Jauch–Piron states

Gottfried T. Rüttimann

Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01002
(Received 9 April 1976)

We argue that a finite proposition system in the sense of Jauch and Piron that admits a unital set of states is necessarily purely classical. Based on this result, we investigate the extensibility of σ-additive states on projection lattices to all projections of a separable Hilbert space.

1. INTRODUCTION AND COMMENTS

The starting point of most quantum logic approaches to quantum mechanics is the assumption that the proposition system L (set of yes–no experiments) has the structure of an orthomodular poset. The structural terms such as order and orthogonality are interpreted in a more or less operational manner. One distinguishes the cases where the center of L is equal to $\{0, 1\}$, not equal to L, or equal to L and refers to L as a purely quantal, quantal, or purely classical proposition system. The next concept introduced is that of a state of a physical system which mathematically turns out to be a probability measure on the orthomodular poset of propositions.

One of the purposes of this note is to examine the following condition imposed by Jauch[1] and Piron[2] on the states of a proposition system (in their case, at least an orthomodular lattice):

$$\text{If} \quad \omega(x) = \omega(y) = 1, \quad \text{then} \quad \omega(x \wedge y) = 1.$$

This condition which relates the quantal to the classical "and" has been advocated by these authors in numerous papers.

In Sec. 4 of this paper we show that a finite proposition system in the sense of Jauch and Piron that admits a reasonable set of states (unital set of states) is purely classical (i.e., a Boolean lattice). Likewise, a quantal proposition system in the sense of Jauch and Piron with a unital set of states must be necessarily infinite.

Theorem 4.3 which establishes this result is based upon a "Hilfssatz" (Theorem 3.5) whose proof is in Sec. 3. Both theorems will be of importance in the theory of strong polytopes.[3] This (mathematical) theory is partially designed to make the connection between the empirical logic approach[4,5] and the convex set approach[6,7] to quantum mechanics in the case of a finite system, i.e., finitely many yes–no experiments.

In Sec. 5 we investigate the extensibility of σ-additive states from projection lattices to all projections of a separable complex Hilbert space. Based on the results in Sec. 4 we then show (Theorem 5.3) that a finite quantal proposition system which can be "realized" in Hilbert space always has at least one state that is not induced by a "Hilbert space state." In a sense, Hilbert space as the state space of a finite quantal proposition system is too small.

2. PRELIMINARIES

Let $\{L, \leq\}$ be a poset with a least element (0) and a greatest element (1). We say *y covers x*, in symbols $x \prec y$, if $x < z \leq y$ implies that $z = y$, $x,y,z \in L$. An element $x \in L$ is called an *atom* if $0 \prec x$. L is called *atomic* if for every $x \in L - \{0\}$ there exists an atom $y \in L$ such that $y \leq x$. L is called *atomistic* if $x = \vee\{y \in L \mid y \leq x,\ y \text{ atom}\}$ for all $x \in L - \{0\}$.

An *orthocomplementation* on L is a mapping $x \to x'$ on L such that (i) $x'' = x$, (ii) $x \vee x'$ exists and is equal to 1, and (iii) if $x \leq y$, then $y' \leq x'$. A poset admitting an orthocomplementation is called *orthocomplemented.* Note that if $x \vee y$ exists, then $x' \wedge y'$ exists and $x' \wedge y' = (x \vee y)'$. A pair $x, y \in L$ is said to be *orthogonal*, denoted $x \perp y$, if $x \leq y'$. An orthocomplemented poset $\{L, \leq, '\}$ is called an *orthomodular poset* if (i) $x \perp y$ implies that $x \vee y$ exists, and (ii) $x \leq y$ implies that there exists $z \in L$ such that $x \perp z$ and $x \vee z = y$. One can actually show that $z = x' \wedge y$. An orthomodular poset that is indeed a lattice is called an *orthomodular lattice.*[8] It is a well-known fact that an atomic orthomodular lattice is atomistic.[9]

Let $\{L, \leq, '\}$ be an orthomodular poset. A mapping $\nu : L \to R$ (R the real numbers) satisfying (i) $\nu(0) = 0$, (ii) if $x \perp y$, then $\nu(x \vee y) = \nu(x) + \nu(y)$ is called a *signed state.* If we define $(t_1\nu_1 + t_2\nu_2)\ (x) = t_1\nu_1(x) + t_2\nu_2(x)$, $t_1, t_2 \in R$, ν_1, ν_2 signed states, $x \in L$, then the set of signed states W becomes a real vector space. A signed state ω for which $\omega(1) = 1$ and $\omega(L) \subseteq [0, 1] \subseteq R$ is called a *state.* Clearly, the set of states Ω is a convex subset of W. Ω is said to be *strong* for L if $\{\omega \in \Omega \mid \omega(x) = 1\} \subseteq \{\omega \in \Omega \mid \omega(y) = 1\}$ implies that $x \leq y$. Ω is called *unital* for L if for every $x \in L - \{0\}$ there exists a state $\omega \in \Omega$ such that $\omega(x) = 1$. The implication "strong $\Rightarrow$ unital" holds true.

Let $\{L, \leq, '\}$ be an orthomodular lattice. A state ω is said to be *Jauch–Piron* if $\omega(x) = \omega(y) = 1$ implies that $\omega(x \wedge y) = 1$. The set Ω is said to have the *Jauch–Piron property* if every element $\omega \in \Omega$ is a Jauch–Piron state.

Theorem 2.1: Let $\{L \leq, '\}$ be an atomic orthomodular lattice and assume that Ω has the Jauch–Piron property. Then Ω is strong if and only if it is unital for L.

Proof: Assume that $\{\omega \in \Omega \mid \omega(x) = 1\} \subseteq \{\omega \in \Omega \mid \omega(y) = 1\}$ for $x, y \in L$. If $x = 0$ then clearly $x \leq y$, so assume that $x \neq 0$. Let z be an atom such that $z \leq x$. Since Ω is unital there exists a state ω such that $\omega(z) = 1$, thus $\omega(x) = 1$ and therefore $\omega(y) = 1$. The state ω being Jauch–Piron, we get $\omega(z \wedge y) = 1$. This implies that $z \wedge y \neq 0$. Now, z is an atom and $0 \neq z \wedge y \leq z$, hence $z \wedge y = z$ or equivalently $z \leq y$. This is true for all atoms z with $z \leq x$. L is atomistic, thus $x = \vee \{z \in L \mid z \leq x, z \text{ an atom}\} \leq y$. The converse is obvious.

3. A "HILFSSATZ"

We list several definitions and facts concerning polytopes[10] that will be used in the sequel.

Let W be an Euclidean space and P be a convex subset of W. An element $\omega \in P$ is called an *extreme point* of P provided $\omega = t\omega_1 + (1 - t)\omega_2$, $\omega_1, \omega_2 \in P$, $t \in (0, 1)$ implies that $\omega = \omega_1 = \omega_2$. The set of extreme points of P is denoted by ext P. A convex, compact subset P of W is called a *polytope* if ext P is finite. A polytope may be equivalently defined as the convex hull of a finite subset or as a bounded set which is the intersection of finitely many closed half- spaces. By the theorem of Minkowski–Carathéodory, $P = \text{con ext } P$. Note that $\dim P + 1 \leq \#\text{ext } P$ ($\dim P$ = affine dimension of the affine span of P in W; we put $\dim \emptyset = -1$). If $\dim P + 1 = \#\text{ext } P$, then P is said to be a *simplex*. Clearly, if P is a simplex then ext P is an affinely independent set in W; if $V \subseteq W$ is a finite affinely independent subset then con V is a simplex with $\dim \text{con } V = \# V - 1$.

Let P be a polytope. A subset $a \subseteq P$ is called a *face* of P provided that for $\omega_1, \omega_2 \in P$, $t \in (0, 1)$, $t\omega_1 + (1 - t)\omega_2 \in a \Leftrightarrow \omega_1, \omega_2 \in a$. A face is a polytope in its own right. If b is a face of a which is a face of P, then b is a face of P; if b is a face of P and $b \subseteq a$, then b is a face of a. Note that $a = \text{aff } a \cap P$. The set-intersection of a family of faces of P is again a face of P. Therefore, the set of faces of P ordered by set inclusion, denoted by $\{F(P), \geq\}$, is a lattice with 0 ($= \emptyset$) as the least, and 1 ($= P$) as the largest element (as usual $\wedge$ denotes infimum, $\vee$ denotes supremum). Note that $F(P)$ is a finite set. An element $a \in F(P)$ covered by 1 is called a *facet* of P; an element $a \in F(P)$ that covers 0 is called a *vertex* of P. A face $a \neq 1$ is equal to the infimum of all facets containing it. The set of vertices of a face a is denoted by $V(a)$. Note that $V(a) = V(1) \cap a$ and that $V(1) = \{\{\omega\} \mid \omega \in \text{ext } P\}$. If $0 \prec b_0 \prec b_1 \prec \cdots \prec b_r \prec 1$, $b_i \in F(P)$ ($i = 0, 1, 2, \cdots, r$), then $\dim b_i = i$.

Let $\omega \in P$, the face $a(\omega) = \wedge \{b \in F(P) \mid \omega \in b\}$ is called face *generated* by ω. Clearly, $a \vee b = a(a \cup b)$. The *interior* of P is defined to be the set $P^I = \{\omega \in P \mid a(\omega) = 1\}$. Note that $P^I = P^{0_r}$ (= topological interior of P in the Euclidean topology relativized to aff P). If $P \neq \emptyset$ then $P^I \neq \emptyset$.

If P is a simplex, then $S \subseteq \text{ext } P$ implies that con $S \in F(P)$. Using this fact one easily shows that $\{F(P), \leq\}$ is a distributive lattice. We get immediately:

Lemma 3.1: Let P be a simplex and assume that $V' \subseteq V(1)$. If $\vee \{v \mid v \in V'\} = 1$, then $V' = V(1)$.

Lemma 3.2: Let P and Q be polytopes. If (i) $P \subseteq Q$, (ii) $\dim P = \dim Q$, and (iii) $F(P) - \{1\} \subseteq F(Q)$, then $P = Q$.

Proof: Note that aff P = aff Q, since $\dim Q = \dim P$ and $P \subseteq Q$. Therefore, given $\nu \in Q$, there exists $\omega_1, \omega_2 \in P$ and $t \geq 0$ such that $\nu = t\omega_1 + (1 - t)\omega_2$.

Due to compactness of P, $\sup\{s \in R \mid s\omega_1 + (1 - s)\omega_2 \in P\}$ is finite and attained, denoted by s_0. Clearly $s_0 \geq 1$. Since $s\omega_1 + (1 - s)\omega_2 \notin P$ for all $s \geq s_0$, we have $\omega_0 = s_0\omega_1 + (1 - s_0)\omega_2 \in P - P^{0_r}$. Thus $a(\omega_0) \in F(P) - \{1\}$.

Now, if $0 \leq t \leq s_0$, then clearly $\nu \in P$. If $t > s_0$, then $\omega_0 = s_0(\nu/t - (1 - t)\omega_2/t) + (1 - s_0)\omega_2 = s_0 \nu/t + (1 - s_0/t)\omega_2$ with $s_0/t \in (0, 1)$. Since $a(\omega_0) \in F(Q)$, $\nu, \omega_2 \in Q$, we get $\nu, \omega_2 \in a(\omega_0)$. Hence $\nu \in P$.

Lemma 3.3: Let P be a polytope and assume that its face-lattice admits an orthocomplementation $a \to a'$ such that $\{F(P), \leq, '\}$ is an orthomodular lattice. Then for any face a, $\dim a + 1$ = number of elements in a maximal orthogonal set of elements of $V(a)$.

Proof: We can assume that $a \neq 0$. Let $\{v_1, v_2, \cdots, v_k\}$ be a maximal orthogonal set in $V(a)$. We extend it to a maximal orthogonal set in $V(1)$, say $\{v_1, v_2, \cdots, v_k, v_{k+1}, \cdots, v_m\}$. Denote $b_j = \vee_{i=1}^{j+1} v_i$, for $0 \leq j \leq m - 1$.

Suppose that $b_j < c \leq b_j \vee v_{j+2} = b_{j+1}$. By orthomodularity, there exists $d \in F(P)$ such that $d \neq 0$, $d \perp b_j$ and $c = b_j \vee d$, but $v_{j+2} \leq b_j'$, $d \leq b_j'$, hence $v_{j+2} = b_j' \wedge (b_j \vee v_{j+2})$ and $d = b_j' \wedge (b_j \vee d)$. Having $b_j \vee d = c \leq b_j \vee v_{j+2}$ we get $0 \neq d \leq v_{j+2}$; v_{j+2} being an atom, we conclude that $d = v_{j+2}$. Thus $c = b_{j+1}$. By orthomodularity and maximality of the set in question, we also get $b_{m-1} = 1$ and $b_{k-1} = a$. Therefore, $0 \prec b_0 \prec b_1 \prec \ldots \prec b_{m-1} = 1$, thus $\dim a = \dim b_{k-1} = k - 1$.

Lemma 3.4: Same assumption as in Lemma 3.3. Let $a \in F(P)$. If every face b with $b < a$ is a simplex, then a is a simplex.

Proof: We can assume that $\dim a > 0$, so $\# V(a) \geq 2$. Let $\{v_1, v_2, \cdots, v_m, v_{m+1}\}$ be a maximal orthogonal set in $V(a)$. By Lemma 3.3, $\dim a = m$. We introduce the following notation: $F = \{1, 2, 3, \cdots, m + 1\}$, $S(I) = \text{con}\{\omega_i \mid i \in I\}$ and $b(I) = \vee_{i \in I} v_i$, where $I \subseteq F$ and $\{\omega_i\} = v_i$ ($i \in F$).

If $I \subset F$, we get by orthomodularity (see proof of the foregoing lemma), $b(I) < a$. Thus for all $I \subset F$, $b(I)$ is a simplex.

Assume that $I \subset F$. Since $\{v_i | i \in I\} \subseteq V\ (b(I))$ and $[0, b(I)]$ is a sublattice of $F(P)$, thus $\vee^{F}_{i\in I} v_i = \vee_{i\in I} v_i = b(I)$, we get by Lemma 3.1, $V(b(I)) = \{v_i | i \in I\}$. Therefore, $b(I) = S(I)$ for all $I \subset F$.

Clearly $S(F) \subseteq a$ is a polytope and every face of $S(F)$ is of the form $S(I)$ for some $I \subseteq F$. Thus every face of $S(F)$ different from $S(F)$ is a face of a.

In view of Lemma 3.2 we are going to show that $S(F)$ is an m-dimensional simplex; to do so, it is enough to prove that $\{\omega_1, \omega_2, \cdots, \omega_m, \omega_{m+1}\}$ is affinely independent. Clearly, $\{\omega_1, \omega_2, \cdots, \omega_m\}$ is affinely independent since $S(I)$ ($I = \{1, 2, \cdots, m\}$) is a simplex and $V(S(I)) = \{\omega_i | i \in I\}$. If $\omega_{m+1} \in \operatorname{aff}\{\omega_1, \cdots, \omega_m\}$, then $\operatorname{aff} S(I) \cap P = b(I)$. Hence $v_{m+1} \leq \vee_{i\in I} v_i$ or $b(I) = \vee_{i\in F} v_i = a$, which is a contradiction.

Therefore, $S(F)$ is a simplex and $\dim S(F) = m = \dim a$. Using Lemma 3.2 we conclude that $S(F) = a$, thus a is a simplex.

Theorem 3.5: Let P be a polytope. If its face lattice admits an orthocomplementation $a \to a'$ such that $\{F(P), \leq, '\}$ is an orthomodular lattice, then P is a simplex.

Proof (by induction): All faces of dimension less than or equal to zero are simplices. Assume now, that all faces $a \in F(P)$ with $\dim a \leq m - 1$ are simplices. Let $b \in F(P)$ with $\dim b = m$. If $a < b$, then $\dim a \leq m - 1$. By Lemma 3.4, b is a simplex. Thus all faces with dimension less than or equal to m are simplices for $-1 < m \leq \dim P$. Therefore, P is a simplex.

4. THE JAUCH-PIRON PROPERTY

Recall that the set of signed states W on an orthomodular poset $\{L, \leq, '\}$ is a real vector space containing the set of states Ω as a convex subset.

With every $x \in L$ we associate a linear functional on W as follows:

$$f_x(\nu) = \nu(x), \nu \in W.$$

Note that $f_{x'} = f_1 - f_x$ and that the set of linear functionals obtained in this fashion is total, i.e., $f_x(\nu) = 0$ for all $x \in L$ implies that $\nu = 0$. If $\{L, \leq, '\}$ is a finite orthomodular poset, then $\{f_x | x \in L\}$ is a finite total set, hence the dual W^*, and finally W is finite dimensional.

Lemma 4.1: Let $\{L, \leq, '\}$ be a finite orthomodular poset. Then Ω is a polytope.

Proof: One verifies immediately that $\Omega = \cap_{x\in L} f_x^{-1} [0, 1] \cap f_1^{-1} [1, \infty)$. Thus Ω is the intersection of finitely many closed half-spaces.

Recall that W is finite dimensional. Since $\{f_x | x \in L\}$ is a total set, a local base for the unique Hausdorff topology on W (e.g., Euclidean topology) is given by the sets $\{N(\varepsilon, x) | x \in L, \varepsilon > 0\}$, where $N(\varepsilon, x) = \{\nu \in W | \, |f_x(\nu)| < \varepsilon\}$ together with their finite intersections.

Now, $\Omega \subseteq 2/\varepsilon\, N(\varepsilon, x)$ for all $x \in L$ and $\varepsilon > 0$. This shows that Ω is bounded. Hence Ω is a polytope.

Let $\{L, \leq, '\}$ be an orthomodular poset. Denote $M = \{x \in L |$ exists $\omega \in \Omega$ such that $\omega(x) \neq 0, 1\}$. Clearly, if $x \in M$, then $x' \in M$. If $x \notin M$, then either $\omega(x) = 0$ for all $\omega \in \Omega$ or $\omega(x) = 1$ for all $\omega \in \Omega$. For if there exist $\omega_1, \omega_2 \in \Omega$ such that $\omega_1(x) \neq 0$ and $\omega_2(x) \neq 1$, then $\omega_1(x) = 1$ and $\omega_2(x) = 0$ since $x \notin M$. Now $(\omega_1/2 + \omega_2/2)(x) = \omega_1(x)/2 + \omega_2(x)/2 = 1/2$. But $\omega_1/2 + \omega_2/2 \in \Omega$, thus $(\omega_1/2 + \omega_2/2)(x) = 0$ or 1, which is a contradiction.

Given $x \in L$, we define $\sigma(x) = \{\omega \in \Omega | \omega(x) = 1\}$. If $t\omega_1 + (1 - t)\omega_2 \in \sigma(x)$, $t \in (0, 1)$, $\omega_1, \omega_2 \in \Omega$, then $(t\omega_1 + (1 - t)\omega_2)(x) = t\omega_1(x) + (1 - t)\omega_2(x) = 1$. Since $\omega_1(x)$, $\omega_2(x) \in [0, 1]$ we conclude that $\omega_1(x) = \omega_2(x) = 1$. Hence $\omega_1, \omega_2 \in \sigma(x)$. The set $\sigma(x) \subseteq \Omega$ is clearly convex, hence $\sigma(x)$ is a face of Ω. Furthermore, note that $\sigma(0) = 0$, $\sigma(1) = 1$ $(= \Omega)$, and $\sigma(x) = f_x^{-1}(1) \cap \Omega$.

Theorem 4.2: Let $\{L, \leq, '\}$ be a finite orthomodular poset. Then for every facet a of Ω there exists an element $x \in L$ such that $\sigma(x) = a$.

Proof: (i) Assume that $M = \varnothing$. Furthermore, assume that $\Omega \neq \varnothing$. If there exists $\omega_1, \omega_2 \in \Omega$ such that $\omega_1 \neq \omega_2$, then $(t\omega_1 + (1 - t)\omega_2)(x) = 1$ or 0 for all $x \in L$ and for all $t \in R$ (see the remark made above). We have $(t\omega_1 + (1 - t)\omega_2)(1) = 1$. Therefore, $t\omega_1 + (1 - t)\omega_2 \in \Omega$ for all $t \in R$, which is a contradiction since Ω is bounded (Lemma 4.1). Hence $\Omega = \{\omega\}$. This shows that if $M = \varnothing$ then $\dim \Omega \leq 0$. Clearly, for these cases the assertion holds true.

(ii) Assume that $M \neq \varnothing$. First we show that $\Omega = \cap_{x\in M} [f_x^{-1}(-\infty, 1] \cap \operatorname{aff}\Omega]$. Let $\nu \in W$ and assume that $\nu(x) \leq 1$, for all $x \in M$, and that $\nu = t\omega_1 + (1 - t)\,\omega_2$ for some $\omega_1, \omega_2 \in \Omega$, $t \in R$. Since $\nu(1) = 1$ and $\nu(x') = \nu(1) - \nu(x) \leq 1$, we get $\nu(x) \geq 0$ for all $x \in M$. If $x \notin M$ then $\nu(x) = t\omega_1(x) + (1 - t)\omega_2(x) = 0$ or 1. Therefore, ν is a state on L. The converse is obvious.

Now, if $a = \varnothing$, we are done. If $a \neq \varnothing$ then $a^I \neq \varnothing$. Select $\omega \in a^I$. We claim that there exists $x \in M$ such that $f_x(\omega) = 1$. If it is not so, then $\omega \in \cap_{x\in M} [f_x^{-1}(-\infty, 1) \cap \operatorname{aff}\Omega] \subseteq \Omega$. Thus $\omega \in \Omega^{0_r}$ since ω is contained in the intersection of finitely many open sets in the Euclidean topology relativized to $\operatorname{aff}\Omega$ and this intersection is contained in Ω. Since $\Omega^{0_r} = \Omega^I$ we get $1 = a(\omega) \leq a$ which contradicts the assumption that a is a facet.

Now let x be an element of M such that $f_x(\omega) = 1$. Since $\omega \in f_x^{-1}(1) \cap \Omega = \sigma(x)$, we conclude that $a = a(\omega) \leq \sigma(x)$. But $\sigma(x) \neq 1$, or else $x \notin M$. Since a is a facet, we get $a = \sigma(x)$.

Theorem 4.3: Let $\{L, \leq, '\}$ be a finite orthomodular lattice. Its set of states Ω is unital and has the Jauch–Piron property if and only if $\{L, \leq\}$ is a Boolean lattice.

Proof: A finite orthomodular lattice is both atomic and atomistic. Hence, if Ω is unital then, by Theorem 2.1, Ω is also strong. Therefore, the mapping $x \in L \to \sigma(x) \in F(\Omega)$ is an order isomorphism, i.e., $x \leq y \Leftrightarrow \sigma(x) \leq \sigma(y)$.

Note that in this case the above mapping is one-to-one. Using again the Jauch–Piron property of Ω, we get $\sigma(\wedge_{i\in I}x_i) = \wedge_{i\in I}\sigma(x_i)$, the infimum of faces being equal to their set intersection.

Let $a \in F(\Omega) - \{1\}$ and let $\{b_1, b_2, \cdots, b_m\}$ be the set of all facets each of which contains a. Then $a = \wedge_{i=1}^m b_i$. By Theorem 4.2, to every $1 \le i \le m$, there exists an element $x_i \in L$ such that $b_i = \sigma(x_i)$. Now $\sigma(\wedge_{i=1}^m x_i) = \wedge_{i=1}^m \sigma(x_i) = \wedge_{i=1}^m b_i = a$. Therefore, the maping $x \to \sigma(x)$ is an onto order isomorphism.

Given $a \in F(\Omega)$ define $a' = \sigma(\sigma^{-1}(a))'$. One immediately verifies that $a \to a'$ is an orthocomplementation on $\{F(\Omega), \le\}$ that makes $\{F(\Omega), \le, '\}$ into an orthomodular lattice ortho-order isomorphic to $\{L, \le, '\}$. Using Theorem 3.5, we conclude that Ω is a simplex. The face-lattice of a simplex being distributive, we conclude that $\{L, \le, '\}$ is a distributive, complemented lattice, hence a Boolean lattice.

The converse of the theorem is easily shown.

5. PROJECTION LATTICES WITH THE EXTENSION PROPERTY

Let H be a separable complex Hilbert space of dimension greater than or equal to three. $B(H)$ denotes the set of bounded linear operators on H and $P(H)$ denotes the set of orthogonal projections on H. We define a partial ordering in $P(H)$ as follows: $E \le F \Leftrightarrow EF = E$ ($\Leftrightarrow$ $(\psi, E\psi) \le (\psi, F\psi)$ for all $\psi \in H$). Then the poset $\{P(H), \le\}$ is a complete lattice with the identity operator as the greatest element and the zero operator as the least element. The mapping $E \to E^{\perp} = 1 - E$ is an orthocomplementation that makes $\{P(H), <, ^{\perp}\}$ into a complete orthomodular lattice.

A subcomplete sublattice $\{L, \le\}$ of $\{P(H), \le\}$ that contains 1 and is closed under the mapping $E \to E^{\perp}$ is called a *projection lattice*. Clearly, $\{L, \le, ^{\perp}\}$ is a complete orthomodular lattice. Note that the set of projections of a von Neumann algebra equipped with the induced order is a projection lattice.[11]

We are going to consider σ-additive states on projection lattices. Generally speaking, a state ω on a orthomodular poset L is said to be *σ-additive* provided that for every countable set $\{x_i\}_{i=1}^{\infty} \subseteq L$ of pairwise orthogonal elements for which $\vee_{i=1}^{\infty} x_i$ exists, $\omega(\vee_{i=1}^{\infty} x_i) = \sum_{i=1}^{\infty}\omega(x_i)$ holds true. It is obvious that, by restriction, every σ-additive state on $P(H)$ induces a σ-additive state on every projection lattice. A projection lattice $\{L, \le\}$ is said to have the *extension property* provided every σ-additive state on L can be extended to a σ-additive state on $P(H)$. The issue of this section is to give a characterization of those finite projection lattices that have the extension property.

We need the following lemmas:

Lemma 5.1: Let D be a von Neumann density operator and E be a projection on H. Then $\mathrm{tr}(DE) = 1$ if and only if $DE = D$.

Proof: Assume that $\mathrm{tr}(DE) = 1$. Then $\mathrm{tr}(DE^{\perp}) = \mathrm{tr}(D(1 - E)) = 0$. Since $0 \le D = D^*$, we get $0 = \mathrm{tr}(DE^{\perp}) = \mathrm{tr}(E^{\perp} DE^{\perp}) = \mathrm{tr}(E^{\perp}\sqrt{D}\sqrt{D}E^{\perp}) = \mathrm{tr}((\sqrt{D}E^{\perp})^* (\sqrt{D}E^{\perp}))$. Having $0 \le (\sqrt{D}E^{\perp})^* (\sqrt{D}E^{\perp})$ we conclude that $(\sqrt{D}E^{\perp})^* (\sqrt{D}E^{\perp}) = 0$, thus $\sqrt{D}E^{\perp} = 0$. Hence, $0 = \sqrt{D}\sqrt{D}E^{\perp} = DE^{\perp} = D - DE$. Therefore, $D = DE$. Conversely, if $DE = D$, then $\mathrm{tr}(DE) = \mathrm{tr}(D) = 1$.

Lemma 5.2: Let $F, E \in P(H)$ and D be a von Neumann density operator. Then $\mathrm{tr}(DE) = \mathrm{tr}(DF) = 1$ implies that $\mathrm{tr}(D(E \wedge F)) = 1$.

Proof: $\mathrm{tr}(DE) = \mathrm{tr}(DF) = 1$ implies that $DE = DF = D$. Multiplication is continuous in the strong operator topology, thus $D(E \wedge F) = D\ \text{s-lim}_{n\to\infty} (EF)^n = \text{s-lim}_{n\to\infty} D(EF)^n = D$. Thus, $\mathrm{tr}(D(E \wedge F)) = 1$.

Theorem 5.3: Let $\{L, \le\}$ be a finite projection lattice. The following statements are mutually equivalent:

(i) L has the extension property;

(ii) $\{L, \le\}$ is a Boolean lattice;

(iii) the elements of L commute pairwise.

Proof: (ii) $\Leftrightarrow$ (iii): A well-known fact.

(i) $\Rightarrow$ (ii): Assume that for every state $\omega \in \Omega(L)$ there exists a σ-additive extension $\omega^{\sim}$ to $P(H)$. By Gleason's theorem[12] there exists a von Neumann density operator D such that $\omega^{\sim}(E) = \mathrm{tr}(DE)$ for all $E \in P(H)$. Now if $\omega(E) = \omega(F)$ for $E, F \in L$, then $\mathrm{tr}(DE) = \mathrm{tr}(DF) = 1$. Thus, $\omega(E \wedge F) = \omega^{\sim}(E \wedge F) = \mathrm{tr}(D(E \wedge F)) = 1$ by Lemma 5.2. Thus $\Omega(L)$ has the Jauch–Piron property.

Given $E \in L - \{0\}$, select $\phi \in E(H)$ with $\|\phi\| = 1$. Then the linear operator D_ϕ, defined by $D_\phi\psi = (\phi, \psi)\,\phi$, is a one-dimensional projection, hence a von Neumann density operator having the property that $D_\phi E = D_\phi$. Then the mapping $F \in L \to \mathrm{tr}(D_\phi F)$ is a state on L and $\mathrm{tr}(D_\phi E) = \mathrm{tr}(D_\phi) = 1$. Thus $\Omega(L)$ is unital for L. Therefore, by Theorem 4.3, $\{L, \le\}$ is a Boolean lattice.

(ii) $\Rightarrow$ (i): Let $\{L, \le\}$ be a finite projection lattice and assume that it is Boolean. Let $\{E_1, E_2, \cdots, E_m\}$ be the set of atoms in L. We have $E_i \perp E_j$ for $i \ne j$ since L is Boolean and also $\vee_{i=1}^m E_i = \sum_{i=1}^m E_i = 1$.

Now, let ω be a state on L. For every $1 \le i \le m$ select a normed vector $\phi_i \in E_i(H)$ and define $\psi = \sum_{i=1}^m \sqrt{\omega(E_i)}\,\phi_i$. Then $\|\psi\| = 1$ since $\sum_{i=1}^m E_i = 1$ and D_ψ becomes a von Neumann density operator. The mapping $F \to \mathrm{tr}(D_\psi F)$ is a σ-additive state on $P(H)$. One easily verifies that $\mathrm{tr}(D_\psi E_i) = \omega(E_i)$.

Since every nonzero element in $\{L, \le\}$ is the sum of a subset of the atoms $\{E_1, E_2, \cdots, E_m\}$ and the trace-functional is linear, we get the assertion.

One final observation:

Theorem 5.4: The projection lattice of a von Neumann algebra, acting on a separable complex Hilbert space H, and not containing a type I_2 factor as direct summand has the extension property.

Proof: Let N be a von Neumann algebra not containing a type I_2 factor. We denote by $\{L(N), \leq\}$ the projection lattice of N. Let $A \in N$ and ω be a σ-additive state on $L(N)$. Let $\{E_\lambda\}$, resp. $\{F_\lambda\}$, be the spectral family of $(A + A^*)/2$, resp. $(A - A^*)/2i$. We have $\{E_\lambda\}$, $\{F_\lambda\} \subseteq L(N)$.

Lodkin has shown[13] ("generalized Gleason theorem") that the positive functional

$$f(A) = \int \lambda \, d\omega(E_\lambda) + i \int \lambda \, d\omega(F_\lambda),$$

is linear in N. Clearly, $f|L(N) = \omega$. Recall that if H is separable and g is a positive linear functional on a von Neumann algebra N, then $g|L(N)$ is a σ-additive state on $L(N)$ if and only if g is ultraweakly continuous on N and $g(1) = 1$.[14] Furthermore, to every positive, ultraweakly continuous linear functional g on N there exists a positive, ultraweakly continuous linear functional $g^\sim$ on $B(H)$ such that $g^\sim|N = g$ and $g^\sim(1) = g(1)$.[15]

Thus, f is a positive, ultraweakly continuous linear functional on N. Let $f^\sim$ be a positive, ultraweakly continuous extension to $B(H)$. Then $f^\sim|P(H)$ is a σ-additive state on $P(H)$ and $f^\sim|L(N) = f|L(N) = \omega$.

[1]J.M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[2]C. Piron, Helv. Phys. Acta **37**, 439 (1964).
[3]G.T. Rüttimann, Bull. Am. Math. Soc. **82**, 314 (1976).
[4]D.J. Foulis and C.H. Randall, J. Math. Phys. **13**, 1667 (1972).
[5]C.H. Randall and D.J. Foulis, J. Math. Phys. **14**, 1472 (1973).
[6]C.M. Edwards, Commun. Math. Phys. **16**, 207 (1970).
[7]B. Mielnik, Commun. Math. Phys. **9**, 55 (1968).
[8]D.J. Foulis, Port. Math. **21**, 65 (1962).
[9]F. Maeda and S. Maeda, *Theory of Symmetric Lattices* (Springer, New York, 1970), p. 30.
[10]B. Grünbaum, *Convex Polytopes* (Wiley, New York, 1967).
[11]S.S. Holland, Jr. in *Trends in Lattice Theory*, edited by J.C. Abbott (Van Nostrand, New York, 1970).
[12]A.M. Gleason, J. Math. Mech. **6**, 885 (1957).
[13]A.A. Lodkin, Funct. Anal. Appl. **8**, 318 (1974); **9**, 271 (1975) Erratum.
[14]J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1969), p. 68, Example 9.
[15]See Dixmier, Ref. 14, p. 68, Example 10.

Nonuniqueness in the inverse source problem in acoustics and electromagnetics*

Norman Bleistein and Jack K. Cohen

Department of Mathematics, University of Denver, Denver, Colorado 80210
(Received 18 May 1976; revised manuscript received 2 August 1976)

A recently developed formulation of the inverse source problem as a Fredholm integral equation of the first kind provides motivation for the development of analytical characterizations of the nonuniqueness in the inverse source problem. Nonradiating sources, i. e., sources for which the field is identically zero outside a finite region, are introduced. It is then shown that the null space of the Fredholm integral equation is exactly the class of nonradiating sources.

1. INTRODUCTION

Recently, Bleistein and Bojarksi[1] presented a new formulation of the inverse source problem for the scalar wave equation. The source function is shown to be a solution of a Fredholm integral equation of the first kind. Here we consider that equation and the extension of the formulation to Maxwell's equations, as well. We show that the solutions to these integral equations are not unique. We further relate this nonuniqueness to features of the direct radiation problems.

More specifically, we give analytic characterizations of sources which produce no radiated field. We show by example that such nonradiatiating sources do exist. Furthermore, we exploit our representation of the class of nonradiating sources to show that they are also the elements of the null space of the integral operator which arises in our formulation of the inverse source problem. This is true both in the scalar and in the electromagnetic case. Thus, if nonuniqueness is to be viewed as a flaw, in these cases it is a flaw of the direct radiation phenomena, rather than a flaw of our formulation of the inverse problem.

For sources known *a priori* to be impulsive with known time of impulse, we find a unique solution to the scalar inverse source problem. For the vector problem, information about the vector nature of the source is required, as well. Research is presently in progress on the other types of additional information that suffices to make the solution of the inverse source problem unique.

We note that the existence of nonradiating sources has been demonstrated earlier in the literature. For example, one can use Green's theorem to replace a source distribution in a domain by a monopole-dipole distribution over any surface bounding that domain (see, for example, Ref. 2, p. 192) such that each yields the same field outside the bounding surface. This demonstrates nonuniqueness. The "difference" of these source distributions then yields zero field outside the bounding surface.

An alternative demonstration of nonradiating sources proceeds as follows.[3] Let us suppose that we seek a nonradiating source for the wave equation, for example. Let f be a function which is zero outside a finite domain. Apply the wave operator to f and let this new function be a source for the wave equation. The solution to the inhomogeneous wave equation with this source is f, itself, which is zero outside a finite domain. Thus, the source is nonradiating.

In an earlier paper, Müller[4] also discussed (far field) approximately nonradiating sources from the point of view of efficiency of radiating systems.

We repeat that the additional characterizations of nonuniqueness which we present here are relevant to our analysis of the inverse source integral equation. We remark that Friedlander presents one of our characterizations as an asymptotic result. That is, for the wave equation, he shows that the radiated field is zero to order $1/r$, with r being the distance from the origin, if the Fourier transform of the source distribution is zero on the hypercone, $\omega = ck$. Here ω is frequency, c is sound speed, and k is wavenumber.

2. SOME REMARKS ABOUT DIRECT PROBLEMS

Here we shall discuss the direct problem. Our objective is to bring out certain features of the direct problem which are relevant to the uniqueness or determinacy of solutions of the inverse problem.

To begin, let us consider the wave equation

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) U(\mathbf{r},t) = -F(\mathbf{r},t), \quad \mathbf{r} = (r_1, r_2, r_3) = (x, y, z), \tag{2.1}$$

with

$$U(\mathbf{r},t) \equiv 0, \quad F(\mathbf{r},t) \equiv 0, \quad t < t_0. \tag{2.2}$$

Here ∇^2 is the three-dimensional Laplacian and t_0 is finite. Furthermore, we shall assume that the source distribution is confined to (assumed to be nonzero only in) a finite domain $\mathcal{D}_0$. This domain is assumed to be contained in a large domain $\mathcal{D}$ (see Fig. 1). Ultimately, in the inverse problem, we shall assume that the radiated field is observed on $\partial\mathcal{D}$.

We introduce the time transform,

$$u(\mathbf{r},\omega) = \int_{-\infty}^{\infty} U(\mathbf{r},t)\exp(i\omega t)\,dt \tag{2.3}$$

and the space–time transform

$$\tilde{u}(\mathbf{k},\omega)=\iiiint U(\mathbf{r},t)\exp\{i[\omega t-\mathbf{k}\cdot\mathbf{r}]\}\,dt\,d^3r$$
$$=\iiint u(\mathbf{r},\omega)\exp(-i\mathbf{k}\cdot\mathbf{r})\,d^3r. \tag{2.4}$$

Here, the domain of integration is all of space in the latter, all of space–time in the former. Throughout, we shall adhere to this convention of denoting functions in space–time by capital letters, their temporal Fourier transforms by lower case letters and temporal-spatial transforms by lower case letters with tildes (~) over them.

The time reduced problem equivalent to (2.1) is

$$(\nabla^2+\omega^2/c^2)u(\mathbf{r},\omega)=-f(\mathbf{r},\omega) \tag{2.5}$$

with u outgoing, which we represent as

$$u(\mathbf{r},\omega)\sim\frac{\exp(i\omega r/c)}{4\pi r}\,u_0(\hat{r},\omega),\quad r\to\infty. \tag{2.6}$$

Here, we have introduced the notation

$$r=|\mathbf{r}|,\quad \hat{r}=\mathbf{r}/r. \tag{2.7}$$

We shall refer to $u_0(\hat{r},\omega)$ as the *phase and range normalized far field amplitude.*

We introduce the outgoing Green's function

$$g(R,\omega)=\frac{\exp(i\omega R/c)}{4\pi R},\quad R=|\mathbf{r}-\mathbf{r}'|, \tag{2.8}$$

and can then express the solution to (2.5) and (2.6) as

$$u(\mathbf{r},\omega)=\iiint_{D_0} f(\mathbf{r}',\omega)g(R,\omega)\,d^3r'. \tag{2.9}$$

Let us define a positive number "a" such that the sphere of radius a centered at the origin contains the domain D_0 and that this sphere, in turn, is contained in D. For $r>a$, we may represent u by

$$u(\mathbf{r},\omega)=\frac{i\omega}{c}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}h_l^{(1)}(\omega r/c)Y_{lm}(\theta,\phi)$$
$$\times\int_0^a f_{lm}(r',\omega)j_l(\omega r'/c)r'^2dr', \tag{2.10}$$
$$r>a.$$

Here, we have used the representation of g valid for $r>a$, $r'<a$ (Ref. 5, p. 541). In this equation

$$f_{lm}(r',\omega)=\int_0^{\pi}\sin\theta'\,d\theta'\int_0^{2\pi}d\phi' f(\mathbf{r}',\omega)Y^*_{lm}(\theta',\phi'),$$
$$l=0,1,2,\cdots,\quad |m|\le l. \tag{2.11}$$

These functions are also the coefficients of f in the expansion

$$f(\mathbf{r},\omega)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}f_{lm}(r,\omega)Y_{lm}(\theta,\phi) \tag{2.12}$$

with respect to the complete set of functions $Y_{lm}(\theta,\phi)$. From the form of (2.10), we see that the radiated field (i.e., u for $r>a$) depends only on the *projections* of the coefficients f_{lm} on the functions $r^2j_l(\omega r/c)$, $|m|\le l$. Thus, we can create a source for which there is no radiated field by setting these projections equal to zero.

We introduce the following:

Definition 1: Let $f(\mathbf{r},\omega)$ be a source function which is nonzero only inside a finite domain D_0. Such a source will be called *nonradiating* if the solution (2.9) is zero outside of some sphere containing D_0.

By studying the spherical harmonic expansion of the solution we have proven the following.

Lemma 1: A source f, nonzero only inside a finite domain D_0, is nonradiating if and only if the integrals

$$c_{lm}=\int_0^a f_{lm}(r,\omega)j_l(\omega r/c)r^2\,dr \tag{2.13}$$

are zero for all $l=0,1,2,\cdots$, $|m|\le l$, with the sphere of radius a sufficiently large to contain D_0.

We shall now generate a nonradiating source by using this idea of projection. Let f be given by

$$f(\mathbf{r},\omega)=\begin{cases}\delta(\mathbf{r})-\dfrac{j_0(\omega r/c)}{\int_0^a r^2j_0^2(\omega r/c)\,dr}, & r\le a,\\ 0, & r>a.\end{cases} \tag{2.14}$$

Here $\delta(\mathbf{r})$ is the Dirac delta function. The function f was constructed merely by seeking a function exactly of the form (2.12) but having the same projections c_{lm} as the delta function. One can now verify by direct substitution into (2.9) or (2.10) that, in fact, for this source, the radiated field is identically zero.

We note here that the integration in (2.15) can be carried out explicitly to yield

$$f(\mathbf{r},\omega)=\delta(\mathbf{r})-\left(\frac{2\omega}{c}\right)^3\frac{j_0(\omega r/c)}{[2\omega a/c-\sin(2\omega a/c)]},\quad r\le a. \tag{2.15}$$

Indeed, we can even obtain $F(\mathbf{r},t)$ explicitly. The result after much calculation is

$$F(\mathbf{r},t)$$
$$=\delta(\mathbf{r})\delta(t)-\frac{3}{rc^2}H(a-r)\cdot\left\{\delta'\left(t-\frac{2a+r}{c}\right)\left[\frac{c}{5a}-\frac{1}{2}\left(\frac{c}{a}\right)^3\right.\right.$$
$$\left.\times\left(t-\frac{r}{c}\right)^2\right]-\delta'\left(t-\frac{2a-r}{c}\right)\left[\frac{c}{5a}-\frac{1}{2}\left(\frac{c}{a}\right)^3\left(t+\frac{r}{c}\right)^2\right]$$
$$-\left(\frac{c}{a}\right)^3\left[\left(t-\frac{r}{c}\right)\delta\left(t-\frac{2a+r}{c}\right)-\left(t+\frac{r}{c}\right)\delta\left(t-\frac{2a-r}{c}\right)\right.$$
$$\left.\left.+H\left(t-\frac{2a+r}{c}\right)-H\left(t-\frac{2a-r}{c}\right)\right]\right\},\quad r\le a,$$
$$F(\mathbf{r},t)\equiv0,\quad r>a. \tag{2.16}$$

The main point of this result is to explicitly exhibit that the source has finite extent both in space and time, i.e., the source is physically reasonable.

We see in this example that by observing the radiated field outside $r=a$, we cannot distinguish between the impulsive source $\delta(\mathbf{r})$ and its "equivalent source" implicit in (2.13). Indeed, if we replace a by $a'<a$ in (2.13), we obtain a whole continuum of equivalent sources indistinguishable from $\delta(\mathbf{r})$ *on the basis of observations of radiated field data, alone*. We emphasize this latter point since, as we shall indicate below, certain types of additional pieces of information will suffice to eliminate such nonuniqueness.

We shall now derive an alternative characterization of nonradiating sources. To do so, we consider the space—time Fourier transform of the source, namely,

$$\tilde{f}(\mathbf{k},\omega)=\iiint_{r\leq a} f(\mathbf{r},\omega)\exp(-i\mathbf{k}\cdot\mathbf{r})\,d^3r, \tag{2.17}$$

and substitute in this formula the identity (Ref. 5, p. 567),

$$\exp(-i\mathbf{k}\cdot\mathbf{r})=4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-i)^l j_l(kr)Y_{lm}(\alpha,\beta)Y_{lm}^*(\theta,\phi). \tag{2.18}$$

Here (θ,ϕ) are the polar angles of $\mathbf{r}$ and (α,β) are the polar angles of $\mathbf{k}$. We find that

$$\tilde{f}(\mathbf{k},\omega)=4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-i)^l Y_{lm}(\alpha,\beta)\int_0^a f_{lm}(r,\omega)j_l(kr)\,r^2dr, \tag{2.19}$$

with f_{lm} defined by (2.11).

If f is a nonradiating source, then the integrals on the right in (2.19) are all zero when

$$\omega=ck. \tag{2.20}$$

This is a four-dimensional cone in $(\underset{\sim}{\mathbf{k}},\omega)$-space and we see that on this cone the transform $\tilde{f}(\mathbf{k},ck)$ is zero. Alternatively, if $\tilde{f}(\mathbf{k},ck)=0$, then, by the completeness of the functions $Y_{lm}(\alpha,\beta)$, the integrals in (2.19) must all be zero for $\omega=ck$. Thus we have proven the following:

Theorem 1: Let $f(\mathbf{r},\omega)$ be a function regular enough to have an expansion in spherical harmonics and nonzero only for $r\leq a$ for some finite a. Then $f(\mathbf{r},\omega)$ is a nonradiating source if and only if

$$\tilde{f}(\mathbf{k},ck)=0. \tag{2.21}$$

The energy radiated at frequency $\underset{\sim}{\omega}$ can be shown to be proportional to the integral of $|\tilde{f}|^2$ over the surface $k=\omega/c$; see, for example, Ref. 6, Sec. 1.2. Thus, the sources we have defined as *nonradiating* indeed do not contribute to the radiated power.

We turn now to the case of Maxwell's equation, in which there arise new features not present in the scalar case. We consider electromagnetic fields $\mathbf{E}$ and $\mathbf{H}$ which arise due to a current density $\mathbf{J}$. Again we shall use lower case letters for temporal transforms and (an over tilde) on lower case letters for full spatial temporal transforms.

We quote the following[7]:

$$\mathbf{e}(\mathbf{r},\omega)=i\omega\left[I+\frac{c^2}{\omega^2}\nabla\nabla\right]\cdot\mathbf{a}(\mathbf{r},\omega), \tag{2.22}$$

$$h(\mathbf{r},\omega)=\mu^{-1}\nabla\times\mathbf{a}(\mathbf{r},\omega). \tag{2.23}$$

Here, μ is the magnetic permeability, $\mathbf{a}(\mathbf{r},\omega)$ the vector potential,

$$\mathbf{a}(\mathbf{r},\omega)=\mu\int g(R,\omega)j(\mathbf{r}',\omega)\,dV', \tag{2.24}$$

with g defined by (2.8), I is the 3×3 dyadic identity operator, and $\nabla\nabla$ is the dyadic differential operator with elements

$$(\nabla\nabla)_{\lambda\nu}=\frac{\partial}{\partial r_\lambda}\frac{\partial}{\partial r_\nu},\quad \lambda,\nu=1,2,3. \tag{2.25}$$

As above, our interest is in nonradiating sources. For this purpose, we use (2.24) to rewrite (2.22) as

$$\mathbf{e}(\mathbf{r},\omega)=-i\omega\mu\int g(R,\omega)\left[I+\frac{c^2}{\omega^2}\nabla\nabla\right]\mathbf{j}(\mathbf{r},\omega)\,dV'. \tag{2.26}$$

We now introduce the notion of a nonradiating current density.

Definition 2: Let $\mathbf{j}(\mathbf{r},\omega)$ be a current density which is nonzero only inside a finite domain D_0. Such a source will be called *nonradiating* if the solution (2.26) is zero outside some sphere containing D_0.

We note that outside of D_0, $\mathbf{e}$ and $\mathbf{h}$ are related through a curl. Thus, if $\mathbf{e}$ is zero, so is $\mathbf{h}$.

It follows from (2.26) that $\mathbf{j}$ is nonradiating if and only if the three components of

$$\left[I+\frac{c^2}{\omega^2}\nabla\nabla\right]\mathbf{j}$$

are nonradiating in the sense of Definition 1. Thus, from Lemma 1, we conclude the following:

Lemma 2: A source $\mathbf{j}$ nonzero only inside a finite domain D_0 is nonradiating if and only if the integrals

$$\mathbf{c}_{lm}=\int_0^a dr\int_0^\pi \sin\theta\,d\theta\int_0^{2\pi}d\phi\, r^2 j_l(\omega r/c)\times Y_{lm}^*(\theta,\phi)\left[I+\frac{c^2}{\omega^2}\nabla\nabla\right]\mathbf{j}(\mathbf{r},\omega) \tag{2.27}$$

are all zero, $l=0,1,2,\cdots$, $|m|\leq l$.

Paralleling the discussion of the scalar case, we introduce the Fourier transform

$$\iiint_{r\leq a}\left[I+\frac{c^2}{\omega^2}\nabla\nabla\right]\mathbf{j}(\mathbf{r},\omega)\exp(-i\mathbf{k}\cdot\mathbf{r})\,d^3r=\left[I-\frac{c^2}{\omega^2}\mathbf{k}\mathbf{k}\right]\tilde{\mathbf{j}}(\mathbf{k},\omega). \tag{2.28}$$

Here $\mathbf{kk}$ is the dyad with components

$$(\mathbf{kk})_{\lambda\nu}=k_\lambda k_\nu,\quad \lambda,\nu=1,2,3. \tag{2.29}$$

We now apply Theorem 1 to each component of the transform here. Alternatively, we can expand the exponential according to (2.18) and compare each coefficient to the $\mathbf{c}_{lm}$'s in (2.27). Either of these methods leads to the following conclusion.

Lemma 3: A source $\mathbf{j}$, nonzero only inside a finite domain D_0, is nonradiating if and only if

$$[I-\hat{k}\hat{k}]\cdot\tilde{\mathbf{j}}(\mathbf{k},ck)=0. \tag{2.30}$$

Here $\hat{k}\hat{k}$ is the unit dyad formed by dividing by k^2 in (2.29). We note that $I - \hat{k}\hat{k}$ is a projector since

$$[I - \hat{k}\hat{k}]^2 = I - \hat{k}\hat{k}, \quad (I - \hat{k}\hat{k})_{\nu\lambda} = (I - \hat{k}\hat{k})_{\lambda\nu}. \tag{2.31}$$

This projector annihilates the "radial" component of a vector in $\mathbf{k}$ space,

$$[I - \hat{k}\hat{k}] \cdot \hat{k} = \hat{k} - \hat{k}(\hat{k} \cdot \hat{k}) = 0, \tag{2.32}$$

while leaving the "angular" part of the vector unchanged.

Thus, the restriction that $\mathbf{j}$ be nonradiating means that on the four-dimensional cone, $\omega = ck$, $\tilde{\mathbf{j}}$ is wholly in the $\hat{k}$ direction. (In particular, it might vanish completely.) The amplitude of this vector on $\omega = ck$ can be expressed in terms of the charge density through the continuity equation,

$$\nabla \cdot \mathbf{j}(\mathbf{r}, \omega) + i\omega\rho(\mathbf{r}, \omega) = 0. \tag{2.33}$$

Upon applying the Fourier transform we have

$$\hat{k} \cdot \tilde{\mathbf{j}}(\mathbf{k}, \omega) = \omega\tilde{\rho}(\mathbf{k}, \omega)/k. \tag{2.34}$$

Now, if (2.30) holds, i.e., if $\mathbf{j}$ is nonradiating, then

$$\tilde{\mathbf{j}}(\mathbf{k}, ck) = c\tilde{\rho}(\mathbf{k}, ck)\hat{k} \tag{2.35}$$

which explicitly exhibits $\tilde{\mathbf{j}}$ as a vector along $\hat{k}$ when $\omega = ck$.

It is interesting to further study the implications of (2.34). To this end, we introduce

$$\tilde{\mathbf{j}}_N = (\tilde{\mathbf{j}} \cdot \hat{k})\hat{k} = k^{-2}\omega\tilde{\rho}(\mathbf{k}, \omega)\mathbf{k}. \tag{2.36}$$

That is, $\tilde{\mathbf{j}}_N$ is a particular solution of the transformed continuity equation. To invert the spatial transform here we employ the following:

(i) multiplication by $i\mathbf{k}$ is the transform of ∇;

(ii) k^{-2} is the transform of $(4\pi r)^{-1}$;

(iii) the transform of a product is the convolution of the transforms. Thus, we conclude that

$$\mathbf{j}_N = \frac{-i\omega}{4\pi} \iiint_{D_0} \frac{\nabla'\rho(\mathbf{r}', \omega)}{R} dV'$$

$$= \frac{-i\omega\nabla}{4\pi} \iiint_{D_0} \frac{\rho(\mathbf{r}', \omega)}{R} dV',$$

$$R = |\mathbf{r} - \mathbf{r}'|. \tag{2.37}$$

That $\mathbf{j}_N$ is nonradiating may be checked directly by substituting (2.37) into (2.24), observing that $\mathbf{a}$ is then the gradient of a scalar and hence $\mathbf{a}$ in (2.23) is identically zero. Then $\mathbf{e}$ is zero, as well, outside of D_0. The remaining part of $\tilde{\mathbf{j}}$,

$$[I - \hat{k}\hat{k}] \cdot \tilde{\mathbf{j}} = \tilde{\mathbf{j}} - \tilde{\mathbf{j}}_N, \tag{2.38}$$

may be written as the cross product of $\hat{k}$ with a vector whose radial component is arbitrary. The inverse transform of this cross product is a curl of a vector. It is this part of $\mathbf{j}$ which produces a radiated field, except when Lemma 3 is satisfied.

Again, from Ref. 6, Sec. 1.2, we note that the power radiated at frequency ω is proportional to $\{[I - \hat{k}\hat{k}] \cdot \tilde{\mathbf{j}}\}^2$ integrated over the surface $k = \omega/c$. Thus, as in the

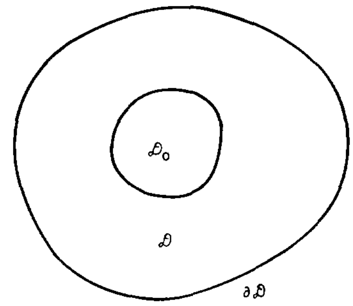

FIG. 1.

scalar case, our definition of a nonradiating current distribution is consistent with this definition of radiated power.

To recapitulate: For the scalar wave equation we have demonstrated a whole class of source distributions which are nonradiating. For Maxwell's equations, the class of nonradiating sources was even richer since it included sources determined from the scalar characterization and still others which arise as a consequence of the specific vector nature of Maxwell's equations.

3. THE WAVE EQUATION—FAR FIELD OBSERVATIONS

We shall consider here the problem defined by (2.1) and (2.2). Our objective is to derive a fundamental identity relating the Fourier transform, $\tilde{f}(\mathbf{k}, \omega)$, of the source distribution, F, to the phase and range normalized far field amplitude, u_0, defined in (2.6). We shall then discuss some of the implications of this identity for inverse source problems. The constraints on F imposed below Eq. (2.2) still apply. The signal $U(\mathbf{r}, t)$ is observed over the entire boundary ∂D in Fig. 1. (If the observations are not made in the far field, $\partial U/\partial n$ must be observed, as well.) We seek information about F in terms of the values of U observed on ∂D.

We turn immediately to the time transformed problem (2.4) and (2.5) and the solution representation (2.9). We are interested in values of $\mathbf{R}$ on ∂D. As our definition of "far field" ($r \gg r'$) observation, we require that ∂D and D_0 are such that the expansion,

$$R \sim r - \hat{r} \cdot \mathbf{r}', \quad \mathbf{r} \in \partial D, \quad \mathbf{r}' \in D_0, \tag{3.1}$$

is valid. We use (3.1) in the phase of the Green's function (2.8) and replace R by r in the amplitude of the Green's function. The solution formula (2.9) then becomes

$$u(\mathbf{r}, \omega) \sim \iiint_{D_0} \frac{\exp\{i\omega/c[r - \hat{r} \cdot \mathbf{r}']\}}{4\pi r} f(\mathbf{r}', \omega)\, d^3r', \quad \mathbf{r} \in \partial D. \tag{3.2}$$

A comparison of (2.6) and (3.2) yields the result

$$u_0(\hat{r}, \omega) = \iiint_{D_0} \exp(-i\omega\hat{r} \cdot \mathbf{r}'/c) f(\mathbf{r}', \omega)\, d^3r'. \tag{3.3}$$

Since $f = 0$ outside of D_0, the integral here can be expressed in terms of the spatial temporal transform, (2.17),

$$u_0(\hat{r}, \omega) = \tilde{f}(\omega\hat{r}/c, \omega). \tag{3.4}$$

If we identify the direction of observation with the direction of $\mathbf{k}$,

$$\hat{r}=\hat{k}, \tag{3.5}$$

then (3.4) yields information about $\tilde{f}$,

$$\tilde{f}(\mathbf{k}, ck)=u_0(\hat{k}, ck). \tag{3.6}$$

By observing everywhere on ∂D, we obtain (3.6) for all directions of $\mathbf{k}$. Thus we obtain the Fourier transform $\tilde{f}(\mathbf{k},\omega)$ on the four-dimensional cone

$$\omega = ck. \tag{3.7}$$

Unfortunately, (3.6) provides insufficient information about the transform of the source distribution to allow for the determination of F through Fourier inversion. However, it should be noted that (3.7) is the same cone on which we found that the transform was zero for nonradiating sources. Thus, were we to obtain information about $\tilde{f}$ here elsewhere but on the cone, we would be obtaining information about the nonradiating part of F from radiated field observations, alone. We note, further, that if $\tilde{f}(\mathbf{k}, ck)=0$, i.e., the source is nonradiating, then $u_0(\hat{k}, ck)=0$. This is Friedlander's result.

We close this section with a single example in which additional information about the source distribution allows us to determine it completely. Let us suppose, then, that we know *a priori* that the source is impulsive, i.e., that F has the form

$$F(\mathbf{r}, t)=F_0(\mathbf{r})\delta(t) \tag{3.8}$$

and we wish to determine F_0. In this case, the temporal transform defined by (2.17) is given by

$$f(\mathbf{r},\omega)=F_0(r), \tag{3.9}$$

and from (3.6)

$$\tilde{F}_0(\mathbf{k})=u_0(\hat{k}, ck), \tag{3.10}$$

which allows us to completely determine F_0 by

$$F_0(\mathbf{r})=\frac{1}{(2\pi)^3}\iiint \exp(i\mathbf{k}\cdot\mathbf{r})u_0(\hat{k}, ck)\, d^3k. \tag{3.11}$$

For example, if the actual source was

$$F(r,t)=\delta(\mathbf{r}-\mathbf{r}_0)\delta(t), \tag{3.12}$$

then the actual solution to the reduced problem would be

$$u(\mathbf{r},\omega)=g(|\mathbf{r}-\mathbf{r}_0|,\omega)\sim\exp(i\omega r/c - i\omega\hat{r}\cdot\mathbf{r}_0/c)/4\pi r. \tag{3.13}$$

Thus, ideally, we would observe on ∂D,

$$u_0(\hat{r},\omega)=\exp(-i\omega\hat{r}\cdot\mathbf{r}_0/c) \tag{3.14}$$

or

$$u_0(\hat{k}, ck)=\exp(i\mathbf{k}\cdot\mathbf{r}_0). \tag{3.15}$$

By inserting this result in (3.11) we obtain

$$F_0(\mathbf{r})=\frac{1}{(2\pi)^3}\iiint \exp[i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)]d^3k=\delta(\mathbf{r}-\mathbf{r}_0) \tag{3.16}$$

which is, indeed, the correct source.

We remark that *a priori* knowledge of a multiplicative time dependence $\Phi(t)$ presents little added difficulty here. In this case, we replace (3.8) by

$$F(\mathbf{r},t)=F_0(\mathbf{r})\Phi(t), \tag{3.8$'$}$$

and correspondingly,

$$f(\mathbf{r},\omega)=F_0(r)\phi(\omega), \tag{3.9$'$}$$

$$\tilde{F}_0(\mathbf{k})=u_0(\hat{k}, ck)/\phi(ck), \tag{3.10$'$}$$

$$F_0(\mathbf{r})=\frac{1}{(2\pi)^3}\iiint \exp(i\mathbf{k}\cdot\mathbf{r})u_0(\hat{k}, ck)/\phi(ck)\, d^3k. \tag{3.11$'$}$$

The fact that the assumption (3.8) leads to a unique determination of the source, shows that if, in (2.5), the source $f(\mathbf{r},\omega)$ is replaced by $f(\mathbf{r})$, then the inverse problem has a unique solution. This could cause the unwary to conclude that the full inverse problem has a unique solution, which is, of course, not true.

4. THE WAVE EQUATION—GENERAL CASE

In this section we remove the restriction of far field observations by use of an inverse source integral equation developed by Bleistein and Bojarski.[1] The basic idea is to use two "independent" Green's functions, say, the outgoing function (2.8) and the incoming Green's function

$$g^*(R,\omega)=\exp(-i\omega R/c)/4\pi R \tag{4.1}$$

to form the Green's identities

$$-u(\mathbf{r},\omega)\gamma(D;\mathbf{r})+\iiint_D g(R,\omega)f(\mathbf{r}',\omega)d^3r' = \iint_{\partial D} n'\cdot(u\nabla' g - g\nabla' u)dS', \tag{4.2}$$

$$-u(\mathbf{r},\omega)\gamma(D;\mathbf{r})+\iiint_D g^*(R,\omega)f(\mathbf{r}',\omega)\, d^3r' = \iint_{\partial D} n'\cdot(u\nabla' g^* - g^*\nabla' u)dS'. \tag{4.3}$$

Here, $R=|\mathbf{r}-\mathbf{r}'|$, $\gamma(D;r)$ denotes the characteristic function of D, i.e.,

$$\gamma(D;\mathbf{r})=\begin{cases}1 & \mathbf{r}\in D,\\ 0 & \mathbf{r}\notin D,\end{cases} \tag{4.4}$$

$\hat{n}'$ is the unit outward normal to D and ∇' denotes the gradient in the prime coordinates (integration variables). We note that

$$g(R,\omega)-g^*(R,\omega)=\frac{i}{2\pi R}\sin(\omega R/c)=\frac{i\omega}{2\pi c}j_0\left(\frac{\omega R}{c}\right), \tag{4.5}$$

with j_0 the spherical Bessel function of order zero. We subtract (4.3) from (4.2) to obtain the integral equation

$$\iiint_{D_0} j_0(\omega R/c)f(\mathbf{r}',\omega)d^3r' = \Theta(\mathbf{r},\omega),\quad R=|\mathbf{r}-\mathbf{r}'| \tag{4.6}$$

with

$$\Theta(\mathbf{r},\omega)=\iint_{\partial D}\hat{n}'\cdot[u(\mathbf{r}',\omega)\nabla' j_0(\omega R/c) - j_0(\omega R/c)\nabla' u(r',\omega)]\,dS'. \tag{4.7}$$

Here, the right side is a function of the observed values of u and $\partial u/\partial n$ on ∂D. Thus, this is a Fredholm integral equation of the first kind for the unknown source f.

As in the previous sections, we assume that f vanishes outside a domain $D_0 \subset D$. Thus, on the left side, we replace the domain of integration by all of 3-space. We then take the spatial Fourier transform (2.4) in (4.6), taking advantage of the convolution form on both sides of the equation to obtain

$$\tilde{j}_0(\omega\mathbf{k}/c)\tilde{f}(\mathbf{k},\omega) = -\tilde{j}_0(\omega\mathbf{k}/c)\int_{\partial D}\int \exp(-i\mathbf{k}\cdot\mathbf{r}')\hat{n}' \times[i\mathbf{k}u(\mathbf{r}',\omega)+\nabla' u(\mathbf{r}',\omega)]dS'. \quad (4.8)$$

At this point, it appears that we can solve for the spatial temporal transform of F by cancelling the common factor $\tilde{j}_0$. Unfortunately, this is not possible since

$$\tilde{j}_0\ \frac{\omega\mathbf{k}}{c} = \frac{2\pi^2c^2}{\omega^2}\,\delta\left(k-\frac{\omega}{c}\right) \quad (4.9)$$

which is zero, except on the four-dimensional cone, $\omega = ck$. Once again, we can only determine $\tilde{f}(\mathbf{k},\omega)$ on this cone as

$$\hat{f}(\mathbf{k},ck) = -\int_{\partial D}\int \exp(-i\mathbf{k}\cdot\mathbf{r}')\hat{n}'\cdot[i\mathbf{k}u(\mathbf{r}',ck)+\nabla' u(\mathbf{r}',ck)]\,dS' \quad (4.10)$$

Thus, we again obtain an equation of the type (3.6), except that the right side is a more complicated function of $\mathbf{k}$. To recapture (3.6) from this result, one must use the far field approximation (2.6) and then calculate the integral here by two-dimensional stationary phase.

Let us return now to the example of Sec. 3 in which we assume that F is impulsive and given by (3.8). Then in place of (3.10) we find that

$$\tilde{F}_0(\mathbf{k}) = -\int_{\partial D}\int \exp(-i\mathbf{k}\cdot\mathbf{r}')\hat{n}'\cdot[i\mathbf{k}u(\mathbf{r}',ck)+\nabla' u(\mathbf{r}',ck)]\,dS' \quad (4.11)$$

and $F_0(\mathbf{r})$ is given by the inverse transform of this function. One can further carry out the calculations in the special case (3.12) to reproduce the source by inverting the transform here when u is given by (3.13).

Again, our comment at the end of the previous section about nonuniqueness in general despite uniqueness in the specific example should be noted.

5. INVERSE SOURCE PROBLEMS FOR MAXWELL'S EQUATIONS

We shall consider here the inverse source problem for Maxwell's equations. Our discussion shall parallel that of Secs. 2 and 3. Our objective is to derive the analog of the identity (3.6) for the case of far field observations, then to derive an exact integral equation such as (4.6) for this vector case and finally to derive the extension of (4.10) to this case.

We assume that (time-transformed) fields $\mathbf{e}$ and $\mathbf{h}$ arise due to a current distribution $\mathbf{j}$ confined to D_0 of Fig. 1 and that these fields are observed on D. The fields are outgoing, which we characterize, in analogy with (2.6) by

$$\mathbf{e}(\mathbf{r},\omega) \sim \frac{\exp(i\omega r/c)}{4\pi r}\,\mathbf{e}_0(\hat{r},\omega), \quad r\to\infty, \quad (5.1)$$

$$\mathbf{h}(\mathbf{r},\omega) \sim \frac{\exp(i\omega r/c)}{4\pi r}\,\mathbf{h}_0(\hat{r},\omega), \quad r\to\infty. \quad (5.2)$$

We shall first consider the case of far field observations. In analogy with (2.9) we begin with the solution formula (2.22).

We use the same assumptions and criteria for far field here as were used in Sec. 3, thereby obtaining in analogy with (3.2)

$$\mathbf{e}(\mathbf{r},\omega) \sim \frac{i\omega\mu\exp(i\omega r/c)}{4\pi r}(I-\hat{r}\hat{r}) \times \iiint_{D} \exp(-i\omega\hat{r}\cdot r'/c)\mathbf{j}(\mathbf{r}',\omega)\,d^3r'. \quad (5.3)$$

Here, $\hat{r}\hat{r}$ is the three-dimensional dyadic formed from the components of $\hat{r}$. By comparing (5.1) and (5.3) we conclude that

$$\mathbf{e}_0(\hat{r},\omega) = i\omega\mu(I-\hat{r}\hat{r})\cdot\tilde{\mathbf{j}}(\omega\hat{r}/c,\omega), \quad (5.4)$$

with $\tilde{\mathbf{j}}$ the fourfold Fourier transform of $\mathbf{j}$ with each component transformed as in (2.4). This result is the analog of (3.4). If, again, we identity $\hat{r}$ with $\hat{k}$ (3.5), then in analogy with (3.6), we obtain

$$\mathbf{e}_0(\hat{k},ck) = ick\mu(I-\hat{k}\hat{k})\cdot\tilde{\mathbf{j}}(\mathbf{k},ck). \quad (5.5)$$

Thus, we obtain the nonradial portion, $\tilde{\mathbf{j}}-\hat{k}(\hat{k}\cdot\tilde{\mathbf{j}})$, of the transform of $\tilde{\mathbf{j}}$ only on the four-dimensional cone, $\omega = ck$. We have seen why this is to be expected from the point of view of the direct problem. We further note here that the presence of the projector $I-\hat{r}\hat{r}$ in (5.4) and, consequently, $I-\hat{k}\hat{k}$ in (5.5) reflects the property that $\mathbf{e}_0$ is transverse to the direction of propagation.

In order to parallel the discussion of Sec. 4, we begin from a vector generalization of (4.2) for Maxwell's equations. To do so, we introduce the Green's dyadic

$$\mathbb{G}(R,\omega) = \left(I+\frac{c^2}{\omega^2}\nabla\nabla\right)g(R,\omega), \quad R = |\mathbf{r}-\mathbf{r}'|. \quad (5.6)$$

The scalar Green's function g is defined in (2.8) and the dyadic operator appearing here is defined below (2.24). Similarly,

$$\mathbb{G}^*(R,\omega) = \left(I+\frac{c^2}{\omega^2}\nabla\nabla\right)g^*(R,\omega), \quad (5.7)$$

with $*$ denoting complex conjugate.

The Green's identity corresponding to (4.2) is[7]

$$\mathbf{e}(r,\omega)\gamma(D;\mathbf{r}) - i\omega\mu\int\int_D\int \mathbb{G}(R,\omega)\cdot\mathbf{j}(\mathbf{r}',\omega)\,d^3r' = \int_\partial\int_D \hat{n}'\cdot\{\mathbb{G}\times(\nabla'\times\mathbf{e})-\mathbf{e}\times(\nabla'\times\mathbb{G})\}\,d^3r'. \quad (5.8)$$

The function $\gamma(D;\mathbf{r})$ is defined in (4.4). If $\mathbb{G}$ is replaced by $\mathbb{G}^*$ here, the equation remains valid, as well. The difference of these two equations leads to the analog of (4.6) and (4.7), namely

$$-i\omega\mu\iiint_D\left[\left(I+\frac{c^2}{\omega^2}\nabla'\nabla'\right)j_0\left(\frac{\omega R}{c}\right)\right]\cdot\mathbf{j}(\mathbf{r}',\omega)\,d^3r' = \Theta(\mathbf{r},\omega), \quad (5.9)$$

$$\Theta(\mathbf{r},\omega)=\iint_{\partial D}\hat{n}'\cdot\left\{\left[\left(I+\frac{c^2}{\omega^2}\nabla'\nabla'\right)\cdot j_0\left(\frac{\omega R}{c}\right)\right]\times(\nabla'\times\mathbf{e}) - \mathbf{e}\times\nabla'\times\left[\left(I+\frac{c^2}{\omega^2}\nabla'\nabla'\right)\cdot j_0\left(\frac{\omega R}{c}\right)\right]\right\}\, d^3r'. \quad (5.10)$$

Integration by parts in (5.9) yields the following alternative form:

$$-i\omega u\iiint_D j_0\left(\frac{\omega R}{c}\right)\left(I+\frac{c^2}{\omega^2}\nabla'\nabla'\right)\cdot\mathbf{j}(\mathbf{r}',\omega)\, d^3r'=\Theta(\mathbf{r},\omega). \quad (5.11)$$

Equation (5.9) or (5.11), along with (5.10) provides the generalization of (4.6) and (4.7) to the inverse source problem for Maxwell's equations. As in Sec. 4, these are Fredholm integral equations of the first kind. By taking the spatial Fourier transform in either equation, we obtain the generalization of (4.10),

$$\begin{aligned}&-ick\mu(I-\hat{k}\hat{k})\tilde{\mathbf{j}}(\mathbf{k},ck)\\&=\int_{\partial D}\int\exp(-i\mathbf{k}\cdot\mathbf{r}')\hat{n}'\cdot\{(I-\hat{k}\hat{k})\times[\nabla\times\mathbf{e}(\mathbf{r}',ck)]\\&\quad+\mathbf{e}(\mathbf{r}',ck)\times[i\mathbf{k}\times(I-\hat{k}\hat{k})]\}\,dS'.\end{aligned} \quad (5.12)$$

Thus, on the four-dimensional cone $\omega=ck$, we obtain an expression for that part of $\tilde{\mathbf{j}}$ which gives rise to the radiated field. This type of result is consistent with our analysis of the direct problem in Sec. 2.

We close this section with a simple example in which a unique solution is obtained for $\mathbf{j}$ despite the inherent nonuniqueness in (5.17). Thus, let us suppose that the original time dependent source is impulsive and restricted to being horizontal. Therefore,

$$\mathbf{J}(\mathbf{r},t)=\mathbf{J}_0(\mathbf{r})\delta(t),\quad \mathbf{J}_0=\mathbf{J}_{01}\hat{r}_1+\mathbf{J}_{02}\hat{r}_2. \quad (5.13)$$

In this case, we would seek a solution of (5.12) of the form

$$\tilde{\mathbf{j}}(\mathbf{k},ck)=\tilde{\mathbf{j}}(\mathbf{k}),\quad \tilde{\mathbf{j}}(\mathbf{k})\cdot\hat{r}_3=0. \quad (5.14)$$

In addition, from (5.12), $\tilde{\mathbf{j}}(\mathbf{k})$ must satisfy the equation

$$(I-\hat{k}\hat{k})\cdot\tilde{\mathbf{j}}(\mathbf{k})=(I-\hat{k}\hat{k})\cdot\tilde{\mathbf{j}}_0(\mathbf{k}). \quad (5.15)$$

This equation has as a particular solution

$$(I-\hat{k}\hat{k})\cdot\tilde{\mathbf{j}}_0(\mathbf{k}),$$

and as a general solution

$$\tilde{\mathbf{j}}(\mathbf{k})=(I-\hat{k}\hat{k})\cdot\tilde{\mathbf{j}}_0(\mathbf{k})+\psi(\mathbf{k})\hat{k} \quad (5.16)$$

with $\psi(\mathbf{k})$ an arbitrary scalar function of $\mathbf{k}$. We now use the constraint in (5.14) to conclude that

$$\psi(\mathbf{k})\hat{k}\cdot\hat{r}_3=-[\hat{r}_3-(\hat{r}_3\cdot\hat{k})\hat{k}]\cdot\tilde{\mathbf{j}}_0(\mathbf{k})=(\hat{r}_3\cdot\hat{k})\hat{k}\cdot\tilde{\mathbf{j}}_0(\mathbf{k}). \quad (5.17)$$

Thus we can solve for $\psi(\mathbf{k})$ and conclude that, indeed,

$$\tilde{\mathbf{j}}(\mathbf{k})=\tilde{\mathbf{j}}_0(\mathbf{k}). \quad (5.18)$$

We note that we have used the fact that the source was impulsive—information of the type used in the scalar problem—and, in addition, we used information about the vector nature of the source distribution. Together, this added information sufficed for the unique determination of $\tilde{\mathbf{j}}$ and hence $\mathbf{J}(\mathbf{r},t)$. As in Sec. 4, we could as well solve the problem if $\delta(t)$ were replaced by a known time dependence $\Phi(t)$.

6. NONUNIQUENESS IN INVERSE SOURCE PROBLEMS

We shall discuss here, the integral equations (4.6) and (5.11). We shall show that both integral operators have nontrivial null spaces, i.e., that in both cases the homogeneous equation has nontrivial solutions. We shall also show that, in fact, the null spaces are the sets of nonradiating sources defined in Sec. 2.

We begin by defining the null space for the scalar problem.

Definition 3: The source function

$$f(\mathbf{r},\omega)=\int F(\mathbf{r},t)\exp(i\omega t)\,dt \quad (6.1)$$

is said to be in the null space $N(a)$ if f vanishes for $r>a$ and

$$\iiint_{r'<a} j_0(\omega R/c)f(\mathbf{r}',\omega)\,d^3r'=0,\quad R=|\mathbf{r}-\mathbf{r}'|. \quad (6.2)$$

We note that $j_0(\omega R/c)$ has the spherical harmonic expansion

$$j_0\left(\frac{\omega R}{c}\right)=4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}j_l\left(\frac{\omega r}{c}\right)j_l\left(\frac{\omega r'}{c}\right)Y_{lm}(\theta,\phi)Y^*_{lm}(\theta',\phi'). \quad (6.3)$$

By comparing (6.3) and (2.10) we see that j_0 and g have the same $\mathbf{r}'$ dependence at each order l,m. Thus we conclude the following.

Lemma 4: $f(\mathbf{r},\omega)$ is nonradiating and vanishes outside $r=a$ if and only if f is in $N(a)$.

Proof: Let f vanish outside $r=a$. Substitute (6.3) into (6.2) and conclude that such a function f is in $N(a)$ if and only if the coefficients c_{lm} defined by (2.14) are all zero. However, the vanishing of these coefficients is exactly the requirement for a source which vanishes outside $r=a$ to be nonradiating. This completes the proof.

We conclude from this that the nonuniqueness in the inverse source problem is a direct consequence of the fact that the direct problem admits nonradiating sources. Thus, if this "defect" is to be overcome, it must be done by finding additional information about the source over and above observations of the radiated field.

Let us now turn to the question of eigenfunctions. Thus, we consider the equation

$$\iiint_{r'\leq a} j_0(\omega R/c)\psi(\mathbf{r}',\omega)\,d^3r'=\lambda\psi(\mathbf{r},\omega),\quad r\leq a \quad (6.4)$$

with $\psi\equiv 0$ for $r>a$. Since j_0 satisfies the homogeneous Helmholtz equation, we conclude that

$$\lambda[\nabla^2+\omega^2/c^2]\psi(\mathbf{r},\omega)=0. \quad (6.5)$$

Here $\lambda=0$ yields the null space $N(a)$ while for $\lambda\neq 0$ we find that the normalized eigenfunctions are

$$\psi_{lm}(\mathbf{r},\omega)=N_l^{-1}j_l(\omega r/c)Y_{lm}(\theta,\phi)H(a-r),\quad l=0,1,\cdots,\quad |m|\leq l, \quad (6.6)$$

and

$$\lambda_{lm}=4\pi N_l^2,\quad l=0,1,2,\cdots,\quad |m|\leq l. \quad (6.7)$$

Here, H denotes the unit step function and

$$N_l^2=\int_0^a j_l^2(\omega r/c)r^2\,dr. \tag{6.8}$$

The eigenvalues are seen to be positive. Actually, the integral in (6.8) is given in Ref. 8. The eigenvalues are then given by

$$\lambda_{l,m}=\frac{a^3}{2}\left\{\left[j_l\left(\frac{a\omega}{c}\right)\right]^2-j_{l-1}\left(\frac{a\omega}{c}\right)j_{l+1}\left(\frac{a\omega}{c}\right)\right\},$$
$$l=0,1,2,\cdots,\quad |m|\le l. \tag{6.9}$$

Furthermore, for fixed frequency and increasing l we find that

$$\sim\frac{\pi a^3}{4l^3}\left(\frac{ea\omega}{2cl}\right)^{2l}\left[1+O\left(\frac{1}{l}\right)\right]=O\left(\frac{1}{l^{2l+3}}\right),\quad l\to\infty. \tag{6.10}$$

Therefore, the eigenvalues decrease rapidly to zero. Thus, the inverse problem is confirmed to be both ill-posed and ill-conditioned, as is well known.

Since $N(a)$ is not empty [cf. (2.15)], the eigenfunctions are not complete. Indeed, for any function $f(\mathbf{r},\omega)$ which vanishes outside $r=a$, if we subtract its projection on the eigenfunctions, we obtain a function in the null space. We denote the projection on the null space by

$$P_0f=f-\sum_{l=0}^{\infty}\sum_{m=-l}^{l}d_{lm}\psi_{lm}(\mathbf{r},\omega). \tag{6.11}$$

Here

$$d_{lm}=\int\int\int f(\mathbf{r},\omega)\psi_{lm}^*(\mathbf{r},\omega)\,d^3r=N_l^{-1}c_{lm},$$
$$l=0,1,2,\cdots,\quad |m|\le l, \tag{6.12}$$

with the c_{lm}'s defined by (2.12) and (2.14). In fact, the source (2.15) is just the function $P_0\delta(\mathbf{r})$.

The analogous results for the electromagnetic case follow almost immediately, because the kernel of the integral operator in (5.11) is exactly the same as it was for the scalar case. Thus we state the following without proof:

Lemma 5: Let the current density $\mathbf{j}$ be confined to a finite domain $D_0\subset D$. Then the null space for the integral equation (5.11) is exactly the class of nonradiating sources.

7. CONCLUDING REMARKS

We have shown that our formulations of the inverse source problem in acoustics or electromagnetics admit nonunique solutions. When this nonuniqueness is characterized by a null space of source functions, that collection of sources turned out to be exactly the set of "nonradiating" sources for the corresponding *direct* problem. Thus, we concluded that unique determination of a source distribution requires other information in addition to observations of the radiated field.

One example of such "additional information" was provided for the scalar problem, namely, the case of a known multiplicative time dependence. (In the vector problem, additional information of the vector nature of the source was required, as well.) Research is now in progress on a systematic investigation of such " additional information." We cite some preliminary results, below.

If the radial dependence can be characterized by a known multiplicative factor in the source, then explicit solutions of the scalar problem can again be determined much as in the case of known multiplicative time dependence.

If the (r,t) dependence can be split off as a known multiplicative factor, this will suffice for unique determination of the source.

If the source is known to have no component in the null space, then, in fact, it is exactly given by its eigenfunction expansion.

We may alternatively take the point of view that we seek only an "equivalent" source. That is, for a given radiated field, we might seek a source with a prescribed temporal dependence. For example, (3.10) or (4.11) are expressions for $\tilde{F}_0(\mathbf{k})$ when $F_0(\mathbf{x})$ is an *equivalent source* with impulsive time dependence no matter what the actual source distribution happens to be.

Finally, we mention the synthesis problem. Here, we seek a source distribution which is to provide a prescribed radiated field. In the scalar problem, let us denote the unknown spatial array by $F_0(\mathbf{r})$ and the response to a pure frequency, ν, by $U(\mathbf{r},\nu)$. That is, corresponding to the source distribution,

$$F_0(\mathbf{r})\exp(-i\nu t),$$

the wave equation (2.5) has solution

$$U(\mathbf{r},\nu)\exp(-i\nu t).$$

If $U(\mathbf{r},\nu)$ were given for all ν, then (3.10) or (4.10) provides a solution for $\tilde{F}_0(\mathbf{k})$ when u is replaced by the function U appearing here.

The solution presented here avoids many real and practical issues. Typically, one is not given the radiated field for all frequencies. Indeed, one is more likely to be given the radiated power density in a limited band width, rather than the field itself. Furthermore, this solution has a nonradiating component which we could extract via use of the projection (6.9). Presumably, we might use the added freedom of nonradiating sources to generate sources satisfying other constraints, such as being confined as nearly as possible to a region in space, such as a horn or a surface. Research on this approach to synthesis is presently in progress.

*This work partially supported by the Office of Naval Research.

[1]N. Bleistein and N.N. Bojarski recently developed formulations of the inverse problem in acoustics and electromagnetics, DRI Report #MS-R-7501, NTIS #AD/A-003 588 (1974).
[2]J.A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).
[3]F.G. Friedlander, Proc. London Math. Soc. 3, 27, 551–576 (1973).
[4]C. Müller, IRE Trans. Antennas Propag. **AP-4**, 3, 224–32 (1956).
[5]J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).
[6]L.B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, Englewood Cliffs, New Jersey, 1973).
[7]Chen-To Tai, *Dyadic Green's Functions in Electromagnetic Theory* (Intext Educational Publishers, Scranton, 1971).
[8]G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1958), 2nd ed.

Subalgebras of the similitude algebra and their invariants

Anthony Peccia*

Centre de Recherches Mathématiques, Université de Montréal, Montreal, Canada
(Received 27 August 1976)

The subalgebras of the similitude algebra have previously been classified into conjugacy classes; in this article these classes are classified into isomorphism classes. For each conjugacy class of subalgebras, the invariants are also calculated. All the results are summarized in tables.

1. INTRODUCTION

A general method for obtaining the subalgebras of a given Lie algebra was given in Ref. 1. The method consists of an iterative procedure for reducing the problem of finding the subalgebras of L with nontrivial ideal N, to that of finding the subalgebras of the ideal N and those of the factor algebra. If the algebra L is simple, then matrix realizations of the algebra are used to obtain its subalgebras. In a later paper,[2] this method was used to obtain all the subalgebras of the similitude algebra (the semidirect product of the Poincaré algebra and the dilatation operator).

The subalgebras were classified into conjugacy classes under the connected component of the similitude group, and a representative algebra for each class was also listed. Two algebras L and L' are conjugate under a group G if $\exists g \in G$ such that $gLg^{-1} = L'$. The virtue of such a classification is clarified by considering the physical interpretation of such a class.

Identifying the elements of the subalgebras with the infinitesimal transformations on space—time (i.e., rotations, boosts, translations, and etc.), the statement that two algebras are conjugate is equivalent to the statement that the two algebras describe the same set of transformations (or observables) as viewed from different coordinate systems, while, unconjugate algebras describe physically distinct operations; for example, the algebra of rotations is physically distinct from the algebra of translations.

Below, the algebras are reclassified into isomorphism classes. Each isomorphism class corresponds to an orbit of $\mathrm{GL}(n,R)$ acting on the subalgebras of dimension n. In terms of the structure constants, the statement can be rewritten in a more explicit form. Two algebras L and L' of dimension n and structure constants c_{ij}^{k} c'^{k}_{ij} respectively, are isomorphic if $\exists g \in \mathrm{GL}(n,R)$ such that

$$g_{il} c_{lm}^{n} g_{mj} g_{nk}^{-1} = c'^{k}_{ij}.$$

Since conjugate algebras have the same structure constants, they are trivially isomorphic, and so only the conjugate classes of subalgebras are classified into isomorphism classes.

The main interest in this type of classification rests on the fact that through isomorphism all algebraic properties obtained for one algebra are immediately transferable to all algebras within the same class. Further, a knowledge of a suitable choice of structure constant within the orbit may greatly reduce the difficulties and complications arising in the actual calculation of a particular result.

For each representative algebra, a basis for the set of invariants is obtained. Here, by invariant is understood a function of the elements of the algebra such that the function commutes with all the elements of the algebra. The function is assumed to be at least first differentiable in all variables, so that any invariant with a finite range of eigenvalues is ignored, like the sign of the energy in the Poincaré algebra.

This paper constitutes a sequel to a paper by Patera *et al*[3] in which the conjugacy classes of the subalgebras of the Poincaré algebra were classified into isomorphism classes, and the respective invariants, for each representative of a conjugacy class, were calculated. Since some of the conjugacy classes of the Poincaré algebra are also conjugacy classes of the similitude algebra, parts of their results are contained in this article, providing an independent check for these results.

In Sec. 2, the general method employed in classifying the conjugacy classes of subalgebras into isomorphism classes, and in calculating the invariants is discussed. The calculation of the invariants and the role which these might play in labeling the irreps of an arbitrary Lie algebra has been the subject of considerable investigation recently.[3,4,5,6] From Schur's lemma, it follows, that within the irreps, the invariants are of the form λI where $\lambda \in C$ and I is the identity operator for the representation; showing that λ is distinct within the different irreps would lead to applications analogous to those of the Casimir operators for these (generalized) invariants. Two special types of invariants, the Casimir operators, and the rational invariants were systematically treated by Abellanos *et al.*[4] Rational invariants are elements of the quotient field of the enveloping algebra. In other words, they are ratios of homogeneous polynomials on the algebra. Considerations on the more general invariants along with additional references can be found in the above mentioned paper by Patera *et al*[3]; also a short treatment will appear in the conclusion. For a discussion of operator caculus see Ref. 7.

Section 3, containing the main results, consists of a list of representative algebras for the conjugacy classes of the similitude algebra and their respective invariants. The algebras are organized first by dimension and then for each dimension, they are grouped into isomorphism classes. Each isomorphism class is designated by a notation which refers to a standard basis that characterizes that class, whenever such a basis exists. This notation is consistent with the one used by Patera,[5] who listed the standard basis of the Lie algebra characterizing each isomorphism class. Comments on the classes and the invariants also appear in Sec. 3. Section 4 is reserved for the conclusion.

2. METHODS

All the conjugacy classes of the subalgebras of the similitude algebra, listed in Ref. 2, are reorganized first by dimension and then for each class the invariants are determined. A knowledge of the invariants, often, aids the mapping of conjugacy classes into isomorphism classes.

An algorithm for obtaining the invariants is achieved by reducing the problem to that of solving a system of linear first order partial differential equations. This method has been discussed in detail in the literature.[4,5,6]

In one paper,[5] for each of the isomorphism classes of dimension $n \leq 5$, the method was used to obtain the basis for the set of its invariants. (However, here, for each representative algebra the invariants were independently determined.) The method consists in identifying the adjoint representation of a Lie algebra with a set of c-number first order linear differential operators. That is, let L be an n-dimensional Lie algebra, then

$$\forall X_i, X_j \in L, \quad \text{let}\, \mathrm{ad}(X_i)X_j \equiv [X_i, X_j] = c_{ij}^k X_k$$

and replace

$$\mathrm{ad}X_i \to x_k c_{ij}^k \frac{\partial}{\partial x_j},$$

where x_i are c-numbers. The equation for the invariant is

$$[X_i, F(X_1, X_2, \ldots, X_n)] = 0 \quad \forall X_i \in L \tag{1}$$

and reduces to

$$x_k c_{ij}^k \frac{\partial}{\partial x_j} F(x_1, x_2, \ldots, x_n) = 0, \quad i = 1, 2, 3, \ldots, n.$$

The invariant thus formed must then be converted from a function of c-number variables to an operator function. Simple identification of $X_i \leftrightarrow x_i$ will be sufficient, in the sense that the resulting function will indeed be an operator invariant, i.e., satisfy (1), only if all the x_i appearing in the function mutually commute. In the case where the function is a homogeneous polynomial (a Casimir operator) it is well known that the invariant operator must be fully symmetrized in all its variables. The reason being, that if u and v are polynomials in the x_i and

$$x_k c_{ij}^k \frac{\partial u}{\partial x_j} = v, \quad \text{then} \quad [X_i, U] = V \tag{2}$$

provided U and V are the polynomials with the same coefficients as u and v and are fully symmetrized in the corresponding $X_i \in L$.

All the invariants found for these subalgebras are of the form

$$\exp(U_{m+1}/U_m) \prod_{i=1}^{m} U_i^{\alpha_i},$$

where the U_i's are relatively prime homogeneous polynomials in the x_i. The exponents as well as the coefficients may be complex numbers. Then replacing x_i by X_i in each polynomial and symmetrizing each individual polynomial will lead to an operator invariant. The proof essentially follows from (2), as for Casimir operators, and is based on the assumption that $[[X_i, U_j], U_j] = 0 \;\forall\; X_i \in L$, for all polynomials U_j featured in the invariant. This assumption was indeed satisfied for all invariants found here and elsewhere.[3,5] It can be shown that for semisimple and nilpotent algebras, the Casimir operators are sufficient to define a basis for the set of all invariants. Further, the number of independent invariants for an algebra of dimension n is $n$mod(2).

As was mentioned earlier, in order to classify the conjugacy classes into isomorphism classes, the representative subalgebras, for each conjugacy class, are first organized by dimension. Then for each dimension they are further reorganized by the dimension of the derived algebra; this last step can be repeated as many times as is required.

All isomorphism classes for real Lie algebras of dimension $n \leq 5$ have been listed by Mubarakzyanov[8] and later reproduced in a more accessible paper.[5] For algebras of dimension ≤ 5, the classification is reduced to identifying each representative subalgebra with an abstract representative of the isomorphism classes listed in Ref. 5. The classes are denoted by $A_{n,m}^{\alpha_1 \cdots \alpha_i}$, where n identifies the dimension of the subalgebra, m is used to index the different isomorphism classes for the same dimension, and the parameters $\alpha_1 \cdots \alpha_n$ are used to select a particular class out of an infinite set of isomorphism classes. The structure constants are functions of the parameters $\alpha_1 \cdots \alpha_i$. For example, $A_{3,5}^h$ with standard basis c_1, c_2, c_3 such that $[c_1, c_2] = c_1$, $[c_2, c_3] = hc_2$ with $-1 \leq h \leq 1$ is such an infinite set of isomorphism classes.

Since no complete classification of isomorphism classes for real Lie algebras of dim ≥ 6 exists,[9] a slightly modified approach is used. The algebras are first reordered as before, and then for each set consisting of say an algebra of dimension n with a derived algebra of dimension l and second derived algebra of dimension l' and etc., the invariants are used to group the algebra into possible isomorphism classes. At this point, it often is evident which algebras are not isomorphic. Then from the so chosen candidates for a specific class (the number of candidates is not large), their largest Abelian ideals are identified and by comparing the respective factor algebras, the algebras are classified into isomorphism classes. As a further test, the explicit transformation, which transforms the structure constants of one algebra into those of the other algebras within the class, is constructed. For the similitude subalgebras, isomorphism classes of dim ≥ 6 are found to contain at most two sets of conjugacy classes.

3. ISOMORPHISM CLASSES AND INVARIANTS

The usual basis for the similitude algebra is used, with the L_i's representing the rotation generators, the K_i's representing the boost generators, D representing the dilatation generator, and the P_μ's representing the space–time translations. Then the commutation relations are

$$[L_i,L_j]=\epsilon_{ijk}L_k,\quad [K_i,K_j]=-\epsilon_{ijk}L_k,$$
$$[L_i,P_j]=\epsilon_{ijk}P_k,\quad [L_i,K_j]=\epsilon_{ijk}K_k,$$
$$[K_i,P_j]=\delta_{ij}P_0\quad [K_i,P_0]=P_i,$$
$$[D,P_u]=2P_u,\quad [L_i,P_0]=0,$$
$$[D,L_i]=0,\quad [D,K_i]=0,$$
$$i,j,k,=1,2,3.$$

The invariants of the Poincaré algebra $S_{1,1}$ are well known,

$$m_2=P_0^2-P_1^2-P_2^2-P_3^2,\quad W^2=W_\mu W^\mu,$$

where $W_\mu=\epsilon_{\mu\nu\delta\lambda}M_{\nu\delta}P_\lambda$ is the Pauli—Lubanski spin operator with $M_{0i}=K_i$ and $M_{ij}=\epsilon_{ijk}L_k$, while the invariant of the similitude algebra $S_{1,3}$ consists of the ratio of these two operators which reduces to the spin operator in the frame with the eigenvalues of $P_i=0$ (the rest frame).

The results of the classification and the basis for the invariants are summarized in Tables I—IX. The notation $S_{i,j}$ refers to the conjugacy classes as found in Ref. 2. When listing the elements of a subalgebra, the semicolon is used to indicate that all elements to its right belong to the derived algebra.

All one-dimensional subalgebras are isomorphic and appear in Table I. The n-dimensional subalgebras appear in Table n, except for those of dimension 10 and 11 which are listed in Table IX.

If an infinite set of isomorphism classes contains an infinite set of conjugacy classes, then the parameters of the set of isomorphism classes are functions of the parameters of the set of conjugacy classes. The function need not be one to one, and such a case arises in classifying the set $S_{5,4}^c$; for each c, the corresponding class of $S_{5,4}^c$ is associated with an element of $A_{3,7}^P$ such that $P=\tan c$. The range of c is $0<c<\pi$, $c\neq\pi/2$ which implies that $-\infty<\tan c<\infty$, and $\tan c\neq 0$; however, the range of the parameter P is $P>0$. This suggests that classes with $|\tan c|$ equal are isomorphic. This can be verified by a slight rearrangement of the basis of the representative algebra of $S_{5,4}^c$. In this case, as for all those of dim ≤ 5, the range of the parameters for the set of isomorphism classes has been defined.[8] For the higher dimensional case, even though the infinite sets of isomorphism classes have not been constructed yet, the problem can be treated analogously. For example, consider $S_{10,19}^{a,b}$ with $a^2+b^2\neq 0$ from Table VII; since the invariant has the exponent $1+b$, it follows that algebras with distinct b's are not isomorphic. However, $S_{10,19}^{a,b}$ is isomorphic to $S_{10,19}^{-a,b}$; this can be seen by interchanging P_2 with P_1 and L_2+K_1 with L_1-K_2 in the ordered basis of $S_{10,19}^{a,b}$. All other cases can be approached in a similar manner.

4. CONCLUSION

As was mentioned earlier, the greatest utility of such a classification of subalgebras lies in its capacity to remove redundancies in computations. The availability of a suitable choice of structure constants for a particular class can reduce the complexities in a particular calculation.

It is expected that the invariants obtained here will provide a useful tool in the representation theory of these algebras, much like the Casimir operators do, in the case of semisimple Lie algebras. The extension of special function theory, via group theory, in order to include these invariants should produce new and useful results.

On the physical aspect, the eight-dimensional algebra $S_{2,1}$ is importantly contained in the "infinite momentum frame"[10,11] calculations, in Dirac's "front frame" dynamics,[12] and in the investigations of "Galilean subdynamics."[13,14] One of the invariants of this algebra is

$$L_3-\frac{P_2}{P_0-P_3}(L_2+K_1)-\frac{P_1}{P_0-P_3}(L_1-K_2)\tag{3}$$

and because of its somewhat extensive use in physics, it has already been named the lightlike helicity or the null-plane helicity.[15,16] The name is appropriate, since for zero mass particles with discrete spin, L_2+K_1 and L_1-K_2 are both zero and so (3) reduces to L_3. The similitude algebra is another example of an algebra with nonpolynomial invariant which has reached prominence in physical applications.[17]

TABLE I. One-dimensional subalgebras.

Class	Notation	Generators	Range of parameters	Invariants
A_1	$S_{11,6}$	$L_3-\tan cK_3$;	$0<c<\pi,\ c\neq\pi/2$	generator
	$S_{12,10}$	L_3;		"
	$S_{12,20}$	$4L_3+P_0+P_3$;		"
	$S_{12,21}$	$2L_3+P_0$;		"
	$S_{12,22}$	$2L_3-P_3$;		"
	$S_{13,9}$	K_3;		"
	$S_{13,15}$	$2K_3-P_2$;		"
	$S_{14,9}$	L_2+K_1;		"
	$S_{14,20}$	$2L_2+2K_1-P_0-P_3$;		"
	$S_{14,21}$	$L_2+K_1-P_2$;		"
	$S_{15,8}$	P_0-P_3;		"
	$S_{15,9}$	P_0;		"
	$S_{15,10}$	P_3;		"
	$S_{15,22}$	D;		"
	$S_{15,41}$	$D+2a\cos cL_3-2a\sin cK_3$;	$0\leq c<\pi,\ a>0$	"
	$S_{15,42}$	$D-L_2-K_1$;		"

Further investigations along these lines should reveal the relevant and operational properties of these more general invariants, despite their present rather precarious mathematical status.

ACKNOWLEDGMENTS

I would like to extend my gratitude to J. Patera and P. Winternitz of the Centre de Recherches Mathématiques, Université de Montréal and to R. T. Sharp of the McGill University for their helpful comments.

TABLE II. Two-dimensional subalgebras.

Class	Notation	Generators	Range of parameters	Invariants
$2A_1$	$S_{9,6}$	$L_3,\ K_3;$		both generators
	$S_{10,5}$	$L_2+K_1,\ L_1-K_2;$		"
	$S_{10,13}$	$L_1-K_2+P_2,\ L_2+K_1;$		"
	$S_{11,12}$	$D,\ \cos cL_3-\sin cK_2;$	$0<c<\pi,\ c\neq\pi/2$	"
	$S_{11,18}$	$D+2aL_3,\ \cos cL_3-\sin cK_3;$	" $a>0$	"
	$S_{12,7}$	$L_3,\ P_0-P_3;$		"
	$S_{12,8}$	$L_3,\ P_3;$		"
	$S_{12,9}$	$L_3,\ P_0;$		"
	$S_{12,17}$	$4L_3+P_0+P_2,\ P_0-P_3;$		"
	$S_{12,18}$	$2L_3+P_0,\ P_3;$		"
	$S_{12,19}$	$2L_3-P_3,\ P_0;$		"
	$S_{12,32}$	$D,\ L_3;$		"
	$S_{12,41}$	$D-2aK_3,\ L_3;$	$a>0$	"
	$S_{12,42}$	$2D-4K_3+P_0+P_3,\ L_3;$		"
	$S_{12,46}$	$2D-4K_3+x(P_0+P_3),\ 4L_3+(P_0+P_3);$	$-\infty<x<\infty$	"
	$S_{13,8}$	$K_3,\ P_2;$		"
	$S_{13,11}$	$2K_3-P_2,\ P_1;$		"
	$S_{13,24}$	$D,\ K_3;$		"
	$S_{13,30}$	$D+2aL_3,\ K_3;$	$a>0$	"
	$S_{14,7}$	$L_2+K_1,\ P_0-P_3;$		"
	$S_{14,8}$	$L_2+K_1,\ P_2;$		"
	$S_{14,17}$	$2L_2+2K_1-P_0-P_3,\ P_0-P_3;$		"
	$S_{14,18}$	$L_2+K_1-P_2,\ P_0-P_3;$		"
	$S_{14,19}$	$2L_2+2K_1-P_0-P_3,\ P_2;$		"
	$S_{14,30}$	$D,\ L_2+K_1;$		"
	$S_{14,51}$	$D+L_1-K_2,\ L_2+K_1;$		"
	$S_{15,5}$	$P_0-P_3,\ P_2;$		"
	$S_{15,6}$	$P_0,\ P_3;$		"
	$S_{15,7}$	$P_1, P_2;$		"
	$S_{15,35}$	$D+2a(\cos cL_3-\sin cK_3),\ P_0-P_3$	$a=1,\ c=3\pi/2$	"
A_2	$S_{8,9}$	$K_3;\ L_2+K_1$		none
	$S_{8,17}$	$2K_3-P_2;\ L_2+K_1$		"
	$S_{11,5}$	$\cos cL_3-\sin cK_3;\ P_0-P_3$	$0<c<\pi,\ c\neq\pi/2$	"
	$S_{13,7}$	$K_3;\ P_0-P_3$		"
	$S_{13,13}$	$2K_3-P_2;\ P_0-P_3$		"
	$S_{14,49}$	$D-2aK_3;\ L_2+K_1$	$a\neq 0$	"
	$S_{14,50}$	$2D+4K_3+P_0-P_3;\ L_2+K_1$		"
	$S_{14,62}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3$		"
	$S_{14,63}$	$D-2K_3+b(L_1-K_2+P_1);\ L_2+K_1-P_2$	$b\geq 0$	"
	$S_{15,19}$	$D;\ P_0-P_3$		"
	$S_{15,20}$	$D;\ P_0$		"
	$S_{15,21}$	$D;\ P_3$		"
	$S_{15,35}$	$D+2a(\cos cL_3-\sin cK_3);\ P_0-P_3$	$a>0,\ a\neq 1$ $0\leq c<2\pi,\ c\neq 3\pi/2$	"
	$S_{15,36}$	$D-L_2-K_1;\ P_0-P_3$		"
	$S_{15,37}$	$D+2aL_3;\ P_0$	$a>0$	"
	$S_{15,38}$	$D+2aL_3;\ P_3$	$a>0$	"
	$S_{15,39}$	$D-2aK_1;\ P_3$	$a>0$	"
	$S_{15,40}$	$D+2L_3+2K_1;\ P_3$		"

TABLE III. Three-dimensional subalgebras.

Class	Notation	Generators	Range of parameters	Invariants
$3A_1$	$S_{9,12}$	$D,\ L_3,\ K_3;$		all generators
	$S_{10,4}$	$L_2+K_1,\ L_1-K_2,\ P_0-P_3;$		"
	$S_{10,10}$	$L_2+K_1,\ L_1-K_2+P_2,\ P_0-P_3;$		"
	$S_{10,18}$	$D,\ L_2+K_1,\ L_1-K_2;$		"

TABLE III. (*Continued*).

Class	Notation	Generators	Range of parameters	Invariants
	$S_{12,5}$	$L_3,\ P_0,\ P_3;$		all generators
	$S_{12,39}$	$D-2aK_3,\ L_3,\ P_0-P_3;$	$a=-1$	"
	$S_{13,6}$	$K_3,\ P_1,\ P_2;$		"
	$S_{14,4}$	$L_2+K_1,\ P_0-P_3,\ P_2;$		"
	$S_{14,12}$	$2L_2+2K_1-P_0-P_3,\ P_0-P_3,\ P_2;$		"
	$S_{15,2}$	$P_0-P_3,\ P_2,\ P_1;$		"
	$S_{15,3}$	$P_1,\ P_2,\ P_3;$		"
	$S_{15,4}$	$P_0,\ P_1,\ P_2;$		"
$A_1\oplus A_2$	$S_{8,8}$	$(P_2)\oplus(K_3;L_2+K_1)$		P_2
	$S_{8,26}$	$(D)\oplus(K_3;L_2+K_1)$		D
	$S_{9,5}$	$(L_3)\oplus(K_3;P_0-P_3)$		L_3
	$S_{11,11}$	$(D-2\cot cL_3+2K_3)\oplus(\cot cL_3-K_3;P_0-P_3)$	$0<c<\pi,\ c\neq\pi/2$	$D-2\cot cL_3+2K_3$
	$S_{11,17}$	$(D+2(a-\cot c)L_3+2K_3)$	$a\neq 0$	
		$\oplus(\cot cL_3-K_3;P_0-P_3)$	$0<c<\pi,\ c\neq\pi/2$	$D-2(a\cot c)L_3+2K_3$
	$S_{12,29}$	$(L_3)\oplus(D;P_0-P_3)$		L_3
	$S_{12,30}$	$(L_3)\oplus(D;P_3)$		L_3
	$S_{12,31}$	$(L_3)\oplus(D;P_0)$		L_3
	$S_{12,39}$	$(L_3)\oplus(D-2aK_3;P_0-P_3)$	$a\neq 0,-1$	L_3
	$S_{12,40}$	$(L_3)\oplus(2D-4K_3+P_0+P_3;P_0-P_3)$		L_3
	$S_{12,45}$	$(4L_3+P_0+P_3)\oplus(2D-4K_3+x[P_0+P_3];P_0-P_3)$	$-\infty<x<\infty$	$4L_3+P_0+P_3$
	$S_{13,5}$	$(P_2)\oplus(K_3;P_0-P_3)$		P_2
	$S_{13,11}$	$(P_1)\oplus(2K_3-P_2;P_0-P_2)$		P_1
	$S_{13,22}$	$(D+2K_3)\oplus(D-2K_3;P_0-P_3)$		$D+2K_3$
	$S_{13,23}$	$(K_3)\oplus(D;P_2)$		K_3
	$S_{13,29}$	$(K_3)\oplus(D+2aL_3;P_0-P_3)$	$a\neq 0$	K_3
	$S_{14,28}$	$(L_2+K_1)\oplus(D;P_0-P_3)$		L_2+K_1
	$S_{14,29}$	$(L_2+K_1)\oplus(D;P_2)$		L_2+K_1
	$S_{14,45}$	$(P_0-P_3)\oplus(D-2aK_3;L_2+K_1)$	$a=-1$	P_0-P_3
	$S_{14,46}$	$(L_2+K_1)\oplus(D+L_1-K_2;P_0-P_3)$		L_2+K_1
	$S_{15,31}$	$(P_0-P_3)\oplus(D-2aK_3;P_2)$	$a=-1$	P_0-P_3
	$S_{15,33}$	$(P_0-P_3)\oplus(D+2a[\cos cL_3-\sin cK_3];P_0-P_3)$	$a=1,\ c=\pi/2$	P_0-P_3
$A_{3,1}$	$S_{10,11}$	$L_2+K_1-P_2,\ L_1-K_2+bP_2-P_1;P_0-P_3$	$b\neq 0$	P_0-P_3
	$S_{10,12}$	$L_2+K_1-P_2,\ L_1-K_2-P_1;P_0-P_3$		"
	$S_{14,5}$	$L_2+K_1,\ P_1;\ P_0-P_3$		"
	$S_{14,6}$	$L_2+K_1,\ P_2-bP_1;P_0-P_3$	$b\neq 0$	"
	$S_{14,13}$	$2L_2+2K_1-P_0-P_3,\ P_1;P_0-P_3$		"
	$S_{14,14}$	$L_2+K_1-P_2,\ P_1;\ P_0-P_3$		"
	$S_{14,15}$	$2L_2+2K_1-P_0-P_3,\ P_2-bP_1;\ P_0-P_3$	$b=0$	"
	$S_{14,16}$	$L_2+K_1-P_2,\ P_2-bP_1;\ P_0-P_3$	$b\neq 0$	"
$A_{3,2}$	$S_{8,15}$	$2K_3+P_1;\ L_2+K_1,\ P_0-P_3$		$(P_0-P_3)\exp[(L_2+K_1)/(P_3-P_0)]$
	$S_{8,16}$	$2K_3-P_2+bP_1;\ L_2+K_1,\ P_0-P_3$	$b\neq 0$	$(P_0-P_3)\exp[(L_2+K_1)/b(P_3-P_0)]$
	$S_{15,32}$	$D-\cos c(L_2+K_1)+\sin c(L_1-K_2);\ P_2,\ P_0-P_3$	$0<c<\pi$	$(P_0-P_3)\exp[P_2/(P_3-P_0)]$
$A_{3,3}$	$S_{7,5}$	$K_3;\ L_2+K_1,\ L_1-K_2$		$(L_1-K_2)/(L_2+K_1)$
	$S_{8,7}$	$K_3;L_2+K_1,\ P_0-P_3$		$(L_2+K_1)/(P_0-P_3)$
	$S_{8,14}$	$2K_3-P_2;\ L_2+K_1,\ P_0-P_3$		"
	$S_{10,24}$	$D+2aL_3-2bK_3;\ L_2+K_1,\ L_1-K_2$	$a=0,\ b\neq 0$	$(L_2+K_1)/(L_1-K_2)$
	$S_{10,25}$	$2D+4aL_3+4K_3+P_0-P_3;\ L_2+K_1,L_1-K_2$	$a=0$	"
	$S_{10,32}$	$D-2K_3;\ L_2+K_1,\ L_1-K_2+P_2$		$(L_1-K_2+P_2)/(L_2+K_1)$
	$S_{15,16}$	$D;\ P_2,\ P_0-P_3$		$(P_0-P_3)/P_2$
	$S_{15,17}$	$D;\ P_0,\ P_3$		$(P_0-P_3)/(P_0+P_3)$
	$S_{15,18}$	$D;\ P_1,\ P_2$		P_1/P_2
	$S_{15,32}$	$D-\cos c(L_2+K_1)+\sin c(L_1-K_2);\ P_2,\ P_0-P_3$	$c=0$	$(P_0-P_3)/P_2$
$A_{3,4}$	$S_{11,4}$	$L_3-\tan cK_3;\ P_0,\ P_3$	$0<c<\pi,\ c\neq\pi/2$	$P_0^2-P_3^2$
	$S_{13,4}$	$K_3;\ P_0,\ P_3$		"
	$S_{13,12}$	$2K_3-P_2;\ P_0,\ P_3$		"
	$S_{14,48}$	$2D+4K_3+P_0-P_3;\ L_2+K_1,\ P_2$		$P_2(L_2+K_1)$
$A_{3,5}^h$				
$h=a/(1+a)$	$S_{14,45}$	$D-2aK_3;\ L_2+K_1,\ P_0-P_3$	$a\neq 0,\ -1$	$(L_2+K_1)^{1+a}/(P_0-P_3)^a$
$=a$	$S_{14,47}$	$D-2aK_3;L_2+K_1,\ P_2$	$a\neq 0$	$(L_2+K_1)/P_2^a$
$=\frac{1}{3}$	$S_{14,59}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3;\ P_0-P_3$		$(2L_2+2K_1-P_0-P_3)^3/(P_0-P_3)$
$=\frac{1}{2}$	$S_{14,60}$	$D-2K_3;\ L_2+K_1-P_2;\ P_0-P_3$		$(L_2+K_1-P_2)^2/(P_0-P_3)$
$=\frac{1}{2}$	$S_{14,61}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3;\ P_2$		$(2L_2+2K_1-P_0-P_3)^2/P_2$
$=1+a$	$S_{15,31}$	$D-2aK_3;\ P_2,\ P_0-P_3$	$a\neq 0,\ -1$	$(P_0-P_3)^{1+a}/P_2$
$=\dfrac{1-a\sin c}{1+a\sin c}$	$S_{15,33}$	$D+2a\cos cL_3-2a\sin cK_3;\ P_0+P_3,\ P_0-P_3$	$0\leq c<2\pi,\ c\neq 3\pi/2$ $a>0,\ a\neq 1$	$(P_0+P_3)^{1+a\sin c}/(P_0-P_3)^{1-a\sin c}$

Class	Notation	Generators	Range of parameters	Invariants
	$S_{12,5}$	$L_3,\ P_0,\ P_3;$		all generators
	$S_{12,39}$	$D-2aK_3,\ L_3,\ P_0-P_3;$	$a=-1$	''
	$S_{13,6}$	$K_3,\ P_1,\ P_2;$		''
	$S_{14,4}$	$L_2+K_1,\ P_0-P_3,\ P_2;$		''
	$S_{14,12}$	$2L_2+2K_1-P_0-P_3,\ P_0-P_3,\ P_2;$		''
	$S_{15,2}$	$P_0-P_3,\ P_2,\ P_1;$		''
	$S_{15,3}$	$P_1,\ P_2,\ P_3;$		''
	$S_{15,4}$	$P_0,\ P_1,\ P_2;$		''
$A_1 \oplus A_2$	$S_{8,8}$	$(P_2)\oplus(K_3;L_2+K_1)$		P_2
	$S_{8,26}$	$(D)\oplus(K_3;L_2+K_1)$		D
	$S_{9,5}$	$(L_3)\oplus(K_3;P_0-P_3)$		L_3
	$S_{11,11}$	$(D-2\cot cL_3+2K_3)\oplus(\cot cL_3-K_3;P_0-P_3)$	$0<c<\pi,\ c\neq\pi/2$	$D-2\cot cL_3+2K_3$
	$S_{11,17}$	$(D+2(a-\cot c)L_3+2K_3)$ $\oplus(\cot cL_3-K_3;P_0-P_3)$	$a\neq 0$ $0<c<\pi,\ c\neq\pi/2$	$D-2(a\cot c)L_3+2K_3$
	$S_{12,29}$	$(L_3)\oplus(D;P_0-P_3)$		L_3
	$S_{12,30}$	$(L_3)\oplus(D;P_3)$		L_3
	$S_{12,31}$	$(L_3)\oplus(D;P_0)$		L_3
	$S_{12,39}$	$(L_3)\oplus(D-2aK_3;P_0-P_3)$	$a\neq 0,-1$	L_3
	$S_{12,40}$	$(L_3)\oplus(2D-4K_3+P_0+P_3;P_0-P_3)$		L_3
	$S_{12,45}$	$(4L_3+P_0+P_3)\oplus(2D-4K_3+x[P_0+P_3];P_0-P_3)$	$-\infty<x<\infty$	$4L_3+P_0+P_3$
	$S_{13,5}$	$(P_2)\oplus(K_3;P_0-P_3)$		P_2
	$S_{13,11}$	$(P_1)\oplus(2K_3-P_2;P_0-P_2)$		P_1
	$S_{13,22}$	$(D+2K_3)\oplus(D-2K_3;P_0-P_3)$		$D+2K_3$
	$S_{13,23}$	$(K_3)\oplus(D;P_2)$		K_3
	$S_{13,29}$	$(K_3)\oplus(D+2aL_3;P_0-P_3)$	$a\neq 0$	K_3
	$S_{14,28}$	$(L_2+K_1)\oplus(D;P_0-P_3)$		L_2+K_1
	$S_{14,29}$	$(L_2+K_1)\oplus(D;P_2)$		L_2+K_1
	$S_{14,45}$	$(P_0-P_3)\oplus(D-2aK_3;L_2+K_1)$	$a=-1$	P_0-P_3
	$S_{14,46}$	$(L_2+K_1)\oplus(D+L_1-K_2;P_0-P_3)$		L_2+K_1
	$S_{15,31}$	$(P_0-P_3)\oplus(D-2aK_3;P_2)$	$a=-1$	P_0-P_3
	$S_{15,33}$	$(P_0-P_3)\oplus(D+2a[\cos cL_3-\sin cK_3];P_0-P_3)$	$a=1,\ c=\pi/2$	P_0-P_3
$A_{3,1}$	$S_{10,11}$	$L_2+K_1-P_2,\ L_1-K_2+bP_2-P_1;P_0-P_3$	$b\neq 0$	P_0-P_3
	$S_{10,12}$	$L_2+K_1-P_2,\ L_1-K_2-P_1;P_0-P_3$		''
	$S_{14,5}$	$L_2+K_1,\ P_1;\ P_0-P_3$		''
	$S_{14,6}$	$L_2+K_1,\ P_2-bP_1;P_0-P_3$	$b\neq 0$	''
	$S_{14,13}$	$2L_2+2K_1-P_0-P_3,\ P_1;P_0-P_3$		''
	$S_{14,14}$	$L_2+K_1-P_2,\ P_1;\ P_0-P_3$		''
	$S_{14,15}$	$2L_2+2K_1-P_0-P_3,\ P_2-bP_1;\ P_0-P_3$	$b=0$	''
	$S_{14,16}$	$L_2+K_1-P_2,\ P_2-bP_1;\ P_0-P_3$	$b\neq 0$	''
$A_{3,2}$	$S_{8,15}$	$2K_3+P_1;\ L_2+K_1,\ P_0-P_3$		$(P_0-P_3)\exp[(L_2+K_1)/(P_3-P_0)]$
	$S_{8,16}$	$2K_3-P_2+bP_1;\ L_2+K_1,\ P_0-P_3$	$b\neq 0$	$(P_0-P_3)\exp[(L_2+K_1)/b(P_3-P_0)]$
	$S_{15,32}$	$D-\cos c(L_2+K_1)+\sin c(L_1-K_2);\ P_2,\ P_0-P_3$	$0<c<\pi$	$(P_0-P_3)\exp[P_2/(P_3-P_0)]$
$A_{3,3}$	$S_{7,5}$	$K_3;\ L_2+K_1,\ L_1-K_2$		$(L_1-K_2)/(L_2+K_1)$
	$S_{8,7}$	$K_3;L_2+K_1,\ P_0-P_3$		$(L_2+K_1)/(P_0-P_3)$
	$S_{8,14}$	$2K_3-P_2;\ L_2+K_1,\ P_0-P_3$		''
	$S_{10,24}$	$D+2aL_3-2bK_3;\ L_2+K_1,\ L_1-K_2$	$a=0,\ b\neq 0$	$(L_2+K_1)/(L_1-K_2)$
	$S_{10,25}$	$2D+4aL_3+4K_3+P_0-P_3;\ L_2+K_1,L_1-K_2$	$a=0$	''
	$S_{10,32}$	$D-2K_3;\ L_2+K_1,\ L_1-K_2+P_2$		$(L_1-K_2+P_2)/(L_2+K_1)$
	$S_{15,16}$	$D;\ P_2,\ P_0-P_3$		$(P_0-P_3)/P_2$
	$S_{15,17}$	$D;\ P_0,\ P_3$		$(P_0-P_3)/(P_0+P_3)$
	$S_{15,18}$	$D;\ P_1,\ P_2$		P_1/P_2
	$S_{15,32}$	$D-\cos c(L_2+K_1)+\sin c(L_1-K_2);\ P_2,\ P_0-P_3$	$c=0$	$(P_0-P_3)/P_2$
$A_{3,4}$	$S_{11,4}$	$L_3-\tan cK_3;\ P_0,\ P_3$	$0<c<\pi,\ c\neq\pi/2$	$P_0^2-P_3^2$
	$S_{13,4}$	$K_3;\ P_0,\ P_3$		''
	$S_{13,12}$	$2K_3-P_2;\ P_0,\ P_3$		''
	$S_{14,48}$	$2D+4K_3+P_0-P_3;\ L_2+K_1,\ P_2$		$P_2(L_2+K_1)$
$A_{3,5}^h$				
$h=a/(1+a)$	$S_{14,45}$	$D-2aK_3;\ L_2+K_1,\ P_0-P_3$	$a\neq 0,\ -1$	$(L_2+K_1)^{1+a}/(P_0-P_3)^a$
$=a$	$S_{14,47}$	$D-2aK_3;L_2+K_1,\ P_2$	$a\neq 0$	$(L_2+K_1)/P_2^a$
$=\frac{1}{3}$	$S_{14,59}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3;\ P_0-P_3$		$(2L_2+2K_1-P_0-P_3)^3/(P_0-P_3)$
$=\frac{1}{2}$	$S_{14,60}$	$D-2K_3;\ L_2+K_1-P_2;\ P_0-P_3$		$(L_2+K_1-P_2)^2/(P_0-P_3)$
$=\frac{1}{2}$	$S_{14,61}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3;\ P_2$		$(2L_2+2K_1-P_0-P_3)^2/P_2$
$=1+a$	$S_{15,31}$	$D-2aK_3;\ P_2,\ P_0-P_3$	$a\neq 0,\ -1$	$(P_0-P_3)^{1+a}/P_2$
$=\dfrac{1-a\sin c}{1+a\sin c}$	$S_{15,33}$	$D+2a\cos cL_3-2a\sin cK_3;\ P_0+P_3,\ P_0-P_3$	$0\leq c<2\pi,\ c\neq 3\pi/2$ $a>0,\ a\neq 1$	$(P_0+P_3)^{1+a\sin c}/(P_0-P_3)^{1-a\sin c}$

TABLE IV. (*Continued*).

Class	Notation	Generators	Range of parameters	Invariants
$A_1\oplus A^h_{3,5}$				
$h=\frac{1+a}{1-a}$	$S_{12,36}$	$(L_3)\oplus(D-2aK_3;\,P_0,P_3)$	$a>0$	$L_3;\ (P_0-P_3)^{1+a}/(P_0+P_3)^{1-a}$
$=\frac{1}{2}$	$S_{15,29}$	$(P_0+P_2)\oplus(D-2aK_2,P_1,P_0-P_2)$	$a=1$	$P_0+P_2;\ (P_0-P_2)^2/P_1$
$A_1\oplus A_{3,6}$	$S_{6,3}$	$(P_0-P_3)\oplus(L_3;L_2+K_1,L_1-K_2)$		$P_0-P_3;\ (L_2+K_1)^2+(L_1-K_2)^2$
	$S_{6,11}$	$(D)\oplus(L_3;L_2+K_1,L_1-K_2)$		$D;$ "
	$S_{9,3}$	$(K_3)\oplus(L_3;\,P_1,P_2)$		$K_3;\ P_1^2+P_2^2$
	$S_{12,2}$	$(P_0-P_3)\oplus(L_3;P_1,P_2)$		$P_0-P_3;$ "
	$S_{12,3}$	$(P_3)\oplus(L_3;P_2,P_1)$		$P_3;\ P_1^2+P_2^2$
	$S_{12,4}$	$(P_0)\oplus(L_3;P_2,P_1)$		$P_0;$ "
	$S_{12,11}$	$(P_0-P_3)\oplus(4L_3+P_0+P_3;P_2,P_1)$		$P_0-P_3;\ P_1^2+P_2^2$
	$S_{12,12}$	$(P_3)\oplus(2L_3+P_0;P_2;P_1)$		$P_3;\ P_1^2+P_2^2$
	$S_{12,13}$	$(P_0)\oplus(2L_3-P_3;P_2,P_1)$		P_0 "
$A_1\oplus A^P_{3,7}$				
$P=\tan c$	$S_{5,8}$	$(D)\oplus(L_3-\tan cK_3;L_2+K_1,L_1-K_2)$	$0<c<\pi,\ c\neq\pi/2$	$D;\ \{(L_1-K_2)^2+(L_2+K_1)^2 \times\left(\frac{L_1-K_2-i(L_2+K_1)}{L_1-K_2+i(L_2+K_1)}\right)^{i\tan c}$
$=1/a$	$S_{10,23}$	$(P_0-P_3)\oplus(D+2aL_3-2bK_3;\ L_2+K_1,\ L_1-K_2)$	$b=-1,\ a>0$	$P_0-P_3;$ same with $\tan c\to 1/a$
$=1/a$	$S_{13,28}$	$(K_3)\oplus(D+2aL_3;P_1,P_2)$	$a>0$	$K_3;\ (P_1^2+P_2^2)\,\{(P_1-iP_2)/(P_1+iP_2)\}^{i/a}$
$A_1\oplus A_{3,8}$	$S_{4,3}$	$(P_3)\oplus(;L_3,K_1,K_2)$		$P_3;\ L_3^2-K_1^2-K_2^2$
	$S_{4,8}$	$(D)\oplus(;L_3,K_1,K_2)$		$D;\ L_3^2-K_1^2-K_2^2$
$A_1\oplus A_{3,9}$	$S_{3,3}$	$(P_0)\oplus(;L_3,L_1,L_2)$		$P_0;\ L_1^2+L_2^2+L_3^2$
	$S_{3,8}$	$(D)\oplus(;L_3,L_1,L_2)$		$D;\ L_1^2+L_2^2+L_3^2$
$A_{4,1}$	$S_{10,7}$	$L_1-K_2-P_1,\ 2L_2+2K_1-P_0-P_3;\ P_2,\ P_0-P_3$		$P_0-P_3;\ P_2^2+(P_0-P_3)(2L_2+2K_1-P_0-P_3+2P_2)$
	$S_{10,8}$	$L_1-K_2,\ 2L_2+2K_1-P_0-P_3;\ P_2,\ P_0-P_3$		$P_0-P_3;\ P_2^2+(P_0-P_3)(2L_2+2K_1-P_0-P_3)$
	$S_{14,3}$	$L_2+K_1,\ P_0+P_3;\ P_1,\ P_0-P_3$		$P_0-P_3;\ P_1^2+P_3^2-P_0^2$
	$S_{14,11}$	$L_2+K_1-P_2,\ P_0+P_3;\ P_1,\ P_0-P_3$		" ; "
$A^\alpha_{4,2}$				
$\alpha=1$	$S_{7,7}$	$2K_3-P_2;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$		$(L_1-K_2)/(P_0-P_3);(P_0-P_3)\times\exp((2L_2+2K_1)/(P_0-P_3))$
$=1$	$S_{15,26}$	$D-L_2-K_1;\ P_1,\ P_0-P_3,\ P_2$		$P_2/(P_0-P_3);\ (P_0-P_3)\exp(2P_1/(P_0-P_3))$
$A_{4,4}$	$S_{15,30}$	$D+2L_3-2K_1;\ P_0,\ P_1,\ P_2$		$(P_0-P_2)\exp(P_1/(P_0-P_2));\ (P_0^2-P_2^2-P_1^2)/(P_0-P_2)^2$
$A^{B,\gamma}_{4,5}$				
$B=1\quad\gamma=1$	$S_{7,4}$	$K_3;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$		$(P_0-P_3)/(L_1-K_2);\ (P_0-P_3)/(L_2+K_1)$
$=\frac{b}{1+b}\quad=\frac{b}{1+b}$	$S_{10,23}$	$D+2aL_3-2bK_3;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$	$a=0,\ b\neq 0,\ -1$	$(P_0-P_3)^b/(L_1-K_2)^{1+b};\ (P_0-P_3)^b/(L_2+K_1)^{1+b}$
$=\frac{1}{2}\quad=\frac{1}{2}$	$S_{10,29}$	$D-2K_3;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$		$(P_0-P_3)/(L_1-K_2)^2;\ (P_0-P_3)/(L_2+K_1)^2$
$B=\frac{1}{1+a}\ \gamma=\frac{a}{1+a}$	$S_{14,38}$	$D-2aK_3;\ L_2+K_1,\ P_2,\ P_0-P_3$	$a\neq 0,\ -1$	$(P_0-P_3)/P_2^{1+a};\ (P_0-P_3)^a/(L_2+K_1)^{1+a}$
$=\frac{2}{3}\quad=\frac{1}{3}$	$S_{14,54}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3,\ P_2,\ P_0-P_3$		$(P_0-P_3)^2/P_2^3;\ (P_0-P_3)/(2L_2+2K_1-P_0-P_3)^3$
$=1\quad=1$	$S_{15,13}$	$D;\ P_1,\ P_2,\ P_3-P_0$		$(P_0-P_3)/P_2;\ (P_0-P_3)/P_1$
$=1\quad=1$	$S_{15,14}$	$D;\ P_1,\ P_2,\ P_3$		$P_3/P_2;\ P_3/P_1$
$=1\quad=1$	$S_{15,15}$	$D;\ P_1,\ P_2,\ P_0$		$P_0/P_2;\ P_0/P_1$
$=\frac{1}{1+a}\quad=\frac{1}{1+a}$	$S_{15,25}$	$D-2a\cos cL_3+2a\sin cK_3;\ P_1,\ P_2,\ P_0-P_3$	$c=\pi/2,\ a>0$	$(P_0-P_3)/P_2^{1+a};\ (P_0-P_3)/P_1^{1+a}$
$=\frac{1}{1-a}\quad=\frac{1}{1-a}$	$S_{15,25}$	$D-2a\cos cL_3+2a\sin cK_3;\ P_1,\ P_2,\ P_0-P_3$	$c=3\pi/2,\ a>0,\ a\neq 1$	$(P_0-P_3)/P_2^{1-a};\ (P_0-P_3)/P_1^{1-a}$
$=\frac{1+a}{1-a}\quad=\frac{1}{1-a}$	$S_{15,29}$	$D-2aK_2;\ P_1,\ P_2,\ P_0$	$a>0$	$(P_0+P_2)^{1+a}/(P_0-P_2)^{1-a};\ (P_0+P_2)/P_1^{1-a}$

TABLE IV. (*continued*).

