Generalizations of the quaternion algebra and Lie algebras

J. F. Plebański[a)]
Departamento de Física, Centro de Investigación y de Estudios Avanzados del I.P.N., México 14, D.F., Mexico
M. Przanowski
Institute of Physics, Technical University of Łódź, Wólczańska 219, 93-005 Łódź, Poland

(Received 29 June 1987; accepted for publication 30 September 1987)

Natural generalizations of the quaternion algebra called quaternionlike algebras (ql algebras) are considered. The notion of a Lie algebra induced by a ql algebra is introduced and a classification of such Lie algebras is presented. It is shown that local Lie groups defined by Lie algebras induced by ql algebras exhaust, with accuracy to local isomorphisms, all local Lie groups endowed with some simple composition laws of their local parameters.

I. INTRODUCTION

It is well known[1] that the group SO(3;R) [SO(2,1;R)] is locally isomorphic to the local Lie group[2,3] (R^3,φ), where φ is a mapping of some sufficiently small open neighborhood $U \subset R^3 \times R^3$ of the point $0 \in R^3 \times R^3$ into R^3 defined as follows:

$$\varphi:\ U \ni (x,y) \to z \in R^3,\quad z^i = \frac{x^i + y^i + \epsilon^{ijk} x_j y_k}{1 + g_{lm} x^l y^m},\qquad (1.1)$$

$i,j,... = 1,2,3;$ $x = (x^1,x^2,x^3)$, $y = (y^1,y^2,y^3)$, $z = (z^1,z^2,z^3) \in R^3$, ϵ^{ijk} is the totally antisymmetric Levi-Civita symbol in three dimensions; the matrix $(g_{ij}) := \text{diag}(-1,-1,-1)$ [for SO(2,1;R), $(g_{ij}) := \text{diag}(1,1,-1)$] enables one to lower the indices $i,j,...$ according to the rule $x_i := g_{ij} x^j$, $y_i := g_{ij} y^j$; the summational convention applies.

The composition law (1.1) is very attractive because of its simplicity. Therefore the natural question arises whether the formula (1.1) is "rigid," or whether, by changing ϵ^{ijk} and g_{ij} into more general real or complex objects, assuming also that x, y, z are elements of R^n or C^n, respectively, with a suitable n, one can find the composition laws for some other local Lie groups. In the present paper we answer this question.

In Sec. II we consider the associative algebras which generalize the quaternion algebra in a natural manner and due to this fact we call them quaternionlike algebras, or, briefly, ql algebras. In particular they contain the quaternion algebra and the generalized quaternion algebra.[4,5]

If Q is an $(n+1)$-dimensional ql algebra ($n \geqslant 1$) over F ($=R$ or C), then, as we shall see in Sec. II, there exists a unique decomposition $Q = Fe_0 \oplus V$, where e_0 is the unity of Q and V is an n-dimensional vector subspace of Q such that for each vector $v \in V$, $vv \in Fe_0$. Then, for any $v,w \in V$ their commutator $[v,w] := vw - wv$ belongs to V. Thus the pair $(V[\cdot,\cdot])$ is an n-dimensional Lie algebra over F which we call a Lie algebra induced by Q and denote by $\widetilde{Q}_L$. It appears that the local Lie group defined by $\widetilde{Q}_L$ is locally isomorphic to some local Lie group with composition law being the modification of the formula (1.1) as it has been described above. To establish this isomorphism we define some n-dimensional Lie group H_Q which is a suitable subset of Q and we show that the Lie algebra of H_Q can be identified with $\widetilde{Q}_L$. We find a local coordinate system in a neighborhood of the identity of H_Q such that the composition law expressed in terms of this local coordinate system is just the appropriate modification of (1.1). In this way one finds a mapping from the class of all nonisomorphic ql algebras onto the class of all locally nonisomorphic local Lie groups the composition laws of which are some definite modifications of the law (1.1). We will prove also that for ql algebras of dimension $n+1>2$ this mapping is 1:1. The local coordinate system we have just spoken about enables us to write down in a concise form the Baker–Campbell–Hausdorff series[1,2,5,6] for the elements of any $\widetilde{Q}_L$ algebra. Our formula is an obvious generalization of the analogous formulas given for the Lie algebras SO(3;R) and SO(2,1;R).[1,6,7] Moreover, it can be found that if Q is the quaternion algebra, then the group H_Q appears to be the group of all quaternions of norm 1 and one has the well-known isomorphism $H_Q \simeq \text{SU}(2)$.

In Sec. III the classification of Lie algebras induced by ql algebras is given. Thus, of course, the classification of local Lie groups endowed with the composition laws being simple modifications of (1.1) is also given. We study real and complex algebras separately. Employing the results of Bianchi,[8] Behr *et al.*,[9] Ellis and MacCallum,[10] MacCallum,[11,12] Mubarakzyanov,[13] Morozov,[14] and Patera *et al.*[15,16] we list all real Lie algebras of dimension $2 \leqslant n \leqslant 6$ induced by real ql algebras.

Concluding remarks close our paper.

II. QUATERNIONLIKE ALGEBRAS, INDUCED LIE ALGEBRAS, AND LOCAL LIE GROUPS

Let Q be an $(n+1)$-dimensional ($n \geqslant 1$) algebra with unity e_0 over F ($=R$ or C) for which there exists a decomposition

$$Q = Fe_0 \oplus V,\qquad (2.1)$$

where V is an n-dimensional vector subspace of Q such that for each vector $v \in V$, $vv \in Fe_0$.

We have the following.

Proposition 2.1: The decomposition (2.1) is unique.

Proof: Let $Q = Fe_0 \oplus V_1$. We will show that $V_1 = V$. Indeed, if v_1 is any nonzero vector in V_1, then, by (2.1), $v_1 = ae_0 + v$ with $a \in F$ and $0 \neq v \in V$. From the assumption we have $v_1 v_1 \in Fe_0$ and $vv \in Fe_0$. Hence it follows that $av = 0$, and

[a)] On leave of absence from the University of Warsaw, Warsaw, Poland.

 0022-2488/88/030529-07$02.50

in consequence, as $v \neq 0$, $a = 0$. It means that $v_1 \in V$. If $v_1 = 0$, then, of course, $v_1 \in V$. Thus one finds that $V_1 \subset V$. Analogously we prove the inclusion $V \subset V_1$. Finally $V_1 = V$. ■

Let $e_1,...,e_n$ be a basis of V. Then

$$e_i e_j = \tfrac{1}{4} K_{ij} e_0 + \tfrac{1}{2} C^k{}_{ij} e_k , \qquad (2.2)$$

where small Latin indices i, j, k (as well as $l, m, p,...$ in our further considerations) are assumed to run through $1,...,n$ and $K_{ij}, C^k{}_{ij} \in F$; the summational convention applies. [The factors $\frac{1}{4}$ and $\frac{1}{2}$ in (2.2) are taken for further convenience.]

We intend to establish that

$$C^k{}_{ij} = - C^k{}_{ij} . \qquad (2.3)$$

We first prove the following.

Proposition 2.2: If the vectors $v,w \in V$, then $vw + wv \in Fe_0$.

Proof: By the definition of V one infers that

$$(v+w)(v+w) = vv + ww + vw + wv \in Fe_0 ,$$

$vv \in Fe_0$ and $ww \in Fe_0$. Hence, our assertion holds. ■

Utilizing Proposition 2.2 one can easily verify (2.3). Indeed, from Eq. (2.2) and Proposition 2.2 we get

$$e_i e_j + e_j e_i = \tfrac{1}{4}(K_{ij} + K_{ji}) e_0 + \tfrac{1}{2}(C^k{}_{ij} + C^k{}_{ji}) e_k \in Fe_0 .$$

Thus $C^k{}_{ij} + C^k{}_{ji} = 0$ and (2.3) holds.

Now the question arises under what assumptions the algebra Q appears to be associative. We give an answer to this question by establishing three theorems which are fundamental for our further purposes.

Theorem 2.1: Here Q is an associative algebra iff

$$C^l{}_{jm} C^m{}_{ki} = K_{jk} \delta^l{}_i - K_{ij} \delta^l{}_k , \qquad (2.4)$$

$$K_{il} C^l{}_{jk} = K_{lk} C^l{}_{ij} . \qquad (2.5)$$

Proof: Clearly Q is an associative algebra iff

$$(e_i e_j) e_k = e_i (e_j e_k) \qquad (2.6)$$

for arbitrary $i,j,k = 1,...,n$. Employing Eqs. (2.2) and (2.3) one can easily find that the requirement (2.6) is fulfilled iff the equations

$$- C^l{}_{im} C^m{}_{kj} - C^l{}_{km} C^m{}_{ji} = K_{ij} \delta^l{}_k - K_{jk} \delta^l{}_i \qquad (2.7)$$

and (2.5) hold. Executing the antisymmetrization $[ijk]$ in (2.7) one obtains the Jacobi identity

$$C^m{}_{[ij} C^l{}_{k]m} = 0 . \qquad (2.8)$$

Using (2.3) and (2.8) to the left-hand side of (2.7) we get (2.4). This completes the proof. ■

If $n = 1$ then Eqs. (2.4) and (2.5) are satisfied for every K_{11} (we have of course $C^1{}_{11} = 0$). Hence for $n = 1$ the algebra Q is associative for an arbitrary K_{11}. If $n > 1$, then K_{ij} is defined in terms of $C^i{}_{jk}$. In fact we have the following.

Theorem 2.2: If $n > 1$, then Eq. (2.4) necessitates the following formula:

$$K_{ij} = K_{ji} = [1/(n-1)] C^m{}_{il} C^l{}_{jm} . \qquad (2.9)$$

Proof: It is a straightforward matter to show that Eq. (2.4) yields (2.8) and then, by (2.3), also (2.7). Contracting (2.7) with respect to the indices l and k, and then with respect to l and i one gets

$$- C^l{}_{im} C^m{}_{lj} - C^l{}_{lm} C^m{}_{ji} = n K_{ij} - K_{ji} , \qquad (2.10)$$

$$- C^l{}_{lm} C^m{}_{kj} - C^l{}_{km} C^m{}_{jl} = K_{kj} - n K_{jk} . \qquad (2.11)$$

Changing the index $k \to i$ in formula (2.11) and adding the result to (2.10), employing also (2.3), we obtain $(n+1) \times (K_{ij} - K_{ji}) = 0$. Thus

$$K_{ij} = K_{ji} . \qquad (2.12)$$

Contracting Eq. (2.4) with respect to the indices l and k, utilizing (2.12) and (2.3), one gets (2.9). ■

Our analysis of the conditions under which Q is an associative algebra is closed by the following theorem.

Theorem 2.3: For $n \neq 3$, Eqs. (2.4) and (2.5) yield the following formula:

$$K_{il} C^l{}_{jk} = 0 . \qquad (2.13)$$

Proof: For $n = 1$, Eq. (2.13) holds. If $n \neq 1$, then contracting (2.4) with $C^j{}_{lp}$, using also (2.3), (2.5), and (2.9), we have $(n-3) K_{pm} C^m{}_{ki} = 0$. Thus the theorem holds. ■

The most distinguished example of our algebras is the quaternion algebra which is realized when $n = 3$ and $C^k{}_{ij} = 2\epsilon^{kij}$ ($\Rightarrow K_{ij} = -4\delta^i{}_j$). Then taking $n = 3$ and $C^k{}_{ij} = 2a^k \epsilon^{kij}$ ($\Rightarrow K_{ij} = -4a^i \delta^i{}_j$; of course, there is no summation over i or k) with $a^3 = 1$ one constructs the so-called generalized quaternion algebra (see van der Waerden,[4] §93, Jacobson,[5] Sec. X, §7). The cited examples make it reasonable to call our algebras quaternionlike algebras. Thus we arrive at the definition.

Definition 2.1: A quaternionlike algebra (ql algebra) is an associative algebra Q with unity admitting the decomposition (2.1).

If Q is an $(n+1)$-dimensional ql algebra over F and $Q = Fe_0 \oplus V$ is the decomposition (2.1), then the pair $(V,[\cdot,\cdot])$, where

$$[\cdot,\cdot]: V \times V \ni (v,w) \mapsto [v,w] := vw - wv \in V ,$$

is an n-dimensional Lie algebra over F which we call a Lie algebra induced by Q and we denote it by $\tilde{Q}_L$. From (2.2) it follows that the numbers $C^k{}_{ij}$ are the structure constants of $\tilde{Q}_L$ with respect to the basis $e_1,...,e_n$. Then from (2.9) one finds that the numbers $(n-1) K_{ij}$ constitute the components of the Killing tensor of $\tilde{Q}_L$ (Refs. 2 and 5) with respect to the basis $e_1,...,e_n$.

Let $q = q^0 e_0 + q^i e_i$ be an element of an $(n+1)$-dimensional ql algebra Q. Then, the vector $\bar{q} := q^0 e_0 - q^i e_i \in Q$ is said to be a conjugate vector to the vector q. Define the following subset of Q:

$$H_Q := \{q \in Q: q\bar{q} = 1 e_0\} . \qquad (2.14)$$

Employing the formulas (2.4) and (2.5) one can easily check that the set H_Q together with the multiplication inherited from Q constitute a group. Moreover, as Q possesses a differentiable structure of F^{n+1}, then H_Q is an n-dimensional submanifold of Q. Finally, H_Q is an n-dimensional (real or complex) Lie group. [Notice that if Q is the quaternion algebra, then we have a well-known isomorphism $H_Q \simeq SU(2)$.] Let now $W \subset H_Q$ be an open neighborhood of e_0 in H_Q such that the pair (W,ψ) is an allowable chart of H_Q, where $\psi: W \to F^n$ is a mapping which sends $q = q^0 e_0 + q^i e_i \in W$ into $(q^1/q^0,...,q^n/q^0) \in F^n$. Let $p \in W$, $q \in W$, and $p,q \in W$. Then, denoting $x^i := p^i/p^0$, $y^i := q^i/q^0$, $z^i := (pq)^i/(pq)^0$, using (2.2), one finds

$$\pi^i_Q(x,y):=z^i=\frac{x^i+y^i+\frac{1}{2}C^i{}_{jk}x^jy^k}{1+\frac{1}{4}K_{lm}x^ly^m}, \qquad (2.15)$$

where x stands for $(x^1,...,x^n)\in\psi(W)\subset F^n$ and analogously y stands for $(y^1,...,y^n)\in\psi(W)\subset F^n$. Henceforth we call the coordinates x^i, y^i, z^i, etc., the projective coordinates. Thus one arrives at the following.

Theorem 2.4: Let Q be an $(n+1)$-dimensional ql algebra over F and let H_Q be an n-dimensional Lie group defined by the formula (2.14). Let $U\subset F^n\times F^n$ be an open neighborhood of the point $0\in F^n\times F^n$ such that for every $(x,y)\in U$,

$$1+\tfrac{1}{4}K_{ij}x^iy^j\neq 0, \qquad (2.16)$$

where K_{ij} is defined according to (2.2); moreover, $(0,y)$ and $(x,0)$ are elements of U for all $x,y\in F^n$. Then the pair (F^n,π_Q) is a local Lie group locally isomorphic to H_Q, where π_Q: $U\to F^n$ is a mapping defined as in (2.15). ■

From (2.15) we find immediately

$$\left(\frac{\partial^2\pi^i(x,y)}{\partial x^j\,\partial y^k}\right)_{\substack{x=0\\y=0}}-\left(\frac{\partial^2\pi^i(x,y)}{\partial x^k\,\partial y^j}\right)_{\substack{x=0\\y=0}}=C^i{}_{jk}. \qquad (2.17)$$

Therefore, the numbers $C^i{}_{jk}$ are the structure constants of the Lie algebra of H_Q [and, of course, of (F^n,π_Q)] with respect to a suitable chart (W,ψ) [$(\psi(W),\mathrm{id})$, respectively].[3] Thus we can identify these Lie algebras with the Lie algebra $\tilde{Q}_L$. The formula (2.15) resembles closely (1.1). In fact (2.15) is a "natural" modification of (1.1), that we have spoken about in the Introduction. But, for completeness, one should solve the following problem: Let K_{ij}, $C^i{}_{jk}=-C^i{}_{kj}$ be some numbers in F and let $U\subset F^n\times F^n$ be an open neighborhood of the point $0\in F^n\times F^n$ defined analogously as in Theorem 2.4. Finally, let π: $U\to F^n$ be a mapping defined as in (2.15). The question is what the conditions are for the pair (F^n,π) to be a local Lie group.

The answer is the following.

Theorem 2.5: (F^n,π) is a local Lie group iff Eqs. (2.4) and (2.5) hold.

Proof: First, $\pi(x,0)=0=\pi(0,x)$ for every $x\in F^n$; moreover, if $(-x,x)\in U$, then $\pi(-x,x)=0$. The mappings π: $U\to F$ and $F^n\ni x\mapsto -x\in F^n$ are analytic. Therefore it remains only to prove that if $(x,y)\in U$, $(y,z)\in U$, $(\pi(x,y),z)\in U$, and $(x,\pi(y,z))\in U$, then

$$\pi(\pi(x,y),z)=\pi(x,\pi(y,z)). \qquad (2.18)$$

Simple manipulations show that (2.18) is satisfied iff Eqs. (2.4) and (2.5) hold. Thus the proof is complete. ■

From Theorem 2.5 and our previous considerations it follows that there exists a 1:1 correspondence between the class of all nonisomorphic ql algebras of dimension >2 and the class of all locally nonisomorphic local Lie groups of dimension >1 endowed with the composition laws of the form (2.15), where K_{lm} and $C^i{}_{jk}=-C^i{}_{kj}$ are the elements of F. Every Lie algebra induced by a ql algebra can be identified with the Lie algebra of the corresponding local Lie group. One easily finds that in the case of the quaternion algebra [$\tilde{Q}_L\simeq\mathrm{so}(3;R)$] the formula (2.15) turns into (1.1) with $(g_{ij})=\mathrm{diag}(-1,-1,-1)$; in the case of the generalized quaternion algebra such that $C^1{}_{ij}=-2\epsilon^{1ij}$, $C^2{}_{ij}=-2\epsilon^{2ij}$, $C^3{}_{ij}=2\epsilon^{3ij}$ [$\tilde{Q}_L\simeq\mathrm{so}(2,1;R)$] the formula (2.15) yields (1.1) with $(g_{ij})=\mathrm{diag}(1,1,-1)$. The latter results in a slightly different formalism than was found by Plebański[1,7] many years ago. It is now evident that the composition law (2.15) appears to be a natural generalization of (1.1).

We close the present section with some considerations on the Baker–Campbell–Hausdorff formula for the elements of the Lie algebras induced by the ql algebras.

Let Q be an $(n+1)$-dimensional ql algebra over F and $\tilde{Q}_L$ the Lie algebra induced by Q. If $v=v^ie_i\in\tilde{Q}_L$, then using (2.2) one gets

$$e^v=(\cosh\Delta)e_0+[(\sinh\Delta)/\Delta]v^ie_i\in H_Q, \qquad (2.19)$$

where $\Delta:=\sqrt{(v|v)}$, $(v|v):=-\frac{1}{4}K_{ij}v^iv^j$. If v belongs to a sufficiently small open neighborhood of the vector $0\in\tilde{Q}_L$, then v^i are the canonical coordinates of the first kind (see Ref. 2, Sec. III, Chap. 4) of the point $e^v\in H_Q$. From (2.19) and the notion of the projective coordinates one easily finds the relation between the canonical coordinates v^i and the corresponding projective coordinates x^i,

$$x^i=\frac{\sinh\Delta}{\Delta\cosh\Delta}v^i=\frac{\tanh\Delta}{\Delta}v^i. \qquad (2.20)$$

This is the "tangential" parametrization[1,6,7] generalized on an arbitrary H_Q.

Let $v,w\in\tilde{Q}_L$. Then

$$e^ve^w=e^{v\#w}, \qquad (2.21)$$

where $v\#w$ is the Baker–Campbell–Hausdorff series.[1,2,5–7] To express $v\#w$ in a concise form we proceed as follows. If v and w are elements of a sufficiently small open neighborhood of the vector $0\in\tilde{Q}_L$, then we have the projective coordinates

$$x^i=\frac{\tanh\sqrt{(v|v)}}{\sqrt{(v|v)}}v^i,\quad y^i=\frac{\tanh\sqrt{(w|w)}}{(w|w)}w^i,$$
$$z^i=\frac{\tanh\sqrt{(v\#w|v\#w)}}{(v\#w|v\#w)}(v\#w)^i \qquad (2.22)$$

for e^v, e^w, and $e^{v\#w}$, respectively. Define

$$\mathbf{x}:=x^ie_i,\quad \mathbf{y}:=y^ie_i,\quad \mathbf{z}:=z^ie_i. \qquad (2.23)$$

Utilizing (2.15), (2.22), and (2.23) one gets

$$v\#w=\frac{\operatorname{arctanh}\sqrt{(\mathbf{z}|\mathbf{z})}}{\sqrt{(\mathbf{z}|\mathbf{z})}}\mathbf{z}, \qquad (2.24)$$

$$\mathbf{z}=\frac{\mathbf{x}+\mathbf{y}+\mathbf{x}\wedge\mathbf{y}}{1-(\mathbf{x}|\mathbf{y})},\quad \mathbf{x}=\frac{\tanh\sqrt{(v|v)}}{\sqrt{(v|v)}}v,$$
$$\mathbf{y}=\frac{\tanh\sqrt{(w|w)}}{\sqrt{(w|w)}}w, \qquad (2.25)$$

where $(\mathbf{x}|\mathbf{y}):=-\frac{1}{4}K_{ij}x^iy^j$, $\mathbf{x}\wedge\mathbf{y}:=\frac{1}{2}C^i{}_{jk}x^jy^ke_i$. Then inserting (2.25) into (2.24) and understanding that the right-hand side of (2.24) is a "formal sum" of the Baker–Campbell–Hausdorff series we find the Baker–Campbell–Hausdorff formula in a concise form for arbitrary $v,w\in\tilde{Q}_L$. This result is an obvious generalization of the one given for the Lie algebras so(3;R) and so(2,1;R) (see Refs. 1, 6, and 7).

III. CLASSIFICATION OF LIE ALGEBRAS INDUCED BY QUATERNIONLIKE ALGEBRAS

In this section we present the classification of Lie algebras induced by ql algebras. Thus, at one stroke, we get also the classification of the local Lie groups endowed with the composition laws of the form (2.15). From Theorem 2.3 it follows that the case of $n = 3$ is rather a particular one and it should be examined separately. First we consider real Lie algebras and then complex Lie algebras.

A. Real Lie algebras

1. n=3

All nonisomorphic three-dimensional real Lie algebras were found by Bianchi.[8] Then Bianchi's classification has been reformulated by Behr *et al.*[9] and Ellis and MacCallum.[10] We follow them (see also MacCallum[11,12] and Spindel[17]).

The structure constants of a three-dimensional Lie algebra can be written as follows:

$$C^i{}_{jk} = M^{il}\epsilon_{ljk} + N_l\delta^{li}_{jk}\,, \tag{3.1}$$

where $M^{il} = M^{li}$, ϵ_{ljk} is the totally antisymmetric Levi-Civita symbol, and

$$\delta^{li}_{jk} := \delta^l{}_j\delta^i{}_k - \delta^l{}_k\delta^i{}_j\,.$$

Then the Jacobi identity (2.8) is equivalent to the relation

$$M^{ij}N_j = 0\,. \tag{3.2}$$

From (2.9) and (3.1), utilizing also (3.2), one gets

$$K_{ij} = -\tfrac{1}{2}M^{lk}M^{pr}\epsilon_{ilp}\epsilon_{jkr} + N_iN_j\,. \tag{3.3}$$

Inserting (3.1) and (3.3) into (2.5), employing (3.2), we conclude that Eq. (2.5) is satisfied iff

$$\epsilon_{ijk}\det(M^{lm}) = \epsilon_{kij}\det(M^{lm})\,. \tag{3.4}$$

As (3.4) holds true, the condition (2.5) is fulfilled automatically without any further assumptions.

Consider now the consequences of (2.4). From (3.1) and (3.2) we find

$$C^l{}_{jm}C^m{}_{ki} = -M^{ml}M^{pr}\epsilon_{jmp}\epsilon_{kir} + 2M^{lm}N_j\epsilon_{mki} + N_j(N_k\delta^l{}_i - N_i\delta^l{}_k)\,. \tag{3.5}$$

Substituting (3.3) and (3.5) into (2.4) we arrive at the conclusion that Eq. (2.4) is satisfied iff

$$M^{ik}N_j = 0 \Leftrightarrow M^{ik} = 0 \quad \text{or} \quad N_j = 0\,. \tag{3.6}$$

Gathering the present results we can see that a three-dimensional real or complex Lie algebra is induced by a ql algebra iff $M^{ik} = 0$ or $N_j = 0$.

For each three-dimensional (real or complex) Lie algebra there exists a basis e_1, e_2, e_3 such that

$$M^{ij} = \operatorname{diag}(M^1, M^2, M^3), \quad N_i = (0,0,N)\,. \tag{3.7}$$

Then from (3.2) and (3.7) we have

$$M^3N = 0\,. \tag{3.8}$$

The commutators of the basic vectors are

$$[e_1,e_2] = M^3e_3, \quad [e_2,e_3] = M^1e_1 - Ne_2, \quad [e_3,e_1] = Ne_1 + M^2e_2\,. \tag{3.9}$$

Equations (3.3) and (3.7) yield

$$K_{ij} = -M^1M^2\delta^3{}_i\delta^3{}_j - M^3M^1\delta^2{}_i\delta^2{}_j - M^2M^3\delta^1{}_i\delta^1{}_j + N^2\delta^3{}_i\delta^3{}_j\,. \tag{3.10}$$

Assume first that $M^1 = M^2 = M^3 = 0$, $N \neq 0$ [compare with (3.6)]. In the real case these conditions define a three-dimensional real Lie algebra of class B and type V (the Bianchi–Behr classification). Rescaling, if necessary, the basic vectors e_i one can make $N = 1$.

Let now $N = 0$. In the real case this condition characterizes all three-dimensional real Lie algebras of class A. Then rescaling, if necessary, the basic vectors e_i we can make all nonzero M^i either 1 or -1. Thus one arrives at the following nonisomorphic three-dimensional real Lie algebras with $N = 0$ (we apply the Bianchi–Behr classification):

$$\begin{aligned}
&M^1 = M^2 = M^3 = 0 \leftrightarrow \mathrm{I}\,,\\
&M^1 = 1, \quad M^2 = M^3 = 0 \leftrightarrow \mathrm{II}\,,\\
&M^1 = M^2 = 1, \quad M^3 = 0 \leftrightarrow \mathrm{VII}_0\,,\\
&M^1 = 1, \quad M^2 = -1, \quad M^3 = 0 \leftrightarrow \mathrm{VI}_0\,,\\
&M^1 = M^2 = M^3 = 1 \leftrightarrow \mathrm{IX}\,,\\
&M^1 = M^2 = 1, \quad M^3 = -1 \leftrightarrow \mathrm{VIII}\,.
\end{aligned} \tag{3.11}$$

Concluding, type V and all the types in (3.11) exhaust all three-dimensional real Lie algebras induced by real ql algebras (see also Table I). One finds immediately that the Lie algebras so(3;R) and so(2,1;R) are of types IX and VIII, respectively.

2. n≠3

For $n = 1$ we have $C^1{}_{11} = 0$ and our one-dimensional real Abelian Lie algebra appears to be induced by a two-dimensional real ql algebra. Evidently the latter assertion holds true for the complex case, too.

Let $n > 1$ and $n \neq 3$. Contracting Eq. (2.4) with K_{pl} and utilizing (2.13) one gets

$$K_{i[j}K_{p]k} = 0\,, \tag{3.12}$$

where $[jp]$ stands for the antisymmetrization with respect to the indices j, p.

From (3.12) it follows that K_{ij} is of the form

$$K_{ij} = \lambda K_iK_j, \quad \lambda \in R \text{ and } K_i \in R\,. \tag{3.13}$$

Consider first the case of the rank $(K_{ij}) = 1$. Then, we can always choose the basic vectors e_i so that

$$K_{ij} = \epsilon\delta^n{}_i\delta^n{}_j, \quad \epsilon^2 = 1\,. \tag{3.14}$$

The condition (2.13) with the use of (3.14) gives

$$C^n{}_{ij} = 0\,. \tag{3.15}$$

The condition (2.4) amounts presently to

$$C^\alpha{}_{j\sigma}C^\sigma{}_{ki} = \epsilon\delta^n{}_j\delta^{n\alpha}_{ki} \tag{3.16}$$

(from now on lowercase Greek indices α, β, σ,... are assumed to run through $1,\ldots,n-1$). Equation (3.16) for $j = \beta$ gives

$$C^\alpha{}_{\beta\sigma}C^\sigma{}_{ki} = 0\,. \tag{3.17}$$

Inserting $j = n$, $k = n$, $i = \beta$, and then $j = n$, $k = \beta$, $i = \gamma$ into (3.16) we obtain

TABLE I. Real Lie algebras of dimension $2\leqslant n \leqslant 6$ induced by ql algebras. The terminology of Patera *et al.*[15,16] has been used. In the parentheses $(\cdots)$ the Bianchi–Behr type is given.

Dimension	Name	Nonzero commutation relations	Comments
$n=2$	$A_{2,1}$	$[e_2,e_1]=e_1$	solvable
	$A_{3,1}$ (II)	$[e_2,e_3]=e_1$	nilpotent
	$A_{3,3}$ (V)	$[e_3,e_1]=e_1$, $[e_3,e_2]=e_2$	solvable
	$A_{3,4}$ (VI_0)	$[e_3,e_1]=e_1$, $[e_3,e_2]=-e_2$	solvable
	$A_{3,6}$ (VII_0)	$[e_3,e_1]=e_2$, $[e_3,e_2]=-e_1$	solvable
$n=3$	$A_{3,8}$ (VIII)	$[e_1,e_2]=-e_3$, $[e_2,e_3]=e_1$, $[e_3,e_1]=e_2$	simple $\mathrm{so}(2,1;R)$ $\simeq \mathrm{sl}(2;R)$
	$A_{3,9}$ (IX)	$[e_1,e_2]=e_3$, $[e_2,e_3]=e_1$, $[e_3,e_1]=e_2$	simple $\mathrm{so}(3;R)$ $\simeq \mathrm{su}(2)$
$n=4$	$A_{4,5}^{1,1}$	$[e_4,e_1]=e_1$, $[e_4,e_2]=e_2$, $[e_4,e_3]=e_3$	solvable
	$A_{4,5}^{1,-1}$	$[e_4,e_1]=e_1$, $[e_4,e_2]=e_2$, $[e_4,e_3]=-e_3$	solvable
$n=5$	$A_{5,1}$	$[e_3,e_5]=e_1$, $[e_4,e_5]=e_2$	nilpotent
	$A_{5,4}$	$[e_2,e_4]=e_1$, $[e_3,e_5]=e_1$	nilpotent
	$A_{5,7}^{1,1,1}$	$[e_5,e_1]=e_1$, $[e_5,e_2]=e_2$, $[e_5,e_3]=e_3$, $[e_5,e_4]=e_4$	solvable
	$A_{5,7}^{1,1,-1}$	$[e_5,e_1]=e_1$, $[e_5,e_2]=e_2$ $[e_5,e_3]=e_3$, $[e_5,e_4]=-e_4$	solvable
	$A_{5,7}^{1,-1,-1}$	$[e_5,e_1]=e_1$, $[e_5,e_2]=e_2$, $[e_5,e_3]=-e_3$, $[e_5,e_4]=-e_4$	solvable
	$A_{5,17}^{\pm 1,0,0}$	$[e_5,e_1]=e_3$, $[e_5,e_2]=e_4$ $[e_5,e_3]=-e_1$, $[e_5,e_4]=-e_2$	solvable
$n=6$	$A_{6,3}$	$[e_1,e_2]=e_6$, $[e_1,e_3]=e_4$, $[e_2,e_3]=e_5$	nilpotent
	$A_{6,4}$	$[e_1,e_2]=e_5$, $[e_1,e_3]=e_6$, $[e_2,e_4]=e_6$	nilpotent
	$A_{6,5}^{a}$	$[e_1,e_3]=e_5$, $[e_1,e_4]=e_6$, $[e_2,e_3]=ae_6$, $[e_2,e_4]=e_5$ $(a\neq 0)$	nilpotent
		$[e_6,e_1]=e_1$, $[e_6,e_2]=e_2$, $[e_6,e_3]=e_3$, $[e_6,e_4]=e_4$, $[e_6,e_5]=e_5$	solvable
		$[e_6,e_1]=e_1$, $[e_6,e_2]=e_2$, $[e_6,e_3]=e_3$, $[e_6,e_4]=e_4$, $[e_6,e_5]=-e_5$	solvable
		$[e_6,e_1]=e_1$, $[e_6,e_2]=e_2$, $[e_6,e_3]=e_3$, $[e_6,e_4]=-e_4$, $[e_6,e_5]=-e_5$	solvable

$$C^{\alpha}{}_{\sigma}C^{\sigma}{}_{\beta}=\epsilon\delta^{\alpha}{}_{\beta}\,, \tag{3.18}$$

where $C^{\alpha}{}_{\sigma}:=C^{\alpha}{}_{n\sigma}$ and

$$C^{\alpha}{}_{\sigma}C^{\sigma}{}_{\beta\gamma}=0\,. \tag{3.19}$$

Contracting (3.19) with $C^{\rho}{}_{\alpha}$ and employing (3.18) one finds

$$C^{\rho}{}_{\beta\gamma}=0 \tag{3.20}$$

[Notice that Eq. (3.17) is a consequence of (3.20).] Thus the only nonzero structure constants are of the form $C^{\alpha}{}_{n\beta}$.

Assume first

$$\epsilon=1\,. \tag{3.21}$$

Hence the eigenvalues of the matrix $(C^{\alpha}{}_{\beta})$ are ± 1. It is a well-known result in the linear algebra (see Ref. 4, Sec. 12, §88) that by some linear transformation the matrix $(C^{\alpha}{}_{\beta})$ can be brought to the following form:

$$(C^{\alpha}{}_{\beta})=\begin{pmatrix} A_1 & & \\ 0 & \ddots & 0 \\ & & A_r \end{pmatrix}, \tag{3.22}$$

where the matrices $A_1,\ldots,A_r$ are of the form

$$A_{\tilde{\nu}}=\begin{pmatrix} 0 & 0 & \cdots & 0 & a_1 \\ 1 & 0 & \cdots & 0 & a_2 \\ 0 & 1 & \cdots & 0 & a_3 \\ \cdots & & \cdots & & \cdots \\ 0 & 0 & \cdots & 1 & a_{f\tilde{\nu}} \end{pmatrix}, \tag{3.23}$$

where $\tilde{\nu} = 1,...,r, f_1 + \cdots + f_r = n - 1$. From (3.18) with (3.21) one infers that the matrices $A_{\tilde{\nu}}$ are of the forms

$$A_{\tilde{\nu}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{or} \quad A_{\tilde{\nu}} = 1 \quad \text{or} \quad A_{\tilde{\nu}} = -1 . \tag{3.24}$$

It is easy to check that a suitable linear transformation brings the leftmost matrix in (3.24) to the diagonal form $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Gathering the above results, we can see that if, in Eq. (3.14), $\epsilon = 1$, then there exists a basis e_i such that

$$C^n{}_{ij} = 0 = C^\alpha{}_{\beta\gamma}, \quad C^\alpha{}_{n\beta} = \epsilon_\beta \delta^\alpha{}_\beta, \quad \epsilon_\beta = \pm 1, \tag{3.25}$$

and also (3.14) with $\epsilon = 1$ holds true. (For three-dimensional real Lie algebras this is the case of the Bianchi–Behr type V, $\epsilon_1 = 1 = \epsilon_2$, or of the type VI_0, $\epsilon_1 = 1 = -\epsilon_2$.)

Assume now

$$\epsilon = -1 . \tag{3.26}$$

Then the eigenvalues of $(C^\alpha{}_\beta)$ are $\pm i$. Utilizing (3.18) with (3.26) one finds that the matrices $A_{\tilde{\nu}}$, defined by (3.22) and (3.23), are of the form

$$A_{\tilde{\nu}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.27}$$

Therefore, by some obvious changing of a basis, the matrix $(C^\alpha{}_\beta)$ can be brought to the following form:

$$(C^\alpha{}_\beta) = \begin{pmatrix} 0, & -I_{(1/2)(n-1)} \\ I_{(1/2)(n-1)}, & 0 \end{pmatrix}, \tag{3.28}$$

where $I_{(1/2)(n-1)}$ is the identity matrix of degree $\frac{1}{2}(n-1)$. Equations (3.15), (3.20), and (3.14) with $\epsilon = -1$ hold true. Note that $(C^\alpha{}_\beta)$ satisfying Eq. (3.18) with $\epsilon = -1$ defines an almost complex structure[18] on a real vector space generated by $e_1,...,e_{n-1}$. (One can easily check that for three-dimensional real Lie algebras, precisely the type VII_0 belongs to the just considered algebras.)

Finally let us remark that real Lie algebras induced by ql algebras and such that

$$\text{rank}(K_{ij}) = 1 \tag{3.29}$$

are solvable but non-nilpotent Lie algebras.[2,5,19,20]

Consider now the case of

$$K_{ij} = 0 . \tag{3.30}$$

From (2.4) with (3.30) it follows that

$$C^i{}_{jm} C^m{}_{kl} = 0 . \tag{3.31}$$

Hence our real Lie algebras appear to be either Abelian Lie algebras (when $C^i{}_{jk} = 0$) or nilpotent Lie algebras of the nilpotency class 2.[2,5,19,20] [For three-dimensional real Lie algebras this is the case of type I (Abelian) or of type II (nilpotent of class 2).] To close the considerations found in this subsection (III A) we list (Table I) all nonisomorphic real Lie algebras of dimension $2 \leqslant n \leqslant 6$ induced by ql algebras omitting the Lie algebras which are algebraic sums of algebras of lower dimension. We follow the works of Behr *et al.*,[9] Ellis and MacCallum,[10] MacCallum,[11,12] and Patera *et al.*[15,16]

B. Complex Lie algebras

In the complex case the considerations are very similar to those concerning real Lie algebras. The only difference is that some algebras that are nonisomorphic on the real level appear to be isomorphic after complexification. Thus one finds that for $n = 3$ the complex Lie algebras VIII and IX are isomorphic to $\text{sl}(2;C)$ and also the complex types VI_0 and VII_0 overlap. Generally we conclude that for every $n > 1$ an n-dimensional complex Lie algebra induced by a complex ql algebra for which $\text{rank}(K_{ij}) = 1$ possesses a basis $e_1,...,e_n$ such that the structure constants with respect to it take the form (3.25).

IV. CONCLUDING REMARKS

The main results of the present paper can be summarized as follows.

An n-dimensional real Lie algebra appears to be the one induced by a ql algebra if and only if it belongs to one of the following types.

(1) An Abelian Lie algebra.

(2) A nilpotent Lie algebra of the nilpotency class 2, i.e., a non-Abelian Lie algebra for which $C^i{}_{jm} C^m{}_{kl} = 0$, $i,j,... = 1,...,n$.

(3) A solvable Lie algebra for which there exists a basis $e_1,...,e_n$ such that $C^n{}_{ij} = 0 = C^\alpha{}_{\beta\gamma}$, $C^\alpha{}_{n\beta} = \epsilon_\beta \delta^\alpha{}_\beta$; $\epsilon_\beta = \pm 1$; $i,j, = 1,...,n$; $\alpha, \beta, \gamma = 1,...,n-1$.

(3′) A solvable Lie algebra for which there exists a basis $e_1,...,e_n$ such that $C^n{}_{ij} = 0 = C^\alpha{}_{\beta\gamma}$, and the $(n-1) \times (n-1)$ matrix

$$(C^\alpha{}_\beta) = \begin{pmatrix} 0, & -I_{(1/2)(n-1)} \\ I_{(1/2)(n-1)}, & 0 \end{pmatrix},$$

where $C^\alpha{}_\beta := C^\alpha{}_{n\beta}$, and $I_{(1/2)(n-1)}$ is the identity matrix of degree $\frac{1}{2}(n-1)$; $i,j = 1,...,n$; $\alpha, \beta, \gamma = 1,...,n-1$.

(4) A Lie algebra isomorphic to $\text{su}(2)$.

(4′) A Lie algebra isomorphic to $\text{sl}(2;R)$.

In the complex case one has the types (1), (2), (3) and the following.

(4″) A Lie algebra isomorphic to $\text{sl}(2;C)$.

The above presented list gives also a classification of all real or complex local Lie groups endowed with the composition laws of the form (2.15).

Finally let us note that our considerations on ql algebras appear to be closely related to the problem of a definition of a cross product in a vector space of an arbitrary dimension. (We are indebted to Professor J. Adem for this suggestion. The paper on this subject has been submitted to the Journal of Mathematical Physics[21]; see also Ref. 22.)

ACKNOWLEDGMENTS

One of us (M.P.) wishes to thank the staff of the Department of Physics at Centro de Investigación y de Estudions Avanzados del I.P.N., México 14, D.F., for the warm hospitality during his stay at this department where our paper was conceived. He is especially indebted to Dr. A. Zepeda and Dr. Alberto García for their assistance. M. P. also wishes to express his gratitude to Dr. E. Campesino, Secretario Académico of CINVESTAV for his assistance.

The work of one of us (M.P.) was supported in part by the CONACYT, México, D.F., and by the Centro de Investigación y de Estudios Avanzados del I.P.N., México 14, D.F., Mexico.

[1] J. Plebański, "On the generators of unitary and pseudo-orthogonal groups," CIEA Report, México, 1966.
[2] N. Bourbaki, *Groups et Algébras de Lie* (Hermann, Paris, 1972).
[3] L. Pontrjagin, *Topological Groups* (Gordon–Breach, New York, 1966).
[4] B. L. van der Waerden, *Algebra I and II* (Springer, Berlin, 1967).
[5] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[6] B. Mielnik and J. Plebański, Ann. Inst. H. Poincaré **XII**, 215 (1970).
[7] J. F. Plebański, "On the linear unitary transformations of two canonical variables," CIEA Report, México, 1986 (unpublished).
[8] L. Bianchi, Mem. Soc. Ital. Sci. **11**, 267 (1897).
[9] F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, J. Math. Phys. **9**, 497 (1968).
[10] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[11] M. A. H. MacCallum, "The mathematics of anisotropic spatially-homogeneous cosmologies," lectures given at the First Cracow International Summer School of Cosmology 1978, preprint 1979.
[12] M. A. H. MacCallum, "On the classification of the real four-dimensional Lie algebras," preprint.
[13] G. M. Mubarakzyanov, Izv. Vyss. Uch. Zav. Mat. **32**, 114 (1963); **34**, 99 (1963); **35**, 104 (1963).
[14] V. V. Morozov, Isz. Vyss. Uch. Zav. Mat. **5**, 161 (1958).
[15] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[16] J. Patera and P. Winternitz, J. Math. Phys. **18**, 1449 (1977).
[17] P. Spindel, "Gravity before supergravity," in *Supersymmetry* (Plenum, New York, 1985).
[18] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, London, 1963), Vol. 2, Chap. IX.
[19] J. A. Bahturin, *Lectures of Lie Algebras* (Academie, Berlin, 1978).
[20] J. Dixmier, *Enveloping Algebras* (Academie, Berlin, 1977).
[21] J. F. Plebański and M. Przanowski, "Notes on a cross product of vectors," submitted to J. Math. Phys.
[22] W. S. Massey, Am. Math. Mon. **90**, 697 (1983).

Partial *-algebras of matrices and operators

Giuseppina Epifanio and Camillo Trapani
Istituto di Fisica dell'Università, Via Archirafi 36, I-90123 Palermo, Italy

(Received 9 June 1987; accepted for publication 7 October 1987)

The set $\mathscr{M}_\infty$ of infinite matrices and the set Q_∞ of squarable matrices are considered as partial *-algebras. The connection between Q_∞ and two partial *-algebras of closed operators is studied. Conditions for a matrix representation in "von Neumann's sense" of a family of closed operators are given.

I. INTRODUCTION

Since the beginning of the theory of linear operators, the matrix calculus has played so relevant a role as to achieve the status of a classical argument.

This growth of importance, mainly due to the quantum mechanical custom of replacing indifferently matrices with operators, has given rise to a wide literature (some papers date back to the pioneer age of quantum mechanics), essentially concerned with the study of the relationship between matrices and operators, which has emphasized the fact that a theory of matrix representation of unbounded operators is not at all a slight modification of the bounded case.

From the algebraic point of view the situation was clear as far as one had to deal with finite or bounded matrices. As a matter of fact, the operators occurring in the applications are often unbounded and unbounded matrices may exhibit a behavior as singular as that of operators. For instance, the set of all infinite matrices does not carry any usual algebraic structure because of two well known features: the multiplication of two matrices is not always defined and, even if it is, the associative property may fail to be true.

The question arises whether any help from this point of view can be given by some partial algebraic structure like those introduced and studied in the last years by some authors (Refs. 1–3).

In Sec. II, we show that the answer is affirmative: the set $\mathscr{M}_\infty$ of all infinite matrices and the set Q_∞ of squarable matrices are in fact partial *-algebras.

In Sec. III, we study the correspondence between Q_∞ and the sets $\mathfrak{C}(\mathscr{D}_0)$, and $\mathfrak{C}^*(\mathscr{D}_0)$ of, respectively, $\mathscr{D}_0$-minimal and $\mathscr{D}_0$-maximal closed operators on the linear hull $\mathscr{D}_0$ of an arbitrary orthonormal basis.

In Sec. IV, the problem of the matrix representation in "von Neumann's sense" of families of closed operators is examined.

II. PARTIAL ALGEBRAIC STRUCTURE IN THE SET OF INFINITE MATRICES

The main concept we have to deal with is that of a partial *-algebra due to Borchers (Ref. 2) and studied by Antoine and Karwowski [Ref. 1(a)].

For the reader's convenience we recall the basic definitions.

Definition 2.1: A partial *-algebra is a vector space $\mathscr{A}$ with involution $x \to x^\dagger$ [i.e., $(x+\lambda y)^\dagger = x^\dagger + \bar{\lambda} y^\dagger$; $x^{\dagger\dagger} = x$] and a subset $\Gamma \subseteq \mathscr{A} \times \mathscr{A}$ such that (i) $(x,y) \in \Gamma$ implies $(y^\dagger, x^\dagger) \in \Gamma$; (ii) (x,y) and $(x,z) \in \Gamma$ imply $(x, y+\lambda z) \in \Gamma$; and (iii) if $(x,y) \in \Gamma$, then there exists an element $x \circ y \in \mathscr{A}$ and for this multiplication the distributive property holds in the following sense: if $(x,y) \in \Gamma$ and $(x,z) \in \Gamma$ then

$$x \circ y + x \circ z = x \circ (y+z).$$

Furthermore, $(x \circ y)^\dagger = y^\dagger \circ x^\dagger$.

Notice that it is not required that the $\circ$-product be associative.

The partial *-algebra $\mathscr{A}$ is said to have a unit if there exists an element $1 \in \mathscr{A}$ (necessarily unique) such that $1^\dagger = 1$, $(1,x) \in \Gamma$, and $1 \circ x = x \circ 1 = x \ \forall x \in \Gamma$.

Whenever $(x,y) \in \Gamma$, we say that x is a left multiplier of y [and write $x \in L(y)$] or y is a right multiplier of x [$y \in R(x)$].

If $S \subseteq \mathscr{A}$ we put $LS = \cap_{x \in S} L(x)$, $RS = \cap_{x \in S} R(x)$; $MS = LS \cap RS$. If $s = \mathscr{A}$, $M\mathscr{A}$ is called the set of universal multipliers of $\mathscr{A}$.

A particularly interesting situation occurs when

$$\{(x,y) \in \mathscr{A} \times \mathscr{A} \mid x \in \mathscr{A}_0 \text{ or } y \in \mathscr{A}_0\} \subseteq \Gamma,$$

where $\mathscr{A}_0 \subseteq \mathscr{A}$ is a *-algebra.

In this case we say, following Lassner (Ref. 3) that $\mathscr{A}$ is a quasi-*-algebra with distinguished *-algebra $\mathscr{A}_0$.

A quasi-*-algebra $(\mathscr{A}, \mathscr{A}_0)$ is said to be a topological quasi-*-algebra if $\mathscr{A}$ is endowed with a locally convex topology τ such that (i) $\mathscr{A}_0$ is dense in $\mathscr{A}$; (ii) the multiplications $x \to x \circ y$ and $x \to y \circ x$ are continuous for every $x \in \mathscr{A}_0$; and (iii) the map $x \to x^\dagger$ is continuous.

We will now show that the set $\mathscr{M}_\infty$ of all infinite matrices

$$\mathscr{M}_\infty = \{(A_{\mu\nu}),\ A_{\mu\nu} \in \mathbb{C},\ \mu,\nu \in \mathbb{N}\}$$

carries a very natural structure of partial *-algebra.

Proposition 2.2: (i) In $\mathscr{M}_\infty$, the map $(A_{\mu\nu}) \to (A^*_{\mu\nu})$, where $A^*_{\mu\nu} = \overline{A_{\nu\mu}}$, defines an involution and the usual rows–columns product defines a partial multiplication on the set

$$\Gamma = \{((A_{\mu\nu}),(B_{\mu\nu})) \colon \Sigma_\rho A_{\mu\rho} B_{\rho\nu} < \infty \quad \forall \mu,\nu \in \mathbb{N}\}.$$

Thus $(\mathscr{M}_\infty, \Gamma)$ is a partial *-algebra with unit $1 = (\delta_{\mu\nu})$.

(ii) The set $R\mathscr{M}_\infty$ of the right universal multipliers of $\mathscr{M}_\infty$ consists exactly of the matrices with a finite number of nonzero elements in each column. Analogously, the set $L\mathscr{M}_\infty$ of the left-universal multipliers of $\mathscr{M}_\infty$ consists of those matrices having a finite number of nonzero elements in each row.

(iii) The set $M\mathscr{M}_\infty$ of the universal multipliers of $\mathscr{M}_\infty$, i.e., $M\mathscr{M}_\infty = L\mathscr{M}_\infty \cap R\mathscr{M}_\infty$, is a *-algebra; then $\mathscr{M}_\infty$ is a

 0022-2488/88/030536-05$02.50

quasi-*-algebra with distinguished *-algebra $M\mathscr{M}_\infty$.

Proof: (i) is straightforward.

(ii) We prove only that a matrix $(B_{\mu\nu})$ having an infinite number of nonzero elements in some column does not belong to $R\mathscr{M}_\infty$. Let, in fact, $(A_{\mu\nu})$ be a matrix such that for some $\sigma\in\mathbb{N}$, $\Sigma_\nu A_{\sigma\nu}$ does not converge. Without loss of generality we may suppose $B_{\mu\rho}\neq 0$ $\forall\mu$, for fixed ρ. Put $C_{\sigma\nu} = A_{\sigma\nu}B_{\nu\rho}^{-1}$. Then

$$\sum_\nu C_{\sigma\nu}B_{\nu\rho} = \sum_\nu A_{\sigma\nu}B_{\nu\rho}^{-1}B_{\nu\rho} = \sum_\nu A_{\sigma\nu} = \infty.$$

(iii) Let $(A_{\mu\nu})$, $(B_{\mu\nu})\in M\mathscr{M}_\infty$ and

$$C_{\mu\nu} = \sum_\rho A_{\mu\rho}B_{\rho\nu};$$

suppose that there exists μ such that $C_{\mu\nu}\neq 0$ $\forall\nu\in\mathbb{N}$. Let ρ_0 be the minimal number such that $A_{\mu\rho} = 0$, $\rho>\rho_0$; necessarily some of the $B_{\rho\nu}$'s are not zero for $\rho\leqslant\rho_0$. Then the matrix $(B_{\rho\nu})$ has infinite nonzero elements in the σth row with $1\leqslant\sigma\leqslant\rho_0$. This is a contradiction.

Remark: In $\mathscr{M}_\infty$ a topology τ_0 can be introduced by means of the set of seminorms

$$p_{\mu\nu}[(A_{\rho\sigma})] = |A_{\mu\nu}| \quad \forall\mu,\nu\in\mathbb{N}.$$

It is easily proved that $\mathscr{M}_\infty[\tau_0]$ is a Fréchet space and $M\mathscr{M}_\infty$ is τ_0 dense in $\mathscr{M}_\infty$; moreover operations are continuous. By the previous proposition we get that $\mathscr{M}_\infty$ is a topological quasi-*-algebra over $M\mathscr{M}_\infty$.

One of the most remarkable subsets of $\mathscr{M}_\infty$ is the set $\mathscr{M}_b$ of bounded matrices

$$\mathscr{M}_b = \Big\{(A_{\mu\nu})\in\mathscr{M}_\infty: \sum_\mu\Big|\sum_\nu A_{\mu\nu}\xi_\nu\Big|^2\leqslant C^2\sum_\mu|\xi_\mu|^2 \ \ \forall(\xi_\mu)_\mu\in l_2\Big\}.$$

As is well known, $\mathscr{M}_b$ is a *-algebra isomorphic to the *-algebra $B(\mathscr{H})$ of bounded operators in Hilbert space. But $\mathscr{M}_\infty$ contains also many *-algebras of unbounded matrices, like the set $\mathscr{M}_d$, considered in Ref. 4, which is isomorphic to the *-algebra $C_\mathscr{D} = \mathscr{L}^\dagger(\mathscr{D})$ of unbounded operators. Here we are interested in the following subset of $\mathscr{M}_\infty$:

$$Q_\infty = \Big\{(A_{\mu\nu})\in\mathscr{M}_\infty: \sum_\mu|A_{\mu\nu}|^2<\infty,\ \sum_\nu|A_{\mu\nu}|^2<\infty\Big\}.$$

Following von Neumann (Ref. 5) we call elements of Q_∞ squarable ("quadrierbar") matrices. Clearly, $\mathscr{M}_b\subseteq Q_\infty$ and $\mathscr{M}_d\subseteq Q_\infty$.

As is known Q_∞ plays an important role in the study of the correspondence between matrices and operators in scalar product space. Actually, a necessary condition for a matrix $(A_{\mu\nu})$ to represent an operator is that $(A_{\mu\nu})\in Q_\infty$.

Proposition 2.3: (i) Q_∞ is stable under involution.

(ii) Q_∞ is a complex vector space under the usual operations.

Proof: (i) is obvious. (ii) follows from the inequalities

$$\sum_\mu|A_{\mu\nu}+B_{\mu\nu}|^2\leqslant\sum_\mu(|A_{\mu\nu}|+|B_{\mu\nu}|)^2 \leqslant 2\Big(\sum_\mu|A_{\mu\nu}|^2+\sum_\mu|B_{\mu\nu}|^2\Big)<\infty.$$

Thus Q_∞ inherits from $\mathscr{M}_\infty$ the structure of vector space stable under involution; but for the multiplication the situation is more complicated. In fact, the product of two elements of Q_∞ is always defined, as a consequence of the inequality

$$\Big|\sum_\rho A_{\mu\rho}B_{\rho\nu}\Big|^2\leqslant\sum_\rho|A_{\mu\rho}|^2\sum_\rho|B_{\rho\nu}|^2<\infty,$$

but it need not belong to Q_∞. Then Q_∞ is neither an algebra nor a partial *-subalgebra of $\mathscr{M}_\infty$. Nevertheless for a suitable choice of the set Γ_Q, (Q_∞,Γ_Q) becomes a partial *-algebra.

Proposition 2.4: (Q_∞,Γ_Q) is a partial *-algebra with unity $\mathbf{1} = (\delta_{\mu\nu})$ if

$$\Gamma_Q = \Big\{((A_{\mu\nu}),(B_{\mu\nu}))\in Q_\infty\times Q_\infty: \sum_\mu\Big|\sum_\nu A_{\mu\nu}B_{\nu\rho}\Big|^2<\infty,\ \sum_\mu\Big|\sum_\nu A_{\rho\nu}B_{\nu\mu}\Big|^2<\infty\Big\}.$$

Proof: We already proved that Q_∞ is a vector space stable under involution. By the definition itself it follows that if $((A_{\mu\nu}),(B_{\mu\nu}))\in\Gamma_Q$ then $((B^*_{\mu\nu}),(A^*_{\mu\nu}))\in\Gamma_Q$.

Let now $((A_{\mu\nu}),(B_{\mu\nu}))\in\Gamma_Q$ and $((A_{\mu\nu}),(C_{\mu\nu}))\in\Gamma_Q$. We have

$$\sum_\mu\Big|\sum_\nu A_{\mu\nu}(B_{\nu\rho}+\lambda C_{\nu\rho})\Big|^2 = \sum_\mu\Big|\sum_\nu A_{\mu\nu}B_{\nu\rho}+\lambda\sum_\nu A_{\mu\nu}C_{\nu\rho}\Big|^2 \leqslant 2\Big(\sum_\mu\Big|\sum_\nu A_{\mu\nu}B_{\nu\rho}\Big|^2+|\lambda|^2\sum_\mu\Big|\sum_\nu A_{\mu\nu}C_{\nu\rho}\Big|^2\Big)<\infty.$$

In an analogous way the other condition can be proved and therefore $((A_{\mu\nu}),(B_{\mu\nu}+\lambda C_{\mu\nu}))\in\Gamma_Q$.

At this point the equality $((A_{\mu\nu})\cdot(B_{\mu\nu}))^* = (B^*_{\mu\nu})(A^*_{\mu\nu})$ and the distributive property, in the sense of Definition 2.1, can be easily proved.

III. MATRICES AND OPERATORS IN SCALAR PRODUCT SPACES

The wide gap existing between the matrix representation of bounded operators and that of unbounded ones has been already mentioned in the Introduction. Let us recall shortly the terms of the question.

If $\{A_i\}$ is a family of closed operators such that there is a dense linear manifold Δ with $\Delta\subset\cap_i[D(A_i)\cap D(A_i^*)]$ and $A_i\restriction\Delta = A_j\restriction\Delta$ $\forall i,j$, it is always possible to find a matrix $(A_{\mu\nu})$ such that for any vector $\varphi = \Sigma_\nu\xi_\nu e_\nu\in D(A_i)$ its image $\psi = A_i\varphi$ can be determined by the matrix $(A_{\mu\nu})$ [the $e_\nu\in\Delta$ $\forall\nu\in\mathbb{N}$ and $A_{\mu\nu} = (Ae_\nu,e_\mu)$; from now on the Hilbert space is always supposed to be separable]. Nevertheless, the matrix $(A_{\mu\nu})$ has in Hilbert space a domain, in general, larger than the $D(A_i)$'s; therefore it defines an operator $T\supseteq A$ and no general connection between the A_i's and T is known.

For this reason, we will use the words "matrix representation" when a prescription to find the domain is also given. One possible way to do this is to use the notion of matrix representation in von Neumann's sense (for a more detailed discussion of this point, see Ref. 6).

No problems arise clearly if the domain of the operators can be considered to be the whole space as it happens for bounded operators or for the operators of $C_{\mathscr{D}}$ as shown in Ref. 4.

Here we will show that some one-to-one correspondence (preserving operations) between some class of operators and matrices of $\mathscr{M}_\infty$ can be established.

We will deal with two partial *-algebras of closed operators introduced by Antoine and Karwowski in Ref. 1(a). A detailed study has been made by Antoine and Mathot in Ref. 1(b).

Let $\overline{C}(\mathscr{D},\mathscr{H})$ be the set of closed operators A in $\mathscr{H}$ such that $\mathscr{D}\subseteq D(A)\cap D(A^*)$. Given $A\in\overline{C}(\mathscr{D},\mathscr{H})$ define $A^\ddagger = \overline{A^*\restriction\mathscr{D}}$ and $A^\dagger = [A\restriction\mathscr{D}]^*$. The operators $A^\ddagger, A^{\ddagger\ddagger}$ are called $\mathscr{D}$ minimal (i.e., $\mathscr{D}$ is a core for them). The operators $A^\dagger$, $A^{\dagger\dagger}$ are called $\mathscr{D}$ maximal (i.e., they are the adjoints of $\mathscr{D}$-minimal operators; in fact $A^{\dagger\dagger}=A^{\ddagger *}$ and $A^\dagger = A^{\ddagger\ddagger *}$). Let us denote by $\mathfrak{C}(\mathscr{D})$ the set of $\mathscr{D}$-minimal operators and by $\mathfrak{C}^*(\mathscr{D})$ the set of $\mathscr{D}$-maximal ones. Here $\mathfrak{C}(\mathscr{D})$ has a partial *-algebra structure when one defines the operations as

$$A\hat{+}B=\overline{(A+B)\restriction\mathscr{D}},\quad \lambda A=\overline{\lambda A\restriction\mathscr{D}},$$

$$A\rightarrow A^\ddagger=\overline{A^*\restriction\mathscr{D}},\quad A\square B=(A*B)^{\ddagger\ddagger},$$

where $A*B=[B^\dagger(A^\dagger\restriction\mathscr{D})]^*$ defined whenever $B\mathscr{D}\subseteq D(A^{\dagger\dagger})$ and $A^\dagger\mathscr{D}\subseteq D(B^\dagger)$.

Analogously, the set $\mathfrak{C}^*(\mathscr{D})$ can be considered as a partial *-algebra with the following operations: $A\check{+}B = [(A^*+B^*)\restriction\mathscr{D}]^*$, $\lambda A=[\bar\lambda A^*\restriction\mathscr{D}]^*$; $A\rightarrow A^\dagger = [A\restriction\mathscr{D}]^*$ and partial multiplication $A*B$ defined as above.

We have examined in Sec. II the partial *-algebra of matrices Q^∞. The question arises now: does it correspond to some partial *-algebra of operators?

Let us first remark that given a matrix $(A_{\mu\nu})\in Q_\infty$ and a basis (e_ν) in $\mathscr{H}$ two closed operators can be determined in an easy way. The first one is the operator $R(A)$ which is the closure of the operator $R_0(A)$ defined by the matrix $A=(A_{\mu\nu})$ on the linear hull $\mathscr{D}_0$ of the basis vectors.

The second one is the operator $S(A)$ defined by $(A_{\mu\nu})$ on the domain

$$D(S(A))=\left\{\varphi=\sum_\nu \xi_\nu e_\nu\in\mathscr{H}\;\middle|\;\sum_\mu\left|\sum_\nu A_{\mu\nu}\xi_\nu\right|^2<\infty\right\}.$$

In order to show that $S(A)$ is closed we prove the following.

Lemma 3.1:

$$S(A)=R(A^*)^*=R_0(A^*)^*.$$

Proof: Let

$$\varphi=\sum_1^m{}_\nu\,\xi_\nu e_\nu\in\mathscr{D}_0,\quad \psi=\sum_1^\infty{}_\nu\,\eta_\nu e_\nu\in D(S(A));$$

we get

$$(R_0(A^*)\varphi,\psi)=\sum_1^m{}_\nu\,\xi_\nu\sum_1^\infty{}_\mu\,\overline{A_{\nu\mu}}\,\eta_\mu=\sum_\nu\xi_\nu\,\overline{(S(A)\psi)_\nu}$$
$$=(\varphi,S(A)\psi),$$

i.e., $S(A)\subseteq R_0(A^*)=R(A^*)^*$.

Conversely, let us suppose that $\psi\in D(R_0(A^*)^*)$ with $\psi=\Sigma_\mu\eta_\mu e_\mu$ then

$$(R_0(A^*)e_\nu,\psi)=(e_\nu,R_0(A^*)^*\psi)=\sum_\mu A^*_{\mu\nu}\bar\eta_\mu$$
$$=\overline{\sum_\mu A_{\nu\mu}\eta_\mu}.$$

Therefore $\Sigma_\mu A_{\mu\nu}\eta_\mu$ represents the νth component of the vector $R_0(A^*)^*\psi$; then it must be

$$\sum_\nu\left|\sum_\mu A_{\nu\mu}\eta_\mu\right|^2<\infty.$$

Thus $\psi\in D(S(A))$.

Since, by definition, $R(A)$ is a $\mathscr{D}_0$-minimal operator, $S(A)$ is $\mathscr{D}_0$ maximal.

This suggests the possibility to find two classes of operators corresponding to Q_∞ for a given basis (e_ν). We can, in fact, define for a fixed basis (e_ν) the following two maps [$\mathscr{D}_0$ being the linear hull of (e_ν)]:

$$R: A=(A_{\mu\nu})\in Q_\infty\rightarrow R(A)\in\mathfrak{C}(\mathscr{D}_0),$$
$$S: A=(A_{\mu\nu})\in Q_\infty\rightarrow S(A)\in\mathfrak{C}^*(\mathscr{D}_0).$$

Let us now discuss the question whether R and S are *-isomorphisms of partial *-algebras.

We recall first the following definition [Ref. 1(a)].

Definition 3.2: A homomorphism of a partial *-algebra $\mathfrak{M}$ into another one $\mathfrak{N}$ is a linear map σ: $\mathfrak{M}\rightarrow\mathfrak{N}$ such that

(i) $\sigma(x^\dagger)=[\sigma(x)]^\dagger$;

(ii) if $x\in L(y)$ in $\mathfrak{M}$, then $\sigma(x)\in L(\sigma(y))$ in $\mathfrak{N}$ and $\sigma(x)\cdot\sigma(y)=\sigma(xy)$.

Clearly a *-isomorphism of the partial *-algebras $\mathfrak{M}$ and $\mathfrak{N}$ is a homomorphism of $\mathfrak{M}$ into $\mathfrak{N}$ which is one-to-one and onto.

Proposition 3.3: The map R: $A\in Q_\infty\rightarrow R(A)\in\mathfrak{C}(\mathscr{D})$ is a *-isomorphism of the partial *-algebras $(Q_\infty,\Gamma_Q,\circ)$ and $(\mathfrak{C}(\mathscr{D}_0),\Gamma,\square)$.

Proof: One readily checks that

$$R(A^*)=R(A)^\ddagger$$

and

$$R(A+\lambda B)=R(A)\hat{+}\lambda R(B).$$

Let us now show that if $(A,B)\in\Gamma_Q$ then $(R(A),R(B))\in\Gamma$. It is clearly enough to prove that

$$R(A)e_\mu\in D(S(A))$$

and

$$R(A^*)e_\mu\in D(S(B^*))\quad\forall\mu\in\mathbb{N}.$$

We have in fact $(R(B)e_\mu)_\nu=B_{\nu\mu}$, $(R(A^*)e_\mu)_\nu=\overline{A_{\mu\nu}}$. Since $(A,B)\in\Gamma_Q$ we get

$$\sum_\mu\left|\sum_\nu A_{\mu\nu}B_{\nu\rho}\right|^2<\infty$$

and

$$\sum_\mu\left|\sum_\nu B_{\nu\mu}A_{\rho\nu}\right|^2<\infty;$$

these, respectively, mean that $R(B)e_\nu\in D(S(A))$ and $R(A^*)e_\nu\in D(S(B^*))$. Now

$R(A)\Box R(B)$

$$= (R(A)*R(B))^{\ddagger\ddagger}$$
$$= [R_0(B)^*R_0(A^*)]^{*\ddagger\ddagger} = [S(B^*)R_0(A^*)]^{*\ddagger\ddagger}$$
$$= [R_0(B^*A^*)]^{*\ddagger\ddagger} = [S(AB)]^{\ddagger\ddagger} = R(AB).$$

The map R is a *-isomorphism since $\mathscr{D}_0$ is a core for all the operators of $\mathfrak{C}(\mathscr{D})$.

Proposition 3.4: The map $S: A\in Q_\infty \to S(A)\in\mathfrak{C}^*(\mathscr{D})$ is a *-isomorphism of the partial *-algebras $(Q_\infty,\Gamma_Q,\circ)$ and $(\mathfrak{C}^*(\mathscr{D}),\Gamma_*,*)$.

Proof: One readily checks that $S(A^*) = S(A)^\dagger$ and $S(A+\lambda B) = S(A) \check{+} \lambda S(B)$. Let now $(A,B)\in\Gamma_Q$, we need to prove $(S(A),S(B))\in\Gamma_*$. Since $S(B)\mathscr{D}_0 = R_0(B)\mathscr{D}_0$ and $S(A^*)\mathscr{D}_0 = R_0(A^*)\mathscr{D}_0$ the statement $(S(A),S(B))\in\Gamma_*$ is equivalent to the statement $(R(A),R(B))\in\Gamma$ already proved in Proposition 3.3. It remains only to prove that $S(AB) = S(A)*S(B)$. But

$$S(A)*S(B) = [S(B^*)R_0(A^*)]^*$$
$$= R(B^*A^*) = S(AB).$$

Since, as remarked in Ref. 1, $\mathfrak{C}(\mathscr{D}_0)$ and $\mathfrak{C}^*(\mathscr{D}_0)$ are isomorphic and $R: Q_\infty \to \mathfrak{C}(\mathscr{D}_0)$ is a *-isomorphism, so is S: $Q^\infty \to \mathfrak{C}^*(\mathscr{D}_0)$.

IV. MATRIX REPRESENTATION IN von NEUMANN SENSE

We return now to the problem of matrix representation. We have already seen that a matrix $A = (A_{\mu\nu})\in Q_\infty$ and a basis (e_ν) in $\mathscr{H}$ identify two closed operators in Hilbert space, namely, $R(A)$ and $S(A)$, which are, respectively, $\mathscr{D}_0$ minimal and $\mathscr{D}_0$ maximal, where $\mathscr{D}_0$ is the linear hull of the basis vectors.

If A is a closed operator in $\mathscr{H}$ with $\Delta(A) = D(A)\cap D(A^*)$ dense in $\mathscr{H}$ and (e_ν) is a basis in $\Delta(A)$, then the matrix $(A_{\mu\nu})$ belongs to Q_∞; thus it identifies the two operators $R(A)$ and $S(A)$. According to von Neumann, the "basis for a matrix representation of A" is a basis (e_ν) such that $R(A) = A$ and he proved that such a basis always exists for a closed symmetric operator. In Ref. 6 we gave a necessary and sufficient condition for a closed operator to have a matrix representation in this sense. Clearly if $A = R(A)$ then $A^* = S(A^*)$. We will then split the definition of matrix representation into two parts.

Definition 4.1: Let A be a closed operator with dense domain in $\mathscr{H}$ such that $\Delta(A)$ is also dense and (e_ν) a basis in $\Delta(A)$. Put $A_{\mu\nu} = (Ae_\nu,e_\mu)$.

We say that A admits a matrix representation of the first kind if $A = R(A)$.

We say that A admits a matrix representation of the second kind if $A = S(A)$.

Proposition 4.2: The operator A admits a representation of the first kind if, and only if, A^* admits a representation of the second kind.

Clearly, the map $R: Q_\infty \to \mathfrak{C}(\mathscr{D}_0)$ is a matrix representation of the first kind and the map $S: Q_\infty \to \mathfrak{C}^*(\mathscr{D}_0)$ is a matrix representation of the second kind.

We wish now to discuss the following question. Given a family $\mathscr{A}$ of closed operators defined together with their adjoints on a dense domain $\mathscr{D}$ [i.e., a subset of $\overline{C}(\mathscr{D},\mathscr{H})$], is it possible to find an orthonormal basis (e_ν) for a matrix representation of the first kind of all operators of $\mathscr{A}$?

Let us first consider the following two families of operators defined by $\mathscr{A}$:

$$\mathscr{A}_m = \{A^{\ddagger\ddagger} = \overline{A\restriction D}\,|A\in\mathscr{A}\},$$
$$\mathscr{A}_M = \{A^{\dagger\dagger} = [A^*\restriction D]^*|A\in\mathscr{A}\}.$$

By the definitions it follows that $\mathscr{A}_m\subseteq\mathfrak{C}(\mathscr{D})$ and $\mathscr{A}_M\subseteq\mathfrak{C}^*(\mathscr{D})$. Clearly, if $\mathscr{A}$ is *-invariant, $\mathscr{A}_M = \mathscr{A}_m^*$. Thus if $\mathscr{A}_m$ admits a matrix representation of the first kind then $\mathscr{A}_M$ admits a matrix representation of the second kind. So we can confine ourselves to consider the question whether $\mathscr{A}_m$ admits a matrix representation of the first kind. Therefore, from now on, we will take directly $\mathscr{A}\subseteq\mathfrak{C}(\mathscr{D})$. In this condition, since each element of $\mathscr{A}$ is $\mathscr{D}$ minimal, for each $A\in\mathscr{A}$ there exists a basis (e_ν) for a matrix representation (of the first kind) of A. But, of course, nothing enables us to say that the basis is the same for all operators of $\mathscr{A}$.

It is clear that if $\mathscr{A}$ admits a matrix representation (of the first kind) with respect to (e_ν) then $\mathscr{A}\subseteq\mathfrak{C}(\mathscr{D}_0)$, where $\mathscr{D}_0$ is the linear hull of the basis vectors. The converse is also true.

The following proposition gives some conditions for the matrix representation of a set $\mathscr{A}\subseteq\mathfrak{C}(\mathscr{D})$.

Proposition 4.3: Let $\mathscr{A}\subseteq\mathfrak{C}(\mathscr{D})$, (e_ν) be an orthonormal basis in $\mathscr{D}$, and $\mathscr{D}_0$ be the linear hull of the e_ν's. Then for the statements (1) (e_ν) is a basis for the matrix representation of the first kind of $\mathscr{A}$; (2) $\mathscr{A}\subseteq\mathfrak{C}(\mathscr{D}_0)$; (3) $\mathscr{D}_0$ is a common core for all elements of $\mathscr{A}$; (4) $\mathscr{D}_0$ is dense in $\mathscr{D}$ with the $\mathscr{A}$-graph topology defined by the seminorms $\varphi\to\|A\varphi\|$ $A\in\mathscr{A}$; and (5) $\mathscr{D}$ is separable for the $\mathscr{A}$-graph topology, we get

$$(1)\Leftrightarrow(2)\Leftrightarrow(3);\quad (4)\Rightarrow(3);\quad (4)\Rightarrow(5).$$

If $\mathscr{A}$ is directed (i.e., $\forall A,B\in\mathscr{A}$ $\exists C\in\mathscr{A}$: $\|A\varphi\|,\|B\varphi\|\leqslant\|C\varphi\|$) then the statements (1)–(4) are equivalent.

Proof: (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) follows easily from the definitions.

(4) $\Rightarrow$ (5). The set of finite linear combinations with rational coefficients of the basis vectors is contained in $\mathscr{D}$ and is dense in it for the $\mathscr{A}$-graph topology.

(4) $\Rightarrow$ (3). Let $\mathscr{D}_0$ be dense in $\mathscr{D}$ for the $\mathscr{A}$-graph topology. Then $\forall\varphi\in\mathscr{D}$ there exists a net $\{\varphi_\alpha\}\subseteq\mathscr{D}_0$ such that

$$\varphi_\alpha\to\varphi \quad\text{and}\quad A\varphi_\alpha\to A\varphi \quad \forall A\in\mathscr{A},$$

this implies that $A = \overline{A\restriction\mathscr{D}_0}$ $\forall A\in\mathscr{A}$ and therefore $\mathscr{D}_0$ is a common core for $\mathscr{A}$.

Now, for directed $\mathscr{A}$, it is proved in Ref. 7 [Theorem 1(4)] that the completion $\hat{\mathscr{D}}_0$ of $\mathscr{D}_0$ for the $\mathscr{A}$-graph topology is

$$\hat{\mathscr{D}}_0 = \bigcap_{A\in\mathscr{A}} D(\overline{A\restriction\mathscr{D}_0}) = \bigcap_{A\in\mathscr{A}} D(A) \supseteq \mathscr{D},$$

then $\mathscr{D}_0$ is dense in $\mathscr{D}$ for the $\mathscr{A}$-graph topology.

Proposition 4.4: Let $\mathscr{A}\subseteq\mathfrak{C}(\mathscr{D})$ and assume that the $\mathscr{A}$-graph topology on $\mathscr{D}$ is separable. Then there exists in $\mathscr{D}$ a basis for a matrix representation of the first kind of $\mathscr{A}$.

Proof: By the assumption, there is in $\mathscr{D}$ a sequence $\{f_n\}$

dense in it for the $\mathscr{A}$-graph topology. Let (e_ν) be the basis obtained from $\{f_n\}$ by orthonormalization and $\mathscr{D}_0$ the linear hull of (e_ν). Here $\mathscr{D}_0$ is clearly dense in $\mathscr{D}$ for the $\mathscr{A}$-graph topology. The statement follows from Proposition 4.3.

ACKNOWLEDGMENT

This work was supported by Fondi (MPI), Università di Palermo.

[1](a) J.-P. Antoine and W. Karwowski, Publ. RIMS, Kyoto University **21**, 205 (1985); **22**, 507 (E) (1986); (b) J.-P. Antoine and F. Mathot, Ann. Inst. H. Poincaré **46**, 299 (1987).
[2]H. J. Borchers, in RCP 25 (Strasbourg) **22**, 26 (1975).
[3]G. Lassner, Physica A **124**, 471 (1984).
[4]G. Epifanio, J. Math. Phys. **17**, 1688 (1976).
[5]J. von Neumann, Math. Ann. **102/1**, 49 (1929); J. Math. **161**, 208 (1929).
[6]G. Epifanio and C. Trapani, J. Math. Phys. **20**, 148 (1979).
[7]S. Gudder and W. Scruggs, Pac. J. Math. **70**, 369 (1977).

Noncanonical groups of transformations, anomalies, and cohomology

J. F. Cariñena
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain
Luis A. Ibort[a)]
Department of Mathematics, University of California, Berkeley, California 94702

(Received 21 July 1987; accepted for publication 7 October 1987)

Two cohomology classes associated to groups of transformations (symplectic or not) of Hamiltonian and Lagrangian systems are studied. A geometrical interpretation of the family of cocycles arising from a class of nonsymplectic actions is given in terms of the Poisson structure of the phase space of the system. These ideas are used to study nongauge (i.e., anomalous) groups of transformations of (locally or globally defined) Lagrangian systems. In particular, well-known results about the magnetic monopole system are described in this context and some hints relating Yang–Mills anomalies with nonsymplectic groups of transformations are given.

I. INTRODUCTION

Physics has been increasingly concerned about anomalies, i.e., classical symmetries broken at the quantum level—for example, the loss of gauge invariance in a Yang–Mills theory with massless Weyl fermions is called the Yang–Mills or non-Abelian anomaly. Different families of anomalies have been computed using powerful techniques from global analysis, and some of its implications, showing, for example, the inconsistency of the canonical quantization procedure for the Yang–Mills anomaly, have been pointed out.[1] In a different way Dirac pointed out that consistency in the quantum representation of the translation group for the magnetic monopole system implies the quantization condition for the electric charge[2] (see also, for example, Ref. 3 and references therein).

The main goal of this paper is to provide a common background for both phenomena in terms of a general cohomological structure associated to noncanonical action of Lie groups on (pre)symplectic manifolds. We will study the classical structure of groups of transformations in Hamiltonian and Lagrangian systems and we will show that there exists a natural way (similar to the descent equations of Faddeev but inspired in a different choice of double complex) of constructing a family of cocycles with values in forms on the (pre)symplectic manifold (phase space of the system). This family has a geometrical interpretation in terms of the Poisson bracket of the theory. The physical meaning of the discussion is displayed step by step through the detailed description of groups of transformations for classical finite-dimensional Lagrangian systems mimicking the notion of anomalous systems.

This paper is organized as follows: Section II will be devoted to the statement of usual properties of gauge groups of transformations of Lagrangian systems and the introduction of some obvious generalizations of such concepts. We will describe the descent method for a noncanonical action of a Lie group in Sec. III and in Sec. IV the relation of the family of cocycles obtained with symplectic geometry is studied. In particular, we will provide a Poisson bracket interpretation of them. Some applications, remarks, and implications of the ideas described in this paper will be considered in Sec. V and, in particular, the application of this approach to anomalies in quantum field theory is sketched.

II. GROUPS OF TRANSFORMATIONS AND LAGRANGIAN SYSTEMS

Groups of gauge transformations of Lagrangian systems have been studied for a long time and their cohomological implications analyzed in different contexts.[4–6] As has been pointed out in the Introduction, nongauge groups of transformations are relevant to understanding the geometric structure of anomalous systems. We will assume in the following discussion that the configuration space of the system is a C^∞-differentiable (finite- or infinite-dimensional) manifold Q, and L is a Lagrangian function defined on TQ, the tangent bundle of Q. Let G be a group of fiber preserving transformations of TQ. The action of G may either preserve the canonical tensor field S of TQ or not.[7] In the former case the action of G on TQ corresponds to a lifted action of a group H on Q times (semidirect product) a group of translations along the fibers of TQ. We will call these transformations natural and we will be restricted to this simpler case in what remains of this section.

Let us recall some well-known results about gauge transformations before introducing some generalizations. A group G is said to be a group of gauge transformations of (TQ,L) if $g^*L = L + \hat{\alpha}_g$, where α_g is a closed one-form on Q, and $\hat{\alpha}_g$ the associated function on TQ. The set of one-forms α_g satisfy the one-cocycle condition

$$\alpha_{g_1 g_2} = g_2^* \alpha_{g_1} + \alpha_{g_2}. \tag{2.1}$$

If the group G preserves the vertical endomorphism S, $[g_*, S] = 0\ \forall g \in G$, then

$$g^*\theta_L = g^*(dL \circ S) = dL \circ S \circ g_*$$
$$= d(L \circ g) \circ S = d(g^*L) \circ S = \theta_{g^*L}.$$

Consequently,

$$\theta_{g^*L} = \theta_L + \pi^* \alpha_g, \tag{2.2}$$

[a)] On leave of absence from the Departamento de Física Teórica, Universidad de Zaragoza, 50009, Spain.

where we have just used that $\theta_{\hat{\alpha}} = \pi^*\alpha \ \forall\alpha\in B^1(Q)$.[4] The infinitesimal version of (2.1) is

$$\alpha_{[a,b]} = d\langle X_a,\alpha_b\rangle - d\langle X_b,\alpha_a\rangle \quad \forall a,b\in \mathfrak{g}, \tag{2.3}$$

where $\alpha_a = (d/dt)\alpha_{\exp ta}|_{t=0}$, and X_a denotes the vector field associated to the infinitesimal generator a. Then if G acts by natural transformations we have that the Poincaré–Cartan one-form θ_L transforms as shown in (2.2) and correspondingly the Lagrange two-form ω_L is invariant because of the closedness of α_g.

If L is a regular Lagrangian we find that G acts by symplectomorphisms of ω_L, and if $H^1(Q) = 0$, there exists a global Hamiltonian function associated to each infinitesimal generator a of G defined by the formula[4]

$$f_a = \theta_L(X_a) - h_a,$$

with $dh_a = \alpha_a \ \forall a\in\mathfrak{g}$. The main implication of the nontriviality of the cocycle α_a is the appearance of a nontrival two-cocycle c, with coefficients in $\mathbf{R}$ in the commutation relation of the Hamiltonian functions f_a. More explicitly, we get

$$\{f_a,f_b\} = f_{[a,b]} + c(a,b) \quad \forall a,b\in\mathfrak{g}.$$

A natural generalization of the above structure appears removing the closed character of the cocycle α_a. The first change is that the group of transformations is not any longer a group of symplectomorphisms of ω_L. Let us make more precise these assertions.

We will call a group of transformations G quasigauge if $g^*L = L + \hat{\alpha}_g$, where α_g is a family of (non-necessarily closed) one-forms on Q. If G acts by natural transformations on TQ, we have $g^*\omega_L = \omega_L + \pi^* d\alpha_a$ or, infinitesimally,

$$L_{X_a}\omega_L = \pi^* d\alpha_a \quad \forall a\in\mathfrak{g}. \tag{2.4}$$

There is an important remark related with the noncanonical character of the quasigauge transformations. There is no local Hamiltonian function associated to the infinitesimal generators of the groups G, because $i_{X_a}\omega_L$ is not closed and can be written, in a nonunique way, as a sum

$$i_{X_a}\omega_L = \beta_a + \alpha_a, \tag{2.5}$$

where β_a is a closed one-form. Locally there will exist a function f_a such that $\beta_a = df_a$ and the previous equation shows that X_a has a canonical part $(\hat{\omega}_L^{-1}\beta_a)$ and a noncanonical one $(\hat{\omega}_L^{-1}\alpha_a)$. We will analyze this structure in detail in Sec. IV.

Example: Perhaps the easiest example for this corresponds to a system with a broken symmetry. Let L_0 be a Lagrangian in $T\mathbf{R}^n$, that is, invariant under translations along k directions on $\mathbf{R}^n$, and let $L_A = L_0 + \hat{A}$, where $A = A_i\, dq^i$ is a generic one-form on $\mathbf{R}^n$ (in physical terms that corresponds to couple the original system with a magnetic potential A). In general L_A is not any longer $\mathbf{R}^k$ invariant, and for any infinitesimal generator $a\in\mathbf{R}^k$, we have $\hat{\alpha}_a = X_a L = a^i(\partial Aj/\partial q^i)\dot{q}^j$. This gives for α_a the expressions $\alpha_a = a^i(\partial A_j/\partial q^i)dq^j$. Notice that $\omega_{L_A} = \omega_{L_0} + \pi^* dA$, then

$$L_{X_a}\omega_{L_A} = dL_{X_a}\pi^*A = \pi^* d\alpha_a = a^i(A_{j,ik} - A_{k,ij})dq^k\wedge dq^j = a^i\frac{\partial F}{\partial q^i},$$

where $F = dA$ is the magnetic field associated to A. The nontrivial cocycle α_a reflects the transformation properties of L_A and by previous commentaries we could say that the Lagrangian L_A is "anomalous" with respect to the translation group $\mathbf{R}^k$.

III. NONCANONICAL COCYCLES

In this section we are going to establish the cohomological foundations of some of the ideas encountered in the commentaries of the previous section. Afterwards we will discuss the interpretation of some of these objects in terms of symplectic geometry.

From the discussion in the previous section it is clear that the relevant geometric structure to be studied is an action (noncanonical in general) of a Lie group G on a manifold M equipped with a closed two-form Ω (possibly degenerate). As we noticed before, if the action leaves Ω invariant we are in the well-known case of (pre)symplectic actions. By the contrary, if G does not leave Ω invariant we get a family of closed two-forms ω_a defined as follows:

$$L_{X_a}\Omega = \omega_a \quad \forall a\in\mathfrak{g}. \tag{3.1}$$

The family of two-forms ω_a, obviously satisfies the one-cocycle relation

$$L_{X_a}\omega_b - L_{X_b}\omega_a - \omega_{[a,b]} = 0 \tag{3.2}$$

or, equivalently,

$$d(\langle X_a,\omega_b\rangle - \langle X_b,\omega_a\rangle) - \omega_{[a,b]} = 0 \quad \forall a,b\in\mathfrak{g}. \tag{3.3}$$

It happens that the existence of this one-cocycle has physically relevant consequences on the transformation properties of (G-invariant) systems described on (M,Ω). Some of these properties are related with a family of cohomological objects associated to ω_a.

The best way to describe the origin and structure of these objects is dealing with the cohomologies involved here, i.e., the cohomology on the group G and the cohomology on the manifold M.

Let us consider an open neighborhood U on M such that there exists a family of one-forms α_a satisfying $d\alpha_a = \omega_a$ (for example, U contractible). Consider the double complex $\oplus_{p,q\geq 0}\Omega^{p,q}(\mathfrak{g},U)$ of left-invariant p-forms on G with values in forms on U, i.e., $\Omega^{p,q}(\mathfrak{g},U) = \Lambda^p(\mathfrak{g}^*)\otimes\Omega^q(U)$, where $\Lambda^p(\mathfrak{g}^*)$ represents linear p-forms on $\mathfrak{g}$ and $\Omega^q(U)$ differential q-forms on U. This double complex could be represented as a grid with the (p,q) entry given by the (p,q) factor $\Omega^{p,q}(\mathfrak{g},U)$. The exterior differential d maps $\Omega^{p,q}(\mathfrak{g},U)$ into $\Omega^{p,q+1}(\mathfrak{g},U)$ and the exterior differential on the group ∂ maps $\Omega^{p,q}$ into $\Omega^{p+1,q}$. Because of the action of G on M there is another cohomology operator, denoted by δ, defined as follows:

$$\delta\alpha(a_1,\ldots,a_{k+1}) = \sum_{i=1}^{k+1}(-1)^{i+1}L_{X_{a_i}}\alpha(a_1,\ldots,\hat{a}^i,\ldots,a_{k+1}) + \sum_{i<j}(-1)^{i+j}\alpha([a_i,a_j],a_1,\ldots,\hat{a}_i,\ldots,\hat{a}_j,\ldots,a_{k+1}),$$

where α is an element in $\Omega^{k,q}$ and $a_1,\ldots,a_{k+1}$ is a family of elements in g. The Lie derivative commutes with the exterior differential and then we have that δ commutes with d as well as ∂, i.e.,

$$\delta d = d\delta, \quad \partial d = d\partial.$$

The one-cocycle ω_a defined previously is an element in $\Omega^{1,2}(\mathbf{g},U)$, and satisfies $\delta\omega_a = 0 = d\omega_a$. From $d\omega_a = 0$ we get that there exists a family α_a in $\Omega^{1,1}(\mathbf{g},U)$ such that $d\alpha_a = \omega_a$. The element λ in $\Omega^{2,1}(\mathbf{g},U)$ defined by $\lambda = \delta\alpha$ is a two-cocycle with respect to δ—namely $\delta\lambda = \delta^2\alpha = 0$, and is again a cocycle with respect to d because $d\lambda = d\delta\alpha = \delta\, d\alpha = \delta\omega = 0$. Notice that the explicit formula for λ is given by

$$\begin{aligned}\lambda(a,b) &= L_{X_a}\alpha_b - L_{X_b}\alpha_a - \alpha_{[a,b]} \\ &= i_{X_a}\omega_b - i_{X_b}\omega_a + d\langle X_a,\alpha_b\rangle \\ &\quad - \langle X_b,\alpha_a\rangle - \alpha_{[a,b]} \quad \forall a,b \in \mathbf{g}.\end{aligned} \tag{3.4}$$

Again $\lambda(a,b)$ is a family of closed one-forms on U, then there exists a family of functions $h(a,b)$ such that $dh(a,b) = \lambda(a,b)$. The family $h(a,b)$ defines a cochain element in $\Omega^{2,0}(\mathbf{g},U)$. The image under δ of h provides a new element τ in $\Omega^{3,0}(\mathbf{g}U)$ satisfying $\delta\tau = 0$, i.e., τ is three-cocycle in g. In addition we have that τ has values in locally constant functions because of the relation $d\tau = d\delta h = \delta\, dh = \delta\lambda = 0$. Then, as U is connected $\tau(a,b,c)$ is a constant function for any elements a,b,c in g and this means that τ is a three-cocycle in g with values in **R**. This construction is a particular case of the tic-tac-toe lemma for double complexes[8] (see also, Ref. 9 for a different example of the use of the tic-tac-toe lemma).

