin{aligned} R_{M+1,N+1} &= \sum_{k+l=M} R_{k,N+1} L R^*_{l,N} \\ &\leqslant \sum_{k+l=M} P_k L P^*_l = P_{M+1}. \end{aligned} \tag{24}$$

Inequalities (23) and (24) together imply $Q_{M+1,N+1} \leqslant R_{M+1,N+1} \leqslant P_{M+1}$. This completes the proof.

[1] S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).
[2] T.W. Mullikin and H.D. Victory, J. Math. Anal. Appl. **58**, 605 (1977).
[3] C.T. Kelley, J. Math. Phys. **18**, 764 (1977).
[4] C.T. Kelley, J. Math. Phys. **19**, 494 (1978).
[5] C.T. Kelley, "Operator-valued Chandrasekhar H-functions," (to appear in J. Math. Anal. Appl).
[6] C.T. Kelley and T.W. Mullikin, J. Math. Phys. **19**, 500 (1978).
[7] C.T. Kelley, "Solution of H-equations by Iteration" (to appear in SIAM J. Math. Anal.).
[8] R.L. Bowden and P.F. Zweifel, Astrophys. J. **210**, 178 (1976).
[9] R.L. Bowden, R. Menikoff, and P.F. Zweifel, J. Math. Phys. **17**, 1722 (1976).
[10] L.B. Rall, Rend. Circ. Math. Palermo **10**, 314 (1961).
[11] T.W. Mullikin, "Wiener-Hopf Factorizations: Computational Methods for a Class of Matrix Functions" (to appear).
[12] R.L. Bowden, private communication.
[13] T.W. Mullikin, Mathematics of Radiative Transfer, in *Theoretical Principles in Astrophysics and Relativity*, edited by N.R. Lebovitz, W.M. Ried, and P.O. Vandervoort (Univ. of Chicago Press, Chicago, 1978), Chap. 6.