Class	Notation	Generators	Range of parameters	Invariants
$A_{4,6}^{\alpha,P}$				
$\alpha, P=\tan c$	$S_{5,3}$	$L_3-\tan c K_3;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$	$0<c<\pi,\ c\neq\pi/2$	$(P_0-P_3)^2/\{(L_2+K_1)^2+(L_1-K_2)^2\}$; $\{(L_2+K_1)^2+(L_1-K_2)^2\}\times\left(\frac{L_1-K_2-i(L_2+K_1)}{L_1-K_2+i(L_2+K_1)}\right)^{i\tan c}$
$\alpha=\frac{1+b}{a}\ P=\frac{b}{a}$	$S_{10,23}$	$D+2aL_3-2bK_3;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$	$a\neq 0,\ b\neq -1$	$\frac{(P_0-P_3)^{2b/(1+b)}}{(L_1-K_2)^2+(L_2+K_1)^2}$; $\{(L_1-K_2)^2+(L_2+K_1)^2\}\times\left(\frac{L_1-K_2-i(L_2+K_1)}{L_1-K_2+i(L_2+K_1)}\right)^{ib/a}$
$\alpha=\tan c\ P=0$	$S_{11,2}$	$L_3-\tan c K_3;\ P_1,\ P_2,\ P_0-P_3$	$0<c<\pi,\ c\neq\pi/2$	$P_1^2+P_2^2;\ (P_1^2+P_2^2)[(P_1-iP_2)/(P_1+iP_2)]^{i\tan c}$
$\alpha=1/a\cos c$ $P=\frac{1+a\sin c}{a\cos c}$	$S_{15,25}$	$D+2a\cos c L_3-2b\sin c K_3;\ P_1,\ P_2,\ P_0-P_3$	$a>0,\ c\neq\pi/2,\ 3\pi/2$ $0\leq c<2\pi$	$\frac{(P_0-P_3)^2}{(P_2^2+P_1^2)^{1+a\sin c}};\ (P_1^2+P_2^2)\left(\frac{P_1-iP_2}{P_1+iP_2}\right)^{i/a\cos c}$
$\alpha=\frac{1}{a}\ P=\frac{1}{a}$	$S_{15,27}$	$D+2aL_3;\ P_1,\ P_2,\ P_3$	$a>0$	$P_3^2/(P_1^2+P_2^2);\ (P_1^2+P_2^2)[(P_1-iP_2)/(P_1+iP_2)]^{i/a}$
$\alpha=\frac{1}{a}\ P=\frac{1}{a}$	$S_{15,28}$	$D+2aL_3;\ P_1,\ P_2,\ P_0$	$a>0$	$P_0^2/(P_1^2+P_2^2)$; "
$A_{4,7}$	$S_{14,41}$	$2D-4K_3+P_0+P_3;\ L_2+K_1,\ P_1,\ P_0-P_3$		none
	$S_{14,56}$	$2D-4K_3+x(P_0-P_3);\ L_2+K_1-P_2,\ P_0-P_3$	$-\infty<x\infty,\ x\neq 0$	"
$A_{4,8}$	$S_{14,40}$	$D-2aK3;\ L_2+K_1,\ P_1,\ P_0-P_3$	$a=-1$	$(D+2K_3)(P_3-P_0)+2P_1(L_2+K_1);\ P_0-P_3$
	$S_{14,43}$	$D-2aK_3;\ L_2+K_1,\ P_2-cP_1,\ P_0-P_3$	$a=-1,\ c\neq 0$	$c(D+2K_3)(P_3-P_0)+2(cP_1-P_2)(L_2+K_1);\ P_0-P_3$
	$S_{14,58}$	$D-2bK_3+a(b-1)(P_0-P_3);\ L_2+K_1-P_2,\ P_2-aP_1,\ P_0-P_3$	$a\neq 0,\ b=-1$	$a[D+2K_3-2a(P_0+P_3)](P_3-P_0)+2(aP_1-P_2)(L_2+K_1-aP_1);\ P_0-P_3$
$A_{4,9}^{h}$				
$h=0$	$S_{8,5}$	$K_3,\ P_1;\ L_2+K_1,\ P_0-P_3$		none
$=0$	$S_{8,6}$	$K_3,\ P_2-bP_1,\ L_2+K_1,\ P_0-P_3$	$b\neq 0$	"
$=0$	$S_{8,12}$	$2K_3-P_2,\ P_1;\ L_2+K_1,\ P_0-P_3$		"
$=0$	$S_{8,13}$	$2K_3-P_2,\ P_2-bP_1;\ L_2+K_1,\ P_0-P_3$	$b\neq 0$	"
$=1$	$S_{10,30}$	$D-2K_3;\ L_2+K_1-P_2,\ L_1-K_2+bP_2-P_1,\ P_0-P_3$	$b\neq 0$	"
$=1$	$S_{10,31}$	$D-2aL_3-2K_3;\ L_2+K_1-P_2,\ L_1-K_2-P_1,\ P_0-P_3$	$a=0$	"
$=0$	$S_{14,26}$	$D,\ L_2+K_1;\ P_1,\ P_0-P_3$		"
$=0$	$S_{14,27}$	$D,\ L_2+K_1;\ P_2-bP_1,\ P_0-P_3$	$b\neq 0$	"
$=a$	$S_{14,40}$	$D-2aK_3;\ L_2+K_1,\ P_1,\ P_0-P_3$	$a\neq 0,\ -1$	"
$=0$	$S_{14,42}$	$D+L_1-K_2,\ L_2+K_1;\ P_1,\ P_0-P_3$		"
$=a$	$S_{14,43}$	$D-2aK_3;\ L_2+K_1,\ P_2-cP_1,\ P_0-P_3$	$a\neq 0,\ -1,\ c\neq 0$	"
$=0$	$S_{14,44}$	$D+L_1-K_2,\ L_2+K_1;\ P_2-cP_1,\ P_0-P_3$	$c\neq 0$	"
$=\frac{1}{2}$	$S_{14,45}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3,\ P_1\ P_0-P_3$		"
$=1$	$S_{14,56}$	$2D-4K_3+x(P_0-P_3);\ L_2+K_1-P_2,\ P_1,\ P_0-P_3$	$x=0$	"
$=\frac{1}{2}$	$S_{14,57}$	$2D-2K_3+2b(L_1-K_2)+ba(P_0-P_3);\ 2L_2+2K_1+P_0+P_3\ P_2-aP_1,\ P_0-P_3$	$a\neq 0$ $-\infty<b<\infty$	"
$=b$	$S_{14,58}$	$D-2bK_3+a(b-1)(P_0-P_3);\ L_2+K_1-P_2,\ P_2-aP_1,\ P_0-P_3$	$a\neq 0,\ b\neq -1$ $-\infty<b<\infty$	"
$A_{4,10}$	$S_{6,6}$	$L_3;\ L_2+K_1-P_2;\ L_1-K_2-P_1,\ P_0-P_3$		$P_0-P_3;\ 2(P_0-P_3)L_3-(L_1+K_2-P_2)^2-(L_1-K_2-P_1)^2$
	$S_{10,31}$	$D-2aL_3-2K_3;\ L_2+K_1-P_2,\ L_1-K_2-P_1,\ P_0-P_3$	$a=0$	$P_0-P_3;\ 2(P_0-P_3)(D-2K_3)-(L_1+K_2-P_2)^2-(L_1-K_2-P_1)^2$
$A_{4,11}^{P}$				
$P=\frac{1}{a}$	$S_{10,31}$	$D+2aL_3-2K_3;\ L_2+K_1-P_2,\ L_1-K_2-P_1,\ P_0-P_3$	$a\neq 0$	none

TABLE IV. (*continued*).

Class	Notation	Generators	Range of parameters	Invariants
$A_{4,12}$	$S_{2,4}$	$L_3,\ K_3;\ L_2+K_1,\ L_1-K_2$		none
	$S_{5,12}$	$D-2aK_3,\ L_3-\tan c K_3;\ L_2+K_1,\ L_1-K_2$	$a\neq 0$ $0<c<\pi,\ c\neq\pi/2$	"
	$S_{6,16}$	$D-2aK_3,\ L_3;\ L_2+K_1,\ L_1-K_2$	$a\neq 0$	"
	$S_{6,17}$	$2D+4K_3+(P_0-P_3),\ L_3;\ L_2+K_1,\ L_1-K_2$		"
	$S_{6,20}$	$2D+4K_3+x(P_0-P_3),\ 4L_3+P_0-P_3;\ L_2+K_1,\ L_1-K_2$	$-\infty<x<\infty$	"
	$S_{7,17}$	$D+2aL_3,\ K_3;\ L_2+K_1,\ L_1-K_2$	$a\neq 0$	"
	$S_{11,9}$	$L_3-\tan c K_3,\ D;\ P_1,\ P_2$	$0<c<\pi,\ c\neq\pi/2$	"
	$S_{11,15}$	$D+2aL_3,\ L_3-\tan c K_3;\ P_1,\ P_2$	$a\neq 0$ $0<c<c\pi,\ c\neq\pi/2$	"
	$S_{12,28}$	$D,\ L_3;\ P_1,\ P_2$		"
	$S_{12,37}$	$D-2aK_3,\ L_3;\ P_1,\ P_2$	$a>0$	"
	$S_{12,38}$	$2D-4K_3+P_0+P_3;\ L_3;\ P_1,\ P_2$		"
	$S_{12,44}$	$2D+4K_3+x(P_0+P_3);\ 4L_3+P_0+P_3,\ P_1,\ P_2$	$-\infty<x<\infty$	"

TABLE V. Five-dimensional subalgebras.

Class	Notation	Generators	Range of parameters	Invariants
$A_1\oplus 2A_2$	$S_{9,10}$	$(L_3)\oplus(D+2K_3;\ P_0+P_3)\oplus(D-2K_3;\ P_0-P_3)$		L_3
$A_2\oplus A_{3,3}$	$S_{7,11}$	$(D;\ P_0-P_3)\oplus(D+2K_3;\ L_1-K_2,\ L_2+K_1)$		$(L_2+K_1)/(L_1-K_2)$
$2A_1\oplus A_{3,4}$	$S_{13,1}$	$(P_1)\oplus(P_2)\oplus(K_3;\ P_0+P_3,\ P_0-P_3)$		$(P_0-P_3)/(P_0+P_3),\ P_1,\ P_2$
$2A_1\oplus A_{3,6}$	$S_{12,1}$	$(P_0)\oplus(P_3)\oplus(L_3;\ P_1,\ P_2)$		$P_1^2+P_2^2$
$A_2\oplus A_{3,6}$	$S_{6,10}$	$(D,\ P_0-P_3)\oplus(L_3;\ L_1-K_2,\ L_2+K_1)$		$(L_1-K_2)^2+(L_2+K_1)^2$
	$S_{9,2}$	$(K_3;\ P_0-P_3)\inplus(L_3;\ P_1,\ P_2)$		$P_1^2+P_2^2$
$A_2\oplus A_{3,7}^p$ $p=\tan c$	$S_{5,7}$	$(D;\ P_0-P_3)\oplus(2L_3-\tan c[2K_3+D];\ L_2+K_1,\ L_1-K_2)$	$0<c<\pi,\ c\neq\pi/2$	$\{(L_1-K_2)^2+(L_2+K_1)^2\}\left[\dfrac{L_1-K_2-i(L_2+K_1)}{L_1-K_2+i(L_2+K_1)}\right]^{i\tan c}$
$p=1/a$	$S_{13,26}$	$(K_3;\ P_0-P_3)\oplus(D+2K_3+2aL_3;\ P_1,P_2$	$a\neq 0$	$(P_1^2+P_2^2)\{(P_1-iP_2)/(P_1+iP_2)\}^{i/a}$
$A_2\oplus A_{3,8}$	$S_{4,7}$	$(D;\ P_3)\oplus(;\ L_3,\ K_1,\ K_2)$		$L_3^2-K_1^2-K^2$
$A_2\oplus A_{3,9}$	$S_{3,7}$	$(D;\ P_0)\inplus(;\ L_1,\ L_2,\ L_3)$		$L_1^2+L_2^2+L_3^2$
$A_1\oplus A_{4,1}$	$S_{14,1}$	$(P_2)\oplus(L_2+K_1;\ P_0-P_3,\ P_1,\ P_0+P_3)$		$P_0-P_3;\ P_2;\ P_1^2-(P_0-P_3)\times(L_2+K_1)$
$A_1\oplus A_{4,5}$	$S_{15,23}$	$(P_0+P_3)\oplus(D+2a(\cos c L_3-\sin c K_3);\ P_0-P_3,\ P_1,\ P_2))$	$c=\pi/2,\ a=1$	$P_0+P_3;\ (P_0-P_3)/P_2^2;\ (P_0-P_3)/P_1^2$
$A_1\oplus A_{4,1}^h$				D
$h=0$	$S_{8,2}$	$(P_2)\oplus(K_3,\ P_1;\ L_2+K_1,\ P_0-P_3)$		P_2
$h=0$	$S_{10,16}$	$(L_2+K_1)\oplus(D,\ L_1-K_2,\ P_2,\ P_0-P_3)$		L_2+K_1
$A_1\oplus A_{4,12}$	$S_{2,8}$	$(D)\oplus(L_3,\ K_3;\ L_2+K_1,\ L_1-K_2)$		D
	$S_{6,15}$	$(P_0-P_3)\oplus(L_3,\ D-2aK_4;\ L_2+K_1,\ L_1-K_2)$	$a=-1$	P_0-P_3
	$S_{9,9}$	$(K_3)\oplus(L_3,\ D;\ P_2,\ P_1)$		K_3
	$S_{12,34}$	$(P_0-P_3)\inplus(L_3,\ D-2aK_3;\ P_2,\ P_1)$	$a=-1$	P_0-P_3
$A_{5,4}$	$S_{10,2}$	$L_2+K_1,\ L_1-K_2,\ P_1,\ P_2;\ P_1,P_2;\ P_0-P_3,$		P_0-P_3
$A_{5,5}$	$S_{10,6}$	$L_1-K_2,\ 2L_2+2K_1-P_0-P_3,\ P_1,P_2;\ P_0-P_3$		P_0-P_3
$A_{5,7}^{\alpha,b,c}$ $\alpha=1,\ c=1$ $b=1$	$S_{15,12}$	$D;\ P_1,\ P_2,\ P_3,\ P_0$		$P_2/(P_0-P_3);\ (P_0-P_3)/P_1,\ (P_0-P_3)/(P_0+P_3)$
$\alpha=1/(1+a)$ $b=1/(1+a)$ $c=(1-a)/(1+a)$	$S_{15,23}$	$D+2a(\cos c L_3-\sin c K_3);\ P_1,\ P_2,\ P_3,\ P_0$	$c=\pi/2$ $a>0,\ a\neq 1$	$(P_0-P_3)/P_2^{1+a};\ (P_0-P_3)/P_1^{1+a};\ (P_0-P_3)^{1-a}/(P_0+P_3)^{1+a}$

TABLE V. (*Continued*).

Class	Notation	Generators	Range of parameters	Invariants
$A^{c}_{5,11}$				
$c=1$	$S_{15,24}$	$D-L_2-K_1;\ P_2,\ P_1,\ P_3,\ P_0$		$(P_0-P_3)/P_2;\ (P_0-P_3)\exp\{(P_0+P_3)/(P_3-P_0)\}$ $4P_1/(P_0-P_3)+(P_0+P_3)^2/(P_0-P_3)^2$
$A^{b,p,q}_{5,13}$				
$b=-1\quad p=0$ $q=\cot c$	$S_{11,1}$	$\cos c L_3-\sin c K_3;\ P_1,\ P_2,\ P_3,\ P_0$	$0<c<\pi,\ c\neq\pi/2$	$P_0^2-P_3^2;\ P_1^2+P_2^2;\ (P_0-P_3)^{2\cot c}\left\{\frac{P_1-iP_2}{P_1+iP_2}\right\}^{i}$
$b=\frac{1-a\sin c}{1+a\sin c}$ $p=\frac{1}{1+a\sin c}$ $q=\frac{a\cos c}{1+a\sin c}$	$S_{15,23}$	$D-2a(\cos c L_3-\sin c K_3);$ $P_1,\ P_2,\ P_3,\ P_0$	$0\leq c<\pi,\ c\neq\pi/2$ $a>0$	$\frac{(P_0-P_3)^b}{P_0+P_3};\ \frac{(P_0-P_3)^{2P}}{P_1^2+P_2^2};$ $(P_0-P_3)^{2q}\left\{\frac{P_1-iP_2}{P_1+iP_2}\right\}^{i}$
$A^{c,b}_{5,19}$				
$c=1\quad b=1$	$S_{7,3}$	$K_3;\ L_2+K_1,\ L_1-K_2,\ P_2,\ P_0-P_3$		$(P_0-P_3)/(L_2+K_1)$
$=a+1\quad =a$	$S_{10,22}$	$D-2aK_3;\ L_2+K_1,\ L_1-K_2,\ P_2,$ P_0-P_3	$a\neq0$	$(P_0-P_3)^a/(L_2+K_1)^{1+a}$
$=2\quad =1$	$S_{10,28}$	$D-2K_3;\ L_2+K_1,\ L_1-K_2-P_1,$ $P_2,\ P_0-P_3$		$(P_0-P_3)/(L_2+K_1)^2$
$=1\quad =1$	$S_{14,23}$	$D;\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$		$(P_0-P_3)/P_2$
$=a+1\quad =1$	$S_{14,33}$	$D-2aK_3;\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$	$a\neq0$	$(P_0-P_3)/P_2^{a+1}$
$=3\quad =2$	$S_{14,52}$	$D-K_3;\ 2L_2+2K_1-P_0-P_3,\ P_1,\ P_2$ P_0-P_3		$(P_0-P_3)^2/P_2^3$
$A^{b}_{5,20}$				
$b=1$	$S_{7,6}$	$2K_3-P_1;\ L_2+K_1,\ L_1-K_2,$ $P_0-P_3,\ P_2$		$(P_0-P_3)\exp\{(L_2+K_1)/(P_3-P_0)\}$
$b=1$	$S_{14,35}$	$D+L_1,\ -K_2;\ L_2+K_1,\ P_1,\ P_2$ P_3-P_0		$(P_0-P_3)\exp\{2P_2/(P_0-P_3)\}$
$A^{b}_{5,23}$				
$b=1$	$S_{14,34}$	$2D-4K_3+P_0+P_3;\ L_2+K_1,\ P_1,\ P_2$ P_0-P_3		$(P_0-P_3)/P_2^2$
$A^{h}_{5,30}$				
$h=0$	$S_{8,3}$	$K_3,\ P_1;\ L_2+K_1,\ P_3,\ P_0$		$P_1^2+P_3^2-P_0^2$
$=0$	$S_{8,10}$	$2K_3-P_2,\ P_1;\ L_2+K_1,\ P_3,\ P_0$		"
$=2$	$S_{19,27}$	$D-K_3;\ L_1-K_2,\ 2L_2+2K_1-P_0$ $-P_3,\ P_2,\ P_0-P_3$		$(P_2^2+P_3^2-P_0^2-2(P_3-P_0)(L_2+K_1))^3/(P_3-P_0)^4$
$=1/a$	$S_{14,37}$	$D-2aK_3;\ L_2+K_1,\ P_1,\ P_3,\ P_0$	$a\neq0$	$(P_1^2+P_3^2-P_0^2)^{1+a}/(P_0-P_3)^2$
$=1$	$S_{14,53}$	$D-2K_3;\ L_2+K_1-P_2,\ P_1,\ P_3,\ P_0$		$(P_1^2+P_3^2-P_0^2)/(P_0-P_3)$
$A^{h}_{5,32}$				
$h=0$	$S_{14,24}$	$D,\ L_2+K_1;\ P_1,\ P_3,\ P_0$		$(P_1^2+P_3^2-P_0^2)/(P_0-P_3)^2$
$A^{a,b}_{5,33}$				
$a=1,\ b=1$	$S_{8,21}$	$K_3,\ D;\ L_2+K_1,\ P_2,\ P_0-P_3$		$P_2(L_2+K_1)/(P_0-P_3)$
$=0\quad =1$	$S_{13,17}$	$K_3,\ D;\ P_1,\ P_2,\ P_0-P_3$		P_2/P_1
$=-1\quad =2$	$S_{13,18}$	$K_3,\ D;\ P_1,\ P_3,\ P_0$		$(P_0-P_3)^2/P_1(P_0+P_3)$
$A^{\gamma,b}_{5,35}$				
$b=1,\ \gamma=0$	$S_{2,3}$	$L_3,\ K_3;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$		$(P_0-P_3)^2/\{(L_1-K_2)^2+(L_2+K_1)^2\}$
$b=\frac{1+a}{a}$ $\gamma=-\frac{\tan c}{a}$	$S_{5,11}$	$D-2aK_3,\ \cos c L_3-\sin c K_3;$ $L_2+K_1,\ L_1-K_2,\ P_0-P_3$	$a\neq0$ $c<c<\pi,\ c\neq\pi/2$	$\frac{(P_0-P_3)^{2a}}{((L_1-K_2)^2+(L_2+K_1)^2)^{1+a}}\times\left(\frac{L_1-K_2-i(L_2+K_1)}{L_1-K_2+i(L_2+K_1)}\right)^{i\tan c}$
$b=\frac{1+a}{a}$ $\gamma=0$	$S_{6,15}$	$D-2aK_3,\ L_3;\ L_2+K_1,\ L_1-K_2,$ P_0-P_3	$a\neq0,\ -1$	$(P_0-P_3)^{2a}/\{(L_1-K_2)^2+(L_2+K_1)^2\}^{1+a}$

TABLE V. (*Continued*).

Class	Notation	Generators	Range of parameters	Invariants
$b=1$ $\gamma=1/a$	$S_{7,16}$	$D-2aL_3,\ K_3;\ L_2+K_1,\ L_1-K_2$ P_0-P_3	$a\neq 0$	$\dfrac{(P_0-P_3)^2}{(L_1-K_2)^2+(L_2+K_1)^2} \times \left[\dfrac{L_1-K_2-i(L_2+K_1)}{L_1-K_2+i(L_2+K_1)}\right]^{i/a}$
$b=1$ $\gamma=-\tan c$	$S_{11,8}$	$D,\ L_3-\tan cK_3;\ P_1,\ P_2,\ P_0-P_3$	$0<c<\pi$ $c\neq\pi/2$	$\dfrac{(P_0-P_3)^2}{P_1^2+P_2^2}\left[\dfrac{P_2-iP_1}{P_2+iP_1}\right]^{-i\tan c}$
$b=1-a\tan c$ $\gamma=-\tan c$	$S_{11,14}$	$D-2aL_3,\ L_3-\tan cK_3;\ P_1,\ P_2,$ P_0-P_3	$a\neq 0$ $0<c<\pi,\ c\neq\pi/2$	$\dfrac{(P_0-P_3)^2}{(P_1^2+P_2^2)^{1-a\tan c}}\left(\dfrac{P_2-iP_1}{P_2+iP_1}\right)^{-i\tan c}$
$b=1,\ \gamma=0$	$S_{12,24}$	$D,\ L_3;\ P_1,\ P_2,\ P_0-P_3$		$(P_0-P_3)^2/(P_1^2+P_2^2)$
$=1\quad =0$	$S_{12,25}$	$D,\ L_3;\ P_1,\ P_2,\ P_3$		$P_3^2/(P_1^2+P_2^2)$
$=1\quad =0$	$S_{12,26}$	$D,\ L_3;\ P_1,\ P_2,\ P_0$		$P_0^2/(P_1^2+P_2^2)$
$b=1+a,\ \gamma=0$	$S_{12,34}$	$D-2aK_3,\ L_3;\ P_1,\ P_2,\ P_0-P_3$	$a\neq 0,\ -1$	$(P_0-P_3)^2/(P_1^2+P_2^2)^{1+a}$
$=2\quad =0$	$S_{12,35}$	$2D-4K_3+P_0+P_3;\ L_3;\ P_1,\ P_2,$ P_0-P_3		$(P_0-P_3)^2/(P_1^2+P_2^2)$
$=2\quad =0$	$S_{12,36}$	$2D-4K_3+x(P_0-P_3),\ 4L_3+P_0$ $+P_3;\ P_1,\ P_2,\ P_0-P_3$	$-\infty<x<\infty$	"
$A_{5,36}$	$S_{8,22}$	$D,\ K_3;\ L_2+K_1,\ P_1,\ P_0-P_3$		$D-2K_3-2(L_2+K_1)P_1/(P_0-P_3)$
	$S_{8,23}$	$D,\ K_3;\ L_2+K_1,\ P_2-cP_1,\ P_0-P_3$	$c\neq 0$	$D-2K_3+2(L_2+K_1)(P_2-cP_1)/$ (cP_0-cP_3)
$A_{5,37}$	$S_{6,19}$	$D-2K_3,\ L_3;\ L_2+K_1-P_2,$ $L_1-K_2-P_1,\ P_0-P_3$		$\{(L_2+K_1-P_1)^2+(L_1-K_2-P_1)^2\}/$ $(2P_0-2P_3)+2L_3$

TABLE VI. Six-dimensional subalgebras.

Dim. of derived L	Notation	Generators	Range of parameters	Invariants
SL(2, C) 6	$S_{1,2}$	$;L_1,\ L_2,\ L_3,\ K_1,\ K_2,\ K_3$		$(\bar{L})^2-(\bar{K})^2;\ \bar{L}\cdot\bar{K}$
E(3) 6	$S_{3,2}$	$;L_1,\ L_2,\ L_3,\ P_1,\ P_2,\ P_3$		$P_1^2+P_2^2+P_3^2;\ \bar{P}\cdot\bar{L}$
E(21) 6	$S_{4,2}$	$;K_1,\ K_2,\ L_3,\ P_1,\ P_2,\ P_0$		$P_1^2+P_2^2-P_0^2;\ P_0L_3+(\bar{P}\times\bar{K})_3$
5	$\begin{cases} S_{5,2} \\ S_{10,20} \end{cases}$	$L_3-\tan cK_3;\ L_2+K_1,\ L_1-K_2,$ $P_1,\ P_2,\ P_0-P_3$ $D+2aL_3-2bK_3;\ L_2+K_1,\ L_1-K_2$ $P_1,\ P_2,\ P_0-P_3$	$0<c<\pi,\ c\neq\pi/2$ $b=0,\ a\neq 0$	none "
5	$S_{6,2}$	$L_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,$ P_0-P_3		$P_0-P_3;\ P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L}$
5	$S_{6,5}$	$4L_3+P_0+P_3;\ L_2+K_1,\ L_1-K_2,$ $P_1,\ P_2,\ P_0-P_3$		$P_0-P_3;\ P_0L_3+(\bar{P}\times\bar{K})_3$ $-\bar{P}\cdot\bar{L}+(P_0^2-P_3^2)/4$
5	$S_{10,20}$	$D-2aL_3-2bK_3;\ L_2+K_1,$ $L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$	$b\neq 0,\ -1$ $-\infty<a<\infty$	none
5	$S_{10,20}$	$D-2aL_3-2bK_3;\ L_2+K_1,$ $L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$	$b=-1$ $-\infty<a<\infty$	$P_0-P_3,\ D-2\{a(P_0L_3-\bar{P}\cdot\bar{L}$ $+(\bar{P}\times\bar{K})_3+P_0K_3-\bar{P}\cdot\bar{K}$ $-(\bar{P}\times\bar{L})_3\}/(P_0-P_3)$
5	$S_{10,21}$	$2D+4aL_3-4K_3+P_0+P_3;\ L_2+K_1,$ $L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$	$-\infty<a<\infty$	none
5	$S_{10,26}$	$D-K_3;\ L_1-K_2,\ 2L_2+2K_1$ $-P_0-P_3,\ P_1,\ P_2,\ P_0-P_3$		"
5	$S_{14,31}$	$D-2aK_3;\ L_2+K_1,\ P_1,\ P_2,\ P_3,\ P_0$	$a\neq 0$	$P_2^{a+1}/(P_0-P_3);\ P_2^2/(P_1^2+P_3^2-P_0^2)$
4	$S_{7,10}$	$D,\ K_3;\ L_2+K_1,\ L_1-K_2,\ P_2,$ P_0-P_3		none
$A_1\oplus A_5$ 4	$S_{8,1}$	$(P_2)\oplus(L_3;\ L_2+K_1,\ P_1,\ P_3,\ P_0)$		$P_2;\ P_1^2+P_3^2-P_0^2$
4	$S_{8,19}$	$D,\ K_3;\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$		none
4	$S_{8,20}$	$D,\ K_3;\ L_2+K_2,\ P_1,\ P_3,\ P_0$		"
$A_{3,4}\oplus A_{3,6}$ 4	$S_{9,1}$	$(K_3;\ P_0,\ P_3)\oplus(L_3;\ P_1,\ P_2)$		$P_0^2-P_3^2;\ P_1^2+P_2^2$
4	$S_{11,7}$	$D,\ L_3-\tan cK_3;\ P_1,\ P_2,\ P_3,\ P_0$	$0<c<\pi,\ c\neq\pi/2$	$(P_2^2+P_1^2)/(P_0^2-P_3^2);$ $(P_1^2+P_2^2)/(P_0-P_3)^2\{(P_2-iP_1)/$ $(P_2+iP_1)\}^{i\tan c}$

TABLE VI. (*Continued*).

Dim of derived L	Notation	Generators	Range of parameters	Invariants
4	$S_{11,13}$	$D+2aL_3,\ L_3-\tan c K_3;\ P_1,\ P_2,\ P_3,\ P_0$	$a>0$; $0<c<\pi,\ c\neq\pi/2$	"; $\dfrac{(P_1{}^2+P_2{}^2)^{1-a\tan c}}{(P_0-P_3)}\left(\dfrac{P_2-iP_1}{P_2+iP_1}\right)^{i\tan c}$
4	$S_{12,23}$	$D,\ L_3;\ P_1,\ P_2,\ P_3,\ P_0$		$(P_2^2+P_1^2)/(P_0+P_3)^3;\ (P_2+P_1{}^2)/(P_0-P_3)^2$
4	$S_{12,33}$	$D-2aK_3;\ P_1,\ P_2,\ P_3,\ P_0$	$a>0$	$(P_2^2+P_1^2)^{1-a}/(P_0+P_3)^2;\ (P_2^2+P_1^2)^{a+1}/(P_0-P_3)^2$
4	$S_{13,16}$	$D,\ K_3;\ P_1,\ P_2,\ P_3,\ P_0$		$P_2^2/(P_0{}^2-P_3{}^2);\ P_1/P_2$
4	$S_{13,25}$	$D+2aL_3,\ K_3;\ P_1,\ P_2,\ P_3,\ P_0$	$a>0$	$(P_0{}^2-P_3^2)/(P_1{}^2+P_2^2);\ (P_2{}^2+P_1{}^2)\{(P_2-iP_1)/(P_2+iP_1)\}^{i/a}$
4	$S_{14,22}$	$D,\ L_2+K_1;\ P_1,\ P_2,\ P_3,\ P_0$		$(P_1{}^2+P_3^2-P_0{}^2)/P_2^2;\ (P_0-P_3)/P_2$
4	$S_{14,32}$	$D+L_1-K_2,\ L_2+K_1;\ P_1,\ P_2,\ P_3,\ P_0$		$((P_0-P_3)\exp(2P_2/(P_0-P_3));\ (P_1{}^2+P_2^2+P_3^2-P_0{}^2)/(P_0-P_3)^2$
$A_2\oplus A_{14,12}$ 3	$S_{2,7}$	$(D;\ P_0-P_3)\oplus(L_3,\ D+2K_3;\ L_2+K_1,\ L_1-K_2)$		none
3	$S_{7,2}$	$K_3,\ P_1,\ P_2;\ L_2+K_1,\ L_1-K_2,\ P_0-P_3$		"
	$S_{10,15}$	$D,\ L_1-K_2,\ L_2+K_1;\ P_1,\ P_2,\ P_0-P_3$		"
3	$S_{9,8}$	$D,\ K_3,\ L_3;\ P_1,\ P_2,\ P_0-P_3$		"
3	$S_{10,1}$	$L_2+K_1,\ L_1-K_2,\ P_0+P_3;\ P_1,\ P_2,\ P_0-P_3$		$P_0-P_3;\ P_1{}^2+P_2^2+P_3{}^2-P_0^2$

TABLE VII. Seven-dimensional subalgebras.

Dim. of derived L	Notation	Generators	Range of Parameters	Invariants
$D\oplus$ SL (2, C) 6	$S_{1,4}$	$(D)\oplus C;\ L_1,\ L_2,\ L_3,\ K_1,\ K_2,\ K_3)$		$D;\ \overline{L}^2-\overline{K}^2;\ \overline{L}\cdot\overline{K}$
$A_1\oplus$ E(3)	$S_{3,1}$	$(P_0)\oplus(;L_1,\ L_2,\ L_3,\ P_1,\ P_2,\ P_3)$		$P_0,\ \overline{P}\cdot\overline{L},\ \overline{P}^2-P_0{}^2$
$D\,\square\,$E(3) 6	$S_{3,6}$	$D;\ P_1,\ P_2,\ P_3,\ L_1,\ L_2,\ L_3$		$\overline{P}\cdot\overline{L}/(P_1{}^2+P_2^2+P_3{}^2)$
$A_1\oplus$ E(2, 1) 6	$S_{4,1}$	$(P_3)\oplus(;P_1,\ P_2,\ P_0,\ K_1,\ K_2,\ L_3)$.		$P_3;\ P_0L_3+(\overline{P}\times\overline{K})_3;\ P_1{}^2+P_2{}^2-P_0{}^2$
$D\,\square\,$E(2, 1) 6	$S_{4,6}$	$D;\ P_1,\ P_2,\ P_0,\ K_1,\ K_2,\ K_3$		$(P_0L_3+(\overline{P}\times\overline{K})_3)^2/(P_1{}^2+P_2{}^2-P_0{}^2)$
6	$S_{7,1}$	$K_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_3,\ P_0$		$\overline{P}^2-P_0{}^2$
6	$S_{10,19}$	$D-2aL_3-2bK_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_3,\ \mathrm{P}_0$	$a^2+b^2\neq 0$	$(\mathrm{P}_0{}^2-\overline{P}^2)^{1+b}/(P_0-P_3)^2$
5	$S_{2,2}$	$K_3,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$		$(P_0L_3+(\overline{P}\times\overline{K})_3-\overline{P}\cdot\overline{L})/(P_0-P_3)$
	$S_{6,9}$	$D,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$		"
5	$S_{5,6}$	$D,\ L_3-\tan c K_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$	$0<c<\pi,\ c\neq\pi/2$	$D\sin c+2\{\cos c(\overline{P}\cdot\overline{L}-(\overline{P}\times\overline{K})_3-P_0L_3)+\sin c(P_0K_3-(\overline{P}\times\overline{L})_3-\overline{K}\cdot\overline{P})\}/(P_0-P_3)$
5	$S_{7,14}$	$D-2aL_3,\ K_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$	$a\neq 0$	$D+2\{a(\overline{P}\cdot\overline{L}-(\overline{P}\times\overline{K})_3-P_0L_3)+(P_0K_3-(\overline{P}\times\overline{L})_3-\overline{K}\cdot\overline{P})\}/(P_0-P_3)$
5	$S_{5,10}$	$D-2aK_3,\ L_3-\tan c K_3,\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$	$a\neq 0$; $0<c<\pi,\ c\neq\pi/2$	$D\sin c+2\{(a+D\cos c)(\overline{P}\cdot\overline{L}-(\overline{P}\times\overline{K})_3-P_0L_3)+\sin c(P_0K_3-(\overline{P}\times\overline{L})-\overline{P}\cdot\overline{K})\}/(P_0-P_3)$
5	$S_{6,1}$	$L_3,\ P_0+P_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$		$P_0-P_3;\ P_0^2-\overline{P}^2,\ P_0L_3+(\overline{P}\times\overline{K})_3-\overline{P}\cdot L$
5	$S_{6,13}$	$D-2aK_3,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$	$a\neq 0,\ -1$	$(P_0L_3+(\overline{P}\times\overline{K})_3-\overline{P}\cdot\overline{L})/(P_0-P_3)$

TABLE XII. (*Continued*).

Dim of derived L	Notation	Generators	Range of parameters	Invariants
5	$S_{6,13}$	$D-2aK_3,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3$	$a=-1$	$P_0-P_3;\ P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L};\ 2(P_0K_3-(\bar{P}\times\bar{L})_3-\bar{K}\cdot\bar{P})+(P_0-P_3)D$
5	$S_{6,14}$	$2D-4K_3+P_0+P_3,\ L_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$		$(P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L})/(P_0-P_3)$
5	$S_{6,18}$	$2D-4K_3+x(P_0-P_3),\ 4L_3+P_0+P_3;\ L_1-K_2,\ L_2+K_1,\ P_1,\ P_2,\ P_0-P_3\ P_1,\ P_2,\ P_0-P_3$	$-\infty<x<\infty$	$\{P_2^2+P_1^2+P_3^2-P_0^2+4[(P_0-P_3)L_3+P_1(L_1-K_2)+P_2(L_2+K_1)]\}/(P_0-P_3)$
5	$S_{7,9}$	$D,\ K_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$		$2(P_0K_3-(\bar{P}\times\bar{L})_3-\bar{P}\cdot\bar{K})/(P_0-P_3)+D$
5	$S_{8,18}$	$D,\ K_3;\ L_2+K_1,\ P_1,\ P_2,\ P_3,\ P_0$		$(P_1^2+P_3^2-P_0^2)/P_2^2$
4	$S_{9,7}$	$D,\ K_3;\ L_3;\ P_1,\ P_2;\ P_3,\ P_0$		$(P_1^2+P_2^2)/(P_0^2-P_3^2)$
4	$S_{10,14}$	$D,\ L_1-K_2,\ L_2+K_1;\ P_1,\ P_2,\ P_3,\ P_0$		$(P_0^2-\bar{P}^2)/(P_0-P_3)^2$

TABLE VIII. Eight-dimensional subalgebras.

Dim. of derived L	Notation	Generators	Range of Parameters	Invariants
6	$S_{2,1}$	$K_3,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_0,\ P_1,\ P_2,\ P_3$		$P_0^2-\bar{P}^2;\ (P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L})/(P_0-P_3)$
6	$S_{3,5}$	$D,\ L_3;\ L_1,\ L_2,\ P_0,\ P_1,\ P_2,\ P_3$		$(\bar{P}\cdot\bar{L})^2/P_0^2,\ \bar{P}^2/P_0^2$
6	$S_{4,5}$	$D,\ L_3;\ K_1,\ K_2,\ P_0,\ P_1,\ P_2,\ P_3$		$(P_0L_3+(\bar{P}\times\bar{K})_3)/P_3;\ (P_1^2+P_2^2-P_0^2)/P_3^2$
6	$S_{5,5}$	$D,\ L_3-\tan c\,K_3;\ L_2+K_1,\ L_1-K_2,\ P_0,\ P_1,\ P_2,\ P_3$	$0<c<\pi,\ c\neq\pi/2$	none
6	$S_{5,9}$	$D-2aK_3,\ L_3-\tan c\,K_3;\ L_2+K_1,\ L_1-K_2,\ P_0,\ P_1,\ P_2,\ P_3$	$a\neq 0$; $0<c<\pi,\ c\neq\pi/2$	"
6	$S_{6,8}$	$D,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_0,\ P_1,\ P_2,\ P_3$		$(P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L})/(P_0-P_3);\ (P_0^2-\bar{P}^2)/(P_0-P_3)^2$
6	$S_{6,12}$	$D-2aK_3,\ L_3;\ L_1-K_2,\ L_2+K_1,\ P_0,\ P_1,\ P_2,\ P_3$	$a\neq 0$	"; $(P_0^2-\bar{P}^2)^{1+a}/(P_0-P_3)^2$
6	$S_{7,8}$	$D,\ K_3;\ L_1-K_2,\ L_2+K_1,\ P_0,\ P_1,\ P_2,\ P_3$		none
6	$S_{7,13}$	$D+aL_3,\ K_3;\ L_2+K_1,\ L_1-K_2,\ P_0,\ P_1,\ P_2,\ P_3$	$a\neq 0$	"
5	$S_{2,6}$	$D,\ L_3,\ K_3;\ L_2+K_1,\ L_1-K_2,\ P_1,\ P_2,\ P_0-P_3$		$\frac{P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L}}{P_0-P_3};\ D-\frac{2}{P_0-P_3}\{\bar{P}\cdot\bar{K}+(\bar{P}\times\bar{L})_3-P_0K_3\}$

TABLE XI. Nine-, ten-, and eleven-dimensional subalgebras.

Dim. of derived L	Notation	Generators	Range of Parameters	Invariants
(6)	$S_{2,5}$	$D,\ L_3,\ K_3;\ L_1-K_2,\ L_2+K_1,\ P_0,\ P_1,\ P_2,\ P_3$		$(P_0L_3+(\bar{P}\times\bar{K})_3-\bar{P}\cdot\bar{L})/(P_0-P_3)$
(10)	$S_{1,1}$	$;\ L_1,\ L_2,\ L_3,\ K_1,\ K_2,\ K_3,\ P_1,\ P_2,\ P_3,\ P_0$		$\bar{P}^2-P_0^2;\ W^2$
(10)	$S_{1,3}$	$D;\ L_1,\ L_2,\ L_3,\ K_1,\ K_2,\ K_3,\ P_1,\ P_2,\ P_3,\ P_0$		$W^2/(\bar{P}^2-P_0^2)$

*Permanent address: Physics Department, McGill University, Montreal, Canada.

[1]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597 (1975).

[2]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1615 (1975).

[3]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977 (1976).

[4]L. Abellanos and L. Martinez Alonso, J. Math. Phys. **16**, 1580 (1975).

[5]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).

[6]M. Pauri and G.M. Prosperi, Nuovo Cimento **43**, 533 (1966).

[7]V.P. Maslov, *Operator Methods* (Nauka, Moscow, 1973).

[8]G.M. Mubarakzyanov, Izv. Vyssh. Ucbeben. Zaved. Mat. 32, 114 (1963); **34**, 99 (1963); **35**, 104 (1963).

[9]All nilpotent Lie algebras of dim = 6 were classified and appear in V.V. Morozov, Izv. Vyssh. Ucbeben. Zaved. Mat. **5**, 161 (1958); the solvable ones also appear in Ref. 5.

[10]J. Kogut and L. Susskind, Phys. Rep. **8**, 75 (1973).

[11]H. Bacry and N.P. Chang, Ann. Phys. (N.Y.) **47**, 407 (1968).

[12]P.A.M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).

[13]P.A.M. Dirac, Proc. R. Soc. London Ser. A **322**, 435 (1971), **328**, 1 (1972).

[14]L.C. Biedenharn and H. Van Dam, Phys. Rev. D **9**, 471 (1974).

[15]H. Leutwyler and J. Stern, Nucl. Phys. B **20**, 77 (1970).

[16]S.P. de Alwis and J. Stern, Nucl. Phys. B **77**, 509 (1974).

[17]For further examples and references see Ref. 3.

Exact occupation statistics for two-dimensional lattices of single particles*

Eizo Miyazaki and Iwao Yasumori

Department of Chemistry, Tokyo Institute of Technology, O-okayama, Meguro-ku, Tokyo 152, Japan
(Received 14 September 1976)

A general expression is developed which describes exactly the ensemble average number of one- or two-dimensional structures per arrangement, created when indistinguishable single particles are arranged on a two-dimensional lattice. The expression obtained is applied to the calculation of some physically important structures which appear on a rectangular and a closest-packed hexagonal lattice. The problem of nearest-neighbor pairs is then solved as a special case from the general expression.

I. INTRODUCTION

A statistical mechanical treatment of adsorption, heterogeneous catalysis, elasticity, alloys, magnetism, and other cooperative phenomena requires knowledge of the number of occupied, mixed and vacant nearest-neighbor pairs. It also requires knowledge of the (partial) structures with more complex configuration which are formed on an ordered two-dimensional lattice, e.g., mixed or vacant structures composed of so-called "B_n-sites"[1] and of a stepped surface with some "terraces."[2] These structures are known as active sites in the sense that they may play important role on adsorption and catalysis on a solid surface. Several simple structures among them formed on a rectangular lattice are shown by dotted lines in Fig. 1.

The question of the nearest-neighbor pairs has previously been considered,[3–5] and the exact relationships have been developed which describe the exact occupation statistics for one-dimensional array of λ-bells[3] and for two-dimensional rectangular[4] or other lattices[5] of single particles ($\lambda = 1$; here λ refers to the number of contiguous lattice sites occupied by a particle).

In the present paper, a general expression is developed which describes exactly the ensemble number of the structures created when indistinguishable single particles are arranged on a two-dimensional lattice. The expression obtained is then applied to calculation of some structures which appear on a rectangular and a closest-packed hexagonal lattices. There the problem of the nearest-neighbor pairs is solved as a special case from the general expression.

II. GENERAL EXPRESSION

As is known, the quantity $W(q,N)$, the number of distinguishable ways in which q indistinguishable single particles are arranged on a lattice of N equivalent sites, is given by

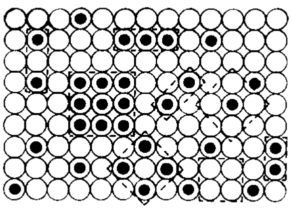

FIG. 1. A representation of several simple structures created when single particles (●) are placed on a rectangular lattice (○). The structures are shown by dotted lines.

$$W(q,N) = \binom{N}{q} = N!/(N-q)!q! \tag{1}$$

because there are $N-q$ vacant sites and $N-q+q=N$ individuals to be permuted, of which q are of one kind and $N-q$ of another. The total number of particles which appear in all possible arrangements, N_1, is then written by

$$N_1 = q\,W(q,N) = q\binom{N}{q} \tag{2}$$

since there are q particles in each of the $W(q,N)$ arrangements. The quantity N_1 is also derived on the basis of the following argument: When q particles are arranged in all possible ways on a lattice of N sites, a particle at a particular site will occur $\binom{N-1}{q-1}$ times since the rest of the particles, $q-1$, can be arranged in all possible ways on the rest of the sites, $N-1$. There are N distinguishable ways of placing a particle on the lattice; hence we obtain

$$N_1 = N\binom{N-1}{q-1}, \tag{3}$$

which is equivalent to the $q\binom{N}{q}$ in Eq. (2).

These arguments will be generalized to the calculation of the total number of structures with a one- or two-dimensional configuration, instead of the particles in the above case, created when q single particles are placed in all possible ways on N sites. We define α as the number of particles included in a structure having a definite configuration and consisting of z sites, and we define σ_z as the number of distinguishable ways of placing such a structure on a lattice composing of N sites. If q particles are arranged in all possible ways on the lattice, the structure occupying a particular position on the lattice occurs $\binom{N-z}{q-\alpha}$ times because the $q-\alpha$ remaining particles can be arranged in all possible ways on the remaining $N-z$ sites. For the case of single particle ($\lambda = 1$) the quantity, $\binom{N-z}{q-\alpha}$, is independent of the position on the lattice, therefore N_z, the total number of such structures created when q particles are arranged in all possible ways on N sites is given by

$$N_z = \sigma_z\binom{N-z}{q-\alpha}, \tag{4}$$

and the ensemble average number of the structures per arrangement, n_z, is

$$n_z = N_z/W(q,N) = \sigma_z\binom{N-z}{q-\alpha}\Big/\binom{N}{q}$$

$$= \sigma_z q(q-1)(q-2)\cdots(q-\alpha+1)(N-q)(N-q-1) \times\cdots(N-q-z+\alpha+1)[N(N-1)(N-2)\cdots(N-z+1)]^{-1} \tag{5}$$

or P_z, the ensemble average number of the structures per sites is

$$P_z = n_z/N = \sigma_z q(q-1)(q-2)\cdots(q-\alpha+1)(N-q)(N-q-1)\cdots \times(N-q-z+\alpha+1)[NN(N-1)(N-2)\cdots(N-z+1)]^{-1}. \tag{6}$$

In the following sections, we will consider the quantity, σ_z, for evaluating these quantities, N_z, n_z, and P_z, for some structures formed on a rectangular and a closest-packed hexagonal lattice. The structures considered here are important for physical and chemical problems.

III. STRUCTURES ON A RECTANGULAR LATTICE

A. $r \times s$ rectangular structure

We first consider the case of a rectangular structure consisting of r columns and s rows which appears on a similar rectangular lattice composing of R columns and S rows. There are $(R-r+1)(S-s+1)$ distinguishable ways of placing the structure on the lattice of which $R-r+1$ are of one kind along a column and $S-s+1$ are of the other along a row. When $r \neq s$, there are $(R-s+1)(S-r+1)$ additional distinguishable ways which may be obtained by interchanging r and s. Hence the number of such the structures created when q particles are arranged in all possible ways on a $R\times S$-rectangular lattice is obtained from Eq. (4) as

$$[(R-r+1)(S-s+1) + (R-s+1)(S-r+1)]\binom{RS-rs}{q-\alpha} \tag{7}$$

for $r\neq s$, and

$$(R-r+1)(S-r+1)\binom{RS-r^2}{q-\alpha} \tag{8}$$

for $r=s$.

The number of occupied nearest-neighbor pairs (N_{11}), number of occupied third-nearest-neighbor pairs with or without intervening particles (N_{111} or N_{101}, respectively) created when q particles are arranged in all possible ways may be derived respectively by setting $[r=2,\ s=1,\ \alpha=2]$, $[r=3,\ s=1,\ \alpha=3]$, or $[r=3,\ s=1,\ \alpha=2]$ in Eq. (7) as

$$N_{11} = (2RS-S-R)\binom{RS-2}{q-2}, \tag{9}$$

$$N_{111} = (2RS-2S-2R)\binom{RS-3}{q-3}, \tag{10}$$

or

$$N_{101} = (2RS-2S-2R)\binom{RS-3}{q-2}, \tag{11}$$

and N_{10}, the number of mixed nearest-neighbor pairs where one site is occupied and another is vacant, is given by setting $r=2$, $s=1$, and $\alpha=1$ in Eq. (7) as

$$N_{10} = 2(2RS-S-R)\binom{RS-2}{q-1}, \tag{12}$$

where the factor 2 prior to the bracket in Eq. (12) arises from interchangeability between the occupied site and the vacant site in the mixed nearest-neighbor pairs. The ensemble average number of these nearest-neighbor pairs per arrangement, n_{11}, n_{10}, n_{111}, and n_{101}, are calculated from Eq. (5) and are given by

$$n_{11} = (2RS-S-R)\frac{q(q-1)}{RS(RS-1)}, \tag{13}$$

$$n_{10} = 2(2RS-S-R)\frac{q(RS-1)}{RS(RS-1)}, \tag{14}$$

$$n_{111} = (2RS-2S-2R)\frac{q(q-1)(q-2)}{RS(RS-1)(RS-2)}, \tag{15}$$

and

$$n_{101} = (2RS-2S-2R)\frac{q(q-1)(RS-1)}{RS(RS-1)(RS-2)}. \tag{16}$$

The results (13), (15), and (16) are equivalent to the equations derived by previous methods.[4,5] Thus, the numbers of nearest-neighbor pairs may easily be calculated as special cases of Eq. (7).

B. Linear array along a diagonal

Here we consider the case of a linear array consisting of r sites along a diagonal. There are $2(R-r+1)(S-r+1)$ distinguishable ways of placing the linear array on a $R\times S$-rectangular lattice where the factor 2 arises because there are two distinguishable directions of diagonals on the lattice. Hence, the number of the linear arrays created when q particles are arranged in all possible ways on the $R\times S$-rectangular lattice is given by

$$2(R-r+1)(S-r+1)\binom{RS-r}{q-\alpha}, \tag{17}$$

so that N'_{11}, the number of occupied next-nearest-neighbor pairs, i.e., the number of occupied nearest-neighbor pairs along a diagonal, is obtained by setting $r=2$ and $\alpha=2$ in Eq. (17) as

$$N'_{11} = 2(R-1)(S-1)\binom{RS-2}{q-2} \tag{18}$$

or the ensemble average number of next-nearest-neighbor pairs per arrangement, n'_{11}, is

$$n'_{11} = 2(R-1)(S-1)\,q(q-1)/RS(RS-1). \tag{19}$$

Equations (18) and (19) are also equivalent to the previously derived results.[4]

C. Isolated $r \times s$ rectangular structure with filled particles

In this section we consider the case of an isolated $r\times s$-rectangular structure with filled particles, i.e., the structure that is completely occupied by particles and surrounded by vacant sites of nearest neighbors. In this case three different configurations should be considered separately due to the structures forming at (a) corners, (b) edges, and (c) other positions on a $R\times S$-rectangular lattice: For the respective case, there are (a) eight distinguishable ways of placing the structure on four corners, of which four ways are obtained by interchanging r and s when $r \neq s$, (b) $2(R-r-1) + 2(S-s-1) + 2(R-s-1) + 2(S-r-1)$ distinguishable ways on the four edges, of which the last two terms are of additional distinguishable ways created by interchanging r and s when $r \neq s$, and (c) $(R-r-1)(S-s-1) + (R-s-1)(S-r-1)$ ways which are obtained by referring to Eq. (7). Hence, the total number of isolated $r \times s$-rectangular structures with filled particles created when q particles are arranged in all possible ways is given by summing the above three cases as

$$8\binom{RS-rs-r-s}{q-rs} + 2(R+S-2r-2)\binom{RS-rs-2s-r}{q-rs} + 2(R+S-2s-2)\binom{RS-rs-2r-s}{q-rs} + [2(RS-R-S+rs+r+s+1) - (R+S)(r+s)]\binom{RS-rs-2r-2s}{q-rs} \quad (20)$$

for $r \neq s$, and

$$4\binom{RS-r^2-2r}{q-r^2} + 2(R+S-2r-2)\binom{RS-r^2-3r}{q-r^2} + (R-r-1)(S-r-1)\binom{RS-r^2-4r}{q-r^2} \quad (21)$$

for $r=s$. When we set $r=s=1$ in Eq. (21), we obtain the number of isolated particles created when q particles are arranged in all possible ways on a $R \times S$-rectangular lattice as

$$4\binom{RS-3}{q-1} + 2(R+S-4)\binom{RS-4}{q-1} + (R-2)(S-2)\binom{RS-5}{q-1} \quad (22)$$

and the probability that a site is occupied and isolated is

$$\frac{(RS-q)(RS-q-1)q}{(RS)(RS)(RS-1)(RS-2)}\left[4 + \frac{2(R+S-4)(RS-q-2)}{RS-3} + \frac{(R-2)(S-2)(RS-q-2)(RS-q-3)}{(RS-3)(RS-4)}\right]. \quad (23)$$

Equation (23) leads to $\theta(1-\theta)^4$ in the limit as R and S approach infinity, where θ denotes surface coverage defined by $\theta = q/RS$. The quantity given by Eq. (23) is important for some catalysis problems on metal surfaces.[5]

IV. STRUCTURES ON A CLOSEST-PACKED HEXAGONAL LATTICE

Here we shall consider a few examples of structures which appear on a closest-packed hexagonal lattice containing M sites along an edge, as shown in Fig. 2. The total number of sites, N_h, is then given as a function of M by

$$N_h = \sum_{k=1}^{M-1} 6(M-k) + 1 = 3M(M-1) + 1.$$

A. Linear array of r-contiguous sites

In this case, there are three different directions on the lattice, and for each direction there are

$$2\sum_{k=M}^{2M-2} (k-r+1) + (2M-1-r+1) = (M-1)(3M-2r) + 2M - r$$

distinguishable ways of placing the structure on the lattice for $r \leq M+1$ and

$$2\sum_{k=r}^{2M-2} (k-r+1) + (2M-1-r+1) = 3(2M-r)^2$$

ways for $r \geq M$, where the factor 2 prior to the first terms in the left-hand side of the equations arises from the symmetry property associated with the axis denoted by an arrow in Fig. 2 and the second terms are due to the longest linear array consisting of $2M-1$ sites. Hence, the number of linear arrays of r-contiguous sites created when q particles are arranged in all possible ways is given by

$$3[(M-1)(3M-2r) + 2M - r]\binom{N_h - r}{q - \alpha} \quad (24)$$

for $r \leq M+1$, and

$$3(2M-r)^2\binom{N_h - r}{q-\alpha} \quad (25)$$

for $2M-1 \geq r \geq M$. Thus, $N_{11,h}$, $N_{111,h}$, and $N_{101,h}$, the numbers of occupied nearest-neighbor, next-nearest-neighbor pairs with and without intervening particles, are obtained by setting $[r=2,\ \alpha=2]$, $[r=3,\ \alpha=3]$, and $[r=3,\ \alpha=2]$ in (24), respectively, as

$$N_{11,h} = 3(3M^2 - 5M + 2)\binom{N_h-2}{q-2}, \quad (26)$$

$$N_{111,h} = 3(3M^2 - 7M + 3)\binom{N_h-3}{q-3}, \quad (27)$$

and

$$N_{101,h} = 3(3M^2 - 7M + 3)\binom{N_h-3}{q-2}, \quad (28)$$

and the ensemble average numbers per arrangement, $n_{11,h}$, $n_{111,h}$, and $n_{101,h}$

$$n_{11,h} = 3(3M^2 - 5M + 2)\frac{q(q-1)}{N_h(N_h-1)}, \quad (29)$$

$$n_{111,h} = 3(3M^2 - 7M + 3)\frac{q(q-1)(q-2)}{N_h(N_h-1)(N_h-2)}, \quad (30)$$

and

$$n_{101,h} = 3(3M^2 - 7M + 3)\frac{q(q-1)(N_h-1)}{N_h(N_h-1)(N_h-2)}. \quad (31)$$

The ensemble average numbers per site [see Eq. (6)], $P_{11,h}$, $P_{111,h}$, and $P_{101,h}$, lead to $3\theta^2$, $3\theta^3$, and $3\theta^2(1-\theta)$ respectively, in the limit as N_h or M approaches infinity.

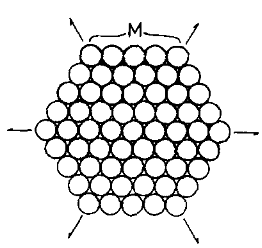

FIG. 2. Closest-packed hexagonal lattice with M sites along an edge. The arrows denote symmetry axes.

B. Closest-packed hexagonal structure with m sites along an edge

There are

$$2\sum_{k=M}^{2M-m-1}(k-m+1)+2M-2m+1$$
$$=3(M-m)(M-m+1)+1$$

distinguishable ways of placing the structure on the lattice; hence the number of such the structures created when q particles are arranged in all possible ways is obtained from Eq. (4) as

$$[3(M-m)(M-m+1)+1]\binom{N_h-z}{q-\alpha}, \qquad (32)$$

where $z=3m(m-1)+1$. Thus, for example, the number of structures in which a vacant site is surrounded by six filled sites, created when q particles are arranged in all possible ways, is obtained by setting $m=2$ and $\alpha=6$ in (32) as

$$(3M^2-9M+7)\binom{N_h-7}{q-6}, \qquad (33)$$

or the probability that a site is vacant and surrounded by six filled sites is

$$(3M^2-9M+7)\frac{q(q-1)(q-2)\cdots(q-5)(N_h-q)}{N_h^2(N_h-1)(N_h-2)\cdots(N_h-6)}, \qquad (34)$$

which is also equivalent to the probability of success when attempting to place, in a random manner, an additional particle on the vacant site in the structure. The quantity shown in (34) leads to $\theta^6(1-\theta)$ in the limit as N_h or M approaches infinity.

V. CONCLUSION

We have derived a general expression which describes exactly the ensemble average number of structures per arrangement, created when single particles are arranged on a two-dimensional lattice. The number of nearest-neighbor pairs of various types is then obtained easily from the general expression.

*This work supported in part by RCA Research Laboratories, Inc.

[1]E. F. G. Herington and E. K. Rideal, Trans. Faraday Soc. 40, 505 (1944); R. van Hardveld and A. van Montfoort, Surface Sci. 4, 396 (1966); 15, 189 (1969); G. C. Bond, Proc. Inter. Cong. Catalysis, 4th, 1968, 2, 266 (1969); I. Yasumori, T. Kabe, and Y. Inoue, J. Phys. Chem. 78, 583 (1974); Y. Inoue, I. Kojima, S. Moriki and I. Yasumori, Proc. Inter. Cong. Catalysis, 6th, 1976, in press; P. J. Estrup, Physics Today 28(4), 33 (1975) and the references therein.

[2]B. Lang, R. W. Joyner, and G. A. Somorjai, Surface Sci. 30, 440, 454 (1972); S. L. Bernasek and G. A. Somorjai, *ibid.* 48, 204 (1975); L. C. Isett and J. M. Blakely, J. Vac. Sci. Tech. 12, 237 (1975).

[3]R. B. McQuistan, Nuovo Cimento B 58, 86 (1968).

[4]R. B. McQuistan, J. Math. Phys. 12, 2113 (1971).

[5]E. Miyazaki, The Sixth International Congress on Catalysis, 1976, in London, Paper No. A-31.

Systems of imprimitivity and representations of quantum mechanics on fuzzy phase spaces*

S. Twareque Ali and E. Prugovečki

Department of Mathematics, University of Toronto, Toronto, Canada M5S 1A1
(Received 18 June 1976)

The problem of expressing quantum mechanical expectation values as averages with respect to nonnegative density functions on phase space, by analogy with classical mechanics, is reexamined in the light of some earlier work on fuzzy phase spaces. It is shown that such phase space representations are possible if ordinary phase space is replaced by a so-called fuzzy phase space, on which the usual marginal distribution functions are redefined to conform to the fact that arbitrarily precise simultaneous measurements on position and momentum are not compatible with quantum mechanics. In the process a generalization of Wigner's theorem on the nonexistence of phase space representations of quantum mechanics, which also satisfy the standard (classical) marginality conditions in position and momentum, is obtained. It is shown that a (continuous) representation of quantum mechanics exists on a given fuzzy phase space if an only if the corresponding confidence functions for position and momentum measurements satisfy the Heisenberg uncertainty relations.

1. INTRODUCTION

In classical statistical mechanics the state of a system is represented by a probability density $\rho(q,p)$, which is a normalized nonnegative real-valued function on a $6n$-dimensional phase space $\Gamma = \mathbb{R}^{6n}$. The expectation value $\langle A\rangle_\rho$ of any observable A in the state ρ is given by the integral

$$\langle A\rangle_\rho = \int_\Gamma A(q,p)\rho(q,p)\,dq\,dp,$$

where $A(q,p)$ is the measurable function on Γ which represents the observable A. Furthermore, the marginal density functions

$$\rho'(q) = \int_{\mathbb{R}^{3n}} \rho(q,p)\,dp$$
$$\rho''(p) = \int_{\mathbb{R}^{3n}} \rho(q,p)\,dq$$

represent, respectively, the configuration and momentum space probability densities of the system.

In quantum mechanics, on the other hand, the state of a system is represented in general by a normalized density matrix ρ which is a positive trace-class operator on a Hilbert space $\mathcal{H}$, and the expectation value of an observable A (which is now a self adjoint operator on $\mathcal{H}$) is given by

$$\langle A\rangle_\rho = \mathrm{Tr}[A\rho].$$

Attempts have often been made (cf. Refs. 1 and 2, and references cited therein) to write quantum mechanical expectation values also as phase space averages with respect to some probability density $\rho(q,p)$. In other words, one tries to find a density function $\rho(q,p)$ for each state ρ and a generalized function $A(q,p)$ for each observable A such that one would have

$$\mathrm{Tr}[A\rho] = \int_\Gamma A(q,p)\rho(q,p)\,dq\,dp.$$

However, it was shown by Wigner (cf., for example, Ref. 3) that if, for all states ρ, the corresponding densities $\rho(q,p)$ are chosen to be positive semidefinite [$\rho(q,p) \geq 0$ for all (q,p) in Γ], then the marginality conditions

$$\int_{\mathbb{R}^{3n}} \rho(q,p)\,dp = \langle q|\rho|q\rangle, \tag{1.1a}$$

$$\int_{\mathbb{R}^{3n}} \rho(q,p)\,dq = \langle p|\rho|p\rangle, \tag{1.1b}$$

cannot be satisfied in general without violating the canonical commutation relations

$$[Q_j, P_k] = i\delta_{jk}$$

for position and momentum. On the other hand, if one drops the positivity condition, Wigner's result no longer applies. Instead, however, one is faced with a tremendous choice[1] of functions $\rho(q,p)$, which, in an attempt to preserve formal appearances, are then labeled as "quasiprobabilities." The nonpositive nature of these quasiprobabilities is then explained away as a quantum effect brought about by the uncertainty relations.[1]

In contradistinction, we propose to show that a physically meaningful and logically consistent way out of this impasse can be found by making use of the very root of the problem—namely, the impossibility of measuring both position and momentum precisely and simultaneously. In other words, it seems plausible that if we do not insist on being able to measure sharp and simultaneous values for position and momentum, and if we therefore replace conventional phase space by a fuzzy phase space,[4,5] we would still be able to obtain a positive semidefinite density $\rho(q,p)$ which would then satisfy a modified marginality condition, appropriate to the fuzzy phase space being used. This is the approach that we shall adopt in this paper. For details on fuzzy spaces, and in particular on fuzzy phase spaces, we refer the reader to Refs. 4–7, and only collect here a few basic facts about fuzzy phase spaces, to set up the notation as well as for future reference.

A fuzzy phase space (Γ, ν) is obtained from the ordinary phase space Γ by replacing each point $(q,p) \in \Gamma$ by a fuzzy point $(\hat{q},\hat{p})$ defined as the four-tuple $((q,\nu_q), (p,\nu'_p))$, where for each q and p, ν_q and ν'_p are normalized *confidence measures*. For any $q \in \mathbb{R}^{3n}$ (resp. $p \in \mathbb{R}^{3n}$) and any Borel set Δ_1 (resp. Δ_2) of $\mathbb{R}^{3n}$, $\nu_q(\Delta_1)$ [resp. $\nu'_p(\Delta_2)$] represents the probability that when the observable Q (resp. P) is measured with a realistic, i.e., imperfectly accurate, apparatus and a reading q (resp. p) is obtained, the actual value of Q (resp. P) is within the set Δ_1 (resp. Δ_2). Thus, a fuzzy phase space is not associated to a system in isolation from the methods and instruments used in performing measurements on that system, but rather to a system *and* a class of instruments whose accuracy calibrations[4,5] yield the chosen confidence measures at each point $(q,p)\in\Gamma$. If ν_q (resp. ν'_p) has a density function χ_q (resp. χ'_p) such that

$$\nu_q(\Delta_1)=\int_{\Delta_1}\chi_q(q')\,dq', \tag{1.2a}$$

$$\nu'_p(\Delta_2)=\int_{\Delta_2}\chi'_p(p')\,dp', \tag{1.2b}$$

for all Borel sets Δ_1, Δ_2 in $\mathbb{R}^{3n}$, we call χ_q (resp. χ'_p) a *confidence function* for q (resp. p). In the present notation, ordinary phase space $\Gamma=\mathbb{R}^{6n}$ is to be associated with the fuzzy phase space (Γ,δ), consisting of four-tuples $[(q,\delta_q),(p,\delta_p)]$, δ_q (resp. δ_p) being the delta measure on $\mathbb{R}^{3n}$ which is centered at q (resp. p). The fuzzy phase space (Γ,ν) is equipped with the same Borel structure as $\Gamma=\mathbb{R}^{6n}$ through the association

$$(q,p)\mapsto((q,\nu_q),(p,\nu'_p)).$$

Our fuzzy phase spaces will be assumed to be transformation spaces under the Galilean group $\mathcal{G}$, acting independently on each particle in the system. A typical element g of $\mathcal{G}$ is given by the transformations

$$\begin{aligned}\mathbf{r}_i&\to\mathbf{r}'_i=R_i\mathbf{r}_i+\mathbf{v}_it+\mathbf{d}_i,\\ \mathbf{k}_i&\to\mathbf{k}'_i=R_i\mathbf{k}_i+m_i\mathbf{v}_i,\\ t&\to t'=t+b,\end{aligned}\tag{1.3}$$

of the position $\mathbf{r}_i=(q_{3i+1},q_{3i+2},q_{3i+3})$ and momentum $\mathbf{k}_i=(p_{3i+1},p_{3i+2},p_{3i+3})$ of the ith particle ($i=0,1,2,\ldots,n-1$), characterized by a rotation R_i, velocity increment $\mathbf{v}_i$, and space translation $\mathbf{d}_i$, and can be considered as mapping the corresponding phase space element (q,p) into another element $([q]g,[p]g)$ computed in accordance with (1.3). The action of g on (Γ,ν) is then assumed to be

$$((q,\nu_q),(p,\nu'_p))\to(([q]g,\nu_{[q]g}),([p]g,\nu'_{[p]g})), \tag{1.4}$$

and, as a consequence,[6]

$$\nu_{[q]g}(\Delta_1)=\nu_q([\Delta_1]g^{-1}), \tag{1.5a}$$

$$\nu'_{[p]g}(\Delta_2)=\nu'_p([\Delta_2]g^{-1}) \tag{1.5b}$$

for all Borel sets Δ_1, Δ_2 in $\mathbb{R}^{3n}$; here $[\Delta_1]g^{-1}$ (resp. $[\Delta_2]g^{-1}$) is the translate of the set Δ_1 (resp. Δ_2) through g^{-1}. Further, $\nu_q=\nu_{q'}$ or $\nu_p=\nu_{p'}$ if and only if $q=q'$ or $p=p'$, respectively.

Once we are prepared to replace the standard phase space by a fuzzy space, we have to replace the marginality conditions (1.1) by the new relations

$$\int_{\mathbb{R}^{3n}}\rho(q,p)\,dp=\int_{\mathbb{R}^{3n}}\nu_q(dq')\langle q'|\rho|q'\rangle, \tag{1.6a}$$

$$\int_{\mathbb{R}^{3n}}\rho(q,p)\,dq=\int_{\mathbb{R}^{3n}}\nu'_p(dp')\langle p'|\rho|p'\rangle, \tag{1.6b}$$

where the right-hand sides of (1.6) represent the probability distributions in the random variables q and p observed with apparatuses whose accuracy calibrations[4,5] at given points q_0 and p_0 are specified by the confidence measures ν_{q_0} and ν'_{p_0}, respectively.

After examining in Sec. 2 an abstract concept of phase space representations of quantum mechanics and relating it to the existence of informationally complete systems of imprimitivity on the phase space Γ, we tackle in Sec. 3 the problem of the existence of representations which obey the marginality conditions (1.6). We obtain a kind of extension of Wigner's result in Theorem 4 by proving that there are no (continuous) representations if the canonically conjugate spreads[4] of the confidence measures ν_q and ν'_p do not satisfy the uncertainty relations, while solutions do exist as soon as these relations are satisfied.

Another crucial result is contained in Theorem 2, which shows that if one starts with an abstract phase space representation of quantum mechanics, which then leads to a system of imprimitivity, specified in terms of a positive operator valued (POV) measure $a(\Delta)$ on Γ, one is mathematically led to a fuzzy phase space structure as soon as one imposes the physically mandatory requirement that the marginal Q and P components of a be informationally equivalent to the conventional spectral measures E^Q and E^P of Q and P, respectively.

These and other points will be further discussed in the concluding Sec. 4.

2. PHASE SPACE REPRESENTATIONS AND SYSTEMS OF IMPRIMITIVITY

In this section we give the formal definitions of a phase space representation of quantum mechanics and of a phase space system of imprimitivity, and derive, in Theorem 1, the relationship between the two. The proof of the theorem is to be found in Appendix A.

Since we shall be dealing with nonrelativistic quantum mechanics for n spinless particles, we shall take for our Hilbert space $\mathcal{H}=L^2(\mathbb{R}^{3n},\,dq)$, and denote by $\mathcal{B}_1(\mathcal{H})$ the normed (under the "trace norm") linear space of all trace class operators on it. The positive cone of $\mathcal{B}_1(\mathcal{H})$, i.e., the set of all positive semidefinite trace class operators on $\mathcal{H}$, will be denoted by $\mathcal{B}_1(\mathcal{H})^+$. By $\mathcal{M}(\Gamma)$ we shall denote the (normed linear) space of all bounded complex measures on the phase space $\Gamma=\mathbb{R}^{6n}$, and by $\mathcal{M}(\Gamma)^+$ the cone of all positive semidefinite bounded measures on Γ. Let $g\to U_g$ be a strongly continuous, unitary, irreducible representation on $\mathcal{H}$ of the Galilean group $\mathcal{G}$ on Γ. We next introduce two concepts which will be basic to the sequel.

Definition 1: A *phase space representation of* the (nonrelativistic) *quantum mechanics* of a spinless, n-particle system is a linear map (over $\mathbb{R}^+$)

$$\pi:\rho\in\mathcal{B}_1(\mathcal{H})^+\to\mu_\rho\in\mathcal{M}(\Gamma)^+ \tag{2.1}$$

which satisfies:

$$\text{(i)}\ \int_\Gamma\mu_\rho(dq\,dp)=\mathrm{tr}\rho,\quad\text{for all }\ \rho\in\mathcal{B}_1(\mathcal{H})^+; \tag{2.2}$$

$$\text{(ii)}\ \mu_{\rho_1}\neq\mu_{\rho_2}\ \text{ iff }\ \rho_1\neq\rho_2; \tag{2.3}$$

$$\text{(iii)}\ \pi[U_g^*\rho U_g]=[\mu_\rho]g, \tag{2.4}$$

where, for any Borel set Δ in Γ,

$$([\mu_\rho]g)(\Delta) = \mu_\rho([\Delta]g^{-1}). \tag{2.5}$$

Definition 2: A *phase space system of imprimitivity* is a positive operator valued (POV) measure a defined on the Borel sets Δ of Γ [and assuming values in the set $\mathcal{B}(\mathcal{H})^+$ of positive bounded operators on $\mathcal{H}$] which satisfies

(i) $a(\Gamma) = I$, (2.6)

where I is the identity element in $\mathcal{B}(\mathcal{H})$,

(ii) $a([\Delta]g) = U_g^* a(\Delta) U_g$, for all $g \in \mathcal{G}$. (2.7)

A phase space system of imprimitivity is said to be *informationally complete*[7] iff the only trace-class operator ρ satisfying $\mathrm{Tr}[a(\Delta)\rho] = 0$ for all Borel sets Δ in Γ is $\rho = 0$; it is said to have a *continuous spectral density* if for each $(q,p) \in \Gamma$ there exists a bounded positive definite operator $F(q,p)$ in $\mathcal{B}(\mathcal{H})^+$ satisfying

$$a(\Delta) = \int_\Delta F(q,p)\, dq\, dp, \tag{2.8}$$

for all Borel sets Δ in Γ, and such that under the Galilean translations $(q,p) \in \mathcal{G}$,

$$U^*_{(q,p)} F(0,0) U_{(q,p)} = F(q,p). \tag{2.9}$$

We shall state now the main result of this section, namely the connection between a phase space representation of quantum mechanics and a phase space system of imprimitivity.

Let $L^1(\Gamma)$ denote the set of absolutely integrable (with respect to the Lebesgue measure $dq\,dp$) complex functions on Γ and $L^1(\Gamma)^+$, the nonnegative functions in this set.

Theorem 1: Every phase space representation of quantum mechanics π determines an informationally complete phase space system of imprimitivity a such that, for all $\rho \in \mathcal{B}_1(\mathcal{H})^+$ and all Borel sets Δ in Γ,

$$[\pi(\rho)](\Delta) = \mathrm{tr}[a(\Delta)\rho] = \int_\Delta f_\rho(q,p)\, dq\, dp, \tag{2.10}$$

where f_ρ is an element in $L^1(\Gamma)^+$. The map $\rho \mapsto f_\rho$ is a linear isometry. Conversely, every informationally complete phase space system of imprimitivity determines, canonically through (2.10), a phase space representation of quantum mechanics.

Furthermore, a has a continuous spectral density $F(q,p)$, if and only if f_ρ in (2.10) is a continuous function in $L^1(\Gamma)^+$ for each $\rho \in \mathcal{B}_1(\mathcal{H})^+$.

In view of this result we shall adopt the following definition.

Definition 3: The phase space representation of quantum mechanics π will be said to be *continuous* if and only if its canonically associated system of imprimitivity [Eq. (2.10)] has a continuous spectral density.

Having thus succeeded in setting up a one-to-one correspondence between phase space representations of quantum mechanics and informationally complete phase space systems of imprimitivity, we shall exploit, in the next section, the mathematical properties of the latter to relate phase space systems of imprimitivity to fuzzy phase spaces. The analysis is intended to show how physical restrictions, namely the marginality conditions, lead to the existence of fuzzy phase spaces which are Borel isomorphic to Γ (Theorem 2). Otherwise, in the absence of such marginality conditions, we could not venture to make any suggestions as to a physical interpretation of the measure μ_ρ on Γ associated with a state ρ, since there would be no assurance that such an interpretation would conform to the conventional interpretation of quantum mechanics, in the realm of validity of the latter. In fact, our concept of phase space representations of quantum mechanics essentially coincides with that proposed by Agarwal and Wolf,[1] except that the Agarwal–Wolf Ω rules of association, translated into our terms, would only require μ_ρ to be a signed (real) measure, when $\rho \in \mathcal{B}_1(\mathcal{H})^+$, rather than a positive semidefinite one. To make the correspondence with the Agarwal–Wolf approach clear we note that the existence of the POV measure $a(\Delta)$ enables us to associate to any complex valued Borel measurable function $g(q,p)$ the operator (in the notation of Refs. 1 and 2)

$$\hat{g}(\hat{q},\hat{p}) = \int_\Gamma g(q,p)\, a(dq\, dp), \tag{2.11}$$

which in our case is positive semidefinite if $g(q,p) \geq 0$.

To end this section we mention two examples of normalized POV measures on Γ, which would lead to linear maps $\pi : \mathcal{B}_1(\mathcal{H})^+ \to \mathcal{M}(\Gamma)^+$ through Eq. (2.10). However, in the first case one of the marginality conditions in (1.6) is violated, whereas in the second, the covariance condition (2.7), and hence (2.4), is not satisfied (such a POV measure could arise if, for example, one tried to compute joint probabilities for the outcome of successive measurements on the observables Q and P, following the suggestion of Davies and Lewis[8]).

Let χ_q and χ'_p be two sets of confidence functions on $\mathbf{R}^{3n}$ which satisfy

$$\chi_q(q') = \chi_0(q' - q), \tag{2.12a}$$

$$\chi'_p(p') = \chi'_0(p' - p). \tag{2.12b}$$

Let E^Q and E^P be the usual spectral measures for the operators Q and P, respectively, and let

$$E^Q(\chi_q) = \int_{\mathbf{R}^{3n}} \chi_q(q') E^Q(dq'), \tag{2.13a}$$

$$E^P(\chi'_p) = \int_{\mathbf{R}^{3n}} \chi'_p(p') E^P(dp'). \tag{2.13b}$$

(i) Let $\chi_q^{1/2}$ be the function defined as

$$\chi_q^{1/2}(q') = [\chi_q(q')]^{1/2}, \tag{2.14}$$

for all $q' \in \mathbf{R}^{3n}$. Then the spectral density

$$F(q,p) = E^Q(\chi_q^{1/2})\, E^P(\chi'_p)\, E^Q(\chi_q^{1/2}) \tag{2.15}$$

defines, through (2.8), a POV measure on Γ for which the associated linear map π leads to the marginality condition (1.6a) but *not* (1.6b). On the other hand, $F(q,p)$ satisfies the covariance condition (2.9).

(ii) Let $\rho_0 \in \mathcal{B}_1(\mathcal{H})^+$, and suppose that $\mathrm{Tr}\rho_0 = 1$. Then the spectral density

$$F(q,p)=\mathrm{Tr}[E^Q(\chi_q)\rho_0]E^P(\chi'_p) \qquad (2.16)$$

defines a POV measure which violates the covariance condition (2.9), but satisfies both the marginality conditions in (1.6).

3. QUANTUM MECHANICS ON FUZZY PHASE SPACES

In the Introduction we had defined fuzzy phase spaces without assuming any relationship to exist between the ν_q's and the ν'_q's. However, it is clear that the quantum mechanical uncertainty relations will set at least some limitations upon the distribution of values for the ν_q's in relation to those for the ν'_p's, and vice versa. A first result of this nature is proved in Theorem 3, which puts a restriction on the spectral density $F(q,p)$ for a continuous phase space representation of quantum mechanics satisfying either one of the two marginality conditions in Eqs. (1.6). This result then leads easily to Theorem 4, which discusses the existence and uniqueness of continuous phase space representations of quantum mechanics satisfying both the marginality conditions in (1.6). Theorem 4 also achieves a generalization of Wigner's result[2,3] (under the additional hypothesis of covariance under the Galilean group) mentioned in the Introduction and says, in particular, that no continuous representations exist, on fuzzy phase spaces, which violate the quantum mechanical uncertainty principle. Again we defer proofs of mathematical results to Appendix B. We begin with some definitions.

Definition 4: The *marginal Q component* of a phase space system of imprimitivity a is the normalized POV measure a^Q defined on the Borel sets Δ_1 of $\mathbb{R}^{3n}$ by

$$a^Q(\Delta_1)=\int_{\mathbb{R}^{3n}} a(\Delta_1\times dp). \qquad (3.1a)$$

Similarly, the *marginal P* component of a is the normalized POV measure a^P on $\mathbb{R}^{3n}$ defined as

$$a^P(\Delta_2)=\int_{\mathbb{R}^{3n}} a(dq\times\Delta_2). \qquad (3.1b)$$

Clearly, $a^Q(\Delta_1)$ is a system of imprimitivity on the configuration space with respect to the translations q in that space, while $a^P(\Delta_2)$ is a system of imprimitivity on the momentum space with respect to the translations p in that space, i.e.,

$$a^Q([\Delta_1]q)=U^*_{(q,0)}a^Q(\Delta_1)U_{(q,0)}, \qquad (3.2a)$$

for all $(q,0)\in\mathcal{G}$, and Borel sets Δ_1 in $\mathbb{R}^{3n}$; and

$$a^P([\Delta_2]p)=U^*_{(0,p)}a^P(\Delta_2)U_{(0,p)}, \qquad (3.2b)$$

for all $(0,p)\in\mathcal{G}$, and Borel sets Δ_2 in $\mathbb{R}^{3n}$.

The POV measures a^Q and a^P are examples of fuzzy position and momentum observables.[9] Let Q and P be the usual position and momentum operators, respectively, on $\mathcal{H}$ for our n-particle quantum mechanical system. Let E^Q and E^P be their corresponding spectral measures. Following Refs. 6 and 7 we introduce the following definition.

Definition 5: a^Q is said to be *informationally equivalent* to E^Q if and only if for any pair of vectors $\phi,\psi\in\mathcal{H}$, the equality

$$\langle\phi\,|\,a^Q(\Delta_1)\phi\rangle=\langle\psi\,|\,a^Q(\Delta_1)\psi\rangle,$$

for all Borel sets Δ_1 in $\mathbb{R}^{3n}$ implies the equality

$$\langle\phi\,|\,E^Q(\Delta_1)\phi\rangle=\langle\psi\,|\,E^Q(\Delta_1)\psi\rangle$$

for all Borel sets Δ_1 in $\mathbb{R}^{3n}$.

An analogous statement defines the informational equivalence of a^P and E^P.

The next theorem shows that the informational equivalence of a^Q to E^Q and of a^P to E^P (which has to be imposed on physical grounds if a^Q and a^P are to differentiate between states equally effectively as E^Q and E^P, respectively, do) automatically leads to a fuzzy-phase-space structure. In stating this theorem we limit ϕ to the [dense in $L^2(\mathbb{R}^{3n})$] Schwartz space $\mathcal{S}(\mathbb{R}^{3n})$ in order to insure that the integrals in (3.3) are well defined even if ν_q and ν'_p are not absolutely continuous with respect to the Lebesgue measure. However, it will be established in Theorem 3 that this contingency does not occur for continuous representations, so that for this case of exclusive physical interest (cf. Sec. 4) one can allow ϕ to vary over the entire space $L^2(\mathbb{R}^{3n})$.

Theorem 2: Let a be a phase space system of imprimitivity and a^Q and a^P its marginal Q and P components, respectively. Suppose a^Q (resp. a^P) is informationally equivalent to E^Q (resp. E^P). Then a determines a fuzzy phase space (Γ,ν) that is Borel isomorphic to Γ, and it satisfies the marginality conditions

$$\langle\phi\,|\,a^Q(\Delta_1)\phi\rangle=\int_{\Delta_1}dq\int_{\mathbb{R}^{3n}}\nu_q(dq')\,|\,\phi(q')\,|^2 \qquad (3.3a)$$

$$\langle\phi\,|\,a^P(\Delta_2)\phi\rangle=\int_{\Delta_2}dp\int_{\mathbb{R}^{3n}}\nu'_p(dp')\,|\,\tilde{\phi}(p')\,|^2, \qquad (3.3b)$$

for all $\phi\in\mathcal{S}(\mathbb{R}^{3n})$, where $\tilde{\phi}$ denotes the Fourier transform of ϕ.

It follows, therefore, that any phase space representation of quantum mechanics π, whose canonically associated phase space system of imprimitivity has marginal Q and P components which are informationally equivalent to E^Q and E^P, respectively, is actually a representation on a fuzzy phase space and satisfies the marginality conditions (1.6).

The question now arises as to what extent a $\mathcal{G}$ covariant fuzzy phase space (Γ,ν) which is Borel isomorphic to Γ canonically determines an informationally complete phase space system of imprimitivity, and hence its own, possibly unique, phase space representation of quantum mechanics. This question is partially answered in Theorem 4 below. In Theorem 3 we first get a restriction, to be put on the ν_q's and ν'_p's provided they are to come from a continuous phase space representation of quantum mechanics.

Theorem 3: Let a be a phase space system of imprimitivity with a continuous spectral density $F(q,p)$, and suppose that a^Q satisfies the marginality relation

$$\langle\phi\,|\,a^Q(\Delta_1)\phi\rangle=\int_{\Delta_1}dq\int_{\mathbb{R}^{3n}}\nu_q(dq')\,|\,\phi(q')\,|^2$$

for all vectors ϕ in $\mathcal{S}(\mathbb{R}^{3n})$ and all Borel sets Δ_1 in $\mathbb{R}^{3n}$. Then a^Q is informationally equivalent to E^Q, the operator $F(q,p)$ is of trace class and ν_q is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{3n}$, so that for all Borel sets Δ_1 in $\mathbb{R}^{3n}$

$$\nu_q(\Delta_1)=\int_{\Delta_1}\chi_q(q')\,dq',$$

where $\chi_q\in L^1(\mathbb{R}^{3n})$ for each $q\in\mathbb{R}^{3n}$, and

$$\chi_q(q')=\chi_0(q'-q), \tag{3.4}$$

for almost all q' and all q in $\mathbb{R}^{3n}$; furthermore,

$$\langle q\,|\,F(0,0)\,|\,q\rangle=(2\pi)^{-3n}\chi_0(q). \tag{3.5}$$

Similarly, if a^P satisfies the marginality relation

$$\langle\phi\,|\,a^P(\Delta_2)\phi\rangle=\int_{\Delta_2}dp\int_{\mathbb{R}^{3n}}\nu_p'(dp')\,|\,\tilde\phi(p')\,|^2,$$

for all $\phi\in\mathcal{S}(\mathbb{R}^{3n})$ (whose Fourier transforms are given by $\tilde\phi$), and all Borel sets Δ_2 in $\mathbb{R}^{3n}$, then a^P is informationally equivalent to E^P, the operator $F(q,p)$ is of trace class and ν_p' is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{3n}$. Again, for all Borel sets Δ_2 in $\mathbb{R}^{3n}$,

$$\nu_p'(\Delta_2)=\int_{\Delta_2}\chi_p'(p')\,dp',$$

where $\chi_p'\in L^1(\mathbb{R}^{3n})$,

$$\chi_p'(p')=\chi_0'(p'-p) \tag{3.6}$$

for almost all p' and all p in $\mathbb{R}^{3n}$, and

$$\langle p\,|\,F(0,0)\,|\,p\rangle=(2\pi)^{-3n}\chi_0'(p). \tag{3.7}$$

We see that Wigner's theorem, in the form stated in Ref. 3, and under the additional hypothesis of covariance under the Galilean group, is already contained in this result, even apart from the restriction imposed upon $F(q,p)$ by the uncertainty principle; for according to the hypothesis of that theorem a^Q would have to satisfy a marginality condition of the type

$$\begin{aligned}\langle\phi\,|\,a^Q(\Delta_1)\phi\rangle&=\int_{\Delta_1}dq\int_{\mathbb{R}^{3n}}\delta_q(dq')\,|\,\phi(q')\,|^2dq'\\&=\int_{\Delta_1}dq\,|\,\phi(q)\,|^2.\end{aligned} \tag{3.8}$$

However, the δ measure is not absolutely continuous with respect to the Lebesgue measure. In other words, *either one* of the two marginality conditions

$$\int_{\mathbb{R}^{3n}}\rho(q,p)\,dp=\langle q\,|\,\rho\,|\,q\rangle,$$

$$\int_{\mathbb{R}^{3n}}\rho(q,p)\,dq=\langle p\,|\,\rho\,|\,p\rangle,$$

is enough to preclude the existence of a continuous phase space representation of quantum mechanics.

From Theorem 3 we can easily derive the following result, whose first part represents an extension of Wigner's theorem.

Theorem 4: Suppose $\chi_0(q)$ and $\chi_0'(p)$ are nonnegative normalized functions from $L^1(\mathbb{R}^{3n})$ whose spreads[4]

$$s_j=\Big[2\int_{\mathbb{R}^{3n}}(q_j-\bar q_j)^2\chi_0(q)\,dq\Big]^{1/2}, \tag{3.9a}$$

$$r_j=\Big[2\int_{\mathbb{R}^{3n}}(p_j-\bar p_j)^2\chi_0'(p)\,dp\Big]^{1/2}, \tag{3.9b}$$

exist for $j=1,2,\ldots,3n$, where $\bar q_j$ and $\bar p_j$ denote the respective mean values of q_j and p_j. If $s_{j_0}r_{j_0}<1$, for some $j=j_0$, then there exists no continuous phase space system of imprimitivity satisfying the marginality conditions (3.3). If

$$r_{3i+1}=r_{3i+2}=r_{3i+3}\geq s_{3i+1}^{-1}=s_{3i+2}^{-1}=s_{3i+3}^{-1}, \tag{3.10}$$

$i=0,1,2,\ldots,n-1$, there is at least one pair χ_0 and χ_0' for which such a system of imprimitivity exists. When (3.10) is an equality, the only such pair is

$$\chi_q(q')=\pi^{-3n/2}\prod_{j=1}^{3n}s_j^{-1}\exp[-s_j^{-2}(q_j'-q_j)^2], \tag{3.11a}$$

$$\chi_p'(p')=\pi^{-3n/2}\prod_{j=1}^{3n}s_j\exp[-s_j^2(p_j'-p_j)^2], \tag{3.11b}$$

and the corresponding system of imprimitivity is unique, namely it is the one having the spectral density

$$F(q,p)=(2\pi)^{-3n}\,|\,\phi_{q,p}^{(s)}\rangle\langle\phi_{q,p}^{(s)}\,|, \tag{3.12}$$

where

$$\phi_{q,p}^{(s)}(x)=\pi^{-3n/4}\prod_{j=1}^{3n}s_j^{-1/2}\times\exp\left[-\frac{(x_j-q_j)^2}{2s_j^2}+ip_j\left(x_j-\frac{q_j}{2}\right)\right]; \tag{3.13}$$

furthermore, this system of imprimitivity is informationally complete.

The above theorem is an immediate consequence of (3.5) and (3.7) and the well-known fact[10] that there is no trace class operator on $\mathcal{H}$ which would satisfy both these relations if the product $2^{-1}s_jr_j$ of its standard deviations $2^{-1/2}s_j$ and $2^{-1/2}r_j$ in the canonically conjugate variables q_j and p_j is smaller than one-half, while for the case where (3.10) is satisfied there is such a trace class operator, which is unique, namely (3.12), when (3.10) is an equality. (The equalities $r_{3i+1}=r_{3i+2}=r_{3i+3}$ are a consequence of the rotational symmetry intrinsic to Galilean invariance.)

Implicit in Theorem 3 is the method of construction of all those continuous phase space representations of n-particle quantum mechanics whose marginal Q and P components are informationally equivalent to the configuration and momentum representations, respectively (i.e., in accordance with Theorem 2, of *fuzzy phase space representations*). These representations are obtained by taking any density matrix ρ_0, which is rotationally invariant, and writing

$$a(\Delta)=(2\pi)^{-3n}\int_\Delta\exp(-ipQ)\exp(-iqP)\rho_0\times\exp(iqP)\exp(ipQ)\,dq\,dp. \tag{3.14}$$

The resulting POV measure $a(\Delta)$ will obviously give rise to a system of imprimitivity on Γ, whose informational completeness has to be checked in each individual case. In the special case of (3.12) this informational completeness may be derived[5] by using, for example, the well-known analyticity properties displayed by the inner products of arbitrary vectors and coherent states. More generally if ρ_0 is of the form

$$\rho_0=|\,\phi_0\rangle\langle\phi_0\,|, \tag{3.15}$$

where ϕ_0 is a vector in $L^2(\mathbb{R}^{3n})$ for which the "reproducing kernel"

$$K(q,p\,;\,q',p')=\langle\phi_0|\,U_{(q-q',p-p')}\phi_0\rangle \tag{3.16}$$

never vanishes, then the corresponding system of imprimitivity is informationally complete.[11]

Theorem 3 also implies that if the confidence functions $\chi_0(q)$ and $\chi_0'(p)$ are given *a priori* (e.g., by the accuracy calibrations of a class of instruments for measuring Q and P), then the choice of ρ_0 will be unique if and only if there is a unique trace class operator $F(0,0)=(2\pi)^{-3n}\rho_0$ which satisfies both (3.5) and (3.7). If $\chi_0(q)$ and $\chi_0'(p)$ are optimal [i.e., have minimal spreads (3.10) in relation to the uncertainty principle], then this uniqueness is implicitly stated in Theorem 4. However, this uniqueness is not a feature that is common to all choices of χ_0 and χ_0', as can be shown by counterexamples (cf. Ref. 7, Sec. 2).

4. DISCUSSION

According to Definition 1, a phase space representation of the quantum mechanics of a given system assigns to each quantum mechanical state (i.e., density operator) ρ a unique normalized measure μ_ρ on Γ. There is, however, no physical input in that observation as yet. One might be tempted to interpret $\mu_\rho(\Delta)$ as the probability of a measurement outcome (q,p) falling within $\Delta\subset\Gamma$, when the system is in the state ρ. But even if we disregard the unfeasibility of experimentally determining sharp values $(q,p)\in\Gamma$, we are still left with a question of consistency: Since there is an (uncountable) infinity of possible phase space representations, each such representation assigning, in general, a different value $\mu_\rho(\Delta)$ to the same set Δ, for the system in the same state ρ, which one of these distinct probabilities is the "correct" one?

Following this line of thought, one can seek a way out of the impasse by looking for guidance from the Q marginal values $\mu_\rho(\Delta_1\times\mathbb{R}^{3n})$ and P marginal values $\mu_\rho(\mathbb{R}^{3n}\times\Delta_2)$ of μ_ρ, since that is familiar territory, where one is dealing with either Q measurements or with P measurements separately. But according to Wigner's theorem (contained in Theorem 4 above), no continuous phase space representation of quantum mechanics would provide us with marginal probabilities which coincide with the conventional ones, namely $\mathrm{Tr}[\rho E^Q(\Delta_1)]$ and $\mathrm{Tr}[\rho E^P(\Delta_2)]$, respectively, obtained by making precise and separate measurements of Q alone and P alone. Faced with this fact, one can still persist in the attempt of assigning some physical interpretation to the marginal values of μ_ρ by considering the POV measure $a(\Delta)$ that is unambiguously attached (according to Theorem 1) to each phase space representation, and then looking for probabilistic interpretations which would make its marginal components a^Q and a^P informationally equivalent (Definition 5) to E^Q and E^P, respectively. This last demand is certainly not only reasonable, but also mandatory, since, whatever the new interpretation, it should be at least as effective in distinguishing between the physical properties of different states as the conventional interpretation was.

In carrying out this last step of the program one is led to singling out from among all phase space representations a species of special ones, namely, those which satisfy the above mentioned criterion of informational equivalence. Using Theorem 2, we arrive thus at the conclusion that the Q and P marginal values of the measure μ_ρ on Γ, assigned to ρ by each one of these special representations can be interpreted as probabilities for *fuzzy* measurements of Q and P, respectively.

Armed with this fact, one can go back and reconsider μ_ρ globally (on Γ) for each one of these *fuzzy* phase space representations. It is then natural to interpret $\mu_\rho(\Delta)$ as being the probability for the outcome of a *fuzzy simultaneous measurement* of Q and P, with the system in the state ρ, to "fall within the fuzzy set Δ" (cf. Ref. 4 for details). The compatibility of such an interpretation with the uncertainty relations is then confirmed by Theorem 4, which states that there are no fuzzy representations which violate the uncertainty principle. Also, it follows from the analysis of fuzzy position and momentum operators in Ref. 6 that the fuzzy position operators $\tilde{Q}$ associated to a^Q and the fuzzy momentum operators $\tilde{P}$ associated to a^P,

$$\tilde{Q}_j=\int_{\mathbb{R}^{3n}}q_j\,a^Q(dq),\quad \tilde{P}_j=\int_{\mathbb{R}^{3n}}p_j\,a^P(dp),$$

satisfy the same canonical commutation relations as do the standard (or sharp) operators Q and P, whenever the confidence measures ν_0 and ν_0' have finite spreads (i.e., whenever $\tilde{Q}$ and $\tilde{P}$ exist). Furthermore, an analysis of the operational meaning of the simultaneous (fuzzy) measurement of Q and P confirms the experimental feasibility of backing up the present interpretation with empirical observations (cf. Ref. 4, and the references cited therein).

Among all fuzzy phase space representations of the quantum mechanics of a given system, a prominent role is played by the *optimal* ones, i.e., the ones which correspond to confidence measures ν_q and ν_p' which have minimal spreads (3.10) in relation to the uncertainty principle. According to the second part of Theorem 4, such optimally fuzzy phase space representations exist if and only if ν_q and ν_p' are derivable from confidence functions χ_q and χ_p', respectively, which are the Gaussians in (3.11). Furthermore, for given ν_q and ν_p' the corresponding phase space system of imprimitivity is unique—being given by the spectral density (3.12). It is precisely these optimal representations that have been studied in some detail in Refs. 4 and 5, and have been shown to give rise to the $L^2(\Gamma)$ spaces which can be interpolated in a natural manner between the configuration space and the momentum space representations of a given quantum mechanical system. In addition, in relation to these optimal representations, the nonoptimal ones play the same role as a fuzzy phase space representation of classical mechanics plays vis-à-vis the conventional (sharp) representation (cf. Ref. 5 for details).

There is another very significant feature that is shared by all phase space representations of quantum mechanics, and which establishes their superiority from the point of view of state resolution: the POV measure $a(\Delta)$ associated with any such representation (Theorem 1) is informationally complete, while, on the other hand, neither $E^Q(\Delta_1)$ nor $E^P(\Delta_2)$ [and therefore, by Theorem 2, neither $a^Q(\Delta_1)$ nor $a^P(\Delta_2)$] come even close to having this important property. Moreover, as shown in Ref. 7 by a counter example, not even $E^Q(\Delta_1)$ and $E^P(\Delta_2)$ taken together are informationally complete! Thus, if we impose the condition that quantum mechanical theories be not redundant (in the sense of not possessing states which cannot be distinguished, even in principle, by means of empirical observations), then position and momentum observations, performed separately, cannot ensure that. However, simultaneous position *and* momentum observations do eliminate the possibility of redundancy, when related to ρ by means of the above discussed interpretation of $\mu_\rho(\Delta)$ for fuzzy phase space representations.

As a final comment, we ought to mention that our specialization to the case of *continuous* phase space representations, from Theorem 3 onwards, might appear somewhat artificial. However, the following serves as a physical justification for imposing this restriction: By Eqs. (2.8)–(2.10) and the fact that, as a trace-class operator, $F(0,0)$ has a completely discrete spectrum it follows that for a continuous representation

$$|f_\rho(q,p)| \leq \mathrm{tr}F(0,0) = (2\pi)^{-3n} \tag{4.1}$$

for all density matrices ρ with unit trace [cf. Eq. (3.5)]. Conversely, such a relationship implies the continuity of the corresponding representation. The physical desirability of the above inequality becomes apparent if we put back factors of $\hbar$, for then we get

$$|f_\rho(q,p)| \leq h^{-3n}. \tag{4.2}$$

As a matter of fact, after recalling the traditional role of h^{3n} in the statistical mechanics considerations of a system of $3n$ degrees of freedom[12] and our interpretation of $f_\rho(q,p)$ as a probability density on (Γ,ν), it becomes evident that (4.2) expresses the impossibility of locating a quantum system in a phase space cell of volume less than h^{3n}.

APPENDIX A

The proof of Theorem 1 will depend on the following three lemmas:

Lemma 1: The map π may be extended to a linear isometry

$$\pi: \mathcal{B}_1(\mathcal{H}) \to \mathcal{M}(\Gamma) \tag{A1}$$

which is bijective onto its image in $\mathcal{M}(\Gamma)$. [$\mathcal{B}_1(\mathcal{H})$ is considered as a Banach space with respect to the "trace norm" $\|\cdot\|_1$, and $\mathcal{M}(\Gamma)$ as a Banach space with the norm it acquires as the dual of $C_\infty(\Gamma)$—the set of all bounded, complex, continuous functions on Γ which vanish at infinity.]

Proof: Both $\mathcal{B}_1(\mathcal{H})$ and $\mathcal{M}(\Gamma)$ are generated by their positive cones $\mathcal{B}_1(\mathcal{H})^+$ and $\mathcal{M}(\Gamma)^+$, respectively. Consider first

$$\pi: \mathcal{B}_1(\mathcal{H})^+ \to \mathcal{M}(\Gamma)^+. \tag{A2}$$

Since $\mathrm{Tr}\rho = \|\rho\|_1$ for $\rho \in \mathcal{B}_1(\mathcal{H})^+$ and $\mu(\Gamma) = \|\mu\|_{\mathcal{M}}$ for $\mu \in \mathcal{M}(\Gamma)^+$, and in view of the condition that $\pi(\rho_1)=\pi(\rho_2)$ iff $\rho_1=\rho_2$, it follows that π is a bijective isometry of $\mathcal{B}_1(\mathcal{H})^+$ onto its image $\pi[\mathcal{B}_1(\mathcal{H})^+]$ in $\mathcal{M}(\Gamma)^+$ under π. Hence $\pi[\mathcal{B}_1(\mathcal{H})^+]$ is closed. Next, for any $\rho \in \mathcal{B}_1(\mathcal{H})$, let $\rho = \rho_1 - \rho_2 + i(\rho_3 - \rho_4)$, where $\rho_1, \rho_2, \rho_3, \rho_4 \in \mathcal{B}_1(\mathcal{H})^+$. We define

$$\pi(\rho) = \pi(\rho_1) - \pi(\rho_2) + i\pi(\rho_3) - i\pi(\rho_4), \tag{A3}$$

to extend π to the whole of $\mathcal{B}_1(\mathcal{H})$. Then $\pi[\mathcal{B}_1(\mathcal{H})] \subset \mathcal{M}(\Gamma)$ is a closed linear subspace of $\mathcal{M}(\Gamma)$, and so π extends to a one-to-one positive linear map from $\mathcal{B}_1(\mathcal{H})$ onto $\pi[\mathcal{B}_1(\mathcal{H})]$. Hence π is continuous,[13] and since for $\rho \in \mathcal{B}_1(\mathcal{H})^+$, $\|\pi(\rho)\|_{\mathcal{M}} = \|\rho\|_1$, it follows that for all $\rho \in \mathcal{B}_1(\mathcal{H})$

$$\|\pi(\rho)\|_{\mathcal{M}} \leq \|\rho\|_1. \tag{A4}$$

On the other hand, the inverse map π^{-1} from the Banach space $\pi[\mathcal{B}_1(\mathcal{H})]$ to $\mathcal{B}_1(\mathcal{H})$ is also continuous, being a positive linear map. Thus,

$$\|\rho\|_1 \leq \|\pi(\rho)\|_{\mathcal{M}}, \tag{A5}$$

so that π is isometric.

Lemma 2: Let a be a phase space system of imprimitivity. Then $a(\Delta)=0$ iff Δ is of Lebesgue measure zero.

Proof: Let $\psi_1, \psi_2, \cdots$ be an orthonormal basis of vectors in $\mathcal{H}$. Consider the probability measure

$$\mu(\Delta) = \sum_{n=1}^{\infty} \frac{1}{2^n} \langle \psi_n | a(\Delta)\psi_n \rangle, \tag{A6}$$

defined on the Borel sets Δ of Γ. Then if g is an element of the Galilean group $\mathcal{G}$, we have

$$\mu([\Delta]g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \langle \psi_n | a([\Delta]g)\psi_n \rangle = \sum_{n=1}^{\infty} \frac{1}{2^n} \langle U_g\psi_n | a(\Delta) U_g\psi_n \rangle, \tag{A7}$$

by virtue of the imprimitivity relation (2.7). Thus, $\mu(\Delta)=0$ implies $a(\Delta)=0$, which in turn implies, because of (A7), that $\mu([\Delta]g)=0$, for all $g\in\mathcal{G}$. Hence, the null sets of μ are invariant under g, and it follows that μ is equivalent to the invariant Lebesgue measure on Γ. But by (A6), $\mu(\Delta)=0$ iff $a(\Delta)=0$, whence the result.

Lemma 3: Every informationally complete phase space system of imprimitivity a defines an isometric positive linear map

$$\hat{\pi}: \mathcal{B}_1(\mathcal{H}) \to L^1(\Gamma), \tag{A8}$$

which is a bijection onto its image $\hat{\pi}[\mathcal{B}_1(\mathcal{H})] \subset L^1(\Gamma)$; $\hat{\pi}[\mathcal{B}_1(\mathcal{H})]$ is thus a closed subspace of the Banach space $L^1(\Gamma)$. Furthermore, a has a continuous spectral density iff $\hat{\pi}(\rho)$ defines a continuous function in $L^1(\Gamma)$ for all $\rho\in\mathcal{B}_1(\mathcal{H})$.

Proof: For any $\rho\in\mathcal{B}_1(\mathcal{H})$, consider the measure $\mu_\rho(\Delta) = \mathrm{tr}[\rho a(\Delta)]$. By Lemma 2, $\mu_\rho(\Delta)$ is absolutely continuous with respect to the Lebesgue measure. Hence there is an $f_\rho \in L^1(\Gamma)$ such that

$$\mathrm{Tr}[\rho a(\Delta)] = \int_\Delta f_\rho(q,p)\, dq\, dp, \tag{A9}$$

and

$$\mathrm{Tr}\rho = \int_\Gamma f_\rho(q,p)\, dq\, dp. \tag{A10}$$

Set

$$\hat{\pi}(\rho) = f_\rho. \tag{A11}$$

By informational completeness $\hat{\pi}$ is one-to-one. It is obviously also linear and positive, and in fact, in view of (A10), for $\rho\in\mathcal{B}_1(\mathcal{H})^+$

$$\|\hat{\pi}(\rho)\|_{L^1}=\|\rho\|_1. \tag{A12}$$

We may now prove, exactly as we did for π in Lemma 1, that the isometry $\hat{\pi}$ also extends to the whole of $\mathcal{B}_1(\mathcal{H})$.

To prove the rest of the lemma, we show first that the map $g\to U_g^*\rho U_g$ is continuous in the trace norm topology of $\mathcal{B}_1(\mathcal{H})$. Let $\mathcal{B}(\mathcal{H})$ be the set of all bounded operators on $\mathcal{H}$, equipped with the usual operator norm. Then it is well known[14] that $\mathcal{B}(\mathcal{H})$ is the dual of $\mathcal{B}_1(\mathcal{H})$, so that, in particular, for $\rho\in\mathcal{B}_1(\mathcal{H})$,

$$\|\rho\|_1=\sup_{\substack{\|A\|\le 1\\ A\in\mathcal{B}(\mathcal{H})}}|\mathrm{Tr}[A\rho]|. \tag{A13}$$

Let $\phi,\psi\in\mathcal{H}$ and $A\in\mathcal{B}(\mathcal{H})$. Then,

$$\begin{aligned}|\langle\psi|[U_gAU_g^*-A]\phi\rangle| &\le |\langle\psi|U_gA[U_g^*-I]\phi\rangle| \\ &\quad+|\langle\psi|[U_g-I]A\phi\rangle| \\ &\le\|\psi\|\|A\|\|(U_g^*-I)\phi\| \\ &\quad+\|\phi\|\|A\|\|(U_g^*-I)\psi\|\end{aligned} \tag{A14}$$

and the last expression converges to zero as $g\to e$, by virtue of the strong continuity of $g\to U_g$. From (A14) it also follows that the weak continuity of $g\to U_gAU_g^*$ is actually uniform in A for $\|A\|\le 1$. Thus,

$$\sup_{\substack{\|A\|\le 1\\ A\in\mathcal{B}(\mathcal{H})}}|\langle\psi|[U_gAU_g^*-A]\phi\rangle|\to 0,\quad g\to e. \tag{A15}$$

But since the weak and the ultraweak topologies on $\mathcal{B}(\mathcal{H})$ coincide on bounded sets, we have, for all $\rho\in\mathcal{B}_1(\mathcal{H})$, the result that

$$\sup_{\substack{\|A\|\le 1\\ A\in\mathcal{B}(\mathcal{H})}}|\mathrm{Tr}[(U_gAU_g^*-A)\rho]|\to 0,\quad g\to e, \tag{A16}$$

i.e., $\|U_g^*\rho U_g-\rho\|_1\to 0,\quad g\to e.$ (A17)

Since $\mathcal{G}$ is a group, (A17) proves the continuity of $g\mapsto U_g^*\rho U_g$ at all points $g\in\mathcal{G}$.

Next suppose that a has the continuous spectral density $F(q,p)$ and for any $\rho\in\mathcal{B}_1(\mathcal{H})$ define f_ρ through (A11). Then, by (A9) and (2.8)

$$f_\rho(q,p)=\mathrm{Tr}[F(q,p)\rho], \tag{A18}$$

so that, by (2.9)

$$\begin{aligned}&|f_\rho(q,p)-f_\rho(0,0)| \\ &=|\mathrm{Tr}[F(q,p)\rho]-\mathrm{Tr}[F(0,0)\rho]| \\ &=|\mathrm{Tr}[(U_{(q,p)}^*\rho U_{(q,p)}-\rho)F(0,0)]| \\ &\le\|U_{(q,p)}^*\rho U_{(q,p)}-\rho\|_1\|F(0,0)\|,\end{aligned} \tag{A19}$$

which converges to zero as $(q,p)\to(0,0)$ by virtue of (A17).

Conversely, suppose that for all $\rho\in\mathcal{B}_1(\mathcal{H})$, f_ρ is a continuous function in $L^1(\Gamma)$. For fixed (q,p) define a linear functional on $\mathcal{B}_1(\mathcal{H})$ by

$$\rho\mapsto f_\rho(q,p). \tag{A20}$$

This is clearly a positive linear mapping and therefore defines a continuous functional on $\mathcal{B}_1(\mathcal{H})$. Thus there exists an operator $F(q,p)$ in $\mathcal{B}(\mathcal{H})^+$ such that

$$\mathrm{Tr}[F(q,p)\rho]=f_\rho(q,p). \tag{A21}$$

Finally (A9) implies that the relation (2.8) holds.

Proof of Theorem 1: Let π be a phase space representation of quantum mechanics, extended to the whole of $\mathcal{B}_1(\mathcal{H})$ by Lemma 1. Let $\pi(\rho)=\mu_\rho\in\mathcal{M}(\Gamma)$. Then, for a fixed Borel set Δ in Γ, $\rho\mapsto\mu_\rho(\Delta)$ is a positive linear functional on $\mathcal{B}_1(\mathcal{H})$. Thus there is an operator $a(\Delta)\in\mathcal{B}(\mathcal{H})^+$ such that

$$\mu_\rho(\Delta)=\mathrm{tr}[a(\Delta)\rho]. \tag{A22}$$

Since $\mu_\rho\in\mathcal{M}(\Gamma)$ for each ρ and $\mu_\rho(\Gamma)=\mathrm{Tr}\rho$, it follows[15] that $\Delta\mapsto a(\Delta)$ is a normalized POV measure on Γ. Also, $\pi[U_g^*\rho U_g]=[\mu_\rho]_g$, and therefore a is a phase space system of imprimitivity. Finally, the condition $\pi(\rho_1)=\pi(\rho_2)$ iff $\rho_1=\rho_2$ implies that a is informationally complete.

The converse follows upon defining $\pi(\rho)$ by

$$[\pi(\rho)](\Delta)=\mathrm{Tr}[a(\Delta)\rho], \tag{A23}$$

while the rest of the theorem is a consequence of Lemma 3.

APPENDIX B

Proof of Theorem 2: Consider first a^Q. Since a^Q and E^Q give the same information, the von Neumann algebras $\mathcal{A}(a^Q)$ and $\mathcal{A}(E^Q)$ generated by them are identical.[6] Furthermore, $\mathcal{A}(E^Q)$ is isometrically isomorphic, in the C^* algebraic sense, to $L^\infty(\mathbb{R}^{3n})$ (i.e., the set of bounded Lebesgue measurable functions under the "essential sup" norm). Then, for $\psi\in\mathcal{H}$ and $f\in C_\infty(\mathbb{R}^{3n})$,

$$(a^Q(f)\psi)(q)=F_f(q)\psi(q), \tag{B1}$$

where,

$$a^Q(f)=\int_{\mathbb{R}^{3n}}f(q)a^Q(dq),$$

and F_f is some measurable function in $L^\infty(\mathbb{R}^{3n})$. We show next that $F_f(q)$ may actually be chosen to be continuous in q. Indeed,

$$\|a^Q(f)\|=\operatorname*{ess.sup}_{q\in\mathbb{R}^{3n}}|F_f(q)|, \tag{B2}$$

and

$$\|a^Q(f)\|\le K\|f\|_{C_\infty}, \tag{B3}$$

for some positive number K, independent of f.[15] [The norm in $C_\infty(\mathbb{R}^{3n})$ is the supremum norm.] Thus,

$$|F_f(q)|\le K\|f\|_{C_\infty}, \tag{B4}$$

for almost all q in $\mathbb{R}^{3n}$. Let $\gamma(q)F_f$ denote the translated function

$$(\gamma(q)F_f)(q')=F_f(q'-q). \tag{B5}$$

Then,

$$\begin{aligned}&\operatorname*{ess.sup}_{q'\in\mathbb{R}^{3n}}|\gamma(q)F_f(q')-F_f(q')| \\ &=\operatorname*{ess.sup}_{q'\in\mathbb{R}^{3n}}|F_f(q'-q)-F_f(q')| \\ &=\operatorname*{ess.sup}_{q'\in\mathbb{R}^{3n}}|F_{q[f]}(q')-F_f(q')|,\end{aligned} \tag{B6}$$

by virtue of (3.2a), where $(q[f])(q')=f(q+q')$. Therefore, by linearity (B6) equals

$$\operatorname{ess.sup}_{q'\in\mathbb{R}^{3n}} |F_{q[f]-f}(q')| \leq K\|q[f]-f\|, \tag{B7}$$

by (B4). Since the function f is uniformly continuous, the right-hand side of (B7) converges to zero as $q\to 0$. Thus the map $q\mapsto \gamma(q)F_f$ is continuous in the norm topology of $L^\infty(\mathbb{R}^{3n})$ and hence[16] F_f may be replaced almost everywhere by a continuous function. We shall henceforth assume that this has been done.

Thus, for fixed $q\in\mathbb{R}^{3n}$, $f\mapsto F_f(q)$ defines a positive linear functional on $C_\infty(\mathbb{R}^{3n})$, so that there exists a probability measure [note that $a^Q(\mathbb{R}^{3n})=I$] ν_q for each q in $\mathbb{R}^{3n}$ such that

$$F_f(q)=\nu_q(f). \tag{B8}$$

Hence, for any Borel set Δ_1 in $\mathbb{R}^{3n}$ Eq. (B1) may be written as

$$(a^Q(\Delta_1)\psi)(q)=\nu_q(\Delta_1)\psi(q), \tag{B9}$$

for almost all q in $\mathbb{R}^{3n}$. The marginality condition (3.3a) is straightforward to verify.

To prove that the set $\{(q,\nu_q)\,|\,q\in\mathbb{R}^{3n}\}$ constitutes a fuzzy configuration space, we only need to verify further that $q\neq q'$ iff $\nu_q\neq\nu_{q'}$. But this is obvious since the ν_q's are all probability measures on the locally compact space $\mathbb{R}^{3n}$ which satisfy $\nu_q(\mathbb{R}^{3n})=1$ for all $q\in\mathbb{R}^{3n}$.

The construction of a fuzzy momentum space, using a^P, is similar. Combining the two we get the desired fuzzy phase space (Γ,ν).

Proof of Theorem 3: We first prove that a^Q and E^Q are informationally equivalent. Indeed, the marginality condition

$$\langle\phi\,|\,a^Q(\Delta_1)\phi\rangle=\int_{\Delta_1}dq\int_{\mathbb{R}^{3n}}\nu_q(dq')\,|\phi(q')|^2$$

for all $\phi\in\mathcal{S}(\mathbb{R}^{3n})$ implies that the operator $a^Q(\Delta_1)$ has the form

$$(a^Q(\Delta_1)\psi)(q)=\nu_q(\Delta_1)\psi(q)$$

for all $\psi\in\mathcal{H}$ and almost all q in $\mathbb{R}^{3n}$. Since a^Q satisfies the imprimitivity relation (3.2a), ν_q satisfies the covariance condition

$$\nu_q(\Delta_1)=\nu_0([\Delta_1]q^{-1}), \tag{B10}$$

for all $q\in\mathbb{R}^{3n}$ and all Borel sets Δ_1 in $\mathbb{R}^{3n}$. Also, since ν_q is a probability measure, $\nu_q=\nu_{q'}$ iff $q=q'$. Thus it follows from Theorem 2 in Ref. 6 that a^Q and E^Q are informationally equivalent.

Next let $\psi_1,\psi_2,\cdots$ be an orthonormal basis of vectors in $\mathcal{H}$ and let

$$P_N=\sum_{k=1}^{N}|\psi_n\rangle\langle\psi_n|,\quad 1\leq N\leq\infty. \tag{B11}$$

Consider for $\phi\in\mathcal{H}$ the nonnegative expression,

$$I_N(\phi)=\int_{\mathbb{R}^{3n}}\langle\exp(ipQ)\phi\,|\,P_NF(0,0)P_N\exp(ipQ)\phi\rangle\,dp. \tag{B12}$$

We observe that according to (2.9)

$$I_\infty(\phi)=\int_{\mathbb{R}^{3n}}\langle\phi\,|\,F(0,p)\phi\rangle\,dp. \tag{B13}$$

Choosing ϕ to be of compact support, and using the unitarity of the Fourier transform in $L^2(\mathbb{R}^{3n})$ we get for $N<\infty$,

$$\begin{aligned} I_N(\phi) &=\sum_{n,k=1}^{N}\langle\psi_m\,|\,F(0,0)\psi_k\rangle\int_{\mathbb{R}^{3n}}dp\int_{\mathbb{R}^{3n}}dx\,\exp(-ipx)\phi(x)\\ &\quad\times\psi_m(x)\int_{\mathbb{R}^{3n}}dy\,\exp(ipy)\,\overline{\psi_k(y)}\phi(y)\\ &=(2\pi)^{3n}\sum_{m,k=1}^{N}\langle\psi_m\,|\,F(0,0)\psi_k\rangle\int_{\mathbb{R}^{3n}}|\phi(x)|^2\,\overline{\psi_k(x)}\\ &\quad\times\psi_m(x)\,dx. \end{aligned} \tag{B14}$$

If $\phi_1,\phi_2,\cdots\in L^2(\mathbb{R}^{3n},dx)$ is a sequence of continuous functions of compact support for which

$$\sum_{l=1}^{\infty}|\phi_l(x)|^2\equiv 1, \tag{B15}$$

then we have by (B14),

$$\sum_{l=1}^{\infty}I_N(\phi_l)=(2\pi)^{3n}\sum_{k=1}^{N}\langle\psi_k\,|\,F(0,0)\psi_k\rangle. \tag{B16}$$

Since $\langle\psi_k\,|\,F(0,0)\psi_k\rangle\geq 0$, the above expression increases monotonically as $N\to\infty$. On the other hand, by (B14) and (B15)

$$\sum_{l=1}^{\infty}I_N(\phi_l)=\int_{\mathbb{R}^{3n}}dp\sum_{l=1}^{\infty}\langle\exp(ipQ)\phi_l\,|\,P_NF(0,0)P_N\times\exp(ipQ)\phi_l\rangle. \tag{B17}$$

We shall prove later that $F(0,0)$ has no continuous spectrum. Consequently, by choosing $\{\psi_n\}$ to consist of eigenvectors of $F(0,0)$ we achieve that

$$P_NF(0,0)P_N\leq F(0,0),\quad N=1,2,3,\cdots. \tag{B18}$$

Thus, by Lebesgue's dominated convergence theorem, in the limit $N\to\infty$ the expression (B17) converges to

$$\begin{aligned} &\int_{\mathbb{R}^{3n}}dp\sum_{l=1}^{\infty}\langle\exp(ipQ)\phi_l\,|\,F(0,0)\exp(ipQ)\phi_l\rangle\\ &=\sum_{l=1}^{\infty}\int_{\mathbb{R}^{3n}}\langle\phi_l\,|\,F(0,p)\phi_l\rangle\,dp\\ &=\sum_{l=1}^{\infty}\int_{\mathbb{R}^{3n}}\nu_0(dx)\,|\phi_l(x)|^2, \end{aligned} \tag{B19}$$

so that the series in (B16) converges in the limit $N\to\infty$,

$$\begin{aligned} \sum_{k=1}^{\infty}\langle\psi_k\,|\,F(0,0)\psi_k\rangle &=(2\pi)^{-3n}\int_{\mathbb{R}^{3n}}\nu_0(dx)\\ &=(2\pi)^{-3n}. \end{aligned} \tag{B20}$$

Since $F(0,0)\geq 0$, the above sum is independent of the chosen orthonormal sequence $\psi_1,\psi_2,\cdots$. Hence we conclude that $F(0,0)$ is of trace class (cf. Ref. 10, Chap. II, Sec. 11), and has a spectral decomposition of the form

$$F(0,0)=\sum_k|\psi_k\rangle\lambda_k\langle\psi_k|,\quad\lambda_k\geq 0, \tag{B21}$$

$$\sum_k\lambda_k=\mathrm{Tr}[F(0,0)]. \tag{B22}$$

Inserting this result in (B14) and letting $N\to\infty$ we obtain for all $\phi\in\mathcal{S}(\mathbb{R}^{3n})$

$$\int_{\mathbb{R}^{3n}} \langle \phi | F(0,p)\phi\rangle dp = (2\pi)^{3n} \int \langle q' | F(0,0) | q'\rangle \times |\phi(q')|^2 dq'. \qquad \text{(B23)}$$

On the other hand, according to (3.3a)

$$\int_{\mathbb{R}^{3n}} \langle \phi | F(0,p)\phi\rangle dp = \int \nu_0(q') |\phi(q')|^2 dq'. \qquad \text{(B24)}$$

It follows that the right-hand sides of (B23) and (B24) must be equal for arbitrary $\phi \in \mathcal{S}(\mathbb{R}^{3n})$ so that ν_0 has density

$$\chi_0(q') = (2\pi)^{3n} \langle q' | F(0,0) | q'\rangle$$

with respect to the Lebesgue measure and, therefore, (3.5) must be true.