The preceding discussion has a local character in M. Nevertheless, the conclusion is global; that means that if we change the neighborhood U we get cocyles τ' and λ' in the same cohomology classes as τ and λ respectively. Let U,V be two nondisjoint neighborhoods of M and α_a^U,α_a^V two families of one-forms satisfying $d\alpha_a^U = \omega_{a|U}$ and $d\alpha_a^V = \omega_{a|V}$. On the intersection $U \cap V$, there exists a family of closed one-forms β_a such that $\alpha_a^U = \alpha_a^V + \beta_a$. The two-cocyles λ^U and λ^V obtained, respectively, from α^U and α^V differ by $\delta\beta_a$. But there exists a family of functions g_a such that $\beta_a = dg_a$; then $\lambda^U = \lambda^V + \delta\, dg_a = \lambda^V + d(\delta g_a)$. But $\lambda^U = dh^U$ and $\lambda^V = dh^V$, therefore $h^U = h^V + \delta g + c$, where $c(a,b)$ is a constant on M depending only on a,b, and finally $\tau^U = \delta h^U = \delta(h^V + \delta g + c) = \delta h^V + \delta c = \tau^V + \partial c$. The conclusion is that the cohomology class of the three-cocycle τ^U is an invariant of the action of G on M.

This discussion can be summarized in the theorem below.

Theorem 1: Any action of Lie group G on a (pre)symplectic manifold (M,Ω) has associated a cohomology class $[\tau] \in H^3(\mathbf{g},\mathbf{R})$ and a cohomology class $[\lambda] \in H^2(\mathbf{g},Z^1(M))$. Furthermore, if the action is symplectic the classes $[\tau]$ and $[\lambda]$ vanish, and h becomes a two-cocycle in the δ cohomology. Here λ is given by formula (3.4) and τ is defined by

$$\begin{aligned}\tau(a,b,c) &= \underset{(a,b,c)}{\mathbf{S}} (\langle X_a,\lambda(b,c)\rangle - h([a,b],c)) \\ &= \underset{(a,b,c)}{\mathbf{S}} (X_a h(b,c) - h([a,b],c)) \quad \forall a,b,c \in \mathbf{g},\end{aligned} \tag{3.5}$$

with $dh(a,b) = \lambda(a,b)$ and the symbol $\underset{(a,b,c)}{\mathbf{S}}$ denotes cyclic sum over the indices (a,b,c).

Proof: The construction of $[\tau]$ and $[\lambda]$ was discussed previously as well as its independence of the covering chosen for defining them. Notice that if the action is symplectic that means that $\omega_a = 0$ and then $d\alpha_a = 0$. Thus there exists a family of functions k_a such that $dk_a = \alpha_a$. Because of the commutativity of δ and d and the vanishing of λ we get $h = \delta k$ and $dh = 0$. In consequence $[h]$ is a well-defined class in $H^2(\mathbf{g},\mathbf{R})$. It is important to remark that the cohomology operator δ collapses to ∂ when it is restricted to multilinear forms on g with coefficients in locally constant functions and this is the reason τ is a three-cocycle in g with coefficients in **R**, i.e., with respect to ∂.

As an immediate consequence of the previous theorem we get the following corollary.

Corollary 1: Any action of a Lie group G on an exact (pre)symplectic manifold (M,Ω) has vanishing $[\lambda]$ and $[\tau]$ cohomology classes.

Proof: It is obvious because we can define in any neighborhood a family of one-forms α_a such that $d\alpha_a = \omega_a$ and $\delta\alpha = 0$ by $L_{X_a}\theta = \alpha_a$. It follows that h is going to be a locally constant cochain, hence τ is trivial in the ∂ cohomology, $\tau = \partial h$.

From this remark follows the important conclusion.

Corollary 2: Any Lagrangian system (TQ,L) with a quasigauge group of transformations G has trivial cohomology classes $[\lambda]$ and $[\tau]$.

Proof: The reason is that any Lagrangian system has an exact (pre)symplectic structure $\omega_L = d\theta_L$, where θ_L is the pullback of the canonical Liouville one-form on T^*Q by the Legendre transformation.

Notice that this remark does not contradict the possible existence of nontrivial three-cocycles for Lagrangian systems not globally defined. In such a case the symplectic structure in the canonical formalism is not exact. This is what happens with a particle moving in a monopole magnetic field F (Ref. 10) because the two-form giving the field strength F is not exact. We will proceed along with this discussion in the last section.

IV. POISSON BRACKETS AND COHOMOLOGY

In the previous section we have shown how a noncanonical action of a Lie group in a (pre)symplectic manifold causes a family of cocycles of g with coefficients in the ring of differential forms on M. In this section we will provide a symplectic interpretation of the second cohomology class $[\lambda]$ described before, formula (3.3), in terms of the Poisson structure induced in the ring of functions on M by the symplectic structure Ω (if Ω were presymplectic we should use the ring of first class functions, i.e., those invariant along char Ω) and a new formula for computing τ in some particular cases.

From the definition of the one-cocycle ω_a, formula (3.1), we get that locally we can find families of one-forms α_a and β_a such that

$$i_{X_a}\Omega = \beta_a + \alpha_a \quad \forall a \in \mathbf{g}, \tag{4.1}$$

where β_a is a family of closed one-forms. We will call the locally Hamiltonian vector field H_a associated with β_a, i.e., satisfying the equation $i_{X_a}\Omega = \beta_a$, the Hamiltonian part of X_a. The vector field A_a defined by $i_{A_a}\Omega = \alpha_a$ is called the noncanonical part of X_a with respect to the decomposition (4.1). Clearly, $X_a = H_a + A_a \quad \forall a \in \mathfrak{g}$ and, as it is obvious from the definitions, both H_a and A_a are not uniquely defined. Because of this there is no real reason to talk about the Hamiltonian of X_a. In spite of this one has reason to ask about what happens with the "representation" of g obtained for each decomposition (4.1) using only the canonical or Hamiltonian part H_a. The canonical part of X_a is only part of the fundamental vector field representation of g in the Lie algebra of vector fields of M. Because of that, it is expected that the Hamiltonian vector fields associated with the infinitesimal generators of G do not provide a representation of g. The cocycle λ is defined by the commutation relations of the Hamiltonians associated with the infinitesimal generators of g only for special decompositions of the fundamental representation of g. Thus the link between the cocycle λ and the Poisson bracket commutator is partial.

We will say that a decomposition $X_a = A_a + H_a$ of the fundamental vector fields of a noncanonical action of G on M is Abelian if $\Omega(A_a, A_b) = 0 \quad \forall a,b \in \mathfrak{g}$, or in other words, if the distribution generated by the noncanonical part of the fundamental vector fields X_a is isotropic. Two important cases in which we have Abelian decompositions are given by the following.

Examples: (1) Let G be a group of quasigauge transformations of a Lagrangian system with Lagrangian L. Defining the decomposition of X_a as given by formula (2.4) it is obvious that A_a is a vertical vector field because α_a is the pullback of a one-form on the base space Q. The vertical distribution of TQ is Lagrangian and then clearly $\omega_L(A_a, A_b) = 0 \quad \forall a,b \in \mathfrak{g}$. Because of this any quasigauge group of transformations admits an Abelian decomposition.

(2) Let G be a group acting by diffeomorphism on a manifold Q. There is a natural action of TG on TQ. Let L be a Lagrangian function invariant under the complete lifting of G. The Lie algebra of TG is $\mathfrak{g}^c \oplus \mathfrak{g}^v$. The complete lifting part $\mathfrak{g}^c$ is Hamiltonian, and the vertical part $\mathfrak{g}^v$ is noncanonical. This (trivial) decomposition is clearly Abelian.

The main theorem relating Poisson brackets and cocycles is the following.

Theorem 2: Let G be a group of noncanonical transformations of the symplectic manifold (M,Ω) such that there exists an Abelian decomposition $X_a = H_a + A_a \ \forall a \in \mathfrak{g}$ of its fundamental realization. If the Hamiltonian associated to X_a with respect this decomposition is f_a, we have

$$d\{f_a, f_b\} = df_{[a,b]} + \lambda(a,b) \quad \forall a \in \mathfrak{g}, \tag{4.2}$$

where λ is the two-cocycle associated to the action of G as given in formula (3.4) and

$$\tau(a,b,c) = \underset{(a,b,c)}{\mathrm{S}} \langle A_a, \lambda_{[b,c]} \rangle \quad \forall a,b,c \in \mathfrak{g}, \tag{4.3}$$

where τ is the three-cocycle defined in formula (3.5).

Proof: Computing $d\{f_a, f_b\}$ we get

$$\begin{aligned} d\{f_a,f_b\} &= d\Omega(H_a,H_b) = d\Omega(X_a - A_a, X_b - A_b) \\ &= d\Omega(X_a,X_b) + d\Omega(A_a,A_b) - d\Omega(X_a,A_b) - d\Omega(A_a,X_b) \\ &= d(i_{X_b}(i_{X_a}\Omega)) + d\langle X_a,\alpha_b\rangle - d\langle X_b,\alpha_a\rangle \\ &= L_{X_b}(i_{X_a}\Omega) - i_{X_b^d}(i_{X_a}\Omega) + d(\langle X_a,\alpha_b\rangle - \langle X_b,\alpha_a\rangle) \\ &= i_{[X_a,X_b]}\Omega + i_{X_a}L_{X_b}\Omega - i_{X_b}\,d\alpha_a + d(\langle X_a,\alpha_b\rangle - \langle X_b,\alpha_a\rangle) \\ &= i_{X_{[a,b]}}\Omega + i_{X_a}\,d\alpha_b - i_{X_b}\,d\alpha_a + d(\langle X_a,\alpha_b\rangle - \langle X_b,\alpha_a\rangle) \\ &= df_{[a,b]} - \alpha_{[a,b]} + i_{X_a}\,d\alpha_b - i_{X_b}\,d\alpha_a + d(\langle X_a,\alpha_b\rangle - \langle X_b,\alpha_a\rangle) \\ &= df_{[a,b]} + \lambda(a,b). \end{aligned}$$

Noticing the $dh(a,b) = \lambda(a,b)$ we get that the formula (4.2) from before can be written as

$$\{f_a, f_b\} = f_{[a,b]} + h(a,b) + c(a,b),$$

where $c(a,b)$ is just a constant depending only on a,b. Then a straightforward computation gives us

$$\begin{aligned} \underset{(a,b,c)}{\mathrm{S}} h([a,b],c) &= \underset{(a,b,c)}{\mathrm{S}} \{f_{[a,b]}, f_c\} - \underset{(a,b,c)}{\mathrm{S}} f_{[[a,b],c]} \\ &= \underset{(a,b,c)}{\mathrm{S}} \langle H_a, \lambda(b,c)\rangle, \end{aligned}$$

and from formula (3.5) we get

$$\begin{aligned} \tau(a,b,c) &= \underset{(a,b,c)}{\mathrm{S}} \langle X_a, \lambda(a,b)\rangle - \underset{(a,b,c)}{\mathrm{S}} h([a,b],c) \\ &= \underset{(a,b,c)}{\mathrm{S}} \langle X_a, \lambda(b,c)\rangle - \underset{(a,b,c)}{\mathrm{S}} \langle H_a, \lambda(b,c)\rangle \\ &= \underset{(a,b,c)}{\mathrm{S}} \langle A_a, \lambda(b,c)\rangle. \end{aligned}$$

Notice that in the trivial case, i.e., when we are dealing with canonical actions, $\alpha_a = 0$ and $\lambda = 0$, then $h(a,b)$ is a two-cocycle in the ∂ cohomology, that can be assimilated to the constant $c(a,b)$ reproducing the classical results.[11] It is also important to notice that the decomposition used to obtain the relation (4.2) is not unique. Not all possible decompositions are Abelian, so formula (4.2) is only true for Hamiltonians corresponding to Abelian decompositions. Formula (4.3) deserves some commentaries too. First it is convenient to remark that this formula is true only for Abelian decompositions and it shows once again that for symplectic actions τ is equal to zero because we can choose $A_a = 0\ \forall a \in \mathfrak{g}$ and then trivially we get $\tau = 0$.

Example: Continuing the example started in Sec. II, we get

$$i_{X_a}\omega_{L_A} = i_{X_a}\omega_{L_0} + i_{X_a}\, dA = df_a + \alpha_a.$$

Assuming, for example, that the Lagrangian L_0 is the kinetic energy corresponding to a metric $\langle\ ,\ \rangle$ we get that

$$f_a(q,\dot q) = \langle a,\dot q\rangle - \langle a,A(q)\rangle,$$

and

$$H_a = a_i\frac{\partial}{\partial q^i} - a_i\frac{\partial A_j}{\partial q^i}\frac{\partial}{\partial \dot q^j}.$$

Then

$$A_a = a_i\frac{\partial A_j}{\partial q^i}\frac{\partial}{\partial \dot q^j}$$

is a familiy of vertical fields and obviously they define an Abelian decomposition of X_a. The Poisson bracket of f_a and f_b is easily computed and it gives

$$d\{f_a,f_b\} = \omega_{L_A}(H_a,H_b) = d(F(X_a,X_b)).$$

This result is in complete agreement with formula (4.2) because $[a,b] = 0\ \forall a,b \in \mathfrak{g}$ and

$$\begin{aligned}\lambda(a,b) &= i_{X_a}\omega_b - i_{X_b}\omega_a - \alpha_{[a,b]}\\ &= (a^k b^i - a^i b^k)(\partial^2_{ik}A_j - \partial^2_{ij}A_k)dq^j\\ &= d(F(X_a,X_b)).\end{aligned}$$

Finally the three-cocycle τ vanishes because the Lagrangian L_A is globally defined. As we will show later, even in the monopole system the three-cocycle τ is still zero.

V. SOME APPLICATIONS, REMARKS, AND COMMENTARIES

During the general discussion we have been studying the example of the translation group acting in the system of a particle moving in a magnetic field $F = dA$. This example lead us to trivial results in the sense that both cohomology classes, $[\lambda]$ and $[\tau]$, were trivial. We can modify slightly this example considering a charged particle moving in the field of a magnetic monopole. A magnetic monopole of strenth g at the origin in $\mathbf{R}^3$ creates a magnetic field given by a two-form F on $\mathbf{R}^3 - 0$ (the field is singular at the origin) that is closed but not exact. Its integral over S^2, the unit sphere centered at the origin, is $4\pi g$. There is no globally defined vector potential for F, hence there is no globally defined Lagrangian for the system.[10] The two-form F is completely determined by its restriction to S^2, where it coincides with $g\sigma$, σ being the area form on S^2. However, there are local one-forms, A_S defined on $S^2 - \{\text{north pole}\}$ and A_N defined on $S^2 - \{\text{south pole}\}$, differing by an exact form on their common domain, whose exterior derivatives are both F. Choosing these forms we can reproduce the computations we did for the globally defined potential example and we get again that the two-cocycle λ is given by the formula $\lambda(a,b) = d(F(X_a,X_b))$. Here τ is still zero because $\tau(a,b,c) = dF(X_a,X_b,X_c)$, and as far as F is closed the three-cocycle τ will be zero. For the monopole system the two-cocycle λ is not trivial but τ still is. Thus the Dirac's quantization condition for the electric charge does not appear from the existence of a three-cocycle on the system but from integrability conditions of the system in the geometric quantization scheme, namely integer class of the charge symplectic form of the system, as was pointed out in Ref. 12, which is equivalent to the requirement that the path integral quantization of the system be well defined.[9]

Finally, a brief discussion about Yang–Mills anomalies and the ideas described previously is in order (a forthcoming paper is devoted to a thorough discussion on the subject). Let Q be the space of Yang–Mills potentials, irreducible connections on the principal fiber bundle $P(G,X)$ with X a $2n$-dimensional space-time. Here L_{eff} is the effective Lagrangian of the theory obtained from the generating functional $Z[A] = \int d\mu[A]\exp(-S_{\text{YM}}[A])$ by the formula $\int L_{\text{eff}}\, dt = W[A] = -\ln Z[A]$, where S_{YM} is the classical Yang–Mills action with Lagrangian density $L(A,\psi,\bar\psi) = (i/2)\bar\psi D_A\psi$, $\psi,\bar\psi$ are Weyl massless fermions, and D_A is the covariant Dirac operator associated to the connection A. The group of gauge transformations of the theory will be denoted by $\mathbf{G}$ and it is well known that the effective action $W[A]$ of the theory is not gauge invariant. The anomaly of the theory is defined as the infinitesimal variation of the effective action under the group of gauge transformations and can be easily seen to be a one-cocycle on $\mathbf{G}$ (Wess–Zumino consistency condition). From the previous discussion we know that there exists a two-cocycle λ that modifies the commutation relations of the Hamiltonians associated to the infinitesimal generators of the group of gauge transformations. In this sense the cochain h such that $dh = \lambda$ is precisely the Schwinger term computed in Ref. 1.

ACKNOWLEDGMENTS

We would like to acknowledge the referee for the proper references on the anomalous current commutators. One of us (L.A.I.) would like to thank O. Alvarez for very stimulating discussions in San Lorenzo del Escorial.

[1] D. Gross and R. Jackiw, Phys. Rev. D **6**, 447 (1972); L. D. Faddeev, Phys. Lett. B **145**, 81 (1984).

[2] P. A. M. Dirac, Proc. R. Soc. London Ser. A **133**, 60 (1931).

[3] R. Jackiw, in *Current Algebras and Anomalies*, edited by S. B. Treiman, R. Jackiw, B. Zumino, and E. Witten (Princeton U. P., Princeton, NJ, 1985), Vol. 211.

[4] J. F. Cariñena and L. A. Ibort, Nuovo Cimento B **87**, 41 (1985).

[5] J. Houard, J. Math. Phys. **18**, 502 (1977).

[6] J. M. Lévy-Leblond, Commun. Math. Phys. **12**, 64 (1969).

[7] J. F. Cariñena, M. Crampin, and L. A. Ibort, "Groups of transformations and gauge symmetries," preprint, 1987.

[8] R. Bott and L. W. Tu, *Differential Forms in Algebraic Topology* (Springer, Berlin, 1980).

[9] O. Alvarez, Commun. Math. Phys. **100**, 279 (1985).

[10] A. P. Balachandran, G. Marmo, B. S. Skagerstan, and A. Stern, "Gauge symmetries and fiber bundles," *Lecture Notes in Physics*, Vol. 188 (Springer, Berlin, 1983).

[11] N. Woodhouse, *Geometric Quantization* (Oxford U. P., New York, 1980).

[12] M. Crampin, J. Phys. A: Math. Gen. **14**, 3407 (1981).

Irreducible *-representations of the Lie superalgebras $B(0,n)$ with finite-degenerated vacuum. II

J. Blank and M. Havlíček
Nuclear Center of the Charles University Prague, V Holešovičkách 2, 180 00 Praha 8, Czechoslovakia

(Received 4 February 1987; accepted for publication 23 September 1987)

The method presented in the first part of this work is applied to the superalgebra $B(0,2)$. Two families of irreducible *-representations of this superalgebra and its real form osp(1,4) are constructed explicitly in terms of differential operators on the Hilbert space $L^2(\widetilde{M})\otimes\mathbb{C}^N$ of N-component vector functions Ψ: $\widetilde{M}\to\mathbb{C}^N$: (i) the family $\{\pi_J: J=0,1,...\}$ of massless representations with $N=2$, $\widetilde{M}=\mathbb{R}^+\times(-\pi,\pi)\times\mathbb{R}^+$, the dimension of the vacuum subspace of π_J being $J+1$; (ii) the family $\{\pi_0^{(\vartheta)}: \vartheta>0\}$ of massive representations such that $\pi_0^{(\vartheta)}\restriction$ so(3,2) equals the direct sum of three irreducible representations of so(3,2). This family is characterized by $N=4$, $\widetilde{M}=\mathbb{R}^+\times(0,\pi)\times\mathbb{R}^+$ and nondegenerated vacuum. It is also shown that all the remaining massive representations form a system of families $\{\pi_J^{(\vartheta)}: \vartheta>J/2\}$, $J=1,2,...$, with $N=4(J+1)$, $(J+1)$-fold degenerated vacuum and common $\widetilde{M}=\mathbb{R}^+\times(0,\pi)\times\mathbb{R}^+$.

I. INTRODUCTION

In the first part of this study[1] we have presented a method for constructing *-representations of complex Lie superalgebras $B(0,n)$ $(n=1,2,...)$ whose real forms are osp$(1,2n)$. In the present paper the method is applied for obtaining a family of irreducible Hilbert space *-representations of $B(0,2)$ and osp(1,4).

We shall start by recalling basic features of the method and specifying it for the case $n=2$. A basis of $B(0,2)$ is used in which the odd and even generators are denoted by a_j and b_{jk}, respectively, with $j,k=\pm1,\pm2$ and $b_{jk}=b_{kj}$. The relevant commutation and anticommutation relations read

$$[b_{jk},a_l]=g_{jl}a_k+g_{kl}a_j,\quad \{a_j,a_k\}=2b_{jk},\tag{1.1}$$

where $g_{jk}:=\mathrm{sgn}(j)\delta_{j+k}$. The (unique) involution on $B(0,2)$ can be defined by $a_j^*:=a_{-j}$, which implies $b_{jk}^*=b_{-j-k}$.

The construction is based on the following assumption: an infinite-dimensional linear representation Ω of $B(0,2)$ is given and there exists involution $T\mapsto T^\#$ on a subspace $\Lambda(V)$ of all linear operators on the representation space V such that for all $z\in B(0,2)$, one has $\Omega(z)\in\Lambda(V)$ and

$$\Omega(z^*)=\Omega(z)^\#.\tag{1.2}$$

The problem consists in finding an Ω-invariant subspace $\mathscr{D}\subset V$ and defining scalar product $(\cdot,\cdot)$ on $\mathscr{D}$ such that $z\mapsto\pi(z):=\Omega(z)\restriction\mathscr{D}$ becomes an algebraically irreducible representation of $B(0,2)$ on $\mathscr{D}$ satisfying for all $\Phi,\Psi\in\mathscr{D}$ the following *-condition:

$$(\Phi,\pi(z)\Psi)=(\pi(z^*)\Phi,\Psi).\tag{1.3}$$

In addition, we require that the vacuum subspace

$$\mathscr{D}_{\rm vac}:=\mathscr{D}\cap\{\Phi\in V: \Omega(a_r)\Phi=0,\ r=1,2\}\tag{1.4a}$$

be finite dimensional,

$$1\leqslant\dim\mathscr{D}_{\rm vac}<\infty.\tag{1.4b}$$

Remark 1.1: The real linear hull of even generators b_{jk} equals sp$(4,\mathbb{R})\sim$so(3,2); this Lie algebra can also be expressed as the real linear hull of $x_{jk}:=\frac{1}{2}(b_{jk}-b_{-j-k})+(i/2)(b_{-jk}+b_{j-k})$, $j\geqslant k$, for which one has $x_{jk}^*=-x_{jk}$. Then (1.3) implies that the representation $x\mapsto\pi(x)$ of sp$(4,\mathbb{R})$ is skew symmetric, and as sp$(4,\mathbb{R})$ is noncompact, $\mathscr{D}$ must be infinite dimensional, at least one of operators $\pi(x)$ being unbounded.[2]

Suppose that for a given linear representation Ω on V there exists a subspace $\mathscr{D}\subset V$ such that $\pi\equiv\Omega\restriction\mathscr{D}$ has all the required properties. We have seen that a necessary condition for it is $\dim\mathscr{D}=\infty$. Some further necessary conditions are implied by the following properties of operators

$$\widetilde{E}:=\pi(b_{2-1}),\quad \widetilde{F}:=\pi(b_{1-2}),\qquad \widetilde{H}:=\pi(b_{2-2})-\pi(b_{1-1}),\tag{1.5a}$$

and

$$\widetilde{N}:=\pi(b_{1-1}+b_{2-2})=\frac{1}{2}\sum_{r=1}^{2}\{\pi(a_r),\pi(a_r^*)\}\tag{1.5b}$$

(see Ref. 1): (a) the vacuum subspace $\mathscr{D}_{\rm vac}$ is invariant under $\widetilde{N}$, $\widetilde{E}$, $\widetilde{F}$, and $\widetilde{H}$; (b) the restrictions of $\widetilde{E}$, $\widetilde{F}$, and $\widetilde{H}$ to $\mathscr{D}_{\rm vac}$ form a finite-dimensional representation of sl$(2,\mathbb{C})$; and (c) $\widetilde{N}$ is positive and commutes with $\widetilde{E}$, $\widetilde{F}$, and $\widetilde{H}$.

Well-known properties of finite-dimensional representations of sl$(2,\mathbb{C})$ now imply existence of a subspace $V_{\rm HW}\subset\mathscr{D}_{\rm vac}$ on which $\widetilde{N}$ is a multiple of identity and the operators $\widetilde{E}$, $\widetilde{F}$, and $\widetilde{H}$ form an irreducible highest-weight representation of sl$(2,\mathbb{C})$. In other words, a non-negative integer J and a vector $\Psi_J\in\mathscr{D}_{\rm vac}$ exist such that $\{\widetilde{E}^k\Psi_J: k=0,1,...,J\}$ is a basis in $V_{\rm HW}$ and the following relations hold:

$$\widetilde{E}^{J+1}\Psi_J=0,\quad \widetilde{F}\Psi_J=0,\quad \widetilde{H}\Psi_J=J\Psi_J,\tag{1.6a}$$

$$\widetilde{N}\Psi_J=\nu\Psi_J\tag{1.6b}$$

for some $\nu\geqslant0$. These equations are completed by $\Psi_J\in\mathscr{D}_{\rm vac}$, i.e.,

$$\widetilde{A}_r\Psi_J=0,\quad \widetilde{A}_r:=\pi(a_r),\quad r=1,2.\tag{1.6c}$$

Equations (1.6) represent the main part of necessary condi-

 0022-2488/88/030546-14$02.50

tions as mentioned above. One more condition is implied by algebraic irreducibility of π that is equivalent to requiring that Ψ_J (and any other $\Psi\in\mathscr{D}$) be cyclic, i.e.,

$$\mathscr{D} = \mathscr{U}(\tilde{A}_1,\tilde{A}_2,\tilde{A}_{-1},\tilde{A}_{-2})\Psi_J, \tag{1.7}$$

where $\mathscr{U}(\cdots)$ denotes the linear hull of the set of all operators

$$T_n = \prod_{r=1}^{n} (\tilde{A}_1^{j_r}\tilde{A}_2^{k_r}\tilde{A}_{-1}^{l_r}\tilde{A}_{-2}^{m_r}),$$

with $n = 0,1,\dots$, and any non-negative integers j_r, k_r, l_r, and m_r.[3] We have argued in Ref. 1 that explicit knowledge of a basis $\mathscr{E}$ of $\mathscr{D}$ would be useful for proving algebraic irreducibility and for introducing a scalar product of $\mathscr{D}$ for which Eq. (1.3) will hold. It has been further suggested that such a basis should consist of vectors Φ_K satisfying

$$\tilde{N}_r\Phi_K = (\nu_r + n_r(K))\Phi_K, \quad r = 1,2, \tag{1.8a}$$

where

$$\tilde{N}_r := \tfrac{1}{2}(\tilde{N} + (-1)^r\tilde{H}) = \pi(b_{r,-r}),$$
$$\nu_r := \tfrac{1}{2}(\nu + (-1)^r J), \tag{1.8b}$$

and $n_r(K)$ are some integers. In particular, the HW vector Ψ_J is in $\mathscr{E}$ and $n_1(K) = n_2(K) = 0$ for $\Phi_K = \Psi_J$.

The paper is organized as follows. Section II deals with solving Eqs. (1.6) for a family of representations $\{\Omega\}\equiv\{\Omega_N\}$ $(N = 2,4,\dots)$ of $B(0,2)$ in terms of linear differential operators on the space $C_N^\infty(\tilde{M})$ of infinitely differentiable vector functions $\tilde{M}\ni x\mapsto\Phi(x)\in\mathbb{C}^N$.[4] In the third section the family $\{\Omega_2\}$ is considered in detail; for each $J = 0,1,\dots$, the HW vector Ψ_J is found and an infinite set $\mathscr{E}_J\subset C_2^\infty$ containing Ψ_J is obtained. It is further shown that $z\mapsto\pi_J(z) := \Omega_2(z) \restriction (\mathscr{E}_J)_{\text{lin}}$ is an algebraically irreducible representation of $B(0,2)$ and that the *-condition (1.3) holds for the usual L^2-scalar product; the completion of $(\mathscr{E}_J)_{\text{lin}}$ under this scalar product equals $L^2(\mathbb{R}^+\times(-\pi,\pi)\times\mathbb{R}^+)\otimes\mathbb{C}^2$. The dimension of the vacuum subspace of π_J equals $J+1$ and the family $\{\pi_J$: $J = 0,1,\dots\}$ is just the set of all the massless representations of osp(1.4).[5]

In Sec. IV we show that each element of $\mathscr{E}_J$ is an analytic vector for each of the operators $\pi_J(a_j)$ and $\pi_J(b_{jk})$ and hence a basis in $B(0,2)$ can be chosen such that all the generators are represented by essentially self-adjoint operators. Reduction of each π_J with respect to so(3,2) is performed and weight diagrams of the resulting irreducible components are found.

The next section deals with the remaining families $\{\Omega_N\}$, $N = 4,8,12,\dots$. An analysis similar to that in Sec. III is performed; detailed results are given for the case $N = 4$ that covers just all the massive representations belonging to the second class of Ref. 5. Finally, in the last section our results are compared to those of earlier works.

II. ANALYSIS OF NECESSARY CONDITIONS

In Ref. 4 we have presented for $N = 2,4,\dots$ families $\{\Omega\}\equiv\{\Omega_N\}$ of linear representations of $B(0,2)$ in terms of linear differential operators on the vector space $C_N^\infty(M)$ of $\mathbb{C}^N$-valued functions that are infinitely differentiable on $M := \mathbb{R}\times(0,\infty)\times(0,\infty)$. For a given even N the family $\{\Omega_N\}$ is labeled by one real parameter $\varkappa$ that takes values in some interval $\mathscr{K}_N$. Explicit formulas given below for the odd generators $\tilde{A}_j := \Omega(a_j)$ are related to expressions for the operators $\hat{Y}_j$ of Ref. 4 by

$$\tilde{A}_j = 2^{-1/2}V(\hat{Y}_j - i\hat{Y}_{-j})V^{-1}. \tag{2.1}$$

Here V is a bijection of spaces $C^\infty(M)$ and $C^\infty(\tilde{M})$ with $\tilde{M} := (0,\infty)\times(0,\pi)\times(0,\infty)$ given by

$$(V\psi)(\rho,\varphi,z) := \rho^{1/2}\psi(\rho\cos\varphi, z, \rho\sin\varphi).$$

Expressing the V images of the partial derivatives $p_k\psi\equiv\partial_k\psi$, $k = 1,2,3$, in terms of $\partial_\rho V\psi$, $\partial_\varphi V\psi$, and $\partial_z V\psi$ yields

$$Vp_1V^{-1} = \rho^{1/2}\cos\varphi\,\partial_\rho\,\rho^{-1/2} - (1/\rho)\sin\varphi\,\partial_\varphi,$$
$$Vp_2V^{-1} = \partial_z, \tag{2.2}$$
$$Vp_3V^{-1} = \rho^{1/2}\sin\varphi\,\partial_\rho\,\rho^{-1/2} + (1/\rho)\cos\varphi\,\partial_\varphi.$$

Then, by Theorem III.3 of Ref. 4 one gets

$$\tilde{A}_1 = 2^{-1/2}\bar{\eta}[(\rho + \partial_\rho - 1/2\rho)(\cos\varphi\otimes A - i\sin\varphi\otimes B) + (i/\rho)(-\cos\varphi\,\partial_\varphi\otimes B + i\sin\varphi\,\partial_\varphi\otimes A + (1/2\sin\varphi)\otimes C)], \tag{2.3}$$

$$\tilde{A}_2 = 2^{-1/2}\bar{\eta}\Big[(z + \partial_z)\otimes A + \frac{1}{z}\Big(i\,\partial_\varphi\otimes B - \frac{i}{2}\cot\varphi\otimes C - \frac{1}{2}\otimes D\Big)\Big],$$

where $\eta := \exp(i\pi/4)$. The $N\times N$ matrices A, B, C, and D (Ref. 6) satisfy the anticommutation relations as given in Ref. 4; A and C are Hermitian and the remaining two anti-Hermitian.

The involutive map $\hat{D}\mapsto\hat{D}^\#$ defined in Ref. 1 for linear differential operators on $C_N^\infty(M)$ is transformed through V to the involution $\tilde{D}\mapsto\tilde{D}^\#$ on the space $\tilde{\Lambda}_N$ of linear differential operators on $C_N^\infty(\tilde{M})$:

$$\tilde{D}^\# := V\hat{D}^\# V^{-1}. \tag{2.4}$$

One then has, for $r = 1,2$,

$$\tilde{A}_r^\# = \tilde{A}_{-r} \tag{2.5a}$$

and $\partial_\rho^\# = -\partial_\rho$, $\partial_\varphi^\# = -\partial_\varphi$, $\partial_z^\# = -\partial_z$; consequently, Eqs. (2.3) yield

$$\tilde{A}_r^\# = -i\tilde{A}_r + \tilde{\delta}_r, \tag{2.5b}$$

with

$$\tilde{\delta}_1 := 2^{1/2}\eta\rho(\cos\varphi\otimes A - i\sin\varphi\otimes B), \quad \tilde{\delta}_2 := 2^{1/2}\eta z\otimes A. \tag{2.5c}$$

For the particle-number operator one has by Eq. (1.8b) $\tilde{N}_r = \frac{1}{2}\{A_r,\tilde{A}_{-r}\}$, which can be expressed via even generators $\hat{X}_{jk}$ of Ref. 4 as follows: $\tilde{N}_r = -(i/2)V(\hat{X}_{rr} + \hat{X}_{-r-r})V^{-1}$ (cf. Remark 1.1). Theorem III.3 of Ref. 4 now gives

$$\tilde{N}_1 = \frac{1}{2}\Big[\rho^2 - \partial_\rho^2 - \rho^{-2}\Big(\partial_\varphi^2 + \frac{1}{4} - \frac{1}{\sin^2\varphi}\otimes T\Big)\Big],$$
$$T := \Big[\frac{1}{2}(BC + 1)\Big]^2 - \frac{1}{4}, \tag{2.6}$$
$$\tilde{N}_1 = \frac{1}{2}\Big[z^2 - \partial_z^2 - z^{-2}\Big(\frac{1}{4} + \Big(\partial_\varphi\otimes AB - \frac{i}{2}\otimes(I - AD) - \frac{1}{2}\cot\varphi\otimes AC\Big)^2\Big)\Big].$$

In addition, one has $\widetilde{N}_r^{\#} = \widetilde{N}_r$ for $r = 1,2$.

For analyzing Eqs. (1.6) it is convenient to rewrite operators $\widetilde{E}, \widetilde{F}, \widetilde{H}$, and $\widetilde{N}$ with the help of Eq. (2.5a) as follows:

$$\widetilde{E} = \tfrac{1}{2}\{\widetilde{A}_2,\widetilde{A}_1^{\#}\}, \quad \widetilde{F} = \tfrac{1}{2}\{\widetilde{A}_1,\widetilde{A}_2^{\#}\},$$
$$\widetilde{H} = \widetilde{N}_2 - \widetilde{N}_1, \quad \widetilde{N} = \widetilde{N}_1 + \widetilde{N}_2.$$

Then, in view of $\widetilde{A}_r\Psi = 0$ for $\Psi\in\mathscr{D}_{\rm vac}$ and Eq. (2.5b), we get

$$\{\widetilde{A}_r,\widetilde{A}_s^{\#}\}\Psi = \widetilde{A}_r\widetilde{A}_s^{\#}\Psi = \widetilde{A}_r\widetilde{\delta}_s\Psi = \{\widetilde{A}_r,\widetilde{\delta}_s\}\Psi, \quad r,s = 1,2.$$

Further simplification of Eqs. (1.6) is achieved via transformation $\widetilde{D}\mapsto\widetilde{D}^{(f)} := T_f\widetilde{D}T_f^{-1}$ on the space $\widetilde{\Lambda}_N$, where T_f is the operator of multiplication by the function $[\rho,\varphi,z]\mapsto f(\rho,\varphi,z) := (\rho z)^{-1/2}\exp\tfrac{1}{2}(\rho^2 + z^2)$. Let $\Psi\in\mathscr{D}_{\rm vac}$ and $\Phi := T_f\Psi$; then one gets, with the help of Eqs. (2.3), (2.5c), and anticommutation relations for matrices A, B, C, D,

$$T_f\widetilde{F}\Psi = \widetilde{F}^{(f)}\Phi = (z/\rho)\sin\varphi(\rho\cot\varphi\,\partial_\rho - \partial_\varphi)\Phi, \tag{2.7}$$

$$T_f\widetilde{H}\Psi = \widetilde{H}^{(f)}\Phi = (z\,\partial_z - \rho\,\partial_\rho)\Phi, \tag{2.8a}$$

$$T_f\widetilde{N}\Psi = \widetilde{N}^{(f)}\Phi = (2 + \rho\,\partial_\rho + z\,\partial_z)\Phi, \tag{2.8b}$$

$$T_f\widetilde{E}\Psi = \widetilde{E}^{(f)}\Phi = (\rho/z)\sin\varphi\,[z\cot\varphi\,\partial_z - (\partial_\varphi + i\otimes U)]\Phi, \tag{2.9}$$

with $U := -\tfrac{1}{4}(\{B,D\} + 2AB)$. The expressions (2.7) and (2.8) do not contain matrices and hence each of equations $\widetilde{F}^{(f)}\Phi_J = 0$ and $\widetilde{H}^{(f)}\Phi_J = J\Phi_J$, where

$$\Phi_J := T_f\Psi_J, \tag{2.10}$$

represents an uncoupled system of N identical equations for components $(\Phi_J)_\alpha$ of the vector function Φ_J. The general solution can be found by the method of characteristics.[7] A function $\Phi_J\in C_N^\infty(\widetilde{M})$ satisfies $\widetilde{F}^{(f)}\Phi_J = 0$ iff

$$\Phi_J(\rho,\varphi,z) = u_J(\rho\sin\varphi,z), \tag{2.11a}$$

where u_J is any $\mathbb{C}^{(N)}$-valued function whose components belong to $C^\infty((0,\infty)\times(0,\infty))$. Let us now insert this solution into $\widetilde{H}^{(f)}\Phi_J = J\Phi_J$; by using Eq. (2.8a) and setting $x := \rho\sin\varphi$, we get $(z\,\partial_z - x\,\partial_x - J)u_J = 0$. The general solution reads

$$u_J(x,z) = z^J v_J(xz), \tag{2.11b}$$

v_J being any function of the single variable $y := xz = \rho\sin\varphi$ such that $v_J\in C_N^\infty(0,\infty)$. Further specification of v_J will be obtained from Eq. (1.6b) which can be replaced in view of (1.8b) by

$$\widetilde{N}_1\Psi_J = \tfrac{1}{2}(\widetilde{N} - \widetilde{H})\Psi_J = \nu_1\Psi_J,$$

where ν_1 is some non-negative number. Using Eqs. (2.8) and denoting $y := xz = \rho z\sin\varphi$, we find that the function v_J satisfies the following equation:

$$yv_J'(y) = (\nu_1 - 1)v_J(y).$$

The solution reads

$$v_J(y) = y^{\nu_1 - 1}\otimes c_J,$$

where c_J is a constant vector from $\mathbb{C}^N$.

Next we pass to the condition (1.6c); we insert

$$\Psi_J(\rho,\varphi,z) = (\rho z)^{1/2}\exp(-(\rho^2 + z^2)/2)z^J \times(\rho z\sin\varphi)^{\nu_1 - 1}\otimes c_J \tag{2.12}$$

into $(\widetilde{A}_1 \pm \widetilde{A}_2)\Psi_J = 0$ and make use of the relation $S = BC - AD + 1$, where S is a diagonal matrix.[6] This leads to the following pair of equations for c_J:

$$Sc_J = 4(1 - \nu_1 - J/2)c_J, \tag{2.13a}$$

$$(S + 2AD - 2J - 2)c_J = 0. \tag{2.13b}$$

The second equation can be transformed with the help of $2D = [S,A]$ and $A^{-1} = A$ (see Ref. 4) to

$$(S - 2J - 2)Ac_J = 0. \tag{2.14}$$

Let us finally consider the first of conditions (1.6a), i.e., $(\widetilde{E}^{(f)})^{J+1}\Phi_J = 0$. From (2.9) and (2.12) one finds, for $k = 1, 2,\ldots,$

$$(\widetilde{E}^{(f)})^k\Phi_J = \rho^k z^{J-k}(\rho z\sin\varphi)^{\nu_1 - 1}\otimes M_k^{(J)}(\varphi)c_J, \tag{2.15a}$$

where the matrix-valued functions $\varphi\mapsto M_k^{(J)}(\varphi)$ are given by $M_1^{(J)}(\varphi) := J\cos\varphi - \sin\varphi\,(\partial_\varphi + iU)$ and

$$M_k^{(J)}(\varphi) := [(J - k + 1)\cos\varphi - \sin\varphi(\partial_\varphi + iU)] \times M_{k-1}^{(J)}(\varphi), \quad k = 2,3,\ldots.$$

With the help of the relation

$$[\partial_\varphi + i(U + \alpha)]\{(\beta + 1)\cos\varphi - \sin\varphi\,[\partial_\varphi + i(U + \alpha + \beta)]\}$$
$$= \{\beta\cos\varphi - \sin\varphi\,[\partial_\varphi + i(U + \alpha + \beta + 1)]\} \times[\partial_\varphi + i(U + \alpha - 1)],$$

which holds for any $\alpha, \beta\in\mathbb{C}$, we get

$$M_{J+1}^{(J)}(\varphi) = (-i\sin\varphi)^{J+1}\prod_{l=0}^{J}(U + J - 2l).$$

The condition $(\widetilde{E}^{(f)})^{J+1}\Phi_J = 0$ is thus equivalent to

$$\prod_{l=0}^{J}(U + J - 2l)c_J = 0. \tag{2.15b}$$

By Eq. (2.12) one sees that Ψ_J is nonzero if and only if $c_J \neq 0$. Then the vectors $\widetilde{E}^k\Psi_J$, for $k = 0,1,\ldots,J$, are linearly independent; this assertion easily follows from $\widetilde{H}\Psi_J = J\Psi_J$, $\widetilde{F}\Psi_J = 0$, and commutation relations for $\widetilde{E}$, $\widetilde{F}$, and $\widetilde{H}$. Now $c_J\neq 0$ if and only if $Ac_J\neq 0$ and thus we can conclude the analysis of necessary conditions as follows.

Proposition 2.1: For given non-negative integer J and positive even N the vector function Ψ_J fulfills Eqs. (1.6) if and only if it has the form (2.12), where $\nu_1\geqslant 0$, $c_J\in\mathbb{C}^N$ is an eigenvector of S with eigenvalue $4(1 - \nu_1 - J/2)$ satisfying Eq. (2.15b), and Ac_J is an eigenvector of S with eigenvalue $2J + 2$.

This proposition represents restriction for the parameter $\varkappa$, on which the matrices A, U, and S depend, and selects in this way from among the members of the family $\{\Omega_N\}$ a subset $\{\Omega_N\}_J$ of admissible representations: a nontrivial HW vector Ψ_J does not exist unless $\Omega\in\{\Omega_N\}_J$. Since the matrices A, U and S are known for any even N, the subset of admissible representations can be explicitly specified for all values of N and J.

One immediately sees that the HW vector does not exist for any $\Omega\in\{\Omega_{4m+2}\}$, $m = 1,2,\ldots$. In fact, in this case the only positive eigenvalues of S are $2m - 1 \pm \vartheta(\varkappa)$, where

$|\vartheta(\varkappa)| < 1$ (see §3 of the second part of Ref. 4), which is not equal to $2J+2$ for any $J=0,1,\dots$. In this way the families $\{\Omega_N\}$ are excluded for $N=6,10,14,\dots$. The case $N=2$ will be considered in the next two sections and the remaining cases $N=4m$, $m=1,2,\dots$, in Sec. V.

III. IRREDUCIBLE REPRESENTATIONS ON THE SPACE OF TWO-COMPONENT VECTOR FUNCTIONS

Here and in the next section only the family $\{\Omega_2\}\equiv\{\Omega_2^{(\varkappa)}: \varkappa\in[-\frac{9}{2},\infty)\}$ will be considered. Each representation $\Omega_2^{(\varkappa)}$ is determined by eight matrices that are expressed via Pauli matrices and a real ϑ related to $\varkappa$ by $\vartheta^2=2\varkappa+9$ (see Ref. 4, Appendix to the second part):

$$A=-\sigma_2,\quad B=-i\sigma_1,\quad C=0,\quad D=-i\vartheta\sigma_1,$$
$$T=V=0,\quad U=(\vartheta-\sigma_3)/2,\quad S=I-\vartheta\sigma_3. \tag{3.1}$$

Let us find by Proposition 2.1 the set of admissible representations $\{\Omega_2\}_J$ for $J=0,1,\dots$. The spectrum of S is degenerated just for $\varkappa=-\frac{9}{2}$, i.e., for $\vartheta=0$; then $S=I$, so that $2J+2$ is not its eigenvalue for any J and, consequently, $\Omega_2^{(-9/2)}$ is not admissible. If $\vartheta\neq 0$, then S has two nondegenerated eigenvalues $1-\vartheta$ and $1+\vartheta$. Accordingly, there are two cases: in the first one we set $4(1-\nu_1-J/2)=1-\vartheta$ which implies $c_J=\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ (up to a nonzero factor). Then $Ac_J=\left[\begin{smallmatrix}0\\-i\end{smallmatrix}\right]$ and Eq. (2.14) yields $\vartheta=2J+1$. By inserting this value into $4(1-\nu_1-J/2)=1-\vartheta$, we get $\nu_1=1$; further we find $Uc_J=Jc_J$, i.e., Eq. (2.15b) is satisfied. The corresponding HW vector, which will be denoted $\Psi_J^{(+)}$, is given by

$$\Psi_J^{(+)}(\rho,\varphi,z)=(\rho z)^{1/2}z^J\exp\left(-\frac{z^2+\rho^2}{2}\right)\otimes\begin{bmatrix}1\\0\end{bmatrix}. \tag{3.2}$$

Now one gets from Eq. (2.15a), by induction for $k=0,1,\dots,J$,

$$(\widetilde{E}^k\Psi_J^{(+)})(\rho,\varphi,z)$$
$$=[J!/(J-k)!](\rho/z)^k e^{-ik\varphi}\Psi_J^{(+)}(\rho,\varphi,z). \tag{3.3}$$

Similarly, the second case $4(1-\nu_1-J/2)=1+\vartheta$ leads to $\vartheta=-(2J+1)$,

$$\Psi_J^{(-)}(\rho,\varphi,z)=(\rho z)^{1/2}z^J\exp\left(-\frac{z^2+\rho^2}{2}\right)\otimes\begin{bmatrix}0\\1\end{bmatrix}, \tag{3.2'}$$

and $\widetilde{E}_k\Psi_J^{(-)}=[J!/(J-k)!](\rho/z)^k e^{ik\varphi}\Psi_J^{(-)}$. Thus for each $J=0,1,\dots$ the set $\{\Omega_2\}_J$ consists of two representations with $\vartheta=\pm(2J+1)$.

Let us denote by $\widetilde{A}_r(\pm J)$ the operators we obtain by inserting matrices (3.1) with $\vartheta=\pm(2J+1)$ into Eqs. (2.3). It is convenient to introduce for $\mu=\pm1$ matrices $\sigma^\mu:=\frac{1}{2}(\sigma_1-i\mu\sigma_2)$; then

$$\widetilde{A}_1(J)=\widetilde{A}_1(-J)=-2^{-1/2}\eta\sum_{\mu=\pm1}\mu e^{-i\mu\varphi}\times\left(\partial_\rho+\rho-\frac{\frac{1}{2}+i\mu\partial_\varphi}{\rho}\right)\otimes\sigma^\mu,$$

$$\widetilde{A}_1^\#(J)=\widetilde{A}_1^\#(-J)=-2^{-1/2}\bar{\eta}\sum_{\mu=\pm1}\mu e^{-i\mu\varphi}\times\left(\partial_\rho-\rho-\frac{\frac{1}{2}+i\mu\partial_\varphi}{\rho}\right)\otimes\sigma^\mu, \tag{3.4a}$$

$$\widetilde{A}_2(\pm J)=-2^{-1/2}\eta\sum_{\mu=\pm1}\mu\times\left[\partial_z+z-\frac{\mu}{z}\left(\pm\left(J+\frac{1}{2}\right)-i\partial_\varphi\right)\right]\otimes\sigma^\mu,$$

$$\widetilde{A}_2^\#(\pm J)=-2^{-1/2}\bar{\eta}\sum_{\mu=\pm1}\mu\times\left[\partial_z-z-\frac{\mu}{z}\left(\pm\left(J+\frac{1}{2}\right)-i\partial_\varphi\right)\right]\otimes\sigma^\mu. \tag{3.4b}$$

These formulas and Eqs. (3.2) and (3.2′) imply that any $\Psi\in\mathscr{D}_J^{(\pm)}=\mathscr{U}\{A_r(\pm J), A_r^\#(\pm J): r=1,2\}\Psi_J^{(\pm)}$ [cf. Eq. (1.7)] depends on the variable φ as follows:

$$\Psi(\rho,\varphi,z)=\sum_{m\in Z_\psi}e^{im\varphi}\sum_{\mu=\pm1}\Psi_m^{(\mu)}(\rho,z)\otimes|\mu\rangle, \tag{3.5a}$$

where Z_ψ is a finite set of integers, $\Psi_m^{(\mu)}\in C^\infty((0,\infty)\times(0,\infty))$ and

$$|+\rangle:=\begin{bmatrix}1\\0\end{bmatrix},\quad |-\rangle:=\begin{bmatrix}0\\1\end{bmatrix}. \tag{3.5b}$$

Thus $\mathscr{D}_J^{(\pm)}$ is a subspace not only in $C_2^\infty(\widetilde{M})$ but also in $C_2^\infty(\widetilde{M}_{\rm ext})$ with $\widetilde{M}_{\rm ext}:=\{[\rho,\varphi,z]: \rho,z\in(0,\infty), \varphi\in(-\pi,\pi)\}$. This extension of $\widetilde{M}$ is possible owing to special properties of matrices (3.1): they cause that the functions $\Psi_J^{(\pm)}$ do not depend on φ and in Eqs. (3.4) no terms containing $\cot\varphi$ occur [cf. (2.3)].

The inclusion $\mathscr{D}_J^{(\pm)}\subset C_2^\infty(\widetilde{M}_{\rm ext})$ has the following important consequence.

Proposition 3.1: Let J be any non-negative integer and $\pi_J^{(\pm)}$ be the representation of $B(0,2)$ that arises if one restricts Ω_2 for $\vartheta=\pm(2J+1)$ to the subspace $\mathscr{D}_J^{(\pm)}$. Then $\pi_J^{(-)}=\widetilde{T}\pi_J^{(+)}\widetilde{T}^{-1}$, where $\widetilde{T}=-iR_\varphi\otimes\sigma_2$ and R_φ is the operator of reflection with respect to the variable φ.

Proof: It is sufficient to verify $\Psi_J^{(-)}=\widetilde{T}\Psi_J^{(+)}$, $\widetilde{T}\widetilde{A}_r(J)\widetilde{T}^{-1}=\widetilde{A}_r(-J)$, and $\widetilde{T}\widetilde{A}_r^\#(J)\widetilde{T}^{-1}=\widetilde{A}_r^\#(-J)$, for $r=1,2$. The first relation immediately follows from Eqs. (3.2) and (3.2′); the remaining ones are obtained by Eqs. (3.4) if one uses $\sigma_2\sigma^\mu\sigma_2=-\sigma^{-\mu}$. ∎

It thus suffices to consider representations $\pi_J^{(+)}$; the upper index will hereafter be dropped, i.e., $\pi_J\equiv\pi_J^{(+)}$, and similarly we set $\Psi_J\equiv\Psi_J^{(+)}$ and $\mathscr{D}_J\equiv\mathscr{D}_J^{(+)}$. Besides Eqs. (3.4) we shall also need expressions for even generators. Explicit formulas will be given for $x_{jk}:=\frac{1}{2}(b_{jk}-b_{-j-k})+(i/2)(b_{-jk}+b_{j-k})$, with $j,k=\pm1,\pm2$ and $j\geqslant k$, rather than for b_{jk}, since the elements x_{jk} also span the even subalgebra of $B(0,2)$ and the result is for them simpler. By using Theorem III.3 of Ref. 4, inserting for the matrices $A,B,\dots$ from Eq. (3.1) with $\vartheta=2J+1$ and denoting [cf. Eqs. (2.2)]

$$\tilde p_1 := \rho^{1/2}\cos\varphi\,\partial_\rho\rho^{-1/2} - (1/\rho)\sin\varphi\,\partial_\varphi,$$
$$\tilde p_3 := \rho^{1/2}\sin\varphi\,\partial_\rho\rho^{-1/2} + (1/\rho)\cos\varphi\,\partial_\varphi,$$

we find that the operators $\widetilde{X}_{jk}(J) \equiv \widetilde{X}_{jk} := V\widehat{X}_{jk}V^{-1}$ are as follows:

$$\widetilde{X}_{-2-2} = iz^2\otimes I,\quad \widetilde{X}_{-1-2} = i\rho z\cos\varphi\otimes I,$$
$$\widetilde{X}_{1-2} = [z(\partial_\rho - 1/2\rho)\cos\varphi - (z/\rho)\sin\varphi\,\partial_\varphi]\otimes I,$$
$$\widetilde{X}_{2-2} = (z\,\partial_z + \tfrac12)\otimes I,$$
$$\widetilde{X}_{-1-1} = i\rho^2\otimes I,\quad \widetilde{X}_{1-1} = (\rho\,\partial_\rho + \tfrac12)\otimes I,$$
$$\widetilde{X}_{2-1} = \rho(\partial_z - 1/2z)\cos\varphi\otimes I - i(\rho/z)\times\sin\varphi\,[(J+\tfrac12 - i\,\partial_\varphi)\otimes I - \tfrac12\otimes\sigma_3], \tag{3.6}$$
$$\widetilde{X}_{11} = -i[\partial_\rho^2 + \rho^{-2}(\partial_\varphi^2 + \tfrac14)]\otimes I,$$
$$\widetilde{X}_{21} = -i\tilde p_1(\partial_z - 1/2z)\otimes I - (1/z)\,\tilde p_3\times[(J+\tfrac12 - i\,\partial_\varphi)\otimes I - \tfrac12\otimes\sigma_3],$$
$$\widetilde{X}_{22} = -i[\partial_z^2 - z^{-2}(J+\tfrac12 - i\,\partial_\varphi)^2]\otimes I - iz^{-2}(J+\tfrac12 - i\,\partial_\varphi)\otimes\sigma_3.$$

According to what has been argued in Sec. I, a basis in $\mathscr{D}_J$ consisting of common eigenfunctions of the operators $\widetilde{N}_1$ and $\widetilde{N}_2$ should be found. The corresponding explicit expressions can be obtained from Eqs. (3.6) using $\widetilde{N}_r = -(i/2)(\widetilde{X}_{rr} + \widetilde{X}_{-r,-r})$. By Eqs. (2.8b) and (3.2) we find $\widetilde{N}\Psi_J = (2+J)\Psi_J$ and then (1.8b) implies that the eigenvalues of $\widetilde{N}_r$ and $\nu_r + n_r$ with $\nu_1 = 1$ and $\nu_2 = J+1$. Now, in view of Eq. (3.5), the sought eigenfunctions can be written as $e^{im\varphi}\psi^{(m,\mu)}_{n_1n_2}(\rho,z)\otimes|\mu\rangle$, where the functions $\psi^{(m,\mu)}_{n_1n_2}\in C^\infty((0,\infty)\times(0,\infty))$ satisfy

$$\tfrac12[\rho^2 - \partial_\rho^2 + \rho^{-2}(m^2 - \tfrac14)]\psi^{(m,\mu)}_{n_1n_2} = (1+n_1)\psi^{(m,\mu)}_{n_1n_2},$$
$$\tfrac12[z^2 - \partial_z^2 + z^{-2}((J+m+(1-\mu)/2)^2 - \tfrac14)]\psi^{(m,\mu)}_{n_1n_1} = (J+1+n_2)\psi^{(m,\mu)}_{n_1n_2}. \tag{3.7}$$

We will select for each J a set $\mathscr{E}_J$ of these functions such that $(\mathscr{E}_J)_{\rm lin} = \mathscr{D}_J$ and $\Psi_J\in\mathscr{E}_J$. For getting it the same functions as in Ref. 1 will be used, viz.

$$x\mapsto f_n^{(\alpha)}(x) := c_n^{(\alpha)}x^{\alpha+1/2}e^{-x^2/2}L_n^{(\alpha)}(x^2),\quad x>0, \tag{3.8a}$$

where $\alpha > -1$, $n = 0,1,\ldots$, $c_n^{(\alpha)} := (2n!/\Gamma(\alpha+n+1))^{1/2}$, and $L_n^{(\alpha)}$ are the Laguerre polynomials (Ref. 8, §8.97). They fulfill

$$\tfrac12[-d_x^2 + x^2 + x^{-2}(\alpha^2 - \tfrac14)]f_n^{(\alpha)} = (2n+\alpha+1)f_n^{(\alpha)} \tag{3.8b}$$

and obey the following recurrence relations:

$$(d_x + x - (\tfrac12+\alpha)/x)f_n^{(\alpha)} = -2n^{1/2}f_{n-1}^{(\alpha+1)},$$
$$(d_x - x - (\tfrac12+\alpha)/x)f_n^{(\alpha)} = -2(\alpha+n+1)^{1/2}f_n^{(\alpha+1)}, \tag{3.9a}$$

$$(d_x + x - (\tfrac12-\alpha)/x)f_n^{(\alpha)} = 2(n+\alpha)^{1/2}f_n^{(\alpha-1)},$$
$$(d_x - x - (\tfrac12-\alpha)/x)f_n^{(\alpha)} = 2(n+1)^{1/2}f_{n+1}^{(\alpha-1)}. \tag{3.9b}$$

Notice that the relations (3.9b) make sense for $\alpha>0$ only.

Let us now introduce for any integer m, non-negative integers k,l and $\mu = \pm1$ the linearly independent functions $|klm;\mu\rangle_J\in C_2^\infty(\widetilde{M}_{\rm ext})$,

$$|klm;\mu\rangle_J(\rho,\varphi,z) := f_k^{|m|}(\rho)\,f_l^{|J+m+(1-\mu)/2|}(z)\times\frac{e^{im\varphi}}{(2\pi)^{1/2}}\otimes|\mu\rangle, \tag{3.10a}$$

and denote

$$\mathscr{E}_J := \{|klm;\mu\rangle_J:\ k,l = 0,1,\ldots,\ m = 0,\pm1,\ldots,\ \mu = \pm1\}. \tag{3.10b}$$

From Eqs. (3.7) and (3.8b) it immediately follows that each element of $\mathscr{E}_J$ fulfills

$$\widetilde{N}_1|klm;\mu\rangle_J = (2k+|m|+1)|klm;\mu\rangle_J \tag{3.11a}$$

and

$$\widetilde{N}_2|klm;\mu\rangle_J = (2l + |J+m+(1-\mu)/2| + 1)|klm;\mu\rangle_J\,. \tag{3.11b}$$

Thus $|klm;\mu\rangle_J$ are common eigenfunctions of $\widetilde{N}_1$ and $\widetilde{N}_2$ with integer eigenvalues. With the help of Eqs. (3.9) we shall now find for $r = 1,2$ the action of $\widetilde{A}_r(J)$ and $\widetilde{A}_r^\#(J)$ on any $|klm;\mu\rangle_J$. By taking into account that $\sigma^\mu|\nu\rangle = \delta_{\mu-\nu}|-\nu\rangle$ for $\mu,\nu = \pm1$, we have

$$\frac{\bar\eta}{2^{1/2}}\widetilde{A}_1(J)|klm;\mu\rangle_J = \begin{cases} -\mu(k+\mu m)^{1/2}|klm-\mu;-\mu\rangle_J, & \mu m\geqslant 1,\\ \mu k^{1/2}|k-1lm-\mu;-\mu\rangle_J, & \mu m\leqslant 0,\end{cases} \tag{3.12a}$$

$$\frac{\eta}{2^{1/2}}\widetilde{A}_1^\#(J)|klm;\mu\rangle_J = \begin{cases} -\mu(k+1)^{1/2}|k+1lm-\mu;-\mu\rangle_J, & \mu m\geqslant 1,\\ \mu(k+1-\mu m)^{1/2}|klm-\mu;-\mu\rangle_J, & \mu m\leqslant 0,\end{cases} \tag{3.12b}$$

$$\frac{\bar\eta}{2^{1/2}}\widetilde{A}_2(J)|klm;\mu\rangle_J = \begin{cases} \mu l^{1/2}|kl-1m;-\mu\rangle_J, & \mu(J+m+\tfrac12)\geqslant\tfrac12,\\ -\mu(l+\tfrac12-\mu(J+m+\tfrac12))^{1/2}|klm;-\mu\rangle_J, & \mu(J+m+\tfrac12)\leqslant-\tfrac12,\end{cases} \tag{3.12c}$$

$$\frac{\eta}{2^{1/2}}\widetilde{A}_2^\#(J)|klm;\mu\rangle_J = \begin{cases} \mu(l+\tfrac12+\mu(J+m+\tfrac12))^{1/2}|klm;-\mu\rangle_J, & \mu(J+m+\tfrac12)\geqslant\tfrac12,\\ -\mu(l+1)^{1/2}|kl+1m;-\mu\rangle_J, & \mu(J+m+\tfrac12)\leqslant-\tfrac12\,.\end{cases} \tag{3.12d}$$

These formulas show that the subspace $(\mathscr{E}_J)_{\text{lin}} \subset C_2^\infty(\widetilde{M}_{\text{ext}})$ is invariant under all the operators $\widetilde{A}_r(J)$ and $\widetilde{A}_r^\#(J)$, $r=1,2$. Next we shall find with the help of them the intersection of $(\mathscr{E}_J)_{\text{lin}}$ with the vacuum subspace $V_J := \{\Phi\in C_2^\infty(\widetilde{M}_{\text{ext}}): \widetilde{A}_r(J)\Phi = 0, r=1,2\}$.

Lemma 3.2: The subspace $\mathscr{D}_J^{\text{vac}} := (\mathscr{E}_J)_{\text{lin}} \cap V_J$ is spanned by linearly independent functions

$$\widetilde{E}^m \Psi_J = J!\left(\frac{\pi m!}{2(J-m)!}\right)^{1/2} |00-m;+\rangle_J, \qquad (3.13)$$

for $m=0,1,...,J$, i.e., $\dim \mathscr{D}_J^{\text{vac}} = J+1$.

Proof: By Eqs. (3.12a) and (3.12c) one sees that for $m=0,1,...,J$ the functions $|00-m;+\rangle_J$ are in $\mathscr{D}_J^{\text{vac}}$. On the other hand, let $\Psi\in\mathscr{D}_J^{\text{vac}}$, i.e.,

$$\Psi = \sum_{klm} c_{klm}|klm;+\rangle_J + \sum_{klm} d_{klm}|klm;-\rangle_J$$

and $\widetilde{A}_1(J)\Psi = \widetilde{A}_2(J)\Psi = 0$. The first of these conditions implies, with the help of (3.12a), (3.12c), and linear independence of $|klm{:}\mu\rangle_J$,

$$\Psi = \sum_l \sum_{m<0} c_{0lm}|0lm;+\rangle_J + \sum_l \sum_{m>0} d_{0lm}|0lm;-\rangle_J,$$

and then $\widetilde{A}_2(J)\Psi = 0$ by (3.12b) and (3.12d) yields

$$\Psi = \sum_{m=0}^{J} c_{00-m}|00-m;+\rangle_J. \qquad \blacksquare$$

We shall use shortened notation

$$\mathscr{U}_J \equiv \mathscr{U}(\widetilde{A}_1(J), \widetilde{A}_1^\#(J), \widetilde{A}_2(J), \widetilde{A}_2^\#(J)).$$

Proposition 3.3: (a) One has $\mathscr{D}_J \equiv \mathscr{U}_J\Psi_J \subset (\mathscr{E}_J)_{\text{lin}}$. (b) To each nonzero $\Psi\in\mathscr{D}_J^{\text{vac}}$ there exist $\widetilde{T},\widetilde{S}\in\mathscr{U}_J$ such that $\Psi = \widetilde{T}\Psi_J$ and $\Psi_J = \widetilde{S}\Psi$.

Proof: The first statement is due to $\Psi_J = ((\pi/2)J!)^{1/2}|000;+\rangle_J \in \mathscr{E}_J$ and to invariance of $(\mathscr{E}_J)_{\text{lin}}$ under $\widetilde{A}_r(J)$ and $\widetilde{A}_r^\#(J)$ for $r=1,2$. By Proposition 3.4 and Lemma 3.1 of Ref. 1 and by Lemma 2 one sees that $\widetilde{B}_{r-s} := \frac{1}{2}\{\widetilde{A}_r(J), \widetilde{A}_s^\#(J)\}$, $r,s=1,2$, generate an irreducible representation of gl(2,$\mathbb{C}$) on $\mathscr{D}_J^{\text{vac}}$, which is equivalent to (b). $\blacksquare$

Lemma 3.4: The projections $\widetilde{P}_\pm := \frac{1}{2}\otimes(I\pm\sigma_3)$ belong to $\mathscr{U}_J$.

Proof: It is sufficient to show $I\otimes\sigma_3\in\mathscr{U}_J$. Consider the second-order Casimir element c_2 of sp(4,$\mathbb{R}$); it can be expressed as a biquadratic polynomial function of the odd elements a_r and $a_{-r} = a_r^*$, $r=1,2$; hence $\pi_J(c_2)\in\mathscr{U}_J$. On the other hand, by (3.1) and Theorem II.3 of Ref. 4 one finds $\pi_j(c_2) = \frac{1}{2}\otimes(2J+1-\sigma_3)^2 - 8$ and thus $I\otimes\sigma_3$ is a linear combination of the identity and $\pi_J(c_2)$. $\blacksquare$

Now it is not difficult to show that $\mathscr{E}_J$ is a basis of the subspace $\mathscr{D}_J$. In view of Proposition 3(a) we only must find for each $|klm;\mu\rangle_J$ an operator $\widetilde{T}\in\mathscr{U}_J$ such that $|klm;\mu\rangle_J = \widetilde{T}|000;+\rangle_J$, which can easily be done with the help of Eqs. (3.12) (see Appendix A).

We will further prove that $\mathscr{D}_J$ has no proper subspaces invariant under $\widetilde{A}_r(J)$ and $\widetilde{A}_r^\#(J)$ for $r=1,2$. This assertion holds true if any nonzero $\Psi\in\mathscr{D}_J$ is a cyclic vector; by taking into account that $\Psi_J\sim|000;+\rangle_J$ is cyclic by the very definition of $\mathscr{D}_J$, and making use of $\mathscr{D}_J = (\mathscr{E}_J)_{\text{lin}}$ and of the second statement of Proposition 3, one sees that it suffices to find for each nonzero $\Psi\in(\mathscr{E}_J)_{\text{lin}}$ an operator $\widetilde{S}\in\mathscr{U}_J$ that transforms Ψ into a nonzero element of the vacuum subspace $\mathscr{D}_J^{\text{vac}}$. Moreover, because of Lemma 4, we can assume that Ψ belongs to one of the subspaces $\widetilde{P}_\pm\,\mathscr{D}_J$. Then the sought operator $\widetilde{S}$ can again be found by applying Eqs. (3.12) as is shown in Appendix A.

The fact that $\mathscr{E}_J$ is a basis of $\mathscr{D}_J$ facilitates introducing a scalar product on $\mathscr{D}_J$ such that the *-condition (1.3) will hold. Consider the Hilbert space $\mathscr{H} := L^2(\widetilde{M}_{\text{ext}}, d\rho\, d\varphi\, dz)\otimes\mathbb{C}^2 \equiv L^2(\widetilde{M}_{\text{ext}})\otimes\mathbb{C}^2$ and denote by $\langle\cdot,\cdot\rangle$ the scalar product on $\mathscr{H}$. As $\mathscr{E}_J$ is an orthonormal basis in $\mathscr{H}$, the relation $\mathscr{D}_J = (\mathscr{E}_J)_{\text{lin}}$ means that $\mathscr{D}_J$ is a dense subspace in $\mathscr{H}$. Furthermore, invariance of $\mathscr{D}_J$ under all operators in $\mathscr{U}_J$ implies that the *-condition (1.3) is equivalent to

$$\langle klm;\mu_J|\widetilde{A}_r(J)|k'l'm';\mu_j'\rangle = \overline{\langle k'l'm';\mu_j'|\widetilde{A}_r^\#(J)|klm;\mu_J\rangle}, \qquad (3.14)$$

for $r=1,2$ and all vectors $|klm;\mu\rangle_J, |k'l'm';\mu'\rangle_J\in\mathscr{E}_J$. This condition is indeed fulfilled, as can be checked by a direct calculation with the help of Eqs. (3.12) and orthonormality relations.

We have thus derived basic properties of representations π_J that can be summarized as follows.

Theorem 3.5: For each $J=0,1,...$, the map $z\mapsto\pi_J(z)$, defined via the operators (3.4) and the Racah basis of $B(0,2)$ by

$$\pi_J(a_r) := \widetilde{A}_r(J), \quad \pi_J(a_{-r}) := \widetilde{A}_r^\#(J) \quad (r=1,2)$$

and

$$\pi_J(b_{jk}) := \tfrac{1}{2}\{\pi_J(a_j), \pi_J(a_k)\} \quad (j,k=\pm1,\pm2),$$

is a *-representation of $B(0,2)$ on $L^2(\mathbb{R}^+\times(-\pi,\pi)\times\mathbb{R}^+)\otimes\mathbb{C}^2$ with domain $\mathscr{D}_J = (\mathscr{E}_J)_{\text{lin}}$ given by Eqs. (3.10) and projection $\widetilde{P}_+ = \frac{1}{2}\otimes(I+\sigma_3)$. Moreover, π_J is algebraically irreducible, its vacuum is $(J+1)$-fold degenerate, and the representations π_J and π_J' are nonequivalent if $J\neq J'$.

Proof: Only the last assertion has not been proved. Suppose $\pi_{J'} = \widetilde{U}\pi_J\widetilde{U}^{-1}$; then one has $\mathscr{D}_{J'}^{\text{vac}} = \widetilde{U}\mathscr{D}_J^{\text{vac}}$ (see Lemma 2) and hence

$$J'+1 = \dim\mathscr{D}_{J'}^{\text{vac}} = \dim\mathscr{D}_J^{\text{vac}} = J+1,$$

i.e., $J=J'$. $\blacksquare$

IV. ADDITIONAL PROPERTIES OF REPRESENTATIONS π_J

A. Essential self-adjointness

We have seen that the operators $\widetilde{A}_r(J)$ and $\widetilde{A}_{-r}(J) = \widetilde{A}_r^\#(J)$, $r=1,2$, can be regarded as densely defined operators on $L^2(\widetilde{M}_{\text{ext}})\otimes\mathbb{C}^2$ with the common invariant domain $\mathscr{D}_J = (\mathscr{E}_J)_{\text{lin}}$. Consequently, any $\widetilde{T}\in\mathscr{U}_J$ is also a densely defined operator, which is symmetric if $\widetilde{T}^\# = \widetilde{T}$; this immediately follows by the *-condition (1.3). It turns out that the results of Ref. 1 concerning essential self-adjointness can be generalized for polynomial functions of operators $\widetilde{A}_j(J)$ with $j=\pm1,\pm2$.

Lemma 4.1: Let $p=1,2,...$, and $\widetilde{B}_p$ be a product of any p elements of the set $\{\widetilde{A}_j(J): j=\pm1,\pm2\}$. Then one has for each $|klm;\mu\rangle_J\in\mathscr{E}_J$,

$$\|\widetilde{B}_p|klm;\mu\rangle_J\|^2 \leqslant 2^p \prod_{j=1}^{p} (n_{kl} + |m| + J + 1 + j)$$
$$\equiv 2^p \Pi_{klm}^{(p)}, \qquad (4.1)$$

where $n_{kl} := \max\{k,l\}$.

Proof: Let $\widetilde{B}_0 := I$; then, for $p = 0,1,\ldots,$ one has $\widetilde{B}_{p+1} = \widetilde{B}_p \widetilde{A}_{j_p}$ [hereafter we write $\widetilde{A}_j$ instead of $\widetilde{A}_j(J)$ and similarly $|klm;\mu\rangle \equiv |klm;\mu\rangle_J$], where $j_p \in \{\pm 1, \pm 2\}$, and the assertion can easily be proved by induction using Eqs. (3.12). For example, if $\widetilde{B}_{p+1} = \widetilde{B}_p \widetilde{A}_1$, then we get for $\mu m \geqslant 1$

$$\|\widetilde{B}_{p+1}|klm;\mu\rangle\|^2 = 2(k + |m|)\|\widetilde{B}_p|klm-\mu;-\mu\rangle\|^2$$
$$\leqslant 2^{p+1}(n_{kl} + |m|)\Pi_{klm-\mu}^{(p)}.$$

Now for $\mu m \geqslant 1$ we have $|m-\mu| = \mu m - 1 = |m| - 1$ and

$$\Pi_{klm-\mu}^{(p)} < \prod_{j=1}^{p} (n_{kl} + |m| + J + 1 + j)$$
$$= \frac{\Pi_{klm}^{(p+1)}}{n_{kl} + |m| + J + p + 2};$$

hence

$$\|\widetilde{B}_{p+1}|klm;\mu\rangle\|^2$$
$$= 2^{p+1}(n_{kl} + |m|)\frac{\Pi_{klm}^{(p+1)}}{n_{kl} + |m| + J + p + 2}$$
$$< 2^{p+1}\prod_{klm}^{(p+1)}. \qquad \blacksquare$$

Clearly, there are at most 4^p different operators $\widetilde{B}_p^{(r)} \equiv \widetilde{B}_p$ $(1 \leqslant r \leqslant 4^p)$; the above lemma combined with the argument we have used for proving Proposition 4.5 of Ref. 1 yields the following assertion.

Proposition 4.2: For $p = 1,2,4$ and arbitrary complex $\alpha_1,\ldots,\alpha_{4^p}$, let

$$\widetilde{P}_p := \sum_{r=1}^{4^p} \alpha_r \widetilde{B}_p^{(r)}.$$

Then each $|klm;\mu\rangle$ is an analytic vector of $\widetilde{P}_1$ and $\widetilde{P}_2$ and a semianalytic vector of $\widetilde{P}_4$. In addition, a sufficient condition for $\widetilde{P}_p$ to be essentially self-adjoint (e.s.a.) reads $\widetilde{P}_p = \widetilde{P}_p^{\#}$ for $p = 1,2$ and $\widetilde{P}_p \geqslant 0$ for $p = 4$, respectively.

As an important example consider, for $j,k = \pm 1, \pm 2$, the operators [cf. Eq. (3.6)]

$$\widetilde{P}_2^{(jk)} \equiv i\widetilde{X}_{jk} := (i/4)(\{\widetilde{A}_j,\widetilde{A}_k\} - \{\widetilde{A}_{-j},\widetilde{A}_{-k}\})$$
$$- \tfrac{1}{4}(\{\widetilde{A}_{-j},\widetilde{A}_k\} + \{\widetilde{A}_j,\widetilde{A}_{-k}\}) \qquad (4.2)$$

and

$$\widetilde{P}_1^{(j)} \equiv \bar{\eta}\widetilde{Y}_j := \bar{\eta}2^{-1/2}(\widetilde{A}_j + i\widetilde{A}_{-j}).$$

In view of $\widetilde{A}_j^{\#} = \widetilde{A}_{-j}$ one has $(\widetilde{P}_2^{(jk)})^{\#} = \widetilde{P}_2^{(jk)}$ and $(\widetilde{P}_1^{(j)})^{\#} = \widetilde{P}_1^{(j)}$, i.e., all these operators are e.s.a. Consequently, there is a basis in $B(0,2)$ such that the π_J images of all its elements are e.s.a. operators (see Sec. II of Ref. 1).

With the help of operators $\widetilde{P}_2^{(jk)}$ we can further construct

$$\Delta \equiv \widetilde{P}_4 := \sum_{\substack{j,k=\pm 1,\pm 2\\ j>k}} (\widetilde{P}_2^{(jk)})^2 = -\sum_{\substack{j,k=\pm 1,\pm 2\\ j>k}} \widetilde{X}_{jk}^2;$$

as $\langle(\widetilde{P}_2^{(jk)})^2\Psi,\Psi\rangle = \|\widetilde{P}_2^{(jk)}\Psi\|^2 \geqslant 0$ for any $\Psi \in \mathscr{D}_J$, we have $\Delta \geqslant 0$ and, consequently, Δ is e.s.a. Now Δ is the Nelson operator for the representation of $\mathrm{sp}(4,\mathbb{R}) \sim \mathrm{so}(3,2)$ that arises by restricting π_J to the even subalgebra of osp(1,4). In view of the Nelson theorem[9] this representation can be integrated, which yields a unitary representation of the universal covering group for the component of unity of SO(3,2).

B. Reduction of π_J with respect to so(3,2)

In this subsection the π_J's are regarded as representations of the real Lie superalgebra osp(1,4). By Theorem 3.5 the operator $\pi_J(x)$ commutes with the projection $\widetilde{P}_+$ for each even element $x \in \mathrm{osp}(1,4)$, which means that the restriction of π_J to the even subalgebra so(3,2) of osp(1,4) is reducible. In order to get the components of $\pi_J \restriction \mathrm{so}(3,2)$ corresponding to the projections $\widetilde{P}_+$ and $\widetilde{P}_- := I - \widetilde{P}_+$, let us introduce matrices $\epsilon_\pm := (I \pm \sigma_3)/2$, so that $\widetilde{P}_\pm = 1 \otimes \epsilon_\pm$; now each $x \in \mathrm{so}(3,2)$ equals a real linear combination of x_{jk} and by inserting $\sigma_3 = \Sigma_{\mu=\pm 1}\mu\epsilon_\mu$ into Eq. (3.6), we get

$$\pi_J(x) = \sum_{\mu=\pm 1} \tau_J^{(\mu)}(x) \otimes \epsilon_\mu. \qquad (4.3)$$

Here $\tau_J^{(\mu)}(x)$ is an operator on $L^2(\widetilde{M}_{\mathrm{ext}})$ with the domain $D_{J+(1-\mu)/2}$, where, for $p = 0,1,\ldots,$ we have denoted

$$D_p := \{|klm\rangle_p:\ l,k = 0,1,\ldots,\ m = 0,\pm 1,\ldots\}_{\mathrm{lin}} \qquad (4.4a)$$

with

$$|klm\rangle_p(\rho,\varphi,z) := f_k^{|m|}(\rho) f_l^{|m+p|}(z)[e^{im\varphi}/(2\pi)^{1/2}] \qquad (4.4b)$$

[notice that for each p the set $\{|klm\rangle_p:\ k,l = 0,\ldots,$ $m = 0,\pm 1,\ldots\}$ is an orthonormal basis in $L^2(\widetilde{M}_{\mathrm{ext}})$]. By Eq. (4.3) one sees that $x \mapsto \tau_J^{(\mu)}(x)$ is a representation of so(3,2) on $L^2(\widetilde{M}_{\mathrm{ext}})$ for both $\mu = \pm 1$.

With the help of Eqs. (3.6) we easily find the operators $\tau_J^{(\mu)}(x)$; e.g., we get

$$\tau_J^{(\mu)}(x_{22}) = -i[\partial_z^2 - z^{-2}(J + \tfrac{1}{2} - i\,\partial\varphi)^2]$$
$$- i\mu z^{-2}(J + \tfrac{1}{2} - i\,\partial\varphi).$$

In general, one has, for any $x \in \mathrm{so}(3,2)$,

$$\tau_J^{(\mu)}(x) = \alpha(x) + (J + (1-\mu)/2)\beta(x)$$
$$+ (J + (1-\mu)/2)^2\gamma(x), \qquad (4.5a)$$

where $\alpha(x)$, $\beta(x)$, and $\gamma(x)$ are differential operators on $L^2(\widetilde{M}_{\mathrm{ext}})$ that do not depend on J and μ. It thus holds, for $J = 1,2,$ that

$$\tau_J^{(+)} = \tau_{J-1}^{(-)} =: \tau_J. \qquad (4.5b)$$

Let us further set $\tau_0 := \tau_0^{(+)}$; then τ_J is, for each $J = 0,1,\ldots,$ a representation of so(3,2) on $L^2(\widetilde{M}_{\mathrm{ext}})$ with the domain D_J, which is related to τ_J by [cf. Eq. (4.3)]

$$\tau_J(x)|klm\rangle_J \otimes |+\rangle = \pi_J(x)|klm;+\rangle_J, \quad x \in \mathrm{so}(3,2).$$

This relation implies that τ_J is a skew-symmetric representation since all the operators $\widetilde{X}_{jk} = \pi_J(x_{jk})$ are skew symmetric.

Now the Hilbert space $L^2(\widetilde{M}_{\mathrm{ext}}) \otimes \mathbb{C}^2$ can be identified with $L^2(\widetilde{M}_{\mathrm{ext}}) \oplus L^2(\widetilde{M}_{\mathrm{ext}})$; accordingly, the sought decomposition of $\pi_J \restriction \mathrm{so}(3,2)$ reads

$$\pi_J \upharpoonright so(3,2) = \tau_J \oplus \tau_{J+1}, \quad J = 0,1,\ldots .$$

Are the representations τ_J algebraically irreducible? The following proposition shows that this is not the case for τ_0.

Proposition 4.3: One has $\tau_0 = \tau_{\rm even} \oplus \tau_{\rm odd}$, where $\tau_{\rm even}$ and $\tau_{\rm odd}$ are skew-symmetric representations of so(3,2) on $L^2(\widetilde{M}_{\rm ext})$ with domains $D_{\rm even} := [(I+R)/2]D_0$ and $D_{\rm odd} := [(I-R)/2]D_0$, respectively, and R is the unitary operator of reflection with respect to φ.

Proof: Equation (4.3b) yields $|klm\rangle_0(\rho,\varphi,z) = f_k^{|m|}(\rho)f_l^{|m|}(z)[e^{im\varphi}/(2\pi)^{1/2}]$; thus

$$R\,|klm\rangle_0 = |kl-m\rangle_0,$$

for all k,l,m. Consequently, D_0 is R invariant; further, $R^2 = I$ implies that $(I \pm R)/2$ are orthogonal to each other and hence $D_0 = D_{\rm even} \oplus D_{\rm odd}$. It remains to verify that $R\tau_0(x_{jk})R^{-1} = \tau_0(x_{jk})$, for $j,k = \pm 1, \pm 2$ and $j \geqslant k$; however, this is obvious by Eqs. (3.6). ■

The problem of reduction of π_J with respect to so(3,2) is completely solved as follows.

Theorem 4.4: With the above notation one has

$$\pi_J \upharpoonright so(3,2) = \tau_J \oplus \tau_{J+1}, \tag{4.6a}$$

for $J = 1,2,\ldots$, and

$$\pi_0 \upharpoonright so(3,2) = \tau_{\rm even} \oplus \tau_{\rm odd} \oplus \tau_1, \tag{4.6b}$$

where $\tau_{\rm even}$, $\tau_{\rm odd}$, and τ_J for $J = 1,2,\ldots$ are algebraically irreducible skew-symmetric representations of so(3,2) on $L^2(\widetilde{M}_{\rm ext})$.

Proof: Since the decomposition (4.6) has been derived above, it remains to prove algebraic irreducibility. This can be done in the same way as in Theorem 3.5, i.e., for each pair of vectors φ,ψ belonging to the domain of the representation under consideration one finds an operator

$$\widetilde{T}_{\varphi\psi} \in \mathscr{U}(\tau(b_{jk}) \colon j,k = \pm 1, \pm 2,\ j \geqslant k)$$

such that $\varphi = \widetilde{T}_{\varphi\psi}\psi$. However, the problem is more complicated because the "odd" operators $\widetilde{A}_j$ are no more available; in fact, constructing $\widetilde{T}_{\varphi\psi}$ explicitly is a tedious business and we shall not reproduce it here.

There is another approach for proving algebraic irreducibility, which is based on essential self-adjointness of the Nelson operator Δ for each of the representations $\pi_J \upharpoonright so(3,2)$. As Δ equals the direct sum of the Nelson operators Δ_τ for individual irreducible components τ of $\pi_J \upharpoonright so(3,2)$, essential self-adjointness of Δ implies that each Δ_τ is e.s.a., so that τ is integrable. Then, with the help of one theorem due to Harish-Chandra,[10] one can show that τ is algebraically irreducible if the only bounded operators on $L^2(\widetilde{M}_{\rm ext})$ that commute with $\tau(x)$ for each $x \in so(3,2)$ are multiples of identity. Details will be given elsewhere. ■

Remark 4.5: Let τ be any of representations $\tau_{\rm even}$, $\tau_{\rm odd}$, $\tau_1,\tau_2,\ldots$ and denote the domain of τ by D. Now τ is skew symmetric and irreducible and for the central element

$$z := \frac{i}{2}(b_{1-1} + b_{2-2}) = \frac{1}{2}\sum_{r=1}^{2}(x_{rr} + x_{-r-r})$$

of the maximal compact subalgebra $k \subset so(3,2)$, which is isomorphic to u(2) (see Remark 3.2 of Ref. 1), one has

$$\tau(-iz) = \tfrac{1}{2}\widetilde{N} \upharpoonright D \geqslant 0.$$

Hence each τ is in the Evans list[11] and in order to write it down in the Evans notation we only have to find the corresponding weight diagram. To this purpose is needed in the first place the direct-sum decomposition of the domain D in terms of eigenspaces $\mathscr{N}(\lambda)$ of $\frac{1}{2}\widetilde{N} \upharpoonright D$. Since $\widetilde{N} = N_1 + \widetilde{N}_2$, one finds by Eqs. (3.11), that for $J = 0,1,\ldots$ the eigenvalues of $\tau_J(-iz)$ read

$$\lambda_s^{(J)} \equiv J/2 + s + 1 \colon\ s = 0,1,\ldots,$$

and

$$\dim \mathscr{N}(\lambda_s^{(J)}) = (s+1)(J+s+1).$$

In the second place each $\mathscr{N}(\lambda_s^{(J)})$ has to be expressed via direct sum of representation spaces of irreducible representations of su(2). To this end it suffices to determine the spectrum of the operator

$$\tfrac{1}{2}\tau_J(b_{2-2} - b_{1-1}) \upharpoonright \mathscr{N}(\lambda_s^{(J)}) = \tfrac{1}{2}\widetilde{H} \upharpoonright \mathscr{N}(\lambda_s^{(J)})$$

including multiplicities. This can easily be done and the sought decomposition reads

$$\mathscr{N}(\lambda_s^{(J)}) = \bigoplus_{j=J/2}^{J/2+s} V_j,$$

each V_j being a $(2j+1)$-dimensional space that carries the irreducible representation of su(2) with the highest weight $2j$. Now the weight diagrams of the representations $\tau_{\rm even}$, $\tau_{\rm odd}$, and τ_J, for $J = 1,2,\ldots$, can immediately be determined, and using them we find that these representations appear in the Evans list as ρ_{10}^+, ρ_{20}^+, and $\rho_{J/2+1,J/2}^+$, respectively.

V. FOUR- AND MORE-COMPONENT REPRESENTATIONS

This section deals with the remaining families $\{\Omega_{4m}\}$, $m = 1,2,\ldots$. Let us start with recalling some of their basic properties as given in Ref. 4. The representations in $\{\Omega_{4m}\}$ are labeled by a real ϑ (Ref. 12) taking values in

$$\mathscr{T}_m := ((m-1)/2, +\infty). \tag{5.1}$$

For each $\Omega^{(\vartheta)} \in \{\Omega_{4m}\}$ the odd generators $\widetilde{A}_j(\vartheta) \equiv \widetilde{A}_j := \Omega^{(\vartheta)}(a_j)$, $j = \pm 1, \pm 2$, are expressed via four $4m \times 4m$ matrices A,B,C,D [see Eqs. (2.3) and (2.5)] that depend on ϑ; the even generators

$$\widetilde{B}_{jk}(\vartheta) \equiv \widetilde{B}_{jk} := \Omega^{(\vartheta)}(b_{jk}) = \tfrac{1}{2}\{\widetilde{A}_j, \widetilde{A}_k\}$$

contain only the following quadratic polynomials of A,B,C,D:

$$S{:}BC - AD + 1, \quad T := \tfrac{1}{4}[(BC+1)^2 - 1],$$

$$U := -\tfrac{1}{4}(\{B,D\} + 2AB), \quad V := -\tfrac{1}{4}(\{C,D\} + 2AC).$$

The matrices S,T,U,V are block diagonal and their block structure is determined by four orthogonal projections F_α on $\mathbb{C}^{4m}$ that satisfy $\Sigma_{\alpha=1}^4 F_\alpha = I$; their dimensions $m_\alpha := \dim \operatorname{Ran} F_\alpha$ read

$$m_1 = m+1, \quad m_2 = m-1, \quad m_3 = m_4 = m. \tag{5.2}$$

One has $F_\alpha Z F_\beta = \delta_{\alpha-\beta} Z^\alpha$, where $Z = S,T,U,V$ and $\alpha,\beta = 1,2,3,4$. In particular

$$S^{(\alpha)} = s_\alpha F_\alpha, \tag{5.3a}$$

with

$$s_1 = -s_2 = 2m, \quad s_3 = -s_4 = 4\vartheta, \tag{5.3b}$$

and $U^{(\alpha)}$ has nondegenerated spectrum $\sigma(U^{(\alpha)})$

$= \{2j-1-m_\alpha : j=1,2,...,m_\alpha\}$. Explicit knowledge of the block structure of A will also be needed. It appears that

$$A^{\alpha,\beta} = A^{\alpha+2,\beta+2} = 0, \quad \alpha,\beta = 1,2, \tag{5.4}$$

all the remaining blocks $A^{(\gamma,\delta)}$ being $n\times(n+1)$ or $(n+1)\times n$ matrices of rank $n := \min(m_\gamma, m_\delta)$ such that the following implications hold for the solution of the equation $A^{(\gamma,\delta)}c = 0$:

$$m_\gamma = m_\delta + 1 \Rightarrow c = 0, \tag{5.5a}$$

$$m_\gamma = m_\delta - 1 \Rightarrow \text{unique normalized solution}$$

$$c \in \text{Ran } F_\delta \text{ exists.} \tag{5.5b}$$

Now we are prepared to analyze the necessary conditions according to Proposition 2.1. Let $J = 0,1,...$ and $m = 1,2,...$ be given; we are looking for all $\vartheta \in \mathscr{T}_m$ for which a vector $c \in \mathbb{C}^{4m}$ exists such that $Ac \neq 0$ and the components $c^{(\alpha)} := F_\alpha c$, $1 \leqslant \alpha \leqslant 4$, satisfy

(i) $s_\alpha c^{(\alpha)} = 4(1 - \nu_1 - J/2)c^{(\alpha)}$, for some $\nu_1 \geqslant 0$,

(ii) $s_\alpha (Ac)^{(\alpha)} = (2J+2)(Ac)^{(\alpha)}$,

(iii) $\prod_{l=0}^{J} (U^{(\alpha)} + J - 2l)c^{(\alpha)} = 0.$

The first condition is fulfilled iff $4(1-\nu_1 - J/2)$ is in the spectrum of S. According to Eq. (5.3b) there are four possibilities; let us discuss, e.g., the case

$$4(1-\nu_1 - J/2) = -4\vartheta. \tag{5.6}$$

As ϑ must be positive [see (5.1)], one gets by (i) and (5.3b)

$$c^{(1)} = c^{(3)} = 0.$$

Further, condition (ii) always gives

$$(Ac)^{(4)} = (Ac)^{(2)} = 0.$$

By Eq. (5.4) the first of these equations becomes $A^{(4,2)}c^{(2)} = 0$ (since $c^{(1)} = 0$), which, with the help of (5.2) and (5.5a) yields $c^{(2)} = 0$. On the other hand, the second equation, which can be rewritten as $A^{(2,4)}c^{(4)} = 0$, has by (5.5b) nonzero solution $c^{(4)}$. Now, inversing the implication (5.5a) gives $c^{(4)} \neq 0 \Rightarrow 0 \neq (Ac)^{(1)} = A^{(1,4)}c^{(4)}$, i.e., $Ac \neq 0$, and by using condition (ii) for $\alpha = 1$, one has

$$J = m - 1.$$

Hence in the case (5.6) a nonzero c fulfills the conditions (i), (ii), and $Ac \neq 0$ iff the only nonzero component of c is $c^{(4)}$ and $J = m-1$. Inserting into (5.6) and using (5.1) yields $\nu_1 = 1 + \vartheta - (m-1)/2 > 1$; for the corresponding eigenvalue ν of $\widetilde{N}(\vartheta) := \Omega^{(\vartheta)}(b_{1-1} + b_{2-2})$ one finds by Eqs. (1.8),

$$\nu = 2\nu_1 + J = 2 + 2\vartheta.$$

Finally, as $c^{(1)} = c^{(2)} = c^{(3)} = 0$, condition (iii) becomes

$$\prod_{l=0}^{m-1} (U^{(4)} + m - 1 - 2l)c^{(4)} = 0.$$

However, the numbers $2l + 1 - m$, $l = 0,1,...,m-1$ are just all the eigenvalues of $U^{(4)}$ and thus (iii) is fulfilled identically.