This completes the proof of the first half of Theorem 3, with the exception of the claim that $F(0,0)$ has a pure point spectrum, which was used when we argued that (B18) could be satisfied by an appropriate choice of the vectors $\psi_1, \psi_2, \cdots$.

To verify that this claim is true, we note that the positive-definiteness of $F(0,0)$ implies by itself that $F(0,0)$ has a spectral decomposition of the form

$$F(0,0) = \int_0^\infty \lambda\, dE_\lambda. \qquad \text{(B25)}$$

For every integer k, let

$$\alpha_0^{(k)} = 0, \quad \alpha_1^{(k)} = k^{-1}, \quad \cdots, \quad \alpha_{n+1}^{(k)} = \alpha_n^{(k)} + k^{-1}, \qquad \text{(B26)}$$

so that,

$$F_k = \sum_{n=0}^{\infty} \alpha_n^{(k)} (E_{\alpha_{n+1}^{(k)}} - E_{\alpha_n^{(k)}}) \qquad \text{(B27)}$$

is a positive definite operator with a pure point spectrum, and

$$F_k \leq F(0,0), \quad k = 1, 2, \cdots, \qquad \text{(B28)}$$

$$\|F(0,0) - F_k\| \leq k^{-1}. \qquad \text{(B29)}$$

Define now, in analogy with (B12)

$$I_N^{(k)}(\phi) = \int_{\mathbb{R}^{3n}} \langle \exp(ipQ)\phi | P_N^{(k)} F_k P_N^{(k)} \exp(ipQ)\phi\rangle. \qquad \text{(B30)}$$

We note that because of (B28),

$$I_\infty^{(k)}(\phi) \leq I_\infty(\phi) < \infty, \qquad \text{(B31)}$$

so that the entire argument used in deriving (B16) and (B17) can be repeated verbatim to derive analogous relations for $\sum_l I_N^{(k)}(\phi_l)$. In these relations, however, the role of $F(0,0)$ is taken over by F_k. On the other hand, F_k is constructed in such a manner that it does have a pure point spectrum, and therefore, a choice of an orthonormal basis $\psi_1^{(k)}, \psi_2^{(k)}, \cdots$ can be effected so that

$$P_N^{(k)} F_k P_N^{(k)} \leq F_k, \quad N = 1, 2, \cdots \qquad \text{(B32)}$$

would hold true. Hence we conclude that F_k is of trace class. Since, according to (B29), $F(0,0)$ is the uniform limit of $F_1, F_2, \cdots$, it follows that $F(0,0)$ is a compact operator. This establishes the fact that $F(0,0)$ must have a pure point spectrum, and concludes the proof of the first half of Theorem 3.

The second half of this theorem is proven in exactly the same manner by working in the momentum rather than the configuration representation.

*Supported in part by the National Research Council of Canada.

[1]G.S. Agarwal and E. Wolf, Phys. Rev. D 2, 2187 (1970).
[2]M.D. Srinivas and E. Wolf, Phys. Rev. D 11, 1477 (1975).
[3]E.P. Wigner, *Perspectives in Quantum Theory*, edited by W. Yourgrau and A. Vander Merwe (M.I.T. Press, Cambridge, Mass., 1971), pp. 25–36.
[4]E. Prugovečki, J. Math. Phys. 17, 517 (1976).
[5]E. Prugovečki, J. Math. Phys. 17, 1673 (1976).
[6]S.T. Ali and H.D. Doebner, J. Math. Phys. 17, 1105 (1976).
[7]E. Prugovečki, "Information-Theoretical Aspects of Quantum Measurement (University of Toronto preprint).
[8]E.B. Davies and J.T. Lewis, Commun. Math. Phys. 17, 239 (1970).
[9]S.T. Ali and G.G. Emch, J. Math. Phys. 15, 176 (1974).
[10]J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated from the German edition by R.T. Beyer (Princeton U.P., Princeton, N.J., 1967), Chap. III, Sec. 4.
[11]J.R. Klauder, J. Math. Phys. 5, 177 (1964).
[12]R.C. Tolman, *The Principles of Statistical Mechanics* (Oxford U.P., Oxford, 1962), Sec. 20.
[13]L. Nachbin, *Topology and Order* (Van Nostrand, Princeton, N.J., 1965).
[14]S. Sakai, *C*-Algebras and W*-Algebras*, Ergebnisse der Mathematik und ibre Grenzgebiete, Band 60 (Springer, Berlin, 1961).
[15]S.K. Berberian, *Notes on Spectral Theory* (Van Nostrand, Princeton, N.J., 1966).
[16]J. Dieudonné, *Treatise on Analysis* (Academic, New York and London, 1970), Vol. II, p. 281.

On the definition of scattering subspaces in nonrelativistic quantum mechanics

John D. Dollard

Department of Mathematics, University of Texas, Austin, Texas 78712
(Received 22 March 1976)

A physically motivated definition of scattering subspaces is given for problems in nonrelativistic quantum mechanics. The definition is more stringent than earlier similar definitions. It is applicable to potential scattering and to n-body problems.

INTRODUCTION

Recently, some attempts have been made[1,2] to give a physically intuitive characterization of scattering subspaces for quantum-mechanical systems. For a nonrelativistic quantum-mechanical particle, which is described by a normalized element of $\mathcal{L}^2(\mathbf{R}^3)$, the characterizations essentially say this: Let H be the Hamiltonian for the particle and let B_n denote the ball of radius n about the origin in $\mathbf{R}^3$. Let $\psi \in \mathcal{L}^2(\mathbf{R}^3)$. Then ψ is in the positive time scattering subspace if and only if for each $n = 1, 2, 3, \cdots$ we have

$$\lim_{t\to+\infty} \int_{B_n} |[\exp(-iHt)\psi](\mathbf{x})|^2 \, d\mathbf{x} = 0. \tag{1}$$

(We work with units in which $\hbar = 1$.) The negative time scattering subspace is defined similarly, with $+\infty$ being replaced by $-\infty$ above. The interpretation of this definition is simple: If ψ is normalized, then $|[\exp(-iHt)\psi] \times (\mathbf{x})|^2$ is the position probability density at time t for the particle which has state ψ at time $t=0$. Thus Eq. (1) says that the probability for finding the particle in B_n approaches zero as $t \to +\infty$. If this is true for all n, then it is asymptotically correct to think of the particle as being "far" from the origin. This permits one to think of the particle as being asymptotically outside the range of a center of force (presumably localized near the origin) and hence in some sense "free." However, our knowledge of the behavior of free quantum-mechanical particles is quite detailed, and it is legitimate to ask in what sense the behavior described by a ψ in the scattering subspace above approximates the usual behavior of a quantum-mechanical free particle. We now discuss this point. (The discussion will be partly heuristic. Strict mathematical argument will begin in Sec. I.)

One feature of the behavior of a quantum-mechanical free particle is this: Not only does it flee the origin in the sense described by Eq. (1), but one can make statements about the *manner* in which it flees. It would not be correct to say that the particle asymptotically flees with a fixed radial velocity, because of course the particle will have a spectrum of possible velocities. What *is* correct is to say that the integral over all v_r of the probability for asymptotic flight from the origin with radial velocity v_r is unity. Intuitively this is because the component of the wavefunction corresponding to (unchanging) velocity $\mathbf{v}$ will asymptotically represent departure from the origin with fixed radial velocity v_r equal to the magnitude of $\mathbf{v}$, and integrating over all velocities accounts for all possible motions of the particle. Since the probability for v_r to be exactly zero will be zero, we can say that the integral of the probabilities for all velocities v_r *other than zero* must be unity. Thus the free particle will flee the origin with some nonzero velocity, although this velocity may be quite small. We shall refer to this type of flight from the origin as *flight with velocity*. If we now think of an interacting quantum-mechanical particle, then the velocity of such a particle will not be constant, so the above definition of flight with velocity is not applicable, but it is easy to repair this by thinking of the average radial velocity of the particle in a given time interval. If the particle does asymptotically become free, then its velocity should eventually approach a constant value, and (thus) so should its average radial velocity. For each v_r we can compute the probability that the asymptotic average radial velocity will be v_r, and if the integral of all these probabilities (excluding the point $v_r = 0$) is unity, we will describe the behavior of the particle as *asymptotic flight with velocity*. (Remark: exclusion of the point $v_r = 0$ is now significant. If the wavefunction of the particle in question is a bound state, then the asymptotic average radial velocity of the particle will be zero with probability one, representing no flight from the origin at all. Thus it is now a nontrivial requirement to ask that the integral over all probabilities excluding the point $v_r = 0$ should be unity.)

In the definition of scattering subspaces given in Eq. (1), no attempt is made to characterize the manner in which the particle in question leaves the origin. The departure could become more and more leisurely as time goes on, representing a behavior uncharacteristic of asymptotically free particles. In this paper we attempt to sharpen the definition of scattering subspaces by capturing mathematically the notion of asymptotic flight with velocity and defining the scattering subspaces to be the collections of wavefunctions ψ such that $\exp(-iHt)\psi$ exhibits asymptotic flight with velocity. If ψ belongs to a scattering subspace as so defined, then the particle described by $\exp(-iHt)\psi$ can be more confidently described as "asymptotically free" than particles concerning which it is only known that eventually they escape from any bounded set. By comparison with the usual time-dependent nonrelativistic potential scattering theory, we shall see that our subspaces coincide with the standard scattering subspaces in familiar cases.

I. DEFINITION OF THE SCATTERING SUBSPACES

Our entire discussion will be carried out within the framework of nonrelativistic quantum mechanics in

Hilbert space. We shall give definitions and theorems for the case $t \to +\infty$. The corresponding definitions and theorems for the case $t \to -\infty$ can be obtained from our discussion by making obvious replacements, as we occasionally note.

We consider a nonrelativistic particle of mass m with self-adjoint Hamiltonian H acting on the space $\mathcal{L}^2(\mathbf{R}^3)$ of complex-valued functions square-integrable over three-dimensional Euclidean space. The form of H is not important for now. In quantum mechanical discussions it is slightly more natural to deal with momentum than velocity, and we shall construct the scattering subspace by discussing the momentum rather than the velocity of the particle. Naturally the two differ only by the factor m. Now a particle starting from the origin with momentum $\mathbf{k}$ will be on the sphere of radius kt/m after time t. (k denotes the magnitude of $\mathbf{k}$.) We set $k = 1/n$ where n is a positive integer, and for $\psi \in \mathcal{L}^2(\mathbf{R}^3)$ we define $\|\psi\|_{snt}$ (s stands for "scattering") by defining its square,

$$\|\psi\|^2_{snt} = \int_{x \geq t/nm} |[\exp(-iHt)\psi](\mathbf{x})|^2 \, d\mathbf{x}. \tag{2}$$

For a normalized ψ, the number $\|\psi\|^2_{snt}$ is the probability that the particle with wavefunction $\exp(-iHt)\psi$ will be found outside the sphere of radius t/nm at time t, and we can think of this number as measuring the probability that the particle was fleeing the origin with average radial momentum exceeding $1/n$ during the time interval $[0, t]$. (Of course, the particle was not *at* the origin at time $t = 0$, but would be essentially localized in some sphere about the origin. If t is large enough so that t/nm is much greater than the radius of this initial sphere, then our statement about the interpretation of $\|\psi\|^2_{snt}$ is justified.) The number $\|\psi\|^2_{snt}$ may not have a limit as $t \to +\infty$, but it will have an inferior limit, which we denote by $\|\psi\|^2_{sn}$:

$$\|\psi\|^2_{sn} = \varliminf_{t \to +\infty} \|\psi\|^2_{snt}. \tag{3}$$

Following the interpretation above, if ψ is normalized, then $\|\psi\|^2_{sn}$ can be regarded as a lower bound for the probability that the particle with wavefunction $\exp(-iHt)\psi$ was fleeing the origin with average radial momentum exceeding $1/n$ during the infinite time interval $[0, \infty)$. We shall not attempt to decide whether the probabilities for asymptotic flight with various average radial momenta actually become constant for large positive time t, as suggested in the partially heuristic introduction. Instead, we shall build the concept of flight with velocity on the numbers defined in Eq. (2), which give an unambiguous description of the actual motion of the particle and clearly provide the kind of information we are looking for. Observe that

$$\|\psi\|^2_{s1} \leq \|\psi\|^2_{s2} \leq \cdots \leq \|\psi\|^2 \tag{4}$$

so that the sequence $\|\psi\|^2_{sn}$ has a limit as $n \to \infty$. We write

$$\|\psi\|^2_s = \lim_{n \to \infty} \|\psi\|^2_{sn}. \tag{5}$$

If now ψ is a normalized function and if $\|\psi\|^2_s$ equals unity, we will say that the flight from the origin described by $\exp(-iHt)\psi$ is flight with velocity, since $\|\psi\|^2_s$ is a lower bound for the probability that the particle fled the origin with *some* nonzero average velocity. We have avoided inclusion of zero velocity by considering only states whose average radial momentum exceeded $1/n$ for some n.

We now define the positive time scattering subspace $\mathcal{H}^+_{sc}$ as follows:

$$\mathcal{H}^+_{sc} = \{\psi \in L^2(\mathbf{R}^3) \,|\, \|\psi\|^2_s = \|\psi\|^2\}. \tag{6}$$

The companion subspace $\mathcal{H}^-_{sc}$ is defined using the condition $x \geq |t|/nm$ in Eq. (2) and the limit as $t \to -\infty$ in Eq. (3). We now justify the terminology "subspace."

Theorem 1: $\mathcal{H}^+_{sc}$ is a closed subspace of $\mathcal{L}^2(\mathbf{R}^3)$.

Proof: We give the proof for $\mathcal{H}^+_{sc}$. $\mathcal{H}^+_{sc}$ is obviously closed under scalar multiplication. To prove closure under addition, we define for $\psi \in \mathcal{L}^2(\mathbf{R}^3)$ the numbers (b stands, somewhat optimistically, for "bound")

$$\|\psi\|^2_{bnt} = \|\psi\|^2 - \|\psi\|^2_{snt} = \int_{x \leq t/nm} |[\exp(-iHt)\psi](\mathbf{x})|^2 \, d\mathbf{x}. \tag{7}$$

We define

$$\|\psi\|^2_{bn} = \varlimsup_{t \to +\infty} \|\psi\|^2_{bnt} = \|\psi\|^2 - \varliminf_{t \to +\infty} \|\psi\|^2_{snt} = \|\psi\|^2 - \|\psi\|^2_{sn}. \tag{8}$$

We also put

$$\|\psi\|^2_b = \lim_{n \to \infty} \|\psi\|^2_{bn} = \|\psi\|^2 - \|\psi\|^2_s \tag{9}$$

so that

$$\psi \in \mathcal{H}^+_{sc} \iff \|\psi\|^2_b = 0. \tag{10}$$

Now the number $\|\psi\|_{bnt}$, by its definition in terms of the integral in Eq. (7), has some properties of a Hilbert-space norm. In particular it satisfies the parallelogram law. Thus we have

$$\begin{aligned}\|\psi_1 + \psi_2\|^2_{bnt} &\leq \|\psi_1 + \psi_2\|^2_{bnt} + \|\psi_1 - \psi_2\|^2_{bnt} \\ &= 2\|\psi_1\|^2_{bnt} + 2\|\psi_2\|^2_{bnt}.\end{aligned} \tag{11}$$

If ψ_1 and ψ_2 are in $\mathcal{H}^+_{sc}$ and thus satisfy the right-hand condition of the equivalence (10), then by Eq. (11) we have

$$\begin{aligned}\|\psi_1 + \psi_2\|^2_{bn} &= \varlimsup_{t \to +\infty} \|\psi_1 + \psi_2\|^2_{bnt} \\ &\leq 2\varlimsup_{t \to +\infty} \|\psi_1 + \psi_2\|^2_{bnt} + 2\varlimsup_{t \to +\infty} \|\psi_2\|^2_{bnt} \\ &= 2\|\psi_1\|^2_{bn} + 2\|\psi_2\|^2_{bn},\end{aligned} \tag{12}$$

so that

$$\|\psi_1 + \psi_2\|^2_b \leq 2\|\psi_1\|^2_b + 2\|\psi_2\|^2_b = 0, \tag{13}$$

and $\psi_1 + \psi_2 \in \mathcal{H}^+_{sc}$. Thus $\mathcal{H}^+_{sc}$ is closed under addition. To see that $\mathcal{H}^+_{sc}$ is a closed subspace of $\mathcal{L}^2(\mathbf{R}^3)$, let $\psi_l \in \mathcal{H}^+_{sc}$, $l = 1, 2, \cdots$, and let ψ_l converge strongly to ψ. Using some obvious properties of $\|\psi\|_{bnt}$, we have for any l

$$\|\psi\|_{bnt} \leq \|\psi_l\|_{bnt} + \|\psi - \psi_l\|_{bnt} \leq \|\psi_l\|_{bnt} + \|\psi - \psi_l\|. \tag{14}$$

Thus

$$\|\psi\|_b = \lim_{n \to \infty} \varlimsup_{t \to +\infty} \|\psi\|_{bnt} \leq \|\psi - \psi_l\| \quad \text{for any } l, \tag{15}$$

so that the left-hand side of Eq. (15) must be zero, and $\psi \in \mathcal{H}^+_{sc}$. This completes the proof.

The scattering subspaces $\mathcal{H}^{\pm}_{sc}$ are defined in terms of the asymptotic behavior of the function $\exp(-iHt)\psi$. Thus if $\psi\in\mathcal{H}^{\pm}_{sc}$ and u is a real number, $\exp(-iHu)\psi$ should belong to $\mathcal{H}^{\pm}_{sc}$, since $\exp(-iHt)\psi$ and $\exp[-iH(t+u)]\psi$ will have similar asymptotic behavior. Thus $\mathcal{H}^{\pm}_{sc}$ should be invariant under application of $\exp(-iHu)$ for all real u, which is equivalent to saying that the subspace $\mathcal{H}^{\pm}_{sc}$ should reduce the Hamiltonian H.

Theorem 2: The subspaces $\mathcal{H}^{\pm}_{sc}$ reduce the Hamilton H.

Proof for $\mathcal{H}^{+}_{sc}$: Let u be a real number, and suppose $\psi\in\mathcal{H}^{+}_{sc}$. If $u\geq 0$, then

$$\|\exp(-iHu)\psi\|^2_{bnt}=\int_{x\leq t/nm}|[\exp(-iH(t+u))\psi](\mathbf{x})|^2\,d\mathbf{x}$$
$$=\int_{x\leq (r-u)/nm}|[\exp(-iHr)\psi](\mathbf{x})|^2\,d\mathbf{x}, \tag{16}$$

where $r=t+u$. Thus

$$\|\exp(-iHu)\psi\|^2_{bnt}\leq\int_{x\leq r/nm}|[\exp(-iHr)\psi](\mathbf{x})|^2\,d\mathbf{x}$$
$$=\|\psi\|^2_{bnr}=\|\psi\|^2_{bn(t+u)}. \tag{17}$$

Now the superior limit as $t\to+\infty$ of $\|\psi\|^2_{bn(t+u)}$ is identical with that of $\|\psi\|^2_{bnt}$, so

$$\|\exp(-iHu)\psi\|^2_b=\lim_{n\to\infty}\overline{\lim_{t\to+\infty}}\|\exp(-iHu)\psi\|^2_{bnt}$$
$$\leq\lim_{n\to\infty}\overline{\lim_{t\to+\infty}}\|\psi\|^2_{bn(t+u)}=\|\psi\|^2_b=0, \tag{18}$$

so that $\exp(-iHu)\psi\in\mathcal{H}^{+}_{sc}$. If $u<0$ then (16) still holds but $r-u>r$ so that (17) is false in general. However, replacing n by $2n$ in (16), as soon as r is greater than $|u|$ we have

$$\frac{r-u}{(2n)m}\leq\frac{r}{nm} \tag{19}$$

and hence

$$\|\exp(-iHu)\psi\|^2_{b(2n)t}\leq\int_{x\leq r/nm}|[\exp(-iHr)\psi](\mathbf{x})|^2\,d\mathbf{x}$$
$$=\|\psi\|^2_{bnr}. \tag{20}$$

Arguing as above, we have that

$$\lim_{n\to\infty}\overline{\lim_{t\to+\infty}}\|\exp(-iHu)\psi\|^2_{b(2n)t}=0 \tag{21}$$

and since the limit as $n\to\infty$ of $\overline{\lim}_{t\to+\infty}\|\exp(-iHu)\psi\|^2_{bnt}$ is known to exist, this limit must also be zero, so again $\exp(-iHu)\psi\in\mathcal{H}^{+}_{sc}$, completing the proof.

Since the subspaces $\mathcal{H}^{\pm}_{sc}$ reduce H, it is natural to consider the properties of the spectrum of H in $\mathcal{H}^{\pm}_{sc}$. Because of their interpretation, the subspaces $\mathcal{H}^{\pm}_{sc}$ should certainly be orthogonal to the subspace $\mathcal{H}_{pp}$ containing the "pure point" spectrum of H, i.e., the subspace spanned by the bound states (eigenfunctions) of H. This is in fact the case, and it follows that the parts of H in $\mathcal{H}^{\pm}_{sc}$ have continuous spectrum.

Theorem 3: The subspaces $\mathcal{H}^{\pm}_{sc}$ are orthogonal to the subspace $\mathcal{H}_{pp}$ spanned by the bound states of H.

Proof for $\mathcal{H}^{+}_{sc}$: We will show that if $\psi\in\mathcal{H}^{+}_{sc}$ then ψ satisfies the condition expressed by Eq. (1). The conclusion of the theorem then follows from a theorem of Wilcox (Ref. 1, Theorem 2.1). Let $\psi\in\mathcal{H}^{+}_{sc}$ and let l and n be positive integers. As soon as t is large enough so that $l\leq t/mn$, we have

$$\int_{x\leq l}|[\exp(-iHt)\psi](\mathbf{x})|^2\,d\mathbf{x}\leq\int_{x\leq t/mn}|[\exp(-iHt)\psi](\mathbf{x})|^2\,d\mathbf{x}$$
$$=\|\psi\|^2_{bnt}. \tag{22}$$

Thus

$$\overline{\lim_{t\to+\infty}}\int_{x\leq l}|[\exp(-iHt)\psi](\mathbf{x})|^2\,d\mathbf{x}\leq\overline{\lim_{t\to+\infty}}\|\psi\|^2_{bnt}=\|\psi\|^2_{bn} \tag{23}$$

and since this holds for all $n=1,2,3,\cdots$ we have

$$\overline{\lim_{t\to+\infty}}\int_{x\leq l}|[\exp(-iHt)\psi](\mathbf{x})|^2\,d\mathbf{x}\leq\lim_{n\to\infty}\|\psi\|^2_{bn}=\|\psi\|^2_b=0. \tag{24}$$

Thus ψ satisfies the required condition, and we are done.

II. CONNECTION WITH USUAL SCATTERING THEORY

To establish the connection of the subspaces $\mathcal{H}^{\pm}_{sc}$ with the usual formulation of scattering theory, we first take H to be the free Hamiltonian $H_0=-\Delta/2m$ where Δ is the (natural self-adjoint extension of the) Laplacian operator. We will verify some results that are easy to anticipate. As in Ref. 3 we write

$$\exp(-iH_0t)=C_tQ_t, \tag{25}$$

where

$$(Q_t\varphi)(\mathbf{x})=\exp(imx^2/2t)\varphi(\mathbf{x}) \tag{26}$$

and, letting $\tilde{\varphi}$ denote the Fourier transform of φ, C_t is defined by

$$(C_t\varphi)(\mathbf{x})=(m/it)^{3/2}\exp(imx^2/2t)\tilde{\varphi}(m\mathbf{x}/t). \tag{27}$$

Then for $\psi\in\mathcal{L}^2(\mathbf{R}^3)$ we have

$$\|\psi\|^2_{snt}=\int_{x\geq t/nm}|[\exp(-iH_0t)\psi](\mathbf{x})|^2\,d\mathbf{x}$$
$$=\int_{x\geq t/nm}\left|\frac{m}{t}\right|^3\left|(\widetilde{Q_t\psi})\left(\frac{m\mathbf{x}}{t}\right)\right|^2\,d\mathbf{x}. \tag{28}$$

Letting $\mathbf{k}=m\mathbf{x}/t$, we have

$$\|\psi\|^2_{snt}=\int_{k\geq 1/n}|(\widetilde{Q_t\psi})(\mathbf{k})|^2\,d\mathbf{k}. \tag{29}$$

Now Q_t clearly converges strongly to 1 as $t\to+\infty$, and as a result

$$\|\psi\|^2_{sn}=\lim_{t\to+\infty}\|\psi\|^2_{snt}=\int_{k\geq 1/n}|\tilde{\psi}(\mathbf{k})|^2\,d\mathbf{k}. \tag{30}$$

Since $|\tilde{\psi}(\mathbf{k})|^2$ is the momentum probability density for the particle, $\|\psi\|^2_{sn}$ is just the probability that the particle has magnitude of the momentum exceeding $1/n$, in exact correspondence with our earlier interpretation. (Note that if a particle has momentum $\mathbf{k}$, then asymptotically its radial momentum is the magnitude k, so $\|\psi\|^2_{sn}$ is the probability for asymptotic radial momentum exceeding $1/n$.) We also have for any $\psi\in\mathcal{L}^2(\mathbf{R}^3)$,

$$\lim_{n\to\infty}\|\psi\|^2_{sn}=\int_{\mathbf{R}^3}|\tilde{\psi}(\mathbf{k})|^2\,d\mathbf{k}=\|\psi\|^2, \tag{31}$$

so that $\psi\in\mathcal{H}^{+}_{sc}$. The corresponding argument also applies to $\mathcal{H}^{-}_{sc}$, so that

$$\mathcal{H}^{\pm}_{sc}=\mathcal{L}^2(\mathbf{R}^3), \tag{32}$$

as expected.

Next let us consider a theory with a Hamiltonian H such that the Møller wave matrices $W_\pm$ exist, where $W_\pm$ is defined by

$$W_\pm = s\text{-}\lim_{t\to\pm\infty} \exp(iHt)\exp(-iH_0t). \tag{33}$$

We denote by $R^\pm$ the range of $W_\pm$. If $\psi \in R^+$, then as $t \to +\infty$ the wavefunction $\exp(-iHt)\psi$ converges strongly to $\exp(-iH_0t)W_+\psi$ (see Ref. 3), so that

$$\begin{aligned}\|\psi\|^2_{sn} &= \overline{\lim_{t\to+\infty}}\|\psi\|^2_{snt}\\ &= \overline{\lim_{t\to+\infty}}\int_{x\geq t/nm}|[\exp(-iHt)\psi](\mathbf{x})|^2\,d\mathbf{x}\\ &= \lim_{t\to+\infty}\int_{x\geq t/nm}|[\exp(-iH_0t)W_+\psi](\mathbf{x})|^2\,d\mathbf{x}\\ &= \int_{k\geq 1/n}|(\widetilde{W_+\psi})(\mathbf{k})|^2\,d\mathbf{k}.\end{aligned} \tag{34}$$

Thus $\|\psi\|^2_{sn}$ is just the probability that the asymptotic freely propagating state $\exp(-iH_0t)W_+\psi$ will have magnitude of momentum exceeding $1/n$. This is just what might have been expected intuitively. Clearly we also have

$$\|\psi\|^2_s = \lim_{n\to\infty}\|\psi\|^2_{sn} = \|W_+\psi\|^2 = \|\psi\|^2, \tag{35}$$

so that $\psi \in \mathcal{H}^+_{sc}$. Similar remarks apply to functions $\psi \in R^-$, so that we can conclude

$$R^\pm \subseteq \mathcal{H}^\pm_{sc}. \tag{36}$$

In particular, if the theory is asymptotically complete in the strong sense, so that

$$R^+ = R^- = \mathcal{H}^\perp_{pp}, \tag{37}$$

where $\perp$ denotes orthogonal complement, then because $\mathcal{H}^\pm_{sc}$ is orthogonal to $\mathcal{H}_{pp}$, we have

$$\mathcal{H}^\pm_{sc} = R^\pm = \mathcal{H}^\perp_{pp}. \tag{38}$$

For conditions under which the Møller wave matrices exist or under which the theory is complete in the strong sense, see Refs. 4 or 5.

It can be shown that in the case of a Coulomb potential,

$$V_c(\mathbf{x}) = e_1e_2/x, \tag{39}$$

Eq. (38) still holds. In this case, the subspaces $\mathcal{H}^\pm_{sc}$ are defined exactly as before. The definition of the subspaces $R^\pm$ requires some modification, because the usual Møller wave matrices of Eq. (33) do not exist.[4] However, if the Møller wave matrices are defined as in Ref. 4 making use of a "distorted" free Hamiltonian, then it is found that Eq. (37) holds and that for $\psi \in R^\pm$ the asymptotic behavior of $\exp(-iHt)\psi$ is enough like that of a free particle so that the above argument leading to Eq. (38) can be pushed through without substantial modification. Similar remarks apply to the case of a Coulomb-like potential (short-range plus Coulomb) when the theory is known to be asymptotically complete. (See Ref. 5.) Thus we have found a definition of scattering subspaces which is physically intuitive and which reproduces the usual scattering subspaces both for short-range and Coulomb potentials. Even in theories in which the existence of the Møller wave matrices is in question, one can be confident that the elements of these subspaces represent particles which flee the origin in somewhat the same manner as free particles.

As a closing remark, we note that the above definition of scattering subspaces can be transferred to the setting of n-body scattering processes. If there are static potentials at the origin, so that center-of-mass momentum is not conserved, one merely replaces the variable $\mathbf{x}$ in Eq. (2) by the center-of-mass coordinate for the n particles, and replaces R^3 by R^{3n}. The number $\|\psi\|^2_{snt}$ then gives the probability that during the time interval $[0, t]$, the center of mass of the n-body system was fleeing the region of the static potentials with an average radial momentum exceeding $1/n$. The number $\|\psi\|^2_s$ should equal unity (for a normalized ψ) only if at least some of the component particles of the system asymptotically flee the static potentials with velocity, in which case it makes sense to say that scattering is being described. If there are no static potentials present then one must factor out the center-of-mass coordinate in the well-known way, with a corresponding modification of Eq. (2). In any case, one then has a description of the n-body scattering subspace in which it is not necessary to mention explicitly the various channels of the system.

[1]Calvin H. Wilcox, J. Func. Anal. **12**, 257 (1973).
[2]W.O. Amrein and V. Georgescu, Helv. Phys. Acta **46**, 635 (1973).
[3]J. Dollard, Commun. Math. Phys. **12**, 193 (1969).
[4]J. Dollard, Rocky Mt. J. Math. **1**, 5 (1971).
[5]B. Simon, Thesis, Princeton University (1970).

A new family of solutions of the Einstein field equations

Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 22 March 1976)

An ordinary differential equation is presented, from solutions of which may be constructed solutions of the Einstein field equations. The study of these solutions may shed light upon the still obscure systematics of the Tomimatsu–Sato spinning mass fields.

In this author's formulation[1] of the stationary axially symmetric gravitational field problem, solutions of Einstein's vacuum field equations are constructed from complex solutions of the single nonlinear field equation

$$(\mathrm{Re}\,\mathcal{E})\nabla^2\mathcal{E} = \nabla\mathcal{E}\cdot\nabla\mathcal{E}. \tag{1}$$

Until now it appears to have escaped notice that this equation has solutions of the form

$$\mathcal{E} = r^k Y_k(\cos\theta), \tag{2}$$

where r, θ, ϕ are spherical polar coordinates, and the function $Y_k(y)$ satisfies the *ordinary* differential equation

$$(\mathrm{Re}\,Y_k)\left(k(k+1)Y_k + \frac{d}{dy}(1-y^2)\frac{d}{dy}\,Y_k\right) = k^2Y_k^2 + (1-y^2)\left(\frac{d}{dy}\,Y_k\right)^2, \tag{3}$$

which has not yet been solved in complete generality.

Employing the well-known symmetry group[2] of Eq. (1), one can construct from solutions of the form (2) solutions of a more general nature. Furthermore, since these spacetimes possess not one but two commuting Killing vector fields, the method of Geroch[3] may be employed to generate additional vacuum spacetimes. Thus, the solution of the ordinary differential equation (3) provides many vacuum spacetimes.

Thus far we have enjoyed only limited success in solving Eq. (3). It should be observed that it is unnecessary to consider separately solutions with $k<0$, for if $\mathcal{E}$ is a solution of Eq. (1), then so is $\mathcal{E}^{-1}$. When $k=0$ the real and imaginary parts of the complex potential $\mathcal{E}$ are functionally related. Therefore, the complete solution for $k=0$ is implicit in the work of Papapetrou.[4] When $k=1$ there is a trivial subcase in which $\mathrm{Im}\,Y_1 = \mathrm{const}$. The corresponding spacetimes are not stationary but *static*, and were anticipated in the work of Lewis.[5] To construct new *stationary* solutions one must turn to other values of k, or if $k=1$, then demand that $\mathrm{Im}\,Y_1 \neq \mathrm{const}$.

Efforts to find the general solution of Eq. (3) may be spurred on by the observation that a series of specific solutions can be constructed without inordinate difficulty. These particular solutions may be expressed in the form

$$Y_k(y) = \frac{1}{p-iq}\,\frac{N_k(y)}{D_k(y)}, \quad k\geq 1, \tag{4}$$

where $p^2+q^2=1$, $D_k = N_{k-1}^*$, and the N_k are the following polynomials:

$$N_0 = 1,$$
$$N_1 = pv + (iq)w,$$
$$N_2 = p^2v^3 + 3p(iq)v^2w + 3p(iq)vw^2 + (iq)^2w^3,$$
$$N_3 = p^3v^6 + 6p^2(iq)v^5w + 15p^2(iq)v^4w^2 + 10[p^2(iq)+p(iq)^2]v^3w^3 + 15p(iq)^2v^2w^4 + 6p(iq)^2vw^5 + (iq)^3w^6,$$
$$N_4 = p^4v^{10} + 10p^3(iq)v^9w + 45p^3(iq)v^8w^2 + 10[7p^3(iq) + 5p^2(iq)^2]v^7w^3 + 35[p^3(iq)+5p^2(iq)^2]v^6w^4 + 252p^2(iq)^2v^5w^5 + 35[5p^2(iq)^2+p(iq)^3]v^4w^6 + 10[5p^2(iq)^2+7p(iq)^3]v^3w^7 + 45p(iq)^3v^2w^8 + 10p(iq)^3vw^9 + (iq)^4w^{10},$$
$$N_5 = p^5v^{15} + 15p^4(iq)v^{14}w + 105p^4(iq)v^{13}w^2 + 35[8p^4(iq)+5p^3(iq)^2]v^{12}w^3 + 105[3p^4(iq) + 10p^3(iq)^2]v^{11}w^4 + 21[6p^4(iq)+137p^3(iq)^2]v^{10}w^5 + 35[129p^3(iq)^2+14p^2(iq)^3]v^9w^6 + 45[94p^3(iq)^2 + 49p^2(iq)^3]v^8w^7 + 45[49p^3(iq)^2+94p^2(iq)^3]v^7w^8 + 35[14p^3(iq)^2+129p^2(iq)^3]v^6w^9 + 21[137p^2(iq)^3 + 6p(iq)^4]v^5w^{10} + 105[10p^2(iq)^3+3p(iq)^4]v^4w^{11} + 35[5p^2(iq)^3+8p(iq)^4]v^3w^{12} + 105p(iq)^4v^2w^{13} + 15p(iq)^4vw^{14} + (iq)^5w^{15},$$

where $v=(1+y)/2$ and $w=(1-y)/2$.

The spacetime corresponding to Y_2 is a good approximation[6] to the $\delta=2$ Tomimatsu–Sato (T–S) field near its poles. The spacetimes corresponding to Y_1–Y_4 also result from contractions[7] performed upon the Kerr and T–S solutions. Presumably the spacetime corresponding to Y_5 could be obtained by a contraction performed on a yet-to-be-constructed T–S field. Note, however, that it would be easier to construct Y_k for much higher values of k than it would be to construct the $\delta=5$ T–S solution. One may even hope to infer from the early members of the series the form of N_k for an arbitrary value of k. It is obvious, for example, that the coefficients in N_k are closely associated with the binomial expansion of $(v+w)^K$, where $K=k(k+1)/2$.

CONCLUSIONS

It would be very encouraging if one were to discover how to solve the ordinary differential equation (3) in complete generality. There is some evidence[6] that the particular series of solutions presented in this paper may be relevant to the study of the singularity structure of the T–S spinning mass fields. Furthermore, the relative simplicity of our series of solutions of Eq. (1) suggests that one may sooner comprehend the systematics of this series than the systematics of the T—S series of solutions, to which our series is so closely related. The eventual discovery of a more direct method of generating the T—S solutions may open the door to the discovery of yet other types of spinning mass solutions, just as the development[1] of a direct method of generating the Kerr solution led to the discovery[8] of the Tomimatsu—Sato series of solutions.

ACKNOWLEDGMENT

The author would like to acknowledge several useful discussions with John E. Economou.

[1]F. J. Ernst, Phys. Rev. **167**, 1175 (1968); J. Math. Phys. **15**, 1409 (1974).
[2]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[3]R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).
[4]A. Papapetrou, Ann. Physik **12**, 309 (1953).
[5]T. Lewis, Proc. R. Soc. A **136**, 176 (1932).
[6]F. J. Ernst, J. Math. Phys. **17**, 1091 (1976); J. E. Economou, J. Math. Phys. **17**, 1095 (1976).
[7]The concept of contraction has been illustrated in detail by J. Plebański, Ann. Phys. (N. Y.) **90**, 196 (1975).
[8]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972); Progr. Theor. Phys. 50, 95 (1973).

Percolation theory on directed graphs

D. K. Arrowsmith and J. W. Essam

Department of Mathematics, Westfield College, University of London, England
(Received 16 March 1976)

The pair connectivity P_{uv} of a directed graph G between vertices u and v is the probability that there is a path from u to v when each edge and vertex has a given probability of being deleted, deletions being made independently. We consider the coefficient $\vec{d}$ in the expansion $P_{uv}(G) = \sum_{A' \subseteq A} \vec{d}_{uv}(G') \prod_{a \in A'} p_a \prod_{w \in V'} p_w$, where A and V are respectively the arc and vertex sets of G, and $p_a(p_w)$ is the probability that the arc a (vertex w) is not deleted. G' is the arc set A' together with its set of incident vertices V'. It is shown that $\vec{d}_{uv}(G')$ is nonzero if and only if G' is coverable by some set of (directed) paths from u to v and has no circuit. When these conditions are satisfied, $\vec{d}_{uv}(G') = (-1)^{t_{uv}+1}$ where the number of independent paths from u to v is t_{uv}. Moreover, t_{uv} is shown to have the value of $\nu(G)+1$, $\nu(G)$ being the cyclomatic number of the graph G.

1. INTRODUCTION

The pair connectedness for undirected graphs has been considered previously.[1] The present work generalizes results to the directed case and the relation between the two cases is discussed. In particular, it is shown that the coefficient $d_{uv}(G)$ for an undirected graph G is given by

$$d_{uv}(G) = \sum_{H \in \mathcal{D}(G)} \vec{d}_{uv}(H),$$

where $\mathcal{D}(G)$ is the set of directed graphs obtained by directing G in all possible ways. The coefficients d_{uv} and $\vec{d}_{uv}$ will be known as undirected and directed d-weights respectively and the suffices representing the root points will not always be made explicit.

The pair connectedness determines other properties of this graph. For example the expected number of vertices which may be reached from u is

$$S_u(G) = \sum_{v \in V \setminus u} P_{uv},$$

where V is the vertex set of G.

Both P_{uv} and S_u are of interest in the theory of ferromagnets[2] where V is the set of spins located at lattice sites and the edges are the possible interactions. P_{uv} is related to the spin–spin correlation function and S_u to the susceptibility.

Another application is to blocking probabilities in telephone networks[3,4] where the vertices are the switches and the edges are the lines connecting them. A deleted edge corresponds to a busy line, a deleted vertex corresponds to a blocked switch, and the complement of P_{uv} is the probability of a call from u to v being blocked. The nature of the switches in a multistage network means that the graph must be directed, although alternative networks which may be undirected have been recently considered. The assumption of independent deletions which is basic to the percolation model is not valid in a telephone network since whole paths become busy simultaneously. However, the theory serves as a first approximation which is good for low traffic densities.

2. PRELIMINARY DEFINITIONS AND FORMULAS

Consider a graph G which has a vertex set V and an edge set E in the undirected case or arc set A in the directed case. Let p_α denote the probability that a particular element α of the graph is not deleted and deletions are assumed to be made independently. Let $\mathcal{S}_{uv}$ be the set of all self-avoiding paths from u to v. A path must follow the arrows for a directed graph. The probability that at least one path from u to v remains in the partially deleted graph is given by inclusion and exclusion.

$$P_{uv}(G) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{\substack{S \subseteq \mathcal{S}_{uv} \\ |S|=r}} pr\{S\}, \tag{2.1}$$

where n is the number of self-avoiding paths between u and v. If $g(S)$ is the subgraph obtained by taking the union of all paths in S we may rearrange the sum in the form

$$P_{uv}(G) = \sum_{E' \subseteq E} d_{uv}(G') \prod_{e \in E'} p_e \prod_{w \in V'} p_w \tag{2.2}$$

when G is undirected, or as in the abstract when G is directed, (i.e., d is replaced by $\vec{d}$ and the edge set E' is replaced by the arc set A'). G' is the graph with edge set E' and vertex set V' which consists of the subset of V incident with E'. In both the directed and undirected cases

$$d_{uv}(G') = \sum_{\substack{S \subseteq \mathcal{S}_{uv} \\ g(S)=G'}} (-1)^{|S|+1}. \tag{2.3}$$

If $g(S) = G'$, S is said to cover G'. Clearly $d(G')$ and $\vec{d}(G')$ are zero when G' is not coverable by paths (e.g., if G is disconnected).

It has been shown that[5] the d-weights satisfy a deletion-contraction rule

$$d(G) = d(G^\gamma) - d(G^\delta) \tag{2.4}$$

which enables them to be calculated recursively. The graphs G^γ and G^δ are obtained from G by contracting and deleting any edge of the graph. In order to establish our main result we shall require a similar result for directed d-weights. The rule is most easily established from a different formula for $\vec{d}(G)$. Thus with $p_e = p_w = 1$

in (2.2)

$$\vec{\gamma}_{uv}(G)=\sum_{A'\subseteq A}\vec{d}_{uv}(G'), \tag{2.5}$$

where $\vec{\gamma}(G)$ is one if there is a path from u to v on G, and zero otherwise. Since the set of all subsets of the arc set A form a lattice, (2.5) may be inverted[6] to yield

$$\vec{d}_{uv}(G)=\sum_{A'\subseteq A}(-1)^{|A\setminus A'|}\vec{\gamma}_{uv}(G'). \tag{2.6}$$

Now consider a particular arc $a\in A$ and divide the sum (2.6) according as $a\in A'$ or not,

$$\vec{d}_{uv}(G)=\sum_{\substack{A'\subseteq A\\ a\in A'}}+\sum_{\substack{A'\subseteq A\\ a\notin A'}}. \tag{2.7}$$

The subsets in the first sum are in one-to-one correspondence with those of A^{γ}. If (condition A) a either leads out of a source (vertex with in-degree zero) or into a sink (vertex with out-degree zero), then $\vec{\gamma}_{uv}(A')$ is the same for both G and G^{γ}. Thus in this case

$$\sum_{\substack{A'\subseteq A\\ a\in A'}}(-1)^{|A\setminus A'|}\vec{\gamma}_{uv}(G')=\vec{d}_{uv}(G^{\gamma}). \tag{2.8}$$

Notice that if condition A is not satisfied, then the contraction of a may introduce paths in G^{γ} with no corresponding one in G. This is not the case for undirected graphs and (2.4) applies to any edge.

Similarly the subsets A' in the second sum are in one-to-one correspondence with those of A^{δ}, and $\vec{\gamma}(G')$ is always unchanged by deletion since $a\notin A'$. Thus

$$\sum_{\substack{A'\subseteq A\\ a\notin A'}}(-1)^{|A\setminus A'|}\gamma_{uv}(G')=\vec{d}_{uv}(G^{\delta}) \tag{2.9}$$

and finally subject to condition A

$$\vec{d}_{uv}(G)=\vec{d}_{uv}(G^{\gamma})-\vec{d}_{uv}(G^{\delta}). \tag{2.10}$$

The d weight of a rooted graph is a topological invariant since it is unchanged by the insertion of vertices of degree two.[1] Another topological invariant $k(G)$, this time for an unrooted graph, may be obtained by considering the expected number of independent circuits (closed paths), thus

$$\langle c(G)\rangle=\sum_{A'\subseteq A}\vec{k}(G')\prod_{a\in A'}p_a\prod_{w\in V'}p_w \tag{2.11}$$

with a similar formula for the undirected case.[7] For the undirected problem

$$c(G)=\nu(G),$$

the cyclomatic number[3] which is defined by $\nu(G)=|E|-|V|+n(G)$, where n is the number of components in G. For the directed case this is only true for strongly connected graphs.[8] Setting $p_a=p_w=1$ in (2.11),

$$c(G)=\sum_{A'\subseteq A}\vec{k}(G') \tag{2.13}$$

and on inversion

$$\vec{k}(G)=\sum_{A'\subseteq A}(-1)^{|A\setminus A'|}c(G'). \tag{2.14}$$

The properties of the "k-weights" may be deduced from those of the d-weights by the device of adding an extra edge (arc) connecting the root points to form a *derived* graph. Thus let G be a graph with root points u and v and let $\bar{G}$ be the derived graph. In the directed case the extra arc a is from v to u. Now

$$\begin{aligned}\vec{k}(\bar{G})&=\sum_{A'\subseteq\bar{A}}(-1)^{|\bar{A}\setminus A'|}c(G')\\&=\sum_{\substack{A'\subseteq\bar{A}\\ a\in A'}}(-1)^{|\bar{A}\setminus A'|}c(G')+\sum_{\substack{A'\subseteq\bar{A}\\ a\notin A'}}(-1)^{|\bar{A}\setminus A'|}c(G')\\&=\sum_{A'\subseteq A}(-1)^{|A\setminus A'|}(c(\bar{G}')-c(G'))\\&=\sum_{A'\subseteq A}(-1)^{|A\setminus A'|}\vec{\gamma}(G'),\end{aligned}$$

where $\vec{\gamma}(G')$ is one if there is a path in G' from u to v, since then the addition of a increases the number of independent circuits by one, but zero otherwise. Comparing this with (2.6), we obtain the desired relation

$$\vec{k}(\bar{G})=\vec{d}(G), \tag{2.16}$$

which also clearly holds for the undirected weights.[1] The undirected k weight is related[1] to the topological invariant β defined by Crapo[9] by

$$k(G)=(-1)^{c(G)+1}\beta(G). \tag{2.17}$$

3. STATEMENTS AND PROOFS OF RESULTS

The d-weights of directed and undirected graphs are related as follows.

Theorem 1: The undirected d-weight of a graph G_{uv} is equal to the sum of the directed d-weights of G_{uv} over all possible orientations.

It remains to classify the properties of directed d-weights which are given in the following results. We omit the trivial case of a noncoverable graph where the directed d-weight is always zero.

Theorem 2: The directed d-weight of a coverable directed graph G is ± 1 or 0.

Theorem 3: The directed d-weight of a coverable directed graph G is zero if and only if G has a circuit.

A collection $\mathcal{C}=\{\pi_i\mid i=1,\ldots,k\}$ of (directed) paths in a coverable directed graph G_{uv} is said to be *independent* if the matrix $M=[m_{ij}]$ has maximal row rank where $\pi_i=\sum_{j=1}^{n}m_{ij}a_j$, $\pi_i\in\mathcal{C}$, and a_j, $j=1,\ldots,n$ is the collection of arcs in G_{uv}. A collection $\mathcal{C}$ is said to be *maximally independent* if every path not in $\mathcal{C}$ is dependent on $\mathcal{C}$. Obviously the number of paths in such a class is an invariant of G_{uv}.

Theorem 4: If G is a coverable graph with no circuit, then

$$\vec{d}_{uv}(G)=(-1)^{t_{uv}+1}, \tag{3.1}$$

where t_{uv} is the maximal number of independent paths from u to v.

Remark: The term coverable always refers to coverings of G by paths from u to v, the assumed root vertices.

The following lemma enables us to compute simply the directed d-weight.

Lemma 5: For a coverable graph G, $t_{uv}(G)=|E|-|V|+2$.

We note that the combination of these theorems together with the undirected form of (2.16) and (2.17) yields a theorem of Greene,[10]

$\beta(G) = |\mathcal{A}_{uv}(G)|$.

Here $\mathcal{A}_{uv}(G)$ is the set of all directed graphs which may be obtained by directing G in such a way that all circuits contain the arc a, $\partial a = (u - v)$ (arbitrarily chosen), and the rooted graph $G\backslash a$ is coverable.

Proof of Theorem 1: Let $\mathcal{D}(G)$ be the set of orientations of the undirected graph G with roots u and v. We note that the result follows easily when G is parallel or disconnected.

Therefore we consider a graph which is not of these types. We can choose an edge e_0 with boundary vertices u and w such that $w \neq v$. Let $\mathcal{D}^+(G)$ be the set of orientations of G such that u is a source. It follows that

$$\sum_{H\in\mathcal{D}(G)} \vec{d}_{uv}(H) = \sum_{H\in\mathcal{D}^+(G)} \vec{d}_{uv}(H). \tag{3.2}$$

The result (2.10) can be applied to the edge e_0 of every subgraph $H \in \mathcal{D}^+(G)$ to obtain

$$\vec{d}_{uv}(H) = \vec{d}_{uv}(H^\gamma) - \vec{d}_{uv}(H^\delta), \tag{3.3}$$

where $H^\delta \in \mathcal{D}^+(G^\delta)$ and either $H^\gamma \in \mathcal{D}^+(G^\gamma)$ or $\vec{d}_{uv}(H^\gamma) = 0$. Also the set of graphs obtained by deletion and contraction of e_0 from all elements of $\mathcal{D}^+(G)$ includes each element of $\mathcal{D}^+(G^\gamma)$ and $\mathcal{D}^+(G^\delta)$ exactly once.

Therefore,

$$\sum_{H\in\mathcal{D}^+(G)} \vec{d}_{uv}(H) = \sum_{H\in\mathcal{D}^+(G)} \vec{d}_{uv}(H^\gamma) - \sum_{H\in\mathcal{D}^+(G)} \vec{d}_{uv}(H^\delta)$$

$$= \sum_{H^\gamma\in\mathcal{D}^+(G^\gamma)} \vec{d}_{uv}(H^\gamma) - \sum_{H^\delta\in\mathcal{D}^+(G^\delta)} \vec{d}_{uv}(H^\delta). \tag{3.4}$$

We now observe that if the theorem were true for G^γ and G^δ, then by (3.2) and (3.4) it would also be true for G. Thus we proceed inductively using the deletion-contraction rule until we reach graphs which are either parallel or disconnected. This must occur at some stage and we have noted the result to be true for such graphs.

Proof of Theorem 2: We take the root u to be a source of the directed graph G, otherwise $\vec{d}_{uv}(G) = 0$. We assume G is not a parallel graph as it is easy to compute in this case that $\vec{d}_{uv}(G) = \pm 1$. We apply the deletion-contraction rule to an arc a attached to u with another vertex w, $(\neq v)$. Then

$\vec{d}_{uv}(G) = \vec{d}_{uv}(G^\gamma) - \vec{d}_{uv}(G^\delta)$.

The following cases have to be considered at the vertex w:

(I) $G\backslash\{a\}$ has in-degree zero at w; G^δ not coverable $\Rightarrow \vec{d}_{uv}(G^\delta) = 0$, hence $\vec{d}_{uv}(G) = \vec{d}_{uv}(G^\gamma)$.

(II) $G\backslash\{a\}$ has both in- and out-degree nonzero at w and G^δ is coverable; G^γ not coverable $\Rightarrow \vec{d}_{uv}(G^\gamma) = 0$, hence $\vec{d}_{uv}(G) = -\vec{d}_{uv}(G^\delta)$.

(III) As in II, except that G^δ is not coverable, $\vec{d}_{uv}(G^\delta) = 0$.

This classification gives us an algorithm for obtaining from the nonparallel graph G either $\vec{d}_{uv}(G) = 0$ [(III)] or a graph G' [$= G^\gamma$ (I) or G^δ (II)], such that $|\vec{d}_{uv}(G')| = |\vec{d}_{uv}(G)|$.

In cases (I) and (II) the graph G' may be a parallel graph in which case $|\vec{d}_{uv}(G)| = |\vec{d}_{uv}(G')| = +1$, or we can apply the above argument for G to G'.

If the algorithm does not reduce to case (III), then after a finite number of reductions the contraction–deletion rule can no longer be applied. This will occur if the last contraction–deletion gives rise to parallel graph. Therefore, $|\vec{d}_{uv}(G)| = 0$ or 1.

Proof of Theorem 3: ($\Leftarrow$) By the hypotheses u is a source and G is not parallel. Applying the contraction–deletion rule to an arc a incident out of u we have the cases (I), (II), and (III) as in Theorem 2.

If we have cases (I) or (II) rather than (III), which gives $\vec{d}_{uv}(G) = 0$ as required, then G' ($= G^\gamma$ or G^δ) is a directed graph with a circuit and $|\vec{d}_{uv}(G)| = |\vec{d}_{uv}(G')|$. Now G' satisfies the conditions on G given above and so the contraction-deletion rule can be applied again. However, the process will terminate after a finite number of steps. This follows because by the rules for (I) and (II) an edge adjacent to, and oriented towards, a vertex of the circuit is never contracted but always deleted. This implies that the graph is noncoverable when the last but one of these edges has been deleted and thus we have case (III).

($\Rightarrow$) We prove that a coverable directed graph G with no circuit implies $|\vec{d}_{uv}(G)| = 1$. This is obviously true for the special case of G being a parallel graph. Excluding this case we apply the contraction–deletion rule, as in Theorem 2, to an appropriate arc a adjacent to u and w $(\neq v)$ and we obtain the following two cases:

(1) $\vec{d}_{uv}(G) = \vec{d}_{uv}(G^\gamma)$ and G^γ has no circuit,

(2) $\vec{d}_{uv}(G) = -\vec{d}_{uv}(G^\delta)$ and G^δ has no circuit.

It is easy to see that in Case 1 G^γ is coverable. It is also true that in Case 2 G^δ is coverable. Let (i) $\{\pi_i\}_1^r$, (ii) $\{\pi_i'\}_1^s$, and (iii) $\{\pi_i''\}_1^t$, be the set of paths in G such that they (a) do not pass through w, (b) are from u to w not containing a, (c) are from w to v. The sets (ii) and (iii) are nonempty. The fact that G^δ does not contain a circuit implies that each composite path $\pi_i' \circ \pi_j''$ is a self avoiding path from u to v.

The collection of paths $\{\pi_i\}_1^r \cup \{\pi_i' \circ \pi_j''\}_1^{s,t}$, is a covering for G^δ.

Thus we obtain a graph G' ($= G^\gamma$ or G^δ) which is either parallel in which case the result is proved or we can apply the above arguments to G'. Eventually the graph G is reduced to a parallel graph which is coverable. Thus $|\vec{d}_{uv}(G)| = 1$.

Proof of Theorem 4: The result follows easily for parallel graphs. Thus as in the proof of Theorem 3 we can apply the contraction–deletion rule to an arc a of G, with vertices u and w $(w \neq v)$, and obtain two cases

(1) $\vec{d}_{uv}(G^\gamma) = \vec{d}_{uv}(G)$, G^γ has no circuit,

(2) $\vec{d}_{uv}(G^\delta) = -\vec{d}_{uv}(G)$, G^δ has no circuit.

We show that the number of independent paths in G and G^γ are equal and differs from the number of independent paths in G^δ by one. Thus we see that the result will be true if it is true for the derived graphs.

In both cases the graphs are coverable. First of all we show that in Case (1) G and G^γ have the same maximal number of independent paths. Given the natural one—one correspondence between paths of G and G^γ we see a maximal independent set $\mathcal{C}^\gamma$ of directed paths on G^γ gives an independent set $\mathcal{C}$ for G. Suppose $\mathcal{C}$ is not maximal. Then there exists a path π_0 such that $\mathcal{C}\cup\{\pi_0\}$ is independent. However, if π_0^γ is the corresponding path in G^γ the set $\mathcal{C}^\gamma\cup\{\pi_0^\gamma\}$ is not independent. Hence

$$\pi_0^\gamma=\sum_{i=1}^n \alpha_i\pi_i^\gamma,\quad \pi_i^\gamma\in\mathcal{C}^\gamma,\quad \alpha_i\in\mathbb{Q}. \qquad (*)$$

In the cases $\pi_0\ni a$, $\pi_0\not\ni a$ the coefficients α_i at the vertex w in $(*)$ sum to 1 and 0 respectively and therefore

$$\pi_0=\sum_{i=1}^n \alpha_i\pi_i.$$

Thus $\mathcal{C}\cup\{\pi_0\}$ is not independent.

In Case (2) we show the maximal number of independent paths for G^δ differs by one from that of G. Let $\mathcal{C}^\delta$ be a maximal independent set of paths for G^δ. Then if π_0 is any path in G with initial arc a we claim $\mathcal{C}=\mathcal{C}^\delta\cup\{\pi_0\}$ is maximal for G. Independence is obvious. For maximality suppose $\exists$ a path π in G such that $\mathcal{C}\cup\{\pi\}$ is independent. We have

(i) $a\notin\pi$. π is a path in G^δ and therefore $\mathcal{C}\cup\{\pi\}=\mathcal{C}^\delta\cup\{\pi_0,\pi\}$ is not independent.

(ii) $a\in\pi\ (\neq\pi_0)$. Let π_1 be any fixed path from u to w in G^δ and $\pi'=\pi\,|\,G^\delta$ and $\pi_0'=\pi_0\,|\,G^\delta$. By the no-circuit property we have directed paths $\pi_1\circ\pi'$ and $\pi_1\circ\pi_0'$ from u to v. Moreover,

$$\pi-\pi_0=\pi_1\circ\pi'-\pi_1\circ\pi_0'=\sum\alpha_i\pi_i^\delta,\quad \pi_i^\delta\in\mathcal{C}^\delta,$$

for some $\alpha_i\in\mathbb{Q}$, and thus $\mathcal{C}\cup\{\pi\}$ is not independent.

Thus the result can be proved inductively by reducing G after a finite number of steps to a parallel graph for which the result is trivially true.

Proof of Lemma 5: Let $\pi_1,\ldots,\pi_n$ be a maximal independent set (MIS) of paths for G.

Define a sequence of subgraphs $G_i=\cup_{j=1}^i\pi_j$. Moreover, define $\mu(G_i)=E_i-V_i+2$, where $E_j(V_j)$ are the number of edges (vertices) of the graph G_j. We note that $\mu(G_1)=+1$ which trivially equals the number of elements in a MIS of paths for G_1.

We assume the result that $\mu(G_j)$ equals the number of elements in a MIS of paths in G_j and show that it is also true for G_{j+1}.

The graph $G_{j+1}=\pi_{j+1}\cup G_j$ and therefore $G_{j+1}\backslash G_j$ is a union of k distinct chains, each chain connected to two distinct vertices of G_j.

The degenerate case of $G_{j+1}=G_j$ $(k=0)$ is trivial because $\mu(G_{j+1})=\mu(G_j)$. If $k\neq 0$, then for each of the chains we can find a path from u to v containing that chain with all other edges and vertices in G_j.

Let $\rho_1',\ldots,\rho_k'$ be the paths so obtained. We claim that these paths together with a MIS for G_j, say $\{\rho_1,\ldots,\rho_{\mu(j)}\}$, form a MIS for G_{j+1}. It is obvious that the collection is an independent set of paths. To prove that the collection is maximal, let π be a path in G_{j+1}. Then if $\pi\subseteq G_j$, obviously π is not independent of the collection.

If $\pi\nsubseteq G_j$, then $\exists$ a collection of paths $\rho_{i_1},\ldots,\rho_{i_l}$ such that

$$\pi-\sum_{j=1}^{l}\rho_{i_j}=\text{a union of paths in } G_j=\sum_{i=1}^{\mu(j)}\alpha_i\rho_i$$

which implies that π is not independent of the set $\{\rho_1',\ldots,\rho_k',\ \rho_1,\ldots,\rho_{\mu(j)}\}$.

Also, $\mu(G_{j+1})=\mu(G_j)+k$. Therefore $\mu(G_{j+1})$ equals the number of elements in a MIS for G_{j+1}.

Thus $\mu(G)=\mu(G_n)=$ number of elements in a MIS of paths for G.

If an additional arc a_0 is attached to G oriented from v to u the derived graph $\bar{G}$ is strongly connected and from (12), $\mu(G)=\nu(\bar{G})$.

We note that the independent paths $\{\pi_1,\ldots,\pi_n\}$ give rise to a set of independent circuits $(\pi_1,a_0),\ldots,(\pi_n,a_0)$. However, for strongly connected graphs, cycles can be generated by circuits and so a maximal collection of independent circuits has $\nu(\bar{G})$ elements and therefore a maximal set of independent "external" circuits (containing arc a_0) must constitute a basis for the cycle space on $\bar{G}$.

[1]J.W. Essam, J. Math. Phys. 12, 874—82 (1971).
[2]A.G. Dunn, J.W. Essam, and J.M. Loveluck, J. Phys. C. 8 743—50 (1975).
[3]K.W. Cattermole, Electron. Lett. 8, 316—18 (1972).
[4]C. Clos, Bell Syst. Tech. J. 32, 406—24 (1953).
[5]P.W. Kasteleyn and C.M. Fortuin, J. Phys. Soc. Jpn. Suppl. 26, 11—13 (1969).
[6]G.C. Rota, Z. Wahrsheinlichkeitstheorie 2, 340—68 (1964).
[7]In Ref. 1 the expected number of components for the undirected case $\langle n(G)\rangle$ was considered. However, $n(G)=c(G)-|E|+|V|$, where $\langle|V|\rangle=\sum_{w\in V}p_w$ and $\langle|E|\rangle=\sum_{a\in A}p_ap_{w_+}p_{w_-}$, $\partial a=w_+-w_-$, and so $\langle c(G)\rangle$ is trivially related to $\langle n(G)\rangle$. The "k weights" are the same as defined in Ref. 1 except that here $k(\cdot)=k(\longrightarrow)=0$ whereas in Ref. 1 $k(\cdot)=1$ and $k(\longrightarrow)=-1$.
[8]C. Berge, *The Theory of Graphs*, Methuen (London) 1962, p. 29.
[9]H.H. Crapo, J. Comb. Theory 2, 406—17 (1967).
[10]C. Greene private communication with G.C. Rota (1975).

The Killing form for graded Lie algebras

Nigel B. Backhouse

Department of Applied Mathematics and Theoretical Physics, University of Liverpool, Liverpool, England
(Received 17 December 1975)

A Killing form with nice symmetry and invariance properties is constructed for an arbitrary graded Lie algebra. When this form is nondegenerate, Casimir operators can be constructed, and the graded Lie algebra possesses properties analogous to those possessed by semisimple ordinary Lie algebras.

1. INTRODUCTION

Recent years have seen the appearance of so-called super-symmetry models in particle physics, combining bosons and fermions, for which the underlying symmetry defines a mathematical structure known as a graded Lie algebra (GLA). Crudely speaking, such an algebra possesses a binary operation which is a commutator for some pairs of elements but is an anticommutator for some other pairs of elements. Details of the applications of GLA to particle physics can be found in the excellent review article of Corwin, Ne'eman, and Sternberg.[1] Independently, Allcock[2,3] has defined structures in a generalized classical mechanics which satisfy the axioms for GLA. The mathematics of GLA, which were originally defined by Nijenhuis,[4] are briefly discussed in Ref. 1 and further considered in a very recent paper of Pais and Rittenberg.[5] The latter propose a definition for semisimple GLA and then classify all such GLA. It is the purpose of the present paper to develop some of the mathematical properties of GLA beyond those revealed in Refs. 1 and 5. Hopefully we have made a start on some sort of structure theory for GLA.

First we define a GLA (we only consider $\mathbf{Z}_2$ gradings). Let $\mathbf{L}=\mathbf{L}_0\oplus\mathbf{L}_1$ be a direct sum of real or complex vector spaces $\mathbf{L}_0$ (the even space) and $\mathbf{L}_1$ (the odd space). Define a sign function σ on $\mathbf{L}_0$ and $\mathbf{L}_1$ by $\sigma(l)=0,1$, according as $l\in\mathbf{L}_0$, $\mathbf{L}_1$, respectively. Then $\mathbf{L}$, together with a bilinear product $[\ ,\ ]$, becomes a GLA provided

$$[\mathbf{L}_0,\mathbf{L}_0]\subseteq\mathbf{L}_0,\quad [\mathbf{L}_0,\mathbf{L}_1]=[\mathbf{L}_1,\mathbf{L}_0]\subseteq\mathbf{L}_1,\quad [\mathbf{L}_1,\mathbf{L}_1]\subseteq\mathbf{L}_0; \tag{A1}$$

$$[l,m]=-(-1)^{\sigma(l)\sigma(m)}[m,l]\quad\text{for all } l,m\in\mathbf{L}_0 \text{ or } \mathbf{L}_1; \tag{A2}$$

$$[l,[m,n]]=[[l,m],n]+(-1)^{\sigma(l)\sigma(m)}[m,[l,n]], \tag{A3}$$

for all $l,\ m,\ n\in\mathbf{L}_0$ or $\mathbf{L}_1$.

The axiom (A2) implies that $[\ ,\ ]$ is anticommutative (like a commutator) on $\mathbf{L}_0\times\mathbf{L}_0$, $\mathbf{L}_0\times\mathbf{L}_1$, and $\mathbf{L}_1\times\mathbf{L}_0$, but is commutative (like an anticommutator) on $\mathbf{L}_1\times\mathbf{L}_1$. The axiom (A3), which is the graded Jacobi identity, can be rewritten in the more symmetrical form

$$(-1)^{\sigma(l)\sigma(m)}[[l,m]n]+(-1)^{\sigma(m)\sigma(n)}[[m,n],l]$$
$$+(-1)^{\sigma(n)\sigma(l)}[[n,l],m]=0,\quad\text{for all } l,m,n\in\mathbf{L}_0 \text{ or } \mathbf{L}_1.$$

It is evident that the even space $\mathbf{L}_0$ is an ordinary Lie algebra (LA). Although it turns out that GLA possesses properties analogous to those for LA, it is the case that the structure and properties peculiar to any given GLA do not correspond nicely to those of its even space. For this reason it is often rewarding to deemphasize the structure of the even space and to consider the GLA as a whole.

A most important tool in the study of LA is the Killing form K. If $\mathbf{L}$ denotes an LA and $l\to\operatorname{ad}l$, $l\in\mathbf{L}$, the adjoint representation defined by $(\operatorname{ad}l)m=[l,m]$, $m\in\mathbf{L}$, then $K:\mathbf{L}\times\mathbf{L}\to$ field of scalars is given by $K(l,m)=\operatorname{Tr}(\operatorname{ad}l\operatorname{ad}m)$ for $l,m\in\mathbf{L}$. The importance of K stems from its nice symmetry and invariance properties which in particular help in the unfolding of the beautiful theory of semisimple LA. In this theory the nondegeneracy of K provides, among other things, an inner product in the root space and permits the construction of Casimir operators of arbitrary degree. It would be very pleasing if an analogous theory could be developed for a certain class of GLA. Corwin *et al*[1] showed by example that it is not helpful to define a form for a GLA in precisely the same way as done above for LA. They did not, however, say that a slight modification yields a bilinear form with satisfactory invariance and symmetry properties. It is our first job in this paper to establish the existence of such a graded Killing form and generalizations to higher order. In this we have been guided by the definition of the metric tensor given in Ref. 5.

When the Killing form of a GLA is nondegenerate we can construct Casimir operators in the universal enveloping algebra, so answering a query of Pais and Rittenberg.[5] Under the same hypothesis we obtain a result which is analogous to the easy part of Cartan's criterion for semisimple LA, namely the nonexistence of Abelian ideals. Although the converse is not true, we are able to mimic for GLA with nondegenerate form (called nondegenerate GLA) a number of the classic results from the theory of semisimple LA.

We have drawn quite freely from the texts on LA and in particular from Refs. 6–8. The LA expert will find several places where we could have been more economical in our proofs, however, we have aimed at variety to show what seems possible within the theory of GLA.

2. INVARIANT FORMS

We construct invariant forms for $\mathbf{L}$, a GLA, by means of the following device: Let T be any linear transformation of $\mathbf{L}$ (as a vector space) into itself, then define a new linear transformation T' by setting $T'l=l_0-l_1$, where $Tl=l_0+l_1$ for $l_0\in\mathbf{L}_0$, $l_1\in\mathbf{L}_1$. We can clearly write $T'=P_0T-P_1T$ where P_0, P_1 are the orthogonal projections onto $\mathbf{L}_0$, $\mathbf{L}_1$, respectively. Let $l\in\mathbf{L}$, then $\operatorname{ad}l$, defined by $(\operatorname{ad}l)m=[l,m]$ is a linear transformation of L into itself. We define the multilinear maps $K_n: \mathbf{L}\times\mathbf{L}\times\cdots\times\mathbf{L}$ (n factors) → field of scalars by

$$K_n(l_1, l_2, \ldots, l_n) = \mathrm{tr}(\mathrm{ad}l_1\mathrm{ad}l_2\cdots\mathrm{ad}l_n)' \tag{2.1}$$

for all $l_1, l_2, \ldots, l_n \in \mathbf{L}$. To establish the symmetry of K_n we need the following result:

Lemma 1: (a) If $l\in\mathbf{L}_0$, then $P_0\mathrm{ad}l = P_0\mathrm{ad}lP_0 = \mathrm{ad}lP_0$, and $P_1\mathrm{ad}l = P_1\mathrm{ad}lP_1 = \mathrm{ad}lP_1$;

(b) If $l\in\mathbf{L}_1$, then $P_0\mathrm{ad}l = P_0\mathrm{ad}lP_1 = \mathrm{ad}lP_1$ and $P_1\mathrm{ad}l = P_1\mathrm{ad}lP_0 = \mathrm{ad}lP_0$.

Proof: The result is an easy consequence of axiom (A1) and the fact that P_0, P_1 annihilates $\mathbf{L}_1$, $\mathbf{L}_0$, respectively.

Now we can state

Theorem 1:

$$K_n(l_1, l_2, l_3, \ldots, l_n) = (-1)^{\sigma(l_n)}K_n(l_n, l_1, \ldots, l_{n-1}),$$

for all $l_1, l_2, \ldots, l_n \in \mathbf{L}_0$ or $\mathbf{L}_1$.

Proof:

$$\begin{aligned}K_n(l_1, l_2, \ldots, l_n) &= \mathrm{Tr}(P_0\mathrm{ad}l_1\mathrm{ad}l_2\cdots\mathrm{ad}l_n - P_1\mathrm{ad}l_1\mathrm{ad}l_2\cdots\mathrm{ad}l_n)\\ &= \mathrm{Tr}(\mathrm{ad}l_nP_0\mathrm{ad}l_1\cdots\mathrm{ad}l_{n-1}\\ &\quad - \mathrm{ad}l_nP_1\mathrm{ad}l_1\cdots\mathrm{ad}l_{n-1}),\end{aligned}$$

using the symmetry of the trace operation. If $l_n\in\mathbf{L}_0$, then $\mathrm{ad}l_nP_0 = P_0\mathrm{ad}l_n$ and $\mathrm{ad}l_nP_1 = P_1\mathrm{ad}l_n$, by Lemma 1(a). If $l_n\in\mathbf{L}_1$, then $\mathrm{ad}l_nP_0 = P_1\mathrm{ad}l_n$ and $\mathrm{ad}l_2P_1 = P_0\mathrm{ad}l_n$, by Lemma 1(b). The conclusion of the theorem follows.

Corollary: $K_n(l_1, l_2, \ldots, l_n) = 0$ if the set $\{l_1, l_2, \ldots, l_n\}$ contains an odd number of elements of $\mathbf{L}_1$, where $l_1, l_2, \ldots, l_n \in \mathbf{L}_0$ or $\mathbf{L}_1$.

Proof: By iterating Theorem 1 we obtain

$$\begin{aligned}&K_n(l_1, l_2, \ldots, l_n)\\ &\quad= (-1)^{[\sigma(l_n)+\sigma(l_{n-1})+\cdots+\sigma(l_1)]}K_n(l_1, l_2, \ldots, l_n),\end{aligned}$$

from which the result follows—recall that $\sigma(l) = 0, 1$ according as $l\in\mathbf{L}_0$ or $\mathbf{L}_1$.

Before stating and proving the invariance properties of K_n we need the following lemma.

Lemma 2:

$$\mathrm{ad}[l, m] = \mathrm{ad}l\mathrm{ad}m - (-1)^{\sigma(l)\sigma(m)}\mathrm{ad}m\mathrm{ad}l,$$

for all $l, m\in\mathbf{L}_0$ or $\mathbf{L}_1$.

Proof:

$$\begin{aligned}(\mathrm{ad}[l, m])n = [[l, m], n] &= [l, [m, n]]\\ &\quad - (-1)^{\sigma(l)\sigma(m)}[m, [l, n]]\end{aligned}$$

by (A3). The conclusion is now immediate.

This has been leading up to

Theorem 2:

$$\begin{aligned}&K_n([l, l_1], l_2, \ldots, l_n) + (-1)^{\sigma(l)\sigma(l_1)}K_n(l_1, [l, l_2], l_3, \ldots, l_n)\\ &\quad + \cdots + (-1)^{\sigma(l)[\sigma(l_1)+\sigma(l_2)+\cdots+\sigma(l_{n-1})]}K_n(l_1, l_2, \ldots,\\ &\quad \times l_{n-1}, [l_1, l_n]) = 0 \quad \text{for all } l, l_1, l_2, \ldots, l_n \in \mathbf{L}_0 \text{ or } \mathbf{L}_1.\end{aligned}$$

Proof: Using Lemma 2 the first term becomes

$$K_{n+1}(l, l_1, l_2, \ldots, l_n) - (-1)^{\sigma(l)\sigma(l_1)}K_{n+1}(l_1, l, l_2, \ldots, l_n).$$

The second term becomes

$$\begin{aligned}&(-1)^{\sigma(l)\sigma(l_1)}K_{n+1}(l_1, l, l_2, \ldots, l_n)\\ &\quad - (-1)^{\sigma(l)[\sigma(l_1)+\sigma(l_2)]}K_{n+1}(l_1, l_2, l, \ldots, l_n).\end{aligned}$$

The last term becomes

$$\begin{aligned}&(-1)^{\sigma(l)[\sigma(l_1)+\cdots+\sigma(l_{n-1})]}K_{n+1}(l_1, l_2, \ldots, l, l_n)\\ &\quad - (-1)^{\sigma(l)[\sigma(l_1)+\cdots+\sigma(l_n)]}K_{n+1}(l_1, l_2, \ldots, l_n, l).\end{aligned}$$

On addition of all of these terms we get cancellation, leaving

$$\begin{aligned}&K_{n+1}(l, l_1, l_2, \ldots, l_n) - (-1)^{\sigma(l)[\sigma(l_1)\cdots\sigma(l_n)]}\\ &\quad\times K_{n+1}(l_1, l_2, \ldots, l_n, l),\end{aligned}$$

which can be rewritten as

$$[1 - (-1)^{\sigma(l)[\sigma(l)+\sigma(l_1)+\cdots+\sigma(l_n)]}]K_{n+1}(l, l_1, l_2, \ldots, l_n) \tag{2.2}$$

by Theorem 1, bearing in mind that $\sigma(l) = \sigma(l)^2$. If $\sigma(l) = 0$, (2.2) vanishes. If $\sigma(l) = 1$ and $\sigma(l) + \cdots + \sigma(l_n)$ is even, (2.2) vanishes. If $\sigma(l) = 1$ and $\sigma(l) + \cdots + \sigma(l_n)$ is odd, $K_{n+1}(l, l_1, \ldots, l_n)$ vanishes by the corollary to Theorem 1, so again (2.2) vanishes. This concludes the proof.

We have shown that there exist invariant forms of arbitrary degree (unless some of them vanish identically) for any GLA. In particular there exists a bilinear form K_2 from which we obtain the Killing form K defined by

$$K(l, m) = K_2(m, l) = \mathrm{Tr}(\mathrm{ad}m\mathrm{ad}l)' = \mathrm{Tr}(\mathrm{ad}l(\mathrm{ad}m)'),$$

for all $l, m\in\mathbf{L}$. We are not being perverse in defining K so, but it eases slightly the proof of Theorem 3 on Casimir operators. The Killing form K satisfies certain symmetry and invariance equations.

Lemma 3: (a) $K(l, m) = K(m, l)$ for all $l, m\in\mathbf{L}_0$;

(b) $K(l, m) = K(m, l) = 0$ for all $l\in\mathbf{L}_0$, $m\in\mathbf{L}_1$;

(c) $K(l, m) = -K(m, l)$ for all $l, m\in\mathbf{L}_1$;

(d) $K([l, m], n) + (-1)^{\sigma(l)\sigma(n)}K(m, [l, n]) = 0$

for all $l, m, n\in\mathbf{L}_0$ or $\mathbf{L}_1$.

Proof: (a)–(d) are easy deductions from Theorems 1 and 2.

So any GLA does indeed possess a respectable Killing form, which differs from the Killing form for an ordinary LA by having more complicated symmetries. Whereas the Killing form for an LA is symmetric, for a GLA it is the direct sum of a symmetric form on the even space and a skew-symmetric form on the odd space. It is as if a GLA combines Riemannian with symplectic structure. We must warn, however, that the restriction to $\mathbf{L}_0$ of the Killing form of L is not the Killing form of $\mathbf{L}_0$. It is this which makes it difficult correlating properties of $\mathbf{L}_0$ with properties of L.

Invariant forms are of great importance in the theory of LA. However, in physical applications, their main use is in the construction of Casimir operators. Let us now assume that the graded Lie algebra L has a nondegenerate Killing form K. We take a basis $\{l_\alpha\}$, α

$=1, 2, \ldots, t$, for L, where the first s of the l_α lie in L_0 and the last $t-s$ lie in L_1. Define a dual basis $\{l^\alpha\}$, $\alpha = 1, 2, \ldots, t$, by the condition $K(l^\alpha, l_\beta) = \delta_{\alpha\beta}$ for $\alpha, \beta = 1, 2, \ldots, t$. Note that since K is nondegenerate and Lemma 3(b) holds, it is separately nondegenerate on L_0 and L_1; hence $l^\alpha \in L_0$, $\alpha = 1, \ldots, s$, and $l^\alpha \in L_1$, $\alpha = s+1, \ldots, t$. Then we show presently that

$$C_n = \sum_{\alpha_1, \alpha_2, \cdots, \alpha_n = 1}^{t} K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, l_{\alpha_n}) l^{\alpha_n} l^{\alpha_{n-1}}, \ldots, l^{\alpha_1}$$

is invariant under L. The first thing to decide is just what we mean by invariance under L. Just as for ordinary LA, so for L a GLA, we can construct a so-called universal enveloping algebra U(L), which can be thought of as consisting of linear combinations of monomials in the elements of L. Ordinary multiplication makes U(L) into an associative algebra. Furthermore, we can define a bracket operation on L×U(L) by inductively extending the differentiation property of the adjoint representation, namely $(\mathrm{ad} l)mn = [l, mn] = [l, m]n + (-1)^{\sigma(l)\sigma(m)} \times m[l, n]$, where $l, m \in L_0$ or L_1 and $n \in U(L)$. We also have the interpretation $[l, m] = lm \pm ml$, where the plus sign occurs if $l, m \in L_1$, the minus sign for $l, m \in L_0$, $l \in L_0$, $m \in L_1$, or $l \in L_1$, $m \in L_0$. This is useful in reducing expressions. We say that $u \in U(L)$ is invariant under L if $(\mathrm{ad} l)u = 0$ for all $l \in L$. We remark that Corwin *et al.*[1] have proved a graded version of the Poincaré–Birkhoff–Witt theorem which says that a basis for U(L) consists of a commutative identity together with monomials of the form $(l_1)^{i_1}(l_2)^{i_2} \cdots (l_t)^{i_t}$ where $i_1, \ldots, i_s$ are arbitrary integral exponents but $i_{s+1}, \ldots, i_t$ can only assume the values 0 or 1. In this we have l_α^0 = the identity. Before proving Theorem 3 we need:

Lemma 4: If $[l, l^\alpha] = \sum_\beta L_{\alpha\beta} l^\beta$ and $[l, l_\beta] = \sum M_{\alpha\beta} l_\alpha$, then $M_{\alpha\beta} = -(-1)^{\sigma(l)\sigma(l_\beta)} L_{\alpha\beta}$.

Proof: Since $K(l^\alpha, l_\beta) = \delta_{\alpha\beta}$, we have $M_{\alpha\beta} = K(l^\alpha, [l, l_\beta]) = -(-1)^{\sigma(l)\sigma(l_\beta)} K([l, l^\alpha], l_\beta)$, by Lemma 3(d). The latter is $-(-1)^{\sigma(l)\sigma(l_\beta)} L_{\alpha\beta}$, as required.

Now we have:

Theorem 3:

$$C_n = \sum_{\alpha_1, \alpha_2, \cdots, \alpha_n = 1}^{t} K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, l_{\alpha_n}) l^{\alpha_n} l^{\alpha_{n-1}} \cdots l^{\alpha_1}$$

is invariant under L.

Proof: Write $\sigma(l_\alpha) = \sigma(l^\alpha) = \sigma(\alpha)$. Now $(\mathrm{ad} l)C_n$, for $l \in L_0$ or L_1, is

$$\begin{aligned}
&\sum K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, l_{\alpha_n})[l, l^{\alpha_n}] l^{\alpha_{n-1}} \cdots l^{\alpha_1} \\
&\quad + \sum K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, l_{\alpha_n}) l^{\alpha_n} (-1)^{\sigma(l)\sigma(\alpha_n)} [l, l^{\alpha_{n-1}}] \cdots l^{\alpha_1} \\
&\quad + K_n(l_\alpha, \ldots, l_{\alpha_n}) l^{\alpha_n} l^{\alpha_{n-1}} (-1)^{\sigma(l)[\sigma(\alpha_n)+\sigma(\alpha_{n-1})]} \\
&\quad \times [l, l^{\alpha_{n-2}}] \cdots l^{\alpha_1} + \cdots \\
&\quad + \sum K_n(l_{\alpha_1}, \ldots, l_{\alpha_n}) l^{\alpha_n} l^{\alpha_{n-1}} \cdots l^{\alpha_2} \\
&\quad \times (-1)^{\sigma(l)[\sigma(\alpha_n)+\cdots+\sigma(\alpha_2)]} [l, l^{\alpha_1}].
\end{aligned} \tag{2.3}$$

Now each $[l, l^{\alpha_i}]$ is expanded in the dual basis, the matrix elements are taken inside $K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, l_{\alpha_n})$ and the sum over l^{α_i} performed. By Lemma 4 this gives us terms like $-(-1)^{\sigma(l)\sigma(\beta_i)} K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, [l, l_{\beta_i}], \ldots, l_{\alpha_n})$ with l^{β_i} in the $(n-i)$th place in the monomial. We now rename β_i as α_i, to give us from (2.3) the expression

$$\begin{aligned}
&-\sum\{(-1)^{\sigma(l)\sigma(\alpha_n)} K_n(l_{\alpha_1}, l_{\alpha_2}, \ldots, [l, l_{\alpha_n}]) \\
&\quad + (-1)^{\sigma(l)[\sigma(\alpha_n)+\sigma(\alpha_{n-1})]} K_n(l_{\alpha_1}, \ldots, [l, l_{\alpha_{n-1}}], l_{\alpha_n}) \\
&\quad + \cdots + (-1)^{\sigma(l)[\sigma(\alpha_n)+\cdots+\sigma(\alpha_1)]} K_n([l, l_{\alpha_1}], l_{\alpha_2} \cdots l_{\alpha_n})\} \\
&\quad \times l^{\alpha_n} l^{\alpha_{n-1}} \cdots l^{\alpha_1}.
\end{aligned} \tag{2.4}$$

If we multiply the expression in (2.4) within braces by $\sigma(l)[\sigma(\alpha_1) + \sigma(\alpha_2) + \cdots + \sigma(\alpha_n)]$, then because of Theorem 2 the expression vanishes. So the whole expression (2.4) vanishes, which concludes the proof.

Theorem 3 answers question 3 of Pais and Rittenberg—see Sec. 7 of Ref. 5. Now consider as an example the di-spin algebra defined in Ref. 1. L_0 has basis e, h, f and L_1 has basis x, y, where

$$\begin{aligned}
&[h, e] = 2e, \quad [h, x] = x, \quad [x, x] = e, \\
&[h, f] = -2f, \quad [h, y] = -y, \quad [y, y] = -f, \\
&[e, f] = h, \quad [f, x] = y, \quad [x, y] = -\tfrac{1}{2}h, \quad [e, y] = x,
\end{aligned} \tag{2.5}$$

and all other brackets are zero. This is essentially the algebra denoted GSU(2) in Ref. 5. It is easy to check that K_1 vanishes identically—a good reason for this will be given later—and moreover, that there is no linear Casimir invariant (because the di-spin algebra is centerless). The Killing form, in the given basis, has matrix

$$\begin{pmatrix} 0 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix}. \tag{2.6}$$

The dual basis is $\{\frac{1}{3}f, \frac{1}{6}h, \frac{1}{3}e, -\frac{1}{3}y, \frac{1}{3}x\}$ and hence the second order Casimir invariant is

$$C_2 = \tfrac{1}{3}(fe + ef) + \tfrac{1}{6}h^2 + \tfrac{1}{3}(xy - yx). \tag{2.7}$$

This can be written as $\frac{2}{3}(ef + xy) + \frac{1}{6}(h^2 - h)$ if we use $ef - fe = h$ and $xy + yx = -\frac{1}{2}h$. It is easy to show that in the representation given in Ref. 1, the above Casimir invariant assumes the value $\frac{1}{6}n(n+1)$ times the identity matrix. This agrees with the corresponding result in Ref. 5. Higher order invariant forms are more tedious to obtain, but use can be made of Theorem 1 and the following lemma.

Lemma 5:

$$\begin{aligned}
&[K_{n-1}([l_1, l_2], l_3, \ldots, l_n) \\
&\quad = K_n(l_1, l_2, \ldots, l_n) - (-1)^{\sigma(l_1)\sigma(l_2)} \\
&\quad \times K_n(l_2, l_1, l_3, \ldots, l_n).
\end{aligned}$$

Proof: This an easy deduction from Lemma 2.

For the cubic form of the di-spin algebra we find that the only nonzero matrix elements are $K_3(h, e, f) = 3$, $K_3(f, x, x) = \frac{3}{2}$, $K_3(e, y, y) = -\frac{3}{2}$, $K_3(h, x, y) = -\frac{3}{2}$, and that related through Lemma 5 and Theorem 1. Thus $K_3(e, h, f) = -3$, $K_3(x, f, x) = -\frac{3}{2}$, $K_3(y, e, y) = \frac{3}{2}$, and $K_3(x, h, y) = \frac{3}{2}$. When we form the third order Casimir C_3 we find that it can be reduced to the second order expression $\frac{1}{4}C_2$. One suspects that C_2 is the only independent invariant operator.

3. NONDEGENERATE GRADED LIE ALGEBRAS

This section consists of a first listing of properties of certain GLA suggested by the classical theory of LA. We begin with some definitions. **L** will always denote a GLA $\mathbf{L}_0 \oplus \mathbf{L}_1$.

(1) A linear subset **M** of **L** is an ideal if $[\mathbf{L}, \mathbf{M}] \subseteq \mathbf{M}$.

(2) An ideal **M** of **L** is a graded ideal if wherever $m \in \mathbf{M}$ is written as $m = m_0 + m_1$, where $m_0 \in \mathbf{L}_0$, $m_1 \in \mathbf{L}_1$, we have also $m_0, m_1 \in \mathbf{M}$. Equivalently $\mathbf{M} = (\mathbf{M} \cap \mathbf{L}_0) \oplus (\mathbf{M} \cap \mathbf{L}_1)$.

(3) An ideal **M** is Abelian if $[\mathbf{M}, \mathbf{M}] = \mathbf{O}$.

The notion of a graded ideal, as given in Ref. 1, would seem to be the significant one for GLA. In particular we note without proof that the set of graded ideals of a GLA is closed under addition, intersection, and the bracket operation (by the graded Jacobi identity). Graded ideals occur naturally as the kernels of structure preserving transformations between GLA.

It is tempting to say that a GLA is semisimple if it contains no nontrivial graded Abelian ideals. However, at this early stage in the game, and especially, as we show later, such GLA do not correspond nicely with the semisimple GLA already defined by Pais and Rittenberg[5]; we refrain from making such a definition and so causing terminological confusion. It turns out, however, that some rather nice properties do hold for GLA whose Killing form is nondegenerate—for these we offer the name nondegenerate GLA.

If the Killing form K is degenerate it possesses a nonzero kernel $\mathbf{M} = \{l \in \mathbf{L} : K(l, m) = 0 \text{ for all } m \in \mathbf{L}\}$. More generally, if **N** is a subset of **L**, its left orthogonal complement $\mathbf{N}_\lambda^\perp$ can be defined as $\{l \in \mathbf{L} : K(l, n) = 0 \text{ for all } n \in \mathbf{N}\}$. Evidently $\mathbf{N}_\lambda^\perp$ is a linear subset of **L** even if **N** is not. The right orthogonal complement $\mathbf{N}_\rho^\perp = \{l \in \mathbf{L} : K(n, l) = 0 \text{ for all } n \in \mathbf{N}\}$. In general there is no reason to suppose $\mathbf{N}_\rho^\perp = \mathbf{N}_\lambda^\perp$. However, given a subset **N**, define $\mathbf{N}' = \{n_0 - n_1 : \text{for all } n_0 + n_1 \in \mathbf{N}, \text{ where } n_0 \in \mathbf{L}_0, n_1 \in \mathbf{L}_1\}$. Then

Lemma 6: $(\mathbf{N}')_\rho^\perp = \mathbf{N}_\lambda^\perp$.

Proof: Suppose $l \in (\mathbf{N}')_\rho^\perp$ and $l = l_0 + l_1$ where $l_0 \in \mathbf{L}_0$, $l_1 \in \mathbf{L}_1$. Then $K(n_0 - n_1, l_0 + l_1) = 0$ for all $n_0 - n_1 \in \mathbf{N}'$ with $n_0 \in \mathbf{L}_0$, $n_1 \in \mathbf{L}_1$ and $n_0 + n_1 \in \mathbf{N}$. By Lemma 3 we deduce $K(l_0 + l_1, n_0 + n_1) = 0$, for all $n_0 + n_1 \in \mathbf{N}$. It follows that $l \in \mathbf{N}_\lambda^\perp$; hence $(\mathbf{N}')_\rho^\perp \subseteq \mathbf{N}_\lambda^\perp$. The reverse inclusion also holds, hence the result.

We say that a subset **N** is graded if $\mathbf{N}' = \mathbf{N}$. Evidently graded ideals are graded, but in general an ideal is not graded. For a graded subset Lemma 6 tells us that we can define a unique orthogonal complement $\mathbf{N}^\perp$ which is equal to both the left and right orthogonal complements. In particular there was no need to define a left and right kernel of the Killing form. We now have:

Lemma 7: The orthogonal complement of a graded ideal is a graded ideal.

Proof: Let **N** be a graded ideal of **L**. By Lemma 3(d) we can write $K([l, n^\perp], n) - K(n^\perp, [n, l]) = 0$ for all $n \in \mathbf{N}$, $n^\perp \in \mathbf{N}^\perp$, $l \in \mathbf{L}$. The second term vanishes because $[n, l] \in \mathbf{N}$ which is orthogonal to $n^\perp$. Therefore $[l, n^\perp]$ is orthogonal to n for all $n \in \mathbf{N}$, hence $[l, n^\perp] \in \mathbf{N}^\perp$ for all $l \in \mathbf{L}$, $n^\perp \in \mathbf{N}^\perp$. This ensures that $\mathbf{N}^\perp$ is an ideal.

Now let $n^\perp \in \mathbf{N}^\perp$ be decomposed as $n^\perp = n_0^\perp + n_1^\perp$, where $n_0^\perp \in \mathbf{L}_0$ and $n_1^\perp \in \mathbf{L}_1$. We have that $K(n_0^\perp + n_1^\perp, n) = 0$ for all $n \in \mathbf{N}$. In particular, since **N** is graded, $K(n_0^\perp + n_1^\perp, n_0) = 0$ for all $n_0 \in \mathbf{N} \cap \mathbf{L}_0$. By Lemma 3(b), $K(n_1^\perp, n_0) = 0$, so $K(n_0^\perp, n_0) = 0$ for all $n_0 \in \mathbf{N} \cap \mathbf{L}_0$. Also by Lemma 3(b), $K(n_0^\perp, n_1) = 0$ for all $n_1 \in \mathbf{N} \cap \mathbf{L}_0$. Since $\mathbf{N} = (\mathbf{N} \cap \mathbf{L}_0) \oplus (\mathbf{N} \cap \mathbf{L})$, we deduce that $K(n_0^\perp, n) = 0$ for all $n \in \mathbf{N}$. It follows that $n_0^\perp \in \mathbf{N}^\perp$. Similarly $n_1^\perp \in \mathbf{N}^\perp$. Hence $\mathbf{N}^\perp$ is graded.

Lemma 8: The kernel of the Killing form is a graded ideal.

Proof: This follows from Lemma 7 since the kernel of the Killing form is the orthogonal complement of the graded ideal **L**.

If K is nondegenerate, its kernel is trivial. However, we can deduce a much stronger statement:

Theorem 4: A nondegenerate GLA has no nontrivial Abelian graded ideals.

Proof: Let A be an Abelian graded ideal of **L**, and B a complementary vector subspace. Thus $\mathbf{L} = \mathbf{A} \oplus \mathbf{B}$ and $\mathbf{A} \cap \mathbf{B} = \{0\}$. If $a \in \mathbf{A}$, then ad$a$ is the zero transformation on **A**, since **A** is Abelian, and maps **B** into **A**, since **A** is an ideal. Since **A** is graded, $(\text{ad}l\,\text{ad}a)'$ is the zero transformation on **A** and maps **B** into **A**. It follows that $K(l, a) = 0$ for all $l \in \mathbf{L}$. Hence a belongs to the kernel of K, which we know to be trivial. Therefore $\mathbf{A} = \{0\}$. This concludes the proof.

The converse to Theorem 4 is not true. For example, it is easy to check that the (f, d) algebra described in Sec. II(D) of Ref. 1 has degenerate Killing form yet has no nontrivial ideals. We must not, however, be perturbed by this, for as we shall see, Theorem 4 alone implies some quite strong results. We need another technical result.

Lemma 9: Let **M** be a graded ideal of **L**. Then the restriction to **M** of the Killing form of **L** coincides with the Killing form of **M**.

Proof: The argument is similar to that involved in Theorem 4. We can now prove:

Theorem 5: Let **L** be a nondegenerate GLA. Then **L** can be written as a direct sum of graded ideals $\oplus_\alpha \mathbf{M}_\alpha$, where each $\mathbf{M}_\alpha$ is nondegenerate and has no nontrivial graded ideals.

Proof: Suppose **N** is a proper graded ideal of **L** of greatest dimension, then $\mathbf{N}^\perp$ is a graded ideal by Lemma 7. From Lemma 3(d) we can write $K([n, n^\perp], l) - K(n^\perp, [l, n]) = 0$ for all $l \in \mathbf{L}$, $n \in \mathbf{N}$, $n^\perp \in \mathbf{N}^\perp$. But **N** is an ideal so $[l, n] \in \mathbf{N}$ which is orthogonal to $n^\perp$. Hence $K([n, n^\perp], l) = 0$ for all $l \in \mathbf{L}$, which, by nondegeneracy, implies $[n, n^\perp] = 0$. So **N** commutes with $\mathbf{N}^\perp$.

Now $\mathbf{N} \cap \mathbf{N}^\perp$ is a graded ideal of **L** and also a subspace of **N**. If $\mathbf{N} \cap \mathbf{N}^\perp \neq \{0\}$ then the center Z(N) of **N** is nontrivial. It is not hard to check that Z(N) is a graded ideal of **N**. But the amusing thing is that Z(N) is also a graded ideal of **L**. This follows by applying the graded

Jacobi identity. This makes Z(N) a nontrivial Abelian graded ideal of the nondegenerate algebra **L**. The contradiction with Theorem 4 shows that $\mathbf{N}\cap\mathbf{N}^{\perp}=\{0\}$.

$\mathbf{N}+\mathbf{N}^{\perp}$ is a graded ideal of **L** which must coincide with **N** or **L** by maximality. Since $\mathbf{N}\cap\mathbf{N}^{\perp}=\{0\}$, we deduce that $\mathbf{N}^{\perp}=\{0\}$ or $\mathbf{L}=\mathbf{N}\oplus\mathbf{N}^{\perp}$. We now show that $\mathbf{N}^{\perp}=\{0\}$ is impossible. The Killing form of L restricted to N is nondegenerate since its kernel is $\mathbf{N}\cap\mathbf{N}^{\perp}=\{0\}$. The Killing form of L restricted to $\mathbf{L}_0$ is symmetric and nondegenerate; since **N** is graded the same is true of $\mathbf{N}\cap\mathbf{L}_0$. The Killing form of L restricted to $\mathbf{L}_1$ is skew symmetric and nondegenerate; since **N** is graded the same is true of $\mathbf{N}\cap\mathbf{L}_1$. We can find a basis for $\mathbf{N}\cap\mathbf{L}_0$, $\{n_\alpha\}$, $\alpha=1,\ldots,p$ say, for which $K(n_\alpha,n_\beta)=K(n_\beta,n_\alpha)=\lambda_\alpha\delta_{\alpha\beta}$, where $\lambda_\alpha\neq 0$. Let $l\in\mathbf{L}_0$, then consider $l'=l-\sum_{\alpha=1}^{p}\times[K(l,n_\alpha)n_\alpha]/\lambda_\alpha$. Evidently $K(l',n_\beta)=0$. Also l' is orthogonal to all of $\mathbf{N}\cap\mathbf{L}_1$, so $l'\in\mathbf{N}^{\perp}$ and hence $l'=0$. We can also find a basis for $\mathbf{N}\cap\mathbf{L}_1$, $\{m_\alpha^{\pm}\}$, $\alpha=1,2,\ldots,q$, say, for which $K(m_\alpha^+,m_\beta^+)=K(m_\alpha^-,m_\beta^-)=0$ for all α,β and $K(m_\alpha^+,m_\beta^-)=-K(m_\alpha^-,m_\beta^+)=\mu_\alpha\delta_{\alpha\beta}$ (this follows from the canonical form for a real skew-symmetric nondegenerate bilinear form). Let $l\in\mathbf{L}_1$; then consider

$$l'=l-\sum_{\alpha=1}^{q}\frac{1}{\mu_\alpha}(K(l,m_\alpha^-)m_\alpha^+-K(l,m_\alpha^+)m_\alpha^-).$$

Evidently $K(l',m_\beta^{\pm})=0$ for all β. Also l' is orthogonal to all of $\mathbf{N}\cap\mathbf{L}_0$, so $l'\in\mathbf{N}^{\perp}$ and $l'=0$. It follows that all elements of $\mathbf{L}_0\oplus\mathbf{L}_1$ can be expressed as linear combinations of basis elements of N, hence $\mathbf{N}=\mathbf{L}_0\oplus\mathbf{L}_1$, contrary to the proper maximality of **N**. We are left with $\mathbf{L}=\mathbf{N}\oplus\mathbf{N}^{\perp}$.

We have already observed that the Killing form of L restricted to N is nondegenerate. If the Killing form of L restricted to $\mathbf{N}^{\perp}$ were degenerate, then, since **N** is orthogonal to $\mathbf{N}^{\perp}$ and we have $\mathbf{L}=\mathbf{N}\oplus\mathbf{N}^{\perp}$, we would have that K itself is degenerate. Finally, by Lemma 9, we have that N, $\mathbf{N}^{\perp}$ have nondegenerate Killing forms. Induction completes the proof.

Corollary: If **L** is a nondegenerate GLA then $[\mathbf{L},\mathbf{L}]=\mathbf{L}$.

Proof: In the above theorem the ideals $\mathbf{M}_\alpha$, $\mathbf{M}_\beta$ commute if $\alpha\neq\beta$. It follows that $[\mathbf{L},\mathbf{L}]=\oplus_\alpha[\mathbf{M}_\alpha,\mathbf{M}_\alpha]$. Now $[\mathbf{M}_\alpha,\mathbf{M}_\alpha]$ is an ideal of $\mathbf{M}_\alpha$, so we either have $[\mathbf{M}_\alpha,\mathbf{M}_\alpha]=\{0\}$ or $\mathbf{M}_\alpha$. The former case would imply that $\mathbf{M}_\alpha$ is an Abelian graded ideal of L, contrary to Theorem 4. It follows that $[\mathbf{L},\mathbf{L}]=\oplus_\alpha\mathbf{M}_\alpha=\mathbf{L}$.

Using an obvious definition of solvability of GLA the above corollary implies that a nondegenerate GLA is not solvable. Conversely a solvable GLA must have a degenerate Killing form. On the other hand a GLA with degenerate Killing form [e.g., the (f,d) algebra of Ref. 1] does not necessarily possess ideals, so in particular is not necessarily solvable.

The above corollary also implies that the first invariant form of a nondegenerate GLA vanishes identically—Theorem 2 gives $K_1([l,m])=0$. Furthermore, there is no way of constructing a linear Casimir invariant, for such an element would lie in the (trivial) center of L. This explains why the di-spin algebra has no linear invariants.

The above results encourage one to check through some of the other theorems about semisimple LA to see if they work for nondegenerate GLA. We now turn our attention to a study of graded derivations with the aim of showing that all graded derivations of a nondegenerate GLA are inner derivations. We say that a linear transformation of L to itself is an even derivation if it preserves $\mathbf{L}_0$ and $\mathbf{L}_1$ and if $d([l,m])=[dl,m]+[l,dm]$ for all $l,m\in\mathbf{L}$. d is an odd derivation if it interchanges $\mathbf{L}_0$ and $\mathbf{L}_1$ and if $d[l,m]=[dl,m]+(-1)^{\sigma(l)}[l,dm]$ for all $m\in\mathbf{L}$, $l\in\mathbf{L}_0$, or $\mathbf{L}_1$. If $\mathbf{D}_0$, $\mathbf{D}_1$ are the spaces of even and odd derivations of L, respectively, then $\mathbf{D}=\mathbf{D}_0\oplus\mathbf{D}_1$ can be given the structure of a GLA. All we have to do is to insist that $[d,d']=dd'-(-1)^{\sigma(d)\sigma(d')}d'd$ for all d, $d'\in\mathbf{D}_0$ or $\mathbf{D}_1$.

If $l\in\mathbf{L}_0$, then $\mathrm{ad}l\in\mathbf{D}_0$ and if $l\in\mathbf{L}_1$, $\mathrm{ad}l\in\mathbf{D}_1$. These follow from the graded Jacobi identity. It follows that $\mathrm{ad}\mathbf{L}=\{\mathrm{ad}l:l\in\mathbf{L}\}$ is a graded subspace of D. Furthermore, we have:

Lemma 10: adL is a graded ideal of D.

Proof: We write $(\mathrm{ad}\mathbf{L})_0=\{\mathrm{ad}l:l\in\mathbf{L}_0\}$ and $(\mathrm{ad}\mathbf{L})_1=\{\mathrm{ad}l:l\in\mathbf{L}_1\}$. Then $\mathrm{ad}\mathbf{L}=(\mathrm{ad}\mathbf{L})_0\oplus(\mathrm{ad}\mathbf{L})_1$. To prove the result, it suffices to show that $[\mathbf{D}_0,(\mathrm{ad}\mathbf{L})_0]\subseteq(\mathrm{ad}\mathbf{L})_0$, $[\mathbf{D}_0,(\mathrm{ad}\mathbf{L})_1]\subseteq(\mathrm{ad}\mathbf{L})_1$, $[\mathbf{D}_1,(\mathrm{ad}\mathbf{L})_0]\subseteq(\mathrm{ad}\mathbf{L})_1$, $[\mathbf{D}_1,(\mathrm{ad}\mathbf{L})_1]\subseteq(\mathrm{ad}\mathbf{L})_0$.

Suppose $d\in\mathbf{D}_0$, and $\mathrm{ad}l\in(\mathrm{ad}\mathbf{L})_0$ or $(\mathrm{ad}\mathbf{L})_1$, then

$$\begin{aligned}[d,\mathrm{ad}l](m)&=d\,\mathrm{ad}l(m)-\mathrm{ad}l\,dm\\&=d[l,m]-[l,dm]\\&=[dl,m]\\&=(\mathrm{ad}\,dl)(m).\end{aligned}$$

It follows that $[d,\mathrm{ad}l]\in(\mathrm{ad}\mathbf{L})_0$ or $(\mathrm{ad}\mathbf{L})_1$ according as $l\in\mathbf{L}_0$ or $\mathbf{L}_1$.

Suppose $d\in\mathbf{D}_1$ and $\mathrm{ad}l\in(\mathrm{ad}\mathbf{L})_0$ or $(\mathrm{ad}\mathbf{L})_1$, then

$$\begin{aligned}[d,\mathrm{ad}l](m)&=d\,\mathrm{ad}l(m)-(-1)^{\sigma(l)}\mathrm{ad}l\,dm\\&=d[l,m]-(-1)^{\sigma(l)}[l,dm]\\&=[dl,m]+(-1)^{\sigma(l)}[l,dm]-(-1)^{\sigma(l)}[l,dm]\\&=(\mathrm{ad}\,dl)(m).\end{aligned}$$

It follows that $[d,\mathrm{ad}l]\in(\mathrm{ad}\mathbf{L})_1$ or $(\mathrm{ad}\mathbf{L})_0$ according to $l\in\mathbf{L}_0$ or $\mathbf{L}_1$. This concludes the proof.

Theorem 6: A nondegenerate GLA has no outer derivations.

Proof: We are asked to prove $\mathrm{ad}\mathbf{L}=\mathbf{D}$.

Now $l\to\mathrm{ad}l$ is a graded homomorphism of L into D with kernel the center of L. But L is nondegenerate so its kernel is trivial. It follows that L is embedded as adL in D, so that adL has nondegenerate Killing form K. Furthermore, since adL is a graded ideal of D, we know that K is the restriction to adL of the Killing form K' of D. If $(\mathrm{ad}\mathbf{L})^{\perp}$ is the orthogonal complement of adL with respect to K', then the nondegeneracy of K ensures that $(\mathrm{ad}\mathbf{L})\cap(\mathrm{ad}\mathbf{L})^{\perp}$ is zero. This further means that $\mathbf{N}=[\mathrm{ad}\mathbf{L},(\mathrm{ad}\mathbf{L})^{\perp}]$ is zero, for, since both adL and $(\mathrm{ad}\mathbf{L})^{\perp}$ are graded ideals, N is contained in both adL and $(\mathrm{ad}\mathbf{L})^{\perp}$. If now $d\in(\mathrm{ad}\mathbf{L})^{\perp}$, then $0=[d,\mathrm{ad}l]=\mathrm{ad}\,dl$ for all $l\in\mathbf{L}$, from which $dl=0$ for all $l\in \mathbf{L}$, from which $d=0$. Thus $(\mathrm{ad}\mathbf{L})^{\perp}=\{0\}$. To complete the proof, we take $d\in \mathbf{D}$ and note that $l\to K'(\mathrm{ad}l, d)$ is a linear functional on $\mathbf{L}$. Since K is nondegenerate there exists $m\in \mathbf{L}$ such that $K(l,m) = K'(\mathrm{ad}l, d)$ for all $l\in\mathbf{L}$. Write $\tilde{d}=d-\mathrm{ad}m$, then $K(\mathrm{ad}l,\tilde{d}) = K'(\mathrm{ad}l, d) - K'(\mathrm{ad}l, \mathrm{ad}m) = K(l,m) - K(l,m) = 0$ for all $l\in\mathbf{L}$. It follows that $\tilde{d}\in(\mathrm{ad}\mathbf{L})^{\perp}=\{0\}$. Hence $d=\mathrm{ad}m$, an inner derivation.

Again we have a result which works for nondegenerate GLA but does not work for all GLA with no nontrivial Abelian ideals (as pointed out in Ref. 1). Further problems which suggest themselves are (a) representation theory with a view to proving Schur's Lemma and complete reducibility for nondegenerate GLA; (b) root space analysis for nondegenerate GLA—perhaps this is too much to expect; and (c) classification of low dimensional GLA. There are many other obvious problems.

Note added in proof: After the author had submitted this paper he was alerted (Ref. 9) to the existence of some other preprints which mention and use the Killing form for GLA's as defined above. The major effort in these works has been to classify all simple GLA's (Refs. 10, 11).

The author is now preparing a paper on the classification of GLA's of small dimension.

[1]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. 47, 573 (1975).
[2]G. R. Allcock, Philos. Trans. R. Soc. London A 279, 487 (1975).
[3]G. R. Allcock, Proc. R. Soc. London A 344, 175 (1975).
[4]A. Nijenhuis, Proc. K. Ned. Akad. Wet. A 58, 3 (1955).
[5]A. Pais and V. Rittenberg, J. Math. Phys. 16, 2062 (1975).
[6]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[7]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[8]R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[9]I. Kaplansky, "Newsletter on graded Lie algebras, nos. 1—4" (Dept. Math., University of Chicago).
[10]P. Freund and I. Kaplansky, J. Math. Phys. 17, 228 (1976).
[11]V. G. Kac, Functional Anal. Appl. 9, 91 (1975).

Plane wave solutions in scalar tensor theories and solutions of source-free Einstein–Maxwell theory

Dipankar Ray

Department of Physics, New York University, New York, New York 10003

(Received 26 January 1976)

Vacuum solutions of the field equations that satisfy the original notion of the plane wave, namely $g_{\mu\nu}=g_{\mu\nu}(Z)$, $Z=\Sigma_\mu a\mu x^\mu$, a_μ's being constants and $g^{\mu\nu}Z_{,\mu}Z_{,\nu}=0$, are sought for both Brans–Dicke scalar–tensor theory and the more recent scalar–tensor theory due to Sen and Dunn. (Here Greek letters range from 1 to 4 and Latin letters from 1 to 3.) A complete set of solutions are obtained for both cases. Although not required at the outset, it turns out that, in both cases, the scalar field is also function of Z alone. As a by-product, one gets the complete set of solutions of the Einstein–Maxwell equations for the null electromagnetic field for the cases when $g_{\mu\nu}$ satisfies the above-mentioned requirement.

1. INTRODUCTION

Bondi, Pirani, and Robinson[1] have defined a plane gravitational wave as nonflat solution of the vacuum Einstein equation that admits a five-parameter group of motion. A question has been raised[1–3] as to whether a plane gravitational wave as defined above is really as "plane" as a plane electromagnetic wave in Maxwell's theory, i.e., for the plane gravitational wave as defined above, can $g_{\mu\nu}$ be expressed as follows:

$$g_{\mu\nu}=g_{\mu\nu}(Z),\quad Z=\sum_\mu a_\mu x^\mu,\quad g^{\mu\nu}Z_{,\mu}Z_{,\nu}=0. \tag{1.1}$$

According to Ebner,[2] there exists a subclass of plane gravitational waves as defined by Bondi, Pirani, and Robinson, namely "homogeneous plane gravitational waves" for which $g_{\mu\nu}$ can be put in the form (1.1). Also, Takeno[3] has suggested that instead of the more familiar definition due to Bondi, Pirani, and Robinson, Eqs. (1.1) should be used for defining a plane gravitational wave.

Under these circumstances, it seems worthwhile to look into other theories of gravitation, scalar tensor theories for instance, for the solution of the vacuum field equations that can be put in the form (1.1).

In this paper, we look into vacuum solutions of the field equations with $g_{\mu\nu}$ of the form (1.1) in Brans–Dicke[4] theory and also in a more recent scalar tensor theory by Sen and Dunn.[5]

2. OBSERVATIONS ON $g_{\mu\nu}$ AND $R_{\mu\nu}$

Obviously (1.1) can be rewritten as

$$g_{\mu\nu}=g_{\mu\nu}(x^4),^{6}\quad g^{44}=0. \tag{2.1}$$

From (2.1),

$$g_{i3}=\lambda g_{i1}+\mu g_{i2}, \tag{2.2}$$

λ, μ being some functions of x^4.

Also from (2.2),

$$\det|g_{\mu\nu}|=-(g_{34}-\lambda g_{14}-\mu g_{24})^2(g_{11}g_{22}-g_{12}^2).$$

Thus to have $\det|g_{\mu\nu}|<0$, we must have

$$g_{11}g_{22}-g_{12}^2>0. \tag{I}$$

Therefore, let

$$(g_{11}g_{22}-g_{12}^2)^{1/2}=\alpha, \tag{2.3}$$

where α is some real positive function of x^4. From (2.1) one gets after a little calculation

$$\begin{aligned}R_{ij}&=\frac{1}{2}\left(g^{4k}\frac{\partial g_{ik}}{\partial x^4}\right)\left(g^{4l}\frac{\partial g_{jk}}{\partial x^4}\right),^{7}\\ R_{4i}&=\frac{1}{2}\left(g^{4k}\frac{\partial g_{ik}}{\partial x^4}\right)\left(g^{4l}\frac{\partial g_{4l}}{\partial x^4}-\frac{\partial\log\sqrt{-g}}{\partial x^4}\right)\\ &\quad-\frac{1}{2}\frac{\partial}{\partial x^4}\left(g^{4k}\frac{\partial g_{ik}}{\partial x^4}\right).\end{aligned} \tag{2.4}$$

From (2.4),

$$R_i^4=0, \tag{2.5a}$$

$$R=-\tfrac{1}{2}g_{ij}\frac{\partial g^{4i}}{\partial x^4}\frac{\partial g^{4j}}{\partial x^4}, \tag{2.5b}$$

$$R_4^4=-\tfrac{1}{2}g_{ij}\frac{\partial g^{4i}}{\partial x^4}\frac{\partial g^{4j}}{\partial x^4}. \tag{2.5c}$$

With these purely mathematical observations, we proceed to study two scalar tensor theories one by one.

3. SEN–DUNN THEORY

Here the vacuum field equations are

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=\sigma_{,\mu}\sigma_{,\nu}-\tfrac{1}{2}g_{\mu\nu}\sigma_{,\alpha}\sigma^{,\alpha} \tag{3.1}$$

where $e^{\sigma\sqrt{2/3}}$ is the scalar field in Sen–Dunn theory.

Equation (3.1) can be rewritten as

$$R_{\mu\nu}=\sigma_{,\mu}\sigma_{,\nu}. \tag{3.2}$$

Considering (3.2) for $\mu=i$, $\nu=j$, and using (2.4),

$$\sigma=\tfrac{1}{2}\sum_{i=1}^{3}g^{4k}\frac{\partial g_{ik}}{\partial x^4}x^i+\text{some function of }x^4. \tag{3.3}$$

However, since R_{44} is a function of x^4 alone, we get from (3.2) that $\sigma_{,4}$ is a function of x^4 alone. But in view of (3.3), $\sigma_{,4}$ can be a function of x^4 alone if and only if

$$g^{4k}\frac{\partial g_{ik}}{\partial x^4}=c_i,\quad i=1,2,3, \tag{3.4}$$

where c_1, c_2, c_3 are constants. From (3.4) and (2.1)

$$g_{ij}\frac{\partial g^{4i}}{\partial x^4}\frac{\partial g^{4j}}{\partial x^4}=0. \qquad (3.5)$$

Equations (2.1), (2.2), and (3.5), together with inequality (I), give

$$\frac{d\lambda}{dx^4}=0, \quad \frac{d\mu}{dx^4}=0, \qquad (3.6)$$

i.e., λ, μ are constants.

From (2.1), (2.2), and (3.6),

$$g^{4k}\frac{\partial g_{ik}}{\partial x^4}=0, \quad \text{i.e.,} \quad c_i=0. \qquad (3.7)$$

From (3.3) and (3.7),

$$\sigma=\sigma(x^4). \qquad (3.8)$$

Also from (2.4) and (3.8),

$$R_{ij}=0, \quad R_{4i}=0.$$

Thus (2.1), (2.2), (3.6), and (3.8) together, satisfy (3.2) except for $\mu=\nu=4$. Therefore (3.2) for all μ and ν are satisfied if along with (2.1), (2.2), (3.6), and inequality (I) we have

$$\sigma=\int\sqrt{R_{44}}\,dx^4, \qquad (3.9)$$

provided $R_{44}>0$.

This provides the complete set of solutions of the vacuum field equations of Sen–Dunn theory, with a metric of the form (1.1).

4. NULL ELECTROMAGNETIC FIELD IN EINSTEIN THEORY

It is of some interest to note that if on the other hand we have $R_{44}<0$, then (2.1), (2.2), and (3.6), together with inequality (I) provide the complete set of solutions of the Einstein–Maxwell equations for null electromagnetic fields for $g_{\mu\nu}$ of the form (1.1); in other words, given (2.1), (2.2), and (I) (which are the corollary of (1.1) and $\det|g_{\mu\nu}|<0$), the necessary and sufficient condition that $g_{\mu\nu}$ provide a solution for the coupled Einstein Maxwell equation for the null electromagnetic field is (3.6).

Proof: (3.6) *is necessary*: For null electromagnetic fields, the Einstein equations are[8]

$$R_{\mu\nu}=-k_\mu k_\nu, \qquad (4.1a)$$

$$k_\mu k^\mu=0. \qquad (4.1b)$$

Considering (4.1a) for $\mu=\nu=i$ and using (2.4) [note that (2.4) follows from (2.1)],

$$\frac{1}{2}\left(g^{4k}\frac{\partial g_{ik}}{\partial x^4}\right)^2+k_i^2=0.$$

Thus

$$k_i=0$$

and

$$g^{4k}\frac{\partial g_{ik}}{\partial x^4}=0.$$

which along with (2.1), (2.2), and (I) leads to (3.6).

(3.6) *is sufficient*: As before, (2.1), (2.2), (3.6), and (I) lead to

$$R_{ij}=0, \quad R_{4i}=0.$$

Thus (4.1a) is satisfied if along with (2.1), (2.2), (3.6), (I), and $R_{44}<0$ we have

$$k_\mu=\alpha_{,\mu}, \qquad (4.2a)$$

and

$$\alpha=\int\sqrt{-R_{44}}\,dx^4. \qquad (4.2b)$$

Also since σ depends only on x^4, k_μ has only one nonvanishing component, namely k_4. Thus, since $g^{44}=0$. (4.1b) is also satisfied.

Now, Maxwell's equations are

$$F^{\mu\nu}{}_{;\nu}=0, \quad {}^*F^{\mu\nu}{}_{;\nu}=0, \qquad (4.3)$$

where

$$F^{\mu\nu}=k^\mu p^\nu-k^\nu p^\mu, \quad {}^*F^{\mu\nu}=\tfrac{1}{2}\sqrt{-g}\,[\mu\nu\alpha\beta]F^{\alpha\beta},$$

$[\mu\nu\alpha\beta]=0$ when (μ,ν,α,β) are not all different,

$=1$ when (μ,ν,α,β) are all different and form an even permutation of (1,2,3,4),

$=-1$ when (μ,ν,α,β) are all different and form an odd permutation of (1,2,3,4),

and p^μ is a vector satisfying

$$p_\mu p^\mu=-1,^{9} \quad p_\mu k^\mu=0. \qquad (4.4)$$

A little calculation shows that given (2.1), (2.2), (3.6), and (I), and k_μ being defined by (4.2), if p_μ is chosen consistent with (4.4) and if p_μ is a function of x^4 alone, then (4.3) are satisfied. Thus (3.6) provides the sufficient condition on $g_{\mu\nu}$ as well.

5. BRANS-DICKE THEORY

If ϕ is the scalar field in Brans–Dicke theory, ω the Brans–Dicke constant, $\Box\phi\equiv\phi_\mu{}^{;\mu}$, then vacuum field equations in the Brans–Dicke theory are given by

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=-(\omega/\phi^2)(\phi_{,\mu}\phi_{,\nu}-\tfrac{1}{2}g_{\mu\nu}\phi_{,\alpha}\phi^{,\alpha})-(1/\phi)(\phi_{,\mu;\nu}-g_{\mu\nu}\Box\phi),$$

which can be rewritten as

$$R_{\mu\nu}=-(\omega/\phi^2)\phi_{,\mu}\phi_{,\nu}-(1/\phi)(\phi_{,\mu;\nu}+\tfrac{1}{2}g_{\mu\nu}\Box\phi). \qquad (5.1)$$

Setting

$$\bar g_{\mu\nu}\equiv\phi g_{\mu\nu}, \qquad (5.2a)$$

$$\bar R_{\mu\nu}\equiv \text{Ricci tensor formed out of } \bar g_{\mu\nu}, \qquad (5.2b)$$

and

$$\Phi=\log\phi, \qquad (5.2c)$$

Eq. (5.1) reduces to

$$\bar R_{\mu\nu}=-(\omega+\tfrac{3}{2})\Phi_{,\mu}\Phi_{,\nu}. \qquad (5.2d)$$

We note that (5.2) is quite similar to the Sen–Dunn field equation (3.2). We shall use this similarity later. Before that we prove that the scalar field ϕ (and hence Φ) is a function of x^4 alone. [For $\omega\neq-\frac{3}{2}$, application of the Bianchi identities on (5.1) directly gives $\Box\phi=0$.

This considerably simplifies the proof of ϕ being a function of x^4 alone. However, the proof given here applies to both $\omega \neq -\frac{3}{2}$ as well as to $\omega = -\frac{3}{2}$.]

From (2.1), (2.5a), and (5.1),

$$\frac{\partial \phi'^4}{\partial x^j} + \phi'^4\left(\frac{\omega}{\phi}\,\phi_{,j} + \frac{\phi}{2}\,g^{4i}\frac{\partial g_{ij}}{\partial x^4}\right) = 0.$$

Therefore, either

$$\phi'^4 = 0 \tag{5.3a}$$

or

$$(\log\phi'^4 + \omega\log\phi)_{,j} = (\phi/2)\chi_{,j}, \tag{5.3b}$$

where

$$\chi \equiv -\left[\left(g^{4i}\frac{\partial g_{i1}}{\partial x^4}\right)x^1 + \left(g^{4i}\frac{\partial g_{i2}}{\partial x^4}\right)x^2 + \left(g^{4i}\frac{\partial g_{i3}}{\partial x^4}\right)x^3\right]. \tag{5.3c}$$

However, we shall just now see that (5.3b) also leads to (5.3a).

From (5.3b) we see that, treating x^4 as a constant, the derivatives of $\log\phi'^4 + \omega\log\phi$ with respect to x^1, x^2, and x^3, are proportional to the derivatives of χ with respect to x^1, x^2, and x^3. For physical functions, this means that $\log\phi'^4 + \omega\log\phi$ and χ are functionally dependent when x^4 is treated as a constant, i.e.,

$$\log\phi'^4 - \omega\log\phi = \psi(\chi, x^4),$$

where ψ is some function; then from (5.3b),

$$\frac{\phi}{2} = \frac{\partial\psi(\chi, x^4)}{\partial\chi}.$$

Therefore, ϕ is also a function of χ and x^4, which owing to (2.1), (5.3c) leads to

$$\phi'^4 = 0. \tag{5.4}$$

Thus (5.4) is true either way.

From (2.1), (2.2), and (5.4),

$$\phi = \phi(x^1 + \lambda x^3, x^2 + \mu x^3, x^4). \tag{5.5}$$

From (5.5), $\phi_{,i}$'s are also functions of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4 only and $\phi_{,4}$ is given by

$$\phi_{,4} = \xi x^3 + \eta, \tag{5.6}$$

where ξ and η are functions of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4, ξ being given by

$$\xi = \phi_{,1}\lambda^1 + \phi_{,2}\mu^1, \quad \lambda' \equiv \frac{d\lambda}{dx^4}, \quad \mu' \equiv \frac{d\mu}{dx^4}, \tag{5.7}$$

and η is the derivative of ϕ with respect to x^4, treating $x^1 + \lambda x^3$, $x^2 + \mu x^3$ as constants. ($\phi_{,4}$ denotes the derivative of ϕ, with respect to x^4, treating x^1, x^2, and x^3 as constants.)

Using (2.1), (2.2), and (5.5), a little calculation shows that $\Box\phi$ depends only on $g_{\mu\nu}$ and $\phi_{,i}$ but not on $\phi_{,4}$. Thus $\Box\phi$ also depends only on $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4.

By using (2.1), (2.2), and (5.5), Eqs. (5.1) for $\mu = 4$, $\nu = i$ reduce to

$$\frac{\partial\phi_{,4}}{\partial x^i} + \phi_{,4}\theta_{,i} - U = 0, \tag{5.8}$$

where

$$\theta \equiv \log\phi + \tfrac{1}{2}\chi, \tag{5.9}$$

$$U \equiv \left(\phi_{,k}\begin{Bmatrix} k \\ 4i \end{Bmatrix} - g_{4i}\frac{\Box\phi}{2} - R_{4i}\phi\right).$$

χ is given by (5.3c).

Since ϕ, $\phi_{,k}$, and $\Box\phi$ are functions of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4 only, and g_{4i} and R_{4i} are functions of x^4 only, it follows that U is a function of x^4 only. Also from (2.1), (2.2), and (5.3c), a little calculation shows that χ is also a function of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4 only and therefore, from (5.9), θ is also a function of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4 only.

Putting (5.6) into (5.8) for $i = 1, 2$,

$$x^3\alpha_p + \beta_p = 0, \quad p = 1, 2, \tag{5.10}$$

where

$$\alpha_p = \frac{\partial\xi}{\partial x^p} + \xi\theta_{,p}, \tag{5.11a}$$

$$\beta_p = \frac{\partial\eta}{\partial x^p} + \eta\theta_{,p} - U. \tag{5.11b}$$

Since ξ and θ are functions of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4 only, it follows from (5.11) that α_p, β_p ($p = 1, 2$) are also functions of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4 only. Since x^3 obviously cannot be expressed as a function of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4, we get from (5.10)

$$\alpha_p = 0, \quad \beta_p = 0, \quad p = 1, 2. \tag{5.12}$$

From (5.12),

$$\frac{\partial(\xi e^\theta)}{\partial x^p} = 0, \quad p = 1, 2.$$

However, since ξe^θ depends on x^3 only through $x^1 + \lambda x^3$ and $x^2 + \mu x^3$, we must have

$$\frac{\partial(\xi e^\theta)}{\partial x^i} = 0, \quad i = 1, 2, 3$$

or

$$\xi = \zeta(x^4)e^{-\theta}, \quad \text{where } \zeta \text{ is some function.} \tag{5.13}$$

Considering (5.1) for $\mu = \nu = 4$, using (2.2), (5.5), (5.6), and (5.13) and proceeding as before, we have

$$Ax^{3^2} + Bx^3 + C = 0,$$

where A, B, and C are functions of $x^1 + \lambda x^3$, $x^2 + \mu x^3$, and x^4; specifically

$$A = -(\omega/\phi)\xi^2 - (\xi_{,1}\lambda^1 + \xi_{,2}\mu^1).$$

Again for similar reasons

$$A = 0,$$

i.e.,

$$(\omega/\phi)\xi^2 + (\xi_{,1}\lambda^1 + \xi_{,2}\mu^1) = 0. \tag{5.14}$$

Putting (5.13) into (5.14) and using (5.7), (5.9), and (5.13) either

$$\phi_{,1}\lambda^1 + \phi_{,2}\mu^1 = 0, \tag{5.15a}$$

$$\chi_{,1}\lambda^1 + \chi_{,2}\mu^1 = 0. \tag{5.15b}$$

However, we shall now see that either of (5.15) essentially leads to $\lambda'=0$, $\mu'=0$.

Case (I): (5.15a) is true: Using (2.1), (2.2), (5.4), (5.5), (5.15a), and (5.3c),

$$\phi^4{}_{;4}=0, \tag{A}$$

$$\chi_{,i}\phi^i=0, \tag{B}$$

$$g^{il}\phi_{,i;j}\chi_{,l}=0. \tag{C}$$

From (5.1), (5.4), and (A),

$$R_4^4=-\frac{\Box\phi}{2\phi}. \tag{D}$$

Considering $g^{ij}\chi_{,i}R_{jk}$, using (2.4), (5.1), (5.3c), (B), and (C),

$$g^{ij}\chi_{,i}\chi_{,j}=-\Box\phi/4\phi. \tag{E}$$

Using (2.5c), (5.3c), (D), and (E),

$$g_{ij}\frac{\partial g^{4i}}{\partial x^4}\frac{\partial g^{4j}}{\partial x^4}=0. \tag{F}$$

Using (2.1), (2.2), (F), and inequality (I),

$$\lambda'=0,\quad \mu'=0. \tag{G}$$

Case (II): (5.15b) is true: (5.15b) together with (5.3c), (2.1), and (2.2) lead to (F) and as before, (2.1), (2.2), and (F) together with inequality (I) lead to (G). Thus in any case we have

$$\lambda'=0,\quad \mu'=0, \tag{5.16}$$

i.e., λ and μ are constants.

From (2.1), (2.2), (2.4), (5.3c), and (5.16),

$$R_{ij}=0, \tag{5.17a}$$

$$R_{i4}=0, \tag{5.17b}$$

$$\phi^4{}_{;4}=0. \tag{5.17c}$$

Considering R_4^4 with the help of (2.1), (2.2), (5.4), (5.16), and (5.17),

$$\Box\phi=0. \tag{5.18}$$

Considering $g^{\mu\nu}R_{\mu\nu}$ from (3.2) and using (2.1), (3.4), (3.17), and (3.18)

$$g_{ij}\phi^{,i}\phi^{,j}=0. \tag{5.19}$$

Equations (2.1), (2.2), (5.19), and inequality (I) lead to

$$\phi=\phi(x^4). \tag{5.20}$$

From (5.2a) and (5.20), we note that if $g_{\mu\nu}$ satisfies (2.1), (2.2), (3.6), and (I), so does $\bar{g}_{\mu\nu}$, i.e.,

$$\bar{g}_{\mu\nu}=\bar{g}_{\mu\nu}(x^4),\quad \bar{g}^{44}=0,\quad \bar{g}_{i3}=\lambda\bar{g}_{i1}+\mu\bar{g}_{i2}, \tag{5.21}$$

where

$$\frac{d\lambda}{dx^4}=0,\quad \frac{d\mu}{dx^4}=0.$$

As before, (5.22) satisfies (5.2d) except for $\mu=\nu=4$. Thus following Sec. 3, the complete set of solutions of (5.2d) and hence of (5.1) are given by $\bar{g}_{\mu\nu}$ that satisfy (5.21) if for $\omega\neq-\frac{3}{2}$,

$$\frac{\bar{R}_{44}}{\omega+\frac{3}{2}}<0.$$

and

$$\phi=\exp\left[\int\left(\frac{-R_{44}}{\omega+\frac{3}{2}}\right)^{1/2}dx^4\right].$$

for $\omega=-\frac{3}{2}m$,

$$\bar{R}_{44}=0$$

and ϕ is an arbirary function of x^4.

Then $g_{\mu\nu}$ can be determined from $\bar{g}_{\mu\nu}$ and ϕ.

6. CONCLUSION

A. Solutions

Thus (2.1), (2.2), (3.6), and (I) together provide the complete set of solutions for $g_{\mu\nu}$ of the form (1.1) for:

(a) vacuum field equations for the Sen–Dunn theory if

$$R_{44}>0$$

and scalar field σ is given by

$$\sigma=\int\sqrt{R_{44}}dx^4;$$

(b) the null electromagnetic field in the Einstein theory if

$$R_{44}<0$$

and the propagation vector k_μ [defined in (4.1)] is given by

$$k_\mu=\alpha_{,\mu},\quad \alpha=\int\sqrt{-R_{44}}dx^4;$$

(c) the vacuum field equations in Brans–Dicke theory for $\omega\neq-\frac{3}{2}$ if

$$\frac{\bar{R}_{44}}{\omega+\frac{3}{2}}<0\quad [\bar{R}_{\mu\nu}\text{ being defined in (5.2)}]$$

and ϕ the scalar field is given by

$$\phi=\exp\left[\int\left(\frac{-R_{44}}{\omega+\frac{3}{2}}\right)^{1/2}dx^4\right];$$

(d) the vacuum field equations in the Brans–Dicke theory for $\omega=-\frac{3}{2}$ if

$$\bar{R}_{44}=0.$$

From the above it is obvious that a simple procedure for making R_{44} positive, zero, or negative as desired is needed for the present situation. This is given as follows.

From (2.1), (2.2), (3.6), and (I) it is obvious that by a suitable choice of coordinates, one can set $\lambda=0$, $\mu=0$, without any loss of generality and without any violation of the previous equations. $g_{\mu\nu}$ thus takes the form

$$g_{\mu\nu}=\begin{vmatrix} g_{11} & g_{12} & 0 & g_{14}\\ g_{12} & g_{22} & 0 & g_{24}\\ 0 & 0 & 0 & g_{34}\\ g_{14} & g_{24} & g_{34} & g_{44}\end{vmatrix}, \tag{6.1}$$

where as before, $g_{\mu\nu}$ are functions of x^4 alone. Then,

$$R_{44} = \frac{\partial^2 \log}{\partial x^{4^2}} - \frac{\partial \log g_{34}}{\partial x^4} \frac{\partial \log \alpha}{\partial x^4} - \frac{1}{4}\left(\frac{\partial g^{11}}{\partial x^4} \frac{\partial g_{11}}{\partial x^4} + 2\frac{\partial g^{12}}{\partial x^4}\frac{\partial g_{12}}{\partial x^4} + \frac{\partial g^{22}}{\partial x^4}\frac{\partial g_{22}}{\partial x^4} \right), \quad (6.2)$$

where α is given by (2.3).

From (6.1) and (2.3) we see that g^{11}, g^{12}, g^{22}, and α are completely specified by g_{11}, g_{12}, and g_{22}. Thus from (6.2) we note that, for preassigned g_{11}, g_{12}, g_{22}, one can easily make R_{44} as positive definite, zero, or negative definite as desired only by suitably assigning g_{34}.

Also from (6.2) we see that instead of solving the scalar field (or the propagation vector in the case of the null electromagnetic field) in terms of $g_{\mu\nu}$ as has been done previously, the field equations can be solved for a preassigned (but dependent on x^4 only) scalar field or propagation vector by suitably choosing g_{34}.

B. Covariant reformulation of results

The original intuitive notion of plane wave, as stated in Sec. 1 is based on a special choice of coordinates. Thus, the solutions obtained in this paper have been specified only in a special system of coordinates. However, it is interesting to note that the notion of plane wave as given in Sec. 1, as well as the solutions obtained, can be restated in a coordinate independent manner by using the language of group theory as can be seen through the two following statements.

Statement 1: A necessary and sufficient condition that $g_{\mu\nu}$ can be put in the form (1.1) is that space defined by $g_{\mu\nu}$ admits a three-parameter Abelian group of motion with isotropic hypersurface of transitivity. (By isotropic hypersurface, we mean a hypersurface σ = const, such that $\sigma_{,\mu}\sigma^{,\mu} = 0$.)

Necessary condition: (1.1) can be transformed into (2.1), and (2.1) admits a three-parameter Abelian group G_3 with generators, $\partial/\partial x^1$, $\partial/\partial x^2$, $\partial/\partial x^3$, and x^4 = const as an isotropic hypersurface of transitivity.

Sufficient condition: If a space admits a three-parameter Abelian group, then by proper choice of coordinates, its generators can be expressed as[10]

$$\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}.$$

Obviously x^4 = const gives a hypersurface of transitivity, which if isotropic, leads to $g^{44} = 0$.

Statement 2: A space that admits a three-parameter Abelian group G_3 with isotropic hypersurface of transitivity will satisfy $R_{\mu\nu} = \pm \sigma_{,\mu}\sigma_{,\nu}$ with appropriate choice of σ, if and only if the three Killing vectors ξ^μ, η^μ, ζ^μ corresponding to G_3 are such that

$$\zeta_\mu \zeta^\mu = 0, \quad \eta_\mu \zeta^\mu = 0, \quad \xi_\mu \zeta^\mu = 0 \quad (6.3)$$

and for any direction element dx^μ

$$R_{\mu\nu}\, dx^\mu\, dx^\nu > 0 \quad (\text{for } R_{\mu\nu} = \sigma_{,\mu}\sigma_{,\nu}),$$

$$R_{\mu\nu}\, dx^\mu\, dx^\nu < 0 \quad (\text{for } R_{\mu\nu} = -\sigma_{,\mu}\sigma_{,\nu}).$$

Proof: With the special choice of coordinates used in (2.1), ξ^μ, η^μ, ζ^μ are $\delta_1{}^\mu$, $\delta_2{}^\mu$, $\delta_3{}^\mu$ respectively. Thus (6.3) is equivalent to

$$g_{31} = 0 = g_{32} = g_{33}.$$

The metric thus takes the form of (6.1) and hence the result follows.

It may however be noted that although only a three-parameter Abelian group was postulated at the outset, the metric (6.1) actually admits a five-parameter intransitive group with an isotropic hypersurface of transitivity.

ACKNOWLEDGMENT

The author thanks the referee for suggesting the need for the covariant formulation of the problem.

[1] H. Bondi, F.A.E. Pirani, and I. Robinson, Proc. Roy. Soc. **251**, 519 (1959).
[2] D.W. Ebner, J. Math. Phys. **15**, 166 (1974).
[3] H. Takeno, "The Mathematical Theory of Plane Gravitational Waves in General Relativity," Scientific Report No. 1, Research Institute of Theoretical Physics, Hiroshima University (1961).
[4] C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
[5] D.K. Sen and K.A. Dunn, J. Math. Phys. **12**, 578 (1971).
[6] x^4 is some coordinate, not the time coordinate.
[7] $R_{\mu\nu} \equiv \partial^2 \log\sqrt{-g}/\partial x^\mu \partial x^\nu - \{^{\lambda}_{\mu\nu}\}\partial \log\sqrt{-g}/\partial x^\lambda - \partial/\partial x^\lambda\{^{\lambda}_{\mu\nu}\} + \{^{\lambda}_{\mu\alpha}\}\{^{\alpha}_{\nu\lambda}\}$.
[8] C.W. Misner and J.A. Wheeler, Ann. Phys. 2, 525 (1957).
[9] The Minkowski metric is of the form (+ − − −) here.
[10] L.P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1949), p. 249.

Closure in anisotropic cosmological models

Gregory J. Galloway

Department of Mathematics, University of California, La Jolla, California 92037
(Received 26 January 1976)

The closure result of Friedmann cosmology is briefly reviewed. A new closure result is presented for nonrotating, dust filled, but otherwise anisotropic, inhomogeneous cosmological models. The isotropy assumptions are replaced by much milder physical assumptions while the crucial Friedmann condition, $8\pi\kappa\rho-3h^2>0$, is replaced by the only slightly stronger $4\pi\kappa\rho-3h^2>0$.

I. INTRODUCTION

One of the many intriguing features of Friedmann cosmology is the prediction, under certain conditions, of the closure (finiteness) of the universe. The Friedmann models we have in mind are dust filled, with space—time topology $M^4=\mathbb{R}\times V^3$ in which the spatial sections $V_t^3: t=\text{const}$ (t representing proper time) are othogonal to the fluid flow (and hence the fluid is nonrotating). The Friedmann models are characterized by stringent symmetry conditions; all *local* observations made by an observer comoving with the cosmic fluid are isotropic. As a consequence, the spatial sections V_t^3 (corresponding to our spatial universe at time t) are locally isotropic (and, hence, locally homogeneous) spaces, i.e., are spaces of constant sectional curvature, and all physical parameters (e.g., the density of matter ρ and the Hubble expansion parameter h) are constant on each of these spatial sections.

Now if V^3 is complete and if on V^3

$$\tfrac{8}{3}\pi\kappa\rho-h^2>0 \tag{1}$$

holds (where κ is the gravitational constant and units have been chosen so that the speed of light $c=1$), then V^3 is a space of constant *positive* curvature and, in fact, is *covered* by the 3-sphere. It follows that V^3 is compact, i.e., the universe is "finite." It is worth pointing out that V^3 need not be open (infinite) if the quantity in (1) is zero or negative.[1]

In this paper we wish to consider in what way the above closure result relies on the strong symmetry assumptions of the Friedmann models. A new closure theorem is presented in which the isotropy assumptions of the Friedmann models (which, of course, are not precisely satisfied in our own universe) are replaced by much milder physical assumptions. The proof will require a modest strengthening of the crucial condition (1).

II. THE CLOSURE THEOREM

In the cosmological models now to be considered, the mass—energy content of the universe is again represented by an incoherent fluid or dust (no pressures) with the dust particles representing galaxies or clusters of galaxies. This representation seems reasonable for the present epoch of the universe. The energy—momentum tensor for a dust is given by

$$T_{ij}=\rho u_i u_j, \tag{2}$$

where ρ is the density of matter and $\mathbf{u}$ is the unit timelike tangent field to the world lines of the fluid. By the equations of motion, these world lines are geodesics. As in the Friedmann models we assume the topology of space—time to be of the form

$$M^4=\mathbb{R}\times V^3 \tag{3}$$

and again we make the simplication that the spatial sections V_t^3, parameterized by proper time, be orthogonal to the fluid flow.

Introducing coordinates x^α ($\alpha=1,2,3$) in V^3, the metric in comoving coordinates takes the form

$$ds^2=-dt^2+g_{\alpha\beta}(x,t)\,dx^\alpha\,dx^\beta, \tag{4}$$

where $g_{\alpha\beta}(x,t)$ is the induced metric on the section V_t^3 and $\mathbf{u}=\partial/\partial t$. No assumption of isotropy or homogeneity of the section V_t^3 with metric $g_{\alpha\beta}$ is imposed. Furthermore, there is no requirement that the matter density ρ be constant on any section, nor that expansion about any point be isotropic.

By letting

$$g_{\alpha\beta}(x,t)\,dx^\alpha\,dx^\beta=G^2(t)\,d\sigma^2, \tag{5}$$

where $d\sigma^2$ is a metric of constant sectional curvature on V^3 and the function $G(t)$ is constant on each spatial section we obtain the form of the metric of the Friedmann models.

Let $\mathbf{X}$ be a vector tangent to a section V_t^3 at point p. Extend $\mathbf{X}$ along the flow line through p by making it invariant under the flow generated by $\mathbf{u}=\partial/\partial t$,

$$[\mathbf{u},\mathbf{X}]=\nabla_{\mathbf{u}}\mathbf{X}-\nabla_{\mathbf{X}}\mathbf{u}=0. \tag{6}$$

Here [,] is the Lie bracket and ∇ is the connection associated with the space—time metric. This vector field may be interpreted as a position vector tracking a nearby fluid molecule from the fluid molecule at $x=p$.[2]

Let $\|\mathbf{X}\|=[g_{\alpha\beta}X^\alpha X^\beta]^{1/2}$ be the length of $\mathbf{X}$. Note that, by construction, the X^α are constant along the flow line. A positive time derivative, $(\partial/\partial t)\|\mathbf{X}\|>0$, indicates a recession of nearby fluid molecules in the direction of $\mathbf{X}$, and a negative second derivative, $(\partial^2/\partial t^2)\|\mathbf{X}\|<0$, indicates a deceleration of the recession in the direction of $\mathbf{X}$.

The overall expansion behavior of a small fluid element is determined by the expansion scalar $\theta=\operatorname{div}\mathbf{u}$.[3] In our case θ is given by

$$\theta=\frac{1}{\sqrt{\det(g_{\alpha\beta})}}\,\frac{\partial}{\partial t}\,\sqrt{\det(g_{\alpha\beta})}=-g^{\alpha\beta}b_{\alpha\beta}, \tag{7}$$

where $b_{\alpha\beta}=-\tfrac{1}{2}(\partial/\partial t)g_{\alpha\beta}$ is the α,βth component in the comoving coordinate system of the *second fundamental form* B defined by

$$B(\mathbf{X},\mathbf{Y}) = -\langle \nabla_{\mathbf{X}}\mathbf{u}, \mathbf{Y}\rangle$$

for vectors $\mathbf{X}, \mathbf{Y}$ tangent to the spatial section V_t^3. (Here $\langle\ ,\ \rangle$ denotes the space–time metric.) We are now in the position to state the following theorem.

Theorem: Consider the cosmological model characterized by Eqs. (2), (3), and (4) above. Suppose at each point p of some section $V_{t_0}^3$,

(i) there is recession in all directions, i.e.,

$$\frac{\partial}{\partial t}\Big|_{t_0} \|\mathbf{X}\| \geq 0$$

for all vector fields $\mathbf{X}$ defined along the flow line through p, satisfying Eq. (6) and perpendicular to $\mathbf{u}$,

(ii) the rate of recession is decreasing in all directions, i.e.,

$$\frac{\partial^2}{\partial t^2}\Big|_{t_0} \|\mathbf{X}\| \leq 0$$

for all vector fields $\mathbf{X}$ as in (i).

Then if V^3 is complete and

(iii) $\inf_{V_{t_0}^3} (\frac{4}{3}\pi\kappa\rho - h^2) = \lambda > 0,$

where $h = \frac{1}{3}\theta$ is the averaged Hubble parameter,[4] V^3 is compact and

$$\mathrm{diam}(V_{t_0}^3) \leq \pi\sqrt{2/3\lambda}.$$

A couple of remarks before giving the proof. First, we emphasize that conditions (i) and (ii), which seem satisfied at least at one point of the actual universe, do not demand that the recession or rate of recession be the same in all directions. Also, by introducing the *mean length* L,

$$\frac{\dot{L}}{L} = \frac{1}{3}\theta \quad \left(\cdot \equiv \frac{\partial}{\partial t}\right),$$

the Raychaudhuri equation[5] becomes for the particular cosmological model under consideration

$$3\frac{\ddot{L}}{L} = -\tfrac{4}{3}\pi\rho - \sigma^2 \leq 0 \tag{8}$$

(σ, the shear, is a scalar measure of the anisotropy of expansion). Thus, from strictly formal considerations, there must be an overall deceleration of expansion at each point of any spatial section. Finally, note that the matter density term in (iii) is half that appearing in (1) so that condition (iii) is slightly stronger than the analogous condition for the Friedmann model case.

The *proof* employs a theorem of Myers, which states that if M^n is a complete Riemannian manifold with Ricci tensor R_{ij} satisfying $\mathrm{Ric}(\mathbf{V},\mathbf{V}) \equiv R_{ij}V^iV^j \geq a > 0$ for all unit vectors $\mathbf{V} = V^i(\partial/\partial x^i)$, then M^n is compact and $\mathrm{diam}(M^n) \leq \pi\sqrt{n-1}/\sqrt{a}$.[6]

The Ricci tensor of space–time, R_{ij}, is related in comoving coordinates to the Ricci tensor of $V_{t_0}^3$, $P_{\alpha\beta}$, by the equation[7]

$$R_{\alpha\beta} = \frac{1}{2}\frac{\partial^2}{\partial t^2} g_{\alpha\beta} + \tfrac{1}{2}\theta\frac{\partial}{\partial t} g_{\alpha\beta} - 2b_\alpha^\gamma b_{\beta\gamma} + P_{\alpha\beta}, \tag{9}$$

where from Einstein's equations and Eq. (2),

$$R_{\alpha\beta} = 4\pi\kappa\rho g_{\alpha\beta}. \tag{10}$$

Let ξ be a unit vector at some point p of $V_{t_0}^3$ and extend ξ along the flow line through this point by making it invariant under the flow. In coordinates $\xi = \xi^\alpha(\partial/\partial x^\alpha)$, $(\partial/\partial t)\xi^\alpha = 0$. Then from (9) and (10) we find

$$\begin{aligned}\mathrm{Ric}_{V_{t_0}}(\xi,\xi) &\equiv P_{\alpha\beta}\xi^\alpha\xi^\beta \\ &= 4\pi\kappa\rho - \frac{1}{2}\frac{\partial^2}{\partial t^2}\|\xi\|^2 - \tfrac{1}{2}\theta\frac{\partial}{\partial t}\|\xi\|^2 + 2\|\nabla_\xi \mathbf{u}\|^2.\end{aligned} \tag{11}$$

Let ξ, $\mathbf{e}_2$, $\mathbf{e}_3$ be orthonormal vectors at p and extend $\mathbf{e}_2$ and $\mathbf{e}_3$ along the flow line through p in the usual way. A simple calculation shows that $B(\xi,\xi) = -(\partial/\partial t)\|\xi\|$ at p, so

$$\|\nabla_\xi \mathbf{u}\|^2 = \left(\frac{\partial}{\partial t}\|\xi\|\right)^2 + B^2(\xi,\mathbf{e}_2) + B^2(\xi,\mathbf{e}_3).$$

By substituting this equation into (11) and simplifying we are lead to

$$\begin{aligned}\mathrm{Ric}_{V_{t_0}}(\xi,\xi) = 4\pi\kappa\rho - \frac{\partial^2}{\partial t^2}\|\xi\| + \left(\frac{\partial}{\partial t}\|\xi\|\right)^2 - \theta\frac{\partial}{\partial t}\|\xi\| \\ + 2B^2(\xi,\mathbf{e}_2) + 2B^2(\xi,\mathbf{e}_3).\end{aligned} \tag{12}$$

Since, as is easily shown,

$$\theta = \frac{\partial}{\partial t}\|\xi\| + \frac{\partial}{\partial t}\|\mathbf{e}_2\| + \frac{\partial}{\partial t}\|\mathbf{e}_3\| \quad \text{at } p,$$

$$\begin{aligned}&\left(\frac{\partial}{\partial t}\|\xi\|\right)^2 - \theta\frac{\partial}{\partial t}\|\xi\| \\ &= -\left(\frac{\partial}{\partial t}\|\xi\|\,\frac{\partial}{\partial t}\|\mathbf{e}_2\| + \frac{\partial}{\partial t}\|\xi\|\,\frac{\partial}{\partial t}\|\mathbf{e}_3\|\right).\end{aligned} \tag{13}$$

Using assumption (i) and the Schwarz inequality one checks that

$$\begin{aligned}&\frac{\partial}{\partial t}\|\xi\|\frac{\partial}{\partial t}\|\mathbf{e}_2\| + \frac{\partial}{\partial t}\|\xi\|\frac{\partial}{\partial t}\|\mathbf{e}_3\| \\ &\leq \frac{\partial}{\partial t}\|\xi\|\frac{\partial}{\partial t}\|\mathbf{e}_2\| + \frac{\partial}{\partial t}\|\xi\|\frac{\partial}{\partial t}\|\mathbf{e}_3\| + \frac{\partial}{\partial t}\|\mathbf{e}_2\|\frac{\partial}{\partial t}\|\mathbf{e}_3\| \\ &\leq \tfrac{1}{3}\theta^2 = 3h^2.\end{aligned} \tag{14}$$

Furthermore, by assumption (ii) the inequality

$$\frac{\partial^2}{\partial t^2}\|\xi\| \leq 0 \tag{15}$$

holds. By combining (12)–(15) we obtain

$$\mathrm{Ric}_{V_{t_0}}(\xi,\xi) \geq 4\pi\kappa\rho - 3h^2 \geq 3\lambda.$$

Our result now follows as a consequence of Myers' theorem.

In the Friedmann case, with metric given by (4) and (5), $\|\xi\| = G(t)/G(t_0)$. One finds that

$$\left(\frac{\partial}{\partial t}\|\xi\|\right)^2 - \theta\frac{\partial}{\partial t}\|\xi\| = -2h^2$$

[compare with (13) and (14)],

$$\frac{\partial^2}{\partial t^2}\|\xi\| = -\tfrac{4}{3}\pi\kappa\rho$$

[compare with (15)], and $B(\xi, e_i) = 0$, $i = 1, 2$. Then (12) becomes

$$\mathrm{Ric}_{V_{t_0}}(\xi, \xi) = 2(\tfrac{8}{3}\pi\kappa\rho - h^2)$$

as expected.

If the fluid flow is not geodesic (as in the case of a perfect fluid with a nonzero spatial pressure gradient) but is still nonrotating, the metric can be expressed, in comoving coordinates, in the form

$$ds^2 = -\varphi^2 dt^2 + g_{\alpha\beta}dx^\alpha\, dx^\beta.$$

The Ricci quadratic form on the spatial section V_t^3 is then given by the equation

$$\begin{aligned}\mathrm{Ric}_{V_t}(\xi, \xi) = \mathrm{Ric}_M(\xi, \xi) - \mathbf{u}^2\|\xi\| + (\mathbf{u}\|\xi\|)^2 - \theta\mathbf{u}\|\xi\| \\ + 2B^2(\xi, e_2) + 2B^2(\xi, e_3) + (1/\varphi)\,\xi^2(\varphi) \\ - \langle\overline{\nabla}_\xi\, \xi, \nabla_{\mathbf{u}}\mathbf{u}\rangle, \qquad (16)\end{aligned}$$

where

$$\mathbf{u} = \frac{1}{\varphi}\,\frac{\partial}{\partial t},$$

$\overline{\nabla}$ is the connection associated with the induced metric $g_{\alpha\beta}$, and ξ has been extended arbitrarily to some neighborhood of V_t^3 and is invariant under the flow generated by $\partial/\partial t$. This equation differs from (12) only in the addition of the last two terms on the right-hand side.

To prove the compactness of V^3 in this case, it would suffice to show that there is a bound on the lengths of the minimal geodesics from some fixed point in V^3 to all other points of V^3. Along any such geodesic we may extend ξ so that $\overline{\nabla}_\xi\, \xi = 0$, making the last term on the right-hand side of (16) equal to zero. Then by applying Myers' theorem directly to each of these geodesics, we could conclude the compactness of V^3 under the same assumptions as before (with $\partial/\partial t$ replaced by $\mathbf{u}$) if we knew that the term $\xi^2(\varphi)$ were nonnegative. In fact, it need not be; thus the term $(1/\varphi)\xi^2(\varphi)$ in (16) appears to spoil the compactness result in this nongeodesic case.

We remark, in closing, that it would be desirable to obtain some generalization of our closure result which does not require the spatial sections to be orthogonal to the fluid flow (and, hence, the fluid to be nonrotating).

Note added in proof: Recently we obtained a very satisfactory generalization of the closure theorem to the nongeodesic case. By replacing condition (iii) by the more general

$$\inf_{\substack{V_t^3 \\ \|\xi\|=1}} (\mathrm{Ric}(\xi, \xi) - 3h^2) = \lambda > 0,$$

and by making the additional requirement that the 4-acceleration $\nabla_{\mathbf{u}}\mathbf{u}$ have an upper bound on V_t^3,

$$\sup_{V_t^3}\|\nabla_{\mathbf{u}}\mathbf{u}\| = \mu < \infty,$$

we arrive at a new diameter estimate for V_t^3 depending on μ and λ and, again, are able to conclude the compactness of V^3. The proof requires a generalization of Myers' theorem.

ACKNOWLEDGMENT

I wish to express my sincere thanks to Professor Theodore T. Frankel for many valuable discussions and, in particular, for focusing my attention on Myers' theorem.

[1]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 725.

[2]See, for example, M.P. Ryan, Jr. and L.C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, N.J., 1975), pp. 48, 49.

[3]G. F. R. Ellis, "Relativistic Cosmology" in *General Relativity and Cosmology*, edited by R.K. Sachs (Academic, New York, 1971), pp. 109–13.

[4]See Ref. 3, p. 113.

[5]See Ref. 3, p. 127.

[6]The proof of Myers' theorem shows that if along a minimal geodesic joining two points with unit tangent V the inequality $\mathrm{Ric}(V, V) \ge a > 0$ is satisfied, then the length of the geodesic must be less than or equal to $\pi\sqrt{n-1/a}$. For a proof of Myers' theorem see, for example, J. Milnor, *Morse Theory* (Princeton U.P., Princeton, N.J., 1963), pp. 104, 105.

[7]L. Landau and E. Lifschitz, *Classical Theory of Fields* (Addison-Wesley, Reading, Massachusetts, 1971), 3rd ed., p. 292.

Higher spin states in the stochastic mechanics of the Bopp–Haag spin model

Thad Dankel, Jr.

Department of Mathematical Sciences, University of North Carolina at Wilmington, Wilmington, North Carolina 28401

(Received 14 July 1976)

The author has previously shown that the dynamically natural random variable representing angular momentum in the stochastic mechanics of the Bopp–Haag spin model has the expectation values predicted by quantum mechanics for spin = 1/2. The result is generalized to all higher values of the spin.

1. INTRODUCTION

The history of attempts to reformulate quantum mechanics in terms of motions of classical bodies and fields at the microscopic level is a long one, extending back to the origins of quantum mechanics itself. One of the more recent of these attempts has been called stochastic mechanics.[1-3] This theory assumes universal Brownian motion and Newton's second law $\mathbf{F}=m\mathbf{a}$; it shows these assumptions to be mathematically equivalent to the Schrödinger equation for a broad class of physical systems.

One advantage that stochastic mechanics has over other "hidden variable" theories is that it has no need for ad hoc quantum forces.[4] This advantage stems from the definition of the stochastic mechanical acceleration $\mathbf{a}$, based on considerations of time-reversal symmetry and the theory of Markov processes. However, the main advantage of the stochastic mechanical viewpoint is that physical quantities are no longer operators, but random variables. To say this is not to disparage the massive body of understanding yielded by operator theory, but to point out that there are potentially important insights to be gained by alternatively treating quantum mechanical systems in terms of the quite extensive and highly developed mathematical theory of random processes.

One of the quantum mechanical systems which has been studied from the viewpoint of stochastic mechanics is the spin model of Bopp and Haag.[5] The present author has shown[3] that when the Bopp–Haag version of the freely spinning electron is cast into stochastic mechanical form, the random vector field $\mathbf{L}$ which is naturally interpreted as angular momentum has expectation values as required by quantum mechanics,

$$\langle \mathbf{L}\rangle = mh\mathbf{k}, \tag{1}$$

$$\langle \mathbf{L}^2\rangle = l(l+1)h^2, \tag{2}$$

where $\mathbf{k}$ is the unit vector along the z axis, h is Planck's constant divided by 2π, $l=\frac{1}{2}$, and $m=\pm\frac{1}{2}$, depending on whether one starts with a spin-up state or a spin-down state. The purpose of the present paper is to prove (1) and (2) for all allowable values of l and m: l nonnegative integral or half-integral, and $m=-l,-l+1,\ldots,l-1,l$. Thus, the compatibility of the stochastic mechanical and quantum mechanical treatments will be demonstrated for free particles of arbitrary spin.

2. PRELIMINARIES

In this section we review the definition of stochastic mechanical velocities, some facts about the rotation group, and the basic features of the Bopp–Haag spin model. Since we are concerned only with free systems, many formulas are simpler than the corresponding ones for motion in a force field. The terminology and notation are those of Ref. 3.

Let the configuration space of a physical system be a Riemannian manifold M. Let ψ be a wavefunction satisfying the free Schrödinger equation

$$hi\frac{\partial\psi}{\partial t} = -c\Delta\psi, \tag{3}$$

where Δ is the Laplace–Beltrami operator on M, and c is some constant. Put $\psi=\exp(R+iS)$, and define u and v by

$$u=2hdR, \tag{4}$$

$$v=2hdS. \tag{5}$$

Then v is called the current velocity, and is identified with the ordinary classical velocity of the system whose quantum version we are examining; u is called the osmotic velocity, and arises from the assumption of universal Brownian motion superimposed on the classical motion. The latter name derives from the theory of Brownian motion of particles suspended in a liquid,[2,6] in which u is the velocity required of a particle to offset osmotic effects. We shall see that both u and v contribute to the angular momentum—in particular, to its expectation values.

Of course, the basic result[1-3] of stochastic mechanics is that the kinematics of Markov processes and $\mathbf{F}=m\mathbf{a}$ (with $\mathbf{a}$ properly defined!) are equivalent to (3). However, we do not need this result here, so we do not discuss it further.

We will be concerned in particular with $M=\mathrm{SO}(3)$, the group of rotations in three-dimensional space. We parameterize SO(3) with the Euler angle coordinates θ,ϕ,χ, following the convention of Refs. 3 and 5. Any rotation T may be written in the form $T=T_\chi T_\theta T_\phi$, where T_ϕ is a rotation through angle ϕ about the z axis, T_θ is a rotation through angle θ about the line into which T_ϕ takes the x axis, and T_χ is a rotation through angle χ about the line into which T_θ takes the z axis. We denote unit vectors along the axes of these rotations by

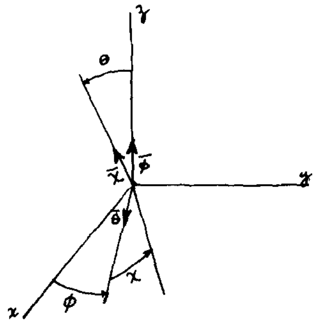

FIG. 1.

$\bar{\phi}, \bar{\theta}$, and $\bar{\chi}$, respectively. See Fig. 1. We shall also denote θ, ϕ, χ and $\bar{\theta}, \bar{\phi}, \bar{\chi}$ by $\theta^1, \theta^2, \theta^3$ and $\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3$, respectively.

If k is any positive constant, we may endow SO(3) with the Riemannian metric g_{ij}, where

$$[g_{ij}] = \begin{bmatrix} k & 0 & 0 \\ 0 & k & k\cos\theta \\ 0 & k\cos\theta & k \end{bmatrix}. \tag{6}$$

Then the inverse matrix $[g^{ij}]$ is

$$[g^{ij}] = \begin{bmatrix} 1/k & 0 & 0 \\ 0 & \dfrac{1}{k\sin^2\theta} & -\dfrac{\cos\theta}{k\sin^2\theta} \\ 0 & -\dfrac{\cos\theta}{k\sin^2\theta} & \dfrac{1}{k\sin^2\theta} \end{bmatrix}. \tag{7}$$

Later we shall put $k = 2\Theta$, where Θ is the moment of inertia of a spinning ball.

We shall require means of passing from the current and osmotic velocities [with values in cotangent spaces of SO(3)] to the angular momentum $\mathbf{L}$ (with values in R^3). We do this via the map U from any cotangent space of SO(3) to R^3 defined by

$$U(a_i\, d\theta^i) = g^{ji} a_i \bar{\theta}_j, \tag{8}$$

where the summation convention applies.

The Haar measure $d\Omega$ on SO(3) is given by[7] $d\Omega = \sin\theta\, d\theta\, d\phi\, d\chi$. All probability densities and integrations will be with respect to this measure. The integration $\int_{SO(3)} f\, d\Omega$, f a function on SO(3), means $\int_{\chi=0}^{2\pi}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} f(\theta, \phi, \chi)\sin\theta\, d\theta\, d\phi\, d\chi$, the limits being clear from Fig. 1. In Ref. 3, where $l = \frac{1}{2}$, the measure $d\theta\, d\phi\, d\chi$ was used. However, it is easy to see that in this simplest case the two choices of measure yield identical results for $\langle \mathbf{L}\rangle$ and $\langle \mathbf{L}^2\rangle$, and that the latter choice does not yield the correct (i.e., quantum mechanical) expectation values for $l > \frac{1}{2}$.

In their spin model, Bopp and Haag[5] consider a rigid, rigidly charged spherical ball. In the case of no external field, the Schrödinger equation is

$$hi\frac{\partial\psi}{\partial t} = \left(\frac{1}{2m}\mathbf{p}^2 + \frac{1}{2\Theta}\mathbf{M}^2\right)\psi, \tag{9}$$

where m is the mass of the ball, Θ its moment of inertia, and $\mathbf{p}$ and $\mathbf{M}$ the generalized linear and angular momentum operators, respectively. Separating variables, we are led to the eigenvalue problem

$$\mathbf{M}^2\psi = \lambda\psi. \tag{10}$$

The eigenvalue problem for $\mathbf{p}$ governs the motion of the center of mass of the ball, which doesn't concern us here. The solutions of (10) have been found by Bopp and Haag[5]: The eigenvalues are $\lambda = l(l+1)$, with $l = \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \ldots$, and for each such l, there are $(2l+1)^2$ eigenfunctions of the form

$$\psi^l_{\nu\mu} = \exp[i(\mu\phi + \nu\chi)]\, d^l_{\nu\mu}(\theta), \tag{11}$$

where $\mu, \nu = -l, -l+1, \ldots, l-1, l$, and the $d^l_{\nu\mu}(\theta)$ are closely related to the matrix elements of the lth representation of SU(2). The eigenvalue of the z component of angular momentum is μ, which plays the role of m in (1). Explicitly, we have[5,8]

$$d^l_{\nu\mu}(\theta) = [\cos(\theta/2)]^{\nu+\mu}[\sin(\theta/2)]^{\nu-\mu} P^{\nu-\mu,\nu+\mu}_{l-\nu}[\cos(\theta)],$$
$$\nu - \mu \geq 0, \quad \nu + \mu \geq 0, \tag{12}$$

where the P's are the standard Jacobi polynomials.[9] Now, $P^{\alpha\beta}_n$ is defined only for $\alpha, \beta \geq 0$; however, we use the following symmetry properties[8] of the $d^l_{\nu\mu}$ to yield a formula similar to (12) in other cases:

$$d^l_{\nu\mu}(\theta) = (-1)^{\nu-\mu} d^l_{-\nu-\mu}(\theta), \qquad \nu - \mu \leq 0, \quad \nu + \mu \leq 0, \tag{13}$$

$$d^l_{\nu\mu}(\theta) = (-1)^{\nu-\mu} d^l_{\mu\nu}(\theta), \qquad \nu - \mu \leq 0, \quad \nu + \mu \geq 0, \tag{14}$$

$$d^l_{\nu\mu}(\theta) = (-1)^{l-\nu} d^l_{-\mu\nu}(\theta + \pi), \quad \nu - \mu \geq 0, \quad \nu + \mu \leq 0. \tag{15}$$

We shall need the differential equation[9] satisfied by the $P^{\alpha\beta}_n$,

$$(1 - x^2)\frac{d^2}{dx^2}P^{\alpha\beta}_n(x) + [\beta - \alpha - (\alpha + \beta + 2)x]\frac{d}{dx}P^{\alpha\beta}_n(x)$$
$$+ n(n + \alpha + \beta + 1)P^{\alpha\beta}_n(x) = 0. \tag{16}$$

3. THE MAIN RESULT

Theorem: Consider the wavefunction $\psi^l_{\nu\mu}$ given by (11). Let $\rho = K|\psi^l_{\nu\mu}|^2$, where K is a normalization constant chosen so that $\int_{SO(3)}\rho\, d\Omega = 1$. Form the osmotic velocity u and the current velocity v via (4) and (5), and define the osmotic angular velocity $\boldsymbol{\omega}_u$ and the current angular velocity $\boldsymbol{\omega}_v$ by $\boldsymbol{\omega}_u = U(u)$ and $\boldsymbol{\omega}_v = U(v)$, where U is given by (8) and the constant k of (7) is taken to be $k = 2\Theta$, Θ being the moment of inertia of the spinning ball. Define the angular momentum $\mathbf{L}$ by

$$\mathbf{L} = \Theta(\boldsymbol{\omega}_u + \boldsymbol{\omega}_v). \tag{17}$$

Let $\langle\ \rangle$ denote expectation with respect to the probability measure $\rho\, d\Omega$. Then

$$\langle \mathbf{L}\rangle = \mu h\mathbf{k} \quad \text{and} \quad \langle \mathbf{L}^2\rangle = l(l+1)h^2,$$

where $\mathbf{k}$ is the unit vector along the z axis.

Proof: We first treat the case $\nu - \mu \geq 0$, $\nu + \mu \geq 0$, so that representation (12) holds. Using (4) and (5), we find

$$\boldsymbol{\omega}_u = (h/\Theta)[-\mu\csc(\theta) + \nu\cot(\theta) - \sin(\theta)(P'/P)]\bar{\theta}, \tag{18}$$

$P \equiv P^{\nu-\mu,\nu+\mu}_{l-\nu}[\cos(\theta)]$, and

$$\boldsymbol{\omega}_v = \frac{h}{\Theta}\left[\left(\frac{\mu - \nu\cos\theta}{\sin^2\theta}\right)\bar{\phi} + \left(\frac{\nu - \mu\cos\theta}{\sin^2\theta}\right)\bar{\chi}\right]. \tag{19}$$

Now, explicitly in terms of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, the standard basis vectors in R^3, we have

$$\bar{\theta}=\cos(\phi)\mathbf{i}+\sin(\phi)\mathbf{j}, \tag{20}$$

$$\bar{\phi}=\mathbf{k}, \tag{21}$$

$$\bar{\chi}=\sin(\theta)\sin(\phi)\mathbf{i}-\sin(\theta)\cos(\phi)\mathbf{j}+\cos(\theta)\mathbf{k}. \tag{22}$$

Performing the integration with respect to ϕ first, we find from (18) and (20) that

$$\langle\omega_u\rangle=0. \tag{23}$$

Also, (19), (21), and (22) yield

$$\langle\omega_v\rangle=\mu\hbar\mathbf{k}/\Theta. \tag{24}$$

By (23), (24), and (17), the first conclusion of the theorem holds.

For the second part, note that ω_u and ω_v are orthogonal, so that

$$\mathbf{L}^2=\Theta^2(\omega_u^2+\omega_v^2).$$

By a straightforward calculation with (18)–(22), this is

$$\mathbf{L}^2=\hbar^2\left[\frac{2\mu^2+\nu^2(1+\cos^2\theta)-4\mu\nu\cos\theta}{\sin^2\theta}+2(\mu-\nu\cos\theta)\frac{P'}{P}+\sin^2\theta\left(\frac{P'}{P}\right)^2\right]. \tag{25}$$

Now, the normalization constant $K=(4\pi^2/2^{2\nu})C$, where

$$C=\int_{-1}^{1}(1+x)^{\nu+\mu}(1-x)^{\nu-\mu}(P_{l-\nu}^{\nu-\mu,\nu+\mu}(x))^2dx, \tag{26}$$

as is easily seen from (11) and (12) by making the change of variable

$$x=\cos(\theta) \tag{27}$$

in the integral $\int_{SO(3)}|\psi_{\nu\mu}^l|^2d\Omega$. The exact value of C is known,[9] but we do not require it here. Hence,

$$\rho=2^{2\nu}|\psi_{\nu\mu}^l|^2/4\pi^2C. \tag{28}$$

In terms of the variable x, we find from (25) and (28) that

$$\begin{aligned}\langle\mathbf{L}^2\rangle=\frac{\hbar^2}{C}\bigg\{&\int_{x=-1}^{1}[2\mu^2+\nu^2(1+x^2)-4\mu\nu x](1+x)^{\nu+\mu-1}\\&\times(1-x)^{\nu-\mu-1}P^2dx+\int_{x=-1}^{1}2(\mu-\nu x)(1+x)^{\nu+\mu}(1-x)^{\nu-\mu}\\&\times PP'dx+\int_{x=-1}^{1}(1+x)^{\nu+\mu+1}(1-x)^{\nu-\mu+1}[P'(x)]^2dx\bigg\}\\\equiv{}&\frac{\hbar^2}{C}(I_1+I_2+I_3),\end{aligned} \tag{29}$$

where the prime denotes differentiation with respect to x, and $P\equiv P_{l-\nu}^{\nu-\mu,\nu+\mu}(x)$. Let $\alpha=\nu-\mu$, $\beta=\nu+\mu$, $n=l-\nu$. Then an integration by parts shows that

$$I_3=-\int_{-1}^{1}P(x)\frac{d}{dx}[(1+x)^{\beta+1}(1-x)^{\alpha+1}P'(x)]dx.$$

Carrying out the indicated differentiation and using (16), we find that this simplifies to

$$I_3=n(n+\alpha+\beta+1)C=(l-\nu)(l+\nu+1)C. \tag{30}$$

Integrating by parts in I_2, we find

$$I_2=-\int_{-1}^{1}P(x)\frac{d}{dx}\{[(\beta-\alpha)-(\beta+\alpha)x](1+x)^{\beta}\times(1-x)^{\alpha}P(x)\}dx.$$

Upon carrying out the differentiation under the integral sign, we see that I_2 appears again on the right-hand side with a minus sign. Solving for I_2 and simplifying yields

$$I_2=-\tfrac{1}{2}\int_{-1}^{1}P^2(x)(1+x)^{\beta-1}(1-x)^{\alpha-1}[-(\beta+\alpha)+(\beta-\alpha)^2-2(\beta^2-\alpha^2)x+(\beta+\alpha)(\beta+\alpha+1)x^2]dx.$$

Expressing I_1 in terms of α and β and adding it to this last equation, we find

$$I_1+I_2=\left(\frac{\beta+\alpha}{2}+\frac{(\beta+\alpha)^2}{4}\right)C=(\nu+\nu^2)C. \tag{31}$$

From (29), (30), and (31), the second part of the theorem follows. This concludes the proof when $\nu-\mu\geq 0$ and $\nu+\mu\geq 0$.

In the general case, first use (11), (12), and the appropriate one of (13)–(15) to express $\psi_{\nu\mu}^l$ in terms of a Jacobi polynomial P. Defining α, β, and n by $P\equiv P_n^{\alpha\beta}$, we find in each of the three cases (13)–(15) that the integral expressions for I_1, I_2, and I_3 and those of (29) above are identical when expressed in terms of α, β, and n [in the case (15), the formula[9] $P_n^{\alpha\beta}(-x)=(-1)^nP_n^{\alpha\beta}(x)$ and the change of variable $x\to -x$ are necessary]. Hence, as before,

$$I_1+I_2+I_3=\left(n(n+\alpha+\beta+1)+\frac{\alpha+\beta}{2}+\frac{(\alpha+\beta)^2}{4}\right)C.$$

This simplifies to the desired $\langle\mathbf{L}^2\rangle=l(l+1)\hbar^2$ in each case. The verification of the formula for $\langle\mathbf{L}\rangle$ is even easier: Since $\langle\omega_u\rangle=0$, as before, we see that $\langle\mathbf{L}\rangle$ depends only on ω_v, and therefore only on the factor $\exp[i(\mu\phi+\nu\chi)]$ in $\psi_{\nu\mu}^l$, which has the same form in all cases. The proof is complete.

[1]E. Nelson, Phys. Rev. **150**, 1079–85 (1966).

[2]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, 1967).

[3]T.G. Dankel, Jr., Arch. Ration. Mech. Anal. **37**, 192–221 (1970).

[4]See, e.g., D. Bohm, R. Schiller, and J. Tiomno, Nuovo Cimento Suppl. (Ser. 10) **1**, 48–66 (1955).

[5]F. Bopp and R. Haag, Z. Naturforsch. **5a**, 644–653 (1950).

[6]A. Einstein, *Investigations on the Theory of the Brownian Movement* (Dover, New York, 1956).

[7]I.M. Gel'fand and Z.Ya. Shapiro, *American Mathematical Society Translations, Series* 2 (American Mathematical Society, Providence, 1956), Vol. 2.

[8]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1960).

[9]See, e.g., M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

Group theoretic aspects of conservation laws of nonlinear dispersive waves: KdV type equations and nonlinear Schrödinger equations*

Sukeyuki Kumei

Department of Physics, University of the Pacific, Stockton, California 95204
(Received 5 January 1976; revised manuscript received 3 September 1976)

Group theoretic properties of nonlinear time evolution equations have been studied from the standpoint of a generalized Lie transformation. It has been found that with each constant of motion of the KdV type equation $f_{xxx}+a(f)f_x+f_t=0$ and of the coupled nonlinear Schrödinger equation $f_{xx}+a(f,g)+if_t=0$, $g_{xx}+a(g,f)-ig_t=0$ one invariance group of the equations is always associated. The well-known series of constants of motion of the KdV equation and the cubic Schrödinger equation will be recovered from the invariance groups of the equations. The doublet solution of the KdV equation will be characterized as the invariant solution of one of the groups. In a more general context, it will be shown that the well-known equation of quantum mechanics $(d/dt)\langle U\rangle=\langle[iH,U]+\partial U/\partial t\rangle$ can be generalized to a class of nonlinear time evolution equations and that if U is a generator of an invariance group of the equation then $(d/dt)\langle U\rangle=0$. The class includes equations such as the KdV, the cubic Schrödinger, and the Hirota equations.

INTRODUCTION

In this paper, we study group theoretic aspects of time evolution equations of nonlinear waves, particularly of the Korteweg–de Vries (KdV) equation $f_{xxx}+ff_x+f_t=0$ and of the cubic Schrödinger equation $f_{xx}+f^2f^*+if_t=0$.

Some time ago, Anderson, Kumei, and Wulfman proposed a generalization[1] of the Lie–Ovsjannikov[2–4] theory of invariance groups of differential equations, and applied it to a number of quantum mechanical systems to systematically study dynamical groups.[5] Recently it has been shown by Ibragimov and Anderson[6] that this generalized transformation is an infinite dimensional contact transformation.

It has been shown in the preceding paper[7] that the sine-Gordon equation $f_{xt}-\sin f=0$ admits an infinite number of one-parameter invariance groups of this new type, with each of which one can associate a series of conservation laws. Although the generalization appears to broaden the usefulness of group theoretic analysis of differential equations, particularly of nonlinear ones, the physical implications of the new type of symmetry are still unclear in many respects.

The aim of the present paper is to investigate some of the well studied equations of nonlinear waves[8] from the standpoint of the generalized theory, and to gain a clearer insight into the physical significance of the presence of the new kind of symmetry. It will be shown that some of the fundamental properties of the KdV and the cubic Schrödinger equations are the direct results of the existence of new groups.

In Sec. I, we briefly review a few basic ideas of infinitesimal invariance transformations to fix notations.

In Sec. II, we investigate group theoretic properties of the KdV equation and the related equations. The main results are: (1) With each constant of motion of the KdV type equation $f_{xxx}+a(f)f_x+f_t=0$, one invariance group is associated, hence the KdV equation admits an infinite number of invariance groups. (2) The doublet solution, as well as the singlet solution, of the KdV equation is the invariant solution (or generalized similarity solution) of one of the groups.

In Sec. III, we prove that with each conservation law of the coupled nonlinear Schrödinger equation $f_{xx}+a(f,g)+if_t=0$, $g_{xx}+a(g,f)-ig_t=0$ one can associate one invariance group. The constants of motion of the cubic Schrödinger equation due to Zakharov and Shabat[9] will be recovered from the invariance group of the equation.

In Sec. IV, we investigate some general properties of generators of invariance groups of time evolution equations $H(t,x^i,f,f_i,f_{ij},\cdots)+f_t=0$. It will be shown that (1) A generator U of an invariance group of $H+f_t=0$ always satisfies the relation $[\mathcal{H},U]+\partial U/\partial t=0$, where $\mathcal{H}$ is a Lie operator associated with H; (2) For a class of nonlinear time evolution equations, the equation $(d/dt)\langle U\rangle=\langle[\mathcal{H},U]+\partial U/\partial t\rangle$ can be generalized; in particular, if U is a generator, then $(d/dt)\langle U\rangle=0$.

I. INFINITESIMAL INVARIANCE TRANSFORMATIONS

We denote m-dimensional real and complex vector space by R^m and C^m, respectively and we consider the following infinite direct sum of the spaces; by denoting $C^{(N+1)^k}$ by $\underset{k}{C}$,

$$V=R^N\oplus\underset{0}{C}\oplus\underset{0}{C'}\oplus\underset{1}{C}\oplus\underset{1}{C'}\oplus\cdots\oplus\underset{k}{C}\oplus\underset{k}{C'}\oplus\cdots. \tag{1}$$

The prime is to distinguish two spaces of the same dimensions. We denote the elements of $\underset{k}{C}$ and $\underset{k}{C'}$ by $\underset{k}{u}$ and $\underset{k}{v}$, thus the elements of V are

$$z=(x,u,v,\underset{1}{u},\underset{1}{v},\ldots,\underset{k}{u},\underset{k}{v},\cdots),\quad x\in R^{N+1}. \tag{2}$$

The components of $\underset{k}{u}$, $\underset{k}{v}$ are written as $u_{p_1p_2\cdots p_k}$, $v_{p_1p_2\cdots p_k}$ where each index runs from 0 through N[10]:

$$u=(u),\quad v=(v),\quad \underset{1}{u}=(u_0,u_1,\ldots,u_N),\quad \underset{1}{v}=(v_0,v_1,\ldots,v_N),$$

$$\underset{2}{u}=(u_{00},u_{01},\ldots,u_{0N},\ldots,u_{N0},u_{N1},\ldots,u_{NN}),$$

$$\underset{2}{v}=(v_{00},v_{01},\ldots,v_{0N},\ldots,v_{N0},v_{N1},\ldots,v_{NN}),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots. \tag{3}$$

Now we consider an infinitesimal transformation in V

$$\bar{z}=z+\epsilon Z,\quad Z=(0,\eta,\zeta,\underset{1}{\eta},\underset{1}{\zeta},\ldots,\underset{k}{\eta},\underset{k}{\zeta},\cdots),\tag{4}$$

where

$$\eta=\eta(z;c),\quad \zeta=\zeta(z;c),\tag{5}$$

and the components of $\underset{k}{\eta}$ and $\underset{k}{\zeta}$ are to be determined by the formula

$$\eta_{p_1p_2\cdots p_k}=D_{p_1p_2\cdots p_k}\eta,\quad \zeta_{p_1p_2\cdots p_k}=D_{p_1p_2\cdots p_k}\zeta,\tag{6}$$

where $D_{ij\cdots m}=D_iD_j\cdots D_m$ with

$$D_i=\partial_{x^i}+(u_i\partial_u+v_i\partial_v)+(u_{ij}\partial_{u_j}+v_{ij}\partial_{v_j})+\cdots + (u_{ij\cdots m}\partial_{u_{j\cdots m}}+v_{ij\cdots m}\partial_{v_{j\cdots m}})+\cdots.\tag{7}$$

In this paper, the summation rule will be assumed for repeated indices. In (5), c denotes a collection of all the real and complex numbers appearing in the expression of η or ζ. We write (4) compactly in the usual way as

$$\bar{z}=(1+\epsilon U)z,\tag{8}$$

with

$$U=(\eta\partial_u+\zeta\partial_v)+(\eta_i\partial_{u_i}+\zeta_i\partial_{v_i})+\cdots+(\eta_{i\cdots k}\partial_{u_{i\cdots k}}+\zeta_{i\cdots k}\partial_{v_{i\cdots k}})+\cdots.\tag{9}$$

The operator U has the following property (see Appendix A for the proof):

Lemma 1: If a function $A(z)$ is twice differentiable with respect to all the variables, then $(D_iU-UD_i)A(z)=0$ for $i=0,1,\ldots,N$.

We consider a set of differential equations for functions $f(x)$ and $g(x)$,

$$F^i(z;c)=0,\quad i=1,2,\tag{10a}$$

$$u=f(x),\quad v=g(x),\quad \underset{k}{u}=\underset{k}{f}(x),\quad \underset{k}{v}=\underset{k}{g}(x),\quad k=1,2,\ldots,\infty,\tag{10b}$$

where $\underset{k}{f}(x)$ and $\underset{k}{g}(x)$ are functions of the $(N+1)^k$-tuple

$$\begin{aligned}&\underset{1}{f}(x)=(f_0,f_1,\ldots,f_N),\quad \underset{1}{g}(x)=(g_0,g_1,\ldots,g_N),\\&\underset{2}{f}(x)=(f_{00},\ldots,f_{NN}),\quad \underset{2}{g}(x)=(g_{00},\ldots,g_{NN}),\\&\cdots\cdots\cdots\cdots\cdots\cdots,\end{aligned}\tag{11}$$

with $f_{i\cdots j}=\partial_{x^i}\cdots\partial_{x^j}f(x)$, $g_{i\cdots j}=\partial_{x^i}\cdots\partial_{x^j}g(x)$. C in (10a) represents a set of parameters (real or complex) appearing in the differential equation. Each solution of Eq. (10) defines a manifold in V which we call a solution manifold.

It is well known[2–4] that a group transformation e^{aU} maps a solution manifold of (10) into another (or the same) solution manifold if and only if

$$UF^i(z;c)\big|_{\mathcal{J}}=0,\quad i=1,2,\tag{12}$$

where $(\cdot\cdot)\big|_{\mathcal{J}}$ indicates to evaluate the quantity under the conditions

$$F^i=0,\quad D_{p_1\cdots p_k}F^i=0,\quad i=1,2,\quad k=1,2,\ldots,\infty.\tag{13}$$

The operator U is then a generator of an invariance group.

We define #-conjugation of a quantity $A(z;c)=A(x,u,v,\ldots,\underset{k}{u},\underset{k}{v};c)$ by

$$A(z;c)^{\#}=A(x,v,u,\ldots,\underset{k}{v},\underset{k}{u};c^*),\tag{14}$$

where the asterisk represents a complex conjugation. An important subclass of Eq. (10) is

$$\begin{aligned}&F^i(z;c)=0\quad\text{with } F^2=(F^1)^{\#},\\&u=f(x),\quad v=f(x)^*,\cdots.\end{aligned}\tag{15}$$

For this equation, the generator U takes the form

$$U=\eta\partial_u+\eta^{\#}\partial_v+\eta_i\partial_u+\eta_i^{\#}\partial_v+\cdots+\eta_{i\cdots k}\partial_{u_{i\cdots k}}+(\eta_{i\cdots k})^{\#}\partial v_{i\cdots k}+\cdots.\tag{16}$$

In this paper, we consider the infinitesimal transformations of the type (4) which involves no transformation in x. This transformation, however, is not as special as it might look. Let us consider an infinitesimal transformation of a more general type[2–6]

$$\bar{z}=z+\epsilon Z,\quad Z=(\hat{\xi},\hat{\eta},\hat{\zeta},\underset{1}{\hat{\eta}},\underset{1}{\hat{\zeta}},\cdots),\quad \hat{\xi}=(\hat{\xi}^0,\hat{\xi}^1,\ldots,\hat{\xi}^N),\tag{17}$$

where

$$\begin{aligned}&\hat{\eta}_{p_1\cdots p_{k-1}p_k}=D_{p_k}\hat{\eta}_{p_1\cdots p_{k-1}}-u_{p_1\cdots p_{k-1}q}D_{p_k}\hat{\xi}^q,\\&\hat{\zeta}_{p_1\cdots p_{k-1}p_k}=D_{p_k}\hat{\zeta}_{p_1\cdots p_{k-1}}-v_{p_1\cdots p_{k-1}q}D_{p_k}\hat{\xi}^q.\end{aligned}\tag{18}$$

It can be proved[11] that if we know the transformations of type (4), then we can also obtain the more general type (17):

Lemma 2: If (4) is an infinitesimal invariance transformation of (10), then for an arbitrary choice of $\hat{\xi}$, $\hat{\eta}$, and $\hat{\zeta}$ subjected to the conditions $\hat{\eta}-\hat{\xi}^iu_i=\eta$, $\hat{\zeta}-\hat{\xi}^iv_i=\zeta$, the transformation (17) is also an invariance transformation of Eq. (10). Conversely, if (17) is an invariance transformation of (10), then so is (4) for $\eta=\hat{\eta}-\hat{\xi}^iu_i$, $\zeta=\hat{\zeta}-\hat{\xi}^iv_i$.

In the following sections, we write the operators (9) and (16) as

$$U=\eta\partial_u+\zeta\partial_v,\quad U=\eta\partial_u+\eta^{\#}\partial_v.\tag{19}$$

They, however, must be always interpreted as their infinite prolongation. Also, we use the following abbreviation:

$$[A(z)]_{\substack{u=f(x)\\v=g(x)}}=[A(u,v)]_{f,g}$$

and

$$\int[A(u,v)]_{f,g}\,dx=\int A(u,v)\,dx.$$

II. A GROUP THEORETIC ANALYSIS OF THE KdV EQUATION

The equation of our interest is $f_{111}+ff_1+f_0=0$.[12–16] The equation is a particular case of (10) for which $F^2=0$, $g=0$. In this section, we use t, x for x^0, x^1, and write coordinates such as u_0, u_{10}, $\cdots$ as u_t, u_{xt}, $\cdots$. Similarly, we write η_0, η_{10}, $\cdots$ as η_t, η_{xt}, $\cdots$. Thus, by definition $\eta_t = D_t\eta$, $\eta_{xt} = D_x D_t \eta$, etc. Also, because the equation involves a single real function, all the v's in the first section are to be ignored.

A. A Lie algebra of an invariance group of the KdV equation

We write the equation as

$$F = u_{xxx} + uu_x + u_t = 0,$$
$$u = f(x,t), \quad u_x = f_x(x,t), \quad u_t = f_t(x,t), \cdots. \tag{20}$$

We look for an operator $U = \eta\partial_u$ which satisfies condition (12) for this equation. We assume the transformation to be a generalized Lie type[1] with $\eta = \eta(x, t, u, u_x, u_{xx}, u_{xxx}, u_{xxxx}, u_{xxxxx})$. The absence in η of coordinates corresponding to t derivatives may be justified for time evolution type equations in which the only t derivative contained is f_t.

The application of Lie's algorithm[3,4] for finding generators leads to the following results:

$$\begin{aligned} &U^1 = (tu_x - 1)\,\partial_u, \\ &U^2 = \tfrac{1}{3}\{xu_x - 3t(u_{xxx} + uu_x) + 2u\}\,\partial_u, \\ &U^3 = u_x\partial_u, \quad U^4 = (u_{xxx} + uu_x)\,\partial_u, \\ &U^5 = (\tfrac{3}{5}u_{xxxxx} + uu_{xxx} + 2u_xu_{xx} + \tfrac{1}{2}u^2u_x)\,\partial_u. \end{aligned} \tag{21}$$

The generators form a nonsemisimple algebra (see Appendix B for the definition of a commutator)

$$\begin{aligned} &[U^1, U^2] = \tfrac{2}{3}U^1, \quad [U^1, U^3] = 0, \quad [U^1, U^4] = U^3, \\ &[U^1, U^5] = U^4, \quad [U^2, U^3] = \tfrac{1}{3}U^3, \quad [U^2, U^4] = U^4, \\ &[U^2, U^5] = \tfrac{5}{3}U^5, \quad [U^3, U^4] = 0, \\ &[U^3, U^5] = 0, \quad [U^4, U^5] = 0. \end{aligned} \tag{22}$$

By making use of Eq. (20), and by applying Lemma 2, one can cast the first four generators into "genuine" Lie generators: They are equivalent to

$$\begin{aligned} &\bar{U}^1 = -t\partial_x - \partial_u, \quad \bar{U}^2 = \tfrac{1}{3}(-x\partial_x - 3t\partial_t + 2u\partial_u), \\ &\bar{U}^3 = -\partial_x, \quad \bar{U}^4 = \partial_t. \end{aligned} \tag{23}$$

This set of generators is well known.[14,17] The generator U^5, however, is new and its properties will be analyzed later.

Let us consider operators $\partial U/\partial t = (\partial_t\eta)\,\partial_u$ and $\mathcal{H} = (u_{xxx} + uu_x)\,\partial_u = U^4$. It is remarkable that all the U^i of (21) satisfy the relation $[\mathcal{H}, U^i] + \partial U^i/\partial t = 0$. In Sec. IV, it will be shown that a generator of an invariance group of time evolution equations always satisfies such a relation.

It is well known[13] that the KdV equation admits an infinite number of conservation laws. To study a possible connection between the present groups and the conservation laws, we need to know effects of infinitesimal invariance transformations on constants of motion.

In his analysis of constants of motion of the time evolution equation $H(x, t, u, u_x, u_{xx}, \ldots, u^{(n)}) + u_t = 0$, $u^{(n)} = u_{\underbrace{xx\cdots x}_{n}}$, $u = f(x,t)$, Lax[15] considered an infinitesimal transformation of a solution $f(x,t)$ into a solution $u = f(x,t) + \epsilon\phi(x,t)$. The function ϕ must satisfy the linear equation

$$H_u(f)\,\phi + H_{u_x}(f)\,\phi_x + \cdots + H_{u^{(n)}}(f)\,\phi^{(n)} + \phi_t = 0, \tag{24}$$

where $H_u(f) = (\partial_u H)_{u=f}$, etc. and $\phi^{(n)} = (\partial_x)^n\phi(x,t)$. We note that $\eta(f)$ of a generator of an invariance group of the equation $H(f) + f_t = 0$ is a special realization of ϕ. The effect of the transformation on the constant of motion $I(f)$ is

$$\begin{aligned} &I(f + \epsilon\phi) = I(f) + \epsilon(\Gamma(f), \phi), \\ &(\Gamma(f), \phi) = \partial_\epsilon I(f + \epsilon\phi)\,|_{\epsilon=0}. \end{aligned} \tag{25}$$

The function $\Gamma(f)$ is a gradient of the functional $I(f)$.[15] For the constant of motion of integral type, i.e., $I(f) = \int \rho(f)\,dx$, the gradient has a simple expression: Assuming $\rho(u) = \rho(x, t, u, \ldots, u^{(k)})$,

$$\Gamma(u) = \rho_u - D_x\rho_{u_x} + D_x^2\rho_{u_{xx}} + \cdots + (-D_x)^k\rho_{u^{(k)}} \tag{26}$$

In this case, we have

$$(\Gamma(f), \phi) = \int \Gamma(f)\,\phi\,dx. \tag{27}$$

Lax observed

$$(\Gamma(f), \phi) \text{ is a constant of the motion.} \tag{28}$$

B. Constants of motion of $f_{xxx} + a(f)f_x + f_t = 0$ and its groups

Now we prove a theorem which establishes a relationship between a constant of motion of the KdV type equation and its invariance group. We consider an equation

$$f_{xxx} + a(f)f_x + f_t = 0, \tag{29}$$

where $a(f)$ is a function of f. We assume that an initial value problem for this equation is well posed for a periodic boundary condition $f(x,t) = f(x + x_0, t)$ or for a condition $f(-\infty, t) = f(\infty, t) = 0$. Let us suppose that the system has a constant of motion of integral type $I(f) = \int \rho(f)\,dx$. The limits of the integration are either over the period or from $-\infty$ to ∞. We prove:

Theorem 1. If $\Gamma(u)$ is the gradient of a constant of motion $I(f) = \int \rho(f)\,dx$ associated with the equation $f_{xxx} + a(f)f_x + f_t = 0$, then the operator $U = \eta\partial_u$ which has $\eta(u) = D_x\Gamma(u)$ is a generator of an invariance group of the equation.

Proof: It is sufficient if we prove $\{U(u_{xxx} + a(u)u_x + u_t)\}_f = 0$. We consider a transformation of a solution f to a solution $f + \epsilon\phi$. Then, by (24), $\phi_{xxx} + a(f)\phi_x + a_u(f)f_x\phi + \phi_t = 0$. Thus, $0 = \int \Gamma(f)(\phi_{xxx} + a\phi_x + a_uf_x\phi + \phi_t)\,dx$. Integrating this by parts and assuming null contribution from the boundary terms, we obtain $0 = \int \{-D_x^3\Gamma - D_x(a\Gamma) + \Gamma a_uu_x - D_t\Gamma\}_f\,\phi\,dx + (d/dt)\int \Gamma\,\phi\,dx$. The second term vanishes because of (28). Because we can prescribe an arbitrary admissible function for ϕ at initial time t_0, this equation implies $\{D_x^3\Gamma + D_x(a\Gamma) - \Gamma a_uu_x + D_t\Gamma\}_f = 0$. Differentiating this with respect to x, and defining $\eta = D_x\Gamma$, we find $\{D_x^3\eta + \eta a_uu_x + (D_x\eta)\,a + D_t\eta\}_f = \{U(u_{xxx} + au_x + u_t)\}_f = 0$.

This theorem establishes a relationship between constants of motion and invariance groups of Eq. (29),

$$I(f) = \int \rho(f)\,dx \longrightarrow \Gamma(u) \longleftrightarrow \{U = \eta\partial_u, \ \eta = D_x\Gamma\}. \tag{30}$$

The process from U to I involves an integration process and not all the generators are integrable to I. In Sec. IV, we provide another scheme to connect a group to a constant of motion which can supplement such a nonintegrable case.

The application of the theorem to the generators (21) leads to (within constant factors),

$$I^1 = \int (\tfrac{1}{2} tu^2 - xu)\, dx, \quad I^3 = \int \tfrac{1}{2} u^2\, dx,$$
$$I^4 = \int (\tfrac{1}{3} u^3 - u_x^2)\, dx, \tag{31}$$
$$I^5 = \int (\tfrac{1}{4}u^4 - 3uu_x^2 + \tfrac{9}{5} u_{xx}^2)\, dx.$$

The generator U^2 is not integrable. The constants (31) coincide with members of the set of constants of motion due to Miura, Gardner, and Kruskal.[13] The simplest constant $I = \int u\, dx$ is missing; the reason is that it gives $\Gamma = 1$, hence $U = 0$. In the last section, however, we show that one can associate this with the generator U^2. Thus, we write $I^2 = \int u\, dx$.

The fact that there exist an infinite number of constants of motion for the KdV equation means that the equation is invariant under an infinite number of groups; the situation is similar to the case of the sine-Gordon equation $f_{xt} - \sin f = 0$.[7]

Now, we study properties of the groups associated with constants of motion of the KdV equation. First we review a few important properties of the gradient found by Lax[15] and Gardner.[16]

C. Properties of gradients (Lax and Gardner)

Lax has proved that the gradients associated with the constants of motion of the KdV equation has the following unique properties:

(1) If $\Gamma^i(u)$ is a gradient of $I^i = \int \rho^i(f)\, dx$, $i, j \geq 2$, then $\Gamma^i(u) D_x \Gamma^j(u) = J^{ij}$ with J^{ij} = polynomial in $u, u_x, u_{xx}, \cdots$.

(2) Every solitary wave solution

$$u = 3c\, \mathrm{sech}^2 \tfrac{1}{2}\sqrt{c}\,(x - ct) \equiv s(x - ct) \tag{32}$$

is an eigenfunction of the gradients

$$\Gamma(s) = \gamma(c) s, \quad \gamma(c) = \text{eigenvalue}. \tag{33}$$

In the study of doublet solutions of the KdV equation, Lax, as well as Kruskal and Zabusky,[12] focused his attention on three constants I^3, I^4, and I^5. For these constants, the gradients are

$$\Gamma^3 = u, \quad \Gamma^4 = u^2 + 2u_{xx},$$
$$\Gamma^5 = u^3 + 3u_x^2 + 6uu_{xx} + \tfrac{18}{5} u_{xxxx}, \tag{34}$$

and correspondingly,

$$\Gamma^3(s) = s, \quad \Gamma^4(s) = 2cs, \quad \Gamma^5(s) = \tfrac{18}{5} c^2 s. \tag{35}$$

Another remarkable property of $\Gamma(u)$ of the KdV equation is due to Gardner,

(3) If we define an operator W^i associated with $\Gamma^i(u)$ of I^i, $i > 2$, by

$$W^i = (D_x \Gamma^i)\, \partial_u + (D_x^2 \Gamma^i)\, \partial_{u_x} + (D_x^3 \Gamma^i)\, \partial_{u_{xx}} + \cdots, \tag{36}$$

then $[W^i, W^j] = 0$.

D. Properties of U^i, $i > 2$

We note the similarity between the generator $U^i = (D_x \Gamma^i)\, \partial_u$ and Gardner's operator W^i. They, however, are different in that the prolonged U^i involves terms such as $(\cdot)\, \partial_{u_t}$, $(\cdot)\, \partial_{u_{xt}}$ whereas W^i does not. Nevertheless Gardner's result implies that two generators U^i and U^j associated with I^i and I^j commute,

$$[U^i, U^j] = 0, \quad i, j > 2. \tag{37}$$

This is obviously the reflection of the fact that the KdV equation is a completely integrable Hamiltonian system.[18,19]

Incidentally, it is often useful to note that: If $I(f) = \int \rho(f)\, dx$ is a constant of motion associated with the differential equation $F(x, t, f, f_x, f_t, f_{xx}, f_{xt}, f_{tt}, \cdots) = 0$, and if U is a generator of an invariance group of $F = 0$, then the quantity $I' = \int \{U\rho(u)\}_f\, dx$ is also a constant of motion of the same equation. The application of this scheme to the KdV equation, however, fails to generate a constant; indeed, by making use of Eq. (26), Lax's result (1), and Lemma 1, we find

$$\int U^i \rho^j\, dx = \int \sum_k (D_x^k \eta^i)\, \rho^j_{u(k)}\, dx = \int \sum_k \eta^i (-D_x)^k \rho^j_{u(k)}\, dx$$
$$= \int \eta^i \Gamma^j\, dx = \int (D_x \Gamma^i)\, \Gamma^j\, dx = \int D_x J^{ij}\, dx = 0. \tag{38}$$

Although the method fails to generate a string of constants of motion, it has been found that U^4 gives rise to the following recursive relation:

$$0 = \int U^4 \rho^i\, dx = c\, \frac{d}{dt}\, I^{i+1}. \tag{39}$$

This relation has been checked up to $i = 4$.

E. Properties of e^{aU^i}

If $u = f(x, t)$ is a solution of the KdV equation, then, by construction, a function $u = \bar{f}(x, t; a) = \{e^{aU^i} u\}_f$ is also a solution provided a series $\sum_{k=0}^{\infty} \{(a^k/k^*)(U^i)^k\}_f$ exists. First, we show that this group transformation does not alter the values of the constants of motion I^j,

$$\int \{\rho^j(u)\}_f\, dx = \int \{\rho^j(u)\}_{\bar{f}}\, dx, \quad j \geq 2. \tag{40}$$

Proof: First, by (38), $\int \{U^i \rho^j\}_f\, dx = 0$. This must hold at initial time for it is a constant of motion: $\int \{U^i \rho^j\}_{\mu(x)}\, dx = 0$ for any admissible initial condition $f(x, 0) = \mu(x)$. It can be proved that this is possible only if $U^i \rho^j = D_x h^{ij}(u)$, h^{ij} = polynomial in $u, u_x, u_{xx}, \cdots$. Then, by using Lemma 1, $(U^i)^k \rho^j = (U^i)^{k-1} D_x h^{ij} = D_x (U^i)^{k-1} h^{ij}$. Thus, $\int \{\rho^j\}_{\bar{f}}\, dx = \int \{\rho^j + D_x \sum_{k=1}^{\infty} (a^k/k!)(U^i)^{k-1} h^{ij}\}_f\, dx = \int \{\rho^j\}_f\, dx$.

This result reminds us of quantum mechanics where group operations e^{iaA}, e^{ibB} do not alter the values of observables $\langle A \rangle$ and $\langle B \rangle$ provided $[A, B] = 0$. Here operators U^i and observables I^i are related by (30) and in fact the U^i's commute by (37).

The relation (40) indicates that both solutions $f(x, t)$ and $\bar{f}(x, t; a)$ will break up into the same set of solitons. To prove this we start from Lax's result (2). We suppose Γ to be a linear combination of Γ^i associated with the constants of motion I^i of integral type. Differentiating Eq. (33) by x and using the relationship between Γ and U, we obtain

$$\{Uu\}_s = \gamma(c)s_x, \quad s = s(x-ct). \tag{41}$$

This and Lemma 1 give rise to

$$\{(U)^n u\}_s = (\gamma\partial_x)^n s = \{(\gamma D_x)^n u\}_s. \tag{42}$$

This relation implies that: For the solitary wave solution (32), we have the operator identity $U = \gamma D_x$. Consequently, the group operation e^{aU} has the effect of translation in x when it is operated on the solitary wave solution,

$$\{e^{aU}u\}_{s(x-ct)} = s(x - ct + a\gamma(c)). \tag{43}$$

Now let us assume that the solution $f(x,t)$ splits into N well separated solitons as $t\to\infty$,

$$f(x,t) \sim \sum_{i=1}^{N} s_i(x - c_i t + \delta_i) \quad \text{as } t\to\infty. \tag{44}$$

For such a wave profile, interactions between solitons are small, hence at least for small a we may assume

$$\{e^{aU}u\}_{f(x,t)} \sim \sum_{i=1}^{N} \{e^{aU}u\}_{s_i(x-c_it+\delta_i)} \quad \text{as } t\to\infty. \tag{45}$$

In view of (43), we can write this as

$$\{e^{aU}u\}_{f(x,t)} \sim \sum_{i=1}^{N} s_i(x - c_i t + \delta_i + a\gamma(c_i)) \quad \text{as } t\to\infty. \tag{46}$$

Thus, two solutions $f(x,t)$ and $\bar f(x,t;a) = \{e^{aU}u\}_{f(x,t)}$ of the KdV equation have the same asymptotic profile as $t\to\infty$ except that the phase of each soliton is shifted by the amount $a\gamma(c_i)$.

F. Invariant solutions of the KdV equation

One curious question would be whether there exists a solution which is mapped onto itself under the transformation e^{aU}. Speaking in a more general context, a solution, of a differential equation $F=0$, which is mapped onto itself by the invariance group of the equation is called an invariant solution (or generalized similarity solution).[4] The necessary and sufficient condition for f to be the invariant solution of e^{aU} is obviously

$$\{Uu\}_f = 0. \tag{47}$$

One of the best known invariant solutions will be the Green's function of the heat equation $f_{xx} - f_t = 0$, $f = (4\pi t)^{-1/2}\exp(-x^2/4t)$. Here the group involved is the dilation group generated by $U = (xu_x + 2tu_t + u)\,\partial_u$ (or equivalently $U' = -x\partial_x - 2t\partial_t + u\partial_u$).

It is well known that the singlet solution of the KdV equation (32) is the invariant solution for $U = U^4 - c^{-1}U^3$ $(=\partial_t + c^{-1}\partial_x)$. The simplest generalization of this is to consider a group generated by $U = U^5 + pU^4 + qU^3$, p, q constants. Then the condition (47) yields

$$\tfrac{3}{5} f_{xxxxx} + ff_{xxx} + 2f_x f_{xx} + \tfrac{1}{2}f^2 f_x + p(f_{xxx} + ff_x) + qf_x = 0.$$

An integration of this equation with respect to x, assuming $f(\pm\infty, t) = 0$, leads to the fourth order equation obtained by Kruskal and Zabusky,[12] and Lax.[15] The nature of the solution was carefully studied by Lax, and the solution was shown to be the doublet solution. From a group theoretic viewpoint, therefore, *the doublet solution of the KdV equation is the invariant solution of the group* $e^{a(U^5+pU^4+qU^3)}$.

The idea here is precisely parallel to Lax's; Lax uses a condition $\Gamma(f) = 0$ to characterize the doublet solution whereas we use $\{Uu\}_f = 0$; but they are related by (30).

III. INVARIANCE GROUPS AND CONSERVATION LAWS OF NONLINEAR SCHRODINGER EQUATIONS

The cubic Schrödinger equation $-if_{xx} - if^2 f^* + f_t = 0$ is another well studied nonlinear equation. It is known to share many common properties with the KdV equation.[9,18,19] In this section we study group theoretic aspects of conservation laws associated with a class of nonlinear Schrödinger equations.

A. Conservation laws of nonlinear Schrödinger equations

We consider a coupled nonlinear Schrödinger equation

$$\begin{aligned} &u_{xx} + a(u,v;c) + iu_t = 0, \quad v_{xx} + a(u,v;c)^{\#} - iv_t = 0, \\ &u = f(x,t), \quad u_x = f_x(x,t), \quad u_t = f_t(x,t), \cdots, \\ &v = g(x,t), \quad v_x = g_x(x,t), \quad v_t = g_t(x,t), \cdots, \end{aligned} \tag{48}$$

where a function a is subject to the condition

$$a_u(f,g;c) = [a_u(f,g;c)]^{\#}, \quad a_u = \partial_u a. \tag{49}$$

[See (14) for the notation #.] Condition (50) amounts to requiring that the equation can be written as a Hamiltonian system,

$$\frac{\delta \hat H}{\delta g} = -if_t, \quad \frac{\delta \hat H}{\delta f} = ig_t, \tag{50}$$

where $\delta\hat H/\delta g$ and $\delta\hat H/\delta f$ are Frechet derivatives of $\hat H = \int E(f,g)\,dx$, E = energy density. Equation (48) reduces to the cubic Schrödinger equation for the special case of $a = u^2 v$ and $g = f^*$.

We assume that an initial value problem is well posed either for a periodic condition $f(x,t) = f(x+x_0,t)$, $g(x,t) = g(x+x_0,t)$ or for a boundary condition $f(\pm\infty,t) = 0$, $g(\pm\infty,t) = 0$. Let us suppose that the system described by (48) has a constant of motion $I(f,g) = \int \rho(f,g)\,dx$ where the integration is over the period or from $-\infty$ to $+\infty$. The following theorem establishes the relationship between the I and an invariance group of the equation. In the following, quantities $\delta I/\delta u$ and $\delta I/\delta v$ represent $\{\delta I/\delta f\}_{f=u,g=v}$ and $\{\delta I/\delta g\}_{f=u,g=v}$.

Theorem 2: If $\delta I/\delta f$ and $\delta I/\delta g$ are Frechet derivatives of a constant of motion $I(f,g) = \int \rho(f,g)\,dx$ associated with Eq. (48), then the operator $U = i(\delta I/\delta v)\,\partial_u - i(\delta I/\delta u)\,\partial_v$ is a generator of an invariance group of the equation.

Proof: We consider infinitesimal transformations of solutions f, g into solutions $f + \epsilon\phi$, $g + \epsilon\psi$. ϕ and ψ must satisfy the equations $A = \phi_{xx} + a_u(f,g;c)\phi + a_v(f,g;c)\psi + i\phi_t = 0$, $B = \psi_{xx} + a_u(g,f;c^*)\psi + a_v(g,f;c^*)\phi - i\psi_t = 0$. The effect of this transformation on I can be found easily; by integration by parts, we arrive at

$$I(f+\epsilon\phi, g+\epsilon\psi) = I(f,g) + \epsilon\int\left(\frac{\delta I}{\delta f}\phi + \frac{\delta I}{\delta g}\psi\right)dx$$
$$\equiv I(f,g) + \epsilon\delta I.$$

Thus, $d/dt\,\delta I=0$. Next obviously,

$$\int\left(i\frac{\delta I}{\delta f}A-i\frac{\delta I}{\delta g}B\right)dx=0.$$

On integrating by parts this yields $0=\int(P\phi+Q\psi)\,dx+(d/dt)\,\delta I$ where

$$P=-\{U[v_{xx}+a(u,v;c)^{\#}-iv_t]\}_{f,g}+i[a_u(f,g;c)-a_u(g,f;c^*)]\frac{\delta I}{\delta f},$$

$$Q=-\{U[u_{xx}+a(u,v;c)+iu_t]\}_{f,g}+i[a_u(f,g;c)-a_u(g,f;c^*)]\frac{\delta I}{\delta g}.$$

Because $(d/dt)\,\delta I=0$, we obtain

$$\int(P\phi+Q\psi)\,dx=0. \tag{*}$$

One can prescribe arbitrary admissible functions for ϕ and ψ at an initial time. Thus, the Eq. (*) implies that P and Q are identically zero. Furthermore, the second terms of P and Q are zero because of condition (49), hence, $P=0$ and $Q=0$ yield the equations to be proved.

This theorem enables us to find constants of motion if we know the invariance groups of Eq. (48); the process involves a straightforward integration process $(\delta I/\delta f,\ \delta I/\delta g)\to I$. However, we note that there may be a generator which is not integrable to a constant of motion. This theorem can be extended to a general Hamiltonian system.[20]

B. Invariance groups of the cubic Schrödinger equation and its conservation laws

We look for the operator of the form (16) which satisfies the invariance condition (12) for $F^1=f_{xx}+f^2f^*+if_t=0$ and $F^2=(F^1)^{\#}=0$. Assuming the transformation to be the generalized type with $\eta=\eta(x,t,u,v,u_x,v_x,\ldots,u_{xxxxx},v_{xxxxx})$, and carrying out Lie's algorithm, we arrive at the following eight generators [writing only the first term of (16)]:

$$\begin{aligned}
&U_1=(-\tfrac{1}{2}ixu+tu_x)\,\partial_u,\\
&U_2=(itu_{xx}+itu^2v+\tfrac{1}{2}xu_x+\tfrac{1}{2}u)\,\partial_u,\quad U_3=iu\partial_u,\\
&U_4=u_x\partial_u,\quad U_5=i(u_{xx}+u^2v)\,\partial_u,\\
&U_6=(u_{xxx}+3uvu_x)\,\partial_u,\\
&U_7=i(u_{xxxx}+u^2v_{xx}+4uvu_{xx}+2uu_xv_x+3vu_x^2+\tfrac{3}{2}u^3v^2)\,\partial_u,\\
&U_8=[u_{xxxxx}+5(uvu_{xxx}+uu_xv_{xx}+2vu_xu_{xx}+uv_xu_{xx}+u_x{}^2v_x)+\tfrac{15}{2}u^2v^2u_x]\,\partial_u.
\end{aligned}\tag{51}$$

The first five generators can be cast into "genuine" Lie type operators by Lemma 2:

$$\bar U^1=-t\partial_x-\tfrac{1}{2}ixu\partial_u,\quad \bar U^2=-x\partial_x-2t\partial_t+u\partial_u,$$

$$\bar U^3=iu\partial_u,\quad \bar U^4=i\partial_x,\quad \bar U^5=-\partial_t.$$

The effects of the group transformation e^{aU^i}, a real, on a solution $f(x,t)$ can be found easily for $i<6$,

$$\begin{aligned}
&f^1=\exp[-i(ax+a^2t/2)/2]f(x+at,t),\\
&f^2=af(ax,a^2t),\quad f^3=\exp(ia)f(x,t),\\
&f^4=f(x+a,t),\quad f^5=f(x,t+a).
\end{aligned}\tag{52}$$

The remaining three generators are of the generalized type, and there exists, at present, no analytic method of finding corresponding global transformations.

The constants of motion associated with the generators (51) can be found by the simple integration process; they are $I^i=\int\rho^i\,dx$ where

$$\begin{aligned}
&\rho^1=xuv-it(u_xv-uv_x),\quad \rho^3=uv,\quad \rho^4=iu_xv,\\
&\rho^5=\tfrac{1}{2}(u_xv_x-\tfrac{1}{2}u^2v^2),\quad \rho^6=i(u_{xxx}v+\tfrac{3}{2}uu_xv^2),\\
&\rho^7=u_{xx}v_{xx}+\tfrac{1}{2}u^3v^3-2(u_xv+uv_x)^2-3u_xv_xuv,\\
&\rho^8=u_{xxxxx}v+5(uu_{xxx}v+uu_xv_{xx}+2u_xu_{xx}v+uu_{xx}v_x+u_x^2v_x)+\tfrac{1}{3}u^2u_xv^3.
\end{aligned}\tag{53}$$

The operator U^2 is not integrable. These constants of motion, except the first one, agree with the ones obtained by Zakharov and Shabat.[9] The phase shift operator U^3, the x translation operator U^4, and the t translation operator U^5 have given rise to the probability density ρ^3, the momentum density ρ^4, and the energy density ρ^5. The first constant I^1 also has a simple meaning if we consider the cubic Schrödinger equation as the Schrödinger equation for a particle with negative mass $-\frac{1}{2}$: The I^1 represents the initial position of the particle, $\langle x_0\rangle=\langle x-tV\rangle=\int f^*\cdot\left(x-t\frac{p}{m}\right)f\,dx=I^1$, $V=$velocity.

Let us define the Lie Hamiltonian by

$$\mathcal{H}=\left(i\frac{\delta\hat H}{\delta v}\right)\partial_u-\left(i\frac{\delta\hat H}{\delta u}\right)\partial_v=U^5,\quad \hat H=\text{energy}=I^5. \tag{54}$$

Then, we find that the operator U^i of (51) satisfies the relation $[\mathcal{H},U^i]+\partial U^i/\partial t=0$ with

$$\frac{\partial U}{\partial t}=i\left(\partial_t\frac{\delta I^i}{\delta v}\right)\partial_u-i\left(\partial_t\frac{\delta I^i}{\delta_u}\right)\partial_v.$$

We note that the second generator U^2 which is not related to a constant of motion also satisfies the relation. A general analysis of this property of the generators will be given in the next section. Some of the other commutation relations among U^i are $[U^i,U^j]=0$ for $3\le i,j\le 8$.

IV. GENERAL PROPERTIES OF GENERATORS OF INVARIANCE GROUPS OF TIME EVOLUTION EQUATIONS

Let us assume that Eq. (15) is a time evolution type: $x^0=$time coordinate,

$$\begin{aligned}
&F^1(z;c)=H(z;c)+u_0=0,\\
&F^2(z;c)=[H(z;c)]^{\#}+v_0=0.
\end{aligned}\tag{55}$$

To carry out a consistent analysis, we must take into account the relation (13),

$$D_{p_1p_2\cdots p_k}(H+u_0)=0,\quad k=1,2,\ldots,\infty. \tag{56}$$

We define two operators associated with H^{21} and U by

$$\mathcal{H}=H\partial_u+H^{\#}\partial_v, \tag{57}$$

$$\frac{\partial U}{\partial x^0}=(\partial_{x^0}\eta)\,\partial_u+(\partial_{x^0}\eta)^{\#}\partial_v. \tag{58}$$

As was mentioned in the first section, they must be interpreted as their infinite prolongation.

By the definition of a time evolution equation, H is not a function of the coordinates corresponding to x^0-derivatives such as u_{01}, v_{120}. In such a case, we can always express any coordinate of x^0-derivatives in terms of other coordinates by making use of the relations (55) and (56). Thus, we assume, without a loss of generality, that η is free of these coordinates.

A key in the present analysis is to write Eq. (55) as

$$(\mathcal{H}+D_0)u=0. \tag{59}$$

We first prove:

Lemma 3: If U is a generator of an invariance group of the equation $H+u_0=0$, then under condition (56) we have $[U,\mathcal{H}]+\partial U/\partial x^0=0$.

Proof: We have $[U,\mathcal{H}]+\partial U/\partial x^0=a\partial_u+a^{\#}\partial_v$ with $a=UH-\mathcal{H}\eta+\partial_{x^0}\eta$. It is sufficient if we prove that a vanishes under (56). Indeed, $0=U(H+u_0)=UH+D_0\eta=UH+\partial_{x^0}\eta+u_0\eta_u+v_0\eta_v+\cdots=UH+\partial_{x^0}\eta-H\partial_u-H^{\#}\partial_v-\cdots=UH+\partial_{x^0}-\mathcal{H}\eta$.

Now, we define the following quantity:

$$\langle U\rangle=\mathrm{Re}\int (vUu)_{\substack{u=f(x)\\ v=f^*(x)}}\, dx^1\,dx^2\cdots dx^N,\quad \mathrm{Re}=\text{real part}, \tag{60}$$

where the integration should be taken over the whole space of interest. Obviously $\langle U\rangle$ is a function of x^0 only. The following lemma describes how it develops in time for a class of nonlinear systems:

Lemma 4: If H of the equation $H+u_0=0$ satisfies the equation

$$\begin{aligned}&H^{\#}+vH_u+u(H_v)^{\#}-D_i[vH_{u_i}+u(H_{v_i})^{\#}]+\cdots\\ &\quad+(-1)^r D_{p_1\cdots p_r}[vH_{u_{p_1\cdots p_r}}+u(H_{v_{p_1\cdots p_r}})^{\#}]=0\end{aligned} \tag{61}$$

and if all the boundary integrals

$$\int_S [u\eta_{j\cdots k}H_{u_{ij\cdots k}}]_{\substack{u=f(x)\\ v=f^*(x)}}\ \nu^i\, d\Omega$$

and

$$\int_S [v\eta^{\#}_{j\cdots k}H_{u_{ij\cdots k}}]_{\substack{u=f(x)\\ v=f^*(x)}}\ \nu^i\, d\Omega$$

vanish for $\nu=(\nu^1,\ldots,\nu^N)=$ normal vector on the boundary surface, then

$$\frac{d}{dx^0}\langle U\rangle=\left\langle [U,\mathcal{H}]+\frac{\partial U}{\partial x^0}\right\rangle . \tag{62}$$

Proof: For brevity, we write (60) as $\langle U\rangle=\mathrm{Re}\int vUu\,d\mathbf{x}$. Then, we have $d/dx^0\langle U\rangle=\mathrm{Re}\int(v_0Uu+vD_0Uu)\,d\mathbf{x}=\mathrm{Re}\int[-H^{\#}\eta+v(\partial U/\partial x^0-\mathcal{H}U)u]\,d\mathbf{x}$. Here, we have used the relations (55) and (56). On the other hand, we have $\langle U\mathcal{H}\rangle=\mathrm{Re}\int vU\mathcal{H}u\,d\mathbf{x}=\mathrm{Re}\int v\eta_{i\cdots j}H_{u_{i\cdots j}}\,d\mathbf{x}$. Applying Green's theorem repeatedly, and using the hypotheses, we find $\langle U\mathcal{H}\rangle=\mathrm{Re}\int(-H^{\#}\eta)\,d\mathbf{x}$. Putting these two together, we obtain $(d/dx^0)\langle U\rangle=\langle[U,\mathcal{H}]+\partial U/\partial x^0\rangle$.

The combination of Lemma 3 and 4 leads to a method to associate a conserved quantity with an invariance group of the equation:

Theorem 3: If the operator U defined by (16) is a generator of an invariance group of the equation $H+u_0=0$, and if H satisfies all the conditions in Lemma 4, then the quantity $\langle U\rangle$ defined by (60) is a constant of motion, i.e., $d/dx^0\langle U\rangle=0$.

We note that in proving this we did not assume the quantity $\int f^*(x)f(x)\,dx^1\cdots dx^N$ to be independent of time.

Lemma 3 can be generalized to a set of nonlinear time evolution equations of the form

$$H^i+u_0^i=0,\quad H^i=H^i(x,\mathbf{u},\underset{1}{\mathbf{u}},\underset{2}{\mathbf{u}},\ldots,\underset{r}{\mathbf{u}}),\quad i=1,2,\ldots,M, \tag{63}$$

where $\mathbf{u}=(u^1,u^2,\ldots,u^M)$, $\underset{k}{\mathbf{u}}=(\underset{k}{u}^1,\underset{k}{u}^2,\ldots,\underset{k}{u}^M)$ and $u^i=f^i(x)$, etc. In this case we have

Lemma 3′: If $U=\eta^i\partial_{u^i}$ is a generator of an invariance group of Eq. (63), then we have $[U,\mathcal{H}]+\partial U/\partial x^0=0$ where $\mathcal{H}=H^i\partial_{u^i}$ and $\partial U/\partial x^0=(\partial_{x^0}\eta^i)\,\partial_{u^i}$.

For Hamilton's equations of a field $[u^1=P=p(x),\ u^2=Q=q(x)]$

$$\frac{\delta\hat{H}}{\delta Q}+P_0=0,\quad -\frac{\delta\hat{H}}{\delta P}+Q_0=0,$$

we obtain the familiar expression

$$[U,\mathcal{H}]+\frac{\partial U}{\partial x^0}=0,\quad \text{with}\quad \mathcal{H}=\frac{\partial\hat{H}}{\partial Q}\partial_p-\frac{\partial H}{\partial p}\partial_Q.$$

The theorem above can be specialized to a real differential equation: If $H(x,u,\underset{1}{u},\underset{2}{u},\ldots,\underset{r}{u})$, in the equation $H+u_0=0$, satisfies an equation

$$H+uH_u-D_i(uH_{u_i})+\cdots+(-1)^rD_{p_1\cdots p_r}(uH_{u_{p_1\cdots p_r}})=0 \tag{64}$$

and if all the surface integrals $\int_S[u\eta_{j\cdots k}H_{u_{ij\cdots k}}]_{u=f(x)}\,\nu^i\,d\Omega$ vanish for $S=$ boundary, then the quantity $\langle U\rangle=\int_v[uUu]_{u=f}\,dx^1\cdots dx^N$ is a constant of motion. Here, $v=$ the whole space inside S.

The following equations which have been attracting considerable attention in the study of propagation of nonlinear waves satisfy the condition (61) or (64):

generalized Korteweg–de Vries equation

$$(\partial_x)^{2n+1}f+f^m\partial_xf+\partial_tf=0,$$

cubic Schrödinger equation in n dimensions

$$-i\left[\sum_{k=1}^{n}(\partial_{x^k})^2f+f^2f^*\right]+\partial_tf=0,$$

Hirota equation[8]

$$a(\partial_x)^3f+ib(\partial_x)^2f+cff^*\partial_xf+idf^2f^*+\partial_tf=0.$$

However, the heat equation $f_{xx}-f_t=0$ and Burgers equation $f_{xx}+ff_x-f_t=0$, both of which represent a dissipative system, do not satisfy Eq. (64).

The application of Theorem 3 to the KdV equation and to the cubic Schrödinger equation has turned out to produce only a few constants of motion,

KdV equation:

$\langle U^1\rangle = -\int u\,dx,\quad \langle U^2\rangle = \int \frac{1}{2}u^2\,dx,$

$\langle U^i\rangle = 0,\quad \text{for } i>2.$

cubic Schrödinger equation:

$\langle U^2\rangle = \int \frac{1}{2}uu^*\,dx,\quad \langle U^i\rangle = 0 \text{ for } i>2.$

V. CONCLUDING REMARKS

We have shown that provided one considers the group transformation which is more general than the one considered by Lie, one can associate one invariance group with each constant of motion of a class of physical systems. Thus for such a system one can derive the constants of motion by finding the invariance groups of the equation. One of the best known methods of finding conservation laws is to use Noether's theorem. The difference between the two is that the groups in the present approach leave the differential equation invariant whereas the groups in Noether's theorem leave an action integral invariant.

In the following communication, a generalization of Theorems 1 and 2 will be discussed.

ACKNOWLEDGMENTS

I am sincerely grateful to Professor Carl E. Wulfman for many helpful discussions and for a number of valuable comments on the manuscripts. I also thank the Research Corporation for supporting this research.

APPENDIX A: PROOF OF LEMMA 1

It is sufficient if we prove $D_0UA = UD_0A$. To avoid complex indices, we represent a set of indices $i\cdots k$ appearing in the expressions (7) and (9) of D_0 and U by a circle $\circ$ or by a dot $\bullet$, and write D_0 and U as

$$D_0 = \partial_{x^0} + \sum_\circ (u_{0\circ}\partial_{u_\circ} + v_{0\circ}\partial_{v_\circ}),$$

$$U = \sum_\circ (\eta_\circ \partial_{u_\circ} + \zeta_\circ \partial_{v_\circ})$$

where the sign $\sum_\circ$ indicates a summation over all the parenthesized quantities in (7) and (9). Then, by the definitions of D_0 and U,

$$D_0UA = D_0\sum_\circ(\eta_\circ A_{u_\circ} + \zeta_\circ A_{v_\circ})$$
$$= \sum_\circ(\eta_\circ D_0A_{u_\circ} + \zeta_\circ D_0A_{v_\circ}) + \sum_\circ[(D_0\eta_\circ)A_{u_\circ} + (D_0\zeta_\circ)A_{v_\circ}].$$

Using $D_0\eta_\circ = \eta_{0\circ} = Uu_{0\circ},\ D_0\zeta_0 = \zeta_{0\circ} = Uv_{0\circ}$,

$$D_0UA = \sum_\circ(\eta_\circ D_0A_{u_\circ} + \zeta_\circ D_0A_{v_\circ}) + \sum_\circ[(Uu_{0\circ})A_{u_\circ} + (Uv_{0\circ})A_{v_\circ}]. \qquad (*)$$

The first term is

$$\sum_\circ(\eta_\circ D_0A_{u_\circ} + \zeta_\circ D_0A_{v_\circ})$$
$$= \sum_\circ\{\eta_\circ[A_{0u_\circ} + \sum_\bullet(u_{0\bullet}A_{u_\circ u_\bullet} + v_{0\bullet}A_{u_\circ v_\bullet})]$$
$$+ \zeta_\circ[A_{0v_\circ} + \sum_\bullet(u_{0\bullet}A_{v_\circ u_\bullet} + v_{0\bullet}A_{v_\circ v_\bullet})]\}$$
$$= \sum_\circ(\eta_\circ A_{0u_\circ} + \zeta_\circ A_{0v_\circ}) + \sum_\bullet u_{0\bullet}\sum_\circ(\eta_\circ A_{u_\circ u_\bullet}$$
$$+ \zeta_\circ A_{u_\bullet v_\circ}) + \sum_\bullet v_{0\bullet}\sum_\circ(\eta_\circ A_{v_\bullet u_\circ} + \zeta_\circ A_{v_\bullet v_\circ})$$
$$= UA_0 + \sum_\bullet u_{0\bullet}UA_{u_\bullet} + \sum_\bullet v_{0\bullet}UA_{v_\bullet}.$$

Hence, $(*)$ gives $D_0UA = U[A_0 + \sum_\circ(u_{0\circ}A_{u\circ} + v_{0\circ}A_{v\circ})] = UD_0A$.

APPENDIX B: A COMMUTATOR OF GENERALIZED LIE TYPE OPERATORS

We consider two operators of the form (19),

$$U^1 = \eta^1\partial_u + \zeta^1\partial_v,\quad U^2 = \eta^2\partial_u + \zeta^2\partial_v.$$

We must interpret these as simplified representations of (9). The commutator of the two is defined as

$$[U^1, U^2] = [(U^1\eta^2) - (U^2\eta^1)]\,\partial_u + [(U^1\zeta^2) - (U^2\zeta^1)]\,\partial_v + \cdots$$
$$+ [(U^1\eta^2_{i\cdots k}) - (U^2\eta^1_{i\cdots k})]\,\partial_{u_{i\cdots k}}$$
$$+ [(U^1\zeta^2_{i\cdots k} - (U^2\zeta^1_{i\cdots k})]\,\partial_{v_{i\cdots k}} + \cdots.$$

We write this as

$$U = [U^1, U^2] = \eta\partial_u + \zeta\partial_v + \cdots + \eta_{i\cdots k}\partial_{u_{i\cdots k}}$$
$$+ \zeta_{i\cdots k}\partial_{v_{i\cdots k}} + \cdots.$$

We prove that this satisfies the condition imposed on (9), i.e., the condition (6). In fact, by applying Lemma 1,

$$\eta_{i\cdots k} = U^1\eta^2_{i\cdots k} - U^2\eta^1_{i\cdots k} = U^1D_{i\cdots k}\eta^2 - U^2D_{i\cdots k}\eta^1$$
$$= D_{i\cdots k}(U^1\eta^2 - U^2\eta^1) = D_{i\cdots k}\eta.$$

Similarly $\zeta_{i\cdots k} = D_{i\cdots k}\zeta$. Therefore, the operator obtained from the commutator of two operators of the form (9) also assumes the same form.

*This study has been supported by a Research Corporation grant.

[1]R. L. Anderson, S. Kumei, and C. E. Wulfman, Phys. Rev. Lett. 28, 988 (1972). I have learned that the idea of the derivative dependent infinitesimal transformation is not new. For instance, H. Johnson, Proc. Am. Math. Soc. 15, 675 (1964). I am grateful to Professor A. Kumpera of University of Montreal for bringing this paper to my attention.

[2]S. Lie, *Transformationgruppen* (Chelsea, New York, 1970), 3 Vols. (Reprints of 1888, 1890, 1893 eds.); *Differentialgreichungen* (Chelsea, New York, 1967) (Reprint of 1897 ed.); *Continuierliche Gruppen* (Chelsea, New York, 1967) (Reprint of 1893 ed.).

[3]L. V. Ovsjannikov, *Group Theory of Differential Equation* (Siberian Section of the Academy of Sciences, Novosibirsk, USSR, 1962). [This book has been translated into English by G. W. Bluman (unpublished).]
[4]G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[5]R. L. Anderson, S. Kumei, and C. E. Wulfman, Rev. Mex. Fis. **21**, 1, 35 (1972); J. Math. Phys. **14**, 1527 (1973).
[6]N. H. Ibragimov and R. L. Anderson, "Lie–Bäcklund Tangent Transformations." J. Math. Anal. Appl. (accepted for publication). (See Sov. Math. Dokl. 17, 437 (1976) for excerption.)
[7]S. Kumei, J. Math. Phys. **16**, 2461 (1975).
[8]A. C. Scott, F. Y. F. Chu, and D. W. MacLaughlin, Proc. IEEE **61**, 1443 (1973).
[9]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys.-JETP **34**, 62 (1972)].
[10]The notation $\underset{k}{u}$ is from Ovsjannikov (Ref. 3).
[11]S. Kumei (unpublished). Also, L. V. Ovsjannikov (Ref. 3).
[12]N. J. Zabusky, in *Nonlinear Partial Differential Equations*, edited by W. Ames (Academic, New York, 1967).
[13]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).
[14]M. D. Kruskal, R. M. Miura, C. S. Gardner, and N. J. Zabusky, J. Math. Phys. **11**, 952 (1970).
[15]P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968); **28**, 141 (1975).
[16]C. S. Gardner, J. Math. Phys. **12**, 1548 (1971).
[17]H. Shen and W. F. Ames, Phys. Lett. A 49, 313 (1974). I owe this reference to Professor R. L. Anderson.
[18]V. E. Zakharov and S. V. Manakov, Teor. Mat. Fiz. **19**, 332 (1974).
[19]Y. Kodama, Prog. Theor. Phys. **54**, 669 (1975).
[20]S. Kumei (unpublished).
[21]H and the energy $\hat{H}$ in Sec. III should not be confused.

On plane symmetric Einstein–Maxwell fields

A. Banerjee

Department of Physics, Jadavpur University, Calcutta-32, India

N. Chakrabarty

Department of Mathematics, St. Anthony's College, Shillong, India

(Received 26 February 1976)

It is shown that the results given by Letelier and Tabensky [J. Math. Phys. **15**, 594 (1974)] regarding the solution of Einstein–Maxwell equations for plane symmetry will be modified after suitable correction of an error in their paper. Further appropriate exterior solutions, which satisfy the conditions of fit at the boundaries of plane symmetric charged dust distributions, are obtained. One such exterior nonstatic solution can be transformed to a static one, which is a member of the general class.

INTRODUCTION

In a recent paper Letelier and Tabensky[1] have given the general solution of Einstein Maxwell equations in the charge-free empty space, where both the metric tensors $g_{\mu\nu}$ and the Maxwell field tensors $F_{\mu\nu}$ remain invariant under the group of motions that characterize plane symmetry in the sense of Taub.[2] The work of Letelier and Tabensky, however, contains an error in sign in the field equations and the results given therein are modified to a significant extent after suitable corrections. It is found that a class of the solutions, which is apparently time dependent in their case, is indeed a class of static solutions.

In the second part we have considered the problem of matching of some of the solutions given earlier by De[3,4] for the interior of plane symmetric distributions of incoherent charged dust to their corresponding exterior metric. One such exterior solution obtained in the same comoving system used for the interior, is found to satisfy the conditions of fit at both the boundaries of the plane symmetric charged dust distribution, which will finally collapse.

It is also interesting to note that a suitable exterior nonstatic metric matching with the corresponding interior one in another case given by De[4] can be transformed to a purely static metric, which is a member of the general class of solutions mentioned earlier.

1. DISCUSSION OF THE GENERAL SOLUTIONS

The final forms of the Eqs. (10) and (11) of Letelier and Tabensky, after a suitable correction of sign should be

$$\mu_{,01} + W_{,01} + \tfrac{1}{2}\mu_{,0}\mu_{,1} = -\tfrac{1}{4}ke^{(W-2\mu)}(C_1^2 + C_2^2) \tag{1}$$

and

$$\mu_{,01} + \mu_{,0}\mu_{,1} = \tfrac{1}{4}ke^{(W-2\mu)}(C_1^2 + C_2^2). \tag{2}$$

Their Eq. (21) must then be written in the form

$$2e^{\mu}(e^{\mu/2})' + \tfrac{1}{2}k(C_1^2 + C_2^2) + C_3 e^{\mu/2} = 0. \tag{3}$$

It can be shown that for $C_3 \geq 0$, the metric reduces to a static one. Particularly for $C_3 = 0$, the solution is

$$e^{\mu(x)} = [C_4 - \tfrac{3}{4}k(C_1^2 + C_2^2)x]^{2/3}, \tag{4}$$

contrary to the time dependent solution obtained by Letelier and Tabensky. On the other hand, when $C_3 < 0$, there are, however, either time dependent spatially homogeneous solutions or static solutions discussed in Ref. 1.

2. MATCHING OF THE PLANE SYMMETRIC CHARGED DUST METRIC WITH THE CORRESPONDING CHARGE FREE ELECTROVAC SOLUTION

1. Exact solutions corresponding to the interior of the plane symmetric charged dust distributions were given by De.[3,4] He had further shown that such distributions would finally collapse. We consider one such solution which is given in the form[3]

$$ds^2 = (R+T)^{-4/5}dt^2 - (R+T)^{-2/5}dx^2 - (R+T)^{4/5}(dy^2 + dz^2), \tag{5}$$

$$F^{01} = \tfrac{2}{5}(R+T)^{-1/5}R', \quad 4\pi\rho = 4\pi|\sigma| = \tfrac{2}{5}(R+T)^{-3/5}[-R''], \tag{6}$$

where $T = (at + b)$, a and b being arbitrary constants. R is any function of x subject to the condition $R'' < 0$ in the interior region, so that the matter density may remain positive. The symbol $'$ stands for the differentiation with respect to the x coordinate. We note that since in this case R'' is negative everywhere in the interior region, the charge density is of the same sign throughout. One should note further that here we take F^{01} as the only nonvanishing component of the Maxwell tensor without loss of generality in view of the duality transformation. Now if we choose $R' = \text{const}$, the solution reduces to that of empty space–time containing only the nonnull electromagnetic field. It is thus worthwhile to find a suitable function $R(x)$, so that R'' vanishes at the boundaries. We choose

$$R'' = (x^2 - \alpha^2), \tag{7}$$

α being a constant and at the boundary $x = \pm\alpha$. The form (7) satisfies the requirement $R'' < 0$ for $x^2 < \alpha^2$ that is within the interior of the charged dust. The relation (7) immediately gives

$$R' = [x^3/3 - \alpha^2 x + \beta]\ . \tag{8}$$

Further, if we assume that the electric field vanishes on the y-z plane passing through the origin of the spatial coordinates or, in other words, the central plane of symmetry $F^{01}=0$ at $x=0$ so that $\beta=0$. One then gets for the interior

$$R=(x^4/12-\alpha^2x^2/2+C), \tag{9}$$

where C is the integration constant.

Now at

$$x=+\alpha, \quad R=-\tfrac{2}{3}\alpha^3$$

and at (10)

$$x=-\alpha, \quad R=+\tfrac{2}{3}\alpha^3,$$

so that the matter-free and charge-free space with nonvanishing electromagnetic field tensor F^{01} is separated into two regions by the bounded charged dust distribution. The corresponding line element can be written in the same form as (5) with

$$R(x)=(-\tfrac{2}{3}\alpha^3x+\gamma) \tag{11}$$

in region I and

$$R(x)=(\tfrac{2}{3}\alpha^3x+\gamma) \tag{12}$$

in region II. The constants γ, α, and C appearing in (9), (11), and (12) are related in view of the continuity of $R(x)$ at $x\pm\alpha$ and the necessary relation is

$$\alpha^4=4(\gamma-C).$$

We can thus go from region I to region II by the transformation $x\to -x$, and the metric along with its first derivatives are seen to be continuous across the boundaries $x=\pm\alpha$.

2. Following almost the same procedure one can find the exterior solution corresponding to another simple interior solution given for a plane symmetric nonstatic distribution of charged dust in the form (De[4])

$$ds^2=(X+G)^{-2/3}(dt^2-dx^2)-(X+G)^{2/3}(dy^2+dz^2) \tag{13}$$

with

$$F^{01}=(X'/3)(1-9k_1^2/(X')^2)^{1/2},$$

$$8\pi\rho=-\frac{2}{3}\left[\frac{X''}{(X+G)^{1/3}}\right],$$

and

$$8\pi|\sigma|=\frac{2/3}{(X+G)^{1/3}}\left|\frac{d}{dx}\{X'(1-9k_1^2/(X')^2)^{1/2}\}\right|, \tag{14}$$

where X is an arbitrary function of the spatial coordinate x, subject to the restriction that $X''<0$ in the interior and G is a function of time written as

$$G(t)=(3k_1t+k_2), \tag{15}$$

k_1 and k_2 being arbitrary constants.

By the same arguments given previously one can write the corresponding exterior metric in matter-free and charge-free space with a nonvanishing electric field by putting $X'=\text{const}$ and this in turn will be determined from the field equation [cf. Eqs. (2.7c) of De[4]]

$$R_0^0+R_1^1=2F^{01}F_{01} \tag{16}$$

which gives the relation

$$9k_1^2-X'^2=-9A^2. \tag{17}$$

Here the electromagnetic field tensor F^{01} is expressed in view of the Maxwell equations

$F^{01}=A/(-g)^{1/2}$, A being an integration constant.

The relation (17) gives directly

$$X'=\pm 3(k_1^2+A^2)^{1/2}. \tag{18}$$

We write $3k_1=\alpha$ and $\pm 3(k_1^2+A^2)^{1/2}=\beta$, so that $\beta^2>\alpha^2$.

Thus the desired metric is expressed by the line element

$$ds^2=(\alpha t+\beta x+\gamma)^{-2/3}(dt^2-dx^2)-(\alpha t+\beta x+\gamma)^{2/3}(dy^2+dz^2), \tag{19}$$

where γ is a constant.

The metric (19) can now be easily transformed to a purely static one by a coordinate transformation of the type

$$\bar{x}=\frac{x+(\alpha/\beta)t}{(1-\alpha^2/\beta^2)^{1/2}} \quad\text{and}\quad \bar{t}=\frac{t+(\alpha/\beta)x}{(1-\alpha^2/\beta^2)^{1/2}}. \tag{20}$$

Such a transformation, which is apparently a Lorentz transformation is allowed because in this case $\beta^2>\alpha^2$. The transformed metric is thus, omitting the bars over the coordinates,

$$\begin{aligned} ds^2=&[\beta(1-\alpha^2/\beta^2)^{1/2}x+\gamma]^{-2/3}(dt^2-dx^2)\\ &-[\beta(1-\alpha^2/\beta^2)^{1/2}x+\gamma]^{2/3}(dy^2+dz^2) \end{aligned} \tag{21}$$

which is a member of the general class of solutions in (4) given earlier (Patnaik[5]).

ACKNOWLEDGMENT

The authors are grateful to Professor A. K. Raychaudhuri for his suggestions.

[1]P.S. Letelier and R.R. Tabensky, J. Math. Phys. **15**, 594 (1974).
[2]A.H. Taub, Ann. Math. **53**, 472 (1951).
[3]U.K. De, Nuovo Cimento B **15**, 61 (1973).
[4]U.K. De, Prog. Theor. Phys. **49**, 1546 (1973).
[5]S.N. Patnaik, Proc. Cambridge Philos. Soc. **67**, 127 (1970).

Causal solutions of nonlinear wave and spinor equations obtained by Gel'fand–Shilov regularization

D. Baumgarten, G. Braunss, and O. Wagner

Mathematisches Institut der Justus Liebig-Universität, D6300 Giessen, Federal Republic of Germany
(Received 9 February 1976)

By extending the Gel'fand–Shilov regularization method to products of locally integrable functions it can be shown that the nonlinear Klein–Gordon equations $\partial^2 u/\partial t^2 - \partial^2 u/\partial x_1^2 - \cdots - \partial^2/\partial x_n^2 + ku^{2p+1} = 0$, where $k = \text{const} > 0$, $n \geq 3$, $p = \text{integer} \geq 1$, have causal solutions which have no δ singularities. It is further shown that this also can be expected for the nonlinear Dirac equation $\gamma_\lambda \partial\psi/\partial x_\lambda + l^2\psi(\bar{\psi}\psi) = 0$.

INTRODUCTION

The following notation will be used:

$$\Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \cdots - \frac{\partial^2}{\partial x_n^2},$$

$$x = (t, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}, \quad r = (x_1^2 + \cdots + x_n^2)^{1/2},$$

$$dx = dt\, dx_1 \cdots dx_n,$$

$K = \{x \mid t^2 - r^2 \geqslant 0\}$ (light cone),

$K_+ = \{x \mid t \geqslant r\}$ (forward light cone),

$K_- = \{x \mid t \leqslant -r\}$ (backward light cone).

We shall consider wave equations

$$\Box u + ku^{2p+1} = 0, \quad k = \text{const} > 0, \qquad (*)$$

for $n \geqslant 3$, $p = \text{integer} \geqslant 1$. A causal solution of (*) is a distribution with support in K which is a weak solution of (*) defined in a suitable sense. Causal solutions of (*) have been studied by applying either a limit process (cf. Ref. 1) or a class calculus of distributions (cf. Refs. 2 and 3). Both methods have led to the same result: If (*) has a solution depending on $s = t^2 - r^2$ which is analytic in a neighborhood of $s = 0$, then (*) permits a causal solution with support in K which has a δ term concentrated on $s = 0$. The same condition granted it had been shown that then the equation

$$\Box u + ku^{2p+1} = \delta_{x=0} \qquad (**)$$

permits a causal solution with support in K_+ (in K_-) which has a δ term concentrated on $t = r$ (on $t = -r$). In particular for $n = 3$, $p = 1$, it has been shown that such solutions exist [they exist for Eq. (*) in general]. Now, Eq. (*) also has solutions which are singular at $s = 0$. These solutions are const$s^{-1/2p}$ or $s^{-1/2p} f(s)$ where f is bounded for $s \geqslant 0$ but oscillates increasingly if $s \to 0$ or $s \to \infty$ [that is, the zeros of f have an accumulation point at $s = 0$ and $s = \infty$, the function f behaves, for example, like the function cos(log$s$)]. Our paper deals with these solutions. By extending the Gel'fand–Shilov regularization method to products of locally integrable functions and applying it to Eq. (*) we shall prove that Eq. (*) has causal solutions with support in K, K_+, and K_- which have no δ singularities but singularities of the above type, i.e., singularities which are algebraic (branching points). Apart from the different type of singularities these solutions differ from the δ type solutions by the following properties: First, their asymptotic behavior is like $s^{-1/2}$ whereas the δ type behaves like s^{-1} for $s \to \infty$. Second, the retarded (advanced) solutions with support in K_+ (in K_-) do not require an additional $\delta_{x=0}$ term on the right-hand side of Eq. (*), i.e., they satisfy Eq. (*) instead of (**). Third, the sum of an advanced and a retarded solution yields a solution with support in K—despite the nonlinearity of (*).

1. AN EXTENSION OF THE GEL'FAND–SHILOV REGULARIZATION METHOD

Definition 1: Let $u: M \times G_0 \to \mathbb{C}$, $M \subset R^{n+1}$, $G_0 \subset \mathbb{C}$, and $F: u(M \times G_0) \to \mathbb{C}$ be such that $F[u(x;\lambda)]$, $x \in M$, $\lambda \in G_0$, is locally integrable with respect to x on M for all λ in G_0. Then the distribution $F[u_M(\cdot;\lambda)]$ shall be defined by

$$\langle F[u_M(\cdot;\lambda)], \varphi\rangle = \int_M F[u(x;\lambda)]\,\varphi(x)\,dx, \quad \varphi \in C_c^\infty(\mathbb{R}^{n+1}).$$

If the above expression is analytic with respect to λ on G_0 for all $\varphi \in C_c^\infty(\mathbb{R}^{n+1})$ and has an analytic continuation into some $G \subset \mathbb{C}$ then the distribution $F[u_M(\cdot;\lambda)]$, $\lambda \in G$, is defined by analytic continuation (≡a.c.),

$$\langle F[u_M(\circ;\lambda)], \varphi\rangle_{\lambda\in G} = \text{a.c.}\,\langle F[u_M(\circ;\lambda)], \varphi\rangle_{\lambda\in G_0}, \quad \varphi \in C_c^\infty(\mathbb{R}^{n+1}).$$

Remark 1: If $u(x;\lambda)$ is locally integrable then for $F = 1$ (= identity) our definition yields the usual Gel'fand–Shilov regularization (cf. Ref. 4). Therefore we call the above construction an *extended GS-regularization*. Formally it can be looked upon as a multiplication rule for certain distributions, so one might be tempted by it to construct algebras of these distributions. This is perfectly all right as long as one does not require these algebras to be derivation algebras (under differentiation). In this case some additional rules would have to be set up. However, for the problems treated here no such algebraic finesse is needed.

Definition 2: Let $u(\cdot;\lambda)$ and F be such that the distributions $u_M(\cdot;\lambda)$ and $F[u_M(\cdot;\lambda)]$ exist in the sense of Definition 1 where M is one of the following sets: K, K_+, K_-. If for some λ_0 depending possibly on M and all $\varphi \in C_c^\infty(\mathbb{R}^{n+1})$,

$$\langle \Box u_M(\circ;\lambda_0), \varphi\rangle + \langle F[u_M(\cdot;\lambda_0)], \varphi\rangle = 0,$$

then $u_M(\circ;\lambda_0)$ shall be called a (weak) causal solution of

$$\Box u + F(u) = 0$$

in the sense of an extended GS regularization; $u_M(\cdot;\lambda_0)$ is called retarded or advanced in case $M = K_+$ or $M = K_-$ accordingly.

2. CAUSAL SOLUTIONS OF WAVE EQUATIONS $\Box u + ku^{2p+1} = 0$

Theorem 1: Consider

$$\Box u + ku^{2p+1} = 0, \quad k = \text{const} > 0, \tag{1}$$

where $n \geq 3$ $p = \text{integer} \geq 1$. Let the distribution P_M^λ, $\text{Re}\lambda > -1$, be defined by

$$\langle P_M^\lambda, \varphi \rangle = \int_M (t^2 - r^2)^\lambda \varphi(x)\,dx, \quad \varphi \in C_c^\infty(\mathbb{R}^{n+1}).$$

Then

$$u_M(\circ; -1/2p) = [(np - p - 1)/kp^2]^{1/2p} P_M^{-1/2p},$$

for $M = K, K_+, K_-$, is a causal solution of (1) in the sense of an extended GS regularization.

Proof: Let $\lambda \in \mathbb{C}$, $-j-1 \neq \lambda \neq -j-(n+1)/2$ $(j = 0, 1, 2, \cdots)$. For $M = K, K_+, K_-$ there holds the identity (cf. Ref. 5)

$$\langle \Box P_M^{\lambda+1}, \varphi \rangle = 4(\lambda+1)(\lambda+(n+1)/2)\langle P_M^\lambda, \varphi \rangle, \quad \varphi \in C_C^\infty(\mathbb{R}^{n+1}). \tag{2}$$

Let $u(x;\lambda) = c(t^2 - r^2)^\lambda$, $c = \text{const}$. Then $u(x;\lambda)$ is locally integrable for $\text{Re}\lambda > -1$. Hence by Definition 1

$$u_M(\cdot;\lambda) = cP_M^\lambda$$

is defined by analytic continuation for all $\lambda \in \mathbb{C}$, $-j-1 \neq \lambda \neq -j-(n+1)/2$ $(j = 0, 1, 2, \cdots)$. Using the identity (2) we obtain for $\lambda = -1/2p$, $p = \text{integer} \geq 1$, $n \geq 3$, and $c = [(np-p-1)/kp^2]^{1/2p}$ after a short calculation,

$$\langle (\Box u_M)(\circ; -1/2p), \varphi \rangle = -kc^{2p+1}\langle P_M^{-1-1/2p}, \varphi \rangle, \quad \varphi \in C_c^\infty(\mathbb{R}^{n+1}).$$

Now let $F(u) = ku^{2p+1}$. By Definition 1

$$F[u_M(\circ;\lambda)] = kc^{2p+1} P_M^{\lambda(2p+1)}$$

exists for $-j-1 \neq \lambda(2p+1) \neq -j-(n+1)/2$ $(j = 0, 1, 2, \cdots)$, so we may choose $\lambda = -1/2p$, $n \geq 3$, $p = \text{integer} \geq 1$. Hence for $\varphi \in C_c^\infty(\mathbb{R}^{n+1})$

$$\langle (\Box u_M)(\circ; -1/2p), \varphi \rangle + \langle k[u_M(\circ; -1/2p)]^{2p+1}, \varphi \rangle = 0.$$

This proves the assertion.