One can proceed similarly in the other cases when the rhs of Eq. (5.6) is replaced by $2m$, $-2m$, and 4ϑ, respectively. Our analysis of existence of HW vectors Ψ_J for the families $\{\Omega_{4m}\}$, $m = 1,2,...$, can then be summarized as follows.

Proposition 5.1: (a) For each $\Omega^{(\vartheta)} \in \{\Omega_{4m}\}$, i.e., $\theta > (m-1)/2$, there exists a HW vector $\Psi_{m-1}^{(\vartheta)} \equiv \Psi_{m-1}$ given by

$$\Psi_{m-1}(\rho,\varphi,z) = (\rho z)^{1/2}\exp(-(\rho^2+z^2)/2) \times z^{m-1}(\rho z \sin\varphi)^{\vartheta-(m-1)/2} \otimes c, \tag{5.7a}$$

where $c^{(1)} = c^{(2)} = c^{(3)} = 0$, and $c^{(4)}$ satisfies $A^{(2,4)}c^{(4)} = 0$. The corresponding eigenvalue of $\widetilde{N}(\vartheta)$ reads

$$\nu = 2 + 2\vartheta. \tag{5.7b}$$

(b) If $m \geqslant 2$, then no further HW vectors exist for any $\Omega^{(\vartheta)} \in \{\Omega_{4m}\}$. For $\Omega^{(\vartheta)} \in \{\Omega_4\}$ there are two more possibilities:

$$\text{(i)}\ \Psi_0^*(\rho,\varphi,z) = (\rho z)^{1/2}\exp\left(-\frac{\rho^2+z^2}{2}\right) \times (\rho z \sin\varphi)^{-\vartheta} \otimes \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad 0 < \vartheta \leqslant 1, \tag{5.8a}$$

with $\widetilde{N}(\vartheta)\Psi_0^* = (2-2\vartheta)\Psi_0^*$; and (5.8b)

$$\text{(ii)}\ \Psi_1(\rho,\varphi,z) = (\rho z)^{1/2}\exp\left(-\frac{\rho^2+z^2}{2}\right) \times (\rho \sin\varphi)^{-1} \otimes \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vartheta = 1, \tag{5.9}$$

with $\widetilde{N}(1)\Psi_1 = \Psi_1$.

The following simple argument enables us to reduce the case (5.8) and exclude (5.9). Suppose that for given $\Omega^{(\vartheta)}$ a HW vector Ψ_J is known and construct the subspace $\mathscr{D}_J$ according to Eq. (1.7). If a scalar product on $\mathscr{D}_J$ exists such that $\Omega^{(\vartheta)} \upharpoonright \mathscr{D}_J$ satisfies (1.3), then for any $\widetilde{T} \in \mathscr{U}(\widetilde{A}_{\pm 1}(\vartheta), \widetilde{A}_{\pm 2}(\vartheta))$ one has, for the norm of $\Psi := \widetilde{T}\Psi_J$,

$$\|\Psi\|^2 = (\Psi_J, \widetilde{T}^\# \widetilde{T}\Psi_J). \tag{5.10}$$

As $\widetilde{T}^\# \widetilde{T}\Psi_J = \Sigma_k \widetilde{M}_k \Psi_k$, where each $\widetilde{M}_k$ is a monomial in $\widetilde{A}_r^\# \equiv \widetilde{A}_{-r}$, $r = 1,2$ and $\Psi_k \in \mathscr{D}_J^{\text{vac}} \equiv \mathscr{D}_J \cap \mathscr{D}_{\text{vac}}$,[3] one finds by Eqs. (1.1), (1.6), and (1.8) that $(\Psi_J, \widetilde{T}^\# \widetilde{T}\Psi_J) = c(\vartheta,J)\|\Psi_J\|^2$. The necessary condition (5.10) is not fulfilled if $c(\vartheta,J) < 0$. For $c(\vartheta,J) = 0$ this condition is violated if $\widetilde{T}\Psi_J \neq 0$.

Consider first the case (5.9). Equations (1.8) yield, for $r = 1,2$,

$$\widetilde{N}_r(\vartheta)\Psi_1 \equiv \widetilde{N}_r\Psi_1 = \tfrac{1}{2}\widetilde{A}_r\widetilde{A}_r^\#\Psi_1 = \nu_r\Psi_1, \quad \nu_1 := 0, \quad \nu_2 := 1.$$

Further, $\widetilde{F}\Psi_1 \equiv \tfrac{1}{2}\{\widetilde{A}_1, \widetilde{A}_2^\#\}\Psi_1 = 0$ and $\widetilde{F}\widetilde{E}\Psi_1 = \Psi_1$, where $\widetilde{E} \equiv \tfrac{1}{2}\{\widetilde{A}_2, \widetilde{A}_1^\#\}$ [see (1.5) and (1.6)]. Then the relations (1.1) yield, for $\widetilde{T} = \widetilde{A}_2^\#\widetilde{A}_1^\#$,

$$\tilde{T}^{\#}\tilde{T}\Psi_1 = \tilde{A}_1(2\tilde{N}_2 - \tilde{A}_2^{\#}\tilde{A}_2)\tilde{A}_1^{\#}\Psi_1$$
$$= 4\tilde{N}_2\tilde{N}_1\Psi_1 - \tilde{A}_1\tilde{A}_2^{\#}(2\tilde{E} - \tilde{A}_1^{\#}\tilde{A}_2)\Psi_1$$
$$= -(2\tilde{F} - \tilde{A}_2^{\#}\tilde{A}_1)2\tilde{E}\Psi_1$$
$$= -4\Psi_1 + 2\tilde{A}_2^{\#}(\tilde{E}\tilde{A}_1 + \tilde{A}_1)\Psi_1 = -4\Psi_1.$$

The necessary condition (5.10) is thus violated; consequently, the case (5.9) cannot yield any *-representation.

Let us pass to the case (5.8), for which one has $\tilde{N}_1(\vartheta)\Psi_0^* = \tilde{N}_2(\vartheta)\Psi_0^* = (1-\vartheta)\Psi_0^*$, and let $\tilde{T} = \frac{1}{2}[\tilde{A}_1^{\#}(\vartheta),\tilde{A}_2^{\#}(\vartheta)]$. Proceeding as above, we find that the rhs of (5.10) now equals $2(1-\vartheta)(1-2\vartheta)\|\Psi_0^*\|^2$; hence the case (5.8) yields no *-representation for $\vartheta\in(\frac{1}{2},1]$. Finally, if $\vartheta = \frac{1}{2}$, a direct calculation using explicit formulas for $\tilde{A}_r^{\#}(\frac{1}{2})$ [see Eqs. (2.3), (2.5), and (5.11) below] gives

$$(\tilde{T}\Psi_0^*)(\rho,\varphi,z) = 4i\rho z \exp\left(-\frac{\rho^2+z^2}{2}\right)(\sin\varphi)^{1/2}\otimes\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

so that no *-representation can be obtained for $\vartheta = \frac{1}{2}$, either.

By adding these results to Proposition 5.1, we arrive at the following conclusion [for a given $\Omega^{(\vartheta)}\in\{\Omega_{4m}\}$ we set $\mathscr{U}_m^{(\vartheta)}(A) \equiv \mathscr{U}(\tilde{A}_{\pm 1}(\vartheta),\tilde{A}_{\pm 2}(\vartheta))$].

Theorem 5.2: If for some $\Omega^{(\vartheta)}\in\{\Omega_{4m}\}$, $m = 1,2,\ldots$, a subspace $\mathscr{D}$ of $4m$-component vector functions belonging to $C^\infty(\mathbb{R}^+\times(0,\pi)\times\mathbb{R}^+)$ exists such that $\pi\equiv\Omega^{(\vartheta)}\restriction\mathscr{D}$ is an irreducible representation of $B(0,2)$ with finite-degenerated vacuum and satisfies the *-condition (1.3) for some scalar product on $\mathscr{D}$, then π equals to one of the following:

(a) $\tau_J^{(\mu)}\equiv\Omega^{(\vartheta)}\restriction\mathscr{U}_{J+1}^{(\vartheta)}(A)\Psi_J$,

$J = 0,1,\ldots,\vartheta > J/2$,

with the HW vector Ψ_J given by (5.7a) and $(J+1)$dimensional vacuum subspace;

(b) $\rho^{(\vartheta)}\equiv\Omega^{(\vartheta)}\restriction\mathscr{U}_1^{(\vartheta)}(A)\Psi_0^*$, $0<\vartheta<\frac{1}{2}$,

with the HW vector (5.8a) and nondegenerated vacuum.

Remark 5.3: The representations listed in (a) are exactly all the massive representations of Ref. 5, the parameters E_0 and j being related to our ϑ and J by $E_0 = 1+\vartheta$ and $j = J/2$.

The last problem to be solved is conversing the preceding theorem and reducing the representations $\pi_J^{(\vartheta)}$ and $\rho^{(\vartheta)}$ with respect to the sublagebra so(3,2). Only the representations $\pi_0^{(\vartheta)}$ will be considered; the remaining cases can be treated similarly, however, detailed calculations have not been finished yet. In the rest of this section it is thus assumed $J = 0$, i.e., $m = 1$, and $\vartheta > 0$.

It turns out to be convenient to transform the 4×4 matrices $A,B,\ldots$, which are explicitly given in Ref. 4, by a unitary matrix R such that $T' := RTR^{-1}$ becomes diagonal. The result is expressed via Pauli matrices as follows:

$$A' = \sigma_1\otimes\sigma_3,\quad B' = i\sigma_1\otimes\sigma_2,$$
$$C' = 2\vartheta\sigma_1\otimes\sigma_1,\quad D' = i\sigma_2\otimes(\sigma_3 - 2\vartheta),$$
$$S' = (1+\sigma_3)\otimes I + 2\vartheta(1-\sigma_3)\otimes\sigma_3, \tag{5.11}$$
$$T' = \vartheta I\otimes(\vartheta+\sigma_3),$$
$$U' = -\tfrac{1}{2}(1+\sigma_3)\otimes\sigma_1,\quad V' = -i\vartheta(1+\sigma_3)\otimes\sigma_3.$$

The matrices S', T', U', V' are block diagonal, the dimensions of the blocks being 2,1,1 [see (5.2)]. The operators we obtain by inserting the matrices (5.11) into Eqs. (2.3) and (2.5) will be denoted $\tilde{A}_r(\vartheta)\equiv\tilde{A}_r$ and $\tilde{A}_r^{\#}$, $r = 1,2$. Similarly,

$$\tilde{B}_{jk}(\vartheta)\equiv\tilde{B}_{jk} = \tfrac{1}{2}\{\tilde{A}_j,\tilde{A}_k\},\quad j,k = \pm 1, \pm 2,$$

where $\tilde{A}_{-r} = \tilde{A}_r^{\#}$ for $r = 1,2$. In particular, for the operators $\tilde{N}_r\equiv\tilde{B}_{r,-r}$ Eq. (2.6) yields

$$\tilde{N}_1 = \tfrac{1}{2}[\rho^2 - \partial_\rho^2 + \rho^{-2}\Theta_1],$$
$$\tilde{N}_2 = \tfrac{1}{2}[z^2 - \partial_z^2 + z^{-2}\Theta_2], \tag{5.12a}$$

with

$$\Theta_1 := -\partial_\varphi^2 - \tfrac{1}{4} + (\sin\varphi)^{-2}\otimes T',$$
$$\Theta_2 := \Theta_1 - 2i\,\partial_\varphi\otimes U' + i\cot\varphi\otimes V' - 2T' + \tfrac{1}{2}S' + 2\vartheta^2. \tag{5.12b}$$

Our next goal is finding common eigenfunctions of operators Θ_1 and Θ_2 on the space $C_4^\infty(0,\pi)$ of four-component vector functions that are C^∞ on $(0,\pi)$. Consider for $m = 0,1,\ldots$ and $1\leqslant n\leqslant 4$ the following elements of $C_4^\infty(0,\pi)$:

$$|m,n\rangle_\vartheta := \begin{cases} v_m^{(\vartheta+1/2)}\otimes e_n, & n = 1,3,\\ v_m^{(\vartheta-1/2)}\otimes e_n, & n = 2,4.\end{cases} \tag{5.13a}$$

Here the vectors $e_n\in\mathbb{C}^4$ are related to the basis (3.5b) by

$$e_1 := |+\rangle\otimes|+\rangle,\quad e_2 := |+\rangle\otimes|-\rangle,$$
$$e_3 := |-\rangle\otimes|+\rangle,\quad e_4 := |-\rangle\otimes|-\rangle, \tag{5.13b}$$

and the functions $v_m^{(\beta)}\in C^\infty(0,\pi)$ are defined for $\beta > -1$ via the spherical functions $\mathbb{P}_\nu^{(\mu)}$ or Jacobi polynomials $P_n^{(\alpha,\beta)}$ (see Ref. 8):

$$(-i)^m v_m^{(\beta)}(\varphi) := \left(\frac{\sin\varphi}{2}\right)^{1/2}\gamma_m^{(\beta)}\mathbb{P}_{\beta+m}^{(-\beta)}(\cos\varphi)$$
$$= \left(\frac{\sin\varphi}{2}\right)^{\beta+1/2}\frac{\gamma_m^{(\beta)}m!}{\Gamma(\beta+m+1)}$$
$$\times P_m^{(\beta,\beta)}(\cos\varphi), \tag{5.13c}$$

where

$$\gamma_m^{(\beta)} := \left[\frac{(2\beta+2m+1)\Gamma(2\beta+m+2)}{(2\beta+m+1)m!}\right]^{1/2}.$$

One has for each $m = 0,1,\ldots$ and $\beta > -1$,

$$[-d_\varphi^2 + (\beta^2 - \tfrac{1}{4})/(\sin\varphi)^2]v_m^{(\beta)} = (m+\beta+\tfrac{1}{2})^2 v_m^{(\beta)}$$

and

$$T'e_n = [(\vartheta - (-1)^n/2)^2 - \tfrac{1}{4}]e_n,\quad 1\leqslant n\leqslant 4.$$

Then the first of Eqs. (5.12b) gives, for $m = 0,1,\ldots$ and $1\leqslant n\leqslant 4$,

$$\Theta_1|m,n\rangle_\vartheta = [(m+\vartheta-(-1)^n/2+\tfrac{1}{2})^2 - \tfrac{1}{4}]|m,n\rangle_\vartheta. \tag{5.14}$$

By using Eq. (5.11), we see that this relation holds also for Θ_2 if $n = 3,4$, whereas for $n = 1,2$, functional relations satisfied by the functions $v_m^{(\beta)}$ (see Appendix B) yield

$$\Theta_2|m,1\rangle_\vartheta = [(m+\vartheta+1)^2 + \tfrac{3}{4} - 2\vartheta]|m,1\rangle_\vartheta$$
$$+ 2[(m+1)(2\vartheta+m+1)]^{1/2}|m+1,2\rangle_\vartheta,$$

$$\Theta_2|m+1,2\rangle_\vartheta = 2[(m+1)(2\vartheta+m+1)]^{1/2}|m,1\rangle_\vartheta + [(m+\vartheta+1)^2+\tfrac{3}{4}+2\vartheta]|m+1,2\rangle_\vartheta.$$

Thus the subspace $\{|m,1\rangle_\vartheta, |m+1,2\rangle_\vartheta\}_{\text{lin}}$ is invariant under Θ_2; at the same time this is an eigenspace of Θ_1 [for the eigenvalue $[(m+\vartheta+1)^2-\frac{1}{4}]$, see (5.14)]. By taking proper linear combinations of $|m,1\rangle_\vartheta$ and $|m+1,2\rangle_\vartheta$, we find that besides $|m,3\rangle_\vartheta$ and $|m,4\rangle_\vartheta$, $m=0,1,\ldots$, there are further common eigenfunctions of Θ_1 and Θ_2:

$$|m,+\rangle_\vartheta := \left(\frac{m}{2\vartheta+2m}\right)^{1/2}|m-1,1\rangle_\vartheta + \left(\frac{2\vartheta+m}{2\vartheta+2m}\right)^{1/2}|m,2\rangle_\vartheta,$$

$$|m,-\rangle_\vartheta := \left(\frac{2\vartheta+m+1}{2(\vartheta+m+1)}\right)^{1/2}|m,1\rangle_\vartheta + \left(\frac{m+1}{2(\vartheta+m+1)}\right)^{1/2}|m+1,2\rangle_\vartheta,$$

$$m=0,1,\ldots .$$

The corresponding eigenvalues of Θ_r, $r=1,2$, are given by

$$\lambda_1(m,\mu) = (m+\vartheta+\delta_{\mu+1})^2-\tfrac{1}{4},$$
$$\lambda_2(m,\mu) = (m+\vartheta+\delta_{\mu-1})^2-\tfrac{1}{4}, \quad \mu=\pm 1,$$
$$\lambda_r(m,n) = (m+\vartheta+4-n)^2-\tfrac{1}{4},$$
$$\text{for } r=1,2, \quad n=3,4.$$

Now one finds by Eqs. (3.8b) and (5.12a) that for $k,l,m=0,1,\ldots$ the vector functions

$$|klm;\mu\rangle_\vartheta(\rho,\varphi,z) := f_k^{(m+\vartheta+\delta_{\mu+1})}(\rho) f_l^{(m+\vartheta+\delta_{\mu-1})}(z) \otimes |m,\mu\rangle_\vartheta(\varphi), \quad \mu=\pm 1,$$

$$|klm;n\rangle_\vartheta(\rho,\varphi,z) := f_k^{(m+\vartheta+4-n)}(\rho) f_l^{(m+\vartheta+4-n)}(z) \otimes |m,n\rangle_\vartheta(\varphi), \quad n=3,4, \tag{5.15a}$$

satisfy

$$[\widetilde{N}_r - (2j_r+m+\vartheta+3\delta_{\mu+1}+\mu r)]|klm;\mu\rangle_\vartheta = 0,$$
$$[\widetilde{N}_r - (2j_r+m+\vartheta+5-n)]|klm;n\rangle_\vartheta = 0, \tag{5.15b}$$

where $r=1,2$ and $j_1:=k$, $j_2:=l$. Notice that the HW vector (5.7a) can be written as follows:

$$\Psi_0 = 2^{-\vartheta-1}(\pi\Gamma(2\vartheta+1))^{1/2}|000;4\rangle_\vartheta. \tag{5.16}$$

The action of the operators $\widetilde{A}_r(\vartheta)$ and $\widetilde{A}_r^{\#}(\vartheta)$, $r=1,2$, on the vector functions (5.15a) can be found by direct calculation. The resulting formulas given in Appendix B show that the subspace

$$\mathscr{D}_\vartheta = (\mathscr{E}_\vartheta)_{\text{lin}},$$
$$\mathscr{E}_\vartheta := \{|klm;\mu\rangle_\vartheta, |klm;n\rangle_\vartheta : k,l,m=0,1,\ldots, \mu=\pm 1, n=3,4\} \tag{5.17}$$

is invariant under all of them and, as $\Psi_0\in\mathscr{D}_\vartheta$, the domain of $\pi_0^{(\vartheta)}$ fulfills

$$\mathscr{U}_1^{(\vartheta)}(A)\Psi_0 \subset \mathscr{D}_\vartheta, \quad \vartheta>0. \tag{5.18}$$

Moreover, each of the vector functions (5.15a) is in

$$L_4^2(\widetilde{M}) := L^2(\mathbb{R}^+\times(0,\pi)\times\mathbb{R}^+)\otimes\mathbb{C}^4.$$

This can be checked with the help of Eqs. (3.8a) and (5.13c) showing that

$$v_m^{(\beta)}(\varphi) \sim (\sin\varphi)^{\beta+1/2}P_m(\cos\varphi),$$

where P_m is a polynomial of degree m; hence $v_m^{(\beta)}\in L^2(0,\pi)$ for any $\beta>-1$. In fact, the sets $\{v_m^{(\beta)}: m=0,1,\ldots\}$ and $\{f_k^{(\alpha)}: k=0,1,\ldots\}$ are orthonormal bases in $L^2(0,\pi)$ and $L^2(\mathbb{R}^+)$, respectively, for all $\alpha,\beta>-1$ (see Ref. 8, §8.904); hence $\mathscr{E}_\vartheta$ is an orthonormal basis in $L_4^2(\widetilde{M})$. By using this fact and the explicit formulas of Appendix B, one easily verifies that $\Omega^{(\vartheta)}\upharpoonright\mathscr{D}_\vartheta$ fulfills the *-condition (1.3) for the $L_4^2(\widetilde{M})$ scalar product. The formulas of Appendix B also show that for any $\Psi\in\mathscr{D}_\vartheta$ such that $\widetilde{A}_r\Psi=0$, $r=1,2$, one has $\Psi\sim|000;4\rangle_\vartheta$ (cf. Lemma 3.2), i.e., the representation $\Omega^{(\vartheta)}\upharpoonright\mathscr{D}_\vartheta$ has a nondegenerated vacuum.

Let us finally examine questions of irreducibility [notice that if one proves algebraic irreducibility of $\Omega^{(\vartheta)}\upharpoonright\mathscr{D}_\vartheta$, then equality will hold in (5.18), i.e., $\Omega^{(\vartheta)}\upharpoonright\mathscr{D}_\vartheta = \pi_0^{(\vartheta)}$]. Consider the subspaces $\mathscr{D}_\vartheta^{(\alpha)}\subset\mathscr{D}_\vartheta$, $\alpha=1,3,4$, where

$$\mathscr{D}_\vartheta^{(1)} := \{|klm;\mu\rangle_\vartheta : k,l,m=0,1,\ldots, \mu=\pm 1\}_{\text{lin}},$$
$$\mathscr{D}_\vartheta^{(\alpha)} := \{|klm;\alpha\rangle_\vartheta : k,l,m=0,1,\ldots\}_{\text{lin}}, \quad \alpha=3,4.$$

As the matrices S', T', U', V' leave invariant the corresponding subspaces in $\mathbb{C}^4$, each of $\mathscr{D}_\vartheta^{(\alpha)}$ is invariant under all the operators $\widetilde{B}_{jk}(\vartheta)$, $j,k=\pm 1,\pm 2$; the representations $x\mapsto\Omega^{(\vartheta)}(x)\upharpoonright\mathscr{D}_\vartheta^{(\alpha)}$, $x\in\text{so}(3,2)$, will be denoted $\tau_\vartheta^{(\alpha)}$.

Proposition 5.4: The representations $\tau_\vartheta^{(\alpha)}$, $\alpha=1,3,4$, are algebraically irreducible.

Proof: Let $|000;1\rangle_\vartheta := |000;+\rangle_\vartheta$ and

$$\mathscr{U}_1^{(\vartheta)}(B) := \mathscr{U}(\widetilde{B}_{jk}(\vartheta), \; j,k=\pm 1,\pm 2)$$

[cf. Eq. (1.7)]. It suffices to find to each nonzero $\Psi\in\mathscr{D}_\vartheta^{(\alpha)}$ two operators $\widetilde{S}_\alpha$ and $\widetilde{T}_\alpha\in\mathscr{U}_1^{(\vartheta)}(B)$ such that

$$\Psi = \widetilde{S}_\alpha|000;\alpha\rangle_\vartheta, \quad T_\alpha\Psi = |000;\alpha\rangle_\vartheta. \tag{5.19}$$

This can be done with the help of Appendix B [cf. the proof of assertions (i) and (ii) in Appendix A]. ■

The relations (5.19) also cover the essential part of proving irreducibility of $\Omega^{(\vartheta)}\upharpoonright\mathscr{D}_\vartheta$, if one can show that the projections $\widetilde{P}_\alpha$ onto the subspaces $\mathscr{D}_\vartheta^{(\alpha)}$ satisfy

$$\widetilde{P}_\alpha\in\mathscr{U}_1^{(\vartheta)}(A), \quad \alpha=1,3,4. \tag{5.20}$$

To this end we make use of the following property of the second-order Casimir operator $\Omega^{(\vartheta)}(c_2)$ of so(3,2) that obviously belongs to $\mathscr{U}_1^{(\vartheta)}(B)$ (cf. Lemma 3.4):

$$\Omega^{(\vartheta)}(c_2) = I\otimes(S'+\varkappa-4) = I\otimes(S'+4\vartheta^2-8).$$

(See Ref. 12.) One has $\widetilde{P}_\alpha = I\otimes F_\alpha$, where F_1 projects onto the subspace of $\mathbb{C}^4$ spanned by e_1, e_2 and F_3, F_4 onto subspaces $\{e_3\}_{\text{lin}}$, $\{e_4\}_{\text{lin}}$, respectively. Now Eqs. (5.3) yield $S' = 2F_1 + 4\vartheta(F_3-F_4)$ and, due to this relation, each F_α can be expressed as a polynomial function of S' provided that the eigenvalues of S' are nondegenerated, i.e., $\vartheta\neq\frac{1}{2}$. Then $\widetilde{P}_\alpha$ equals the same polynomial function of $\Omega^{(\vartheta)}(c_2)-4\vartheta^2-8$ so that $\widetilde{P}_\alpha\in\mathscr{U}_1^{(\vartheta)}(B)\subset\mathscr{U}_1^{(\vartheta)}(A)$. However, for $\vartheta=\frac{1}{2}$ we only find that $\widetilde{P}_4$ and $\widetilde{P}_1+\widetilde{P}_3$ are in $\mathscr{U}_1^{(\vartheta)}(B)$. Unfortunately, the fourth-order Casimir operator of so(3,2) does not help since its eigenvalues corresponding to $\widetilde{P}_1$ and $\widetilde{P}_3$ also coincide for $\vartheta=\frac{1}{2}$.

Let us conclude this section by summarizing basic properties of representations $\pi_0^{(\vartheta)} \equiv \Omega^{(\vartheta)} \upharpoonright \mathscr{U}_1^{(\vartheta)}(A)$ for $\vartheta > 0$.

Theorem 5.5: (a) Each $\pi_0^{(\vartheta)}$ is a *-representation of $B(0,2)$ on $L^2(\mathbb{R}^+ \times (0,\pi) \times \mathbb{R}^+) \otimes \mathbb{C}^4$ and its domain equals $\mathscr{D}_\vartheta$ given by Eq. (5.17), the corresponding grading of End $\mathscr{D}_\vartheta$ (see Ref. 1) being determined by the projection $\widetilde{P}_1$ onto $\mathscr{D}_\vartheta^{(1)}$.

(b) The vacuum subspace of $\pi_0^{(\vartheta)}$ is one dimensional, spanned by the vector function

$$\Psi_0(\rho,\varphi,z) = (\rho z)^{\vartheta + 1/2} \times \exp(-(\rho^2 + z^2)/2)(\sin\varphi)^\vartheta \otimes e_4.$$

(c) Each $\pi_0^{(\vartheta)}$ is algebraically irreducible and with respect to so(3,2) reduces into three algebraically irreducible skew-symmetric representations.

(d) The representations $\pi_0^{(\vartheta)}$ and $\pi_0^{(\vartheta')}$ are nonequivalent if $\vartheta \neq \vartheta'$.

Proof: Only irreducibility of $\pi_0^{(\vartheta)}$ and assertion (d) have not been proved yet. Irreducibility can be verified as in Proposition 5.4: to each $\Psi \in \mathscr{D}_\vartheta$ one finds $\widetilde{S}$ and $\widetilde{T} \in \mathscr{U}_1^{(\vartheta)}(A)$ such that $\Psi = \widetilde{S}|000;4\rangle_\vartheta$ and $\widetilde{T}\Psi = |000;4\rangle_\vartheta$. This can be achieved by using Eqs. (5.19) and (5.20) and (see Appendix B)[13]

$$\widetilde{A}_2^{\#}|000;4\rangle_\vartheta = -\eta(2\vartheta+2)^{1/2}|000;+\rangle_\vartheta,$$

$$\widetilde{A}_2|000;+\rangle_\vartheta = -\bar{\eta}(2\vartheta+2)^{1/2}|000;4\rangle_\vartheta,$$

$$[\widetilde{A}_1^{\#},\widetilde{A}_2^{\#}]|000;4\rangle_\vartheta = 2[(2\vartheta+2)(2\vartheta+1)]^{1/2}|000;3\rangle_\vartheta,$$

$$\widetilde{A}_2\widetilde{A}_1|000;3\rangle_\vartheta = -[(2\vartheta+2)(2\vartheta+1)]^{1/2}|000;4\rangle_\vartheta,$$

$$\eta := \exp(i\pi/4).$$

Finally, assertion (d) follows from Eq. (5.7b), which gives the lowest eigenvalue of $\widetilde{N}(\vartheta) \equiv \widetilde{B}_{1-1}(\vartheta) + \widetilde{B}_{2-2}(\vartheta)$.[3] ∎

VI. CONCLUDING REMARKS

To our knowledge, infinite-dimensional representations of osp(1,4) have been treated in Refs. 5 and 14. In the former work irreducible *-representations have been classified according to how they reduce with respect to so(3,2), and divided into four classes. The set $\{\pi_J: J = 1,2,...\}$ is identical with the third class, whereas the representation π_0 belongs to the second class, which is labeled by a continuous parameter $E_0 > \frac{1}{2}$; π_0 corresponds to $E_0 = 1$. The whole set $\{\pi_J: J = 0,1,...\}$ just covers all the massless representations; the representation space of each of them are two-component vector functions. Similarly, the four-component representations $\pi_0^{(\vartheta)}$, $\vartheta > 0$, of Sec. V are just all the massive representations in the second class. Finally, the fourth class is completely covered by representations $\{\pi_J^{(\vartheta)}: \vartheta > J/2\}$, $J = 1,2,...$, that can be obtained from the families $\{\Omega_{4(J+1)}\}$ using the HW vector (5.7a). Construction of these representations is in progress.

In Ref. 5 explicit form of the representations under consideration is not given. Progress in this direction has been attained in Ref. 14, where special attention has been given to representations with nondegenerated vacuum characterized by the order of parastatistics $p > 0$. The authors of Ref. 14 have constructed all the representations belonging to the first and second Heidenreich's class by giving explicit formulas for matrix elements of the odd generators in a concrete basis. Their case I is for $p = 1$ equivalent to our π_0 and for $p > 2$ to $\pi_0^{(2p-4)}$; further the case III, which contains representations with degenerated vacuum labeled by $q = 1,\frac{3}{2},2,...$, is equivalent to π_{2q} (a counterpart to π_1, the first member of Heidenreich's third class, is missing). The basis used in Ref. 14 is related to the reduction of so(3,2) with respect to so(2,1) ⊕ so(2,1) and is quite different from the bases (3.10) and (5.17). The resulting formulas for the odd generators are much simpler in the latter bases; moreover, we give basis-independent expressions [see Eqs. (2.3), (2.5), (3.1), and (5.11)].

ACKNOWLEDGMENTS

A part of this work was done during our stay at the International Centre for Theoretical Physics in Trieste. We would like to thank Professor Abdus Salam, The International Atomic Energy Agency, and UNESCO for hospitality at the ICTP.

APPENDIX A: ALGEBRAIC IRREDUCIBILITY OF REPRESENTATIONS π_J

In order to complete the proof of algebraic irreducibility sketched in Sec. III it remains to check the following two assertions (the notation of Sec. III is used, however we omit writing down dependence of $\widetilde{A}_r$, $\widetilde{A}_r$ and $|klm;\mu\rangle$ on J).

(i) For each $|klm;\mu\rangle \in \mathscr{E}_J$ there exists $\widetilde{T} \in \mathscr{U}_J$ such that $|klm;\mu\rangle = \widetilde{T}|000;+\rangle$.

(ii) For $\mu = \pm 1$ and any linear combination

$$\Psi_\mu = \sum_{klm} c_{klm}|klm;\mu\rangle \neq 0 \tag{A1}$$

there exists $\widetilde{S} \in \mathscr{U}_J$ such that $\widetilde{S}\Psi_\mu$ is a nonzero vector of $\mathscr{D}_J^{\text{vac}}$, i.e., it equals a linear combination of $|00m;+\rangle$, where $-J \leqslant m \leqslant 0$.

Explicit knowledge of the action on any $|klm;\mu\rangle$ of $\widetilde{A}_r^2$ and $\widetilde{A}_r^{\#2}$ will be needed. With the help of Eqs. (3.12) one finds

$$-(i/2)\widetilde{A}_1^2|klm;\mu\rangle = [k(k+|m|)]^{1/2}|k-1\,lm;\mu\rangle, \tag{A2}$$

$$-(i/2)\widetilde{A}_2^2|klm;\mu\rangle = [l(l+|\mu(J+m+\tfrac{1}{2})-\tfrac{1}{2}|)]^{1/2}|kl-1\,m;\mu\rangle, \tag{A3}$$

$$(i/2)\widetilde{A}_1^{\#2}|klm;\mu\rangle = [(k+1)(k+1+|m|)]^{1/2}|k+1\,lm;\mu\rangle, \tag{A4}$$

$$(i/2)\widetilde{A}_r^{\#2}|klm;\mu\rangle = [(l+1)(l+1+|\mu(J+m+\tfrac{1}{2})-\tfrac{1}{2}|)]^{1/2} \times |kl+1\,m;\mu\rangle. \tag{A5}$$

Now we get, by (A4) and (A5),

$$\widetilde{A}_1^{\#2k}\widetilde{A}_2^{\#2l}|00m;\mu\rangle = \alpha(k,l,m,\mu)|klm;\mu\rangle,$$

with some nonzero $\alpha(k,l,m,\mu)$. Hence (i) will hold if we find, for each $\mu = \pm 1$ and $m = 0, \pm 1,...$, an operator $\widetilde{T}_m^{(\mu)} \in \mathscr{U}_J$ such that

$$|00m;\mu\rangle = \widetilde{T}_m^{(\mu)}|000;+\rangle.$$

Let us consider first the case $-J \leqslant m \leqslant 0$. One has

$$|00m;-\rangle = \eta[2(J+m+1)]^{-1/2}\widetilde{A}_2^{\#}|00m;+\rangle \tag{A6}$$

and then $\widetilde{T}_m^{(+)}$ can be found using Proposition 3.3(b) since $|00m;+\rangle \in \mathscr{D}_J^{\rm vac}$. If $m \leqslant -J-1$, we use

$$|00m;-\rangle = -\bar{\eta}(2|J+m|)^{-1/2}\widetilde{A}_2|00m;+\rangle$$

and further

$$|00m;+\rangle = -(i/2)(|m|\,|m+J|)^{-1/2}\widetilde{A}_2^{\#}\widetilde{A}_1^{\#}|00m+1;+\rangle.$$

With the help of these relations we find that $\widetilde{T}_m^{(+)}$ is proportional to $(\widetilde{A}_2^{\#}\widetilde{A}_1^{\#})^{|J+m|}\widetilde{T}_{-J}^{(+)}$ and $\widetilde{T}_m^{(-)} = -\bar{\eta}(2|j+m|)^{-1/2}\widetilde{A}_2\widetilde{T}_m^{(+)}$. Finally, for $m>0$ we use m times

$$|00m;-\rangle = -(i/2)(m(J+m+1))^{-1/2}\widetilde{A}_2^{\#}\widetilde{A}_1^{\#}|00m-1;-\rangle$$

and then (A6) with $m=0$, which gives $\widetilde{T}_m^{(-)} \sim (\widetilde{A}_2^{\#}\widetilde{A}_1^{\#})^m\widetilde{A}_2^{\#}$. As for $\widetilde{T}_m^{(+)}$, Eq. (3.12c) shows that $\widetilde{A}_2$ transforms $|00m;-\rangle$ into $|00m;+\rangle$, therefore $\widetilde{T}_m^{(+)} = \widetilde{A}_2\widetilde{T}_m^{(-)}$.

For proving (ii) let us denote by $\bar{k}$ the largest k in (A1) and by $\bar{l}$ the largest l for which $c_{\bar{k}lm} \neq 0$. Further let m_1 (m_2) be the minimum (maximum) of the set $\{m: c_{\bar{k}\bar{l}m} \neq 0\}$. Then, by using Eqs. (A2) and (A3) and setting $\widetilde{S}_{\bar{k}\bar{l}} := \widetilde{A}_2^{2\bar{l}}\widetilde{A}_1^{2\bar{k}}$, one gets

$$\Psi_\mu^{(0)} := \widetilde{S}_{\bar{k}\bar{l}}\Psi_\mu = \sum_{m=m_1}^{m_2} c_m|00m;\mu\rangle, \tag{A7}$$

the coefficients c_m being proportional to $c_{\bar{k}\bar{l}m}$. Let us consider first the case $\mu = +1$.

(a) If $m_2 > 0$, then the relation

$$-\frac{i}{2}\widetilde{A}_2\widetilde{A}_1|00m;+\rangle = \begin{cases} -(m(J+m))^{1/2}|00m-1;+\rangle, & m \geqslant 1,\\ 0, & m \leqslant 0,\end{cases}$$

yields $(-(i/2)\widetilde{A}_2\widetilde{A}_1)^{m_2}\Psi_+^{(0)} = (-1)^{m_2}(m_2!(J+m_2)!/J!)^{1/2}c_{m_2}|00m;+\rangle$, i.e., one can choose $\widetilde{S} = (\widetilde{A}_2\widetilde{A}_1)^{m_2}\widetilde{S}_{\bar{k}\bar{l}}$.

(b) For $m_2 \leqslant 0$ and $m_1 \geqslant -J$ we see that $\Psi_+^{(0)} \in \mathscr{D}_J^{\rm vac}$ and thus only the case $m_1 \leqslant -J-1$ remains. By using

$$-\frac{i}{2}\widetilde{A}_1\widetilde{A}_2|00m;+\rangle = \begin{cases} -(|(J+m||m|)^{1/2}|00m+1;+\rangle, & m \leqslant -J-1,\\ 0, & m \geqslant -J,\end{cases}$$

one sees that $(A_1A_2)^{|J+m_1|}\Psi_+^{(0)}$ is proportional to $|00-J;+\rangle \in \mathscr{D}_J^{\rm vac}$, and thus we set $\widetilde{S} = (\widetilde{A}_1\widetilde{A}_2)^{|J+m_1|}\widetilde{S}_{\bar{k}\bar{l}}$.

Let us now suppose $\mu = -1$ in (A7). If $\widetilde{A}_r\Psi_-^{(0)} = 0$ for both $r = 1,2$, then $\Psi_-^{(0)} \in \mathscr{D}_J^{\rm vac}$, whence $\widetilde{S} = \widetilde{S}_{\bar{k}\bar{l}}$. If $\widetilde{A}_1\Psi_-^{(0)} \neq 0$, we use

$$\widetilde{A}_1\Psi_-^{(0)} = \sum_{m=m_1}^{\min(-1,m_2)} c_m|m|^{1/2}|00m+1;+\rangle,$$

and $\widetilde{S}$ is obtained by applying (b) to $\widetilde{A}_1\Psi_-^{(0)}$. Finally, if $\widetilde{A}_2\Psi_-^{(0)} \neq 0$, then one has

$$\widetilde{A}_2\Psi_-^{(0)} = \sum_{m=\max(-J,m_1)}^{m_2} c_m(J+m+1)^{1/2}|00m+1;+\rangle.$$

Now $\widetilde{S} = \widetilde{A}_2\widetilde{S}_{\bar{k}\bar{l}}$ for $m_2 \leqslant 0$; otherwise one gets $\widetilde{S}$ by applying (a) to $\widetilde{A}_2\Psi_-^{(0)}$.

APPENDIX B: EXPLICIT FORMULAS FOR REPRESENTATIONS $\pi_0^{(\vartheta)}$

The action of operators $\widetilde{A}_j \equiv \pi_0^{(\vartheta)}(a_j)$ and $\widetilde{B}_{jk} \equiv \pi_0^{(\vartheta)}(b_{jk})$, $j,k = \pm1, \pm2$, on the vector functions (5.15a) has been found by using Eqs. (3.9) and the following relations for the functions $v_m^{(\beta)}$ (see Ref. 8, §§8.733 and 8.735):

$$i(2\beta+2m+1)^{1/2}\cos\varphi\, v_m^{(\beta)}(\varphi) = \left[\frac{(m+1)(2\beta+m+1)}{2\beta+2m+3}\right]^{1/2} v_{m+1}^{(\beta)}(\varphi) - \left[\frac{m(2\beta+m)}{2\beta+2m-1}\right]^{1/2} v_{m-1}^{(\beta)}(\varphi),$$

$$(2\beta+2m+1)^{1/2}\sin\varphi\, v_{m+1}^{(\beta)}(\varphi) = \left[\frac{(2\beta+m)(2\beta+m+1)}{2\beta+2m+3}\right]^{1/2} v_{m+1}^{(\beta)}(\varphi) + \left[\frac{m(m+1)}{2\beta+2m-1}\right]^{1/2} v_{m-1}^{(\beta)}(\varphi),$$

$$(2\beta+2m+1)^{1/2}\sin\varphi\, v_{m-1}^{(\beta)}(\varphi) = \left[\frac{(2\beta+m)(2\beta+m+1)}{2\beta+2m-1}\right]^{1/2} v_{m-1}^{(\beta)}(\varphi) + \frac{m(m+1)}{2\beta+2m+3}^{\,1/2} v_{m+1}^{(\beta)}(\varphi),$$

$$[d_\varphi + (\beta-\tfrac{1}{2})\cot\varphi]\,v_m^{(\beta)}(\varphi) = -i[(m+1)(2\beta+m)]^{1/2}v_{m+1}^{(\beta-1)}(\varphi),$$

$$[d_\varphi - (\beta+\tfrac{1}{2})\cot\varphi]\,v_m^{(\beta)}(\varphi) = -i[m(2\beta+m+1)]^{1/2}v_{m-1}^{(\beta+1)}(\varphi).$$

It is convenient to calculate the action of operators $\widetilde{A}_1$ and $\widetilde{A}_1^{\#}$ in the following orthonormal basis:

$$\mathscr{E}_\vartheta^{(1)} \equiv \{|klm;\mu\lambda)_\vartheta:\ k,l,m = 0,1,\ldots,\ \mu,\lambda = \pm 1\} \subset L_4^2(\widetilde{M}),$$

where

$$|klm;\mu+)_\vartheta := |klm;\mu\rangle_\vartheta,$$

$$|klm;\mu-)_\vartheta := (2\vartheta+2m+1-\mu)^{-1/2} \times [(2\vartheta+m+\delta_{\mu+1})^{1/2}|klm;3+\delta_{\mu-1}\rangle_\vartheta + \mu(m+\delta_{\mu+1})^{1/2}|klm-\mu;3+\delta_{\mu+1}\rangle_\vartheta].$$

Similarly, the action of $\widetilde{A}_2$ and $\widetilde{A}_2^{\#}$ will be expressed via vectors

$$|klm;\mu+]_\vartheta := |klm;\mu\rangle_\vartheta,$$

$$|klm;\mu-]_\vartheta := (2\vartheta+2m+1+\mu)^{1/2} \times [(m+\delta_{\mu-1})^{1/2}|klm+\mu;3+\delta_{\mu-1}\rangle_\vartheta + \mu(2\vartheta+m+\delta_{\mu-1})^{1/2}|klm;3+\delta_{\mu+1}\rangle_\vartheta],$$

which also form an orthonormal basis $\mathscr{E}_\vartheta^{(2)} \subset L_4^2(\widetilde{M})$. By introducing

$$f_{m,\vartheta}(j,\mu) := \eta \begin{cases} (2j)^{1/2}, & \mu = 1, \\ [2(j+m+\vartheta+1)]^{1/2}, & \mu = -1, \end{cases}$$
$$j = 0,1,...,$$

with $\eta := \exp(i\pi/4)$ [notice that $f_{m,\vartheta}(j,\mu) = 0$ iff $j = 0$ and $\mu = 1$], one has

$$\widetilde{A}_1|klm;\mu\lambda)_\vartheta = f_{m,\vartheta}(k,\mu)|k-\delta_{\mu-1},lm;-\mu,-\lambda)_\vartheta,$$
$$\widetilde{A}_1^\#|klm;\mu\lambda)_\vartheta = \overline{f_{m,\vartheta}}(k+\delta_{\mu+1},-\mu)|k+\delta_{\mu+1},lm;-\mu,-\lambda)_\vartheta,$$
$$\widetilde{A}_2|klm;\mu\lambda]_\vartheta = \mu\,\overline{f_{m,\vartheta}}(l,-\mu)|k,l-\delta_{\mu+1},m;-\mu,-\lambda]_\vartheta,$$
$$\widetilde{A}_2^\#|klm;\mu\lambda]_\vartheta = -\mu f_{m,\vartheta}(l+\delta_{\mu-1},\mu) \times |k,l+\delta_{\mu-1},m;-\mu,-\lambda]_\vartheta. \qquad \text{(B1)}$$

For proving algebraic irreducibility in Proposition 5.4 and Theorem 5.5 the following formulas for $\widetilde{B}_{rs} \equiv \frac{1}{2}\{\widetilde{A}_r,\widetilde{A}_s\}$, $r,s = 1,2$, which directly follow from (B1), have been used.

$$\widetilde{B}_{11}|klm;\mu\lambda)_\vartheta = f_{m,\vartheta}(k,\mu)f_{m,\vartheta}(k-\delta_{\mu-1},-\mu)|k-1,lm;\mu\lambda)_\vartheta,$$
$$\widetilde{B}_{11}^\#|klm;\mu\lambda)_\vartheta = \overline{f_{m,\vartheta}}(k+1,\mu)\,\overline{f_{m,\vartheta}}(k+\delta_{\mu+1},-\mu) \times |k+1,lm;\mu\lambda)_\vartheta,$$
$$\widetilde{B}_{22}|klm;\mu\lambda]_\vartheta = f_{m,\vartheta}(l-\delta_{\mu+1},\mu)f_{m,\vartheta}(l,-\mu)|k,l-1,m;\mu\lambda]_\vartheta,$$
$$\widetilde{B}_{22}^\#|klm;\mu\lambda]_\vartheta = \overline{f_{m,\vartheta}}(l+\delta_{\mu-1},\mu)\,\overline{f_{m,\vartheta}}(l+1,-\mu) \times |k,l+l,m;\mu\lambda]_\vartheta$$
$$\widetilde{B}_{12}|00m;\mu)_\vartheta = \left[\frac{m(\vartheta+m+1)(2\vartheta+m)}{\vartheta+m}\right]^{1/2}|00m-1;\mu)_\vartheta,$$
$$\widetilde{B}_{12}|00m;n\rangle_\vartheta = \left[\frac{m(\vartheta+m+\delta_{3-n})(2\vartheta+m-1+2\,\delta_{3-n})}{\vartheta+m-1+\delta_{3-n}}\right]^{1/2} \times|00m-1;n\rangle_\vartheta, \quad n=3,4.$$

[1]J. Blank and M. Havlíček, J. Math. Phys. **27**, 2823 (1986).
[2]H. D. Doebner and O. Melsheimer, Nuovo Cimento **49**, 73 (1967).
[3]In fact, for a cyclic vector Ψ of π one has $\mathscr{D} = \mathscr{U}(\widetilde{A}_j,\widetilde{B}_{jk};j,k = \pm 1, \pm 2)\Psi$, where $\widetilde{B}_{jk} := \pi(b_{jk})$, and Eq. (1.7) follows by using $\widetilde{B}_{jk} = \frac{1}{2}\{\widetilde{A}_j,\widetilde{A}_k\}$. With the help of relations (1.1), (1.4a), and invariance of $\mathscr{D}_{\rm vac}$ under $\widetilde{B}_{r-s}$, for $r,s = 1,2$, one finds that $\mathscr{D} = \mathscr{U}(\widetilde{A}_1^\#,\widetilde{A}_2^\#)\mathscr{D}_{\rm vac}$. This further implies (in view of the relation $\widetilde{N}\Psi = \nu\Psi$, which holds for any $\Psi\in\mathscr{D}_{\rm vac}$) that ν is the lowest eigenvalue of $\widetilde{N}$.
[4]M. Bednář *et al.*, JINR E2-82-771, Dubna, 1982; JINR E2-83-150, Dubna, 1983.
[5]W. Heidenreich, Phys. Lett. B **110**, 461 (1982).
[6]The notation of $N \times N$ matrices introduced in Ref. 4 has been changed as follows: instead of Z, $\widetilde{Z}$, $\widetilde{W}$, and H we write C, D, S, and U, respectively.
[7]E. Kamke, *Differentialgleichungen Reeler Funktionen* (Akad. Verlaggesellschaft, Leipzig, 1956), §31 (in German).
[8]I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products* (Fizmatgiz, Moscow, 1963) (in Russian).
[9]G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups I* (Springer, Berlin, 1972), §4.4.6.
[10]Harish-Chandra, Trans. Am. Math. Soc. **75**, 185 (1953), Theorem 5.
[11]N. T. Evans, J. Math. Phys. **8**, 170 (1967).
[12]This parameter is related to $\varkappa$ [the eigenvalue of the second-order Casimir operator of $B(0,2)$] *by* $\vartheta = \frac{1}{2}(\varkappa + 5 - m^2)^{1/2}$. For the parameter μ used in Ref. 4 one has $\mu = 4\vartheta$.
[13]For $\vartheta = \frac{1}{2}$ this argument can be applied only if $\widetilde{P}_4\Psi\neq 0$; however, if $\Psi\in\mathscr{D}_\vartheta^{(1)}\oplus\mathscr{D}_\vartheta^{(3)}$, then one finds $\widetilde{T}$ satisfying $\widetilde{T}\Psi = |00m;4\rangle_\vartheta$ directly, using Appendix B.
[14]I. Inaba, T. Maekawa, and T. Yamamoto, Lett. Nuovo Cimento **32**, 415 (1981); J. Math. Phys. **23**, 954 (1982).

Casimir operators for massless representations of the super-Poincare algebra and the reduction of the ten-dimensional massless scalar superfield

P. Kwon and M. Villasante
Department of Physics, University of California, Los Angeles, California 90024

(Received 11 August 1987; accepted for publication 23 September 1987)

Invariant operators for the massless little algebra of SP_d (super-Poincaré algebra in d dimensions) are given. They are used to decompose the scalar (massless) superfield in ten dimensions. Explicit expressions for the irreducible pieces are obtained after exploiting a relevant Cartan subalgebra.

I. INTRODUCTION

The obtaining of explicit expressions for irreducible representations of the $N=1$ super-Poincaré algebra in higher dimensions SP_d by reduction of general superfields has been accomplished in the massive case thanks to a complete understanding of the Casimir operators involved.[1] An exception to this is the case $d=10$ whose special difficulties have been pointed out in Ref. 2. The massive representations in four dimensions had been understood long before due to the efforts of a number of authors.[3] The massless representations of SP_4 for any N have also been completely understood long ago[4] by use of the method of induced representations.

In this paper we analyze the massless representations of SP_d and concentrate on the case $d=10$, which is especially important in the authors' opinion given the current interest in superstring theories. Aside from that, it shares the special difficulties of the massive case in $d=10$ and, since the off-shell extension of the massless case is the massive one, it is the only one whose off-shell extension is at present completely unknown.

The paper is structured as follows. In Sec. II we establish our conventions and give the commutation relations of the super-Poincaré algebra. Then, in Sec. III, we derive the little algebra whose representations will be the actual concern of the rest of the paper. The formulas given there will be familiar from our knowledge of the four-dimensional case.[4] Section IV is devoted to giving the complete set of Casimir operators corresponding to the general case. In Sec. V we start our analysis of the massless scalar superfield in ten dimensions by giving the particular form of the Casimir operators suitable for this case as well as their eigenvalues and the corresponding SO(8) irreducible representations. We also give there some necessary field contents.

We digress a bit in Sec. VI to explain how the Wigner method of induced representations works in this case. In Sec. VII we stop to give simplified forms of the relevant higher-order Casimir invariants as well as a derivation of their eigenvalues. Finally, we return in Sec. VIII to the main object of this paper in order to give explicit expressions for the irreducible pieces contained in the massless scalar superfield in ten dimensions.

II. THE SUPER-POINCARÉ ALGEBRA SP_d

We will start this section by recapitulating some necessary formulas from Ref. 1. The super-Poincaré algebra in d dimensions SP_d is defined by the commutation relations

$$[P_A,P_B]=0, \tag{2.1a}$$

$$[J_{AB},P^C]=-2i\delta^C_{[A}P_{B]}, \tag{2.1b}$$

$$[J_{AB},J^{CD}]=-4i\delta_{[A}{}^{[C}J_{B]}{}^{D]}, \tag{2.1c}$$

$$[P_A,Q^\alpha]=0, \tag{2.1d}$$

$$[J_{AB},Q^\alpha]=-(i/2)\Gamma_{AB}{}^\alpha{}_\beta Q^\beta, \tag{2.1e}$$

$$\{Q^\alpha,Q^\beta\}=(\not{P}C^{-1})^{\alpha\beta}, \tag{2.1f}$$

where the Latin indices A,B,C,D run from 0 to $d-1$ while the Greek indices run from 0 to $2^{[d/2]}$, which is the dimension of the representation of the Dirac (Clifford) algebra in d dimensions [corresponding to the basic spinorial representation of SO$(1,d-1)$ for d odd or the sum of the two basic spinorial representations of SO$(1,d-1)$ for d even]. Brackets enclosing a set of indices will denote complete antisymmetrization with strength 1, as usual. The Dirac algebra in our conventions is

$$\{\Gamma_A,\Gamma_B\}=2\eta_{AB}=2\,\mathrm{diag}(+\ -\cdots-),$$

and Γ tensors, again as usual, are

$$\Gamma_{A_1\cdots A_n}=\Gamma_{[A_1}\Gamma_{A_2}\cdots\Gamma_{A_n]}.$$

Here Q is a translationally invariant Majorana spinor, which can exist in all dimensions except 5, 6, 7 mod 8 (Ref. 5) and C is the charge conjugation matrix, $\bar{Q}=Q^\dagger\Gamma_0=Q^TC$.

The generators of SP_d are realized in superspace as follows:

$$P_A=-i\frac{\partial}{\partial x^A},$$

$$J_{AB}=-(x_AP_B-x_BP_A)+\frac{i}{2}\bar{\theta}\Gamma_{AB}\frac{\partial}{\partial\bar{\theta}}+\Sigma_{AB}, \tag{2.2}$$

$$Q^\alpha=i\left(\frac{\partial}{\partial\bar{\theta}_\alpha}+\frac{1}{2}\not{P}^\alpha{}_\beta\theta^\beta\right),$$

where Σ_{AB} represents the external spin operator which acts on the external indices of the corresponding superfield and $(i/2)\bar{\theta}\Gamma_{AB}(\partial/\partial\bar{\theta})$ is necessary to describe the internal spin part.

In addition to these generators, in superspace we can introduce a covariant derivative which, like Q, is a translationally invariant Majorana spinor,

$$D^\alpha=i\left(\frac{\partial}{\partial\bar{\theta}_\alpha}-\frac{1}{2}\not{P}^\alpha{}_\beta\theta^\beta\right), \tag{2.3}$$

 0022-2488/88/030560-12$02.50

but which anticommutes with it,

$$\{D^\alpha, Q^\beta\} = 0. \tag{2.4a}$$

Also

$$\{D^\alpha, D^\beta\} = -(\not{P}C^{-1})^{\alpha\beta}. \tag{2.4b}$$

In 2 mod 8 dimensions we can have Majorana–Weyl spinors.[5] Thus, in ten dimensions using the Weyl projectors,

$$\Pi^{(\pm)} = \tfrac{1}{2}(I \pm \Gamma_{(11)}), \quad \Gamma_{(11)} = \Gamma_0\Gamma_1\cdots\Gamma_9, \tag{2.5}$$

we can split Q in two pieces,

$$Q^{(\pm)} = \Pi^{(\pm)}Q \tag{2.6a}$$

and similarly for D,

$$D^{(\pm)} = \Pi^{(\pm)}D. \tag{2.6b}$$

Here $Q^{(\pm)}$ and $D^{(\pm)}$ satisfy the Majorana condition. Hence we can define two simpler superalgebras,[2] $SP_{10}^{(+)}$ and $SP_{10}^{(-)}$, whose gradings are, respectively, provided by $Q^{(+)}$ and $Q^{(-)}$. An obvious Casimir operator for the algebra (2.1) is the square of the momentum operator,

$$P^2 = P_A P^A. \tag{2.7}$$

Depending on the eigenvalue of P^2, the irreducible representations of SP_d, just like those of the Poincaré algebra P_d, can be separated into four categories (we discard the case $P_0 < 0$):

(i) $P^2 = M^2 > 0$;

(ii) $P^2 = 0$ but P_A not identically zero;

(iii) $P^2 = -M^2 < 0$;

(iv) $P_A = 0, \forall A$.

The representations that are interesting for physical applications fall into categories (i) and (ii). Category (i) is composed of the so-called massive representations.

III. LITTLE ALGEBRA

We now proceed to look for the little algebra corresponding to categories (i) and (ii). That is, we look for the generators of SP_d that leave invariant a particular form of P_A which we choose to be the one corresponding to a collinear frame,[4]

$$P_A = P_0(1,0,\ldots,0,z) \tag{3.1}$$

with

$$z = \sqrt{1-(M/P_0)^2}, \quad 0 \leqslant z \leqslant 1,$$

which clearly satisfies $P_A P^A = M^2$. The massless limit corresponds to

$$M \to 0 \Rightarrow z \to 1,$$

while for $z \neq 1$, we are in the massive case. The even generators of SP_d which leave invariant Eq. (3.1) are

$$\begin{aligned} &L_{ij} = J_{ij}, \\ &L_{i,d-1} = J_{i,d-1} + zJ_{0i}, \quad i,j = 1,\ldots,d-2. \end{aligned} \tag{3.2}$$

They satisfy the commutation relations

$$\begin{aligned} &[L_{ij}, L^{kl}] = -4i\delta_{[i}{}^{[k}L_{j]}{}^{l]}, \\ &[L_{ij}, L^{k,d-1}] = -2i\delta_{[i}{}^{k}L_{j]}{}^{d-1}, \\ &[L_{i,d-1}, L^{j,d-1}] = -i(1-z^2)L_i{}^j. \end{aligned} \tag{3.3}$$

When $z \neq 1$, we can normalize the $L_{j,d-1}$ generators to write

$$\begin{aligned} &[L_{ij}, L^{kl}] = -4i\delta_{[i}{}^{[k}L_{j]}{}^{l]}, \\ &\left[L_{ij}, \frac{L^{k,d-1}}{\sqrt{1-z^2}}\right] = -4i\delta_{[i}{}^{[k}L_{j]}{}^{d-1]}\frac{1}{\sqrt{1-z^2}}, \\ &\left[\frac{L_{i,d-1}}{\sqrt{1-z^2}}, \frac{L^{j,d-1}}{\sqrt{1-z^2}}\right] = -4i\delta_{[i}{}^{[j}L_{d-1]}{}^{d-1]}, \end{aligned} \tag{3.4}$$

which are the commutation relations of $SO(d-1)$. When $z=1$, we cannot do that, rather we get from (3.3)

$$\begin{aligned} &[L_{ij}, L^{kl}] = -4i\delta_{[i}{}^{[k}L_{j]}{}^{l]}, \\ &[L_{ij}, L^{k,d-1}] = -2i\delta_{[i}{}^{k}L_{j]}{}^{d-1}, \\ &[L_{i,d-1}, L_{j,d-1}] = 0, \end{aligned} \tag{3.5}$$

which is the Lie algebra of the Euclidean group in $d-2$ dimensions $E(d-2)$. This is the Wigner–Inönu contraction of Eq. (3.4).

Now we turn to the odd part of the little algebra. Since the Majorana charges Q commute with the momentum, they will also be part of the little algebra. In order to clarify the meaning of the anticommutation relations (2.1f), we turn to the light-cone coordinates,

$$\not{P} = P_A\Gamma^A = \tfrac{1}{2}(P_+\Gamma^+ + P_-\Gamma^- + P_i\Gamma^i),$$

with

$$P_\pm = P_0 \pm P_{d-1}, \quad \Gamma^\pm = \Gamma^0 \pm \Gamma^{d-1}. \tag{3.6}$$

Then we can introduce the projection operators,

$$\Pi_\pm = \tfrac{1}{2}(I \pm \Gamma_{0,d-1}). \tag{3.7}$$

These projection operators play a very important role in the massless case. They satisfy the following relations:

$$\begin{aligned} &\Gamma^\pm\Gamma^\mp = 4\Pi_\pm, \quad \Pi_\pm\Gamma_\pm = \Gamma_\pm\Pi_\mp = \Gamma_\pm, \\ &\Pi_\pm\Gamma_\mp = \Gamma_\pm\Pi_\pm = 0. \end{aligned} \tag{3.8}$$

Making use of $\Pi_\pm$ we can split Q in two pieces,

$$Q_\pm = \Pi_\pm Q, \tag{3.9}$$

which also satisfy the Majorana condition,

$$\bar{Q}_\pm = Q_\pm{}^\dagger\Gamma_0 = Q_\pm{}^T C = \bar{Q}\Pi_\mp \tag{3.10}$$

because of the following properties of $\Pi_\pm$:

$$C^{-1}\Pi_\pm{}^T C = \Pi_\mp, \quad \Pi_\pm{}^\dagger = \Pi_\pm. \tag{3.11}$$

From (2.1f) we can derive the anticommutation relations for $Q_\pm$,

$$\begin{aligned}\{Q^\alpha_+,Q^\beta_+\}&=\tfrac{1}{2}P_+(\Gamma^+C^{-1})^{\alpha\beta},\\ \{Q^\alpha_-,Q^\beta_-\}&=\tfrac{1}{2}P_-(\Gamma^-C^{-1})^{\alpha\beta},\\ \{Q^\alpha_+,Q^\beta_-\}&=P_i(\Gamma^i\Pi_+C^{-1})^{\alpha\beta},\\ \{Q^\alpha_-,Q^\beta_+\}&=P_i(\Gamma^i\Pi_-C^{-1})^{\alpha\beta}.\end{aligned} \tag{3.12}$$

This implies that in the collinear reference frame (3.1) we have two mutually anticommuting sets of Majorana charges,

$$\begin{aligned}\{Q^\alpha_+,Q^\beta_+\}&=\tfrac{1}{2}P_+(\Gamma^+C^{-1})^{\alpha\beta},\\ \{Q^\alpha_-,Q^\beta_-\}&=\tfrac{1}{2}P_-(\Gamma^-C^{-1})^{\alpha\beta},\\ \{Q^\alpha_+,Q^\beta_-\}&=\{Q^\alpha_-,Q^\beta_+\}=0.\end{aligned} \tag{3.13}$$

In the massless limit we have

$$P_-=0\Rightarrow\{Q^\alpha_-,Q^\beta_-\}=0 \tag{3.14}$$

and since $Q^\dagger_{-\gamma}=Q^\alpha_-(C\Gamma_0)_{\alpha\gamma}$ we have

$$\{Q_-,Q^\dagger_-\}=0, \tag{3.15}$$

which implies

$$Q_-=0. \tag{3.16}$$

So in a massless representation, we have only half as many odd generators in the little algebra as in the massive case. This is precisely what happens in four dimensions.[4]

Applying the projectors $\Pi_\pm$ to the superspace coordinates

$$\theta_\pm=\Pi_\pm\theta\Leftrightarrow\bar\theta_\pm=\bar\theta\Pi_\mp \tag{3.17}$$

which implies

$$\frac{\partial}{\partial\bar\theta_\mp}=\Pi_\pm\frac{\partial}{\partial\bar\theta}. \tag{3.18}$$

Then, from (2.2), (3.9), (3.17), and (3.18) we get the following superspace representations for $Q_\pm$:

$$Q_\pm=i\left(\frac{\partial}{\partial\bar\theta_\mp}+\frac{1}{4}P_\pm\Gamma^\pm\theta_\mp\right), \tag{3.19}$$

which in the massless case

$$P_-=0\Rightarrow Q_-=i\frac{\partial}{\partial\bar\theta_+} \tag{3.20}$$

implies that the representation of the little algebra is given by a superfield which does not depend on θ_+,

$$\phi(x,\theta)=\phi(x,\theta_-),\quad P_-\phi(x,\theta_-)=0. \tag{3.21}$$

In order to complete the structure of the little algebra, we have to look at the commutation relations of the even with the odd part. In the massive case they are simply the ones corresponding to the fact that the Q transforms according to the spinorial representation of SO$(d-1)$.[1] In the massless case we have

$$\begin{aligned}[][J_{ij},Q_\pm]&=-(i/2)\Gamma_{ij}Q_\pm,\\ [J_{0i}+J_{i,d-1},Q_+]&=-(i/2)\Gamma_+\Gamma_iQ_-,\\ [J_{0i}+J_{i,d-1},Q_-]&=0.\end{aligned} \tag{3.22}$$

So the condition $Q_-=0$ is respected by the E$(d-2)$ generators and Q_+ provides the grading which is quite similar to the one of the SP$_d$ algebra since Q_+ corresponds to the spinor representation of the SO$(d-2)$ generators J_{ij} and commutes with the noncompact ones,

$$\begin{aligned}[][J_{ij},Q_+]&=-(i/2)\Gamma_{ij}Q_+,\\ [J_{0i}+J_{i,d-1},Q_+]&=0,\\ \{Q_+,Q_+\}&=\tfrac{1}{2}P_+\Gamma^+C^{-1}.\end{aligned} \tag{3.23}$$

The representations of the algebra (3.5), (3.23) are labeled by a representation of the underlying E$(d-2)$ algebra which serves as a Clifford vacuum for the action of the Q_+. As indicated before, this representation is in general infinite dimensional. Again the interesting representations will be the finite dimensional ones corresponding to vanishing "little mass." This means that all the noncompact generators vanish and therefore they are described by a SO$(d-2)$ representation instead, acting as a Clifford vacuum. These are the representations one obtains from taking the massless limit of massive representations. When we talk of massless representations we mean this type, where not only the mass P^2 vanishes but the little mass as well.

As mentioned in Sec. II, in the particular case of ten dimensions, we have Weyl projectors which respect the Majorana condition. We can apply them also in the massless case since they commute with the light-cone projectors

$$[\Pi^{(\pm)},\Pi_\pm]=0 \tag{3.24}$$

so that we can define the Majorana spinors,

$$Q^{(\pm)}_\pm=\Pi^{(\pm)}\Pi_\pm Q. \tag{3.25}$$

Since the $Q^{(+)}$ and the $Q^{(-)}$ are mutually anticommuting,[2] to keep them both would amount to working with an $N=2$ extended super-Poincaré algebra, which is not our intention here. We will only keep $Q^{(+)}$. So our massless superspace in ten dimensions will include only the anticommuting coordinates,

$$\theta^{(-)}_-=\Pi^{(-)}\Pi_-\theta \tag{3.26}$$

and we have the representations

$$\begin{aligned}Q^{(+)}_+&=i\left(\frac{\partial}{\partial\bar\theta^{(-)}_-}+\frac{1}{4}P_+\Gamma^+\theta^{(-)}_-\right),\\ D^{(+)}_+&=i\left(\frac{\partial}{\partial\bar\theta^{(-)}_-}-\frac{1}{4}P_+\Gamma^+\theta^{(-)}_-\right),\end{aligned} \tag{3.27}$$

and the anticommutation relations

$$\begin{aligned}\{Q^{(+)\alpha}_+,Q^{(+)\beta}_+\}&=\tfrac{1}{2}P_+(\Pi^{(+)}\Gamma^+C^{-1})^{\alpha\beta},\\ \{D^{(+)\alpha}_+,D^{(+)\beta}_+\}&=-\tfrac{1}{2}P_+(\Pi^{(+)}\Gamma^+C^{-1})^{\alpha\beta},\\ \{Q^{(+)}_+,D^{(+)}_+\}&=0.\end{aligned} \tag{3.28}$$

We will call SP$^{(+)}_{10+}$ the ten-dimensional super-Poincaré algebra whose grading is provided just by $Q^{(+)}_+$.

IV. CASIMIR OPERATORS

The basic object for the construction of Casimir operators for the massive representations of SP$_d$ is[1]

$$U_{AB}=J_{AB}+(2/P^2)P^EJ_{E[A}P_{B]}+(1/4P^2)\,\bar Q\not{P}\Gamma_{AB}Q. \tag{4.1}$$

This object is singular when $P^2=0$. One could consider the nonsingular object

$$V_{AB}=\lim_{P^2\to 0}P^2U_{AB}\,, \tag{4.2}$$

but there are two problems with this operator. First, the contribution from the odd part of the algebra disappears, as we can see by using light-cone coordinates and going to a collinear frame,

$$(1/4P^2)\overline{Q}\not{P}\Gamma_{AB}Q = (1/P_+)\overline{Q}_+\Gamma^-\Gamma_{AB}Q_+ + (1/P_-)\overline{Q}_-\Gamma^+\Gamma_{AB}Q_-\,, \tag{4.3}$$

which in the limit becomes, by Eq. (3.16),

$$\lim_{P^2\to 0}\tfrac{1}{4}\overline{Q}\not{P}\Gamma_{AB}Q=P_+\overline{Q}_-\Gamma^+\Gamma_{AB}Q_-=0\,. \tag{4.4}$$

So V_{AB} only carries contribution from the Poincaré subalgebra P_d. Second, all the Casimirs,

$$\mathrm{Tr}\,V^n=V_{A_1}{}^{A_2}V_{A_2}{}^{A_3}\cdots V_{A_n}{}^{A_1} \tag{4.5}$$

vanish and they do not allow us to distinguish between different massless representations. The same is true for the squares of the generalized Pauli–Lubanski tensors,[1]

$$W_{A_i\cdots A_{2k+1}}=P_{[A_1}V_{A_2A_3}\cdots V_{A_{2k}A_{2k+1}]}\,. \tag{4.6}$$

This is so because

$$V_{AB}=2P^EJ_{E[A}P_{B]} \tag{4.7}$$

is orthogonal to the momentum,

$$P^AV_{AB}=0\,, \tag{4.8}$$

when $P^2=0$.

So these operators are not useful to describe the massless representations of SP_d as defined in Sec. III.

What we can do is to describe these representations by means of the Casimir operators of the little algebra. Since we are interested in representations whose little mass vanishes, we would encounter the same problem mentioned above should we try to use the Casimir operator for the graded $\mathrm{E}(d-2)$ algebra (3.5), (3.23). Instead we will look directly at the invariant operators of the graded $\mathrm{SO}(d-2)$ algebra. In analogy with the massive case, we start by constructing an operator which commutes with Q_+,

$$U_{ij}=J_{ij}-(i/8P_+)\overline{Q}_+\Gamma^-\Gamma_{ij}Q_+\,,\qquad [U_{ij},Q]=0\,. \tag{4.9}$$

Therefore any scalar or pseudoscalar constructed out of the U_{ij} will be an invariant operator for the graded $\mathrm{SO}(d-2)$ algebra.

In order to know how many independent operators we need, we have to look at the algebra satisfied by the U_{ij},

$$[U_{ij},U^{kl}]=-4i\delta_{[i}{}^{[k}U_{j]}{}^{l]}\,, \tag{4.10}$$

which is precisely the algebra of $\mathrm{SO}(d-2)$. This is the way it should be, because we know that the irreducible representations of a graded Lie algebra are characterized by an irreducible representation of the even part; therefore, since here the representations of the graded $\mathrm{SO}(d-2)$ algebra are described solely by objects constructed out of the U_{ij}, it is only natural that they obey a $\mathrm{SO}(d-2)$ algebra. Casimir operators for the algebra are the traces

$$\mathrm{Tr}\,U^n=U_{i_1}{}^{i_2}U_{i_2}{}^{i_3}\cdots U_{i_n}{}^{i_1}\,, \tag{4.11}$$

which are all scalars while in even dimensions we also have the Pfaffian,

$$\mathrm{Pf}(U)=\frac{1}{2^{(d-2)/2}((d-2)/2)!}\times\epsilon^{i_1\cdots i_{d-2}}U_{i_1i_2}\cdots U_{i_{d-3}i_{d-2}}\,, \tag{4.12}$$

which is a pseudoscalar. A complete set of independent operators[6] is the following.

(a) For d odd,

$$C_p=\mathrm{Tr}\,U^{2p}\,,\quad p=1,2,\ldots,[(d-2)/2]\,.$$

(b) For d even,

$$C_p=\mathrm{Tr}\,U^{2p}\,,\quad p=1,2,\ldots,(d-2)/2-1\,;$$

$$C'_{(d-2)/2}=2^{(d-2)/2}((d-2)/2)!\,\mathrm{Pf}(U)\,.$$

For the next section we need expressions which are appropriate for $d=10$. They are

$$U_{ij}=J_{ij}-(i/8P_+)\overline{Q}{}^{(+)}_+\Gamma^-\Gamma_{ij}Q^{(+)}_+ \tag{4.13}$$

and the complete set of Casimirs is

$$C_p=\mathrm{Tr}\,U^{2p}\,,\quad p=1,2,3\,;\quad C'_4=2^44!\,\mathrm{Pf}(U)\,. \tag{4.14}$$

Finally, let us close this section by mentioning that the eigenvalues of these operators for any irreducible representation in terms of the highest weight vector, for all the classical groups, have been given in a beautiful series of papers by Perelomov and Popov.[7] A translation of their results suitable for our needs can be found in Ref. 1.

V. THE MASSLESS SCALAR SUPERFIELD IN TEN DIMENSIONS

What we can call the Casimir approach to decompose a general superfield consists of computing the eigenvalues of the Casimir operators by writing the appropriate representation for the operators J_{AB} [i.e., Σ_{AB} in Eq. (2.2)] and then finding the representations corresponding to those eigenvalues.[3] Different irreducible pieces can be separated by either using appropriate projection operators or solving equivalent differential equations. In four dimensions these procedures have been explained in the papers of Ref. 3, for instance, while in Ref. 1 they have been applied in detail to 11-dimensional case.

The simplest case corresponds to the scalar superfield, which means

$$\Sigma_{AB}=0 \tag{5.1}$$

in Eq. (2.2). Then from Eqs. (4.9), (2.2), and (2.3), one finds (let us remember that $P_i=0$ in our reference frame)

$$U_{ij}=-(i/8P_+)\overline{D}_+\Gamma^-\Gamma_{ij}D_+ \tag{5.2}$$

in the general case while, by using Eq. (4.13) instead of (4.9), we get

$$U_{ij}=-(i/8P_+)\overline{D}{}^{(+)}_+\Gamma^-\Gamma_{ij}D^{(+)}_+ \tag{5.3}$$

in ten dimensions.

The first Casimir is then

$$C_1 = \mathrm{Tr}\, U^2 = \left(-\frac{i}{8P_+} \right)^2 \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{ij} D^{(+)}_{+} \times \bar{D}^{(+)}_{+} \Gamma^- \Gamma^{ji} D^{(+)}_{+} . \tag{5.4}$$

This operator can be manipulated by successive Fierz transformations, as described in Appendix A to obtain

$$C_1 = -14 . \tag{5.5}$$

There are three irreducible representations of SO(8) which correspond to this eigenvalue which are precisely the three eight-dimensional ones. They are tabulated in Table I along with the eigenvalues of all the Casimir operators (4.14).

For dimensional reasons, only two of these representations are included in the massless scalar superfield. It is also clear that one of the representations must be bosonic (true) while the other must be fermionic (spinorial). So we either have $[1] \oplus [\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ or $[1] \oplus [\frac{1}{2}\frac{1}{2}\frac{1}{2} -\frac{1}{2}]$. We will see below how to resolve this ambiguity.

Here again we encounter the fact that C_1 is a number, just like in the massive case.[2] This implies that C_1 cannot be used to separate the irreducible pieces and one must resort to higher Casimir operators or some other method to separate them. This is a reflection of the fact that there is no scalar or pseudoscalar operator quadratic in $D^{(+)}_{+}$, unlike the 11-dimensional case where we have $\bar{D}D$ for instance.

In order to resolve the ambiguity in the decomposition of $\phi(x,\theta^{(-)}_{-})$, we will derive a relation between C_2 and C'_4. By a Fierz transformation, one can derive the identity

$$\begin{aligned} &\bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_1 i_2} D^{(+)}_{+} \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_3 i_4} D^{(+)}_{+} \\ &\quad = -3P_+ \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_1 i_2} \Gamma_{i_3 i_4} D^{(+)}_{+} \\ &\qquad + \tfrac{1}{16} \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_1 i_2} \Gamma^{mn} \\ &\qquad \times \Gamma_{i_3 i_4} D^{(+)}_{+} \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{mn} D^{(+)}_{+} . \end{aligned} \tag{5.6}$$

Antisymmetrizing in the indices $i_1,...,i_4$ the first term vanishes due to the Majorana condition. Furthermore, the identity

$$\Gamma_{[i_1 i_2} \Gamma^{mn} \Gamma_{i_3 i_4]} = \Gamma_{i_1 i_2 i_3 i_4}{}^{mn} + 4\delta_{[i_1}{}^{m} \delta_{i_2}{}^{n} \Gamma_{i_3 i_4]} \tag{5.7}$$

implies

$$\begin{aligned} &\tfrac{3}{4} \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{[i_1 i_2} D^{(+)}_{+} \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_3 i_4]} D^{(+)}_{+} \\ &\quad = \tfrac{1}{16} \bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_1 \cdots i_4 mn} D^{(+)}_{+} \bar{D}^{(+)}_{+} \Gamma^- \Gamma^{mn} D^{(+)}_{+} . \end{aligned} \tag{5.8}$$

Now from the definition of $\Gamma_{(11)}$ we get

$$\begin{aligned} &\bar{D}^{(+)}_{+} \Gamma^- \Gamma_{i_1 \cdots i_4 mn} D^{(+)}_{+} \\ &\quad = \tfrac{1}{2} \epsilon_{i_1 \cdots i_4 mn j_1 j_2} \bar{D}^{(+)}_{+} \Gamma^- \Gamma^{j_1 j_2} D^{(+)}_{+} \end{aligned} \tag{5.9}$$

and substituting in (5.8) we get the self-duality condition for $U_{[i_1 i_2} U_{i_3 i_4]}$,

$$U_{[i_1 i_2} U_{i_3 i_4]} = (1/4!) \epsilon_{i_1 \cdots i_4 j_1 \cdots j_4} U^{j_1 j_2} U^{j_3 j_4} . \tag{5.10}$$

This self-duality property allows us to express the pseudoscalar operator C'_4 in terms of the scalar ones C_1 and C_2,

$$\begin{aligned} C'_4 &= \epsilon^{i_1 \cdots i_8} U_{i_1 i_2} U_{i_3 i_4} U_{i_5 i_6} U_{i_7 i_8} \\ &= \tfrac{1}{24} \epsilon^{i_1 \cdots i_8} \epsilon_{i_1 \cdots i_4 j_1 \cdots j_4} U^{j_1 j_2} U^{j_3 j_4} U_{i_5 i_6} U_{i_7 i_8} \\ &= 8C_1{}^2 - 16C_2 + 80i\, \mathrm{Tr}\, U^3 , \end{aligned} \tag{5.11}$$

where we have expanded the product of the two Levi-Civita symbols and made use of the commutation relations (4.10). The Casimir $\mathrm{Tr}\, U^3$ is not independent. In fact, for the group SO(p), we have the identity

$$\mathrm{Tr}\, U^3 = [(p-2)/2] i\, \mathrm{Tr}\, U^2 . \tag{5.12}$$

So for us $\mathrm{Tr}\, U^3 = 3i\, C_1$ and we get finally

$$\begin{aligned} C'_4 &= 8C_1{}^2 - 240C_1 - 16C_2 \\ &= 4928 - 16C_2 . \end{aligned} \tag{5.13}$$

According to Table I, Eq. (5.13) is satisfied by the representations [1] and $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ but not by $[\frac{1}{2}\frac{1}{2}\frac{1}{2} -\frac{1}{2}]$. So $\phi(x,\theta^{(-)}_{-})$ contains [1] and $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$. The field content of each of the irreducible pieces is given in Table II. As will be clear in the next section, the result is obtained by reducing the Kronecker products shown in the table. Let us note that the irreducible superfield [1] contains precisely the physical states of the supergravity multiplet in ten dimensions and nothing else.

For completeness, we will give the field content of the general massless chiral superfield,

$$\phi(x,\theta^{(-)}_{-}) = \sum_{n=0}^{8} \theta^{(-)\alpha_1}_{-} \cdots \theta^{(-)\alpha_n}_{-} F_{\alpha_1 \cdots \alpha_n}(x) \tag{5.14}$$

separated by powers of $\theta^{(-)}_{-}$ in Table III.

For $n > 4$, we have the same representations as for $8-n$. Comparing Tables II and III, we see that the main difference between the two irreducible pieces lies in the $\theta^{(-)\alpha_1}_{-} \cdots \theta^{(-)\alpha_4}_{-}$ sector, where one of them inherits the field [2] of $\phi(x,\theta^{(-)}_{-})$ and the other one inherits [1111]. There are four possible choices of $\phi(x,\theta^{(\pm)}_{\pm})$. If we consider $\phi(x,\theta^{(+)}_{+})$, the field content is again given by Tables II and III, where now entries correspond to $\theta^{(+)n}_{+}$, and Eqs. (5.10), (5.13) remain unchanged. Instead, if we consider

TABLE I. Casimir eigenvalues for irreducible representations of SO(8).

	Dimension	C_1	C_2	C_3	C'_4
[1]	8	−14	308	−14 714	0
$[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$	8	−14	$\frac{301}{2}$	$-\frac{14\,707}{8}$	2520
$[\frac{1}{2}\frac{1}{2}\frac{1}{2} -\frac{1}{2}]$	8	−14	$\frac{301}{2}$	$-\frac{14\,707}{8}$	−2520

TABLE II. Field content of the irreducible pieces in $\phi(x,\theta^{(-)}_{-})$.

Irreducible superfield		Fields
[1]	[1]×[1]	[2],[11],[0]
	$[1]\times[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$	$[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$, $[\frac{1}{2}\frac{1}{2}\frac{1}{2} -\frac{1}{2}]$
$[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$	$[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]\times[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$	[1111], [11], [0]
	$[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]\times[1]$	$[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$, $[\frac{1}{2}\frac{1}{2}\frac{1}{2} -\frac{1}{2}]$

TABLE III. Field content of $\phi(x,\theta^{(-)}_{-})$.

n	SO(8) representations in $F_{\alpha_1\cdots\alpha_n}$
0	[0]
1	$[\frac{1}{2}\frac{1}{2}\frac{1}{2}-\frac{1}{2}]$
2	[11]
3	$[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$
4	[2], [1111]

$\phi(x,\theta^{(+)}_{-})$ and/or $\phi(x,\theta^{(-)}_{+})$ we have to make in Tables II and III the interchanges $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]\leftrightarrow[\frac{1}{2}\frac{1}{2}\frac{1}{2}-\frac{1}{2}]$, $[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]\leftrightarrow[\frac{3}{2}\frac{1}{2}\frac{1}{2}-\frac{1}{2}]$, $[1111]\leftrightarrow[111-1]$ and there is an overall minus sign in (5.10) and (5.13). Finally, let us mention that, in spite of the above identification of fields power by power in $\theta^{(-)}_{-}$, we are still unable to write explicit expressions for the irreducible pieces since we do not know how the fields in the jth power of $\theta^{(-)}_{-}$ ($j<4$) are related to those in the $(8-j)$th power. We will address this question in Sec. VIII.

VI. WIGNER METHOD OF INDUCED REPRESENTATIONS

It is very instructive to look at the Wigner method of constructing the states of the representation in order to gain a better understanding of the problem. The first step is to divide the $Q^{(+)}_{+}$ ($D^{(+)}_{+}$) operators in Eq. (3.27) in two groups $q,q^\dagger$ ($d,d^\dagger$) whose anticommutation relations correspond to a Clifford algebra in standard form. Since $Q^{(+)}_{+}$ has eight independent components, we will have four q and four $q^\dagger$ operators and similarly in the D sector. Then one proceeds to construct a representation of the little algebra (3.5),(3.23) as usual: from an irreducible SO(8) representation $|\Omega\rangle$ acting as a Clifford vacuum, $q^\alpha|\Omega\rangle=0$, one obtains the remaining states by repeated application of the "creation" operators $q^\dagger$ until all possibilities are exhausted. These states form an irreducible representation of (3.5), (3.23) which we call $|\overline{\Omega}\rangle$,

$$|\overline{\Omega}\rangle:\ |\Omega\rangle, q^\dagger_\alpha|\Omega\rangle, q^\dagger_{\alpha_1}q^\dagger_{\alpha_2}|\Omega\rangle,\ldots, q^\dagger_{\alpha_1}q^\dagger_{\alpha_2}q^\dagger_{\alpha_3}q^\dagger_{\alpha_4}|\Omega\rangle\,. \tag{6.1}$$

The dimension of $|\overline{\Omega}\rangle$ is

$$\dim\overline{\Omega}=\sum_{j=0}^{4}\binom{4}{j}\dim\Omega=2^4\dim\Omega\,. \tag{6.2}$$

When we refer to the scalar superfield, then $|\Omega\rangle$ is just the trivial representation of SO(8) and $\dim\overline{\Omega}=16$.

Now one can proceed to apply the creation operators $d^\dagger$ to an irreducible representation $|\overline{\Omega}\rangle$ of the little algebra, which now satisfies the new vacuum condition $d^\alpha|\overline{\Omega}\rangle=0$, in order to generate a general superfield $|\overline{\overline{\Omega}}\rangle$,

$$|\overline{\overline{\Omega}}\rangle:\ |\overline{\Omega}\rangle, d^\dagger_\alpha|\overline{\Omega}\rangle, d^\dagger_{\alpha_1}d^\dagger_{\alpha_2}|\overline{\Omega}\rangle,\ldots, d^\dagger_{\alpha_1}\cdots d^\dagger_{\alpha_4}|\overline{\Omega}\rangle\,, \tag{6.3}$$

whose dimension is, of course,

$$\dim\overline{\overline{\Omega}}=2^8\dim\Omega\,. \tag{6.4}$$

We can see, just as in the massive case in Ref. 2, that we cannot obtain the irreducible representations involved by taking antisymmetrized Kronecker powers since the operators $q^\dagger_\alpha$ (or $d^\dagger_\alpha$) do not form an irreducible representation of SO(8) (this group has no four-dimensional irreducible representations). So the alternative procedure followed in Ref. 1 does not work here and, even though one might guess the answer by playing with dimensions, only the Casimir approach provides an unambiguous solution.

The complete set of states of $\phi(x,\theta^{(-)}_{-})$ will be given by the above procedure taking $|\Omega\rangle$ as the trivial representation of SO(8): $|0\rangle$, and it is displayed in Table IV. We have separated the states corresponding to the irreducible pieces [1] and $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$, which, of course, must be done at the "d level." Since $d^\dagger_\alpha$ is a fermionic operator, it is clear that the bosonic representation described by the "superweight" [1] [1] must contain only states with an even number of $d^\dagger_\alpha$ while the fermionic one, with superweight $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$, must contain the states with an odd number of $d^\dagger_\alpha$. This will become relevant in Sec. VIII.

The even powers of $d^\dagger_\alpha$ (eight operators in total) span the representation [1] of SO(8) while the odd powers (also eight in total) span the representation $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ instead. On the other hand, the set of powers of $q^\dagger_\alpha$ (even and odd, 16 operators in all) span the reducible representation $[1]\oplus[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ of SO(8). Thus in Table IV we just have the states corresponding to the Kronecker products given in Table II.

VII. SIMPLIFIED FORMS OF THE CASIMIR OPERATORS

In this section we will present some simplifications of the Casimirs C_2 and C'_4 which will give us the means to compute their eigenvalues directly. Let us call

$$\begin{aligned}\mathscr{M}_i{}^j&=(-8P_+/i)U_i^j\\&=\overline{D}^{(+)}_{+}\Gamma^-\Gamma_i{}^jD^{(+)}_{+}\\&=(C\Gamma^-\Gamma_i{}^j\Pi^{(+)})_{\alpha_1\alpha_2}D^{(+)\alpha_1\alpha_2}_{+}\,,\end{aligned} \tag{7.1}$$

where we have used the definition $D^{(+)\alpha_1\cdots\alpha_n}_{+}=D^{(+)[\alpha_1}_{+}\cdots D^{(+)\alpha_n]}_{+}$. Using the formulas of Appendix A of Ref. 1 to reduce the $D^{(+)}_{+}$ tensors we get

$$\mathscr{M}_i{}^k\mathscr{M}_k{}^j=\mathscr{M}_{(2)i}{}^j-24P_+\mathscr{M}_i{}^j+112P_+{}^2\delta_i^j \tag{7.2}$$

if we define, in general,

$$\begin{aligned}\mathscr{M}_{(n)i}{}^j=&(C\Gamma^-\Gamma_i{}^{i_2}\Pi^{(+)})_{\alpha_1\alpha_2}\times\cdots\\&\times(C\Gamma^-\Gamma_{i_n}{}^j\Pi^{(+)})_{\alpha_{2n-1}\alpha_{2n}}D^{(+)\alpha_1\cdots\alpha_{2n}}_{+}\,.\end{aligned} \tag{7.3}$$

TABLE IV. States of $\phi(x,\theta^{(-)}_{-})$ separated into the irreducible pieces.