Remark 2: Both $P_{K_+}^\lambda$ and $P_{K_-}^\lambda$ for $\text{Re}\lambda > -1$ are locally integrable and the intersection of their supports is the singleton $\{0\}$. Consequently $P_{K_+}^\lambda P_{K_-}^\lambda$ for $\text{Re}\lambda > -1$ is the zero function; its analytic continuation is trivially the zero function for all $\lambda \in \mathbb{C}$. Since on the other hand, $P_K^\lambda = P_{K_+}^\lambda + P_{K_-}^\lambda$, we have—despite the nonlinearity of Eq. (1)—a linear superposition: $u_K = u_{K_+} + u_{K_-}$.

Theorem 2: Consider

$$\Box u + ku^3 = 0, \quad k = \text{const} > 0, \tag{3}$$

for $n = 3$. Let

$$u(x;\lambda) = cs^\lambda \text{cn}(a\log s + b), \quad s = t^2 - r^2,$$

where a, b, c are real constants with b arbitrary, $a \neq 0 \neq c$, and cn means cosinus amplitudinis with modulus κ, where

$$8a^2\kappa^2 = c^2k = 1 + 4a^2, \quad 0 < \kappa \leq 1.$$

Then $u_M(\cdot; -\frac{1}{2})$ for $M = K, K_+, K_-$ is a causal solution of (3) in the sense of an extended GS regularization.

Remark 3: The condition $0 < \kappa \leq 1$ is not obligatory. If $\kappa' = (1-\kappa^2)^{1/2}$ and if we replace, for example, $g = a\log s + b$ either by $\kappa' g$ or $(1+\kappa)g$ and κ by $i\kappa/\kappa'$ or $2\kappa^{1/2}/(1+\kappa)$ correspondingly, then the resulting functions can be treated exactly as we treat the example considered here.

Proof of Theorem 2: The proof will be in three steps for $M = K$; for $M = K_+$ and $M = K_-$ the proofs are completely analogous and therefore will be omitted.

(i) Let

$$v(x;\lambda) = s^\lambda w(z;\lambda), \quad z = a\log s + b,$$

and assume

$$w(z;\lambda) = \sum_{m=-\infty}^{\infty} c_m(\lambda)\exp(im\alpha z), \quad \alpha = \text{real const} \neq 0,$$

to be uniformly convergent on the real line with respect to z for each complex λ and the $c_m(\lambda)$ to be entire functions of λ. If $\text{Re}\lambda > -1$ then $v(x;\lambda)$ is locally integrable on K with respect to x. Let $\varphi \in C_c^\infty(\mathbb{R}^4)$ and

$$\overline{\varphi}(t, r) = \int_{\Omega_3} \varphi(x)\,d\Omega_3,$$

where Ω_3 is the surface of the unit sphere in $\mathbb{R}^3$. Let

$$\psi(t, r) = \tfrac{1}{2}[\overline{\varphi}(t, r) + \overline{\varphi}(-t, r)].$$

Substituting $t = r(1+\tau)^{1/2}$, we get for $\text{Re}\lambda > -1$

$$\begin{aligned}
&\langle v_K(\circ;\lambda), \varphi \rangle \\
&= \int_K v(x;\lambda)\varphi(x)\,dx \\
&= \int_{\tau=0}^\infty \int_{r=0}^\infty r^{2\lambda+3}\tau^\lambda(1+\tau)^{-1/2} w(z;\lambda)\psi(r\sqrt{1+\tau}, r)\,dr\,d\tau \\
&= \sum_{m=-\infty}^{\infty} \int_{\tau=0}^\infty \tau^{\lambda+im\beta}(1+\tau)^{-1/2} \int_{r=0}^\infty C_m(r;\lambda)\psi(r\sqrt{1+\tau}, r)\,dr\,d\tau,
\end{aligned}$$

where

$$C_m(r;\lambda) = c_m(\lambda) r^{2\lambda+3}\exp\{im\alpha[a\log(r^2) + b]\}, \quad \beta = \alpha a.$$

Let

$$\phi_m(\tau;\lambda) = (1+\tau)^{-1/2}\int_{r=0}^\infty C_m(r;\lambda)\psi(r\sqrt{1+\tau}, r)\,dr.$$

Then $\phi_m(\cdot;\lambda)$ is C^∞ on $(-1, \infty)$ and $\phi_m(\tau;\lambda) = 0$ if $\tau \geq \tau_0(\psi)$ for $\text{Re}\lambda > -2$ and all integers m. It follows that

$$\langle v_K(\cdot;\lambda), \varphi \rangle = \sum_{m=-\infty}^{\infty} \int_{\tau=0}^\infty \tau^{\lambda+im\beta}\phi_m(\tau;\lambda)\,d\tau$$

exists by analytic continuation for $\text{Re}\lambda > -2$, $\lambda \neq 1 + im\beta$, $m = \text{integer}$.

(ii) Let $w(z) = \text{cn}z$. Then (cf. Ref. 6)

$$w(z) = \sum_{m=-\infty}^{\infty} c_m \exp(im\alpha z), \ \alpha = \text{real const} \neq 0,$$

uniformly. Let

$$T_1 : w \to [a^2D^2 + a(2\lambda+1)D + \lambda(\lambda+1)]w, \quad D = d/dz,$$

$$T_2 : w \to w^3.$$

Then $(T_1 w)(z;\lambda)$ and $(T_2 w)(z)$ too can each be expanded in a uniformly convergent Fourier series for all $\lambda \in \mathbb{C}$,

$$w_1(z;\lambda) \equiv (T_1 w)(z;\lambda) = \sum_{m=-\infty}^{\infty} c_m^{(1)}(\lambda)\exp(im\alpha z),$$

$$c_m^{(1)}(\lambda) = [-a^2\alpha^2 m^2 + ia\alpha m(2\lambda+1) + \lambda(\lambda+1)]c_m,$$

$$w_2(z) \equiv (T_2 w)(z) = \sum_{m=-\infty}^{\infty} c_m^{(2)} \exp(im\alpha z).$$

Let $u(x;\lambda) = s^\lambda w(z)$. A short calculation yields

$$(\Box u)(x;\lambda) = 4s^{\lambda-1} w_1(z;\lambda).$$

Further,

$$[u(x;\lambda)]^3 = s^{3\lambda} w_2(z).$$

Let

$$\phi_m^{(j)}(\tau;\lambda) = (1+\tau)^{-1/2} \int_{r=0}^{\infty} C_m^{(j)}(r;\lambda)\, \psi(r\sqrt{1+\tau}, r)\, dr \quad (j=1,2),$$

$$C_m^{(1)}(r;\lambda) = 4c_m^{(1)}(\lambda)\, r^{2\lambda+1} \exp\{im\alpha[a\log(r^2)+b]\},$$

$$C_m^{(2)}(r;\lambda) = c_m^{(2)} r^{6\lambda+3} \exp\{im\alpha[a\log(r^2)+b]\}.$$

Then, for all $\varphi \in C_c^\infty(\mathbb{R}^4)$,

$$\langle(\Box u)_K(\circ;\lambda), \varphi\rangle = \sum_{m=-\infty}^{\infty} \int_{\tau=0}^{\infty} \tau^{\lambda-1+im\beta} \phi_m^{(1)}(\tau;\lambda)\, d\tau, \quad \mathrm{Re}\lambda > 0, \tag{4}$$

$$\langle[u_K(\cdot;\lambda)]^3, \varphi\rangle = \sum_{m=-\infty}^{\infty} \int_{\tau=0}^{\infty} \tau^{3\lambda+im\beta} \phi_m^{(2)}(\tau;\lambda)\, d\tau, \quad \mathrm{Re}\lambda > -\tfrac{1}{3}. \tag{5}$$

By Definition 1 and the results of (i) it follows that (4) exists by analytic continuation for $\mathrm{Re}\lambda > -1$, $\lambda \neq im\beta$, m = integer, whereas (5) exists by analytic continuation for $\mathrm{Re}\lambda > -\frac{2}{3}$, $\lambda \neq -\frac{1}{3} + im\beta$, m = integer.

By verification one proves that $u(x;-\frac{1}{2})$ is a solution of (3). Hence we must have

$$4c_m^{(1)}(-\tfrac{1}{2}) + kc_m^{(2)} = 0$$

and consequently

$$\phi_m^{(1)}(\tau;-\tfrac{1}{2}) + k\phi_m^{(2)}(\tau;-\tfrac{1}{2}) = 0, \quad \tau \geq 0$$

for all integers m. Thus for all $\varphi \in C_c^\infty(\mathbb{R}^4)$,

$$\langle(\Box u)_K(\circ;-\tfrac{1}{2}), \varphi\rangle + \langle k[u_K(\cdot;-\tfrac{1}{2})]^3, \varphi\rangle = 0. \tag{6}$$

(iii) Let ∂K denote the boundary of K, dS the surface measure on ∂K, and $\partial/\partial n$ the derivative in the direction of the inward normal of K. Then for $\mathrm{Re}\lambda > 1$:

$$u(x,\lambda)\Big|_{\partial K} = 0, \quad \frac{\partial u}{\partial n}(x;\lambda)\Big|_{\partial K} = 0.$$

Applying Green's formula we get for $\mathrm{Re}\lambda > 1$ and $\varphi \in C_c^\infty(\mathbb{R}^4)$,

$$\begin{aligned}
&\langle(\Box u_K)(\circ;\lambda), \varphi\rangle \\
&\quad = \langle u_K(\circ;\lambda), \Box\varphi\rangle \\
&\quad = \int_K u(x;\lambda)(\Box\varphi)(x)\, dx \\
&\quad = \int_K (\Box u)(x;\lambda)\, \varphi(x)\, dx \\
&\qquad + \int_{\partial K} [\varphi(x) \frac{\partial u}{\partial n}(x;\lambda) - u(x;\lambda)\frac{\partial\varphi}{\partial n}(x)]\, dS \\
&\quad = \int_K (\Box u)(x;\lambda)\, \varphi(x)\, dx = \langle(\Box u)_K(\circ;\lambda), \varphi\rangle.
\end{aligned}$$

By analytic continuation we obtain with (4) for $\mathrm{Re}\lambda > -1$, $\lambda \neq im\beta$, m = integer,

$$\langle(\Box u_K)(\cdot;\lambda), \varphi\rangle = \langle(\Box u)_K(\cdot;\lambda), \varphi\rangle.$$

From this equation and Eq. (6) the assertion follows.

3. THE NONLINEAR DIRAC EQUATION $\gamma_\lambda\, \partial\psi/\partial x_\lambda + l^2\psi(\bar\psi\psi) = 0$

We conclude with a few remarks on the nonlinear Dirac equation

$$\gamma_\lambda \frac{\partial\psi}{\partial x_\lambda} + l^2\psi(\bar\psi\psi) = 0 \tag{7}$$

(summation runs from 1 to 4 over indices occurring twice, $x_4 = it$). This equation has been considered by Heisenberg (cf. Ref. 7), who studied solutions of the type

$$\psi(x) = [x_\lambda\gamma_\lambda\chi(s) + \varphi(s)]a, \quad s = t^2 - r^2, \tag{8}$$

where a is a constant spinor and χ and φ are real-valued functions. Without going into a detailed investigation of causal solutions of (7) we want to show that the singularities of (8) fit in our scheme. The following calculations are somewhat simpler and easier to discuss than those given in Ref. 7. As in Ref. 7, we start with the substitutions

$$\chi(s) = |A|^{-1/2} s^{-3/4} f(z), \tag{9}$$

$$\varphi(s) = (\mathrm{sign}A)|A|^{-1/2} s^{-1/4} g(z), \tag{10}$$

where $z = \log s$ and A is a real constant (depending on a and l^2) which may be positive or negative. Then (7) splits into the following two equations ($f' \equiv df/dz$, $g' \equiv dg/dz$):

$$\tfrac{5}{2}f + 2f' + g(f^2+g^2) = 0, \tag{11}$$

$$\tfrac{1}{2}g - 2g' + f(f^2+g^2) = 0. \tag{12}$$

From here on our calculations will differ from those given in Ref. 7. Let

$$f = \rho\cos\phi, \quad g = \rho\sin\phi,$$

where ρ and ϕ are functions of $z = \log s$. Substituting these expressions in (11) and (12), then multiplying (11) by $\sin\phi$ (by $\cos\phi$) and (12) by $\cos\phi$ (by $-\sin\phi$), and adding the resulting equations we get ($'$ means differentiation with respect to z)

$$\rho^2 = 2\phi' - \tfrac{3}{2}\sin 2\phi, \quad 2\rho' = \rho(\tfrac{1}{2} - 3\cos^2\phi).$$

Differentiation of the first of these equations and substitution of $2\rho\rho'$ from the second and ρ^2 from the first equation yields

$$\phi'' + \phi' = \tfrac{3}{4}(\sin 2\phi + \tfrac{3}{4}\sin 4\phi) \equiv G(\phi).$$

It follows by integration that

$$\phi'(z) = \exp(-z)\{C + \int_{-\infty}^{z} \exp(\zeta)\, G[\phi(\zeta)]\, d\zeta\},$$

where C is a constant of integration. Let

$$h(s) = 2\exp(-z)\int_{-\infty}^{z} \exp(\zeta)\, G[\phi(\zeta)]\, d\zeta - \tfrac{3}{2}\sin[2\phi(z)] \quad (z = \log s).$$

Then

$$\rho(z) = \{2\phi'(z) - \tfrac{3}{2}\sin[2\phi(z)]\}^{1/2} = s^{-1/2}[C + sh(s)]^{-1/2}.$$

Since

$$|G(\phi)| \le \tfrac{3}{4}(1 + \tfrac{3}{4}), \quad \phi \in \mathbb{R},$$

we have $|h(s)| \le 12$, $s \ge 0$, and thus

$$|\rho(z)| \le |Cs^{-1} + 12|^{1/2}, \quad s \ge 0.$$

Going back to (9) and (10) we get

$$\chi(s) = |A|^{-1/2} s^{-5/4}[C + sh(s)]^{1/2} \cos[\phi(\log s)],$$

$$\varphi(s) = (\operatorname{sgn} A)|A|^{-1/2} s^{-3/4}[C + sh(s)]^{1/2} \sin[\phi(\log s)].$$

Evidently the singularities of χ and φ fit in our scheme, so it can be expected that the corresponding causal solutions of (7) do not contain δ singularities.

[1]G. Braunss, J. Diff. Eqn. **9**, 86 (1971).
[2]G. Braunss and R. Liese, J. Diff. Eqn. **16**, 399 (1974).
[3]G. Braunss and R. Liese, "Retarded and advanced causal solutions of nonlinear wave equations," to appear in J. Diff. Eqn.
[4]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I.
[5]E. M. De Jager, *Mathematics Applied to Physics* (Springer, Berlin, 1970).
[6]I. M. Ryzhik and I. S. Gradstein, *Tables of Series, Products, and Integrals* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1957).
[7]W. Heisenberg, Z. Naturforsch. **9a**, 292 (1954).

Lie theory and the wave equation in space–time. 3. Semisubgroup coordinates

E. G. Kalnins

Mathematics Department, University of Waikato, Hamilton, New Zealand

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 22 March 1976)

We classify and study those coordinate systems which permit R separation of variables for the wave equation in four-dimensional space–time and such that at least one of the variables corresponds to a one-parameter symmetry group of the wave equation. We discuss over 100 such systems and relate them to orbits of triplets of commuting operators in the enveloping algebra of the conformal group SO(4,2).

1. INTRODUCTION

This paper is an introduction to the problem of R-separation of variables for the wave equation

$$(\partial_{00}-\partial_{11}-\partial_{22}-\partial_{33})\Psi(x)=0. \tag{1.1}$$

As is well known,[1] the symmetry group of (1.1) is locally isomorphic to the fifteen-parameter group SO(4, 2). In this and subsequent papers we will show explicitly that every known separable coordinate system for (1.1) (as well as some systems which we derive for the first time) corresponds to a three-dimensional commuting subspace of the space of second-order elements in the enveloping algebra of so(4, 2). [We consider the elements of so(4, 2) as first-order differential operators which map solutions of (1.1) into solutions.] If the commuting operators S_1, S_2, S_3 form a basis for such a subspace then the associated R-separable solutions Ψ of (1.1) are characterized by the eigenvalue equations $S_j\Psi=\lambda_j\Psi$, $j=1, 2, 3$, where the eigenvalues λ_j are the separation constants. The group SO(4, 2) acts on the enveloping algebra of so(4, 2) via the adjoint representation and decomposes the set of three-dimensional commuting subspaces of second-order elements into SO(4, 2)-orbits. We regard coordinate systems associated with subspaces on the same orbit as equivalent.

Several earlier papers of the authors and collaborators can be considered as preparation for the problem we tackle directly here. In particular, the Helmholtz,[2] Klein–Gordon,[3] and Euler–Poisson–Darboux,[4] equations are special cases of (1.1) as are the eigenvalue equations for the Laplace operator on the sphere S_3[5] and the hyperboloids of one and two sheets.[6] The same is true for the time-dependent Schrödinger equations for the free-particle, free-fall, harmonic oscillator[7] and hydrogen atom. Our procedure will follow closely the analogous study of the three-variable wave equation in Refs. 3, 8, and 9. The difference consists mainly in the greater complexity of the four-variable case (although a number of computations turn out to be easier in four dimensions than in three). In this paper we proceed as in Ref. 8 and present a group theoretic analysis of (1.1) as well as a rough classification of semisubgroup systems for this equation. Our future (much more detailed) results will be fitted into the framework established here.

In Sec. 2 we define the symmetry algebra so(4, 2) of (1.1) in two distinct bases and construct a Fourier-transform Hilbert space $\mathcal{H}$ as well as a Hilbert space of positive energy solutions of this equation. On $\mathcal{H}$, the elements of so(4, 2) exponentiate to yield a unitary irreducible representation of the covering group SU(2, 2) of the identity component in SO(4, 2). In Sec. 3 we determine explicitly the action of SU(2, 2). Most of the results in this section were obtained in Ref. 10 by another method. [However, Eq. (6.6) for the action of a lightlike special conformal transformation appears to be new.]

The remainder of the paper is concerned with separation of variables. In analogy with Ref. 8 we say that R-separable coordinates $\{u_j\}$ associated with a three-dimensional commuting subspace of symmetry operators are *semisubgroup coordinates* if the subspace has a basis S_1, S_2, S_3 such that $S_1=A^2$ where $A\in$ so(4, 2) and $[A, S_j]=0$, $j=2, 3$. A particular $A\in$ so(4, 2) may correspond to several (or to no) semisubgroup systems. If Ψ satisfies (1.1) and the equation $A\Psi=i\lambda\Psi$, then, since A is a symmetry of (1.1), we can use standard Lie theory and introduce new variables y_0, y_1, y_2, y_3 such that $A=\partial_{y_0}+f(y)$ and $\Psi(y)=r(y)\exp(i\lambda y_0)\Phi_\lambda(y_j)$, where r is a fixed function satisfying $\partial_{y_0}r+fr=0$. Then (1.1) reduces to a second-order partial differential equation (†) for Φ_λ in the three variables y_j. The possible semisubgroup systems A^2, S_2, S_3 thus correspond to the possible coordinate systems such that the reduced equation (†) separates.

In Secs. 4–8 we examine the possible semisubgroup systems for which S_2 and S_3 belong to the symmetry enveloping algebra of (†). They are of seven types corresponding to seven choices for A. Using the notation introduced in Sec. 2, we find the types are:

1] $A=\Gamma_{56}$. In this case (†) is the eigenvalue equation for the Laplace operator on the sphere S_3. There are six coordinate systems.[5]

2] $A=P_0$ and (†) is the Helmholtz equation (5.1) which separates in 11 coordinate systems.[2]

3] $A=P_3$ and (†) is the Klein–Gordon equation (5.4) which separates in 53 orthogonal coordinate systems.[3]

4] $A=D$ and (†) is the eigenvalue equation (5.8) for the Laplace operator on the hyperboloid which separates in 35 coordinate systems.[6]

5] $A = P_0 + P_1$ and (†) is the free particle Schrödinger equation (6.1). There are 17 coordinate systems.[7]

6] $A = \Gamma_{43}$ and (†) is a generalized EPD equation (7.1). The number of coordinate systems for this case has not yet been determined.

7] $A = \Gamma_{12} + \Gamma_{34} - \Gamma_{56}$ and (†) is (8.3). The number of coordinate systems for this case is still unknown.

Cases 6] and 7] will be discussed in detail in future papers. For each case we show how to pass from the Fourier-transform Hilbert space $\mathcal{H}$ to the space of positive energy solutions of (1.1).

Our classification includes all known semisubgroup coordinates for (1.1) with one principal exception. Diagonalization of the operator $A = P_0 + P_1$ in Case 5] does not uniquely determine the variable y_0 which is split off to obtain the reduced equation (†). Thus (†) is not unique in this case. The possibilities for the nonorthogonal coordinates which can arise will be classified in a future paper. [See Ref. 9 where a similar classification was carried out for $(\partial_{tt} - \partial_{xx} - \partial_{yy})\Psi = 0$.] In all other cases there is an identity analogous to (2.24) which uniquely determines the reduced equation. The variable y_0 is still not unique, but new nonorthogonal coordinates so obtained are rather trivial modifications of the coordinates we have listed.

Finally, in Sec. 9 we classify the orbits in so(4, 2) under the adjoint action of SO(4, 2) to see why not every $A \in$ so(4, 2) belongs to a semisubgroup system.

The next two papers in this series will be devoted to an explicit classification of all orthogonal R-separable coordinate systems (semisubgroup or not) whose coordinate surfaces are families of confocal cyclides. The classification will proceed in analogy to that in Ref. 3. However, the number of coordinate systems involved is approximately 300. Later we will classify the nonorthogonal systems. Future work will concern the results in special function theory which follow from separation of variables in (1.1). Equation (1.1) is the most important equation in special function theory and it is no accident that Bateman[11,12] devoted so much energy to its solution by separation of variable methods.

2. SO(4, 2) AND THE WAVE EQUATION

The symmetry algebra of the wave equation

$$(\partial_{00} - \partial_{11} - \partial_{22} - \partial_{33})\Psi(x) = 0, \quad x = (x_0, x_1, x_2, x_3) \tag{2.1}$$

is the set of all linear differential operators

$$L = \sum_{j=0}^{3} a_j(x)\partial_j + b(x)$$

such that $L\Psi$ is a (local) solution of (1.1) whenever Ψ is a (local) solution.

As is well known, the possible symmetry operators L form a 15-dimensional Lie algebra, isomorphic to so(4, 2), where the commutator is the usual Lie bracket.[13] A convenient basis for this model of so(4, 2) is provided by the linear momentum operators

$$P_\alpha = \partial_\alpha, \quad \alpha = 0, 1, 2, 3, \tag{2.2}$$

the generators of homogeneous Lorentz transformations

$$M_{21} = x_2\partial_1 - x_1\partial_2, \quad M_{13} = x_1\partial_3 - x_3\partial_1, \quad M_{32} = x_3\partial_2 - x_2\partial_3,$$
$$M_{01} = x_0\partial_1 + x_1\partial_0, \quad M_{02} = x_0\partial_2 + x_2\partial_0, \quad M_{03} = x_0\partial_3 + x_3\partial_0,$$
$$M_{jk} = -M_{kj}, \quad M_{0j} = M_{j0}, \quad j, k = 1, 2, 3, \tag{2.3}$$

the generator of dilatations

$$D = -(1 + x_0\partial_0 + x_1\partial_1 + x_2\partial_2 + x_3\partial_3), \tag{2.4}$$

and the generators of special conformal transformations

$$K_0 = -2x_0 + (x\cdot x - 2x_0^2)\partial_0 - 2x_0x_1\partial_1 - 2x_0x_2\partial_2 - 2x_0x_3\partial_3,$$
$$K_1 = 2x_1 + (x\cdot x + 2x_1^2)\partial_1 + 2x_1x_0\partial_0 + 2x_1x_2\partial_2 + 2x_1x_3\partial_3,$$
$$K_2 = 2x_2 + (x\cdot x + 2x_2^2)\partial_2 + 2x_2x_0\partial_0 + 2x_2x_1\partial_1 + 2x_2x_3\partial_3,$$
$$K_3 = 2x_3 + (x\cdot x + 2x_3^2)\partial_3 + 2x_3x_0\partial_0 + 2x_3x_1\partial_1 + 2x_3x_2\partial_2, \tag{2.5}$$

where

$$x\cdot y = x_0y_0 - x_1y_1 - x_2y_2 - x_3y_3 = x_0y_0 - \mathbf{x}\cdot\mathbf{y} = \sum_{\alpha\beta=0}^{3} g_{\alpha\beta}x_\alpha y_\beta. \tag{2.6}$$

The commutation relations will follow from relations (2.22) to be derived later.

The symmetry operators can be exponentiated to yield a local Lie transformation group of symmetries of (2.1).[13,14] Indeed, the momentum and Lorentz operators generate the Poincaré group of symmetries

$$\Psi(x) \to \Psi(\Lambda^{-1}(x-a)), \quad a = (a_0, a_1, a_2, a_3), \quad \Lambda \in SO(1,3), \tag{2.7}$$

while the dilatation operators generate

$$\exp(\lambda D)\Psi(x) = \exp(-\lambda)\Psi(\exp(-\lambda)x), \quad \lambda \in R \tag{2.8}$$

and the K_α generate the special conformal tranformations

$$\exp(a_0K_0 + a_1K_1 + a_2K_2 + a_3K_3)\Psi(x) = [1 + 2x\cdot a + (a\cdot a)(x\cdot x)]^{-1}\Psi\left(\frac{x + a(x\cdot x)}{1 + 2x\cdot a + (a\cdot a)(x\cdot x)}\right). \tag{2.9}$$

We shall also consider the inversion operator

$$R\Psi(x) = (x\cdot x)^{-1}\Psi(-x/x\cdot x) \tag{2.10}$$

which is a symmetry not generated by the local Lie symmetries (2.2)–(2.5).

As is well known from quantum field theory,[1,13] by formally taking the Fourier transform in the variables x_α we can express the positive energy solutions of (2.1) in the form

$$\Psi(x) = \frac{1}{(2\pi)^{3/2}} \iiint_{-\infty}^{\infty} \exp(ik\cdot x) f(\mathbf{k})\, d\mu(\mathbf{k}), \tag{2.11}$$
$$k_0 = (k_1^2 + k_2^2 + k_3^2)^{1/2}, \quad d\mu(\mathbf{k}) = dk_1\, dk_2\, dk_3/k_0.$$

Let $\mathcal{H}$ be the Hilbert space of all complex Lebesgue measurable functions $f(\mathbf{k})$ such that

$$\iiint_{-\infty}^{\infty} |f|^2\, d\mu(\mathbf{k}) < \infty, \tag{2.12}$$

and with inner product

$$\langle f, g\rangle = \int\int\int f\bar{g}\, d\mu(\mathbf{k}), \quad f, g \in \mathcal{H}. \tag{2.13}$$

As is well known,[1,13] the functions Ψ, Φ related to f, g, respectively, by (2.11) satisfy

$$\langle \Psi, \Phi\rangle \equiv \langle f, g\rangle = i \int\int_{x_0=t}\int \Psi(x)\partial_0\bar{\Phi}(x)\, dx_1\, dx_2\, dx_3 = -i\int\int_{x_0=t}\int(\partial_0\Psi(x))\bar{\Phi}(x)\, dx_1\, dx_2\, dx_3, \tag{2.14}$$

independent of t. [Note that (2.14) is easily derived from (2.13) for f, g belonging to the dense subspace $\mathcal{D}$ of $\mathcal{H}$ consisting of C^∞ functions with compact support bounded away from $\mathbf{k}=0$, and then considering the closure of $\mathcal{D}$. For $f\in\mathcal{H}$ the corresponding Ψ is a solution of (2.1) in the sense of distribution theory.]

The operators (2.2)–(2.5) acting on solutions of (2.1) induce, via (2.11), corresponding operators on $\mathcal{H}$,

$$P_0 = ik_0, \quad P_j = -ik_j, \quad j = 1, 2, 3, \tag{2.15}$$

$$M_{21} = k_2\partial_{k_1} - k_1\partial_{k_2}, \quad M_{13} = k_1\partial_{k_3} - k_3\partial_{k_1}, \quad M_{32} = k_2\partial_{k_3} - k_3\partial_{k_2}, \quad M_{01} = k_0\partial_{k_1}, \quad M_{02} = k_0\partial_{k_2}, \quad M_{03} = k_0\partial_{k_3} \tag{2.16}$$

$$D = 1 + k_1\partial_{k_1} + k_2\partial_{k_2} + k_3\partial_{k_3}, \tag{2.17}$$

$$\begin{aligned} K_0 &= ik_0(\partial_{k_1k_1} + \partial_{k_2k_2} + \partial_{k_3k_3}),\\ K_1 &= i(k_1\partial_{k_1k_1} - k_1\partial_{k_2k_2} - k_1\partial_{k_3k_3} + 2k_2\partial_{k_1k_2} + 2k_3\partial_{k_1k_3} + 2\partial_{k_1}),\\ K_2 &= i(k_2\partial_{k_2k_2} - k_2\partial_{k_1k_1} - k_2\partial_{k_3k_3} + 2k_1\partial_{k_2k_1} + 2k_3\partial_{k_2k_3} + 2\partial_{k_2}),\\ K_3 &= i(k_3\partial_{k_3k_3} - k_3\partial_{k_1k_1} - k_3\partial_{k_2k_2} + 2k_1\partial_{k_3k_1} + 2k_2\partial_{k_3k_2} + 2\partial_{k_3}). \end{aligned} \tag{2.18}$$

In Ref. 13 it is shown that $\mathcal{H}$ is invariant under R and

$$Rf(\mathbf{k}) = \frac{1}{4\pi}\int\int\int J_0((2k\cdot l)^{1/2}) f(\mathbf{l})\, d\mu(\mathbf{l}), \tag{2.19}$$

where $J_0(z)$ is a Bessel function. Furthermore, $R^2 = E$ (the identity operator on $\mathcal{H}$) and R is a unitary self-adjoint operator on this space. Also we have the relations

$$RK_\alpha R^{-1} = -P_\alpha, \quad RDR^{-1} = -D, \quad RM_{\alpha\beta}R^{-1} = M_{\alpha\beta}. \tag{2.20}$$

Note: There is a minus sign error in the corresponding expression in Ref. 8 which propagates through several equations. The error is corrected in Ref. 15.

Now we introduce a new basis for the symmetry algebra of (2.1) which makes apparent the isomorphism with so(4, 2). We define so(4, 2) as the 15-dimensional Lie algebra of 6×6 real matrices $\mathcal{A}$ such that $\mathcal{A}G + G\mathcal{A}^t = 0$, where 0 is the zero matrix and

$$G = \begin{bmatrix} 1 & & & & & 0\\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & -1 & \\ 0 & & & & & -1 \end{bmatrix} = (G_{\alpha\beta}).$$

Let $\mathcal{E}_{ij}$ be the 6×6 matrix with a 1 in row i, column j, and zeros everywhere else. It is straightforward to check that the matrices

$$\begin{aligned} \Gamma_{ab} &= \mathcal{E}_{ab} - \mathcal{E}_{ba} = -\Gamma_{ba}, \quad 1\le a < b \le 4,\\ \Gamma_{aA} &= \mathcal{E}_{aA} + \mathcal{E}_{Aa} = -\Gamma_{Aa}, \quad 1 \le a \le 4, \quad A = 5, 6,\\ \Gamma_{56} &= \mathcal{E}_{65} - \mathcal{E}_{56} = -\Gamma_{65} \end{aligned} \tag{2.21}$$

form a basis for so(4, 2) with commutation relations

$$[\Gamma_{\alpha\beta}, \Gamma_{\gamma\varphi}] = G_{\beta\gamma}\Gamma_{\alpha\varphi} + G_{\alpha\varphi}\Gamma_{\beta\gamma} - G_{\varphi\beta}\Gamma_{\alpha\gamma} - G_{\alpha\gamma}\Gamma_{\beta\varphi}. \tag{2.22}$$

This basis can be related to the operators (2.2)–(2.5) as follows:

$$\begin{aligned} &P_0 = \Gamma_{15} + \Gamma_{56}, \quad P_1 = \Gamma_{12} + \Gamma_{26}, \quad P_2 = \Gamma_{13} + \Gamma_{36},\\ &P_3 = \Gamma_{14} + \Gamma_{46}, \quad K_0 = \Gamma_{15} - \Gamma_{56}, \quad K_1 = \Gamma_{12} - \Gamma_{26},\\ &K_2 = \Gamma_{13} - \Gamma_{36}, \quad K_3 = \Gamma_{14} - \Gamma_{46}, \quad M_{21} = \Gamma_{32},\\ &M_{13} = \Gamma_{24}, \quad M_{32} = \Gamma_{43}, \quad M_{01} = \Gamma_{25},\\ &M_{02} = \Gamma_{35}, \quad M_{03} = \Gamma_{45}, \quad D = \Gamma_{16}. \end{aligned} \tag{2.23}$$

(That is, the appropriate commutation relations are satisfied if these identifications are made.)

For our models of so(4, 2), [acting on the solution space of (2.1) or on the Hilbert space $\mathcal{H}$] we have the identities

$$\begin{aligned} &\text{(i)}\ P_0^2 - P_1^2 - P_2^2 - P_3^2 = K_0^2 - K_1^2 - K_2^2 - K_3^2 = 0,\\ &\text{(ii)}\ \Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{14}^2 + \Gamma_{23}^2 + \Gamma_{24}^2 + \Gamma_{34}^2 = \Gamma_{56}^2 + 1,\\ &\text{(iii)}\ \Gamma_{32}^2 + \Gamma_{24}^2 + \Gamma_{43}^2 - \Gamma_{25}^2 - \Gamma_{35}^2 - \Gamma_{45}^2 = -D^2 + 1,\\ &\text{(iv)}\ \Gamma_{12}^2 + \Gamma_{56}^2 - \Gamma_{15}^2 - \Gamma_{16}^2 - \Gamma_{26}^2 - \Gamma_{25}^2 = \Gamma_{43}^2 + 1. \end{aligned} \tag{2.24}$$

If $\{\Psi_\alpha(x)\}$ is an orthonormal (ON) basis for the Hilbert space of positive energy solutions of (2.1) then (in the sense of distributions)

$$\sum_\alpha \bar{\Psi}_\alpha(x)\Psi_\alpha(x') = \Delta_+(x - x') = \frac{1}{(2\pi)^3}\int\int\int \times \exp[ik\cdot(x - x')]\, d\mu(\mathbf{k}), \tag{2.25}$$

where the distribution Δ_+ is given explictly by[16]

$$\Delta_+(x) = \frac{1}{(2\pi)^2}\frac{1}{r^2 - t^2} + \frac{i}{4\pi r}[\delta(r + t) - \delta(r - t)], \quad r = (x_1^2 + x_2^2 + x_3^2)^{1/2}, \quad t = x_0. \tag{2.26}$$

Note that

$$\Psi(x) = \langle\Psi, \Delta_+(x' - x)\rangle, \tag{2.27}$$

where the integration is carried out over $\mathbf{x}'$.

3. THE ACTION OF THE CONFORMAL GROUP

As is well known, the representation of so(4, 2) on $\mathcal{H}$ defined by the operators (2.15)–(2.18) can be extended to a unitary irreducible representation of the covering group SU(2, 2) of the identity component of SO(4, 2).[10] The maximal compact connected subgroup of SU(2, 2) is SO(4)×SO(2), where SO(4) is generated by the Lie algebra operators Γ_{ij}, $1\le i < j \le 4$ and SO(2) by Γ_{56}. We will explicitly determine the action of this subgroup on $\mathcal{H}$ as well as the actions of other interesting subgroups.

The operators $M_{\alpha\beta}$ generate a subgroup of SU(2, 2) isomorphic to the homogeneous Lorentz group SO(3, 1). The action of this subgroup is determined by

$$T(0)f(\mathbf{k}) = f(\mathbf{k}0), \quad 0 \in \mathrm{SO}(3),$$
$$\exp(aM_{01})f(\mathbf{k}) = f(k_1(a), k_2, k_3), \tag{3.1}$$
$$k_1(a) = k_1 \cosh a + k_0 \sinh a,$$

where M_{21}, M_{13}, M_{32} generate SO(3), and the results for M_{02}, M_{03} follow easily from that for M_{01}. The P_α generate a translation subgroup of operators

$$\exp(\Sigma\, a_\alpha P_\alpha)f(\mathbf{k}) = \exp(ia\cdot k)f(\mathbf{k}). \tag{3.2}$$

The unitary operators $\exp(\Sigma a_\alpha K_\alpha)$ are more difficult to compute. Since the SO(3, 1) subgroup transforms a via the adjoint action, one has to consider only three distinct cases : (1) $a=(a_0, 0, 0, 0)$, $a_0 \neq 0$, timelike; (2) $a=(0, a_1, 0, 0)$, $a_1 \neq 0$, spacelike; (3) $a=(a_1, a_1, 0, 0)$, $a_1 \neq 0$, lightlike. All other cases can be obtained by composing these three operators with the operators (3.1), (3.2).

Starting with the timelike case, we introduce the basis $\{f_{l_j}\}$ for H consisting of generalized eigenvectors of the commuting operators P_α,

$$\begin{aligned} &f_{l_j}(\mathbf{k}) = \delta(k_1 - l_1)\delta(k_2 - l_2)\delta(k_3 - l_3)k_0, \quad -\infty < l_j < \infty \\ &P_h f_{l_j} = -il_h f_{l_j}, \quad h = 1, 2, 3, \quad P_0 f_{l_j} = il_0 f_{l_j}, \\ &\langle f_{l_j}, f_{l'_j}\rangle = \delta(l_1 - l_1')\delta(l_2 - l_2')\delta(l_3 - l_3')l_0, \\ &l_0 = (l_1^2 + l_2^2 + l_3^2)^{1/2}. \end{aligned} \tag{3.3}$$

It follows that the functions $g_{l_j} = Rf_{l_j}$,

$$g_{l_j}(\mathbf{k}) = \frac{1}{4\pi} J_0((2k\cdot l)^{1/2}), \tag{3.4}$$

form a basis for H consisting of generalized eigenvectors of the commuting operators K_α,

$$\begin{aligned} &K_h g_{l_j} = il_h g_{l_j}, \quad K_0 g_{l_j} = -il_0 g_{l_j}, \\ &\langle g_{l_j}, g_{l'_j}\rangle = \delta(l_1 - l_1')\delta(l_2 - l_2')\delta(l_3 - l_3')l_0, \end{aligned} \tag{3.5}$$

as follows from the fact that R is unitary.

Now we have for $f \in H$ that

$$\exp(aK_0)f(\mathbf{s}) = \int\int\int G(a, \mathbf{l}, \mathbf{s})f(\mathbf{l})\, d\mu(\mathbf{l}), \tag{3.6}$$

$$\begin{aligned} G(a, \mathbf{l}, \mathbf{s}) &= \langle \exp(aK_0)f_{l_j}, f_{s_j}\rangle \\ &= \langle R\exp(-aP_0)Rf_{l_j}, f_{s_j}\rangle = \langle \exp(-aP_0)g_{l_j}, g_{s_j}\rangle \\ &= \frac{1}{16\pi^2}\iiint \exp(-iak_0)J_0((2k\cdot l)^{1/2})\, d\mu(\mathbf{k}) \\ &= \frac{-1}{4\pi|a|}\exp[i(l_0 + s_0)a^{-1}] \\ &\quad \times J_0(a^{-1}[2(s_0l_0 + s_1l_1 + s_2l_2 + s_3l_3)]^{1/2}), \\ a \neq 0. \end{aligned} \tag{3.7}$$

To compute the action of $\exp(aK_1)$, we choose a basis of eigenfunctions of the commuting operators P_0, P_1, M_{32},

$$\begin{aligned} &h_{\rho\lambda m}(\mathbf{k}) = (2\pi)^{-1/2}\delta(k_0 - \rho)\delta(k_1 - \lambda)\exp(im\theta), \\ &0 \leq \rho, \quad -\rho \leq \lambda \leq \rho, \quad m = 0, \pm 1, \ldots, \\ &d\mu(\mathbf{k}) = dk_0\, dk_1\, d\theta, \quad k_2 = (k_0^2 - k_1^2)^{1/2}\sin\theta, \end{aligned} \tag{3.8}$$

$$k_3 = (k_0^2 - k_1^2)^{1/2}\cos\theta.$$

Here,

$$\begin{aligned} &P_0 h = i\rho h, \quad P_1 h = -i\lambda h, \quad M_{32}h = imh, \\ &\langle h_{\rho\lambda m}, h_{\rho'\lambda' m'}\rangle = \delta(\rho - \rho')\delta(\lambda - \lambda')\delta_{mm'}. \end{aligned} \tag{3.9}$$

It follows that the functions $g_{\rho\lambda m} = Rh_{\rho\lambda m}$,

$$\begin{aligned} g_{\rho\lambda m}(\mathbf{k}) &= \frac{\exp(im\varphi)}{(8\pi)^{1/2}}\; J_m[(rs\exp(\alpha - \beta))^{1/2}] \\ &\quad \times J_m[(rs\exp(\beta - \alpha))^{1/2}] \end{aligned} \tag{3.10}$$

$$\begin{aligned} &\rho = r\cosh\alpha, \quad \lambda = r\sinh\alpha, \\ &k_0 = s\cosh\beta, \quad k_1 = s\sinh\beta, \\ &d\mu(\mathbf{k}) = s\, ds\, d\beta\, d\theta, \quad s \geq 0, \quad -\infty < \beta < \infty, \quad -\pi \leq \theta \leq \pi, \end{aligned}$$

form a basis for H and satisfy relations

$$\begin{aligned} &K_0 g = -i\rho g, \quad K_1 g = i\lambda g, \quad M_{32}g = img, \\ &\langle g_{\rho\lambda m}, g_{\rho'\lambda' m'}\rangle = \delta(\rho - \rho')\delta(\lambda - \lambda')\delta_{mm'}. \end{aligned} \tag{3.11}$$

Now,

$$\exp(aK_1)f(\mathbf{k}) = \int H(a, \mathbf{l}, \mathbf{k})f(\mathbf{l})\, d\mu(\mathbf{l}), \tag{3.12}$$

where

$$\begin{aligned} H(a, \mathbf{l}, \mathbf{k}) &= \langle \exp(aK_1)f_{\mathbf{l}}, f_{\mathbf{k}}\rangle \\ &= \sum_{m=-\infty}^{\infty}\int_0^\infty\int_{-\rho}^{\rho}\langle f_{\mathbf{l}}, \exp(-aK_1)g_{\rho\lambda m}\rangle\langle g_{\rho\lambda m}, f_{\mathbf{k}}\rangle\, d\lambda\, d\rho \\ &= \sum_{m=-\infty}^{\infty}\int_0^\infty\int_{-\rho}^{\rho}\exp(i\lambda a)\,\overline{g_{\rho\lambda m}}(\mathbf{l})g_{\rho\lambda m}(\mathbf{k})\, d\lambda\, d\rho \\ &= \frac{1}{4\pi a^2}\exp[i(k_1 + l_1)/a] \\ &\quad \times J_0(a^{-1}\sqrt{2}\,(k_0l_0 + k_1l_1 - k_2l_2 - k_3l_3)^{1/2}). \end{aligned} \tag{3.13}$$

We will compute $\exp a(K_0 + K_1)$ in Sec. 6.

The dilatation operator generates the symmetries

$$\exp(aD)f(\mathbf{k}) = e^a f(e^a\mathbf{k}). \tag{3.14}$$

Using these results we can exponentiate the compact generator $\Gamma_{56} = \frac{1}{2}(P_0 - K_0)$. The operators P_0, D, K_0 generate a SL(2, R) subgroup of SU(2, 2) and we have

$$\begin{aligned} \exp(2\theta\Gamma_{56}) &= \exp(\tan\theta P_0)\exp(-\sin\theta\cos\theta K_0) \\ &\quad \times \exp(-2\ln\cos\theta D) \end{aligned} \tag{3.15}$$

on SL(2, R). Evaluating the right-hand side of this expression we find

$$\begin{aligned} &\exp(2\theta\Gamma_{56})f(\mathbf{k}) \\ &\quad = -\frac{|\cot\theta|}{4\pi}\int \exp[-i(l_0 + k_0)\cot\theta]J_0(\csc\theta[2(k_0l_0 + k_1l_1 \\ &\qquad + k_2l_2 + k_3l_3)]^{1/2})f(\mathbf{l})\, d\mu(\mathbf{l}), \quad \theta \neq n\pi. \end{aligned} \tag{3.16}$$

The operators P_1, D, K_1 generate another SL(2, R) subgroup of SU(2, 2) and there follows the relation

$$\exp(2\theta\Gamma_{12}) = \exp(\tan\theta P_1)\exp(\sin\theta\cos\theta K_1)\exp(-2\ln\cos\theta D)$$

or

$$\begin{aligned} &\exp(2\theta\Gamma_{12}) \\ &\quad = \frac{1}{4\pi\sin^2\theta}\int \exp[i(k_1 + l_1)\cot\theta]J_0(\csc\theta[2(k_0l_0 + k_1l_1 \\ &\qquad - k_2l_2 - k_3l_3)]^{1/2})f(\mathbf{l})\, d\mu(\mathbf{l}), \quad \theta \neq n\pi. \end{aligned} \tag{3.17}$$

4. DIAGONALIZATION OF Γ_{56}

The restriction of the unitary irreducible representation T of SU(2, 2) on $\mathcal{H}$, discussed above, to the compact subgroup SO(4) decomposes into a direct sum of SO(4)-irreducible representation D_F, $\dim D_F=(2F+1)^2$. A basis of eigenvectors for the commuting operators $\Gamma_{56}, \Gamma_{43}, \Gamma_{12}$ can be used to exhibit this decomposition,

$$\Gamma_{56}f=i\lambda f,\quad \Gamma_{43}f=imf,\quad \Gamma_{12}f=i\rho f,$$
$$-i\Gamma_{56}=\tfrac{1}{2}k_0(-\Delta_3+1),\quad \Gamma_{43}=k_3\partial_{k_2}-k_2\partial_{k_3},$$
$$-i\Gamma_{12}=\frac{k_1}{2}\left(\partial_{k_1k_1}-\partial_{k_2k_2}-\partial_{k_3k_3}-1\right)+k_2\partial_{k_1k_2}+k_3\partial_{k_1k_3}+\partial_{k_1}. \tag{4.1}$$

Setting

$$k_1=(\xi^2-\eta^2)/2,\quad k_2=\xi\eta\sin\theta,$$
$$k_3=\xi\eta\cos\theta,\quad d\mu(\mathbf{k})=2\xi\eta\,d\xi\,d\eta\,d\theta,$$

we find that the ON basis consists of eigenfunctions

$$f_{Fab}(\xi,\eta,\theta)=\left(\frac{(F-a)!\,(F-b)!}{\pi(F+b)!\,(F+a)!}\right)^{1/2}\exp[-(\xi^2+\eta^2)/2](\xi\eta)^{a+b}$$
$$\times L_{F-a}^{(a+b)}(\xi^2)L_{F-b}^{(a+b)}(\eta^2)\exp[i(a+b)\theta], \tag{4.2}$$

$$\lambda=2F+1,\quad m=a+b,\quad \rho=a-b,$$
$$2F=0,1,2,\cdots,\quad a,b=F,F-1,\ldots,-F.$$

[Here $L_n^{(\alpha)}(x)$ is a generalized Laguerre polynomial.]

The $(2F+1)^2$ functions f_{Fab} for fixed F form a basis for D_F, so $T\,|\,\mathrm{SO}(4)\approx\sum_{2F=0}^{\infty}\oplus D_F$.

The known recurrence relations for Laguerre polynomials imply

$$\begin{aligned}2\Gamma_{16}f_{Fab}={}&[(F-a+1)(F+b+1)]^{1/2}f_{F+1/2,a-1/2,b+1/2}-[(F-a)(F+b)]^{1/2}f_{F+1/2,a+1/2,b-1/2}\\&+[(F+a+1)(F-b+1)]^{1/2}f_{F+1/2,a+1/2,b-1/2}-[(F+a)(F-b)]^{1/2}f_{F-1/2,a-1/2,b+1/2},\\2iP_3f_{F,a,b}={}&[(F+a+1)(F+b+1)]^{1/2}f_{F+1/2,a+1/2,b+1/2}-[(F-b)(F+b+1)]^{1/2}f_{F,a,b+1}-[(F-a)(F+a+1)]^{1/2}f_{F,a+1,b}\\&+[(F-a)(F-b)]^{1/2}f_{F-1/2,a+1/2,b+1/2}+[(F+a)(F+b)]^{1/2}f_{F-1/2,a-1/2,b-1/2}-[(F+b)(F-b+1)]^{1/2}f_{F,a,b-1}\\&-[(F+a)(F-a+1)]^{1/2}f_{F,a-1,b}+[(F-a+1)(F-b+1)]^{1/2}f_{F+1/2,a-1/2,b-1/2}.\end{aligned} \tag{4.3}$$

Expressions (4.1), (4.3), and the commutation relations (2.22) suffice to determine the action of any $\Gamma_{\alpha\beta}$ on f_{Fab}.

There is a close connection between the quantum Kepler problem in three-space,

$$H\Phi=E\Phi,\quad H=-\partial_{x_1x_1}-\partial_{x_2x_2}-\partial_{x_3x_3}+e/r,$$
$$r=(x_1^2+x_2^2+x_3^2)^{1/2},\quad \int\int\int_{R^3}|\Phi|^2dx_1\,dx_2\,dx_3<\infty, \tag{4.4}$$

and the equation $\Gamma_{56}f=i\lambda f$. Indeed the equations can be identified if we set $k_j=x_j\sqrt{-E}$, $E=-e^2/4\lambda^2$. (Although the eigenvalue problems are defined on Hilbert spaces with different inner products, it follows from the Virial theorem, Ref. 17, p. 51, that if E belongs to the point spectrum of H, and Φ is the corresponding eigenvector, then Φ has finite norm in $\mathcal{H}$. Conversely, if f is an eigenvector of Γ_{56} then $\int\int\int_{R^3}|f|^2dx_1\,dx_2\,dx_3<\infty$ and f corresponds to an energy eigenvalue E in the point spectrum of H.) Since the eigenvalues of $-i\Gamma_{56}$ are $\lambda=2F+1$, $2F=0,1,2\cdots$, it follows that the point spectrum of H consists of the eigenvalues $E=-e^2/4(2F+1)^2$. (Similarly, the continuous spectrum of H is related to the operator Γ_{15}.)

Applying the transformation (2.11) to the basis $\{f_{Fab}\}$ we can determine the corresponding ON basis $\{\Psi_{Fab}\}$ of positive energy solutions of (2.1),

$$\Psi_{Fab}(x)=(2\pi)^{-3/2}\int\int\int_{-\infty}^{\infty}\exp(ik\cdot x)f_{Fab}(\mathbf{k})\,d\mu(\mathbf{k})$$
$$=\pi^{-1}\sqrt{2}\exp[i(a+b)(\varphi-\pi/2)]\left(\frac{(F-a)!\,(F-b)!}{(F+a)!\,(F+b)!}\right)^{1/2}$$
$$\times\int_0^\infty\int_0^\infty\exp[\tfrac{1}{2}(ix_0-1)(\xi^2+\eta^2)-\frac{i}{2}x_1(\xi^2-\eta^2)](\xi\eta)^{a+b+1}$$
$$\times L_{F-a}^{(a+b)}(\xi^2)L_{F-b}^{(a+b)}(\eta^2)J_{a+b}(r\xi\eta)\,d\xi\,d\eta, \tag{4.5}$$
$$x_2=r\sin\varphi,\quad x_3=r\cos\varphi.$$

In terms of the coordinates

$$x_0=\frac{\sin\psi}{y_0-\cos\psi},\quad x_1=\frac{y_1}{y_0-\cos\psi},\quad x_2=\frac{y_2}{y_0-\cos\psi},$$
$$x_3=\frac{y_3}{y_0-\cos\psi},\quad y_0=\cos\alpha\cos\sigma,\quad y_1=\cos\alpha\sin\sigma,$$
$$y_2=\sin\alpha\sin\varphi,\quad y_3=\sin\alpha\cos\varphi, \tag{4.6}$$

we have

$$\Psi_{Fab}(x)=(\cos\alpha\cos\sigma-\cos\psi)\exp[i(m\varphi+\rho\sigma-(2F+1)\Psi)]$$
$$\times(2\pi)^{-1/2}(-1)^{F-a+1}\left(\frac{(F+a)!\,(F+b)!}{\pi(F-a)!\,(F-b)!}\right)^{1/2}(\sin\alpha)^{a+b}$$
$$\times\frac{(\cos\alpha)^{2F-a-b}}{\Gamma(a+b+1)}\,{}_2F_1\left(\begin{matrix}b-F,a-F\\a+b+1\end{matrix}\,\middle|\,-\tan^2\alpha\right). \tag{4.7}$$

Indeed, direct computation shows

$$\Gamma_{56}=-\partial_\psi+\sin\psi(\cos\alpha\cos\sigma-\cos\psi)^{-1},\quad \Gamma_{43}=\partial_\varphi,$$
$$\Gamma_{12}=\partial_\sigma+\cos\alpha\sin\sigma(\cos\alpha\cos\sigma-\cos\psi)^{-1} \tag{4.8}$$

on the solution space of (2.1). Hence

$$\Psi_{Fab}(x)=(\cos\alpha\cos\sigma-\cos\psi)\exp\{i[m\varphi+\rho\sigma-(2F+1)\psi]\}g(\alpha)$$

and substitution into (2.1) yields R-separation of variables. It follows from this that $g(\alpha)$ must be a multiple of $(\sin\alpha)^{a+b}(\cos\alpha)^{2F-a-b}\,{}_2F_1\,(b-F,a-F;a+b+1;-\tan^2\alpha)$. The constant is determined by explicitly computing (4.6) for convenient values of the variables.

There is another model of this irreducible representation which is very convenient for computations involving eigenfunctions of Γ_{56}, Γ_{43}, and Γ_{12}. The representation space $\mathcal{J}$ consists of functions h of three complex variables u, v, w such that $h(-u, -v, -w) = h(u, v, w)$. More precisely, $\mathcal{J}$ is the complex Hilbert space with ON basis

$$h_{Fab} = \left[\frac{(F+b)!\,(F-b)!}{(F+a)!\,(F-a)!}\right]^{1/2} u^{F+a} v^{F+b} w^{2F-a-b},$$

$$a, b = F, F-1, \ldots, -F, \quad 2F = 0, 1, \cdots. \tag{4.9}$$

The operators

$$\Gamma_{56} = i(u\partial_u + v\partial_v + w\partial_w + 1), \quad \Gamma_{43} = \frac{i}{2}(u\partial_u + v\partial_v - w\partial_w),$$

$$\Gamma_{12} = i(u\partial_u - v\partial_v),$$

$$2\Gamma_{16} = vw(v\partial_v + 1) - \frac{v^{-1}w^{-1}}{2}(-u\partial_u + v\partial_v + w\partial_w) \tag{4.10}$$

$$+ \frac{uw}{2}(u\partial_u - v\partial_v + w\partial_w + 2) - u^{-1}w^{-1}(v\partial_v),$$

$$2iP_3 = (uv - vw^{-1})(v\partial_v + 1) - \frac{1}{2}\left(\frac{u}{w} + \frac{v}{w^2}\right)(-u\partial_u + v\partial_v + w\partial_w)$$

$$+ (u^{-1}v^{-1} - u^{-1}w)(u\partial_u) + \tfrac{1}{2}(w^2 - v^{-1}w)$$

$$\times (u\partial_u - v\partial_v + w\partial_w + 2),$$

acting on this basis satisfy relations (4.1), (4.3) and completely determine the action of so(4, 2). The three variable model appears to be the simplest in which to compute matrix elements of the SU(2, 2) operators with respect to the $\{\Gamma_{56}, \Gamma_{43}, \Gamma_{12}\}$ eigenbasis. For some examples of matrix elements computed with this model see Ref. 18. (Indeed in this reference it is shown that one can choose another basis for the complexification of so(4, 2) for which the differential operators take a much simpler form. The action of the Lie algebra on the basis $\{f_{Fab}\}$ corresponds exactly to the 12 known differential recurrence relations for the functions ${}_2F_1$.)

We can see from (4.6) and (4.8) how one characterizes those solutions Ψ of (2.1) such that $\Gamma_{56}\Psi = i\lambda\Psi$. It follows from these expressions that $\Psi = (y_0 - \cos\psi)\Phi(y)\exp(-i\lambda\psi)$ where $y = (y_0, y_1, y_2, y_3)$ is an element of the sphere $S_3: y_0^2 + y_1^2 + y_2^2 + y_3^2 = 1$. Moreover, Eq. (2.1) for Ψ reduces to the eigenvalue equation

$$(\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{14}^2 + \Gamma_{23}^2 + \Gamma_{24}^2 + \Gamma_{34}^2)\Phi = (1 - \lambda^2)\Phi. \tag{4.11}$$

Here (4.11) is the eigenvalue equation for the Laplace—Beltrami operator on S_3. Indeed, the symmetry algebra of this equation is so(4) with basis $\{\Gamma_{ij},\ 1 \le i < j \le 4\}$. The operators

$$\Gamma_{12} = y_0\partial_{y_1} - y_1\partial_{y_0}, \quad \Gamma_{34} = y_2\partial_{y_3} - y_3\partial_{y_2},$$
$$\Gamma_{23} = y_1\partial_{y_2} - y_2\partial_{y_1} \tag{4.12}$$

acting on S_3 generate this symmetry algebra.

Thus, the effect of diagonalizing Γ_{56} is to reduce the separation of variables problem for (2.1) to the corresponding problem for (4.11). The latter equation was studied in Ref. 5 where it was shown that (4.11) separates in exactly six orthogonal coordinate systems, each corresponding to a commuting pair of symmetric second order symmetry operators from the enveloping algebra of so(4). Briefly, the list is

1] $\Gamma_{43}^2, \quad \Gamma_{12}^2$ (cylindrical)

2] $\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{23}^2, \quad \Gamma_{12}^2$ (spherical)

3] $\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{23}^2, \quad \Gamma_{23}^2 + k^2\Gamma_{13}^2, \quad 0 < k < 1$ (spheroelliptic)

4] $\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{23}^2 + a(\Gamma_{12}^2 - \Gamma_{34}^2), \quad \Gamma_{12}^2, \quad -\infty < a < -1$ (elliptic cylindrical, Type I)

5] Same as 4] with $-1 < a < 0$ (elliptic cylindrical, Type II)

6] $\Gamma_{23}^2 - \Gamma_{14}^2 + \left(\frac{a-b-1}{1-a-b}\right)\left(\Gamma_{13}^2 - \Gamma_{24}^2\right) + \left(\frac{b-a-1}{1-a-b}\right) \times \left(\Gamma_{12}^2 - \Gamma_{34}^2\right)$,
$2b(1-a)(\Gamma_{13}^2 + \Gamma_{24}^2) + 2a(1-b)(\Gamma_{12}^2 + \Gamma_{34}^2) + 2b(1-a)(\Gamma_{13}^2 - \Gamma_{24}^2) + 2a(1-b)\left(\frac{a-b-1}{1-a-b}\right)(\Gamma_{12}^2 - \Gamma_{34}^2)$ (ellipsoidal).

The names of the separable coordinates are listed in parentheses. These systems are studied in detail in Ref. 5 and related to the hydrogen atom eigenvalue equation.

5. DIAGONALIZATION OF P_0, $P_3{}'$, AND D

Next we search for coordinate systems permitting separation of variables in (2.1) such that the corresponding basis functions Ψ are eigenfunctions of $P_0: P_0\Psi = i\omega\Psi$. In this case we can set $\Psi(x) = \exp(i\omega x_0)\Phi(x_1, x_2, x_3)$ where

$$(\partial_{11} + \partial_{22} + \partial_{33} + \omega^2)\Phi = 0. \tag{5.1}$$

It follows that the reduced equation for the eigenfunctions is the Helmholtz equation. The symmetry algebra for (5.1) is $\mathcal{E}(3)$, the Lie algebra of the Euclidean group in three-space. A basis for $\mathcal{E}(3)$ is $\{P_1, P_2, P_3, M_{21}, M_{13}, M_{32}\}$. It is well known[2,19] that this equation separates in exactly 11 orthogonal coordinate systems, each system corresponding to a pair of commuting second order symmetric operators in the enveloping algebra of $\mathcal{E}(3)$. Briefly, the separable systems are

1] $P_2^2, \quad P_3^2$ (Cartesian),

2] $M_{21}^2, \quad P_3^2$ (cylindrical),

3] $\{M_{21}, P_2\}, \quad P_3^2$ (parabolic cylindrical),

4] $M_{21}^2 + d^2P_1^2, \quad P_3^2, \quad d > 0$ (elliptic cylindrical),

5] $M_{21}^2 + M_{13}^2 + M_{32}^2, \quad M_{21}^2$ (spherical),

6] $M_{21}^2 + M_{13}^2 + M_{32}^2 - a^2(P_1^2 + P_2^2), \quad M_{21}^2, \quad a > 0$ (prolate spheroidal),

7] $M_{21}^2 + M_{13}^2 + M_{32}^2 + a^2(P_1^2 + P_2^2), \quad M_{21}^2, \quad a > 0$ (oblate spheroidal),

8] $\{M_{32}, P_2\} - \{M_{13}, P_1\}, \quad M_{21}^2$ (parabolic),

9] $M_{21}^2 - c^2P_3^2 + c(\{M_{13}, P_1\} + \{M_{32}, P_2\})$,
$c(P_2^2 - P_1^2) + \{M_{13}, P_1\} - \{M_{32}, P_2\}$ (paraboloidal),

10] $P_1^2 + aP_2^2 + (a+1)P_3^2 + M_{21}^2 + M_{13}^2 + M_{32}^2$,
$M_{13}^2 + a(M_{32}^2 + P_3^2), \quad a > 1$ (ellipsoidal),

11] $M_{21}^2 + M_{13}^2 + M_{32}^2, \quad M_{32}^2 + bM_{13}^2, \quad 1 > b > 0$ (conical). Here $\{A, B\} = AB + BA$.

On $\mathcal{H}$ the condition $P_0 f = i\omega f$ implies $f(\mathbf{k}) = \delta(k_0 - \omega) g_\omega(\hat{\mathbf{k}})$ where $\omega > 0$ and $\mathbf{k} = (k_1, k_2, k_3)$ ranges over the unit sphere $S_2 : \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 1$, $\mathbf{k} = \omega\hat{\mathbf{k}}$. To determine the functions g_ω one uses the Hilbert space $L_2(S_2)$ of square integrable functions on S_2 [with measure $d\Omega(\hat{\mathbf{k}}) = d\hat{k}_1 \cdot d\hat{k}_2/\hat{k}_3$] on which $\mathcal{E}(3)$ acts via

$$P_j = -i\omega\hat{k}_j, \quad M_{jl} = \hat{k}_j\partial_{\hat{k}_l} - \hat{k}_l\partial_{\hat{k}_j}, \quad 1 \le j, l \le 3, \quad j \ne l. \tag{5.2}$$

These operators determine a unitary irreducible representation of E(3) on $L_2(S_2)$.[20] Once the eigenfunctions $g_{\alpha\beta}(\hat{\mathbf{k}})$ of the operator pairs 1]—11] have been determined, the corresponding separable solutions $\Psi_{\omega\alpha\beta}(x)$ of (2.1) can be obtained from the integral transform

$$\Psi_{\omega\alpha\beta}(x) = \frac{\omega \exp(i\omega x_0)}{(2\pi)^{3/2}} \int_{S_2} \exp(i\omega\mathbf{x}\cdot\hat{\mathbf{k}}) g_{\alpha\beta}(\hat{\mathbf{k}})\, d\Omega(\hat{\mathbf{k}}). \tag{5.3}$$

All the eigenfunctions $g_{\alpha\beta}$ and integrals (5.3) have been computed in Ref. 2.

Now we study coordinate systems permitting variable separation in (2.1) such that the basis functions Ψ are eigenfunctions of $P_3 : P_3\Psi = -i\lambda\Psi$. Here we can set $\Psi(x) = \exp(-i\lambda x_3)\Phi(x_0, x_1, x_2)$, where

$$(\partial_{00} - \partial_{11} - \partial_{22} + \lambda^2)\Phi = 0. \tag{5.4}$$

The symmetry algebra of the Klein—Gordon equation (5.4) is $\mathcal{E}(2,1)$ with basis $\{P_0, P_1, P_2, M_{21}, M_{01}, M_{02}\}$. Furthermore, the pseudo-Euclidean (or Poincaré) group E(2, 1) is the symmetry group of (5.4). In Ref. 3 it is shown in detail that variables separate in (5.4) for 53 orthogonal coordinate systems, each system characterized by a pair of commuting second-order symmetric operators in the enveloping algebra of $\mathcal{E}(2,1)$. [Of course the coordinates 1]—4] for (5.1) are counted again in the list of 53 systems for (5.4).]

On $\mathcal{H}$ the requirement $P_3 f = -i\lambda f$ implies $f(\mathbf{k}) = \delta(k_3 - \lambda) g(k_1, k_2)$ where $-\infty < \lambda < \infty$. The search for eigenfunctions reduces to a study of the Hilbert space $L_2(H)$ of square integrable functions with respect to the measure $d\zeta = dk_1\, dk_2/k_0$, where $k_0 = (k_1^2 + k_2^2 + \lambda^2)^{1/2}$. The inner product is

$$(h, h') = \int \int_{-\infty}^{\infty} h(k_1, k_2)\bar{h}'(k_1, k_2)\, d\zeta, \quad h, h' \in L_2(H), \tag{5.5}$$

and the action of $\mathcal{E}(2,1)$ on $L_2(H)$ is given by

$$P_0 = ik_0, \quad P_1 = -ik_1, \quad P_2 = -ik_2,$$
$$M_{21} = k_2\partial_{k_1} - k_1\partial_{k_2}, \quad M_{01} = k_0\partial_{k_1}, \quad M_{02} = k_0\partial_{k_2}. \tag{5.6}$$

As is well known,[1,20] these operators define a unitary irreducible representation of E(2, 1) on $L_2(H)$. Once the eigenfunctions $g_{\alpha\beta}$ corresponding to each of the 53 separable systems have been determined, the associated separable solutions of (2.1) follow from

$$\Psi_{\lambda\alpha\beta}(x) = \frac{\exp(-i\lambda x_3)}{(2\pi)^{3/2}} \int\int_{-\infty}^{\infty} \exp[i(k_0x_0 - k_1x_1 - k_2x_2)] g_{\alpha\beta}(k_1, k_2)\, d\zeta. \tag{5.7}$$

A detailed study of the basis functions $g_{\alpha\beta}$ and the integrals (5.7) has not yet been undertaken.

Now we search for separable coordinate systems for (2.1) such that the corresponding basis functions Ψ are eigenfunctions of $D: D\Psi = -i\nu\Psi$. Then we have $\Psi(x) = \rho^{i\nu-1}\Phi(s_0, s_1, s_2, s_3)$, where $x_\alpha = \rho s_\alpha$, $\rho \ge 0$, $s\cdot s = \epsilon$, and $\epsilon = +1, -1$, or 0 depending on whether $x\cdot x > 0$, < 0, or $= 0$. From Eq. (2.24ii) we see that the reduced equation for Φ is

$$(M_{21}^2 + M_{13}^2 + M_{32}^2 - M_{01}^2 - M_{02}^2 - M_{03}^2)\Phi(s) = (\nu^2 + 1)\Phi(s). \tag{5.8}$$

The operator D commutes with the subalgebra so(3, 1) with basis $\{M_{21}, M_{13}, M_{32}, M_{01}, M_{02}, M_{03}\}$ and, in fact, so(3, 1) is the symmetry algebra of (5.8).

As discussed in Ref. 6, (5.8) separates in 34 orthogonal coordinate systems for the case $\epsilon = +1$, each system characterized by a pair of commuting second-order symmetric operators in the enveloping algebra of so(3, 1). Some results for the case $\epsilon = -1$ are also presented in Ref. 6.

On $\mathcal{H}$ the requirement $Df = -i\nu f$ implies $f(\mathbf{k}) = k_0^{i\nu-1} h_\nu(\hat{\mathbf{k}})$, $-\infty < \nu < \infty$, where $\mathbf{k} = k_0\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}$ ranges over the unit sphere S_3. The eigenfunction problem reduces to a study of the Hilbert space $L_2(S_3)$ on which so(3, 1) acts via

$$M_{21} = \hat{k}_2\partial_{\hat{k}_1} - \hat{k}_1\partial_{\hat{k}_2}, \quad M_{13} = -\hat{k}_3\partial_{\hat{k}_1}, \quad M_{32} = \hat{k}_3\partial_{\hat{k}_2},$$
$$M_{01} = -(1 + i\nu)\hat{k}_1 + (1 - \hat{k}_1^2)\partial_{\hat{k}_1} - \hat{k}_1\hat{k}_2\partial_{\hat{k}_2},$$
$$M_{02} = -(1 + i\nu)\hat{k}_2 - \hat{k}_1\hat{k}_2\partial_{\hat{k}_1} + (1 - \hat{k}_2^2)\partial_{\hat{k}_2}, \tag{5.9}$$
$$M_{03} = -(1 + i\nu)\hat{k}_3 - \hat{k}_1\hat{k}_3\partial_{k_1} - \hat{k}_1\hat{k}_3\partial_{\hat{k}_2},$$

where we have chosen $\hat{k}_1, \hat{k}_2$ as the independent variables. These operators determine an irreducible unitary global representation of SO(3, 1) which belongs to the principal series. Once the eigenfunctions $h_{\nu\alpha\beta}(\hat{\mathbf{k}})$ for each of the 34 separable systems have been determined the corresponding separable solutions $\Psi_{\nu\alpha\beta}$ of (2.1) can be obtained from

$$\Psi_{\nu\alpha\beta}(x) = \frac{\rho^{i\nu-1}}{(2\pi)^{3/2}}\Gamma(1 - i\nu) \times \int\int_{S_3} \exp[\pm\pi(i + \nu)/2] |1 - \hat{k}_1s_1 - \hat{k}_2s_2 - \hat{k}_3s_3|^{i\nu-1} \times h(\hat{\mathbf{k}})\, d\Omega(\hat{\mathbf{k}}), \tag{5.10}$$

where the plus sign occurs when $1 - \hat{\mathbf{k}}\cdot\mathbf{s} > 0$ and the minus sign occurs when $1 - \hat{\mathbf{k}}\cdot\mathbf{s} < 0$. For the case $\epsilon = +1$, $x_0 > 0$, these integrals are evaluated in Ref. 6. A number of cases for $\epsilon = -1$ are also computed.

6. THE SCHRÖDINGER EQUATION

Now we consider the separable coordinate systems for (2.1) such that the basis functions Ψ are eigenfunctions of $P_0 + P_1 : (P_0 + P_1)\Psi = i\beta\Psi$. Setting $\Psi(x) = \exp(is\beta) \times \Phi(t, x_2, x_3)$, where $2s = x_0 + x_1$, $2t = x_1 - x_0$, we find that the reduced equation satisfied by Φ is the free particle Schrödinger equation

$$(i\beta\partial_t + \partial_{22} + \partial_{33})\Phi(t, x_2, x_3) = 0, \tag{6.1}$$

which admits as symmetries the operators

$$\mathcal{P}_1 = P_2, \quad \mathcal{P}_2 = P_3, \quad \mathcal{E} = P_0 + P_1, \quad \mathcal{K}_{-2} = P_1 - P_0,$$
$$\mathcal{K}_2 = -\tfrac{1}{2}(K_0 + K_1), \quad \mathcal{M} = -M_{32}, \quad \mathcal{B}_1 = \tfrac{1}{2}(M_{02} + M_{21}),$$
$$\mathcal{B}_2 = \tfrac{1}{2}(M_{03} - M_{13}), \quad \mathcal{D} = -(D + M_{01}). \tag{6.2}$$

All these operators commute with $\mathcal{E} = P_0 + P_1$ and they form a basis for the nine-dimensional symmetry algebra $\mathcal{G}_2$ of (6.1). This algebra is discussed in Refs. 7 and 21. In Ref. 7 it is shown that (6.1) admits R-separable solutions in 17 (nonorthogonal) coordinate systems. Each system is characterized by a commuting pair of symmetry operators from the enveloping algebra of $\mathcal{G}_2$, one operator first-order and one second-order.

On $\mathcal{H}$ the requirement $(P_0+P_1)f = i\beta f$ implies $f(\mathbf{k}) = u\delta(u-\beta)l_\beta(v,w)$, where $\beta > 0$, $u = k_0 - k_1$, $v = k_2$, $w = k_3$. The search for eigenfunctions l_β reduces to a study of the Hilbert space $L_2(R^2)$ on which the Schrödinger algebra acts via

$$\mathcal{P}_1 = -iv, \quad \mathcal{P}_2 = -iw, \quad \mathcal{E} = i\beta, \quad \mathcal{K}_{-2} = \frac{-i}{\beta}(v^2+w^2),$$
$$\mathcal{K}_2 = \frac{-i\beta}{4}(\partial_{vv}+\partial_{ww}), \quad \mathcal{M} = v\partial_w - w\partial_v, \tag{6.3}$$
$$\mathcal{B}_1 = \tfrac{1}{2}\beta\partial_v, \quad \mathcal{B}_2 = \tfrac{1}{2}\beta\partial_w, \quad \mathcal{D} = -(1+v\partial_v+w\partial_w).$$

It is known[7,22] that these operators induce a unitary irreducible representation of the Schrödinger group G_2 on $L_2(R^2)$. Once the eigenfunctions $l_{\beta\alpha\rho}(v,w)$ corresponding to each separable system have been determined, the corresponding separable solutions $\Psi_{\beta\alpha\rho}(x)$ of (2.1) follow from

$$\Psi_{\beta\alpha\rho}(x) = \frac{\exp(i\beta s)}{(2\pi)^{3/2}} \iint_{-\infty}^{\infty} \exp\left[\frac{-it}{\beta}(v^2+w^2) - i(x_2v+x_3w)\right] l_{\beta\alpha\rho}(v,w)\,dv\,dw. \tag{6.4}$$

Using the u,v,w coordinates we can now compute the operator $\exp[a(K_0+K_1)]$ in $\mathcal{H}$. Indeed the well-known expression

$$\exp[it(\partial_{xx}+\partial_{yy})]f(x,y) = \text{l.i.m.}\ \frac{1}{4\pi t}\iint_{-\infty}^{\infty}\exp\left\{-\frac{1}{4it}[(x-s_1)^2 + (y-s_2)^2]\right\} f(s_1,s_2)\,ds_1\,ds_2, \tag{6.5}$$

for time translation of solutions of the free-particle Schrödinger equation, e.g.,[7] together with expressions (6.2) and (6.3) for $\mathcal{K}_{-2}$ leads to

$$\exp[a(K_0+K_1)]f(\mathbf{k}) = \frac{1}{4\pi i a}\iint_{-\infty}^{\infty}\exp\left\{\frac{1}{-4ai(k_0-k_1)}[(k_2-s_2)^2+(k_3-s_3)^2]\right\} \times f\left(\frac{s_2^2+s_3^2-(k_0-k_1)^2}{2(k_0-k_1)}, s_2, s_3\right)ds_2\,ds_3, \quad f\in\mathcal{H}.$$

7. THE GENERALIZED EPD EQUATION

We next look for solutions Ψ of (2.1) such that $\Gamma_{43}\Psi = im\Psi$. Then $\Psi(x) = \exp(im\varphi)\Phi(x_0,x_1,r)$ where

$$x_3 = r\cos\varphi, \quad x_2 = r\sin\varphi$$

and Φ satisfies the reduced equation

$$\left(\partial_{00} - \partial_{rr} - \frac{1}{r}\partial_r + \frac{m^2}{r^2} - \partial_{11}\right)\Phi = 0. \tag{7.1}$$

If Φ is independent of x_1 then (7.1) reduces to the Euler–Poisson–Darboux (EPD) equation. Expression (7.1) can be written in the operator form (2.24iv),

$$(\Gamma_{12}^2 + \Gamma_{56}^2 - \Gamma_{15}^2 - \Gamma_{16}^2 - \Gamma_{26}^2 - \Gamma_{25}^2)\Phi = (\Gamma_{43}^2+1)\Phi = (1-m^2)\Phi. \tag{7.2}$$

The symmetry algebra of (7.1) is so(2, 2) with basis $\{\Gamma_{12},\Gamma_{56},\Gamma_{15},\Gamma_{16},\Gamma_{26},\Gamma_{25}\}$ or alternate basis $\{P_0,K_0,P_1,K_1,D,M_{01}\}$. Separable coordinate systems for this interesting equation will be classified in a later publication.

On $\mathcal{H}$ the requirement $\Gamma_{43}f = imf$ implies $f(\mathbf{k}) = \exp(im\theta)\times j(l,k)$ where $m = 0, \pm1, \pm2, \cdots$, $l \geq 0$, $k_3 = l\cos\theta$, $k_2 = l\sin\theta$, $k_1 = k$. The eigenfunction problem reduces to a study of the Hilbert space $\mathcal{L}_2$ of functions $j(l,k)$ Lebesgue square integrable with respect to the measure $d\rho(l,k) = l(l^2+k^2)^{-1/2}dl\,dk$. The inner product is

$$(j_1,j_2) = \int_0^\infty\int_{-\infty}^{\infty} j_1\bar{j}_2\,d\rho(l,k), \quad j_1,j_2\in\mathcal{L}_2.$$

The symmetry algebra so(2, 2) acts on $\mathcal{L}_2$ via

$$\Gamma_{12} = \frac{ik}{2}(\partial_{kk} - \partial_{ll} - l^{-1}\partial_l + m^2l^{-2} - 1) + il\partial_{lk} - \partial_k,$$
$$\Gamma_{56} = \frac{i}{2}(k^2+l^2)^{1/2}(-\partial_{ll} - l^{-1}\partial_l + m^2l^{-2} - \partial_{kk} + 1),$$
$$\Gamma_{15} = \frac{i}{2}(k^2+l^2)^{1/2}(\partial_{ll} + l^{-1}\partial_l - m^2l^{-2} + \partial_{kk} + 1),$$
$$\Gamma_{16} = 1 + l\partial_l + k\partial_k, \tag{7.3}$$
$$\Gamma_{26} = \frac{ik}{2}(-\partial_{kk} + \partial_{ll} + l^{-1}\partial_l - m^2l^{-2} - 1) - il\partial_{lk} - i\partial_k,$$
$$\Gamma_{25} = (k^2+l^2)^{1/2}\partial_k.$$

A third basis of so(2, 2) for which the structure of the Lie algebra becomes more transparent is

$$A_1 = \Gamma_{56}+\Gamma_{12}, \quad A_2 = \Gamma_{52}+\Gamma_{16}, \quad A_3 = \Gamma_{26}+\Gamma_{15},$$
$$B_1 = \Gamma_{56}-\Gamma_{12}, \quad B_2 = \Gamma_{52}-\Gamma_{16}, \quad B_3 = \Gamma_{26}-\Gamma_{15}, \tag{7.4}$$

which commutation relations

$$[A_1,A_2] = -2A_3, \quad [A_2,A_3] = 2A_1, \quad [A_1,A_3] = 2A_2,$$
$$[B_1,B_2] = -2B_3, \quad [B_2,B_3] = 2B_1, \quad [B_1,B_3] = 2B_2, \tag{7.5}$$
$$[A_i,B_j] = 0.$$

With respect to this basis the isomorphism so(2, 2) $\approx$ sl(2)$\times$sl(2) is obvious. Moreover, it follows from (4.1) and (4.2) that in the eigenspace $\mathcal{H}_m$ of $\mathcal{H}$ corresponding to the eigenvalue m of $-i\Gamma_{43}$ there is an ON basis $\{f_{\alpha\beta}\}$ such that

$$A_1f_{\alpha\beta} = i(|m|+2\alpha+1)f_{\alpha\beta}, \quad B_1f_{\alpha\beta} = i(|m|+2\beta+1)f_{\alpha\beta}:$$
$$\alpha,\beta = 0,1,2,\cdots.$$

Also $A_1^2 - A_2^2 - A_3^2 = B_1^2 - B_2^2 - B_3^2 = 1 - m^2$ on $\mathcal{H}_m$. It follows that this action of sl(2)$\times$sl(2) on $\mathcal{H}_m$ is irreducible and extends to a unitary irreducible representation $D^-_{(|m|-1)/2}\otimes D^-_{(|m|-1)/2}$ of SL(2, R)$\times$SL(2, R) on $\mathcal{H}_m$. Here D_k^- is a representation of SL(2, R) belonging to the negative discrete series.

Once the eigenbasis $\{g_{\xi\eta}\}$ in $\mathcal{H}_m$ corresponding to a separable system for (7.1) has been constructed, the associated separable solutions of (7.1) can be obtained from the transform

$$\Psi_{m t \eta}(x)=\frac{\exp[im(\varphi-\pi/2)]}{(2\pi)^{1/2}}\int_0^\infty l\,dl J_m(lr)\int_{-\infty}^\infty dk \cdot \exp[i(x_0(k^2+l^2)^{1/2}-x_1 k)]j(l,k)/(k^2+l^2)^{1/2}. \tag{7.6}$$

8. DIAGONALIZATION OF $\Gamma_{12}+\Gamma_{34}-\Gamma_{56}$

Finally we study the separable solutions of (2.1) for which the basis functions Ψ are eigenfunctions of $L=\Gamma_{12}+\Gamma_{34}-\Gamma_{56}$: $L\Psi=-i(2\kappa+1)\Psi$. By a tedious computation one can verify that (2.1) is equivalent to the equation

$$(\tfrac{1}{3}A_0^2+A_1^2+A_2^2+A_3^2-B_1^2-B_2^2-C_1^2-C_2^2)\Psi=(\tfrac{1}{3}L^2+3)\Psi, \tag{8.1}$$

where

$$A_0=\Gamma_{34}+\Gamma_{12}+2\Gamma_{56},\quad A_1=\Gamma_{13}+\Gamma_{24},\quad A_2=\Gamma_{12}-\Gamma_{34},$$
$$A_3=\Gamma_{14}-\Gamma_{23},\quad B_1=\Gamma_{25}-\Gamma_{15},\quad B_2=\Gamma_{45}-\Gamma_{36}, \tag{8.2}$$
$$C_1=\Gamma_{15}+\Gamma_{26},\quad C_2=\Gamma_{35}+\Gamma_{46}.$$

Thus, the reduced equation is

$$(\tfrac{1}{3}A_0^2+A_1^2+A_2^2+A_3^2-B_1^2-B_2^2-C_1^2-C_2^2)\Psi=(-\mu^2/3+3)\Psi,\quad \mu=2\kappa+1. \tag{8.3}$$

The operators (8.2) satisfy the commutation relations for su(2, 1) and the expression on the left-hand side of (8.3) is the Casimir operator for su(2, 1).

The usual model for su(2, 1) is the space of 3×3 complex matrices $\mathcal{A}$ such that

$$\mathcal{A}^t\mathcal{G}^{2,1}+\mathcal{G}^{2,1}\mathcal{A}=0,$$

where[23]

$$\mathcal{G}^{2,1}=\begin{bmatrix}1 & & 0\\ & 1 & \\ 0 & & -1\end{bmatrix}.$$

This real Lie algebra is eight-dimensional with basis

$$\mathcal{A}_0=\begin{bmatrix}i&0&0\\0&i&0\\0&0&-2i\end{bmatrix},\quad \mathcal{A}_1=\begin{bmatrix}0&1&0\\-1&0&0\\0&0&0\end{bmatrix},\quad \mathcal{A}_2=\begin{bmatrix}i&0&0\\0&-i&0\\0&0&0\end{bmatrix},$$

$$\mathcal{A}_3=\begin{bmatrix}0&i&0\\i&0&0\\0&0&0\end{bmatrix},\quad \mathcal{B}_1=\begin{bmatrix}0&0&1\\0&0&0\\1&0&0\end{bmatrix},\quad \mathcal{B}_2=\begin{bmatrix}0&0&0\\0&0&1\\0&1&0\end{bmatrix}, \tag{8.4}$$

$$\mathcal{C}_1=\begin{bmatrix}0&0&i\\0&0&0\\-i&0&0\end{bmatrix},\quad \mathcal{C}_2=\begin{bmatrix}0&0&0\\0&0&i\\0&-i&0\end{bmatrix},$$

and the basis elements satisfy the same commutation relations as the corresponding operators (8.2). The symmetry algebra of Eq. (8.3) is also su(2, 1).

It follows from (4.1) and (4.2) that the possible values of κ are 0, 1, 2, ⋯ and for fixed κ, the solution space of (8.3) has an ON basis $\{\Psi_{l,s}: l=0,1,2,\cdots,\ s=0,1,\ldots,\kappa+l\}$ such that

$$L\Psi_{ls}=-i(2\kappa+1)\Psi_{ls},\quad A_0\Psi_{ls}=i(\kappa+3l+2)\Psi_{ls},$$
$$A_2\Psi_{ls}=i(2s-\kappa-l)\Psi_{ls}. \tag{8.5}$$

The solution space of (8.3) transforms irreducibly under this action of su(2, 1) and the Lie algebra representation lifts to a global unitary irreducible representation of SU(2, 1), see Ref. 24.

The problem of separation of variables for (8.3) is far from settled. The variables in (8.3) are intertwined in an extremely complicated manner and the standard techniques for separating variables in the wave equation, e.g., Ref. 25, yield no nontrivial separable systems for this case. However, it follows from standard Lie theory, Ref. 26, p. 49, that every pair of commuting operators in su(2, 1) leads to a separable coordinate system. It is not yet known whether there exist separable systems corresponding to second-order operators in the enveloping algebra of su(2, 1).

9. CONCLUDING REMARKS

For completeness we classify the orbits in so(4, 2) under the adjoint action of SO(4, 2). This classification has been given by Zassenhaus[27] and later by many others but we present the results here in an explicit form adapted to our notation. (This orbit analysis is useful because we know that coordinate systems whose defining operators can be mapped into one another under an action of the adjoint group are equivalent.) We list the possible eigenvalues of a 6×6 matrix $A\in$ so(4, 2) such that $\Gamma=TAT^{-1}$ for some $T\in SO(4,2)$, i.e., we list an element on each SO(4, 2) orbit. It is easy to show that if $\lambda\neq 0$ is an eigenvalue then so are $-\lambda$ and $\bar\lambda$. We use the notation $\lambda(n)$, $n=2,\ldots,5$, to signify that λ corresponds to a generalized eigenvector x of rank n, i.e., n is the smallest integer m such that $(A-\lambda E)^m x=0$ where E is the 6×6 identity matrix.

Possible eigenvalues	Canonical form Γ
1. $\pm\alpha\pm i\beta$, $\pm i\gamma$, $\alpha,\beta\neq 0$	$\gamma\Gamma_{12}+\beta(\Gamma_{34}+\Gamma_{65})+\alpha(\Gamma_{35}+\Gamma_{46})$
2. $\pm i\alpha$, $\pm\beta$, $\pm\gamma$, $\beta^2+\gamma^2>0$	$\alpha\Gamma_{12}+\beta\Gamma_{35}+\gamma\Gamma_{46}$
3. $\pm i\alpha$, $\pm i\beta$, $\pm i\gamma$, $\alpha,\beta,\gamma\neq 0$	$\alpha\Gamma_{12}+\beta\Gamma_{34}+\gamma\Gamma_{56}$
3a. $\pm i\alpha$, $\pm i\beta$, 0, 0, $\alpha,\beta\neq 0$	$\alpha\Gamma_{12}+\beta\Gamma_{34}$ or $\alpha\Gamma_{12}+\beta\Gamma_{56}$
3b. $\pm i\alpha$, 0, 0, 0, 0	$\alpha\Gamma_{12}$ or $\alpha\Gamma_{56}$

4. $\alpha(2)$, $-\alpha(2)$, $\pm i\beta$, $\alpha\neq 0$ — $\alpha(\Gamma_{35}+\Gamma_{46})+\beta\Gamma_{12}+\frac{1}{2}(\Gamma_{34}+\Gamma_{36}+\Gamma_{45}+\Gamma_{65})$

5. $i\alpha(3)$, $-i\alpha(3)$ — $\alpha(\Gamma_{12}+\Gamma_{34}+\Gamma_{65})+\frac{1}{2}(\Gamma_{13}+\Gamma_{24}+\Gamma_{15}+\Gamma_{26})$

6. $i\alpha(2)$, $-i\alpha(2)$, $\pm i\beta$ — $\alpha(\Gamma_{34}+\Gamma_{65})+\beta\Gamma_{12}+\frac{1}{2}(\Gamma_{34}+\Gamma_{56}+\Gamma_{36}+\Gamma_{54})$

7. $\pm\alpha$, $0(3)$, 0, $\alpha\neq 0$ — $\alpha\Gamma_{46}+\frac{1}{\sqrt{2}}(\Gamma_{21}+\Gamma_{25})$

8. $\pm i\alpha$, $0(3)$, 0 — $\alpha\Gamma_{12}+\frac{1}{\sqrt{2}}(\Gamma_{35}+\Gamma_{65})$ or $\alpha\Gamma_{12}+\frac{1}{\sqrt{2}}(\Gamma_{43}+\Gamma_{45})$

9. $0(5)$, 0 — $\frac{1}{2}(\Gamma_{43}+\Gamma_{36}+\Gamma_{45}+\Gamma_{56})+\frac{1}{\sqrt{2}}(\Gamma_{24}+\Gamma_{26})$

From these results we can see why many operators $\Gamma\in so(4,2)$ do not directly correspond to a semisubgroup coordinate system. For example, it is easy to check that an element of so(4, 2) which commutes with Γ (case 1], $\alpha,\beta,\gamma\neq 0$) also commutes with each of the (commuting) operators $\Gamma_{12}, \Gamma_{34}+\Gamma_{65}$, and $\Gamma_{35}+\Gamma_{46}$. The coordinate system associated with these operators is equivalent to a separable system for Eq. (7.1). By interpreting the remaining cases in a similar fashion one can show that each case is in fact associated with at least one semisubgroup coordinate system.

[1] L. Gross, J. Math. Phys. **5**, 687—95 (1964).
[2] C. Boyer, E.G. Kalnins, and W. Miller, Jr., Nagoya Math. J. **60**, 35—80 (1976).
[3] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, 331 (1976).
[4] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, 369 (1976).
[5] E.G. Kalnins, W. Miller, Jr., and P. Winternitz, SIAM J. Appl. Math. **30**, 630—64 (1976).
[6] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **18**, 1 (1977).
[7] C. Boyer, E.G. Kalnins, and W. Miller, Jr., J. Math. Phys. **16**, 499—511 (1975).
[8] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **16**, 2507—16 (1975).
[9] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, 356 (1976).
[10] G. Post, J. Math. Phys. **17**, 24—32 (1976).
[11] H. Bateman, *Electrical and Optical Wave-Motion* (Dover, New York, 1955), reprint of 1914 ed.
[12] H. Bateman, *Partial Differential Equations of Mathematical Physics* (Cambridge U.P., Cambridge, 1969), reprint of 1932 ed.
[13] H. Kastrup, Phys. Rev. **140**, B183 (1965).
[14] W. Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972).
[15] W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., to appear).
[16] P. Dirac, Proc. Cambridge Philos. Soc. 30, 150—63 (1934).
[17] A. Davydov, *Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1965).
[18] W. Miller, Jr., J. Math. Phys. **13**, 1393—99 (1972).
[19] A. Makarov, J. Smorodinsky, K. Valiev, and P. Winternitz, Nuovo Cimento A **52**, 1061—84 (1967).
[20] N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (AMS Trans., Am. Math. Soc., Providence, R.I., 1968).
[21] G. Burdet, M. Perrin, and P. Sorba, Commun. Math. Phys. **34**, 85—90 (1973).
[22] C. Boyer, Helv. Phys. Acta **47**, 589 (1974).
[23] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[24] T. Hirai, Proc. Jpn. Acad. **42**, 907—12 (1966).
[25] M. Bôcher, *Die Reihentwickelungen der Potentialtheorie* (Leipzig, 1894).
[26] L. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).
[27] H. Zassenhaus, Can. Math. Bull. **1**, 31—39, 101—11, 183—91 (1958).

Conservative neutron transport theory*

R. L. Bowden, W. L. Cameron,† and P. F. Zweifel

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060

(Received 30 August 1976)

A functional analytic development of the Case full-range and half-range expansions for the neutron transport equation for a conservative medium is presented. A technique suggested by Larsen is used to overcome the difficulties presented by the noninvertibility of the transport operator K^{-1} on its range. The method applied has considerable advantages over other approaches and is applicable to a class of abstract integro-differential equations.

I. INTRODUCTION

The neutron transport equation for a "conservative" half-space ($c=1$ in one-speed theory) presents special complications for essentially technical reasons. The orthodox Case approach to the one-speed situation was originally worked out by Shure and Natelson,[1] while Greenberg and Zweifel[2] used the Larsen–Habetler resolvent integration technique[3] to treat the same equation. We restrict our attention in this paper to the resolvent integration method and point out that the special difficulties encountered for $c=1$ (cf. Ref. 2 for a detailed discussion) occur because the transport operator K^{-1} is not invertible on its range for that situation. The standard technique, originally introduced by Lekkerkerker[4] is to restrict K^{-1} to a suitable subspace of its domain on which it is invertible, deal with the restricted operator of the standard Larsen–Habetler scheme, and eventually extend the result to the full domain. While this technique in fact works, it is somewhat cumbersome and introduces notational complexities, especially in the conservative multigroup case[5] which is, of course, a generalization of the one-speed situation and has been treated by the same technique. (We should point out that the solutions to the conservative transport equation are of considerable physical importance, especially in obtaining asymptotic solutions to ordinary transport equations in the boundary layer.[6,7])

Recently, we have been studying some problems in plasma oscillations and rarefied gas dynamics where the ordinary Larsen–Habetler technique is not directly applicable because the operator corresponding to $K=(K^{-1})^{-1}$ of the neutron transport equation is unbounded. (In the neutron case for $c=1$ the operator K^{-1} is unbounded but K is bounded. In the plasma and gas case both operators are unbounded.) It is in fact possible to integrate the resolvent about an unbounded spectrum, as has been done by Bareiss,[8] but the technique involves approximating the transport operator by a sequence of bounded operators and is somewhat cumbersome. Larsen suggested another approach, namely to define an operator $S=(K-zI)^{-1}$, where z is some complex number not in the spectrum of K.[9] Then S is a bounded, invertible operator, and the whole machinery of the resolvent integration technique can be applied to S. This technique has proved extremely fruitful in treating the plasma and gas problems and has, in fact been generalized to treat a class of abstract integro-differential equations.[10] In the process of writing out these cases, we suddenly realized that the same technique could be applied to the conservative neutron transport case, with considerable simplification over the treatments of Refs. 2 and 5. We present the analysis in this paper omitting many of the calculational details because they have already appeared in the above cited references. In Sec. II we treat the one-speed case and Sec. III the multigroup equations.

II. THE CASE EIGENFUNCTION EXPANSION FOR A CONSERVATIVE MEDIUM

We follow the notation of Refs. 2 and 3 to write the transport equation for $c=1$ in the form

$$\frac{\partial\psi}{\partial x}(x,\mu)+K^{-1}\psi(x,\mu)=\frac{q(x,\mu)}{\mu},\quad \mu\neq 0, \tag{1a}$$

with

$$(K^{-1}\psi)(\mu)=\frac{1}{\mu}\left[\psi(\mu)-\tfrac{1}{2}\int_{-1}^{+1}\psi(s)\,ds\right]. \tag{1b}$$

We note that K^{-1} is not invertible on its range. In fact, the vectors $e_0(\mu)=1$ and $e_1(\mu)=\mu$, $-1<\mu<1$, span the $\lambda=0$ root linear manifold of K^{-1}.[2] Furthermore, as is well known, the spectrum of K^{-1} is confined to the real line. Thus, the operator $S=(K^{-1}-iI)^{-1}$ is a bounded invertible operator. We easily compute

$$(S\psi)(\mu)=\frac{\mu}{1-i\mu}\psi(\mu)+\frac{(1-i\mu)^{-1}}{2\Lambda(i)}\int_{-1}^{+1}\frac{s\psi(s)\,ds}{1-is}. \tag{2}$$

As in Ref. 2, we work in the space $X_p=\{f\mid \mu f\in L_p(-1,1)\}$ but restrict S to the space $H_p=\{f\in X_p\mid f \text{ is of class } H^*\}$; the final results can then be extended to X_p by continuity. Here $\Lambda(z)$ is the usual dispersion function for $c=1$:

$$\Lambda(z)=1-\frac{z}{2}\int_{-1}^{+1}\frac{ds}{z-s}. \tag{3}$$

We now proceed to deal with S by the technique of Ref. 3, i.e., we compute the resolvent and by contour integration of the resolvent about the spectrum of S we obtain the desired Case eigenfunction expansion. The resolvent is seen to be

$$(zI-S)^{-1}\psi(\mu)=\frac{(1+iz)^{-1}}{t^{-1}(z)-\mu}\left\{(1-i\mu)\psi(\mu)+\frac{(1+iz)^{-1}}{2\Lambda(t^{-1}(z))}\int_{-1}^{+1}\frac{s\psi(s)\,ds}{t^{-1}(z)-s}\right\}, \tag{4a}$$

with

$$t(z) = z/(1 - iz) \tag{4b}$$

and

$$t(t^{-1}(z)) = z, \tag{4c}$$

so that

$$t^{-1}(z) = z/(1 + iz) \tag{4d}$$

The spectrum of S can be computed by studying the analytic structure of the resolvent or one can use the spectral mapping theorem to transform the spectrum of K^{-1}. In either case one finds

$$\sigma(S) = P\sigma(S) \cup C\sigma(S), \tag{5a}$$

with

$$P\sigma(S) = \{i\} \tag{5b}$$

and

$$C\sigma(S) = \{z \mid z = \tfrac{1}{2}(i + e^{i\theta}), \quad -\pi \leq \theta \leq 0\}. \tag{5c}$$

$C\sigma(S)$ is, of course, a semicircle. Furthermore, the point i is an eigenvalue of multiplicity 2.

We now utilize the identity

$$\frac{1}{2\pi i} \int_C (zI - S)^{-1} dz = I, \tag{5d}$$

where the contour C surrounds the spectrum of S. As usual C is "squeezed" down into a contour Γ surrounding $C\sigma(S)$ and a contour Γ_i surrounding the eigenvalue i. We compute the two contributions to (5d) separately. First consider

$$\frac{1}{2\pi i} \int_\Gamma (zI - S)^{-1}\psi(\mu)\, dz = \frac{1}{2\pi i} \int_{\Gamma'} \frac{1}{z' - \mu} \left(\psi(\mu) + \frac{\Lambda^{-1}(z')}{2} \int_{-1}^{+1} \frac{s\psi(s)\, ds}{z' - s}\right) dz'. \tag{6}$$

Here Γ' is any contour surrounding the cut $[-1, 1]$. Equation (6) was obtained simply by integrating (4a) around the semicircle $C\sigma(S)$ and introducing the change of variable $z' = t^{-1}(z) = z/(1 + iz)$. This is precisely the result of Ref. 2 for the branch cut integration. Thus we are led directly to the standard formula[2,3]

$$\frac{1}{2\pi i} \int_\Gamma (zI - S)^{-1}\psi(\mu)\, dz = \int_{-1}^{+1} A(\nu)\, \phi_\nu(\mu)\, d\nu, \tag{7a}$$

with

$$A(\nu) = \frac{1}{N(\nu)} \int_{-1}^{+1} \mu\psi(\mu)\, \phi_\nu(\mu)\, d\mu, \tag{7b}$$

$$\phi_\nu(\mu) = \frac{\nu}{2} P \frac{1}{z - \mu} + \tfrac{1}{2}[\Lambda^+(\nu) + \Lambda^-(\nu)], \tag{7c}$$

and

$$N(\nu) = \nu\Lambda^+(\nu)\Lambda^-(\nu). \tag{7d}$$

As usual we denote

$$\Lambda^\pm(\nu) = \lim_{\epsilon \to 0} \Lambda(\nu \pm i\epsilon), \quad -1 < \nu < 1.$$

The integration around Γ_i of (4a) involves the evaluation of a residue at a second order pole [since $\Lambda(t^{-1}(i)) = \Lambda'(t^{-1}(i)) = 0$]. Using the standard residue formula

$$\frac{1}{2\pi i} \oint \frac{p(z)}{q(z)}\, dz = \frac{2}{3[q''(z_0)]^2} [3p'(z_0)\, q''(z_0) - p(z_0)\, q'''(z_0)], \tag{8a}$$

if $q(z_0) = q'(z_0) = 0$, and identifying

$$p(z) = \tfrac{1}{2}[z(1 - i\mu) - \mu]^{-1} \int_{-1}^{+1} \frac{s\psi(s)\, ds}{z(1 - is) - s} \tag{8b}$$

and

$$q(z) = \Lambda(t^{-1}(z)), \tag{8c}$$

one easily finds

$$\frac{1}{2\pi i} \int_{\Gamma_i} (zI - S)^{-1}\psi(\mu)\, dz = \frac{3}{2}\left[\mu \int_{-1}^{+1} s\psi(s)\, ds + \int_{-1}^{+1} s^2\psi(s)\, ds\right]. \tag{9}$$

If one now combines Eqs. (6) and (9), one obtains Eq. (10) of Ref. 2, i.e., the Case full-range expansion formula for $c = 1$,

$$\psi(\mu) = \tfrac{1}{2}a_0 - \tfrac{1}{2}a_1\mu + \int_{-1}^{+1} A(\nu)\, \phi_\nu(\mu)\, d\nu, \tag{10}$$

where the expansion coefficients a_i are defined by

$$a_i = 3 \int_{-1}^{+1} (-\mu)^{2-i}\, \psi(\mu)\, d\mu. \tag{11}$$

We now sketch the procedure that can be used to obtain the Case half-range expansion. As usual, we define an operator $E: X_p' \to X_p$, where X_p' is the space of functions $f: [0, 1] \to \mathbb{C}$ with

$$\|f\|_{p'} = [\int_0^1 |uf(u)|^p\, du]^{1/p} < \infty,$$

and we require

(i) $(E\psi)(\mu) = \psi(\mu), \quad \mu > 0,$

(ii) $(zI - S)^{-1}E\psi$ is analytic for $\mathrm{Re}\, z < 0$,

(iii) $(zI - S)^{-1}E\psi$ has at worst a simple pole at $z = i$. (12)

Condition (ii) will guarantee that in the integral of $(zI - S)^{-1}E\psi$ around a contour containing the spectrum of S there will be no contribution from the portion of $C\sigma(S)$ with $\mathrm{Re}\, z < 0$. Because the transformation $S \to K^{-1}$ maps $C\sigma(S)$, $\mathrm{Re}\, z < 0$ into $[-1, 0)$, this assures that no negative Case continuum modes will occur in the full range expansion of $E\psi$, i.e., the half-range expansion of $\psi \in X_p'$. Condition (iii) guarantees that the discrete coefficient a_1 does not enter into the half-range expansion of ψ. These conditions could be used to derive the operator E, but the result would be the same as that used in Ref. 2. Therefore, we shall only verify that the operator E as given in Ref. 2,

$$Ef(\mu) = \begin{cases} \dfrac{1}{X(\mu)} \dfrac{3}{2} \displaystyle\int_0^1 \dfrac{sf(s)\, ds}{X(-s)(s - \mu)}, & \mu < 0, \\ f(\mu), & \mu > 0, \end{cases} \tag{13}$$

has the correct properties. Here $X(z)$ provides the

Wiener–Hopf factorization of $\Lambda(z)^2$:

$$X(z)\,X(-z) = 3\Lambda(z),$$

where $X(z)$ is analytic in $\mathbb{C}\setminus[0,1]$ and vanishes as $1/z$ as $|z|\to\infty$.

When we substitute Eq. (13) into Eq. (4a), we find after simplification that $(zI-S)^{-1}E\psi$ satisfies

$$[(zI-S)^{-1}E\psi](\mu) = [z(1-i\mu)-\mu]^{-1}\quad (1-i\mu)\,\psi(\mu) + \frac{3}{2}\,\frac{1}{X(t^{-1}(z))} \times \int_0^1 \frac{s\psi(s)\,ds}{X(-s)[z(1-is)-s]}\Bigg\},\quad \mu>0 \qquad (14a)$$

$$[(zI-S)^{-1}E\psi](\mu) = \frac{3}{2[z(1-i\mu)-\mu]}\Bigg\{\int_0^1 \frac{s\psi(s)}{X(-s)}\left[\frac{1-i\mu}{X(\mu)(s-\mu)} + \frac{1}{X(t^{-1}(z))[z(1-is)-s]}\right.\; ds\Bigg\},\quad \mu<0. \qquad (14b)$$

Equation (14) can be used to quickly verify that $(zI-S)^{-1}E\psi$ satisfies properties (ii) and (iii). To see this, note that t^{-1} maps the left half complex plane into itself and is analytic except for a simple pole at $z=i$. Thus $X(t^{-1}(z))$ is analytic for $\mathrm{Re}\,z<0$. Moreover, for $\mu>0$ and $\mathrm{Re}\,z<0$, $z(1-i\mu)-\mu$ does not vanish. Therefore, from Eq. (14a) we have that $(zI-S)^{-1}E\psi(\mu)$ is analytic in z for $\mathrm{Re}\,z<0$ and $\mu>0$. To see that $(zI-S)^{-1}E\psi$ is analytic for $\mathrm{Re}\,z<0$ when $\mu<0$, we need only check that $z=\mu/(1-i\mu)$ is not a singularity of $(zI-S)^{-1}E\psi$. This is done by recalling from Eq. (4d) that $t^{-1}(\mu/(1-i\mu))=\mu$. Thus $(zI-S)^{-1}E\psi$ is analytic for $\mathrm{Re}\,z<0$. At $z=i$, we note from Eq. (14) that $(zI-S)^{-1}E\psi$ has a simple pole induced by the zero of $X(t^{-1}(z))$.

Integrating $(zI-S)^{-1}E\psi(\mu)$ on z along a contour containing the point i and the semicircle $\{z\,|\,z=\frac{1}{2}(i+e^{i\theta}),\ -\pi/2<\theta<0\}$ yields the Case half-range eigenfunction expansion.

III. CONSERVATIVE MULTIGROUP TRANSPORT

We now derive the result of Ref. 5 in the same simple manner used in Sec. II. We define

$$(K^{-1}\psi)(x,\mu) = (1/\mu)[\Sigma\psi(x,\mu) - C\textstyle\int_{-1}^{+1}\psi(x,s)\,ds],\quad \mu\neq 0. \qquad (15)$$

Here ψ is an N-component vector where the ith component represents the neutron angular densities in the ith group, Σ is the diagonal cross-section matrix, and C the group–group transfer matrix. The appropriate space to seek a solution is, as in Ref. 5, the space

$$X_p^N = \bigotimes_{j=1}^{N} X_p.$$

As in the one-speed case, the computations are done in a dense subspace of Hölder continuous functions, and can be extended to X_p^n by continuity.[11]

We have the dispersion function

$$\Lambda(z) = (\Sigma-2C)\,C^{-1}\Sigma - \textstyle\int_{-1}^{+1}\mu D(z,\mu)\,d\mu, \qquad (16a)$$

where

$$D(z,\mu) = (zI-\mu\Sigma^{-1})^{-1}. \qquad (16b)$$

As in Ref. 5, we consider the conservative case for which $\det(\Sigma-2c)=0$. In this case K^{-1} given by Eq. (14) is not invertible on its range. Thus defining S as before, i.e., $S=(K^{-1}-iI)^{-1}$, we find

$$S\eta(\mu) = B(\mu)\{\mu\eta(\mu) + \Sigma[C^{-1} - \textstyle\int_{-1}^{+1}B(s)\,ds]^{-1}\int_{-1}^{+1}sB(s)\eta(s)\,ds\}, \qquad (17a)$$

where

$$B(\mu) = (\Sigma - i\mu I)^{-1}. \qquad (17b)$$

We have assumed $z=i$ is in the resolvent set of K^{-1}. If not, any other point could be chosen assuming the spectrum of K does not consist of the entire complex plane. Furthermore, we have assumed that $\det\Lambda(z)$ vanishes as $1/z^2$ as $|z|\to\infty$.

It is convenient to define

$$F(z,\mu) = (zI-\mu B(\mu))^{-1}. \qquad (18)$$

Then a direct computation gives

$$(zI-S)^{-1}\psi(\mu) = F(z,\mu)\{\psi(\mu) + B(\mu)R^{-1}(z) \times[C^{-1} - \textstyle\int_{-1}^{+1}B(s)\,ds]^{-1}\int_{-1}^{+1}tB(t)F(z,t)\psi(t)\,dt\}. \qquad (19a)$$

Here we have defined

$$R(z) = I - [C^{-1} - \textstyle\int_{-1}^{+1}B(s)\,ds]^{-1}\int_{-1}^{+1}tB^2(t)F(z,t)\,dt. \qquad (19b)$$

R is actually related to the dispersion matrix $\Lambda(z)$, Eq. (16a), by

$$R(z) = [C^{-1} - \textstyle\int_{-1}^{+1}B(s)\,ds]^{-1}\Sigma^{-1}\Lambda(t^{-1}(z))\,\Sigma^{-1}. \qquad (20)$$

Since $\det\Lambda(z)$ has a double zero at infinity, it follows that $\det R(z)$ will have a double zero at $t(\infty)=i$. The continuous spectrum of K transforms into the semicircle given by Eq. (5c) and the additional eigenvalues of K [zeros of $\Lambda(z)$] transform by $\nu_i\to t(\nu_i)$.

The eigenfunction expansion is again obtained by integrating the resolvent around the spectrum. The integration around the continuous spectrum can be transformed into the identical form found in Ref. 5 (or see the result for the subcritical situation which is also identical)[12] by the change of variable $z'=t^{-1}(z)$. Similarly the integrals about the isolated point eigenvalues ν_i can, by the same change of variables, be transformed into the expansions met in Refs. 13 and 5. Only the contribution from the double pole at $+i$ remains to be evaluated. Again the appropriate residue for a second order pole must must be used.

We proceed to evaluate $(1/2\pi i)\int_{\Gamma_i}(zI-S)^{-1}\psi(\mu)\,dz = I_1$. We have from Eqs. (19a) and (20)

$$I_1 = \frac{1}{2\pi i}\int_{\Gamma_i}\left(F(z,\mu)\,B(\mu)\,\frac{\Lambda_c(t^{-1}(z))}{\det\Lambda(t^{-1}(z))} \times\Sigma\int_{-1}^{+1}sB(s)F(z,s)\psi(s)\,ds\right)dz. \qquad (21)$$

From the diagonal expansion of the $\det\Lambda(z)$,[14] we find

$$\det\Lambda(z)=\det(\Sigma C^{-1}\Sigma-2\Sigma)+\frac{2}{3z^2}\operatorname{Tr}\Sigma^{-1}\Lambda_c(z)+O(1/z^4), \tag{22}$$

where $\Lambda_c(z)$ is the cofactor matrix of $\Lambda(z)$. Note by definition of the critical multigroup problem, the first term of the rhs of Eq. (22) vanishes. The second term gives the residue which we need. The result is

$$I_1=\tfrac{3}{2}[\operatorname{Tr}\Sigma^{-1}\Lambda_c(\infty)]^{-1}\{\mu\Sigma^{-1}\Lambda_c(\infty)\int_{-1}^{+1}s\psi(s)\,ds + \Lambda_c(\infty)\Sigma^{-1}\int_{-1}^{+1}s^2\psi(s)\,ds\}. \tag{23}$$

For use in solving transport problems, it is convenient, if not essential, to recast this result as expansion coefficients multiplying eigenvectors of K (or K^{-1}). This, in fact, is the form in which the result was expressed in Ref. 5. This is accomplished by representing $\Lambda_c(\infty)$ as (details found in Ref. 15)

$$\Lambda_c(\infty)=\tfrac{2}{3}\operatorname{Tr}[\Sigma^{-1}\Lambda_c(\infty)]\,\Sigma\xi\otimes\xi, \tag{24a}$$

where $\hat{\xi}$ and ξ are certain null vectors introduced by Ref. 5,

$$\Lambda(\infty)\xi=0 \tag{24b}$$

and

$$\Lambda^T(\infty)\Sigma\hat{\xi}=0. \tag{24c}$$

The normalization $\hat{\xi}^T\xi=\frac{3}{2}$ has been imposed. Using this representation, we obtain finally the eigenfunction expansion of Ref. 5, which is

$$\psi(\mu)=\sum_{i=1}^{2n}\psi_{\nu_i}+\psi_\Gamma+\int_{-1}^{+1}d\mu\,\mu^2[\psi(\mu),\hat{\xi}]\,\xi + \left(\int_{-1}^{+1}d\mu\,\mu[\psi(\mu),\Sigma\hat{\xi}]\right)\mu\Sigma^{-1}\xi. \tag{25}$$

The first term on the rhs is surely the contribution from the finite eigenvalues of K. This, along with ψ_I, is identical with the subcritical result obtained in Ref. 12. Only the contribution from the eigenvalue at infinity is essentially different in the critical case.

For the half-space expansion, again an "albedo operator" E must be introduced. This operator has precisely the same properties as in the one-speed case, Sec. II. The appropriate E is

$$(E\psi)_i(\sigma_i\mu)=\begin{cases}-[X^{-1}(\mu)\int_0^1 s(\mu-s)^{-1}Y^{-1}(-s)\Sigma^2\psi_\Gamma(s)\,ds]_i, & -1\le\sigma_i\mu\le 0,\\ \psi_i(\sigma_i\mu), & \mu>0,\end{cases} \tag{26}$$

where X and Y provide the Wiener–Hopf factorization of Λ, as in Ref. 5:

$$Y(-z)X(z)=\Lambda(z). \tag{27}$$

We now compute

$$(1/2\pi i)\int(zI-S)^{-1}E\psi(\mu)\,dz \tag{28}$$

about the spectrum of S and thereby obtain the half-range expansion formula,

$$\psi(\mu)=\sum_{i=1}^{n}\psi_{\Gamma_i}+\psi_\Gamma+\left(\int_{-1}^{+1}s^2\,ds[E\psi(s),\hat{\xi}]\right)\xi, \tag{29}$$

where ψ_{Γ_i} and ψ_Γ are defined in Ref. 12.

IV. CONCLUSION

We feel that the results of Refs. 2 and 5 have been obtained obtained in the present paper in a much simpler and thereby more elegant fashion. In particular we have avoided the introduction of subspaces Y_p and restrictions of operators, etc. However, we point out that the method described here is quite general and will permit us to study large classes of unbounded and/or noninvertible operators. The problems posed by critical neutron transport is that the point spectrum extends to infinity. The "Larsen transform" utilized here reduces both classes of problems to tractable form.

*Supported by the National Science Foundation, Grant ENG 75-15882.

†Based on a portion of a Ph.D thesis to be submitted to the Graduate School, Virginia Polytechnic Institute and State University.

[1]F. Shure and M. Natelson, Ann. Phys. (N.Y.) 26, 274 (1964).
[2]W. Greenberg and P.F. Zweifel, J. Math. Phys. 17, 163 (1976).
[3]E.W. Larsen and G.J. Habetler, Commun. Pure Appl. Math. 26, 525 (1973).
[4]C.G. Lekkerkerker, "The Linear Transport Equation, The Degenerate Case $c=1$," submitted for publication.
[5]R.L. Bowden, William Greenberg, and P.F. Zweifel, "Critical Multigroup Transport," to be published in SIAM J. Appl. Math.
[6]E.W. Larsen, "Hydrodynamic Approximation to Transport Problems," oral presentation to the Conference on Recent Developments in Kinetic Theories (June 1976).
[7]E.W. Larsen and J.B. Keller, J. Math. Phys. 15, 75 (1974).
[8]E. Bareiss, "A Spectral Theory for the Stationary Transport Operator in Slab Geometry," A.N.L.-6940 (1964).
[9]E.W. Larsen, private communication.
[10]Papers describing all of this work are currently in preparation by the present authors along with M. Arthur, L. Garbanati, and W. Greenberg. We acknowledge helpful discussions with these authors in preparation of this manuscript.
[11]W. Greenberg and S. Sancaktar, J. Math. Phys. 17, 2092 (1976)
[12]R.L. Bowden, S. Sancaktar, and P.F. Zweifel, J. Math. Phys. 17, 76, 82 (1976).
[13]K.M. Case and P.F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[14]F.E. Hohn, *Elementary Matrix Algebra* (Collier-Macmillan, London, 1964), p. 88.
[15]William L. Cameron, thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia (1977).

Prolongation structures and nonlinear evolution equations in two spatial dimensions. II. A generalized nonlinear Schrödinger equation

H. C. Morris

School of Mathematics, Trinity College, Dublin, Ireland
(Received 26 February 1976)

The prolongation structure approach of Wahlquist and Estabrook is used to determine an inverse scattering formulation for a generalization of the nonlinear Schrödinger equation to two spatial dimensions.

1. INTRODUCTION

In earlier work[1] we have shown how to generalize a prolongation structure[2] for an evolution equation in one spatial dimension to one for a generalized equation in two spatial dimensions. The equations

$$-i\frac{\partial}{\partial t}A=\left[\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)A-2A(\Phi-\Psi)\right], \quad (1.1)$$

$$\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)\Phi=-\frac{1}{2}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)(AA^*), \quad (1.2)$$

$$\left(\frac{\partial}{\partial y}+\frac{\partial}{\partial x}\right)\Psi=\frac{1}{2}\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)(AA^*), \quad (1.3)$$

represent a generalization of the nonlinear Schrödinger equation

$$i\frac{\partial}{\partial t}A=-\frac{\partial^2 A}{\partial x^2}+2A|A|^2 \quad (1.4)$$

to several spatial dimensions.

We will determine an inverse scattering problem for Eqs. (1.1)–(1.3) by adopting the following strategy. We will first develop, in Sec. 2, a previously unknown prolongation structure for the two-dimensional system

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)A=2A(\Phi-\Psi), \quad (1.5)$$

$$\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)\Phi=-\frac{1}{2}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)(AA^*), \quad (1.6)$$

$$\left(\frac{\partial}{\partial y}+\frac{\partial}{\partial x}\right)\Psi=\frac{1}{2}\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)(AA^*). \quad (1.7)$$

The general procedure previously developed[1] will then be used to extend that prolongation structure into one for the system (1.1)–(1.3). The Eqs. (1.5)–(1.7) are an example of a general class of equations which will be determined and developed in detail in the next of this series of papers. The method used is independent of that used by Ablowitz and Haberman[3] to determine Eqs. (1.1)–(1.3).

2. A PROLONGATION STRUCTURE FOR THE TWO-DIMENSIONAL SYSTEM

To determine a prolongation structure for the equations

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)A=+2A(\Phi-\Psi), \quad (2.1)$$

$$\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)\Phi=-\frac{1}{2}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)(AA^*), \quad (2.2)$$

$$\left(\frac{\partial}{\partial y}+\frac{\partial}{\partial x}\right)\Psi=\frac{1}{2}\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)(AA^*), \quad (2.3)$$

we must first settle on an appropriate closed set of 2-forms. It proves convenient to introduce the variables R and L defined by

$$R=\frac{1}{2}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)A, \quad (2.4)$$

$$L=\frac{1}{2}\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)A, \quad (2.5)$$

in which case Eqs. (2.1)–(2.3) are expressed in the form

$$\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)\Phi=-(RA^*+AR^*), \quad (2.6)$$

$$\left(\frac{\partial}{\partial y}+\frac{\partial}{\partial x}\right)\Psi=(LA^*+AL^*), \quad (2.7)$$

$$L_y=-R_x+A(\Phi-\Psi), \quad (2.8)$$

$$R_y^*=L_x^*+A^*(\Phi-\Psi), \quad (2.9)$$

where (2.9) is equivalent to the conjugate of (2.1).

These equations together with Eqs. (2.4) and (2.5), which define R and L, have an equivalent expression in terms of the closed ideal of eight 2-forms, α_1, α_2, α_1^*, α_2^*, α_3, α_4, α_5, α_6, defined by

$$\alpha_1=dA\wedge dy-(R-L)\,dx\wedge dy, \quad (2.10)$$

$$\alpha_2=dA\wedge dx+(R+L)\,dx\wedge dy, \quad (2.11)$$

$$\alpha_3=d\Phi\wedge(dx+dy)-(RA^*+AR^*)\,dx\wedge dy, \quad (2.12)$$

$$\alpha_4=d\Psi\wedge(dx-dy)+(LA^*+AL^*)\,dx\wedge dy, \quad (2.13)$$

$$\alpha_5=dL\wedge dx-dR\wedge dy+A(\Phi-\Psi)\,dx\wedge dy, \quad (2.14)$$

$$\alpha_6=dR^*\wedge dx+dL^*\wedge dy+A^*(\Phi-\Psi)\,dx\wedge dy. \quad (2.15)$$

Seeking a prolongation structure,

$$\Omega=d\zeta+F(A,A^*,R,R^*,L,L^*,\Phi,\Psi,\zeta)\,dx+G(A,A^*,R,R^*,L,L^*,\Phi,\Psi,\zeta)\,dy, \quad (2.16)$$

in the usual way, we discover that we can choose F and G to have the forms

$$F=x_1+x_2A+x_3A^*+Lx_4+R^*x_5+\Phi x_6+\Psi x_7, \quad (2.17)$$

$$G=x_8+x_9A+x_{10}A^*+L^*x_5-Rx_4+\Phi x_6-\Psi x_7, \quad (2.18)$$

and that they must satisfy the Lie bracket constraint

$$
\begin{aligned}
&[F,G]\\
&\quad=(R-L)GA-(R+L)F_A+(R^*-L^*)G_{A^*}-(R^*+L^*)F_{A^*}\\
&\qquad-(LA^*+LA^*)F_\Psi+(RA^*+AR^*)F_\Phi-A^*(\Phi-\Psi)F_{R^*}\\
&\qquad-A(\Phi-\Psi)F_L,
\end{aligned}
\tag{2.19}
$$

where we have used the notation $H_p=\partial H/\partial p$ for partial derivatives. Substitution of (2.17) and (2.18) into this relationship produces the bracket relations on the vector fields x_i. These consist of the simple bracket relations

$$
\begin{aligned}
&[x_1,x_8]=0,\quad [x_1,x_5]=-(x_{10}+x_3),\quad [x_1,x_4]=x_2-x_9,\quad [x_2,x_{11}]=0,\quad [x_2,x_4]=0,\quad [x_2,x_5]=-x_7,\quad [x_3,x_{10}]=0,\\
&[x_3,x_5]=0,\quad [x_3,x_4]=-x_6,\quad [x_4,x_5]=0,\quad [x_4,x_6]=0,\quad [x_4,x_7]=0,\quad [x_4,x_9]=0\quad [x_4,x_{10}]=-x_7,\quad [x_5,x_6]=0,\\
&[x_5,x_7]=0,\quad [x_5,x_8]=x_{10}-x_3,\quad [x_5,x_{10}]=0,\quad [x_6,x_7]=0,
\end{aligned}
\tag{2.20}
$$

together with the more complicated constraints

$$
\begin{aligned}
&[x_1,x_9]+[x_2,x_8]=0,\quad [x_1,x_{10}]+[x_3,x_8]=0,\quad [x_1,x_6]+[x_6,x_8]=0,\quad [x_7,x_8]-[x_1,x_7]=0,\quad [x_2,x_{10}]+[x_3,x_9]=0,\\
&[x_2,x_6]+[x_6,x_9]+x_4=0,\quad [x_2,x_7]+[x_9,x_7]+x_4=0,\quad [x_3,x_6]+[x_6+x_{10}]+x_5=0,\quad [x_3,x_7]+[x_{10},x_7]+x_5=0.
\end{aligned}
\tag{2.21}
$$

In order to determine a representation of this algebraic structure we adopt the procedure of completing the structure into a Lie algebra. In order to do this it is necessary to introduce additional vectors x_{11}, x_{12}, x_{13}, x_{14}, and x_{15} defined by

$$
[x_1,x_9]=-x_{11},\quad [x_1,x_{10}]=-x_{12},\quad [x_1,x_6]=-x_{13},\quad [x_7,x_8]=-x_{14},\quad [x_2,x_3]=-x_{15}.
$$

We are able to embed the algebraic structure we have found previously into the Lie algebra defined by the nonzero Lie bracket relations,

$$
\begin{aligned}
&[x_1,x_2]=x_{11},\quad [x_1,x_3]=-x_{12},\quad [x_1,x_4]=(x_2-x_9),\quad [x_1,x_5]=-(x_3+x_{10}),\quad [x_1,x_6]=-x_{13},\quad [x_1,x_7]=-x_{14},\\
&[x_1,x_9]=-x_{11},\quad [x_1,x_{10}]=-x_{12},\quad [x_1,x_{13}]=(x_1+x_8),\quad [x_1,x_{14}]=(x_8-x_1),\quad [x_1,x_{15}]=x_8,\quad [x_2,x_3]=-x_{15},\\
&[x_2,x_5]=-x_7,\quad [x_2,x_6]=-x_4,\quad [x_2,x_8]=x_{12},\quad [x_2,x_{12}]=\tfrac{1}{2}(x_1+x_8),\quad [x_2,x_{13}]=x_2,\quad [x_2,x_{14}]=x_2,\quad [x_2,x_{15}]=2x_2,\\
&[x_3,x_4]=-x_6,\quad [x_3,x_7]=-x_5,\quad [x_3,x_8]=x_{12},\quad [x_3,x_{11}]=\tfrac{1}{2}(x_8-x_1),\quad [x_3,x_{13}]=-x_3,\quad [x_3,x_{14}]=-x_3,\quad [x_3,x_{15}]=-2x_3,\\
&[x_4,x_{10}]=-x_7,\quad [x_4,x_{12}]=(x_{13}-x_{15}),\quad [x_{13}-x_{15}),\quad [x_4,x_{13}]=-x_4,\quad [x_4,x_{14}]=x_4,\quad [x_4,x_{15}]=x_4,\quad [x_5,x_8]=(x_{10},x_3),\\
&[x_5,x_9]=x_6,\quad [x_5,x_{11}]=(x_{14}+x_{15}),\quad [x_5,x_{13}]=-x_5,\quad [x_5,x_{14}]=x_5,\quad [x_5,x_{15}]=-x_5\quad [x_6,x_8]=x_{13},\\
&[x_6,x_{10}]=-x_5,\quad [x_6,x_{11}]=x_{11},\quad [x_6,x_{12}]=-x_3,\quad [x_6,x_{13}]=-2x_6,\quad [x_6,x_{15}]=-x_6,\quad [x_7,x_8]=-x_{14},\quad [x_7,x_9]=x_4,\\
&[x_7,x_{11}]=-x_2,\quad [x_7,x_{12}]=x_{12},\quad [x_7,x_{14}]=2x_7,\quad [x_7,x_{15}]=x_7,\quad [x_8,x_9]=-x_{11},\quad [x_8,x_{10}]=x_{12},\\
&[x_8,x_{13}]=(x_1+x_8),\quad [x_8,x_{14}]=(x_1-x_8),\quad [x_8,x_{15}]=x_1,\quad [x_9,x_{10}]=(x_{15}-x_{14}-x_{13}),\quad [x_9,x_{12}]=\tfrac{1}{2}(x_8-x_1),\\
&[x_9,x_{13}]=x_9,\quad [x_9,x_{14}]=x_9,\quad [x_{10},x_{11}]=-\tfrac{1}{2}(x_8+x_1),\quad [x_{10},x_{13}]=-x_{10},\quad [x_{10},x_{14}]=x_{10},\quad [x_{11},x_{13}]=x_{11},\\
&[x_{11},x_{14}]=-x_{11},\quad [x_{11},x_{15}]=x_{11},\quad [x_{12},x_{13}]=x_{12},\quad [x_{12},x_{14}]=-x_{12},\quad [x_{12},x_{15}]=-x_{12},
\end{aligned}
$$

with all other Lie brackets being zero.