	$\vert\overline{\overline{\Omega}}\rangle$	$\vert\overline{\Omega}\rangle$	$\vert\Omega\rangle$
[1]	1 $d^\dagger_{\alpha_1}d^\dagger_{\alpha_2}$ $d^\dagger_{\alpha_1}d^\dagger_{\alpha_2}d^\dagger_{\alpha_3}d^\dagger_{\alpha_4}$	1 $q^\dagger_\alpha$ $q^\dagger_{\alpha_1}q^\dagger_{\alpha_2}$	$\vert 0\rangle$
$[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$	$d^\dagger_\alpha$ $d^\dagger_{\alpha_1}d^\dagger_{\alpha_2}d^\dagger_{\alpha_3}$	$q^\dagger_{\alpha_1}q^\dagger_{\alpha_2}q^\dagger_{\alpha_3}$ $q^\dagger_{\alpha_1}q^\dagger_{\alpha_2}q^\dagger_{\alpha_3}q^\dagger_{\alpha_4}$	

Taking the trace of Eq. (7.2) and using (5.12), we get

$$\mathcal{M}_{(2)i}{}^{i}=0\,. \tag{7.4}$$

Multiplying (7.2) by $\mathcal{M}_j{}^i$ and using (5.12) again we get also

$$\mathcal{M}_{(2)i}{}^{j}\mathcal{M}_j{}^{i}=\mathcal{M}_j{}^{i}\mathcal{M}_{(2)i}{}^{j}=0\,. \tag{7.5}$$

Squaring (7.2), taking trace of the result, and making use of (7.4) and (7.5), we arrive at

$$\operatorname{Tr}\mathcal{M}^4=\operatorname{Tr}\mathcal{M}^2_{(2)}+616\,448P^4_+ \tag{7.6a}$$

or

$$\operatorname{Tr}U^4=(-i/8P_+)^4\operatorname{Tr}\mathcal{M}^2_{(2)}+\tfrac{301}{2}\,. \tag{7.6b}$$

Note that the constant $\frac{301}{2}$ appears in Table I. Again we can reduce the product of $D^{(+)}_+$ tensors appearing in $\operatorname{Tr}\mathcal{M}^2_{(2)}$ to get after some algebra

$$\operatorname{Tr}\mathcal{M}^2_{(2)}=\operatorname{Tr}\mathcal{M}_{(4)}+8\times 8!\,P^4_+\,. \tag{7.7}$$

Therefore we have for C_2,

$$C_2=(-i/8P_+)^4\operatorname{Tr}\mathcal{M}_{(4)}+\tfrac{917}{4}\,, \tag{7.8}$$

where, according to (7.3),

$$\operatorname{Tr}\mathcal{M}_{(4)}=(C\Gamma^-\Gamma_{i_1}{}^{i_2}\Pi^{(+)})_{\alpha_1\alpha_2}\times\cdots\times(C\Gamma^-\Gamma_{i_4}{}^{i_1}\Pi^{(+)})_{\alpha_7\alpha_8}D^{(+)\alpha_1\cdots\alpha_8}_+\,. \tag{7.9}$$

A similar treatment for the operator C'_4 gives in turn the result

$$C'_4=(-i/8P_+)^4\mathcal{M}'_{(4)}+1260 \tag{7.10}$$

with

$$\mathcal{M}'_{(4)}=\epsilon^{i_1\cdots i_8}(C\Gamma^-\Gamma_{i_1i_2}\Pi^{(+)})_{\alpha_1\alpha_2}\times\cdots\times(C\Gamma^-\Gamma_{i_7i_8}\Pi^{(+)})_{\alpha_7\alpha_8}D^{(+)\alpha_1\cdots\alpha_8}_+\,. \tag{7.11}$$

There are only eight nonvanishing operators $D^{(+)\alpha}_+$ that by choosing an appropriate representation of the Dirac algebra (see Appendix B for details), we can make correspond to the values $\alpha=1,...,8$. In that case all objects with eight totally antisymmetrized α indices will be proportional to the Levi-Civita symbol. Thus we have

$$D^{(+)\alpha_1\cdots\alpha_8}_+=\epsilon^{\alpha_1\cdots\alpha_8}D^{(+)123\cdots 8}_+\,, \tag{7.12a}$$

$$(C\Gamma^-\Gamma_{i_1}{}^{i_2}\Pi^{(+)})_{\alpha_1\alpha_2}\cdots(C\Gamma^-\Gamma_{i_4}{}^{i_1}\Pi^{(+)})_{\alpha_7\alpha_8}=A\epsilon_{\alpha_1\cdots\alpha_8}\,, \tag{7.12b}$$

$$\epsilon_{i_1\cdots i_8}(C\Gamma^-\Gamma^{i_1i_2}\Pi^{(+)})_{[\alpha_1\alpha_2}\cdots(C\Gamma^-\Gamma^{i_7i_8}\Pi^{(+)})_{\alpha_7\alpha_8]}=A'\epsilon_{\alpha_1\cdots\alpha_8}\,. \tag{7.12c}$$

The constants A and A' can be calculated and their values are

$$A=2^7\,,\quad A'=2^{11}\,. \tag{7.13}$$

Therefore we have

$$C_2=(1/8P_+)^4\,2^7\times 8!D^{(+)123\cdots 8}_+ +\tfrac{917}{4}=(1260/P_+{}^4)D^{(+)123\cdots 8}_+ +\tfrac{917}{4} \tag{7.14}$$

and

$$C'_4=(1/8P_+)^4\,2^{11}\times 8!D^{(+)123\cdots 8}_+ +1260=(20\,160/P_+{}^4)D^{(+)123\cdots 8}_+ +1260\,. \tag{7.15}$$

The eigenvalues of $D^{(+)123\cdots 8}_+$ can be obtained from very simple considerations. First, we note that if two operators $d,d^\dagger$ satisfy

$$\{d,d^\dagger\}=P_+\,,\quad \{d,d\}=\{d^\dagger,d^\dagger\}=0\,,$$

then the operator $C=\frac{1}{2}[d,d^\dagger]$ satisfies

$$C^2=(P_+/2)^2\,. \tag{7.16}$$

Since $D^{(+)123\cdots 8}_+$ is the product of four such operators made out of four independent, mutually anticommuting such d^α, it satisfies

$$(D^{(+)123\cdots 8}_+)^2=(P_+/2)^8 \tag{7.17}$$

and from this we obtain the eigenvalues of C_2 and C'_4

$$C_2=\begin{cases}308\,,\\ \frac{301}{2}\,;\end{cases}\quad C'_4=\begin{cases}2520\,,\\ 0\,;\end{cases} \tag{7.18}$$

which are precisely the values of Table I. Expressions (7.14), (7.15) for C_2 and C'_4 are considerable simplifications over the initial definition of these operators.

VIII. IRREDUCIBLE COMPONENTS OF $\phi(x,\theta^{(-)}_-)$

We devote this section to the irreducible pieces of $\phi(x,\theta^{(-)}_-)$ and give explicit expressions for them. In principle this is straightforward if we use projection operators[3] constructed out of Casimir operators since we know the eigenvalues. The result of this procedure is

$$\phi_{[1]}=\frac{C_2-\frac{301}{2}}{308-\frac{301}{2}}\phi(x,\theta^{(-)}_-)=\frac{C'_4-2520}{-2520}\phi(x,\theta^{(-)}_-)\,, \tag{8.1}$$

$$\phi_{[\frac12\frac12\frac12\frac12]}=\frac{C_2-308}{\frac{301}{2}-308}\phi(x,\theta^{(-)}_-)=\frac{C'_4}{2520}\phi(x,\theta^{(-)}_-)\,.$$

The projection operators in terms of C_2 or C'_4 are simple since they contain only one factor. But even so, this procedure is not practical given the fact that both C_2 and C'_4 are very complicated operators containing eight covariant derivatives, even with the simplifications of the previous section. The usual alternative procedure of solving equivalent differential equations is not at all different here, since the equations to solve are

$$C_2\phi_{[1]}=308\phi_{[1]}\ \text{ or }\ C'_4\phi_{[1]}=0 \tag{8.2a}$$

and

$$C_2\phi_{[\frac12\frac12\frac12\frac12]}=\tfrac{301}{2}\phi_{[\frac12\frac12\frac12\frac12]}\ \text{ or }\ C'_4\phi_{[\frac12\frac12\frac12\frac12]}=2520\,\phi_{[\frac12\frac12\frac12\frac12]}\,. \tag{8.2b}$$

We recall again the fact that there is no simpler operator in this case, like $\overline{D}D$ in other dimensions.

In this paper we will follow a different approach which makes use of the Cartan subalgebra of (4.10), whose generators are

$$H_I=U_{2I-1,2I}\,,\quad I=1,...,4\,. \tag{8.3}$$

From basic group theory, we know that simultaneous eigenstates of the operators H_I will also be eigenstates of the Casimir operators. These eigenstates $|\psi\rangle$ satisfy

$$\mathbf{H}|\psi\rangle = \mathbf{w}|\psi\rangle\,, \tag{8.4}$$

where $\mathbf{H} = (H_1,H_2,H_3,H_4)$ is a vector operator and $\mathbf{w} = (w_1,w_2,w_3,w_4)$ is called a weight. The eight weights of the vector representation $[1]$ of SO(8) are

$$(\pm 1,0,0,0),\ (0,\pm 1,0,0),\ (0,0,\pm 1,0),\ (0,0,0,\pm 1) \tag{8.5}$$

and the weights of the spinorial representation $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ are

$$(\pm\tfrac{1}{2},\pm\tfrac{1}{2},\pm\tfrac{1}{2},\pm\tfrac{1}{2}) \tag{8.6}$$

with an even number of $+$ and/or $-$ signs.

Let us start with the eigenstate corresponding to the highest weight $(1,0,0,0)$,

$$(H_1,H_2,H_3,H_4)|\psi\rangle = (1,0,0,0)|\psi\rangle\,. \tag{8.7}$$

Following Ref. 1 we will use a Grassman–Gaussian as ansatz:

$$e^{\bar\theta_-^{(-)}\Gamma\theta_-^{(-)}}, \tag{8.8}$$

with $\Gamma = \lambda^{ij}\Gamma^+\Gamma_{ij}$, the exponent is the most general bilinear in $\theta_-^{(-)}$ as defined in (3.26).

Applying $D_+^{(+)}$ [given in (3.27)] to (8.8) we get

$$D_+^{(+)\alpha}e^{\bar\theta_-^{(-)}\Gamma\theta_-^{(-)}} = 2i(\Gamma\theta_-^{(-)})^\alpha e^{\bar\theta_-^{(-)}\Gamma\theta_-^{(-)}} - (i/4)P_+(\Gamma^+\theta_-^{(-)})^\alpha e^{\bar\theta_-^{(-)}\Gamma\theta_-^{(-)}}. \tag{8.9}$$

We will now show that the solution of (8.7) is given by (8.8) with $\lambda^{ij} = a\delta_1^{[i}\delta_2^{j]}$ and will determine the value of a. In this case Eq. (8.9) becomes

$$\begin{aligned} D_+^{(+)\alpha}e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} &= 2ia(\Gamma^+\Gamma_{1,2}\theta_-^{(-)})^\alpha e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} \\ &\quad - (i/4)P_+(\Gamma^+\theta_-^{(-)})^\alpha e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}}. \end{aligned} \tag{8.10}$$

Next we will split $D_+^{(+)\alpha}$ with the projection operators,

$$\Pi_1{}^\pm = \tfrac{1}{2}(I \pm i\Gamma_{1,2}) \tag{8.11}$$

and define

$$d^\pm = \Pi_1{}^\pm D_+^{(+)}. \tag{8.12}$$

They satisfy the anticommutation relations

$$\begin{aligned} \{d^{+\alpha},d^{-\beta}\} &= -\tfrac{1}{2}P_+\{\Pi_1{}^+\Pi^{(+)}\Gamma^+C^{-1}\}^{\alpha\beta}, \\ \{d^{-\alpha},d^{+\beta}\} &= -\tfrac{1}{2}P_+\{\Pi_1{}^-\Pi^{(+)}\Gamma^+C^{-1}\}^{\alpha\beta}, \\ \{d^{+\alpha},d^{+\beta}\} &= \{d^{-\alpha},d^{-\beta}\} = 0\,, \end{aligned} \tag{8.13}$$

which show that d^- and d^+ are, respectively, the creation (raising) and annihilation (lowering) operators of a fermionic Clifford algebra. Now if we apply $\Pi_1{}^\pm$ to Eq. (8.10) we obtain

$$\begin{aligned} &d^-e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} \\ &\quad = (-2a - (i/4)P_+)\Gamma^+\Pi_1{}^-\theta_-^{(-)}e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}}, \\ &d^+e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} \\ &\quad = (2a - (i/4)P_+)\Gamma^+\Pi_1{}^+\theta_-^{(-)}e^{a\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}}. \end{aligned} \tag{8.14}$$

Choosing $a = \mp(i/8)P_+$ we derive

$$\begin{aligned} d^-e^{-(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} &= 0\,, \\ d^+e^{(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} &= 0\,. \end{aligned} \tag{8.15}$$

Therefore

$$\begin{aligned} H_1e^{\mp(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} &= \pm e^{\mp(i/8)P_+\bar\theta_-^{(-)}\Gamma_{1,2}\theta_-^{(-)}}, \\ H_Ie^{\mp(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} &= 0\,,\quad I = 2,3,4\,, \end{aligned} \tag{8.16}$$

so $e^{-(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}}$ and $e^{(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}}$ are the eigenstates corresponding to the highest and lowest weight, respectively. They are also the Clifford highest and lowest state, respectively, of the algebra spanned by d^+ and d^-. One can use either state to generate the rest of the representation. For definiteness we will use the highest state $e^{-(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}}$. We can obtain the remaining states by applying an even number of d^- to the highest state for a total of

$$\sum_{j=0}^{2}\binom{4}{2j} = 1 + 6 + 1 = 8$$

states. This is equivalent to use the lowering operators E of the SO(8) algebra in the Cartan basis since they are quadratic in d^-, as explained in Appendix C. We note that if we apply an odd number of d^- to the highest state we will get

$$\sum_{j=1}^{2}\binom{4}{2j-1} = 4 + 4 = 8$$

fermionic states (at the d level) which together will form the $[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ representation (see Sec. VI).

Of course in order to have a representation of the super-Poincaré algebra $SP_{10+}^{(+)}$, we must also apply all possible operators $Q_+^{(+)}$. These can also be split

$$q^\pm = \Pi_1{}^\pm Q_+^{(+)}, \tag{8.17}$$

which satisfy

$$q^\pm e^{\mp(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} = 0\,, \tag{8.18}$$

as opposed to Eq. (8.15). Therefore we have the expressions

$$\begin{aligned} \Phi_{[1]} = \sum_{k=0}^{4}\sum_{j=0}^{2} d^{+\alpha_1}\cdots d^{+\alpha_{2j}}q^{-\beta_1}\cdots \\ \times q^{-\beta_k}e^{-(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} \\ \times F_{\alpha_1\cdots\alpha_{2j}\beta_1\cdots\beta_k}(x) \end{aligned} \tag{8.19a}$$

and

$$\begin{aligned} \Phi_{[\frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]} = \sum_{k=0}^{4}\sum_{j=1}^{2} d^{+\alpha_1}\cdots d^{+\alpha_{2j-1}}q^{-\beta_1}\cdots \\ \times q^{-\beta_k}e^{-(i/8)P_+\bar\theta_-^{(-)}\Gamma^+\Gamma_{1,2}\theta_-^{(-)}} \\ \times F_{\alpha_1\cdots\alpha_{2j-1}\beta_1\cdots\beta_k}(x)\,, \end{aligned} \tag{8.19b}$$

each with $2^4 \times 8 = 128$ components.

In order to proceed further we note that

$$d^{\pm\alpha} = i\Pi_1^{\pm\alpha}{}_\beta\, e^{\pm(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-} \times \frac{\partial}{\partial\bar\theta^{(-)}_{-\beta}} e^{\mp(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-} \tag{8.20a}$$

and

$$q^{\pm\alpha} = i\Pi_1^{\pm\alpha}{}_\beta\, e^{\mp(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-} \times \frac{\partial}{\partial\bar\theta^{(-)}_{-\beta}} e^{\pm(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-}. \tag{8.20b}$$

Replacing this in Eq. (8.19a), we get

$$\Phi_{[1]} = \sum_{j=0}^{2}\sum_{k=0}^{4} i^{k+2j}\Pi_1^{+\alpha_1}{}_{\gamma_1}\cdots\Pi_1^{+\alpha_{2j}}{}_{\gamma_{2j}} \times\Pi_1^{-\beta_1}{}_{\delta_1}\cdots\Pi_1^{-\beta_k}{}_{\delta_k} e^{-(i/8)P_+\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-} \times H^{\gamma_1\cdots\gamma_{2j}\delta_1\cdots\delta_k}(\theta^{(-)}_-)\, F_{\alpha_1\cdots\alpha_{2j}\beta_1\cdots\beta_k}(x), \tag{8.21}$$

where here the Grassmann Hermite polynomials[1] are

$$H^{\gamma_1\cdots\gamma_n}(\theta^{(-)}_-) = e^{(i/4)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-}\frac{\partial}{\partial\bar\theta^{(-)}_{-\gamma_1}}\cdots\frac{\partial}{\partial\bar\theta^{(-)}_{-\gamma_n}} \times e^{-(i/4)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-}, \tag{8.22}$$

and, finally, after an obvious field redefinition, we can write

$$\Phi_{[1]} = e^{-(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-}\sum_{j=0}^{2}\sum_{k=0}^{4} i^{k+2j} \times H^{\alpha_1\cdots\alpha_{2j}\beta_1\cdots\beta_k}(\theta^{(-)}_-)\psi_{\alpha_1\cdots\alpha_{2j};\beta_1\cdots\beta_k}(x), \tag{8.23}$$

where the $\psi_{\alpha_1\cdots\alpha_{2j};\beta_1\cdots\beta_k}(x)$ satisfy

$$\begin{aligned}&\Pi_1^{-\alpha_p}{}_\gamma\psi_{\alpha_1\cdots\alpha_p\cdots\alpha_{2j};\beta_1\cdots\beta_k}(x) = 0, \quad p=1,\ldots,2j,\\ &\Pi_1^{+\beta_r}{}_\gamma\psi_{\alpha_1\cdots\alpha_{2j};\beta_1\cdots\beta_r\cdots\beta_k}(x) = 0, \quad r=1,\ldots,k.\end{aligned} \tag{8.24}$$

In a similar way

$$\Phi_{[\frac12\frac12\frac12\frac12\frac12]} = e^{-(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{1,2}\theta^{(-)}_-}\sum_{j=1}^{2}\sum_{k=0}^{4} i^{k+2j-1} \times H^{\alpha_1\cdots\alpha_{2j-1};\beta_1\cdots\beta_k}(\theta^{(-)}_-) \times\psi_{\alpha_1\cdots\alpha_{2j-1};\beta_1\cdots\beta_k}(x), \tag{8.25}$$

where the ψ fields satisfy the restrictions above.

From formulas (8.23) and (8.25) we see that also in ten dimensions the irreducible superfield are expansions in terms of the eigenfunctions of a Grassmann oscillator.[8]

The expressions (8.19) and (8.23) are asymmetrical in the sense that one of the elements of the Cartan subalgebra (namely H_1) is privileged. In order to obtain more symmetrical expressions we look for more general Gaussians. If we take

$$\chi_\pm^{(K)} = e^{\lambda^{ij}_{(K,\pm)}\bar\theta^{(-)}_-\Gamma^+\Gamma_{ij}\theta^{(-)}_-} \tag{8.26}$$

with

$$\lambda^{ij}_{(K,\pm)} = \mp(i/8)P_+\delta^{[i}_{2K-1}\delta^{j]}_{2K}, \quad K=1,2,3,4,$$

it is easy to show that

$$H_I\chi_\pm^{(K)} = U_{2I-1,2I}\chi_\pm^{(K)} = \pm\delta_I{}^K\chi_\pm^{(K)}. \tag{8.27}$$

Thus the $\chi_\pm^{(K)}$ give us all the eigenstates of the Cartan subalgebra corresponding to the weights in (8.5) and therefore span the whole representation [1].

The appropriate projection operators now are

$$\Pi_K{}^\pm = \tfrac12(I \pm i\Gamma_{2K-1,2K}), \tag{8.28}$$

which we can use to split $Q^{(+)}_+$.

$$q_K{}^\pm = \Pi_K{}^\pm Q^{(+)}_+ = i\Pi_K{}^\pm e^{\mp(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{2K-1,2K}\theta^{(-)}_-}\frac{\partial}{\partial\bar\theta^{(-)}_-} \times e^{\pm(i/8)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{2K-1,2K}\theta^{(-)}_-}. \tag{8.29}$$

The $q_K{}^\pm$ satisfy

$$q_K{}^\pm\chi_\pm^{(K)} = 0 \tag{8.30}$$

for every $K=1,\ldots,4$. There is no point in splitting $D^{(+)}_+$ since we have already considered all the states of the representation [1] whose weights are (8.5) and not only the one corresponding to the highest weight. Then the superfield $\Phi_{[1]}$ will be given by

$$\Phi_{[1]} = \sum_{K=1}^{4}\sum_{n=0}^{4} q_K^{-\alpha_1}\cdots q_K^{-\alpha_n}\chi_+^{(K)}F^{(K,+)}_{\alpha_1\cdots\alpha_n}(x) + \sum_{K=1}^{4}\sum_{n=0}^{4} q_K^{+\alpha_1}\cdots q_K^{+\alpha_n}\chi_-^{(K)}F^{(K,-)}_{\alpha_1\cdots\alpha_n}(x) \tag{8.31}$$

and using Eq. (8.29)

$$\Phi_{[1]} = \sum_{K=1}^{4}\sum_{n=0}^{4} i^n\Pi_K^{-\alpha_1}{}_{\beta_1}\cdots\Pi_K^{-\alpha_n}{}_{\beta_n}\chi_+^{(K)} \times H_+^{(K)\beta_1\cdots\beta_n}(\theta^{(-)}_-)F^{(K,+)}_{\alpha_1\cdots\alpha_n}(x) + \sum_{K=1}^{4}\sum_{n=0}^{4} i^n\Pi_K^{+\alpha_1}{}_{\beta_1}\cdots\Pi_K^{+\alpha_n}{}_{\beta_n}\chi_-^{(K)} \times H_-^{(K)\beta_1\cdots\beta_n}(\theta^{(-)}_-)F^{(K,-)}_{\alpha_1\cdots\alpha_n}(x), \tag{8.32}$$

where

$$H_\mp^{(K)\beta_1\cdots\beta_n}(\theta^{(-)}_-) = e^{\mp(i/4)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{2K-1,2K}\theta^{(-)}_-}\frac{\partial}{\partial\bar\theta^{(-)}_{-\beta_1}}\cdots\frac{\partial}{\partial\bar\theta^{(-)}_{-\beta_n}} \times e^{\pm(i/4)P_+\bar\theta^{(-)}_-\Gamma^+\Gamma_{2K-1,2K}\theta^{(-)}_-}. \tag{8.33}$$

Redefining indices

$$(K,\pm)\to k = \begin{cases}K, & \text{for } (K,+),\quad K=1,\ldots,4,\\ 8-K+1, & \text{for } (K,-),\quad K=1,\ldots,4,\end{cases} \tag{8.34}$$

we get finally

$$\Phi_{[1]} = \sum_{K=1}^{8}\sum_{j=0}^{4} i^n\chi^{(k)}H^{(k)\beta_1\cdots\beta_n}(\theta^{(-)}_-)\psi^{(k)}_{\beta_1\cdots\beta_n}(x), \tag{8.35}$$

where the fields $\psi^{(k)}_{\beta_1\cdots\beta_n}(x)$ must satisfy

$$\Pi_{(k)}{}^{\beta_p}{}_{\gamma}\psi^{(k)}_{\beta_1\cdots\beta_p\cdots\beta_n}(x)=0\,,\quad p=1,\ldots,n\,. \qquad (8.36)$$

IX. CONCLUSION

In this paper we have analyzed the massless representations of the super-Poincaré algebra with particular emphasis on the ten-dimensional case. We have used the Casimir approach to obtain explicit expressions for superfields which are irreducible under the corresponding little algebra.

Due to the fact that we have used the massless condition $P^2=0$ from the beginning, these are on-shell superfields: they carry physical field components but no auxiliary fields. A simple comparison of the number of components of $\phi_{[1]}$ with the number of degrees of freedom of the supergravity multiplet in ten dimensions[9] quickly illustrates this point. This is a well known fact.[10] In order to study the auxiliary field structure, one must relax the condition $P^2=0$, i.e., one must look at the massive case. In doing so, one must enlarge our superspace to include $\theta^{(-)}_{+}$ and in the expansion

$$\Phi(x,\theta^{(-)}_{-},\theta^{(-)}_{+}) = \sum_{n=0}^{8}\theta^{(-)\alpha_1}_{+}\cdots\theta^{(-)\alpha_n}_{+}\Phi^{(n)}_{\alpha_1\cdots\alpha_n}(x,\theta^{(-)}_{-})\,,$$

the superfield $\Phi^{(0)}(x,\theta^{(-)}_{-})$ carries the physical fields while the other ones carry the auxiliary fields.[10]

The techniques we have applied here to decompose the massless scalar chiral superfield $\phi(x,\theta^{(-)}_{-})$ into its irreducible components can be extended to the massive case and the results will be shown elsewhere. However, knowing the structure of the massless field must be important in its own right if one wants to know which irreducible pieces of the massive case are the off-shell extension of each massless piece and also if we want to write Lagrangians for them.

ACKNOWLEDGMENT

We appreciate helpful discussions with Professor R. Finkelstein.

APPENDIX A: FIERZ REARRANGEMENTS

We have the general Fierz identity,

$$\begin{aligned}\overline{Q}_1MQ_2\overline{Q}_3NQ_4 &= -\frac{1}{\Delta}\sum_j\lambda(j)\overline{Q}_1MO_jNQ_4\overline{Q}_3O_jQ_2\\ &+\frac{1}{\Delta}\sum_j\lambda(j)\overline{Q}_1MO_jNQ_4\,\mathrm{Tr}[O_j(2\xi_{23})C^{-1}]\\ &-\frac{1}{\Delta}\sum_j\lambda(j)\lambda'(j)\overline{Q}_1MO_jN(2\xi_{24})CO_jQ_3\\ &+\frac{1}{\Delta}\sum_j\lambda(j)\overline{Q}_1MO_jN(2\xi_{34})CO_jQ_2\,, \end{aligned}\qquad (\mathrm{A}1)$$

where $Q_1,\ldots,Q_4$ are Majorana spinors which satisfy $\{Q^{\alpha}_a,Q^{\beta}_b\}=2\xi^{\alpha\beta}_{ab}$, $a,b=1,2,3,4$, O_j is a complete set of $\Delta\times\Delta$ matrices which are orthogonal under trace,

$$\mathrm{Tr}[O_iO_j]=\Delta\lambda(j)\delta_{ij}\,,$$

and finally

$$(CO_j)^T=\lambda'(j)CO_j\,.$$

If we take the basis $\{O_j\}$ to be made out of Γ tensors, we have, in ten dimensions,

$$\overline{D}^{(+)}_{+}O_jD^{(+)}_{+}=0$$

unless

$$O_j=\Gamma_0,\Gamma_{d-1},\Gamma_0\Gamma_{ij},\Gamma_{d-1}\Gamma_{ij}$$

or $\Gamma_{(11)}$ times the same matrices. Therefore one can write

$$\begin{aligned}&\overline{D}^{(+)}_{+}MD^{(+)}_{+}\overline{D}^{(+)}_{+}ND^{(+)}_{+}\\ &=(1/2\Delta)\overline{D}^{(+)}_{+}M\Pi^{(+)}\Gamma^{+}\Gamma_{i_1i_2}ND^{(+)}_{+}\\ &\quad\times\overline{D}^{(+)}_{+}\Gamma^{-}\Gamma^{i_1i_2}D^{(+)}_{+}\\ &\quad-(P_+/4)\overline{D}^{(+)}_{+}M\Pi^{(+)}\Gamma^{+}ND^{(+)}_{+}\\ &\quad+(2P_+/\Delta)\overline{D}^{(+)}_{+}M\Pi^{(+)}\Gamma^{+}\Gamma_{i_1i_2}N\Gamma^{i_1i_2}D^{(+)}_{+}\,,\end{aligned}\qquad (\mathrm{A}2)$$

where we have used

$$\overline{D}^{(+)}_{+}\Gamma^{-}D^{(+)}_{+}=-(P_+/4)\Delta\,. \qquad (\mathrm{A}3)$$

Here Δ is the dimension of the Dirac matrices,

$$\Delta=32\,.$$

If we put $M=N=\Gamma^-$ in (A2), we get, after some arithmetic,

$$\overline{D}^{(+)}_{+}\Gamma^{-}\Gamma_{i_1i_2}D^{(+)}_{+}\overline{D}^{(+)}_{+}\Gamma^{-}\Gamma^{i_1i_2}D^{(+)}_{+} = -14\times(8P_+)^2 \qquad (\mathrm{A}4)$$

which immediately gives

$$C_1=-14\,. \qquad (\mathrm{A}5)$$

APPENDIX B: A PARTICULAR REPRESENTATION OF THE DIRAC ALGEBRA

Let us start by denoting with a subindex to the left the space-time dimension of the Γ algebra and the associated charge conjugation matrix. Thus

$${}_n\Gamma,\ {}_nC \text{ are } 2^{[n/2]}\times2^{[n/2]} \text{ matrices.}$$

Then a particular representation in ten dimensions is

$$C={}_8C\otimes i\sigma_2\,,\quad \Gamma_0={}_8I\otimes\sigma_1\,,$$
$$\Gamma_j={}_8\Gamma_j\otimes i\sigma_3\,,\quad j=1,\ldots,8\,,$$
$$\Gamma_9={}_8I\otimes(-i)\sigma_2\,,$$

where, if we keep descending,

$${}_8C={}_6C\otimes I\,,\quad {}_8\Gamma_1=I\otimes\sigma_1\,,$$
$${}_8\Gamma_l={}_6\Gamma_l\otimes i\sigma_2,\quad l=2,\ldots,7\,,$$
$${}_8\Gamma_8={}_6\Gamma_8\otimes(-i)\sigma_2,\quad {}_6\Gamma_8={}_6\Gamma_{2,\ldots,7}\,,$$

and

$${}_6C={}_4C\otimes(-i)\sigma_2\,,\quad {}_6\Gamma_2={}_4I\otimes i\sigma_1\,,$$
$${}_6\Gamma_m={}_4\Gamma_m\otimes\sigma_2\,,\quad m=3,\ldots,6\,,$$
$${}_6\Gamma_7={}_4I\otimes i\sigma_3\,,$$

with

$$ {}_4C = {}_2C \otimes I, \quad {}_2C = i\sigma_2, $$
$$ {}_4\Gamma_3 = {}_2I \otimes (-i)\sigma_1, \quad {}_4\Gamma_{\substack{4\\5}} = \sigma_{\substack{1\\2}} \otimes i\sigma_2, $$
$$ {}_4\Gamma_6 = {}_2I \otimes i\sigma_3 . $$

In this decomposition we have

$$ {}_8\Gamma_j{}^2 = {}_8I, \quad j = 1,...,8, $$
$$ {}_6\Gamma_l{}^2 = -{}_6I, \quad l = 2,...,7, $$
$$ {}_4\Gamma_m{}^2 = -{}_4I, \quad m = 3,...,6, $$
$$ {}_{2k}C_{2k}\Gamma_{2k}C^{-1} = (-)^k{}_{2k}\Gamma^T, \quad k = 1,...,5 . $$

The constants A and A' in

$$ (C\Gamma^-\Gamma_{i_1}{}^{i_2}\Pi^{(+)})_{[\alpha_1\alpha_2} \cdots (C\Gamma^-\Gamma_{i_4}{}^{i_1}\Pi^{(+)})_{\alpha_7\alpha_8]} $$
$$ = A\epsilon_{\alpha_1\cdots\alpha_8}, $$
$$ \epsilon_{i_1\cdots i_8}(C\Gamma^-\Gamma^{i_1i_2}\Pi^{(+)})_{[\alpha_1\alpha_2} \cdots (C\Gamma^-\Gamma^{i_7i_8}\Pi^{(+)})_{\alpha_7\alpha_8]} $$
$$ = A'\epsilon_{\alpha_1\cdots\alpha_8}, $$

can be computed by contracting the indices with

$$ ({}_6\Gamma_8\,{}_6C^{-1})^{\alpha_1\alpha_2} \cdots ({}_6\Gamma_8\,{}_6C^{-1})^{\alpha_7\alpha_8}. $$

The right-hand side gives then

$$ A \times 2^4 4!\,\mathrm{Pf}({}_6\Gamma_8\,{}_6C^{-1}) = 2^4 4!\,A $$

in the first case and

$$ A' 2^4 4!\,\mathrm{Pf}({}_6\Gamma_8\,{}_6C^{-1}) = 2^4 4!\,A' $$

in the second.

The left-hand sides can be computed independently by expanding the antisymmetrization. The calculation is quite tedious, but one finally gets

$$ 2^{11} \times 4! $$

in the first case and

$$ 2^{15} \times 4! $$

in the second, so that

$$ A = 2^7, \quad A' = 2^{11} . $$

Now we turn to $D^{(+)}_+$. Since it is a Majorana spinor, we have

$$ D^{(+)\alpha}_+ C_{\alpha\beta} = D^{(+)\dagger}_+{}_\gamma \Gamma_0{}^\gamma{}_\beta $$
$$ \Leftrightarrow D^{(+)\dagger}_+{}_\gamma = D^{(+)\alpha}_+ (C\Gamma_0)_{\alpha\gamma} . $$

This implies

$$ \{D^{(+)\alpha}_+, D^{(+)\dagger}_+{}_\beta\} = -P_+(\Pi^{(+)}\Pi_+)^\alpha{}_\beta . $$

In our representation

$$ \Pi_+ = \begin{pmatrix} {}_8I & 0 \\ 0 & 0 \end{pmatrix}, $$
$$ \Pi^{(+)} = \begin{pmatrix} \frac{1}{2}({}_8I + {}_8\Gamma_{(9)}) & 0 \\ 0 & \frac{1}{2}({}_8I - {}_8\Gamma_{(9)}) \end{pmatrix}, $$
$$ {}_8\Gamma_{(9)} = \begin{pmatrix} {}_6I & 0 \\ 0 & -{}_6I \end{pmatrix}, \quad \Pi_+\Pi^{(+)} = \begin{pmatrix} {}_6I & 0 & \\ 0 & 0 & 0 \\ & 0 & 0 \end{pmatrix}, $$
$$ C\Gamma_0 = \begin{pmatrix} {}_6I & 0 & \\ 0 & {}_6C & 0 \\ & 0 & 0 \end{pmatrix}. $$

So

$$ D^{(+)\dagger}_{+1} = -D^{(+)6}_+, \quad D^{(+)\dagger}_{+2} = D^{(+)5}_+, $$
$$ D^{(+)\dagger}_{+3} = -D^{(+)8}_+, \quad D^{(+)\dagger}_{+4} = D^{(+)7}_+ . $$

Therefore

$$ D^{(+)1\cdots 8}_+ = D^{(+)16}_+ D^{(+)25}_+ D^{(+)38}_+ D^{(+)47}_+ $$
$$ = \prod_{i=1}^{4} \frac{1}{2}[D^{(+)i}_+, D^{(+)\dagger}_{+i}], $$

as claimed in the text.

APPENDIX C: EXPRESSIONS FOR THE SO(8) OPERATOR IN THE CARTAN BASIS

Let us start by defining the operators

$$ \Pi_J^{(\eta)} = \tfrac{1}{2}(I + i\eta\Gamma_{2J-1,2J}), \quad J = 1,...,4, \quad \eta = \pm 1, \tag{C1} $$

and the product

$$ P_{\eta_1\eta_2\eta_3\eta_4} = \prod_{J=1}^{4} \Pi_J^{(-\eta_J)} . \tag{C2} $$

With these we can define

$$ d_{\eta_1\eta_2\eta_3\eta_4} = P_{\eta_1\eta_2\eta_3\eta_4} D^{(+)}_+ . \tag{C3} $$

Then, from the commutation relation,

$$ [U_{ij}, D^{(+)\alpha}_+] = -(i/2)(\Gamma_{ij}D^{(+)}_+)^\alpha \tag{C4} $$

one can easily show

$$ [H_J, d_{\eta_1\cdots\eta_4}] = \tfrac{1}{2}\eta_J d_{\eta_1\cdots\eta_4}, \quad J = 1,...,4, \tag{C5} $$

where H_J are the operators of the Cartan subalgebra (8.3). Clearly $d_{\eta_1\cdots\eta_4}$ increases or decreases the eigenvalues of H_J by $\frac{1}{2}$ depending on the sign of η_J. In fact, from (8.11), we see that $d_{-\pm\pm\pm}$ can be considered the components of d^- and $d_{+\pm\pm\pm}$ the components of d^+. But these are too many since we know that d^+ and d^- have four components each. So half of $d_{\eta_1\cdots\eta_4}$ must vanish identically in our $\theta^{(-)}_-$ superspace. Indeed

$$ \Gamma^{1\cdots 8} = -\Gamma_{(11)}\Gamma_0\Gamma_9 \Leftrightarrow \Gamma^{1\cdots 8}\Pi^{(+)}\Pi_+ = -\Pi^{(+)}\Pi_+ . \tag{C6} $$

On the other hand

$$ \Gamma^{1\cdots 8}P_{\eta_1\cdots\eta_4} = \eta P_{\eta_1\cdots\eta_4}, \quad \eta = \eta_1\eta_2\eta_3\eta_4 . \tag{C7} $$

Therefore

$$ d_{\eta_1\cdots\eta_4} = P_{\eta_1\cdots\eta_4}\Pi^{(+)}\Pi_+ D $$
$$ = \tfrac{1}{2}(1 + \eta\Gamma^{1\cdots 8})\tfrac{1}{2}(1 - \Gamma^{1\cdots 8})P_{\eta_1\cdots\eta_4}\Pi^{(+)}\Pi_+ D \tag{C8} $$

and we conclude

$$ d_{\eta_1\cdots\eta_4} = 0 \text{ when } \eta = +1 . \tag{C9} $$

The operators corresponding to the root vectors in the Cartan basis of the SO(8) algebra are given by

$$ E(\eta\hat{e}_I + \eta'\hat{e}_J) $$
$$ = \tfrac{1}{2}(U_{2I-1,2J-1} + i\eta U_{2I,2J-1} + i\eta' U_{2I-1,2J} - \eta\eta' U_{2I,2J}), \quad I, J = 1,2,3,4 . \tag{C10} $$

Making use of Eq. (5.3), (C10) then becomes

$$E(\eta\hat{e}_I+\eta'\hat{e}_J) = -(i/8P_+)\bar{D}^{(+)}_{+}\Gamma^-\tilde{\Gamma}_I^{(\eta)}\tilde{\Gamma}_J^{(\eta')}D^{(+)}_{+}, \tag{C11}$$

where

$$\tilde{\Gamma}_I^{(\eta)}=(1/\sqrt{2})(\Gamma_{2I-1}+i\eta\Gamma_{2I}),\quad \eta=\pm 1,\quad I=1,...,4. \tag{C12}$$

The $\tilde{\Gamma}_I^{(\eta)}$ matrices anticommute among themselves and also satisfy

$$\tilde{\Gamma}_I^{(\eta)}\Pi_I^{(\xi)}=\delta_{\eta,-\xi}\tilde{\Gamma}_I^{(\eta)},\quad \Pi_I^{(\xi)}\tilde{\Gamma}_I^{(\eta)}=\delta_{\eta,\xi}\tilde{\Gamma}_I^{(\eta)}. \tag{C13}$$

Using (C13) and the orthogonality of $\Pi_I^{(\xi)}$, we can express $E(\eta\hat{e}_I+\eta'\hat{e}_J)$ in terms of $d_{\xi_1\cdots\xi_4}$. Let us work out, for instance, $E(\eta\hat{e}_1+\eta'\hat{e}_2)$. Note that

$$D^{(+)}_{+}=\sum_{\xi_i=\pm 1}\Pi_1^{(\xi_1)}\cdots\Pi_4^{(\xi_4)}D^{(+)}_{+}. \tag{C14}$$

We have

$$\begin{aligned}E(\eta\hat{e}_1+\eta'\hat{e}_2) &= -\frac{i}{8P_+}\left[\sum_{\xi_i}\Pi_1^{(\xi_1)}\cdots\Pi_4^{(\xi_4)}D^{(+)}_{+}\right]^T C\Gamma^-\tilde{\Gamma}_1^{(\eta)}\tilde{\Gamma}_2^{(\eta')}\left[\sum_{\xi_i'}\Pi_1^{(\xi_1')}\cdots\Pi_4^{(\xi_4')}D^{(+)}_{+}\right]\\ &= -\frac{i}{8P_+}\sum_{\xi_i}\sum_{\xi_i'}\bar{D}^{(+)}_{+}\Pi_1^{(\xi_1)}\cdots\Pi_4^{(\xi_4)}\Gamma^-\tilde{\Gamma}_1^{(\eta)}\tilde{\Gamma}_2^{(\eta')}\Pi_1^{(\xi_1')}\cdots\Pi_4^{(\xi_4')}D^{(+)}_{+}\\ &= -\frac{i}{8P_+}\sum_{\xi_i}\sum_{\xi_i'}\delta_{\eta,\xi_1}\delta_{\eta',\xi_2}\delta_{\eta,-\xi_1'}\delta_{\eta',\xi_2'}\delta_{\xi_3,\xi_3'}\delta_{\xi_4,\xi_4'}\\ &\quad\times\bar{D}^{(+)}_{+}\Pi_1^{(\xi_1)}\cdots\Pi_4^{(\xi_4)}\Gamma^-\tilde{\Gamma}_1^{(\eta)}\tilde{\Gamma}_2^{(\eta')}\Pi_1^{(\xi_1')}\cdots\Pi_4^{(\xi_4')}D^{(+)}_{+}\\ &= -\frac{i}{8P_+}\sum_{\xi_3=\pm 1}\sum_{\xi_4=\pm 1}\bar{d}_{\eta\eta'\xi_3\xi_4}\Gamma^-\tilde{\Gamma}_1^{(\eta)}\tilde{\Gamma}_2^{(\eta')}d_{\eta\eta'-\xi_3-\xi_4}.\end{aligned} \tag{C15}$$

Let us consider in particular $\eta=1$, $\eta'=+1$. Then, recalling (C9), we obtain

$$\begin{aligned}E(-\hat{e}_1+\hat{e}_2)=E(-1,1,0,0,0) &= -(i/8P_+)(C\Gamma^-\tilde{\Gamma}_1^{(-)}\tilde{\Gamma}_2^{(+)})_{\alpha\beta}(d^\alpha_{-+++}\,d^\beta_{-+--}+d^\alpha_{-+--}\,d^\beta_{-+++})\\ &= -(i/8P_+)(C\Gamma^-\tilde{\Gamma}_1^{(-)}\tilde{\Gamma}_2^{(+)})_{\alpha\beta}[d^\alpha_{-+++},d^\beta_{-+--}].\end{aligned} \tag{C16}$$

The E operators corresponding to other root vectors can be computed similarly. Schematically, they are given by

$$E(r_1r_2r_3r_4)\sim[d_{\eta_1\eta_2\eta_3\eta_4},d_{\xi_1\xi_2\xi_3\xi_4}],\quad r_I=\tfrac{1}{2}(\eta_I+\xi_I),\quad I=1,...,4. \tag{C17}$$

[1]R. Finkelstein and M. Villasante, J. Math. Phys. **27**, 1595 (1986).
[2]M. Villasante, Lett. Math. Phys. **11**, 351 (1986).
[3]A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); E. Sokatchev, *ibid.* **99**, 96 (1975); J. G. Taylor, *ibid.* **169**, 484 (1980); J. Kim, J. Math. Phys. **25**, 2037 (1984).
[4]S. Ferrara and C. Savoy, in *Supergravity 1981*, Proceedings of Conference held in Trieste, Italy, 1981, edited by S. Ferrara and J. G. Taylor (Cambridge U.P., Cambridge, 1982), p. 151. See also the review: P. Fayet and S. Ferrara, Phys. Rep. **32**, 249 (1977).
[5]J. Scherk, in *Recent Developments in Gravitation*, Proceedings of a Summer Institute held in Cargèse, France, 1978, edited by M. Lévy and S. Deser, NATO Advanced Study Institute, Series B, Vol. 44 (Plenum, New York, 1979), p. 479; P. van Nieuwenhuizen, in *Supergravity 1981*, Proceedings of Conference held in Trieste, Italy, 1981, edited by S. Ferrara and J. G. Taylor (Cambridge U.P., Cambridge, 1982), p. 151; R. Finkelstein and M. Villasante, Phys. Rev. D **31**, 425 (1985).
[6]The original article is in Italian and is G. Racah, "Sulla Caratterizzazione delle Rappresentazioni Irriducibili dei Gruppi Semisemplici di Lie," Lincei Rend. Sci. Fis. Mat. Nat. **8**, 108 (1950); see also, B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974); H. Bacry, *Lectures on Group Theory and Particle Theory* (Gordon and Breach, New York, 1977).
[7]A. M. Perelomov and V. S. Popov, JETP Lett. **1**, 160 (1965); **2**, 20 (1965); Sov. Mat. Dokl. **8**, 631 (1967).
[8]R. Finkelstein and M. Villasante, Phys. Rev. D **33**, 1666 (1986).
[9]B. E. W. Nilsson, Nucl. Phys. B **188**, 176 (1981).
[10]S. J. Gates, M. T. Grisaru, M. Roček, and W. Siegel, *Superspace or One Thousand and One Lessons in Supersymmetry* (Benjamin/Cummings, Reading, MA, 1983).

Dynamic structure and embedded representation in physics: The group theory of the adiabatic approximation

D. J. Rowe and P. Rochford
Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7
J. Repka
Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

(Received 11 August 1987; accepted for publication 11 November 1987)

The group theoretical concepts of embedded representations and dynamical structure groups, distinct from dynamical symmetry groups, are introduced in order to describe the common physical situation in which collective bands of states of a many-body system are well described by an algebraic collective model even though the states may not span an invariant subspace of the many-body Hilbert space.

I. INTRODUCTION

A symmetry group of a quantum mechanical system is a group made up of transformations of the Hilbert space that commute with the Hamiltonian. Thus the degenerate eigenspaces of the Hamiltonian carry representations (not necessarily irreducible) of the symmetry group.

The value of symmetry groups in physics is well understood. It is also well-known that group and algebraic structures play vital roles under more general circumstances. In particular, dynamical symmetry groups, which do not commute with the Hamiltonian, have recently attracted widespread interest.[1]

A dynamical symmetry group of a quantum mechanical system is a group of transformations of the Hilbert space whose irreducible subspaces are invariant under the action of the Hamiltonian. Dynamical symmetry groups of interest often contain full symmetry groups as subgroups and hence irreps that, in general, contain sequences (bands) of subirreps of the symmetry subgroup. Thus the irreducible subspaces of a dynamical symmetry group need not be eigenspaces of the Hamiltonian. A familiar example is the n-dimensional harmonic oscillator which has $\mathrm{Sp}(n,\mathbb{R})$ as dynamical group and $\mathrm{su}(n)$ as symmetry group.

We point out here that even this generalization can be further extended with advantage. We shall define what, for want of a better name, we call simply a *dynamical structure* group. The essential ideas underlying the concept are familiar in physics in the context of the adiabatic approximation, but, as far as we are aware, the associated group theory has not been discussed. To illustrate, consider the rotational states of a diatomic molecule. The question arises as to whether or not an observed band of rotational states of the molecule spans an irreducible representation of a suitably defined rotor algebra. It will be shown in this paper that nonequivalent irreps of a rotor algebra are distinguished by distinct deformation shapes of the system. In the simple rotor model, the deformation of a system is defined by its quadrupole moments. A system with a well-defined deformation is then one with fixed quadrupole moments in the body-fixed (principal axes) frame. However, in practical situations there are inevitably vibrational shape fluctuations and, as a consequence, the physical states of the system have a distribution of deformations. It follows that they straddle a corresponding distribution of irreps. Nevertheless, if the vibrational frequencies are large in comparison with the rotational frequencies, it can happen that the distribution of deformation shapes of the system is very slightly perturbed by the rotational motion, and that, in the adiabatic limit, a simple rotational structure is maintained. One may then observe sets of states of constant intrinsic structure (i.e., constant distributions of deformation shapes) that are meaningfully described as rotational bands and that, in isolation, are indistinguishable from states of an irrep of the rotor algebra.

In such a situation, it is evident that the rotor algebra is playing a vital dynamical role in describing the relative properties of rotational states even though the structure of every state may be described very poorly by the states of any single irrep of the rotor algebra. We therefore introduce the concept of an *embedded representation* which expresses this phenomenon in precise algebraic terms.

The admittance of embedded representations opens up the possibility of applying algebraic techniques to much more general situations than was hitherto recognized, e.g., to situations where one has neither a full symmetry nor a dynamical symmetry of the Hamiltonian, but where there is an adiabatic decoupling of collective and intrinisic degrees of freedom to such an extent that it is a good approximation to freeze the intrinsic structure in a description of the relative properties of low-lying collective states.

The analysis of such dynamical structure and embedded representations also throws light on the interpretation of the significance of a model's success in explaining a limited set of physical data. The traditional approach to the interpretation of physical phenomena is to make models that fit the known data and then to use them to make further predictions which can be subjected to experimental test. In doing this it is clearly important to focus on data which provide significant tests and which can distinguish different models. The existence of embedded representations implies that many predictions follow from the relative dynamical structure of states and, hence, that agreement of an algebraic model's predictions with the data may not imply the existence of a full dynamical symmetry. In other words, the relative properties of a number of states in isolation may be the same as they would be if the states were to belong to an irreducible subrepresentation

 0022-2488/88/030572-06$02.50

of a Lie algebra, whereas, in fact, they only belong to an embedded representation.

Similar observations have been used to formulate successful theories of relative dynamical structure, such as the equations-of-motion formalism.[2] They also explain why models are often much more successful than they superficially have any right to be. For example, many theories based on independent-particle approximations, such as the Hartree–Fock and random-phase approximations, are remarkably successful even in situations where the independent-particle approximation has little reason to be good.

The main objective of this paper is to analyze the circumstances under which a Lie algebra can exhibit embedded representations. We shall analyze in some detail the basic rotor and vibrator algebras, which are of fundamental importance in the theory of many-particle collective structure.

II. COLLECTIVE MOTIONS OF A MANY-BODY SYSTEM

Enormous simplification can be achieved in the treatment of the collective states of a many-body system if the variables can be separated into subsets of collective and orthogonal intrinsic coordinates and the many-particle Hilbert space factored into a direct product of collective and intrinsic Hilbert subspaces. This is possible, for example, for center-of-mass motion and, as a consequence, the treatment of many-particle center-of-mass motion is trivial.

However, for collective motions in general it is not possible. Nevertheless, one frequently observes bands of collective states that are well described by collective models expressed in terms of relatively small numbers of collective degrees of freedom. Such models are usually justified by the argument that collective motions are slow (adiabatic) in comparison to the more rapid intrinsic motions. As a consequence, the intrinsic structure of the system may be very little perturbed by the collective motions.

For the purposes of this analysis, a collective model is defined as a triple $(H^{\text{COLL}},\mathbf{g},\Gamma^{\text{COLL}})$ of a model Hamiltonian H^{COLL} acting on a Hilbert space $\mathbb{H}^{\text{COLL}}$, a dynamical Lie algebra $\mathbf{g}$ of collective observables, and a unitary representation (Γ^{COLL}) of $\mathbf{g}$ carried by $\mathbb{H}^{\text{COLL}}$.

A model will be called "simple" if the representation (Γ^{COLL}) is irreducible.

The spectrum of H^{COLL} is said to consist of collective bands, where a band is a set of states belonging to a common irrep of the dynamical collective algebra. Thus, by definition, a simple collective model features a single collective band.

To understand the success of a collective model, one seeks to embed the states of the collective model in the microscopic many-particle Hilbert space of the system.

Let $\mathbb{H}$ be the microscopic Hilbert space for the system and suppose that it carries a unitary representation Γ of $\mathbf{g}$. Let $\Gamma = \Sigma_\lambda \Gamma^{(\lambda)}$ be a direct sum of irreps. Let H be the microscopic Hamiltonian for the system and let $H^{(\lambda)}$ be its projection to the irreducible subspace $\mathbb{H}^{(\lambda)}$ for the irrep $\Gamma^{(\lambda)}$. Let $|\lambda\alpha\rangle$ be an eigenstate of $H^{(\lambda)}$ of eigenvalue $E^0_{\lambda\alpha}$. We may assume that the set of states $\{|\lambda\alpha\rangle\}$ defines an orthonormal basis for the Hilbert space. The Hamiltonian can then be expressed

$$H = H_0 + V,$$

where

$$H_0 = \sum_\lambda H^{(\lambda)}$$

and V is defined by its matrix elements

$$V_{\lambda\alpha,\lambda'\beta} = \begin{cases} \langle\lambda\alpha|H|\lambda'\beta\rangle, & \lambda \neq \lambda', \\ 0, & \lambda = \lambda'. \end{cases}$$

Since the irreducible $\mathbb{H}^{(\lambda)}$ subspaces are invariant under the action of the Hamiltonian H_0, it follows, by definition, that $\mathbf{g}$ is a dynamical algebra for H_0. Furthermore, if ϵ^0_λ is some conveniently defined intrinsic energy for $H^{(\lambda)}$ (e.g., for a ladder representation, the expectation value of H_0 in the lowest weight state), then collective energies $\mathscr{E}_{\lambda\alpha}$ can be defined by

$$E^0_{\lambda\alpha} = \epsilon^0_\lambda + \mathscr{E}_{\lambda\alpha},$$

and the Hamiltonian H_0 can be expressed

$$H_0 = H^{\text{COLL}} + H^{\text{INTR}},$$

where H^{COLL} and H^{INTR} are defined by their matrix elements

$$\langle\lambda\alpha|H^{\text{COLL}}|\lambda'\beta\rangle = \delta_{\lambda\lambda'}\delta_{\alpha\beta}\mathscr{E}_{\lambda\alpha},$$

$$\langle\lambda\alpha|H^{\text{INTR}}|\lambda'\beta\rangle = \delta_{\lambda\lambda'}\delta_{\alpha\beta}\,\epsilon^0_\lambda.$$

The full Hamiltonian is then

$$H = H^{\text{COLL}} + H^{\text{INTR}} + V.$$

It is noteworthy that this decomposition of the Hamiltonian is obtained without reference to collective or intrinsic variables and without factorization of the Hilbert space.

When there is a multiplicity of equivalent irreps, V depends on the particular combinations selected. If V can be made negligible, $\mathbf{g}$ is a dynamical symmetry algebra for the system and we derive the collective model.

An interesting question now arises as to whether or not the absence of irrep mixing is an essential condition for the observation of pure collective states. Is it possible to embed collective model states in the microscopic Hilbert space in a way that admits the possibility of large irrep mixing interactions?

III. DEFINITION OF AN EMBEDDED REPRESENTATION

Let $\mathbf{g}$ be a Lie algebra and Γ a (reducible) representation of $\mathbf{g}$ on a Hilbert space $\mathbb{H}$. Let Γ_P be the projection of Γ onto a subspace $\mathbb{H}_0$ of $\mathbb{H}$; i.e., if $X\in\mathbf{g}$ and $\{|\alpha\rangle\}$ is an orthonormal basis for $\mathbb{H}$, defined such that a subset of basis vectors is a basis for $\mathbb{H}_0$, then, for any $|\alpha\rangle\in\mathbb{H}_0$,

$$\Gamma_P(X)|\alpha\rangle = \sum_{\beta\in\mathbb{H}_0} |\beta\rangle\langle\beta|\Gamma(X)|\alpha\rangle.$$

If Γ_P is a representation of $\mathbf{g}$, we call it an embedded representation. Note, however, that an embedded representation is not generally a subrepresentation.

For example, given a representation Γ of $\mathbf{g}$, we may define a set of submatrices Γ_P; i.e., for each matrix $\Gamma(X)$, $X\in\mathbf{g}$, we define a submatrix $\Gamma_P(X)$ such that

$$\Gamma(X)=\begin{pmatrix}\Gamma_P(X) & * & * \\ * & * & * \\ * & * & *\end{pmatrix}.$$

If Γ_P is a representation, then it is an embedded representation.

If, relative to some basis, Γ is of the block form

$$\Gamma(X)=\begin{pmatrix}\Gamma_P(X) & * & * \\ 0 & * & * \\ 0 & 0 & *\end{pmatrix},$$

for all $X\in\mathbf{g}$, then Γ_P is a subrepresentation of Γ.

If Γ is of the block form

$$\Gamma(X)=\begin{pmatrix}* & * & * \\ 0 & \Gamma_P(X) & * \\ 0 & 0 & *\end{pmatrix},$$

for all $X\in\mathbf{g}$, then Γ_P is the quotient of two subrepresentations of Γ, sometimes called a *subquotient representation*.

Any subrepresentation or subquotient is an embedded representation. It is possible, however, to find examples of embedded representations which are not subrepresentations or subquotients, but the requirement that the submatrices $\{\Gamma_P(X);\ X\in\mathbf{g}\}$ be a representation is a strong condition.

IV. THE ROTOR GROUP AND ITS LIE ALGEBRA

A dynamical group for the rigid rotor is the semidirect product $[\mathbb{R}^5]\mathrm{SO}(3)$ of an Abelian normal subgroup $\mathbb{R}^5$ and the rotation group SO(3). A group element is a pair $(e^{i\omega},\Omega)$, where ω is an element of the $\mathbb{R}^5$ Lie algebra and Ω is in SO(3).

The $\mathbb{R}^5$ Lie algebra is spanned by a set $\{Q_\nu;\ \nu=0,\pm 1,\pm 2\}$ of quadrupole moments which transform under rotations as the components of an $L=2$ spherical tensor, i.e.,

$$Q_\nu\to\Omega\cdot Q_\nu=\sum_\mu Q_\mu\,\mathscr{D}^2_{\mu\nu}(\Omega),\quad \Omega\in\mathrm{SO}(3),$$

where $\mathscr{D}^2$ is an $L=2$ Wigner rotation matrix.

The group product is given by

$$(e^{i\omega_1},\Omega_1)\cdot(e^{i\omega_2},\Omega_2)=(e^{i(\omega_1+\Omega_1\cdot\omega_2)},\Omega_1\Omega_2).$$

A. Unitary irreps of the rotor algebra

Unitary irreps of the rotor group and its Lie algebra are easily derived by Mackey's theory of induced representations,[3] as first shown explicitly, to our knowledge, by Weaver, Biedenharn, and Cusson.[4]

Let the elements of $\mathbb{R}^5$ act multiplicatively on the wave functions of a Hilbert space of square integrable functions of the coordinates of a many-body system by

$$Q_\nu\Psi(x)=Q_\nu(x)\Psi(x),$$

where $Q_\nu(x)$ is a quadrupole moment of the many-particle configuration x. Let SO(3) act by a representation $\mathscr{R}$.

Since $\mathbb{R}^5$ is Abelian, its irreps are all one-dimensional and are carried by a single eigenstate $|q\rangle$ of the commuting $\mathbb{R}^5$ basis operators, i.e.,

$$Q_\nu|q\rangle=q_\nu|q\rangle.$$

Let

$$|q^\Omega\rangle=\mathscr{R}(\Omega)|q\rangle,\quad \Omega\in\mathrm{SO}(3).$$

Since

$$\mathscr{R}(\Omega)Q_\nu|q\rangle=\sum_\mu Q_\mu\,\mathscr{D}^2_{\mu\nu}(\Omega)|q^\Omega\rangle=q_\nu|q^\Omega\rangle,$$

it follows that

$$Q_\mu|q^\Omega\rangle=q^\Omega_\mu\,|q^\Omega\rangle,$$

with

$$q^\Omega_\mu=\sum_\nu \mathscr{D}^{2}_{\mu\nu}{}^*(\Omega)q_\nu.$$

The set of states $\{|q^\Omega\rangle;\ \Omega\in\mathrm{SO}(3)\}$ constitutes an SO(3) orbit. We pick a representative point on this orbit having the property that $q_{\pm 1}=0$, $q_2=q_{-2}$. We let this point be the reference state $|q\rangle$ and refer to it as the intrinsic state. The nonvanishing quadrupole moments q_0, q_2 at the representative point are likewise referred to as the intrinsic quadrupole moments.

The isotropy subgroup of the intrinsic state is the subset of SO(3) rotations that leave $|q\rangle$ invariant up to phase. If q_2 is nonzero, the isotropy subgroup is D_2, the group generated by rotations through angle π about the principal axes. If $q_2=0$ but $q_0\neq 0$, the isotropy subgroup is D_∞, i.e., the group of all rotations about the symmetry axis plus rotations though π about any perpendicular axis. The only other possibility is that all the quadrupole moments are zero, in which case the intrinsic state is rotationally invariant and the isotropy subgroup is the full SO(3) group.

We consider first the generic case. It is well-known that D_2 has four irreps, all one-dimensional, and labeled by ϵ_1, $\epsilon_2=\pm 1$. Thus the corresponding one-dimensional irreps of the semidirect product group $[\mathbb{R}^5]D_2$ have four labels (q_0, q_2, ϵ_1, ϵ_2) and satisfy

$$\mathscr{R}(\pi,0,0)|q\rangle=\epsilon_1|q\rangle,\quad \mathscr{R}(0,\pi,0)|q\rangle=\epsilon_2|q\rangle.$$

Orthonormal basis vectors for the generic irreps of $[\mathbb{R}^5]\mathrm{SO}(3)$ are given by the set of all state vectors of the form

$$|qJMK\rangle=\int d\Omega|q^\Omega\rangle\Psi^J_{MK}(\Omega),$$

where

$$\langle q^\Omega|q'^{\Omega'}\rangle=\delta(q-q')\delta(\Omega-\Omega'),$$

$$\Psi^J_{MK}(\Omega)=\sqrt{\frac{2J+1}{16\pi^2(1+\delta_{K0})}}\times[\mathscr{D}^{J}_{MK}{}^*(\Omega)+\epsilon_2(-1)^{J+K}\mathscr{D}^{J}_{M-K}{}^*(\Omega)],$$

with J, M, K integers and K restricted to either all even or all odd values such that $(-1)^K=\epsilon_1$. Wave functions representing these state vectors are given, in Dirac notation, by

$$\Psi^J_{MK}(\Omega)=\langle q^\Omega|qJMK\rangle.$$

Finally, from the actions of $\mathbb{R}^5$ and SO(3) on $|q^\Omega\rangle$, given above, we obtain the induced representations of the $[\mathbb{R}^5]\mathrm{SO}(3)$ operators,

$$\Gamma(Q_\mu)\Psi^J_{MK}(\Omega) = \sum_\nu q_\nu \mathscr{D}^2_{\mu\nu}{}^*(\Omega)\Psi^J_{MK}(\Omega),$$

$$\Gamma(\omega)\Psi^J_{MK}(\Omega) = \Psi^J_{MK}(\omega^{-1}\Omega).$$

For an axially symmetric representation, i.e., $q_2 = 0$, we observe that the irreps of $[\mathbb{R}^5]D_\infty$ are labeled by (q_0, K, ϵ_2), where K is a positive or zero integer and $\epsilon_2 = \pm 1$. Basis functions for the corresponding induced irrep of $[\mathbb{R}^5]\mathrm{SO}(3)$ are then given by the set of all wave functions of the form $\Psi^J_{MK}(\Omega)$, but with K now held fixed.

The constancy of K for an axially symmetric irrep is easily understood, because if $q_2 = 0$, then

$$\Gamma(Q_\mu) = q_0 \mathscr{D}^2_{\mu 0}{}^*(\Omega),$$

and there is no mechanism for connecting wave functions Ψ^J_{MK} and $\Psi^{J'}_{M'K'}$ with $K \neq K'$.

For the same reason, one easily shows that the spherically symmetric irreps, for which all quadrupole moments are zero, are characterized by fixed values of J. A spherically symmetric irrep of the $[\mathbb{R}^5]\mathrm{SO}(3)$ group is an irrep of SO(3) and a trivial identity representation of $\mathbb{R}^5$.

B. Embedded representations of the rotor algebra

If we start with a generic irrep $(q_0, q_2, \epsilon_1, \epsilon_2)$ of $[R^5]\mathrm{SO}(3)$ and restrict to a subspace of states of $K = \text{const}$, we immediately see by inspection of the results of the last section that $\Gamma(Q_\mu)$ projects

$$\Gamma(\mathrm{Q}_\mu) \to \Gamma_P(\mathrm{Q}_\mu) = q_0 \mathscr{D}^2_{\mu 0}{}^*(\Omega).$$

Thus the subspace carries an embedded representation (q_0, K, ϵ_2) of $[\mathbb{R}^5]\mathrm{SO}(3)$.

This extremely simple result is already of considerable physical significance. For it expresses the known result that, in isolation, bands of states of a triaxial rotor having K as a good quantum number are indistinguishable from axially symmetric rotor states. In this context we recall that bands of states of constant K naturally occur for a triaxial rotor if two of its principal moments of inertia accidentally happen to be equal, which can happen, as pointed out by Meyer ter Vehn,[5] even though $q_2 \neq 0$.

Consider next the situation in which, instead of a single $[\mathbb{R}^5]\mathrm{SO}(3)$ irrep, we have a direct sum of a distribution of irreps and a basis for the carrier space given by state vectors

$$|nJMK\rangle = \int d\Omega \int dq\, \phi_n(q)|q^\Omega\rangle \Psi^J_{MK}(\Omega),$$

where $\{\phi_n(q)\}$ is a set of weight functions and dq is any suitable measure such that

$$\int dq\, \phi^*_m(q)\phi_n(q) = \delta_{mn}.$$

We now have basis wave functions, again in Dirac notation, of the form

$$\Psi^J_{nMK}(q,\Omega) = \phi_n(q)\Psi^J_{MK}(\Omega).$$

The action of $[\mathbb{R}^5]\mathrm{SO}(3)$ on these wave functions is defined by

$$\Gamma(Q_\mu)\Psi^J_{nMK}(q,\Omega) = \sum_\nu q_\nu \mathscr{D}^2_{\mu\nu}{}^*(\Omega)\Psi^J_{nMK}(q,\Omega),$$

$$\Gamma(w)\Psi^J_{nMK}(q,\Omega) = \Psi^J_{nMK}(q,\omega^{-1}\Omega).$$

Now restrict to the subspace of states of $n = \text{const}$. This restriction can be expressed as a projection

$$\Psi(q,\Omega) \to \Psi(\Omega) = \int dq\, \phi^*_n(q)\Psi(q,\Omega),$$

under which

$$\Psi^J_{nMK}(q,\Omega) \to \Psi^J_{MK}(\Omega).$$

Under this projection, $\Gamma(Q_\mu)$ projects to $\Gamma_P(Q_\mu)$, where

$$\Gamma_P(Q_\mu)\Psi^J_{MK}(\Omega) = \sum_\nu \langle q_\nu\rangle \mathscr{D}^2_{\mu\nu}{}^*(\Omega)\Psi^J_{MK}(\Omega),$$

with

$$\langle q_\nu\rangle = \int dq\, q_\nu |\phi_n(q)|^2.$$

Thus, although the subspace of states of fixed n does not carry a subrepresentation of Γ, the projection Γ_P is nevertheless an irrep of the $[\mathbb{R}^5]\mathrm{SO}(3)$ Lie algebra. Therefore Γ_P is an embedded representation.

This demonstrates explicitly the known result that, in the adiabatic limit, a soft vibrational rotor can exhibit bands of states which, in isolation, are indistinguishable from those of a rigid rotor.

The structure of the carrier space for an embedded representation of the rotor algebra is remarkably similar to that of a standard sub-irrep. Whereas the carrier space for a sub-irrep is spanned by the SO(3) orbit

$$\{|q^\Omega\rangle = \mathscr{R}(\Omega)|q\rangle;\ \Omega \in \mathrm{SO}(3)\},$$

the carrier space for an embedded representation is spanned by the SO(3) orbit

$$\{|\phi^\Omega\rangle = \mathscr{R}(\Omega)|\phi\rangle;\ \Omega \in \mathrm{SO}(3)\},$$

where

$$|\phi\rangle = \int dq\, \phi_n(q)|q\rangle.$$

The characteristic feature of the rotor algebra is that the matrix elements of its irreps depend linearly on some of its irrep labels. Consequently, it becomes possible to mix states from irreps with different values of these labels in a way that preserves a parallel linear relationship with some average representation labels. Evidently, this possibility exists for other semidirect sum Lie algebras with Abelian ideals.

It is of interest to discover if physically significant embedded representations can occur under more general (nonlinear) circumstances. It would seem to be unlikely or, at least, that the circumstances would have to be artificially contrived. However, for physical applications, one is also very interested in situations that closely approximate embedded representations. These appear to be much more widespread and relevant for the construction of tractable theories based on the adiabatic approximation.

Consider, for example, a sequence of irreps of a Lie algebra in which the matrix elements vary smoothly as functions of the representation labels. It may then be a good approximation, in some situations, to make a linear approximation for the dependence over some relatively narrow range of irreps in order to construct approximate embedded representations, as for the rotor algebra.

In the following we shall illustrate this possibility for the Heisenberg–Weyl and symplectic Lie algebras.

V. THE HEISENBERG–WEYL LIE ALGEBRA

For simplicity, we consider the first Heisenberg–Weyl Lie algebra hw(1), which is the most familiar dynamical algebra for vibrations in a single degree of freedom. The extension to higher Heisenberg–Weyl algebras is straightforward.

A basis for the complex extension of hw(1) is given by the set of operators $\{A,B,\Lambda\}$ having the commutation relations

$$[B,A] = \Lambda, \quad [\Lambda,A] = [\Lambda,B] = 0.$$

A. Irreps of hw(1)

A lowest weight state $|\lambda\rangle$ for a unitary irrep is defined by

$$\Gamma^{(\lambda)}(B)|\lambda\rangle = 0, \quad \Gamma^{(\lambda)}(\Lambda)|\lambda\rangle = \lambda|\lambda\rangle,$$

where λ serves as a label for the irrep. An orthonormal basis for the irrep is defined recursively by the equation

$$\Gamma^{(\lambda)}(A)|\lambda,n\rangle = \sqrt{\lambda(n+1)}\,|\lambda,n+1\rangle.$$

Thus we obtain the matrix elements for the λ irrep,

$$\langle\lambda,m+1|\Gamma^{(\lambda)}(A)|\lambda,n\rangle = \sqrt{\lambda\,(n+1)}\delta_{mn},$$

$$\langle\lambda,n|\Gamma^{(\lambda)}(B)|\lambda,m+1\rangle = \langle\lambda,m+1|\Gamma^{(\lambda)}(A)|\lambda,n\rangle^*,$$

$$\langle\lambda,m|\Gamma^{(\lambda)}(\Lambda)|\lambda,n\rangle = \lambda\delta_{mn}.$$

The $\Gamma^{(\lambda)}$ irrep is seen to be related to the more familiar $\Gamma^{(\lambda=1)}$ irrep by

$$\Gamma^{(\lambda)}(A) = \sqrt{\lambda}\,a^\dagger, \quad \Gamma^{(\lambda)}(B) = \sqrt{\lambda}\,a, \quad \Gamma^{(\lambda)}(\Lambda) = \lambda I,$$

where

$$a^\dagger = \Gamma^{(1)}(A), \quad a = \Gamma^{(1)}(B),$$

and $I = \Gamma^{(1)}(\Lambda)$ is the identity operator.

B. Approximate embedded representations of hw(1)

Consider now the situation in which, instead of a single hw(1) irrep, we have a direct sum $\Gamma = \Sigma_\lambda \Gamma^{(\lambda)}$ of irreps and an orthonormal basis for the carrier space given by state vectors

$$|\phi_\nu,n\rangle = \sum_\lambda \phi_{\nu\lambda}|\lambda,n\rangle,$$

where $\{\phi_{\nu\lambda}\}$ is a set of coefficients such that

$$\sum_\lambda \phi^*_{\mu\lambda}\phi_{\nu\lambda} = \delta_{\mu\nu}.$$

If we restrict to a subspace of states of ν = const and put

$$\lambda = \lambda_0 + \epsilon,$$

where

$$\lambda_0 = \langle\lambda\rangle = \sum_\lambda |\phi_{\nu\lambda}|^2\lambda,$$

we obtain

$$\langle\phi_\nu,m+1|\Gamma(A)|\phi_\nu,n\rangle = \delta_{mn}\sqrt{\lambda_0(n+1)}\,(1 - \langle\epsilon^2\rangle/8\lambda_0^2 + \cdots),$$

$$\langle\phi_\nu,n|\Gamma(B)|\phi_\nu,m+1\rangle = \langle\phi_\nu,m+1|\Gamma(A)|\phi_\nu,n\rangle^*,$$

$$\langle\phi_\nu,m|\Gamma(\Lambda)|\phi_\nu,n\rangle = \delta_{mn}\lambda_0,$$

where

$$\langle\epsilon^2\rangle = \sum_\lambda |\phi_{\nu\lambda}|^2(\lambda-\lambda_0)^2.$$

It follows that we have an approximate embedded representation of the hw(1) Lie algebra provided $\langle\epsilon^2\rangle \ll 8\lambda_0^2$.

An alternative, and possibly more useful, construction of an approximate embedded representation for the Lie group HW(1) and its Lie algebra hw(1), which is equivalent to the above to within the limits of the approximation, is to construct the carrier space of the embedded representation as the span of an orthonormal basis of states which satisfy the equations

$$\Gamma(B)|\phi_\nu,0\rangle = 0, \quad \langle\phi_\nu,0|\Gamma(\Lambda)|\phi_\nu,0\rangle = \lambda_0,$$

$$|\phi_\nu,n+1\rangle = \frac{\Gamma(A)|\phi_\nu,n\rangle}{\langle\phi_\nu,n|\Gamma(BA)|\phi_\nu,n\rangle^{1/2}}.$$

Although in quantum mechanics one is accustomed to considering only $\lambda = 1$ irreps of the Heisenberg–Weyl algebras hw(n), corresponding, with a suitable choice of A, B, and Λ, to the commutation relations

$$[x,p] = i\hbar I,$$

with $\hbar$ fixed at the value given by Planck's constant, other representations with $\lambda \neq 1$ (or, equivalently, with $\hbar$ different from the Planck value) exist mathematically and could exist in physics. It is conceivable that what one sees in physics are embedded representations with only mean value of $\hbar$ given by the Planck value, and that higher energy representations exist with orthogonal distributions of $\hbar$.

Note also that, although the Heisenberg–Weyl algebras are the simplest dynamical algebras for describing vibrational dynamics, they are not the most appropriate in all situations. The normal mode vibrations of a composite system with internal degrees of freedom may be more appropriately described by a symplectic algebra, for example, as we now consider.

VI. THE sp(1,R)~su(1,1) VIBRATOR ALGEBRA

For example, Sp(1,R) is the appropriate dynamical group for a theory of monopole (breathing mode) vibrations of nuclei.[6]

A basis for the complex extension of the Lie algebra sp(1,R) is given by a raising operator A, a lowering operator B, and a u(1) operator C with the commutation relations

$$[C,A] = 2A, \quad [C,B] = -2B, \quad [B,A] = 4C.$$

A lowest weight state for a unitary irrep of sp(1,R) is defined by

$$\Gamma^{(N)}(B)|N,0\rangle = 0, \quad \Gamma^{(N)}(C)|N,0\rangle = N|N,0\rangle,$$

where N, a positive integer or (for a spinor irrep) a positive half integer, serves as a label for the irrep. One easily shows that orthonormal basis states are defined recursively for such an irrep by the equation

$$\Gamma^{(N)}(A)|N,n\rangle = [4(N+n)(n+1)]^{1/2}|N,n+1\rangle.$$

The matrix elements of the sp(1,ℝ) operators are given in this basis by

$$\langle N,m+1|\Gamma^{(N)}(A)|N,n\rangle = \delta_{mn}[4(N+n)(n+1)]^{1/2},$$

$$\langle N,n|\Gamma^{(N)}(B)|N,m+1\rangle = \langle N,m+1|\Gamma^{(N)}(A)|N,n\rangle^*,$$

$$\langle N,m|\Gamma^{(N)}(C)|N,n\rangle = \delta_{mn}(N+2n).$$

Now consider a band of states,

$$|\psi_n\rangle = \sum_N C_N|N,n\rangle,$$

that straddle a distribution of sp(1,ℝ) irreps. Write

$$N = N_0 + \lambda,$$

where

$$N_0 = \langle N\rangle = \sum_N |C_N|^2 N.$$

We then easily find that

$$\langle\psi_{m+1}|\Gamma(A)|\psi_n\rangle = \delta_{mn}[4(N_0+n)(n+1)]^{1/2} \times(1-\langle\lambda^2\rangle/8(N_0+n)^2+\cdots),$$

$$\langle\psi_n|\Gamma(B)|\psi_{m+1}\rangle = \langle\psi_{m+1}|\Gamma(A)|\psi_n\rangle^*,$$

$$\langle\psi_n|\Gamma(C)|\psi_n\rangle = \delta_{mn}(N_0+2n),$$

where

$$\langle\lambda^2\rangle = \sum_N |C_N|^2(N-N_0)^2.$$

It follows that, in the limit of $8N_0^2$ large compared with $\langle\lambda^2\rangle$, $\langle\lambda^2\rangle/8(N_0+n)^2\to 0$ and the set of states $\{|\psi_n\rangle\}$ carries an embedded representation $\Gamma^{(N_0)}$ of sp(1,ℝ).

We again observe that an alternative, and possibly more useful, definition of the (approximate) embedded representation is to define its carrier space as the span of the orthonormal basis states which satisy

$$\Gamma(B)|\phi_0\rangle = 0, \quad \langle\phi_0|\Gamma(C)|\phi_0\rangle = N_0,$$

$$|\phi_{n+1}\rangle = \frac{\Gamma(A)|\phi_n\rangle}{\langle\phi_n|\Gamma(BA)|\phi_n\rangle^{1/2}},$$

where Γ is the reducible representation of Sp(1,ℝ) given by the direct sum of the irreps $\Gamma^{(N)}$.

These are significant results because they mean that one can use sp(1,ℝ) as a dynamical structure algebra for the description of nuclear monopole vibrations, even though it may be overly restrictive to assume that the physical states belong to a single irrep. More important, however, is the fact that the sp(1,ℝ) Lie algebra is prototypical of richer dynamical structure Lie algebras, such as sp(3,ℝ), which has featured widely in the theory of nuclear collective states.[7]

VII. DISCUSSION

We have shown that the success of an algebraic model in describing subsets of observable properties of a many-body system does not necessarily imply the existence of a corresponding dynamical symmetry group for the system. It may only imply the existence of a dynamical structure group. In particular, we have shown that the observation of bands of states that accurately obey the predictions of the rotational model does not imply that the states of the band belong to an irreducible subspace of the rotor algebra. They may belong only to an embedded representation.

It will, of course, be recognized that, given sufficient experimental data so that one can extract matrix elements of a basis for a Lie algebra between all physical states of a system, it is possible to distinguish between an embedded representation and a subrepresentation. The indistinguishability arises in practice when one considers a restricted set of data involving, for example, the states of a single collective band, and ignores the often small matrix elements connecting these states to other possibly higher energy states.

Although perhaps not recognized explicitly, these concepts have been implicitly used in physics both in the context of the adiabatic approximation, as we have already discussed, and in what is often referred to as *renormalization*. For example, to take into account the corrections to a model due to coupling to neglected states, one often assigns renormalized values to the parameters of the model, such as effective masses or effective charges, different from their physical values.

However, the explicit recognition that one can use dynamical structure groups, which are neither symmetry groups nor even dynamical symmetry groups, opens up the possibility of more extensive applications of group theory in physics than hitherto. For example, it has long been maintained that the application of Elliott's SU(3) model of nuclear rotations[8] should be restricted to light nuclei because one knows that, in heavy nuclei, the spin-orbit interaction mixes SU(3) irreps strongly. As we show elsewhere,[9] one can in fact admit very large mixing of SU(3) irreps by the spin-orbit interaction and still retain the essential properties of SU(3) bands.

In a forthcoming paper[9] we investigate the application of the concepts of dynamical structure and embedded representations to the microscopic description of nuclear collective structure.

[1] *Dynamical Groups and Spectrum Generating Algebras*, edited by A. Barut, A. Bohm, and Y. Ne'eman (World Scientific, Singapore, 1986).

[2] D. J. Rowe, Rev. Mod. Phys. **40**, 153 (1968).

[3] G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).

[4] L. Weaver, L. Ç. Biedenharn, and R. Y. Cusson, Ann. Phys. (NY) **77**, 250 (1973).

[5] J. Meyer ter Vehn, Nucl. Phys. A **249**, 111 (1975).

[6] I. P. Okhrimento and A. I. Steshenko, Yad. Fiz. **32**, 381 (1980) [Sov. Phys. J. **32**, 197 (1980)]; J. Broeckhove, Physica A **114**, 454 (1982).

[7] G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **38**, 10 (1977); D. J. Rowe, Rep. Prog. Phys. **48**, 1419 (1985).

[8] J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958).

[9] P. Rochford and D. J. Rowe, "The survival of rotor and SU(3) bands under strong spin-orbit symmetry mixing," preprint.

Cauchy surfaces in a globally hyperbolic space-time

Jan Dieckmann
Fachbereich Mathematik, TU Berlin, D-1000 Berlin 12, West Germany

(Received 16 June 1987; accepted for publication 7 October 1987)

With the help of volume functions it is shown that a globally hyperbolic space-time possesses a Cauchy surface which is a three-dimensional, connected, spacelike C^∞ hypersurface.

I. INTRODUCTION

It is a widely used theorem that every globally hyperbolic space-time M possesses a Cauchy surface which also is a C^∞ hypersurface of M. Sachs and Wu[1] called it "one of the folk theorems of the subject," since an elegant proof of this statement is still missing. (In fact, there is a proof that every globally hyperbolic space-time possesses a Cauchy surface in Geroch,[2] and also a sketch of a proof of the announced result in Ref. 3.) In this paper we will give a proof of the above result. We shall use the terminology and notation of O'Neill.[4]

II. GLOBALLY HYPERBOLIC SPACE-TIMES

In the following proposition we call a function $t: M \to \mathbb{R}$ a *time function*, if

$$p < q \Rightarrow t(p) < t(q) \quad (p,q \in M).$$

If σ is a finite Borel measure on M with

(i) $\sigma(U) > 0 \quad (\varnothing \neq U \subseteq M)$,

(ii) $\sigma(\mathrm{Bd}\, I^{\pm}(p)) = 0 \quad (p \in M)$,

we call the functions

$$t^{\pm}(p) := \mp \sigma(I^{\pm}(p)) \quad (p \in M)$$

(*future or past*) *volume functions*. In Ref. 5 it is shown that such a measure always exists.

Proposition: Let t^-, t^+ be volume functions in M and let $\tilde{t} := \ln(-t^-/t^+)$. The following are equivalent: (i) M is globally hyperbolic; (ii) $\tilde{t}$ is a continuous time function and for all causal, inextendible curves γ we have $\mathrm{Ran}(\tilde{t}\circ\gamma) = \mathbb{R}$; and (iii) $\tilde{t}^{-1}(\{a\})$ is for all $a \in \mathbb{R}$ a Cauchy surface.

The proof may be found in Ref. 6.

Now we are able to prove the announced theorem.

Theorem: A space-time (M,g) is globally hyperbolic iff a C^∞ manifold N_0 and a diffeomorphism $\Psi: N_0 \times \mathbb{R} \to M$ exist with $\Psi(N_0 \times \{a\})$ being for all $a \in \mathbb{R}$ a spacelike, connected Cauchy surface.

Proof: "$\Rightarrow$" Let $t^{\pm}$ be volume functions. Due to the proposition the function

$$\tilde{t} := \ln(-t^-/t^+)$$

is a continuous time function. In Ref. 7 it is shown that we can smooth a continuous time function receiving a C^∞-time function t with

$$|(\tilde{t} - t)(p)| < 1 \quad (p \in M),$$

and dt is timelike. It follows that the level surfaces of t are spacelike hypersurfaces of M. In fact, they even are Cauchy surfaces; for let $a \in \mathbb{R}$, $N_a := t^{-1}(\{a\})$, and γ be an inextendible, timelike curve. Because t is a time function, γ can meet N_a at most once. Because

$$t(p) \leqslant |t(p) - \tilde{t}(p)| + \tilde{t}(p) < a \quad [p \in \tilde{t}^{-1}(\{a-1\})],$$

$$t(p) \geqslant \tilde{t}(p) - |\tilde{t}(p) - t(p)| > a \quad [p \in \tilde{t}^{-1}(\{a+1\})],$$

the fact that $\tilde{t}^{-1}(\{a \pm 1\})$ are Cauchy surfaces, and the continuity of t, γ must meet N_a at least once.

To construct Ψ we need a lemma.

Lemma: The map $\rho: M \to N_0$, which takes each $p \in M$ to the unique point on the Cauchy surface N_0, at which the integral curve of the timelike C^∞-vector field grad t through p meets N_0, is a submersion.

Proof of the lemma: Let $p \in M$. The integral curve α_p through p meets N_0 in a unique point $\alpha_p(u)$. Because the maps

$$\Phi_{\pm u}(\cdot) := \Phi(\cdot, \pm u),$$

where Φ denotes the flow of grad t, are diffeomorphisms, $\tilde{N} := \Phi_{-u}(N_0)$ is a hypersurface through p. Moreover, $\tilde{N}$ is transversal to grad t since

$$\begin{aligned} d\Phi_u(\mathrm{grad}\, t_p) &= [\Phi_u \circ \alpha_p]'(0) \\ &= [\Phi(\Phi(p,u),\cdot)]'(0) \\ &= \alpha'_{\rho(p)}(0) = \mathrm{grad}\, t_{\rho(p)}\,. \end{aligned}$$

Since we also have grad $t_p \neq 0$, there is a chart (x,U) at p, such that grad $t = \partial_1$ on U and $x(\tilde{N}) \subset (u^1)^{-1}(\{0\})$, where u^1 is a natural coordinate function on $\mathbb{R}^4$ (see O'Neill[8]). Obviously we have

$$\rho \restriction_U = \Phi_u \circ x^{-1} \circ \pi \circ x,$$

where $\pi: \mathbb{R}^4 \to (u^1)^{-1}(\{0\})$ denotes the natural projection, which shows that ρ is a submersion. ▼

Now we construct the inverse of Ψ: The C^∞ map

$$\tilde{\Psi}: M \to N_0 \times \mathbb{R}, \quad p \to (\rho(p), t(p))$$

is bijective and, because t and ρ are of maximum rank, even a diffeomorphism. Since M is connected and

$$N_0 = \pi \circ \tilde{\Psi}(M),$$

where $\pi: N_0 \times \mathbb{R} \to N_0$ denotes the projection, N_0 is also connected. The map $\Psi := \tilde{\Psi}^{-1}$ has all required properties.

"$\Leftarrow$" A space-time possessing a Cauchy surface is globally hyperbolic.[9] ■

[1]R. K. Sachs and H. Wu, Bull. Am. Math. Soc. 83, 1155 (1977).
[2]R. Geroch, J. Math. Phys. 11, 437 (1970).

 0022-2488/88/030578-02$02.50

[3]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, 1973).
[4]B. O'Neill, *Semi-Riemannian Geometry* (Academic, New York, 1983).
[5]J. Dieckmann, Ph.D. thesis, TU Berlin, 1987.
[6]See Ref. 5, p. 56.
[7]H. J. Seifert, Gen. Relativ. Gravit. **8**, 815 (1977).
[8]See Ref. 4, p. 30.
[9]See Ref. 4, p. 422.

Diffeomorphisms, orientation, and pin structures in two dimensions

L. Dąbrowski[a] and R. Percacci
International School for Advanced Studies (ISAS), Trieste, Italy

(Received 1 July 1987; accepted for publication 23 September 1987)

A set of generators for the modular group of a surface with boundary, both in the orientable and nonorientable cases is given. All inequivalent pin structures are constructed and their transformations under these generators are computed.

I. INTRODUCTION

In Ref. 1 we solved the following problem: given a compact, connected, orientable, two-dimensional Riemannian manifold Σ without boundary, compute the action of the orientation preserving diffeomorphisms of Σ on the set of inequivalent spin structures on Σ. Since isotopic diffeomorphisms induce the same transformation on spin structures, it is sufficient to determine the action of one particular representative in each isotopy class. The group of isotopy classes $\Omega^+(\Sigma) = \pi_0(\mathscr{D}\mathit{iff}^+\Sigma)$, which we call the modular group of Σ, has a finite set of generators, for which representative diffeomorphisms are known (the Dehn twists). Therefore the problem of determining the action of an arbitrary diffeomorphism on spin structures is reduced to that of determining the action of a finite number of diffeomorphisms.

In this paper we address the analogous problem for an arbitrary compact, connected manifold (henceforth called a surface). Since we are going to discuss orientation reversing diffeomorphisms and nonorientable surfaces, we do not work with spin structures, but rather with pin structures [i.e., prolongations of the bundle of orthonormal frames to the double covering of the full group O(2)]. This is explained in more detail in Sec. II, where we also collect some basic facts on the topology of surfaces. In Sec. III we consider the case of an orientable surface with boundary and compute the action of the full diffeomorphism group on pin structures. In the rest of the paper we discuss the case of a nonorientable surface N, possibly with boundary. In Sec. IV, we give explicit representatives for the generators of the modular group $\Omega(N)$. In Sec. V we give a complete description of all pin structures on N. The group O(2) has two inequivalent double coverings $\mathrm{Pin}^+(2)$ and $\mathrm{Pin}^-(2)$, which have to be treated separately. For instance, we will find that depending on the topology of N, pin structures exist for one of them, but not for the other. In Sec. VI we compute the action of the generators of $\Omega(N)$ on pin structures and we find the orbits of this action. In the two appendices we collect some supplementary topological results on nonorientable surfaces.

The motivation for this work came from the problem of modular invariance in superstring theory. Surfaces with boundary occur in the Feynman integral representation of the vacuum amplitude for open strings and of the scattering amplitude of closed strings. Nonorientable surfaces appear in the theory of nonoriented strings.

[a] On leave of absence from I.F.T., Uniwersytet Wrocławski, Poland.

II. PRELIMINARIES

In this section we recall some basic facts on the topology of surfaces, their diffeomorphism groups, the double coverings of the linear and orthogonal groups in two dimensions, and pin structures.

We denote Σ_g an orientable surface without boundary of genus g; it is homeomorphic to a sphere if $g = 0$ or to the connected sum of g tori if $g \geqslant 1$. Removing from Σ_g n disjoint open disks $D_1,...,D_n$ we obtain a surface $\Sigma_{g,n}$ with boundary consisting of n circles $d_1,...,d_n$. Every smooth compact, connected, orientable surface is homeomorphic, and actually diffeomorphic, to $\Sigma_{g,n}$ for some g,n. We will always work with a specific realization of Σ_g as a surface embedded in $\mathbb{R}^3$, symmetrically with respect to the plane reflections K_i ($i = 1,2,3$) which invert the ith axis (see Fig. 1). The first homology group is $H_1(\Sigma_g,\mathbb{Z}) = \mathbb{Z}^{2g}$ and $H_1(\Sigma_{g,n},\mathbb{Z}) = \mathbb{Z}^{2g+n-1}$ for $n \geqslant 1$; the generators are the cycles a_A, b_A for $A = 1,...,g$ drawn in Fig. 1 and the cycles d_h for $h = 1,...,n$, with the relation

$$\sum_{h=1}^{n} d_h = 0. \tag{2.1}$$

We denote N_g ($g \geqslant 1$) a nonorientable surface without boundary of genus g; it is homeomorphic to the connected sum of g real projective planes. Removing from N_g n disjoint open disks $D_1,...,D_n$ we obtain a surface $N_{g,n}$ with a boundary consisting of n circles $d_1,...,d_n$. Every compact, connected, nonorientable surface is homeomorphic,[2] and actually diffeomorphic, to $N_{g,n}$ for some g,n. For our purposes it will

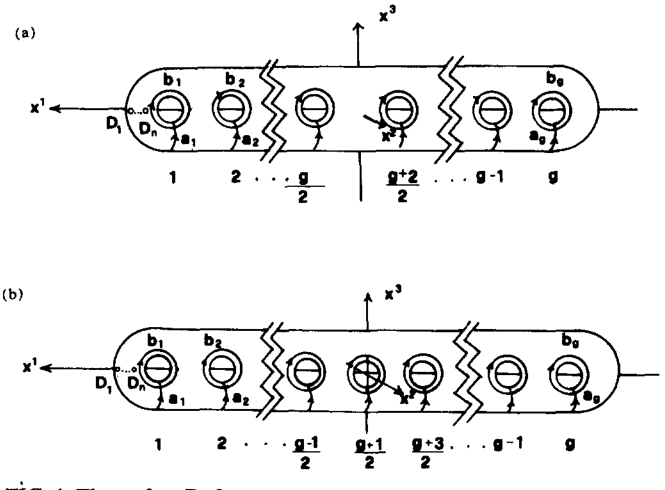


FIG. 1. The surface Σ_g for (a) g even and (b) g odd.

 0022-2488/88/030580-14$02.50

be convenient to use an alternative picture of a nonorientable surface. Let $J = K_1K_2K_3$ be the total inversion in $\mathbb{R}^3$, i.e., $J(x^1,x^2,x^3) = (-x^1,-x^2,-x^3)$. Consider the surface Σ_{g-1} ($g \geqslant 1$) embedded in $\mathbb{R}^3$ as before (see Fig. 2) and remove $2n$ disks D_h with $h = 1,\dots,2n$ such that $J(D_h) = D_{2n-h+1}$. Then J restricts to an orientation-reversing, fixed point-free diffeomorphism of $\Sigma_{g-1,2n}$, which generates a group $\mathbb{Z}_2 = \{\mathrm{Id}, J\}$. The quotient $\Sigma_{g-1,2n}/\mathbb{Z}_2$ is a compact, nonorientable surface with boundary consisting of n circles. It can be thought of as the part of $\Sigma_{g-1,2n}$ with $x^1 \geqslant 0$ subject to the identifications $(0,x^2,x^3) = (0,-x^2,-x^3)$. This is the connected sum of $\Sigma_{(g-1)/2}$ and a projective plane, if g is odd, or $\Sigma_{(g-2)/2}$ and a Klein bottle if g is even, with n disks removed in both cases. We show in Appendix A that these spaces are homeomorphic to $N_{g,n}$. Therefore

$$\Sigma_{g-1,2n}/\mathbb{Z}_2 = N_{g,n}\,.$$

The natural projection π: $\Sigma_{g-1,2n} \to N_{g,n}$ is a double covering. The first homology group is $H_1(N_g,\mathbb{Z}) = \mathbb{Z}^{g-1} \oplus \mathbb{Z}_2$ and $H_1(N_{g,n},\mathbb{Z}) = \mathbb{Z}^{g+n-1}$ for $n \geqslant 1$. We shall use the same symbol for a loop on $\Sigma_{g-1,2n}$ and the projection onto $N_{g,n}$ of the part of that loop which lies in the half-space $x^1 \geqslant 0$. Then, the generators of $H_1(N_{g,n},\mathbb{Z})$ can be listed as follows: $a_1,\dots,a_{(g-1)/2}$, $b_1,\dots,b_{(g-1)/2}$, $c_{(g-1)/2}$, $d_1,\dots,d_n$ with the relation

$$2(c_{(g-1)/2} - a_{(g-1)/2}) + \sum_{h=1}^{n} d_h = 0\,, \tag{2.2}$$

if g is odd, and $a_1,\dots,a_{g/2}$, $b_1,\dots,b_{g/2}$, $d_1,\dots,d_n$ with the relation

$$2a_{g/2} + \sum_{h=1}^{n} d_h = 0\,, \tag{2.3}$$

if g is even. The orientation reversing generators are $c_{(g-1)/2}$ for g odd and $b_{g/2}$ for g even.

Let $D(\Sigma_{g,n})$ be the group of diffeomorphisms of $\Sigma_{g,n}$. We have two chains of inclusions

$$D_0(\Sigma_{g,n}) \subset D_\partial(\Sigma_{g,n}) \subset D_B(\Sigma_{g,n}) \subset D(\Sigma_{g,n})$$

and

$$D_\partial(\Sigma_{g,n}) \subset D^+(\Sigma_{g,n}) \subset D(\Sigma_{g,n})\,,$$

where D_0, D_∂, D_B, and D^+ denote the subgroups of D consisting of diffeomorphisms which are isotopic to the identity, leave the boundary pointwise fixed, map each connected component of the boundary to itself, or preserve the orientation, respectively. Here and in the following by isotopy we always mean smooth isotopy, i.e., homotopy through diffeomorphisms. We denote

$$\Omega(\Sigma_{g,n}) = \pi_0(D(\Sigma_{g,n})) = D(\Sigma_{g,n})/D_0(\Sigma_{g,n})$$

and, with obvious notation, we have inclusions

$$\Omega_\partial(\Sigma_{g,n}) \subset \Omega_B(\Sigma_{g,n}) \subset \Omega(\Sigma_{g,n})$$

and

$$\Omega_\partial(\Sigma_{g,n}) \subset \Omega^+(\Sigma_{g,n}) \subset \Omega(\Sigma_{g,n})\,.$$

Similarly, for a nonorientable surface, we have

$$D_0(N_{g,n}) \subset D_\partial(N_{g,n}) \subset D_B(N_{g,n}) \subset D(N_{g,n})$$

and

$$\Omega_\partial(N_{g,n}) \subset \Omega_B(N_{g,n}) \subset \Omega(N_{g,n})\,.$$

In all cases, Ω will be called the modular group.

In two dimensions, every homeomorphism is continuously isotopic to a diffeomorphism and furthermore, if two diffeomorphisms are continuously isotopic, they are also smoothly isotopic. Therefore, the modular group of a surface can be identified with its homeotopy group (the group of continuous isotopy classes of homeomorphisms). This will allow us to use known results on the homeotopy groups of surfaces.

A spin structure on an oriented n-dimensional Riemannian manifold M is a prolongation of the bundle of oriented orthonormal frames to the group $\mathrm{Spin}(n)$, the double covering of $\mathrm{SO}(n)$. As discussed in Ref. 1, in order to define rigorously the transformation of spin structures under orientation-preserving diffeomorphisms, it is necessary to use instead the prolongations of the bundle of all oriented frames to the double covering of $\mathrm{GL}^+(n)$. If f is an orientation reversing diffeomorphism of M, or if M is not orientable, then the derivative Tf is an automorphism of the bundle of all frames. To define spinors on M and their transformation under f in these cases, it is necessary to use a prolongation of the bundle of frames to a double covering of $\mathrm{GL}(n)$. The group $\mathrm{GL}(n)$ is retractable to its maximal compact subgroup $\mathrm{O}(n)$ and its two double coverings are retractable to the double coverings of $\mathrm{O}(n)$, denoted $\mathrm{Pin}^+(n)$ and $\mathrm{Pin}^-(n)$. This is easily established using the Iwasawa decompositions of $\mathrm{GL}(n)$ and its double coverings.

The general discussion in Secs. II and III of Ref. 1 can be repeated in this more general case, the only modification being the nonuniqueness of the double covering of $\mathrm{GL}(n)$. A prolongation of the bundle of frames to a fixed double covering of $\mathrm{GL}(n)$ exists if and only if there exists a prolongation of the bundle of orthonormal frames to the corresponding double covering of $\mathrm{O}(n)$ (i.e., a pin structure). Furthermore, when they exist, there is a bijective correspondence between prolongations of the bundle of frames to a fixed double covering of $\mathrm{GL}(n)$ and prolongations of the bundle of orthonormal frames to the corresponding double covering of $\mathrm{O}(n)$. The topological conditions for the existence of $\mathrm{Pin}^+(2)$— and $\mathrm{Pin}^-(2)$—structures are, respectively,

$$w_2 + w_1^2 = 0 \quad \text{and} \quad w_2 = 0\,, \tag{2.4}$$

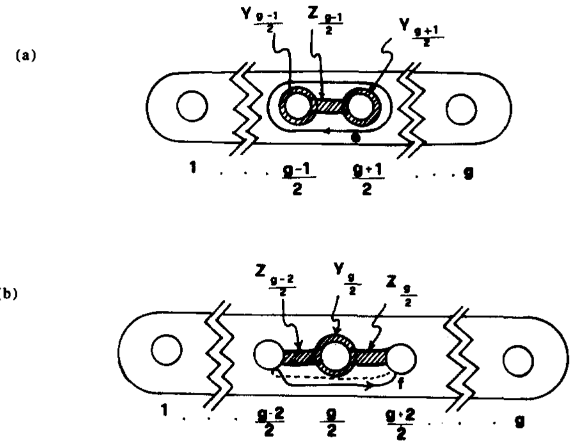


FIG. 2. The surface Σ_{g-1} for (a) g odd and (b) g even.

where w_1 and w_2 are the first and second Stiefel–Whitney classes of M. Furthermore, when pin structures exist, they are classified by $H^1(M,\mathbb{Z}_2) = \mathrm{Hom}(H_1(M,\mathbb{Z}),\mathbb{Z}_2)$. In practice, the transformation of prolongations under diffeomorphisms can be computed working almost all the time with the groups $\mathrm{Pin}^{\pm}(n)$, so in the following we will always talk about pin structures rather than prolongations of the bundle of frames. In this paper we shall use the charts and local trivializations of the bundle of frames of Σ_g which were introduced in Sec. 4 of Ref. 1. In order to be able to give a unified treatment for all genera $g \geqslant 0$ we have to define the chart U also in the case of the sphere [this was not needed in Ref. 1 because S^2 admits only one spin structure and $\Omega^+(S^2) = 1$]. This we achieve by declaring the equator to be the loop c_0 and introducing a coordinate neighborhood Z_0 of c_0 with coordinates (ζ_0, z_0). Then we define an atlas of F on the open covering U, U', U'', where $U = Z_0$, U', and U'' are the northern and southern hemispheres and the fields of frames e, e', e'' are defined by Eqs. 4.10, 6, 7 in Ref. 1, respectively. This ensures that the transition functions of F have even winding number and have lifts to the group Spin (2).

Finally, we describe the double covers of O(2). To the Euclidean space $\mathbb{R}^2$ are associated two Clifford algebras $C^{\pm}(2)$, which are generated by two elements γ_1, γ_2, with the relation

$$\gamma_i \gamma_j + \gamma_j \gamma_i = \pm 2\delta_{ij} \quad (i = 1,2),$$

respectively. We shall use a 2×2 matrix representation with $\gamma_i = \sigma_i$ [for $C^+(2)$] or $\gamma_i = \sqrt{-1}\sigma_i$ [for $C^-(2)$], where σ_i are the first two Pauli matrices. The group $\mathrm{Pin}^{\pm}(2)$ is a subgroup of $C^{\pm}(2)$, respectively, consisting of two connected components. The identity-connected component Spin(2) consists of elements of the form $\exp(s\gamma_1\gamma_2)$ with $0 \leqslant s \leqslant 2\pi$; the other component is obtained by composing elements of Spin(2) with γ_1. The covering homomorphism ρ: $\mathrm{Pin}^{\pm}(2) \to \mathrm{O}(2)$ is defined by

$$\rho(a)\gamma_i = a\gamma_i a^{-1},$$

$$\rho(\gamma_1 a)\gamma_i = -\gamma_1 a \gamma_i a^{-1} \gamma_1^{-1}, \quad \text{for } a \in \mathrm{Spin}\ (2),$$

where the γ_i are regarded here as a basis for $\mathbb{R}^2 \subset C^{\pm}(2)$. Notice that $0 \leqslant s \leqslant \pi$ parametrizes an open path in Spin(2) joining $\mathbb{1}$ to $-\mathbb{1}$, which covers a loop in SO(2) starting and ending at $\mathbb{1}$.

III. DIFFEOMORPHISMS AND PIN STRUCTURES ON $\Sigma_{g,n}$

The group $\Omega^+(\Sigma_g)$ and its action on spin structures have been discussed in Ref. 1. A set of generators for $\Omega(\Sigma_g)$ is given by a set of generators for $\Omega^+(\Sigma_g)$ plus (the isotopy class of) an orientation-reversing diffeomorphism. In the presence of a nontrivial boundary there are additional generators. It is convenient to regard $\Sigma_{g,n}$ as a subset of Σ_g and represent the generators of $\Omega(\Sigma_{g,n})$ by diffeomorphisms of Σ_g; some of them will be isotopic to the identity on Σ_g, but not on $\Sigma_{g,n}$, since an isotopy on $\Sigma_{g,n}$ must consist of diffeomorphisms which map the boundary to itself. Consider the loops a_A, b_A which wind around the A th handle, for $A = 1,\dots,g$; loops c_A which connect the A th and the $(A+1)$-st handle, for $A = 1,\dots,g-1$; loops r_{Ah}, s_{Ah} which intersect D_h and wind around the A th handle, for $A = 1,\dots,g$ and $h = 1,\dots,n$; loops t_h which intersect D_h and D_{h+1}, for $h = 1,\dots,n-1$ (see Fig. 3).

Let ℓ denote anyone of these loops. We introduce local coordinates in a tubular neighborhood $U[\ell]$ of ℓ, by choosing an orientation preserving diffeomorphism of $U[\ell]$ onto the open annulus $\{z \in \mathbb{C} \mid 1 - \varepsilon < |z| < 1 + \varepsilon\}$ in such a way that the oriented loop ℓ is mapped onto the clockwise oriented circle $|z| = 1$. In the case of the loops a_A, b_A, c_A these coordinates are related to those introduced in Ref. 1 by $U[a_A] = X_A$, $U[b_A] = Y_A$, $U[c_A] = Z_A$, $|z| - 1 = x_A, y_A, z_A$, $-\arg z = \chi_A, \vartheta_A, \zeta_A$, respectively. In the case of the loops r_{Ah}, s_{Ah} we assume also that the disk D_h is mapped onto the disk $|z+1| < \varepsilon/2$ and furthermore in the case of the loops t_h the disk D_{h+1} is mapped onto the disk $|z-1| < \varepsilon/2$. For each ℓ we define a diffeomorphism $f[\ell]$ called the twist around ℓ as follows: it is the identity in the complement of $U[\ell]$ and on $U[\ell]$

$$f[\ell]: z \mapsto z \cdot \exp[-2\sqrt{-1}\lambda((|z|-1)/\varepsilon)], \quad \text{for } \ell = a_A, b_A, c_A, \tag{3.1}$$

$$f[\ell]: z \mapsto z \cdot \exp\left[-2\sqrt{-1}\lambda\left(2\frac{|z|-1+\varepsilon}{\varepsilon}\right) - 2\sqrt{-1}\lambda\left(-2\frac{|z|-1-\varepsilon}{\varepsilon}\right)\right], \quad \text{for } \ell = r_{Ah}, s_{Ah}, \tag{3.2}$$

$$f[\ell]: z \mapsto z \cdot \exp\left[-\sqrt{-1}\lambda\left(2\frac{|z|-1+\varepsilon}{\varepsilon}\right) - \sqrt{-1}\lambda\left(-2\frac{|z|-1-\varepsilon}{\varepsilon}\right)\right], \quad \text{for } \ell = t_h, \tag{3.3}$$

where λ is a smooth function such that $\lambda(s) = 0$ for $s \leqslant 0$, $\lambda(s) = \pi$ for $s \geqslant 1$, and $d\lambda/ds \geqslant 0$. This definition agrees with the one given in Ref. 1 for $f[a_A]$, $f[b_A]$, and $f[c_A]$. The effect of these twists on the annulus is shown in Fig. 4: the spiraling lines are the images under the twists of the intersection of the real axis with the annulus.

For the orientation-reversing generator we take (the isotopy class of the restriction to $\Sigma_{g,n}$ of) the reflection K_3, assuming that the disks $D_1,\dots,D_n$ are placed on Σ_g in such a way that $K_3(D_n) = D_n$. Notice that the total inversion J would not work for odd n.

It was proved in Ref. 3 that $\Omega(\Sigma_{g,n})$ is generated by the following set: $\{f[a_A], f[b_A]$ for $A = 1,\dots,g$; $f[c_A]$ for $A = 1,\dots,g-1$; $f[r_{Ah}], f[s_{Ah}]$ for $A = 1,\dots,g$, $h = 1,\dots,n$;

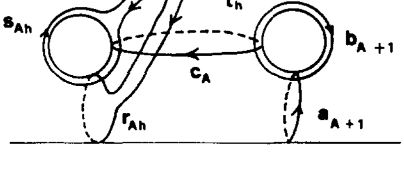


FIG. 3. The twists on $\Sigma_{g,n}$.

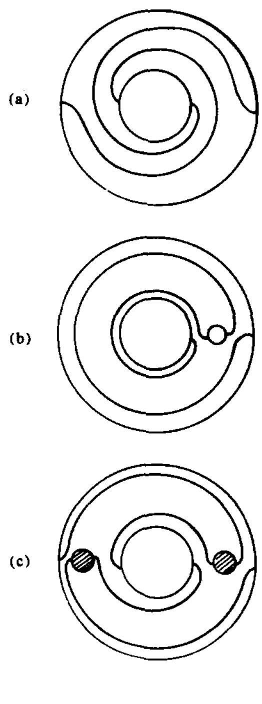


FIG. 4. The twists (3.1), (3.2), and (3.3).

$f[t_h]$ for $h=1,\dots,n-1$; $K_3\}$. In the special cases $g=0$, $g=1$ or $n=0$, $n=1$, some of these generators are not defined and $\Omega(\Sigma_{g,n})$ is generated by the remaining ones. We observe that this is not a minimal set of generators; for instance it was proven in Ref. 4 that $f[a_A]$ with $A=3,\dots,g$ can be expressed as combinations of the remaining $f[a_A]$, $f[b_A]$, and $f[c_A]$. It can also be seen that $f[r_{A,h+1}]$ and $f[s_{A,h+1}]$ are isotopic to $f[t_h]^{-1}f[r_{A,h}]f[t_h]$ and $f[t_h]^{-1}f[s_{Ah}]f[t_h]$.

The generators for the subgroups $\Omega^+(\Sigma_{g,n})$, $\Omega_B(\Sigma_{g,n})$, and $\Omega_\partial(\Sigma_{g,n})$ are obtained from those of $\Omega(\Sigma_{g,n})$ by omitting the generator $\{K_3\}$, the generators $\{f[t_h]$ for $h=1,\dots,n-1\}$ and the generators $\{K_3;\ f[t_h]$ for $h=1,\dots,n-1\}$, respectively.

On $\Sigma_{g,n}$ $(n\geqslant 1)$ there are 2^{2g+n-1} inequivalent spin structures labeled by $(2g+n)$-tuples of numbers $(i_1,\dots,i_g;\ j_1,\dots,j_g;\ k_1,\dots,k_n)$, each equal to 0 or 1 and with the relation

$$\sum_{h=1}^{m} k_h = 0 \bmod 2\,, \tag{3.4}$$

which derives from (2.1). All spin structures on $\Sigma_{g,n}$ have the same Spin(2) bundle $\widetilde{F}$, which is the restriction to $\Sigma_{g,n}$ of the bundle $\widetilde{F}$ of Ref. 1. In place of 4.12–4.16 in Ref. 1 we now have bundle morphisms $\eta_{\mathbf{ijk}}:\widetilde{F}\to F$ defined by

$$\eta_{\mathbf{ijk}}(\tilde{e}(x)) = e(x)r_{\mathbf{ijk}}(x)\,, \tag{3.5}$$

where $r_{\mathbf{ijk}}: U\to \mathrm{SO}(2)$ are such that composed with the loops a_A, b_A, and d_h (regarded as maps $S^1\to\Sigma_g$) they have winding numbers i_A, j_A, and k_h, respectively. This can be achieved by placing the disks $D_1,\dots,D_n$ in Z_0 if $g=0$ and Y_1 if $g\geqslant 1$ (see Fig. 5, where Y_1 has been slightly extended) and modifying the functions $r_{\mathbf{ij}}$ on the shaded strips in such a way that crossing the strip between D_h and D_{h+1}, the function $r_{\mathbf{ijk}}$ rotates by $2\pi q_h$ with $q_h=\Sigma_{m=1}^{h} k_m$.

Every spin structure $(\mathbf{ijk})$ extends uniquely to a

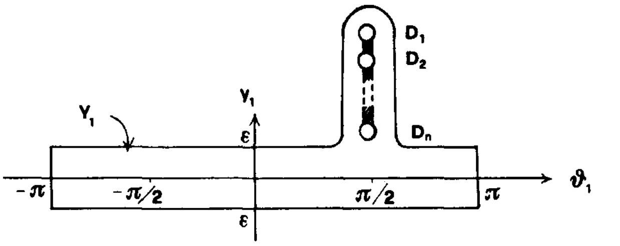


FIG. 5. Extension of the chart Y_1 in the presence of boundary.

$\mathrm{Pin}^+(2)$ and a $\mathrm{Pin}^-(2)$ structure, both of which will be denoted again $(\mathbf{ijk})$. In particular, having fixed the structure group, the pin bundle is the same for all pin structures.

The transformation of pin structures under $\Omega(\Sigma_{g,n})$ can be determined counting the winding numbers of the three terms on the rhs of (4.22) in Ref. 1. For this to work it is necessary that the homology generators and their images under f be contained in U; this is true for $f[a_A]$, $f[b_A]$, $f[c_A]$, and K_3, and will also be true for $f[r_{Ah}]$, $f[s_{Ah}]$, and $f[t_h]$ provided we deform the loops r_{Ah}, s_{Ah}, and t_h in such a way that they are entirely contained in U. We observe that the middle term in (4.22) is the only function which has values in GL(2) and in general not in O(2); however, it can be deformed into a function which has value in the subgroup O(2) without altering its winding numbers. In practice, the winding number of $h\circ\ell$ is determined counting how many times the vector tangent to $f\circ\ell$ winds with respect to the frames e on U.

Most of the labels $(\mathbf{ijk})$ are invariant under the action of the generators, so for each diffeomorphism we list only the labels which are changed,

$$\begin{aligned}
&f[a_B]:\ j_A\mapsto j_A+\delta_{AB}i_B\,,\\
&f[b_B]:\ i_A\mapsto i_A+\delta_{AB}j_B\,,\\
&f[c_B]:\ j_A\mapsto j_A+(\delta_{AB}+\delta_{A,B+1})(i_B+i_{B+1}+1)\,,\\
&f[r_{Bm}]:\ j_A\mapsto j_A+\delta_{AB}k_m\,,\\
&f[s_{Bm}]:\ i_A\mapsto i_A+\delta_{AB}k_m\,,\\
&f[t_m]:\ k_h\mapsto k_h+(\delta_{hm}+\delta_{h,m+1})(k_m+k_{m+1})\,.
\end{aligned}$$

Since K_3 leaves all pin structures invariant the orbits of $\Omega(\Sigma_{g,n})$ are the same as the orbits of $\Omega^+(\Sigma_{g,n})$. For $n=0$, they are characterized by the invariant

$$\varphi(\mathbf{i},\mathbf{j}) = \sum_{A=1}^{g} (i_A+1)(j_A+1)\,; \tag{3.6}$$

the pin structures with $\varphi=0$ (resp. 1) are called even (resp. odd). For $n\geqslant 1$ and $\mathbf{k}=\mathbf{0}$, the orbits are the same as in the case $n=0$. If $\mathbf{k}\neq\mathbf{0}$ there are $[n/2]$ orbits (where [ ] denotes the integer part of a number) which are characterized by the integer

$$K=\sum_{h=1}^{n} k_h\,. \tag{3.7}$$

Because of (3.4) the labels k_h that are equal to 1 must occur in pairs, so K is even.

In Table I we list all the orbits of $\Omega(\Sigma_{g,n})$, together with their invariants, cardinality, and a representative pin structure for each. We recall that $2[n/2]$ denotes the largest even

TABLE I. Orbits of $\Omega(\Sigma_{g,n})$ and $\Omega^+(\Sigma_{g,n})$.

κ	φ	Standard form	Cardinality
0	0	(0,...,0; 1,...,1;0,...,0)	$2^{g-1}(2^g+1)$
0	1	(0,...,0;1,...,1,0;0,...,0)	$2^{g-1}(2^g-1)$
2	—	(0,...,0; 1,...,1;0,...,0,1,1)	$2^{2g}\binom{n}{2}$
...	...	...	...
$2m$	—	(0,...,0; 1,...,1;0,...,0,1,...,1)	$2^{2g}\binom{n}{2m}$
...	...	...	...
$2[n/2]$	—	(0,...,0; 1,...,1;$\theta(n+1)$,1,...,1)	$2^{2g}\binom{n}{2[n/2]}$

integer which is $\leqslant n$ and we use the abbreviation $\theta(n) = n - 2[n/2]$ [i.e., $\theta(n) = 0$ for n even, and 1 for n odd].

The groups $\Omega_B(\Sigma_{g,n})$ and $\Omega_\partial(\Sigma_{g,n})$ have the same orbits, because they have the same set of generators except for K_3, which leaves all pin structures invariant. Here **k** itself is invariant. If $\mathbf{k}=\mathbf{0}$, $\varphi(\mathbf{i},\mathbf{j})$ is also invariant and there are again two orbits corresponding to even and odd pin structures. If $\mathbf{k}\neq\mathbf{0}$, there are 2^{n-1} orbits, characterized by **k**. In Table II we list all the orbits of $\Omega_B(\Sigma_{g,n})$; the invariant vectors **k** are ordered as if they were binary numbers.

IV. GENERATORS FOR $\Omega(N_{g,n})$

We saw in Sec. II that the double covering of $N_{g,n}$ is $\Sigma_{g-1,2n}$. Every diffeomorphism of $N_{n,g}$ lifts to two diffeomorphisms of $\Sigma_{g-1,2n}$ which commute with J. One of these preserves the orientation and the other, being obtained from the first by composition with J, reverses the orientation. Conversely, every diffeomorphism of $\Sigma_{g-1,2n}$ which commutes with J factors to a diffeomorphism of $N_{g,n}$. Thus there is an isomorphism between $D(N_{g,n})$ and $D^+(\Sigma_{g-1,2n}) \cap C(\mathbb{Z}_2)$, where $C(\mathbb{Z}_2)$ denotes the centralizer of $\mathbb{Z}_2 = \{\mathrm{Id}, J\}$ in $D(\Sigma_{g-1,2n})$. There follows that the generators of $\Omega(N_{g,n})$ can be represented by diffeomorphisms of $\Sigma_{g-1,2n}$ which are isotopic to diffeomorphisms which commute with J. On the other hand, if we have a diffeomorphism f such that $\mathrm{supp} f \cap J(\mathrm{supp} f) = \varnothing$ and which does not commute with J, we define a "symmetrized" diffeomorphism $\hat{f}$,

$$\hat{f} = fJfJ\,.$$

Since $\mathrm{supp}\, JfJ = J(\mathrm{supp} f)$, f commutes with JfJ and $\hat{f}$ commutes with J.

Consider first the twists around the following loops (which do not meet the plane $x^1 = 0$): a_A for $A = 1,...,[g/2]$; b_A for $A = 1,...,[(g-1)/2]$; c_A for $A = 1,...,[(g-2)/2]$; r_{Ah} for $A = 1,...,[g/2]$, $h = 1,...,n$; s_{Ah} for $A = 1,...,[(g-1)/2]$, $h = 1,...,n$, and t_h for $h = 1,...,n-1$.

The corresponding symmetrized diffeomorphisms $\hat{f}[a_A]$, $\hat{f}[b_A]$, $\hat{f}[c_A]$, $\hat{f}[r_{Ah}]$, $\hat{f}[s_{Ah}]$, and $\hat{f}[t_h]$ project onto the diffeomorphism of $N_{g,n}$ which will be denoted again $f[a_A]$, $f[b_A]$, $f[c_A]$, $f[r_{Ah}]$, $f[s_{Ah}]$, and $f[t_h]$ since they are twists and can be described in local coordinates on $N_{g,n}$ by Eqs. (3.1)–(3.3). Indeed, $N_{g,n}$ can be identified with the intersection of $\Sigma_{g-1,2n}$ and the half-space $x^1 > 0$ with certain identifications if $x^1 = 0$. Therefore every symmetrized diffeomorphism of $\Sigma_{g-1,2n}$ whose support does not intersect the plane $x^1 = 0$ can be immediately regarded as a diffeomorphism on $N_{g,n}$.

Now consider the loop t_n, which we draw as in Fig. 6. The twist $f[t_n]$ can be chosen to commute with J, but its support crosses the plane $x^1 = 0$ and it does not project to a twist on $N_{g,n}$. The diffeomorphism of $N_{g,n}$ defined by $f[t_n]$ will be denoted σ_n and called a "slide," since it can be described as sliding D_n through a Möbius strip and back to its original position (see Appendix B); notice that σ_n reverses the orientation of the boundary d_n. Similarly the composition of twists $\hat{f}[t_h]^{-1}\cdots\hat{f}[t_{n-1}]^{-1}f[t_n]\hat{f}[t_{n-1}] \times\cdots\hat{f}[t_h]$ projects onto a slide σ_h that reverses the orientation of d_h.

Next, consider the twists around the remaining loops r_{Ah} and s_{Ah}. It can be shown that for $A = [(g+2)/2],...,g-1$; $h = 1,...,n$, $\hat{f}[r_{Ah}]$ and $\hat{f}[s_{Ah}]$ project onto transformations isotopic to $\sigma_h^{-1}\circ\hat{f}[r_{g-A,h}]\circ\sigma_h$ and $\sigma_h^{-1}\circ\hat{f}[s_{g-A,h}]\circ\sigma_h$, respectively. For g even, the projection of $\hat{f}[s_{g/2,h}]$ is isotopic to σ_h^2. So these twists do not produce any new generator of $\Omega(N_{g,n})$.

The only generators of $\Omega(\Sigma_{g-1,2n})$ that have not been used until now are the twists around the loops that cover the orientation-reversing loops, namely $f[c_{(g-1)/2}]$ for g odd and $f[b_{g/2}]$ for g even. They cannot be symmetrized and have to be replaced by a new type of transformation, called "Y diffeomorphisms." Since they appear already in the case when there is no boundary, we begin by assuming $n = 0$. The support of a Y diffeomorphism is any closed subset $W \subset N_g$ homeomorphic to $N_{2,1}$ (a Klein bottle without a disk); in particular this implies that Y diffeomorphisms exist only if $g \geqslant 2$. For any such subset W one can define, up to isotopy, a Y diffeomorphism, in a way which is described in Appendix B. For our purposes, it will be more useful to have a descrip-

TABLE II. Orbits of $\Omega_\partial(\Sigma_{g,n})$ and $\Omega_B(\Sigma_{g,n})$.

k	φ	Standard form	Cardinality
0,...,0	0	(0,...,0; 1,...,1;0,...,0)	$2^{g-1}(2^g+1)$
0,...,0	1	(0,...,0;1,...,10;0,...,0)	$2^{g-1}(2^g-1)$
0,...,0,1,1	—	(0,...,0; 1,...,1;0,...,0,1,1)	2^{2g}
0,...,0,1,0,1	—	(0,...,0; 1,...,1;0,...,0,1,0,1)	2^{2g}
...	...	...	...
1,...,1, $\theta(n+1)$	—	(0,...,0; 1,...,1;1,...,1,$\theta(n+1)$)	2^{2g}

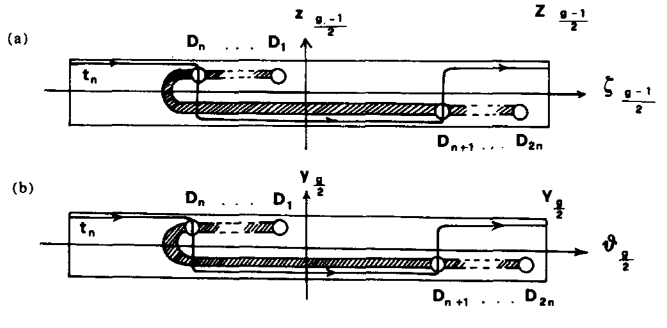


FIG. 6. The loop t_n and the (dashed) region where the functions r_{ijk} have to be modified in the presence of a boundary.

tion of the lift of a Y diffeomorphism to Σ_{g-1}; furthermore, we can restrict our attention to a Y diffeomorphism with a fixed support, since all the others can then be obtained by composition with twists. Consider the subset $\widehat{W}$ of Σ_{g-1} which is bounded by the curves f and $J(f)$, if g is even, and e and $J(e)$ if g is odd (see Fig. 2). These subsets are homeomorphic to $\Sigma_{1,2}$ (a torus without two disks) and project onto subsets W of N_g which are homeomorphic to $N_{2,1}$. We define diffeomorphisms y of Σ_{g-1} with support $\widehat{W}$. Regard $\widehat{W}$ as a subset of Σ_1 embedded in $\mathbb{R}^3$ as in Fig. 7.

The diffeomorphism y is defined by

$$y = J \circ T \circ K_3 ,$$

where T is a rigid rotation by π of the disks D_1 and D_2 in the direction of the arrows, joined smoothly to the identity on the region which is bound by the dashed lines. Since y is the identity on D_1 and D_2, when $\widehat{W}$ is regarded as a subset of Σ_{g-1}, this transformation can be extended smoothly by the identity in the complement of $\widehat{W}$.

The diffeomorphisms y commute with J and therefore project to diffeomorphisms of N_g with support W which are shown in Appendix B to be Y diffeomorphisms in the sense of Ref. 5. It can be seen that y is isotopic to the following combinations of twists:

$$y \simeq (f[c_{(g-1)/2}] \circ f[b_{(g-1)/2}]^{-1} \circ f[b_{(g+1)/2}]^{-1})^2 \circ f[e]^{-1}, \quad g \text{ odd},$$

$$y \simeq (f[b_{g/2}] \circ f[c_{g/2}] \circ f[c_{(g-2)/2}])^2 \circ f[f]^{-1}, \quad g \text{ even}.$$

However, these combinations of twists do not commute with J and y cannot be expressed as a combination of twists on N_g.[6]

In the case of surfaces with nonempty boundary, it is convenient to choose the support of y in such a way that y does not move the boundary. In addition, in order to apply the method of Ref. 1 to the transformation of pin structures on $N_{g,n}$, it is necessary that the support of every diffeomorphism of $\Sigma_{g-1,2n}$ representing a generator of $\Omega(N_{g,n})$ be contained entirely in the domain of a local trivialization of the bundle of frames (specifically, U). These two requirements can be met by deforming the curves e and f of Fig. 2 in such a way that they run parallel and sufficiently close (at a distance $<\varepsilon/2$ in the charts X,Y,Z) to the curves $b_{(g-1)/2}$, $c_{(g-1)/2}$, $b_{(g+1)/2}$ and $c_{(g-2)/2}$, $b_{g/2}$, $c_{g/2}$, respectively.

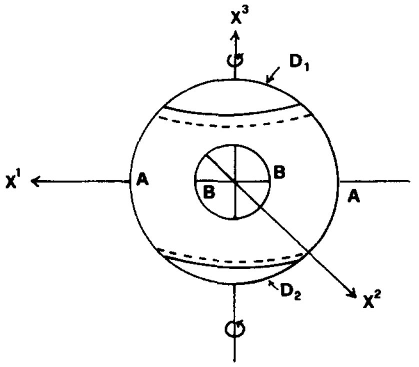


FIG. 7. The support of the diffeomorphism y.

Then, W can be regarded as a tubular neighborhood of these curves.

For $n=0$, it was shown in Refs. 6 and 7 that $\Omega(N_g)$ is generated by the following set $\{f[a_A], f[b_A]$ for $A=1,\dots,(g-1)/2$; $f[C_A]$ for $A=1,\dots,(g-3)/2; y\}$ for g odd$\{f[a_A]$ for $A=1,\dots,g/2$; $f[b_A]$ for $A=1,\dots,(g-2)/2; f[c_A]$ for $A=1,\dots,(g-2)/2; y\}$ for g even.

Using the results of Scott[8] it can be shown that $\Omega(N_{g,n})$ is generated by the sets given above plus the following: $\{f[r_{Ah}]$ for $A=1,\dots,[g/2]$; $h=1,\dots,n$; $f[s_{Ah}]$ for $A=1,\dots,[(g-1)/2], h=1,\dots,n; f[t_h]$ for $h=1,\dots,n-1$; σ_h for $h=1,\dots,n\}$. In the special cases $g=1$ or $n=0$, $n=1$, some of these generators are not defined and $\Omega(N_{g,n})$ is generated by the remaining ones [in particular $\Omega(N_{1,0})$ is trivial].

The generators for the subgroups $\Omega_B(N_{g,n})$ and $\Omega_\partial(N_{g,n})$ are obtained from those of $\Omega(N_{g,n})$ omitting the generators $\{f[t_h]$ for $h=1,\dots,n-1\}$ and the generators $\{f[t_h]$ for $h=1,\dots,n-1$; σ_h for $h=1,\dots,n\}$, respectively.

V. PIN STRUCTURES ON $N_{g,n}$

The main result of this section will be the following.

Proposition: A nonorientable surface of odd genus and without boundary admits only $\text{Pin}^-(2)$ structures. In all other cases $N_{g,n}$ admits both $\text{Pin}^-(2)$ and $\text{Pin}^+(2)$ structures.

In the course of the proof we shall explicitly construct all the inequivalent pin structures on $N_{g,n}$ as quotients of pin structures on the double covering; this will provide a natural way of labeling them and we shall see that their number agrees with the general cohomological result.

Let $p: F \to N_{g,n}$ be the bundle of frames of $N_{g,n}$ and $p_\Sigma: F_\Sigma \to \Sigma_{g-1,2n}$ be the bundle of frames of $\Sigma_{g-1,2n}$. There is a canonical isomorphism of F_Σ to the pullback $\pi^*F = \{(x,e) \in \Sigma_{g-1,2n} \times F \mid \pi(x) = p(e)\}$ given by $e \mapsto (p_\Sigma(e), T\pi(e))$. Let $\mathbb{Z}_2$ be the subgroup of Aut F_Σ generated by TJ. There is a canonical isomorphism of the quotient $F_\Sigma/\mathbb{Z}_2$ to F given by $[e] \mapsto T\pi(e)$. In the following we shall identify objects which are related by these canonical isomorphisms.

If $(\widetilde{F},\eta)$ is a $\text{Pin}^+(2)$—or a $\text{Pin}^-(2)$—structure on $N_{g,n}$, we can construct, respectively, a $\text{Pin}^+(2)$—or a $\text{Pin}^-(2)$—structure $(\widetilde{F}_\Sigma, \eta_\Sigma)$ on $\Sigma_{g-1,2n}$ as follows:

$$\widetilde{F}_\Sigma = \pi^*\widetilde{F} = \{(x,\tilde{e}) \in \Sigma_{g-1,2n} \times \widetilde{F} \mid \pi(x) = \tilde{p}(e)\}$$

with $\tilde{p} = p \circ \eta$ and $\eta_\Sigma(x,\tilde{e}) = (x, \eta(\tilde{e}))$. This pin structure is invariant under J, in the sense that there exists a lift of J to an automorphism $\widetilde{TJ}$ of $\widetilde{F}_\Sigma$ such that $\eta_\Sigma \circ \widetilde{TJ} = TJ \circ \eta_\Sigma$, and furthermore

$$(\widetilde{TJ})^2 = \text{Id}_{\widetilde{F}_\Sigma} . \tag{5.1}$$

For instance, we can take

$$\widetilde{TJ}(x,\tilde{e}) = (J(x), \tilde{e}) . \tag{5.2}$$

We observe that since $(TJ)^2 = \text{Id}_{F_\Sigma}$ the only other possibility is $(\widetilde{TJ})^2 = \gamma$, the automorphism of $\widetilde{F}_\Sigma$ given by right multiplication with $-\mathbf{1}$. Conversely, if a $\text{Pin}^+(2)$—or $\text{Pin}^-(2)$—structure $(\widetilde{F}_\Sigma, \eta_\Sigma)$ on $\Sigma_{g-1,2n}$ is invariant under J and (5.1) holds, we can define, respectively, a $\text{Pin}^+(2)$—

or Pin$^-$(2)—structure $(\tilde{F},\eta)$ on $N_{g,n}$ as follows: we take $\tilde{F}=\tilde{F}_\Sigma/\tilde{\mathbb{Z}}_2$ where $\tilde{\mathbb{Z}}_2=\{\mathrm{Id}_{\tilde{F}_\Sigma},\tilde{T}J\}$ and define η: $\tilde{F}\to F$ by $\eta([\tilde{e}]_{\tilde{\mathbb{Z}}_2})=[\eta_\Sigma(\tilde{e})]_{\mathbb{Z}_2}$. This gives us a method to construct pin structures on $N_{g,n}$ from pin structures on $\Sigma_{g-1,2n}$.

The quotient of the pullback of a pin structure $(\tilde{F},\eta)$ under the group $\tilde{\mathbb{Z}}_2$ generated by (5.2) is identical to $(\tilde{F},\eta)$. It follows that every pin structure on $N_{g,n}$ can be obtained by the method described above from some pin structure on $\Sigma_{g-1,2n}$.

The quotient of inequivalent pin structures on $\Sigma_{g-1,2n}$ yields inequivalent pin structures on $N_{g,n}$. This will be seen in each specific case later on and can be proven using covering space methods.[9] As discussed in general in Ref. 1, there exist precisely two lifts of J, which differ by composition with γ. If (5.1) holds for one of them, it also holds for the other, so they generate two different subgroups of Aut $\tilde{F}_\Sigma$ which we denote $\tilde{\mathbb{Z}}_2^{(\ell)}$ with $\ell\in\{0,1\}$. Taking the quotient of a symmetric pin structure by these two subgroups yields inequivalent pin structures on $N_{g,n}$. It can be seen, however, that the bundles $\tilde{F}_\Sigma/\tilde{\mathbb{Z}}_2^{(\ell)}$ for $\ell=0,1$ are isomorphic as pin bundles. The problem of the isomorphism between the pin bundles corresponding to different pin structures will be discussed elsewhere.

The action of J on pin structures can be determined again using the method of Ref. 1. We have to extend the bundle chart on U to a bundle chart on $U\cup J(U)$. This can be done in a straightforward manner by choosing the frames on $J(X_A)$ to be given by

$$e(J(p))=TJ(e(p))\begin{pmatrix}-1&0\\0&1\end{pmatrix}$$

for $p\in X_A$. For the pin structures we assume $r_{ijk}(p)=1$ for $p\in J(U)\setminus U$. In the chart $Z_{(g-1)/2}$ (for g odd) and $Y_{g/2}$ (for g even) the function r_{ijk} have to be modified in such a way that when composed with the loops $b_{(g-1)/2}$, $b_{(g+1)/2}$, $d_1,\dots,d_{2n}$ they have winding numbers $j_{(g-1)/2}$, $j_{(g+1)/2}$, $k_1,\dots,k_{2n}$ (for g odd) and when composed with the loops $b_{g/2}$, $a_{g/2}$, $d_1,\dots,d_{2n}$ they have winding numbers $j_{g/2}$, $i_{g/2}$, $k_1,\dots,k_{2n}$ (for g even), with the relation $\Sigma_{h=1}^{2n}k_h=0 \bmod 2$ in both cases. This is achieved by changing r_{ijk} on the shaded strip in Fig. 6 in such a way that crossing this strip between d_h and d_{h+1} the function r_{ijk} rotates by $2\pi q_h$ with $q_h=\Sigma_{m=1}^{h}k_m$.

Counting winding numbers, we find that the action of J is $i_A\mapsto i_{g-A}$; $j_A\mapsto j_{g-A}+(\delta_{A,(g-1)/2}+\delta_{A,(g+1)/2})q_n$; $k_h\mapsto k_{2n-h+1}$ for g odd and $i_A\mapsto i_{g-A}+\delta_{A,g/2}q_n$; $j_A\mapsto j_{g-A}$; $k_h\mapsto k_{2n-h+1}$ for g even. Therefore the symmetric pin structures, i.e., those which are invariant under J, are characterized by

$$\begin{aligned}
i_A&=i_{g-A},\\
j_A&=j_{g-A}+(\delta_{A,(g-1)/2}+\delta_{A,(g+1)/2})\sum_{h=1}^{n}k_h,\\
k_h&=k_{2n-h+1},
\end{aligned}\tag{5.3a}$$

for g odd, and

$$\begin{aligned}
i_A&=i_{g-A}+\delta_{A,g/2}\sum_{h=1}^{n}k_h,\\
j_A&=j_{g-A},\\
k_h&=k_{2n-h+1},
\end{aligned}\tag{5.3b}$$

for g even. In particular, the first equation in (5.3b) written for $A=g/2$ implies that for g even

$$q_n=\sum_{h=1}^{n}k_h=0.\tag{5.4}$$

We now examine which invariant pin structures satisfy condition 5.1. This will determine whether Pin$^+$(2)—or Pin$^-$(2)—structures exist on $N_{g,n}$.

In local trivializations of F_Σ and $\tilde{F}_\Sigma$ over U defined by frames e and $\tilde{e}$, the lifts TJ and $\tilde{T}J$ are locally represented by functions h and $\tilde{h}$ which, due to our choices of trivialization, have values in O(2) and Pin$^\pm$(2), respectively. Condition 5.1 then reads

$$\tilde{h}(J(p))\tilde{h}(p)=1\tag{5.5}$$

[the case $(\tilde{T}J)^2=\gamma$ corresponds to having -1 on the rhs]. Since we have a choice between two possibilities, it will suffice to check (5.5) at a single point.

If g is odd, let p be the point whose coordinates in the chart $Z_{(g-1)/2}$ are $(\zeta,z)=(0,0)$ [see Fig. 6(a)]. In this chart J: $(\zeta,z)\mapsto(\zeta+\pi,-z)$, so $J(p)=(\pi,0)$, $h(0,0)=h(\pi,0)=\begin{pmatrix}-1&0\\0&1\end{pmatrix}$, and $\tilde{h}(0,0)=\pm\gamma_1$, $\tilde{h}(\pi,0)=\pm\gamma_1$.

The relation between the sign of h at the two points has to be determined by continuity. We begin by considering the case $n=0$. Using (4.10) in Ref. 1 we find

$$h(\zeta,0)=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\times R\left(-2\lambda\left(\frac{\zeta+\pi-\varepsilon}{\pi-2\varepsilon}\right)-2\lambda\left(\frac{\zeta-\varepsilon}{\pi-2\varepsilon}\right)-\pi\right).\tag{5.6}$$

Using (4.14) and the coordinate transformation (4.2), (4.3) in Ref. 1, we have

$$r_{ij}(\zeta,0)=R\left(2i_{(g-1)/2}\lambda\left(-\frac{\zeta}{\varepsilon}\right)+2i_{(g+1)/2}\lambda\left(\frac{\zeta+\pi}{\varepsilon}\right)\right).$$

Since the pin structure is symmetric, we can put $i_{(g-1)/2}=i_{(g+1)/2}$. Then from (4.21) in Ref. 1 written for $f=J$,

$$\begin{aligned}
\rho(\tilde{h}(\zeta,0))&=r_{ij}(\zeta+\pi,0)^{-1}h(\zeta,0)r_{ij}(\zeta,0)\\
&=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\times R\left(-2\lambda\left(\frac{\zeta+\pi-\varepsilon}{\pi-2\varepsilon}\right)-2\lambda\left(\frac{\zeta-\varepsilon}{\pi-2\varepsilon}\right)-\pi\right)\\
&\quad\times R\left(2i_{(g-1)/2}\left[\lambda\left(\frac{\zeta}{\varepsilon}\right)+\lambda\left(-\frac{\zeta}{\varepsilon}\right)+\lambda\left(\frac{\zeta+\pi}{\varepsilon}\right)+\lambda\left(-\frac{\zeta+\pi}{\varepsilon}\right)\right]\right).
\end{aligned}\tag{5.7}$$

If $i_{(g-1)/2}=0$ the second rotation is the identity matrix; if $i_{(g-1)/2}=1$ it rotates by 2π and back as ζ grows from 0 to π. Therefore, in both cases, this is a path in SO(2) which starts and ends at the identity and is homotopically trivial. However, when ζ grows from 0 to π, the argument of the first rotation decreases by 2π, and therefore $\tilde{h}$ must change sign: if $\tilde{h}(0,0)=\gamma_1$ then $\tilde{h}(\pi,0)=-\gamma_1$ and vice versa. So

$\tilde{h}(\pi,0)\tilde{h}(0,0) = -(\gamma_1)^2$ and condition (5.5) holds for the group $\mathrm{Pin}^-(2)$. Therefore, if g is odd, N_g admits only $\mathrm{Pin}^-(2)$ structures. (See Ref. 10 for $g = 1$.)

Now consider the case $n \geqslant 1$. If $q_n = \Sigma_{h=1}^n k_h = 0$ then (5.6) and (5.7) remain valid and we obtain a $\mathrm{Pin}^-(2)$ structure on $N_{g,n}$. If $q_n = 1$, the second rotation matrix in (5.7) will have to be modified for those values of ζ such that $(\zeta + \pi,0)$ is in the shaded strip in Fig. 6(a); as a result of this modification, it will rotate by 2π when ζ grows from 0 to π. Altogether the rhs of (5.7) rotates by 4π, so $\tilde{h}(\pi,0)\tilde{h}(0,0) = (\gamma_1)^2$ and condition (5.5) holds for the group $\mathrm{Pin}^+(2)$.

If g is even, let p be the point whose coordinates in the chart $Y_{g/2}$ are $(\vartheta,y) = (\pi/2,0)$ [see Fig. 6(b)]. Then $J(p)$ has coordinates $(-\pi/2,0)$ and the function h: $Y_{g/2} \to \mathrm{GL}(2)$ which represents locally TJ is $h(\vartheta,y) = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$. Using (4.14), (4.16), (4.21) of Ref. 1 and (5.4) we have

$$\rho(\tilde{h}(\vartheta,0)) = r_{\mathbf{ijk}}(\vartheta + \pi,0)^{-1} h(\vartheta,0) r_{\mathbf{ijk}}(\vartheta,0)$$

$$= \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} R\left(2j_{g/2}\left[\lambda\left(\frac{\vartheta+\pi}{\varepsilon}\right) + \lambda\left(\frac{\vartheta}{\varepsilon}\right)\right]\right).$$

In particular $\rho(\tilde{h}(\pm\pi/2,0)) = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$ so $\tilde{h}(\pi/2,0) = \pm\gamma_2$ and $\tilde{h}(-\pi/2,0) = \pm\gamma_2$. To see how the sign of $\tilde{h}$ at $\vartheta = -\pi/2$ is related to the sign at $\vartheta = \pi/2$ we use again a continuity argument. If $j_{g/2} = 0$, $\rho(\tilde{h}(\vartheta,0))$ is constant and therefore also $\tilde{h}$ is constant. So $\tilde{h}(-\pi/2,0)\tilde{h}(\pi/2,0) = (\gamma_2)^2$ and condition (5.5) holds for the group $\mathrm{Pin}^+(2)$. If $j_{g/2} = 1$, $\rho(\tilde{h})$ rotates by 2π as ϑ grows continuously from $-\pi/2$ to $\pi/2$ and therefore $\tilde{h}$ must change sign. So if we choose $\tilde{h}(\pi/2,0) = \gamma_2$ we must have $\tilde{h}(-\pi/2,0) = -\gamma_2$ and vice versa; there follows that $\tilde{h}(-\pi/2,0)\tilde{h}(\pi/2,0) = -(\gamma_2)^2$ and the condition (5.5) holds for the group $\mathrm{Pin}^-(2)$. Altogether, we find that if g is even, a symmetric $\mathrm{Pin}^+(2)$ structure on $\Sigma_{g-1,2n}$ projects to a $\mathrm{Pin}^+(2)$ structure on $N_{g,n}$ only if $j_{g/2} = 0$ and a symmetric $\mathrm{Pin}^-(2)$ structure on $\Sigma_{g-1,2n}$ projects to a $\mathrm{Pin}^-(2)$ structure on $N_{g,n}$ only if $j_{g/2} = 1$. This completes the proof of the proposition.

We now count the pin structures that we have constructed. We consider first the case $n \geqslant 1$. If g is odd the symmetric pin structures on $\Sigma_{g-1,2n}$ have labels $(i_1,\ldots,i_{(g-1)/2}, i_{(g-1)/2},\ldots,i_1;\ j_1,\ldots,j_{(g-1)/2},j_{(g-1)/2},\ldots,j_1;\ k_1,\ldots,k_n, k_n,\ldots,k_1)$ with $\Sigma_{h=1}^n k_h = 1$ for $\mathrm{Pin}^+(2)$ and $\Sigma_{h=1}^n k_h = 0$ for $\mathrm{Pin}^-(2)$. Thus there are 2^{g+n-2} symmetric $\mathrm{Pin}^+(2)$ structures and 2^{g+n-2} symmetric $\mathrm{Pin}^-(2)$ structures. If g is even, the symmetric pin structures on $\Sigma_{g-1,2n}$ have labels $(i_1,\ldots,i_{(g-2)/2},i_{g/2},i_{(g-2)/2},\ldots,i_1;\ j_1,\ldots,j_{(g-2)/2},j_{g/2}, j_{(g-2)/2},\ldots,j_1;\ k_1,\ldots,k_n,k_n,\ldots,k_1)$ with $j_{g/2} = 0$ for $\mathrm{Pin}^+(2)$ and $j_{g/2} = 1$ for $\mathrm{Pin}^-(2)$, and with the relation (5.4) holding. Thus there are 2^{g+n-2} symmetric $\mathrm{Pin}^+(2)$ structures and 2^{g+n-2} symmetric $\mathrm{Pin}^-(2)$ structures. From the general discussion earlier in this section, each symmetric pin structure on $\Sigma_{g-1,2n}$ gives rise to two inequivalent pin structures on $N_{g,n}$ which differ in the value of the label ℓ. Therefore in all cases we will get 2^{g+n-1} inequivalent pin structures on $N_{g,n}$. If $n = 0$ there are 2^{g-1} symmetric pin structures of each type on Σ_{g-1} and they give rise to 2^g inequivalent pin structures on N_g.

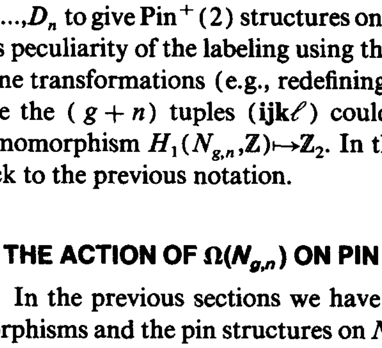
Because of the symmetry (5.3), we use as labels for pin structures on $N_{g,n}$ the following: $(i_1,\ldots,i_{(g-1)/2};\ j_1,\ldots,j_{(g-1)/2};\ k_1,\ldots,k_n;\ell)$ with $\Sigma_{h=1}^n k_h = 0$ for $\mathrm{Pin}^-(2)$ and $\Sigma_{h=1}^n k_h = 1$ for $\mathrm{Pin}^+(2)$, if g is odd, and $(i_1,\ldots,i_{g/2};\ j_1,\ldots,j_{(g-2)/2};k_1,\ldots,k_n;\ell)$ with $\Sigma_{h=1}^n k_h = 0$ for g even. The labels i_A, j_A, k_h, and ℓ are associated to the homology generators on $N_{g,n}$: a_A, b_A, d_h and the orientation reversing loop, respectively. Differences in these labels can be interpreted as the value on the corresponding loop of a homomorphism from $H_1(N_{g,n},\mathbb{Z})$ to $\mathbb{Z}_2$.

The fact that for the group $\mathrm{Pin}^+(2)$ on $N_{g,n}$ with g odd $\Sigma_{h=1}^n k_h \neq 0$ shows that the labels ($\mathbf{ijk}\ell$) cannot be interpreted naively as a homomorphism from $H_1(N_{g,n},\mathbb{Z})$ to $\mathbb{Z}_2$, as is usual. The fact that at least one of the labels k_h must be nonzero is equivalent to the statement that these $\mathrm{Pin}^+(2)$ structures cannot be extended to the interior of the disks $D_1,\ldots,D_n$ to give $\mathrm{Pin}^+(2)$ structures on N_g. We could correct this peculiarity of the labeling using the freedom to perform affine transformations (e.g., redefining $k_1 \mapsto k_1 + 1$). In this case the $(g+n)$ tuples ($\mathbf{ijk}\ell$) could be interpreted as a homomorphism $H_1(N_{g,n},\mathbb{Z}) \mapsto \mathbb{Z}_2$. In the following we shall stick to the previous notation.

VI. THE ACTION OF $\Omega(N_{g,n})$ ON PIN STRUCTURES

In the previous sections we have described the diffeomorphisms and the pin structures on $N_{g,n}$ in terms of diffeomorphisms and pin structures on its orientable double cover $\Sigma_{g-1,2n}$. In this section we shall use the results of Sec. III, together with some additional information, to determine the action of diffeomorphisms on pin structures on $N_{g,n}$.

Let $f \in D(N_{g,n})$ and $f_\Sigma \in D(\Sigma_{g-1,2n})$ be one of its two lifts. Given a pin structure $(\tilde{F},\eta)$ on $N_{g,n}$, there is a pin structure $(\tilde{F}_\Sigma,\eta_\Sigma)$ on $\Sigma_{g-1,2n}$ such that $(\tilde{F},\eta)$ is its quotient under a group $\tilde{\mathbb{Z}}_2$ generated by one of the two lifts $\tilde{T}J$. The map f_Σ transforms $(\tilde{F}_\Sigma,\eta_\Sigma)$ into some other pin structure $(\tilde{F}_\Sigma,\eta'_\Sigma)$; so in the following diagram all solid arrows commute:



(6.1)

The transformed pin structure $(\tilde{F}',\eta')$ is defined by the requirement that there exists a map $\tilde{T}f: \tilde{F}' \to \tilde{F}$ (dashed arrow) which forms a commutative square with η',η and Tf. From the square on the left-hand side of the upper cube, $(\tilde{F}',\eta')$ must be the quotient under a group $\tilde{\mathbb{Z}}_2$ of $(\tilde{F}_\Sigma,\eta'_\Sigma)$; but $(\tilde{F}_\Sigma,\eta'_\Sigma)$ projects to two inequivalent pin structures on $N_{g,n}$ which differ in the value of the label ℓ, so (6.1) does not entirely determine the transform of $(\tilde{F},\eta)$ under f.

To determine the transformation rules of the label ℓ, we observe that since Tf_Σ commutes with TJ we must have

$$\tilde{T}f_\Sigma \circ \tilde{T}J = (\gamma)^{q_0} \tilde{T}J \circ \tilde{T}f_\Sigma \tag{6.2}$$

with $q=0$ or $q=1$. If $\tilde{T}f_\Sigma$ commutes with $\tilde{T}J(q=0)$, then we can define $\tilde{T}f$ by

$$\tilde{T}f([\tilde{e}])=[\tilde{T}f_\Sigma(\tilde{e})],$$

where [ ] denotes equivalence classes with respect to the group $\tilde{\mathbb{Z}}_2$ generated by $\tilde{T}J$; if $\tilde{T}f_\Sigma$ "anticommutes" with $\tilde{T}J$ $(q=1)$ the same definition works provided the equivalence class $[\tilde{e}]$ on the lhs is taken with respect to the group $\tilde{\mathbb{Z}}_2$ generated by the other lift of J, namely, $\gamma\circ\tilde{T}J$. This means that under f

$$\ell\mapsto\ell'=\ell+q. \tag{6.3}$$

We now consider the transformation properties of pin structures on $\Sigma_{g-1,2n}$ (not necessarily symmetric) under the lifts of diffeomorphisms of $N_{g,n}$. The transformation of the labels **ijk** under the twists $\hat{f}[a_A],\hat{f}[b_A],\hat{f}[c_A],\hat{f}[r_{Ah}],\hat{f}[s_{Ah}],\hat{f}[t_h]$ can be immediately obtained from the results of Sec. III (the position of the boundaries on the surface does not affect the transformation properties of pin structures).

The effect of $f[t_n]$ is only to interchange k_n with k_{n+1} and to transform $j_{(g-1)/2}\mapsto j_{(g-1)/2}+k_{n+1}$, $j_{(g+1)/2}\mapsto j_{(g+1)/2}+k_n$ if g is odd and $i_{g/2}\mapsto i_{g/2}+k_n$, $j_{g/2}\mapsto j_{g/2}+k_n+k_{n+1}$ if g is even. The support of y is chosen as in Sec. IV in such a way that y does not move the boundary. Therefore, y leaves **k** invariant and it transforms $i_{(g-1)/2}\mapsto i_{(g+1)/2}+\Sigma_{h=1}^n k_h$, $i_{(g+1)/2}\mapsto i_{(g-1)/2}+\Sigma_{h=1}^n k_h$, $j_{(g-1)/2}\mapsto j_{(g+1)/2}+\Sigma_{h=1}^n k_h$, and $j_{(g+1)/2}\mapsto j_{(g-1)/2}+\Sigma_{h=1}^n k_h$ for g odd and $i_{g/2}\mapsto i_{g/2}+i_{(g-2)/2}+i_{(g+2)/2}$, $j_{(g-2)/2}\mapsto j_{(g-2)/2}+j_{g/2}+1$, and $j_{(g+2)/2}\mapsto j_{(g+2)/2}+j_{g/2}+1$ for g even.

These transformations simplify somewhat on symmetric pin structures. In particular, one finds then that for g even, $j_{g/2}$ is invariant and for g odd $\Sigma_{h=1}^n k_h$ is invariant, in accordance with the fact that the structure group $\mathrm{Pin}^\pm(2)$ cannot be changed.

To determine the transformation of the label ℓ we rewrite (6.2) in the local trivialization of $\tilde{F}_\Sigma$ over U,

$$\tilde{h}_f(J(x))\tilde{h}_J(x)=(-)^q\tilde{h}_J(f_\Sigma(x))\tilde{h}_f(x), \tag{6.4}$$

where $\tilde{h}_f$ and $\tilde{h}_J$ are the local representatives of $\tilde{T}f_\Sigma$ and $\tilde{T}J$, respectively. Since we have a choice between two possibilities, it is sufficient to check this formula at a single point. All our generators are such that there exists a point $p\in U$ which is not in the support of f_Σ and such that $r_{\mathbf{ijk}}(p)=\mathbf{1}$ for all **ijk**. Then $\tilde{h}_f(p)$ and $\tilde{h}_f(J(p))$ must be either $\mathbf{1}$ or $-\mathbf{1}$, and furthermore $\tilde{h}_J(f_\Sigma(p))=\tilde{h}_J(p)$. The value of $\tilde{h}_f$ at p can be fixed arbitrarily to be $\mathbf{1}$; the value of $\tilde{h}_f$ at $J(p)$ is determined by continuity. In particular, if there exists a path in U joining p to $J(p)$ which lies entirely in the complement of the support of f, then $\tilde{h}_f(J(p))=\tilde{h}_f(p)=\mathbf{1}$ and therefore $q=0$. Direct inspection shows that this is the case for all our generators except $\hat{f}[a_{g/2}]$ for g even.

To determine the transformation of ℓ under $\hat{f}[a_{g/2}]$ we choose the point p to have coordinates $(\vartheta,y)=(\pi/2,0)$ in the chart $Y_{g/2}$ as in the discussion of Sec. V. Then we have $\tilde{h}_J(p)=\tilde{h}_J(\hat{f}[a_{g/2}](p))=\gamma_2$ and $\tilde{h}_f(p)=\pm\mathbf{1}$, $\tilde{h}_f(J(p))=\pm\mathbf{1}$. The relative sign of $\tilde{h}_f$ at p and $J(p)$ is the winding number of the function

$$\rho(\tilde{h}_f(x))=r_{\mathbf{ijk}}(\hat{f}[a_{g/2}](x))h_f(x)r_{\mathbf{ijk}}(x)$$

along the path $s\mapsto(\vartheta,y)=(-s,0)$ with $-\pi/2\leqslant s\leqslant\pi/2$. This winding number is $i_{g/2}$, so we get $\tilde{h}_f(J(x))\tilde{h}_J(x)=(-)^{i_{g/2}}\gamma_2$ and $\tilde{h}_J(\hat{f}[a_{g/2}](x))\cdot\tilde{h}_f(x)=\gamma_2$. Comparing with (6.4) we find $q=i_{g/2}$. We now collect the transformation rules for pin structures on $N_{g,n}$ under the generators of $\Omega(N_{g,n})$ which were listed at the end of Sec. IV. We use the labeling of pin structures which was discussed at the end of Sec. V. Since most of the labels are invariant, we only list those which are changed. For g odd,

$$\begin{array}{ll}
f[a_B]: & j_A\mapsto j_A+\delta_{AB}i_B,\quad B=1,\ldots,(g-1)/2,\\
f[b_B]: & i_A\mapsto i_A+\delta_{AB}j_B,\\
f[c_B]: & j_A\mapsto j_A+(\delta_{AB}+\delta_{A,B+1})(i_B+i_{B+1}+1),\quad B=1,\ldots,(g-3)/2,\\
y: & i_{(g-1)/2}\mapsto i_{(g-1)/2}-\sum_{h=1}^n k_h,\\
f[r_{Bm}]: & j_A\mapsto j_A+\delta_{AB}k_m,\quad B=1,\ldots,(g-1)/2,\quad m=1,\ldots,n,\\
f[s_{Bm}]: & i_A\mapsto i_A+\delta_{AB}k_m,\\
f[t_m]: & k_h\mapsto k_h+(\delta_{hm}+\delta_{h,m+1})(k_m+k_{m+1}),\quad m=1,\ldots,n-1,\\
\sigma_h: & j_{(g-1)/2}\mapsto j_{(g-1)/2}+k_h,\quad h=1,\ldots,n,
\end{array} \tag{6.5a}$$

and for g even

$$\begin{array}{ll}
f[a_B]: & j_A\mapsto j_A+\delta_{AB}i_B,\quad B=1,\ldots,(g-2)/2,\\
f[b_B]: & i_A\mapsto i_A+\delta_{AB}j_B,\\
f[a_{g/2}]: & \ell\mapsto\ell+i_{g/2},\\
f[c_B]: & j_A\mapsto j_A+(\delta_{AB}+\delta_{A,B+1})(i_B+i_{B+1}+1),\quad B=1,\ldots,(g-4)/2,\\
f[c_{(g-2)/2}]: & j_{(g-2)/2}\mapsto j_{(g-2)/2}+i_{(g-2)/2}+i_{g/2}+1,\\
y: & j_{(g-2)/2}\mapsto\begin{cases} j_{(g-2)/2}+1, & \text{for } \mathrm{Pin}^+(2),\\ j_{(g-2)/2}, & \text{for } \mathrm{Pin}^-(2),\end{cases}\\
f[r_{Bm}]: & j_A\mapsto j_A+\delta_{AB}k_m,\quad B=1,\ldots,(g-2)/2,\quad m=1,\ldots,n,\\
f[s_{Bm}]: & i_A\mapsto i_A+\delta_{AB}k_m,\\
f[t_m]: & k_h\mapsto k_h+(\delta_{hm}+\delta_{h,m+1})(k_m+k_{m+1}),\quad m=1,\ldots,n-1,\\
\sigma_h: & i_{g/2}\mapsto i_{g/2}+k_h,\quad h=1,\ldots,n.
\end{array} \tag{6.5b}$$

We now discuss the orbits of the action of $\Omega(N_{g,n})$ on pin structures. We begin with the case $n=0$. If g is odd, N_g only admits $\mathrm{Pin}^-(2)$ structures, and they are all quotients of even pin structures on Σ_{g-1}, i.e., pin structures for which $\varphi(\mathbf{i},\mathbf{j})=0$ with $\varphi(\mathbf{i},\mathbf{j})$ defined as in (3.5). So there is no invariant coming from the parity of the pin structure on the double cover. However, ℓ and

$$\psi(\mathbf{i},\mathbf{j})=\sum_{A=1}^{(g-1)/2}(i_A+1)(j_A+1)$$

are invariants. Using the method of Sec. IV, Ref. 1, it can be seen that for the values $(\psi,\ell)=(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$, it is possible to transform the $\mathrm{Pin}^-(2)$ structures to standard forms which are given in Tables III and IV. So for g odd there are four orbits.

If g is even, $i_{g/2}$ is invariant; in fact $i_{g/2}+1=\varphi(\mathbf{ij})$, i.e., the pin structures on N_g with $i_{g/2}=0$ (resp. 1) are quotients of odd (resp. even) pin structures on Σ_{g-1}. In addition, if $i_{g/2}=0$ also ℓ is invariant. For $(i_{g/2},\ell)=(0,0)$ and $(0,1)$ it is possible to bring the pin structures to standard forms. If $i_{g/2}=1$, the situation is different for $\mathrm{Pin}^+(2)$ and $\mathrm{Pin}^-(2)$. All $\mathrm{Pin}^+(2)$ structures with $i_{g/2}=1$ can be transformed into each other. For $\mathrm{Pin}^-(2)$ structures with $i_{g/2}=1$ there is a further invariant

$$\psi(\mathbf{ij})=\sum_{A=1}^{(g-2)/2}(i_A+1)(j_A+1)\,.$$

Every $\mathrm{Pin}^-(2)$ structure with $(i_{g/2},\psi)=(1,0)$ and $(1,1)$ can be brought to a standard form. Altogether, if g is even, there are three orbits for the action of $\Omega(N_g)$ on $\mathrm{Pin}^+(2)$ structures and four orbits for the action on $\mathrm{Pin}^-(2)$ structures.

Next we discuss the case with nonempty boundary ($n\geqslant 1$). If g is odd, every $\mathrm{Pin}^-(2)$ structure must have $\Sigma_{h=1}^{n}k_h=0 \bmod 2$. If $\mathbf{k}=\mathbf{0}$, the structure of the orbits is the same as in the case without boundary. If $\mathbf{k}\neq\mathbf{0}$, $\psi(\mathbf{ij})$ is no longer invariant but ℓ remains invariant. Thus there are $2[n/2]$ orbits with $\mathbf{k}\neq\mathbf{0}$, labeled by ℓ and by κ defined in (3.7). For g odd, every $\mathrm{Pin}^+(2)$ structure must have $\Sigma_{h=1}^{n}k_h=1 \bmod 2$, therefore at least one coefficient k_h must be equal to one. Again ψ is not invariant, but ℓ remains invariant. There are $2[(n+1)/2]$ orbits labeled by ℓ and κ. If g is even and $\mathbf{k}=\mathbf{0}$, the structure of the orbits is the same as in the case without boundary. If $\mathbf{k}\neq\mathbf{0}$, neither $i_{g/2}$ nor ℓ nor ψ are anymore invariant. So there are $[n/2]$ orbits labeled by κ. In Table III we collect the standard pin structure and the cardinality of each orbit. The orbits for $n=0$ are the same as for $\kappa=0$ (i.e., $\mathbf{k}=\mathbf{0}$).

Finally we discuss the subgroups $\Omega_B(N_{g,n})$ and $\Omega_\partial(N_{g,n})$. In $\Omega_B(N_{g,n})$ we lack the twists $f[t_h]$ which generate permutations of the boundaries, hence $\mathbf{k}$ is an invariant. The structure of the other invariants is the same as for the group $\Omega(N_{g,n})$; only the number and the cardinality of the orbits is different. For $\mathbf{k}=\mathbf{0}$ everything is as in the case of $\Omega(N_{g,n})$; for $\mathbf{k}\neq\mathbf{0}$ we have the following situation: for g odd and $\mathrm{Pin}^-(2)$ there are $2(2^{n-1}-1)$ orbits labeled by $\mathbf{k}$ and ℓ, each containing 2^{g-1} pin structures; for g odd and

TABLE III. Orbits of $\Omega(N_{g,n})$ for g odd.

For $\mathrm{Pin}^-(2)$ κ	ℓ	ψ	Standard form	Cardinality
0	0	0	(0,...,0; 1,...,1; 0,...,0;0)	$2^{(g-3)/2}(2^{(g-1)/2}+1)$
0	0	1	(0,...,0; 1,...,1,0; 0,...,0;0)	$2^{(g-3)/2}(2^{(g-1)/2}-1)$
0	1	0	(0,...,0; 1,...,1; 0,...,0;1)	$2^{(g-3)/2}(2^{(g-1)/2}+1)$
0	1	1	(0,...,0; 1,...,1,0; 0,...,0;1)	$2^{(g-3)/2}(2^{(g-1)/2}-1)$
2	0	—	(0,...,0; 1,...,1; 0,...,0,1,1;0)	$2^{g-1}\binom{n}{2}$
2	1	—	(0,...,0; 1,...,1; 0,...,0,1,1;1)	$2^{g-1}\binom{n}{2}$
...	...	...	...	...
$2m$	0	—	(0,...,0; 1,...,1; 0,...,0,1,...,1;0)	$2^{g-1}\binom{n}{2m}$
$2m$	1	—	(0,...,0; 1,...,1; 0,...,0,1,...,1;1)	$2^{g-1}\binom{n}{2m}$
...	...	...	...	...
$2[n/2]$	0	—	(0,...,0; 1,...,1; $\theta(n+1)$,1,...,1;0)	$2^{g-1}\binom{n}{2[n/2]}$
$2[n/2]$	1	—	(0,...,0; 1,...,1; $\theta(n+1)$,1,...,1;1)	$2^{g-1}\binom{n}{2[n/2]}$

For $\mathrm{Pin}^+(2)$ κ	ℓ	Standard form	Cardinality
1	0	(0,...,0; 1,...,1; 0,...,0,1;0)	$2^{g-1}\binom{n}{1}$
1	1	(0,...,0; 1,...,1; 0,...,0,1;1)	$2^{g-1}\binom{n}{1}$
...	...	...	...
$2m+1$	0	(0,...,0; 1,...,1; 0,...,0,1,...,1;0)	$2^{g-1}\binom{n}{2m+1}$
$2m+1$	1	(0,...,0; 1,...,1; 0,...,0,1,...,1;1)	$2^{g-1}\binom{n}{2m+1}$
...	...	...	...
$2[(n+1)/2]$	0	(0,...,0; 1,...,1; $\theta(n)$,1,...,1;0)	$2^{g-1}\binom{n}{2[(n+1)/2]}$
$2[(n+1)/2]$	1	(0,...,0; 1,...,1; $\theta(n)$,1,...,1;1)	$2^{g-1}\binom{n}{2[(n+1)/2]}$

TABLE IV. Orbits of $\Omega(N_{g,n})$ for g even.

For Pin$^-$(2) κ	$i_{g/2}$	ℓ	ψ	Standard form	Cardinality
0	0	0	—	(0,...,0; 1,...,1; 0,...,0;0)	2^{g-2}
0	0	1	—	(0,...,0; 1,...,1; 0,...,0;1)	2^{g-2}
0	1	—	0	(0,...,0,1; 1,...,1; 0,...,0;0)	$2^{(g-2)/2}(2^{(g-2)/2}+1)$
0	1	—	1	(0,...,0,1; 1,...,1,0; 0,...,0;0)	$2^{(g-2)/2}(2^{(g-2)/2}-1)$
2	—	—	—	(0,...,0; 1,...,1; 0,...,0,1,1;0)	$2^g\binom{n}{2}$
...	...	...	...	...	...
$2m$	—	—	—	(0,...,0; 1,...,1; 0,...,0,1,...,1;0)	$2^g\binom{n}{2m}$
...	...	...	...	...	...
$2[n/2]$	—	—	—	(0,...,0; 1,...,1;$\theta(n+1)$,1,...,1;0)	$2^g\binom{n}{2[n/2]}$

For Pin$^+$(2) κ	$i_{g/2}$	ℓ	Standard form	Cardinality
0	0	0	(0,...,0; 1,...,1; 0,...,0;0)	2^{g-2}
0	0	1	(0,...,0; 1,...,1; 0,...,0;1)	2^{g-2}
0	1	—	(0,...,0,1; 1,...,1; 0,...,0;0)	2^{g-1}
2	—	—	(0,...,0; 1,...,1; 0,...,0,1,1;0)	$2^g\binom{n}{2}$
...	...	...	...	...
$2m$	—	—	(0,...,0; 1,...,1; 0,...,0,1,...,1;0)	$2^g\binom{n}{2m}$
...	...	...	...	...
$2[n/2]$	—	—	(0,...,0; 1,...,1;$\theta(n+1)$,1,...,1;0)	$2^g\binom{n}{2[n/2]}$

Pin$^+$(2) there are 2^n orbits labeled by **k** and ℓ, each containing 2^{g-1} pin structures; for g even there are $2^{n-1}-1$ orbits labeled by $\mathbf{k}\neq 0$, each containing 2^g pin structures, independently of the structure group.

In the case of the subgroup $\Omega_\partial(N_{g,n})$ we further lack the slides which reverse the orientation of the boundaries. Everything is as in the case of $\Omega_B(N_{g,n})$ except for g even and $\mathbf{k}\neq 0$. In this case, independently of the structure group, $i_{g/2}$ is invariant; in addition, if $i_{g/2}=0$ also ℓ is invariant. For $i_{g/2}=0$ and 1 there are $2(2^{n-1}-1)$ and $2^{n-1}-1$ orbits, respectively. In Tables V and VI we collect the standard pin structures and the cardinalities of the orbits of $\Omega_\partial(N_{g,n})$; the invariant vectors **k** are ordered as if they were binary numbers.

TABLE V. Orbits of $\Omega_\partial(N_{g,n})$ for g odd.

For Pin$^-$(2) **k**	ℓ	ψ	Standard form	Cardinality
0,...,0	0	0	(0,...,0; 1,...,1; 0,...,0;0)	$2^{(g-3)/2}(2^{(g-1)/2}+1)$
0,...,0	0	1	(0,...,0; 1,...,1,0; 0,...,0;0)	$2^{(g-3)/2}(2^{(g-1)/2}-1)$
0,...,0	1	0	(0,...,0; 1,...,1; 0,...,0;1)	$2^{(g-3)/2}(2^{(g-1)/2}+1)$
0,...,0	1	1	(0,...,0; 1,...,1,0; 0,...,0;1)	$2^{(g-3)/2}(2^{(g-1)/2}-1)$
0,...,0,1,1	0	—	(0,...,0; 1,...,1; 0,...,0,1,1;0)	2^{g-1}
0,...,0,1,1	1	—	(0,...,0; 1,...,1; 0,...,0,1,1;1)	2^{g-1}
...	...	...	...	...
1,...,1,$\theta(n+1)$	0	—	(0,...,0; 1,...,1; 1,...,1,$\theta(n+1)$;0)	2^{g-1}
1,...,1,$\theta(n+1)$	1	—	(0,...,0; 1,...,1; 1,...,1,$\theta(n+1)$;1)	2^{g-1}

For Pin$^+$(2) **k**	ℓ	Standard form	Cardinality
0,...,0,1	0	(0,...,0; 1,...,1; 0,...,0,1;0)	2^{g-1}
0,...,0,1	1	(0,...,0; 1,...,1; 0,...,0,1;1)	2^{g-1}
0,...,0,1,0	0	(0,...,0; 1,...,1; 0,...,0,1,0;0)	2^{g-1}
0,...,0,1,0	1	(0,...,0; 1,...,1; 0,...,0,1,0;1)	2^{g-1}
...	...	...	...
1,...,1,$\theta(n)$	0	(0,...,0; 1,...,1;1,...,1,$\theta(n)$;0)	2^{g-1}
1,...,1,$\theta(n)$	1	(0,...,0; 1,...,1;1,...,1,$\theta(n)$;1)	2^{g-1}

TABLE VI. Orbits of $\Omega_\partial(N_{g,n})$ for g even.

For Pin$^-$(2) k	$i_{g/2}$	ℓ	ψ	Standard form	Cardinality
0,...,0	0	0	—	(0,...,0; 1,...,1; 0,...,0;0)	2^{g-2}
0,...,0	0	1	—	(0,...,0; 1,...,1; 0,...,0;1)	2^{g-2}
0,...,0	1	—	0	(0,...,01; 1,...,1; 0,...,0;0)	$2^{(g-2)/2}(2^{(g-2)/2}+1)$
0,...,0	1	—	1	(0,...,0,1; 1,...,10; 0_1,...,0;0)	$2^{(g-2)/2}(2^{(g-2)/2}-1)$
0,...,0,1,1	0	0	—	(0,...,0; 1,...,1; 0,...,0,1,1;0)	2^{g-2}
0,...,0,1,1	0	1	—	(0,...,0; 1,...,1; 0,...,0,1,1;1)	2^{g-2}
0,...,0,1,1	1	—	—	(0,...,0,1; 1,...,1; 0,...,0,1,1;0)	2^{g-1}
...	...	...	...	...	...
1,...,1,$\theta(n+1)$	0	0	—	(0,...,0; 1,...,1; 1,...,1,$\theta(n+1)$;0)	2^{g-2}
1,...,1,$\theta(n+1)$	0	1	—	(0,...,0; 1,...,1; 1,...,1,$\theta(n+1)$;1)	2^{g-2}
1,...,1,$\theta(n+1)$	1	—	—	(0,...,0,1; 1,...,1; 1,...,1,$\theta(n+1)$;0)	2^{g-1}

For Pin$^+$(2) k	$i_{g/2}$	ℓ	Standard form	Cardinality
0,...,0	0	0	(0,...,0; 1,...,1; 0,...,0;0)	2^{g-2}
0,...,0	0	1	(0,...,0; 1,...,1; 0,...,0;1)	2^{g-2}
0,...,0	1	—	(0,...,0,1; 1,...,1; 0,...,0;0)	2^{g-1}
0,...,0,1,1	0	0	(0,...,0; 1,...,1; 0,...,0,1,1;0)	2^{g-2}
0,...,0,1,1	0	1	(0,...,0; 1,...,1; 0,...,0,1,1;1)	2^{g-2}
0,...,0,1,1	1	—	(0,...,0,1; 1,...,1; 0,...,0,1,1;0)	2^{g-1}
...	...	...	...	...
1,...,1,$\theta(n+1)$	0	0	(0,...,0; 1,...,1;1,...,1,$\theta(n+1)$;0)	2^{g-2}
1,...,1,$\theta(n+1)$	0	1	(0,...,0; 1,...,1;1,...,1,$\theta(n+1)$;1)	2^{g-2}
1,...,1,$\theta(n+1)$	1	—	(0,...,0,1; 1,...,1;1,...,1,$\theta(n+1)$;0)	2^{g-1}

ACKNOWLEDGMENTS

We are grateful to David Chillingworth and Bruno Zimmermann for correspondence.

APPENDIX A: TWO MODELS FOR NONORIENTABLE SURFACES

Given two surfaces M_1 and M_2, one can define a third surface, called the connected sum of M_1 and M_2, by removing an open disk from M_1 and one from M_2 and sewing together the resulting boundaries. All surfaces can be obtained by forming connected sums of simple building blocks.

The basic building block of nonorientable surfaces is the real projective plane, which can be visualized as a disk with the points on the boundary antipodally identified. Removing an open disk from a real projective plane we obtain a surface with boundary which we call a crosscap. It is convenient to visualize a crosscap as an annulus with the points of the interior boundary antipodally identified, as symbolized by the cross in Fig. 8(a). Cutting along the diameter BPA, straightening out the half-circles, and sewing the segments PRP we get a square with two slides oppositely identified, as in Fig. 8(b). Thus a crosscap is homeomorphic to a Möbius strip. Diameters in Fig. 8(a) correspond to vertical lines in Fig. 8(b).

The next simplest nonorientable surface is the Klein bottle, which is usually pictured as a square with the sides identified as in Fig. 9(a). Cutting along the segment AB, sewing the lower dotted triangle to the upper one along the side AA, and then straightening everything we get the alternative picture in Fig. 9(b), in which the vertical segments AB and AMNB are pairwise identified as shown. This figure shows that the Klein bottle consists of two Möbius strips (dotted rectangles) sewn onto the ends of a cylinder and therefore is homeomorphic to a sphere with two crosscaps, or equivalently the connected sum of two real projective planes. Removing an open disk from a Klein bottle, we obtain a surface with boundary which can be drawn in three alternative ways, as in Fig. 10. Figure 10(a) shows a disk with two crosscaps; in Fig. 10(b) the sides of the square are

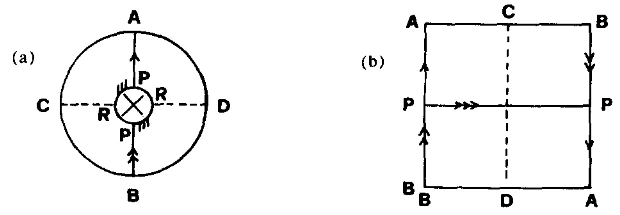


FIG. 8. A crosscap is homeomorphic to a Möbius strip.

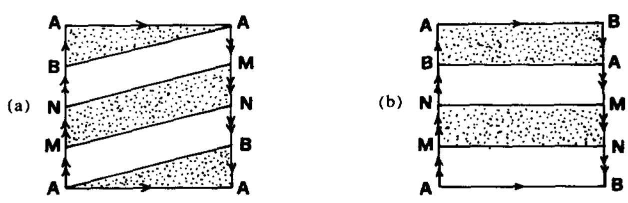


FIG. 9. A Klein bottle is homeomorphic to N_2.

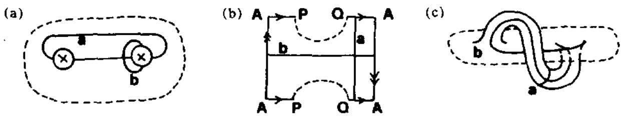


FIG. 10. A Klein bottle with a disk removed.

identified as in Fig. 9(a). In all cases the boundary is the broken line and we have drawn two homology generators a, b.

A nonorientable surface without boundary N_g of genus g, defined as the connected sum of g real projective planes, can also be thought of as a sphere with g crosscaps. If $g \geqslant 3$, N_g can also be regarded as the connected sum of a Klein bottle and N_{g-2}. This is seen by dividing N_g into a region W containing two crosscaps as in Fig. 10(a), and its complement. We can use for W the picture in Fig. 10(c); keeping one end of the tube fixed, slide the other through the gth crosscap. As a result, we obtain an ordinary handle attached to N_{g-2}. This shows that N_g is also the connected sum of a torus and N_{g-2}. This process can be iterated $[(g-1)/2]$ times until one (for g odd) or two (for g even) crosscaps are left. Thus N_g is the connected sum of $(g-1)/2$ tori and a real projective plane, for g odd, and $(g-2)/2$ tori and a Klein bottle, for g even. For more details see Ref. 2.

APPENDIX B: SLIDES AND *Y* DIFFEOMORPHISMS

We discuss in more detail some of the diffeomorphisms of a nonorientable surface which were introduced in Sec. III. In particular, we describe slides and Y diffeomorphisms directly on $N_{g,n}$, viewed as a sphere with g crosscaps and n open disks removed, and relate this to the description of their lifts to $\Sigma_{g-1,2n}$ which was given in the text.

A slide is a diffeomorphism of $N_{g,n}$ whose support is the dashed region in Fig. 11(a), homeomorphic to a Möbius strip with a disk removed. It can be described as sliding the disk through the crosscap and back to its original position. The boundary of the disk is mapped into itself with the orientation reversed. Figure 11(b) shows the effect of the slide on the segments EC and DF. Notice that the slide is the identity on the path AMNB.

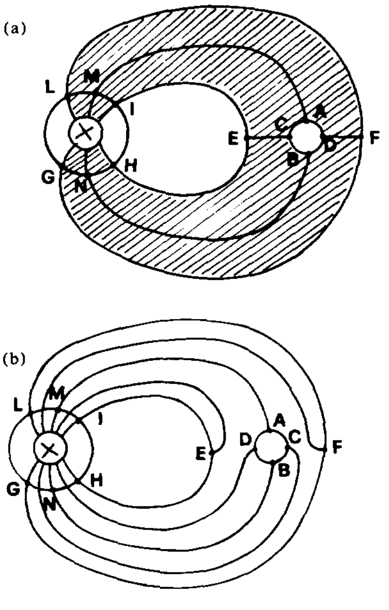


FIG. 11. The slide.

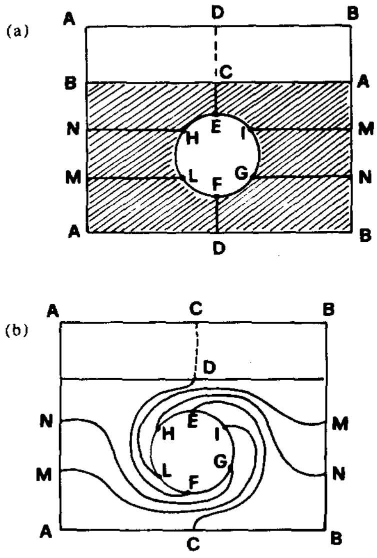


FIG. 12. The Y diffeomorphism.

The support of the diffeomorphism $f[t_n]$ on $\Sigma_{g-1,2n}$ is a cylinder intersecting the plane $x^1 = 0$, with two disks removed, invariant under J, which projects onto a closed subset of $N_{g,n}$ homeomorphic to a Möbius strip with a disk removed. It is easy to see that the diffeomorphism σ_n is a slide, as defined above.

A Y diffeomorphism of N_g ($g \geqslant 2$) has a support W which consists of the shaded region of Fig. 11(a) together with a crosscap as in Fig. 8(a) sewn along the boundary ADBC. It is a slide joined smoothly to the reflection through the diameter AB in the crosscap. In order to relate to the definition given in the text, we observe that W is homeomorphic to a Klein bottle without a disk. Making use of Fig. 9(b), we can identify W with the complement of the disk HEIGFLH in Fig. 12(a). The effect of the Y diffeomorphism is shown in Fig. 12(b): it is the identity on the circle HEIGFLH and on the circle ABNMA and reflection through the center in the upper rectangle ACBADB.

In the case $g = 2$, $n = 0$ (the Klein bottle), Fig. 12 represents the whole surface; the complement of W in $N_{2,0}$ is a disk, with its boundary sewn to W along the circle HEIGFLH. In this case there is a smooth isotopy which rotates the central disk by t in the clockwise direction, with $0 \leqslant t \leqslant \pi$. At $t = \pi$ we obtain a diffeomorphism which consists of reflections through the center in the upper and lower rectangle. Translating back from Fig. 9(b) to Fig. 9(a), this corresponds to a reflection through the center of the whole rectangle. It can be seen that this is further isotopic to a reflection through a vertical axis going through the center, which is the description of the Y diffeomorphism of the Klein bottle given in Ref. 5.

The isotopy of the Y diffeomorphism defined above with the diffeomorphism y defined in Sec. III is easily established when we observe that since the double covering of the Klein bottle is a torus, we can identify the rectangle in Fig. 12 with the part of the torus in Fig. 7 with $x^3 \geqslant 0$.

[1]L. Dąbrowski and R. Percacci, "Spinors and diffeomorphisms," Commun. Math. Phys. **106**, 691 (1986).
[2]W. S. Massey, *Algebraic Topology: An Introduction* (Springer, New York, 1967).
[3]J. S. Birman, "Mapping class groups and their relationship to braid groups," Commun. Pure Appl. Math. **22**, 213 (1969).
[4]S. P. Humphries, *Generator for the Mapping Class Group, Lecture Notes in Mathematics*, Vol. 722 (Springer, Berlin, 1979), pp. 44–47.
[5]W. E. R. Lickorish, Proc. Cambridge Philos. Soc. **59**, 307 (1963).
[6]J. S. Birman and D. R. J. Chillingworth, "On the homeotopy group of a nonorientable surface," Proc. Cambridge Philos. Soc. **71**, 437 (1972).
[7]D. R. J. Chillingworth, "A finite set of generators for the homotopy group of a nonorientable surface," Proc. Cambridge Philos. Soc. **65**, 409 (1969).
[8]G. P. Scott, Proc. Cambridge Philos. Soc. **68**, 605 (1970).
[9]M. Blau and L. Dąbrowski, "Pin structures on spaces quotiented by a discrete group," SISSA preprint 116/E.P./1987.
[10]L. Dąbrowski and A. Trautman, "Spin structures on spheres and projective spaces," J. Math. Phys. **27**, 2022 (1986).