3. FOUR-DIMENSIONAL REPRESENTATIONS

If one seeks a linear representation of the vectors x_i of the form

$$
x_i=\zeta^T R_i b,
\tag{3.1}
$$

where $b_i=\partial/\partial\zeta_i$, one discovers that a four-dimensional representation exists in a special form. Each of the matrices R_i can be expressed in terms of Kronecker products of the four two-dimensional matrices U_1, U_2, U_3, U_4 defined by

$$
U_1=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad U_2=\begin{pmatrix}0&1\\0&0\end{pmatrix},\quad U_3=\begin{pmatrix}0&0\\1&0\end{pmatrix},\quad U_4=\begin{pmatrix}0&0\\0&1\end{pmatrix}.
\tag{3.2}
$$

From the multiplication table of the U_i,

	U_1	U_2	U_3	U_4
U_1	U_1	U_2	0	0
U_2	0	0	U_1	U_2
U_3	U_3	U_4	0	0
U_4	0	0	U_3	U_4

, (3.3)

one can easily check that the R_i defined by

$$
\begin{aligned}
&R_1=-U_3\otimes(U_1-U_4),\quad R_2=-U_1\otimes U_3,\quad R_3=-U_1\otimes U_2\\
&R_4=-U_2\otimes U_3,\quad R_5=-U_2\otimes U_2,\quad R_6=-U_2\otimes U_1,\\
&R_7=-U_2\otimes U_4,\quad R_8=-U_3\otimes(U_1+U_4),\quad R_9=U_4\otimes U_3\\
&R_{10}=-U_4\otimes U_2,\quad R_{11}=-U_3\otimes U_3,\quad R_{12}=-U_3\otimes U_2,\\
&R_{13}=-(U_4-U_1)\otimes U_1,\quad R_{14}=-(U_1-U_4)\otimes U_4,\\
&R_{15}=U_1\otimes(U_1-U_4),
\end{aligned}
\tag{3.4}
$$

provide vectors x_i which form a representation of the algebra (2.22).

We observe that a constant, to play the role of an eigenvalue, in the resulting inverse scattering equations may be included by replacing R_1 by $\tilde{R}_1$ defined by

$$R_1 = R_1 + \lambda I \tag{3.5}$$

More generally, if C and D are any two matrices such that

$$[C, D] = 0, \tag{3.6}$$

$$[F, D] + [C, G] = 0, \tag{3.7}$$

then $F + \lambda C$ and $G + \lambda D$ also give rise to a solution of (2.19) and represent a general way of introducing a parameter. The case (3.5) corresponds to the trivial choice $C = I$, $D = 0$, and proves sufficient for illustratory purposes.

This representation leads to the prolongation structure

$$\begin{aligned}
\Omega^1 &= d\zeta^1 - (\lambda\zeta^1 + A\zeta^2 + \zeta^3)\,dx - (\zeta^3)\,dy,\\
\Omega^2 &= d\zeta^2 - (A^*\zeta^1 + \lambda\zeta^2 - \zeta^4)\,dx - (\zeta^4)\,dy,\\
\Omega^3 &= d\zeta^3 - (\Phi\zeta^1 + L\zeta^2 + \lambda\zeta^3)\,dx - (\Phi\zeta^1 - R\zeta^2 - A\zeta^4)\,dy,\\
\Omega^4 &= d\zeta^4 - (R^*\zeta^1 + \Psi\zeta^2 + \lambda\zeta^4)\,dx - (L^*\zeta^1 - \Psi\zeta^2 + A^*\zeta^3)\,dy.
\end{aligned} \tag{3.8}$$

Sectioning Ω^i onto a solution manifold of Eqs. (2.1)–(2.3) gives the inverse scattering problem

$$\begin{aligned}
\underline{\zeta}_x &= \left[\begin{array}{cc|cc} \lambda & A & 1 & 0 \\ A^* & \lambda & 0 & -1 \\ \hline \Phi & L & \lambda & 0 \\ R^* & \Psi & 0 & \lambda \end{array}\right] \underline{\zeta},\\
\underline{\zeta}_y &= \left[\begin{array}{cc|cc} \multicolumn{2}{c|}{\multirow{2}{*}{0}} & 1 & 0 \\ & & 0 & 1 \\ \hline \Phi & -R & 0 & -A \\ L^* & -\Psi & A^* & 0 \end{array}\right] \underline{\zeta}.
\end{aligned} \tag{3.9}$$

Alternatively if we note that the interchanges $U_1 \longleftrightarrow U_4$, $U_2 \longleftrightarrow U_3$ are an automorphism of the algebra (3.3) we can obtain a second representation of the algebra (2.22) and construct the following alternative prolongation structure:

$$\begin{aligned}
\Omega^1 &= d\zeta^1 + (-\lambda\zeta^1 - \Psi\zeta^3 - R^*\zeta^4)\,dx + (-A^*\zeta^2 + \Psi\zeta^3 - L^*\zeta^4)\,dy,\\
\Omega^2 &= d\zeta^2 + (-\lambda\zeta^2 - L\zeta^3 - \Phi\zeta^4)\,dx + (A\zeta^1 + R\zeta^3 - \Phi\zeta^4)\,dy,\\
\Omega^3 &= d\zeta^3 + (\zeta^1 - \lambda\zeta^3 - A^*\zeta^4)\,dx + (-\zeta^1)\,dy,\\
\Omega^4 &= d\zeta^4 + (-\zeta^2 - A\zeta^3 - \lambda\zeta^4)\,dx + (-\zeta^2)\,dy.
\end{aligned} \tag{3.10}$$

Sectioning this prolongation structure yields the inverse scattering problem

$$\begin{aligned}
\underline{\zeta}_x &= \left[\begin{array}{cc|cc} \lambda & 0 & \Psi & R^* \\ 0 & \lambda & L & \Phi \\ \hline -1 & 0 & \lambda & A^* \\ 0 & 1 & A & \lambda \end{array}\right] \underline{\zeta},\\
\underline{\zeta}_y &= \left[\begin{array}{cc|cc} 0 & A^* & -\Psi & L^* \\ -A & 0 & -R & \Phi \\ \hline 1 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ 0 & 1 & & \end{array}\right] \underline{\zeta}.
\end{aligned} \tag{3.11}$$

4. A PROLONGATION STRUCTURE FOR THE GENERALIZED SCHRÖDINGER EQUATION

Suppose that a two-dimensional evolution equation which can be expressed in terms of a closed set of 2-forms $\{\alpha_i\}$, $i = 1, 2, \ldots, N$, possesses a linear prolongation structure $\{\alpha_i, \Omega^\beta\}$, $i = 1, \ldots, N$, $\beta = 1, \ldots, M$, in which the 1-forms Ω^β are expressed by

$$\Omega^\beta = \sum_{\alpha=1}^{M} (F^\beta_\alpha\, dx + G^\beta_\alpha\, dy)\zeta^\alpha + d\zeta^\beta \tag{4.1}$$

and suppose that

$$d\Omega^\beta = \sum_{i=1}^{N} f^{\beta i}\alpha_i + \sum_{\gamma=1}^{M} \eta^\beta_\gamma \wedge \Omega^\gamma. \tag{4.2}$$

It has been shown[1] that the 2-forms $\{\bar{\Omega}^\beta\}$, defined by

$$\begin{aligned}
\bar{\Omega}^\beta = \Omega^\beta \wedge dt &+ \sum_{\gamma=1}^{M} (GA - FB)^\beta_\gamma \zeta^\gamma\, dx \wedge dy\\
&+ (A^\beta_\alpha\, dx + B^\beta_\alpha\, dy) \wedge d\zeta^\gamma,
\end{aligned} \tag{4.3}$$

where A and B are constant $(M \times M)$ matrices which satisfy the conditions

$$[A, B] = 0, \tag{4.4}$$

$$[G, A] + [B, F] = 0, \tag{4.5}$$

provide a prolongation structure for a three-dimensional evolution equation defined by a set of N 3-forms $\{\bar{\alpha}_i\}$, $i = 1, \ldots, N$, having the structure

$$\bar{\alpha}_i = \alpha_i \wedge dt, \quad i = 1, \ldots, K, \tag{4.6}$$

$$\bar{\alpha}_j = \alpha_j \wedge dt + \beta_i, \quad j = K+1, \ldots, N. \tag{4.7}$$

The forms β_j are defined by the equation

$$\sum_{i=K+1}^{N} f^{\beta i}\beta_i = [(dGA - dFB)\zeta]^\beta \wedge dx \wedge dy. \tag{4.8}$$

For further details of the procedure, the original work should be consulted. We note the similarity of Eqs. (4.4)–(4.5) with Eqs. (3.6)–(3.7), which shows that we could have delayed introducing additional parameters until this final stage. We have called the forms $\{\alpha_i\}$, $i = 1, \ldots, K$, which are basically unaltered, linearization forms and the remaining forms $\{\alpha_j\}$, $j = K+1, \ldots, N$, which are fundamentally generalized, the dynamic forms. We shall use this nomenclature in the remainder of this section.

For the system (2.10)–(2.15), the dynamic forms are α_3, α_4, α_5, α_6 and Eq. (4.8) becomes

$$-\left[\begin{array}{cc|c} \multicolumn{2}{c|}{0} & 0 \\ \hline \beta_3 & \beta_5 & \multirow{2}{*}{0} \\ \beta_6 & \beta_4 & \end{array}\right] = (dGA - dFB) \wedge dx \wedge dy, \tag{4.9}$$

where A and B must satisfy the Eqs. (4.4) and (4.5). The matrices

$$A = +\frac{i}{2}\left[\begin{array}{cc|c} \multicolumn{2}{c|}{0} & 0 \\ \hline 1 & 0 & \multirow{2}{*}{0} \\ 0 & 1 & \end{array}\right] \quad \text{and} \quad B = +\frac{i}{2}\left[\begin{array}{cc|c} \multicolumn{2}{c|}{0} & 0 \\ \hline 1 & 0 & \multirow{2}{*}{0} \\ 0 & -1 & \end{array}\right] \tag{4.10}$$

satisfy (4.4) and (4.5) and give

$$\begin{aligned}
&\beta_3 = 0, \quad \beta_4 = 0,\\
&\beta_5 = -\frac{i}{2}\, dA \wedge dx \wedge dy, \quad \beta_6 = \frac{i}{2}\, dA^* \wedge dx \wedge dy,
\end{aligned} \tag{4.11}$$

corresponding to the generalized dynamic forms

$$\tilde{\alpha}_5 = \alpha_5 \wedge dt - \frac{i}{2} dA \wedge dx \wedge dy,$$
$$\tilde{\alpha}_6 = \alpha_6 \wedge dt + \frac{i}{2} dA^* \wedge dx \wedge dy, \tag{4.12}$$

which yield to the generalized Schrödinger equations (1.1)–(1.3). The generalization of the prolongation structure (3.10) is given by

$$\Omega^1 = d\zeta^1 \wedge dt - (\lambda\zeta^1 + A\zeta^2 + \zeta^3) dx \wedge dt - \zeta^3 dy \wedge dt,$$
$$\Omega^2 = d\zeta^2 \wedge dt - (A^*\zeta^1 + \lambda\zeta^2 - \zeta^4) dx \wedge dt - \zeta^4 dy \wedge dt,$$
$$\Omega^3 = d\zeta^3 \wedge dt - (\Phi\zeta^1 + L\zeta^2 + \lambda\zeta^3) dx \wedge dt - (\Phi\zeta^1 - R\zeta^2 - A\zeta^4) dy \wedge dt + \frac{i}{2}[(\lambda\zeta^1 + A\zeta^2) dx \wedge dy + (dx + dy) \wedge d\zeta^1], \tag{4.13}$$
$$\Omega^4 = d\zeta^4 \wedge dt - (R^*\zeta^1 + \Psi\zeta^2 + \lambda\zeta^4) dx \wedge dt - (L^*\zeta^1 - \Psi\zeta^2 + A^*\zeta^3) dy \wedge dt + \frac{i}{2}[-(A^*\zeta^1 + \lambda\zeta^2) dx \wedge dy + (dx - dy) \wedge d\zeta^2].$$

For the alternative prolongation structure (3.11) we can choose

$$A = \frac{i}{2}\left[\begin{array}{c|cc} 0 & 1 & 0 \\ & 0 & 1 \\ \hline 0 & \multicolumn{2}{c}{0} \end{array}\right] \quad \text{and} \quad B = \frac{i}{2}\left[\begin{array}{c|cc} 0 & -1 & 0 \\ & 0 & 1 \\ \hline 0 & \multicolumn{2}{c}{0} \end{array}\right] \tag{4.14}$$

to yield the nonlinear Schrödinger equation (1.1)–(1.3). The corresponding prolongation structure can easily be calculated from (4.3). However, it is sufficient for our present purposes of constructing an inverse scattering problem to concentrate on the prolongation structure (4.13).

Sectioning onto a solution manifold of Eqs. (1.1)–(1.3) gives the equations

$$\zeta^1_x = \lambda\zeta^1 + A\zeta^2 + \zeta^3, \tag{4.15}$$

$$\zeta^1_y = \zeta^3, \tag{4.16}$$

$$\zeta^2_x = A^*\zeta^1 + \lambda\zeta^2 - \zeta^4, \tag{4.17}$$

$$\zeta^2_y = \zeta^4, \tag{4.18}$$

$$\zeta^3_x = \Phi\zeta^1 + L\zeta^2 + \lambda\zeta^3 - \tfrac{1}{2}i\zeta^1_t, \tag{4.19}$$

$$\zeta^3_y = \Phi\zeta^1 - R\zeta^2 - A\zeta^4 - \tfrac{1}{2}i\zeta^1_t, \tag{4.20}$$

$$\zeta^4_x = R^*\zeta^1 + \Psi\zeta^2 + \lambda\zeta^4 - \tfrac{1}{2}i\zeta^2_t, \tag{4.21}$$

$$\zeta^4_y = L^*\zeta^1 - \Psi\zeta^2 + A^*\zeta^3 + \tfrac{1}{2}i\zeta^2_t. \tag{4.22}$$

If $\zeta^i_t = 0$, then (4.15)–(4.22) reduce to Eqs. (3.11) as expected. If $\zeta^i_x = 0$ the Eqs. (4.15)–(4.18) become

$$\zeta^1_y = -\lambda\zeta^1 - A\zeta^2, \quad \zeta^2_y = A^*\zeta^1 + \lambda\zeta^2. \tag{4.23}$$

Equations (4.19) and (4.20) become

$$\tfrac{1}{2}i\zeta^1_t = -(\tfrac{1}{2}AA^* + \lambda^2)\zeta^1 + (\tfrac{1}{2}A_y - \lambda A)\zeta^2,$$
$$\tfrac{1}{2}i\zeta^2_t = (\tfrac{1}{2}A^*_y + \lambda A^*)\zeta^1 + (\tfrac{1}{2}AA^* + \lambda^2)\zeta^2, \tag{4.24}$$

which is equivalent to the standard Zakharov and Shabat[4,5] form of inverse scattering problem for the nonlinear Schrödinger equation

$$i\frac{\partial}{\partial t}A + \frac{\partial^2}{\partial y^2}A + 2|A|^2 A = 0. \tag{4.25}$$

The case $\zeta^i_y = 0$ can clearly be treated in the same way but is basically the same.

If we use the Eqs. (4.16) and (4.18) to eliminate ζ^3 and ζ^4, we can reduce Eqs. (4.15)–(4.22) to the more compact form

$$\zeta^1_x = \lambda\zeta^1 + A\zeta^2 + \zeta^1_y, \quad \zeta^2_x = A^*\zeta^1 + \lambda\zeta^2 - \zeta^2_y, \tag{4.26}$$

$$\tfrac{1}{2}i\zeta^1_t = \Phi\zeta^1 - R\zeta^2 - A\zeta^2_y - \zeta^1_{yy},$$
$$\tfrac{1}{2}i\zeta^2_t = -L^*\zeta^1 + \Psi\zeta^2 - A^*\zeta^1_y + \zeta^2_{yy}. \tag{4.27}$$

This is the form of the inverse scattering problem for the Eqs. (1.1)–(1.3). An alternative derivation of Eqs. (4.26)–(4.27), starting from a general form of inverse scattering problem and determining the Eqs. (1.1)–(1.3), has just been presented by Ablowitz and Haberman[3] and represents a complementary approach to the same problem.

[1]H.C. Morris, J. Math. Phys. 17, 1870 (1976).
[2]H. Wahlquist and F. Estabrook, J. Math. Phys. 16, 1 (1975).
[3]M. Ablowitz and R. Haberman, Phys. Rev. Lett. 35, 1185 (1975).
[4]V. Zakharov and A. Shabat, Zh. Eksp. Teor. Fiz. 61, 118 (1971) [Sov. Phys.-JETP 34, 62 (1972)].
[5]M. Ablowitz, D. Kaup, A. Newell, and H. Segur, Phys. Rev. Lett. 31, 125 (1973).

Axisymmetric stationary Brans–Dicke vacuum fields

B. K. Nayak* and R. N. Tiwari

Department of Mathematics, Indian Institute of Technology, Kharagpur-721302, India
(Received 22 March 1976)

It is shown that the axisymmetric stationary Brans–Dicke vacuum solutions can be obtained from the solutions of the axisymmetric stationary Einstein vacuum fields and also the axisymmetric static Brans–Dicke vacuum fields.

1. INTRODUCTION

To find the exact solutions to the highly nonlinear Einstein's field equations, the well-known methods of Majumdar,[1] Papapetrou,[2] Ehlers,[3] Buchdahl,[4] and Bonnor[5] have been found to be very much useful. With the help of these methods one can generate exact solutions from the known solutions of the simpler situations. In the Brans—Dicke (BD) theory,[6] the field equations being more complex, it is natural to see if the solutions to these field equations can be generated from the already known simpler solutions either of the BD theory or of the conventional Einstein theory. This, in a way, will not only reduce, to a large extent, the effort of solving much more involved nonlinear field equations, but will also provide a clear basis from the viewpoint of physical interpretation of the solutions of the BD theory. Along this line a few methods have been suggested by Janis *et al*,[7] Buchdahl,[8] and the present authors.[9] These, however, provide the exact, but only static solutions to the BD field equations. Recently, McIntosh[10] extended this further by establishing the method of generating the axisymmetric stationary BD vacuum solutions from the axisymmetric Einstein vacuum solutions. This method, however, is not direct. It, finally, involves a set of three differential equations to obtain the solution of the BD field equations. In the present paper we have established a direct procedure of generating the solutions by the method of identification. This method was used earlier by the present authors[11] to obtain exact plane symmetric BD vacuum solutions. It is interesting to note that the BD vacuum solutions so obtained go over to the Einstein vacuum solutions when the BD coupling parameter ω tends to infinity. This, of course, is in accordance with the requirement of the BD theory. Further, we have also obtained a theorem for generating the axisymmetric stationary BD vacuum solutions from the axisymmetric static BD vacuum fields.

2. FIELD EQUATIONS

In the canonical representation, the BD vacuum field equations are given as

$$R_{ij}=-\frac{\omega}{\phi^2}\,\phi_{,i}\phi_{,j}-\frac{1}{\phi}\,\phi_{;ij} \tag{2.1}$$

and

$$g^{ij}\phi_{;ij}=0, \tag{2.2}$$

where ω is the BD coupling parameter, and comma and semicolon followed by an index denote partial and covariant derivatives, respectively. The axisymmetric stationary metric is taken in the form

$$(ds)^2=\exp(2U)(dt+Wd\Phi)^2 - \exp(2K-2U)[(dx^1)^2+(dx^2)^2]-h^2\exp(-2U)(d\Phi)^2, \tag{2.3}$$

where U, W, K, and h are functions of x^1 and x^2 only. The significance of the choice of the stationary metric in this form has been thoroughly discussed by Matzner and Misner[12] and Misra and Pandey.[13] The surviving and independent field equations, from (2.1) and (2.2), for the line element (2.3) are

$$2[(U_{,2})^2-(U_{,1})^2]+\frac{2K_{,1}h_{,1}}{h}-\frac{2K_{,2}h_{,2}}{h} + \frac{\exp(4U)}{2h^2}[(W_{,1})^2-(W_{,2})^2]+\frac{1}{h}(h_{,22}-h_{,11}) = -(\omega+1)[(p_{,2})^2-(p_{,1})^2]-(p_{,22}-p_{,11}) - 2(K_{,1}-U_{,1})p_{,1}+2(K_{,2}-U_{,2})p_{,2}, \tag{2.4}$$

$$2U_{,1}U_{,2}-\frac{K_{,2}h_{,1}}{h}-\frac{K_{,1}h_{,2}}{h}-\frac{W_{,1}W_{,2}\exp(4U)}{2h^2}+\frac{1}{h}h_{,12} = -(\omega+1)p_{,1}p_{,2}-p_{,12}+(K_{,2}-U_{,2})p_{,1}+(K_{,1}-U_{,1})p_{,2}, \tag{2.5}$$

$$W_{,11}+W_{,22}-\frac{1}{h}(W_{,1}h_{,1}+W_{,2}h_{,2})+4(W_{,1}U_{,1}+W_{,2}U_{,2}) = -(W_{,1}p_{,1}+W_{,2}p_{,2}), \tag{2.6}$$

$$U_{,11}+U_{,22}+\frac{1}{h})(u_{,1}h_{,1}+U_{,2}h_{,2})+\frac{\exp(4U)}{2h^2}[(W_{,1})^2+(W_{,2})^2] = -(U_{,1}p_{,1}+U_{,2}p_{,2}), \tag{2.7}$$

$$h_{,11}+h_{,22}=-(h_{,1}p_{,1}+h_{,2}p_{,2}), \tag{2.8}$$

and

$$p_{,11}+p_{,22}+(p_{,1})^2+(p_{,2})^2=-\frac{1}{h}(h_{,1}p_{,1}+h_{,2}p_{,2}), \tag{2.9}$$

where $\phi=e^p$ and subscripts 1 and 2 following a comma denote partial differentiation with respect to x^1 and x^2 respectively.

3. SOLUTIONS FROM THE AXISYMMETRIC STATIONARY EINSTEIN VACUUM FIELDS

Let us consider Einstein vacuum field equations ($R_{ij}=0$) corresponding to the metric

$$(ds)^2=\exp(2V)(dt+Wd\Phi)^2-\exp(2K-2V) \times[(dx^1)^2+(dx^2)^2]-H^2\exp(-2V)(d\Phi)^2, \tag{3.1}$$

where W and K are the same as those given in (2.3). This set of equations (given in the Appendix and to be referred as A) suggests that H satisfies the equation

$$H_{,11}+H_{,22}=0. \tag{3.2}$$

From (2.8) and (2.9), it can also be found that he^p satisfies a similar equation, viz.,

$$(he^p)_{,11}+(he^p)_{,22}=0. \tag{3.3}$$

This leads us to identify H with he^p as

$$H=he^p. \tag{3.4}$$

In view of (3.4) and the following relation:

$$V=U+\tfrac{1}{2}p, \tag{3.5}$$

the set of equations A reduces in terms of U, p, h, K, and W to

$$2[(U_{,2})^2-(U_{,1})^2]+\frac{2K_{,1}h_{,1}}{h}-\frac{2K_{,2}h_{,2}}{h}+\frac{\exp(4U)}{2h^2}[(W_{,1})^2-(W_{,2})^2]+\frac{1}{h}(h_{,22}-h_{,11})$$
$$=-\tfrac{3}{2}[(p_{,2})^2-(p_{,1})^2]-(p_{,22}-p_{,11})-2(K_{,1}-U_{,1})p_{,1}+2(K_{,2}-U_{,2})p_{,2}+\frac{2}{h}(h_{,1}p_{,1}-h_{,2}p_{,2}), \tag{3.6}$$

$$2U_{,1}U_{,2}-\frac{K_{,2}h_{,1}}{h}-\frac{K_{,1}h_{,2}}{h}-\frac{W_{,1}W_{,2}}{2h^2}\exp(4U)+\frac{1}{h}h_{,12}$$
$$=-\tfrac{3}{2}p_{,1}p_{,2}-p_{,12}+(K_{,2}-U_{,2})p_{,1}+(K_{,1}-U_{,1})p_{,2}-\frac{1}{h}(h_{,1}p_{,2}+h_{,2}p_{,1}), \tag{3.7}$$

$$W_{,11}+W_{,22}-\frac{1}{h}(W_{,1}h_{,1}+W_{,2}h_{,2})+4(W_{,1}U_{,1}+W_{,2}U_{,2})=-(W_{,1}p_{,1}+W_{,2}p_{,2}), \tag{3.8}$$

$$U_{,11}+U_{,22}+\frac{1}{h}(U_{,1}h_{,1}+U_{,2}h_{,2})+\frac{\exp(4U)}{2h^2}[(W_{,1})^2+(W_{,2})^2]+\frac{1}{2}\left[p_{,11}+p_{,22}+(p_{,1})^2+(p_{,2})^2+\frac{1}{h}(h_{,1}p_{,1}+h_{,2}p_{,2})\right]$$
$$=-(U_{,1}p_{,1}+U_{,2}p_{,2}), \tag{3.9}$$

and

$$(h_{,11}+h_{,22})+2(h_{,1}p_{,1}+h_{,2}p_{,2})+h[p_{,11}+p_{,22}+(p_{,1})^2+(p_{,2})^2]=0. \tag{3.10}$$

On comparing the Eqs. (3.6)–(3.9) with (2.4)–(2.7) we observe that they are equivalent if the following relationships are satisfied:

$$(\omega-\tfrac{1}{2})[(p_{,2})^2-(p_{,1})^2]=(2/h)(h_{,2}p_{,2}-h_{,1}p_{,1}), \tag{3.11}$$

$$(\omega-\tfrac{1}{2})p_{,2}p_{,1}=(1/h)(h_{,1}p_{,2}+h_{,2}p_{,1}), \tag{3.12}$$

and

$$p_{,11}+p_{,22}+(p_{,1})^2+(p_{,2})^2=-(1/h)(h_{,1}p_{,2}+h_{,2}p_{,2}). \tag{3.13}$$

The Eq. (3.10) together with (3.13) is equivalent to the set of Eqs. (2.8) and (2.9). Hence the set of Eqs. (3.6)–(3.10) along with (3.13) is equivalent to the set of Eqs. (2.4)–(2.9), provided the conditions (3.11) and (3.12) are satisfied. That is, the BD vacuum axisymmetric stationary solutions are obtainable from the Einstein vacuum axisymmetric stationary solutions when the relations (3.11) and (3.12) are valid. These two conditions, however, suggest a relationship between p and h as

$$p=[4/(2\omega-1)]\ln h, \tag{3.14}$$

which in view of (3.4) determines p and h, in terms of H as

$$p=[4/(2\omega+3)]\ln H \quad\text{and}\quad h=[H]^{(2\omega-1)/(2\omega+3)}. \tag{3.15}$$

Thus, we find: Given any Einstein vacuum axisymmetric stationary solution (U_0, W_0, K_0, H_0), one can always generate a corresponding BD vacuum axisymmetric stationary solution (U_B, W_B, K_B, H_B), where

$$U_B=U_0-\tfrac{1}{2}\ln\phi,\quad W_B=W_0,$$
$$K_B=K_0,\quad H_B=[H_0]^{(2\omega-1)/(2\omega+3)},$$

with the BD scalar ϕ being given as

$$\phi=[H_0]^{4/(2\omega+3)}.$$

A useful application of this theorem is made to obtain the BD vacuum solution corresponding to the Kerr solution. The Kerr solution[14] in the form of the metric (2.3) is given as[13]

$$(ds)^2=(dt)^2-(L^2+a^2\cos^2\theta)[(d\theta)^2+(dR)^2]-(L^2+a^2)\sin^2\theta(d\Phi)^2-\frac{2mL}{L^2+a^2\cos^2\theta}(dt+a\sin^2\theta\, d\Phi)^2, \tag{3.16}$$

where $L=e^R+m+[(m^2-a^2)/4]e^{-R}$. In view of the above mentioned theorem, the BD vacuum solution corresponding to (3.16) is

$$(ds)^2=\frac{L^2-2mL+a^2\cos^2\theta}{L^2+a^2\cos^2\theta}\times(L^2-2mL+a^2)^{-2/(2\omega+3)}\sin^{-4/(2\omega+3)}\theta\left[dt-\frac{2amL}{L^2-2mL+a^2\cos^2\theta}d\Phi\right]^2-(L^2-2mL+a)^{2/(2\omega+3)}$$
$$\times\sin^{4/(2\omega+3)}\theta\times(L^2+a^2\cos^2\theta)[(d\theta)^2+(dR)^2]-(L^2-2mL+a^2)^{(2\omega+5)/(2\omega+3)}\sin^{2(2\omega+5)/(2\omega+3)}\theta$$
$$\times\frac{L^2+a^2\cos^2\theta}{L^2-2mL+a^2\cos^2\theta}(d\Phi)^2 \tag{3.17}$$

with the BD scalar ϕ given as

$$\phi=(L^2-2mL+a^2)^{2/(2\omega+3)}\sin^{4/(2\omega+3)}\theta. \tag{3.18}$$

It can be verified that for $\omega\to\infty$, (3.17) reduces to (3.16) and ϕ approaches unity. This agreement makes this BD vacuum solution physically interesting and hence deserves a critical investigation.

4. SOLUTIONS FROM THE AXISYMMETRIC STATIC BRANS-DICKE VACUUM FIELDS

In the following steps, we show that the set of equations (2.4)–(2.9) can be reduced to the set of BD vacuum axisymmetric static field equations. This is done by introducing an auxiliary function L as

$$e^{-2U}=\lambda e^p\cosh 2L, \tag{4.1}$$

and defining two relations by

$$W_{,1}=-2\lambda he^{p}L_{,2} \quad\text{and}\quad W_{,2}=2\lambda he^{p}L_{,1}, \tag{4.2}$$

where λ is any arbitrary constant. With the help of (4.1) and (4.2), the relation (2.6) is identically satisfied and (2.7) reduces to

$$L_{,11}+L_{,22}+\frac{1}{h}(L_{,1}h_{,1}+L_{,2}h_{,2})=-(L_{,1}p_{,1}+L_{,2}p_{,2}). \tag{4.3}$$

Equation (2.4), in this case, reduces to

$$\begin{aligned}2[(L_{,2})^2-(L_{,1})^2]+\frac{2X_{,1}h_{,1}}{h}-\frac{2X_{,2}h_{,2}}{h}+\frac{1}{h}(h_{,22}-h_{,11})\\ =-(\omega+1)[(p_{,2})^2-(p_{,1})^2]-(p_{,22}-p_{,11})\\ -2(X_{,1}-L_{,1})p_{,1}+2(X_{,2}-L_{,2})p_{,2},\end{aligned} \tag{4.4}$$

where the function X is defined to be such that

$$\begin{aligned}X_{,1}&=K_{,1}+\frac{L_{,1}+\frac{1}{4}p_{,1}}{p_{,1}+h_{,1}/h}\,p_{,1},\\ X_{,2}&=K_{,2}+\frac{L_{,2}+\frac{1}{4}p_{,2}}{p_{,2}+h_{,2}/h}\,p_{,2}.\end{aligned} \tag{4.5}$$

Equation (2.5) reduces to

$$\begin{aligned}2L_{,1}L_{,2}-\frac{X_{,1}h_{,2}}{h}-\frac{X_{,2}h_{,1}}{h}+\frac{1}{h}h_{,12}\\ =-(\omega+1)p_{,1}p_{,2}-p_{,12}+(X_{,1}-L_{,1})p_{,2}+(X_{,2}-L_{,2})p_{,1},\end{aligned} \tag{4.6}$$

when the BD scalar $\phi=e^{p}$ satisfies

$$\phi_{,2}h_{,1}=\phi_{,1}h_{,2}. \tag{4.7}$$

With (4.7) being valid, the integrability condition for Eq. (4.5) requires the BD scalar ϕ to satisfy

$$L_{,1}\phi_{,2}=L_{,2}\phi_{,1}. \tag{4.8}$$

But, in this process, the Eqs. (2.8) and (2.9) are not affected. Hence, with the relation (4.7) and (4.8) being valid, the Eqs. (4.3)—(4.6) along with (2.8) and (2.9), now constitute the set of BD vacuum axisymmetric static field equations corresponding to the metric

$$(ds)^2=e^{2L}(dt)^2-e^{2X-2L}[(dx^1)^2+(dx^2)^2]-h^2e^{-2L}(d\Phi)^2. \tag{4.9}$$

Thus, given any axisymmetric static BD vacuum solution (L,X,h) along with the scalar ϕ satisfying (4.7) and (4.8), one can construct the corresponding axisymmetric stationary BD vacuum solution (U, W, K, h) with the same BD scalar ϕ, where U, W, and K are determined from (4.1), (4.2), and (4.5), respectively.

It may be remarked here, that for $\phi=$const this theorem reduces to that of Misra and Pandey[15] obtained for axisymmetric stationary Einstein vacuum fields.

ACKNOWLEDGMENT

The authors are thankful to Professor G. Bandyopadhyay and Professor J. R. Rao for their constant encouragement during the progress of this work.

APPENDIX

The set of independent equations from the Einstein vacuum field equations $R_{ij}=0$ corresponding to the metric (3.1) are given by

$$\begin{aligned}2[(V_{,2})^2-(V_{,1})^2]+\frac{2K_{,1}H_{,1}}{H}-\frac{2K_{,2}H_{,2}}{H}\\ +\frac{e^{4V}}{2H^2}[(W_{,1})^2-(W_{,2})^2]+\frac{1}{H}(H_{,22}-H_{,11})=0,\end{aligned}$$

$$2V_{,2}V_{,1}-\frac{K_{,2}H_{,1}}{H}-\frac{K_{,1}H_{,2}}{H}-\frac{e^{4V}}{2H^2}W_{,1}W_{,2}+\frac{1}{H}H_{,12}=0,$$

$$W_{,11}+W_{,22}-\frac{1}{H}(W_{,1}H_{,1}+W_{,2}H_{,2})+4(W_{,1}V_{,1}+W_{,2}V_{,2})=0,$$

$$V_{,11}+V_{,22}+\frac{1}{H}(V_{,1}H_{,1}+V_{,2}H_{,2})+\frac{e^{4V}}{2H^2}[(W_{,1})^2+(W_{,2})^2]=0,$$

and

$$H_{,11}+H_{,22}=0.$$

*Present address: Department of Mathematics, Government Evening College, Cuttack-3, Orissa, India.

[1]S. D. Majumdar, Phys. Rev. **72**, 390 (1947).
[2]A. Papapetrou, Proc. R. Irish Acad. A**51**, 191 (1947); Ann. Physik **12**, 309 (1953).
[3]J. Ehlers, thesis, Hamburg (1957).
[4]H. A. Buchdahl, Phys. Rev. **115**, 1325 (1959).
[5]W. B. Bonnor, Z. Phys. **151**, 439 (1961).
[6]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[7]A. I. Janis, D. C. Robinson, and J. Wincour, Phys. Rev. **186**, 1729 (1969).
[8]H. A. Buchdahl, Int. J. Theor. Phys. **6**, 407 (1972).
[9]R. N. Tiwari and B. K. Nayak, "Class of the Brans—Dicke Maxwell fields," Phys. Rev. (in press).
[10]C. B. G. McIntosh, Commun. Math. Phys. **37**, 335 (1974).
[11]R. N. Tiwari and B. K. Nayak, J. Phys. A **9**, 369 (1976).
[12]R. A. Matzner and C. W. Misner, Phys. Rev. **154**, 1229 (1967).
[13]R. M. Misra and D. B. Pandey, J. Math. Phys. **13**, 1538 (1972).
[14]R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[15]R. M. Misra and D. B. Pandey, Phys. Lett. A **34**, 245 (1971).

The n-bubble series in the theory of the classical one-component plasma

Y. Furutani* and C. Deutsch

Laboratoire de Physique des Plasmas, Laboratoire Associé au C.N.R.S., Université de Paris—Sud, Centre d'Orsay, 91405 Orsay, France

(Received 8 March 1976)

With the aid of the Mellin transform, an exact expression for the three-dimensional Fourier transform $G(k)$ of the renormalized sum $\exp[-\Lambda\exp(-r)/r]-1+\Lambda\exp(-r)/r$ for n-bubble graphs is given in two equivalent forms, where the plasma parameter Λ is not necessarily smaller than unity. Its several properties, such as small and large k limits, are discussed in detail.

INTRODUCTION

In the usual nodal expansion[1–4] of the potential of average force $w_2(r)=k_BT\ln g_2(r)$ for the pair correlation function $g_2(r)$ of the classical one-component Coulomb gas, we meet the well-known difficulty[2,3] that the series of $w_2(r)$ with respect to the plasma parameter $\Lambda=e^2/(k_BT\lambda_D)$, where λ_D is the Debye length, cannot be pursued beyond the second-order term, when only the long-range (Debye) resummation of the bare Coulomb interaction $1/r$ is taken into account. This stems from the nonexistence of the Fourier transform of the Debye potential

$$(4\pi/k)\int_0^\infty dr\, r\sin(kr)(e^{-r}/r)^n$$

for $n\geq 3$. To circumvent this difficulty, we have recourse to the well-known trick due to Abe and Meeron[5] which consists in the resummation to all orders of the most diverging n bubbles, at $r=0$, with n Debye lines curved together between the (two) root points, through the expression (Fig. 1)

$$G(r)=\sum_{n=1}^{\infty}\frac{(-\Lambda)^n}{n!}\left(\frac{e^{-r}}{r}\right)^n=\exp(-\Lambda e^{-r}/r)-1+\frac{\Lambda e^{-r}}{r},\qquad(1)$$

where r is measured in units of λ_D. In order that the nodal expansion be worked out for $w_2(r)$ in the Fourier space, the actual quantity of interest is

$$G(k)=\frac{4\pi}{k}\int_0^\infty dr\, r\sin(kr)[\exp(-\Lambda\exp(-\alpha r)/r)-1+\Lambda(\exp(-\alpha r)/r)],\qquad(2)$$

where an arbitrary α ($\mathrm{Re}\,\alpha>0$) is retained for the sake of generality.[6] This problem has apparently been treated by several authors[3,7] in the past. To our knowledge, however, neither an explicit expression of $G(k)$ nor a general discussion of its properties has been given in the literature. Also, let us notice that the technique of the modified Mellin transform, first used by Iwata[7] to study the simpler quadrature

$$J=2\pi\rho^2\int_0^\infty dr\, r^2[\exp(-\beta q(r))-1+\beta q(r)-(\beta^2/2!)q^2(r)],$$
$$q(r)=e^2\exp(-k_Dr)/r,$$

is quite a powerful tool for the evaluation of the type (2).

During the study of the asymptotic behavior of chain graphs in which 2 bubbles are combinatorially replaced by higher-order bridge graphs, graphs which are not taken into account in the usual hypernetted-chain equation, we have had to renormalize a given bridge graph of order n with l Debye lines and k nodal points ($n=l-k$). We also found it necessary to give a qualitative indication on how bridge graphs of lower order ($n=3$ and 4) behave at large distances (for the Fourier transform, this corresponds to the small k limit).

The purpose of the present work is therefore to give mathematical support to our previous work[8] on the asymptotic behavior of the equilibrium pair correlation function in dense electron gas. The paper is organized thus: Analytic derivation of Eq. (2) is given in Sec. 1. Section 2 is devoted to an alternative evaluation of $G(k)$ by direct expansion of $\sin kr$ in Eq. (2). Equivalence of two results, Eqs. (10) and (15), is also discussed. After having studied the convergency of the series (10), in Sec. 3 we examine in detail both small and large k limits of $G(k)$, corresponding respectively to large and short distances of $G(r)$.

I. EVALUATION OF $G(k)$

The three-dimensional Fourier transform of $G(r)$, Eq. (2), can be obtained straightforwardly by means of the Mellin transform

$$e^{-x}-1+x=\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}ds\,\Gamma(s)x^{-s},\quad -2<\sigma<-1,\qquad(3)$$

where $\Gamma(s)$ is the gamma function. Setting $x=\Lambda e^{-r}/r$, we obtain

$$G(k)=\frac{4\pi}{k}\int_0^\infty dr\, r\sin(kr)\frac{1}{2\pi i}\times\int_{\sigma-i\infty}^{\sigma+i\infty}ds\,\Gamma(s)\left(\frac{\Lambda}{r}\exp(-\alpha r)\right)^{-s},\quad -2<\sigma<-1$$
$$=\frac{4\pi}{k}\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}ds\,\Gamma(s)\Lambda^{-s}\times\int_0^\infty dr\, r^{s+1}\sin(kr)\exp(\alpha sr),\quad \alpha<0,\qquad(4)$$

with $\mathrm{Re}\,s\leq\sigma<-1$. The last step is justified by the absolute summability of the integrand with respect to s and r, respectively (Fubini's theorem). Now, using the formula[9]

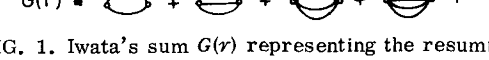

FIG. 1. Iwata's sum $G(r)$ representing the resummation to all orders of the most diverging n bubbles at the origin.

$$\int_0^\infty dt\, t^{\nu-1}\exp(-at)\sin(bt)=\Gamma(\nu)(a^2+b^2)^{-\nu/2}\times\sin\left(\nu\arctan\left(\frac{b}{a}\right)\right),$$

$$b>0,\quad \mathrm{Re}\,a>0,\quad\text{and}\ \mathrm{Re}\,\nu>-1, \tag{5}$$

we obtain

$$G(k)=\frac{4\pi\Lambda^2}{k}\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}ds\Gamma(s)\Gamma(s+2)(a^2s^2+b^2)^{-(s+2)/2}\times\sin\left[(s+2)\arctan\left(\frac{b}{-as}\right)\right], \tag{6}$$

where $a=\alpha\Lambda$ and $b=k\Lambda$. The restriction $\mathrm{Re}(s+2)>-1$, necessary for $\Gamma(s+2)$ to be analytic at the right of $s=-1$, can be easily relaxed by analytic continuation of the integrand onto the left of the Mellin contour, where $\Gamma(s)\Gamma(s+2)$ has a double pole at $s=-n$, $n\geq 2$.

In order to prove that there is no contribution to $G(k)$ from the large semicircle at the left of the Mellin contour, we first expand the integrand for large $|s|$. With the aid of the identities

$$F(s)\equiv\sin\left((s+2)\arctan\left(\frac{b}{-as}\right)\right)\simeq\sin\left(-\frac{b}{a}+\frac{b^3}{3a^3s^2}+\cdots\right)\to-\sin\left(\frac{b}{a}\right),\quad |s|\gg 1$$

and of Sterling's formula

$$\ln[\Gamma(s)\Gamma(s+2)]\simeq(2s+1)\ln s-2s+\ln 2\pi,$$

we have

$$\lim_{|s|\to\infty}\left|\Gamma(s)\Gamma(s+2)(a^2s^2+b^2)^{-(s+2)/2}F(s)\right|=2\pi\frac{s^{s-1}e^{-2s}}{a^{s+2}}\sin\left(\frac{b}{a}\right)\equiv A.$$

Upon putting $s=\rho\exp(i\theta)$, we see

$$\mathrm{Re}\ln A=\ln\left(\frac{2\pi}{a^2}\sin\frac{b}{a}\right)+\rho\left[\ln\left(\frac{\rho}{a}\right)\cos(\theta)-\left[\theta\sin(\theta)+2\cos(\theta)\right]\right]-\ln\rho,$$

$$\pi/2+\eta<\theta<3\pi/2-\eta$$

with $\lim_{\rho\to\infty}\eta\to+0$. Since $\cos\theta<0$, $\mathrm{Re}\ln A\overset{\rho\to\infty}{\to}-\infty$. Thus $\mathrm{Re}A\to 0$. Q.E.D.

We are now ready to evaluate the residues of Eq. (6) at the double pole $s=-n$ of $\Gamma(s)\Gamma(s+2)$. To do this the Cauchy's power series expansion of $\Gamma(s)$[10] is of the order

$$\lim_{s\to -n}\Gamma(s)=\frac{(-)^n}{n!}\left(\frac{1}{s+n}+\psi(n+1)+\tfrac{1}{2}(s+n)A_n+O[(s+n)^2]\right), \tag{7}$$

where $A_n=\pi^2/2+\psi^2(n+1)-\psi'(n+1)$ and $\psi(z)=\Gamma'(z)/\Gamma(z)$ is the di-gamma function. The straightforward calculation then yields

$$\Gamma(s)\Gamma(s+2)=\frac{1}{\Gamma(n+1)\Gamma(n-1)(s+n)^2}\{1+(s+n)[\psi(n+1)+\psi(n-1)]+\tfrac{1}{2}(s+n)^2[A_n+A_{n-2}+2\psi(n-1)\psi(n+1)]+O[(s+n)^3]\}. \tag{8}$$

Now, given an entire function $H(s)$ in the left of the Mellin contour, we have

$$\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}ds\,\Gamma(s)\Gamma(s+2)H(s)=\sum_{n=2}^{\infty}\left(\frac{d}{ds}[\Gamma(s)\Gamma(s+2)H(s)(s+n)^2]\right)_{s=-n}$$

$$=\sum_{n=2}^{\infty}\frac{1}{\Gamma(n-1)\Gamma(n+1)}\left[[\psi(n-1)+\psi(n+1)]H(-n)+\left(\frac{d}{ds}H(s)\right)_{s=-n}\right],\quad -2<\sigma<-1. \tag{9}$$

Putting $H(s)=(a^2s^2+b^2)^{-(s+2)/2}\sin[(s+2)\arctan(b/-as)]$, we obtain successively

$$H(-n)=-\left(\frac{an}{\cos\theta_n}\right)^{n-2}\sin((n-2)\theta_n)$$

and

$$H'(-n)=\left(\frac{an}{\cos(\theta_n)}\right)^{n-2}\left[\ln\left(\frac{an}{\cos(\theta_n)}\right)\sin((n-2)\theta_n)+\theta_n\cos((n-2)\theta_n)+\frac{n-2}{n}\cos\theta_n\sin((n-3)\theta_n)\right],$$

where $\tan\theta_n=b/an$, $0\leq\theta_n<\pi/2$.

Finally we get the desired expression

$$G(k)=\frac{4\pi\Lambda^2}{k}\sum_{n=2}^{\infty}\frac{1}{\Gamma(n-1)\Gamma(n+1)}\left(\frac{an}{\cos(\theta_n)}\right)^{n-2}\cdot\left\{\left[\ln\left(\frac{an}{\cos(\theta_n)}\right)-\psi(n-1)-\psi(n+1)\right]\sin((n-2)\theta_n)+\theta_n\cos((n-1)\theta_n)+\frac{n-2}{n}\cos\theta_n\sin((n-3)\theta_n)\right\}. \tag{10}$$

This expression is equivalent to that given by Del Rio and DeWitt.[3] To see this, it is sufficient to set, in their Eq. (19), $n-ik=(n^2+k^2)^{1/2}\exp(-i\theta_n)$ with $\tan(\theta_n)=k/n$. Then after simple algebra we obtain[11]

$$G(k)_{RW}=\frac{1}{4\pi\Lambda}G(k)_{DF}.$$

II. ALTERNATIVE EVALUATION OF $G(k)$ BY EXPANSION OF sin (kr) IN EQ. (4)

Equation (4) can be rewritten as

$$G(k)=\frac{4\pi}{k}\sum_{l=0}^{\infty}\frac{(-)^l}{(2l+1)!}\int_0^\infty dr\,r(kr)^{2l+1}\times[\exp(-\Lambda e^{-\alpha r}/r)-1+\Lambda e^{-\alpha r}/r]$$

$$=4\pi\sum_{l=0}^{\infty}\frac{(-k^2)^l}{(2l+1)!}\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}ds\,\Gamma(s)\Lambda^{-s}\times\int_0^\infty dr\,r^{2l+2+s}\exp(\alpha sr),\quad \mathrm{Re}s<0,\quad -2<\sigma<-1. \tag{11}$$

Now with the r integration carried out with the change of variable $r=\Lambda t$, we get

$$G(k)=\frac{4\pi\Lambda^2}{k}\sum_{l=0}^{\infty}(-)^l\frac{(k\Lambda)^{2l+1}}{(2l+l)!}\frac{1}{2\pi i}\times\int_{\sigma-i\infty}^{\sigma+i\infty}ds\,\Gamma(s)\Gamma(s+2l+3)(-\alpha\Lambda s)^{-(s+2l+3)}. \tag{12}$$

Next, since in the vicinity of the double pole $s=-n$, $\Gamma(s)\Gamma(s+2l+3)$ can be expressed as

$$\Gamma(s)\Gamma(s+2l+3) = -\frac{1}{\Gamma(n+1)\Gamma(n-2l-2)(s+n)^2}$$
$$\times\{1+(s+n)[\psi(n+1)+\psi(n-2l-2)]$$
$$+\tfrac{1}{2}(s+n)^2[A_n + A_{n-2l-3}$$
$$+2\psi(n+1)\psi(n-2l-2)] + O[(s+n)^3]\}, \tag{13}$$

the s integration straightforwardly gives

$$\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} ds\,\Gamma(s)\Gamma(s+2l+3)(-\alpha\Lambda s)^{-(s+2l+3)}$$
$$=\sum_{n=2}^{\infty}\frac{(an)^{n-2l-3}}{\Gamma(n+1)\Gamma(n-2l-2)}\left(\ln(an)+\frac{n-2l-3}{n}\right.$$
$$\left.-\psi(n+1)-\psi(n-2l-2)\right), \quad a=\alpha\Lambda. \tag{14}$$

Finally $G(k)$ is given by

$$G(k)=\frac{4\pi\Lambda^2}{k}\sum_{l=0}^{\infty}(-)^l\frac{(k\Lambda)^{2l+1}}{(2l+1)!}\sum_{n=2}^{\infty}\frac{an^{n-2l-3}}{\Gamma(n+1)\Gamma(n-2l-2)}$$
$$\times\left(\ln(an)+\frac{n-2l-3}{n}-\psi(n+1)-\psi(n-2l-2)\right). \tag{15}$$

The radius of convergence of the series, associated with that of $\sin kr$, is obviously ∞. The two forms, Eqs. (10) and (15), are not *a priori* the same. In Appendix A it is shown that they are indeed identical.

III. MISCELLANEOUS PROPERTIES OF $G(k)$

A. Convergence of the series (10)

The series is not uniformly but absolutely convergent. To prove this we use for large n the asymptotic expression of the gamma and the di-gamma functions

$$\Gamma(n+1)\simeq\sqrt{2\pi}\,n^{n+1/2}e^{-n}$$

and

$$\psi(n)\simeq \ln n-\frac{1}{2n}-\sum_{l=1}^{\infty}\frac{B_{2l}}{2ln^{2l}},$$

where B_{2l} is the Bernouilli number. Denote by u_n the nth term of the series. Then

$$u_n \overset{n\gg 1}{\simeq} \frac{1}{2\pi(e\Lambda)^4}\left(\frac{e^2\Lambda}{n}\right)^{n+2} n\left[(\ln\Lambda-\ln n)\sin(k)+O\left(\frac{k}{n}\right)\right]$$
$$=\frac{e}{2\pi(e\Lambda)^3}(v_n\ln\Lambda-w_n)\sin(k), \tag{16}$$

with $v_n=(e^2\Lambda/n)^{n+1}$ and $w_n=v_n\ln n$. According to D'Alembert's ratio test[12] for absolute convergence we have successively

$$\lim_{n\to\infty}\frac{v_{n+1}}{v_n}=e^2\Lambda\lim_{n\to\infty}\left(\frac{n}{n+1}\right)^n\frac{n}{(n+1)^2}=e\Lambda\lim_{n\to\infty}\frac{n}{(n+1)^2}\to 0,$$
$$\lim_{n\to\infty}\frac{w_{n+1}}{w_n}=e\Lambda\lim_{n\to\infty}\frac{n}{(n+1)^2}\frac{\ln(n+1)}{\ln n}$$
$$=e\Lambda\lim_{n\to\infty}\frac{n}{(n+1)^2}\left[1+\frac{1}{n\ln n}\left(1-\frac{1}{2n}+\frac{1}{3n^2}-\cdots\right)\right]$$
$$\to 0.$$

Therefore, $\sum_{n=2}^{\infty}u_n$ is absolutely convergent for any value of Λ. However, the series is *not* uniformly convergent.

Proof: Construct the partial sum

$$R_{N,p}(k,\Lambda)=u_{N+1}(k,\Lambda)+\cdots+u_{N+p}(k,\Lambda)$$
$$\simeq\frac{e}{2\pi(e\Lambda)^3}\sin k\sum_{n=N+1}^{N+p}(v_n\ln\Lambda-w_n), \quad N\gg 1. \tag{17}$$

In our case $v_{n+1}/v_n=e^2\Lambda(1+1/n)^{-n}n/(n+1)^2$ and, since $\exp[n/(1+n)]<(1+1/n)^n<e$ for any n, we have the inequality

$$e\Lambda\frac{n}{(n+1)^2}<\frac{v_{n+1}}{v_n}<e\Lambda\frac{n}{(n+1)^2}e^{1/n}.$$

Then

$$\frac{v_{n+1}}{v_n}\simeq\frac{e\Lambda}{n}, \quad \text{for } N+p\geq n\geq N+1\gg 2.$$

The first term in $R_{N,p}$ can thus be approximated by

$$\sum_{n=N+1}^{N+p}v_n\simeq v_n\frac{e\Lambda}{N}\sum_{r=0}^{p-1}\frac{(e\Lambda)^r N!}{(N+r)!}=(e\Lambda)^2\left(\frac{e}{N}\right)^{N+1}(N-1)!$$
$$\times\{e_{N+p-1}(e\Lambda)-e_{N-1}(e\Lambda)\},$$

where $e_n(x)=\sum_{l=0}^{n}x^l/l!$. In the same way, since

$$\frac{w_{n+1}}{w_n}=\frac{v_{n+1}}{v_n}\frac{\ln(n+1)}{\ln n}\simeq\frac{v_{n+1}}{v_n}$$

for sufficiently large n,

$$\sum_{n=N+1}^{N+p}w_n\simeq\sum_{n=N+1}^{N+p}v_n.$$

Therefore,

$$R_{N,p}(k,\Lambda)\simeq\frac{1}{(2\pi)^{1/2}}\sin(k)\ln\left(\frac{\Lambda}{e}\right)^{e/\Lambda}$$
$$\times N^{-3/2}[e_{N+p-1}(e\Lambda)-e_{N-1}(e\Lambda)]. \tag{18}$$

In order that $|R_{N,p}(k,\Lambda)|<\epsilon$, ϵ being an infinitesimally small positive number independent of N, we observe immediately that the smallest integer value N for which the condition is satisfied depends on Λ. The convergence is nonuniform. Q.E.D.

B. Limiting case $\alpha = 0$ of Eq. (10)

Since the ratio $b/a=k/\alpha$ is independent of Λ, two limiting cases have to be checked: $\alpha\to+0$ and $\alpha\to+\infty$.

The limiting case $\alpha=0$ has already been considered by Bowers and Salpeter.[13] In our notation this corresponds to the limit $\theta_n\to\pi/2$ and thus $\cos\theta_n\simeq\alpha n/k$.

Equation (10) then reads

$$G(k)=\frac{\pi}{4}+\sum_{n=1}^{\infty}\frac{(k\Lambda)^n}{\Gamma(n+1)\Gamma(n+3)}$$
$$\times\left[-[\psi(n+1)+\psi(n+3)-\ln(k\Lambda)]\right.$$
$$\left.\times\sin\left(\frac{n\pi}{2}\right)+\frac{\pi}{2}\cos\left(\frac{n\pi}{2}\right)\right]. \tag{19}$$

Now the first term in the sum can be expressed in terms of the Kelvin function

$$\mathrm{ker}_2x=\frac{1}{2}\left(\frac{2}{x}\right)^2\sum_{k=0}^{1}\sin\left(\frac{k\pi}{2}\right)\frac{(1-k)!}{k!}\left(\frac{x^2}{4}\right)^k - \ln\left(\frac{x}{2}\right)\mathrm{ber}_2x+\frac{\pi}{4}\mathrm{bei}_2x+\frac{1}{2}\left(\frac{x}{2}\right)^2 \times\sum_{k=0}^{\infty}\frac{(x^2/4)^k}{\Gamma(k+1)\Gamma(k+3)}\sin\left(\frac{k\pi}{2}\right)[\psi(k+1)+\psi(k+3)],$$

where ber_2x and bei_2x are also Kelvin's functions related to a Bessel function of a complex argument through the relation[14]

$$\mathrm{ber}_2(x)\pm i\,\mathrm{bei}_2(x)=J_2(x\exp(\pm i3\pi/4)).$$

Setting $x^2/4=k\Lambda$ and by the use of the power series expansion of ber_2x and bei_2x,

$$\sum_{l=0}^{\infty}\frac{(k\Lambda)^l}{l!\,\Gamma(l+3)}\begin{pmatrix}\sin\\ \cos\end{pmatrix}\frac{l\pi}{2}=\begin{pmatrix}\frac{1}{k\Lambda}\mathrm{ber}_2(2\sqrt{k\Lambda})\\ -\frac{1}{k\Lambda}\mathrm{bei}_2(2\sqrt{k\Lambda})\end{pmatrix},$$

we finally obtain

$$G(k)_{\alpha=0}=\frac{4\pi\Lambda}{k^2}[1-2\mathrm{ker}_2(2\sqrt{k\Lambda})]. \tag{20}$$

This expression is equivalent to that previously obtained by Bowers and Salpeter.[13]

C. Small k expansion, equivalent to large α limit

The small k expansion ($k^2\ll 1$) of $G(k)$, directly derivable from Eq. (15), is of particular interest for the asymptotic behavior[8] ($r\to\infty$) of a Meeron line, when we take into account the resummation to all orders of the most diverging graphs at $r=0$. In fact it allows us to perform order by order the nodal expansion of the potential of average force with finite graphs.[4,8] The required limit is reached upon setting $b\to 0$, $\cos(\theta_n)\to 0$, and therefore $\theta_n\simeq k/an$ in Eq. (10). To zeroth order in k^2, we recover Iwata's result.[7] This is easily seen when Eq. (10) is given in the form

$$\lim_{k\to 0}\frac{1}{4\pi}\left[G(k)-\frac{4\pi\Lambda^2}{k}\frac{1}{2!}\arctan\left(\frac{k}{2\alpha}\right)\right] =\Lambda^3\left(\sum_{n=3}^{\infty}\frac{(an)^{n-3}}{\Gamma(n+1)\Gamma(n-2)}[\ln(an)-\psi(n+1)-\psi(n-2)] +a\sum_{n=4}^{\infty}\frac{(an)^{n-4}}{\Gamma(n+1)\Gamma(n-3)}\right)=\Lambda^3\sum_{n=3}^{\infty}\frac{(an)^{n-3}}{\Gamma(n+1)\Gamma(n-2)} \times\left(\ln(an)+\frac{n-3}{n}-\psi(n+1)-\psi(n-2)\right),\quad a=\alpha\Lambda. \tag{21}$$

This is identical to Iwata's expression except the multiplicative factor Λ^3. Now, with only the zeroth- and the first-order terms with respect to k^2 in Eq. (15) retained, $G(k)$ assumes the form ($\alpha=1$)

$$G(k)=4\pi\Lambda^2[A(\Lambda)-B(\Lambda)k^2], \tag{22}$$

where

$$A(\Lambda)=\frac{1}{4}+\Lambda\sum_{n=3}^{\infty}\frac{(\Lambda n)^{n-3}}{\Gamma(n+1)\Gamma(n-2)} \times\left(\ln(\Lambda n)+\frac{n-3}{n}-\psi(n+1)-\psi(n-2)\right) \tag{23a}$$

and

$$B(\Lambda)=\frac{1}{6}\left[\frac{1}{8}-\frac{\Lambda}{54}+\frac{\Lambda^2}{96}+\Lambda^3\sum_{n=5}^{\infty}\frac{(\Lambda n)^{n-5}}{\Gamma(n+1)\Gamma(n-4)} \times\left(\ln(\Lambda n)+\frac{n-5}{n}-\psi(n+1)-\psi(n-4)\right)\right]. \tag{23b}$$

Two functions $A(\Lambda)$ and $B(\Lambda)$ are positive and decrease monotonically with Λ. The corresponding $G(r)$, Eq. (2), shows a Debye-like decrease at infinity through the relation

$$G(r)\overset{r\gg 1}{\simeq}\Lambda^2\alpha\exp(-\beta r)/r, \tag{24}$$

with $\alpha=A^2/B$ and $\beta=(A/B)^{1/2}$. Note that $A(\Lambda)$ and $B(\Lambda)$ are also defined as integrals[15] (see Sec. 2),

$$A(\Lambda)=\frac{1}{\Lambda^2}\int_0^{\infty}dr\,r^2[\exp(-\Lambda e^{-r}/r)-1+\Lambda e^{-r}/r] \tag{25}$$

and

$$B(\Lambda)=\frac{1}{3!\,\Lambda^2}\int_0^{\infty}dr\,r^4[\exp(-\Lambda e^{-r}/r)-1+\Lambda e^{-r}/r]. \tag{26}$$

Next, when we relax the condition $|k^2|\ll 1$, Eq. (22) certainly breaks down. In this case we have to include higher order terms in the k^2 expansion of $G(k)$. Slightly modifying Eq. (15), we write $G(k)$ in the following form:

$$G(k)=4\pi\Lambda^2\sum_{l=0}^{\infty}(-)^l\mathcal{A}_l(\Lambda)k^{2l}, \tag{27}$$

where

$$\mathcal{A}_l(\Lambda)=\frac{1}{(2l+1)!\,\Lambda^2}\int_0^{\infty}dr\,r^{2l+2}[\exp(-\Lambda e^{-\alpha r}/r)-1 +\Lambda e^{-\alpha r}/r] \tag{28}$$

$$=\frac{\Lambda^{2l+1}}{(2l+1)!}\left[\sum_{n=2}^{2l+2}\frac{(-)^n}{n!}\frac{\Gamma(2l+3-n)}{(\alpha\Lambda n)^{2l+3}} +\sum_{n=2l+3}^{\infty}\frac{(\alpha\Lambda n)^{n-2l-3}}{\Gamma(n+1)\Gamma(n-2l-2)}\left(\ln(\alpha\Lambda n) +\frac{n-2l-3}{n}-\psi(n+1)-\psi(n-2l-3)\right)\right]. \tag{29}$$

A large Λ limit of $\mathcal{A}_l(\Lambda)$ with l fixed is discussed in Appendix B.

Extensive numerical analysis of $\mathcal{A}_l(\Lambda)$ as a function of Λ has been carried out on the computer UNIVAC, first by summing up the infinite series (23a) and (23b) with a prescribed convergence criterion, and then by direct integration of Eqs. (25) and (26) by means of the Gauss quadratic method[14] (6 subintervals, 24 points in each). We have verified that, for $0.1\leq\Lambda\leq 10$, two methods give the results with absolute error smaller than 10^{-6}. For those values of Λ greater than 10, only the integration method is used, because of a slow convergence of the series. Numerical values of $\mathcal{A}_l(\Lambda)$ for $0\leq l\leq 7$ are given in Table I.

Now it is interesting to see to what extent the truncation of the series (27) may well represent the true $G(k)$. One way of testing a validity of the approximation, Eq. (22), is to evaluate poles of the integrand for the $w_2(r)$ integral

TABLE I. Coefficients $\mathcal{A}_l(\Lambda)$ in the k^2 expansion of $G(k)$.

Λ	$\mathcal{A}_0(\Lambda)$	$\mathcal{A}_1(\Lambda)$	$\mathcal{A}_2(\Lambda)$	$\mathcal{A}_3(\Lambda)$	$\mathcal{A}_4(\Lambda)$	$\mathcal{A}_5(\Lambda)$	$\mathcal{A}_6(\Lambda)$	$\mathcal{A}_7(\Lambda)$
0.5	1.6298(−1)	1.9551(−2)	3.0760(−3)	5.5536(−4)	1.0833(−4)	2.2181(−5)	4.6930(−6)	1.0155(−6)
1.0	1.3356(−1)	1.8570(−2)	3.0311(−3)	5.5277(−4)	1.0816(−4)	2.2169(−5)	4.6929(−6)	1.0168(−6)
1.5	1.1558(−1)	1.7761(−2)	2.9894(−3)	5.5026(−4)	1.0799(−4)	2.2156(−5)	4.6919(−6)	1.0168(−6)
2.0	1.0292(−1)	1.7070(−2)	2.9504(−3)	5.4782(−4)	1.0782(−4)	2.2144(−5)	4.6910(−6)	1.0169(−6)
3.0	8.5724(−2)	1.5929(−2)	2.8791(−3)	5.4314(−4)	1.0749(−4)	2.2119(−5)	4.6890(−6)	1.0167(−6)
4.0	7.4267(−2)	1.5007(−2)	2.8150(−3)	5.3869(−4)	1.0717(−4)	2.2094(−5)	4.6871(−6)	1.0166(−6)
5.0	6.5929(−2)	1.4236(−2)	2.7565(−3)	5.3444(−4)	1.0685(−4)	2.2070(−5)	4.6851(−6)	1.0164(−6)
7.0	5.4403(−2)	1.2998(−2)	2.6530(−3)	5.2646(−4)	1.0624(−4)	2.2022(−5)	4.6812(−6)	1.0161(−6)
10.0	4.3672(−2)	1.1619(−2)	2.5223(−3)	5.1557(−4)	1.0537(−4)	2.1952(−5)	4.6754(−6)	1.0156(−6)
15.0	3.3381(−2)	1.0021(−2)	2.3480(−3)	4.9962(−4)	1.0402(−4)	2.1839(−5)	4.6660(−6)	1.0148(−6)
20.0	2.7273(−2)	8.9034(−3)	2.2094(−3)	4.8573(−4)	1.0276(−4)	2.1732(−5)	4.6568(−6)	1.0140(−6)
25.0	2.3174(−2)	8.0612(−3)	2.0946(−3)	4.7338(−4)	1.0160(−4)	2.1628(−5)	4.6478(−6)	1.0132(−6)
30.0	2.0211(−2)	7.3963(−3)	1.9969(−3)	4.6226(−4)	1.0050(−4)	2.1528(−5)	4.6390(−6)	1.0125(−6)
40.0	1.6183(−2)	6.4003(−3)	1.8375(−3)	4.4283(−4)	9.8493(−5)	2.1338(−5)	4.6218(−6)	1.0109(−6)
50.0	1.3550(−2)	5.6795(−3)	1.7111(−3)	4.2623(−4)	9.6677(−5)	2.1160(−5)	4.6053(−6)	1.0094(−6)
70.0	1.0289(−2)	4.6885(−3)	1.5194(−3)	3.9889(−4)	9.3486(−5)	2.0831(−5)	4.5738(−6)	1.0065(−6)
100.0	7.6138(−3)	3.7706(−3)	1.3197(−3)	3.6735(−4)	8.9487(−5)	2.0393(−5)	4.5299(−6)	1.0023(−6)
150.0	5.3528(−3)	2.8933(−3)	1.1036(−3)	3.2922(−4)	8.4180(−5)	1.9767(−5)	4.4636(−6)	9.9578(−7)

$$w_2(r)=\frac{2\Lambda}{\pi r}\int_0^\infty dk\, k\sin(kr)\left(-\frac{1}{k^2}+\frac{[-1/k^2+H(k)]^2}{1+1/k^2-H(k)}\right),$$

$$H(k)=G(k)/4\pi\Lambda$$

which is just the first step of an iterative process that leads to the HNC equation. We note here that the evaluation of $w_2(r)$ by a contour integral, as was done previously by Del Rio and DeWitt,[3] is one place where an exact analytic form of $H(k)$ has a real utility, although the complicated Eqs. (10) or (15) are certainly unnecessary to obtain numerical results for $w_2(r)$.

When $G(k)$ is replaced by Eq. (22), the equation

$$1+1/k^2-H(k)=0 \tag{30}$$

yields two purely imaginary roots which coalesce at $k=k_c=1.807i$ with the corresponding critical Λ value $\Lambda_c=7.307$. This result $|k_c|>1$, which agrees apparently well with the recent Monte-Carlo data,[16] surely invalidates Eq. (22). In view of estimating how many terms we have to retain in power series (27), we first solve Eq. (30) by using the full expression (10) for $G(k)$. Two purely imaginary roots coalesce this time at $\Lambda_c=4.247$ for $k_c=1.498i$. This is in excellent agreement with Del Rio–DeWitt's result.[3] Next we approximate $H(k)$, successively, by truncating the infinite power series (27) at the order k^6, k^{10}, and k^{14}, i.e.,

$$H^{(\nu)}(k)=\Lambda\sum_{l=0}^{2(2\nu+1)}(-)^l\mathcal{A}_l k^{2l},\quad \nu=1,2,\text{ and }3.$$

Solutions to the equations $1+1/k^2-H^{(\nu)}(k)=0$ are plotted in Fig. 2. As far as the lower part of the $v-\Lambda$ curve ($v=-ik$) is concerned, $H^{(2)}(k)$ is apparently satisfactory for $\Lambda<\Lambda_c$. To evaluate a correct Λ_c value and complex roots for $\Lambda>\Lambda_c$, however, $H^{(3)}(k)$ is found to be a good approximation of practical use, in that curves of complex roots always stem from the coalescent point (Λ_c, v_c). With regards to the upper branches of the $v-\Lambda$ curve, we realize that the curve (a) obtained from the complete expression, Eq. (10), is the only correct results, i.e., an accurate computation of the upper roots requires all powers in k^2 of $H(k)$. Evaluation of two complex roots for $\Lambda>\Lambda_c$ with the use of the exact form, Eq. (10), and analysis of their behavior in the $\Lambda\to\infty$ limit[17] are still left open to further study. As an indication, those complex roots are also plotted on the same figure.

All these technicalities serve as the basis for the evaluation of the onset of short-range order through the appearance of oscillations of the pair correlation function

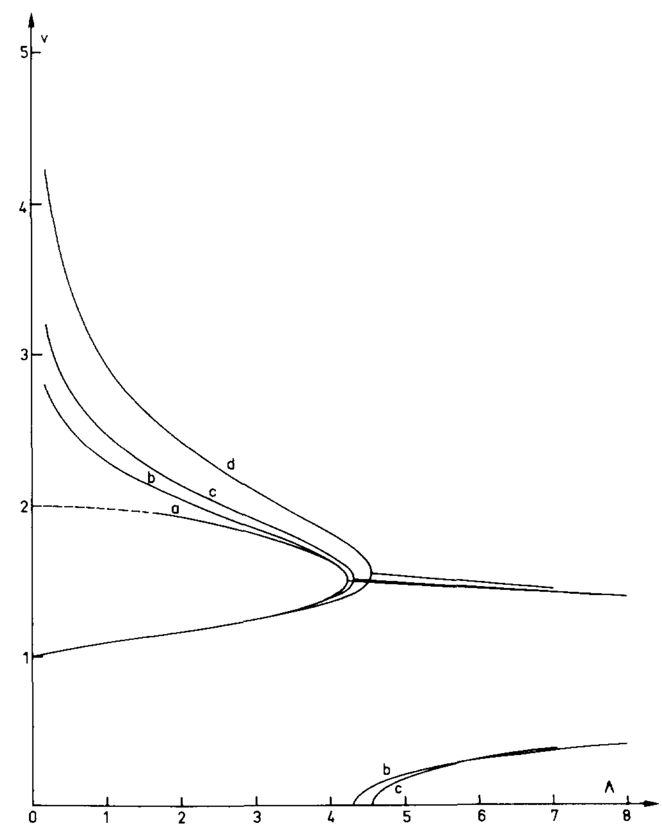

FIG. 2. Numerical solutions to the equation $1+1/k^2-H(k)=0$, (a) with the full $H(k)$, (b) with approximate $H^{(3)}(k)$, (c) with approximate $H^{(2)}(k)$, and (d) with approximate $H^{(1)}(k)$. For $\Lambda>\Lambda_c$, complex roots obtained with $H^{(2)}(k)$ and $H^{(3)}(k)$ are also shown.

TABLE II. Comparison of $G(r)$, Eq. (1), and its asymptotic form, Eq. (24).

	(a) $\Lambda=1.0$		(b) $\Lambda=5.0$		(c) $\Lambda=10.0$	
r	Eq. (1)	Eq. (24)	Eq. (1)	Eq. (24)	Eq. (1)	Eq. (24)
0.5	5.1035(−1)	5.0259(−1)	5.0676	5.2051	11.131	12.453
1.0	6.0080(−2)	6.5741(−2)	9.9831(−1)	8.8736(−1)	2.7040	2.3619
1.5	1.0535(−2)	1.1465(−2)	2.1909(−1)	2.0170(−1)	7.1346(−1)	5.9728(−1)
2.0	2.2387(−3)	2.2496(−3)	5.1292(−2)	5.1578(−2)	1.8498(−1)	1.6992(−1)
2.5	5.3318(−4)	4.7080(−4)	1.2768(−2)	1.4069(−2)	4.8458(−2)	5.1564(−2)
3.0	1.3695(−4)	1.0264(−4)	3.3494(−3)	3.9973(−3)	1.3040(−2)	1.6300(−2)
3.5	3.7113(−5)	2.3014(−5)	9.1725(−4)	1.1682(−3)	3.6172(−3)	5.2996(−3)
4.0	1.0467(−5)	5.2681(−6)	2.6009(−4)	3.4851(−4)	1.0325(−3)	1.7590(−3)
4.5	3.0446(−6)	1.2250(−6)	7.5866(−5)	1.0562(−4)	3.0222(−4)	5.9309(−4)
5.0	9.0759(−7)	2.8843(−7)	2.2649(−5)	3.2412(−5)	9.0393(−5)	2.0247(−4)
6.0			2.1319(−6)	3.1399(−6)	8.5219(−6)	2.4278(−5)
7.0			2.1208(−7)	3.1287(−7)	8.4813(−7)	2.9942(−6)
8.0						
9.0						
10.0						
	$\alpha=0.9606$, $\beta=2.6818$		$\alpha=7.6332$, $\beta=2.1520$		$\alpha=16.415$, $\beta=1.9387$	

$g_2(r)$ around unity, as discussed explicitly for the first time.[8] We must emphasize that the critical value $\Lambda_c = 4.225$, in accord with the previous Del Rio–DeWitt estimate,[3] does correspond only to the pole contributions The branch cut contributions from $2i$ to ∞i will be considered in a forthcoming work.

Finally, we check the validity of Eq. (24) by comparing it with Eq. (1). In contradiction to the asymptotic theory of the inverse Fourier transform, Eq. (24) is not a bad estimate, even for rather smaller values of r, namely, in the region $1<r<3$, whenever $\Lambda \lesssim 10$. (See Table II.)

D. Large k expansion

The simplest way of establishing a large k limit is to start not from Eq. (10) but directly from Eq. (2). We first rewrite this in the form

$$F(k)=\int_0^\infty dr\sin(kr)f(r),$$

with

$$f(r)=r\exp(-\Lambda e^{-r}/r)-r+\Lambda e^{-r}.$$

Obviously $f(r)$ verifies $f(0)=\Lambda$, $f(\infty)=0$, $f'(0) = -(1+\Lambda)$, and $f'(\infty)=-0$. Also $f(r)$ is continuously differentiable for any value of r. Therefore, by the second mean value theorem, we obtain

$$|F(k)|<f(+0)\left|\int_0^\alpha dr\sin(kr)\right|\leq 2\Lambda/k.$$

Hence, $F(k)$ and $G(k)$ consequently decrease to zero as $|k|\to\infty$. Now by successive integration by parts we get

$$F(k)=\frac{1}{k}\left(f(+0)-\frac{1}{k^2}f''(+0)+\frac{1}{k^4}f^{(4)}(+0)+\cdots+\frac{(-)^n}{k^{2n}}f^{(2n)}(+0)+R_n(k)\right),$$

where

$$R_n(k)=\frac{(-)^n}{k^{2n}}\int_0^\infty dr\cos(kr)f^{(2n+1)}(r).$$

Here, as will be shown below, we have set $f''(+\infty) = f^{(3)}(+\infty)=\cdots=0$. The remaining term $R_n(k)$ is bounded, since

$$\left|\int_0^\infty dr\cos(kr)f^{(2n+1)}(r)\right| \leq \int_0^\infty dr\left|f^{(2n+1)}(r)\right| = \left|f^{(2n)}(+0)\right|<\infty.$$

In view of these results we are led to evaluate $f^{(n)}(r)$ for $n\geq 2$. First we have

$$f^{(n)}(r)=[r\exp(-\Lambda e^{-r}/r)]^{(n)}-\delta_{n,1}+(-)^n\Lambda e^{-r},$$

where $\delta_{n,1}$ is the Kronecker delta. To evaluate $[r\exp(-\Lambda e^{-r}/r)]^{(n)}$, we recall the formula

$$(\varphi[\psi(x)])^{(n)}=\sum_{r=0}^{n}\frac{1}{r!}\varphi^{(r)}(\psi(x))\left(\sum_{s=0}^{r}\binom{r}{s}[-\psi(x)]^{r-s}[\psi(x)^s]^{(n)}\right).$$

Upon setting $\varphi=\exp[\psi(r)]$ and $\psi(r)=-\Lambda e^{-r}/r$, we obtain after elementary calculations the following expression:

$$f^{(n)}(r)=\sum_{m=0}^{n}\binom{n}{m}r^{(n-m)}[\exp(-\Lambda e^{-r}/r)]^{(m)} -\delta_{n,1}+(-)^n\Lambda e^{-r}, \tag{31}$$

where

$$\sum_{m=0}^{n}\binom{n}{m}r^{(n-m)}[\exp(-\Lambda e^{-r}/r)]^{(m)} =n!\exp(-\Lambda e^{-r}/r)\left[r\sum_{l=1}^{n}\frac{1}{l!}\left(\Lambda\frac{e^{-r}}{r}\right)^l\sum_{s=1}^{l}\binom{l}{s}(-)^s\frac{(-s)^n}{(s-1)!}\times\sum_{m=0}^{n}\frac{(s+m-1)!}{(n-m)!\,m!}\frac{1}{(sr)^m}+\sum_{l=1}^{n-1}\frac{1}{l!}\left(\Lambda\frac{e^{-r}}{r}\right)^l\sum_{s=1}^{l}\binom{l}{s}(-)^s\times\frac{(-s)^{n-1}}{(s-1)!}\sum_{m=0}^{n-1}\frac{(s+m-1)!}{(n-1-m)!\,m!}\frac{1}{(sr)^m}\right],\quad n\geq 2. \tag{32}$$

We thus observe that $f^{(n)}(+\infty)=0$ for $n\geq 2$ by virtue of the exponential factor e^{-r}/r. On the other hand, we obtain $f^{(n)}(+0)=(-)^n\Lambda$ because of $\exp(-\Lambda e^{-r}/r)$. It then follows that

$$f^{(2n)}(+0)=\Lambda. \tag{33}$$

Since $|R_n(k)| \leq \Lambda/|k|^{2n}$, $G(k\to\infty)$ is calculated to give

$$G(k)=\frac{4\pi}{k}F(k)=\frac{4\pi\Lambda}{k^2}\left(1-\frac{1}{k^2}+\frac{1}{k^4}-\cdots+(-)^n\frac{1}{k^{2n}}\right)$$
$$=\frac{4\pi\Lambda}{1+k^2}\left(1-\frac{(-)^{n+1}}{k^{2(n+1)}}\right). \quad (34)$$

Thus when the integration by parts procedure is repeated sufficiently many times ($n\to\infty$), one can let the remainder be as small as possible. We finally obtain the Debye limit

$$G(k)\overset{k\to\infty}{=}4\pi\Lambda/(1+k^2), \quad (35)$$

which is the obvious limit of Eq. (1) for small r.

Let us remark here that the large k limit is equivalent to the small α limit of $G(k)$ which has already been discussed in Sec. III B, since, in Eq. (10), k enters in the infinite sum through θ_n only. To prove this, we may use an asymptotic formula, due to Titchmarsh,[18] of the Fourier sine and cosine transform $F_s(k)$ and $F_c(k)$ which is valid under rather restricted conditions. Let $\chi(r) = r^{-\alpha}\psi(r)$ where $0<\alpha<1$ and $\psi(r)$ is of bounded variation in the interval $(0,\infty)$. Then

$$F_s(k)=\int_0^\infty dr\sin(kr)r^{-\alpha}\psi(r) \quad (36)$$
$$\sim\psi(+0)\Gamma(1-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)k^{\alpha-1}, \quad |k|\to\infty. \quad (37)$$

$F_c(k)$ is obtained with $\cos(\pi\alpha/2)$ replaced by $\sin(\pi\alpha/2)$. We apply this to Eq. (4) and examine the asymptotic limit of the integral $\int_0^\infty dr\, r^{s+1}\sin(kr)\exp(sr)$, with Re$s<-1$. Setting $\alpha=-(s+1)$ in Eq. (37), we find

$$\lim_{|k|\to\infty}\int_0^\infty dr\, r^{s+1}\sin(kr)\exp(sr)\simeq-\Gamma(s+2)k^{-(s+2)}\sin\left(\frac{\pi s}{2}\right) \quad (38)$$

with $-2<\text{Re}s<-1$, which just corresponds to the Mellin contour. The remaining procedure to obtain $G(k)$ for large k is similar to that given in Sec. 1. The result is

$$G(k)\simeq-\frac{4\pi}{k^3}\frac{1}{2\pi i}\int_M ds\,\Gamma(s)\Gamma(s+2)(k\Lambda)^{-s}\sin\left(\frac{\pi s}{2}\right)$$
$$=\frac{4\pi}{k^3}\sum_{l=0}^\infty\frac{(k\Lambda)^{l+2}}{\Gamma(l+1)\Gamma(l+3)}\left[-[\psi(l+1)+\psi(l+3)-\ln(k\Lambda)]\sin\left(\frac{l\pi}{2}\right)+\frac{\pi}{2}\cos\left(\frac{l\pi}{2}\right)\right],$$

which is Eq. (19). The large k limit of $G(k)$ is thus Bowers–Salpeter's expression, Eq. (20), which reads in the asymptotic limit

$$G(k)\sim\frac{4\pi\Lambda}{k^2}\left[1-\sqrt{\pi}\frac{\exp(-b^{1/4})}{b^{1/4}}\cos\left(2\sqrt{b}+\frac{9\pi}{8}\right)\right],$$
$$b=k\Lambda. \quad (39)$$

Although the correction term is quite different from that given by Eq. (34), the dominant term in two expressions is just the Fourier transform of the bare Coulomb potential.

IV. CONCLUDING REMARKS

With regards to the onset of shortrange order for $g_2(r)$, qualitative and physically meaningful results[3,8] have already been obtained by the method of residues which takes into account only those poles closest to the real axis of the complex k plane. Nonetheless, a more quantitative analysis, which also includes the contribution to $g_2(r)$ of all other poles located farther as well as the branch cut lying on the imaginary axis $2i<k<i\infty$, is still required in order to definitely settle this problem. This will be undertaken in a forthcoming work.

ACKNOWLEDGMENTS

We are gratefully indebted to Professor K. Chadan for many valuable discussions and to Dr. P. Midy for constant assistance in numerical analysis.

APPENDIX A: PROOF OF THE EQUIVALENCE OF EQ. (15) AND EQ. (10)

For the most general proof we start from the expression

$$(1+x^2)^{m/2}\sin(m\arctan x)=\sum_{l=0}^\infty\frac{(-)^l}{(2l+1)!}x^{2l+1}\frac{m!}{(m-2l-1)!},$$
$$|x|\leq 1. \quad (A1)$$

It is readily seen that, for a given m, the infinite sum over l is automatically trucated when $m\leq 2l$ by virtue of the relation

$$\lim_{z\to-n}\frac{1}{\Gamma(z)}=\frac{1}{\Gamma(-n)}=\frac{1}{(-n-1)!}=0, \quad n=0,1,2,\cdots.$$

First, setting $m=n-2$, $n\geq 3$, and $x=k\Lambda/an$ in Eq. (A1), we get

$$\sum_{l=0}^\infty\frac{(-)^l}{(2l+1)!}\left(\frac{k\Lambda}{an}\right)^{2l+1}\frac{1}{\Gamma(n-2l-2)}$$
$$=\frac{1}{\Gamma(n-1)}\frac{\sin[(n-2)\theta_n]}{\cos^{n-2}(\theta_n)}. \quad (A2)$$

Next, differentiating (A1) with respect to x and then setting $m=n-2$, $n\geq 3$, and $x=k\Lambda/an$ in the resulting expression, we obtain

$$\sum_{l=0}^\infty\frac{(-)^l}{(2l+1)!}\left(\frac{k\Lambda}{an}\right)^{2l+1}\frac{2l+1}{n}\frac{1}{\Gamma(n-2l-1)}$$
$$=\frac{n-2}{n}\frac{1}{\Gamma(n-1)}\frac{\sin(\theta_n)\cos((n-3)\theta_n)}{\cos^{n-2}(\theta_n)}.$$

Finally we shall prove the identity

$$\sum_{l=0}^\infty\frac{(-)^l}{(2l+1)!}\left(\frac{k\Lambda}{an}\right)^{2l+1}\frac{\psi(n-2l-2)}{\Gamma(n-2l-2)}$$
$$=\frac{1}{\Gamma(n-1)\cos^{n-2}(\theta_n)}\left[\left(\psi(n-1)+\ln[\cos(\theta_n)]+\frac{n-2}{n}\sin^2(\theta_n)\right)\sin((n-2)\theta_n)-\left(\theta_n-\frac{n-2}{n}\sin(\theta_n)\cos(\theta_n)\right)\times\cos((n-2)\theta_n)+\frac{n-2}{n}\sin(\theta_n)\cos((n-3)\theta_n)\right]. \quad (A3)$$

Proof: Differentiate Eq. (A2) with respect to n, by using the relation

$$\frac{d}{dn}\frac{1}{\Gamma(n-1)}\equiv\frac{d}{dz}\frac{1}{\Gamma(z-1)}\bigg|_{z=n}=-\frac{\psi(n-1)}{\Gamma(n-1)}.$$

We then obtain

$$-\sum_{l=0}^{\infty}\frac{(-)^l}{(2l+1)!}\left(\frac{k\Lambda}{an}\right)^{2l+1}\frac{\psi(n-2l-2)}{\Gamma(n-2l-2)}$$

$$=\frac{d}{dz}\left\{\frac{1}{\Gamma(z-1)}\frac{\sin((z-2)\theta_z)}{\cos^{z-2}(\theta_z)}\right\}\Big|_{z=n}$$

$$+\frac{k\Lambda}{an}\frac{n-2}{n}\frac{1}{\Gamma(n-1)}\frac{\cos(\theta_n)\cos((n-3)\theta_n)}{\cos^{n-2}(\theta_n)}.$$

Evaluation of the first term then yields Eq. (A3). Q.E.D.

Collecting the results obtained, we get

$$\sum_{l=0}^{\infty}\frac{(-)^l}{(2l+1)!}\left(\frac{k\Lambda}{an}\right)^{2l+1}\frac{1}{\Gamma(n-2l-2)}$$

$$\times\left(\ln(an)+\frac{n-2}{n}-\frac{2l+1}{n}-\psi(n+1)-\psi(n-2l-2)\right)$$

$$=\frac{1}{\Gamma(n-1)\cos^{n-2}(\theta_n)}\left\{\left[\ln\left(\frac{an}{\cos(\theta_n)}\right)-\psi(n+1)\right.\right.$$

$$\left.-\psi(n-1)\right]\sin((n-2)\theta_n)+\frac{n-2}{n}\cos^2(\theta_n)$$

$$\times\sin((n-2)\theta_n)+\theta_n\cos((n-2)\theta_n)$$

$$\left.-\frac{n-2}{n}\sin(\theta_n)\cos(\theta_n)\cos((n-2)\theta_n)\right\}.$$

Combination of the second and fourth terms by trigonometric algebra recovers Eq. (10). Q.E.D.

The above derivation does not apply for $n=2$. The $n=2$ term in Eq. (15) can be correctly evaluated when we recall the relation

$$\lim_{z\to -m}\frac{\psi(z)}{\Gamma(z)}=(-)^{m+1}m!.$$

Then the $n=2$ term reads

$$\frac{1}{\Gamma(3)}\sum_{l=0}^{\infty}(-)^l\frac{(k/2\alpha)^{2l+1}}{(2l+1)!}(-)^{2l+2}(2l)!$$

$$=\frac{1}{2!}\sum_{l=0}^{\infty}(-)^l\frac{(k/2\alpha)^{2l+1}}{2l+1}=\frac{1}{2!}\arctan\left(\frac{k}{2\alpha}\right).$$

Thus, when $\alpha=1$, we obtain

$$G(k)=\frac{4\pi\Lambda^2}{k}\left[\frac{1}{2!}\arctan\left(\frac{k}{2}\right)+\sum_{l=0}^{\infty}(-)^l\frac{(k\Lambda)^{2l+1}}{(2l+1)!}\right.$$

$$\times\sum_{n=3}^{\infty}\frac{(\Lambda n)^{n-2l-3}}{\Gamma(n+1)\Gamma(n-2l-3)}\left(\ln(\Lambda n)+\frac{n-2l-3}{n}\right.$$

$$\left.\left.-\psi(n+1)-\psi(n-2l-2)\right)\right]. \qquad \text{(A4)}$$

APPENDIX B: LARGE Λ LIMITS OF $\mathcal{A}_l(\Lambda)$

After integrating Eq. (28) by parts once, we can write $\mathcal{A}_l(\Lambda)$ as

$$\mathcal{A}_l(\Lambda)=\frac{1}{\Lambda}\left(1-\frac{1}{(2l+1)!(2l+3)}\right.$$

$$\left.\times\int_0^{\infty}dr\,r^{2l+1}(1+r)\exp(-r-\Lambda e^{-r}/r)\right). \qquad \text{(B1)}$$

Now, in order to evaluate large Λ limits of $\mathcal{A}_l(\Lambda)$, let us consider the integral

$$I_\alpha(\Lambda)=\int_0^{\infty}dr\,r^\alpha\exp(-r-\Lambda e^{-r}/r). \qquad \text{(B2)}$$

Introducing a new variable $r=\Lambda^{1/2}u$, $I_\alpha(\Lambda)$ becomes

$$I_\alpha(\Lambda)=\Lambda^{(\alpha+1)/2}\int_0^{\infty}du\exp[-\Lambda^{1/2}h(u)],$$

where $h(u)=u+(1/u)\exp(-\sqrt{\Lambda}\,u)-(\alpha/\sqrt{\Lambda})\ln u$. Denote by u_∞ a solution of $h'(u)=0$. Then, dividing the integration range into two subintervals $[0,u_\infty]$ and $[u_\infty,\infty]$, we can apply the Laplace approximation to each of the two integrals, since $\exp[-\sqrt{\Lambda}\,h(u)]$ has a sharp maximum at $u=u_\infty$. The result is

$$I_\alpha(\Lambda)\simeq\frac{\sqrt{\pi}}{2}\frac{x_\alpha^{\alpha+1}}{z_\alpha}\exp(\varphi_\alpha)[1+\mathrm{erf}(z_\alpha)], \qquad \text{(B3)}$$

where $x_\alpha=\sqrt{\Lambda}\,u_\infty$, $z_\alpha=[(1+\alpha)x_\alpha/(1+x_\alpha)+(x_\alpha-\alpha)\times(1+x_\alpha)]^{1/2}/\sqrt{2}$ and $\varphi_\alpha=[\alpha-x_\alpha(2+x_\alpha)]/(1+x_\alpha)$.

Then

$$\mathcal{A}_l(\Lambda)\overset{\Lambda\to\infty}{\simeq}\frac{1}{\Lambda}\left(1-\frac{\sqrt{\pi}}{2}\frac{1}{(2l+1)!(2l+3)}\right)[I_{2l+1}(\Lambda)+I_{2l+2}(\Lambda)]. \qquad \text{(B4)}$$

Accuracy of the above expression depends crucially on a value of x_α which we could obtain from the recurrence relation ($x_\alpha=\lim_{n\to\infty}x_n$)

$$x_n=\ln\left(\frac{\Lambda}{x_{n-1}}\frac{x_{n-1}+1}{x_{n-1}-\alpha}\right),$$

valid when $\sqrt{\Lambda}\gg\alpha$. In this case a starting test value $x_0=\ln(\Lambda/\ln\Lambda)$ ensures a rapid convergence of iteration.

Finally, in order to corroborate Eq. (B3), the use of the Laplace approximation must be justified. To this end, consider the integral $\int_0^\infty dr\,r^\alpha\exp(-r-\Lambda/r)$ which may be considered as a loose bound of $I_\alpha(\Lambda)$, since $\Lambda e^{-r}/r<\Lambda/r$. With the aid of the identity

$$\int_0^{\infty}dr\exp(-\beta r-\Lambda/r)=\frac{4\Lambda}{z}K_1(z),\quad \mathrm{Re}\beta>0,\quad \mathrm{Re}\Lambda\geq 0, \qquad \text{(B5)}$$

where $K_1(z)$ is the modified Bessel function of order 1 and $z=2(\beta\Lambda)^{1/2}$, we obtain

$$\int_0^{\infty}dr\,r^\alpha\exp(-r-\Lambda/r)=(-)^\alpha\,4\Lambda\left[\frac{\partial^\alpha}{\partial\beta^\alpha}\left(\frac{K_1(z)}{z}\right)\right]_{\beta=1}.$$

Using then the formula for the derivative

$$\left[\frac{\partial^\alpha}{\partial\beta^\alpha}\left(\frac{K_1(z)}{z}\right)\right]_{\beta=1}=\left(\frac{2\Lambda}{z}\frac{\partial}{\partial z}\right)^\alpha\left(\frac{K_1(z)}{z}\right)\Big|_{z=2\sqrt{\Lambda}}$$

$$=(-2\Lambda)^\alpha K_{\alpha+1}(2\sqrt{\Lambda})/(2\sqrt{\Lambda})^{\alpha+1},$$

we have

$$\int_0^{\infty}dr\,r^\alpha\exp(-r-\Lambda e^{-r}/r)\leq 2\Lambda^{(\alpha+1)/2}K_{\alpha+1}(2\sqrt{\Lambda}). \qquad \text{(B6)}$$

Then in the limit $\Lambda\to\infty$ with fixed α, the asymptotic expansion for a large argument of $K_{\alpha+1}(2\sqrt{\Lambda})$ is in order and yields

$$\int_0^{\infty}dr\,r^\alpha\exp(-r-\Lambda/r)\overset{\Lambda\to\infty}{\simeq}\sqrt{\pi}\Lambda^{(\alpha+1)/2}\exp(-2\sqrt{\Lambda}), \qquad \text{(B7)}$$

a result which we can recover by using the Laplace approximation.

To finish it seems interesting to give below some properties of $I_\alpha(\Lambda)$: (1) $I_\alpha(\Lambda)$ obeys the recurrence relation

$$\left(1+\frac{\lambda}{\alpha+1}\frac{\partial}{\partial\lambda}\right)I_\alpha(\lambda)=\frac{1}{\alpha+1}\left(1-\lambda\frac{\partial}{\partial\lambda}\right)I_{\alpha+1}(\lambda),$$

$$\lambda=1/\Lambda; \qquad \text{(B8)}$$

(2) When $I_0(\lambda)$ is known, $I_\alpha(\lambda)$ is a solution to the differential equation

$$\frac{1}{\alpha!}\left(1-\lambda\frac{\partial}{\partial\lambda}\right)^\alpha I_\alpha(\Lambda)$$
$$=\left(1+\frac{\lambda}{\alpha}\frac{\partial}{\partial\lambda}\right)\cdots\left(1+\frac{\lambda}{2}\frac{\partial}{\partial\lambda}\right)\left(1+\lambda\frac{\partial}{\partial\lambda}\right)I_0(\lambda), \qquad \text{(B9)}$$

which is not yet solved.

*Present address: Department of Electronics, Okayama University, Okayama 700, Japan.

[1]E. E. Salpeter, Ann. Phys. (N.Y.) **5**, 183 (1958).
[2]E. G. D. Cohen and T. J. Murphy, Phys. Fluids **12**, 1404 (1969).
[3]F. Del Rio and H. E. DeWitt, Phys. Fluids **12**, 791 (1969).
[4]C. Deutsch and M. Lavaud, Phys. Rev. A **9**, 2598 (1974).
[5]E. Meeron, in *Plasma Physics*, edited by J. E. Drummond (McGraw-Hill, New York, 1961); R. Abe, Prog. Theor. Phys. **22**, 213 (1959).
[6]The introduction of α in the Debye screening is just a scale change and affords no new result in comparison with Eq. (19) in the Del Rio—DeWitt's 1969 paper (Ref. 3). However, $G(k)$ as a function of α is useful in order to later discuss the limiting cases, i.e., $\alpha=0$ and $\alpha=\infty$.
[7]G. Iwata, Prog. Theor. Phys. **24**, 1118 (1960).
[8]C. Deutsch and Y. Furutani, J. Phys. A **8**, L83 (1975); also C. Deutsch, Y. Furutani, and M. M. Gombert, Phys. Rev. A **13**, 2244 (1976).
[9]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York and London, 1965).
[10]A. Erdelyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, p. 46.
[11]The Del Rio—DeWitt form, probably the simplest that this complicated expression for $G(k)$ can be given, is derivable directly from Eq. (4) by using $\sin(kr)=[\exp(ikr)-\exp(-ikr)]/2i$ and then working with $\exp[(\alpha s\pm ik)r]$.
[12]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (University Press, Cambridge, 1958).
[13]D. L. Bowers and E. E. Salpeter, Phys. Rev. **119**, 1180 (1960).
[14]A. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1965), Appl. Ser. 55.
[15]Numerical values of a function related to $A(\Lambda)$, Eq. (25), have been published previously. See Appendix C of F. J. Rogers and H. E. DeWitt, Phys. Rev. A **8**, 1061 (1973).
[16]J. P. Hansen, Phys. Rev. A **8**, 3096 (1973); see also S. G. Brush, H. L. Sahlin, and E. Teller, J. Chem. Phys. **45**, 2102 (1966).
[17]In the $\Lambda\to\infty$ limit, a partial result has already been reported by M. S. Cooper and H. E. DeWitt, Phys. Lett. A **40**, 391 (1972).
[18]E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon Press, Oxford, 1948), 2nd ed.

A classical Markov process in nonequilibrium quantum statistical mechanics

P. L. Torres*

Department of Physics, Queen Mary College, London, El 4NS, England
(Received 19 January 1976)

We obtain, from the exact microscopic, reversible, dynamics of an oscillator coupled to a heat bath, a classical, Markov stochastic process, for the energy of the oscillator. Also, the corresponding macroscopic phenomenological equation, describing an irreversible approach to equilibrium, is obtained.

1. INTRODUCTION

A central problem in nonequilibrium statistical mechanics is to explain how a large assembly of particles, whose microscopic laws of motion are reversible, can generally admit a contracted description in which a relatively small number of macroscopic variables evolves according to some self-contained, irreversible phenomenological laws (e.g., hydrodynamics, heat conduction).[1,2]

In Refs. 1 and 2, it was proposed that such laws may be extracted from the microscopic equations of motion by means of a two-stage program. The first stage corresponded to the derivation of a classical, Pauli, master equation for the macroscopic variables; and the second stage, to the derivation, from the master equation, of phenomenological equations of motion.

The object of the present paper is to carry out a similar program in a rigorous way for a particular solvable model. This is the model of Ref. 3, consisting of an oscillator coupled to an infinite heat bath: Models of this type have been treated elsewhere with different purposes (see, e.g., Refs. 4–7). For the macroscopic variable, we consider the energy of the oscillator. Our main result is that this variable evolves according to a law which, in a certain well-defined limit, corresponds to a classical Markov process.[8,9]

In Sec. 2, we describe the model and certain of its properties, then we consider the existence of the Van Hove limit for the time correlation functions of the energy of the oscillator corresponding to a KMS state[10,11] of the composite system (Sec. 3). Through these time correlation functions, we obtain the classical Markov process, with specified diffusion equation, obeyed by the energy variable of the oscillator (Sec. 4). Finally, (Sec. 5), the corresponding phenomenological equation is obtained from the diffusion equation.

Throughout this article, we shall use the following notations. The standard symbols Z_+, R, R_+, and C will be used to denote the positive integers, the real line, the positive reals, and the complex numbers respectively. $\bar{z}$ will denote the complex conjugate of $z \in C$. If H is a Hilbert space, we shall denote by $(\cdot,\cdot)_H$, the corresponding inner product (linear in the second variable). The Fourier transform of a function f on R will be denoted by $\hat{f}$, with the convention that (cf. Ref. 3)

$$\hat{f}(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} dy\, f(y) \exp(-ixy) \qquad \forall\, x \in R.$$

Finally, we shall use $\hbar = 1$.

2. THE MODEL

In this section we set up the model, which is a slightly different (less general) version of the one considered in Ref. 3, together with some of its consequences, most of them worked out in Ref. 3.

Let H be a complex Hilbert space, and T a strongly-continuous unitary representation of R in H, whose infinitesimal generator is ih.

We render $\mathcal{A}(H)$, the set of all complex-valued functions on H with finite point support into a $*$ algebra by equipping it with an involution $A \to A^*$ and a binary multiplication $(A, B) \to AB$, according to the formulas

$$A^*(f) = \overline{A(-f)} \qquad \forall\, f \in H \tag{2.1}$$

and

$$(AB)(f) = \sum_{g \in H} A(g) B(f-g) \exp(i \operatorname{Im}(f,g)_H) \quad \forall\, f \in H. \tag{2.2}$$

The set $\mathcal{J}$ of all faithful Hilbert space representations of $\mathcal{A}(H)$ is nonvoid, and the map $A \to \|A\| = \sup_{\pi \in \mathcal{J}} \|\pi(A)\|$ of $\mathcal{A}(H)$ into R, is a C^*-norm on $\mathcal{A}(H)$.[12] The C^*-algebra of the CCR over H is the completion $\bar{\mathcal{A}}(H)$ of $\mathcal{A}(H)$ with respect to this norm. If, for each $f \in H$, we define $\delta_f \in \mathcal{A}(H)$ by the formula

$$\delta_f(g) = \begin{cases} 1 & \text{for } f = g, \\ 0 & \text{for } f \neq g, \end{cases} \tag{2.3}$$

then $\bar{\mathcal{A}}(H)$ is the norm-closed linear span of $\{\delta_f \mid f \in H\}$ and δ_0 is the unit element of $\bar{\mathcal{A}}(H)$.

Let $S(H)$ be the set of all states on $\bar{\mathcal{A}}(H)$. The characteristic functional $\Gamma : H \to C$, corresponding to the state w is defined by

$$\Gamma(f) = w(\delta_f) \quad \forall\, f \in H. \tag{2.4}$$

Let $\tau_T : R \to \operatorname{Aut}\bar{\mathcal{A}}(H)$ be defined by the formula

$$(\tau_T(t)A)(f) = A(T(-t)f), \quad \forall\, A \in \bar{\mathcal{A}}(H),\ f \in H,\ t \in R. \tag{2.5}$$

The automorphisms $\tau_T(R)$, induced by the transformation $T(R)$ of H, represent a *quasifree evolution* of $\bar{\mathcal{A}}(H)$, (cf. Ref. 13).

Let $\sum(H, T)$ be the triple $(\bar{\mathcal{A}}(H), S(H), \tau_T)$. We consider this triple to represent a physical system, whose dynamical variables, states, and dynamics correspond to the self-adjoint elements of $\bar{\mathcal{A}}(H)$, $S(H)$, and $\tau_T(R)$, respectively. We call such a system a *quasifree system*.

Definition 2.1: We define I to be the class of all quasifree systems $\sum(H^1, T^1)$ for which:

(i) The one-particle Hamiltonian h is a lower semibounded operator in H^1, $h \geq mI_{H^1}$, where $m \in R_+$, and $h^{1/2}$ is invertible. Further, H^1 lies in the domain of $h^{-1/2}$.

(ii) If we define $F: H^1 \times R \to C$ by the formula

$$F(f^1, t) = (f^1, T^1(t)f^1)_{H^1} \quad \forall f^1 \in H^1, \quad t \in R, \tag{2.6}$$

then for each $f^1 \in H^1$, $F(f^1, (\cdot))$ is the inverse Fourier transform of a bounded, continuous, L_1-class function $\hat{F}(f^1, (\cdot))$ on R, which satisfies a Lipschitz condition

$$|\hat{F}(f^1, x+y) - \hat{F}(f^1, y)| \leq A|x|^\delta$$

$$\forall x, y \in R, \text{ with } 0 < \delta < 1, \tag{2.7}$$

where A, δ are constants.

Definition 2.2: We define II to be the class of quasi-free systems $\sum(H^2, T^2)$ such that $H^2 = C$, T^2 is given by the formula

$$T^2(t)z = z\exp(ix_0 t) \quad \forall t \in R, \quad z \in C, \tag{2.8}$$

where x_0 is a positive constant, and where

$$(z, z')_C = \bar{z}z' \quad \forall z, z' \in C. \tag{2.9}$$

Let $\sum^1 \equiv \sum(H^1, T^1)$ and $\sum^2 \equiv \sum(H^2, T^2)$ be quasifree systems of class I and II, respectively.

Definition 2.3: We define $\sum = \sum(H, T_\lambda)$ to be the quasi-free system obtained by compounding $\sum^1$ with $\sum^2$ as follows:

(i) The Hilbert space H is the one obtained through the direct sum $H^1 \oplus C$, with inner product

$$(f_1^1 \oplus z_1, f_2^1 \oplus z_2)_H = (f_1^1, f_2^1)_{H^1} + \bar{z}_1 z_2$$

$$\forall f_1^1, f_2^1 \in H^1, \quad z_1, z_2 \in C. \tag{2.10}$$

(ii) T_λ is defined as the strongly-continuous unitary representation of R in H, whose infinitesimal generator, ih_λ, is given by the formula

$$h_\lambda = h_0 + \lambda v, \quad h_0 = h \oplus x_0, \tag{2.11}$$

where

$$v(f^1 \oplus z) = zg^1 \oplus (g^1, f^1)_{H^1} \quad \forall f^1 \in H^1, \quad z \in C, \tag{2.12}$$

g^1 is a fixed element of H^1 and λ a real positive constant.

(iii) Since we know [3] that for each $f^1 \in H^1$, $\hat{F}(f^1, (\cdot))$ is a real valued, nonnegative function on R, we will assume that

$$\hat{F}(g^1, x) > 0 \qquad \forall x \in [0, x_0]. \tag{2.13}$$

(iv) We assume that $\lambda \in (0, \lambda_0)$, where λ_0 is the positive real number referred to in Lemma 6.1, (ii) of Ref. 3.

(v) If we call m_v the lower bound of the operator v, then we assume that

$$\inf\{m, x_0\} > |\inf\{0, \lambda_0 m_v\}|, \tag{2.14}$$

where m is defined in (i) of Definition 2.1.

This model is solvable, and if we define

$$f(t) = T_\lambda(t)(0 \oplus 1) \qquad t \in R, \tag{2.15}$$

we can define u_λ and k_λ to be the maps of $R_+ \cup \{0\}$ into C and H^1 respectively, given by the formula

$$f(t) = k_\lambda(t) \oplus u_\lambda(t) \qquad t \in R_+ \cup \{0\}. \tag{2.16}$$

Then,[3]

$$u_\lambda(t) = v_\lambda(t) + w_\lambda(t) \qquad \forall t \in R_+, \tag{2.17}$$

where

$$v_\lambda(t) = \exp i[x_0 + i\lambda^2(\pi/2)^{1/2} J(g^1, x_0)]t \quad \forall t \in R_+, \tag{2.18}$$

and

$$w_\lambda(t) = \frac{-\lambda^2}{(8\pi)^{1/2}} \int_{-\infty}^{\infty} dx \exp(-ixt)[J(g^1, x_0) - J(g^1, x)]$$
$$\times \{[x - x_0 - i\lambda^2(\pi/2)^{1/2} J(g^1, x_0)]$$
$$\times [x - x_0 - i\lambda^2(\pi/2)^{1/2} J(g^1, x)]\}^{-1} \quad \forall t \in R_+, \tag{2.19}$$

where

$$J(g^1, x) = \hat{F}(g^1, x) + iG(g^1, x) \qquad \forall x \in R, \tag{2.20}$$

$G(g^1, (\cdot))$ being a real valued, bounded, continuous function on R. Also

$$|J(g^1, x+y) - J(g^1, y)| \leq B|x|^\delta \quad \forall x, y \in R, \tag{2.21}$$

where B is a constant, and δ is the same as the one in Eq. (2.7).

Now, the characteristic functional Γ corresponding to a primary KMS state w (w is unique except for a gauge transformation of the second kind[13]), which characterizes the equilibrium state of Σ at the inverse temperature β, can be taken as

$$\Gamma(f) = \exp[-\tfrac{1}{2}(\coth(\tfrac{1}{2}\beta h_\lambda)f, f)_H] \quad \forall f \in H. \tag{2.22}$$

Let $(\mathcal{H}, \pi, \Omega)$ be the GNS triple[14,15] induced by this initial state w of Σ, and $R(f)$ be the self-adjoint field operator corresponding to the Weyl operator $\pi(\delta_f) = \exp(iR(f))$. Then[12,13]

$$\Gamma(f) = (\Omega, \exp(iR(f))\Omega)_{\mathcal{H}} \equiv \langle \exp(iR(f)) \rangle \quad \forall f \in H. \tag{2.23}$$

We can define $\forall f \in H$, the annihilation and creation operators[13] by

$$a(f) = \tfrac{1}{2}[R(f) + iR(if)], \quad a(f)^* = \tfrac{1}{2}[R(f) - iR(if)] \tag{2.24}$$

which have the following commutation relations:

$$[a(f_1), a(f_2)] = 0$$

and

$$[a(f_1), a(f_2)^*] \subseteq (f_1, f_2)_H I_{\mathcal{H}} \quad \forall f_1, f_2 \in H. \tag{2.25}$$

Using Eqs. (2.22)–(2.25) and the invariance under gauge transformations of the first kind,[13] we can easily deduce that $\forall f_1, \ldots, f_n \in H$, and $\forall n \in Z_+$

$$2\langle a(f_2)^* a(f_1) \rangle = (\coth(\tfrac{1}{2}\beta h_\lambda)f_1, f_2)_H - (f_1, f_2)_H, \tag{2.26}$$

and

$$\langle a(f_1)^* \cdots a(f_n)^* a(f_1) \cdots a(f_n) \rangle = \overset{+}{|} \langle a(f_i)^* a(f_j) \rangle \overset{+}{|}, \tag{2.27}$$

where the rhs stands for the permanent[16] of the $n \times n$ matrix $[\langle a(f_i)^* a(f_j) \rangle]$.

Since,[3] $\forall f = f^1 \oplus z \in H$, $\pi(\delta_f) = \pi^1(\delta_{f^1}) \otimes \pi^2(\delta_z)$, it follows that

$$\begin{aligned} a(f) &= a^1(f^1) \otimes I^2 + I^1 \otimes b(z), \\ a(f)^* &= a^1(f^1)^* \otimes I^2 + I^1 \otimes b(z)^*, \end{aligned} \tag{2.28}$$

where $a^1(f^1)$, $a^1(f^1)^*$ and $b(z) \equiv b\bar{z}$, $b(z)^* \equiv b^* z$ are the annihilation and creation operators of Σ^1 and Σ^2, respectively.

If we take, at time $t=0$, $f(0)\equiv 0\oplus 1$ ($\in H$), then

$$x_0 a(f(0))^* a(f(0)) = I^1 \otimes x_0 b^* b \equiv I^1 \otimes E(0), \qquad (2.29)$$

where $E(0)$ is the energy operator of Σ^2 at $t=0$, whose time evolution, $\forall t\in R$, $E(t)$, is given because of the quasifree evolution, by $f(t)$.

3. THE TIME CORRELATION FUNCTIONS

Definition 3.1: We call, $\forall t_1,\ldots,t_n \in R$, and $\forall n\in Z_+$, $\langle E(t_1)\cdots E(t_n)\rangle$ the time correlation functions for the energy of Σ^2, for the equilibrium state given by Eq. (2.22).

We are now interested in the existence of the Van Hove limit (weak coupling limit), i.e., $\lambda^2 t=\tau$, $\lambda\to 0_+$, $t\to\infty$, τ, fixed, for these time correlation functions. For that, we consider first the Van Hove limit of Eq. (2.27), for $f_k = f(t_k)$, $t_k\in R$, $k=1,\ldots,n$.

Because of Eqs. (2.11), (2.14), and the Rellich theorem (see Ref. 17, p. 1263), we obtain that

$$s\text{-}\lim_{\lambda\to 0_+} \coth(\tfrac{1}{2}\beta h_\lambda) f(0) = \coth(\tfrac{1}{2}\beta h_0) f(0) = 0\oplus \coth(\tfrac{1}{2}\beta x_0). \qquad (3.1)$$

Using Eqs. (2.10), (2.16), and (3.1), it follows that $\forall \tau\in R_+$,

$$\lim_{\lambda\to 0_+} [(\coth(\tfrac{1}{2}\beta h_\lambda) f(0), f(\tau/\lambda^2))_H - \coth(\tfrac{1}{2}\beta x_0) u_\lambda(\tau/\lambda^2)] = 0, \qquad (3.2)$$

and

$$\lim_{\lambda\to 0_+} [(f(0), f(\tau/\lambda^2))_H - u_\lambda(\tau/\lambda^2)] = 0. \qquad (3.3)$$

Because of Eqs. (2.13), (2.20), and (2.21), we obtain through Eq. (2.19) that $\forall\tau\in R_+$,

$$\lim_{\lambda\to 0_+} w_\lambda(\tau/\lambda^2) = 0, \qquad (3.4)$$

uniformly w.r.t. τ.

Using Eqs. (2.17), (2.18), (2.20), (2.26), (2.27), (3.2)–(3.4), and the unitarity of T_λ, it follows that

$$\lim_{\lambda\to 0_+} \left\langle a\left(f\left(\frac{\tau_1}{\lambda^2}\right)\right)^* \cdots a\left(f\left(\frac{\tau_n}{\lambda^2}\right)\right)^* a\left(f\left(\frac{\tau_1}{\lambda^2}\right)\right)\cdots a\left(f\left(\frac{\tau_n}{\lambda^2}\right)\right)\right\rangle = \left(\frac{1}{[\exp(\beta x_0)-1]}\right)^n \overset{+}{|}\exp(-\alpha(x_0)|\tau_i-\tau_j|)\overset{+}{|},$$
$$\forall \tau_1,\ldots,\tau_n\in R, \quad \forall n\in Z_+, \qquad (3.5)$$

where

$$\alpha(x_0) = (\pi/2)^{1/2}\hat{F}(g^1, x_0), \quad (>0). \qquad (3.6)$$

Now, with Eqs. (2.17), (2.18), (2.20), (2.25), (2.29), and (3.3)–(3.5), we can obtain the time correlation functions for the energy in the Van Hove limit, which we will denote by $\langle E_{\tau_1}\cdots E_{\tau_n}\rangle$. However, such expressions are not invariant under the permutations of the indices of the τ's. In order to achieve such invariance, we can proceed in two different ways.

(i) Taking the Wick normal products for the energies (cf. Ref. 6), in which case the mean value of the energy of Σ^2 in the equilibrium state is given by

$$\langle E\rangle \equiv \langle E_\tau\rangle = x_0/[\exp(\beta x_0)-1] \quad \forall\tau\in R. \qquad (3.7)$$

We notice that this is a rather formal procedure.

(ii) Taking the classical limit, $x_0\beta\to 0_+$, in which case $\langle E\rangle = \beta^{-1}$.

With both procedures, we obtain that

$$\langle E_{\tau_1}\cdots E_{\tau_n}\rangle = \overset{+}{|}V\overset{+}{|} \quad \forall\tau_1,\ldots,\tau_n\in R, \quad \forall n\in Z_+, \qquad (3.8)$$

where

$$V = [\langle E\rangle \exp(-\alpha|\tau_i-\tau_j|)], \qquad (\alpha>0), \qquad (3.9)$$

and where α and $\langle E\rangle$ are given by $\alpha(x_0)$ and Eq. (3.7) or by $\alpha(0)$ and β^{-1}, depending on whether we choose the point of view (i), or (ii), respectively [also if procedure (i) is the one chosen, $\langle E_{\tau_1}\cdots E_{\tau_n}\rangle$ really stands for $\langle :E_{\tau_1}\cdots E_{\tau_n}:\rangle$].

The Van Hove limit, as well as the classical one, $x_0\beta\to 0_+$ exists *uniformly* w.r.t. the τ's.

4. THE STOCHASTIC PROCESS

For arbitrary, but fixed $\tau_1,\ldots,\tau_n\in R$, $n\in Z_+$, we are faced through Eq. (3.8) with a moment-problem.[18] That is, we want to know if there is a unique characteristic function $\varphi_{\tau_1,\ldots,\tau_n}(x_1,\ldots,x_n)$, $x_1,\ldots,x_n\in R$, which reproduces these time correlation functions.

Now, we consider

$$\varphi_{\tau_1,\ldots,\tau_n}(x_1,\ldots,x_n) = 1/|I-iVD|, \quad D=[\delta_{ij}x_i], \qquad (4.1)$$

which is the characteristic function corresponding to an n-dimensional gamma-type distribution.[19,20] This distribution can be obtained as the one corresponding to one-half the sum of squares of two independent n-dimensional Gaussian vectors $(X_{\tau_1};\ldots;X_{\tau_n})$, $(Y_{\tau_1};\ldots;Y_{\tau_n})$ with zero mean values and the same covariance matrix V.[19,20]

That this distribution reproduces the time correlation functions of Eq. (3.8), can be seen in the usual way through Eq. (4.1) or by considering the moments of the $2n$-dimensional Gaussian vector $(1/\sqrt{2})(X_{\tau_1}+\beta_1 Y_{\tau_1};\ldots;X_{\tau_{2n}}+\beta_{2n}Y_{\tau_{2n}})$, where $\beta_1,\ldots,\beta_{2n}\in C$, and then taking $\beta_l=i$ for $l=1,\ldots,n$ and $\beta_j=-i$ for $j=n+1,\ldots,2n$.

If $\Gamma_{j_1,\ldots,j_n} = \langle E_{\tau_1}^{j_1}\cdots E_{\tau_n}^{j_n}\rangle$ $\forall j_1,\ldots,j_n\in Z_+\cup\{0\}$, $\tau_1,\ldots,\tau_n\in R$, $n\in Z_+$, and

$$\lambda_{2n} = \Gamma_{2n,0,0,\ldots,0} + \Gamma_{0,2n,0,\ldots,0} + \cdots + \Gamma_{0,0,0,\ldots,2n}, \qquad (4.2)$$

it follows that the series

$$\sum_{n=1}^{\infty} \lambda_{2n}^{-1/2n} \qquad (4.3)$$

diverges (comparing it with the harmonic series). Then,[18] the uniqueness of the characteristic function is asserted.

Then, because of Kolmogorov's fundamental theorem, the family of time correlation functions [Eq. (3.8)], gives us a stochastic process $\{E_\tau;\ \tau\in R\}$ with state space R_+, and with

$$E_\tau = \tfrac{1}{2}(X_\tau^2 + Y_\tau^2), \qquad (4.4)$$

where $\{X_\tau;\ \tau \in R\}$ and $\{Y_\tau;\ \tau \in R\}$ are two independent stationary Ornstein—Uhlembeck velocity processes (colored noises) with zero mean value and the same covariance matrix V.

This energy process which, unlike a Gaussian one, is not specified by its mean value $\langle E\rangle$ and covariance matrix $[\langle E\rangle^2 \exp(-2\alpha|\tau_i - \tau_j|)]$,[20] is continuous, stationary both in the strict and in the wide senses.[21] It is *Markovian in the strict sense*,[22] and also in the wide sense. Moreover,[23] It is a homogeneous diffusion process whose Fokker—Plank equation[26] for the transition probability density, $p = p(\tau, E_0, E)$, is given by

$$\frac{\partial p}{\partial \tau} = 2\alpha\langle E\rangle \frac{\partial^2 Ep}{\partial E^2} - 2\alpha \frac{\partial}{\partial E}[(\langle E\rangle - E)p], \quad E, \tau \in R_+. \quad (4.5)$$

This singular, parabolic, differential equation has a unique fundamental solution,[25] normalized $\forall\, E_0,\ \tau \in R_+$, which is given by[4,23,26]

$$p(\tau, E_0, E) = \frac{1}{\langle E\rangle(1-\exp(-2\alpha\tau))} I_0\left[\frac{2\sqrt{E_0 E}\exp(-\alpha\tau)}{\langle E\rangle(1-\exp(-2\alpha\tau))}\right] \times \exp\left[-\frac{E + E_0\exp(-2\alpha\tau)}{\langle E\rangle(1-\exp(-2\alpha\tau))}\right] \quad \forall\, E, E_0, \tau \in R_+, \quad (4.6)$$

where I_0 is the modified Bessel function of zero order.

We have that[23]

$$p_{eq} \equiv \lim_{\tau\to\infty} p(\tau, E_0, E) = \frac{\exp(-E/\langle E\rangle)}{\langle E\rangle} \quad \forall E, E_0 \in R_+, \quad (4.7)$$

as one might anticipate.

The irreversibility of the process can be characterized through the monotonic increasing entropy,[1] given by

$$S(\tau, E_0) = -\int_0^\infty dE\, p(\tau, E_0, E) \log \frac{p(\tau, E_0, E)}{p_{eq}} \quad \forall\, E_0, \tau \in R_+. \quad (4.8)$$

The contracted nature of the description[1,2] of the stochastic process for the energy variable that we have just formulated becomes particularly manifest when one notes that the Van Hove limit for the time correlation functions for the position and momentum operators of Σ^2 do *not* exist.

5. THE PHENOMENOLOGICAL EQUATION

Using Eq. (4.6) and Eqs. (9.6.3), (11.4.28), (13.4.1), and (13.6.12) of Ref. 27, we obtain, for the conditional mean value $\mathcal{E}(E_\tau | E_0)$, for the energy of Σ^2 at time τ ($\tau \in R_+$), that [cf. Ref. 4]

$$\mathcal{E}(E_\tau | E_0) = (E_0 - \langle E\rangle)\exp(-2\alpha\tau) + \langle E\rangle \quad \forall\, E_0, \tau \in R_+. \quad (5.1)$$

This conditional mean value is the unique continuous solution of the differential equation

$$\frac{d\mathcal{E}(E_\tau | E_0)}{d\tau} + 2\alpha\mathcal{E}(E_\tau | E_0) = 2\alpha\langle E\rangle, \quad \text{with } \mathcal{E}(E_0 | E_0) = E_0 \in R_+. \quad (5.2)$$

With the same set of equations as above, we find that the relative width $R(\tau, E_0)$, is given by

$$R(\tau, E_0) = \left[1 - \left(\frac{E_0}{E_0 - \langle E\rangle}\right)^2 \left(\frac{\mathcal{E}(E_\tau | E_0) - \langle E\rangle}{\mathcal{E}(E_\tau | E_0)}\right)^2\right]^{1/2} \quad \forall\, \tau, E_0 \in R_+. \quad (5.3)$$

From Eq. (5.3), it follows that if $\mathcal{E}(E_\tau | E_0) > \langle E\rangle$, then

$$R(\tau, E_0) < 0\left[\left(\frac{\langle E\rangle}{\mathcal{E}(E_\tau | E_0)}\right)^{1/2}\right],$$

which tells us that the fluctuations of the energy variable remain sufficiently small so that we may consider Eq. (5.2) as a phenomenological one, describing an irreversible approach to equilibrium.[1,2]

ACKNOWLEDGMENTS

The author is very much indebted to Dr. G. L. Sewell who suggested this problem, for many illuminating discussions and comments, and for a critical reading of the manuscript; and to Miss M. Tropper for comments on the presentation. Also, the author would like to thank the Consejo de Desarrollo Cientifico y Humanistico of the Universidad Central de Venezuela, for a grant.

*On leave of absence from [and now permanent address]: Departamento de Fisica, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela.

[1]N. G. van Kampen, *Fundamental Problems in Statistical Mechanics*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1962), pp. 173—202.

[2]G. L. Sewell, *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. X-A, pp. 289—327.

[3]G. L. Sewell, Ann. Phys. (N.Y.) 85, 336 (1974).

[4]Robert J. Rubin and Kurt E. Shuler, J. Chem. Phys. 25, 59 (1956).

[5]P. Ullersma, Physica 32, 27, 56, 74, 90 (1966).

[6]G. W. Ford, M. Kac, and P. Mazur, J. Math. Phys. 6, 504 (1965).

[7]Christian Favre and Philippe A. Martin, Helv. Phys. Acta 41, 333 (1968); P. Mazur and E. W. Montroll, J. Math. Phys. 1, 70 (1960); R. J. Rubin, J. Math. Phys. 1 309 (1960); 2, 373 (1961); R. I. Cukier, Physica **61**, 321 (1972); P. C. Hemmer, thesis (Trondheim, 1959); E. W. Montroll and Kurt E. Shuler, J. Chem. Phys. 26, 544 (1957); E. B. Davies, Commun. Math. Phys. 27, 309 (1972); 33, 171 (1973); 39, 91 (1974); J. V. Pulè, Commun. Math. Phys. 38, 241 (1974); Gérard G. Emch and John C. Wolfe, J. Math. Phys., 13, 1236 (1972); K. Hepp and E. Lieb, *Lecture Notes in Physics*, edited by H. Araki (Springer-Verlag, Berlin, 1975), Vol. 38, p. 178; Ph. A. Martin, "On the Stochastic Dynamics of *I*sing Models," preprint, Lausanne (1976).