Grassmannian σ models and strings

M Arik and F. Neyzi
Boğaziçi University, Istanbul, Turkey

(Received 1 July 1987; accepted for publication 30 September 1987)

The relation between the Nambu–Goto string in D dimensions and the two-dimensional σ model defined on the Grassmannian manifold $SO(D-1,1)/SO(D-2)\times SO(1,1)$ is investigated. For $D=3$ and $D=4$ the Nambu–Goto string is identified with a definite sector of the Grassmannian σ model. For $D=3$ the remaining sector is again a string model with an additional term in the Nambu–Goto action.

I. INTRODUCTION

The recent proposal that all superstring theories[1] are contained in the purely bosonic string theory in 26 dimensions has made the bosonic string a focus of interest once again.[2] The bosonic string can be described as a two-dimensional σ model whose dynamical variables represent the world sheet of the string. The action is then proportional to the area of the world sheet swept out by the string in space-time.[3] Equivalently one can view the action as a harmonic map from the two-dimensional string space into the Minkowski space. In the latter formulation, the metric is varied independently.[4]

This paper is an attempt to describe the string in terms of the orientation of the string area element in Minkowski space. Just like the area swept out by the string, the orientation of an area element on the string is a function of two string coordinates. The orientation of the area element is the same as the orientation of the plane tangent to the string at a given point on the D-dimensional Minkowski space and is characterized by variables belonging to the Grassmannian manifold $SO(D-1,1)/[SO(1,1)\times SO(D-2)]$.[5] Here $SO(1,1)$ is the subgroup of $SO(D-1,1)$ that contains Lorentz transformations on the timelike two-dimensional plane tangent to the string at a given point, and $SO(D-2)$ is the rotation group acting on the subspace of Minkowski space orthogonal to the string area element at that point. Our action is again the standard two-dimensional σ model action, whose dynamical variables now represent the orientation of the area element of the string. Minimizing the action, we obtain the field equations, whose solution set contains all of the Nambu–Goto string solutions and more.

Investigating this model in detail in three dimensions, namely starting from the $SO(2,1)/SO(1,1)$ σ model, we note that the equation of motion can be integrated readily. If some integration constants are chosen to be zero, then our model is equivalent to the Nambu–Goto[3] string model. It is also possible to reverse this process and show this isomorphism starting from the string model. For both models, the integrability condition turns out to be the Liouville equation.[6] On the other hand, when the integration constants are chosen to be nonzero, we get a distinct model, which is distinguished from the Nambu–Goto string model by the presence of an additional term in the action, a term somewhat reminiscent of the Wess–Zumino term. Again the classical isomorphism between these models can also be shown by starting from the modified string model action. This time, the integrability condition is the cosh–Gordon equation.

In an attempt to mimic the same process in four dimensions we consider the $SO(3,1)/[SO(1,1)\times SO(2)]$ σ model. Again integrating the equation of motion and setting some constants of integration to zero, we can identify the dynamical variables of the σ model with the conventional string variables. Since our coset space is isomorphic to the coset space $SO(3,C)/SO(2,C)$, our model has an intrinsic complex structure, which is reflected in the fact that the integrability condition of this model is the complex Liouville equation. It is again possible to reverse this process and obtain the σ model from the string model.

Polyakov[5] has shown that the Euclidean string action when modified by an additional extrinsic curvature term leads to an $SO(D)/[SO(D-2)\times SO(2)]$ σ model provided some integrability conditions are imposed on the model. On the other hand we start from the Minkowskian version of this σ model without any additional constraints and show that in three dimensions it indeed leads to an additional term in the string action.

II. THE BOSONIC STRING IN A LORENTZ INVARIANT GAUGE: A GEOMETRICAL APPROACH

In the classical bosonic string model, the evolution of the string is described by two parameters σ^i, $i=0,1$. The dynamical variable is the position $X^\mu(\sigma^0,\sigma^1)$ of the string. The standard action is

$$S=\frac{\mu^2}{2}\int h^{\alpha\beta}\eta_{\mu\nu}\,\partial_\alpha X^\mu\,\partial_\beta X^\nu\sqrt{-\det h}\;d^2\sigma, \qquad (2.1)$$

where the metric in the string space $h^{\alpha\beta}$ and X are varied independently. Since no derivatives of h appear in the action, it can be eliminated using its equation of motion. The action then becomes the Nambu–Goto action

$$S=\mu^2\int\sqrt{-\det h}\;d^2\sigma, \qquad (2.2)$$

where

$$h_{\alpha\beta}=\eta_{\mu\nu}\,\partial_\alpha X^\mu\,\partial_\beta X^\nu. \qquad (2.3)$$

The invariance of this action under arbitrary reparametrizations can be used to choose coordinates u and v such that

$$X_u^2=X_v^2=0 \qquad (2.4)$$

and

 0022-2488/88/030594-05$02.50

$$h_{\alpha\beta} = e^{\lambda}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \tag{2.5}$$

where we have introduced

$$e^{\lambda} \equiv X_u \cdot X_v. \tag{2.6}$$

Here and hereafter subscripts denote partial derivatives. With the choice in Eq. (2.4), the equation of motion becomes linear

$$X_{uv} = 0. \tag{2.7}$$

The parameters u and v are related to the standard timelike and spacelike parameters[7] τ and σ as

$$\tau = u + v, \quad \sigma = u - v. \tag{2.8}$$

Equations (2.4) and (2.7) retain the following reparametrization invariance:

$$u \to u'(u), \quad v \to v'(v). \tag{2.9}$$

The geometrical meaning of this invariance is as follows: Consider the area swept out by the string in Minkowski space. At each point of the string construct a light cone. The intersection of this light cone with the timelike string sheet defines two directions X_u and X_v, which are arbitrary up to reparametrizations described by Eq. (2.9). In some formulations this invariance is used to choose the light-cone gauge

$$X^{+} \equiv X^{0} + X^{D-1} = u + v. \tag{2.10}$$

However, for our purposes, a choice of coordinates u and v, invariant under Lorentz transformations, will be more convenient.

To this end, consider the dynamical variable $x(u,v)$. The first derivative vectors X_u and X_v, as we recall, are the lightlike directions on the plane tangent to the string at the point (u,v). One of the second derivative vectors X_{uv} is zero due to the equation of motion. The other two, X_{uu} and X_{vv}, have components both parallel and perpendicular to the string sheet. The perpendicular components

$$\xi \equiv X_{uu} - \lambda_u X_u, \quad \eta \equiv X_{vv} - \lambda_v X_v \tag{2.11}$$

are spacelike. The significance of the two vectors ξ and η is further emphasized upon noting that they are the nonvanishing components of the second derivative tensor $\nabla_\alpha \nabla_\beta X$. When ξ and η are reparametrized using Eq. (2.9), they transform in the following way:

$$\xi \to \xi\left(\frac{du}{du'}\right)^2, \quad \eta \to \eta\left(\frac{dv}{dv'}\right)^2. \tag{2.12}$$

On the other hand, using the equation of motion we obtain

$$\xi_v = -\lambda_{uv} X_u, \quad \eta_u = -\lambda_{uv} X_v. \tag{2.13}$$

Using Eq. (2.11) together with Eq. (2.13) we get

$$(\xi^2)_v = 2\xi\cdot\xi_v = 0, \quad (\eta^2)_u = 2\eta\cdot\eta_u = 0. \tag{2.14}$$

Therefore ξ^2 is a function of u, and η^2 is a function of v only. The transformation described in Eq. (2.12) is used at this point to choose ξ and η as unit vectors. This choice uniquely determines the parameters u and v up to an interchange of u and v and up to a sign. To sum up, the invariance in Eq. (2.9) is used to set

$$\xi^2 = -1, \quad \eta^2 = -1. \tag{2.15}$$

With this choice the Lorentz invariance of the gauge is manifest.

III. THE SO $(D-1,1)/[SO(1,1)\times SO(D-2)]$ σ MODEL AND THE BOSONIC STRING

The $M = SO(D-1,1)/[SO(1,1)\times SO(D-2)]$ σ model in two dimensions can be described starting from the action

$$S = \int \mathrm{tr}\, P_u P_v \, du\, dv = \frac{1}{4}\int \mathrm{tr}(\partial_u m^{-2}\, \partial_v m^2)\, du\, dv. \tag{3.1}$$

Here P is the projection operator related to the coset space element m by

$$m^{-1}\sigma m = 1 - 2P, \tag{3.2}$$

where σ is the involutive automorphism of the symmetric space M, and obeys

$$\sigma m = m^{-1}\sigma, \quad \sigma^2 = 1. \tag{3.3}$$

The coset space element m can locally be obtained by exponentiating the component of the Lie algebra of $SO(D-1,1)$ orthogonal to the Lie subalgebra $SO(1,1)\times SO(D-2)$. The equation of motion obtained by varying m in the action Eq. (3.1) can be expressed in terms of P to yield

$$[P_{uv}, P] = 0. \tag{3.4}$$

Since M is a Grassmannian manifold associated with the family of timelike planes through a given point in Minkowski space, the projection operator P can be parametrized in terms of two null vectors r and s,

$$r^2 = s^2 = 0, \quad r\cdot s = 1. \tag{3.5}$$

Then P is given by

$$P^{\mu\nu} = r^\mu s^\nu + r^\nu s^\mu. \tag{3.6}$$

A plane can also be characterized by an antisymmetric bivector. Hence we can define the matrix F,

$$F^{\mu\nu} = r^\mu s^\nu - r^\nu s^\mu \tag{3.7}$$

and

$$F^2 = P. \tag{3.8}$$

The equation of motion [Eq. (3.4)] is then equivalent to

$$[F_{uv}, F] = 0. \tag{3.9}$$

We note that

$$F^{\mu\nu} = e^{-\lambda}(X_u^\mu X_v^\nu - X_u^\nu X_v^\mu) \tag{3.10}$$

satisfies Eq. (3.9) identically, with X_u and X_v obeying Eqs. (2.4) and (2.6). We therefore conclude that the solution set of Eq. (3.9) contains all the Nambu–Goto string solutions and "more."

IV. THE ISOMORPHISM BETWEEN THE σ MODEL AND THE BOSONIC STRING MODEL IN THREE DIMENSIONS

In this section we shall explicitly show how to go from the σ model to the bosonic string model and back via the Liouville equation in three dimensions. To this end we start from the $SO(2,1)/SO(1,1)$ σ model action

$$S = \int du\, dv[\xi_u\cdot\xi_v - e^{-\lambda}(\xi^2 - 1)]. \tag{4.1}$$

Here $e^{-\lambda}$ is just a Lagrange multiplier and we have the corresponding constraint

$$\xi^2 = \xi_0^2 - \xi_1^2 - \xi_2^2 = -1. \tag{4.2}$$

The equation of motion

$$\xi_{uv} = -e^{-\lambda}\xi \tag{4.3}$$

can readily be integrated to yield

$$(\xi_u^2)_v = (\xi_v^2)_u = 0. \tag{4.4}$$

In the first part of this section we shall be dealing with the case

$$\xi_u^2 = \xi_v^2 = 0. \tag{4.5}$$

This choice will lead to the Nambu–Goto string. Using the invariance of the action [Eq. (4.1)] under reparametrizations [Eq. (2.9)], we can choose the right-hand sides of Eq. (4.5) to be constants other than zero. The consequences of this alternative will be investigated later in this section.

Noting that the case where ξ, ξ_u, and ξ_v vectors are linearly dependent is trivial, we take these three vectors as independent. Using a method introduced by Pohlmeyer,[8] we expand the second derivative vectors ξ_{uu} and ξ_{vv} in terms of ξ, ξ_u, and ξ_v:

$$\xi_{uu} = -\lambda_u \xi_u, \quad \xi_{vv} = -\lambda_v \xi_v. \tag{4.6}$$

Here we have used

$$\xi_u \cdot \xi_v = e^{-\lambda}, \tag{4.7}$$

which follows from Eqs. (4.2) and (4.3). Since e^{λ} is an integrating factor for the differential equations (4.6), we find that $e^{\lambda}\xi_u$ is a function of v and $e^{\lambda}\xi_v$ is a function of u only. With the definition

$$\mathbf{X}_v \equiv e^{\lambda}\xi_u, \quad \mathbf{X}_u \equiv -e^{\lambda}\xi_v, \quad \mathbf{X}_{uv} = 0, \tag{4.8}$$

Eqs. (4.6) reduce to the equation of motion of the Nambu–Goto string, Eq. (2.7). So with Eq. (4.8) we relate the σ model dynamical variables ξ_u and ξ_v to the Nambu–Goto string variables $\mathbf{X}_u$ and $\mathbf{X}_v$. Furthermore, substituting Eq. (4.6) and (4.7) in the identity

$$\xi_{uv} \cdot \xi_{uv} = -(\xi_u \cdot \xi_v)_{uv} + \xi_{uu} \cdot \xi_{vv}, \tag{4.9}$$

we obtain the Liouville equation

$$\lambda_{uv} = e^{-\lambda}. \tag{4.10}$$

Now we want to reverse this process and obtain the σ model from the bosonic string model. We use the spacelike vectors η and ξ defined through Eqs. (2.11)–(2.15). We further note that

$$\xi \cdot \eta = \lambda_{uv} e^{\lambda}. \tag{4.11}$$

In three dimensions, since the vectors ξ and η are both perpendicular to the plane tangent to the string at the point (u,v), they have to be either parallel or antiparallel. They turn out to be antiparallel. Since our metric is $(+\ -\ -)$,

$$\xi = -\eta, \quad \xi \cdot \eta = 1. \tag{4.12}$$

So Eq. (4.12) along with Eq. (4.11) yields the Liouville equation. Similarly from Eq. (4.12) and from Eqs. (2.11)–(2.15) we calculate ξ_u, ξ_v, and ξ_{uv}. Then we see that ξ_u and ξ_v are indeed null vectors. Finally using the Liouville equation[6] we obtain Eq. (4.3), whereby we establish the isomorphism between the Nambu–Goto string model and the special case of σ model in which ξ_u and ξ_v are null vectors.

Now let us consider the three-dimensional σ model once more, but this time let us investigate the case where the right-hand sides of Eq. (4.5) equal some constants. For later convenience we start from the following action:

$$S = \int [\xi_u \cdot \xi_v + 2 \sinh \lambda (\xi^2 + 1)] du\, dv. \tag{4.13}$$

The equation of motion

$$\xi_{uv} = 2 \sinh \lambda\, \xi, \tag{4.14}$$

when dotted with ξ_u and ξ_v will yield Eq. (4.4) after using the constraint equation (4.2). Using the invariance of the action in Eq. (4.13) under the transformation of u and v as specified by Eq. (2.9) one has

$$\xi_u^2 = +2, \quad \xi_v^2 = -2. \tag{4.15}$$

Note here that the sign difference between ξ_u^2 and ξ_v^2 has a geometrical meaning. Since ξ_u and ξ_v are both perpendicular to the spacelike vector ξ, if one is timelike, the other one has to be spacelike.

Expanding the vectors ξ_{uu} and ξ_{vv} in terms of the three independent vectors ξ, ξ_u, and ξ_v we obtain

$$\begin{aligned} \xi_{uu} &= 2\xi + \lambda_u \tanh \lambda\, \xi_u - \lambda_u \operatorname{sech} \lambda\, \xi_v, \\ \xi_{vv} &= -2\xi + \lambda_v \operatorname{sech} \lambda\, \xi_u + \lambda_v \tanh \lambda\, \xi_v. \end{aligned} \tag{4.16}$$

When we multiply Eq. (4.16) by sech λ and add, after some manipulations we get

$$((\xi_u + e^{\lambda}\xi_v)/(\cosh \lambda))_u = ((e^{\lambda}\xi_u - \xi_v)/(\cosh \lambda))_v. \tag{4.17}$$

We are finally in a position to establish the correspondence between the σ model variables ξ_u and ξ_v: the term inside the left-hand side bracket in Eq. (4.17) is $2\mathbf{X}_v$ and the term inside the other one is $2\mathbf{X}_u$. When we substitute the expressions for ξ_{uu} and ξ_{vv} defined in Eq. (4.16) in Eq. (4.9), we get the cosh–Gordon equation

$$\lambda_{uv} = 2 \cosh \lambda. \tag{4.18}$$

It turns out that the string action corresponding to this σ model characterized by the choice in Eq. (4.15) is

$$S = \int \left[((\mathbf{X}_u \cdot \mathbf{X}_v)^2 - \mathbf{X}_u^2 \mathbf{X}_v^2)^{1/2} + \frac{2}{3} \mathbf{X} \cdot \mathbf{X}_u \wedge \mathbf{X}_v \right]. \tag{4.19}$$

The additional term is also translationally invariant up to a total divergence. It is a parity violating term whose dimension differs from that of the first term, just like the Wess–Zumino term,[9] first obtained using differential geometry methods.[10] The equation of motion is

$$\mathbf{X}_{uv} = \mathbf{X}_u \wedge \mathbf{X}_v. \tag{4.20}$$

This equation, together with Eq. (2.4) and (2.6), yields

$$\mathbf{X}_{uv}^2 = -e^{2\lambda}. \tag{4.21}$$

Again we define a spacelike unit vector ξ perpendicular to the vectors $\mathbf{X}_u$ and $\mathbf{X}_v$. Since $e^{-\lambda}\mathbf{X}_{uv}$ is also a unit vector perpendicular to $\mathbf{X}_u$ and $\mathbf{X}_v$, in addition to Eq. (2.11) we have

$$e^{-\lambda}\xi_{uv} = \xi = -\eta. \tag{4.22}$$

Using Eqs. (4.21), (2.11), and (2.6) along with the identity

$$\mathbf{X}_{uv}^2 = -(\mathbf{X}_u\cdot\mathbf{X}_v)_{uv} + \mathbf{X}_{uu}\cdot\mathbf{X}_{vv}, \tag{4.23}$$

we obtain Eq. (4.18), the cosh–Gordon equation as expected. When we calculate $\boldsymbol{\xi}_u$, $\boldsymbol{\xi}_v$, and $\boldsymbol{\xi}_{uv}$ from Eq. (4.22) and use the cosh–Gordon equation, we get back Eqs. (4.14) and (4.15), thereby proving the isomorphism between the three-dimensional σ model in which the vectors $\boldsymbol{\xi}_u$ and $\boldsymbol{\xi}_v$ are not null and the modified Nambu–Goto string, this time starting from the string action.

V. THE ISOMORPHISM BETWEEN THE σ MODEL AND THE BOSONIC STRING IN FOUR DIMENSIONS

In this section we explicitly show the correspondence between the σ model and the Nambu–Goto string in four dimensions. We again start from the string model in the Lorentz invariant gauge. However, since we are in four dimensions, the vectors ξ and η, both perpendicular to X_u and X_v, do not have to be antiparallel or parallel anymore. So we define

$$\cos\theta \equiv -\xi\cdot\eta. \tag{5.1}$$

Since Eq. (4.11) still holds, we have

$$\lambda_{uv} = -e^{-\lambda}\cos\theta. \tag{5.2}$$

We know from Eq. (2.13) that ξ_v and η_u are parallel to X_u and X_v, respectively. Therefore we expand the other first derivative vectors, ξ_u and η_v, in terms ξ, η, ξ_v, and η_u and obtain

$$\begin{aligned}\xi_u &= -\theta_u(\csc\theta)\eta + \theta_v(\cot\theta)\xi + (\sec\theta)\eta_u,\\ \eta_v &= \theta_v(\cot\theta)\eta - \theta_v(\csc\theta)\xi + (\sec\theta)\xi_v.\end{aligned} \tag{5.3}$$

When we use the fact that ξ_v is perpendicular to η we have the identity

$$(\xi\cdot\eta)_{uv} = \xi_u\cdot\eta_v - \xi_v\cdot\eta_u. \tag{5.4}$$

Substituting Eq. (5.3) in Eq. (5.4), we get, after some manipulations,

$$\theta_{uv} = e^{-\lambda}\sin\theta. \tag{5.5}$$

Using this equation together with Eq. (5.2) we obtain the complex Liouville equation

$$\chi_{uv} = -e^{-\chi}, \tag{5.6}$$

where

$$\chi \equiv \lambda + i\theta. \tag{5.7}$$

Because the Liouville equation is complex now, it is obvious that the σ model dynamical variable we have to define in terms of the bosonic string variables X_u and X_v has to be complex. To this end we define

$$\mathbf{F} \equiv e^{-\lambda}(t_u\mathbf{X}_v - t_v\mathbf{X}_u + i\mathbf{X}_u\wedge\mathbf{X}_v). \tag{5.8}$$

Now we shall start from the SO(3,1)/[SO(1,1) $\times$SO(2)]$\sim$SO(3,C)/SO(2,C) σ model that can be described by a complex unit vector. We show that the integration of this model leads to Eq. (5.8). The σ model equation of motion is

$$\mathbf{F}_{uv} = e^{-\chi}\mathbf{F}. \tag{5.9}$$

When dotted with $\mathbf{F}_u$ and $\mathbf{F}_v$ each, this equation gives

$$(\mathbf{F}_v^2)_v = (\mathbf{F}_v^2)_u = 0 \tag{5.10}$$

after using the constraint

$$\mathbf{F}^2 = 1. \tag{5.11}$$

We choose

$$\mathbf{F}_u^2 = 0,\quad \mathbf{F}_v^2 = 0. \tag{5.12}$$

Expanding the second derivative vectors $\mathbf{F}_{uu}$ and $\mathbf{F}_{vv}$ in terms of $\mathbf{F}$, $\mathbf{F}_u$, and $\mathbf{F}_v$ we obtain

$$\mathbf{F}_{uu} = -\chi_u\mathbf{F}_u,\quad \mathbf{F}_{vv} = -\chi_v\mathbf{F}_v, \tag{5.13}$$

analogous to Eq. (4.6). This e^{χ} is an integrating factor for Eqs. (5.13) after the use of which Eqs. (5.13) reduce to

$$e^{\chi}\mathbf{F}_u = \mathbf{f}(v),\quad e^{\chi}\mathbf{F}_v = \mathbf{g}(u). \tag{5.14}$$

Here **f** and **g** are complex functions of the real variables u and v, respectively. Because of Eqs. (5.9), (5.11), and (5.12) these functions are not arbitrary but subject to the following constraints:

$$\mathbf{f}^2 = \mathbf{g}^2 = 0 \tag{5.15}$$

and

$$\mathbf{f}_v^2 = \mathbf{g}_u^2 = 1. \tag{5.16}$$

For convenience let us introduce

$$\mathbf{R}\equiv\mathrm{Re}\,\mathbf{f},\quad \mathbf{J}\equiv\mathrm{Re}\,\mathbf{g},\quad \mathbf{I}\equiv\mathrm{Im}\,\mathbf{f},\quad \mathbf{K}=\mathrm{Im}\,\mathbf{g}. \tag{5.17}$$

The **J**, **K**, **R**, and **I** are real three-vectors, in terms of which Eqs. (5.15) and (5.16) become

$$\begin{aligned}&\mathbf{J}^2=\mathbf{K}^2,\quad \mathbf{R}^2=\mathbf{I}^2, \mathbf{J}_u^2=\mathbf{K}_u^2,\quad \mathbf{R}_v^2=\mathbf{I}_v^2,\\ &\mathbf{J}\cdot\mathbf{K}=\mathbf{I}\cdot\mathbf{R}=\mathbf{J}_u\cdot\mathbf{K}_u=\mathbf{I}_v\cdot\mathbf{R}_v=0.\end{aligned} \tag{5.18}$$

Finally we are ready to define the four-dimensional bosonic string variables

$$|\mathbf{R}| = |\mathbf{J}| = t_v,\quad |\mathbf{J}| = |\mathbf{K}| = t_u, \tag{5.19}$$

where t is the time component of the four vector x. The unit vector perpendicular to both **R** and **K** is

$$\frac{\mathbf{X}_v}{t_v} = \frac{\mathbf{R}}{|\mathbf{R}|}\wedge\frac{\mathbf{I}}{|\mathbf{I}|},\quad X_v^2 = 0. \tag{5.20}$$

Similarly

$$\frac{\mathbf{X}_u}{t_u} = \frac{\mathbf{K}}{|\mathbf{K}|}\wedge\frac{\mathbf{J}}{|\mathbf{J}|},\quad X_u^2 = 0. \tag{5.21}$$

Using Eq. (5.14) we realize that

$$\begin{aligned}\mathbf{E}_u &\equiv \mathrm{Re}\,\mathbf{F}_u = e^{-\lambda}((\cos\theta)\mathbf{R} + (\sin\theta)\mathbf{I}),\\ \mathbf{B}_u &\equiv \mathrm{Im}\,\mathbf{F}_u = e^{-\lambda}(\cos\theta)\mathbf{I} - (\sin\theta)\mathbf{R}).\end{aligned} \tag{5.22}$$

Analogously

$$\begin{aligned}\mathbf{E}_v &\equiv \mathrm{Re}\,\mathbf{F}_v = e^{-\lambda}((\cos\theta)\mathbf{J} + (\sin\theta)\mathbf{K}),\\ \mathbf{B}_v &\equiv \mathrm{Im}\,\mathbf{F}_v = e^{-\lambda}((\cos\theta)\mathbf{K} - (\sin\theta)\mathbf{J}).\end{aligned} \tag{5.23}$$

These last four sets of equations, Eq. (5.20)–(5.23), lead to **F** defined in Eq. (5.8). Hence we have shown the correspondence between the σ model variable **F** and the bosonic string variables X_u and X_v in four dimensions. As a result of the complex nature of the variables **F** and χ, it was not possible to construct the string model related to the choice

$$\mathbf{F}_u^2 = p(u),\quad \mathbf{F}_v^2 = q(v), \tag{5.24}$$

where p and q are arbitrary functions. One can use the reparametrization invariance to choose p and q of unit modulus. However the phases remain as functions of u and v. For this case, we could not find a consistent definition for the four-vector X which would enable us to interpret this alternative as a distinct string model in four dimensions.

ACKNOWLEDGMENTS

We are grateful to M. Hortaçsu and Y. Nutku for useful discussions.

This work was partially supported by Turkish Technical and Scientific Research Council.

[1]M. B. Green and J. H. Schwarz, Nucl. Phys. B **181**, 502 (1981); Phys. Lett. B **149**, 117 (1984); D. Gross, J. Havey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **54**, 502 (1985); Nucl. Phys. B **256**, 253 (1985).
[2]P. G. O. Freund, Phys. Lett. B **151**, 237 (1985); A. Casher, F. Englert, H. Nicolai, and A. Taomina, Phys. Lett. B **162**, 121 (1985).
[3]Y. Nambu, *Lectures at the Copenhagen Summer Symposium*, 1970; T. Goto, Prog. Theor. Phys. **46**, 1560 (1971).
[4]See, e.g., J. H. Schwarz, Caltech preprint, CALT-68-1290, 1968.
[5]V. V. Nesterenko, Lett. Math. Phys. **7**, 287 (1983); A. Polyakov, Nucl. Phys. B **268**, 406 (1986).
[6]R. Omnes, Nucl. Phys. B **149**, 269 (1979).
[7]P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, Nucl. Phys. **1356**, 109 (1973).
[8]K. Pohlmeyer, Commun. Math. Phys. **46**, 209 (1976).
[9]J. Wess and B. Zumino, Phys. Lett. B **37**, 95 (1971).
[10]B. M. Barbashov, V. V. Nesterenko, and A. M. Chervjakov, Theor. Math. Phys. **45**, 365 (1980) (in Russian).

A novel class of G function integrals

L. T. Wille[a)]
SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, United Kingdom

(Received 24 April 1987; accepted for publication 11 November 1987)

A number of Euler integrals involving Meijer's G function with arguments $[\alpha^2(1-x)^{-1}+\beta^2x^{-1}]^{-1}$, $\lambda x(1-x)$, and $[\alpha^2x+\beta^2(1-x)]^2/[x(1-x)]$ are evaluated in closed form. These results generalize and extend recent work on Bessel-function integrals. As a by-product some new closed form expressions for Meijer's G function are obtained.

I. INTRODUCTION

The integrals considered in this paper are generalizations of some recent results obtained by Glasser.[1] This author found closed form expressions for the integrals

$$\Phi_\nu(\alpha,\beta)=\int_0^1 u^{-3/2}(1-u)^{-3/2}\times[\alpha^2(1-u)^{-1}+\beta^2u^{-1}]^{-\nu/2}\times J_\nu((\alpha^2(1-u)^{-1}+\beta^2u^{-1})^{1/2})du, \quad (1)$$

$$\Psi_0(\alpha,\beta)=\int_0^1 u^{1/2}(1-u)^{-3/2}\times J_0((\alpha^2(1-u)^{-1}+\beta^2u^{-1})^{1/2})du, \quad (2)$$

$$C_0(\alpha,\beta)=\int_0^1 u^{-1/2}(1-u)^{-1/2}\times J_0((\alpha^2(1-u)^{-1}+\beta^2u^{-1})^{1/2})du, \quad (3)$$

where α and β are positive constants; for the integral

$$A_\nu=\int_0^1[u(1-u)]^{(\nu-3)/2}J_\nu(\lambda(u(1-u))^{-1/2})du,\quad \nu=0,1,2 \quad (4)$$

(where a sign error has been corrected); and for the integral

$$B_0=\int_0^1 u^{1/2}(1-u)^{-3/2}J_0(\lambda(u(1-u))^{-1/2})du. \quad (5)$$

These had occurred in a study of impurity screening at metallic surfaces.

Here the following generalizations will be considered:

$$I_{x,x'}^{m,n,p,q}(\alpha,\beta,a_p,b_q)=\int_0^1 u^x(1-u)^{x'}\times G_{p,q}^{m,n}\left((\alpha^2(1-u)^{-1}+\beta^2u^{-1})^{-1}\middle|\begin{matrix}a_p\\ b_q\end{matrix}\right)du, \quad (6)$$

$$J_{x,x'}^{m,n,p,q}(\lambda,a_p,b_q)=\int_0^1 u^x(1-u)^{x'}G_{p,q}^{m,n}\left(\lambda u(1-u)\middle|\begin{matrix}a_p\\ b_q\end{matrix}\right)du, \quad (7)$$

and

$$K_{x,x'}^{m,n,p,q}(\alpha,\beta,a_p,b_q)=\int_0^1 u^x(1-u)^{x'}G_{p,q}^{m,n}\left(\frac{(\alpha^2u+\beta^2(1-u))^2}{(u(1-u))}\middle|\begin{matrix}a_p\\ b_q\end{matrix}\right)du, \quad (8)$$

where a_p and b_q is a contracted notation for $\{a_1,...,a_p\}$ and $\{b_1,...,b_q\}$, and will be used throughout whenever no confusion is possible. Integrals of this form over the unit interval are known as Euler integrals. The integrals I and J can be seen to generalize (1)–(3) and (4), (5), respectively.

The function in the integrand is Meijer's G function,[2,3] whose properties will be briefly described in the next section. In short, the G function is a generalization of the hypergeometric functions, and virtually all special functions of mathematical physics are particular cases of it. An extensive list can be found in Ref. 2, and some new results have been collected in the Appendix. In this way G functions provide an important unifying concept, in the sense that any identity involving them immediately applies to a large number of special functions. Moreover, since G functions can be expressed as a loop integral, many results are relatively straightforward to prove. It will be seen, for example, that the present results, although of a greater generality than Glasser's, are much easier to obtain. Besides, it will be clear from the derivation exactly which conditions have to hold for the parameters x and x' in order for the integral to be tractable.

In view of the relative simplicity of the integrand, the present results are potentially very useful in physical applications (if necessary after a trigonometric substitution). The methods applied in this paper also illustrate the power of using G functions, defined as a complex contour integral, to evaluate definite integrals. The literature on Meijer's G function and closely related ones is extensive,[2–5] and attention is drawn to related Euler integrals in Ref. 3 (3.4) and Ref. 5 (5.2). The present results, however, are not found in these references, apart from the J integral, which is included here for the sake of completeness.

II. PROPERTIES OF MEIJER'S G FUNCTION

In order to make this paper reasonably self-contained, some of the G function's major properties, which will be needed in subsequent sections, will be listed here. Many more results concerning this class of functions are known and can be found in Refs. 2 and 3.

Meijer's G function is most conveniently defined as a Mellin–Barnes contour integral:

[a)] Present address: Department of Materials Science and Mineral Engineering, University of California, Berkeley, California 94720.

 0022-2488/88/030599-05$02.50

$$G^{m,n}_{p,q}\left(z\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) = (2\pi i)^{-1} \times \int_L \frac{\prod_{j=1}^m \Gamma(b_j - s)\prod_{j=1}^n \Gamma(1-a_j+s)}{\prod_{j=m+1}^q \Gamma(1-b_j+s)\prod_{j=n+1}^p \Gamma(a_j - s)} \times z^s\,ds. \quad (9)$$

In this expression an empty product is to be regarded as unity, $0 \leqslant m \leqslant q$, $0 \leqslant n \leqslant p$, and a_h and b_h are such that no pole in the first product of the numerator coincides with a pole in the second product. The path of integration is to be chosen so that the former poles lie to the right of it and the latter ones lie to its left (for a full discussion of the possible choices, see Ref. 2).

From the definition (9) it can easily be shown that

$$G^{m,n}_{p,q}\left(z\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) = G^{m,n}_{p,q}\left(z^{-1}\left|\begin{matrix}1-b_q\\ 1-a_p\end{matrix}\right.\right), \quad (10)$$

and

$$z^\sigma G^{m,n}_{p,q}\left(z\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) = G^{n,m}_{q,p}\left(z\left|\begin{matrix}a_p+\sigma\\ b_q+\sigma\end{matrix}\right.\right). \quad (11)$$

The integral in (9) can be evaluated by means of the calculus of residues. If no two of the parameters b_h, $h = 1,2,...,m$, differ by an integer or zero, the corresponding poles in (9) are simple, and one finds the expression

$$G^{m,n}_{p,q}\left(z\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) = \sum_{h=1}^m \frac{\prod_{j=1}^m \Gamma(b_j - b_h)^* \prod_{j=1}^n \Gamma(1+b_h-a_j)}{\prod_{j=m+1}^q \Gamma(1+b_h-b_j)\prod_{j=n+1}^p \Gamma(a_j - b_h)} \times z^{b_h}{}_pF_{q-1}\left(\begin{matrix}1+b_h-a_p\\ 1+b_h-b_q^*\end{matrix}\left|(-1)^{p-m-n}z\right.\right) \quad (12)$$

valid when $p<q$, or when $p=q$ and $|z|<1$. The asterisks denote that $j \neq h$, $q \neq h$, respectively.

Relatively few integrals containing G functions are known, but, as already noted, those that have been proved are very important because the parameters can be specialized to yield many results concerning the special functions. A key result is that the Mellin transform of the product of two G functions can be written as a G function (under proper conditions, mainly to ensure that the integral is meaningful; see Ref. 2):

$$\int_0^\infty G^{m,n}_{p,q}\left(\eta x\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) G^{\mu,\nu}_{\sigma,\tau}\left(\omega x\left|\begin{matrix}c_\sigma\\ d_\tau\end{matrix}\right.\right) dx = \frac{1}{\eta} G^{n+\mu,m+\nu}_{q+\sigma,p+\tau}\left(\frac{\omega}{\eta}\left|\begin{matrix}-b_1,...,-b_m,c_\sigma,-b_{m+1},...,-b_q\\ -a_1,...,-a_n,d_\tau,-a_{n+1},...,-a_p\end{matrix}\right.\right)$$
$$= \frac{1}{\omega} G^{m+\nu,n+\mu}_{p+\tau,q+\sigma}\left(\frac{\eta}{\omega}\left|\begin{matrix}a_1,...,a_n,-d_\tau,a_{n+1},...,a_p\\ b_1,...,b_m,-c_\sigma,b_{m+1},...,b_q\end{matrix}\right.\right). \quad (13)$$

Consequently the standard integral transforms involving one G function can also be expressed as a G function. Another important result (again, under appropriate conditions) are the Euler integrals

$$\int_1^\infty y^{-\alpha}(y-1)^{\alpha-\beta-1} G^{m,n}_{p,q}\left(zy\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) dy = \Gamma(\alpha-\beta) G^{m+1,n}_{p+1,q+1}\left(z\left|\begin{matrix}a_p,\alpha\\ \beta,b_q\end{matrix}\right.\right), \quad (14)$$

and

$$\int_0^1 y^{-\alpha}(1-y)^{\alpha-\beta-1} G^{m,n}_{p,q}\left(zy\left|\begin{matrix}a_p\\ b_q\end{matrix}\right.\right) dy = \Gamma(\alpha-\beta) G^{m,n+1}_{p+1,q+1}\left(z\left|\begin{matrix}\alpha,a_p\\ b_q,\beta\end{matrix}\right.\right). \quad (15)$$

These three results illustrate why Meijer's G function constitutes an important unification as far as analytical integration of special functions is concerned. Many integrals which have been evaluated by methods specific to the case at hand turn out to be special cases of the results (13)–(15).

III. INTEGRALS WITH ARGUMENT $[\alpha^2(1-u)^{-1}+\beta^2u^{-1}]^{-1}$

It is easy to prove that the integrals in (6) obey the recurrence relations

$$I^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q) = I^{m,n,p,q}_{x',x}(\beta,\alpha,a_p,b_q), \quad (16)$$

$$I^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q) = I^{m,n,p,q}_{x,x'+1}(\alpha,\beta,a_p,b_q) + I^{m,n,p,q}_{x+1,x'}(\alpha,\beta,a_p,b_q), \quad (17)$$

and [using (11)]

$$I^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q) = \alpha^2 I^{m,n,p,q}_{x,x'-1}(\alpha,\beta,a_p+1,b_q+1) + \beta^2 I^{m,n,p,q}_{x-1,x'}(\alpha,\beta,a_p+1,b_q+1). \quad (18)$$

This means that, starting from a given couple (x,x'), all integrals with parameters $(x+j,x'+k)$, $j,k \in Z$, can be reached. Furthermore, it will soon become clear that only integrals with x and x' equal to a half-integer are tractable. Hence it will only be necessary to calculate one tractable integral, say $x = x' = -\frac{1}{2}$, in order to obtain all other tractable cases.

First, however, the discussion will proceed completely generally. Substituting the definition (9) of the G function in (6) and interchanging the order of integration (assuming that both integrals are absolutely convergent), one obtains

$$\beta^{-2s}\int_0^1 u^{x+s}(1-u)^{x'+s}\left[1-\left(1-\frac{\alpha^2}{\beta^2}\right)u\right]^{-s} du. \quad (19)$$

This can be written as a hypergeometric formula using Euler's formula, so that one finds

$$\beta^{-2s}\frac{\Gamma(x+s+1)\Gamma(x'+s+1)}{\Gamma(x+x'+2s+2)} \times {}_2F_1(x+s+1,s;x+x'+2s+2|1-\alpha^2/\beta^2). \quad (20)$$

In general, this cannot be further reduced, but an important simplification would be possible if one could use the result

$$_2F_1(a,a-\tfrac{1}{2};2a|z) = [\tfrac{1}{2}+\tfrac{1}{2}(1-z)^{1/2}]^{1-2a}, \tag{21}$$

since then the remaining loop integral can be identified as a G function. There are only two cases in which this result can be used directly, namely $x=-\frac{1}{2}$, $x'=-\frac{1}{2}$, and $x=-\frac{3}{2}$, $x'=-\frac{1}{2}$. However, by using the Gauss relations between contiguous hypergeometric functions, any function of the form

$$_2F_1(a+j,b+k;c+l\,|z)$$

can be reduced to a linear combination of $_2F_1(a,b;c|z)$ and one of its contiguous functions, with coefficients that are rational functions of a, b, c, and z. This means that all integrals with half-integer (x,x') values can be reduced to the two cases $(-\frac{1}{2},-\frac{1}{2})$ and $(-\frac{3}{2},-\frac{1}{2})$. However, by using the relations (16) and (18) it can easily be shown that the latter result can be expressed in terms of the former. It is straightforward to prove from (20) and (21) that

$$I^{m,n,p,q}_{-1/2,-1/2}(\alpha,\beta,a_p,b_q) = \sqrt{\pi}G^{m,n+1}_{p+1,q+1}\left((\alpha+\beta)^{-2}\middle|\begin{matrix}\frac{1}{2},a_p\\ b_q,0\end{matrix}\right). \tag{22}$$

In Ref. 1 Glasser conjectured that

$$\int_0^1\left(\frac{1}{u}+\frac{1}{1-u}\right)^{1/2} f(\alpha^2(1-u)^{-1}+\beta^2u^{-1})du = F(\alpha+\beta). \tag{23}$$

The result (22) shows that this conjecture is true for any function f that can be written as a finite (or infinite—provided that the series converges absolutely) linear combination of G functions. Closed form expressions for some of the tractable (x,x') values $(x<x')$ are given in Table I. In particular, these contain Glasser's integrals (1)–(3) as a special case.

From (12) it follows that the behavior of the G function near the origin is determined by

$$z^{\min(b_h)}, \quad h=1,\ldots,m. \tag{24}$$

In order for the integral (6) to be convergent one must have

$$\mathrm{Re}(b_h+x+1)>0, \quad h=1,\ldots,m, \tag{25a}$$

$$\mathrm{Re}(b_h+x'+1)>0, \quad h=1,\ldots,m. \tag{25b}$$

Moreover, the conditions

$$m+n-\tfrac{1}{2}(p+q)\geqslant 0, \tag{26a}$$

$$0\leqslant n\leqslant p<q, \tag{26b}$$

$$1\leqslant m\leqslant q, \tag{26c}$$

must also be fulfilled to justify the previous operations. By appealing to the principle of analytical continuation it is possible to relax these conditions (along similar lines as in Ref. 2, Sec. 5.6).

IV. INTEGRALS WITH ARGUMENT $\lambda u(1-u)$

In this case the analysis is much more straightforward. One first notes that

$$J^{m,n,p,q}_{x,x'}(\lambda,a_p,b_q) = \lambda^{-x'}J^{m,n,p,q}_{x-x',0}(\lambda,a_p+x',b_q+x'), \tag{27}$$

so that it suffices to put $x'=0$ in the following. Next, after substituting (7) in the definition (9) one needs to evaluate

$$\int_0^1 u^{x+s}(1-u)^s\,du = \frac{\Gamma(x+s+1)\Gamma(s+1)}{\Gamma(2s+x+2)}, \tag{28}$$

which is the definition of the B function. Expanding the Γ function in the denominator by the multiplication theorem, one finds that the resulting loop integral can be identified as a G function, so that the final result,

$$J^{m,n,p,q}_{x,0}(\lambda,a_p,b_q) = 2^{-(x+1)}\sqrt{\pi} \times G^{m,n+2}_{p+2,q+2}\left(\frac{\lambda}{4}\middle|\begin{matrix}0,-x,a_p\\ b_q,-x/2,-(x+1)/2\end{matrix}\right), \tag{29}$$

is valid under the same conditions [(25),(26)] as before. Finally, it is worth noting that the J integrals are a special case of the I integrals:

$$J^{m,n,p,q}_{x,x'}(\lambda,a_p,b_q) = I^{m,n,p,q}_{x,x'}(1/\lambda,1/\lambda,a_p,b_q). \tag{30}$$

From this one can conclude that Glasser's second set of integrals is included in the first set: $B_0=\Psi_0(\lambda,\lambda)$ and $A_\nu=\lambda^\nu\Phi_\nu(\lambda,\lambda)$.

V. INTEGRALS WITH ARGUMENT $[\alpha^2u+\beta^2(1-u)]^2/[u(1-u)]$

The evaluation in Sec. III depended on the fact that, after interchanging the integration over the unit interval and the loop integration, the resulting $_2F_1(\cdots)$ could be written

TABLE I. Evaluation of $I^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q)$ [defined in (6)] in terms of G functions. The integrals with $x>x'$, as well as those for other half-integer x,x' values, can be derived from the relations (16)–(18).

x	x'	$I^{m,n,p,q}_{x,x'}$
$-\frac{3}{2}$	$-\frac{3}{2}$	$\frac{\alpha+\beta}{\beta}I^{m,n,p,q}_{-3/2,-1/2}(\alpha,\beta,a_p,b_q)$
$-\frac{3}{2}$	$-\frac{1}{2}$	$\sqrt{\pi}\frac{\alpha+\beta}{\alpha}G^{m,n+1}_{p+1,q+1}\left((\alpha+\beta)^{-2}\middle\vert\begin{matrix}\frac{3}{2},a_p\\ b_q,1\end{matrix}\right)$
$-\frac{3}{2}$	$\frac{1}{2}$	$I^{m,n,p,q}_{-3/2,-1/2}(\alpha,\beta,a_p,b_q)-I^{m,n,p,q}_{-1/2,-1/2}(\alpha,\beta,a_p,b_q)$
$-\frac{3}{2}$	$\frac{3}{2}$	$I^{m,n,p,q}_{-3/2,1/2}(\alpha,\beta,a_p,b_q)-I^{m,n,p,q}_{-1/2,1/2}(\alpha,\beta,a_p,b_q)$
$-\frac{1}{2}$	$-\frac{1}{2}$	$\sqrt{\pi}G^{m,n+1}_{p+1,q+1}\left((\alpha+\beta)^{-2}\middle\vert\begin{matrix}\frac{1}{2},a_p\\ b_q,0\end{matrix}\right)$
$-\frac{1}{2}$	$\frac{1}{2}$	$\sqrt{\pi}\frac{\alpha-\beta}{\alpha+\beta}G^{m,n+1}_{p+1,q+1}\left((\alpha+\beta)^{-2}\middle\vert\begin{matrix}-\frac{1}{2},a_p\\ b_q,-1\end{matrix}\right)+\sqrt{\pi}\frac{\beta^2}{\alpha(\alpha+\beta)}G^{m,n+1}_{p+1,q+1}\left((\alpha+\beta)^{-2}\middle\vert\begin{matrix}\frac{1}{2},a_p\\ b_q,0\end{matrix}\right)$

as a binomial function. One may wonder if this method cannot be extended to deal with other cases. An obvious candidate is the relation

$$_2F_1(-a,b;b\,|z) = {}_1F_0(-a|z) = (1+z)^a. \tag{31}$$

This leads to the K integrals defined in (8). Working along the same lines as before, one obtains the hypergeometric function

$$_2F_1(x-s+1,-2s;x+x'+2-2s|1-\alpha^2/\beta^2). \tag{32}$$

Consequently, when $x' = -x-2$, (31) can be used to give

$$K^{m,n,p,q}_{x,-2-x} = 2\sqrt{\pi}\left(\frac{\beta}{\alpha}\right)^{2x} G^{m+1,n}_{p+1,q+1}\left(4\alpha^2\beta^2\left|\begin{matrix}a_p,\frac{1}{2}\\0,b_q\end{matrix}\right.\right). \tag{33}$$

By using the contiguous relations one finds that the integrals are tractable for $(x'+x)\in\{-2,-3,...\}$. However, it turns out that when $x'+x=-1$, the contiguous relations give rise to an identity $0=0$. In that case the integral is a $_2F_1(-a,b;b+1|\cdots)$ that is proportional to an incomplete beta function.[2] No further simplifications are possible, and the same applies when $(x'+x)\in N$. In the present case the recurrence relations are

$$K^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q) = K^{m,n,p,q}_{x',x}(\beta,\alpha,a_p,b_q), \tag{34}$$

$$K^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q) = K^{m,n,p,q}_{x,x'+1}(\alpha,\beta,a_p,b_q) + K^{m,n,p,q}_{x+1,x'}(\alpha,\beta,a_p,b_q), \tag{35}$$

$$K^{m,n,p,q}_{x,x'}(\alpha,\beta,a_p,b_q) = \alpha^2K^{m,n,p,q}_{x+1/2,x'-1/2}(\alpha\alpha,\beta,a_p-\tfrac{1}{2},b_q-\tfrac{1}{2}) + \beta^2K^{m,n,p,q}_{x-1/2,x'+1/2}(\alpha,\beta,a_p-\tfrac{1}{2},b_q-\tfrac{1}{2}), \tag{36}$$

the first two being similar to (16) and (17). The second relation allows the calculation of a $K_{x,x'}$ integral when that corresponding to $x+x'+1$ is known. The third relation does not permit altering the sum $x+x'$—as opposed to (18). The recurrence relation (35) together with (33) gives rise to the final result

$$K^{m,n,p,q}_{x,-2-x-l} = 2\sqrt{\pi}(\beta/\alpha)^{2x}\left[1+\frac{\beta^2}{\alpha^2}\right]^l G^{m+1,n}_{p+1,q+1}\left(4\alpha^2\beta^2\left|\begin{matrix}a_p,\frac{1}{2}\\0,b_q\end{matrix}\right.\right). \tag{37}$$

Again the conditions (25) and (26) are sufficient to ensure the absolute convergence of the integrals and may in certain specific cases be relaxed.

VI. FINAL REMARKS

The results presented here can be generalized in a number of ways. First, consider the integral

$$\int_0^1 u^x(1-u)^{x'}G^{m,n}_{p,q}\left(\left(\frac{au+b}{u^\mu(1-u)^\nu}\right)^\alpha\left|\begin{matrix}a_p\\b_q\end{matrix}\right.\right)du. \tag{38}$$

It is easy to see that this gives rise to

$$_2F_1(x-\alpha\mu s+1,-2\alpha s;x+x'+2-2\alpha s(\mu+\nu)|\cdots). \tag{39}$$

Therefore the method used in Sec. III can be applied if $\mu+\nu=2$ (x,x' half-integer) while the method used in Sec. V can be applied if $\mu+\nu=1$ (x,x' integer $\leqslant -2$). The resulting formulas are similar to those already found. Second, it is also straightforward to extend the present analysis to deal with the H function of one or more variables.[4,5] Since these cases are less likely to occur in physical applications, no explicit expressions will be given here.

In several places in this paper it has been stated that certain conditions must be fulfilled in order for G function integrals to be meaningful. The list of possible combinations can be made very long and detailed, but they all amount to ensuring convergence of the integral. To this end the integrand must be well-behaved throughout the interval of integration, in particular in the endpoints. In the present case this leads to conditions similar to (25) and (26). For the Mellin transform (13), these would need to be supplemented by an appropriate restriction to impose proper behavior of the integrand at infinity. A very comprehensive discussion of this problem can be found in Luke[2] (Sec. 5.6).

APPENDIX: NEW IDENTIFICATIONS OF *G* FUNCTIONS AS NAMED FUNCTIONS

In the course of this work, some G functions identifiable as known special functions, which are not contained in the standard lists,[2,3] have been obtained. Although these identities follow from a comparison of the present results (for particular values of the parameters) with Glasser's,[1] they have been checked by an independent derivation, which will be briefly outlined for each case below.

The first two identities are

$$G^{2,0}_{1,2}\left(z\left|\begin{matrix}1\\0,\frac{1}{2}\end{matrix}\right.\right) = \sqrt{\pi}\,\mathrm{erfc}(z), \tag{A1}$$

$$G^{2,0}_{1,3}\left(z\left|\begin{matrix}1\\0,\frac{1}{2},0\end{matrix}\right.\right) = -\frac{2}{\sqrt{\pi}}\,\mathrm{si}(2\sqrt{z}), \tag{A2}$$

which follow from the integral representation for the functions in the right-hand side and from (14). Note that these two results imply that

$$L_t^{-1}\{p^{-1}\,\mathrm{erfc}(ap^{-1/2})\} = -(2/\pi)\mathrm{si}(2at^{1/2}) \tag{A3}$$

(Eq. 18 in Ref. 1), since the inverse Laplace transform of a G function is known [Ref. 2, Eq. 5.6.3 (10), where $n+1$ in the left-hand side should read n]. Next, one has

$$G^{3,0}_{2,4}\left(z\left|\begin{matrix}\frac{1}{2},1\\-\frac{1}{2},\frac{3}{2},0,0\end{matrix}\right.\right) = \frac{2}{\sqrt{\pi}}[\mathrm{si}(2z^{1/2}) + z^{-1/2}\cos(2z^{1/2})], \tag{A4}$$

which follows from the definition (9) and the identity

$$\frac{\Gamma(-\frac{1}{2}-s)\Gamma(\frac{3}{2}-s)\Gamma(-s)}{\Gamma(1+s)\Gamma(\frac{1}{2}-s)\Gamma(1-s)} = -\frac{\Gamma(\frac{1}{2}-s)\Gamma(-s)}{\Gamma(1+s)\Gamma(1-s)} + 2\frac{\Gamma(-\frac{1}{2}-s)}{\Gamma(1+s)}. \tag{A5}$$

The resulting G function can be identified by (A2) and the known expression of the cosine as a hypergeometric function.

The final set of identities is

$$G^{1,1}_{1,3}\left(z\middle|\begin{matrix}\frac{1}{2}\\(\mu+\nu)/2,(\mu-\nu)/2,-\frac{1}{2}\end{matrix}\right)$$
$$=2^{1-\mu}[(\mu+\nu-1)J_\nu(2z^{1/2})S_{\mu-1,\nu-1}(2z^{1/2})$$
$$-J_{\nu-1}(2z^{1/2})S_{\mu,\nu}(2z^{1/2})]$$
$$+z^{-1/2}\frac{\Gamma((1+\mu+\nu)/2)}{\Gamma((1-\mu+\nu)/2)},\quad \mathrm{Re}(\mu+\nu)>-1, \tag{A6}$$

$$G^{1,1}_{1,3}\left(z\middle|\begin{matrix}\frac{1}{2}\\\frac{1}{2},\frac{1}{2}-\nu,-\frac{1}{2}\end{matrix}\right)=z^{-1/2}\Gamma(\nu)-z^{-\nu/2}J_{\nu-1}(2z^{1/2}), \tag{A7}$$

$$G^{1,1}_{1,3}\left(z\middle|\begin{matrix}\frac{1}{2}\\\nu,0,-\frac{1}{2}\end{matrix}\right)$$
$$=\sqrt{\pi}\Gamma(\nu+\tfrac{1}{2})[J_\nu(2z^{1/2})\mathbf{H}_{\nu-1}(2z^{1/2})$$
$$-\mathbf{H}_\nu(2z^{1/2})J_{\nu-1}(2z^{1/2})],\quad \mathrm{Re}\,\nu>-\tfrac{1}{2}. \tag{A8}$$

These can be derived from (2), (4), and (5) in Sec. 19.1 of Ref. 6 and the Euler transform (15). It is not unlikely that a search in tables of integrals, together with results such as (13)–(15) might produce more new identifications, but such a task has not be undertaken in this work.

[1]M. L. Glasser, J. Math. Phys. **25**, 2933 (1984); **26**, 2082(E) (1985).
[2]Y. L. Luke, *The Special Functions and their Approximations* (Academic, New York, 1969), Vol. I.
[3]A. M. Mathai and R. K. Saxena, *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences* (Springer, Berlin, 1973).
[4]A. M. Mathai and R. K. Saxena, *The H-function with Applications in Statistics and Other Disciplines* (Wiley, New York, 1978).
[5]H. M. Srivastava, K. C. Gupta, and S. P. Goyal, *The H-functions of One and Two Variables with Applications* (South Asian, New Delhi, 1982).
[6]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. II.

Bi-Hamiltonian formulation of the Kadomtsev–Petviashvili and Benjamin–Ono equations

A. S. Fokas and P. M. Santini[a)]
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676

(Received 11 June 1987; accepted for publication 21 October 1987)

It was shown recently that the Kadomtsev–Petviashvili (KP) equation (an integrable equation in 2 + 1, i.e., in two-spatial and one-temporal dimensions) admits a bi-Hamiltonian formulation. This was achieved by considering KP as a reduction of a (3 + 1)-dimensional system (in the variables x,y_1,y_2,t). It is shown here, using the KP as a concrete example, that equations in 2 + 1 possess *two* bi-Hamiltonian formulations and *two* recursion operators. Both Hamiltonian operators associated with the x direction are local; in contrast only one of the Hamiltonian operators associated with the y direction is local. Furthermore, using the Benjamin–Ono equation as a concrete example, it is shown that intergrodifferential equations in 1 + 1 admit an algebraic formulation analogous to that of equations in 2 + 1.

I. INTRODUCTION

This paper investigates symmetries, conserved quantities, recursion operators, mastersymmetries, and the bi-Hamiltonian formulation of two physically important exactly solvable evolution equations: the Kadomtsev–Petviashvili[1] (KP) and Benjamin–Ono[2,3] (BO) equations. The KP equation is a prototype integrable equation in 2 + 1 (i.e., in two-spatial and in one-temporal dimensions), while the BO equation is a prototype singular integrodifferential equation in 1 + 1. The results presented here fit in the general theory developed in Refs. 4 and 5; however, the following conceptual aspects are novel.

(i) Equations in *two* spatial dimensions (x and y) possess *two* recursion operators and *two* sets of compatible Hamiltonian operators. The set associated with the y direction was considered in Refs. 4–6. Here we investigate the recursion operator and the pair of *local* Hamiltonian operators associated with the x direction.

(ii) Integrodifferential equations in 1 + 1 share many common features with equations in 2 + 1.[7] This is because integrodifferential equations are also formulated in terms of two space operators, for example ∂_x and H (the Hilbert transform) in the case of the BO equation. It is shown here that the algebraic formulation of integrodifferential equations is analogous to that of equations in 2 + 1.

The existence of a double representation, corresponding to two recursion operators and two sets of bi-Hamiltonian operators, is also a property of integrodifferential equations in 1 + 1; this will be shown in a separate paper[8] for two explicit examples: the intermediate long wave[9,10] and the BO equations.

Hierarchies of infinitely many time-independent and time-dependent symmetries and conserved quantities of the KP equation have been obtained in Refs. 11 and 12. A recursion operator and a bi-Hamiltonian formulation of the KP were found in Refs. 4–6. This was achieved by introducing the following *extended* representation of the KP equation:

$$q_{1_t} = \int_{\mathbb{R}} dy_2\, \delta(y_1 - y_2) K_{12}, \quad q_1 = q(x,y_1,t)\,, \tag{1.1}$$

where $\mathbb{R}$ denotes integration along the real axis, δ is the Dirac distribution, and K_{12} is some function of q_1 and $q_2 = q(x,y_2,t)$. The introduction of the above form is naturally motivated considering KP as a reduction of a nonlocal (3 + 1)-dimensional system (in the variables x, y_1, y_2, and t).[5,13] The above extension is necessary in order to bypass the Zakharov–Konopelchenko result on the nonexistence of recursion and bi-Hamiltonian operators in the usual (1 + 1)-dimensional formalism.[14]

Hierarchies of infinitely many time-independent and time-dependent symmetries and conserved quantities of the BO equation have been obtained in Refs. 12 and 15, via the mastersymmetry approach introduced by Fuchssteiner and one of the authors (A.S.F.). This approach was subsequently applied to the KP equation. It was shown in Ref. 5 that the mastersymmetry approach is contained in the general theory developed in Refs. 4 and 5.

A. Basic notions

We consider an evolution equation in its abstract form,

$$q_t = K(q)\,, \tag{1.2}$$

on a normed space M of functions of $\mathbb{R}$; K is a suitable C^∞ vector field on M. We assume that the space of smooth vector fields on M is some space S of C^∞ functions on the real line or on the plane vanishing rapidly at infinity. By $K_f[v]$ we denote the Fréchet derivative of K in the direction v, i.e.,

$$K_f[v] \doteq \frac{\partial}{\partial \epsilon} K(q + \epsilon v)\bigg|_{\epsilon=0}. \tag{1.3}$$

Let S^* be the dual of S with respect to the following bilinear form:

$$(\gamma,\sigma) \doteq \int_{\mathbb{R}} dx\, \gamma\sigma \text{ or } (\gamma,\sigma) \doteq \int_{\mathbb{R}} dx\, dy\, \gamma\sigma\,, \tag{1.4}$$

$\gamma \in S^*$, $\sigma \in S$. Let I: $S \to \mathbb{R}$ be a functional, then its gradient is defined by

$$I_f[v] = (\text{grad } I, v)\,. \tag{1.5}$$

It is well known that the function γ is a gradient of a func-

---
[a)] Permanent Address: Universita Degli Studi–Roma, Istituto di Fisica "Guglielmo Marconi," Piazzale delle Scienze, 5, 1-00185 Roma, Italy.

 0022-2488/88/030604-14$02.50

tional I iff $\gamma_f = \gamma_f^+$, where the adjoint of an operator L is defined by $(L^+\gamma,\sigma) = (\gamma,L\sigma)$.

Definition 1.1: (i) A function $\sigma \in S$ is a *symmetry* of (1.2) iff the flow $q_t = \sigma$ commutes with the flow (1.2). This implies

$$\frac{\partial\sigma}{\partial t} + \sigma_f[K] - K_f[\sigma] = 0 . \tag{1.6}$$

(ii) A functional I is conserved by the flow (1.2) iff $dI/dt = 0$. Hence

$$\frac{\partial I}{\partial t} + (\gamma,K) = 0, \quad \gamma \doteq \operatorname{grad} I ,$$

and $\gamma \in S^*$ is called a *conserved gradient* of (1.2). Differentiating the above equation in the arbitrary direction v it follows that γ satisfies

$$\frac{\partial\gamma}{\partial t} + \gamma_f[K] + K_f^+[\gamma] = 0, \quad \gamma_f = \gamma_f^+ . \tag{1.7}$$

(iii) Equation (1.2) is a *Hamiltonian system* iff it can be written in the form

$$q_t = \Theta f , \tag{1.8}$$

where f is a gradient function, i.e., $f_f = f_f^+$, and Θ is a Hamiltonian operator where

(1) Θ is skew symmetric, $\Theta^+ = -\Theta$,

(2) Θ satisfies a Jacobi identity, (1.9a)

$$(a,\Theta'[\Theta b]c) + \text{cyclic permutation} = 0 . \tag{1.9b}$$

A Hamiltonian operator Θ is associated with the Poisson bracket

$$\{I,H\} = (\operatorname{grad} I, \Theta \operatorname{grad} H) . \tag{1.9c}$$

(iv) An operator Φ is called a *recursion operator* or a *strong symmetry* of (1.2) iff it maps symmetries of (1.2) to symmetries of (1.2). Requiring that σ and $\Phi\sigma$ are symmetries of (1.2), it follows that an operator Φ satisfying the operator equation

$$\frac{\partial\Phi}{\partial t} + \Phi_f[K] + [\Phi,K_f] = 0 \tag{1.10}$$

is a recursion operator of (1.2).

(v) An operator Φ is called *hereditary* or *Nijenhuis* iff it generates an Abelian algebra. Assume that the flow $q_t = \sigma$ commutes with the flows $q_t = v$, $q_t = \Phi v$, and that the flow $q_t = v$ commutes with the flow $q_t = \Phi\sigma$, where σ, v are arbitrary functions. Requiring that the flows $q_t = \Phi\sigma$, $q_t = \Phi v$ also commute it follows that

$$\Phi_f[\Phi\sigma]v - \Phi\Phi_f[\Phi v]\sigma \text{ is symmetric w.r.t. } \sigma,v \tag{1.11}$$

(we have assumed for simplicity that $\partial\Phi/\partial t = 0$).

Exactly solvable evolution equations in 1 + 1 admit infinitely many symmetries. These symmetries are usually generated by a hereditary recursion operator Φ. An alternative approach is to use the notion of a *mastersymmetry*. A function τ is a master symmetry of Eq. (1.2) iff the map

$$[\tau,\cdot\,]_L, \quad \text{where } [\tau,\sigma]_L \doteq \tau_f[\sigma] - \sigma_f[\tau] ,$$

maps symmetries of (1.2). Here τ is called a *gradient mastersymmetry* (with respect to the invertible Hamiltonian operator Θ) iff $\Theta^{-1}\tau$ is a gradient function.

Integrable Hamiltonian systems in 1 + 1 have an exceptionally rich algebraic structure: They are *bi-Hamiltonian* systems. The existence of two Hamiltonian operators $\Theta^{(i)}$, $i = 1,2$, implies the existence of a *recursion operator* $\Phi \doteq \Theta^{(2)}(\Theta^{(1)})^{-1}$, which generates infinitely many symmetries, while Φ^+ generates infinitely many gradients of conserved quantities. For example, the two Hamiltonian operators associated with the Korteweg–de Vries (KdV) equation are given by

$$\Theta^{(1)} = D, \quad \Theta^{(2)} = D^3 + 2Dq + 2qD, \quad D \doteq \partial_x .$$

The KdV can be written as

$$q_t = q_{xxx} + 6qq_x = \Theta^{(1)}\gamma^{(1)} = \Theta^{(2)}\gamma^{(2)} ,$$

where

$$\gamma^{(2)} = q = \operatorname{grad} \int_{\mathbb{R}} dx \frac{q^2}{2} ,$$

$$\gamma^{(1)} = q_{xx} + 3q^2 = \operatorname{grad} \int_{\mathbb{R}} dx \frac{-q_x^2}{2 + q^3} .$$

Furthermore, $\Phi \doteq \Theta^{(2)}(\Theta^{(1)})^{-1}$ is a *recursion operator* for the KdV, i.e., Φ generates symmetries and Φ^+ generates gradients of conserved quantities. The KdV is the second member, $n = 1$, of the following Lax hierarchies:

$$q_t = \Phi^n q_x, \quad n = \text{non-negative integer} \tag{1.12}$$

(throughout this paper n,m,r denote non-negative integers), where q_x is a *starting symmetry.*

Exactly solvable equations in 2 + 1, written in the form (1.1), also admit a bi-Hamiltonian formulation.[4–6] For the KP, the two Hamiltonian operators are given by

$$\theta_{12}^{(1)} = D, \quad \theta_{12}^{(2)} = D^3 + Dq_{12}^+ + q_{12}^+D + q_{12}^-D^{-1}q_{12}^- , \tag{1.13a}$$

where

$$\begin{aligned} &D \doteq \partial_x, \quad q_{12}^{\pm} \doteq q_1 \pm q_2 + \alpha(D_1 \mp D_2), \\ &D_i \doteq \partial_{y_i}, \quad i = 1,2, \end{aligned} \tag{1.13b}$$

and $q_i = q(x,y_i,t)$, $i = 1,2$. Indeed

$$\begin{aligned} q_{1_t} &= q_{1xxx} + 6q_1q_{1_x} + 3\alpha^2 D^{-1} q_{1y_1y_1} \\ &= K_{11} = \int_{\mathbb{R}} dy_2\, \delta_{12}\theta_{12}^{(1)}\gamma_{12}^{(1)} = \int_{\mathbb{R}} dy_2\, \delta_{12}\theta_{12}^{(2)}\gamma_{12}^{(2)} , \end{aligned} \tag{1.14}$$

where $\delta_{12} = \delta(y_1 - y_2)$ and $\gamma_{12}^{(i)}$, $i = 1,2$, are suitable extended gradients, i.e.,

$$I_d^{(i)}[v_{12}] = \langle \gamma_{12}^{(i)}, v_{12} \rangle .$$

In the above the subscript d denotes a suitable directional derivative and $\langle\ ,\ \rangle$ denotes a suitable bilinear form.[4] Furthermore, the recursion operator $\phi_{12} \doteq \theta_{12}^{(2)}(\theta_{12}^{(1)})^{-1}$ generates *extended symmetries* σ_{12}, while the adjoint ϕ_{12}^* of ϕ_{12} with respect to $\langle\ ,\ \rangle$ generates *extended conserved gradients* γ_{12}. Then σ_{11}, γ_{11} are symmetries and conserved gradients of the KP, i.e., they satisfy Eqs. (1.6) and (1.7), respectively, where σ, γ, K are replaced by σ_{11}, γ_{11}, K_{11}, and K_{11} is defined in (1.14).

In analogy with Eq. (1.12), KP is the second member, $n=1$ ($\beta_1=\frac{1}{2}$), of the following hierarchy:

$$q_{1_t}=\beta_n\int_{\mathbf{R}} dy_2\,\delta(y_1-y_2)\phi_{12}^n\widehat{M}_{12}\cdot 1\,, \qquad (1.15)$$

where $\widehat{M}_{12}\cdot 1=(Dq_{12}^+ + q_{12}^- D^{-1}q_{12}^+)\cdot 1$ is a *starting extended symmetry*. Actually the operator ϕ_{12} admits two starting symmetry operators $\widehat{M}_{12}$ and $\widehat{N}_{12}\doteq q_{12}^-$. They give rise to the following two hierarchies of *time-independent* symmetries:

$$(\phi_{12}^n\widehat{M}_{12}\cdot 1)_{11},\quad (\phi_{12}^m\widehat{N}_{12}\cdot 1)_{11}\,. \qquad (1.16)$$

Time-dependent symmetries of order r of the KP are produced by linear combinations of

$$(\phi_{12}^n\widehat{M}_{12}\cdot (y_1+y_2)^r)_{11},\quad (\phi_{12}^m\widehat{N}_{12}\cdot (y_1+y_2)^r)_{11}\,, \qquad (1.17)$$

and are closely related to gradient mastersymmetries. The above hierarchies of time-independent and time-dependent symmetries give rise to time-independent and time-dependent conserved quantities.[4-6] Finally, there exists a simple relationship between ϕ_{12} and a *nongradient mastersymmetry* T_{12}:

$$T_{12}=\phi_{12}^2\cdot\delta(y_1-y_2),\quad C\phi_{12}=T_{12_d}+DT_{12_d}^{*}D^{-1}\,, \qquad (1.18)$$

where C is a constant. The above equations are the two-dimensional analogs of the following formulas, valid for the KdV:

$$T=\phi\cdot 1,\quad C\phi=T_f+DT_f^{*}D^{-1}\,. \qquad (1.19)$$

It is well known that the KP equation is associated with the linear problem

$$w_{xx}+(q(x,y,t)+\alpha\,\partial_y)w=0\,. \qquad (1.20)$$

The recursion operator ϕ_{12} is algorithmically derived from Eq. (1.20).[4,6]

B. New results

(i) The KP equation: In Refs. 4–6 the algebraic properties of KP were investigated by expanding in terms of $\delta(y_1-y_2)$. Now we expand in terms of $\delta(x_1-x_2)$ (Ref. 16) and write KP in the form

$$q_{1_t}=\int_{\mathbf{R}} dx_2\,\delta(x_1-x_2)K_{12},\quad q_1=q(x_1,y,t)\,, \qquad (1.21)$$

where K_{12} is some function of q_1, $q_2=q(x_2,y,t)$. Let subscripts 12 denote dependence on x_1, x_2, y; then for arbitrary functions f_{12}, g_{12} we define the following bilinear form:

$$\langle f_{12},g_{12}\rangle\doteq\int_{\mathbf{R}^3} dx_1\,dx_2\,dy\,f_{21}g_{12}\,. \qquad (1.22)$$

Let the arbitrary operator $\widehat{K}_{12}$ depend on the operators q_{12}^+, q_{12}^-, where

$$q_{12}^{\pm}\doteq q_1\pm q_2+D_1^2\pm D_2^2,\quad D_i=\partial_{x_i}\,,\quad q_i=q(x_i,y,t),\quad i=1,2\,; \qquad (1.23)$$

then the directional derivative of $\widehat{K}_{12}$ in the direction σ_{12} is denoted by $\widehat{K}_{12_d}[\sigma_{12}]$ and is defined by

$$\widehat{K}_{12_d}[\sigma_{12}]f_{12}\doteq\frac{\partial\widehat{K}_{12}}{\partial\epsilon}(q_{12}^+ +\epsilon\sigma_{12}^+,q_{12}^- +\epsilon\sigma_{12}^-)f_{12}\big|_{\epsilon=0}\,, \qquad (1.24a)$$

where

$$\sigma_{12}^{\pm}f_{12}\doteq\int_{\mathbf{R}} dx_3(\sigma_{13}f_{32}\pm\sigma_{32}f_{13})\,. \qquad (1.24b)$$

The two Hamiltonian operators associated with the KP equation (1.21) are given by

$$\Theta_{12}^{(1)}\doteq D_1+D_2,\quad \Theta_{12}^{(2)}=\alpha\,\partial_y+q_{12}^-\,, \qquad (1.25)$$

where $q_{12}^{\pm}$ are defined in (1.23). The operators $\Theta_{12}^{(i)}$, $i=1,2$, are skew symmetric, and satisfy the Jacobi identity

$$\langle a_{12},\Theta_{12_d}^{(i)}[b_{12}]c_{12}\rangle+\text{cyclic permutation}=0\,, \qquad (1.26)$$

where Θ_{12_d} and $\langle\ ,\ \rangle$ are defined by (1.22)–(1.24).

It should be stressed that, in contrast to the Hamiltonian pair (1.13), both of the above Hamiltonian operators are *local*. The KP is a bi-Hamiltonian system,

$$q_{1_t}=q_{1x_1x_1x_1}+6q_1q_{1_{x_1}}+3\alpha^2D_1^{-1}q_{1_{yy}}$$
$$=K_{11}=\int_{\mathbf{R}} dx_2\,\delta_{12}\Theta_{12_d}^{(i)}\gamma_{12}^{(i)},\quad i=1,2\,, \qquad (1.27)$$

where $\gamma_{12}^{(i)}$ are appropriate extended gradients.

KP is the fourth member, $n=3$, of the following Lax hierarchy:

$$q_{1_t}=\beta_n\int_{\mathbf{R}} dx_2\,\delta(x_1-x_2)\Phi_{12}^n\widehat{K}_{12}^0\cdot 1\,, \qquad (1.28)$$

where

$$\Phi_{12}\doteq\Theta_{12}^{(2)}(\Theta_{12}^{(1)})^{-1},\quad \widehat{K}_{12}^0\doteq\alpha\,\partial_y+q_{12}^-\,. \qquad (1.29)$$

The recursion operation Φ_{12} admits only one starting symmetry operator $\widehat{K}_{12}^0$, which generates the time-independent symmetries $(\Phi_{12}^m\widehat{K}_{12}^0\cdot 1)_{11}$. Values of m zero or even correspond to (1.16a), while m odd corresponds to (1.16b). Thus in the new formulation the two different hierarchies obtained in Ref. 4 are unified. Similarly Φ_{12}^{*} generates extended conserved gradients $\gamma_{12}^{(m)}$, which give rise to conserved gradients $\gamma_{11}^{(m)}$.

A nongradient mastersymmetry is given by

$$\Phi_{12}^2\cdot\frac{\partial\delta(x_1-x_2)}{\partial x_1}\,.$$

The recursion operator Φ_{12} can also be algorithmically obtained from the linear equation (1.20).

(ii) The BO equation: The BO equation

$$q_t=2qq_x+Hq_{xx},\quad q=q(x,t)\,, \qquad (1.30a)$$

where H denotes the Hilbert transform (throughout this paper principal value integrals are assumed if needed)

$$(Hf)(x)\doteq\pi^{-1}\int_{\mathbf{R}} d\xi(\xi-x)^{-1}f(\xi)\,, \qquad (1.30b)$$

can be written in the form

$$q_{1_t}=\int_{\mathbf{R}} dx_2\,\delta(x_1-x_2)K_{12}\,,\quad q_1=q(x_1,t)\,, \qquad (1.31)$$

where K_{12} is some function of q_1, $q_2=q(x_2,t)$. Let subscript

12 denote dependence on x_1, x_2; then for arbitrary functions f_{12}, g_{12} we define the following bilinear form:

$$\langle f_{12}, g_{12}\rangle \doteq \int_{\mathbb{R}^2} dx_1\, dx_2\, f_{21} g_{12}. \tag{1.32}$$

Let the arbitrary operator $\hat{K}_{12}$ depend on the operators q_{12}^+, q_{12}^-, where

$$\begin{aligned} &q_{12}^{\pm} \doteq q_1 \pm q_2 + i(D_1 \mp D_2), \quad D_i \doteq \partial_{x_i}, \\ &q_i = q(x_i,t), \quad i = 1,2; \end{aligned} \tag{1.33}$$

then the directional derivative of $\hat{K}_{12}$ in the arbitrary direction σ_{12} is denoted by $\hat{K}_{12_d}[\sigma_{12}]$ and is defined by (1.24).

Two compatible Hamiltonian operators associated with the BO equation are given by

$$\Theta_{12}^{(1)} \doteq q_{12}^-, \quad \Theta_{12}^{(2)} = (q_{12}^+ - i q_{12}^- H_{12}) q_{12}^-, \tag{1.34a}$$

where the operator H_{12} is an extended H operator,

$$(H_{12} f)(x_1,x_2) \doteq \pi^{-1} \int_{\mathbb{R}} d\xi\, [\xi - (x_1 + x_2)]^{-1} \times F(\xi, x_1 - x_2), \tag{1.34b}$$

and $f(x_1,x_2) = F(x_1 + x_2, x_1 - x_2)$. The BO equation is a bi-Hamiltonian system with respect to the above Hamiltonian operators.

The BO equation is a member of the following Lax hierarchy:

$$\begin{aligned} &q_{1_t} = \beta_n \int_{\mathbb{R}} dx_2\, \delta(x_1 - x_2) \Phi_{12}^n q_{12}^- \cdot 1, \\ &\Phi_{12} \doteq q_{12}^+ - i q_{12}^- H_{12}. \end{aligned} \tag{1.35}$$

Indeed, (1.35) with $n = 1$ and $n = 2$ yields

$$q_{1_t} = 2i\beta_0 q_{1_{x_1}}, \quad q_{1_t} = 4i\beta_1 (2 q_1 q_{1_{x_1}} + H_1 q_{1 x_1 x_1}). \tag{1.36}$$

The operator $\Phi_{12} = \Theta_{12}^{(2)} (\Theta_{12}^{(1)})^{-1}$ generates the time-independent symmetries of the BO equation $(\Phi_{12}^m q_{12}^- \cdot 1)_{11}$. Similarly, Φ_{12}^* generates extended conserved gradients $\gamma_{12}^{(m)}$.

The above recursion operator Φ_{12} can be derived algorithmically from the associated linear problem of the BO equation.

This paper is organized as follows. In Sec. II we derive the *second representation* of the KP class and we investigate the algebraic properties of the associated recursion operator and bi-Hamiltonian operators. In Sec. III we derive the extended representation of the BO class and we investigate the algebraic properties of the associated recursion and bi-Hamiltonian operators. In addition we discuss the connection with the mastersymmetries theory of the BO equation and with the complex Burgers hierarchy.

II. THE KP EQUATION

A. Derivation of the second representation

Proposition 2.1: The linear equation

$$-\alpha w_y = \hat{q} w, \quad \hat{q} \doteq q(x,y,t) + \partial_x^2, \tag{2.1}$$

is associated with the Lax hierarchy

$$\begin{aligned} q_{1_t} &= \beta_n \int_{\mathbb{R}} dx_2\, \delta(x_1 - x_2) \Phi_{12}^n \hat{K}_{12}^0 \cdot 1 \\ &= \beta_n \int_{\mathbb{R}} dx_2\, \delta(x_1 - x_2)(D_1 + D_2)\, \Psi_{12}^{n+1} \cdot 1, \end{aligned} \tag{2.2}$$

where β_n are constants, $D_i \doteq \partial_{x_i}$, $i = 1,2$, and the operators $\Phi_{12}, \Psi_{12}, \hat{K}_{12}^0$ are defined by

$$\begin{aligned} &\Phi_{12} \doteq (\alpha\, \partial_y + q_{12}^-)(D_1 + D_2)^{-1}, \\ &(D_1 + D_2)\Psi_{12} = \Phi_{12}(D_1 + D_2), \end{aligned} \tag{2.3a}$$

$$q_{12}^{\pm} = \hat{q}_1 \pm \hat{q}_2, \quad \hat{K}_{12}^0 = \alpha\, \partial_y + q_{12}^-. \tag{2.3b}$$

Remark 2.1: (i) $\hat{q}_2 = \hat{q}_1^*$, where * denotes the adjoint with respect to the bilinear form (1.22).

(ii) $\Psi_{12} = \Phi_{12}^*$.

(iii) Equation (2.2) with $n = 0,1,2,3$ and $\beta_1 = \frac{1}{2}$, $\beta_2 = \frac{1}{4}, \beta_3 = \frac{1}{2}$ implies

$$\begin{aligned} &q_{1_t} = 0, \quad q_{1_t} = q_{1_{x_1}}, \quad q_{1_t} = \alpha q_{1_y}, \\ &q_{1_t} = q_{1_{x_1 x_1 x_1}} + 6 q_1 q_{1_{x_1}} + 3\alpha^2 D_1^{-1} q_{1_{yy}}. \end{aligned} \tag{2.4}$$

Thus both the x-translation and the y-translation hierarchies of the KP are generated by the same extended starting symmetry $\hat{K}_{12}^0 \cdot 1 = q_1 - q_2$.

To derive the above Lax hierarchy we look for compatible flows

$$w_t = V w, \quad V \text{ polynomial in } \partial_x. \tag{2.5}$$

Compatibility of (2.1), (2.5) implies the *operator* equation

$$q_t = -(\alpha V_y + [q + \partial_x^2, V]). \tag{2.6}$$

Assuming the integral representation

$$(Vf)(x_1,y) = \int_R dx_2\, v(x_1,x_2,y) f(x_2,y), \quad v_{12} \doteq v(x_1,x_2,y) \tag{2.7}$$

and noting that

$$\begin{aligned} &(q_1 + D_1^2) V_1 f_1 = \int_{\mathbb{R}} dx_2 \{(q_1 + D_1^2) v_{12}\} f_2, \\ &V_1 (q_1 + D_1^2) f_1 = \int_{\mathbb{R}} dx_2 \{(q_2 + D_2^2) v_{12}\} f_2, \\ &V_y f = \int_{\mathbb{R}} dx_2\, v_{12y} f_2, \end{aligned}$$

we obtain the distribution equation

$$q_{1_t} \delta_{12} = -(\alpha v_{12_y} + q_{12}^- v_{12}). \tag{2.8}$$

Thus

$$\begin{aligned} &q_{1_t} \delta_{12} = -(D_1 + D_2) \Psi_{12} v_{12}, \\ &\Psi_{12} \doteq (D_1 + D_2)^{-1} (\alpha\, \partial_y + q_{12}^-). \end{aligned} \tag{2.9}$$

The operator $(D_1 + D_2)\Psi_{12}$ satisfies the following commutator operator equation:

$$\begin{aligned} &[(D_1 + D_2)\Psi_{12}, h_{12}] = 2h'_{12}(D_1 + D_2), \\ &h_{12} = h(x_1 - x_2), \quad h'_{12} = \frac{d}{dx_1} h_{12}. \end{aligned} \tag{2.10}$$

Using the above equation and assuming the expansion

$$v_{12}=\sum_{j=0}^{n}\delta^{j}_{12}v^{(j)}_{12},\quad \delta^{j}_{12}=\frac{d^{j}}{dx^{j}}\delta_{12}, \tag{2.11}$$

Eq. (2.9) yields

$$q_{1_t}\delta_{12}=\sum_{j=0}^{n}\delta^{j}_{12}(D_1+D_2)\Psi_{12}v^{(j)}_{12}$$
$$+2\sum_{j=1}^{n+1}\delta^{j}_{12}(D_1+D_2)v^{(j-1)}_{12}.$$

Thus

$$(D_1+D_2)v^{(n)}_{12}=0,\quad q_{1_t}\delta_{12}=\delta_{12}(D_1+D_2)\Psi_{12}v^{(0)}_{12},$$
$$-\tfrac{1}{2}\Psi_{12}v^{(j)}_{12}=v^{(j-1)}_{12}.$$

Therefore $v^{(0)}_{12}=(-\tfrac{1}{2})^n\Psi^n_{12}v^{(n)}_{12}$. Hence assuming $v^{(n)}_{12}=1$, the above equations imply

$$q_{1_t}\delta_{12}=\delta_{12}(D_1+D_2)\Psi^{n+1}_{12}\cdot 1$$
$$=\delta_{12}\Phi^n_{12}\Phi_{12}(D_1+D_2)\cdot 1,$$

where

$$(D_1+D_2)\Psi_{12}=\Phi_{12}(D_1+D_2).$$

B. Isospectrality yields a hereditary operator

To make this paper self-contained we first introduce an appropriate directional derivative. Recall the integral representation [Eq. (2.7)],

$$(Vf)(x_1,y)=\int_{\mathbf{R}}dx_3\,v(x_1,x_3,y)f(x_3,y).$$

Also, allowing f to depend on x_2 we obtain $Vf_{12}=\int_{\mathbf{R}}dx_3\,v_{13}f_{32}$. In particular,

$$\hat q_1 f_{12}=(q_1+D_1^2)f_{12}=\int_{\mathbf{R}}dx_3\,q_{13}f_{32}. \tag{2.12}$$

Equation (2.12) is a map between an operator and its kernel and induces the following directional derivative:

$$\hat q_{1_d}[\sigma_{12}]f_{12}=\int_{\mathbf{R}}dx_3\,\sigma_{13}f_{32}. \tag{2.13}$$

Equation (2.12) and the bilinear form (1.22) imply that the adjoint of $\hat q_1$, $\hat q_1^*=q_2+D_2^2$ has the representation

$$\hat q_1^* f_{12}=(q_2+D_2^2)f_{12}=\int_{\mathbf{R}}dx_3\,q_{32}f_{13}. \tag{2.14}$$

Hence

$$\hat q^*_{1_d}[\sigma_{12}]f_{12}=\int_{\mathbf{R}}dx_3\,\sigma_{32}f_{13}. \tag{2.15}$$

Equations (2.12)–(2.15) and $q^{\pm}_{12}\doteq\hat q_1\pm\hat q_1^*$ imply (1.24).

Proposition 2.2: (i) Consider the isospectral equation

$$\hat q v+\alpha v_y=\lambda v, \tag{2.16a}$$

and its adjoint, with respect to the bilinear form (1.4),

$$\hat q v^+-\alpha v^+_y=\lambda v^+. \tag{2.16b}$$

Then

$$(\operatorname{grad}\lambda)_{12}=v_1v_2^+, \tag{2.17}$$

where $(\operatorname{grad}\lambda)_{12}$ denotes the gradient of λ with respect to the bilinear form (1.22).

(ii) Equations (2.16) imply

$$(\alpha\,\partial_y+q^-_{12})v_1v_2^+=0. \tag{2.18}$$

To derive the above results, take the directional derivative of (2.16a) in the arbitrary direction f_{12}, multiply this equation by v_1^+ and integrate over $dx\,dy$ to obtain

$$\lambda_d[f_{12}]=\int_{\mathbf{R}^2}dx_1\,dy\,v_1^+\hat q_{1_d}[f_{12}]v_1.$$

Using (2.13) the above becomes

$$\lambda_d[f_{12}]=\int_{\mathbf{R}^3}dx_1\,dx_2\,dy\,v_1^+v_2f_{12}.$$

But

$$\lambda_d[f_{12}]=\int_{\mathbf{R}^3}dx_1\,dx_2\,dy(\operatorname{grad}\lambda)_{21}f_{12},$$

hence (2.17) follows. Equation (2.18) is a trivial consequence of (2.16).

Equation (2.18) suggests that Φ_{12} is a hereditary (Nijenhuis) operator (see Proposition 4.3 of Ref. 4). Actually it can be easily verified that

$$\Phi_{12_d}[\Phi_{12}f_{12}]g_{12}-\Phi_{12}\Phi_{12_d}[f_{12}]g_{12}$$
$$\text{is symmetric w.r.t. } f_{12}, g_{12}, \tag{2.19}$$

i.e., Φ_{12} is indeed hereditary (see Appendix A).

C. Symmetries and conserved gradients

1. Starting symmetries

We recall that the starting symmetry operators play an important role in the theory developed in Refs. 4 and 5. An operator Φ_{12} algorithmically implies starting symmetry operators: Look for operators $\hat S_{12}$ such that $\hat S_{12}H_{12}=0$, but $\Phi_{12}\hat S_{12}H_{12}\neq 0$. Then a starting symmetry operator $\hat K^0_{12}$ is given by $\hat K^0_{12}H_{12}\doteq\Phi_{12}\hat S_{12}H_{12}$.

Proposition 2.3: Let

$$\hat K^0_{12}\doteq\alpha\,\partial_y+q^-_{12},\quad H_{12}\doteq H(x_1-x_2,y), \tag{2.20}$$

where H is an arbitrary function of the arguments indicated. Then the following statements obtain.

(i) $\hat K^0_{12}\cdot H_{12}$ is a starting symmetry associated with the operator Φ_{12} [defined in (2.3)].

(ii) $\hat K^0_{12}$ satisfies a simple commutator operator equation with $h_{12}=h(x_1-x_2)$,

$$[\hat K^0_{12},h_{12}]=2\frac{\partial h_{12}}{\partial_{x_1}}(D_1+D_2). \tag{2.21}$$

(iii) Φ_{12} is a strong symmetry for $\hat K^0_{12}\cdot H_{12}$, i.e.,

$$\mathscr{L}(\Phi_{12},\hat K^0_{12}H_{12})$$
$$\doteq\Phi_{12d}[\hat K^0_{12}H_{12}]+[\Phi_{12},(\hat K^0_{12}H_{12})_d]=0. \tag{2.22}$$

(iv) The Lie algebra of the starting symmetry operator satisfies

$$[\hat K^0_{12}H^{(1)}_{12},\hat K^0_{12}H^{(2)}_{12}]_d=\hat K^0_{12}[H^{(1)}_{12},H^{(2)}_{12}]_I, \tag{2.23}$$

where

$$[K^{(1)}_{12},K^{(2)}_{12}]_d\doteq K^{(1)}_{12_d}[K^{(2)}_{12}]-K^{(2)}_{12_d}[K_{12}], \tag{2.24a}$$

$$[H^{(1)}_{12},H^{(2)}_{12}]_I\doteq\int_{\mathbf{R}}dx_3(H^{(1)}_{13}H^{(2)}_{32}-H^{(2)}_{13}H^{(1)}_{32}). \tag{2.24b}$$

To derive (i) let $\hat{S}_{12} = D_1 + D_2$; then H_{12} is defined by $(D_1 + D_2)H_{12} = 0$, thus $H_{12} = H(x_1 - x_2, y)$. Also $\hat{K}^0_{12} H_{12} = (\alpha\,\partial_y + q^-_{12})H_{12}$. Part (ii) is a straightforward calculation and part (iii) follows from the definition of a starting symmetry and the fact that Φ_{12} is hereditary (see Lemma 4.2 of Ref. 4). Part (iv) is a tedious calculation [see Appendix A for a direct verification of Eq. (2.22) and (2.23)].

2. Symmetries

We recall that σ_{12} is a time-independent extended symmetry of Eq. (2.2) iff

$$[\delta_{12}\Phi^n_{12}\hat{K}^0_{12} \cdot 1, \sigma_{12}]_d = 0. \tag{2.25}$$

Proposition 2.4:

(i) $\delta_{12}\Phi^n_{12}\hat{K}^0_{12} \cdot 1 = \sum_{l=0}^{n} b_{n,l}\Phi^{n-l}_{12}\hat{K}^0_{12}\delta^l_{12}$,

$b_{n,l}$ constants. (2.26)

(ii) $[\delta_{12}\,\Phi^n_{12}\hat{K}^0_{12} \cdot 1,\ \Phi^m_{12}\hat{K}^0_{12} \cdot H_{12}]_d$

$$= \sum_{l=0}^{n} b_{n,l}\Phi^{n-l+m}_{12}\hat{K}^0_{12}[\delta^l_{12}, H_{12}]_I. \tag{2.27}$$

(iii) $\sigma^{(m)}_{12} \doteq \Phi^m_{12}\hat{K}^0_{12} \cdot 1$ are time-independent extended symmetries of (2.2).