[8]We refer the reader to Ref. 9 for the discussion of the distinction between a Pauli master equation and a Markov process, noticing also their footnote number 6.

[9]Irwin Oppenheim and Kurt E. Shuler, Phys. Rev. 138, 1007B (1965).

[10]This stands for a Kubo, Martin, and Schwinger state. See Ref. 11.
[11]R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957); P.C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959); R. Haag, N.M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).
[12]J. Manuceau, Ann. Inst. Heri Poincaré, **A8**, 139 (1968).
[13]F. Rocca, M. Sirugue, and D. Testard, Commun. Math. Phys. **19**, 119 (1970).
[14]This stands for the Gel'fand, Naimark, and Segal construction. See, e.g., Ref. 15.
[15]Gérard G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[16]A.C. Aitken, *Determinants and Matrices* (Oliver and Boyd, Edinburgh, 1956), 9th ed.
[17]N. Dunford and J.T. Schwartz, *Linear Operators, Part II* (Interscience, New York, 1963).
[18]J.A. Shohat and J.D. Tamarkin, *The Problem of Moments* (Am. Math. Soc., Providence, R.I., 1943).
[19]A.S. Krishnamoorthy and M. Parthasarathy, Ann. Math. Stat. **22**, 549 (1951); **31**, 229 (1960).
[20]P.R. Krishnaiah and M.M. Rao, Am. Math. Mon. **68**, 342 (1961).
[21]J.L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[22]R.C. Griffiths, Sankhya Ser. A **32**, 393 (1970).
[23]Eugene Wong, *Proceedings of Symposia in Applied Mathematics* (Am. Math. Soc., Providence, R.I., 1964), Vol. 16, pp. 264—76.
[24]This equation has been obtained in Refs. 4 and 5 from different premises.
[25]William Feller, Ann. Math. **54**, 173 (1950); **55**, 468 (1952).
[26]S. Karlin and J. McGregor, J. Math. Anal. Appl. **1**, 163 (1960).
[27]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (Dover, New York, 1965).

Exact diagonalization of relativistic Hamiltonians including a constant magnetic field

D. L. Weaver

Department of Physics, Tufts University, Medford, Massachusetts
(Received 8 March 1976)

It is shown how to exactly diagonalize the Dirac and Sakata-Taketani Hamiltonians, including the effect of a constant, external magnetic field, using a unitary transformation. In the latter case, the magnetic moment coupling must be of the Yang-Mills type in order to perform the transformation.

1. INTRODUCTION AND SPIN ½

One of the interesting features of the Dirac equation ($\hbar = c = 1$),

$$H_D\psi_D = \left(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp + \alpha_3 P_3 + m\beta + \frac{iq\kappa B}{4m}\beta\frac{(\boldsymbol{\alpha}\times\boldsymbol{\alpha})}{2}\right)\psi_D = E\psi_D, \tag{1}$$

for a spin-$\frac{1}{2}$ particle of charge q, mass m, and anomalous magnetic moment factor κ, moving in a constant external magnetic field $\mathbf{B} = (0, 0, B)$, so that $\mathbf{A} = (-\frac{1}{2}By, \frac{1}{2}Bx, 0)$ and $\boldsymbol{\pi}_\perp = \mathbf{p}_\perp - q\mathbf{A}_\perp$, is that the exact energy eigenvalues E may be obtained directly from Eq. (1) by a unitary transformation, to a diagonal form. The diagonalization can be seen as two steps. The first step, as noted by Tsai,[1] is

$$U_1 H_D U_1^{-1} = \left\{[m^2 + (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}_\perp)^2]^{1/2} - \frac{q\kappa B}{4m}\sigma_3\right\}\beta + \alpha_3 P_3, \tag{2}$$

where

$$U_1 = \exp[\tfrac{1}{2}\beta\boldsymbol{\alpha}\cdot\hat{\boldsymbol{\pi}}_\perp \tan^{-1}(|\boldsymbol{\pi}_\perp|/m)] \tag{3}$$

and $\boldsymbol{\sigma}$ is a representation of the Pauli spin matrices. It is interesting that U_1 is the generalization of the Melosh transformation[2] to include some external field effects, that U_1 commutes with the anomalous magnetic moment interaction, and that if $P_3 = 0$, a possibility since it is a constant of the motion for this problem, Eq. (2) is manifestly diagonal. In the $P \neq 0$ case, one may still obtain a diagonal Hamiltonian by a further unitary transformation with the result

$$U_2U_1H_DU_1^{-1}U_2^{-1} = \left[\left([m^2 + (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}_\perp)^2]^{1/2} - \frac{q\kappa B}{4m}\sigma_3\right)^2 + p_3^2\right]^{1/2}\beta, \tag{4}$$

where

$$U_2 = \exp\ \tfrac{1}{2}\beta\alpha_3 \tan^{-1}\left(\frac{P_3}{[m^2 + (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}_\perp)^2]^{1/2} - (q\kappa B/4m)\sigma_3}\right). \tag{5}$$

The main purpose of this paper is to look at a Hamiltonian form of spin-1 theory with a six-component wavefunction, the Sakata–Taketani theory,[3] under conditions similar to the above in the spin one-1 case, and to find the unitary transformation that exactly diagonalizes the Hamiltonian. It turns out that exact diagonalization is only possible for $\kappa = 1$, in which case the energy eigenvalues take a particularly simple form, equivalent to the spin-$\frac{1}{2}$ theory with $\kappa = 0$. The choice $\kappa = 1$ for a charged vector particle (to lowest order in the fine structure constant) is the value appropriate to Yang–Mills type coupling,[4] and, in fact, as emphasized by Lee,[5] such a choice should not even be regarded as anomalous.

2. SPIN ONE

Corresponding to Eq. (1), the Sakata–Taketani equation for the six-component wavefunction ψ_{ST} is

$$H_{ST}\psi_{ST} = \left[\left(m + \frac{(\pi_\perp^2 - qBS_3)}{2m}\right)\rho_3 + \left(\frac{(\pi_\perp^2 - qBS_3)}{2m} - \frac{(\mathbf{S}\cdot\boldsymbol{\pi}_\perp)^2}{m}\right)i\rho_2 - \frac{qBS_3}{2m}(1+\kappa)(\rho_3 + i\rho_2)\right]\psi_{ST} = E\psi_{ST}. \tag{6}$$

The matrices $\rho_{1,2,3}$ have the algebra of the Pauli spin matrices, and $\mathbf{S}$ are a representation of spin-1 matrices. Equation (6) may be written in a more convenient way as follows. Defining $\alpha \equiv \pi_\perp^2 - 2qBS_3$, one has

$$H_{ST}\psi_{ST} = \left[\left(m + \frac{\alpha}{2m}\right)\rho_3 + \left(\frac{\alpha}{2m} - \frac{(\mathbf{S}\cdot\boldsymbol{\pi}_\perp)^2}{m}\right)i\rho_2 + \frac{qBS_3}{2m}(1-\kappa)(\rho_3 + i\rho_2)\right]\psi_{ST} = E\psi_{ST}. \tag{7}$$

This is useful because α commutes with S_3 and $(\mathbf{S}\cdot\boldsymbol{\pi}_\perp)^2$. The next step is to expand $(\mathbf{S}\cdot\boldsymbol{\pi}_\perp)^2$ according to

$$(\mathbf{S}\cdot\boldsymbol{\pi}_\perp)^2 = \tfrac{1}{2}[\Sigma_1(\pi_1^2 - \pi_2^2) + \Sigma_2(\pi_1\pi_2 + \pi_2\pi_1)] + \tfrac{1}{2}qBS_3 + (\tfrac{1}{2} - S_3^2/a)\,\alpha, \tag{8}$$

where

$$\Sigma_1 \equiv S_1^2 - S_2^2, \quad \Sigma_2 \equiv S_1S_2 + S_2S_1. \tag{9}$$

Letting Σ_3 be S_3, the Σ_i have the algebra[6]

$$\Sigma_i\Sigma_j = \delta_{ij}S_3^2 + i\epsilon_{ijk}\Sigma_k, \tag{10}$$

the same as the Pauli matrix algebra (there $\sigma_3^2 = 1$ so $\sigma_i\sigma_j = \delta_{ij}\sigma_3^2 + i\epsilon_{ijk}\sigma_k$).

With this form for $(\mathbf{S}\cdot\boldsymbol{\pi}_\perp)^2$, one sees that there are two kinds of anticommutation relations involved in H_{ST}, those among the ρ_i and those among the Σ_i, and in some of the terms in H_{ST} both kinds occur. To unitarily transform H_{ST} to a form that is exactly diagonal, H_{ST} must be composed only of two terms that anticommute with one another plus terms that are already diagonal and commute with the unitary transformation. Examining Eqs. (7) and (8), this is only possible for $\kappa = 1$. In this special case, the Hamiltonian simplifies to

$$H_{ST} = \left(m + \frac{\alpha}{2m}\right)\rho_3 + \left(\frac{\alpha}{2m}(S_3^2 - 1) - \frac{1}{2m}[\Sigma_1(\pi_1^2 - \pi_2^2) + \Sigma_2(\pi_1\pi_2 + \pi_2\pi_1) + qBS_3]\right) i\rho_2. \quad (11)$$

It is convenient for later results to consider the diagonalization as taking place in two steps. The first step is

$$H'_{ST} = UH_{ST}U^{-1} = \left(m + \frac{\alpha}{2m}\right)\rho_3 + \left(\frac{\alpha}{2m}(S_3^2 - 1) - \frac{1}{2m}[\Sigma_+^2 + q^2B^2]^{1/2}S_3\right) i\rho_2, \quad (12)$$

where

$$\Sigma_+ \equiv \Sigma_1(\pi_1^2 - \pi_2^2) + \Sigma_2(\pi_1\pi_2 + \pi_2\pi_1) \text{ and } \Sigma_+^2 = S_3^2(\alpha^2 - q^2B^2).$$

The operator U has the form

$$U = \exp(S_3\Sigma_+\theta/2N), \quad (13)$$

where $N = (-S_3\Sigma_+)^2$ and $\tan\theta = N/(qB)$. The second step is to diagonalize the ρ_i dependence. The result is

$$H''_{ST} = VH'_{ST}V^{-1} = \left[\left(m + \frac{\alpha}{2m}\right)^2 - \left(\frac{\alpha}{2m}(S_3^2 - 1) - \frac{1}{2m}(\Sigma_+^2 + q^2B^2)^{1/2}S_3\right)^2\right]^{1/2}\rho_3 = (m^2 + \alpha)^{1/2}\rho_3, \quad (14)$$

where

$$V = \exp(\rho_2\rho_3\phi/2) \quad (15)$$

with

$$\tan\phi = \frac{-i\{(\alpha/2m)(S_3^2 - 1) - (1/2m)(\Sigma_+^2 + q^2B^2)^{1/2}S_3\}}{m + \alpha/2m} \quad (16)$$

and V is unitary in the Sakata–Taketani metric, $V\rho_3V^\dagger = 1$. As mentioned before, the form $(m^2 + \pi_\perp^2 - 2qBS_3)^{1/2}$ is the same eigenvalue for spin $\frac{1}{2}$, as well, when $\kappa = 0$.

The possibility of the exact diagonalization of Eq. (6) only for the special case $\kappa = 1$ and the similarity of the eigenvalues in that case with the spin $-\frac{1}{2}$ result for $\kappa = 0$, may be considered further circumstantial evidence on the electromagnetic properties of the vector bosons which are thought to mediate the weak interactions.[7]

DISCUSSION

It has been shown that the special case $\kappa = 1$ leads to a simplification of the Sakata–Taketani Hamiltonian. In particular, it allows H_{ST} to be diagonalized by a unitary transformation, and the resulting energy eigenvalues have the same form $(m^2 + \pi_\perp^2 - 2qBS_3)^{1/2}$ as the Dirac energy eigenvalues ($\kappa = 0$). The nonrelativistic limit is, of course, $m + a/2m$, and it is worth noting that the magnetic moment q/m, found from this approach does not agree with the "minimal" modification of the Schrödinger Hamiltonian $P^2/2m$, in contrast to the spin-$\frac{1}{2}$ case. Recall that for spin $\frac{1}{2}$, one can make the following modification:

$$P^2 \to \boldsymbol{\sigma}\cdot\mathbf{P}\,\boldsymbol{\sigma}\cdot\mathbf{P} \to \boldsymbol{\sigma}\cdot\boldsymbol{\pi}\,\boldsymbol{\sigma}\cdot\boldsymbol{\pi} = \pi^2 - 2q\mathbf{S}\cdot\mathbf{B}. \quad (17)$$

For spin 1 [in the representation with $(S_i)_{jk} = -i\epsilon_{ijk}$] the argument goes as follows:

$$P^2 \to \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}[P_1, P_2, P_3] + (\mathbf{S}\cdot\mathbf{P})^2 \to \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix}[\pi_1, \pi_2, \pi_3] + (\mathbf{S}\cdot\boldsymbol{\pi})^2 = \pi^2 - q\mathbf{S}\cdot\mathbf{B}, \quad (18)$$

where the two terms are the helicity squared (0 and 1) projection operators, so that a completely nonrelativistic argument gives $q/2m$ for the magnetic moment of the spin-1 particle.

It is, of course, possible to find the exact eigenvalues of Eq. (6), even when $\kappa \neq 1$. In that case, one first reduces the wave equation to three-component form, defining ψ_U to be the upper three-components of ψ_{ST}. The result in the representation with $\rho_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $i\rho_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is

$$\left[m^2 + \alpha + (1-\kappa)\frac{q^2B^2}{2m^2}S_3^2 + (1-\kappa)qB\left(1 + \frac{\alpha}{2m^2}\right)S_3 + (1-\kappa)\frac{qBi}{2m^2}[\Sigma_a(\pi_1^2 - \pi_2^2) - \Sigma_1(\pi_1\pi_2 + \pi_2\pi_1)]\right]\psi_U = E\psi_U,$$

where Eq. (10) has been used to simplify products. Applying a similarity transformation to the terms linear in Σ_i leads to the exact eigenvalues[6]

$$E^2 = m^2 + \alpha + (1-\kappa)\frac{q^2B^2}{2m^2}S_3^2 + (1-\kappa)qBS_3 \quad (19)$$

$$\times\left[\left(1 + \frac{\alpha}{2m}\right)^2 - \left(\frac{\alpha^2 - q^2B^2}{2m^2}\right)^2 S_3^2\right]^{1/2}; \quad (20)$$

the same ones as for the Proca theory.[8]

[1]W. Tsai, Phys. Rev. D 7, 1945 (1973).
[2]H.J. Melosh, Phys. Rev. D 9, 1095 (1974).
[3]M. Taketani and S. Sakata, Proc. Phys. Math. Soc. Jpn. 22, 757 (1939).
[4]C.N. Yang and R. Mills, Phys. Rev. 96, 191 (1954).
[5]T.D. Lee, Phys. Rev. 140, B967 (1965).
[6]D.L. Weaver, Tufts University preprint.
[7]See, for example, K.J. Kim and Y. Tsai, Phys. Rev. D 7, 3710 (1973), and references contain therein.
[8]A. Proca, C.R. Acad. Sci. Paris 202, 1490 (1936). See Ref. 1 and references contained therein for a considerable discussion of the exact eigenvalue problem.

Stochastic mechanics and dissipative forces

Bo-Sture K. Skagerstam

Institute of Theoretical Physics, Fack, S-402 20 Göteborg 5, Sweden
(Received 1 December 1975, revised manuscript received 14 April 1976)

We analyze a simple velocity dependent potential in the framework of stochastic mechanics. A nonlinear Schrödinger–Langevin equation is obtained. This equation turns out to have solutions with the remarkable property of giving an approach to stationary quantum states. Information theoretical aspects on the irreversible behavior of the model is also briefly discussed.

I. INTRODUCTION

It has been suggested[1,2] that stochastic mechanics, as developed in Refs. 1–5, in the case of dissipative forces can give us some understanding of how one should treat, in terms of quantum dynamics, forces which explicitly depend on velocities. In the theory of nuclear phenomena the dissipation of energy (or what can also be called *nuclear friction*) is not a very well understood topic and certainly there is a need for a theory corresponding to "quantized friction."[6,7]

In the present work we will use the machinery of stochastic mechanics in order to analyze the dynamics of a simple velocity dependent potential. Classically our model corresponds to a damped harmonic oscillator. Upon quantization in terms of stochastic processes we will arrive at a nonlinear dynamical equation, the Kostin equation.[8,9] Solutions in closed form will be given. These solutions will have the remarkable property of approaching stationary states of the harmonic oscillator for large times. Thus our model will exhibit an approach to equilibrium (to a stationary quantum state).

The formal setting is very simple and all calculations will be performed in the next section. In the last section we will give some general remarks and point to some information theoretical aspects concerning the irreversible behavior of the model.

We do not claim mathematical rigor and regard our discussion as heuristic. Some of the mathematical aspects have recently been discussed by Messer[10] and we intend to discuss our model in terms of the Weyl quantization in a future publication.

II. STOCHASTIC MECHANICS WITH DISSIPATIVE FORCES

In stochastic mechanics one assumes that the position $x(t)$ of a particle under consideration can be regarded as a stochastic process. The "quantization" is prescribed by giving the correlation function

$$E(dx(t)\,dx(t)) = 2D\,dt \tag{1}$$

with

$$D = \hbar/2m \tag{2}$$

and where $E(\cdot)$ denotes a (conditional) expectation value.[1] Higher order moments vanish by definition. The particle then performs a Markov process of the Wiener type.[1] Now suppose that there is a classical time-reversible conservative force f acting on the particle (we consider one-dimensional problems only). We can then write

$$f = -\frac{\partial}{\partial x}\phi(x), \tag{3}$$

where $\phi(x)$ is the corresponding potential. General dynamical equations[3,4] can then be derived for the stochastic process in terms of current velocity $v(x,t)$ and the stochastic velocity $u(x,t)$,[11]

$$f = m\left[\frac{\partial v}{\partial t}v + \left(\frac{\partial v}{\partial x}v - \frac{\partial u}{\partial x}u + D\frac{\partial^2 u}{\partial x^2}\right)\right] \tag{4}$$

and

$$\frac{\partial u}{\partial t} + D\frac{\partial^2 v}{\partial x^2} + \frac{\partial vu}{\partial x} = 0, \tag{5}$$

where m is the mass of the particle. If we now introduce the complex function

$$\Psi(x,t) \overset{\text{def}}{=} \exp(R(x,t))\exp\left(i\frac{S(x,t)}{2mD}\right), \tag{6}$$

where, by assumption,

$$mv(x,t) = \frac{\partial}{\partial x}S(x,t), \tag{7}$$

one can prove[1,12] that Ψ satisfies the Schrödinger wave equation

$$ih\frac{\partial\Psi(x,t)}{\partial t} = -\frac{h^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + \phi(x)\Psi(x,t), \tag{8}$$

where the function $R(x,t)$ will be related to the probability density (the kernel of the semigroup corresponding to the relevant Markov process) on the configuration space, $\rho(x,t)$, by

$$\ln\rho(\underline{x},t) = 2R(\underline{x},t). \tag{9}$$

The coupled system of nonlinear equations (4) and (5) can thus be linearized by the transformation (6) and the result is the Schrödinger equation. Hence conventional quantum mechanics can be formulated in terms of stochastic processes. We can then use either of the two mathematical schemes in order to obtain information about quantum dynamics.

We shall now see how one easily can extend stochastic mechanics so as to incorporate velocity dependent forces. We consider the most simple linear velocity dependent force classically given by

$$f = -m\beta v. \tag{10}$$

By incorporating this force into the framework of stochastic mechanics one arrives at the dynamical equations[4]

$$-m\beta v - m\omega^2 x = m\left[\frac{\partial v}{\partial t}v + \frac{\partial v}{\partial x}v - \left(\frac{\partial u}{\partial x}u + D\frac{\partial^2 u}{\partial x^2}\right)\right] \quad (11)$$

and Eq. (5). Here we have assumed an additional harmonic potential with angular frequency ω. By letting the "diffusion constant" D tend to zero, which corresponds to the Newtonian limit, one sees that the corresponding force is that of a damped harmonic oscillator

$$m\ddot{x}(t) = -m\beta\dot{x}(t) - m\omega^2 x(t), \quad (12)$$

where the dot $\cdot$ denotes differentiation with respect to time.

If we still assume that Eq. (7) is valid, one can rewrite Eq. (11) in the following manner:

$$m\ \frac{\partial v}{\partial t}v + \frac{\partial v}{\partial x}v - \frac{\partial u}{\partial x}u + D\frac{\partial^2 u}{\partial x^2} \ = -\frac{\partial}{\partial x}(\beta S + \phi), \quad (13)$$

where

$$\phi(x) = \tfrac{1}{2}m\omega^2 x^2. \quad (14)$$

The same procedure as that which lead to the Schrödinger equation (8), can now be applied and one arrives at the following equation[4]:

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + (\beta S + \phi)\Psi + \alpha(t)\Psi, \quad (15)$$

where $\alpha(t)$ is a time dependent constant which we will fix below. Equation (15) has been derived, in another context, by Kostin[8] using the Heisenberg—Langevin equation for a quantum Brownian particle interacting with a thermal environment as discussed by Ford, Kac, and Mazur.[13] A semiclassical derivation has also, independently, been derived by Kan and Griffin.[14]

We will now adjust the time dependent constant $\alpha(t)$ in such a manner that the sum of potential and kinetic energy equals the total energy at each instant of time. This can be regarded as a suitable "renormalization" of the ground state energy of the damped harmonic oscillator because, physically, it is now imbedded in some environment. This adjustment can now easily be achieved by noticing that in stochastic mechanics, as well as in quantum mechanics, we can use the following operator representation of the energy[15]:

$$E_{\text{op}} = i\hbar\frac{\partial}{\partial t} \quad (16)$$

and

$$E \overset{\text{def}}{=} \langle E_{\text{op}}\rangle = -\frac{\hbar^2}{2m}\left\langle\frac{\partial^2}{\partial x^2}\right\rangle + \langle\phi\rangle. \quad (17)$$

These equations imply that

$$\alpha(t) = -\int \rho(x')S(x',t)\,dx'. \quad (18)$$

Now, using the fact that

$$S(x,t) = (mD/i)\ln[\Psi(x,t)/\Psi^x(x,t)], \quad (19)$$

we can easily obtain the following Schrödinger—Langevin equation:

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + \tfrac{1}{2}m\omega^2 x^2\Psi(x,t) + \frac{\beta\hbar'}{2i}\ln\left[\frac{\Psi(x,t)}{\Psi^x(x,t)}\right]\Psi(x,t) - \frac{\beta\hbar'}{2i}\Psi(x,t) \times \int_{-\infty}^{+\infty} dx'\,|\Psi(x',t)|^2\ln\left[\frac{\Psi(x',t)}{\Psi^x(x',t)}\right]. \quad (20)$$

We notice that Eq. (20) has the remarkable property that every stationary solution of the harmonic oscillator is a solution. Moreover, it can be verified by explicit calculations that Eq. (20) also has the following type of solutions:

$$\Psi(x,t) = \psi_n(x-\xi(t))\exp(-iE_n t/\hbar) \times\exp[if(x,\xi(t),\dot{\xi}(t),t)], \quad (21)$$

where ψ_n is the nth eigenstate of the harmonic oscillator and

$$F_n = (n+\tfrac{1}{2})\hbar\omega. \quad (22)$$

The function $f(x,\xi,\dot{\xi},t)$ is given in closed form by

$$f(x,\xi(t),\dot{\xi}(t),t) = (m/\hbar)x\cdot\dot{\xi}(t) + g(\xi(t),\dot{\xi}(t),t), \quad (23)$$

where $\xi(t)$ satisfies the ordinary differential equation

$$\ddot{\xi}(t) + \beta\dot{\xi}(t) + \omega^2\xi(t) = 0 \quad (24)$$

and $g(\xi,\dot{\xi},t)$ is determined by

$$\hbar\frac{dg(\xi(t),\dot{\xi}(t),t)}{dt} = m\beta\xi(t)\dot{\xi}(t) + \tfrac{1}{2}m\omega^2\xi^2(t) - m\frac{\dot{\xi}^2(t)}{2}. \quad (25)$$

It is now a matter of straightforward calculations to find an explicit expression of the mean energy, which turns out to have the form

$$E(t) = \left(\Psi, i\hbar\frac{\partial}{\partial t}\Psi\right) = (n+\tfrac{1}{2})\hbar\omega + \tfrac{1}{2}m\xi_0^2\exp(-\beta t)[(1+\beta^2)\cos^2(\Omega t+\Omega_0) + \Omega^2\sin^2(\Omega t+\Omega_0) + \beta\Omega\sin(2\Omega t+2\Omega_0)] \quad (26)$$

if the solutions of Eq. (24) are parametrized as follows:

$$\xi(t) = \xi_0\exp(-\tfrac{1}{2}\beta t)\cos(\Omega t+\Omega_0) \quad (27)$$

where

$$2\Omega = (4\omega^2-\beta^2)^{1/2} \quad (28)$$

which we assume is positive. We notice that, for large times t, the solution (21) approaches the nth stationary state of the harmonic oscillator and that Eq. (26) implies that

$$\lim_{t\to\infty} E(t) = (n+\tfrac{1}{2})\hbar\omega. \quad (29)$$

Hence we see that our approach to velocity dependent potentials, in a special case, exhibits a remarkable property of an "approach to equilibrium," where equilibrium corresponds to a stationary state of a harmonic oscillator. The solutions of the form (21) have also been obtained, independently, by Kan and Griffin[14] for the special case when $n=0$, and in general by Kan.[16] The question of uniqueness of these solutions is briefly discussed in Ref. 17.

The current and stochastic velocities can now be calculated. For the ground state of the harmonic oscillator, i.e., $n=0$, one finds that

$$v=-\xi_0\exp(-\tfrac{1}{2}\beta t)[\cos(\Omega t+\Omega_0)-\sin(\Omega t+\Omega_0)] \tag{30}$$

and

$$u=-\omega(x-\xi(t)). \tag{31}$$

We see that (30) and (31) tend to their "equilibrium" values,[4] when the time t tends to infinity.

We can then characterize the corresponding stochastic process completely by the following stochastic differential equation:

$$dx(t)=(-\omega x(t)+\omega\xi(t)+\dot{\xi}(t))\,dt+dw(t), \tag{32}$$

where $\omega(t)$ is a Wiener process with diffusion constant D. For each function ξ with continuous time derivative, (32) has the solution[1,4]

$$x(t)=\xi(t)+\int_0^t\exp[-\omega(t-s)]\,d\omega(s),\quad t\geq 0. \tag{33}$$

The two-point function can then easily be evaluated,

$$E(x(t)x(t'))=\frac{\hbar}{2m\omega}\exp(-\omega\,|t-t'|)+\xi(t)\xi(t'). \tag{34}$$

For large times t and t', this decays into the corresponding correlation for the harmonic oscillator, *independent of the initial conditions of* $\xi(t)$.

As a final remark we stress that the uncertainty relation between momentum, and position[17] is valid for all times. This means that we have quantized a system with energy dissipation, without having any contradiction with the Heisenberg uncertainty principle, which has been a central difficulty in previous works (see, e.g., Ref. 7, where additional references can be found).

III. CONCLUSION

In the preceding paragraph we showed that a "quantized" damped harmonic oscillator exhibits a remarkable structure namely that of asymptotic stationary states.[18] Hence we can have the coexistence of damped and stationary solutions, although the superposition principle, in this special case, is no longer valid. We showed this explicitly for the harmonic oscillator but our method works in other cases. A closed form can, e.g., be given for the free motion and the constant force case. We do not discuss these solutions in the present work since these have been discovered, independently, by other people (see, e.g., Refs. 19–21).

Let us now return to the case of the harmonic oscillator and briefly discuss the apparent irreversible behavior. We restrict ourselves to the ground state of the harmonic oscillator. The mean forward velocity can easily be calculated if one uses, e.g., the stochastic differential equation (32). One finds that

$$v_+(x,t)\overset{\text{def}}{=}\lim_{dt\to 0}E\left(\frac{dx(t)}{dt}\right)=-\omega x(t)+\omega\xi(t)+\dot{\xi}(t). \tag{35}$$

The normalized probability density $\rho(x,t)$ will then satisfy the (forward) Fokker–Planck equation

$$\frac{\partial\rho}{\partial t}+\frac{\partial}{\partial x}(v_+\rho)-D\frac{\partial^2\rho}{\partial x^2}=0. \tag{36}$$

With the choice (35) for the forward mean velocity one finds the following unique solution:

$$\rho(x,t)=\left(\frac{m\omega}{\hbar\pi}\right)^{1/2}\exp\left(-\frac{m\omega}{\hbar}[x-\xi(t)]^2\right) \tag{37}$$

which, of course, also could have been obtained from the solution in the form (21). If the probability density $\rho(x,t)$ is time independent one can show, by using the continuity equation for $\rho(x,t)$, that $v_+(x,t)$ must be time independent. For the ground state of the harmonic oscillator this fact corresponds to the choice

$$v_+(x,t)=-\omega x(t). \tag{38}$$

The Fokker–Planck equation then, of course, gives the following unique normalized solution:

$$\rho_0(x)=\left(\frac{m\omega}{\hbar\pi}\right)^{1/2}\exp\left(-\frac{m\omega}{\hbar}x^2\right). \tag{39}$$

We see that every solution of the form (37), independently of the initial conditions of the $\xi(t)$ function, approaches the stationary solution (39). This is typical irreversible behavior.

According to information theory the quantity

$$I(\rho(t))\overset{\text{def}}{=}\int dx\,\rho(x,t)\ln\rho(x,t) \tag{40}$$

is the unique (to within a positive multiplicative constant) measure of the *information* contained in the probability distribution ρ.[22] Straightforward calculations now show that

$$I(\rho(t))=I(\rho_0(t))\overset{\text{def}}{=}I_0, \tag{41}$$

where I_0 is a constant. Hence our total information about the system is constant in time, which should be obvious, since the distributions ρ_0 and ρ satisfy a continuity equation.

In order to measure the irreversible behavior in terms of information-theoretical concepts, one can introduce the relative entropy, $I(\rho(t)\,|\,\rho_0)$ defined by the following equation[23]:

$$I(\rho(t)\,|\,\rho_0)\overset{\text{def}}{=}\int dx\,\rho(x,t)\ln(\rho(x,t)/\rho_0(x)). \tag{42}$$

In our case, $I(\rho(t)\,|\,\rho_0)$ can easily be evaluated and one arrives at the expression

$$I(\rho(t)\,|\,\rho_0)=(2\omega/D)\xi^2(t) \tag{43}$$

which tends to zero as the time parameter tends to infinity. This will have the consequence that

$$\lim_{t\to\infty}\rho(x,t)=\rho_0(x) \tag{44}$$

in a suitable sense. We also notice that $I(\rho(t)\,|\,\rho_0)$ is a measure of the potential energy of the classical damped harmonic oscillator, and more precisely

$$I(\rho(t)\,|\,\rho_0)=(2\omega/\hbar)[E(t)-E_0-\tfrac{1}{2}m\dot{\xi}^2(t)]. \tag{45}$$

Although many of the statements in the present paper cannot be considered as "rigorous" we believe that stochastic mechanics can be useful in the study of some nonstandard quantum mechanical problems.

ACKNOWLEDGMENTS

A large part of the present work has been carried out during a stay at Bedford College, University of London, and the author wishes to thank Professor R. F. Streater for his kind hospitality. Fruitful communications with Professor L. de la Peña-Auerbach (University of Mexico) and Dr. K. K. Kan and Dr. J. Messer are also gratefully acknowledged.

[1]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, N.J., 1967).
[2]L. de la Pena-Auerbach and A. M. Cetto, Found. Phys. **5**, 355 (1975).
[3]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[4]B. Skagerstam, "Stochastic Processes on a Classical Configuration Space," Institute of Theoretical Physics report 75-21, Gothenburg, Sweden (1975).
[5]For a general discussion of *stochastic mechanics* see, e.g., J. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974).
[6]E. D. Mshelia, W. Scheid, and W. Greiner, Nuovo Cimento A **30**, 589 (1975), and references cited therein.
[7]K. Albrecht, Phys. Lett. B **56**, 127 (1975).
[8]M. D. Kostin, J. Chem. Phys. **57**, 3589 (1972).
[9]M. D. Kostin, J. Stat. Phys. **12**, 145 (1975).
[10]J. Messer, "On the Theory of Quantum Mechanical Langevin Systems," Princeton University (1975) and Diplomarbeit (Darmstadt) (1974), unpublished.
[11]We are following the notation of Refs. 3 and 4.
[12]S. Albeverio and R. Höegh-Krohn, J. Math. Phys. **15**, 1745 (1974).
[13]G. W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6**, 504 (1965).
[14]K. K. Kan and J. J. Griffin, Phys. Lett. B **50**, 241 (1974).
[15]L. de la Pena-Auerbach, J. Math. Phys. **10**, 1620 (1969).
[16]K. K. Kan, private communication.
[17]L. de la Pena-Auerbach and A. M. Cetto, Phys. Lett. A **39**, 65 (1972).
[18]B. Skagerstam, Phys. Lett. B **58**, 21 (1975).
[19]J. D. Immele, K. K. Kan, and J. J. Griffin, Nucl. Phys. A **241**, 47 (1975).
[20]R. W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[21]G. J. Papadopoulos, Phys. Rev. D **11**, 2870 (1975).
[22]F. Reza, *An Introduction to Information Theory* (McGraw-Hill, New York, 1961). See also B. Skagerstam, J. Stat. Phys. **12**, 449 (1975).
[23]See, e.g., G. Lindblad, J. Stat. Phys. **11**, 231 (1974).

Electromagnetic eigenmode perturbations caused by deformation of a plasma core in a spherical cavity resonator*

Arno D. Steiger

Lawrence Livermore Laboratory, University of California, Livermore, California 94550
(Received 12 July 1976)

A perturbation formalism for the electromagnetic eigenmodes of a lossless spherical cavity resonator has been derived from the Boltzmann–Ehrenfest theorem. The perturbations are caused by adiabatic deformations of a plasma core located at the center of the cavity.

1. INTRODUCTION

For a lossless spherical microwave resonator we analyze the eigenmode perturbations caused by surface deformations of a nontransparent plasma core located at the center of the cavity. An immediate consequence of the deviations from spherical symmetry of the plasma-field interface is the complete or partial removal of the azimuthal degeneracy of the characteristic electromagnetic multipole fields. Our perturbation calculation is based on the Boltzmann—Ehrenfest adiabatic theorem.[1–3] The applicability of this intuitively appealing approach hinges on the assumption that the deformations of the plasma boundary take place on a time scale which is long in comparison with the oscillation period of the microwave field. Such deformations may be produced by acoustic surface vibrations or by low frequency electrostatic surface oscillations of the plasma which can exist only in the presence of an external electromagnetic field.[4]

For the sake of simplicity we assume the plasma boundary to be sharp, thereby neglecting the thermal motion of the particles. In conformity with the adiabatic approximation for the plasma surface deformations, we treat the plasma as a perfect conductor. The assumption of infinite plasma conductivity, which can be justified only as long as plasma heating by the microwave field may be disregarded, has also been made by Yankov[5] in a theoretical study of the stability of a homogeneous plasma sphere in external uniform and quasiuniform electromagnetic fields, and by Butler[6] in an equilibrium analysis of a dense spherical plasma core located in the center of a spherical microwave cavity in which there exists a rotating transverse electric multipole field of lowest order together with a uniform d.c. magnetic field. High-conductivity plasma cores have been simulated by copper spheres, spheroids and other configurations in an analog experiment[7] devised by Hatch and Butler to study the conditions for the equilibrium of dense plasma cores contained by microwave cavity fields.

Describing the plasma surface deformations by multipole parameters, which are spherical tensor components, and expressing the characteristic electromagnetic fields in terms of vector spherical harmonics enables us to take advantage of the elegant and powerful techniques supplied by the theory of the irreducible representations of the three-dimensional rotation group. These techniques are used extensively in atomic and nuclear physics. By means of the Clebsch—Gordan reduction and the Racah recoupling transformation we obtain concise expressions for the components of the eigenfrequency perturbation matrices.

In Sec. 2 we discuss the characteristic electromagnetic modes for a domain bounded by two concentric spherical shells of infinite conductivity. In accordance with the approximations mentioned above, these characteristic modes represent the unperturbed resonant electromagnetic field set up in the space between the plasma core and the (superconducting) metallic cavity wall. Expressions for the time-average electromagnetic energy of these eigenmodes are derived in Appendix A. Section 3 contains the formal analysis of the eigenmode perturbations induced by deviations from spherical symmetry of the plasma-field interface; in this section we make use of results derived in the Appendices A, B, and C.

2. EIGENMODES OF A MICROWAVE CAVITY BOUNDED BY TWO PERFECTLY CONDUCTING CONCENTRIC SPHERES

In the present section we analyze the electromagnetic eigenmodes of a lossless spherical cavity resonator which is partially filled with a plasma sphere at its center. The plasma is treated in the limit of infinite conductivity. We therefore deal with the problem of determining the characteristic modes of a cavity bounded by two concentric spheres $r=r_0$ and $r=R_0$ $(r_0<R_0)$ which are perfect conductors. Infinite conductivity of the cavity walls entails that the tangential component of the electric field **E** and the normal component of the magnetic field **B** vanish at $r=r_0$ and $r=R_0$.

For harmonic time-dependence of the electromagnetic field vectors,

$$\mathbf{E}(t,\mathbf{x})=\mathbf{E}(\mathbf{x})\exp(-i\omega t),\quad \mathbf{B}(t,\mathbf{x})=\mathbf{B}(\mathbf{x})\exp(-i\omega t), \tag{2.1}$$

the Maxwell equations in a source-free domain of the vacuum are equivalent to either one of the two sets of equations

$$(\nabla^2+k^2)\mathbf{E}=0, \tag{2.2a}$$

$$\nabla\cdot\mathbf{E}=0, \tag{2.2b}$$

$$\mathbf{B}=-(i/k)\nabla\times\mathbf{E}, \tag{2.2c}$$

or

$$(\nabla^2 + k^2)\mathbf{B} = 0, \tag{2.3a}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{2.3b}$$

$$\mathbf{E} = (i/k)\nabla \times \mathbf{B}, \tag{2.3c}$$

where $k = \omega/c$, with c denoting the vacuum light velocity.

In a spherical coordinate system (r, θ, ϕ) a solution of the vector Helmholtz equation (2.2a), which also satisfies (2.2b) is[8]

$$\mathbf{E}_{lm} = f_l(kr)\mathbf{r} \times \nabla Y_{lm}(\theta, \phi), \quad l = 1, 2, 3, \cdots, \quad -l \leq m \leq l, \tag{2.4a}$$

where $f_l(kr)$ denotes either a particular spherical Bessel function or a superposition of two linearly independent spherical Bessel functions, and where $Y_{lm}(\theta, \phi)$ is the spherical harmonic of order (l, m). The solution (2.4a) together with the corresponding magnetic field

$$\mathbf{B}_{lm} = -\frac{i}{k}\nabla \times \mathbf{E}_{lm} \tag{2.4b}$$

is referred to as the *magnetic multipole field* of order (l, m). From Eq. (2.4a) it is immediate that

$$\mathbf{r} \cdot \mathbf{E}_{lm} = 0. \tag{2.5}$$

Magnetic multipole fields are therefore also called transverse electric (TE) multipole fields. The transverse magnetic (TM), or *electric*, *multipole field* of order (l, m) is defined by[8]

$$\mathbf{B}_{lm} = f_l(kr)\mathbf{r} \times \nabla Y_{lm}(\theta, \phi), \quad l = 1, 2, 3, \cdots, \quad -l \leq m \leq l \tag{2.6a}$$

and

$$\mathbf{E}_{lm} = \frac{i}{k}\nabla \times \mathbf{B}_{lm}. \tag{2.6b}$$

The expression (2.6a) satisfies the vector Helmholtz equation (2.3a) as well as the source equation (2.3b), and Eq. (2.6b) corresponds to Eq. (2.3c); Eq. (2.6a) implies the transversality of $\mathbf{B}_{lm}$:

$$\mathbf{r} \cdot \mathbf{B}_{lm} = 0. \tag{2.7}$$

Since the coordinate origin is excluded from the domain $(r_0 \leq r \leq R_0)$, we represent the radial function $f_l(kr)$ as a linear combination of the Bessel functions of the first and the second kind, $j_l(kr)$ and $n_l(kr)$,

$$f_l(kr) = C_l j_l(kr) + D_l n_l(kr). \tag{2.8}$$

The characteristic TE modes of the electromagnetic field in the domain $r_0 \leq r \leq R_0$, bounded by two perfect conductors, are determined by Eqs. (2.4) in conjunction with the boundary conditions

$$[f_l(kr)]_{r=r_0} = 0, \quad [f_l(kr)]_{r=R_0} = 0, \tag{2.9}$$

which ensure that the tangential components of $\mathbf{E}$ and the normal component of $\mathbf{B}$ vanish at the plasma-field interface and on the cavity wall. On account of (2.8) the boundary conditions (2.9) lead to the system of homogeneous equations

$$C_l j_l(kr_0) + D_l n_l(kr_0) = 0, \quad C_l j_l(kR_0) + D_l n_l(kR_0) = 0, \tag{2.10}$$

for which a nontrival solution exists if and only if

$$j_l(kr_0)n_l(kR_0) = j_l(kR_0)n_l(kr_0). \tag{2.11}$$

The nth root of this transcendental equation determines the characteristic frequency

$$\omega_{nl}^{\mathrm{TE}} = ck_{nl}^{\mathrm{TE}} \tag{2.12}$$

of the transverse electric multipole mode TE_{nlm}. Since for each value of l there are $(2l+1)$ values of m, it is obvious that each characteristic frequency ω_{nl} is $(2l+1)$-fold. On account of (2.4) and (2.10)–(2.12) the electromagnetic field in the TE_{nlm} mode is

$$\mathbf{E}_{nlm} = f_l(k_{nl}r)\mathbf{r} \times \nabla Y_{lm}, \quad \mathbf{B}_{nlm} = -(i/k_{nl})\nabla \times \mathbf{E}_{nlm}, \tag{2.13a}$$

with

$$f_l(k_{nl}r) = C_{nl}[j_l(k_{nl}r) + \alpha_{nl}n_l(k_{nl}r)], \tag{2.13b}$$

where

$$\alpha_{nl} = \alpha_{nl}^{\mathrm{TE}} = -[j_l(k_{nl}r_0)/n_l(k_{nl}r_0)] = -[j_l(k_{nl}R_0)/n_l(k_{nl}R_0)], \tag{2.13c}$$

and where

$$k_{nl} = k_{nl}^{\mathrm{TE}}. \tag{2.13d}$$

The relationship between the normalization constant C_{nl} and the total electromagnetic energy of the TE_{nlm} mode is established by Eq. (A10).

The characteristic TM modes of the electromagnetic field in the domain $r_0 \leq r \leq R_0$ are determined by Eqs. (2.6) together with the boundary conditions

$$\left[\frac{\partial f_l(kr)}{\partial r}\right]_{r=r_0} = 0, \quad \left[\frac{\partial f_l(kr)}{\partial r}\right]_{r=R_0} = 0, \tag{2.14}$$

which according to (2.6) ensure the vanishing on the cavity walls of the tangential components of E and the normal component of B. With (2.8) and (2.14) we obtain the system of homogeneous equations

$$\begin{aligned} C_l \frac{d}{dr}[rj_l(kr)]_{r=r_0} + D_l \frac{d}{dr}[rn_l(kr)]_{r=r_0} &= 0, \\ C_l \frac{d}{dr}[rj_l(kr)]_{r=R_0} + D_l \frac{d}{dr}[rn_l(kr)]_{r=R_0} &= 0, \end{aligned} \tag{2.15}$$

for which nontrivial solutions exist if and only if

$$\left(\frac{d}{dr}[rj_l(kr)]_{r_0}\right)\left(\frac{d}{dr}[rn_l(kr)]_{R_0}\right) = \left(\frac{d}{dr}[rj_l(kr)]_{R_0}\right)\left(\frac{d}{dr}[rn_l(kr)]_{r_0}\right). \tag{2.16}$$

The characteristic frequency

$$\omega_{nl}^{\mathrm{TM}} = ck_{nl}^{\mathrm{TM}} \tag{2.17}$$

of the transverse magnetic multipole mode TM_{nlm} is then determined by the nth root of the transcendental equation (2.16); it is clearly again a $(2l+1)$-fold eigenfrequency. With (2.6), (2.15)–(2.17), and (A14) we write for the electromagnetic field in the TM_{nlm} mode,

$$\mathbf{B}_{nlm} = f_l(k_{nl}r)\mathbf{r} \times \nabla Y_{lm}, \quad \mathbf{E}_{nlm} = (i/k_{nl})\nabla \times \mathbf{B}_{nlm} \tag{2.18a}$$

with

$$f_l(k_{nl}r) = C_{nl}[j_l(k_{nl}r) + \alpha_{nl}n_l(k_{nl}r)], \tag{2.18b}$$

where

$$\alpha_{nl} = \alpha_{nl}^{\mathrm{TM}} = -\frac{k_{nl}r_0 j_{l-1}(k_{nl}r_0) - l j_l(k_{nl}r_0)}{k_{nl}r_0 n_{l-1}(k_{nl}r_0) - l n_l(k_{nl}r_0)}$$

$$= -\frac{k_{nl}R_0 j_{l-1}(k_{nl}R_0) - l j_l(k_{nl}R_0)}{k_{nl}R_0 n_{l-1}(k_{nl}R_0) - l n_l(k_{nl}R_0)} \tag{2.18c}$$

and where

$$k_{nl} = k_{nl}^{\mathrm{TM}}. \tag{2.18d}$$

The relationship between the normalization constant C_{nl} and the time-average electromagnetic energy of the TM_{nlm} mode is given by (A11).

3. SHAPE PERTURBATIONS OF THE PLASMA-FIELD INTERFACE

A deformation of the plasma-field interface modifies the electromagnetic eigenmodes which can exist in the domain bounded by the plasma surface and the metallic cavity wall. Changes of the characteristic field frequencies which are caused by small adiabatic shape perturbations of the plasma surface may be determined by means of the Boltzmann–Ehrenfest adiabatic theorem.[1-3] This theorem is applicable provided that the shape perturbations occur on a time scale which is long in comparison with the field oscillation period and that these perturbations are not large enough to induce transitions between adjacent modes. For the present purpose, the Boltzmann–Ehrenfest theorem asserts that the total time-average energy $\bar{U}$ of the electromagnetic field in the domain defined by the plasma boundary and the cavity wall divided by the characteristic frequency ω associated with this domain remains unchanged.

$$\bar{U}/\omega = \text{invariant}, \tag{3.1}$$

if the plasma boundary deformations are continuous and sufficiently small; the latter requirement means that the radial deviations δr of the deformed surface from the spherical surface are small compared to the wavelength, $2\pi c/\omega_{nl}$, of the unperturbed eigenmode. From (3.1) it is immediate that

$$\frac{\delta\omega}{\omega} = \frac{\delta\bar{U}}{\bar{U}}, \tag{3.2}$$

where $\delta\omega$ and $\delta\bar{U}$ denote the changes of ω and $\bar{U}$ which are caused by the deformation of the plasma surface. The force, $\mathbf{F}_p$, exerted by the electromagnetic field on a unit area of the undeformed plasma boundary surface is

$$\mathbf{F}_p = -\mathbf{e}_r\cdot\boldsymbol{\sigma} = -(\sigma_{rr}\mathbf{e}_r + \sigma_{r\theta}\mathbf{e}_\theta + \sigma_{r\phi}\mathbf{e}_\phi), \tag{3.3}$$

where $\mathbf{e}_r$, $\mathbf{e}_\theta$, $\mathbf{e}_\phi$ are the unit vectors of the spherical coordinate system (r,θ,ϕ), and where $\boldsymbol{\sigma}$ denotes the Maxwell stress tensor, whose components are[8,9]

$$\sigma_{\alpha\beta} = \frac{1}{4\pi}[-E_\alpha E_\beta - B_\alpha B_\beta + \tfrac{1}{2}\delta_{\alpha\beta}(E^2+B^2)],$$
$$\alpha = r,\theta,\phi, \quad \beta = r,\theta,\phi. \tag{3.4}$$

Since on the surface of a perfect conductor the tangential components of the electric field and the normal component of the magnetic field vanish, the electromagnetic pressure is always normal to such a surface. Therefore, Eq. (3.3) becomes, in accordance with (3.4),

$$\mathbf{F}_p = -\sigma_{rr}\mathbf{e}_r = -\mathbf{e}_r\frac{1}{8\pi}[B_\theta^2 + B_\phi^2 - E_r^2]_{(r=r_0)}. \tag{3.5}$$

Here, the subscript $(r=r_0)$ indicates that the field components B_θ, B_ϕ, and E_r are to be evaluated at $r=r_0$. The work, $-\delta\bar{U}$, performed by the electromagnetic pressure (3.5) during the adiabatic process of surface deformation is

$$-\delta\bar{U} = \int da(-\delta r\mathbf{e}_r)\cdot\mathbf{F}_p, \tag{3.6}$$

where da denotes the element of surface area. Since the radial deviation δr is assumed to be a continuous function of the angular coordinates θ and ϕ for which

$$|\delta r| \ll r_0, \quad \left|\frac{\partial}{\partial\theta}\delta r\right| \ll |\delta r|, \quad \left|\frac{\partial}{\partial\phi}\delta r\right| \ll |\delta r|, \tag{3.7}$$

we are justified in replacing the unit vector normal to the deformed surface by the unit radial vector $\mathbf{e}_r$. With (3.5) and (3.6) we may then write for the adiabatic change $\delta\bar{U}$ of the time-average of the total electromagnetic energy stored between the plasma core and the cavity wall,

$$\delta\bar{U} = \frac{1}{8\pi}r_0^2\int_0^{2\pi}d\phi\int_0^{\pi}\sin\theta\,d\theta\,\delta r(\theta,\phi)[\overline{E_r^2} - \overline{B_\theta^2} - \overline{B_\phi^2}]_{r_0}. \tag{3.8}$$

An adiabatic and volume conserving deformation of the plasma-field interface is stable if

$$\delta\bar{U} > 0, \tag{3.9}$$

or, by virtue of (3.2), if

$$\delta\omega > 0. \tag{3.10}$$

The assumption of infinite plasma conductivity ensures conservation of the photon number N during the deformation process. The stability conditions (3.9) and (3.10) are therefore directly related by the equation

$$\delta\bar{U} = N\hbar\delta\omega. \tag{3.11}$$

It is convenient to describe the shape of a surface which deviates only slightly from spherical symmetry by means of a multipole parameterization.[10] We therefore characterize the deformed plasma-field interface by an expansion in spherical harmonics of the radius vector $r(\theta,\phi)$ from the center of the sphere to a point with angular coordinates (θ,ϕ) on the deformed surface,

$$r(\theta,\phi) = r_0\left\{1 + \sum_{s=0}^{\infty}\sum_{\sigma=-s}^{s}\beta_{s\sigma}Y_{s\sigma}(\theta,\phi)\right\} + O(\beta^2). \tag{3.12}$$

Since $r(\theta,\phi)$ is a real function of the angular variables, and since $Y_{s\sigma}^* = (-1)^{-\sigma}Y_{s,-\sigma}$, the multipole parameters $\beta_{s\sigma}$ satisfy the relation

$$\beta_{s\sigma}^* = (-1)^\sigma\beta_{s,-\sigma}. \tag{3.13}$$

The surface deformation is measured by the distance

$$\delta r(\theta,\phi) = r_0\sum_{s,\sigma}\beta_{s\sigma}Y_{s\sigma}(\theta,\phi). \tag{3.14}$$

For each multipole of order s there are $(2s+1)$ parameters $\beta_{s\sigma}$ $(-s\le\sigma\le s)$. Equation (3.13) implies that the parameter β_{s0} is real and that the $2s$ complex parameters $\beta_{s\sigma}$ with $\sigma\ne 0$ are not independent but are equivalent to $2s$ independent real parameters. If $s>1$, three of these parameters specify the orientation of the deformed surface, so that $2s-2$ parameters are associated with its intrinsic shape. A mere change in volume of the plasma core without shape distortion of the wall is specified by the monopole parameter β_{00}. The dipole parameters $\beta_{1\sigma}$ ($\sigma=-1,0,1$) represent a displacement of the center of mass of the plasma core and are not related to its shape, as the $|\beta_{1\sigma}|$ are small, in accordance with the first relation in (3.7). We therefore restrict the summation in expression (3.14) to include only terms with $s>1$.

Since for a given TE_{nlm} or TM_{nlm} mode the corresponding eigenfrequency, Eq. (2.12), or Eq. (2.17), as well as the corresponding electromagnetic energy, Eq. (A10), or Eq. (A11), are independent of $m(-l \le m \le l)$, we are confronted with a degenerate perturbation problem when considering distortions of the plasma-field interface. Because of the azimuthal degeneracy of the TE_{nlm} or the TM_{nlm} modes, a particular linear combination of the $2l+1$ independent multipole fields (2.13) or (2.18) which belong to the same value of n and to the same value of l but to different values of m has to be substituted in (3.8). A deformation of the plasma-field interface will either partially or completely remove this degeneracy. As a zero order magnetic or electric multipole solution of Maxwell's equations in the cavity domain bounded by the deformed surface (3.12) and the cavity wall $r=R_0$ we now choose the linear combination

$$\begin{Bmatrix}\mathcal{E}_{nl\mu}\\ \mathcal{B}_{nl\mu}\end{Bmatrix} = \sum_{m=-l}^{l} A(n,l)_{\mu m}\begin{Bmatrix}\mathbf{E}_{nlm}\\ \mathbf{B}_{nlm}\end{Bmatrix}, \quad -l \le \mu \le l, \tag{3.15}$$

where the coefficients $A(n,l)_{\mu m}$ are functions of the deformation parameters $\beta_{s\sigma}$. In order to ensure the same normalization and orthogonality for the electromagnetic field represented by the linear superposition (3.15) as exists for the field vectors $(\mathbf{E}_{nlm}, \mathbf{B}_{nlm})$, the coefficients $A(n,l)_{\mu m}$ must satisfy the unitarity conditions

$$\sum_m A(n,l)^*_{\mu' m} A(n,l)_{\mu m} = \delta_{\mu'\mu}$$
$$\sum_\mu A(n,l)^*_{\mu m'} A(n,l)_{\mu m} = \delta_{m'm}. \tag{3.16}$$

As an immediate consequence of (A1), (A3), and (A5) we have the orthonormality condition

$$\frac{1}{16\pi U_{nl}}\int_{V_{ca}} dV[\mathbf{E}^*_{nlm'}\cdot \mathbf{E}_{nlm} + \mathbf{B}^*_{nlm'}\cdot \mathbf{B}_{nlm}] = \delta_{m'm}. \tag{3.17}$$

In (A10) and (A11) the time-average electromagnetic energies for TE and TM modes are expressed in terms of a single field normalization constant. With (3.15), (3.16), and (3.17) we have

$$\frac{1}{16\pi U_{nl}}\int_{V_{ca}} dV[\mathcal{E}^*_{nl\mu'}\cdot \mathcal{E}_{nl\mu} + \mathcal{B}^*_{nl\mu'}\cdot \mathcal{B}_{nl\mu}] = \delta_{\mu'\mu}. \tag{3.18}$$

The linear combination (3.15) represents the particular electromagnetic field which the perturbed field approaches in the limit of vanishing deformation of the plasma-field interface. By replacing in (3.8) the time-average (real) field components with the corresponding (complex) field components defined by (3.15) we obtain

$$\delta\bar{U}_{nl\mu'\mu} = -\frac{r_0^2}{16\pi}\int d(\theta,\phi)\,[(\mathcal{B}_{nl\mu'})^*_\theta(\mathcal{B}_{nl\mu})_\theta + (\mathcal{B}_{nl\mu'})^*_\phi(\mathcal{B}_{nl\mu})_\phi]_{(r=r_0)}\,\delta r(\theta,\phi) \tag{3.19a}$$

for a TE mode, and for a TM mode

$$\delta\bar{U}_{nl\mu'\mu} = \frac{r_0^2}{16\pi}\int d(\theta,\phi)[(\mathcal{E}_{nl\mu'})^*_r(\mathcal{E}_{nl\mu})_r - (\mathcal{B}_{nl\mu'})^*_\theta(\mathcal{B}_{nl\mu})_\theta - (\mathcal{B}_{nl\mu'})^*_\phi(\mathcal{B}_{nl\mu})_\phi]_{(r=r_0)}\,\delta r(\theta,\phi), \tag{3.19b}$$

where

$$\int d(\theta,\phi)\cdots = \int_0^{2\pi} d\phi \int_0^{\pi}\sin\theta\, d\theta\cdots,$$

and where $\delta r(\theta,\phi)$ is given by Eq. (3.14). The square bracket terms in (3.19a) and (3.19b) are time-average quantities in accordance with the oscillatory time dependence, Eq. (2.1), of the electromagnetic fields. For a given plasma surface deformation, the proper linear combination (3.15) is constructed by selecting the coefficients $A(n,l)_{\mu m}$ such that the matrix (3.19a) or (3.19b) is diagonal in the $(2l+1)$-dimensional subspace associated with a transverse electric or a transverse magnetic (n,l)-mode. If we denote a perturbation $\delta\omega$ of the characteristic frequency ω_{nl} by $\omega^{(1)}_{nl\mu}$, we have for the characteristic frequency $\omega_{nl\mu}$ of a perturbed (n,l)-mode the expression

$$\omega_{nl\mu} = \omega_{nl} + \omega^{(1)}_{nl\mu}, \tag{3.20}$$

which is exact to first order. We then replace (3.2) by

$$\delta_{\mu'\mu}\omega^{(1)}_{nl\mu} = \omega_{nl}\,\delta\bar{U}_{nl\mu'\mu}/\bar{U}_{nl\mu}, \tag{3.21}$$

where

$$\bar{U}_{nl\mu} = \frac{1}{16\pi}\int_{V_{ca}} dV[|\mathcal{E}_{nl\mu}|^2 + |\mathcal{B}_{nl\mu}|^2]. \tag{3.22}$$

From (3.18) and (3.22) it is immediate that

$$\bar{U}_{nl\mu} = U_{nl}. \tag{3.23}$$

With the definitions

$$I(n,l)^{\mathrm{TE}}_{m'm} = -\frac{\omega^{\mathrm{TE}}_{nl}}{16\pi U^{\mathrm{TE}}_{nl}}\, r_0^2 \int d(\theta,\phi)[(B^*_{nlm'})_\theta(B_{nlm})_\theta + (B^*_{nlm'})_\phi(B_{nlm})_\phi]^{\mathrm{TE}}_{r=r_0}\,\delta r(\theta,\phi) \tag{3.24a}$$

or

$$I(n,l)^{\mathrm{TM}}_{m'm} = \frac{\omega^{\mathrm{TM}}_{nl}}{16\pi U^{\mathrm{TM}}_{nl}}\, r_0^2 \int d(\theta,\phi)[(E^*_{nlm'})_r(E_{nlm})_r - (B^*_{nlm'})^*_\theta \times (B_{nlm})_\theta - (B^*_{nlm'})_\phi(B_{nlm})_\phi]^{\mathrm{TM}}_{r=r_0}\,\delta r(\theta,\phi) \tag{3.24b}$$

we obtain from (3.19a) or (3.19b), and (3.21)

$$\sum_{m'm} A(n,l)^*_{\mu'm'}A(n,l)_{\mu m}I(n,l)_{m'm} = \delta_{\mu'\mu}\omega^{(1)}_{nl\mu}. \tag{3.25}$$

If we multiply Eq. (3.25) by $A(n,l)_{\mu'\bar{m}}$, we find after invoking the unitarity relations (3.16) the set of equations

$$\sum_{m=-l}^{l}[I(n,l)_{\bar{m}m} - \omega^{(1)}_{nl\mu}\delta_{\bar{m}m}]A(n,l)_{\mu m} = 0. \tag{3.26}$$

This system of homogeneous equations has nontrivial solutions if and only if its characteristic determinant vanishes,

$$\det[I(n,l)_{m'm} - \omega^{(1)}_{nl\mu}\delta_{m'm}] = 0. \tag{3.27}$$

Since the rank of the characteristic determinant is $2l+1$, the equation (3.27) is an algebraic equation of degree $2l+1$ in the frequency perturbation $\omega_{nl\mu}^{(1)}$. There are then, in general, $2l+1$ different real roots for this equation. These roots are the values of $\omega_{nl\mu}^{(1)}$ to be substituted in (3.26) in order to determine the coefficients $A(n,l)_{\mu m}$. If all $2l+1$ roots $\omega_{nl\mu}^{(1)}$ are different, the azimuthal degeneracy is completely removed.

We now evaluate the integrals in (3.24) by utilizing some of the results derived in Appendices B and C. In accordance with (B15) and (3.14) the integrands in (3.24a) and (3.24b) may be written as

$$[\mathbf{B}_{nlm'}^{*(t)}\cdot\mathbf{B}_{nlm}^{(t)}]_{r=r_0}^{\mathrm{TE}}\delta r$$
$$=\frac{l(l+1)}{(2l+1)^3}[lg_{n,l+1}^{\mathrm{TE}}(r_0)-(l+1)g_{n,l-1}^{\mathrm{TE}}(r_0)]^2$$
$$\times[l\mathbf{Y}_{l,l+1}^{*m'}\cdot\mathbf{Y}_{l,l+1}^{m}+\sqrt{l(l+1)}(\mathbf{Y}_{l,l+1}^{*m'}\cdot\mathbf{Y}_{l,l-1}^{m}+\mathbf{Y}_{l,l-1}^{*m'}\cdot\mathbf{Y}_{l,l+1}^{m})$$
$$+(l+1)\mathbf{Y}_{l,l-1}^{*m'}\cdot\mathbf{Y}_{l,l-1}^{m}]r_0\sum_{s,\sigma}\beta_{s\sigma}Y_{s\sigma}, \qquad (3.28a)$$

and

$$[\mathbf{E}_{nlm'}^{*(r)}\cdot\mathbf{E}_{nlm}^{(r)}-\mathbf{B}_{nlm'}^{*(t)}\cdot\mathbf{B}_{nlm}^{(t)}]_{r=r_0}^{\mathrm{TM}}\delta r$$
$$=(k_{nl}^{\mathrm{TM}}r_0)^{-2}[g_{nl}^{\mathrm{TM}}(r_0)]^2l(l+1)\Big\{\frac{l(l+1)}{2l+1}[(l+1)\mathbf{Y}_{l,l+1}^{*m'}\cdot\mathbf{Y}_{l,l+1}^{m}$$
$$-\sqrt{l(l+1)}(\mathbf{Y}_{l,l+1}^{*m'}\cdot\mathbf{Y}_{l,l-1}^{m}+\mathbf{Y}_{l,l-1}^{*m'}\cdot\mathbf{Y}_{l,l+1}^{m})+l\mathbf{Y}_{l,l-1}^{*m'}\cdot\mathbf{Y}_{l,l-1}^{m}]$$
$$-(k_{nl}^{\mathrm{TM}}r_0)^2\mathbf{Y}_{ll}^{*m'}\cdot\mathbf{Y}_{ll}^{m}\Big\}r_0\sum_{s,\sigma}\beta_{s\sigma}Y_{s\sigma}, \qquad (3.28b)$$

respectively. The vector spherical harmonics $\mathbf{Y}_{l,j}^{m}$ ($j=l,l\pm1$) are defined by (B1). The radial functions $g_{n,l\pm1}^{\mathrm{TE}}(r)$ and $g_{n,l}^{\mathrm{TM}}(r)$ are in accordance with (2.13b) and (2.18b) given by

$$g_{n,l\pm1}^{\mathrm{TE}}(r)=C_{nl}^{\mathrm{TE}}[j_{l\pm1}(k_{nl}^{\mathrm{TE}}r)+\alpha_{nl}^{\mathrm{TE}}n_{l\pm1}(k_{nl}^{\mathrm{TE}}r)] \qquad (3.29a)$$

and

$$g_{nl}^{\mathrm{TM}}(r)=C_{nl}^{\mathrm{TM}}[j_l(k_{nl}^{\mathrm{TM}}r)+\alpha_{nl}^{\mathrm{TM}}n_l(k_{nl}^{\mathrm{TM}}r)]. \qquad (3.29b)$$

By virtue of formula (C6) we find, after combining (3.28a), (3.29a), (A9), and (A10), for the expression (3.24a)

$$I(n,l)_{m'm}^{\mathrm{TE}}=-r_0^3\omega_{nl}^{\mathrm{TE}}(16\pi U_{nl}^{\mathrm{TE}})^{-1}\sum_{s\sigma}\beta_{s\sigma}\int d(\theta,\phi)$$
$$\times[\mathbf{B}_{nlm'}^{*(t)}\cdot\mathbf{B}_{nlm}^{(t)}]_{r_0}^{\mathrm{TE}}Y_{s\sigma}$$
$$=(-)^{m-1}\frac{1}{\sqrt{4\pi}}\omega_{nl}^{\mathrm{TE}}[1-(r_0/R_0)(n_l(k_{nl}^{\mathrm{TE}}r_0)/n_l(k_{nl}^{\mathrm{TE}}R_0))^2]^{-1}$$
$$\times[l\delta_{j',l+1}\delta_{j,l+1}+2\sqrt{l(l+1)}\delta_{j',l+1}\delta_{j,l-1}$$
$$+(l+1)\delta_{j',l-1}\delta_{j,l-1}]\sum_s I(jj'ls;\beta)_{m'm}, \qquad (3.30a)$$

and, after combining (3.28b), (3.29b), (A9), (A11), and (A12), for the expression (3.24b)

$$I(n,l)_{m'm}^{\mathrm{TM}}=r_0^3\omega_{nl}^{\mathrm{TM}}(16\pi U_{nl}^{\mathrm{TM}})^{-1}\sum_{s\sigma}\beta_{s\sigma}\int d(\theta,\phi)[\mathbf{E}_{nlm'}^{*(r)}\cdot\mathbf{E}_{nlm}^{(r)}$$
$$-\mathbf{B}_{nlm'}^{*(t)}\cdot\mathbf{B}_{nlm}^{(t)}]_{(r=r_0)}^{\mathrm{TM}}Y_{s\sigma}$$
$$=(-)^m\frac{1}{\sqrt{4\pi}}\omega_{nl}^{\mathrm{TM}}\Big\{[(k_{nl}^{\mathrm{TM}}r_0)^2-l(l+1)]$$
$$-\left(\frac{r_0}{R_0}\right)^3\left[(k_{nl}^{\mathrm{TM}}R_0)^2-l(l+1)\right]$$
$$\times\left[\frac{(l+1)n_{l-1}(k_{nl}^{\mathrm{TM}}r_0)-ln_{l+1}(k_{nl}^{\mathrm{TM}}r_0)}{(l+1)n_{l-1}(k_{nl}^{\mathrm{TM}}R_0)-ln_{l+1}(k_{nl}^{\mathrm{TM}}R_0)}\right]^2\Big\}^{-1}$$
$$\times\{l(l+1)[(l+1)\delta_{j',l+1}\delta_{j,l+1}-2\sqrt{l(l+1)}$$
$$\times\delta_{j',l+1}\delta_{j,l-1}+l\delta_{j',l-1}\delta_{j,l-1}]-(2l+1)$$
$$\times(k_{nl}^{\mathrm{TM}}r_0)^2\delta_{j'l}\delta_{jl}\}\sum_s I(jj'ls;\beta)_{m'm}, \qquad (3.30b)$$

where

$$I(jj'ls;\beta)_{m'm}=\left(\frac{(2j'+1)(2j+1)}{2s+1}\right)^{1/2}\beta_{s,m'-m}W(ljlj';1s)$$
$$\times\langle j0j'0|s0\rangle\langle l,-m,lm'|s,m'-m\rangle. \qquad (3.30c)$$

In deriving the results (3.30) we have used the symmetry properties

$$\langle a\alpha b\beta|c\gamma\rangle=(-)^{a+b-c}\langle b\beta a\alpha|c\gamma\rangle, \qquad (3.31a)$$
$$\langle a\alpha b\beta|c\gamma\rangle=(-)^{a+b-c}\langle a,-\alpha,b,-\beta|c,-\gamma\rangle, \qquad (3.31b)$$

of the Clebsch–Gordan (C–G) coefficients, as well as the symmetry property

$$W(abcd;ef)=W(cdab;ef)$$

of the Racah W coefficients. From (3.31b) it is immediate that the presence in (3.30c) of the C–G coefficient $\langle j0j'0|s0\rangle$ entails that $j+j'-s\equiv0(\mathrm{mod}\,2)$. Since either $j,j'=l\pm1$ or $j,j'=l$, it is then obvious that only even values of the multipole index s contribute to (3.30a) and (3.30b). Furthermore, from the presence of the C–G coefficient $\langle l,-m,lm'|s\sigma\rangle$ it follows that $|m'-m|\le s\le 2l$. Therefore, only those deformation multipoles $(s\sigma)$ (characterized by the surface parameters $\beta_{s\sigma}$) contribute to (3.30a) and (3.30b) for which

$$\sigma=m'-m, \qquad (3.32a)$$
$$|m'-m|\le s\le 2l, \qquad (3.32b)$$
$$s=0,2,4,\cdots. \qquad (3.32c)$$

Since $s=0$ implies that $m'=m$, the azimuthal degeneracy of a transverse electric or a transverse magnetic (n,l)-mode can be removed (to first order) only by a quadrupole ($s=2$), a hexadecupole ($s=4$), or higher even (deformation) multipoles. The additional condition

$$|j'-j|\le s\le j'+j, \qquad (3.32d)$$

follows from the presence in (3.30c) of the C–G coefficient $\langle j0j'0|s0\rangle$ or of the Racah coefficient $W(ljlj';1s)$. If $s>2l-2$ the condition (3.32d) implies that in (3.30a) and in (3.30b) those terms vanish which are multiplied by $\delta_{j',l-1}\delta_{j,l-1}$. [If $s<2$, it follows from (3.32d) that in (3.30a) and (3.30b) there are no terms containing the factor $\delta_{j',l+1}\delta_{j,l-1}$.] The relation (3.13) for the deformation parameters and the symmetry properties (3.31) for the C–G coefficients ensure that the perturbation matrices (3.30a) and (3.30b) are Hermitian and, thus, that their eigenvalues $\omega_{nl\mu}^{(1)}$ are real. If the rank of the characteristic determinant in (3.27) is $(2l+1-\lambda)$, the multiplicity of the eigenfrequency perturbation $\omega_{nl\mu}^{(1)}$ is λ. This connection between the rank of the characteristic determinant and the multiplicity of the associated eigenvalue is true only for the so-called simple matrics, of which the Hermitian matrices constitute a subclass. Thus, the perturbed eigenfrequencies (3.20) are nondegenerate only if the characteristic determinant in (3.27) has rank $2l$.

The sets of the expansion coefficients in (3.15) constitute the eigenvectors

$$A(n,l)_\mu=\{A(n,l)_{\mu m}:-l\le m\le l\} \qquad (3.33)$$

of the perturbation matrix (3.30a) or (3.30b). Since the matrices (3.30) are Hermitian, their eigenvectors

(3.33) which are associated with distinct eigenvalues $\omega_{nl\mu}^{(1)}$ are orthogonal. This orthogonality property is expressed by the first equation in (3.16). Since for a simple matrix the maximal number of linearly independent eigenvectors associated with the same eigenvalue coincides with the multiplicity of this eigenvalue, it is possible to orthonormalize these eigenvectors and thereby fulfill this orthogonality relation even if the azimuthal degeneracy of a given eigenmode is not completely removed by the deformation of the plasma boundary.

APPENDIX A

The normalization of the characteristic multipole fields (2.13) and (2.18) to unit energy density requires the evaluation of the integrals in the expression

$$U_{nlm} = \frac{1}{16\pi} \int_{V_{ca}} dV[|\mathbf{E}_{nlm}|^2 + |\mathbf{B}_{nlm}|^2], \tag{A1}$$

which represents the time-average of the electromagnetic energy of the cavity resonator volume V_{ca}. The integrations in (A1) are simplified if we take into account that the time-average values of the electric and magnetic field energies in a cavity bounded by perfectly conducting walls (of arbitrary shape) are equal (Ref. 9, p. 291),

$$\int_{V_{ca}} dV(|\mathbf{E}|^2 + |\mathbf{B}|^2) = 2\int_{V_{ca}} dV\,|\mathbf{E}|^2 = 2\int_{V_{ca}} dV\,|\mathbf{B}|^2. \tag{A2}$$

With (A2) and (2.13) or (2.18) we obtain for (A1),

$$U_{nlm} = \frac{1}{8\pi} C_{nl}^2 \int_{r_0}^{R_0} r^2\,dr[j_l(k_{nl}r) + \alpha_{nl} n_l(k_{nl}r)]^2 \int_0^{2\pi} d\phi \times \int_0^{\pi} \sin\theta\,d\theta \left\{\frac{m^2}{\sin^2\theta}\,|Y_{lm}(\theta,\phi)|^2 + \left|\frac{\partial}{\partial\theta} Y_{lm}(\theta,\phi)\right|^2\right\}. \tag{A3}$$

For the evaluation of the these integrals, a few properties of the spherical harmonics, the Legendre functions of the first kind, and the spherical Bessel functions are needed.

The spherical harmonics are defined for $l = 0, 1, 2, \cdots$ and $-l \le m \le l$ by the relation

$$Y_{lm}(\theta,\phi) = \left[\frac{2l+1}{4\pi}\,\frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m(\cos\theta)\exp(im\phi). \tag{A4}$$

By means of the Legendre differential equation and the orthogonality relations for the Legendre functions $P_l^m(\cos\theta)$, which are stated on pp. 198–9 of Ref. 11, it can be shown that

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\,d\theta \left[\frac{\partial}{\partial\theta} Y_{lm'}(\theta,\phi)^* \frac{\partial}{\partial\theta} Y_{lm}(\theta,\phi) + \frac{m'm}{\sin^2\theta} Y_{lm'}(\theta,\phi)^* Y_{lm}(\theta,\phi)\right] = \delta_{m'm} l(l+1). \tag{A5}$$

For the spherical Bessel functions of the first and the second kind,

$$j_l(x) = \left(\frac{\pi}{2x}\right)^{1/2} J_{l+1/2}(x), \quad n_l(x) = \left(\frac{\pi}{2x}\right)^{1/2} N_{l+1/2}(x), \tag{A6}$$

we have the relations

$$\int dr\,r^2 \jmath_l^2(kr) = \tfrac{1}{2}r^3[\jmath_l^2(kr) - \jmath_{l-1}(kr)\jmath_{l+1}(kr)] \tag{A7}$$

where

$$\jmath_l(kr) = \begin{Bmatrix} j_l(kr) \\ n_l(kr) \end{Bmatrix},$$

and

$$\int dr\,r^2 j_l(kr) n_l(kr) = \tfrac{1}{4}r^3[2j_l(kr)n_l(kr) - j_{l-1}(kr)n_{l+1}(kr) - j_{l+1}(kr)n_{l-1}(kr)]. \tag{A8}$$

The Eqs. (A7) and (A8) are immediate consequences of the first formula given on p. 88 of Ref. 11. By virtue of (A3), (A5), (A7), (A8), (2.13c), and the cross product relation

$$j_l(z)n_{l-1}(z) - j_{l-1}(z)n_l(z) = z^{-2}, \tag{A9}$$

stated on p. 439 of Ref. 12, we find for the time-average electromagnetic field energy of a TE_{nlm} mode $(-l \le m \le l)$ in the spherical cavity $r_0 \le r \le R_0$,

$$U_{nl}^{\mathrm{TE}} = (C_{nl}^{\mathrm{TE}})^2 \frac{l(l+1)}{16\pi} (k_{nl}^{\mathrm{TE}})^{-4}\{[R_0 n_l^2(k_{nl}^{\mathrm{TE}}R_0)]^{-1} - [r_0 n_l^2(k_{nl}^{\mathrm{TE}}r_0)]^{-1}\}. \tag{A10}$$

The corresponding expression for a TM_{nlm} mode is

$$U_{nl}^{\mathrm{TM}} = (C_{nl}^{\mathrm{TM}})^2 \frac{l(l+1)}{16\pi} (k_{nl}^{\mathrm{TM}})^{-4}\{[(k_{nl}^{\mathrm{TM}}R_0)^2 - l(l+1)]R_0^{-1} \times [(k_{nl}^{\mathrm{TM}}R_0)n_{l-1}(k_{nl}^{\mathrm{TM}}R_0) - l n_l(k_{nl}^{\mathrm{TM}}R_0)]^{-2} - [(k_{nl}^{\mathrm{TM}}r_0)^2 - l(l+1)]r_0^{-1} \times [(k_{nl}^{\mathrm{TM}}r_0)n_{l-1}(k_{nl}^{\mathrm{TM}}r_0) - l n_l(k_{nl}^{\mathrm{TM}}r_0)]^{-2}\}. \tag{A11}$$

Equation (A11) can be derived from (A3), (A5), (A7), (A8), (2.18c), (A9), and the recursion formula for spherical Bessel functions (Ref. 12, p. 439)

$$\frac{2l+1}{z}\,\jmath_l(z) = \jmath_{l-1}(z) + \jmath_{l+1}(z), \quad \jmath_l(z) = \begin{Bmatrix} j_l(z) \\ n_l(z) \end{Bmatrix}. \tag{A12}$$

In Secs. 2 and 3, and in Appendix C, we refer to the relations

$$\frac{d}{dr}[r\,\jmath_l(kr)] = kr\,\jmath_{l-1}(kr) - l\,\jmath_l(kr) = -kr\,\jmath_{l+1}(kr) + (l+1)\jmath_l(kr), \tag{A13}$$

which follow immediately from the relations

$$\frac{l+1}{z}\,\jmath_l(z) + \frac{d}{dz}\,\jmath_l(z) = \jmath_{l-1}(z), \quad \frac{l}{z}\,\jmath_l(z) - \frac{d}{dz}\,\jmath_l(z) = \jmath_{l+1}(z), \tag{A14}$$

which are stated on p. 439 of Ref. 12.

APPENDIX B

Solid angle integrations involving scalar as well as vector products of multipole field vectors are substantially simplified by expressing the fields in terms of irreducible tensors whose components, known as vector spherical harmonics, are[13,14]

$$\mathbf{Y}^m_{lj}(\theta,\phi)=\sum_{\mu\tau}\langle j\mu 1\tau|lm\rangle Y_{j\mu}(\theta,\phi)\boldsymbol{\epsilon}_\tau. \tag{B1}$$

Here, $Y_{j\mu}$ denotes a spherical harmonic, $\langle j\mu 1\tau|lm\rangle$ a Clebsch–Gordan coefficient,[13,14] and $\boldsymbol{\epsilon}_\tau$ a spherical unit vector, defined in terms of the Cartesian unit vectors $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ by the relations

$$\boldsymbol{\epsilon}_\tau=\begin{cases}-\dfrac{1}{\sqrt{2}}(\mathbf{e}_x+i\mathbf{e}_y), & \tau=1,\\ \mathbf{e}_z, & \tau=0,\\ \dfrac{1}{\sqrt{2}}(\mathbf{e}_x-i\mathbf{e}_y), & \tau=-1.\end{cases} \tag{B2}$$

The definition (1) implies that there are three linearly independent irreducible tensors of rank l; they are associated with $j=l$, $j=l+1$, and $j=l-1$.

From the values of the Clebsch–Gordan coefficients $\langle j\mu 1\tau|lm\rangle$ (given on p. 76 of Ref. 15) it is immediate that the well-known relations

$$L_\pm Y_{lm}=\sqrt{(l\mp m)(l\pm m+1)}\,Y_{l,m\pm1},$$

$$L_z Y_{lm}=mY_{lm},$$

where

$$L_\pm=\exp(\pm i\phi)\left(\pm\frac{\partial}{\partial\theta}+i\cot\theta\frac{\partial}{\partial\phi}\right),$$

$$L_z=-i\frac{\partial}{\partial\phi},$$

may be written as

$$L_\tau Y_{lm}=(-)^\tau\sqrt{l(l+1)}\langle l,m+\tau,1,-\tau|lm\rangle Y_{l,m+\tau}, \tag{B3}$$

where

$$L_\tau=\mathbf{L}\cdot\boldsymbol{\epsilon}_\tau=\begin{cases}-\dfrac{1}{\sqrt{2}}L_+, & \tau=1,\\ L_z & \tau=0,\\ \dfrac{1}{\sqrt{2}}L_-, & \tau=-1,\end{cases}$$

are the spherical components of the angular momentum vector (operator)

$$\mathbf{L}=\frac{1}{i}\mathbf{r}\times\boldsymbol{\nabla}. \tag{B4}$$

From (B1), (B3), and (B4) then follows the relation

$$(\mathbf{r}\times\boldsymbol{\nabla})Y_{lm}(\theta,\phi)=i\sqrt{l(l+1)}\mathbf{Y}^m_{l,l}(\theta,\phi). \tag{B5}$$

From the identity

$$\boldsymbol{\nabla}\times(\mathbf{r}\times\boldsymbol{\nabla})=\mathbf{r}\nabla^2-\boldsymbol{\nabla}\frac{\partial}{\partial r}r$$

and from the Helmholtz equation

$$(\nabla^2+k^2)\,j_l(kr)Y_{lm}(\theta,\phi)=0,$$

where $j_l(kr)$ denotes a spherical Bessel function, it follows that

$$\boldsymbol{\nabla}\times(\mathbf{r}\times\boldsymbol{\nabla}j_lY_{lm})=-\left[k^2\mathbf{r}j_lY_{lm}+\boldsymbol{\nabla}\left(\frac{\partial}{\partial r}rj_l\right)Y_{lm}\right]. \tag{B6}$$

In order to express the first term on the right-hand side of Eq. (B6) in terms of vector spherical harmonics we first represent the position vector $\mathbf{r}$ as the contraction of the two first-rank spherical tensors $rY_{1\tau}$ and $\boldsymbol{\epsilon}_\tau$,[14]

$$\mathbf{r}=r\left(\frac{4\pi}{3}\right)^{1/2}\sum_\tau Y^*_{1\tau}(\theta,\phi)\boldsymbol{\epsilon}_\tau. \tag{B7}$$

Next, we decompose the product of the spherical harmonics $Y^*_{1\tau}$ and Y_{lm} according to the reduction formula

$$Y^*_{1\tau}Y_{lm}=-\left(\frac{3}{4\pi}\right)^{1/2}\left[\left(\frac{l+1}{2l+1}\right)^{1/2}\langle l+1,m-\tau,1\tau|lm\rangle Y_{l+1,m-\tau}-\left(\frac{l}{2l+1}\right)^{1/2}\langle l-1,m-\tau,1\tau|lm\rangle Y_{l-1,m-\tau}\right] \tag{B8}$$

which follows from the Clebsch–Gordan decomposition[13,14]

$$Y^*_{1\tau}Y_{lm}=(-)^\tau Y_{1,-\tau}Y_{lm}$$
$$=(-)^\tau\left(\frac{3}{4\pi}\right)^{1/2}\sum_j\left[\frac{2l+1}{2j+1}\right]^{1/2}\langle 10l0|j0\rangle\times\langle 1,-\tau,lm|j,m-\tau\rangle Y_{j,m-\tau}$$

in virtue of the symmetry relations (3.31a), (3.31b) and

$$\langle a\alpha b\beta|c\gamma\rangle=(-)^{b+\beta}\left(\frac{2c+1}{2a+1}\right)^{1/2}\langle c,-\gamma b\beta|a,-\alpha\rangle. \tag{B9}$$

In view of (B7), (B8), and (B1) we may then write

$$\mathbf{r}Y_{lm}=r\left(\frac{4\pi}{3}\right)^{1/2}\sum_\tau Y^*_{1\tau}Y_{lm}\boldsymbol{\epsilon}_\tau$$
$$=-r\left[\left(\frac{l+1}{2l+1}\right)^{1/2}\mathbf{Y}^m_{l,l+1}-\left(\frac{l}{2l+1}\right)^{1/2}\mathbf{Y}^m_{l,l-1}\right]. \tag{B10}$$

To the second term on the right-hand side of (B6) we apply the gradient formula [Eq. (2.58) of Ref. 13 or the second equation on p. 150 of Ref. 14],

$$\boldsymbol{\nabla}f(r)Y_{lm}(\theta,\phi)=-\left[\left(\frac{l+1}{2l+1}\right)^{1/2}\left(\frac{df}{dr}-\frac{l}{r}f\right)\mathbf{Y}^m_{l,l+1}-\left(\frac{l}{2l+1}\right)^{1/2}\left(\frac{df}{dr}+\frac{l+1}{r}f\right)\mathbf{Y}^m_{l,l-1}\right]. \tag{B11}$$

With $f(r)=(d/dr)[rj_l(kr)]$ and the differential equation

$$\frac{1}{r^2}\frac{d}{dr}\left[r^2\frac{d}{dr}j_l(kr)\right]+\left[k^2-\frac{l(l+1)}{r^2}\right]j_l(kr)=0,$$

we obtain from (B11)

$$\boldsymbol{\nabla}\left(\frac{d}{dr}rj_l\right)Y_{lm}=\left(\frac{l+1}{2l+1}\right)^{1/2}\left[rk^2j_l+l\left(\frac{d}{dr}-\frac{l}{r}\right)j_l\right]\mathbf{Y}^m_{l,l+1}+\left(\frac{l}{2l+1}\right)^{1/2}\left[-rk^2j_l+(l+1)\times\left(\frac{d}{dr}+\frac{l+1}{r}\right)j_l\right]\mathbf{Y}^m_{l,l-1}. \tag{B12}$$

By combining (B6), (B10), and (B12), we arrive at the equation

$$\mathrm{curl}(\mathbf{r}\times\nabla j_l Y_{lm}) = -\left[\left(\frac{l+1}{2l+1}\right)^{1/2} l\left(\frac{d}{dr}-\frac{l}{r}\right) j_l \mathbf{Y}^m_{l,l+1} + \left(\frac{l}{2l+1}\right)^{1/2}(l+1)\left(\frac{d}{dr}+\frac{l+1}{r}\right) j_l \mathbf{Y}^m_{l,l-1}\right. \tag{B13}$$

On account of the relations (A14) we may replace (B13) by

$$\mathrm{curl}\, j_l(kr)(\mathbf{r}\times\nabla)Y_{lm}(\theta,\phi) = \frac{k}{\sqrt{2l+1}}[l\sqrt{l+1}\, j_{l+1}(kr)\mathbf{Y}^m_{l,l+1}(\theta,\phi) - (l+1)\sqrt{l}\, j_{l-1}(kr)\mathbf{Y}^m_{l,l-1}(\theta,\phi)]. \tag{B14}$$

It is obvious that the vector given by (B5) is tangent to a sphere. On the other hand, the vector given by (B14) has a nonvanishing radial part.

Since

$$\nabla\times(\mathbf{r}\times\nabla)\, j_l Y_{lm} = j_l \mathbf{r}\nabla^2 Y_{lm} - \frac{d}{dr}(r j_l)\nabla Y_{lm},$$

we have for the radial part

$$[\nabla\times(\mathbf{r}\times\nabla)\, j_l(kr)Y_{lm}(\theta,\phi)]^{(r)} = -\, j_l(kr)\frac{l(l+1)}{r^2}\mathbf{r}Y_{lm}(\theta,\phi) = \frac{1}{r}\, j_l(kr)\frac{l(l+1)}{\sqrt{2l+1}}[\sqrt{l+1}\mathbf{Y}^m_{l,l+1}(\theta,\phi) - \sqrt{l}\,\mathbf{Y}^m_{l,l-1}(\theta,\phi)] \tag{B15a}$$

and for the tangential part

$$[\nabla\times(\mathbf{r}\times\nabla)\, j_l(kr)Y_{lm}(\theta,\phi)]^{(t)} = -\frac{d}{dr}[r\, j_l(kr)\nabla Y_{lm}(\theta,\phi)] = \frac{k\sqrt{l(l+1)}}{(2l+1)^{3/2}}[l\, j_{l+1}(kr) - (l+1)\, j_{l-1}(kr)] \times[\sqrt{l}\,\mathbf{Y}^m_{l,l+1}(\theta,\phi) + \sqrt{l+1}\,\mathbf{Y}^m_{l,l-1}(\theta,\phi)]. \tag{B15b}$$

The second equation in (B15a) follows immediately from (B10) and the second equation in (B15b) follows from (A13) and the gradient formula (B11).

APPENDIX C

By virtue of the formalism exhibited in the preceding appendix, the solid angle integrations in (3.24) are reduced to the evaluation of an integral of the type

$$I[(l'm',j')^*;lm,j;s\sigma] = \int d(\theta,\phi)\mathbf{Y}^{m'}_{l'j'}(\theta,\phi)^*\cdot\mathbf{Y}^m_{lj}(\theta,\phi)Y_{s\sigma}(\theta,\phi), \tag{C1}$$

where $d(\theta,\phi)$ denotes the solid angle element. The definitions (B1) and (B2) imply that

$$I[(l'm',j')^*;lm,j;s\sigma] = \sum_\tau \langle j',m'-\tau,1\tau|l'm'\rangle\langle j,m-\tau,1\tau|lm\rangle \times\int d(\theta,\phi)Y_{j',m'-\tau}(\theta,\phi)^*Y_{j,m-\tau}(\theta,\phi)Y_{s\sigma}(\theta,\phi). \tag{C2}$$

By applying the reduction formula

$$Y^*_{j',m'-\tau}Y_{j,m-\tau} = (-)^{m-\tau}\sum_{\bar{j}}\left[\frac{(2j'+1)(2j+1)}{4\pi(2\bar{j}+1)}\right]^{1/2}\times\langle j'0j0|\bar{j}0\rangle \langle j',-m'+\tau,j,m-\tau|\bar{j},-m'+m\rangle Y^*_{\bar{j},m'-m}$$

together with the orthonormality relation

$$\int d(\theta,\phi)Y_{\bar{j},m'-m}(\theta,\phi)^*Y_{s\sigma}(\theta,\phi) = \delta_{\bar{j}s}\delta_{m'-m,\sigma}$$

we obtain for (C2)

$$I[(l'm',j')^*;lm,j;s\sigma] = (-)^m\left[\frac{(2j'+1)(2j+1)}{4\pi(2s+1)}\right]^{1/2}\langle j'0j0|s0\rangle S, \tag{C3a}$$

where

$$S \equiv \sum_\tau(-)^\tau\langle j',m'-\tau,1\tau|l'm'\rangle\langle j,m-\tau,1\tau|lm\rangle \times\langle j',-m'+\tau,j,m-\tau|s,m-m'\rangle. \tag{C3b}$$

In this expression the sum over the products of three Clebsch–Gordan (C–G) coefficients can be evaluated by applying the Racah recoupling transformation.[13,14] An immediate consequence of this transformation is the formula

$$\sum_{m_1m_2}\langle j_1m_1j_2m_2|j_{12}m_{12}\rangle\langle j_{12}m_{12}j_3m_3|jm\rangle\langle j_2m_2j_3m_3|j_{23}m_{23}\rangle = \langle j_1m_1j_{23}m_{23}|jm\rangle\sqrt{(2j_{12}+1)(2j_{23}+1)}W(j_1j_2jj_3;j_{12}j_{23}), \tag{C4}$$

which enables us to express the sum (C3b) in terms of Racah W coefficients. In order to bring this sum to the form (C4) we interchange in the first C–G coefficient the angular momentum indices together with their projection numbers $(j',m'-\tau)$ and (l',m') in accordance with the symmetry relation (B9); we further interchange in the second C–G coefficient $(j,m-\tau)$ and $(1,\tau)$ in accordance with the symmetry relation (3.31a). The formula (C4) then yields for the sum (C3b) the expression

$$S = (-)^{j-l}\sqrt{(2l'+1)(2l+1)}W(l'1sj;j'l)\langle l',-m'lm|s,m-m'\rangle. \tag{C5}$$

By combining (C1), (C3a), and (C5) we find

$$\int d(\theta,\phi)\mathbf{Y}^{m'}_{l'j'}(\theta,\phi)^*\cdot\mathbf{Y}^m_{lj}(\theta,\phi)Y_{s\sigma}(\theta,\phi) = (-)^{m-l}\left[\frac{(2j'+1)(2j+1)(2l'+1)(2l+1)}{4\pi(2s+1)}\right]^{1/2} \times W(ljl'j';1s)\langle j0j'0|s0\rangle\langle l,-ml'm'|s\sigma\rangle, \tag{C6}$$

where we have used the symmetry relations (3.31) for the C–G coefficients as well as the symmetry property[13,14]

$$W(abcd;ef) = (-)^{e+f-b-c}W(fdae;bc)$$

for the Racah W coefficient.

*Work performed under the auspices of the U.S. Energy Research and Development Administration under contract No. W-7405-Eng-48.

[1] P. Ehrenfest, Ann. Phys. (Leipzig) **36**, 91 (1911); **51**, 327 (1916); Phil. Mag. 33, 500 (1917).
[2] F.E. Borgnis and C.H. Papas, *Encyclopedia of Physics* (Springer-Verlag, Berlin, 1958), Vol. XVI, p. 412.
[3] C.H. Papas, J. Appl. Phys. **25**, 1552 (1954).
[4] Yu.M. Aliev and E. Ferlengi, Zh. Eksp. Teor. Fiz. **57**, 1623 (1969) [Sov. Phys. JETP **30**, 877 (1970)].
[5] V.V. Yankov, Zh. Eksp. Teor. Fiz. **36**, 560 (1959) [Sov. Phys. JETP 9, 388 (1959)].
[6] J.W. Butler, *Proceedings of the Fourth Internation Conference on Ionization Phenomena in Gases* (North-Holland, Amsterdam, 1960), Vol. 2
[7] A.J. Hatch and J.W. Butler, J. Electron. Control 12, 89 (1962).
[8] J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed.
[9] L.D. Landau and E.M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960).
[10] A. Bohr and B.R. Mottelson, *Nuclear Structure* (Benjamin, New York, 1975), Vol. II.
[11] W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[12] M. Abramowitz and I.A. Stegun, *National Bureau of Standards Applied Mathematics Series No.* 55 (U.S. Government Printing Office, Washington, D.C., 1965).
[13] M.E. Rose, *Multipole Fields*, (Wiley, New York, 1955).
[14] D.M. Brink and G.R. Satchler, *Angular Momentum*, (Clarendon, Oxford, 1968), 2nd ed.
[15] E.U. Condon and G.H. Shortley, *The Theory of Atomic Spectra*, (Cambridge U.P., Cambridge, 1935).

An exact solution for diffraction of a line-source field by a half-plane*

J. Boersma

Department of Mathematics, Technological University, Eindhoven, The Netherlands

S. W. Lee

Department of Electrical Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801
(Received 22 March 1976)

This paper deals with the exact solution of a special electromagnetic diffraction problem, namely, diffraction of a line-source field by a half-plane. The line source is located on the surface of the half-plane, and radiates an E-polarized wave described by $u_n^i = H_n^{(1)}(kr_1)\sin n\phi_1$, where $n = 1,2,3,\cdots$, and (r_1, ϕ_1) are polar coordinates with the origin at the source point. A new, closed-form, exact solution for the total field on the shadow boundary is presented. This exact solution consists of n terms of order k^{-p}, where $p = 1,2,\ldots,n$. Its first two terms, which are of orders $k^{-1/2}$ and $k^{-3/2}$ relative to the incident field, agree with the asymptotic solution derived in a companion paper by the uniform asymptotic theory of edge diffraction.

1. INTRODUCTION

The diffraction of a line-source field by a half-plane was treated by ray techniques in Ref. 1, referred to hereafter as Part I. The diffraction problem considered there has been sketched in Fig. 1. A perfectly conducting half-plane at $x \leq 0$, $y = 0$ is illuminated by a cylindrical wave due to an (anisotropic) line source located at ($x = -d \times \cos\Omega$, $y = d\sin\Omega$). By using the uniform asymptotic theory of edge diffraction, an asymptotic solution for the total field up to and including terms of order $k^{-3/2}$ has been obtained in Part I. That solution was given in (I.2.3) and (I.2.4) for the general case, and in (I.3.6) and (I.4.4) for two special cases. (Equations from Part I are quoted by their numbers preceded by I.) As the uniform asymptotic theory is a formal asymptotic method based on an unproved ansatz, the solution obtained from it in Part I, of course, may or may not check with the asymptotic expansion of the exact solution for the problem under consideration. In the present paper, we will derive the exact solutions for some test cases, and show that they are in complete agreement with the solution obtained by the uniform asymptotic theory.

An arbitrary cylindrical wave emanated from a line source may be considered as a superposition of the multipole fields

$$u^i(r_1, \phi_1) = H_n^{(1)}(kr_1)\begin{bmatrix}\cos n\phi_1 \\ \sin n\phi_1\end{bmatrix}, \quad n = 0, 1, 2, \cdots, \qquad \begin{matrix}(1.1a)\\(1.1b)\end{matrix}$$

where (r_1, ϕ_1) are polar coordinates with the origin at the source point (Fig. 1). For the case $n = 0$ in (1.1a), i.e., when the line source is isotropic, exact solutions to the diffraction problem were first derived by Carslaw and Macdonald around the turn of the century. More easily accessible is the elegant solution due to Clemmow as described in Ref. 2, pp. 580–84. Clemmow's approach is first to decompose the Hankel function $H_0^{(1)}$ as an angular spectrum of plane waves. For each plane wave of the spectrum, the Sommerfeld half-plane solution applies and, then, the total field solution is expressed as a superposition integral with the Sommerfeld half-plane solution weighted by the spectrum of the incident field as its integrand. The same approach can also be applied in principle for the cases $n \neq 0$.[3] However, the superposition integrals in the latter cases become quite complex, and to our knowledge no explicit solution has been obtained.

Since our main purpose is to check the validity of the asymptotic solution given in Part I, we will not solve the diffraction problem of Fig. 1 in its full generality. Instead, our attention will be focused on a test case. In this test case, we assume (i) $u = E_z$ (E-polarized wave), (ii) $\Omega = 0+$ (line source on the upper surface of the half-plane), and (iii) $\phi = \pi$ (observation point on the shadow boundary). This case corresponds to Case A discussed in Part I, Sec. 3. The incident field will be given by (1.1). Thus, the solution to be derived should eventually be compared with (I.3.9) and (I.3.11).

Our method of solution consists of two main steps. In the first one (Secs. 2 and 3), for incident fields in (1.1) with $n = 1$ and 2, the total field solutions are obtained through differentiation of Clemmow's solution for the isotropic line source, and the enforcement of the edge condition. Guided by those results, we then derive in Sec. 4 a recurrence relation for the total field on the shadow boundary due to a general incident field with an index n in (1.1). The recurrence relation is subsequently solved by two methods in Secs. 4 and 5.

Several conventions used in this paper are stated below: (i) The time factor is $\exp(-i\omega t)$ and is suppressed. (ii) Unless explicitly mentioned otherwise in Sec. 2, only the case of E-polarization is considered and $u = E_z$. (iii) Three sets of polar coordinates are employed (Fig. 1): (r, ϕ) has its origin at the edge point ($x = 0$, $y = 0$); (r_1, ϕ_1) at the source point ($x = -d\cos\Omega$, $y = d\sin\Omega$); and (r_{-1}, ϕ_{-1}) at the image source point ($x = -d\cos\Omega$, $y = -d\sin\Omega$). All angles take values between 0 and 2π; ϕ, ϕ_{-1}, and Ω are measured clockwise, and ϕ_1 counterclockwise.

2. DIFFERENTIATION OF SOLUTIONS TO EDGE-DIFFRACTION PROBLEMS

In this section, we deduce a theorem on the differentiation of solutions to half-plane diffraction problems. In Sec. 3, this theorem will be applied to the diffraction

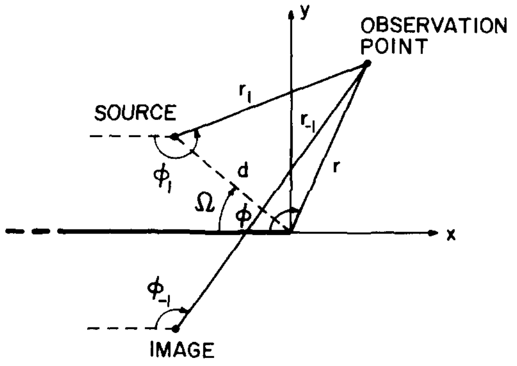

FIG. 1. A half-plane illuminated by a line source at $(x=-d\cos\Omega,\ y=d\sin\Omega)$.

problem of Fig. 1 in the cases of an incident field (1.1) with $n=1$ and 2. The solution to these problems will be obtained by differentiation of the known solution to the diffraction problem for the isotropic line source as presented below.

Referring to Fig. 1, we consider the diffraction of the cylindrical wave

$$u^i(r_1,\phi_1)=H_0^{(1)}(kr_1), \tag{2.1}$$

due to an isotropic line source, by the half-plane $x\le 0$, $y=0$. The two cases of E-polarization and H-polarization are treated simultaneously, and the resulting total field is denoted by $u_1=E_z$ in the case of E-polarization and by $u_2=H_z$ in the case of H-polarization. Then u_1, u_2 must satisfy the reduced wave equation $(\Delta+k^2)u_{1,2}=0$, the radiation condition at infinity, the boundary condition

$$u_1=0,\quad \frac{\partial u_2}{\partial y}=0 \quad \text{on the half-plane}, \tag{2.2}$$

and the edge condition (see Ref. 4, p. 45)

$$u_1(r,\phi)=O(r^{1/2}),\quad u_2(r,\phi)=O(1),\quad r\to 0. \tag{2.3}$$

According to Ref. 2 [Sec. 11.7, Eq. (20)] or Ref. 5 [Eqs. (8.46) and (8.68)], the total fields u_1, u_2 are given by the exact representation

$$u_{1,2}(r,\phi)=\frac{2}{\pi i}[J(k^{1/2}\xi, kr_1)\mp J(k^{1/2}\xi', kr_{-1})], \tag{2.4}$$

where the upper (lower) sign corresponds to u_1 (u_2). The function $J(x,y)$ in (2.4) is defined by

$$J(x,y)=\exp(iy)\int_{-x}^{\infty}\frac{\exp(i\mu^2)}{(\mu^2+2y)^{1/2}}d\mu, \tag{2.5}$$

and the detour parameter $\xi(\xi')$ of the incident (reflected) field is defined by

$$\xi=(r+d-r_1)^{1/2}\,\mathrm{sgn}[\cos\tfrac{1}{2}(\phi-\Omega)]$$
$$=\left(\frac{4rd}{r+d+r_1}\right)^{1/2}\cos\tfrac{1}{2}(\phi-\Omega), \tag{2.6a}$$

$$\xi'=(r+d-r_{-1})^{1/2}\,\mathrm{sgn}[\cos\tfrac{1}{2}(\phi+\Omega)]$$
$$=\left(\frac{4rd}{r+d+r_{-1}}\right)^{1/2}\cos\tfrac{1}{2}(\phi+\Omega). \tag{2.6b}$$

Near the edge $r=0$, the total fields in (2.4) behave as

$$u_1(r,\phi)=\frac{4}{\pi i}\left(\frac{r}{d}\right)^{1/2}\exp(ikd)\sin\tfrac{1}{2}\Omega\sin\tfrac{1}{2}\phi+O(r), \tag{2.7a}$$

$$u_2(r,\phi)=H_0^{(1)}(kd)+\frac{4}{\pi i}\left(\frac{r}{d}\right)^{1/2}\exp(ikd)\cos\tfrac{1}{2}\Omega\cos\tfrac{1}{2}\phi+O(r), \tag{2.7b}$$

which comply with the edge condition (2.3).

Now let us consider the diffraction of an E-polarized wave due to an anisotropic line source and given by

$$u^i(r_1,\phi_1)=\frac{\partial}{\partial x}H_0^{(1)}(kr_1). \tag{2.8}$$

Because $\partial u_1/\partial x$ satisfies the boundary condition on the half-plane $x\le 0$, $y=0$, it would appear that the total field for the present problem is simply given by $\partial u_1/\partial x$, where u_1 is given in (2.4). However, such a result is *incorrect* since $\partial u_1/\partial x$ does not satisfy the edge condition: generally $\partial u_1/\partial x=O(r^{-1/2})$ [compare with (2.7)], whereas the correct total field should be $O(r^{1/2})$ near the edge $r=0$. Therefore, $\partial u_1/\partial x$ must be supplemented with an additional term that should satisfy the reduced wave equation, the radiation condition, and the boundary condition on the half-plane, and should compensate the edge singularity of $\partial u_1/\partial x$. It is easily found that the additional term is a multiple of

$$H_{1/2}^{(1)}(kr)\sin\tfrac{1}{2}\phi=-i\left(\frac{2}{\pi k}\right)^{1/2}\frac{\exp(ikr)}{r^{1/2}}\sin\tfrac{1}{2}\phi. \tag{2.9}$$

The total field v_1 due to the incident field in (2.8) is now given by

$$v_1(r,\phi)=\frac{\partial u_1}{\partial x}+A_1\frac{\exp(ikr)}{r^{1/2}}\sin\tfrac{1}{2}\phi, \tag{2.10}$$

where the constant A_1 is to be determined by the requirement that at the edge $r=0$, the $r^{-1/2}$-singularities in the two terms in (2.10) should cancel. It can easily be shown that (2.10) satisfies all conditions for the present diffraction problem. Hence, by relying on uniqueness, (2.10) does represent the exact total field due to diffraction of the incident field in (2.8).

Next consider the diffraction of an H-polarized wave due to an anisotropic line source and given by

$$u^i(r_1,\phi_1)=\frac{\partial}{\partial y}H_0^{(1)}(kr_1). \tag{2.11}$$

By employing a similar argument as before, it is found that the total field w_2 in this case is given by

$$w_2(r,\phi)=\frac{\partial u_1}{\partial y}+B_2\frac{\exp(ikr)}{r^{1/2}}\cos\tfrac{1}{2}\phi, \tag{2.12}$$

where u_1 is given in (2.4). Because $u_1=0$ on the half-plane, the tangential total electrical field at $x\le 0$, $y=0$, which is proportional to

$$\frac{\partial w_2}{\partial y}=\frac{\partial^2 u_1}{\partial y^2}-B_2\frac{\exp(ikr)}{2r^{3/2}}\sin\tfrac{1}{2}\phi$$
$$=-\left[\frac{\partial^2 u_1}{\partial x^2}+k^2u_1+B_2\frac{\exp(ikr)}{2r^{3/2}}\sin\tfrac{1}{2}\phi\right],$$

also vanishes on the half-plane. The constant B_2 in (2.12) can be determined by enforcing the edge condition $w_2(r,\phi)=O(1)$ as $r\to 0$.

Guided by the two results in (2.10) and (2.12), we can state the following theorem for the differentiation of solutions to the half-plane diffraction problem sketched in Fig. 1.

Theorem: In the two-dimensional diffraction at a perfectly conducting half-plane $x \leq 0,\ y=0$, let $u_1 = E_z$ ($u_2 = H_z$) be the total field due to the incident E-polarized wave (H-polarized wave) u^i. In a similar notation, let v_1 (v_2) be the total field due to diffraction of $\partial u^i/\partial x$, and let w_1 (w_2) be the total field due to diffraction of $\partial u^i/\partial y$. Then

$$v_1 = \frac{\partial u_1}{\partial x} + A_1 \frac{\exp(ikr)}{r^{1/2}} \sin\tfrac{1}{2}\phi, \quad (E\text{-polarization}), \tag{2.13a}$$

$$v_2 = \frac{\partial u_2}{\partial x} + A_2 \frac{\exp(ikr)}{r^{1/2}} \cos\tfrac{1}{2}\phi, \quad (H\text{-polarization}), \tag{2.13b}$$

$$w_1 = \frac{\partial u_2}{\partial y} + B_1 \frac{\exp(ikr)}{r^{1/2}} \sin\tfrac{1}{2}\phi, \quad (E\text{-polarization}), \tag{2.14a}$$

$$w_2 = \frac{\partial u_1}{\partial y} + B_2 \frac{\exp(ikr)}{r^{1/2}} \cos\tfrac{1}{2}\phi, \quad (H\text{-polarization}), \tag{2.14b}$$

where the constants A_1, A_2, B_1, and B_2 are determined by the requirement that the total fields should be free from the $r^{-1/2}$-singularity at the edge $r=0$.

Three remarks are in order: (i) The above theorem is valid not only for the incident field u^i given in (2.1) but also for a general incident field as, e.g., in (1.1). [See the application in connection with (4.8).] (ii) The theorem can be extended to higher-order derivatives. In fact, one such example will be worked out in the next section. (iii) For the special case of plane wave incidence, the present theorem was established by Bouwkamp[6] in 1946. As Bouwkamp points out, a qualitative version of the theorem was already enunciated by Rayleigh in a paper of 1897.

3. DIFFRACTION OF LOWER-ORDER MULTIPOLE FIELDS DUE TO A LINE SOURCE

In this section, we consider the diffraction problem sketched in Fig. 1 when the incident field is an E-polarized wave given by (1.1b) with $n=1$ and 2. By use of the theorem in Sec. 2, we determine the total field solution and this solution is specialized for the case $\Omega = 0+$, that is, the line source is located on the upper surface of the half-plane. Finally, we derive a simple closed-form result for the total field on the shadow boundary $\phi = \pi$. The diffraction problem for an incident field (1.1) with general n will be discussed in Sec. 4.

First, consider the diffraction of the E-polarized wave as given in (1.1b) with $n=1$, viz.,

$$u^i(r_1, \phi_1) = H_1^{(1)}(kr_1)\sin\phi_1. \tag{3.1}$$

If one uses the relation

$$\frac{\partial}{\partial y} = -\sin\phi_1 \frac{\partial}{\partial r_1} - \cos\phi_1 \frac{1}{r_1}\frac{\partial}{\partial \phi_1}, \tag{3.2}$$

(3.1) may be rewritten as

$$u^i(r_1, \phi_1) = \frac{1}{k}\frac{\partial}{\partial y} H_0^{(1)}(kr_1). \tag{3.3}$$

According to the theorem in Sec. 2, the resulting total field u is found to be

$$u(r, \phi) = \frac{1}{k}\left(\frac{\partial u_2}{\partial y} + B_1 \frac{\exp(ikr)}{r^{1/2}} \sin\tfrac{1}{2}\phi\right), \tag{3.4}$$

where u_2 is given in (2.4). To determine B_1, the behavior of $\partial u_2/\partial y$ near the edge should be examined. With the help of (2.7b), we have

$$\frac{\partial u_2}{\partial y} = \frac{2}{\pi i}\frac{\exp(ikd)}{d^{1/2}} \cos\tfrac{1}{2}\Omega\ r^{-1/2}\sin\tfrac{1}{2}\phi + O(1), \quad r \to 0. \tag{3.5}$$

The edge condition requires that u in (3.4) must be free from the $r^{-1/2}$-singularity. In view of (3.5), this requirement is satisfied if B_1 assumes the value

$$B_1 = \frac{2i}{\pi}\frac{\exp(ikd)}{d^{1/2}}\cos\tfrac{1}{2}\Omega. \tag{3.6}$$

Thus, (3.4) and (3.6) give the exact total field (valid for all Ω and ϕ between 0 and 2π) due to diffraction of the incident field (3.1). Specializing the solution in (3.4) for the case $\Omega = 0+$, we have

$$u(r, \phi) = \frac{4}{i\pi k}\frac{\partial}{\partial y}\left(\exp(ikr_1)\int_{-k^{1/2}\xi}^{\infty} \frac{\exp(i\mu^2)}{(\mu^2 + 2kr_1)^{1/2}} d\mu\right) + \frac{2i}{\pi k}\frac{\exp[ik(r+d)]}{(rd)^{1/2}}\sin\tfrac{1}{2}\phi, \tag{3.7}$$

where $\xi = (r + d - r_1)^{1/2}\operatorname{sgn}(\cos\tfrac{1}{2}\phi)$. Along the shadow boundary $\phi = \pi$, it is easily shown that

$$\frac{\partial r_1}{\partial y} = 0, \quad \frac{\partial \xi}{\partial y} = \left(\frac{d}{2r(r+d)}\right)^{1/2} \quad \text{at } \phi = \pi. \tag{3.8}$$

Using (3.8) in (3.7), we obtain

$$u(r, \phi = \pi) = \exp[i(kr + kd + \tfrac{1}{2}\pi)]\frac{2r^{1/2}}{\pi k(r+d)d^{1/2}}. \tag{3.9}$$

This is the exact total field on the shadow boundary due to the incidence of (3.1) with $\Omega = 0+$. When (3.9) is compared with the asymptotic solution in (I.3.9), they coincide.

Next consider the diffraction of the E-polarized wave given by

$$u^i(r_1, \phi_1) = H_2^{(1)}(kr_1)\sin 2\phi_1. \tag{3.10}$$

If one uses (3.2) and the relation

$$\frac{\partial}{\partial x} = -\cos\phi_1\frac{\partial}{\partial r_1} + \sin\phi_1\frac{1}{r_1}\frac{\partial}{\partial \phi_1}, \tag{3.11}$$

(3.10) may be rewritten as

$$u^i(r_1, \phi_1) = \frac{2}{k^2}\frac{\partial^2}{\partial x \partial y} H_0^{(1)}(kr_1). \tag{3.12}$$

On extending the theorem in Section 2, it is found that the total field u in the present diffraction problem may be expressed as

$$u(r, \phi) = \frac{2}{k^2}\left[\frac{\partial^2 u_2}{\partial x \partial y} + A_3\frac{\exp(ikr)}{r^{1/2}}\sin\tfrac{1}{2}\phi + B_3\frac{\exp(ikr)}{r^{1/2}}\left(1 - \frac{1}{ikr}\right)\sin\tfrac{3}{2}\phi\right], \tag{3.13}$$

where u_2 is given by (2.4), and the constants A_3 and B_3 are to be determined by enforcing the edge condition. The second and third terms in (3.13) are multiples of

$$H_{1/2}^{(1)}(kr)\sin\tfrac{1}{2}\phi \quad \text{and} \quad H_{3/2}^{(1)}(kr)\sin\tfrac{3}{2}\phi, \tag{3.14}$$

respectively; these terms do satisfy the wave equation, the radiation condition, and the boundary condition on the half-plane. Near the edge $r=0$, it can be shown that

$$\frac{\partial^2 u_2}{\partial x \partial y} = \frac{1}{\pi i} \frac{\exp(ikd)}{d^{1/2}} \cos\tfrac{1}{2}\Omega \; r^{-3/2} \sin\tfrac{3}{2}\phi + \frac{k}{\pi} \frac{\exp(ikd)}{d^{1/2}} \times \left(1 - \frac{1}{ikd}\right) \cos\tfrac{3}{2}\Omega \; r^{-1/2} \sin\tfrac{1}{2}\phi + O(1), \quad r \to 0. \quad (3.15)$$

The edge condition requires that u in (3.13) must be free from the $r^{-3/2}$-singularity and the $r^{-1/2}$-singularity near the edge. These requirements determine A_3 and B_3 with the results

$$A_3 = -\frac{k}{\pi} \frac{\exp(ikd)}{d^{1/2}} \left(1 - \frac{1}{ikd}\right) \cos\tfrac{3}{2}\Omega, \quad (3.16a)$$

$$B_3 = \frac{k}{\pi} \frac{\exp(ikd)}{d^{1/2}} \cos\tfrac{1}{2}\Omega. \quad (3.16b)$$

Thus, (3.13) and (3.16) give the exact total field (valid for all Ω and ϕ between 0 and 2π) due to diffraction of the incident field (3.10). Specializing this solution for $\Omega = 0+$ and $\phi = \pi$, we obtain

$$u(r, \phi = \pi) = \exp[i(kr + kd + \pi)] \times \frac{4r^{1/2}}{\pi k (r+d) d^{1/2}} \left[1 + i \frac{(r+3d)}{2kd(r+d)}\right]. \quad (3.17)$$

This exact solution again verifies the asymptotic solution derived by the uniform asymptotic theory and given in (I.3.9).

The above procedure can be continued to derive the total field due to diffraction of a higher-order multipole field in (1.1b). However, this is not necessary. In the next two sections, we will derive a recurrence relation for the total field on the shadow boundary, due to the incidence of a general multipole field, and obtain the desired field solution by solving the recurrence relation.

4. DIFFRACTION OF A GENERAL MULTIPOLE FIELD DUE TO A LINE SOURCE

This section deals with the diffraction of the line-source field (1.1) with general n by the half-plane $x \leq 0$, $y = 0$ (Fig. 1). The line source is located on the upper surface of the half-plane ($\Omega = 0+$) and the incident field (1.1) is an E-polarized wave. We will determine the resulting total field on the shadow boundary $\phi = \pi$.

Consider first the case of an incident field (1.1a) which is symmetric with respect to the plane $y = 0$. Then the fields produced by the source at $(x = -d, y = 0+)$ and its image at $(x = -d, y = 0-)$ cancel exactly. Hence, the total field is identically zero everywhere. This result verifies the asymptotic solution derived by the uniform asymptotic theory and given in (I.3.11).

Next consider the diffraction of the asymmetric E-polarized wave as given in (1.1b), viz.,

$$u_n^i(r_1, \phi_1) = H_n^{(1)}(kr_1) \sin n\phi_1, \quad n = 0, 1, 2, \cdots. \quad (4.1)$$

Let the resulting total field on the shadow boundary $\phi = \pi$ be denoted by

$$u(r, \phi = \pi) = g_n(r) \quad \text{for } \Omega = 0+, \quad (4.2)$$

then obviously

$$g_0(r) = 0 \quad (4.3)$$

and, according to (3.9) and (3.17),

$$g_1(r) = \exp[i(kr + kd + \pi/2)] \frac{2}{\pi k} \frac{r^{1/2}}{d^{1/2}(r+d)}, \quad (4.4)$$

$$g_2(r) = \exp[i(kr + kd + \pi)] \times \left[\frac{4}{\pi k} \frac{r^{1/2}}{d^{1/2}(r+d)} + \frac{2i}{\pi k^2} \frac{r^{1/2}(r+3d)}{d^{3/2}(r+d)^2}\right]. \quad (4.5)$$

We will derive a recurrence relation for the functions $\{g_n\}$. For this purpose, we observe that

$$\frac{\partial u_n^i}{\partial x} = -\tfrac{1}{2} k H_{n-1}^{(1)}(kr_1) \sin(n-1)\phi_1 + \tfrac{1}{2} k H_{n+1}^{(1)}(kr_1) \sin(n+1)\phi_1 = -\tfrac{1}{2} k u_{n-1}^i + \tfrac{1}{2} k u_{n+1}^i, \quad n = 1, 2, 3, \cdots, \quad (4.6)$$

where (3.11) and some well-known recurrence relations for the Hankel function have been used. In view of (4.6), the total field on the shadow boundary, due to the incident field $\partial u_n^i / \partial x$, is then equal to

$$-\tfrac{1}{2} k g_{n-1}(r) + \tfrac{1}{2} k g_{n+1}(r). \quad (4.7)$$

On the other hand, referring to the theorem in Sec. 2, the total field is also given by

$$\frac{\partial}{\partial x} g_n(r) - C_n \frac{\exp(ikr)}{r^{1/2}} = g_n'(r) - C_n \frac{\exp(ikr)}{r^{1/2}}. \quad (4.8)$$

Here, the constants $\{C_n\}$ are determined by the requirement that the $r^{-1/2}$-singularity in the total field at the edge should vanish (edge condition); hence,

$$C_n = \lim_{r \to 0} r^{1/2} g_n'(r). \quad (4.9)$$

By equating (4.6) and (4.8), a recurrence relation is obtained,

$$g_{n+1}(r) = g_{n-1}(r) + \frac{2}{k} g_n'(r) - \frac{2C_n}{k} \frac{\exp(ikr)}{r^{1/2}}, \quad n = 1, 2, 3, \cdots. \quad (4.10)$$

It has been verified that g_0, g_1, and g_2 in (4.3)—(4.5) do satisfy (4.10). The field $\{g_n\}$ are now completely specified by (4.10) and the "initial values" g_0 and g_1 in (4.3) and (4.4).

We will now solve the recurrence relation in (4.10). Because g_n obviously consists of n terms of order k^{-p}, $p = 1, 2, \cdots, n$, we can introduce the ansatz

$$g_n(r) = \frac{2i}{\pi} \exp[i(kr + kd + \tfrac{1}{2}n\pi)] \sum_{p=1}^{n} (ik)^{-p} A_{np} G_p(r), \quad (4.11)$$

where the coefficients $\{A_{np}\}$ and the functions $\{G_p\}$ are to be determined. The ansatz is rather special in that $\{G_p\}$ do *not* depend on n, i.e., all $\{g_n\}$ are expressed in terms of the same set of $\{G_p\}$. This choice is suggested by (4.4) and (4.5) where the leading terms contain the same function of r. Without loss of generality we may assume

$$A_{nn} = 1. \quad (4.12)$$

A. Determination of G_n

A comparison of (4.4) and (4.5) with (4.11) yields immediately

$$G_1(r) = \frac{r^{1/2}}{d^{1/2}(r+d)}, \quad G_2(r) = -\frac{r^{1/2}(r+3d)}{d^{3/2}(r+d)^2}. \quad (4.13)$$

By introducing the notation

$$\Gamma_p = \lim_{r\to 0} r^{1/2} G_p'(r), \tag{4.14}$$

C_n in (4.9) becomes

$$C_n = \frac{2i^{n+1}}{\pi} \exp(ikd) \sum_{p=1}^{n} (ik)^{-p} \Gamma_p A_{np}. \tag{4.15}$$

Inserting (4.11) and (4.15) into (4.10) and equating corresponding terms containing the same power of k^{-1}, we obtain

$$A_{n+1,p} G_p(r) = -A_{n-1,p} G_p(r) + 2A_{np} G_p(r) + 2A_{n,p-1}[G_{p-1}'(r) - \Gamma_{p-1} r^{-1/2}]. \tag{4.16}$$

In (4.16) we set $p = n+1$; then, in view of $A_{n+1,n+1} = A_{nn} = 1$, and $A_{n-1,n+1} = A_{n,n+1} = 0$, we have

$$G_{n+1}(r) = 2G_n'(r) - 2\Gamma_n r^{-1/2}, \quad n = 1, 2, 3, \cdots. \tag{4.17}$$

This equation recursively defines the functions $\{G_n\}$. It has been verified that G_2 in (4.13) does satisfy (4.17). From (4.13) we may deduce the series expansions

$$G_1(r) = \frac{1}{d} \sum_{q=0}^{\infty} (-1)^q (r/d)^{q+1/2}, \tag{4.18a}$$

$$G_2(r) = \frac{2}{d^2} \sum_{q=1}^{\infty} (-1)^q (q+\tfrac{1}{2})(r/d)^{q-1/2} \tag{4.18b}$$

which are valid for $0 \le r < d$. Then from (4.17) it is easily found that

$$G_p(r) = \frac{2^{p-1}}{d^p} \sum_{q=p-1}^{\infty} (-1)^q (q+\tfrac{1}{2}) \times (q-\tfrac{1}{2}) \cdots (q-p+\tfrac{5}{2})(r/d)^{q-p+3/2} \tag{4.19}$$

which is valid for $0 \le r < d$. Starting from (4.19), $G_p(r)$ may be written as a hypergeometric function (see Ref. 7, Chapter 15),

$$G_p(r) = (-1)^{p-1} \frac{(2p-1)!}{2^{p-1}(p-1)!} \frac{1}{d^p} \left(\frac{r}{d}\right)^{1/2} \times F(p+\tfrac{1}{2}, 1; \tfrac{3}{2}; -r/d). \tag{4.20}$$

Then, by use of a well-known integral representation for the hypergeometric function [Ref. 7, Eq. (15.3.1)], we obtain

$$G_p(r) = (-1)^{p-1} \frac{(2p-1)!}{2^p (p-1)!} \times \int_0^r (r+d-t)^{-p-1/2} t^{-1/2}\, dt, \quad p = 1, 2, 3, \cdots. \tag{4.21}$$

The latter representation for G_p, which is an analytic continuation of the one in (4.19), is valid for all r between 0 and ∞. Alternatively, by use of a linear transformation formula for the hypergeometric function [Ref. 7, Eq. (15.3.12)], G_p can be reduced to

$$G_p(r) = (-1)^{p-1} 2^{p-1} (p-1)! \left(\frac{r}{d}\right)^{1/2} \times \sum_{q=0}^{p-1} \frac{(2q)!}{2^{2q}(q!)^2} \frac{1}{d^q (r+d)^{p-q}}, \quad p = 1, 2, 3, \cdots. \tag{4.22}$$

The result in (4.22) is the desired final expression for the functions $\{G_p\}$.

B. Determination of A_{np}

The use of (4.17) in (4.16) leads to

$$A_{n+1,p} = -A_{n-1,p} + 2A_{np} + A_{n,p-1}, \tag{4.23}$$

which can be solved in a standard manner subject to the side condition (4.12). The result is

$$A_{np} = \binom{n+p-1}{2p-1}, \tag{4.24}$$

where $(\cdot)$ denotes the binomial coefficient.

In summary, the *exact* total field on the shadow boundary $\phi = \pi$, due to diffraction of an incident E-polarized wave in (4.1) with $\Omega = 0+$, is given by an n-term sum, namely,

$$g_n(r) = \frac{2i}{\pi} \exp[i(kr + kd + \tfrac{1}{2}n\pi)] \times \sum_{p=1}^{n} \binom{n+p-1}{2p-1} (ik)^{-p} G_p(r), \quad n = 1, 2, 3, \cdots, \tag{4.25}$$

where $\{G_p\}$ are given in (4.22). The first two terms of (4.25) are

$$g_n(r) = \exp[i(kr + kd + \tfrac{1}{2}n\pi)] \frac{2nr^{1/2}}{\pi k (r+d) d^{1/2}} \cdot \left[1 + i\frac{(n^2-1)(r+3d)}{6kd(r+d)} + O(k^{-2})\right], \tag{4.26}$$

which agrees with the asymptotic solution in (I.3.9), derived from the uniform asymptotic theory. For the case $n=1$ ($n=2$), there is only one term (two terms) in (4.25); thus, the asymptotic solution in (I.3.9) becomes exact in these cases.

5. ALTERNATIVE SOLUTION OF THE RECURRENCE RELATION

For the diffraction of the E-polarized wave given in (4.1) with $\Omega = 0+$, the resulting total field on the shadow boundary is denoted by $g_n(r)$ as indicated in (4.2). We will now present an alternative method for solving the recurrence relation for $\{g_n\}$ in (4.10), or

$$g_{n+1}(r) = g_{n-1}(r) + \frac{2}{k} g_n'(r) - \frac{2C_n}{k} \frac{\exp(ikr)}{r^{1/2}}, \quad n = 1, 2, 3, \cdots. \tag{5.1}$$

In this method, the integral in (4.21) is obtained in a more natural manner.

Consider first the constants $\{C_n\}$ as defined by (4.9). From (4.4) and (4.5), it may be shown that C_1 and C_2 may be expressed in terms of Hankel functions of half-integral order and of argument kd:

$$C_1 = \frac{i}{\pi k} \frac{\exp(ikd)}{d^{3/2}} = -\left(\frac{k}{2\pi}\right)^{1/2} [H_{3/2}^{(1)}(kd) + iH_{1/2}^{(1)}(kd)], \tag{5.2a}$$

$$C_2 = -\frac{2}{\pi k} \frac{\exp(ikd)}{d^{3/2}} - \frac{3i}{\pi k^2} \frac{\exp(ikd)}{d^{5/2}} = \left(\frac{k}{2\pi}\right)^{1/2} [H_{5/2}^{(1)}(kd) + iH_{3/2}^{(1)}(kd)]. \tag{5.2b}$$

Guided by these results, it is conjectured that

$$C_n = (-1)^n \left(\frac{k}{2\pi}\right)^{1/2} [H_{n+1/2}^{(1)}(kd) + iH_{n-1/2}^{(1)}(kd)]. \tag{5.3}$$

It has been verified that this conjecture also holds true for C_3 and C_4.

Furthermore, observe that (5.1) and (5.3) remain valid when n is a negative integer. As a matter of fact, for negative n, the incident field in (4.1) becomes

$$u^i_{-n}(r_1, \phi_1) = H^{(1)}_{-n}(kr_1)\sin(-n\phi_1) = (-1)^{n+1}u^i_n(r_1, \phi_1). \quad (5.4)$$

Thus, the associated total field g_{-n} and constant C_{-n} satisfy

$$g_{-n}(r) = (-1)^{n+1}g_n(r), \quad C_{-n} = (-1)^{n+1}C_n. \quad (5.5)$$

It is easily seen that (5.1) and (5.3) are consistent with the symmetry relation in (5.5).

The recurrence relation (5.1), valid now for all positive and negative integer n, is solved next by a formal generating-function technique. Introduce the generating function

$$F(r, \theta) = \sum_{n=-\infty}^{\infty} g_n(r)\exp(in\theta), \quad (5.6)$$

then (5.1) implies the following differential equation for F:

$$\frac{\partial}{\partial r}F(r, \theta) + ik\sin\theta F(r, \theta) = C(\theta)\frac{\exp(ikr)}{r^{1/2}}, \quad (5.7a)$$

where

$$C(\theta) = \sum_{n=-\infty}^{\infty} C_n\exp(in\theta). \quad (5.7b)$$

The differential equation (5.7) is solved by variation of parameters. Since $F(r, \theta) = 0$ at $r = 0$ (edge condition), we obtain the solution

$$F(r, \theta) = C(\theta)\int_0^r \exp\{ik[t - (r-t)\sin\theta]\}t^{-1/2}\,dt. \quad (5.8)$$

Using the well-known generating function for Bessel functions [Ref. 7, Eqs. (9.1.42) and (9.1.43)], one has

$$\exp[-ik(r-t)\sin\theta] = \sum_{p=-\infty}^{\infty} J_p[k(r-t)]\exp(-ip\theta) \quad (5.9)$$

and (5.8) can be rewritten as

$$F(r, \theta) = \sum_{q=-\infty}^{\infty} C_q\exp(iq\theta)\sum_{p=-\infty}^{\infty}\exp(-ip\theta) \times \int_0^r J_p[k(r-t)]\exp(ikt)t^{-1/2}\,dt. \quad (5.10)$$

Comparing (5.10) and (5.6), we immediately deduce that

$$g_n(r) = \sum_{p=-\infty}^{\infty} C_{p+n}\int_0^r J_p[k(r-t)]\exp(ikt)t^{-1/2}\,dt. \quad (5.11)$$

Substitute the conjectured values (5.3) for C_{p+n} into (5.11), then by use of the following identities [See Ref. 7, Eq. (9.1.79)]:

$$\sum_{p=-\infty}^{\infty}(-1)^pH^{(1)}_{p+n+1/2}(kd)J_p[k(r-t)] = H^{(1)}_{n+1/2}[k(r+d-t)], \quad (5.12)$$

$$\sum_{p=-\infty}^{\infty}(-1)^pH^{(1)}_{p+n-1/2}(kd)J_p[k(r-t)] = H^{(1)}_{n-1/2}[k(r+d-t)], \quad (5.13)$$

$$t^{-1/2}\exp(ikt) = i\left(\frac{\pi k}{2}\right)^{1/2}H^{(1)}_{1/2}(kt), \quad (5.14)$$

we obtain the desired solution for $g_n(r)$, namely,

$$g_n(r) = \tfrac{1}{2}(-1)^n ik\int_0^r\{H^{(1)}_{n+1/2}[k(r+d-t)] + iH^{(1)}_{n-1/2}[k(r+d-t)]\}H^{(1)}_{1/2}(kt)\,dt, \quad n = 1, 2, 3, \cdots. \quad (5.15)$$

The result in (5.15) is an exact representation of the total field on the shadow boundary. The derivation of (5.15) is based on the conjectured values (5.3) for $\{C_n\}$ and a formal generating-function technique. The convergence of the series involved and the interchange of the order of summation and integration were not seriously studied. Thus, (5.15) requires the following additional verification:

(i) Determine C_n from (4.9) and (5.15), then the conjectured value (5.3) is precisely recovered.

(ii) By direct substitution, the solution in (5.15) has been shown to satisfy the recurrence relation (5.1)

This verification shows that g_n is given by the exact representation (5.15).

To derive a more explicit solution from (5.15), we may express the Hankel function in terms of elementary functions,

$$H^{(1)}_{n+1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2}\exp[i(z - \tfrac{1}{2}n\pi - \tfrac{1}{2}\pi)] \times\sum_{p=0}^{n}\frac{(n+p)!}{(n-p)!\,p!}\frac{(-1)^p}{(2iz)^p}. \quad (5.16)$$

As a result, (5.15) is reduced to

$$g_n(r) = \frac{2i}{\pi}\exp[i(kr + kd + \tfrac{1}{2}n\pi)]\sum_{p=1}^{n}\frac{(n+p-1)!}{(n-p)!\,(p-1)!} \times(-1)^{p-1}(2ik)^{-p}\int_0^r (r+d-t)^{-p-1/2}t^{-1/2}\,dt. \quad (5.17)$$

With the help of the representation (4.21) of $G_p(r)$, (5.17) may be rewritten as

$$g_n(r) = \frac{2i}{\pi}\exp[i(kr + kd + \tfrac{1}{2}n\pi)] \times\sum_{p=1}^{n}\binom{n+p-1}{2p-1}(ik)^{-p}G_p(r) \quad (5.18)$$

which agrees with (4.25), the solution obtained by the first method.

6. DISCUSSION AND NUMERICAL RESULTS

In the present paper, the exact solution to the diffraction of a line-source field by a half-plane is studied by analytical methods. When the incident field given in (4.1) is an E-polarized wave and is due to a line source located on the upper surface of the half-plane, the exact total field on the shadow boundary is given in (4.25), which is an n-term sum (n is an index of the incident field), or in (5.15), which is a finite integral. The first two terms of (4.25) agree with the asymptotic solution determined by the uniform asymptotic theory in Part I.

For a given incident field (fixed n), the total field in (4.25) or (5.15) depends on two parameters d and r, which are the distances from the edge to the source,

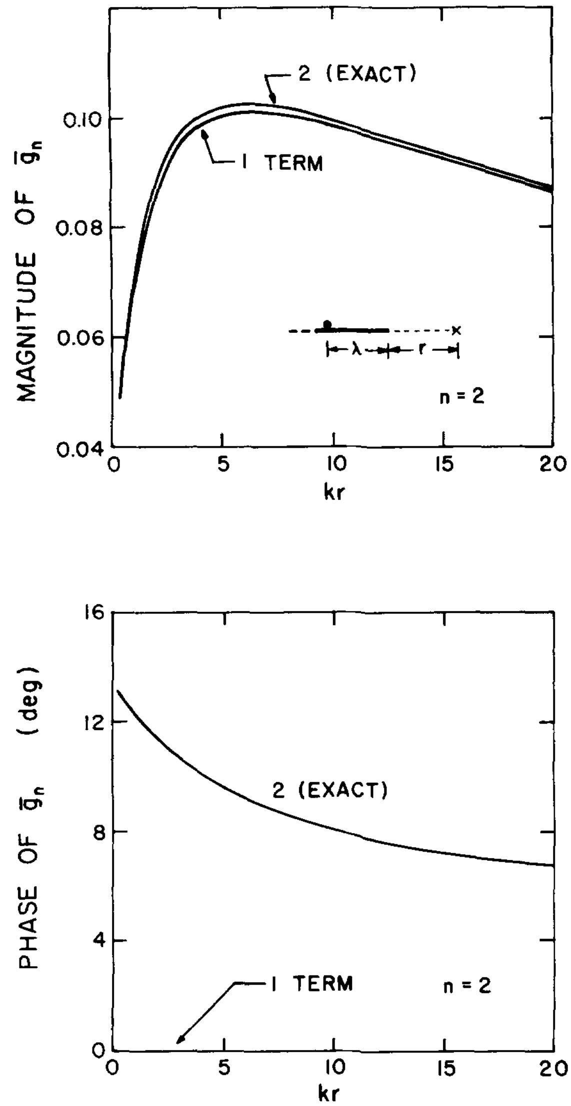

FIG. 2. Normalized total field on the shadow boundary due to an incident field (4.1) from the line source on the half-plane. $\bar{g}_n$ is defined by (4.2) and (6.3), and it is calculated from (4.25) and (4.22) with one, two, . . . , or n terms in the sum.

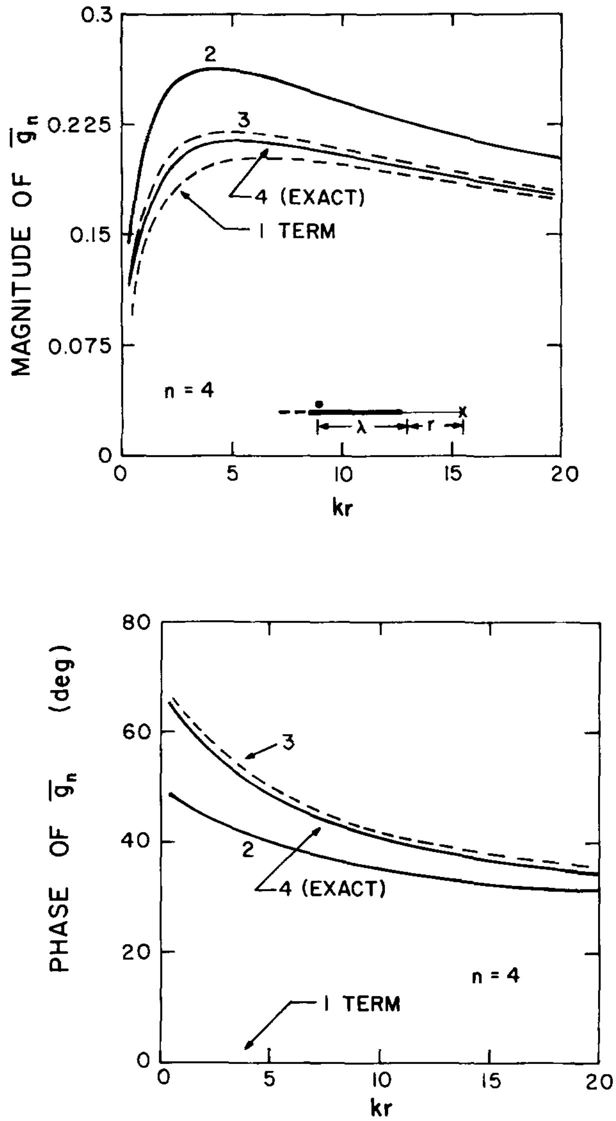

FIG. 3. Same as Fig. 2 except that $n=4$.

and to the observation point, respectively. For the extreme case $(r/d)\to 0$ (near field or faraway source), it is found from (5.15) that

$$g_n(r)=(-1)^n\left(\frac{2kr}{\pi}\right)^{1/2}\times\exp(ikr)[H^{(1)}_{n+1/2}(kd)+iH^{(1)}_{n-1/2}(kd)]\left[1+O\left(\frac{r}{d}\right)\right]$$
$$=\frac{4i}{\pi}\exp[i(kr+kd+\tfrac{1}{2}n\pi)]\left(\frac{r}{d}\right)^{1/2}\times\sum_{p=1}^{n}\frac{(n+p-1)!}{(n-p)!\,(p-1)!}\frac{(-1)^{p-1}}{(2ikd)^p}\left[1+O\left(\frac{r}{d}\right)\right],\quad\left(\frac{r}{d}\right)\to 0. \tag{6.1}$$

For the other extreme case $(d/r)\to 0$ (far field or nearby source), it can be shown from (4.22) and (4.25) that

$$g_n(r)=\frac{4i}{\pi}\exp[i(kr+kd+\tfrac{1}{2}n\pi)]\left(\frac{d}{r}\right)^{1/2}\times\sum_{p=1}^{n}\frac{(n+p-1)!}{(n-p)!\,(p-1)!\,(2p-1)}\frac{(-1)^{p-1}}{(2ikd)^p}\times\left[1+O\left(\frac{d}{r}\right)\right],\quad\left(\frac{d}{r}\right)\to 0. \tag{6.2}$$

The n-term sums appearing in (6.1) and (6.2), are both polynomials in inverse powers of kd. Thus, the use of one or two "dominant terms" in these two extreme cases can give good results only if $kd\gg 1$, and its accuracy is independent of kr.

In Figs. 2 and 3, we fix $kd=2\pi$ (or $d=1\lambda$) and display a normalized total field

$$\bar{g}_n(r)=\exp[-i(kr+kd+\tfrac{1}{2}n\pi)]g_n(r) \tag{6.3}$$

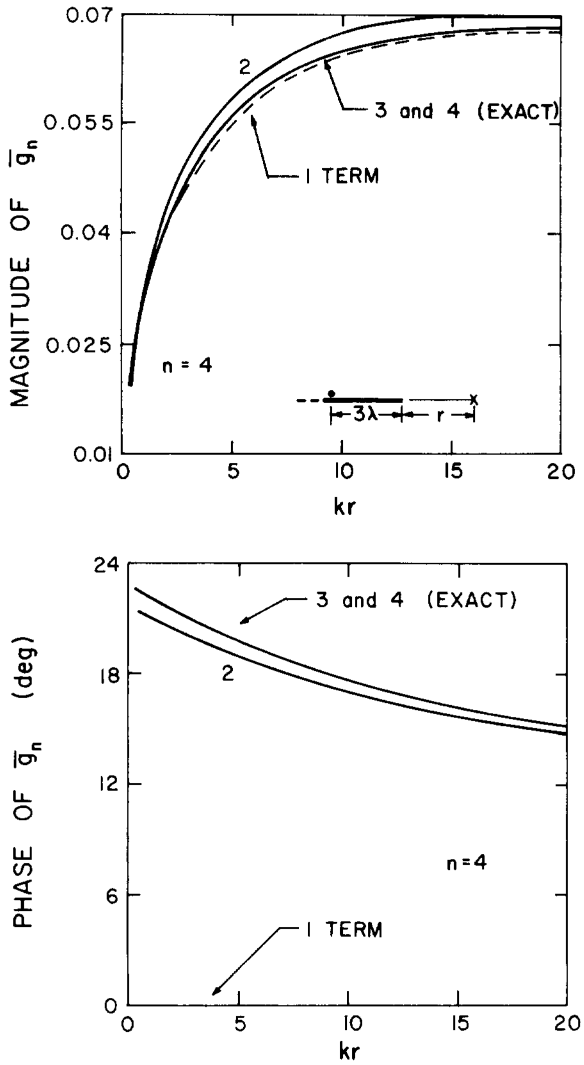

FIG. 4. Same as Fig. 2 except that $n=4$ and $d=3\lambda$.

as a function of kr for two incident fields $n=2$ and $n=4$; $\bar{g}_n$ is calculated from (4.25) and (4.22) with one, two, $\cdots$, or n terms in the sum, where the one with n terms is the exact solution. Since $kd=2\pi$ is relatively small and $n=4$ corresponds to a rapidly varying incident field, the curves calculated with one or two terms in the sum in Fig. 3 do not converge well to the exact solution. In particular, we note in Fig. 3 that the curves calculated with one term show a reasonable magnitude but the phase is far off.

The poor convergence mentioned above becomes less serious as kd is increased, as indicated in (6.1) and (6.2). In Fig. 4, we reconsider the case presented in Fig. 3 but with $kd=6\pi$ (triple the previous value). The curves calculated with two terms already give good results in both magnitude and phase.

*Lee's work was supported by National Science Foundation Grant NSF ENG 73-08218.

[1]J. Boersma and S.W. Lee, "High-frequency diffraction of a line-source field by a half-plane: Solutions by ray techniques," to appear in *IEEE Trans. Antennas Propag.* Mar. 1977.

[2]M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1964), 2nd ed.

[3]R.Kh. Khestanov, *Radio Eng. Electron. Phys. USSR* **15**, 250–257 (1970).

[4]C.J. Bouwkamp, *Rep Prog. Phys.* **17**, 35–100 (1954).

[5]J.J. Bowman, T.B.A. Senior, and P.L.E. Uslenghi, Eds., *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).

[6]C.J. Bouwkamp, *Physica* **12**, 467–474 (1946).

[7]M. Abramowitz and I.A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (National Bureau of Standards, Washington, 1964).

Some aspects of the N-representability problem in finite dimensions. I. Operator endomorphisms which induce necessary conditions

H. Kummer, I. Absar, and A. J. Coleman

Department of Mathematics, Queen's University, Kingston, Canada K7L 3N6
(Received 4 June 1976)

The connection between operator endomorphisms and conditions for N-representability is established. The sets N_N and S_N are defined as the sets of operator endomorphisms which induce necessary conditions and sufficient conditions respectively. A detailed characterization of N_N is presented.

1. INTRODUCTION

The Hamiltonian describing the dynamics of an electron cloud of a molecule in the Born–Oppenheimer approximation is of the form

$$H=\sum_{i=1}^{N}H_1(i)+\sum_{i<k}^{N}H_2(ik). \tag{1.1}$$

The operator is defined on some dense subspace $\tilde{D}$ of the N-fold tensor product $\tilde{H}^N$ of a separable Hilbert space $\tilde{H}$ which comprises all possible pure states of *one* electron and therefore is called the "one electron" or "orbital" space.

The possible energies which the isolated electron cloud is capable of are identical with the numbers contained in the spectrum of the compression[1] of H to the antisymmetric subspace $\tilde{H}^N_\wedge$ of $\tilde{H}^N$. Thus the ground state energy E_0 is given by

$$E_0=\inf \operatorname{tr}(H\tilde{D}), \tag{1.2}$$

where $\tilde{D}$ varies over all positive trace class operators of trace 1 whose range is contained in $\tilde{D}\cap\tilde{H}^N_\wedge$.

Let us call a finite dimensional subspace $H\subseteq\tilde{H}$ *admissible* iff $H^N\subseteq\tilde{D}$. Throughout this paper we assume that a fixed admissible subspace $H\subseteq\tilde{H}$ of dimension $r>3$ has been chosen. For $1\le p\le N\le r$ let S^p be the set of all Hermitian linear operators of $\tilde{H}^p$ whose range is contained in $H^p_\wedge$ and let P^p denote the subset of all positive operators (belonging to S^p) of trace 1, i.e.,

$$P^p \triangleq \{D\in S^p \mid D\ge 0,\ \operatorname{tr}(D)=1\}. \tag{1.3}$$

(Here $\triangleq$ denotes definition)

An element $D\in P^p$ is called a *p density operator* (whose one range is contained in the given admissible subspace of the orbital space $\tilde{H}$). Then clearly we have

$$E_0\le \inf_{D\in P^N} \operatorname{tr}(HD). \tag{1.4}$$

Formula (1.4) will serve us as a basis for obtaining upper bounds to the ground state energy, E_0, for the system.

To proceed further let us introduce the projections P^p and $\tilde{A}^p$ corresponding to the subspaces H^p and $\tilde{H}^p_\wedge$ of $\tilde{H}^p$, respectively. Clearly they commute and $A^p\triangleq P^p\tilde{A}^p$ is the projection corresponding to $H^p_\wedge$. We may write

$$D\in S^p \ \text{ iff } D^*=D \ \text{ and } \ A^pDA^p=D.$$

Furthermore, we have

$$A^NHA^N=A^N(K\otimes I^{N-p})A^N, \tag{1.5}$$

where

$$K=\frac{N}{2}A^2[H_1(1)+H_1(2)+(N-1)H_2(12)]A^2 \tag{1.6}$$

is the operator obtained from the "reduced Hamiltonian"[2] by multiplication on both sides by A^2. In the sequel we shall refer to K as the "truncated reduced Hamiltonian." Obviously $K\in S^2$. If we introduce the (p,N)-expansion operator,

$$\Gamma^N_p : S^p \to S^N,\ ^3$$

devined by $\Gamma^N_p(B)=A^N(B\otimes I^{N-p})A^N$, $B\in S^p$, (1.5) can be rewritten as

$$A^NHA^N=\Gamma^N_2(K). \tag{1.5'}$$

Clearly S^p is a real Hilbert space under the inner product

$$(D,B)\mapsto\langle D\mid B\rangle\triangleq \operatorname{tr}(DB),\quad D,B\in S^p,$$

whose dimension is given by $\binom{r}{p}^2$, where r denotes the dimension of H.

The adjoint operator of the (p,N)-expansion operator relative to this inner product is the (N,p)-contraction operator L^p_N.[2] Thus the defining equation of L^p_N is given by

$$\langle L^p_N(D)\mid B\rangle=\langle D\mid \Gamma^N_p(B)\rangle,\quad D\in S^N,\ B\in S^p.$$

More explictly, L^p_N is given as the partial trace of D over the last $N-p$ factors in the tensor product. Let P^p_N be the image of P^N under L^p_N. An element of P^p_N is called an N-representable p-density operator (whose one range is contained in H). We have the following lemma, crucial to our reasoning:

1.1 *Lemma*: $E_0\le\inf_{D\in P^2_N}\langle K\mid D\rangle$.

Proof: For $D\in P^N$ we have

$$\operatorname{tr}(HD)=\operatorname{tr}(HA^NDA^N)=\operatorname{tr}(A^NHA^ND)=\langle A^NHA^N\mid D\rangle$$
$$=\langle\Gamma^N_2(K)\mid D\rangle=\langle K\mid L^2_N(D)\rangle.$$

Combining this insight with formula (1.4) we obtain

$$E_0\le\inf_{D\in P^N}\operatorname{tr}(HD)=\inf_{D\in P^N}\langle K\mid L^2_N(D)\rangle$$
$$=\inf_{D\in P^2_N}\langle K\mid D\rangle.$$

The formula of Lemma 1.1 is not very practical because a convenient characterization of $\mathcal{P}_N^2$ is unknown, but the *knowledge of any subset of* $\mathcal{P}_N^2$ *will provide us with an upper bound to* E_0. Such a subset is given by the singleton $\{\binom{r}{2}^{-1}A^2\}$. In Part II of this paper we exhibit some other less trivial subsets of $\mathcal{P}_N^2$ as the images of $\mathcal{P}^2$ under certain maps called *operator endomorphisms*. In Part I we focus on *necessary conditions for N representability* induced by such operator endomorphisms.

2. OPERATOR ENDOMORPHISMS AND CONDITIONS FOR *N* REPRESENTABILITY

2.1 *Definition*: Let T be a linear map of S^2 into itself.

(i) T is said to induce a *sufficient condition for N representability* iff $T(\mathcal{P}^2)\subseteq \mathcal{P}_N^2$.

(ii) T is said to induce a necessary condition for N *representability* iff $T(\mathcal{P}_N^2)\subseteq \mathcal{P}^2$.

(iii) T is called *order preserving* iff $T(\mathcal{P}^2)\subseteq \mathcal{P}^2$.

The motivation for this definition is clear. If D is any density operator and T induces a sufficient condition for N representability then $T(D)$ is N representable and the lowest eigenvalue of $T^*(K)$ is an upper bound to the ground state energy E_0. Note that T^* is the adjoint operator to T. Indeed,

$$E_0 \leqslant \inf_{D\in T(\mathcal{P}^2)} \langle K|D\rangle = \inf_{D\in \mathcal{P}^2}\langle K|\,T(D)\rangle$$
$$= \inf_{D\in \mathcal{P}^2}\langle T^*(K)\,|\,D\rangle = \text{lowest eigenvalue of } T^*.$$

Suppose T induces a necessary condition for N representability. Then $T(D)\geqslant 0$ for every N-representable density operator and the lowest eigenvalue of $(T^*)^{-1}(K)$ (we assume T is invertible!) is a lower bound to the lowest eigenvalue of the truncated Hamiltonian which in turn is an upper bound to the ground state energy E_0. Indeed,

$$E_0 \leqslant \inf_{D^N\in \mathcal{P}^N}\langle H|D^N\rangle = \inf_{D\in \mathcal{P}_N^2}\langle K|D\rangle$$
$$\geqslant \inf_{D\in T^{-1}(\mathcal{P}^2)}\langle K|D\rangle = \inf_{D\in \mathcal{P}^2}\langle K|\,T^{-1}(D)\rangle$$
$$= \inf_{D\in \mathcal{P}^2}\langle T^{*-1}(K)\,|\,D\rangle = \text{lowest eigenvalue of } T^{*-1}(K).$$

Also notice the following fact which we shall exploit later on in this paper: If T_1 induces a sufficient condition for N representability and if T_2 induces a necessary condition for N representability then the composite transformation $T_2\circ T_1$ is order preserving.

Finally we remark that for $2\leqslant N\leqslant (r-2)$, $\mathcal{P}^2$ and $\mathcal{P}_N^2$ span S^2 as a linear space. This implies that if a linear map T of S^2 into itself has any of the three properties listed in Def. 2.1, it preserves the trace.

The most interesting and accessible linear transformations of S^2 are the *operator endomorphisms*. Let $\mathcal{U}=\mathcal{U}(\mathcal{H})$ be the group of all unitary transformations of the admissible subspace $\mathcal{H}\subseteq\tilde{\mathcal{H}}$. $\mathcal{U}$ acts in a natural way on $\mathcal{H}^p$ and therefore also on S^p (by similarity transformation). An *operator homomorphism* is a linear map between two carrier spaces of representations of $\mathcal{U}$ which commutes with the action of $\mathcal{U}$. For example, the (N,p)-contraction L_N^p and the (p,N)-expansion operator Γ_p^N are such operator homomorphisms. Since $\mathcal{P}^N$ is invariant under $\mathcal{U}$ and the $(N,2)$-contraction is an operator homomorphism, $\mathcal{P}_N^2$ is invariant under the group $\mathcal{U}$. Similarly if T is an operator *endomorphism* of S^2 (i.e., an operator homomorphism of S^2 into itself) which carries $\mathcal{P}^2$ into $\mathcal{P}_N^2$, then the image $T(\mathcal{P}^2)$ will be a convex subset of $\mathcal{P}_N^2$ which shares with $\mathcal{P}_N^2$ the property of being invariant under $\mathcal{U}$. This is one reason why the operator endomorphisms are of particular interest to us.

The following proposition asserts that the set of all operator endomorphisms of S^2 constitutes a three-dimensional commutative algebra over the field $\mathbf{R}$ of real numbers. We restrict ourselves to the case where the dimension r, of our basic admissible subspace $\mathcal{H}$ of orbital space, satisfies the inequality

$$r\triangleq \dim\mathcal{H} > 3.$$

2.2 *Proposition*: (i) The algebra $\mathcal{A}$ of all operator endomorphisms of S^2 constitutes a three-dimensional commutative algebra over $\mathbf{R}$.

(ii) A linear map T of S^2 into itself belongs to $\mathcal{A}$ iff it has the form

$$D\to T(D)=\alpha\,\mathrm{tr}(D)A^2+\beta(D^1\wedge I^1)+\gamma D.$$

[Here we used the abbreviations D^1 for $L_2^1(D)$ and $D^1\wedge I^1$ for $A^2(D^1\otimes I^1)A^2$.]

Proof: S^2 is the direct sum of three inequivalent irreducible subspaces under $\mathcal{U}$,

$$S^2=\langle A^2\rangle\oplus(\langle A^2\rangle^{\perp}\ominus\mathcal{K})\oplus\mathcal{K}. \tag{2.1}$$

Here $\langle A^2\rangle$ stands for the one-dimensional subspace spanned by A^2, and $\mathcal{K}$ stands for the kernel of L_2^1. Their respective dimensions are given by 1, r^2-1, and $\binom{r}{2}^2-r^2$. Part (i) of the proposition follows at once from Schur's lemma.

In order to prove (ii) let $D\in S^2$ be arbitrary. The three irreducible components of D are given by

$$D_0=\binom{r}{2}^{-1}\mathrm{tr}(D)A^2, \tag{2.2a}$$
$$D_1=4(r-2)^{-1}[D^1\wedge I^1-r^{-1}\mathrm{tr}(D)A^2], \tag{2.2b}$$
$$D_2=D-4(r-2)^{-1}(D^1\wedge I^1)+\binom{r-1}{2}^{-1}\mathrm{tr}(D)A^2. \tag{2.2c}$$

Here $D_0\in\langle A^2\rangle$, $D_2\in\mathcal{K}$, and therefore $D_1\triangleq D-D_0-D_2\in\langle A^2\rangle^{\perp}\ominus\mathcal{K}$.

{In order to verify that $D_2\in\mathcal{K}$ notice that

$$L_2^1(D^1\wedge I^1)=[(r-2)/4]D^1+\tfrac14\mathrm{tr}(D)I^1.\} \tag{2.3}$$

Now let $T\in\mathcal{A}$ be arbitrary. We have

$$T(D)=\epsilon D_0+\lambda D_1+\mu D_2$$

for some triple $(\epsilon,\lambda,\mu)\in\mathbf{R}^3$. Substituting the expressions (2.2) for D_0, D_1, and D_2 we obtain

$$T(D)=\alpha\,\mathrm{tr}(D)A^2+\beta(D^1\wedge I^1)+\gamma D, \tag{2.4}$$

where

$$\alpha=\frac{2}{r(r-1)(r-2)}[(r-2)\epsilon+r\mu-2(r-1)\lambda], \tag{2.5a}$$

$$\beta=\frac{4}{(r-2)}(\lambda-\mu), \tag{2.5b}$$

$$\gamma=\mu. \tag{2.5c}$$

Conversely, suppose T has the form of Proposition 2.2. Then

$$T=\alpha\Gamma_0^2L_2^0+\beta\Gamma_1^2L_2^1+\gamma\mathrm{Id} \tag{2.6}$$

($\mathrm{Id}=$ identity on S^2), which makes it obvious that T is an operator endomorphism.

2.3 *Corollary*: If T is an operator endomorphism of S^2, then T is self-adjoint.

Proof: Clear from formula (2.6).

2.4 *Corollary*: The algebra $\mathcal{A}$ of all operator endomorphisms is isomorphic to the algebra of all *diagonal* 3×3 matrices over $\mathbb{R}$.

Proof: Indeed the map $T\to \mathrm{diag}(\epsilon,\lambda,\mu)$ is such an isomorphism.

Combining the insight expressed in the paragraph following Definition 2.1 with Corollary 2.3 we obtain the following recipe.

Recipe:

(i) If $T\in\mathcal{A}$ is such that $T(\mathcal{P}_N^2)\subseteq\mathcal{P}^2$, then the condition

$T(D)\geq 0$

is a necessary condition for N representability. If, in addition, T is invertible then the lowest eigenvalue of $T^{-1}(K)$ is a lower bound to the lowest eigenvalue of the truncated Hamiltonian (i.e., the compression of the Hamiltonian onto $\mathcal{H}^N$).

(ii) If $T\in\mathcal{A}$ is such that $T(\mathcal{P}^2)\subseteq\mathcal{P}_N^2$, then the lowest eigenvalue of $T(K)$ is an upper bound to the ground state energy E_0 of the N-electron system.

Here K stands for the truncated *reduced* Hamiltonian.[2]

This recipe suggests the introduction of the following sets:

$$\mathcal{M}\triangleq\{T\in\mathcal{A}\mid T(\mathcal{P}^2)\subseteq\mathcal{P}^2\}, \tag{2.7a}$$

$$\mathcal{N}_N\triangleq\{T\in\mathcal{A}\mid T(\mathcal{P}_N^2)\subseteq\mathcal{P}^2\}, \tag{2.7b}$$

$$\mathcal{S}_N\triangleq\{T\in\mathcal{A}\mid T(\mathcal{P}^2)\subseteq\mathcal{P}_N^2\}. \tag{2.7c}$$

In words: $\mathcal{M}$ is the set of all order preserving operator endomorphisms and $\mathcal{N}_N$ and $\mathcal{S}_N$ denote the sets of all operator endomorphisms which induce necessary conditions for N representability and sufficient conditions for N representability, respectively.

$\mathcal{M}$, $\mathcal{N}_N$, $\mathcal{S}_N$ are convex subsets of the semigroup $\mathcal{A}_0$ of all operator endomorphisms which preserve the trace. It follows from Corollary 2.3 that an operator endomorphism T preserves the trace iff $T(A^2)=A^2$ and thus iff $\epsilon=1$. Therefore, $\mathcal{A}_0$ is isomorphic as a semigroup to the semigroup of all diagonal 2×2 matrices $\mathrm{diag}(\lambda,\mu)$. The above mentioned three sets therefore can be thought of *as convex subsets of the* (λ,μ) *plane* [or (α,β) plane]. Note that the eigenvalues (ϵ,λ,μ) of the operator endomorphism $T=T(\alpha,\beta,\gamma)$ are given by

$$\epsilon=\binom{r}{2}\alpha+\frac{r-1}{2}\beta+\gamma, \tag{2.8a}$$

$$\lambda=\frac{r-2}{4}\beta+\gamma, \tag{2.8b}$$

$$\mu=\gamma, \tag{2.8c}$$

[cf. (2.5)], and therefore $T\in\mathcal{A}_0$ iff

$$\binom{r}{2}\alpha+\frac{r-1}{2}\beta+\gamma=1. \tag{2.9}$$

3. OPERATOR ENDOMORPHISMS WHICH INDUCE NECESSARY CONDITIONS

The following theorem gives a complete characterization of the set $\mathcal{N}_N$ in the case where r and N are even numbers.

3.1 *Theorem*: Let r and N be even numbers satisfying $2\leq N\leq r-2$.

(i) A necessary and sufficient condition for $T(\alpha,\beta)\in\mathcal{A}_0$ inducing a necessary condition for N representability is that α,β satisfy the following inequalities:

$$\alpha\geq 0, \tag{3.1a}$$

$$\beta\geq -2N\alpha, \tag{3.1b}$$

$$(r+N-1)\alpha+\beta\leq\frac{2}{r-N}, \tag{3.1c}$$

$$r\alpha+\beta\leq\frac{2(r-N+2)}{(r+1)(r-N)}. \tag{3.1d}$$

(ii) $\mathcal{N}_N$ is the convex hull of the following four extreme points:

$$\mathrm{Id}(D)=D, \tag{3.2a}$$

$$Q_N(D)=\binom{r-N}{r}^{-1}\left[\mathrm{tr}(D)A^2-2N(D^1\wedge I^1)+\binom{N}{2}D\right], \tag{3.2b}$$

$$B_N(D)=\frac{2}{(r-N)(r+1)}\left[\mathrm{tr}(D)A^2-(N-2)(D^1\wedge I^1)-(N-1)D\right], \tag{3.2c}$$

$$C_N(D)=\frac{2}{(r-N)(r+1)}[(r-N+2)(D^1\wedge I^1)-(N-1)D]. \tag{3.2d}$$

Remark: If r or N is odd, the inequalities (3.1) are still sufficient conditions for the corresponding operator endomorphisms to belong to $\mathcal{N}_N$.

Proof: First we choose $D=\binom{N}{2}^{-1}P_U\wedge P_U$, where $U\subseteq\mathcal{H}$ is an N-dimensional subspace and P_U denotes the corresponding projection. $D\in\mathcal{P}_N^2$; in fact $D=L_N^2(P_{[U]})$ where $P_{[U]}\in\mathcal{P}^N$ is the projection onto the one-dimensional subspace of $\mathcal{H}^N$ spanned by the Slater determinant $[U]$ corresponding to U. Let $\{\phi_1,\phi_2,\ldots,\phi_r\}\subseteq\mathcal{H}$ be an orthonormal basis such that $\{\phi_1,\ldots,\phi_N\}\subseteq U$. Then the matrix of $T(D)$ relative to the derived basis $\{[\phi_1,\phi_2],\ldots,[\phi_{r-1},\phi_r]\}$ of $\mathcal{H}^2$ is diagonal; more precisely it is given by

$$\mathrm{diag}\Big(\underbrace{\alpha+N^{-1}\beta+\binom{N}{2}^{-1}\gamma,\ldots,\alpha+N^{-1}\beta+\binom{N}{2}^{-1}\gamma}_{\binom{N}{2}},\underbrace{\alpha+(2N)^{-1}\beta,\ldots,\alpha+(2N)^{-1}\beta}_{N(r-N)},\underbrace{\alpha,\alpha,\ldots,\alpha}_{\binom{r-N}{2}}\Big).$$

From the requirement that $T(D)\in\mathcal{P}^2$ we obtain immediately (3.1a), (3.1b), and

$$\alpha + N^{-1}\beta + \binom{N}{2}^{-1}\gamma \geq 0.$$

Combining this inequality with (2.9) we obtain (3.1c).

Next let $D = D^2(g^N)$ where g is an extreme geminal (cf. Coleman[1,3]). Then

$$T(D) = \left(\alpha + \frac{\beta}{r}\right)A^2 + \gamma D.$$

Since

$$(g|Dg) = \frac{r - N + 2}{r(N-1)}$$

[cf. Coleman,[4] formula (4.5)], we obtain

$$0 \leq (g|T(D)g) = \alpha + \frac{\beta}{r} + \frac{r - N + 2}{r(N-1)}\gamma.$$

Combining this inequality with (2.9) we obtain (3.1d).

The four extreme points of the convex set bound by the four inequalities of (3.1) are obtained by intersecting the four lines:

$$\alpha = 0, \tag{3.3a}$$

$$\beta = -2N\alpha, \tag{3.3b}$$

$$(r + N - 1)\alpha + \beta = \frac{2}{r - N}, \tag{3.3c}$$

$$r\alpha + \beta = \frac{2(r - N + 2)}{(r+1)(r-N)}. \tag{3.3d}$$

The identity Id$(\alpha = \beta = 0)$ is obtained by intersecting (3.3a) and (3.3b); Q_N is obtained by intersecting (3.3b) and (3.3c). B_N is the point of intersection of (3.3c) and (3.3d), and finally C_N is the point of intersection of (3.3d) and (3.3a).

The remainder of the proof consists in showing that the four operator endomorphisms of (3.1) map $\mathcal{P}_N^2$ into $\mathcal{P}^2$. This is trivial for the identity and it is well known for Q_N and B_N (cf. Coleman[5]). In fact we have the following lemma.

3.2 *Lemma*: The map Q_N defined by (3.2b) induces a bijection between $\mathcal{P}_N^2$ and $\mathcal{P}_{r-N}^2$.

Proof: Let Δ be the particle-hole (or Poincaré) isomorphism, i.e., the map of $\oplus_{p=0}^{r} S^p$ into itself whose defining equation is

$$\langle \Delta(D) | B \rangle = \binom{r}{p}\operatorname{tr}(D \wedge B), \quad D \in S^p, \; B \in S^{r-p}, \tag{3.4}$$

where $D \wedge B = A^r(D \otimes B)A^r$. We assert that

$$Q_N(D) = (L_{r-N}^2 \cdot \Delta)(D^N), \quad D \in S^2, \tag{3.5}$$

where $D^N \in S^N$ is any preimage of D relative to the $(N, 2)$ contraction, i.e., $D = L_2^N(D^N)$. Indeed for arbitrary $B \in S^2$ we have

$$\langle L_{r-N}^2 \cdot \Delta(D^N) | B \rangle = \langle \Delta(D^N) | \Gamma_2^{r-N}(B) \rangle$$

$$= \binom{r}{N}\operatorname{tr}(D^N \wedge B \wedge I^{r-N-2}) = \binom{r}{N},$$

$$\langle \Gamma_{N+2}^r (D^N \wedge B) | I^r \rangle = \binom{r}{N}\binom{r}{N+2}^{-1}\operatorname{tr}(D^N \wedge B)$$

$$= \binom{r-N}{2}^{-1}\binom{N+2}{2}\operatorname{tr}[A^{N+2}(D^N \otimes B)].$$

Now using Sasaki's formula,[2]

$$\binom{N+2}{2}A^{N+2} = A^N \otimes A^2\left[I^{N+2} - 2N(13) + \binom{N}{2}(13)(24)\right]A_N \otimes A^2,$$

we obtain

$$\langle L_{r-N}^2 \cdot \Delta(D^N) | B \rangle = \binom{r-N}{2}^{-1}\Big[\operatorname{tr}(D)\operatorname{tr}(B)$$

$$- 2N\operatorname{tr}[L_2^1(D)L_2^1(B)] + \binom{N}{2}\operatorname{tr}(DB)\Big]$$

$$= \binom{r-N}{2}^{-1}\Big[\operatorname{tr}(D)\langle A^2 | B \rangle - 2D\langle D^1 \wedge I^1 | B \rangle$$

$$+ \binom{N}{2}\langle D | B \rangle\Big]$$

$$= \langle Q_N(D) | B \rangle.$$

B being arbitrary this establishes formula (3.5). Formula (3.5) shows that if $D \in \mathcal{P}_N^2$, then $Q_N(D) \in \mathcal{P}_{r-N}^2$ since in this case the preimage D^N can be chosen to belong to $\mathcal{P}^N$.

From (2.9) we obtain that the eigenvalues of Q_N are given by

$$\lambda = -\frac{N}{r-N}, \quad \mu = \binom{r-N}{2}^{-1}\binom{N}{2}. \tag{3.6}$$

Thus for $2 \leq N \leq r - 2$, Q_N is invertible and by formula (3.6)

$$Q_N^{-1} = Q_{r-N}. \tag{3.7}$$

The last formula shows that Q_N^{-1} maps $\mathcal{P}_{r-N}^2$ into $\mathcal{P}_N^2$ and therefore Q_N induces a bijection between $\mathcal{P}_N^2$ and $\mathcal{P}_{r-N}^2$.

Remarks: (1) The second quantization version of this lemma has been proven by Erdahl.[6]

(2) As a corollary of this lemma we obtain the known fact that for $N = r - 2$, $Q_N(D) \geq 0$ is not only a necessary but also a sufficient condition for N representability. Indeed if $N = r - 2$, $Q_N(D) \in \mathcal{P}^2$ is equivalent to $D \in Q_2(\mathcal{P}^2) = \mathcal{P}_{r-2}^2$.

In terms of the eigenvalues (λ, μ) we have

$$Q_N = \left(-\frac{N}{r-N}, \binom{r-N}{2}^{-1}\binom{N}{2}\right). \tag{3.8}$$

Similarly with the help of (2.8) we obtain

$$\mathrm{Id} = (1, 1), \tag{3.9a}$$

$$B_N = \left(-\frac{(r+2)N - 2r}{2(r+1)(r-N)}, -\frac{2(N-1)}{(r+1)(r-N)}\right), \tag{3.9b}$$

$$C_N = \left(\frac{r^2 - (r+2)N}{2(r+1)(r-N)}, \frac{-2(N-1)}{(r+1)(r-N)}\right). \tag{3.9c}$$

From this representation it is immediate that

$$C_N = B_{r-N} \cdot Q_N \tag{3.10}$$

and since B_{r-N} maps $\mathcal{P}_{r-N}^2$ into $\mathcal{P}^2$ (cf. Coleman[5]) we also conclude that C_N induces a necessary condition for N representability. This completes the proof of Theorem 3.1.

Remark: If r or N is odd the proof of inequality (3.1d) fails and therefore it is doubtful that in this case B_N and C_N are extreme in $\mathcal{N}_N$. No such doubt is possible for the operator endomorphisms Id and Q_N.

3.3 *Corollary of Theorem* 3.1: Let r be an even number not smaller than 4.

(i) A necessary and sufficient condition for $T(\alpha,\beta) \in \mathcal{A}_0$ being order preserving is that α,β satisfy the inequalities

$$\alpha \geq 0, \tag{3.11a}$$

$$\beta \geq -4\alpha, \tag{3.11b}$$

$$(r+1)\alpha+\beta \leq \frac{2}{r-2}, \tag{3.11c}$$

$$r\alpha+\beta \leq \frac{2r}{(r+1)(r-2)}. \tag{3.11d}$$

(ii) The semigroup $\mathcal{M}$ of all order preserving operator endomorphisms considered as a convex set is the convex hull of the following four extreme points:

$$\mathrm{Id}(D)=D, \tag{3.12a}$$

$$Q_2(D)=\binom{r-2}{2}^{-1}\left[\mathrm{tr}(D)A^2-4(D^1\wedge I^1)+\binom{r}{2}D\right], \tag{3.12b}$$

$$B_2(D)=\frac{2}{(r-2)(r+1)}[\mathrm{tr}(D)A^2-D], \tag{3.12c}$$

$$C_2(D)=\frac{2}{(r-2)(r+1)}[r(D^1\wedge I^1)-D]. \tag{3.12d}$$

Proof: Since $\mathcal{M}=\mathcal{N}_2$ we obtain the result by putting $N=2$ everywhere in Theorem 3.1.

Notice that Q_2 maps $\mathcal{P}^2$ onto $\mathcal{P}^2_{r-2}\subseteq\mathcal{P}^2_N$ and therefore for $2\leq N\leq(r-2)$, Q_2 induces a sufficient condition for N representability. In Part II of this paper we shall attempt a systematic analysis of the set $\mathcal{S}_N$ of all operator endomorphisms inducing sufficient conditions for N representability. In passing let us note the following interesting formula:

$$C_NQ_2=\frac{r-N-2}{2(r-N)}Q_2+\frac{r-N+2}{2(r-N)}B_2 \tag{3.13}$$

TABLE I. Lower bounds induced by some members of $\mathcal{N}_N$ to the ground state energy (in a.u.) of He_2 in an elliptical orbital basis[5] of cardinality 8.

Condition	λ	μ	Energy (a.u.)
Id	1	1	−7.047
Q_N	−1	1	−8.887
B_N	$-\frac{1}{3}$	$-\frac{1}{6}$	−31.639
C_N	$\frac{1}{3}$	$-\frac{1}{6}$	−35.370
$\frac{1}{2}(\mathrm{Id}+Q_N)=X_1$	0	1	$-\infty$
$\frac{1}{2}(\mathrm{Id}+B_N)=X_2$	$\frac{1}{3}$	$\frac{5}{12}$	−20.606
$\frac{1}{2}(\mathrm{Id}+C_N)=X_3$	$\frac{2}{3}$	$\frac{5}{12}$	−10.686
$\frac{1}{2}(Q_N+B_N)=X_4$	$-\frac{2}{3}$	$\frac{5}{12}$	−13.090
$\frac{1}{2}(Q_N+C_N)=X_5$	$-\frac{1}{3}$	$\frac{5}{12}$	−26.165
$\frac{1}{2}(B_N+C_N)=X_6$	0	$-\frac{1}{6}$	$-\infty$
Y_1	$\frac{3}{14}$	$\frac{1}{2}$	−32.532
Y_2	$-\frac{3}{14}$	$\frac{1}{2}$	−40.930
Y_3	$-\frac{3}{14}$	$-\frac{1}{12}$	−53.205
Y_4	$\frac{3}{14}$	$-\frac{1}{12}$	−62.599

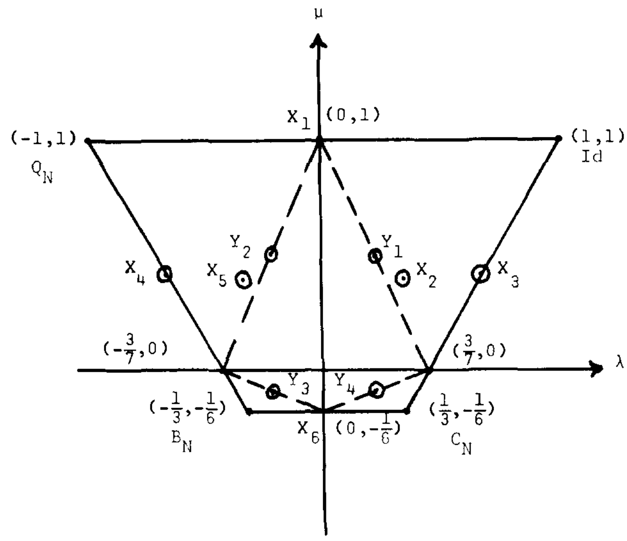

FIG. 1. $\mathcal{N}_N$ for $N=4$ and $r=8$.

which can be easily proved by using the representation of the operator endomorphisms concerned as 2×2 diagonal matrices.

4. SOME NUMERICAL RESULTS

In Table I, we present some lower bounds to the ground state energy for the system of electrons in He_2, assuming an internuclear distance of 1.00 a.u. and using an orbital basis[7] of cardinality $r=8$. Each lower bound corresponds to imposition of the indicated necessary condition for N representability. Each necessary condition is of the operator endomorphism form $(D_0+\lambda D_1+\mu D_2)\geq 0$ and is therefore a member of the set $\mathcal{N}_N$. The set $\mathcal{N}_N$ for $r=8$ and $N=4$ is depicted in Fig. 1.

We are interested in exploring the question as to which of the operator endomorphisms belonging to $\mathcal{N}_N$ will induce the necessary conditions yielding the best lower bound. From Table I it is evident that for the points $\mathcal{N}_N$ tested, Id gives the highest lower bound, namely −7.047 a.u. as compared with the value −4.631 obtained by a complete configuration interaction with the same basis set.

In order to understand this result observe that all *singular* operator endomorphisms (i.e., those for which $\lambda=0$ or $\mu=0$) yield the trivial lower bound $-\infty$. Indeed we have the theorem.

4.1 *Theorem*: If T is singular, then

$$\inf_{D\in T^{-1}(\mathcal{P}^2)}\langle K|D\rangle=-\infty,$$

where D varies over the *full* preimage of $\mathcal{P}^2$ under T.

Proof: Let $K=K_0+K_1+K_2$ be the decomposition of K into its irreducible constituents and let us assume that the orbital space $\mathcal{H}$ is chosen in such a way that $K_1\neq 0$ and $K_2\neq 0$. (This condition is satisfied for our choice of the orbital basis.)

Now let us assume that the operator endomorphism is such that, e.g., $\lambda=0$. Choose an eigenprojection P

of K_1 belonging to a strictly negative eigenvalue ξ. [Since $K_1 \neq 0$ and $\mathrm{tr}(K_1)=0$, such an eigenprojection exists.] Then $\langle K_1 | P_1\rangle = \langle K_1 | P\rangle = \xi < 0$. Now for real η define $D(\eta) \triangleq P_0 + \eta P_1$. Then for all η we have $D(\eta) \in S^2$ and $T(D(\eta)) = P_0 \in \mathcal{P}^2$; therefore $D(\eta) \in T^{-1}(\mathcal{P}^2)$, but

$$\lim_{\eta\to\infty}\langle K | D(\eta)\rangle = \langle K_0 | P_0\rangle + \lim_{\eta\to\infty}\eta\xi = -\infty.$$

Moreover, we have the following proposition.

4.2 *Proposition*: Let $f: \mathcal{N}_N \to R$ be the function defined by

$$f(T) \triangleq \inf_{D\in T^{-1}(\mathcal{P}^2)} \langle K | D\rangle \leq E_0 \text{ for all } T \in \mathcal{N}_N.$$

In case T is nonsingular we also may write

$$f(T) = \text{lowest eigenvalue of } T^{-1}(K).$$

Let T_0 be a singular operator endomorphism, i.e., $\lambda_0 = 0$ or $\mu_0 = 0$. Then

$$\lim_{T\to T_0} f(T) = f(T_0) = -\infty.$$

Proof: Suppose $T = (\lambda, \mu)$. Then $T^{-1}(K) = K_0 + \lambda^{-1}K_1 + \mu^{-1}K_2$. Let $\hat{K} \triangleq K_1 + K_2$ and let

$$\eta_1(T) \leq \eta_2(T) \leq \cdots \leq \eta_\rho(T)$$

be the eigenvalues of $T^{-1}(\hat{K})$. Since

$$\mathrm{tr}[T^{-1}(\hat{K})] = \mathrm{tr}(\hat{K}) = 0$$

we have

$$\sum_{i=1}^{\rho} \eta_i(T) = 0 \quad \text{for all nonsingular } T \in \mathcal{N}_N. \tag{4.1}$$

Moreover, we obtain for the sum of the squares

$$\sum_{i=1}^{\rho} \eta_i^2(T) = \lambda^{-2}\|K_1\|^2 + \mu^{-2}\|K_2\|^2,$$

where $\|K_i\|$ stands for the Hilbert–Schmidt norm of K_i, $\|K_i\| \triangleq \mathrm{tr}(K_i^2)^{1/2}$.

Since $K_i \neq 0$, $i = 1, 2$, it follows that

$$\lim_{T\to T_0} \eta_1(T) = -\infty \tag{4.2}$$

and thus

$$\lim_{T\to T_0} f(T) = -\infty,$$

since $f(T) = \eta_1(T) + \binom{r}{2}^{-1}\mathrm{tr}(K)$.

From Proposition 4.2 it is clear that $f(T)$ will be the better a lower bound for E_0 the "farther away" $T \in \mathcal{N}_N$ is from the cross-shaped figure representing the singular operator endomorphisms. This leads to the conjecture that the best lower bound will be always attained at one of the four vertices of $\mathcal{N}_N$ and most likely at either Id or Q_N, since in terms of the Euclidean distance B_N and C_N will always be closer to the set of singular operator endomorphisms than Id and Q_N. For the same reason we conjecture that of the two values $f(\mathrm{Id})$ and $f(Q_N)$, $f(\mathrm{Id})$ will be the better lower bound if $N < r/2$ and $f(Q_N)$ will be the better lower bound if $N > r/2$. The results of the minimum Slater basis[8] calculations on LiH ($N=4$, $r=12$) and H_2O ($N=10$, $r=14$) as reported in Table II are in agreement with these conjectures.

TABLE II. Lower bounds $f(\mathrm{Id})$ and $f(Q_N)$ (in a.u.) to the ground state energy E_0 of LiH ($N=4$, $r=12$) and H_2O ($N=10$, $r=14$).

Condition T	lower bound $f(T)$ for LiH	lower bound $f(T)$ for H_2O
Id	−9.531	−156.562
Q_N	−10.309	−75.509
B_H	−75.324	−370.200
C_N	−132.474	−1107.291
Complete CI	−7.887	−75.013

ACKNOWLEDGMENT

This work has been supported by N.R.C. Canada (N.R.C. Grants 2990 and 9233). We would like to express our thanks to N.R.C. for this support.

[1]By the compression of a linear operator H on a Hilbert space $\mathcal{H}$ to a subspace $\mathcal{K} \subseteq \mathcal{H}$ we mean the restriction of PH to $\mathcal{K}$ where $P: \mathcal{H} \to \mathcal{K}$ is the orthogonal projection from $\mathcal{H}$ onto $\mathcal{K}$. Cf. P.R. Halmos; *A Hilbert Space Problem Book* (Van Nostrand, Princeton, N.J., 1967), p. 118.

[2]A.J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[3]H. Kummer, J. Math. Phys. **8**, 2063 (1967).
[4]A.J. Coleman, J. Math. Phys. **6**, 1425 (1965).
[5]A.J. Coleman, J. Math. Phys. **13**, 214 (1972).
[6]R.M. Erdahl, J. Math. Phys. 13, 1608 (1972).
[7]G.P. Barnett, Can. J. Phys. **45**, 137 (1967).
[8]W.J. Hehre, W.A. Lathan, R. Ditchfield, M.D. Newton, and J.A. Pople, Gaussian 70, Quantum Chemistry Program Exchange (Indiana University, Bloomington, Indiana).

Some aspects of the *N*-representability problem in finite dimensions. II. Operator endomorphisms which induce sufficient conditions

H. Kummer and I. Absar

Department of Mathematics, Queen's University, Kingston, Canada K7L 3N6
(Received 4 June 1976)

Given the set $\mathcal{S}_N$ of all operator endomorphisms T with the property that whenever D is any 2-density operator, then $T(D)$ is N-representable is considered. We specify a pair $\{Q_2, P_N\}$ of extreme points of $\mathcal{S}_N$ which are related to each other via the particle-hole duality (Theorem 1.2). The rest of the paper is devoted to a complete characterization of the set $\hat{\mathcal{S}}_N$ of all operator endomorphisms T with the property that whenever D is any 2-density operator then $T(D)$ is *quasi*-N-representable, i.e., satisfies the necessary conditions Id, Q_N, B_N, and C_N.

I. INTRODUCTION

This paper (SANP II) is devoted to the characterization of the convex set $\mathcal{S}_N$ of all operator endomorphisms which induce sufficient conditions for N representability. For definitions and notation we ask the reader to consult Secs. 1 and 2 of SANP I,[1] and the references[2,3] therein.

For any $T \in \mathcal{S}_N$ define

$$E(T) \triangleq \inf_{D\in\ ^2} \langle K | T(D)\rangle, \tag{1.1}$$

where K is the truncated reduced Hamiltonian. It follows from Lemma 1.1 of SANP I that $E(T)$ is an upper bound to the ground state energy E_0,

$$E_0 \leqslant E(T) \quad \text{for all } T \in \mathcal{S}_N. \tag{1.2}$$

It is easy to see that $E(T)$ is a concave functional on $\mathcal{S}_N$, i.e.,

$$E(\delta T_1 + (1-\delta)T_2) \geqslant \delta E(T_1) + (1-\delta)E(T_2)$$
$$\text{for } 0 \leqslant \delta \leqslant 1. \tag{1.3}$$

Clearly such a functional attains its minimum on an *extreme* point of $\mathcal{S}_N$. This fact gives us enough motivations to be interested in the set of all extreme points of $\mathcal{S}_N$. Unfortunately we were unable to give a complete characterization of this set. The situation seems to be more complicated than in the case of $\mathcal{N}_N$. However, observe that the family $(\mathcal{S}_N)_{N=2,\ldots,r-2}$ of convex subsets of $\mathcal{A}_0$ (the set of all operator endomorphisms which preserve the trace, c.f. SANP I,[1] Sec. 2) has the following remarkable property:

$$\mathcal{S}_N = Q_N^{-1}(\mathcal{S}_{r-N}) = \{Q_N^{-1}T \mid T \in \mathcal{S}_{r-N}\}, \tag{1.4}$$

i.e., the multiplication by Q_N^{-1} induces a bijection between $\mathcal{S}_{r-N}$ and $\mathcal{S}_N$. We say the family $(\mathcal{S}_N)$ is *invariant under the particle-hole duality*. For such a family of convex subsets of $\mathcal{A}_0$ the following principle holds:

1.1: *Principle of the Particle-Hole Duality*: Let $(\mathcal{C}_N)$ be a family of convex subsets of $\mathcal{A}_0$ which is invariant under the particle-hole duality, i.e.,

$$\mathcal{C}_N = Q_N^{-1}(\mathcal{C}_{r-N}).$$

Then

(i) If $T_N \in \mathcal{C}_N$ for $a \leqslant N \leqslant b$, then $\tilde{T}_N \triangleq Q_N^{-1}T_{r-N} \in \mathcal{C}_N$ for $r-b \leqslant N \leqslant r-a$.

(ii) If T_N is extreme in $\mathcal{C}_N$ for $a \leqslant N \leqslant b$, then $\tilde{T}_N$ is extreme in $\mathcal{C}_N$ for $r-b \leqslant N \leqslant r-a$.

Proof: (i) Suppose $T_N \in \mathcal{C}_N$ for $a \leqslant N \leqslant b$. Then $T_{r-N} \in \mathcal{C}_{r-N}$ for $r-b \leqslant N \leqslant r-a$. But then we obtain by multiplication by Q_N^{-1} that $\tilde{T}_N \triangleq Q_N^{-1}T_{r-N} \in \mathcal{C}_N$ for $r-b \leqslant N \leqslant r-a$.

(ii) Since the multiplication by Q_N^{-1}, $T \to Q_N^{-1}T$, induces an affine bijection between the convex sets $\mathcal{C}_{r-N}$ and $\mathcal{C}_N$ it maps extreme points onto extreme points.

At the end of SANP I we observed that $Q_2 \in \mathcal{S}_N$ for $2 \leqslant N \leqslant (r-2)$. From the principle of the particle-hole duality it follows immediately that $P_N \triangleq \tilde{Q}_2 = Q_N^{-1}Q_2 \in \mathcal{S}_N$ for $2 \leqslant N \leqslant (r-2)$. Q_2 is extreme in $\mathcal{S}_2 \supseteq \mathcal{S}_N$ and hence extreme in $\mathcal{S}_N$. The second part of the principle of the particle-hole duality implies that P_N is extreme in $\mathcal{S}_N$. Thus we have the following theorem.

1.2: *Theorem*: The following two operator endomorphisms are extreme points of $\mathcal{S}_N$:

$$Q_2(D) = \binom{r-2}{2}^{-1}[\mathrm{tr}(D)A^2 - 4(D^1 \wedge I^1) + D], \tag{1.5a}$$

$$P_N(D) = \binom{N}{2}\binom{r-2}{2}^{-1}\left[\binom{N-2}{2}\mathrm{tr}(D)A^2 + 2(N-2)(r-N)(D^1 \wedge I^1) + \binom{r-N}{2}D\right]. \tag{1.5b}$$

In terms of the eigenvalues (λ, μ) we have

$$Q_2 = \left(-\frac{2}{r-2}, \frac{2}{(r-2)(r-3)}\right), \tag{1.6a}$$

$$P_N = \left(\frac{r-N}{N}\,\frac{2}{r-2}, \frac{(r-N)(r-N-1)}{N(N-1)}\cdot\frac{2}{(r-2)(r-3)}\right). \tag{1.6b}$$

What are the other extreme points of $\mathcal{S}_N$? We were not able to settle this question. We have contented ourselves with solving a less difficult problem.

Let $D \in \mathcal{P}^2$ be a density operator. We say D is *quasi-N-representable* iff D satisfies the four necessary conditions for N representability induced by the operator endomorphisms Id, Q_N, B_N, and C_N. In view of Theorem 3.1 of SANP I this means in case of even r and even N that D satisfies all necessary conditions for N representability of the operator endomorphism form $\sigma\,\mathrm{tr}(D)A^2 + \beta(D^1 \wedge I^1) + \gamma D \geqslant 0$.

TABLE I. List of all operator endomorphisms inducing sufficient conditions for quasi-N-representability.

Name	λ	μ
Q_2	$-\frac{2}{r-2}$	$\frac{2}{(r-2)(r-3)}$
$P_N=\tilde{Q}_2$	$\frac{r-N}{N}\frac{2}{r-2}$	$\frac{(r-N)(r-N-1)}{N(N-1)}\frac{2}{(r-2)(r-3)}$
B_2	$-\frac{2}{(r+1)(r-2)}$	$-\frac{2}{(r+1)(r-2)}$
$Q_N^{-1}B_2=\tilde{B}_2$	$\frac{r-N}{N}\frac{2}{(r+1)(r-2)}$	$-\frac{(r-N)(r-N-1)}{N(N-1)}\frac{2}{(r+1)(r-2)}$
J_N	$\frac{2}{(r+1)(r-2)}\left(\frac{r-N}{N}+\frac{(r-2N)r(r-3)}{2N(r-N-1)}\right)$	$-\frac{2}{(r+1)(r-2)}$
$\tilde{J}_N$	$-\frac{2}{(r+1)(r-2)}\left(1+\frac{(2N-r)r(r-3)}{2N(N-1)}\right)$	$-\frac{(r-N)(r-N-1)}{N(N-1)}\frac{2}{(r+1)(r-2)}$
U_N	$\frac{r+4-3N}{3N-4}\frac{2}{(r-2)}$	$\frac{r(r-5)+3N}{3N-4}\frac{2}{(r-2)(r-3)}$
$\tilde{U}_N$	$\frac{2(r-2)-3N}{3(r-N)-4}\frac{r-N}{N}\frac{2}{(r-2)}$	$\frac{(r-N)(r-N-1)}{N(N-1)}\frac{r(r-2)-3N}{3(r-N)-4}\frac{2}{(r-2)(r-3)}$
$B_N^{-1}B_2$	$\frac{r-N}{(r+2)N-2r}\frac{4}{(r-2)}$	$\frac{r-N}{N-1}\frac{1}{r-2}$
$C_N^{-1}B_2=\tilde{B}_N^{-1}B_2$	$-\frac{r-N}{r^2-(r-2)N}\frac{4}{(r-2)}$	$\frac{r-N}{N-1}\frac{1}{r-2}$
V_N	$\frac{(r^2-8)-(3r-8)N}{(N-1)(r-2)(r-4)}$	$\frac{r-N}{N-1}\frac{1}{r-2}$
$\tilde{V}_N$	$\frac{r-N}{N}\frac{2(r-2)^2-3(r-8)N}{(r-N-1)(r-2)(r-4)}$	$\frac{r-N}{N-1}\frac{1}{r-2}$
$F_N=\tilde{F}_N$	$\frac{2(r-2)N}{N(r-4)}$	$\frac{2(r-N-1)}{N(r-3)}$

We say an operator endomorphism T induces a sufficient condition for quasi-N-representability iff whenever $D\in\mathcal{P}^2$, then $T(D)$ is quasi-N-representable, i.e., iff $T\in\mathcal{S}_2$, $Q_NT\in\mathcal{S}_2$, $B_NT\in\mathcal{S}_2$, and $C_NT\in\mathcal{S}_2$.

Let us denote by $\hat{\mathcal{S}}_N$ the set of all operator endomorphisms which induce a sufficient condition for quasi-N-representability. Obviously

$$\mathcal{S}_N\subseteq\hat{\mathcal{S}}_N\triangleq\mathcal{S}_2\cap Q_N^{-1}(\mathcal{S}_2)\cap B_N^{-1}(\mathcal{S}_2)\cap C_N^{-1}(\mathcal{S}_2),$$

where, e.g., $Q_N^{-1}(\mathcal{S}_2)\triangleq\{Q_N^{-1}T\mid T\in\mathcal{S}_2\}$. We are interested in the set of extreme points of $\hat{\mathcal{S}}_N$. Clearly Q_2 and $P_N=\tilde{Q}_2$ belong to this set. The remaining extreme points are operator endomorphisms occuring in the list of Table I.

2. OPERATOR ENDOMORPHISMS WHICH INDUCE SUFFICIENT CONDITIONS FOR QUASI-*N*-REPRESENTABILITY

This section is devoted to the proof of the following theorem.

2.1: *Theorem*: Let r be an even number not smaller than 8. Referring to Table I the following is a complete list of extreme points of $\hat{\mathcal{S}}_N$:

(i) for $2<N\leq r/3$: Q_2, P_N, B_2, J_N, U_N;

(ii) for $r/3\leq N\leq(r+4)/3$: $Q_2, P_N, B_2, J_N, B_N^{-1}B_2, V_N$;

(iii) for $(r+4)/3\leq N\leq r/2$: $Q_2, P_N, B_2, J_N, B_N^{-1}B_2, C_N^{-1}B_2$;

(iv) for $r/2\leq N\leq 2(r-2)/3$: $Q_2, P_N, Q_N^{-1}B_2, \tilde{J}_N, B_N^{-1}B_2, C_N^{-1}B_2$;

(v) for $2(r-2)/3\leq N\leq(2/3)r$: $Q_2, P_N, Q_N^{-1}B_2, \tilde{J}_N, C_N^{-1}B_2, \tilde{V}_N$;

(vi) for $(2/3)r\leq N<(r-2)$: $Q_2, P_N, Q_N^{-1}B_2, \tilde{J}_N, \tilde{U}_N$.

Because of the fact that $\hat{\mathcal{S}}_N=Q_N^{-1}(\hat{\mathcal{S}}_{r-N})$ we may apply the principle of the particle-hole duality, which implies that (iv), (v), and (vi) are simply the dual statements of (iii), (ii), and (i), respectively. It therefore suffices to prove statements (i)–(iii). The proof is based on Theorem A7 (see the Appendix) according to which the set of extreme points of the intersection of two closed convex bodies C_1 and C_2 is given by

$$\text{ext}(C_1\cap C_2)=[(\text{ext}C_1)\cap C_2]\cup[(\text{ext}C_2)\cap C_1]\cup\text{ext}(\partial C_1\cap\partial C_2).$$

Here ∂C denotes the *boundary* of the set C. Now

$$\hat{\mathcal{S}}_N=\mathcal{Q}_N\cap\mathcal{B}_N$$

where

$$\mathcal{Q}_N\triangleq\mathcal{S}_2\cap Q_N^{-1}(\mathcal{S}_2)$$

and

$$\mathcal{B}_N \triangleq B_N^{-1}(\mathcal{S}_2) \cap C_N^{-1}(\mathcal{S}_2).$$

We therefore have to determine the extreme points of $\mathcal{Q}_N$ and $\mathcal{B}_N$. The following proposition gives a complete description of the extreme points of $\mathcal{Q}_N$.

2.2: *Proposition*: Let r be an even number not smaller than 6. Referring to Table I, the following is a complete list of the extreme points of $\mathcal{Q}_N \triangleq \mathcal{S}_2 \cap Q_N^{-1}(\mathcal{S}_2)$:

(i) for $2 < N \leq r/2$: Q_2, P_N, B_2, J_N, F_N;

(ii) for $r/2 \leq N < (r-2)$: $Q_2, P_N, Q_N^{-1}B_2, \tilde{J}_N, F_N$.

Since $\mathcal{Q}_N = Q_N^{-1}(\mathcal{Q}_{r-N})$ the principle of the particle-hole duality again applies and therefore it suffices to prove statement (i), (ii) being the dual statement. For the proof of (i) we need

2.3: *Lemma*: In terms of the eigenvalues λ, μ, $\mathcal{S}_2$ is characterized by the following inequalities:

$$2(r-1)\lambda - r\mu \leq (r-2), \tag{2.1a}$$

$$-(r-1)(r-4)\lambda + r(r-3)\mu \leq 2(r-2), \tag{2.1b}$$

$$2(r-1)\lambda + r(r-3)\mu \geq -2, \tag{2.1c}$$

$$(r+1)(r-2)\mu \geq -2. \tag{2.1d}$$

Proof: Substitute the expression for α and β given by formulas (2.5) of Part I into the inequalities (3.10) of Part I.

This lemma allows us to determine the two sets $\text{ext}(\mathcal{S}_2) \cap Q_N^{-1}(\mathcal{S}_2)$ and $\text{ext}[Q_N^{-1}(\mathcal{S}_2)] \cap \mathcal{S}_2$.

2.4: *Lemma*: Let r be an even number not smaller than 6 and let $2 < N \leq r/2$. Then

(i) $\text{ext}(\mathcal{S}_2) \cap Q_N^{-1}(\mathcal{S}_2) = \{Q_2, B_2\}$,

(ii) $\text{ext}(Q_N^{-1}\mathcal{S}_2) \cap \mathcal{S}_2 = \{P_N\}$.

Proof: With the help of Lemma 2.3 we shall first show that for $2 < N < (r-2)$, $Q_N^{-1} \notin \mathcal{S}_2$ and $Q_N^{-1}C_2 \notin \mathcal{S}_2$. Replacing N by $(r-N)$, this will imply that $\text{Id} \notin Q_N^{-1}(\mathcal{S}_2)$ and $Q_N^{-1}C_2 \notin \mathcal{S}_2$. Next we shall show that $Q_N^{-1}B_2 \in \mathcal{S}_2$ iff $r/2 \leq N < r-2$ and therefore (replacing N by $r-N$) that $B_2 \in Q_N^{-1}(\mathcal{S}_2)$ iff $2 < N \leq r/2$.

Since $P_N \in \mathcal{S}_N \subseteq \mathcal{S}_2$ this will establish (ii). Finally since $Q_2 = Q_N^{-1} P_{r-N} \in Q_N^{-1}(\mathcal{S}_2)$ we obtain (i).

Now we have in terms of λ and μ,

$$Q_N^{-1} = \left(-\frac{r-N}{N}, \frac{(r-N)(r-N-1)}{N(N-1)}\right).$$

Inserting the coordinates of Q_N^{-1} into (2.1b) we obtain, after some algebraic manipulations, $N \geq r-2$. It follows that for $N \leq r-3$, (2.1b) does not hold and therefore $Q_N^{-1} \notin \mathcal{S}_2$. Similarly inserting the coordinates of

$$Q_N^{-1}C_2 = -\frac{2}{(r+1)(r-2)} \frac{r-N}{N}\left(\frac{r^2-2r-4}{4}, \frac{r-N-1}{N-1}\right)$$

into (2.1c) we obtain $(N+1)(N-r+2) \geq 0$, an inequality which fails to hold for $2 < N < (r-2)$. Thus $Q_N^{-1}\mathcal{S}_2 \notin \mathcal{S}_2$.

Next we have

$$Q_N^{-1}B_2 = \frac{2}{(r+1)(r-2)}\left(\frac{r-N}{N}, -\frac{(r-N)(r-N-1)}{N(N-1)}\right).$$

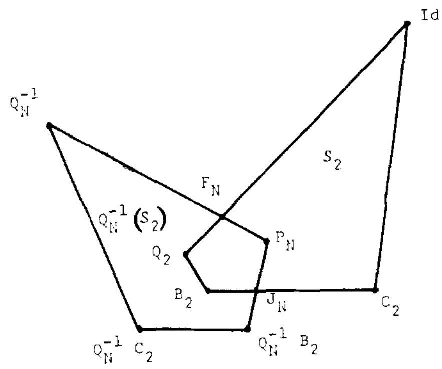

FIG. 1.

Inserting the coordinates of $Q_N^{-1}B_2$ into (2.1a) leads to $N(N-1) \geq 2$ which for $N \geq 2$ is obviously satisfied. Similarly, substitution into (2.1b) leads to $(N-1)^2 \geq (r-3)$ which for $r \geq 3$ is trivially satisfied. Insertion of the coordinates of $Q_N^{-1}B_2$ into (2.1c) and (2.1d) leads to $N \geq (r-1)/2$ and $N \geq r/2$, respectively. Thus $Q_N^{-1}B_2 \in \mathcal{S}_2$ iff $r/2 \leq N < r-2$.

2.5: *Lemma*: Let r be an even number not smaller than 8. Then referring to Table I we have the following:

for $N=3$: $\partial\mathcal{S}_2 \cap \partial Q_N^{-1}(\mathcal{S}_2) = \{F_N, J_N, P_N\}$;

for $3 < N < r/2$: $\partial\mathcal{S}_2 \cap \partial Q_N^{-1}(\mathcal{S}_2) \cap \partial Q_N^{-1}(\mathcal{S}_2) = \{F_N, J_N\}$;

for $N = r/2$: $\partial\mathcal{S}_2 \cap \partial Q_N^{-1}(\mathcal{S}_2) = \{F_N\} \cup B_2 | Q_N^{-1}B_2$.

Here $B_2 | Q_N^{-1}B_2$ stands for the closed segment determined by B_2 and $Q_N^{-1}B_2$. [If $r=6$ and $N=3$, then $\partial\mathcal{S}_2 \cap \partial Q_N^{-1}(\mathcal{S}_2) = \{F_N, P_N\} \cup B_2 | Q_N^{-1}B_2$.] The situation for $3 < N < r/2$ is represented by Fig. 1.

Proof: $\partial\mathcal{S}_2$ is the union of the four closed segments $\text{Id}|Q_2$, $Q_2|B_2$, $B_2|C_2$, and $C_2|\text{Id}$, whereas $\partial Q_N^{-1}(\mathcal{S}_2)$ comprises the four segments $Q_N^{-1}|P_N$, $P_N|Q_N^{-1}B_2$, $Q_N^{-1}B_2|Q_N^{-1}C_2$, and $Q_N^{-1}C_2|Q_N^{-1}$. For $2 < N < (r-2)$, $Q_N^{-1}|P_N$ and $\text{Id}|Q_2$ intersect at the point

$$F_N \triangleq \left(\frac{2}{N}\ \frac{r-2N}{r-4}, \frac{r-N-1}{r-3}\right) = \frac{2(r-N-2)}{N(r-4)}\text{Id} + \frac{(N-2)(r-2)}{N(r-4)}Q_2.$$

Furthermore, for $2 < N \leq r/2$, $P_N | Q_N^{-1}B_2$ and $B_2|C_2$ intersect at

$$J_N = \frac{2}{(r+1)(r-2)}\left(\frac{r-N}{N} + \frac{r(r-3)(r-2N)}{2N(r-N-1)}, -1\right)$$
$$= \frac{2(r-2N+1)}{N(r-N-1)}C_2 + \frac{(N-2)(r-N+1)}{N(r-N-1)}B_2.$$

For $3 < N < r/2$, no other pair of segments from $\partial\mathcal{S}_2$ and $\partial Q_N^{-1}(\mathcal{S}_2)$ respectively intersect so that in this case

$$\partial\mathcal{S}_2 \cap \partial Q_N^{-1}(\mathcal{S}_2) = \{F_N, J_N\}.$$

In the case $N=3$, the point P_N touches the segment $C_2|\text{Id}$ and finally for $N=r/2$, $Q_N^{-1}B_2$ touches the segment $B_2|C_2$ so that the whole closed segment $B_2|Q_N^{-1}B_2$ is a subset of the intersection of the two boundaries.

Combination of the two foregoing lemmas with Theorem A7 establishes statement (i) of Proposition 2.2, statement (ii) being a consequence of the extreme points of $\mathcal{B}_N \triangleq B_N^{-1}(\mathcal{S}_2) \cap C_N^{-1}(\mathcal{S}_2)$.

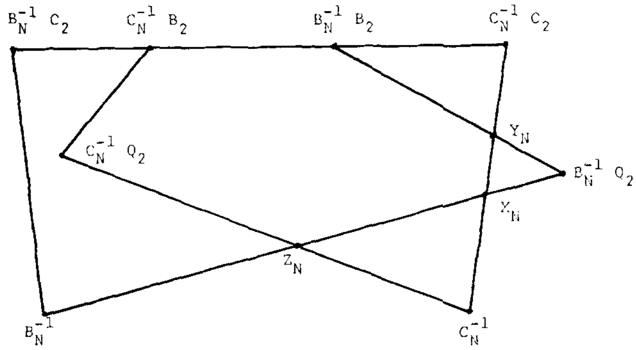

FIG. 2(a). $2(r+2)<(r+2)N\leq 6(r-2)$.

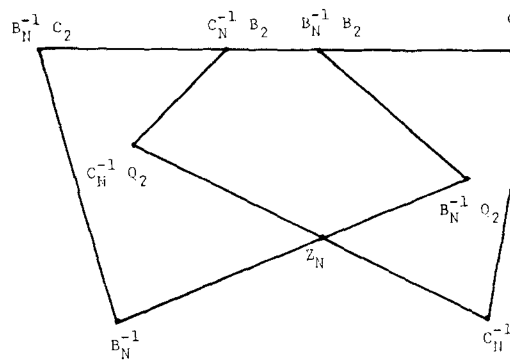

FIG. 2(b). $6(r-2)\leq(r+2)N<r^2-4r+12$.

2.6: *Proposition*: Let r be an even number not smaller than 6. Then the extreme points of $\mathcal{B}_N$ are given by the following:

(i) for $2(r+2)<(r+2)N\leq 6(r-2)$:
$\{B_N^{-1}B_2, C_N^{-1}B_2, C_N^{-1}Q_2, Z_N, X_N, Y_N\}$;

(ii) for $6(r-2)\leq(r+2)N\leq r^2-4r+12$:
$\{B_N^{-1}B_2, C_N^{-1}B_2, B_N^{-1}Q_2, C_N^{-1}Q_2, Z_N\}$

(iii) for $r^2-4r+12\leq(r+2)N<r^2-4$:
$\{B_N^{-1}B_2, C_N^{-1}B_2, B_N^{-1}Q_2, \tilde{X}_N, \tilde{Y}_N, Z_N\}$.

Here,

$$Z_N \triangleq \left(0, -\frac{(r+1)(r-2)(r-N)}{r(r-3)(N-1)}\right),$$

$$X_N \triangleq \left(\frac{2(r+1)(r-N)}{2r(r-2)-(r+2)N}, \frac{(r+1)(r-N)[(r+2)N-2(3N-4)]}{2(N-1)(2r(r-2)-(r+2)N)}\right),$$

$$Y_N \triangleq \left(\frac{(r+1)(r-N)(r-4)}{(r-2)[r(r-1)-(r+2)N]}, \frac{(r+1)(r-N)[2(r-2)-N]}{2(r-2)(N-1)[r(r-1)-(r+2)N]}\right).$$

Note that since $\mathcal{B}_N$ is invariant under the particle hole-duality, the principle of the particle-hole duality applies. Indeed it suffices to prove statements (i) and (ii), since (iii) is the dual of statement (i). Again the proof is based on Theorem A7.

2.7: *Lemma*:

(i) for $2(r+2)<(r+2)N\leq 6(r-2)$ we have:

$\text{ext}[B_N^{-1}(\mathcal{S}_2)]\cap C_N^{-1}(\mathcal{S}_2)=\{B_N^{-1}B_2\}$,

$\text{ext}[C_N^{-1}(\mathcal{S}_2)]\cap B_N^{-1}(\mathcal{S}_2)=\{C_N^{-1}C_2, C_N^{-1}Q_2\}$;

(ii) for $6(r-2)\leq(r+2)N\leq r^2-4r+12$ we have:

$\text{ext}[B_N^{-1}(\mathcal{S}_2)]\cap C_N^{-1}(\mathcal{S}_2)=\{B_N^{-1}B_2, B_N^{-1}Q_2\}$;

$\text{ext}[C_N^{-1}(\mathcal{S}_2)]\cap B_N^{-1}(\mathcal{S}_2)=\{C_N^{-1}C_2, C_N^{-1}Q_2\}$.

Proof: First we show that for $2<N<(r-2)$ we have $B_N^{-1}\notin C_N^{-1}(\mathcal{S}_2)$, $B_N^{-1}C_2\notin C_N^{-1}(\mathcal{S}_2)$, and $B_N^{-1}B_2\notin C_N^{-1}(\mathcal{S}_2)$. Via the particle-hole duality this implies: $C_N^{-1}\notin B_N^{-1}(\mathcal{S}_2)$, $C_N^{-1}C_2\notin B_N^{-1}(\mathcal{S}_2)$, and $C_N^{-1}B_2\in B_N^{-1}(\mathcal{S}_2)$.

Finally we have to show that $B_N^{-1}Q_2\in C_N^{-1}(\mathcal{S}_2)$ iff $6(r-2)\leq(r+2)N$ and thus by duality $C_N^{-1}Q_2\in B_N^{-1}(\mathcal{S}_2)$ iff $(r+2)N\leq r^2-4r=12$. To show that $B_N^{-1}\notin C_N^{-1}(\mathcal{S}_2)$ is equivalent to showing that

$$B_N^{-1}C_N=\left(-\frac{r^2-(r+2)N}{(r+2)N-2r}, 1\right)$$

does not belong to $\mathcal{S}_2$.

Substitution of the coordinates into (2.1b) leads to $(r-1)(r-2)(r-4)\leq 0$, an inequality which is violated for $r\geq 6$. In order to show that $B_N^{-1}C_2\notin C_N^{-1}(\mathcal{S}_2)$ it suffices to verify that

$$B_N^{-1}C_NC_2=-\frac{2}{(r+1)(r-2)}\times\left(\frac{r^2-(r+2)N}{(r+2)N-2r}\frac{r^2-2r-4}{4}, 1\right)$$

does not belong to $\mathcal{S}_2$. Indeed substitution of its coordinates into (2.1c) gives $N\geq r-2$.

Next we have to show that

$$B_N^{-1}C_NB_2=\frac{2}{(r+1)(r-2)}\left(\frac{r^2-(r+2)N}{(r+2)N-2r}, -1\right)$$

belongs to $\mathcal{S}_2$. Substitution of the coordinates into (2.1a) leads to $N\geq 2$. Similarly, substitution of the coordinates into (2.1b) leads to $N\geq 1+2/(r+2)$, an inequality which for $N\geq 2$ certainly holds. Insertion into (2.1c) gives $r(r-1)\geq 0$, and (2.1d) is obviously satisfied.

Finally we have to show that

$$C_NB_N^{-1}Q_2=\frac{2}{(r-2)}\left(\frac{r^2-(r+2)N}{(r+2)N-2r}, \frac{1}{r-3}\right)$$

belongs to $\mathcal{S}_2$ iff $6(r-2)\leq(r+2)N$.

Substitution of the coordinates into (2.1a) leads to $6(r-2)\leq(r+2)N$. Thus the above condition is necessary. That it is sufficient is a consequence of the fact that the remaining three inequalities are trivially satisfied. Indeed insertion of the coordinates into (2.1b) leads to $(r-1)(r-2)(r-4)\geq 0$, whereas insertion into (2.1c) yields $r(r-2)\geq 0$. Finally (2.1d) obviously holds.

2.8: *Lemma*: Let r be an even number not smaller than 6. We have the following:

(i) for $2(r+2)<(r+2)N\leq 6(r-2)$:
$\partial B_N^{-1}(\mathcal{S}_2)\cap\partial C_N^{-1}(\mathcal{S}_2)=B_N^{-1}B_2 | C_N^{-1}C_2\cup\{X_N, Y_N, Z_N\}$;

(ii) for $6(r-2) \leq (r+2)N < r^2 - 4r + 12$:
$\partial B_N^{-1}(\mathcal{S}_2) \cap \partial C_N^{-1}(\mathcal{S}_2) = B_N^{-1}B_2 | C_N^{-1}C_2 \cup \{Z_N\}$.

The situation is represented by Figs. 2(a) and 2(b).

Proof: The segments $B_N^{-1} | B_N^{-1}Q_2$ and $Q_N^{-1} | C_N^{-1}Q_2$ intersect for all $N = 2, \ldots, (r-2)$ at the point

$$Z_N = \frac{2}{r} B_N^{-1} + \frac{r-2}{r} B_N^{-1} Q_2.$$

For $2(r+2) < (r+2)N \leq 6(r-2)$ the segments $B_N^{-1} | B_N^{-1}Q_2$ and $C_N^{-1} | C_N^{-1}C_2$ intersect at

$$X_N = \frac{2(r-2)(3r+4) - r(r+2)N}{(2r(r-2) - (r+2)N)r} C_N^{-1} + \frac{2(r-2)(r-4)(r+1)}{(2r(r-2) - (r+2)N)r} C_N^{-1}C_2$$

Moreover, $B_N^{-1}Q_2 | B_N^{-1}B_2$ and $C_N^{-1} | C_N^{-1}C_2$ intersect at

$$Y_N = \frac{(r+2)(N-2)}{r(r(r-1) - (r+2)N)} C_N^{-1} + \frac{r^3 - r^2 + 2r + 4 - (r+1)(r+2)N}{r(r(r-1) - (r+2)N)} C_N^{-1}C_2.$$

For $(r+2)N = 6(r-2)$, X_N and Y_N coincide with $B_N^{-1}Q_2$. In this case we have

$$X_N = Y_N = B_N^{-1}Q_2 = (r(r-3))^{-1}[4C_N^{-1} + (r+1)(r-4)C_N^{-1}C_2]$$

For $N(r+2) \geq 6(r-2)$, $B_N^{-1}Q_2 \in C_N^{-1}(\mathcal{S}_2)$ and as long as $N(r+2) < r^2 - 4r + 12$, Z_N is the only common boundary point of $B_N^{-1}(\mathcal{S}_2)$ and $C_N^{-1}(\mathcal{S}_2)$ besides the segment $B_N^{-1}B_2 | C_N^{-1}C_2$.

Lemmas 2.7 and 2.8 together with Theorem A7 establish Proposition 2.6.

Finally we have to apply Theorem A7 to the intersection

$$\hat{\mathcal{S}}_N = \mathbf{Q}_N \cap \mathcal{B}_N.$$

The following lemma tells us that all extreme points of $\mathbf{Q}_N$ with the exception of F_N belong to $\mathcal{B}_N$.

2.9: *Lemma*: Let r be an even number not smaller than 6. For $2 < N < (r-2)$ we have

$$(\text{ext}\mathcal{Q}_N) \cap \mathcal{B}_N = (\text{ext}\mathcal{Q}_N) \setminus \{F_N\}.$$

Proof: Since $\mathcal{B}_N$ is invariant under the particle-hole duality it suffices to show that $F_N \notin \mathcal{B}_N$ and $Q_2, B_2 \in \mathcal{B}_N$. From the principle of the particle-hole duality we obtain immediately $P_N, Q_N^{-1}B_2 \in \mathcal{B}_N$. Since $J_N \in P_N | Q_N^{-1}B_2$ and $\mathcal{B}_N$ is convex we obtain $J_N \in \mathcal{B}_N$ and hence $\tilde{J}_N \in \mathcal{B}_N$. Next notice that

$$\mu(T) \leq \mu_0 \triangleq \frac{r-N}{r-1} \frac{1}{r-2}$$

for all $T \in \mathcal{B}_N$. Since $\mu(F_N) \leq \mu_0$ leads to $(N-2)(N-r-2) \geq 0$ we see that for $2 < N < (r-2)$, $F_N \notin \mathcal{B}_N$.

In order to prove that $B_2 \in \mathcal{B}_N$ we have to show that $B_N B_2, C_N B_2 \in \mathcal{S}_2$. Now

$$B_N B_2 = \frac{4}{(r+1)^2(r-2)} \left(\frac{(r+2)N - 2r}{4(r-N)}, \frac{N-1}{r-N} \right).$$

Inserting the coordinates of $B_N B_2$ into (2.1a) yields $N \leq [(r+2)/(r+1)](r-1)$ which is satisfied for $N \leq (r-1)$. Substituting into (2.1b) yields

$$N \leq r + \tfrac{1}{2} - 13(4r+2)^{-1},$$

which for $N \leq r-2$ (and $r \geq 1$) is satisfied. Substitution into (2.1c) leads to $N \leq (r+3)$. Finally insertion of the coordinates into (2.1d) gives $N \leq r+2$.

Similarly, we have

$$C_N B_2 = \frac{4}{(r+1)^2(r-2)} \left(-\frac{r^2 - (r+2)N}{4(r-N)}, \frac{N-1}{r-N} \right).$$

Substitution of these coordinates into (2.1a) yields $N \leq r$. Insertion of the coordinates into (2.1b) leads to $N \leq r - 2/(2r+1)$ which for $N \leq r-1$ (and $r \geq 1$) is satisfied. Substitution of the coordinates of $C_N B_2$ in (2.1c) yields $N \leq r+2$. Finally, the argument concerning (2.1d) is the same as for $B_N B_2$.

Conversely there are few extreme points of $\mathcal{B}_N$ which belong to $\mathcal{Q}_N$. We have

2.10: *Lemma*: Suppose r is an even number not smaller than 8. Then

(i) For $2 < N < (r-2)$ none of the extreme points of $\mathcal{B}_N$ with the possible exceptions of $C_N^{-1}B_2$ and $B_N^{-1}B_2$ belong to $\mathcal{Q}_N$;

(ii) $C_N^{-1}B_2 \in \mathcal{Q}_N$ iff $(r+4)/3 \leq N \leq \frac{2}{3}r$,

$B_N^{-1}B_2 \in \mathcal{Q}_N$ iff $r/3 \leq N \leq [2(r-2)]/3$.

Proof: (i) Because of the invariance of $\mathcal{Q}_N$ under the particle-hole duality it suffices to show that $B_N^{-1}Q_2 \notin \mathcal{S}_2$, $Z_N \notin \mathcal{S}_2$ and $X_N, Y_N \notin \mathcal{S}_2$. Now

$$\mu(B_N^{-1}Q_2) = -\frac{(r+1)(r-N)}{(r-2)(r-3)(N-1)}.$$

Substitution of this value into (2.1d) leads to

$$N \geq (r-2) + \frac{16}{r+5},$$

an inequality which for $N < (r-2)$ is obviously violated. Similarly substitution of $\mu(Z_N)$ into (2.1d) gives

$$N \geq \left(1 - \frac{2(r-3)}{r^3 - r^2 - 2r - 4}\right), \quad r \geq r-2,$$

the second inequality holding for $r \geq 4$. In order to show that $X_N, Y_N \notin \mathcal{S}_2$ it suffices to show that among the inequalities (2.1) there is one which is violated by *both* C_N^{-1} and $C_N^{-1}C_2$. Substitution of the coordinates of C_N^{-1} into (2.1a) yields $(N-r-2)^2 \leq 0$ which is obviously violated. Substitution of the coordinates of

$$C_N^{-1}C_2 = \frac{r-N}{r-2} \left(\frac{r^2 - 2r - 4}{r^2 - (r+2)N}, \frac{1}{N-1} \right)$$

into (2.1a)

$$(N-2)(N-r-2) \geq 0,$$

which for $2 < N < (r-2)$ is clearly violated.

(ii) Notice that the second statement of (ii) is obtained from the first one via the particle-hole duality.

Now let V_N be the point of intersection of the line parallel to the λ axis,

TABLE II. Upper bounds to the ground state energy (in a.u.) of He_2 ($N=4$, $r=8$), LiH ($N=4$, $r=12$), and H_2O ($N=10$, $r=14$) induced by extreme points of $\mathcal{S}_N$ (the set of endomorphisms which induce sufficient conditions for quasi-N-representability).

Condition	He_2 ($N=4, r=8$)	LiH ($N=4, 8=12$)	H_2O ($N=10, r=14$)
Complete CI	−4.631	−7.887	−75.013
Q_2	−3.208	−4.670	−67.238
P_N	−2.643	−7.356	−72.689
B_2	−0.630	−4.157	
J_N	−0.619	−5.619	
U_N		−4.786	
$B_N^{-1}B_2$	−1.078	−4.786	
$\widetilde{C_N^{-1}B_2} = \underset{\sim}{B}{}_N^{-1}B_2$	−1.230		
$Q_N^{-1}B_2 = \tilde{B}_2$	−0.619		−62.180
$\underset{\sim}{\mathcal{J}}_N$	−0.630		−63.721
$\tilde{U}_N$			−63.008
V_N		−4.786	
F_N	−0.684	−5.559	−65.512

$$l_N : \mu = \frac{1}{(r-2)} \frac{r-N}{N-1},$$

with the segment $\mathrm{Id} \mid Q_2$. Explicitly, V_N is given by

$$V_N = \left(\frac{(r^2-8)-(3r-8)N}{(N-1)(r-2)(r-4)}, \frac{1}{(r-2)} \frac{r-N}{N-1} \right).$$

Similarly, let $\tilde{V}_N$ be the point of intersection of the line l_N with the segment $Q_N^{-1} \mid P_N$. Clearly

$$\tilde{V}_N = Q_N^{-1}(V_N) = \left(\frac{r-N}{N} \frac{2(r-2)^2-(3r-8)N}{(r-N-1)(r-2)(r-4)}, \frac{1}{r-2} \frac{r-N}{N-1} \right)$$

is obtained from V_N via the particle-hole duality.

Obviously,

$$C_N^{-1}B_2 \in \mathcal{Q}_N \quad \text{iff} \quad \lambda(V_N) \leq \lambda(C_N^{-1}B_2) \leq \lambda(\tilde{V}_N).$$

The inequality

$$\lambda(V_N) \leq \lambda(C_N^{-1}B_2)$$

leads to $(N-r+2)(3N-r-4) \leq 0$, which for $N < (r-2)$ implies $N \geq \frac{1}{3}(r+4)$. Similarly, the inequality

$$\lambda(C_N^{-1}B_2) \leq \lambda(\tilde{V}_N)$$

simplifies to $(N-r+2)(3N-2r) \geq 0$, which for $N < (r-2)$ implies $N \leq \frac{2}{3}r$.

2.11: *Lemma*: Let r be an even number not smaller than 8. Then (referring to the list in Table I):

(i) for $2 < N \leq r/3 : \partial\mathcal{B}_N \cap \partial\mathcal{Q}_N = \{U_N, P_N, Q_2\}$;

(ii) for $r/3 \leq N \leq (r+4)/3 : \partial\mathcal{B}_N \cap \partial\mathcal{Q}_N = (B_N^{-1}B_2 \mid V_N) \cup \{P_N, Q_2\}$;

(iii) for $(r+4)/3 \leq N \leq \frac{2}{3}(r-2)$: $\partial\mathcal{B}_N \cap \partial\mathcal{Q}_N = (B_N^{-1}B_2 \mid C_N^{-1}B_2) \cup \{P_N, Q_2\}$;

(iv) for $\frac{2}{3}(r-2) \leq N \leq \frac{2}{3}r$: $\partial\mathcal{B}_N \cap \partial\mathcal{Q}_N = (\tilde{V}_N \mid C_N^{-1}B_2) \cup \{P_N, Q_2\}$;

(v) for $\frac{2}{3}r \leq N < (r-2)$: $\partial\mathcal{B}_N \cap \partial\mathcal{Q}_N = \{\tilde{U}_N, P_N, Q_2\}$.

Proof: For all N with $2 < N < (r-2)$ we have $P_N, Q_2 \in \partial\mathcal{B}_N \cap \partial\mathcal{Q}_N$. This is an immediate consequence of Eq. (3.12) of SANP I and its dual [obtained from it by replacing N by $(r-N)$ and subsequent multiplication by Q_N^{-1}].

Next notice that it suffices to prove statements (i), (ii), and (iii) since the remainder is obtained by application of the particle-hole duality. If $2 < N \leq r/3$, then no extreme point of $\mathcal{B}_N$ belongs to $\mathcal{Q}_N$ and the segments $B_N^{-1}Q_2 \mid B_N^{-1}B_2$ and $\mathrm{Id} \mid Q_2$ intersect at

$$U_N = \frac{(r-3N)(N-2)}{2(r-N)(3N-4)} B_N^{-1}Q_2 + \left(1 - \frac{(r-3N)(N-2)}{2(r-N)(3N-4)}\right) B_N^{-1}B_2.$$

If $r/3 \leq N \leq (r+4)/3$ we have $B_N^{-1}B_2 \in \mathcal{Q}_N$ and thus the whole segment $B_N^{-1}B_2 \mid V_N$ belongs to $\mathcal{Q}_N$. Finally, if $(r+4)/3 \leq N \leq \frac{2}{3}(r-2)$ we have $B_N^{-1}B_2 \mid C_N^{-1}B_2 \in \mathcal{Q}_N$. No other pair of segments of $\mathcal{B}_N$ and $\mathcal{Q}_N$ intersect.

Collecting the information from Lemmas 2.9—2.11 and applying Theorem A7 finally leads to the establishment of Theorem 2.1.

3. NUMERICAL RESULTS

Results of calculations on He_2, LiH, and H_2O are presented in Table II. The calculations on He_2 were conducted using an elliptical orbital basis[4] of cardinality 8; The internuclear separation of He_2 for these calculations was 1.00 a.u. The computations on LiH and H_2O were carried out in a Slater type basis, using Poples' STO-3G program[5] to establish the integrals required. An internuclear separation of 3.0 a.u. was for LiH and for H_2O the O—H bond length used was 1.814 a.u. and the H—O—H angle was 104.7°.

For He_2, N is 4 and r is 8, i.e., $N = r/2$. Therefore, this is an example of cases (iii) and (iv) of Theorem 2.1. Note that for $N = r/2$, $B_2 = \tilde{J}_N$, and $J_N = \tilde{B}_2$. From Table II, it can be seen that Q_2 gives the best upper bound that can be obtained using quasi-N-representability criteria and conditions of the operator endomorphism form.

For LiH, N is 4 and r is 12, i.e., $N = r/3$. Thus, this is an example of cases (i) and (ii) of Theorem 2.1. For $N = r/3$, it should be noted that $U_N = B_N^{-1}B_2 = V_N$. In this case P_N turns out to be the best sufficient condition of operator endomorphism form.

For H_2O, N is 10 and r is 14, i.e., $2r/3 < N < r-2$. Hence, calculations on H_2O reported in column 4 of Table II are an example of case (vi) of Theorem 2.1. In this case as well, P_N is the best sufficient condition.

It should be pointed out that in all the cases reported either Q_2 or P_N gave the best upper bounds for quasi-N-representability. In other words: The concave functional $E(T)$ defined by (1.1) considered as a functional on $\mathcal{S}_N$ (the set of all operator endomorphisms which induce sufficient conditions for quasi-N-representability) attains its minimum at an extreme point of $\mathcal{S}_N$ (the set of all operator endomorphisms which induce sufficient conditions for N representability). It follows that the reported upper bounds are the *best* upper bounds which can be obtained using sufficient conditions for N representability induced by operator endomorphisms.

The last line of Table II gives the results for F_N, an extreme point of Q_N (cf. Proposition 2.2). The result shows that even if we allow T to vary over the larger set $Q_N \supseteq \hat{\mathcal{S}}_N \supseteq \mathcal{S}_N$ of all operator endomorphisms T with the property that whenever $D \in \mathcal{P}^2$, then $T(D)$ satisfies the necessary conditions induced by Id and Q_N, $E(T)$ does not provide us with a better upper bound than if we allow T to vary just over $\mathcal{S}_N$. These results seem to suggest that the upper bound $E(T)$ (considered as a functional on $\mathcal{S}_N$) attains its minimum almost always either at the extreme point Q_2 or at the extreme point P_N of $\mathcal{S}_N$ and rarely at any of those other extreme points of S_N which we have not been able to determine.

ACKNOWLEDGMENT

This work has been supported by N.R.C. Canada (N.R.C. Grants 2990 and 9233). We would like to express our thanks to N.R.C. for support.

APPENDIX: SOME THEOREMS ABOUT CONVEXITY

Throughout this appendix we assume L to denote a topological linear Hausdorff space. If $C \subseteq L$ is a subset we use $\bar{C}$ for the closure of C, C^0 for the interior of C and ∂C for the boundary of C.

If $x, y \in L$ and $x \neq y$ we denote by $[xy]$ the closed segment and by (xy) the open segment connecting x and y respectively, i.e.,

$$[xy] = \{z \in L \mid z = \delta x + (1-\delta)y, \quad \delta \in [0, 1]\},$$

$$(xy) = \{z \in L \mid z = \delta x + (1-\delta)y, \quad \delta \in (0, 1)\}.$$

A1: *Definition* (*Ref.* 6, *p.* 11): (i) Let $C \subseteq L$ be a subset. A point $y \in L$ is called *linearly accessible from* C iff there exists $x \in C$, $x \neq y$ such that $(xy) \subseteq C$. The set of all points which are linearly accessible from C we denote by *lina* C. (ii) A point $x \in C \subseteq L$ is called a *core point of* C iff for all $y \in L$ with $y \neq x$ there exists a point $z \in (xy)$ such that $[xz] \subseteq C$. The set of all core points we denote by *core* C.

A2: *Theorem* (*Ref.* 6, *Theorem* 1.14): Let C be a convex subset of L. Let $y \in$ lina C and $x \in$ core C. Then $(xy) \subseteq$ core C.

A3: *Definition* (*Ref.* 6, *p.* 13): A convex subset C of L is called a *convex body* iff it contains an interior point, i.e., if $C^0 \neq \phi$.

A4: *Theorem* (*Ref.* 6, *Theorems* 1.16 *and* 1.17): Let $C \subseteq L$ be a convex body. Then

$$C^0 = \text{core } C, \quad \bar{C} = \text{lina } C.$$

A5: *Lemma*: Let $C \subseteq L$ be a closed convex body. Suppose that for some $x, y \in C$, $x \neq y$ we have $(xy) \cap \partial C \neq \phi$. Then $x, y \in \partial C$.

Proof: Suppose, e.g., $x \in C^0$. Then by Theorem A4 x is a core point. Since by the same theorem $y \in C = \bar{C}$ $=$ lina C, it follows from Theorem A2 that $(xy) \subseteq$ core C $= C^0$. Thus $(xy) \cap \partial C = \phi$.

A6: *Definition*: Let $S \subseteq L$ be a subset. Then $z \in S$ is called *extreme* iff for all $x, y \in L$ with $x \neq y$, $z \in (xy)$ implies $x \notin S$ or $y \notin S$. We denote the set of all extreme points of S by *ext* S.

A7: *Theorem*: Let $C_1, C_2 \subseteq L$ be two closed convex bodies. Then

$$\text{ext}(C_1 \cap C_2) = [(\text{ext } C_1) \cap C_2] \cup [(\text{ext } C_2) \cap C_1] \cup \text{ext}(\partial C_1 \cap \partial C_2).$$

Proof: That (ext C_1) $\cap C_1$ and (ext C_2) $\cap C_1$ are subsets of ext$(C_1 \cap C_2)$ is trivial.

Suppose $z \in \text{ext}(\partial C_1 \cap \partial C_2)$ and assume $z \in (xy)$ with $x, y \in C_1 \cap C_2$. Then $z \in (xy) \cap \partial C_i$, $i = 1, 2$. It follows from Lemma A5 that $x, y \in \partial C_1 \cap \partial C_2$, a statement which contradicts the assumption that $z \in \text{ext}(\partial C_1 \cap \partial C_2)$. Hence $x \notin C_1 \cap C_2$ or $y \notin C_1 \cap C_2$ and therefore $z \in \text{ext}(C_1 \cap C_2)$.

Conversely, suppose $z \in \text{ext}(C_1 \cap C_2)$ but $z \notin$ ext C_1 and $z \notin$ ext C_2. Then $z \in \partial C_1 \cap \partial C_2$. For suppose, e.g., that $z \in C_1^0$. Since $z \in C_2 \setminus \text{ext } C_2$ there exists $x, y \in C_2$, $x \neq y$ with $z \in (xy)$. Furthermore, since z is a core point of C_1 (Theorem A4) there exists $x' \in (zx)$ and $y' \in (zy)$ such that $[zx'] \cup [zy'] \subseteq C_1$. Hence $x', y' \in C_1 \cap C_2$, $x' \neq y'$ and $z \in (x', y')$, a statement which contradicts $z \in \text{ext}(C_1 \cap C_2)$. □

[1]H. Kummer, I. Absar, and A.J. Coleman, J. Math. Phys. 18, 329 (1977).
[2]A.J. Coleman, Rev. Mod. Phys. 35, 668 (1963).
[3]H. Kummer, J. Math. Phys. 8, 2063 (1967).
[4]G.P. Barnett, Can. J. Phys. 45, 137 (1967).
[5]W.J. Nehre, W.A. Lathan, R. Ditchfield, M.D. Newton, and J.A. Pople, Gaussian 70, Quantum Chemistry Program Exchange (Indiana University, Bloomington, Indiana).
[6]F.A. Valentine, *Convex Sets* (McGraw-Hill, New York, 1964).

Remarks on the second order error in the variational calculation of expectation values*

Saul T. Epstein

Theoretical Chemistry Institute and Physics Department, University of Wiconsin, Madison, Wisconsin 53706
(Received 20 May 1976; revised manuscript received 31 August 1976)

A formula due to Gerjuoy [J. Math. Phys. **16**, 761 (1975)] is derived in a straightforward way by use of second order perturbation theory.

Recently, as the result of a rather intricate analysis, Gerjuoy[1] was led to suggest a formula for estimating the second order error in certain variational estimates of expectation values. However, if[2] $(H_t - E_t)\phi_t = 0$, then it is well known[3] that $F - \phi_t^\dagger W \phi_t$ as given in (G1.5) is simply the first order correction to $\phi_t^\dagger W \phi_t$ which one would calculate using first order Rayleigh–Schrödinger perturbation theory with H_t as the zero order Hamiltonian and $(H - H_t)$ as the perturbation. Therefore, it would seem of interest to examine the consequences of applying second order perturbation theory, and indeed, as we will show, this leads in a straightforward way to Gerjuoy's formula.[4]

Let $\phi^{(1)}$ and $\phi^{(2)}$ be the first and second order corrections to ϕ_t produced by the perturbation. Then assuming that our wavefunction is normalized through second order, the second order correction to the expectation value of W is obviously

$$\Delta^{(2)} = \phi^{(1)\dagger} W \phi^{(1)} + \phi^{(2)\dagger} W \phi_t + \phi_t^\dagger W \phi^{(2)}, \tag{1}$$

which since normalization implies that

$$\phi^{(1)\dagger}\phi^{(1)} + \phi^{(2)\dagger}\phi_t + \phi_t^\dagger \phi^{(2)} = 0 \tag{2}$$

we can write as

$$\Delta^{(2)} = \phi^{(1)\dagger}(W + \lambda_t)\phi^{(1)} + \phi^{(2)\dagger}(W + \lambda_t)\phi_t + \phi_t^\dagger (W + \lambda_t)\phi^{(2)} \tag{3}$$

We have spoken of first and second order corrections. In Ref. 1 order is defined variationally with the well-defined but unknown quantity $\delta\phi$ being of first order. On the other hand, in perturbation theory, order is defined by the way in which one splits up the Hamiltonian. Now from our opening remarks it would appear that, to get agreement through first order between these two notions of order, one must treat $(H - H_t)$ as a first order quantity. However, for what follows it is important to note that this is too strong a conclusion. Namely the perturbation results are unaltered through first order if one subtracts from $(H - H_t)$ any Hermitian operator X such that $X\phi_t = 0$ (one must, of course, add it back in higher order) because as one easily sees, such an X makes no contribution to $\phi^{(1)}$.

We now proceed to consider the variational order of $H - \hat{H}_{\text{mod},t}$ and of $H - H'_{\text{mod},t}$. For this purpose it is convenient to write them as much as possible in terms of $(H - E_t)P_t$ since this is a first order quantity (plus higher order corrections) in the variational sense. Doing this, we find

$$\begin{aligned} H - \hat{H}_{\text{mod},t} &= P_t(H - E_t) + (H - E_t)P_t \equiv h, \\ H - H'_{\text{mod},t} &= h + (H - E_t)P_t(H - E_t)/E_t \equiv h + h', \end{aligned} \tag{4}$$

and note that h' has the property that

$$h'\phi_t = 0 \tag{5}$$

so that it should be kept in mind that the remarks at the end of the previous paragraph apply to it. Evidently then in the variational sense h is of first order plus higher order corrections while h' is of second order plus higher order corrections. However, since the higher order corrections we keep mentioning all involve $\delta\phi$, they are actually unknown quantities. Therefore, it would seem that the best we can do if we want to make some contact between our perturbation calculation and Ref. 1 is to treat h and h' as of first and second order respectively in the perturbation sense, and ignore the corrections. This, of course, means that the resulting $\Delta^{(2)}$'s will be only estimates of the second order variational errors; however, this is also true of Gerjuoy's formula.[1]

Having settled on the perturbation order of things, we can finish our task quite quickly. First, using (5), it easily follows from the standard sum over states formula that when $H_t = H'_{\text{mod},t}$, h' makes no contribution to $\phi^{(2)}$. Further, since h is readily seen to have no matrix elements between states orthogonal to ϕ_t, it similarly follows, assuming the standard normalization convention

$$\phi_t^\dagger \phi^{(1)} = 0, \tag{6}$$

that in either case its contribution to $\phi^{(2)}$ is simply a multiple of ϕ_t.[5] Thus for either H_t, $\phi^{(2)}$ is simply a multiple of ϕ_t whence, since $\phi_t^\dagger(W + \lambda_t)\phi_t = 0$ we have as our final result that

$$\Delta^{(2)} = \phi^{(1)\dagger}(W + \lambda_t)\phi^{(1)}. \tag{7}$$

We will now show that this is Gerjouy's formula.[1] First of all we note that $\phi_t^\dagger h \phi_t = 0$ and that $h\phi_t = (H - E_t)\phi_t$. Therefore, we can evidently write the standard formula for $\phi^{(1)}$ namely $\phi^{(1)} = \hat{G}_t(h - \phi_t^\dagger h \phi_t)\phi_t$ as

$$\phi^{(1)} = \hat{G}_t(H - E_t)\phi_t, \tag{8}$$

where

$$(H_t - \hat{E}_t)\hat{G}_t = P_t - 1 \tag{9}$$

with H_t equal to $\hat{H}_{\text{mod},t}$ or $H'_{\text{mod},t}$ as the case may be.

Then, comparing[6] (7), (8), and (9) with (G4.21), (G4.20), and (G4.17a) (and the analog of the latter with $\hat{H}_{\text{mod},t}$ replaced by $H'_{\text{mod},t}$) respectively, we see that as claimed we have derived Gerjuoy's formula.

We conclude with two remarks. (i) In order to follow the variational approach, it is clearly essential to use an H_t which is given as an explicit function of H and ϕ_t, and even so one only gets estimates of the variational corrections. However, the general perturbation approach to calculating corrections is of course more flexible, being indifferent to the nature of H_t.[1] Further, formally at least, it can be readily extended to calculate higher order corrections. (ii) Having chosen H_t it is usually most natural to treat $H - H_t$ as of first order with no higher order corrections.[8] However, if one is willing to entertain other possibilities then there is a multiple infinity to choose from,[9] with rate of convergence of the perturbation series being the only real theoretical criterion of choice among them.

*Aspects of the variation method I. Research supported by National Science Foundation Grant MPS74-17494.

[1]E. Gerjuoy, J. Math. Phys. **16**, 761 (1975). The particular estimate we have in mind is that based on his (4.20) and (4.17a).

[2]All unexplained notation is that of Ref. 1. We will refer to the equations of Ref. 1 by prefixing a G to the equation number.

[3]See, for example, S.T. Epstein, *The Variation Method in Quantum Chemistry* (Academic, New York, 1974), pp. 236–37.

[4]As we have said we must have $(H_t - E_t)\phi_t = 0$. This is obviously true for $H_t = \hat{H}_{\text{mod},t}$ but equally clearly it is *not* true for $H_t = H_{\text{mod},t}$. However, if one simply replaces $H_{\text{mod},t}$ by $H'_{\text{mod},t} = H_{\text{mod},t} + E_t P_t$ then it *is* true, while neither the numerical value of F nor any of the essential results of Ref. 1 are changed in any way. Indeed $H'_{\text{mod},t}$ is much "nicer" than $H_{\text{mod},t}$ in that when $\phi_t = \phi$ then $H'_{\text{mod},t} = H$.

[5]For $H_t = \hat{H}_{\text{mod},t}$ this is a special case of a theorem due to W.H. Adams, J. Chem. Phys. **45**, 3422 (1966); Statement (C) of Sec. II for the case $M=1$, $M=2$. Note that his O is our P_t and his P is $(1-P_t)$.

[6]Note that from (G2.26) it follows that $\Delta^{(2)}$ is to be compared with $-\delta F^{(2)}$.

[7]The choices $H_t = \hat{H}_{\text{mod},t}$ and $H_t = H'_{\text{mod},t}$ were especially singled out for consideration in Ref. 1 because for them conditions have been given [E. Gerjuoy, A.R.P. Rau, L. Rosenberg, and L. Spruch, Phys. Rev. A **9**, 108 (1974); E. Gerjuoy, L. Rosenberg, and L. Spruch, J. Math. Phys. **16**, 455 (1975)] which guarantee that the Hylleraas functional, which is often needed in practice to approximate L_t in calculating the first order correction, approaches its exact value from above, whatever set of trial functions one uses. In this connection it should be also pointed out that if, for example, through a series of linear variation calculations, one can locate E_t in the (discrete) spectrum of *whatever* H_t one is using, then one can easily choose the trial functions so as to guarantee monotonic behavior for that H_t (see Ref. 3, pp. 209–10).

[8]Second order calculations based on this point of view have been reported by S. Hameed and H.M. Foley, Phys. Rev. A **6**, 1399 (1972). However, they wrote $\Delta^{(2)}$ in the computationally more attractive interchanged form

$$\Delta^{(2)} = \phi^{(1)\dagger}(W + \lambda_t)\phi^{(1)} + L_t^\dagger(H - H_t)\phi^{(1)} + \phi^{(1)\dagger}(H - H_t)L_t.$$

[9]Thus one may write $H = \sum_n \nu^n H^{(n)}$, where ν is an order parameter whose numerical value is 1, subject only to the conditions $H^{(0)} = H_t$, $H^{(1)} = H - H_t - X$, where $X\phi_t = 0$, and $\sum_n H^{(n)} = H$. In short, one can ask not only what is $H^{(0)}$, but also what is $H^{(n)}$? [For the first question see S.T. Epstein, in *Perturbation Theory and Its Applications in Quantum Mechanics*, edited by C.H. Wilcox (Wiley, New York, 1966).]

Short derivation of a formula due to Lubkin*

Saul T. Epstein

Theoretical Chemisty Institute and Physics Department, University of Wisconsin, Madison, Wisconsin 53706
(Received 25 May 1976)

We present a short derivation of Lubkin's formula for the ensemble of maximum entropy subject to mean value constraints involving noncommuting operators.

Recently, Lubkin[1] found the ensemble which maximizes the entropy subject to mean value constraints involving noncommuting operators. In this note we will give a shorter and different derivation of his result. We will use the notation of Ref. 1 without further comment.

We first note that by the standard argument, the ensemble which maximizes the entropy subject to the single constraint

$$\sum_j \epsilon_j \operatorname{Tr} PA_j = \sum_j \epsilon_j Q_j \qquad (1')$$

with the ϵ_j real numbers is

$$P = \exp\left[-\lambda(\epsilon)\sum_j \epsilon_j A_j\right] / \operatorname{Tr} \exp\left[-\lambda(\epsilon)\sum_j \epsilon_j A_j\right]$$

where the real constant $\lambda(\epsilon)$ is determined by (1′).

Secondly, we note that the constraints (2) of Ref. 1 can be replaced by the constraints

$$\sum_j \epsilon_j \operatorname{Tr} PA_j = \sum \epsilon_j Q_j, \quad \text{all real } \epsilon_j. \qquad (2')$$

Finally, we note that it then follows that

$$P = \exp\left(-\sum_j \lambda_j A_j\right) / \operatorname{Tr} \exp\left(-\sum_j \lambda_j A_j\right), \qquad (3')$$

with the real numbers λ_j chosen to satisfy the constraints (2) of Ref. 1, will (i) satisfy all the constraints (2′), and (ii) certainly maximize the entropy for one particular choice of the ϵ_j, namely for the ϵ_j determined by $\lambda(\epsilon)\epsilon_j = \lambda_j$.

Therefore, since the maximum subject to one constraint cannot be less than the maximum subject to all constraints, we conclude, in agreement with Lubkin, that (3′) is the ensemble which maximizes the entropy subject to all constraints.

*Work supported by the National Science Foundation Grant MPS74-17494.

[1]E. Lubkin, J. Math. Phys. 17, 753 (1976).

Correction of "Extension of the statistical mechanics of equilibrium to noncummutative constraints" to cover singular constraints

Elihu Lubkin

Department of Physics, University of Wisconsin-Wilwaukee, Milwaukee, Wisconsin 53201
(Received 25 May 1976)

In a previous paper, a Gibbs ensemble formula was given for that ensemble P which maximizes the entropy $S = -\mathrm{Tr}\,P \ln P$ subject to constraints $\mathrm{Tr}\,PA_i = Q_i$, where the A_i need not commute. An oversight, namely, cases where the constraints though consistent force P to be singular, is here properly incorporated into the argument. The result is that one must first regard the constraints as limiting discussion as much as possible to a sub-Hilbert space of constraint C, then the maximum entropy solution is indeed given by a Gibbs formula as before, but referring to matrices over C.

In a previous paper,[1] a Gibbs-ensemble-like expression was given for that density matrix P which maximizes the entropy $S = -\mathrm{Tr}P\ln P$ subject to $c+1$ *constraints*,

$$\mathrm{Tr}PA_i = Q_i, \quad i = 0, 1, \ldots, c, \tag{1}$$

namely,

$$P = \exp\left(-\sum_{i=0}^{n} \lambda_i A_i\right); \tag{2}$$

the A_i are Hermitian matrices which need not commute; $A_0 = I$ (the unit matrix) and $Q_0 = 1$ express normalization, and the matrices are (n by n), n finite; the λ_i are real numbers. I have since found and corrected an error in (I), then also learned that Wichmann had already done the whole thing.[2] Because of the great interest potential in the subject—a thermodynamics of noncommuting observables ("quantum thermodynamics") may develop or is developing[3]—I hope that this presentation of the correction of (I) is not entirely superfluous, especially as the approach of (I) is perhaps more elementary than Wichmann's earlier discussion, even after the present correction is incorporated.

In (I), it was asserted that (2) is the form of the maximum-entropy solution whenever the constraints (1) are consistent, i.e., whenever the set of P satisfying (1) is nonempty. This caution is unfortunately inadequate. The mistake was made by jumping into a Lagrange-multiplier argument without sufficient care about possible solutions on the boundary of the domain. The corrected result is Theorem 1 below; Theorem 2 is a slightly weaker but simpler statement.

Definition: For any Hermitian H, let $E_H = I - O(H)$, where $O(H)$ is the null space of H. Thus, $\mathrm{Tr}E_H = \mathrm{rank}\ H$ E_H is "H's projection."

Theorem 1: $\exists !$ projection E of maximal *rank* r out of those E_P where P satisfies the constraints (1). The ensemble of maximal *entropy* among P satisfying (1) is given by (2) in terms of (r by r) matrices over the r-dimensional constraint space $C = \mathrm{Im}E$.

Definition: If the constraints (1) imply that P is singular, they will be called *singular constraints*; if not, *nonsingular*.

Theorem 2: If the constraints (1) are nonsingular, then the ensemble of maximal entropy satisfying (1) is given by (2).

Proof: Theorem 2 is an evident corollary of Theorem 1.

As for the proof of Theorem 1, only the modification of the argument of Ref. 1 owing to care about the boundary will be discussed.

The geometrical context will be the n^2 real-dimensional vector space of Hermitian matrices. The domain D over which the maximum is sought is the intersection of the affine space of P satisfying the constraints (1) with the nonnegative cone. D is convex. By "boundary" B is meant the set of singular $P \in D$, $\det P = 0$.

First, the original proof correctly gives the result (2) if it is known that the solution does not belong to B. Furthermore, if there is *any* point of D not in B, i.e., $P_0 \in D$, $\det P_0 \neq 0$, then the solution indeed cannot belong to B: This is because the entropy $S = -\mathrm{Tr}P\ln P$ has a $+\infty$ slope at the boundary owing to $(d/dx)(-x\ln x) = +\infty$ at $x = 0$, increasing towards the interior, hence S also increases upon any small motion away from the boundary of D, in particular over a line segment drawn to P_0.

The task remaining is to discuss singular constraints, $D = B$. For this purpose, a minimal projection E is sought such that $PE = P$ for all $P \in D$. If such an $E < I$ exists, then the problem can be reformulated in the sub-Hilbert space $\mathrm{Im}E$, and this iterated until the former case $B \subset_{\neq} D$ is obtained. (Indeed choosing E minimal obviates repetitions.) The desideratum $E < I$ follows from "$E = E_P$ for any P of maximal rank in D" (of course $E_P < I$ since any $P \in D$ is singular). "$E = E_P$" follows from the following lemma:

Lemma: If $P \in D$, $\Sigma \in D$, $\alpha > 0$, $\beta \geq 0$, $\alpha + \beta = 1$, then $E_P \leq E_P \leq E_{\alpha P + \beta\Sigma}$.

Proof of Lemma: We must show that if $(\alpha P + \beta\Sigma)x = 0$ for a vector x, then $Px = 0$. Indeed, $\alpha(x, Px) + \beta(x, \Sigma x) = 0$. Both (x, Px) and $(x, \Sigma x)$ are nonnegative. Hence $(x, Px) = 0$, or $\mathrm{Tr}P|x)(x| = 0$, or $P|x)(x| = 0$,[4] or $Px = 0$, which proves the lemma.

Proof of Theorem 1, *concluded*: Suppose P is of maximal rank in D. Then the lemma shows any line segment from P to any other Σ in D to pass through $\alpha P+\beta\Sigma$ of materially greater rank, a contradiction, unless $E_{\alpha P+\beta\Sigma}=E_P$, which is therefore proven. If Σ is also of maximal rank, then also $E_{\alpha P+\beta\Sigma}=E_\Sigma$, hence $E_P=E_\Sigma\equiv E$ is universal. If $\Sigma\in D$ is however arbitrary, $E=E_{\alpha P+\beta\Sigma}\geq E_\Sigma$ proves that ImE indeed contains ImE_Σ for all $\Sigma\in D$. This allows collapse of the whole problem to the sub-Hilbert space Im$E=C$.

ACKNOWLEDGMENT

I noticed my omission of singular constraints while corresponding with S. T. Epstein about something else: Epstein sent me a very brief argument[5] which attempts to derive the result of (I) and Ref. 2 from Gibbs' original case of only one nontrivial observable, using an argument similar to my discussion of angular momentum in (I). Unfortunately, the brief argument fails without the hypothesis that there exists some ensemble of form (2) which satisfies (1); to know only that the set of ensembles satisfying (1) is nonempty is not enough. This further hypothesis is not needed here. There is also a comment in Ref. 2 to the effect that much work can be omitted if one is willing to be incautious about domains.

[1]E. Lubkin, J. Math. Phys. 17, 753 (1976), hereafter called (I).
[2]E.H. Wichmann, J. Math. Phys. 4, 884 (1963).
[3]W. Bayer and W. Ochs, Z. Naturforsch. 28a, 693—701 (1973) and Refs. 1—4 therein to the "Jaynes" approach, and W. Ochs and W. Bayer, *ibid.*, 1571—85 (1973). Bayer and Ochs are interested in generalizations to infinite-dimensional Hilbert space.
[4]E. Lubkin, J. Math. Phys. 15, 663 (1974), see #59, p. 670.
[5]S. T. Epstein, J. Math. Phys. 18, 344 (1977).

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*