(iv) $\sigma^{(m)}_{11}$ are symmetries of (2.2).

(v) $\sigma^{(m)}_{12} = 0$ are auto-Bäcklund transformations of (2.2), where q_1, q_2 are interpreted to be two different solutions of (2.2).

Part (i) of the above follows from

$$[\Phi_{12}, h_{12}] = 2h'_{12}, \quad [\hat{K}^0_{12}, h_{12}] = 2h'_{12}(D_1 + D_2). \tag{2.28}$$

To derive (ii) note that

$$[\delta_{12}\Phi^n_{12}\hat{K}^0_{12} \cdot 1,\ \Phi^m_{12}\hat{K}^0_{12} \cdot H_{12}]_d$$
$$= \sum_{l=0}^{n} b_{n,l}[\Phi^{n-l}_{12}\hat{K}^0_{12}\delta^l_{12}, \Phi^m_{12}\hat{K}^0_{12}H_{12}]_d$$
$$= \sum_{l=0}^{n} b_{n,l}\Phi^{n-l+m}_{12}[\hat{K}^0_{12}\delta^l_{12}, \hat{K}^0_{12}H_{12}]_d$$
$$= \sum_{l=0}^{n} b_{n,l}\Phi^{n-l+m}_{12}\hat{K}^0_{12}[\delta^l_{12}, H_{12}]_I,$$

where we have used (for the third equality) the fact that Φ is hereditary and a strong symmetry for $\hat{K}^0_{12} \cdot H_{12}$, and the fourth equality follows from Eq. (2.23). Part (iii) follows from (ii) by taking $H_{12} = 1$. Part (iv) follows from (iii) and (2.8) (see Theorem 4.1 of Ref. 4). For part (v) see Theorem 4.2 of Ref. 4.

Remark 2.1: (i) Using Eq. (2.27) with suitable functions H_{12}, it should be possible to show that time-dependent symmetries of (2.2) are generated by linear combinations of $\Phi^m_{12}\hat{K}^0_{12}H_{12}$. See Ref. 5 for the corresponding results associated with the first representation.

(ii) An analysis about conserved gradients should follow closely the methods developed in Refs. 4 and 5. For example, it can be shown that $\Psi^n_{12} \cdot H_{12}$ are extended gradients for all $H_{12} = H(x_1 - x_2, y)$.

3. A nongradient mastersymmetry

Proposition 2.5: (i) $T_{12} \doteq \Phi^2_{12}\delta^1_{12}$, $\delta^1_{12} \doteq \partial\delta(x_1 - x_2)/\partial x_1$ is a nongradient mastersymmetry of the KP class, since

$$[\Phi^n_{12}\hat{K}^0_{12}H_{12}, T_{12}]_d = (n+1)\Phi^{n+1}_{12}\hat{K}^0_{12}H_{12}. \tag{2.29}$$

(ii) T_{12} generates the recursion operator Φ_{12} via

$$2\Phi_{12} = T_{12_d} + (D_1 + D_2)\,T^*_{12_d}(D_1 + D_2)^{-1}. \tag{2.30}$$

(iii) Let

$$\hat{\gamma}^{(n)}_{12} \doteq (\Phi^*_{12})^n\hat{\gamma}^0_{12}, \quad \hat{\gamma}^0_{12} \doteq (D_1 + D_2)^{-1}\hat{K}^0_{12}. \tag{2.31}$$

Then

$$\hat{\gamma}^{(n)}_{12}H_{12} = \text{grad}_{12}\, I_n, \tag{2.32a}$$

$$I_n \doteq 1/(n+2)\langle\hat{\gamma}^{(n+1)}_{12}H_{12}, \delta^1_{12}\rangle. \tag{2.32b}$$

The proof of (i)–(iii) is a consequence of equations $\delta^1_{12_d} = 0$, $\Phi_{12_d}[\delta^1_{12}] = 1$ and of Eq. (4.9), (4.6), and (4.7) of Ref. 5, respectively.

III. THE BO EQUATION

The linear problem associated with the BO equation (1.30) is the following differential Riemann–Hilbert (RH) boundary value problem:

$$\psi^{(-)}(x) = (q(x) + i\,\partial_x)\psi^{(+)}, \tag{3.1}$$

where $\psi^{(+)}$ and $\psi^{(-)}$ are the boundary values on the line $\text{Im}\, x = 0$ of functions holomorphic in the upper and lower half-plane, respectively,[17] and the spectral parameter has been rescaled away.

Equation (3.1) plays a crucial role in the derivation of the recursion and bi-Hamiltonian operators of the BO class.

A. Derivation of the recursion and bi-Hamiltonian operators

Proposition 3.1: The linear problem (3.1) is associated with the hierarchy

$$q_{1_t} = \beta_n \int_{\mathbb{R}} dx_2\, \delta(x_1 - x_2)\, \Phi^n_{12}\hat{K}^0_{12} \cdot 1$$
$$= \beta_n \int_{\mathbb{R}} dx_2\, \delta(x_1 - x_2)\, q^-_{12}\Psi^n_{12} \cdot 1, \tag{3.2}$$

where β_n are constants and the operators Φ_{12}, Ψ_{12}, and $\hat{K}^0_{12}$ are defined by

$$\Phi_{12} \doteq q^+_{12} - iq^-_{12}H_{12}, \quad q^-_{12}\Psi_{12} = \Phi_{12}q^-_{12}, \quad \hat{K}^0_{12} \doteq q^-_{12}, \tag{3.3a}$$

$$H_{12}f_{12} \doteq \pi^{-1}\int_{\mathbb{R}} d\xi\, [\xi - (x_1 + x_2)]^{-1}F(\xi, x_1 - x_2),$$
$$f_{12} \doteq f(x_1, x_2) = F(x_1 + x_2, x_1 - x_2), \tag{3.3b}$$

$$q^\pm_{12} \doteq q_1 \pm q_2 + i(D_1 \mp D_2), \quad q_i = q(x_i, t),$$
$$D_i = \partial_{x_i}, \quad i = 1,2. \tag{3.3c}$$

Remark 3.1: (i) $\Psi_{12} = \Phi^*_{12}$, where * denotes the adjoint with respect to the bilinear form (1.32).

(ii) The first few equations of the BO hierarchy are then

$$q_t = 0, \quad n = 0, \tag{3.4a}$$

$$q_t = q_x, \quad n = 1, \quad \beta_1 = (2i)^{-1} \quad \text{(wave equation)}, \tag{3.4b}$$

$$q_t = 2qq_x + Hq_{xx}, \; n = 2, \; \beta_2 = (4i)^{-1} \quad \text{(BO equation)}, \tag{3.4c}$$

$$q_t = (-q_{xx} + q^3 + \tfrac{3}{2}(qHq_x + Hqq_x))_x ,$$
$$n = 3, \quad \beta_3 = (8i)^{-1} \quad \text{(higher-order BO equation)}, \tag{3.4d}$$

and are obtained from (3.2) using Eqs. (3.16b)–(3.16f).

To derive the representation (3.2) we first seek compatibility between the differential RH problem (3.1) and the evolution equations

$$\psi_t^{(\pm)} = V^{(\pm)} \psi^{(\pm)}, \tag{3.5$_\pm$}$$

where $V^{(\pm)}$ are differential operators of the form

$$V^{(\pm)} = \sum_{j=0}^{n} V_j^{(\pm)}(x) \partial_x^j \tag{3.6}$$

and the coefficients $V_j^{(+)}(x)$ and $V_j^{(-)}(x)$ are holomorphic in the upper and lower half x plane, respectively.

The compatibility condition between (3.1) and (3.5) yields the operator equation

$$q_t = V^{(-)}(q + i\,\partial_x) - (q + i\,\partial_x) V^{(+)}, \tag{3.7}$$

which can be converted into a scalar distribution equation by formally introducing the integral representation

$$(V^{(\pm)} f)(x_1) = \int_{\mathbf{R}} dx_2\, v_{12}^{(\pm)} f(x_2), \quad v_{12}^{(\pm)} \doteq v^{(\pm)}(x_1,x_2). \tag{3.8}$$

For instance, the operator $V_1^{(-)}(q_1 + i\,\partial_{x_1})$ gives rise to the scalar kernel $(q_2 - i\,\partial_{x_2}) v_{12}^{(-)}$, since

$$V_1^{(-)}(q_1 + i\,\partial_{x_1}) f(x_1) = \int_{\mathbf{R}} dx_2\, v_{12}^{(-)}(q_2 + i\,\partial_{x_2}) f(x_2)$$
$$= \int_{\mathbf{R}} dx_2 ((q_2 - i\,\partial_{x_2}) v_{12}^{(-)}) f(x_2). \tag{3.9}$$

Equation (3.7) then corresponds to the following distribution scalar equation:

$$\delta(x_1 - x_2)\, q_{1_t} = -(q_1 + i\,\partial_{x_1}) v_{12}^{(+)} + (q_2 - i\,\partial_{x_2}) v_{12}^{(-)}$$
$$= -\tfrac{1}{2}(q_{12}^{+}(v_{12}^{(+)} - v_{12}^{(-)})$$
$$+ q_{12}^{-}(v_{12}^{+} + v_{12}^{(-)})). \tag{3.10}$$

Equations (3.6), (3.8), and (3.10) imply for $v_{12}^{(\pm)}$ the following expansions in derivatives of δ_{12}:

$$v_{12}^{(\pm)} = \sum_{j=0}^{n} \delta_{12}^{j} v_{12}^{(\pm)j}. \tag{3.11}$$

Combining (3.11), (3.8), and the analyticity properties of $V_j^{(\pm)}(x)$, we obtain that $v_{12}^{(+)}$ and $v_{12}^{(-)}$ are holomorphic in the upper and lower $x_1 + x_2$ plane, respectively. Then, in particular,

$$v_{12}^{(+)} - v_{12}^{(-)} = -iH_{12}(v_{12}^{(+)} - v_{12}^{(-)}) \tag{3.12}$$

[see Eq. (3.19)], and Eq. (3.10) becomes

$$\delta_{12} q_{1_t} = -\tfrac{1}{2}\Phi_{12}\tilde{v}_{12}, \quad \tilde{v}_{12} \doteq v_{12}^{(+)} - v_{12}^{(-)}. \tag{3.13}$$

Remark 3.2: The following operator commutator equations hold:

$$[q_{12}^{-}, h_{12}] = [H_{12}, h_{12}] = 0,$$
$$[q_{12}^{+}, h_{12}] = [\Phi_{12} h_{12}] = 2ih_{12}^{1}, \quad h_{12}^{1} \doteq \frac{\partial h_{12}}{\partial x_1}, \tag{3.14}$$

and hereafter h_{12} indicates an arbitrary function of $x_1 - x_2$. Substituting the expansion $\tilde{v}_{12} = \Sigma_{j=0}^{n} \delta_{12}^{j} \tilde{v}_{12}^{(j)}$ into Eq. (3.13) and using Eqs. (3.14) one obtains

$$\tilde{v}_{12}^{(n)} = 0; \quad \tilde{v}_{12}^{(j-1)} = (i/2)\Phi_{12}\tilde{v}_{12}^{(j)}, \quad 1 \leqslant j \leqslant n-1,$$
$$\delta_{12} q_{1_t} = (i/2)\delta_{12}\Phi_{12}\,\tilde{v}_{12}^{(0)}. \tag{3.15}$$

The iteration (3.15) implies that $\tilde{v}_{12}^{(0)} = (i/2)^{n-1} \times \Phi_{12}^{n-1}\tilde{v}_{12}^{(n-1)}$; to determine $\tilde{v}_{12}^{(n-1)}$ we notice that $\tilde{v}_{12}^{(n)} = v_{12}^{(+)n} - v_{12}^{(-)n} = 0$ implies $v_{12}^{(n)} = v_{12}^{(\pm)n} = c_n = \text{const}$, and then

$$\tilde{v}_{12}^{(n-1)} = (i/2)\,[q_{12}^{+}(v_{12}^{(+)n} - v_{12}^{(-)n})$$
$$+ q_{12}^{-}(v_{12}^{(+)n} + v_{12}^{(-)n})]c_n = ic_n q_{12}^{-} \cdot 1.$$

Equation (3.2) is then obtained defining $\beta_n \doteq i(i/2)^n c_n$.

B. Properties of the extended Hilbert transform

In this subsection we list several interesting and useful properties of the extended Hilbert transform.

Proposition 3.2: The extended Hilbert transform H_{12} enjoys the following properties.

(1) $[H_{12}, h_{12}] = 0$, (3.16a)

(2) $H_{12} a(x_j) = H_j a(x_j), \quad j = 1,2$, (3.16b)

$$H_j f(x_i, x_j) \doteq \pi^{-1} \int_{\mathbf{R}} dy (y - x_j)^{-1} f(x_i, y), \quad i \neq j. \tag{3.16c}$$

(3) $\int_{\mathbf{R}} dx_2\, \delta_{12} H_{12} f_{12} = H_1 f_{11}$, (3.16d)

(4) $\partial_{x_j} H_{12} f_{12} = H_{12}\, \partial_{x_j} f_{12}, \quad j = 1,2$, (3.16e)

(5) $H_{12}^2 = -1$. (3.16f)

Moreover,

(6) $H_{12} f_{12}^{\pm} h_{12} = (H_{12} f_{12})^{\pm} h_{12}$, (3.17a)

(7) $H_{12}(g_{12}^{-} H_{12} f_{12} + (H_{12} g_{12})^{-} f_{12})$
$= -g_{12}^{-} f_{12} + (H_{12} g_{12})^{-} H_{12} f_{12}$, (3.17b)

(8) $H_{12}^{*} = -H_{12}$. (3.17c)

Here H_{12} induces the following analytic properties:

(9) If

$$f_{12}^{(\pm)} \doteq \pm \tfrac{1}{2}(1 \mp iH_{12}) f_{12}$$
$$= (2\pi i)^{-1} \int_{\mathbf{R}} dy (y - (x_1 + x_2 \pm i0))^{-1}$$
$$\times F(y, x_1 - x_2), \tag{3.18}$$

then

(i) $f_{12}^{(+)}$ and $f_{12}^{(-)}$ are holomorphic for $\text{Im}(x_1 + x_2) > 0$ and $\text{Im}(x_1 + x_2) < 0$, respectively.

(ii) $f_{12}^{(+)} + f_{12}^{(-)} = -iH_{12}(f_{12}^{(+)} - f_{12}^{(-)})$. (3.19)

Proof: Equations (3.16f) and (3.17b) are interesting generalizations of well-known identities $H^2 = -1$, $H(gHf + fHg) = -gf + (Hg)(Hf)$, and can be proven using Fourier space. Equations (3.16a)–(3.16e), (3.17a), (3.17c), and (3.18), (3.19) are direct consequences of the definition of H_{12} (see Appendix B for details).

C. Algebraic properties of the BO class

In this section we show that the main algebraic properties of the BO class can be entirely described using the theory developed in Ref. 4; we refer to that paper for details and proofs.

1. Representation of the class

It was shown in Sec. III A that the BO class admits the following representation:

$$q_{1_t} = \beta_n \int_{\mathbf{R}} dx_2\, \delta_{12} \Phi_{12}^n \hat{K}_{12}^0 \cdot 1 \doteq \beta_n \int_{\mathbf{R}} dx_2\, \delta_{12} K_{12}^{(n)} \doteq K_{11}^{(n)}, \tag{3.20}$$

where $\hat{K}_{12}^0 = q_{12}^-$ and Φ_{12} is defined in (3.3a).

The recursion operator Φ_{12} and the "starting" operator $\hat{K}_{12}^0$ enjoy simple commutator relations with $h_{12} = h(x_1 - x_2)$,

$$[\Phi_{12}, h_{12}] = 2i \frac{\partial h_{12}}{\partial_{x_1}}, \quad [\hat{K}_{12}^0, h_{12}] = 0, \tag{3.21}$$

which imply that $\delta_{12} K_{12}^{(n)}$ can be written in the following *alternative form*:

$$\delta_{12} K_{12}^{(n)} = \sum_{l=0}^{n} (-2i)^l \binom{n}{l} \Phi_{12}^{n-l} \hat{K}_{12}^0 \frac{\partial^l \delta(x_1 - x_2)}{\partial_{x_l}}. \tag{3.22}$$

2. The *d* derivative

As in 2 + 1 dimensions, the derivation of the extended algebraic structures of the BO class is based on integral representations of operators depending on q, ∂_x, and H. This mapping between operators and their corresponding kernels induces a mapping between derivatives and leads to the introduction of a new directional derivative, the so-called d derivative.[4] Here we briefly remark that the basic operators $q_{12}^{\pm}$ appearing in the BO formalism are the *same* as for the KP case, replacing x_j by y_j and i by the parameter α [see Eqs. (1.13b) and (1.33)]. Then their d derivative is simply given by

$$q_{12}^{\pm}[g_{12}] f_{12} \doteq g_{12}^{\pm} f_{12}, \tag{3.23}$$

$$g_{12}^{\pm} f_{12} \doteq \int_{\mathbf{R}} dx_3 (g_{13} f_{32} \pm f_{13} g_{32}). \tag{3.24}$$

Since Φ_{12} and $\hat{K}_{12}^0$ are expressed in terms of $q_{12}^{\pm}$, their d derivatives are well defined,

$$\Phi_{12_d}[g_{12}] = g_{12}^+ - i g_{12}^- H_{12}, \quad \hat{K}_{12_d}^0[g_{12}] = g_{12}^-. \tag{3.25}$$

As for the (2 + 1)-dimensional case, the connection between the d derivative and the usual Fréchet derivative is given by the following projective formula:

$$K_{12_d}[\delta_{12} g_{12}] = K_{12_f}[g] \doteq K_{12_{q_1}}[g_{11}] + K_{12_{q_2}}[g_{22}], \tag{3.26}$$

where $K_{12_{q_i}}$ denotes the Fréchet derivative of K_{12} with respect to q_i, i.e.,

$$K_{12_{q_i}}[g_{ii}] \doteq \partial_\epsilon K_{12}(q_i + \epsilon g_{ii}, q_j)|_{\epsilon=0}, \quad i,j = 1,2, \quad i \neq j. \tag{3.27}$$

3. The starting symmetry $\hat{K}_{12}^0 \cdot h_{12}$, its Lie algebra, and its characterization through the recursion operator

The starting symmetry $K_{12}^{(0)} = q_1 - q_2$ of the BO class is written as $q_{12}^- \cdot 1$. As in 2 + 1 dimensions a crucial aspect of this theory is that the operator $\hat{K}_{12}^0 = q_{12}^-$, acting on suitable functions $h_{12} = h(x_1 - x_2)$, solutions of the RH problem $h_{12}^{(+)} - h_{12}^{(-)} = 0$ [$(+)$ and $(-)$ here indicate analyticity in the upper and lower $x_1 + x_2$ half-planes, then $h_{12} = h_{12}^{(+)} = h_{12}^{(-)}$], form a Lie algebra, given by

$$[q_{12}^- h_{12}, q_{12}^- \tilde{h}_{12}]_d = -q_{12}^- [h_{12}, \tilde{h}_{12}]_I, \tag{3.28}$$

where the Lie brackets $[\ ,\ ]_d$, $[\ ,\ ]_I$ are defined by

$$[f_{12}, g_{12}]_d \doteq f_{12_d}[g_{12}] - g_{12_d}[f_{12}], \tag{3.29a}$$

$$[h_{12}, \tilde{h}_{12}]_I \doteq \int_{\mathbf{R}} dx_3 (h_{13} \tilde{h}_{32} - \tilde{h}_{13} h_{32}). \tag{3.29b}$$

As in 2 + 1 dimensions, the starting symmetry $\hat{K}_{12}^0 \cdot h_{12}$ can be characterized through the recursion operator Φ_{12} via the equations

$$\Phi_{12}(h_{12}^{(+)} - h_{12}^{(-)}) = q_{12}^+ (h_{12}^{(+)} - h_{12}^{(-)}) + q_{12}^- (h_{12}^{(+)} + h_{12}^{(-)}) = 2\hat{K}_{12}^0 h_{12}, \tag{3.30a}$$

$$h_{12}^{(+)} = h_{12}^{(-)} = h_{12}, \tag{3.30b}$$

obtained using Eqs. (3.3a) and (3.19).

4. Symmetries, strong and hereditary symmetries

The recursion operator Φ_{12} and the starting operator $\hat{K}_{12}^0 = q_{12}^-$ are the ingredients of the evolution equations

$$q_{1_t} = \int_{\mathbf{R}} \delta_{12} K_{12}^{(n)}. \tag{3.31}$$

They enjoy the following properties.

Proposition 3.3: (i) The recursion operator Φ_{12} is hereditary, namely,

$$\Phi_{12_d}[\Phi_{12} f_{12}] g_{12} - \Phi_{12} \Phi_{12_d}[f_{12}] g_{12}$$

is symmetric w.r.t. f_{12} and g_{12}; (3.32)

(ii) Φ_{12} is a strong symmetry for $K_{12}^0 h_{12}$, namely,

$$\mathcal{L}(\Phi_{12}, \hat{K}_{12}^0 h_{12}) \doteq \Phi_{12_d}[\hat{K}_{12}^0 h_{12}] + [\Phi_{12}, (\hat{K}_{12}^0 h_{12})_d] = 0. \tag{3.33}$$

Proof: Equations (3.32) and (3.33) are verified in Appendix A, although this check is not strictly necessary, for two reasons.

(1) Φ_{12} comes from the isospectral problem (3.1), and an extension of the theorem presented in Ref. 18 should guarantee its hereditariness (see also Ref. 4, §4.E). It is also interesting to remark that a direct proof of the hereditariness of Φ_{12} makes use of Eq. (3.17b).

(2) The hereditariness of Φ_{12} and the characterization (3.30) implies that Proposition 3.3 (ii) holds (see Lemma 4.2 of Ref. 4 and Appendix A for a direct check).

The operator Φ_{12} generates infinitely many commuting symmetries of the BO class; precisely, since Φ_{12} is a hereditary operator and strong symmetry for the starting symmetry $\widehat{K}^0_{12}h_{12}$ that satisfies Eq. (3.28), then Theorem 4.3 of Ref. 4 implies that $\sigma^{(m)}_{12} \doteq \Phi^m_{12} q^-_{12} \cdot 1$ are extended symmetries of every evolution equation of the BO class, namely,

$$\sigma^{(m)}_{12_f}[K^{(n)}] = (\delta_{12}K^{(n)}_{12})_d[\sigma^{(m)}_{12}] \tag{3.34}$$

for every non-negative integer n and m, where, using (3.22),

$$(\delta_{12}K^{(n)}_{12})_d \doteq \sum_{l=0}^{n} (-2i)^l \binom{n}{l} (\Phi^{n-l}_{12}\widehat{K}^0_{12}\delta^l_{12})_d . \tag{3.35}$$

The first three operators $(\delta_{12}K^{(n)}_{12})_d$ of the BO class are explicitly reported below:

$$(\delta_{12}K^{(0)}_{12})_d = 0, \tag{3.36a}$$

$$(\sigma_{12}K^{(1)}_{12})_d = 2i(\partial_{x_1} + \partial_{x_2}), \tag{3.36b}$$

$$\begin{aligned}(\delta_{12}K^{(2)}_{12})_d &= 4i(H_{12}(\partial_{x_1} + \partial_{x_2})^2 + (\partial_{x_1} + \partial_{x_2})(q_1+q_2) \\ &\quad + i((H_1q_1)x_1 - (H_2q_2)x_2) \\ &\quad - i(q_1-q_2)H_{12}(\partial_{x_1} + \partial_{x_2}))\end{aligned} \tag{3.36c}$$

(see Appendix A).

The usefulness of the extended symmetries $\sigma^{(m)}_{12}$ follows from the fact that they give rise to symmetries and Bäcklund transformations; precisely according to Theorem 4.2 of Ref. 4:

If $\sigma^{(m)}_{12}$ is an extended symmetry of Eq. (3.31), then (i) $\sigma^{(m)}_{11} = \sigma^{(m)}_{12}|_{x_2 \to x_1}$ is a symmetry of Eq. (3.31), namely,

$$\sigma^{(m)}_{11_f}[K^{(n)}_{11}] = K^{(n)}_{11_f}[\sigma^{(m)}_{11}]; \tag{3.37}$$

and (ii) the equation

$$\sigma^{(m)}_{12} = \sigma^{(m)}(q_1,q_2) = 0 \tag{3.38}$$

is a Bäcklund transformation for (3.31) where, of course, q_1 and q_2 are now viewed as two different solutions of (3.31).

5. (Bi-) Hamiltonian formalism and constants of motion in involution

Proposition 3.4: (i) If we define

$$\Theta^{(1)}_{12} \doteq q^-_{12}, \quad \Theta^{(2)}_{12} \doteq \Phi_{12}\Theta^{(1)}_{12}, \tag{3.39}$$

then $\Theta_{12} \doteq \Theta^{(1)}_{12} + \kappa\Theta^{(2)}_{12}$ is a Hamiltonian operator for all constants κ, namely,

(a) $\Theta^*_{12} = -\Theta_{12}$, (3.40a)

(b) Θ_{12} satisfy the Jacobi identity w.r.t. the bracket

$$\{a_{12},b_{12},c_{12}\} \doteq \langle a_{12},\Theta_{12_d}[\Theta_{12}b_{12}]c_{12}\rangle . \tag{3.40b}$$

(ii) The adjoint Φ^*_{12} of the recursion operator, given by

$$\Phi^*_{12} = q^+_{12} - iH_{12}q^-_{12}, \tag{3.41}$$

satisfies the following "well-coupling" condition:

$$\Phi_{12}\Theta^{(1)}_{12} = \Theta^{(1)}_{12}\Phi^*_{12} . \tag{3.42}$$

(iii) $\hat{\gamma}^{(1)}_{12} \cdot h_{12} = \Phi^*_{12} \cdot h_{12}$ is an extended gradient, namely,

$$(\hat{\gamma}^{(1)}_{12}h_{12})_d = (\hat{\gamma}^{(1)}_{12}h_{12})^*_d . \tag{3.43}$$

Proof: Equations (3.40)–(3.42) are a direct consequence of the definitions (3.39), of Eqs. (3.17b) and (3.17c), and of the property $q^{\pm *}_{12} = \pm q^{\pm}_{12}$.

Remark 3.3: Using Eq. (3.42) the BO class can be written in the following form:

$$\begin{aligned} q_{1_t} &= \beta_n \int dx_2\, \delta_{12} q^-_{12} (\Phi^*_{12})^n \cdot 1 \\ &= \beta_n \int dx_2\, q^-_{12} \delta_{12} (\Phi^*_{12})^n \cdot 1 \\ &= i\,\partial_{x_1} \int dx_2\, \delta_{12}(\Phi^*_{12})^n \cdot 1 = i\beta_n\,\partial_{x_1}\gamma^{(n)}_{11} . \end{aligned} \tag{3.44}$$

The first Hamiltonian operator $\Theta^{(1)}_{12} = q^-_{12}$ commutes with δ_{12} and reduces to $i\,\partial_{x_1}$. Then ∂_x is the (projected version of the) first Hamiltonian operator of the BO class; this result was already known.[15]

The existence of a compatible pair of Hamiltonian operators is connected to the existence of infinitely many constants of motion in involution. Theorems 4.1–4.5 of Ref. 4 can finally be summarized in the following proposition.

Proposition 3.5: Consider the compatible pair of Hamiltonian operators $\Theta^{(1)}_{12} \doteq q^-_{12}$, $\Theta^{(2)}_{12} \doteq (q^+_{12} - iq^-_{12}H_{12})q^-_{12}$ and define $\Phi_{12} \doteq \Theta^{(2)}_{12}(\Theta^{(1)}_{12})^{-1}$; then the following is true.

(i) Φ_{12} is a hereditary operator.

(ii) $\sigma^{(m)}_{12} \doteq \Phi^{(m)}_{12}q^-_{12} \cdot 1$ and $\gamma^{(m)}_{12} \doteq (\Phi^*_{12})^m \cdot 1$ are extended symmetries and extended gradients of conserved quantities, respectively, for Eqs. (3.2), namely,

$$\sigma^{(m)}_{12_f}[K^{(n)}] = (\delta_{12}K^{(n)}_{12})_d[\sigma^{(m)}_{12}], \tag{3.45a}$$

$$\gamma^{(m)}_{12_f}[K^{(n)}] + (\delta_{12}K^{(n)}_{12})^*_d[\gamma^{(m)}_{12}] = 0, \tag{3.45b}$$

$$((\Phi^*_{12})^m h_{12})_d = ((\Phi^*_{12})^m h_{12})^*_d, \quad h_{12} = h(x_1 - x_2) . \tag{3.45c}$$

(iii) Equations (3.2) are bi-Hamiltonian systems, since they can be written in the following two "extended" Hamiltonian forms

$$q_{1_t} = \beta_n \int_{\mathbf{R}} dx_2\, \delta_{12}\Theta^{(1)}_{12}\gamma^{(n)}_{12} = \beta_n \int_{\mathbf{R}} dx_2\, \delta_{12}\Theta^{(2)}_{12}\gamma^{(n-1)}_{12} . \tag{3.46}$$

(iv) $\sigma^{(m)}_{11}$ and $\gamma^{(m)}_{11}$ are symmetries and gradients of conserved quantities for Eq. (3.2), namely,

$$\sigma^{(m)}_{11_f}[K^{(n)}_{11}] = K^{(n)}_{11_f}[\sigma^{(m)}_{11}], \tag{3.47a}$$

$$\gamma^{(m)}_{11_f}[K^{(n)}_{11}] + K^{(n)}_{11_f}[\sigma^{(m)}_{11}] = 0, \tag{3.47b}$$

$$\gamma^{(m)}_{11_f} = \gamma^{(m)\,+}_{11_f}, \tag{3.47c}$$

where $^+$ denotes the operation of adjoint w.r.t. the bilinear form $(f,g) \doteq \int_{\mathbf{R}} dx\, fg$.

(v) The corresponding conserved quantities I_m, related to $\gamma^{(m)}_{12}$ and $\gamma^{(m)}_{11}$ via equations

$$\gamma_{12}^{(m)} = \mathrm{grad}_{12}\, I_m, \quad I_{m_d}[f_{12}] \doteq \langle \mathrm{grad}_{12}\, I_m, f_{12}\rangle, \quad (3.48a)$$

$$\gamma_{11}^{(m)} = \mathrm{grad}\, I_m, \quad I_{m_f}[f] \doteq (\mathrm{grad}\, I_m, f), \quad (3.48b)$$

are constants of motion of Eqs. (3.2).

(vi) These constants of motion are in involution with respect to the Poisson brackets

$$\{I_n, I_m\} \doteq \langle \delta_{12}\gamma_{12}^{(n)}, \Theta_{12}\gamma_{12}^{(m)}\rangle, \quad \Theta_{12} = \Theta_{12}^{(1)} \text{ and/or } \Theta_{12}^{(2)}, \quad (3.49)$$

namely,

$$\{I_n, I_m\} = 0. \quad (3.50)$$

(vii) The equations $K_{12}^{(m)} = K^{(m)}(q_1,q_2) = 0$ are Bäcklund transformations (BT) for the BO class (3.2), interpreting q_1 and q_2 as to different solutions of (3.2).

Remark 3.4: (i) The first extended symmetries of the BO class are given by

$$\sigma_{12}^{(0)} = q_{12}^- \cdot 1 = q_1 - q_2, \quad (3.51a)$$

$$\begin{aligned}\sigma_{12}^{(1)} &= \Phi_{12}^{(1)} q_{12}^- \cdot 1 \\ &= i(q_{1_{x_1}} + q_{2_{x_2}}) + H_1 q_{1_{x_1}} - H_2 q_{2_{x_2}} \\ &\quad + (q_1+q_2)(q_1-q_2) \\ &\quad - i(q_1-q_2)(H_1q_1 - H_2q_2), \end{aligned} \quad (3.51b)$$

then their projections are the first symmetries of the BO class

$$\sigma_{11}^{(0)} = 0, \quad \sigma_{11}^{(1)} = 2iq_{1_{x_1}}, \quad (3.52)$$

and equations

$$\sigma_{12}^{(0)} = 0, \quad \sigma_{12}^{(1)} = 0, \quad (3.53)$$

are the first two BT's of the class. We remark that the BT's generated by Φ_{12} are polynomial in q_1, q_2, unlike the previously known examples.[17]

D. Connection with the mastersymmetries theory

The mastersymmetry approach was introduced by Fuchssteiner and one of the authors (A.S.F.)[15] as an alternative way of generating symmetries of the BO equation. This approach was subsequently applied to (2 + 1)-dimensional systems like KP,[11] 1 + 1 systems like KdV,[12,18] and finite-dimensional systems like the Calogero–Moser problem.[19]

In this section we briefly show that the existence of a hereditary operator Φ_{12} allows a simple and elegant characterization of the BO mastersymmetries (analogous and more detailed results for KP were reported in Ref. 5).

Proposition 3.6: (i) If

$$K_{12}^{(n)} \doteq \Phi_{12}^n q_{12}^- \cdot 1, \quad (3.54a)$$

$$\tau_{12}^{(m,r)} \doteq \Phi_{12}^m q_{12}^- \cdot (x_1+x_2)^r, \quad (3.54b)$$

then

$$[\delta_{12}K_{12}^{(n)}, \tau_{12}^{(m,1)}]_d = 4inK_{12}^{(n+m-1)}. \quad (3.55)$$

(ii) $\tau_{11}^{(m,1)} \doteq \tau_{12}^{(m,1)}|_{x_2=x_1}$ are mastersymmetries of degree 1 of the BO class, since

$$[K_{11}^{(n)}, \tau_{11}^{(m,1)}]_f = 4inK_{11}^{(n+m-1)}. \quad (3.56)$$

Proof: The derivation of Eq. (3.55), presented in Appendix C, is based on the following important properties:

$$\begin{aligned}(1)\ \ &\Phi_{12d}\,[q_{12}^-(x_1+x_2)] + [\Phi_{12},(q_{12}^-(x_1+x_2))_d] \\ &= iq_{12}^-[H_{12},(x_1+x_2)^-], \end{aligned} \quad (3.57a)$$

$$(2)\ \ [H_{12},(x_1+x_2)]f_{12} = \frac{1}{\pi}\int_{\mathbf{R}} dy\, F(y, x_1-x_2), \quad (3.57b)$$

$$f_{12} = f(x_1,x_2) \doteq F(x_1+x_2, x_1-x_2), \quad (3.57c)$$

$$(3)\ \ iq_{12}^-[H_{12},(x_1-x_2)^-](\delta_{12}^s K_{12}^{(l)}) = 0, \quad \forall s,l \geqslant 0. \quad (3.57d)$$

These follow from the definitions (3.3) and from equation

$$\lim_{|x_2|\to\infty}\left(\sum_{l=0}^{s-1}(-1)^{s-1}\partial_{x_1}^{s-l-1}\partial_{x_2}^{l}K_{12}^{(n)}\right)_{x_1=x_2} = 0 \quad (3.57e)$$

(see Appendix C). Equation (3.56) follows from (3.55) using Theorem 4.1 of Ref. 4.

Remark 3.4: As for the KP case,[5] time-dependent symmetries of the BO hierarchy should be generated via mastersymmetries $\tau_{12}^{(m,r)}$ of degree $r>1$. In this case, an equation analogous to (3.55) should follow from a suitable generalization of Eq. (3.57a) obtained replacing (x_1+x_2) by $(x_1+x_2)^r$, $r>1$.

E. Connection with the complex Burgers hierarchy

It is well known that if $q(x,t)$ is analytic in the upper x plane, then the BO equation (1.30) reduces to the (complex) Burgers equation

$$q_t = 2qq_x + iq_{xx}, \quad (3.58)$$

since

$$Hf^{(\pm)} = \pm if^{(\pm)}, \quad (3.59)$$

where $f^{(+)}(x)$ and $f^{(-)}(x)$ are holomorphic in the upper and lower half x plane, respectively. The same result obviously holds for the whole hierarchy.

Proposition 3.7: If $q(x,t)$ is holomorphic in the upper x plane, then the BO hierarchy (3.2) reduces to the following complex Burgers hierarchy (investigated in Ref. 20):

$$q_t = b_n(i\,\partial_x + \partial_x q\,\partial_x^{-1})^{n-1}q_x, \quad n \geqslant 1, \quad (3.60a)$$

$$b_n \doteq 2^n i\beta_n, \quad \partial_x^{-1} \doteq \int_{-\infty}^{x} dx. \quad (3.60b)$$

Proof: The proof is straightforward and relies on the fact that each gradient $\gamma_{12}^{(n)}$ is a holomorphic function in the upper x_1 and x_2 planes; hence Eq. (3.59) implies that

$$\begin{aligned}\Phi_{12}^*\gamma_{12}^{(n)} &= (q_{12}^+ - iH_{12}q_{12}^-)\gamma_{12}^{(n)} \\ &= (q_{12}^+ + q_{12}^-)\gamma_{12}^{(n)} = 2(q_1 + i\,\partial_{x_1})\gamma_{12}^{(n)}.\end{aligned}$$

Then

$$\begin{aligned}q_{1_t} &= \beta_n\int_{\mathbf{R}} dx_2\,\delta_{12}q_{12}^-(\Phi_{12}^*)^n\cdot 1 \\ &= 2i\beta_n\,\partial_{x_1}\int_{\mathbf{R}} dx_2\,\delta_{12}(\Phi_{12}^*)^n\cdot 1 \\ &= i2^{n+1}\beta_n\,\partial_{x_1}(q_1 + i\,\partial_{x_1})^n\cdot 1\end{aligned}$$

$$= b_n \partial_{x_1} (q_1 + i\partial_{x_1})^{n-1} q_1$$
$$= b_n \partial_{x_1} (q_1 + i\partial_{x_1})^{n-1} \partial_{x_1}^{-1} q_{1_{x_1}}$$
$$= b_n (i\partial_{x_1} + \partial_{x_1} q_1 \partial_{x_1}^{-1})^{n-1} q_{1_{x_1}} .$$

ACKNOWLEDGMENTS

It is a pleasure to acknowledge useful discussions with M. J. Ablowitz, O. Ragnisco, and M. Bruschi. One of the authors (P.M.S.) wishes to thank the friendly hospitality of the Department of Mathematics and Computer Science of Clarkson University.

This work was partially supported by the Office of Naval Research under Grant No. N00014-76-C-0867 and the National Science Foundation under Grant No. DMS-8501325.

APPENDIX A

In this appendix we use the notion of directional derivative and extended bilinear form introduced in (1.24) and (1.22), (1.32), respectively, to prove some of the results presented in this paper. In order to give a self-contained presentation, we first present some results contained in Appendix C of Ref. 4.

The directional derivative of the basic operators $q_{12}^{\pm}$ (1.13b), (1.23), (1.33), is

$$q_{12_d}^{\pm}[f_{12}]g_{12} = f_{12}^{\pm} g_{12} , \tag{A1}$$

where the integral operators $f_{12}^{\pm}$, defined by

$$f_{12}^{\pm} g_{12} \doteqdot \int_{\mathbf{R}} dx_3 (f_{13}\, g_{32} \pm g_{13}\, f_{32}) , \tag{A2}$$

enjoy the following algebraic properties:

$$a_{12}^{\pm} b_{12} = \pm\, b_{12}^{\pm} a_{12} , \tag{A3a}$$

$$(a_{12}^{\pm} b_{12}^{\pm} - b_{12}^{\pm} a_{12}^{\pm}) c_{12} = (a_{12}^{-} b_{12})^{-} c_{12} = -\, c_{12}^{-} a_{12}^{-} b_{12} , \tag{A3b}$$

$$(a_{12}^{+} b_{12}^{-} \mp b_{12}^{\mp} a_{12}) c_{12} = (a_{12}^{\mp} b_{12})^{\pm} c_{12} + \pm\, c_{12}^{\pm} a_{12}^{\mp} b_{12} , \tag{A3c}$$

$$a_{12}^{\pm *} = \pm\, a_{12}^{\pm} . \tag{A3d}$$

Moreover the integral representation

$$q_{12}^{\pm} f_{12} = \int_{\mathbf{R}} dx_3 (q_{13}\, f_{32} \pm f_{13} q_{32}) \tag{A4}$$

implies that $q_{12}^{\pm}$ satisfy Eqs. (A3) as well. Equations (A3) are conveniently used to prove the following properties of the recursion and Hamiltonian operators of the KP and BO equations.

For the KP class, the following is true.

(1) $\Phi_{12} \doteqdot (\alpha\,\partial_y + q_{12}^{-})(D_1 + D_2)^{-1}$ is a strong symmetry of $\hat{K}_{12}^{0} H_{12} = (\alpha\,\partial_y + q_{12}^{-}) H_{12}$. Indeed

$$\Phi_{12_d}[\sigma_{12}] = \sigma_{12}^{-} (D_1 + D_2)^{-1} ,$$
$$(\hat{K}_{12}^{0} H_{12})_d [\sigma_{12}] = \sigma_{12}^{-} H_{12} ,$$

and

$$\begin{aligned} &\mathscr{L}(\Phi_{12}, \hat{K}_{12}^{0} H_{12}) f_{12} \\ &\quad = (\hat{K}_{12}^{0} H_{12})^{-} (D_1 + D_2)^{-1} f_{12} \\ &\qquad - (\Phi_{12} f_{12})^{-} H_{12} + \Phi_{12}\, f_{12}^{-} H_{12} \\ &\quad = ((\alpha\,\partial_y + q_{12}^{-}) H_{12})^{-} g_{12} \\ &\qquad - ((\alpha\,\partial_y + q_{12}^{-}) g_{12})^{-} H_{12} + (\alpha\,\partial_y + q_{12}^{-}) g_{12}^{-} H_{12}, \end{aligned}$$

having introduced $g_{12} \doteqdot (D_1 + D_2)^{-1} f_{12}$ and used $H_{12}(D_1 + D_2)^{-1} = (D_1 + D_2)^{-1} H_{12}$. Using (A3a) we obtain $g_{12}^{-} q_{12}^{-} H_{12} - H_{12}^{-} q_{12}^{-} g_{12} - q_{12}^{-} g_{12}^{-} H_{12}$, which is zero, for (A3b)$^{-}$.

(2) Φ_{12} is a hereditary operator. Indeed

$$\begin{aligned} &\Phi_{12_d}[\Phi_{12} f_{12}] g_{12} - \Phi_{12} \Phi_{12_d}[f_{12}] g_{12} - (\text{sym. w.r.t. } f_{12} \leftrightarrow g_{12}) \\ &= ((\alpha\,\partial_y + q_{12}^{-})(D_1 + D_2)^{-1} f_{12})^{-} (D_1 + D_2)^{-1} g_{12} - (\alpha\,\partial_y + q_{12}^{-})(D_1 + D_2)^{-1} f_{12}^{-} (D_1 + D_2)^{-1} g_{12} - (\text{sym.} \cdots) \\ &\quad = ((D_1 + D_2)^{-1} g_{12})^{-} (\alpha\,\partial_y + q_{12}^{-})(D_1 + D_2)^{-1} f_{12} - ((D_1 + D_2)^{-1} f_{12})^{-} (\alpha\,\partial_y + q_{12}^{-})(D_1 + D_2)^{-1} g_{12} \\ &\qquad - (\alpha\,\partial_y + q_{12}^{-})(D_1 + D_2)^{-1} (f_{12}^{-} (D_1 + D_1)^{-1} g_{12} - g_{12}^{-} (D_1 + D_2)^{-1} f_{12}) = 0 , \end{aligned}$$

using integration by parts,

$$\begin{aligned} &(D_1 + D_2)^{-1} f_{12}^{-} (D_1 + D_2)^{-1} g_{12}^{-} \\ &\quad = ((D_1 + D_2)^{-1} f_{12})^{-} (D_1 + D_2)^{-1} g_{12} - (D_1 + D_2)^{-1} \\ &\qquad \times (((D_1 + D_2)^{-1} f_{12})^{-} g_{12} - g_{12}^{-} (D_1 + D_2)^{-1} f_{12}) \end{aligned}$$

and Eq. (A3b).

(3) $[\hat{K}_{12}^{0} H_{12}^{(1)}, \hat{K}_{12}^{0} H_{12}^{(2)}]_d$

$$\begin{aligned} &= (\hat{K}_{12}^{0} H_{12}^{(2)})^{-} H_{12}^{(1)} - (\hat{K}_{12}^{0} H_{12}^{(1)})^{-} H_{12}^{(2)} \\ &= -H_{12}^{(1)^{-}} (\alpha\,\partial_y + q_{12}^{-}) \\ &\quad + H_{12}^{(2)^{-}} (\alpha\,\partial_y + q_{12}^{-}) H_{12}^{(1)} \\ &= -\hat{K}_{12}^{0} H_{12}^{(1)^{-}} H_{12}^{(2)} , \end{aligned}$$

for (A3a)$^{-}$ and (A3b)$^{-}$.

For the BO class the following is true.

(4) Φ_{12} is a strong symmetry of $q_{12}^{-} h_{12}$, $h_{12} = h(x_1 - x_2)$. Indeed, using (3.43a), we have

$$\begin{aligned} \mathscr{L}(\Phi_{12}, q_{12}^{-} h_{12}) &= (q_{12}^{-} h_{12})^{+} f_{12} - i(q_{12}^{-} h_{12})^{-} H_{12} f_{12} \\ &\quad + (q_{12}^{+} - i q_{12}^{-} H_{12}) f_{12}^{-} h_{12} \\ &\quad - (q_{12}^{+} f_{12} - i q_{12}^{-} H_{12} f_{12})^{-} h_{12} . \end{aligned}$$

Using Eqs. (A3a) and property (3.17a) [see Appendix B (5)] we obtain

$$\begin{aligned} &(f_{12}^{+} q_{12}^{-} h_{12} + q_{12}^{+} f_{12}^{-} h_{12} + h_{12}^{-} q_{12}^{+} f_{12}) \\ &\quad + i((H_{12} f_{12})^{-} q_{12}^{-} h_{12} - q_{12}^{-} (H_{12} f_{12})^{-} h_{12} \\ &\quad - h_{12}^{-} q_{12}^{-} H_{12} f_{12}) , \end{aligned}$$

and the two expressions in parentheses are zero using (A3c) and $(A3b)^-$, respectively.

(5) Φ_{12} is a hereditary operator. Using (3.24a) we have that

$$\Phi_{12_d}[\Phi_{12} f_{12}]g_{12} - \Phi_{12}\Phi_{12_d}[f_{12}]g_{12}$$
$$- (\text{sym. w.r.t. } f_{12}\leftrightarrow g_{12})$$
$$= (q_{12}^+ f_{12} - i q_{12}^- H_{12} f_{12}) g_{12}^+$$
$$- i(q_{12}^+ f_{12} - i q_{12}^- H_{12} f_{12})^- H_{12}\, g_{12}$$
$$- q_{12}^+ (f_{12}^+ g_{12} - i f_{12}^- H_{12}\, g_{12})$$
$$+ i q_{12}^- H_{12}(f_{12}^+ g_{12} - i f_{12}^- H_{12}\, g_{12})$$
$$- (\text{sym. w.r.t. } f_{12}\leftrightarrow g_{12}) .$$

Using (A3c) and (A3b) we obtain

$$q_{12}^- (H_{12}(g_{12}^- H_{12} f_{12} + (H_{12}\, g_{12})^- f_{12})$$
$$+ g_{12}^- f_{12} - (H_{12}\, g_{12})^- H_{12} f_{12}) ,$$

which is zero for Eq. (3.17b).

(6) $q_{12}^{\pm *} = \pm q_{12}^{\pm}$.

These are direct consequences of the definitions (1.32) and (1.33). Their immediate implications are Eqs. (3.17c), (3.40a), and (3.41).

(7) $\Theta_{12}^{(1)} = q_{12}^-$ and $\Theta_{12}^{(2)}$ are Hamiltonian operators. They are skew symmetric, since

$$\Theta_{12}^{(2)*} = ((q_{12}^+ - i q_{12}^- H_{12}) q_{12}^-)^* = q_{12}^{-*}(q_{12}^{+*} - i H_{12}^* q_{12}^{-*})$$
$$= - q_{12}^- (q_{12}^+ - i H_{12} q_{12}^-)$$
$$= - \Theta_{12}^{(2)} \ (\text{being } q_{12}^+ q_{12}^- = q_{12}^- q_{12}^+) .$$

They satisfy the Jacobi identity (3.40b), for instance

$$\langle a_{12}, \Theta_{12_d}^{(1)}[\Theta_{12}^{(1)} b_{12}] c_{12}\rangle$$
$$= \langle a_{12}, q_{12_d}^-[q_{12}^- b_{12}] c_{12}\rangle + \text{cycl. perm. } s$$
$$= \langle a_{12}, (q_{12}^- b_{12})^- c_{12}\rangle + \text{cycl. perm. } s .$$

Using (A3a) and (A3d) we obtain

$$\langle a_{12}, - c_{12}^- q_{12}^- b_{12} + b_{12}^- q_{12}^- c_{12} - q_{12}^- b_{12}^- c_{12}\rangle ,$$

which is zero for any a_{12}, b_{12}, c_{12}, for Eq. (A3b).

(8) The derivation of Eqs. (3.36) is the same as for the corresponding ones of the KP hierarchy (see Appendix C of Ref. 4) and makes extensive use of the equations

$$(\delta_{12}^n)^\pm f_{12} = (\partial_{x_1}^n \pm (-1)^n \partial_{x_2}^n) f_{12} , \tag{A5}$$
$$(\delta_{12} K_{12}^{(0)})_d [f_{12}] = (\delta_{12} q_{12}^- \cdot 1)_d [F_{12}] = (q_{12}^- \delta_{12})_d [f_{12}] = f_{12}^- \delta_{12} = - \delta_{12}^- f_{12} = 0 ,$$
$$(\delta_{12} K_{12}^{(1)})_d [f_{12}] = (\Phi_{12} q_{12}^- \delta_{12})_d [f_{12}] - 2i (q_{12}^- \delta_{12}^1)_d [f_{12}]$$
$$= \Phi_{12_d}[f_{12}] q_{12}^- \delta_{12} + \Phi_{12} q_{12_d}^- [f_{12}] \delta_{12} - 2i q_{12_d}^- [f_{12}] \delta_{12}^1$$
$$= (f_{12}^+ - i f_{12}^- H_{12}) q_{12}^- \delta_{12} + \Phi_{12} f_{12}^- \delta_{12} - 2i f_{12}^- \delta_{12}^1$$
$$= f_{12}^+ q_{12}^- \delta_{12} - i f_{12}^- H_{12} q_{12}^- \delta_{12} - \Phi_{12} \delta_{12}^- f_{12} + 2i (\delta_{12}^1)^- f_{12} = 2i (\partial_{x_1} + \partial_{x_2}) f_{12} ,$$

since

$$f_{12}^+ q_{12}^- \delta_{12} = q_{12}^- f_{12}^+ \delta_{12} - \delta^+ q_{12}^- f_{12} = 2(q_{12}^- - q_{12}^-) f_{12} = 0 ,$$
$$f_{12}^- H_{12} q_{12}^- \delta_{12} = f_{12}^- \delta_{12} H_{12}(q_1 - q_2) = f_{12}^- \delta_{12}(H_1 q_1 - H_2 q_2) = 0 ,$$
$$(\delta_{12}^1)^- f_{12} = (\partial_{x_1} + \partial_{x_2}) f_{12} ,$$
$$(\delta_{12} K_{12}^{(2)})_d [f_{12}] = (\Phi_{12}^2 q_{12}^- \delta_{12})_d [f_{12}] - 4i(\Phi_{12} q_{12}^- \delta_{12}^1)_d [f_{12}] - 4(q_{12}^- \delta_{12}^2)_d [f_{12}]$$
$$= \Phi_{12_d}[f_{12}] \Phi_{12} q_{12}^- \delta_{12} + \Phi_{12} \Phi_{12_d}[f_{12}] q_{12}^- \delta_{12} + \Phi_{12}^2 q_{12_d}^- [f_{12}] \delta_{12}$$
$$- 4i \Phi_{12_d}[f_{12}] q_{12}^- \delta_{12}^1 - 4i \Phi_{12} q_{12_d}^- [f_{12}] \delta_{12}^1 - 4 q_{12_d}^- [f_{12}] \delta_{12}^2$$
$$= (f_{12}^+ - i f_{12}^- H_{12}) \Phi_{12} q_{12}^- \delta_{12} + \Phi_{12}(f_{12}^+ - i f_{12}^- H_{12}) q_{12}^- \delta_{12} + \Phi_{12}^2 f_{12}^- \delta_{12}$$
$$- 4i (f_{12}^+ - i f_{12}^- H_{12}) q_{12}^- \delta_{12}^1 - 4i \Phi_{12} f_{12}^- \delta_{12}^1 - 4 f_{12}^- \delta_{12}^2$$
$$= 4i (H_{12}(\partial_{x_1} + \partial_{x_2})^2 + (\partial_{x_1} + \partial_{x_2})(q_1 + q_2) + i(H_1 q_{1_{x_1}} - H_2 q_{2_{x_2}}) - i(q_1 - q_2) H_{12}(\partial_{x_1} + \partial_{x_2})) ,$$

since, for instance,

$$f_{12}^+ \Phi_{12} q_{12}^- \delta_{12} = f_{12}^+ (\delta_{12} K_{12}^{(1)} + 2i \delta_{12}^1 K_{12}^{(0)}) = f_{12}(K_{22}^{(1)} + K_{11}^{(1)}) - 2i [(\partial_{x_3}(K_{32}^{(0)} f_{13}))_{x_3 = x_2} - (\partial_{x_3}(K_{13}^{(0)} f_{32}))_{x_3 = x_1}]$$
$$= f_{12}(2i(q_{1_{x_1}} + q_{2_{x_2}}) - 2i(q_{1_{x_1}} + q_{2_{x_2}})) = 0 ,$$
$$f_{12}^- H_{12}(\delta_{12} K_{12}^{(1)} + 2i \delta_{12}^1 K_{12}^{(0)}) = f_{12}^- (\delta_{12} H_{12} K_{12}^{(1)} + 2i \delta_{12}^1 H_{12} K_{12}^{(0)})$$
$$= f_{12}(H_2 K_{22}^{(1)} - H_1 K_{11}^{(1)}) - 2i [(\partial_{x_3}(f_{13} H_{32} K_{32}^{(0)}))_{x_3 = x_2} + (\partial_{x_3}(f_{32} H_{13} K_{13}^{(0)}))_{x_3 = x_1}]$$
$$= 2i((H_1 q_{1_{x_1}} - H_2 q_{2_{x_2}}) - (H_1 q_{1_{x_1}} - H_2 q_{2_{x_2}})) f_{12} = 0 ;$$
$$f_{12}^+ q_{12}^- \delta_{12}^1 = q_{12}^- f_{12}^+ \delta_{12}^1 - \delta^{1+} q_{12}^- f_{12} = (q_{12}^- (\partial_{x_1} - \partial_{x_2}) - (\partial_{x_1} - \partial_{x_2}) q_{12}^-) f_{12} = - (q_{1_{x_1}} + q_{2_{x_2}}) f_{12} ;$$
$$f_{12}^- H_{12} q_{12}^- \delta_{12}^1 = f_{12}^- \delta_{12}^1 (H_1 q_1 - H_2 q_2) = - (\partial_{x_3}(f_{13}(H_3 q_3 - H_2 q_2)))_{x_3 = x_2} - (\partial_{x_3} f_{32}(H_1 q_1 - H_3 q_3))_{x_3 = x_1}$$
$$= - H_2 q_{2_{x_2}} + H_1 q_{1_{x_2}} .$$

APPENDIX B

In this appendix we prove some of the properties of the extended Hilbert transform presented in Proposition 3.2.

(1) $\int_{\mathbb{R}} dx_2\, \delta_{12} H_{12}\, g_{12} = H_1 g_{11}$,

since

$$\int_{\mathbb{R}} dx_2\, \delta_{12}\pi^{-1} \int_{\mathbb{R}} dy[y-(x_1+x_2)]^{-1} G(y,x_1-x_2)$$
$$= \pi^{-1}\int_{\mathbb{R}} dy (y-2x_1)^{-1} G(y,0)$$
$$= \pi^{-1}\int_{\mathbb{R}} dy (y-x_1)^{-1} G(2y,0) = H_1 g_{11},$$

$g(x_1,x_2) \doteq G(x_1+x_2,x_1-x_2)$.

(2) $H_{12}a(x_j) = H_j a(x_j), \quad j=1,2$,

since

$$H_{12}a(x_1)$$
$$= \pi^{-1}\int_{\mathbb{R}} dy[y-(x_1+x_2)]^{-1} a\left(\frac{y}{2} + \frac{x_1-x_2}{2}\right)$$
$$= \pi^{-1}\int_{\mathbb{R}} dy (y-x_1)^{-1} a(y) = H_1 a(x_1).$$

APPENDIX C

In order to prove that Eq. (3.55) holds, we must first derive Eqs. (3.57).

(a) Derivation of Eqs. (3.57):

$$\mathscr{L}(\Phi_{12}, q_{12}^-(x_1+x_2))f_{12}$$
$$= (q_{12}^-(x_1+x_2))^+ f_{12} - i(q_{12}^-(x_1+x_2))^- H_{12} f_{12}$$
$$+ (q_{12}^+ - iq_{12}^- H_{12}) f_{12}^-(x_1+x_2)$$
$$- (q_{12}^+ f_{12} - iq_{12}^- H_{12} f_{12})^-(x_1+x_2).$$

Then, using Eqs. (A3a), (A3c), and (A3b), we obtain

$$\mathscr{L}(\Phi_{12}, q_{12}^-(x_1+x_2))f_{12}$$
$$= iq_{12}^-((H_{12}f_{12})^-(x_1+x_2) - H_{12}\, f_{12}^-(x_1+x_2))$$
$$= iq_{12}^-(H_{12}(x_1+x_2)^- - (x_1+x_2)^- H_{12})f_{12},$$

which is Eq. (3.57a)

Equation (3.57b) is a straightforward generalization of equation

$$[H,x]f = \frac{1}{\pi}\int_{\mathbb{R}} dx'\, f(x').$$

In order to prove Eq. (3.57d), we first prove that

$$(H_{12}(x_1+x_2)^- - (x_1+x_2)^- H_{12})(\delta_{12}^s f_{12}) = c_s, \quad \text{(C1a)}$$

$$c_s \doteq \frac{1}{\pi}\int_{\mathbb{R}^2} dx_1\, dx_2\, \delta_{12}(\partial_{x_1} + \partial_{x_2})$$
$$\times \sum_{l=0}^{s-1}(-1)^{s-l}\partial_{x_1}^{s-l-1}\partial_{x_2}^{l} f_{12}$$
$$= \frac{1}{\pi}\int_{\mathbb{R}} dx_2\, \partial_{x_2}\left(\sum_{l=0}^{s-1}(-1)^{s-l}\partial_{x_1}^{s-l-1}\partial_{x_2}^{l} f_{12}\right)_{x_1=x_2}, \quad \text{(C1b)}$$

$$(H_{12}(x_1+x_2)^- - (x_1+x_2)^- H_{12})\delta_{12}^s f_{12}$$
$$= H_{12}\int_{\mathbb{R}} dx_3[(x_1+x_3)\delta_{32}^s f_{32} - \delta_{13}^s f_{13}(x_3+x_2)]$$
$$- \int_{\mathbb{R}} dx_3[(x_1+x_3)H_{32}\delta_{32}^s f_{32}$$
$$- (H_{13}\delta_{13}^s f_{13})(x_3+x_2)]$$
$$= H_{12}((-1)^s(s(\partial_{x_3}^{s-1} f_{32})_{x_3=x_2}$$
$$+ (x_1+x_2)(\partial_{x_3}^{s} f_{32})_{x_3=x_2})$$
$$- s(\partial_{x_3}^{s-1} f_{13})_{x_3=x_1} - (x_1+x_2)(\partial_{x_3}^{s} f_{13})_{x_3=x_1})$$
$$- (-1)^s(s(H_{32}\,\partial_{x_3}^{s-1} f_{32})_{x_3=x_2}$$
$$+ (x_1+x_2)(H_{32}\,\partial_{x_3}^{s} f_{32})_{x_3=x_2}) + s(H_{13}\,\partial_{x_3}^{s-1} f_{13})_{x_3=x_1}$$
$$+ (x_1+x_2)(H_{13}\,\partial_{x_3}^{s} f_{13})_{x_3=x_1},$$

where we have used Eq. (3.16e); using now (3.16d) we obtain

$$[H_{12},(x_1+x_2)]((-1)^s(\partial_{x_3}^{s} f_{32})_{x_3=x_2} - (\partial_{x_3}^{s} f_{13})_{x_3=x_1}),$$

and Eq. (3.57b) finally leads to Eq. (C1).

Equation (3.57d) directly follows from Eq. (C1) when $f_{12} = K_{12}^{(l)}$, since Eq. (3.57e) holds.

(b) Derivation of Eq. (3.55):

$$[\delta_{12}K_{12}^{(n)}, \tau_{12}^{(m,1)}]_d$$
$$= \sum_{l=0}^{n}(-2i)^l\binom{n}{l}[\Phi_{12}^{n-l}q_{12}^-\delta_{12}^l, \Phi_{12}^m q_{12}^-(x_1+x_2)]_d$$
$$= \sum_{l=0}^{n}(-2i)^l\binom{n}{l}\Big(\Phi_{12}^{n+m-l}[q_{12}^-\delta_{12}^l, q_{12}^-(x_1+x_2)]_d$$
$$+ i\Phi_{12}^m\sum_{r=1}^{n-l}\Phi_{12}^{n-l-r}q_{12}^-[H_{12},(x_1+x_2)^-]$$
$$\times \Phi_{12}^{r-1}q_{12}^-\delta_{12}^l\Big),$$

having used the fact that Φ_{12} is a strong symmetry of $q_{12}^- h_{12}$, Eq. (3.57a) and Eq. (2.8) of Ref. 5. Equation (3.28) and equation $[\delta_{12}^l,(x_1+x_2)]_I = 2\delta_{1,l}$, $\delta_{1,l} = 1$ if $l=1$ and 0 if $l \neq 1$, then yield

$$4inK_{12}^{(n+m-1)}$$
$$+ i\sum_{l=0}^{n}\sum_{r=1}^{n-l}\sum_{j=0}^{r-1}(-2i)^l\binom{n}{l}(2i)^j\binom{r-1}{j}\Phi_{12}^{n-l+m-r}q_{12}^-$$
$$\times[H_{12},(x_1+x_2)^-](\delta_{12}^{l+j}K_{12}^{(r-1-j)}) = 4inK_{12}^{(n+m-1)},$$

for Eq. (3.57d).

[1]B. B. Kadomtsev and V. I. Petviashvili, Sov. Phys. Dokl. **15**, 539 (1970).
[2]T. B. Benjamin, J. Fluid Mech. **29**, 559 (1967).
[3]H. Ono, J. Phys. Soc. Jpn. **39**, 1082 (1975).
[4]P. M. Santini and A. S. Fokas, "Recursion operators and bi-Hamiltonian structures in multidimensions I," Commun. Math. Phys. (to be published).
[5]A. S. Fokas and P. M. Santini, "Recursion operators and bi-Hamiltonian structures in multidimensions II," Commun. Math. Phys. (to be published).
[6]A. S. Fokas and P. M. Santini, Stud. Appl. Math. **75**, 179 (1986).
[7]A. S. Fokas and M. J. Ablowitz, Stud. Appl. Math. **68**, 1 (1983).
[8]P. M. Santini, "Bi-Hamiltonian formulations of the intermediate long wave equation," Clarkson University preprint INS #80, May 1987.
[9]R. I. Joseph, J. Phys. A: Math. Gen. **10**, L225 (1977).
[10]T. K. Kubota and D. Dobbs, J. Hydronaut. **12**, 157 (1978).

[11]W. Oevel and B. Fuchssteiner, Phys. Lett. A **88**, 323 (1982); H. H. Chen, Y. C. Lee and J. E. Lin, Physica D **9**, 493 (1983); K. M. Case, J. Math. Phys. **26**, 1158 (1985); A. Yu. Orlof and E. I. Shulman, Lett. Math. Phys. **12**, 171 (1986).
[12]B. Fuchssteiner, Prog. Theor. Phys. **70**, 150 (1983).
[13]P. M. Santini, "Integrable 2 + 1 dimensional equations, their recursion operators and bi-Hamiltonian structures as reduction of multi-dimensional systems," in *Inverse Problems and Interdisciplinary Applications*, edited by P. C. Sabatier (Academic, London, to be published).
[14]V. E. Zakharov and B. G. Konopelchenko, Commun. Math. Phys. **94**, 483 (1984).
[15]A. S. Fokas and B. Fuchssteiner, Phys. Lett. A **86**, 341 (1981).
[16]M. Bruschi has also noticed the possibility of a double representation of the KP hierarchy (private communication).
[17]J. Satsuma, M. J. Ablowitz, and Y. Kodama, Phys. Lett. A **73**, 283 (1979).
[18]A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982).
[19]W. Oevel, "A geometrical approach to integrable systems admitting scaling symmetries," University of Paderborn, preprint, 1986; F. Magri (private communication); I. Y. Dorfman, "Deformations of the Hamiltonian structures and integrable systems," preprint.
[20]W. Oevel, "Mastersymmetries for finite dimensional integrable systems: The Calogero–Moser system," University of Paderborn preprint, 1986.
[21]M. Bruschi and O. Ragnisco, J. Math. Phys. **26**, 943 (1985).

The inverse problem concerning symmetries of ordinary differential equations

F. Gonzalez-Gascon
Departamento de Metotos Matemáticos, Facultad de C. Fisicas, Universidad Complutense, Madrid-3, Spain
A. González-López
Department of Physics, Princeton University, Princeton, New Jersey 08544

(Received 13 August 1987; accepted for publication 21 October 1987)

It is shown that for any local Lie group G of transformations in $R\times R^n$ there exist differential systems of the form $\mathbf{x}^{(m} = \mathbf{f}(t,\mathbf{x},...,\mathbf{x}^{(m-1})$, which are symmetrical under G. The order m of these systems is related to r, the number of essential parameters of G.

I. INTRODUCTION

In a recent paper[1] it was shown that for normal systems of differential equations of type

$$\mathbf{x}^{(m} = \mathbf{f}(t,\mathbf{x},\dot{x},...,\bar{x}^{(m-1})\mathbf{x}\in\mathbb{R}^n, \quad (1)$$

the maximal number of its pointlike symmetry vectors is (i) infinite, when $m=1$; (ii) not greater than n^2+4n+3, when $m=2$; (iii) not greater than $2n^2+nm+2$, when $m>2$. Since the number $2n^2+nm+2$ increases without limit with m, the question arises of whether or not it is possible to find a system of type (1) symmetrical under a given group G, for a sufficiently high value of m. We prove that the reply to this question is affirmative. Our result is local, in the sense that the function f of (1), whose existence we prove, will be, in general, only locally defined.

Note that since a first-order ($m=1$) system always possesses an infinite number of pointlike symmetry vectors, one could naively expect to find for any G a first-order system possessing G among its symmetries. That this is not generally possible is seen if G is, for instance, a group acting transitively on the $(t,\mathbf{x},\dot{x})$ space. For an example see Part (1) of Sec. III.

Note also that the construction given here does not guarantee that G is the maximal group of pointlike symmetries G_M of (1), but only that $G\subset G_M$.

II. MAIN RESULT

Assume that $\mathbf{S}_i(t,\mathbf{x})$, $i=l,...,r$, is a basis of generators of G. Calling $\mathbf{S}_i^e$ the e-order extension of $\mathbf{S}_i$ we have[2]

$$[\mathbf{S}_i^{(e},\mathbf{S}_j^{(e}] = \sum_{k=1}^{r} c_{ijk}\mathbf{S}_k^{(e},$$
$$e=0,1,2,..., \quad i,j=1,...,r, \quad (2)$$

where c_{ijk} are the structure constants of G associated with the basis $\{\mathbf{S}_i(t,\mathbf{x})\}$. On the other hand, the necessary and sufficient condition in order that G be a symmetry group of equations (1) is[2]

$$\mathbf{S}_i^{(m}(\mathbf{x}^{(m}-\mathbf{f})\Big|_{\mathbf{x}^m=\mathbf{f}} = \mathbf{0}, \quad i=1,...,r. \quad (3)$$

Conditions (3) indicate that the manifold M^m of $(t,\mathbf{x},...,\mathbf{x}^{(m})$ space defined by Eq. (1) is invariant under the action of the vector fields $\mathbf{S}_1^{(m},...,\mathbf{S}_r^{(m}$. We are going to prove that given G, one can find a sufficiently high m such that, for a certain $\mathbf{f}$, Eqs. (1) do possess G as a group of symmetries. The idea of the proof is to eliminate the possible transitivity of the action of G^e on $D^e=\{(t,\mathbf{x},...,\mathbf{x}^{(e})\}$ for low values of e by making e bigger and bigger. This is made possible, essentially, due to property (2), implying that at any point of D^e the vector fields $\mathbf{S}_i^{(e}$ generate an involutive distribution $\mathscr{D}^e$ (Ref. 3) of dimension not greater than r. To avoid singularity points where the dimension of the distribution $\mathscr{D}^e$ changes value, we restrict conveniently the domain D^e in order that in this restricted domain $\widetilde{D}^e$, $\dim(\mathscr{D}^e)$ keeps a constant and maximum value d_e. Of course $\dim\mathscr{D}^{e-1}=d_e-1$ in the projection of $\widetilde{D}^e$ along the $\mathbf{x}^{(e}$ axis. See Ref. 4 for details.

Therefore let $\mathbf{S}_1^{(e},...,\mathbf{S}_{d_e}^{(e}$ be a local basis of $\mathscr{D}^e$. Note that it might be necessary to renumber the generators of G for the basis of $\mathscr{D}^e$ to appear in this way. Conditions (3) for the symmetry of (1) under G now take the form

$$\mathbf{S}_i^{(m}(\mathbf{x}^{(m}-\mathbf{f})\Big|_{\mathbf{x}^{(m}=\mathbf{f}} = \mathbf{0}, \quad i=1,...,d. \quad (4)$$

Writing Eq. (1) in the implicit form

$$\mathbf{E}(t,\mathbf{x},...,\mathbf{x}^{(m})=0, \quad (5)$$

where $\mathbf{E}$ is a vector of m components, Eqs. (4) take the form

$$\mathbf{S}_i^{(m}(\mathbf{E})_{\mathbf{E}=\mathbf{0}}=\mathbf{0}, \quad i=1,...,d_m. \quad (6)$$

A sufficient condition necessary for Eqs. (6) to be satisfied is that the m components of the function $\mathbf{E}$ of (6) be local first integrals of $\mathbf{S}_i^{(m}$, that is, if $\mathbf{E}$ satisfies

$$\mathbf{S}_i^{(m}(\mathbf{E})=\mathbf{0}, \quad i=1,...,d_m. \quad (7)$$

But according to the Frobenius theorem[3] the number of locally independent first integrals I of an involutive distribution like $\mathscr{D}^m$ is $d_I=\dim(D^m)-d_m=1+n(1+m)-d_m$. Now, since $d_m\leqslant r$ it follows that $d_I\geqslant n$ for sufficiently large m. Assuming $d_I\geqslant n$, in order to satisfy (7) it is sufficient to choose n first integrals I of $\mathbf{S}_i^{(m}$ such that they satisfy the additional requirement

$$\operatorname{rank}\left(\frac{\partial I_i}{\partial x_k^{(m}}\right)=n, \quad k=1,...,n. \quad (8)$$

Condition (8) guarantees, via the implicit function theorem, that the system of differential equations

$$\begin{aligned} I_1 &= C_1,\\ &\vdots\\ I_n &= C_n \end{aligned} \quad (9)$$

 0022-2488/88/030618-04$02.50

can be locally written in the normal form (1). The symbols $C_1,...,C_n$ in Eqs. (9) are real numbers, and they appear since Eqs. (7) are clearly equivalent to

$$\mathbf{S}_i^{(m}(\mathbf{E}-\mathbf{C}) = \mathbf{0}, \tag{10}$$

for any $\mathbf{C}\in R^n$. Let us see that condition (8) can be satisfied if m is chosen such that

$$\dim(\mathscr{D}^m) = \dim(\mathscr{D}^{m-1}). \tag{11}$$

In fact, if (8) were not satisfied by an appropriate choice of $I_1,...,I_n$ between the d_I first integrals of $\mathscr{D}^m$, we would have

$$\sum_{k=1}^{n} a_k \frac{\partial I_i}{\partial x_k^{(m}} = 0, \quad i = 1,2,...,d_I, \tag{12}$$

where a_k are functions on D^m.

But (12) implies that the vector field $\mathbf{Z}$ defined by

$$\mathbf{Z} = \sum_{x=1}^{n} a_k \frac{\partial}{\partial x_k^{(m}} \tag{13}$$

has $I_1,...,I_{d_I}$ as first integrals. Therefore $\mathbf{Z}\in\mathscr{D}^m$ and we can write

$$\mathbf{Z} = \sum_{i=1}^{d_m} c_i \mathbf{S}_i^{(m}, \tag{14}$$

for certain functions c_i defined on D^m.

Projecting (14) on the vectors

$$\frac{\partial}{\partial t}; \frac{\partial}{\partial x_i}; \frac{\partial}{\partial x_i^{(1}},..., \frac{\partial}{\partial x_i^{(m-1}}, \quad \mathbf{0} = \sum_{i=1}^{d_m} c_i \mathbf{S}_i^{(m-1}. \tag{15}$$

But (15) and (11) are contradictory since from the fact that $\mathbf{S}_1^{(m},...,\mathbf{S}_{d_m}^{(m}$ are a basis of $\mathscr{D}^m$ it immediately follows (note that $\mathbf{S}_i^{m-1}$ does not depend on $\mathbf{x}^{(m)}$ that $\mathbf{S}_1^{(m-1},...,\mathbf{S}_{d_m}^{(m-1}$ generate $\mathscr{D}^{m-1}$. Hence by (16), $\dim \mathscr{D}^{m-1} < d_m$ and (12) is contradicted. Therefore (11) implies (8).

It remains only to prove that (11) can always be satisfied by choosing m conveniently. But this follows from the fact[4] that

$$\dim(\mathscr{D}^r) \leqslant \dim(\mathscr{D}^{r+1}). \tag{16}$$

Indeed, since the dimension is a positive integer, nondecreasing by (16), and bounded by r (the number of parameters of the group) it is obvious that for a certain m (12) holds. Furthermore, m satisfies

$$m \leqslant r - \dim(\mathscr{D}^0) + 1 = r_1, \tag{17}$$

since the worst situation that can occur concerning (11) is that

$$\dim(\mathscr{D}^s) = \dim(\mathscr{D}^{s-1}) + 1, \quad s < m, \tag{18}$$

in which case (17) would hold with the equal sign.

Note that if we require m to be greater than (or equal to) a given k ($k = 1,2,...$) then the above considerations lead to the inequality

$$m \leqslant r - \dim(\mathscr{D}^{k-1}) + k = r_k, \tag{19}$$

the equality sign being valid only when the sequence of dimensions $\dim(\mathscr{D}^{k-1})$, $\dim(\mathscr{D}^k)$,... is strictly increasing by 1 at each step and condition (11) is fulfilled when $\dim(\mathscr{D}^m) = \dim(\mathscr{D}^{m-1}) = r$.

Note also that calling $r(G)$ the minimum integer such that for $k \geqslant r(G)$, $\dim(\mathscr{D}^k)$ maintains a constant value, that is, $\dim(\mathscr{D}^k) = \dim(\mathscr{D}^{k'})$ for every $k,k' \geqslant r(G)$, by (11) we can say that for *every* $m > r(G)$ there are systems of differential equations of order m invariant under G.

Note finally that [see Eq. (11)] the construction given here actually assures the existence of n-parameter families of systems of type (1) invariant, for any value of the parameters, under G. Let us now see, with two examples, that condition (11) is not necessary for the existence of *individual* systems of type (1) invariant under G.

III. EXAMPLES SHOWING THAT CONDITION (11) IS NOT NECESSARY

(1) Let us take as G the Poincaré group in $R \times R^2$ (two spatial dimensions). We shall see that (11) is not satisfied either for $m = 1$ or for $m = 2$. Nevertheless, as has been shown elsewhere[5] $\ddot{x} = \overline{0}$ is a second-order differential system (in fact the only one) invariant under the Poincaré group in $R \times R^2$. In fact the six generators of the group are

$$\mathbf{S}_1 = \frac{\partial}{\partial t}; \quad \mathbf{S}_{1+c} = \frac{\partial}{\partial x_i}; \quad \mathbf{S}_4 = x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1};$$

$$\mathbf{S}_{4+i} = -\left(x_i \frac{\partial}{\partial t} + t \frac{\partial}{\partial x_i}\right); \quad i = 1,2. \tag{20}$$

Since $\dim(\mathscr{D}^0) = 3$ the group acts transitively on $D^0 = (t,x_1,x_2)$.

The corresponding generators of the first extension of G are given by

$$\mathbf{S}_1^{(1} = \frac{\partial}{\partial t}; \quad \mathbf{S}_{1+i}^{(1} = \frac{\partial}{\partial x_i},$$

$$\mathbf{S}_4^{(1} = \mathbf{S}_4 + \dot{x}_1 \frac{\partial}{\partial \dot{x}_2} - \dot{x}_2 \frac{\partial}{\partial \dot{x}_1}, \tag{21}$$

$$\mathbf{S}_{4+i}^{(1} = \mathbf{S}_{4+i} + (\dot{x}_i^2 - 1) \frac{\partial}{\partial \dot{x}_i} + \dot{x}_i \dot{x}_j \frac{\partial}{\partial \dot{x}_j}.$$

One can immediately check that $\dim(\mathscr{D}^1) = 5$. Indeed, the singular points of $\mathscr{D}^1$ are only those satisfying $1 - \dot{x}_1^2 - \dot{x}_2^2 = 0$. Therefore $\tilde{D}_1 = \{(t,x_1,x_2,\dot{x}_1,\dot{x}_2 | 1 - \dot{x}_1^2 - \dot{x}_2^2 \neq 0\}$. Here G^1 acts transitively in each of the two unconnected components of $\tilde{D}_1$ and also in the set $1 - \dot{x}_1^2 - \dot{x}_2^2 = 0$. Accordingly, it is impossible to find a single first-order system of type (1) symmetric under G.

The second-order extension of G is defined by

$$\mathbf{S}_1^{(2} = \frac{\partial}{\partial t}; \quad \mathbf{S}_{1+i}^{(2} = \frac{\partial}{\partial x_i},$$

$$\mathbf{S}_4^{(2} = \mathbf{S}_4^{(1} + \left(-\ddot{x}_2 \frac{\partial}{\partial \ddot{x}_1} + \ddot{x}_1 \frac{\partial}{\partial \ddot{x}_2}\right),$$

$$\mathbf{S}_5^{(2} = \mathbf{S}_5^{(1} + \left(3\dot{x}_1\ddot{x}_1 \frac{\partial}{\partial \ddot{x}_1} + (2\dot{x}_1\ddot{x}_2 + \dot{x}_2\ddot{x}_1) \frac{\partial}{\partial \ddot{x}_2}\right), \tag{22}$$

$$\mathbf{S}_6^{(2} = \mathbf{S}_6^{(1} + \left((2\dot{x}_2\ddot{x}_1 + \dot{x}_1\ddot{x}_2) \frac{\partial}{\partial \ddot{x}_1} + 3\dot{x}_2\ddot{x}_2 \frac{\partial}{\partial \ddot{x}_2}\right).$$

We can see that $\dim \mathscr{D}^2 = 6$. Since $\dim \mathscr{D}^2 = 6 > \dim(\mathscr{D}^1) = 5$ condition (11) is not satisfied. But from this fact one cannot conclude, in general, that there are not systems of type (1), for $m = 2$, symmetrical under G. In fact the system $\dddot{x} = 0$ is a system invariant under G.

Note that since $\dim \mathscr{D}^2 = 6$ and the group has six parameters it is clear that for $k \geqslant 2$ we will have $\dim(\mathscr{D}^k) = 6$. Therefore (12) is satisfied for every $m > 2$ and by the results of Sec. II we can say that there are two-parameter families of m-order ($m > 2$) differential systems symmetrical under G.

(2) Let us take now as G the conformal group in $R \times R$. (Note that the space is now only one dimensional.)

The six generators of this group can be taken as

$$S_1 = \frac{\partial}{\partial t}; \quad S_2 = \frac{\partial}{\partial x}; \quad S_3 = t\frac{\partial}{\partial t} + x\frac{\partial}{\partial x},$$
$$S_4 = t\frac{\partial}{\partial x} + x\frac{\partial}{\partial t}; \quad S_5 = (t^2 + x^2)\frac{\partial}{\partial t} + 2tx\frac{\partial}{\partial x}, \tag{23}$$
$$S_6 = 2tx\frac{\partial}{\partial t} + (t^2 + x^2)\frac{\partial}{\partial x},$$

and therefore the third-order extension of them will be given by

$$S_1^{(3} = \frac{\partial}{\partial t}, \quad S_2^{(3} = \frac{\partial}{\partial x},$$
$$S_3^{(3} = S_3 - \ddot{x}\frac{\partial}{\partial \ddot{x}} - 2\dddot{x}\frac{\partial}{\partial \dddot{x}},$$
$$S_4^{(3} = S_4 + (1 - \dot{x}^2)\frac{\partial}{\partial \dot{x}} - 2\ddot{x}\frac{\partial}{\partial \ddot{x}} + (-4\dot{x}\dddot{x} - 3\ddot{x}^2)\frac{\partial}{\partial \dddot{x}},$$
$$S_5^{(3} = S_5 + (2x - 2x\dot{x}^2)\frac{\partial}{\partial \dot{x}} + (2\dot{x} - 2\dot{x}^3 - 2\ddot{x}t - 6x\dot{x}\ddot{x})\frac{\partial}{\partial \ddot{x}} + \{-6\ddot{x}(2\dot{x}^2 + x\ddot{x}) - 4\dddot{x}(t + 2x\dot{x})\}\frac{\partial}{\partial \dddot{x}}, \tag{24}$$
$$S_6^{(3} = S_6 + 2t(1 - \dot{x}^2)\frac{\partial}{\partial \dot{x}} + \{2(1 - \dot{x}^2) - 2\ddot{x}(x + 3t\dot{x})\}\frac{\partial}{\partial \ddot{x}} + \{-6\ddot{x}(2\dot{x} + t\ddot{x}) - 4\dddot{x}(x + 2t\dot{x})\}\frac{\partial}{\partial \dddot{x}}.$$

From (24) it follows immediately that

$$\dim \mathscr{D}^0 = 2; \quad \dim \mathscr{D}^1 = 3; \quad \dim \mathscr{D}^2 = 4; \quad \dim \mathscr{D}^3 = 5. \tag{25}$$

We see in (25) that condition (11) is not fulfilled for $m = 1$, $m = 2$, and $m = 3$. This implies, as we shall prove in Sec. IV, that there are no one-parameter families of differential equations of first-, second-, or third-order invariant under this group. Nevertheless, as we show now, there is one (and only one) third-order differential equation invariant under G.

Indeed, invariance under S_1 and S_2 implies that the third-order equation will have the form

$$\dddot{x} = f(\dot{x}, \ddot{x}). \tag{26}$$

Invariance under S_3 implies

$$\ddot{x}\frac{\partial f}{\partial \ddot{x}} = 2f, \tag{27}$$

that is,

$$f = a(\dot{x})\ddot{x}^2, \tag{28}$$

where $a(\dot{x})$ is an arbitrary function.

Invariance under S_4 implies

$$(1 - \dot{x}^2)\frac{da}{d\dot{x}} - 6\dot{x}a = -4\dot{x}a - 3, \tag{29}$$

and therefore,

$$a(\dot{x}) = (3\dot{x} - b)/(\dot{x}^2 - 1). \tag{30}$$

Finally invariance under S_5 implies $b = 0$. The resulting third-order differential equation is automatically invariant under S_6.

Therefore we have obtained the differential equation

$$\dddot{x} = \frac{3\dot{x}\ddot{x}^2}{\dot{x}^2 - 1}, \tag{31}$$

which is the only one invariant under the conformal group in $R \times R$.

Note that since $\mathscr{D}^1$ does not act transitively on the whole $(t,x,\dot{x})$ space it is possible to have also first-order differential equations invariant under G. This is precisely what happens with the two differential equations

$$\dot{x} = 1, \quad \dot{x} = -1, \tag{32}$$

which are the only ones (of first order) invariant under G.

Observe that the set $\{(t,x,\dot{x}) | 1 - \dot{x}^2 = 0\}$ defines the singular points of $\mathscr{D}^1$, that is, the points where $\mathscr{D}^1$ has (in this case) dimension 2.

Although $\mathscr{D}^2$ also has singular points, where its dimension does not attain the maximum value, it is easy to check by direct computation that there do not exist second-order equations invariant under the conformal group in $R \times R$. (The generators $S_1^{(2}, S_2^{(2}, S_3^{(2}, S_4^{(2}$ imply $\ddot{x} = 0$, but this equation is incompatible with the two generators $S_5^{(2}$ and $S_6^{(2}$.)

IV. *n*-PARAMETER FAMILIES OF EQUATIONS INVARIANT UNDER *G*

The above examples show that condition (11) is, in general, not necessary for the existence of isolated systems of order $\underline{m}$ invariant under G. We prove here that (11) is necessary and sufficient for the existence of n-parameter families of equations, each of them being invariant under G.

In fact, the necessary and sufficient conditions in order that the n-parameter family defined by

$$\mathbf{E}(t,\bar{x},\dot{\bar{x}},\ldots,\bar{x}^{(m)}) = \bar{c}, \quad \det\left(\frac{\partial \bar{E}}{\partial \bar{x}^{(m)}}\right) \neq 0, \tag{33}$$

be invariant under G are

$$S_i^{(m}(\mathbf{E})_{|\mathbf{E}=\mathbf{C}} = 0, \quad i = 1,\ldots,d_m. \tag{34}$$

Since Eqs. (34) hold as identities in $\mathbf{C} \in R^n$ we must have

$$S_i^{(m}(\mathbf{E}) = 0. \tag{35}$$

Let us see that (35) and

$$\dim(\mathscr{D}^{m-1}) < \dim(\mathscr{D}^m) \tag{36}$$

are contradictory.

In fact, as explained in Sec. II, if $\mathbf{S}_1^{(m},...,\mathbf{S}_{d_m}^{(m}$ is a basis of $\mathscr{D}^m$ then $\mathbf{S}_1^{(m-1},...,\mathbf{S}_{d_m}^{(m-1}$ is a basis of $\mathscr{D}^{m-1}$. Therefore if (36) holds we must have

$$\sum_{i=1}^{d_m} c_i(t,\mathbf{a},...,\mathbf{a}^{(m-1)})\mathbf{S}_i^{(m-1} = \mathbf{0}, \tag{37}$$

where *not* all of the c_i are equal to zero.

Therefore

$$\sum_{i=1}^{d_m} c_i \mathbf{S}_i^{(m}(\mathbf{E}) \overset{(37)}{=} \sum_{i=1}^{d_m} c_i \sum_{j=1}^{n} \psi_{ij}^{(m} \frac{\partial(\mathbf{E})}{\partial x_j^{(m}} \overset{(35)}{=} \mathbf{0}, \tag{38}$$

contradicting the hypothesis of $\det(\partial \mathbf{E}/\partial \mathbf{x}^{(m}) \neq 0$ imposed in (33). Note that this hypothesis concerning the determinant is essential in order to be able to apply the implicit function theorem to the variables $x_1^{(m},...,x_n^{(m}$ and put Eqs. (33) in the normal form (1). To conclude, we give an example of a one-parameter family of second-order differential equations invariant under the Poincaré group in $R \times R$.

In this case, the generators of G can be taken as

$$\mathbf{S}_1 = \frac{\partial}{\partial t}, \quad \mathbf{S}_2 = \frac{\partial}{\partial x}, \quad \mathbf{S}_3 = x\frac{\partial}{\partial t} + t\frac{\partial}{\partial x}. \tag{39}$$

The first and second extensions are given by

$$\mathbf{S}_1^{(1} = \frac{\partial}{\partial t}; \quad \mathbf{S}_2^{(1} = \frac{\partial}{\partial x}; \quad \mathbf{S}_3^{(1} = \mathbf{S}_3 + (1-\dot{x}^2)\frac{\partial}{\partial \dot{x}}, \tag{40}$$

and

$$\mathbf{S}_1^{(2} = \frac{\partial}{\partial t}; \quad \mathbf{S}_2^{(2} = \frac{\partial}{\partial x}; \quad \mathbf{S}_3^{(2} = \mathbf{S}_3^{(1} + (-3\dot{x}\ddot{x})\frac{\partial}{\partial \ddot{x}}. \tag{41}$$

One can immediately see that

$$\dim \mathscr{D}^0 = 2, \quad \dim \mathscr{D}^1 = 3. \tag{42}$$

Since the group has three parameters it is clear that $\dim \mathscr{D}^k = 3$ for any $k \geqslant 1$. Therefore there are one-parameter families of differential equations of order m (for any $m \geqslant 2$) invariant under this group. Taking for simplicity $m = 2$, the symmetry of the equation

$$\ddot{x} = f(t,x,x) \tag{43}$$

under $\mathbf{S}_1$ and $\mathbf{S}_2$ implies

$$\frac{\partial f}{\partial t} = 0; \quad \frac{\partial f}{\partial x} = 0, \tag{44}$$

that is, $f(t,x,x) = g(\dot{x})$. The symmetry under the boosts S_3 implies

$$-3\dot{x}g = \frac{dg}{d\dot{x}}(1-\dot{x}^2) \tag{45}$$

which leads, after integration, to the one-parameter family

$$\ddot{x} = c(1-\dot{x}^2)^{3/2}. \tag{46}$$

[1]F. González-Gascón and A. González-López, J. Math. Phys. **24**, 2006 (1983).

[2]P. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).

[3]F. Warner, *Foundations of Differential Geometry and Lie Groups* (Freeman, San Francisco, 1971).

[4]A. González-López, Ph.D. thesis, Universidad Complutense, Madrid, 1984 (unpublished).

[5]F. González-Gascón and A. González-López, Hadronic J. **6**, 841 (1983).

Integrable forms of the one-dimensional flow equation for unsaturated heterogeneous porous media

P. Broadbridge[a)]
CSIRO Division of Environmental Mechanics, GPO Box 821, Canberra, ACT, 2601, Australia

(Received 25 September 1986; accepted for publication 28 October 1987)

The equation for the horizontal transport of a liquid in an unsaturated scale-heterogeneous porous medium is $\partial\theta/\partial t = \lambda(x)\partial/\partial x[C(\theta)\partial\theta/\partial x] - \lambda'(x)E(\theta)\partial\theta/\partial x - \lambda''(x)\int(C+E)d\theta$. A systematic search for Lie–Bäcklund symmetries leads to the requirement that $C = a(b-\theta)^{-2}$, as in the homogeneous ($\lambda = 1$) case. More generally, (λ, E) may be $((1+mx)^\alpha, (1/\alpha - \frac{3}{2})C)$ or $(\exp(mx), -3C/2)$. In these cases the transport equation may be linearized and solved exactly. Examples of more complicated heterogeneous extensions are presented for the integrable nonlinear diffusion equations and for Burgers' equation.

I. INTRODUCTION

The general form of the nonlinear equation for the one-dimensional horizontal transport of liquid in an unsaturated heterogeneous porous medium is

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[F(\theta,x)\frac{\partial\theta(x,t)}{\partial x} + G(\theta,x)\right], \tag{1}$$

where t and x are time and space coordinates, θ is the volumetric liquid content, and F and G are differentiable functions of two variables. From the general results on Pfaff's problem,[1] there exists a (nonunique) potential function $\Psi(\theta,x)$ and an integrating factor $K(\theta,x)$ such that

$$[F,G] = -K\left[\frac{\partial\Psi}{\partial\theta}, \frac{\partial\Psi}{\partial x}\right].$$

In the context of porous media, Ψ is the (negative) potential energy per unit weight[2,3] of liquid due to capillarity and other interactions between the liquid and the solid medium, and K is the hydraulic conductivity or permeability of the medium for a specific liquid. Thus Eq. (1) is the equation of continuity $\partial\theta/\partial t + \partial v/\partial x = 0$, with the flux v satisfying a generalization of Darcy's law:

$$v = -K(\theta,x)\frac{\partial\Psi(\theta(x,t),x)}{\partial x}. \tag{2}$$

In heterogeneous media, K and Ψ depend explicitly on both the moisture content θ and the position x.

The complicated nonlinear equation [Eq. (1)] is usually simplified by imposing reasonable restrictions on the functions F and G. One commonly used simplification is the assumption that soils at any two different locations have geometrically similar internal structure.[4,5] From scaling analysis[4,6] it then follows that

$$K(\theta,x) = K_*(\theta)[\lambda(x)]^2, \tag{3a}$$

$$\Psi(\theta,x) = \Psi_*(\theta)/\lambda(x), \tag{3b}$$

where $K_*(\theta)$ and $\Psi_*(\theta)$ are the conductivity and potential functions at the surface $x = 0$, and $\lambda(x)$ is the geometrical scaling factor with $\lambda(0) = 1$. Here, we are concerned mainly with *scale-heterogeneous* media,[4] with $\lambda(x)$ twice differentiable. The flow equation [Eq. (1)] then reduces to

$$\frac{\partial\theta}{\partial t} = \lambda(x)\frac{\partial}{\partial x}\left[C(\theta)\frac{\partial\theta}{\partial x}\right] - \lambda'(x)E(\theta)\frac{\partial\theta}{\partial x} - \lambda''(x)\left[\int(C(\theta)+E(\theta))d\theta\right], \tag{4a}$$

where

$$C(\theta) = K_*\frac{d\Psi_*}{d\theta}, \tag{4b}$$

$$E(\theta) = \Psi_*\frac{dK_*}{d\theta}, \tag{4c}$$

and

$$\int(C+E)d\theta = K_*\Psi_*. \tag{4d}$$

Philip[4] found that in the particular case $K_* \propto \Psi_*^{-2}$ and $\log(-\Psi_*) \propto \theta$, Eq. (4) could be transformed to a nonlinear diffusion equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial P}{\partial x}\right], \tag{5}$$

with D an exponential function of P and P a logarithmic function of Ψ. In the physically relevant case of constant potential boundary condition and uniform potential initial condition,

$$t = 0, \quad x \geq 0, \quad \Psi = \Psi_i \tag{6a}$$

and

$$t > 0, \quad x = 0, \quad \Psi = \Psi_0, \tag{6b}$$

nonlinear diffusion equations may be transformed to a single ordinary differential equation (ODE) in which the independent variable is the Boltzmann similarity variable $\varphi = xt^{-1/2}$. Although in this case the ODE must be integrated numerically, we still obtain interesting exact relationships[4] such as the flux at $x = 0$ being proportional to $t^{-1/2}$, independent of λ.

Here we determine the class of model porous media (K_*, Ψ_*, λ) for which Eq. (4) is integrable and for which exact solutions may be found. In Sec. II we assume that the

[a)] Present address: Department of Mathematics, La Trobe University, Bundoora, Victoria, Australia 3083.

flow equation [Eq. (4)] possesses a Lie–Bäcklund symmetry group. This leads to the restriction, as in the case of homogeneous media,[7] that

$$C = a(b-\theta)^{-2}, \quad \text{with } a \text{ and } b \text{ fixed.} \tag{7}$$

However, more generally, λ may be either any power $(1+mx)^{\alpha}$ or any exponential $\exp(mx)$ with m and α fixed. In all such cases, $E(\theta)$ must be a scalar multiple of $C(\theta)$.

In Sec. III we transform each of the integrable flow equations of Sec. II to the exactly solvable nonlinear Fokker–Planck equation

$$\frac{\partial \rho}{\partial \tau} = \frac{\partial}{\partial y}\left|\rho^{-2}\frac{\partial \rho}{\partial y}\right| - \xi \rho^{-2}\frac{\partial \rho}{\partial y}, \tag{8}$$

with ρ a function of Ψ, τ a multiple of t, ξ constant, and y a function of x. The efficacy of Philip's approach depends on the transformed equation [Eq. (5)] being a conservation equation with the dependent variable P being a function of Ψ alone. The same applies to the dependent variable ρ of Eq. (8). Given the initial and boundary conditions (6), Eq. (8) may be solved by the quasianalytic method of Philip,[8] which in this case requires no more than the solution of a sequence of linear ordinary differential equations.

In Sec. IV we consider more general types of heterogeneity and present some examples of exactly solvable heterogeneous deformations of Fujita's nonliner diffusion equation and of Burgers' nonlinear convection–diffusion equation.

II. DETERMINATION OF THE CLASS OF ADMISSIBLE SCALE-HETEROGENEOUS MEDIA

The nonlinear diffusion term in (4) is commonly simplified by the Kirchhoff transformation

$$u = \int C(\theta)d\theta + \text{const.} \tag{9}$$

Then, Eq. (4) becomes

$$\frac{\partial u}{\partial t} = \lambda(x) C \frac{\partial^2 u}{\partial x^2} - \lambda'(x) E \frac{\partial u}{\partial x} - \lambda''(x) C \int \left(1+\frac{E}{C}\right) du. \tag{10}$$

In this section C and E are treated as functions of u; $C = C(\theta(u))$ and $E = E(\theta(u))$, unless other arguments are shown explicitly.

We now assume that Eq. (10) possesses a one parameter Lie–Bäcklund symmetry group,

$$\zeta^{(s)}:\ u \to u_*^{(s)} = \varphi(s;t,x,u_0,u_1,\ldots,u_j,\ldots),$$

where $u_j = [\partial/\partial x]^j u(x,t)$, and such that $\zeta^{(0)}$ is the identity map and $\zeta^{(s)}\zeta^{(u)} = \zeta^{(s+u)}$. Furthermore, we assume that in infinitesimal form,

$$u_* = u + sL(t,x,u_0,u_1,u_2,u_3) + O(s^2). \tag{11}$$

The landmark work of Anderson and Ibragimov[9] recognized that some generality is lost in restricting the generating function L to depend only on derivatives u_j up to some finite order $j=n$. However, for finite n, a direct method exists for determining the full algebra of symmetry generators. Furthermore, in the analysis of Eq. (1) it seems reasonable to consider $n=3$, since for homogeneous media an extension to $n>4$ does not yield any new nonlinear diffusion equations[7] or nonlinear diffusion–convection equations.[10]

Since (11) is presumed to be an infinitesimal symmetry, for every solution $u(x,t)$ of (10),

$$D_t u_* - \lambda(x)C(u_*)(D_x)^2 u_* + \lambda'(x)E(u_*)D_x u_* + \lambda''(x)C(u_*)\int\left(1+\frac{E(u_*)}{C(u_*)}\right)du_* = O(s^2), \tag{12}$$

where D_t and D_x are, respectively, the total t derivative and total x derivative operating on functions $f(t,x,u_0,u_1,\ldots,u_j,\ldots)$ on an infinite-dimensional manifold:

$$D_x f = \frac{\partial f}{\partial x} + \sum_{j=0}^{\infty}\frac{\partial f}{\partial u_j}u_{j+1}, \quad D_t f = \frac{\partial f}{\partial t} + \sum_{j=0}^{\infty}\frac{\partial f}{\partial u_j}u_{jt}.$$

Since u is presumed to be a solution to (10), u_{jt} is taken to be

$$\left(\frac{\partial}{\partial x}\right)^j \Big[\lambda(x)Cu_2 - \lambda'(x)Eu_1 - \lambda''(x)C\int\left(1+\frac{E}{C}\right)du\Big\}.$$

The u_5 terms in (12) are immediately of order s^2, and there remains a polynomial equation in u_4 whose coefficients depend on t, x, and u_j; $j<4$. This and subsequent polynomials have been manipulated using the algebraic software package REDUCE.[11]

Setting u_4^2 terms in (12) to zero [plus $O(s^2)$], we obtain

$$\frac{\partial^2 L}{\partial u_3^2} = 0.$$

Progressively balancing u_4u_3, u_4u_2, and u_4u_1 terms, we obtain

$$0 = \frac{\partial^2 L}{\partial u_3\,\partial u_2} = \frac{\partial^2 L}{\partial u_3\,\partial u_1}$$

and

$$0 = \frac{-2\,\partial^2 L}{\partial u_3\,\partial u}C + 3\frac{dC}{du}\frac{\partial L}{\partial u_3} = 0,$$

implying

$$L = u_3\, g(t,x)C^{3/2} + H(t,x,u,u_1,u_2), \tag{13}$$

for some functions g and H. Setting the remaining u_4 terms to zero, we deduce that $g = \lambda^{3/2}P(t)$ for some function P. We now substitute (13) in (12) and consider (12) as a polynomial equation in u_3. Progressively balancing u_3^2 and u_3u_2 terms, we obtain $\partial^2 H/\partial u_2^2 = 0$ and

$$\frac{\partial^2 H}{\partial u_2\,\partial u_1} = \frac{3}{2}\frac{dC}{du}C^{1/2}P\lambda^{3/2}.$$

Therefore,

$$H = u_2\left(\frac{3}{2}\frac{dC}{du}C^{1/2}P\lambda^{3/2}u_1 + J(t,x,u)\right) + Q(t,x,u,u_1),$$

for some functions J and Q. Now the balance of $u_3u_1^2$ terms in (12) leads to the requirement

$$\frac{2(d^2C/du^2)}{dC/du} = \frac{dC/du}{C},$$

implying $C = \sigma(u+\nu)^2$, with σ and ν constant. By inverting (9), this is equivalent to

$$C = a/(b-\theta)^2 \quad (a,b \text{ constant}). \tag{14}$$

Naturally this form of diffusivity is demanded in the special case of homogeneous ($\lambda = 1$) nonlinear diffusion, as demonstrated by Bluman and Kumei.[7] Previously, exact solutions to the equation

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[\frac{a}{(b-\theta)^2}\frac{\partial\theta}{\partial x}\right]$$

had been developed by Fujita,[12] Knight and Philip,[13] and Rosen.[14] Now, it is known[15] that Lie–Bäcklund symmetries exist also for a scale-heterogeneous porous medium equation [Eq. (4)] with $\lambda = 1 + mx$, and it remains to determine the full class of admissible scale functions $\lambda(x)$.

In the following analysis, we assume that $C = u^2$, since $\sigma = 1$ after a suitable rescaling of the t coordinate and $\nu = 0$ after a suitable choice of the constant of integration in (9). Setting u_3u_1 terms in (12) to zero, we obtain

$$-3P\lambda'\lambda^{1/2}\frac{dE}{du}C^{3/2} - 2\lambda C\frac{\partial J}{\partial u} + 2\frac{dC}{du}J = 0,$$

implying

$$J = -\frac{3}{2}\lambda'\lambda^{1/2}Pu^2\int\frac{1}{u}\frac{dE}{du}du + N(x,t)u^2,$$

for some function N. Then, by balancing the remaining u_3 terms in (12), we obtain

$$0 = -4\lambda^{3/2}\frac{\partial N}{\partial x} + 4\lambda^{1/2}\lambda' N + 2\lambda^2u^{-1}\left(\frac{dP}{dt}\right)$$
$$-3P\lambda''\lambda^2u - \frac{3}{2}P(\lambda')^2\lambda u - 6P\lambda''\lambda^2\frac{E}{u}$$
$$+3P(\lambda')^2\lambda\frac{E}{u} + 6P\lambda''\lambda^2\int\frac{1}{u}\frac{dE}{du}du$$
$$-6P\lambda''\lambda^2\int\frac{E}{u^2}du - 3P(\lambda')^2\lambda\int\frac{1}{u}\frac{dE}{du}du.$$

A balance of u-independent terms in the above leads to the requirement that

$$N = N_0(t)\lambda(x),$$

for some function N_0. A balance of u-dependent terms implies

$$E(u) = -\frac{2}{3}\frac{1}{P}\left(\frac{dP}{dt}\right)\frac{\lambda}{(\lambda')^2} - u^2\left(\frac{\lambda''\lambda}{(\lambda')^2} + \frac{1}{2}\right). \quad (15)$$

Since E is a function of u alone, it follows that

$$\lambda''\lambda/(\lambda')^2 = \alpha_1 \quad \text{(constant)}. \quad (16)$$

Therefore λ is restricted to the form

$$\lambda(x) = (1+mx)^\alpha$$
$$[m \text{ constant and } \alpha = (1-\alpha_1)^{-1}], \quad (17a)$$

or

$$\lambda(x) = \exp(mx) \quad (\text{in the case } \alpha_1 = 1). \quad (17b)$$

Further restrictions are not warranted since an appropriate form of Eq. (1) may be integrated whenever λ satisfies (17a) or (17b).

III. SOLUTION OF ADMISSIBLE SCALE-HETEROGENEOUS FLOW EQUATIONS

A. The case of the exponential scale factor

First we consider the special case $(\lambda,C,E) = (e^{mx}, a(b-\theta)^{-2}, -3/2a(b-\theta)^{-2})$ corresponding to $\alpha_1 = 1$. From the definitions (4b) and (4c), it then follows that

$$\frac{d\ln K_*}{d\ln|\Psi_*|} = -\frac{3}{2},$$

so that $K_* \propto |\Psi_*|^{-3/2}$. Now (4b), (14), and (18) imply

$$-|\Psi_*|^{-3/2}d|\Psi_*| \propto (b-\theta)^{-2}d\theta. \quad (18)$$

Accordingly, we assume

$$K_*(\theta) = \sigma(b-\theta)^{-3} \quad (19)$$

and

$$\Psi_*(\theta) = -\tfrac{1}{2}(a/\sigma)(b-\theta)^2, \quad \text{for some constant } \sigma. \quad (20)$$

Relation (20) could represent a real soil only over a limited range of θ. Following (19) and (20), the flow equation [Eq. (10)] becomes

$$\frac{\partial\theta}{\partial t} = e^{mx}\frac{\partial}{\partial x}\left[a(b-\theta)^{-2}\frac{\partial\theta}{\partial x}\right]$$
$$+\frac{3}{2}me^{mx}a(b-\theta)^{-2}\frac{\partial\theta}{\partial x}$$
$$+\frac{1}{2}am^2e^{mx}(b-\theta)^{-1}. \quad (21)$$

Now, let

$$\rho = -[(2\sigma)^{1/2}/a]|\Psi|^{1/2} \quad (22a)$$
$$= a^{-1/2}e^{-(1/2)mx}(\theta - b). \quad (22b)$$

Then (21) transforms to

$$\frac{\partial\rho}{\partial t} = \frac{\partial\rho}{\partial x}\left[\rho^{-2}\frac{\partial\rho}{\partial x}\right] + \frac{1}{2}m\rho^{-2}\frac{\partial\rho}{\partial x}. \quad (23)$$

Equation (23) is known to possess an infinite hierarchy of Lie–Bäckund symmetries.[10] In fact, it may be transformed to the linear diffusion equation, as shown here in the Appendix. Some exact solutions of Eq. (23) have already been applied to the flushing of oil reservoirs,[10,16,17] to the transport of a solute subject to adsorption,[18] and to rainfall infiltration in unsaturated soil.[19] However, these solutions, when transformed back to a concentration field $\theta(x,t)$ via (22b), do not satisfy boundary and initial conditions which are particularly relevant to heterogenous porous media. Nevertheless, the relevant initial and boundary conditions [Eq. (6)] are expressed easily in terms of ρ:

$$t=0,\quad x\geq 0,\quad \rho=\rho_i = -[(2\sigma)^{1/2}/a]|\Psi_i|^{1/2} \quad (24a)$$

and

$$t>0,\quad x=0,\quad \rho=\rho_0 = -[(2\sigma)^{1/2}/a]|\Psi_0|^{1/2}. \quad (24b)$$

Nonlinear Fokker–Planck equations such as (23), subject to (24), have previously been used to model unsaturated flow in homogenous media. In the latter application, the convective term would be due to gravity[20] rather than heterogeneity. Following the quasianalytic method of Philip,[8] we first reexpress Eq. (23) in the form

$$-\frac{\partial}{\partial t}\int_{\pi_i}^{\rho} x\,d\rho=\frac{\rho^{-2}}{\partial x/\partial\rho}+\frac{1}{2}m(\rho^{-1}-\rho_i^{-1}). \qquad (25)$$

Then, we seek a small-t solution in the form

$$x(\rho,t)=\sum_{j=1}^{\infty}\varphi_j(\rho)t^{j/2}. \qquad (26)$$

The boundary condition (24b) transforms to $\varphi_j(\rho_0)=0$ for all j and the initial condition (24a) is then automatically satisfied. After substituting (26) in (25), the balance of $t^{-1/2}$ terms implies

$$\int_{\rho_i}^{\rho}\varphi_1\,d\rho=\frac{-2\rho^{-2}}{\varphi_1'(\rho)}. \qquad (27)$$

In the absence of the nonlinear convective term in (23), the series (26) would terminate at $j=1$ and the solution of (23) and (24) would be equivalent to the solution of the nonlinear diffusion problem,

$$\frac{\partial\rho}{\partial t}=\frac{\partial}{\partial x}\left[D(\rho)\frac{\partial\rho}{\partial x}\right],\quad\text{with } D=\rho^{-2}, \qquad (28)$$

and subject to (24). This is the nonlinear diffusion problem that (21) and (22) reduce to in the case of a homogenous medium ($m=0$). Unlike the case of general diffusivity,[8] when $D(\rho)=\rho^{-2}$, this problem may be solved exactly in parametric form by transforming (28) to a single linear ordinary differential equation.[12,15] The subsequent balance of $t^{k/2}$ terms ($k=0,1,2,\ldots$) in (25) leads to a sequence of inhomogeneous ordinary intergrodifferential equations,[8] the first two of which are

$$\int_{\rho_i}^{\rho}\varphi_2\,d\rho=\rho^{-2}[\varphi_1'(\rho)]^{-2}\varphi_2'(\rho)-\tfrac{1}{2}m(\rho^{-1}-\rho_i^{-1})$$

and

$$\int_{\rho_i}^{\rho}\varphi_3\,d\rho=\tfrac{2}{3}\rho^{-2}[\varphi_1'(\rho)]^{-2}\varphi_3'(\rho)-\tfrac{2}{3}\rho^{-2}[\varphi_2'(\rho)]^{2}[\varphi_1'(\rho)]^{-3}.$$

Integrating each side of the conservation equation [Eq. (23)] from $x=0$ to $x=\infty$, we obtain

$$\frac{\partial}{\partial t}\int_0^{\infty}(\rho-\rho_i)dx=w_0-w_\infty, \qquad (29)$$

where w_0 and w_∞ are, respectively, the values of $\frac{1}{2}m\rho^{-1}-\rho^{-2}(\partial\rho/\partial x)$ at $x=0$ and at $x=\infty$. From (24a), $w_\infty=-\frac{1}{2}m\rho_i^{-1}$. Now,

$$\begin{aligned}w_0&=-\frac{1}{2}m\rho_0^{-1}-\rho_0^{-2}\frac{\partial\rho}{\partial x}\bigg|_{x=0}\\&=-\frac{1}{2}m\rho_0^{-1}-a^{-1/2}K\frac{\partial\Psi}{\partial x}\bigg|_{x=0}\\&\quad\text{[by (19) and (22)]}\\&=-\frac{1}{2}m\rho_0^{-1}+a^{-1/2}v_0,\end{aligned}$$

where v_0 is the physical flux at $x=0$. We may carry out the integration in (29) by parts and assume (26) to obtain

$$v_0=\tfrac{1}{2}St^{-1/2}+\tfrac{1}{2}a^{1/2}m[\rho_0^{-1}-\rho_i^{-1}]+\sum_{j=2}^{\infty}A_jt^{(1/2)j-1}, \qquad (30)$$

with

$$A_j=\frac{j}{2}a^{1/2}\int_{\rho_i}^{\rho_0}\varphi_j(\rho)d\rho.$$

The sorptivity S is the same as that for the homogeneous ($m=0$) soil[21]

$$S=a^{1/2}\int_{\rho_i}^{\rho_0}\varphi_1\,d\rho. \qquad (31)$$

However, unlike the model [Eqs. (5) and (6)] of Philip,[4] this model predicts that v_0 is not proportional to $t^{-1/2}$, since the order $t^{(1/2)j-1}$ corrections are nontrivial and proportional to the $(j-1)$th power of the strength m of the heterogeneity.

The radius of convergence of the power series in (30) has not yet been established. However, by analogy with the gravitational time scale t_{grav} of gravity-assisted infiltration,[3] for heterogeneity-affected flow, practical convergence is expected until t is of the order of the heterogeneity time scale:

$$t_{\text{het}}=S^2a^{-1}m^{-2}(\rho_i^{-1}-\rho_0^{-1})^{-2}. \qquad (32)$$

The convective term in (23) is of the form $-(dH/d\rho)(\partial\rho/\partial z)$, where $H(\rho)=\frac{1}{2}m\rho^{-1}$. In the case of a porous medium whose texture becomes finer with increasing depth, $m<0$ and $d^2H/d\rho^2>0$ in the domain $\rho_i\le\rho\le\rho_0<0$. It then follows from a general result of Philip[22] that, at large t, $\rho(x,t)$ approaches a traveling wave solution $\rho=g(x-Ut)$, with speed

$$U=\tfrac{1}{2}m(\rho_0^{-1}-\rho_i^{-1})/(\rho_0-\rho_i)=\tfrac{1}{2}m/(\rho_0\rho_i). \qquad (33)$$

In the particular case of Eq. (23), the function g is known exactly.[18,19] Although a traveling potential wave develops in this model, there will not be a large-t asymptotic traveling concentration wave, since in heterogeneous media θ is not a function of Ψ alone.

B. The case of the power law scale factor

We now consider the class of models in which $\lambda=(1+mx)^{\alpha}$ and $C=a(b-\theta)^{-2}$. We assume that $dP/dt=0$, so that (15) reduces to $E=(\alpha^{-1}-\frac{3}{2})C$. Corresponding to (18), we now have

$$K_*\propto|\Psi_*|^{1/\alpha-3/2}. \qquad (34)$$

Now, (34), (4b), and (14) imply

$$-|\Psi_*|^{1/\alpha-3/2}d|\Psi_*|\propto(b-\theta)^{-2}d\theta.$$

We assume

$$K_*=\sigma(b-\theta)^{-\alpha_2} \qquad (35)$$

and

$$\Psi_*=[a/\sigma(1-\alpha_2)](b-\theta)^{\alpha_2-1}, \qquad (36)$$

with σ constant and

$$\alpha_2=(2-3\alpha)/(2-\alpha). \qquad (37)$$

Equation (36), like Eq. (20), could represent a real soil only over a limited range of θ. In applications to soil physics, we require $\alpha > 2$ or $\alpha < 0$, since the physics demands that Ψ_* must be negative and that $dK_*/d\theta$ must be positive as Ψ_* approaches zero.[3] One exception is provided by the case $\alpha = 1$, in which the absence of the final term in (4) ultimately allows more freedom in the function $K_*(\theta)$ than that indicated in (35).[15]

The flow equation [Eq. (10)] now takes the form

$$\frac{\partial\theta}{\partial t} = (1+mx)^\alpha \frac{\partial}{\partial x}\left[a(b-\theta)^{-2}\frac{\partial\theta}{\partial x}\right] + \alpha m a(b-\theta)^{-2}(1+mx)^{\alpha-1}\left[\frac{3}{2}-\frac{1}{\alpha}\right]\frac{\partial\theta}{\partial x} + \alpha(\alpha-1)m^2 a(b-\theta)^{-1}\left[\frac{1}{2}-\frac{1}{\alpha}\right](1+mx)^{\alpha-2}. \tag{38}$$

We define a new dependent variable,

$$\rho = -a^{-1/2}\left[\frac{\sigma}{a}\frac{2\alpha}{\alpha-2}\right]^{1/2-1/\alpha}|\Psi|^{1/2-1/\alpha} \tag{39a}$$

$$= a^{-1/2}(\theta-b)(1+mx)^{1-\alpha/2}. \tag{39b}$$

Equation (38) becomes

$$\frac{\partial\rho}{\partial t} = (1+mx)^{1+\alpha/2}\frac{\partial}{\partial x}\left[(1+mx)^{1-\alpha/2}\rho^{-2}\frac{\partial\rho}{\partial x}\right] + \alpha m(1+mx)\rho^{-2}\frac{\partial\rho}{\partial x}. \tag{40}$$

Now let

$$y = m^{-1}\ln(1+mx). \tag{41}$$

Equation (40) then reduces to

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial y}\left[\rho^{-2}\frac{\partial\rho}{\partial y}\right] + \frac{1}{2}\alpha m\rho^{-2}\frac{\partial\rho}{\partial y}. \tag{42}$$

All of the previously mentioned techniques for solving Eq. (23) also apply to Eq. (42). The initial and boundary conditions [Eq. (6)] transform to

$$t=0,\quad y\geq 0,$$
$$\rho=\rho_i = -a^{-1/2} \times[(\sigma/a)(2\alpha/(\alpha-2))]^{1/2-1/\alpha}|\Psi_i|^{1/2-1/\alpha}$$

and

$$t>0,\quad y=0,$$
$$\rho=\rho_0 = -a^{-1/2} \times[(\sigma/a)(2\alpha/(\alpha-2))]^{1/2-1/\alpha}|\Psi_0|^{1/2-1/\alpha}.$$

A small-t solution $y(\rho,t)$ may be constructed as in (26). For a porous medium whose texture becomes finer with increasing depth, $\alpha m (= (d\lambda/dx)_{x=0}) < 0$, and there will develop an asymptotic large-t traveling potential wave $\Psi = G(y-Ut)$, with $U = \alpha m/\rho_0\rho_i$. From (41), this implies that at large t and for fixed θ, $x(\theta,t)$ increases exponentially in time. The above analysis, which ignores the effect of gravity, can apply to vertical infiltration[3] only up to times of the order of $t_{\text{grav}} = S^2[K_*(\theta_0) - K_*(\theta_i)]^{-2}$, where θ_i and θ_0 are initial and final volumetric moisture contents. The large-t heterogeneity-driven traveling wave or exponentially accelerated profiles could begin to develop without being significantly modified by gravity only if the heterogeneity time scale were significantly less than the gravitational time scale.

IV. MORE COMPLICATED FORMS OF HETEROGENEITY

Equation (1) is much more general than Eq. (4a), since it admits heterogeneity that is more complicated than the scale heterogeneity represented by Eqs. (3a) and (3b). One approach to constructing integrable examples of a more general type is to find smooth deformations of known integrable homogeneous models. This possibility has by no means been fully explored, but some examples are given in this section.

A. Heterogeneous extensions of Fujita's equation

The integrable scale-heterogeneous models introduced in Sec. II may be viewed as smooth deformations of the integrable nonlinear diffusion equation [Eq. (28)] previously studied by Fujita[11] and others.[12-14,23] This same equation also has integrable heterogeneous relatives of a more exotic nature. For example, when

$$F = a[b-(mx+1)^\gamma\theta]^{-2}(mx+1)^2$$

and

$$G = a\gamma m[b-(mx+1)^\gamma\theta]^{-2}(mx+1)\theta,$$

Eq. (1) transforms to (A3) by taking

$$w = [(mx+1)^\gamma\theta - b](mx+1)^{\gamma-1}$$

and

$$y = (1-\gamma)^{-1}m^{-1}[(mx+1)^{1-\gamma}-1].$$

For $\gamma < 0$ and $mx > 0$, none of the above functions have singularities provided b is greater than the water content θ_s at saturation.

B. Heterogeneous Burgers' equation

One integrable heterogeneous deformation of Burgers' equation is

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[(mz+1)^2 D_*\frac{\partial\theta}{\partial z} - k(mz+1)^2(\theta-\theta_i)^2\right], \tag{43}$$

with D_*, m, and k fixed. For $m=0$, (43) is the usual Burgers equation, which has already been applied to field soils with distributed macropores.[24] After defining

$$Z = z + \frac{1}{m},$$

$$w = \frac{k}{2D_*}\left(z+\frac{1}{m}\right)(\theta-\theta_i),\quad \tau = tm^2 D_*,$$

Eq. (43) transforms to

$$\frac{\partial w}{\partial\tau} = Z^2\frac{\partial^2 w}{\partial Z^2} - 4Zw\frac{\partial w}{\partial Z}. \tag{44}$$

Now through the transformation

$$w = -\frac{1}{2}Z\frac{1}{u}\frac{\partial u}{\partial Z}, \tag{45}$$

(44) becomes

$$\left[u\frac{\partial}{\partial Z} - \frac{\partial u}{\partial Z}\right]\left[\frac{\partial u}{\partial\tau} - Z^2\frac{\partial^2 u}{\partial Z^2}\right] = 0. \tag{46}$$

The transformation (45) is a modification of that used by Forsyth,[1] Hopf,[25] and Cole[26] to solve the homogeneous

Burgers equation. From Eq. (46), it is sufficient that u satisfies the linear equation

$$\frac{\partial u}{\partial \tau} - Z^2 \frac{\partial^2 u}{\partial Z^2} = 0. \tag{47}$$

Equation (47) is amenable to solution by standard integral transform techniques.

Similar heterogeneous extensions of higher order equations of the Burgers hierarchy[27] have not yet been systematically investigated. However, from no-go theorems for homogeneous media,[28] an extension to higher spatial dimensions seem improbable.

ACKNOWLEDGMENTS

The author is grateful for the advice and encouragement of Dr. J. H. Knight, Dr. J. R. Philip, Dr. I. White, and Dr. C. Rogers.

APPENDIX: REDUCTION OF THE FOKAS-YORTSOS-ROSEN EQUATION TO THE LINEAR DIFFUSION EQUATION

The convective term may be removed from Eq. (23) by the transformation $(x,\rho) \to (y,w)$:

$$w = \rho e^{-mx/2}, \tag{A1}$$

$$y = (2/m)(e^{mx/2} - 1), \tag{A2}$$

so that

$$\frac{\partial w}{\partial t} = \frac{\partial}{\partial y}\left[D(w)\frac{\partial w}{\partial y}\right], \tag{A3}$$

with

$$D(w) = w^{-2}. \tag{A4}$$

The general nonlinear diffusion equation [Eq. (A3)] may be simplified by applying the Kirchhoff transformation,[3,13,29]

$$W = \int D(w)dw, \tag{A5}$$

so that

$$\frac{\partial W}{\partial t} = D(W)\frac{\partial^2 W}{\partial y^2}. \tag{A6}$$

In the case of $D(w)$ being given by (A4), we take

$$W = -1/w, \tag{A7}$$

so that

$$\frac{\partial W}{\partial t} = W^2 \frac{\partial^2 W}{\partial y^2}. \tag{A8}$$

Equation (A8) may be transformed to the linear diffusion equation

$$\frac{\partial Q}{\partial t} = \frac{\partial^2 Q}{\partial \chi^2}, \tag{A9}$$

using the correspondence of Vein,[23,30]

$$W = \frac{\partial Q}{\partial \chi}, \quad y = Q. \tag{A10}$$

Given any solution Eq. (A9), we may successively apply the inverses of transformations (A10), (A7), (A2), and (A1) to obtain a parametric solution $\chi \to (\chi,\rho)$ to Eq. (23).

[1]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959).

[2]Thus Ψ has the dimensions of length and may be viewed as the nongravitational component of a total hydraulic head, as in Ref. 3.

[3]J. R. Philip, Adv. Hydrosci. **5**, 215 (1969).

[4]J. R. Philip, Aust. J. Soil Res. **5**, 1 (1967).

[5]A. W. Warrick, G. J. Mullen, and D. R. Nielsen, Water Resour. Res. **13**, 355 (1977).

[6]E. E. Miller and R. D. Miller, J. Appl. Phys. **27**, 324 (1956).

[7]G. Bluman and S. Kumei, J. Math. Phys. **21**, 1019 (1980).

[8]J. R. Philip, Soil Sci. **83**, 345 (1957).

[9]R. L. Anderson and N. Ibragimov, *Lie-Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).

[10]A. S. Fokas and Y. C. Yortsos, SIAM J. Appl. Math. **42**, 318 (1982).

[11]*REDUCE User's Manual*, edited by A. C. Hearn (Rand Corp., Santa Monica, CA, 1985).

[12]H. Fujita, Textile Res. J. **22**, 823 (1952).

[13]J. H. Knight and J. R. Philip, J. Eng. Math. **8**, 219 (1974).

[14]G. Rosen, Phys. Rev. B. **19**, 2398 (1979).

[15]P. Broadbridge, Transp. Porous Media **2**, 129 (1987).

[16]C. Rogers, M. P. Stallybrass, and D. L. Clements, Nonlinear Anal. Theory Meth. Appl. 7, 785 (1982).

[17]M. J. King, J. Math. Phys. **26**, 870 (1985).

[18]G. Rosen, Phys. Rev. Lett. **49**, 1844 (1982).

[19]P. Broadbridge and I. White, "Constant rate rainfall infiltration: A versatile nonlinear model I.," Water Resour. Res., in press.

[20]L. A. Richards, Physics **1**, 318 (1931).

[21]J. R. Philip, Soil Sci. **84**, 257 (1957).

[22]J. R. Philip, Soil Sci. **83**, 435 (1957).

[23]A. Munier, J. R. Burgan, J. Gutierrez, E. Fijalkow, and M. R. Feix, SIAM J. Appl. Math. **40**, 191 (1981).

[24]B. E. Clothier, J. H. Knight, and I. White, Soil Sci. **132**, 255 (1981).

[25]E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).

[26]J. D. Cole, Quart. Appl. Math. **9**, 225 (1951).

[27]C. Rogers and P. L. Sachdev, Nuovo Cimento B **83**, 127 (1984).

[28]P. Broadbridge, J. Phys. A, **19**, 1245 (1986).

[29]G. Kirchhoff, *Vorlesungen über die Theorie der Wärme* (Barth, Leipzig, 1894).

[30]W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1965), Vol. II.

A search for bilinear equations passing Hirota's three-soliton condition. IV. Complex bilinear equations

Jarmo Hietarinta
Department of Physical Sciences, University of Turku, 20500 Turku, Finland

(Received 5 August 1987; accepted for publication 23 September 1987)

In this paper the results of a search for complex bilinear equations with two-soliton solutions are presented. The following basic types are discussed: (a) the nonlinear Schrödinger equation $B(D_x,...)G\cdot F=0$, $A(D_x,D_t)F\cdot F=GG^*$, and (b) the Benjamin–Ono equation $P(D_x,...)F\cdot F^*=0$. It is found that the existence of two-soliton solutions is not automatic, but introduces conditions that are like the usual three- and four-soliton conditions. The search was limited by the degree of $A=2$, and by degree of $P\leqslant 4$. The main results are the following: (1) $(iaD_x^3+D_xD_t+iD_y+b)G\cdot F=0$, $D_x^2F\cdot F=GG^*$; (2) $(D_x^2+aD_y^2+iD_t+b)G\cdot F=0$, $D_xD_yF\cdot F=GG^*$; (3) $(iaD_x^3+D_x^2+iD_t)F\cdot F^*=0$; and (4) $(D_xD_t+i(aD_x+bD_t))F\cdot F^*=0$.

I. INTRODUCTION

This is the fourth in a series of papers devoted to searching for bilinear equations having three-soliton solutions (3SS's). We have previously discussed single bilinear equations[1] $P(D_x,D_t,...)F\cdot F=0$, and pairs of equations,[2,3] mainly of type $B(D_x,D_t)G\cdot F=0$, $A(D_x,D_t)(F\cdot F+G\cdot G)=0$, where A is quadratic and B either odd[2] or even.[3]

In this paper we will consider *complex* bilinear equations. The parameters in the previously mentioned systems could also be complex, but their complexity did not play any special role, since complex conjugates never entered. For the present systems complex conjugation is used explicitly. Since the complex parameter has two real degrees of freedom it turns out that we obtain analogs of the three- and four-soliton conditions (3SC, 4SC) already when we try to construct a 2SS.

We will now go through the types of equations that we will discuss in this paper.

A. Nonlinear Schrödinger equation

The most famous complex integrable system is of course the nonlinear Schrödinger equation (NLS)

$$iy_t+y_{xx}+|y|^2y=0. \tag{1}$$

With the dependent variable transformation $y=G/F$, F real, (1) is satisfied if[4]

$$(D_x^2+iD_t)G\cdot F=0, \tag{2a}$$

$$D_x^2F\cdot F=GG^*, \tag{2b}$$

where D_x and D_t are the usual Hirota derivatives. This is an example of the class of bilinear equations

$$B(D_x,D_t,...)G\cdot F=0, \tag{3a}$$

$$A(D_x,D_t,...)F\cdot F=C(D_x,D_t,...)G\cdot G^*, \tag{3b}$$

with the properties

$$[B(X,T,...)]^*=B(-X^*,-T^*,...), \tag{4}$$

$$A(0,0,...)=0. \tag{5}$$

Furthermore, A can be assumed to be even with real coefficients and the polynomial C should also satisfy (4). The overall coefficients of A, B, and C are unimportant.

The 1SS for (3) is given by

$$F=1+Ke^{n+n^*},\quad G=e^n, \tag{6}$$

where

$$n=px+\Omega t+\cdots+n_0, \tag{7}$$

$$K=\frac{C(p-p^*,\Omega-\Omega^*,...)}{2A(p+p^*,\Omega+\Omega^*,...)}, \tag{8}$$

and the dispersion relation between the (complex) parameters $p,\Omega,...$ is given by

$$B(p,\Omega,...)=0. \tag{9}$$

Note that a formally quadratic term appears in F already for the 1SS [cf. (11) and (20) in Ref. 3].

The NLS equation has been generalized in various directions. The result presented in Ref. 4 included a cubic term, the equations can be scaled to

$$iy_t+\beta[y_{xx}+|y|^2y]+i\gamma[y_{xxx}+3|y|^2y_x]=0, \tag{1'}$$

$$(i\gamma D_x^3+\beta D_x^2+iD_t)G\cdot F=0, \tag{2a'}$$

together with (2b). Here β and γ are real constants. The two-dimensional nonlinear Schrödinger equation (2DNLS) (or Benney–Roskes or Davey–Stewartson equation) is given by

$$\begin{aligned}&iu_t-\beta u_{xx}+\gamma u_{yy}+\delta|u|^2u-2uv=0,\\&\beta v_{xx}+\gamma v_{yy}-\beta\delta(|u|^2)_{xx}=0.\end{aligned} \tag{10}$$

After the change of dependent variables,

$$u=G/F,\quad v=2\beta(\ln F)_{xx},\quad F\text{ real}, \tag{11}$$

(10) goes over to[5,6]

$$\begin{aligned}&(iD_t-\beta D_x^2+\gamma D_y^2)G\cdot F=0,\\&(\beta D_x^2+\gamma D_y^2)F\cdot F=-\delta GG^*,\end{aligned} \tag{12}$$

which has N-soliton solutions of the same type as NLS. Redekopp's equations

$$iu_t+u_{xx}=uv,\quad v_t+(|u|^2)_x=0, \tag{13}$$

 0022-2488/88/030628-08$02.50

which are also related to Langmuir waves, yield with (11) ($\beta = -1$) a singular limit of (12),[7]

$$(iD_t + D_x^2)G \cdot F = 0, \quad D_x D_t F \cdot F = GG^*. \tag{14}$$

Tajiri[8] has considered the "coupled Higgs field" equation, which is as in (10) and (12) except that iu_t is replaced by $-cu$ in (10) and iD_t by $-c$ in (12). This system also has N-soliton solutions (NSS). Higher-dimensional generalizations of the above equations have NSS only if extraneous conditions are imposed on the parameters.[9,10]

As an equation with a bilinear form of the above type that does *not* have NSS's we have the Zakharov equations

$$iE_t + E_{xx} = nE, \quad n_{tt} - n_{xx} = (|E|^2)_{xx}, \tag{15}$$

which, after a transformation similar to (11), yield[11]

$$(iD_t - D_x^2 - b)G \cdot F = 0, \quad (D_t^2 + D_x^2)F \cdot F = GG^*. \tag{16}$$

However, this equation does not have 2SS of the standard type,[12] and other tests indicate also that it is probably not integrable.[13] (The other two systems discussed in Ref. 11 are also not integrable.)

When the bilinear equations (2), (12), and (16) are derived there is a possibility of an additional "decoupling constant," so that one can as well take

$$[B(D_x, D_t, \ldots) + d\,]G \cdot F = 0, \tag{17a}$$

$$[A(D_x, D_t, \ldots) + d\,]F \cdot F = GG^*, \tag{17b}$$

as the bilinear form. The (real) constant d is related to the boundary condition of the soliton and if it is nonzero then the 1SS is different from (6): one obtains the so called "envelope-hole soliton"[5,14] by

$$F = 1 + e^n, \quad G = g e^{i\theta}(1 + e^{n + iD}), \tag{18}$$

where n is as in (7), but with real parameters, and

$$\theta = kx + wt + \cdots, \quad d = g^2 = -B(ik, iw, \ldots), \tag{19}$$

where also the constants g, k, and w are real. For D we obtain from (17a),

$$e^{iD} = \frac{B(-p + ik, -\Omega + iw, \ldots) + b}{B(p + ik, \Omega + iw, \ldots) + d}, \tag{20}$$

and from (17b)

$$\cos D = 1 + A(p, \Omega, \ldots)/d. \tag{21}$$

When D is eliminated from these we obtain a rather complicated dispersion relation for the parameters $p, \Omega, \ldots$. This solution shows that it is possible to have real parameters even when the imaginary unit appears in various places in the equations. The envelope-hole solitons are not discussed further in this paper.

B. Hirota–Satsuma equation

In all of the systems above we assumed $[B(X, T, \ldots)]^* = B(-X^*, -T^*, \ldots)$. This is trivially true if B is real and either even or odd (the overall i factor does not matter). The real version of the NLS-type (3) was not studied in Refs. 2 and 3, because we previously assumed that C in (3b) has the property $C(0,0,\ldots) = 0$. Now $C \equiv 1$, so we will include this new type in our analysis. A well-known equation of this type is the Hirota–Satsuma equation

$$\begin{aligned} &u_t + u_{xxx} + 3vu_x = 0, \\ &v_t - a(v_{xxx} + 6vv_x) = 2buu_x. \end{aligned} \tag{22}$$

The substitution (11) with $\beta = 1$ yields now[15]

$$(D_t + D_x^3)F \cdot G = 0, \quad D_x(D_t - aD_x^3)F \cdot F = bG^2, \tag{23}$$

which has N-soliton solutions for $a = \frac{1}{2}$. The 1SS is given by (6)–(8) with $p^* = p$, $\Omega^* = \Omega, \ldots$.

C. Benjamin–Ono equation

One can also have complex bilinear equations with one dependent variable,

$$P(D_x, D_t, \ldots)F \cdot F^* = 0, \tag{24}$$

where P satisfies (4). Such an equation is obtained, e.g., for the Benjamin–Ono (BO) equation

$$u_t + 2uu_x + Hu_{xx} = 0, \tag{25}$$

where H is the Hilbert transform. One now defines a new dependent variable by[16–19]

$$u = i\,\partial_x \log(F^*/F) \tag{26}$$

and then using the property

$$H\,[\,i\,\partial_x \log(F^*/F)\,] = -\partial_x \log(F^*F) \tag{27}$$

one obtains

$$(D_x^2 + iD_t)F \cdot F^* = 0. \tag{28}$$

For the BO equation one uses rational solutions, indeed the polynomial character of F is needed to derive (27). In this paper we will discuss only standard exponential solutions, for which case the analog of (28) is obtained from the Joseph equation

$$u_t + 2uu_x + Gu_{xx} = 0, \tag{29}$$

where the integral transform G is defined by

$$G\,[u(x,t)] = \frac{1}{2}kP\int_{-\infty}^{\infty}\left[\coth\frac{1}{2}\pi k(x' - x) - \operatorname{sgn}(x' - x)\right]u(x,t)dx'. \tag{30}$$

One now has for exponential F,[19]

$$G\,[\,i\,\partial_x \log(F^*/F)\,] = -\partial_x \log(F^*F) + ku, \tag{31}$$

and thus, assuming (26), one obtains[20]

$$(D_x^2 + iD_t + ikD_x)F \cdot F^* = 0. \tag{32}$$

This contains the Benjamin–Ono equation in the limit $k \to 0$, while the KdV equation is obtained (in scaled variables) as $k \to \infty$.[19]

If the system is assumed to be fully complex then the 1SS for (24) can be written as

$$F = 1 + e^n, \tag{33}$$

with n as in (7), and the parameters satisfy the dispersion relation

$$P(p, \Omega, \ldots) = 0. \tag{34}$$

However, it is also possible to assume that the parameters are real, the ansatz is then[18,20] (a slightly different choice was used in Ref. 17)

$$f = 1 + e^{i\phi} e^{n}, \quad \phi \text{ real}, \quad n \text{ as in (7) (real)}, \tag{35}$$

The complex phase $e^{i\phi}$ is determined when (35) is substituted into (24):

$$e^{2i\phi} = -P(-p,-\Omega,...)/P(p,\Omega,...). \tag{36}$$

This 1SS seems peculiar due to the fact that the parameters are not constrained by a dispersion relation. However, (36) becomes a dispersion relation if we insist that $\phi(k,\Omega,...) = dk$ for some *constant* d. This requirement is natural, because then we can write (24) as[18]

$$P(D_x,D_t,...)F_+ \cdot F_- = 0, \quad \text{where } F_\pm = F(x \pm id), \tag{24'}$$

or[19]

$$P(D_x,D_t,...)\exp[idD_x]F\cdot F = 0, \tag{24''}$$

where F is as in (33) with real parameters. In (24″) it is only the real (= even) part of the operator that counts. We will not discuss the real solutions further but consider only the fully complex case.

In this paper we will report the results of a search for bilinear equations of types (3) and (24) having a 2SS. In the next section we will derive the conditions for the existence of a complex 2SS and mixed 1 + 1SS for the NLS equation of type (3). The conditions are given in the general case, while our search is limited to the special case where $C\equiv 1$ and A is quadratic. The real case of Hirota–Satsuma is also discussed for $C\equiv 1$ and A quadratic. In Sec. III we derive the conditions for the existence of 2SS's in equations of the Benjamin–Ono type. For this system our search extends up to degree 4.

II. THE NONLINEAR SCHRÖDINGER TYPE

A. Conditions for complex 2SS's

We start now our discussion of equations of type (3) with (4) and (5). The standard 1SS was given in (6)–(9). [If B and the constants are assumed to be real and $C(0,0,...) = 0$ then we obtain the 1SS discussed in Ref. 3.] In addition to the 1SS (6) the system (3) has also the more trivial 1SS,

$$F = 1 + e^{n}, \quad G = 0, \tag{37}$$

where now A gives the dispersion relation, rather than B as in (9). Obviously such 1SS's can be combined to NSS's while keeping $G = 0$, and thus we find as our first condition that the polynomial A must satisfy the 3SC and 4SC applicable to a bilinear equation of type $AF\cdot F = 0$.[1]

For NSS's composed of 1SS's of type (6) we may assume that they follow the general pattern proposed by Hirota[4]:

$$F = \sum_{\mu=0,1}^{(0)} \exp\left[\sum_{i<j}^{(2N)} \phi(i,j)\mu_i\mu_j + \sum_{i=1}^{2N} \mu_i n_i\right], \tag{38a}$$

$$G = \sum_{\mu=0,1}^{(1)} \exp\left[\sum_{i<j}^{(2N)} \phi(i,j)\mu_i\mu_j + \sum_{i=1}^{2N} \mu_i n_i\right], \tag{38b}$$

$$G^* = \sum_{\mu=0,1}^{(-1)} \exp\left[\sum_{i<j}^{(2N)} \phi(i,j)\mu_i\mu_j + \sum_{i=1}^{2N} \mu_i n_i\right], \tag{38c}$$

where

$$n_i = p_i x + \Omega_i t + \cdots + m_i, \quad \text{for } i = 1,...,N, \tag{39a}$$

$$n_i = n^*_{i-N}, \quad \text{for } i = N+1,...,2N, \tag{39b}$$

$$B(p_i,\Omega_i,...) = 0 \tag{40}$$

and the μ summations $\Sigma^{(a)}$ go over all vectors $\mu = (\mu_1,...,\mu_{2N})$, where $\mu_i = 0$ or 1, and

$$\sum_{i=1}^{N} \mu_i = \sum_{i=1}^{N} \mu_{i+N} + a. \tag{41}$$

Equation (40) defines an affine manifold where the parameters p_i, Ω_i,... (= $\mathbf{p}_i$) belong; we use the notation V_B for it, thus (40) means $\mathbf{p}_i \in V_B$. The functions $\phi(i,j)$ depend on the parameters $p_i,p_j,\Omega_i,\Omega_j,...$. For $i \leqslant N$, $\phi(i,i+N)$ is determined already by the 1SS (6), for other indices ϕ is determined when the 2SS [(38) with $N=2$] is substituted into (3). Let us define the degree of a term by the number of n's in the exponent, then after the substitution we find that (3a) has terms of odd degree, while (3b) has even degree terms. From degree 2 (or 6) terms we find

$$e^{\phi(i,j+N)} = C(p_i - p_j^*,...)/[2A(p_i + p_j^*,...)], \tag{42}$$

and from some of the degree 4 terms

$$e^{\phi(i,j)} = 2A(p_i - p_j,...)/C(p_i + p_j,...), \tag{43}$$

both for all $i<j\leqslant N$. Also $e^{\phi(i+N,j+N)} = [e^{\phi(i,j)}]^*$.

The degree 3 and 5 terms yield a condition, which can be compactly written as

$$\sum_{\sigma=\pm 1}^{(1)} B\left(\sum_{i=1}^{3} \sigma_i p_i,...\right) \prod_{i<j}^{(3)} P(\sigma_i\sigma_j; p_i - \sigma_i\sigma_j p_j,...) = 0, \tag{44}$$

for all $\mathbf{p}_i \in V_B$. Here we have defined $\mathbf{p}_3 = -\mathbf{p}_1^*$ so that parameters satisfy dispersion relation (40) for all indices. The polynomial P is defined by

$$P(+1;\cdots) = A(\cdots), P(-1;\cdots) = C(\cdots). \tag{45}$$

The summation is over all those $\sigma_i \in \{-1,1\}$ such that $\Sigma_i \sigma_i = 1$.

The condition obtained from the remaining part of the degree 4 terms can be written as

$$\sum_{\sigma=\pm 1}^{(0,2)} P\left(\prod_i \sigma_i; \sum_{i=1}^{4} \sigma_i p_i,...\right)[3\sigma_1\sigma_2\sigma_3\sigma_4 - 1] \times \prod_{i<j}^{(4)} P(\sigma_i\sigma_j; p_i - \sigma_i\sigma_j p_j,...) = 0, \quad \forall \mathbf{p}_i \in V_B, \tag{46}$$

where $\mathbf{p}_3 = -\mathbf{p}_1^*$, $\mathbf{p}_4 = -\mathbf{p}_2^*$, and the P's are as in (45). The summation is now over all those σ's for which $\Sigma_i \sigma_i = 0$ or 2.

The conditions (44) and (46) are close analogs to the 3SC and 4SC given in Ref. 3 (especially if we could assume that C is even). The big difference is that these conditions are obtained already when we try to construct a 2SS. The fact that 2SS's are not automatic with complex degrees of freedom is understandable, for two complex sets of parameters have four real degrees of freedom.

It is useful at this point also to consider various special cases of the above conditions. For example, if $C\equiv 1$, which is assumed in our search, (44) simplifies to

$$A(p_1-p_2,...)B(p_1+p_2-p_3,...)$$
$$+A(p_2-p_3,...)B(-p_1+p_2+p_3,...)$$
$$+A(p_3-p_1,...)B(p_1-p_2+p_3,...)=0,$$
$$\forall\ \mathbf{p}_i\in V_B. \qquad (44\text{-}1)$$

If the system was *real* (which implies $\mathbf{p}_3=-\mathbf{p}_1$, $\mathbf{p}_4=-\mathbf{p}_2$) then from both conditions (44) and (46) we could extract a factor $C(0,...)$. Thus if $C(0,...)=0$ then 2SS's would be automatic, in agreement with Ref. 3. In the following we will consider the real case only if $C\equiv 1$, then the conditions are $(\mathbf{p}_1,\mathbf{p}_2\in V_B)$

$$A(p_1-p_2,...)B(p_1+2p_2,...)$$
$$+A(p_1+p_2,...)B(p_1-2p_2,...)=0, \qquad (44\text{-r1})$$
$$A(2p_1-2p_2,...)A(p_1+p_2,...)^2$$
$$+A(2p_1+2p_2,...)A(p_1-p_2,...)^2$$
$$-2[A(2p_1,...)+A(2p_2,...)]$$
$$\times A(p_1+p_2,...)A(p_1-p_2,...)=0. \qquad (46\text{-r1})$$

The 2SS used above was composed of two 1SS's of type (6). But since the system has also 1SS's of type (37) one may ask whether one can combine these different 1SS's as well to construct a 2SS. In fact it can be argued that for a fully integrable system one should be able to combine any kind of solitons to make a multisoliton solution.[21] The natural ansatz for the 2SS is in this case

$$F=1+e^{\phi(1,1)}e^{n_1+n_1^*}+e^{n_A}+e^{\phi(1,1)}L^*Le^{n_1+n_1^*+n_A},$$
$$G=e^{n_1}+Le^{n_1+n_A}, \qquad (47)$$

where the parameters in n_A satisfy the dispersion relation $A(p_A,\Omega_A,...)=0$. When (47) is substituted into (3a) we find from the degree 2 and 4 terms

$$L=-B(p_1-p_A,...)/B(p_1+p_A,...), \qquad (48)$$

while the degree 2 and 4 terms in (3b) yield (42) again. The degree 3 terms in (3b) yield a condition, which can be written as

$$\sum_{\sigma=\pm1}{}' P\Big(-\sigma_1\sigma_2;\sum_{i=1}^{3}\sigma_ip_i,...\Big)P(\sigma_1\sigma_2;p_1-\sigma_1\sigma_2p_2,...)$$
$$\times B(p_1-\sigma_1\sigma_3p_3,...)B(p_2-\sigma_2\sigma_3p_3,...)=0, \qquad (49)$$

where now $\mathbf{p}_2=-\mathbf{p}_1^*$, $\mathbf{p}_3=\mathbf{p}_A$, that is, $\mathbf{p}_1,\mathbf{p}_2\in V_B$, $\mathbf{p}_3\in V_A$, and in the summation $\sigma_1=1$. This is again a typical 3SC and could be written like (44) except that now the polynomials are not determined by the signs of the σ's alone.

Let us again look at some special cases. For $C\equiv 1$ (49) simplifies to

$$\sum_{\sigma=\pm1}[A(p_1-p_2+\sigma p_3,...)B(p_1-\sigma p_3,...)$$
$$+A(p_1-p_2,...)B(p_1+\sigma p_3,...)]$$
$$\times B(p_2+\sigma p_3,...)=0, \qquad (49\text{-}1)$$

where $\mathbf{p}_1,\mathbf{p}_2\in V_B$, $\mathbf{p}_3\in V_A$. If the system is real (i.e., $\mathbf{p}_2=-\mathbf{p}_1$) then C may be assumed to be real and we find that each term in (49) would have a factor $C(0,...)$ or $C(p_A,...)$. If the system is real and $C\equiv 1$ then (49-1) becomes $(\mathbf{p}_1\in V_B,\ \mathbf{p}_3\in V_A)$

$$\sum_{\sigma=\pm1}[A(2p_1+\sigma p_3,...)B(p_1-\sigma p_3,...)$$
$$+A(2p_1,...)B(p_1+\sigma p_3,...)]$$
$$\times B(-p_1+\sigma p_3,...)=0. \qquad (49\text{-r1})$$

In Ref. 3 we assumed that the system is real and $C(0,...)=0$, then what remains from (49) is

$$A(2p_B,...)B(p_B-p_A,...)$$
$$\times B(p_B+p_A,...)C(p_A,...)=0, \qquad (49\text{-r0})$$

for all $\mathbf{p}_B\in V_B$, $\mathbf{p}_A\in V_A$. Obviously this is satisfied if $C\equiv A$. In Ref. 3 we obtained some results which passed the standard 3SC and for which $C\neq A$. However, we did not check whether they also had mixed 2SS's. It turns out that none of the results in Ref. 3 with $C\neq A$ pass (49-r0).

B. Complex results

Our search follows the pattern of previous papers, and we will not discuss the methods in detail here. We assume that $C\equiv 1$, and (as in our previous papers) that A is a quadratic function of X and T. Then A can be rotated and scaled so that $A=X^2$, or XT. In this case the "four-soliton condition" (46) is identically satisfied. The search proceeds as usual from leading monomial to leading homogeneous polynomial to the final result.

1. $A=X^2$. The first problem is to determine the leading monomials in B. We have the freedom of redefining the T variable; we use this to define the highest-order factor (differing from X) of the leading homogeneous polynomial as T. The monomials that satisfy (44-1) and (49-1) are discussed in Appendix A. The results are as follows.

1.1. $B=X^M$, $\sqrt{B}=X^K$, $M,K>0$. Equation (49-1) poses no restrictions, but for (44-1) one needs $K\leqslant[(M+1)/3]+1$ ($[a]$ stands for the integer part of a), except that when $M=3$ it is sufficient to have $K=3$.

1.2. $B=X^MT^N$, $\sqrt{B}=X^KT^L$, $M,N,K,L>0$. For this we find that $M=K=L=1$, $N=2n+1$ is necessary. The only possibly nonlinear case can arise from $B=\sqrt{B}=XT$.

1.3. $B=T^N$, $\sqrt{B}=T^L$, $N,L>0$. In this case (49-1) requires $L=1$, $N=2n+1$, which also passes (44-1).

Next we consider the extension of the above results to homogeneous polynomials. Such a possibility exists only for case 1.2, but the generalizations do not pass the test. After this we tested systematically the nonhomogeneous generalizations fitting to the above; our results are the following.

(i) The most general nonlinear result is

$$B=iaX^3+XT+iY+b,\quad A=X^2,\quad C=1. \qquad (50)$$

As special cases it contains the original NLS (2) and Hirota's generalization (2a').

(ii) Linear dispersion manifolds

1.A. Up to degree 4 any polynomial in X (subject to 1.1 above) passes both (44-1) and (49-1) but at higher degrees we get additional conditions. At degree 5 polynomials with up to two different factors are acceptable; there are also two other possibilities, $i(X+ia)^3(X^2+ibX+c)$ and $i(X+ia)(X^2+ibX+c)^2$, for which we find that $a=b=0$ is necessary. At degree 6 also polynomials with up to two factors are acceptable, the other cases pass the tests in the following special cases: $(X+ia)^4(X^2+ibX+c)$ when $2(2a-b)(b^2+4c)=0$, $(X^3+iaX^2+ibX+c)^2$ when $a^2+3b=0$, and $(X+ia)^3(X+ib)^2(X+ic)$ when $4a^2-2ab-6ac+3b^2-4bc+5c^2=0$. Similar results are expected at higher degrees.

1.B. $B=(X+ia)(T+ib)^{2n+1}$, $A=X^2$, $C=1$.

1.C. $B=i(T+ia)^{2n+1}$, $A=X^2$, $C=1$.

For some of the above cases we find that the parameters p and Ω will be pure imaginary. In such cases p and p^* are related, but in our formulations this was not assumed. Thus the conditions that were used before are in fact too strong, and other such more or less trivial cases may be acceptable.

2. $A=XT$. In this case we have the freedom of reflecting $X\leftrightarrow T$, but no rotational freedom. In Appendix A the following results are obtained for the leading monomial.

2.1. $B=X^M$, $\sqrt{B}=X^K$, $M,K>0$. Condition (44-1) implies $K\leqslant[(M+1)/3]+1$, but (49-1) is more stringent, demanding that $M=2$ or $K=1$. Thus the only interesting case is $B=\sqrt{B}=X^2$.

2.2. $B=X^MT^N$, $\sqrt{B}=X^KT^L$, $M,N,K,L>0$. For these (44-1) is never satisfied, while (49-1) requires ($M=2$ or $K=1$) and ($N=2$ or $L=1$). Since both conditions should be satisfied we find that no monomial of this type is acceptable.

As for the homogeneous extensions we find that $B=X^M$, $\sqrt{B}=X$ extends to $(X+aT)^M$, in fact to Y^M, and $B=\sqrt{B}=X^2$ to X^2+aT^2. The results are as follows.

(iii) The only nonlinear result in this case is

$$B=X^2+aT^2+iY+b,\quad A=XT,\quad C=1. \tag{51}$$

For $a=1$, $b=0$ this can be rotated and scaled to the 2DNLS (12), while for $a=0$, $b=0$ we obtain (14), and for $Y\to 0$ we obtain the coupled Higgs field equation of Tajiri.

(iv) The results with linear dispersion manifold are

2.A. $B=(X+ia)^{2m}$, $A=XT$, $C=1$.

2.B. $B=i(Y+ia)^{2m+1}$, $A=XT$, $C=1$.

C. Real results

We will next discuss the real results with $C=1$, i.e., results of the Hirota–Satsuma type. We take again $A=X^2$ or $A=XT$, which is simpler than in (23). The conditions are given in (44-r1) and (49-r1). Since $\mathbf{p}_2=-\mathbf{p}_1$ when compared to Sec. II B, we find the conditions somewhat easier. For monomials the results are derived in Appendix B.

1. $A=X^2$

1.1. $B=X^M$, $\sqrt{B}=X^K$, $M,K>0$. Equation (44-r1) implies $K\leqslant[M/4]+[(M+1)/4]+2$, expect that there is a special case: for $M=13$, $K=9$ is sufficient.

1.2. $B=X^TT^N$, $\sqrt{B}=X^KT^L$, $M,N,K,L>0$. Equation (44-r1) is passed if $M=2m+1$, $N=2n+1$, $K=L=1$.

1.3. $B=T^N$, $\sqrt{B}=T^L$, $N,L>0$. Now from (49-r1) we obtain $L=2$, $N=2n+1$.

As can be seen the restrictions are less strict than the ones in Sec. B 1. Again the possible homogeneous generalization of 1.2, $B=(X+aT)^MT^N$, does not pass (44-r1) except when $a=0$.

(i) The nonlinear results are

$$B=XT+b,\quad A=X^2,\quad C=1, \tag{52}$$

and

$$B=X^3+T,\quad A=X^2,\quad C=1, \tag{53}$$

which turn out to be just the real and imaginary parts of (50).

(ii) The linear manifold results are as follows.

1.A. Polynomials that depend only on X: Up to degree 7 the acceptable results fall into the sequence $B=(X^2+a)^M$ and $B=(X^2+a)^MX^{2n+1}$.

1.B. $B=X^{2m+1}T^{2n+1}$, $A=X^2$, $C=1$.

1.C. $B=T^{2n+1}$, $A=X^2$, $C=1$.

2. $A=XT$

2.1. $B=X^M$, $\sqrt{B}=X^K$, $M,K>0$. From (49-r1) we obtain $M=2$ or $K=1$.

2.2. $B=X^MT^N$, $\sqrt{B}=X^KT^L$, $M,N,K,L>0$. Now from (44-r1) we get $M=2m$, $N=2n$, $K=L=1$.

Of these 2.1 is as before but 2.2 was not acceptable in the complex case. Statement 2.2 does not have homogeneous generalizations, but 2.1 generalizes to X^2+aT^2, and to $(X+aT)^M$, which generalizes still further to Y^M.

(iii) The nonlinear result,

$$B=X^2+aT^2+b,\quad A=XT,\quad C=1, \tag{54}$$

is again just the real part of (51).

(iv) The linear manifold results are

2.A. $B=X^{2m}$, $A=XT$, $C=1$;

2.B. $B=Y^{2m+1}$, $A=XT$, $C=1$;

2.C. $B=X^{2m}T^{2n}$, $A=XT$, $C=1$;

of which only the last one is a generalization over the complex case.

III. THE BENJAMIN–ONO TYPE

A. Conditions for a complex two-soliton solution

We will now derive the conditions for the Benjamin–Ono-type equation (24) to have a complex 2SS generalizing (33). The general N-soliton ansatz is of the form

$$F=\sum_{\mu=0,1}^{(0,1)}\exp\left[\sum_{i<j}^{(2N)}\phi(i,j)\mu_i\mu_j+\sum_{i=1}^{2N}\mu_i n_i\right], \tag{55}$$

which is a combination of (38a) and (38b), with the conventions (39). The dispersion relation is given by the polynomial P (34).

When the 2SS (55) with $N=2$ is substituted into (33) we obtain

$$e^{\phi(i,j)}=-\tfrac{1}{2}P(p_i-p_j,\ldots)/P_e(p_i+p_j,\ldots),\quad i\leqslant N<j, \tag{56a}$$

$$e^{\phi(i,j)}=-2P_e(p_i-p_j,\ldots)/P(p_i+p_j,\ldots),\quad i,j\leqslant N, \tag{56b}$$

where P_e is the even part of P. In addition to this we get the following conditions:

$$\sum_{\sigma=\pm 1}^{(1)}P\left(\sum_{i=1}^{3}\sigma_ip_i,\ldots\right)\prod_{i<j}^{(3)}P(\sigma_i\sigma_j;p_i-\sigma_i\sigma_jp_j,\ldots)=0, \tag{57}$$

$$\sum_{\sigma=\pm 1}^{(0,2)} P\left(\prod_i^4 \sigma_i;\ \sum_{i=1}^4 \sigma_i p_i,...\right)[3\sigma_1\sigma_2\sigma_3\sigma_4 - 1]$$

$$\times \prod_{i<j}^{(4)} P(\sigma_i\sigma_j; p_i - \sigma_i\sigma_j p_j,...) = 0, \qquad (58)$$

where $\mathbf{p}_3 = -\mathbf{p}_1^*$, $\mathbf{p}_4 = -\mathbf{p}_2^*$ $\forall \mathbf{p}_i \in V_p$, $P(+1;\cdots) = P_e(\cdots)$, and $P(-1;\cdots) = P(\cdots)$.

Note the close analogy of (57) and (58) with (44) and (46): the form is the same, only the meaning of the various polynomials differ from that used in Sec. II A. From the identical form it follows that if (24) has 2SS's then there exists a corresponding NLS-type equation $PG\cdot F = 0$, $P_e F\cdot F + PG\cdot G^* = 0$ which also has a 2SS [for the relative signs, compare, e.g., (42) and (56)]. This is also easy to see directly: substitute $f = F + G$, $f^* = F + G^*$ into $Pf\cdot f^* = 0$ to $[PF\cdot F + PG\cdot G^*] + PG\cdot G + (PG\cdot F)^* = 0$. Due to the type of the ansatz we find that the three groups of terms in the equation must vanish separately, which is equivalent to the pair of equations mentioned earlier.

We also recall the fact that the conditions [(57) and (58)] are obtained for the existence of a 2SS, because no algebraic relation is assumed between the parameters and their complex conjugates. This also means that these conditions do not vanish automatically if P is assumed to be even (and therefore real), it is also necessary to assume that the parameters are real and therefore $\mathbf{p}_3 = -\mathbf{p}_1$, $\mathbf{p}_4 = -\mathbf{p}_2$.

B. Results

The even part of P appears in the denominator of $e^{\phi(i,j)}$ and therefore we will only consider cases where P_e does not vanish identically. We searched for polynomials P of degree up to 4 that satisfied (57); we did not check (58).

If the degree of P is less than 4 we must have a nonvanishing quadratic term. The rotational degrees of freedom are fixed partially by rotating this term to X^2 or XT. When the degree of P is 4 we fix the rotational degrees by the leading term (classified as in Ref. 1). The acceptable leading monomials were $P = \sqrt{P} = X^4$; $P = X^3T$, $\sqrt{P} = X^2T$; $P = X^2T^2$, $\sqrt{P} = XT$. The only possible homogeneous extension $P = X^2T(X + aT)$, $\sqrt{P} = XT(X + aT)$ did not pass (57). We then tested systematically the possibility of additional terms, cubic and linear if $\deg(P) < 4$, and cubic, quadratic, and linear terms if $\deg(P) = 4$. The results are as follows.

(i) The genuinely nonlinear results are

$$P = iaX^3 + X^2 + iT, \qquad (59)$$

which generalizes the Benjamin–Ono equation (28) and (32) by the cubic term, and

$$P = XT + i(aX + bT), \qquad (60)$$

which has $X \leftrightarrow T$ symmetry. The suggested generalization $XT + iY$ did not pass the test.

(ii) As usual we obtained several results with linear dispersion manifold.

A. The one-dimensional results are $P = X^4 + aX^2 + ibX$; $P = X^2(X^2 + iaX + b)$; $P = (X+i)^2(X^2 + iaX)$; and $P = X^4 + iX^3 + aX^2 + ibX$, if $4a^2 - 20ab + a + 24b^2 - 3b = 0$.

B. Up to degree 4 the other results fit into $P = iX^{2n+1}(T+i)^M$, which seems to be allowed also for higher degree.

IV. CONCLUSIONS

In this paper we have discussed certain types of complex bilinear equations. It turned out that the construction of complex two-soliton solutions presented conditions analogous to the usual three- and four-soliton conditions. This is understandable because a two-soliton solution has four sets of real parameters.

Our search revealed generalizations of the bilinear formulations of the known integrable systems of this type. For example, the original nonlinear Schrödinger equation (2) seems to have two (2 + 1)-dimensional generalizations: the well-known (12) or (51) and the apparently new (50). For the Benjamin–Ono type we have the generalization (59) and the new case (60).

We do not know if these new models are completely integrable, but they are the only equations within the class studied that do at least pass the first condition of having two-soliton solutions. It would be interesting to apply other tests of integrability on these systems.

Note added in proof: The Benjamin–Ono-type equation (60) has also been found by Matsuno [see Ref. 22, Eq. (2.3)]. Ito has found[23] that (50) has 3SS for some arbitrarily chosen parameter values. This equation also passes the Painlevé test.[24]

ACKNOWLEDGMENTS

The computations were done with an IBM 3033 at the University of Turku Computing Center using REDUCE 3.2.[25]

The author was supported by the Academy of Finland.

APPENDIX A: CONDITIONS FOR THE LEADING TERMS, NLS EQUATION

In this Appendix we will derive the conditions arising from (44-1) and (49-1). We will only discuss the more stringent of the conditions, for 1.1, 1.2, and 2.2 it is (44-1) while for 1.3 and 2.1 it is (49-1). Since the rewrite rules do not here connect different monomials it means that all monomials must separately vanish. Conditions are derived by a judicious choice of monomials.

1.1. Condition (44-1) reads

$$(p_1 - p_2)^2(p_1 + p_2 - p_3)^M + (\text{cyclic terms}) = 0, \qquad (\text{A1})$$

when $p_1^K \to 0$. Our method is to isolate that monomial $p_1^m p_2^n p_3^k$, where m, n, and k are the smallest (for example, if $M = 3\mu$ it is obtained by $m = k = \mu + 1$, $n = \mu$, and cyclic permutations) and if its coefficient does not vanish identically we can read off the maximum K from the maximum exponent.

The terms where p_3 appears with power k are given by

$$p_3^k\Big\{[(p_1 - p_2)^2(p_1 + p_2)^{M-k}(-1)^k$$

$$+ (p_1^2 + (-1)^{M-k}p_2^2)(p_1 - p_2)^{M-k}]\binom{M}{k}$$

$$- 2(p_1 - (-1)^{M-k}p_2)(p_1 - p_2)^{M-k+1}\binom{M}{k-1}$$

$$+ (p_1-p_2)^{M-k+2}(1+(-1)^{M-k})\binom{M}{k-2}\Big\}\,. \quad \text{(A2)}$$

The terms combine differently depending whether $M-k$ is even or odd. Suppose $M-k$ is odd. Then the center term is

$$p_1^m p_2^{m+1} p_3^k \binom{2m+1}{m}\binom{M+1}{k} 2(-1)^{m+1} \times\{[1+(-1)^{k+m+1}]-k\}/[(2m+1)(M+1)]. \quad \text{(A3)}$$

For most k's this does not vanish identically, so a rewrite rule must be imposed.

(a) If $M=3\mu$ we take $k=\mu+1$, $m=\mu$ then the term in curly brackets in (A3) is $\mu-1$, thus for $\mu>1$ we get the rewrite rule $K\leqslant\mu+1$.

(b) If $M=3\mu+2$ we take again $k=\mu+1$, but $m=\mu+1$, thus $K\leqslant\mu+2$.

(c) For $M=3\mu+1$ the $M-k$ odd terms yield $K\leqslant\mu+2$, but this condition can be improved by the $M-k$ even terms to $K\leqslant\mu+1$.

These results combine to $K\leqslant[(M+1)/3]+1$ for $M>3$. For $M\leqslant 3$ it is sufficient to take $K=M$.

1.2. Now (44-1) reads

$$(p_1-p_2)^2(p_1+p_2-p_3)^M(\Omega_1+\Omega_2-\Omega_3)^N + (\text{cyclic terms})=0, \quad \text{(A4)}$$

when $p_i^K\Omega_i^L\to 0$. From (A4) we choose those terms that have maximum power of Ω_3 and no p_3. They are given by

$$[(p_1-p_2)^2(p_1+p_2)^M(-1)^N+p_2^2(-p_1+p_2)^M + p_1^2(p_1-p_2)^M]\Omega_3^N. \quad \text{(A5)}$$

These terms do not vanish by the rewrite rule $p_i^K\Omega_i^L\to 0$; thus it should vanish identically. This is possible only if N is odd and $M=1$ (and therefore $K=1$), as can be easily seen. To continue in that case let us take the terms with $\Omega_3^{N-1}\Omega_2$,

$$[(p_1-p_2)^2(p_1+p_2-p_3)+(p_2-p_3)^2(-p_1+p_2+p_3) - (p_3-p_1)^2(p_1-p_2+p_3)]N\Omega_3^{N-1}\Omega_2, \quad \text{(A6)}$$

and from these the terms linear in p_3,

$$2p_2p_3(p_1-p_2)N\Omega_3^{N-1}\Omega_2. \quad \text{(A7)}$$

These can vanish only by the rewrite rule, but then $L=1$.

1.3. Condition (49-1) is

$$\sum_{\sigma=\pm1}[(p_1-p_2+\sigma p_3)^2(\Omega_1-\sigma\Omega_3)^N + (p_1-p_2)^2(\Omega_1+\sigma\Omega_3)^N](\Omega_2+\sigma\Omega_3)^N=0, \quad \text{(A8)}$$

for $\Omega_1^L\to 0$, $\Omega_2^L\to 0$, and $p_3\to 0$. Let us take the terms with no p_3 and expand them with decreasing order in Ω_3. We obtain

$$\sum_{\sigma=\pm1}(p_1-p_2)^2\{\Omega_3^N[(-1)^N+1] + N\Omega_1\sigma\Omega_3^{N-1}[(-1)^{N-1}+1]+\cdots\} \times\{\Omega_3^N+N\Omega_2\sigma\Omega_3^{N-1}+\cdots\}. \quad \text{(A9)}$$

The leading term vanishes only if N is odd, in which case the next to leading term is

$$4N^2(p_1-p_2)^2\Omega_3^{2N-2}\Omega_1\Omega_2. \quad \text{(A10)}$$

This can vanish only by a rewrite rule and thus we get $L=1$.

2.1. Condition (49-1) is

$$\sum_{\sigma=\pm1}[(p_1-p_2+\sigma p_3)(\Omega_1-\Omega_2+\sigma\Omega_3)(p_1-\sigma p_3)^M + (p_1-p_2)(\Omega_1-\Omega_2)(p_1+\sigma p_3)^M](p_2+\sigma p_3)^M=0, \quad \text{(A11)}$$

for $p_1^K\to 0$, $p_2^K\to 0$, $p_3\Omega_3\to 0$. Now we take the terms with the maximum number of p_3 and no Ω_3. The terms with $(\sigma p_3)^{2M+1}$ vanish in the σ summation, the next terms are

$$p_3^{2M}(\Omega_1-\Omega_2)\{(p_1-p_2)[(-1)^M+1] + M(-1)^{M-1}p_1+Mp_2\}. \quad \text{(A12)}$$

These terms vanish only if $M=2$ or $K=1$.

2.2. In this case condition (44-1) is more strict, it reads

$$(p_1-p_2)(\Omega_1-\Omega_2)(p_1+p_2-p_3)^M \times(\Omega_1+\Omega_2-\Omega_3)^N+(\text{cyclic terms})=0, \quad \text{(A13)}$$

when $p_i^K\Omega_i^L\to 0$. Take terms with maximal power of Ω_3 and no p_3. They are given by

$$\Omega_3^{N+1}[-p_2(-p_1+p_2)^M-p_1(p_1-p_2)^M], \quad \text{(A14)}$$

and do not vanish by the rewrite rule in question.

APPENDIX B: CONDITIONS FOR THE LEADING MONOMIAL, HIROTA-SATSUMA EQUATION

In this Appendix we derive conditions for monomials for real systems with $C=1$. In each case we discuss only the stronger one of the conditions (44-r1) and (49-r1).

1.1. Equation (44-r1) reads

$$(p_1-p_2)^2(p_1+2p_2)^M+(p_1+p_2)^2(p_1-2p_2)^M=0, \quad \text{(B1)}$$

when $p_i^K\to 0$. The coefficient of $p_1^{M+2-m}p_2^m$ vanishes when m is odd and for even m it is

$$\binom{M}{m}\frac{9m^2-12mM-21m+4M^2+12M+8}{(M-m+1)(M-m+2)}. \quad \text{(B2)}$$

We choose now the optimal m for given M and check whether (B2) vanishes or not.

(a) When $M=4n$, take $m=2n$ to obtain $K\leqslant 2n+2$, because expression (B2) does not vanish for integer n.

(b) For $M=4n+1$ the optimal m is $m=2n+2$. The expression in the curly brackets is $2(2n^2-5n-3)$, which has $n=3$ as the only integer root. This means that $K\leqslant 2n+2$ for all the other cases except for $n=3$, i.e., for $M=13$ it is sufficient that $K\leqslant 9$, which is also necessary for the term with $m=6$.

(c) If $M=4n+2$, take $m=2n+2$ to obtain $K\leqslant 2n+2$.

(d) When $M=4n+3$, take $m=2n+2$ to obtain $K\leqslant 2n+3$. The results combine to $K\leqslant[M/4]+[(M+1)/4]+2$, except for the special case $M=13$, $K=9$.

1.2. Equation (44-r1) reads

$$(p_1-p_2)^2(p_1+2p_2)(q_1+2q_2)^N + (p_1+p_2)^2(p_1-2p_2)(q_1-2q_2)^N=0, \quad \text{(B3)}$$

when $p_i^Kq_i^L\to 0$. The terms $p_1^{2+M}(2q_2)^N$ vanish only if N is odd and similarly for exchanged indices we find M must be odd. Expanding to the next term we obtain

$$4p_1^{1+M}(2q_2)^Np_2(M-1), \quad \text{(B4)}$$

which implies that $K=1$. Similarly we can get the result $L=1$.

1.3. Now (49-r1) is the strongest condition, it reads

$$\sum_{\sigma=\pm 1}[(2p_1+\sigma p_3)^2(\Omega_1-\sigma\Omega_3)^{2N}] + 8p_1^2(\Omega_1^2-\Omega_3^2)^N=0, \quad \text{(B5)}$$

for $\Omega_1^L\to 0, p_3\to 0$. The leading terms in p_1 and Ω_3 vanish only if N is odd. The next to leading terms contain Ω_1^2 so we obtain $N=2n+1, L=2$.

2.1. We use again (49-r1), it is now

$$\sum_{\sigma=\pm 1}[(2p_1+\sigma p_3)(2\Omega_1+\sigma\Omega_3)(p_1-\sigma p_3)^{2M}] + 8p_1\Omega_1(p_1^2-p_3^2)^M=0, \quad \text{(B6)}$$

with $P_3\Omega_3\to 0, p_1^K\to 0$. The p_3^{2M+1} terms vanish by σ summation or by the rewrite rule in index 3. The coefficient of p_3^{2M} is

$$4(2\Omega_1+\sigma\Omega_3)p_1(2-M), \quad \text{(B7)}$$

thus we get the conditions $M=2$ or $K=1$.

2.2. Now we use (44-r1) that reads

$$(p_1-p_2)(\Omega_1-\Omega_2)(p_1+2p_2)^M(\Omega_1+2\Omega_2)^N + (p_1+p_2)(\Omega_1+\Omega_2)(p_1-2p_2)^M(\Omega_1-2\Omega_2)^N, \quad \text{(B8)}$$

for $p_i^K\Omega_i^L\to 0$. The terms $p_1^{M+1}\Omega_2^{N+1}$ and $p_2^{M+1}\Omega_1^{N+1}$ vanish only if N and M, respectively, are even. A next to leading term is

$$2^{M+1}p_2^{M+1}\Omega_1^N\Omega_2(1-2N), \quad \text{(B9)}$$

which implies $L=1$, similarly $K=1$.

[1] J. Hietarinta, J. Math. Phys. **28**, 1732 (1987).
[2] J. Hietarinta, J. Math. Phys. **28**, 2094 (1987).
[3] J. Hietarinta, J. Math. Phys. **28**, 2586 (1987).
[4] R. Hirota, J. Math. Phys. **14**, 805 (1973).
[5] J. Satsuma and M. J. Ablowitz, J. Math. Phys. **20**, 1496 (1979).
[6] A. Nakamura, J. Phys. Soc. Jpn. **52**, 3713 (1983).
[7] Y-C. Ma and L. G. Redekopp, Phys. Fluids **22**, 1872 (1979).
[8] M. Tajiri, J. Phys. Soc. Jpn. **52**, 2277 (1983).
[9] M. Tajiri, J. Phys. Soc. Jpn. **53**, 1221 (1984).
[10] M. Tajiri and M. Hagiwara, J. Phys. Soc. Jpn. **53**, 1634 (1984).
[11] Y-C. Ma, Stud. Appl. Math. **60**, 73 (1979).
[12] J. E. Herrera, J. Phys A: Math. Gen. **16**, L597 (1983).
[13] E. Benilov and S. Burtsev, Phys. Lett. A **98**, 256 (1983); P. Goldstein and L. Infeld, *ibid.* **103**, 8 (1984).
[14] R. Hirota, *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications*, edited by R. M. Miura (Springer, New York, 1976), p. 40.
[15] R. Hirota and J. Satsuma, Phys. Lett. A **85**, 407 (1981).
[16] Y. Matsuno, J. Phys. A: Math. Gen. **12**, 619 (1979).
[17] J. Satsuma and Y. Ishimori, J. Phys. Soc. Jpn. **46**, 681 (1979).
[18] H. H. Chen and Y. C. Lee, Phys. Rev. Lett. **43**, 264 (1979).
[19] Y. Matsuno, *Bilinear Transformation Method* (Academic, Orlando, 1984).
[20] Y. Matsuno, Phys. Lett. A **743**, 233 (1979).
[21] A. Ramani, B. Dorizzi and B. Grammaticos, Phys. Lett. A **99**, 411 (1983).
[22] Y. Matsuno, J. Math Phys. **29**, 49 (1988).
[23] M. Ito (private communication).
[24] B. Grammaticos and A. Ramani (private communication).
[25] A. C. Hearn, *REDUCE User's Manual Version 3.2* (Rand, Santa Monica, 1985), Publ. CP78, Rev. 4/85.

Complete orthonormal sets on the past light cone. I. Functions belonging to mass zero and helicity s

G. H. Derrick
School of Physics, University of Sydney, Sydney, NSW 2006, Australia[a] *and*
Department of Physics, University of St. Andrews, St. Andrews, Fife KY16 9SS, Scotland and
International Centre for Theoretical Physics, 34100 Trieste, Italy

(Received 6 October 1987; accepted for publication 28 October 1987)

This is the first of a series of papers preparing the mathematical framework for a past light-cone formulation for the quantum mechanics of particles of arbitrary mass and spin. The aim of past light-cone quantum theory is to define quantum states solely in terms of data accessible to an observer, i.e., information from within his current past light cone. In order to set up such a theory one needs to define on the past light cone complete orthonormal sets of functions that belong to the appropriate unitary irreducible representation of the Poincaré group. Such functions are interpreted as energy-momentum eigenfunctions. The present paper treats the discrete spin, zero mass case for all values of the helicity $s = 0,\frac{1}{2},\frac{3}{2},...$.

I. NOTATION AND CONVENTIONS

Alphabet conventions: Greek lowercase letters $= 0,1,2,3$, with summation over repeated indices.

Metric tensor: $g_{\lambda\mu} = \mathrm{diag}\,(1,-1,-1,-1)$.

Conjugation operations: A superscript $*$, T, $\dagger$ applied to a quantity denotes, respectively, the complex conjugate, transpose, Hermitian conjugate.

Number fields: $\mathbb{R}$ and $\mathbb{C}$ are, respectively, the real numbers and the complex numbers.

Dirac delta functions: $\delta(x)$ is the ordinary delta function of a real variable x, $\delta(\mathbf{x}) = \delta(x_1)\delta(x_2)\delta(x_3)$ that of a three-vector $\mathbf{x} = (x_1,x_2,x_3)$, and $\delta(\hat{\mathbf{k}},\hat{\mathbf{k}}') = \delta(\cos\theta_k - \cos\theta_{k'})\delta(\varphi_k - \varphi_{k'})$ the surface delta function for two unit vectors $\hat{\mathbf{k}} = (\sin\theta_k\cos\varphi_k,\sin\theta_k\sin\varphi_k,\cos\theta_k)$ and $\hat{\mathbf{k}}' = (\sin\theta'_k\cos\varphi'_k,\sin\theta'_k\sin\varphi'_k,\cos\theta'_k)$.

Generators of the Poincaré group: p^λ and $j^{\mu\kappa}$ are the momentum and the angular momentum operators, respectively, $\mathbf{J} = (j^{23},j^{31},j^{12})$, $\mathbf{K} = (j^{01},j^{02},j^{03})$.

Mass zero, helicity s representation of the Poincaré group[1]*:* The carrier space is the Hilbert space of square integrable functions $\psi(\mathbf{k})$ of a future pointing null vector $k^\lambda = k(1,\sin\theta_k\cos\varphi_k,\sin\theta_k\sin\varphi_k,\cos\theta_k)$ defined by the scalar product

$$(\psi,\psi') = \int \frac{d^3\mathbf{k}}{k}\,\psi^*\psi'.$$

The Poincaré generators are represented by

$$D(p^\lambda) = \hbar k^\lambda, \quad D(\mathbf{J}) = -i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}} + \hbar s\mathbf{T},$$
$$D(\mathbf{K}) = -i\hbar k\frac{\partial}{\partial\mathbf{k}} - \hbar s\hat{\mathbf{k}}\times\mathbf{T}, \tag{1}$$

where

$$\hat{\mathbf{k}} = \mathbf{k}/k,$$
$$\mathbf{T} = [\tan(\tfrac{1}{2}\theta_k)\cos\theta_k,\tan(\tfrac{1}{2}\theta_k)\sin\theta_k,1], \tag{2}$$

and

$$s = 0,\tfrac{1}{2},1,\tfrac{3}{2},... .$$

Mass m, spin-j representation of the Poincaré group[1]*:* For $j = 0,\frac{1}{2},1,\frac{3}{2},...$, the carrier space is the Hilbert space of functions $\psi_s(\mathbf{k})$, where s takes the $2j+1$ values $j, j-1,..., -j$ and k^λ is a future pointing vector on the upper mass shell $k^0 = \varepsilon_k = (\mathbf{k}^2 + m^2c^2/\hbar^2)^{1/2}$. The scalar product is

$$(\psi,\psi') = \sum_s \int \frac{d^3\mathbf{k}}{\varepsilon_k}\,\psi_s^*(\mathbf{k})\psi'_s(\mathbf{k}),$$

and the generator representations are

$$D(p^\lambda) = \hbar k^\lambda, \quad D(\mathbf{J}) = -i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}} + \hbar\mathbf{S}^j,$$
$$D(\mathbf{K}) = -i\hbar\varepsilon_k\frac{\partial}{\partial\mathbf{k}} - \frac{\hbar\mathbf{k}\times\mathbf{S}^j}{\varepsilon_k + mc/\hbar}. \tag{3}$$

Here $\mathbf{S}^j$ is the $(2j+1)\times(2j+1)$ Hermitian matrix representation of the SU(2) generators.

Pauli spin matrices σ^λ and $\bar{\sigma}^\lambda$:

$$\sigma^0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \sigma^1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$
$$\sigma^2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \sigma^3 = \begin{pmatrix}1 & \\ & -1\end{pmatrix}, \tag{4}$$
$$\bar{\sigma}^\lambda = \sigma^2(\sigma^\lambda)^*\sigma^2.$$

$D^{(1/2)0}$ representation of the SO(1,3) *generators:*

$$D^{(1/2)0}(\mathbf{J}) = \tfrac{1}{2}\hbar\boldsymbol{\sigma}, \quad D^{(1/2)0}(\mathbf{K}) = (i\hbar/2)\boldsymbol{\sigma}.$$

II. INTRODUCTION

A. Motivation

In special relativity an observer may be modeled by the following idealized picture. His path through space-time is represented by a timelike trajectory $x^\lambda = z^\lambda(\tau)$, where the $z^\lambda(\tau)$ are the four functions of a parameter τ. The latter is most conveniently taken to be the proper interval $\int(g_{\lambda\mu}\,dz^\lambda\,dz^\mu)^{1/2}$ measured along the trajectory from some arbitrary initial event. Thus $g_{\lambda\mu}v^\lambda v^\mu = 1$, $v^0 \geqslant 1$, where $v^\lambda = dz^\lambda(\tau)/d\tau$ is the observer's four-velocity vector. An ideal clock carried by the observer will then record the proper time τ/c that has elapsed since the initial event. Each value

[a] Permanent address.

 0022-2488/88/030636-06$02.50

of τ corresponds to a "here-now" $z^\lambda(\tau)$, at which the observer has knowledge only of events within his current past light cone, i.e., the region

$$x^0 < z^0(\tau) - |\mathbf{x} - \mathbf{z}(\tau)|. \tag{5}$$

Thus any events that the observer learns about at his current proper time τ/c lie on the past light cone

$$x^0 = z^0(\tau) - |\mathbf{x} - \mathbf{z}(\tau)|. \tag{6}$$

Introducing a past pointing null vector y^λ, (2) may be written in parametric form

$$x^\lambda = z^\lambda(\tau) + y^\lambda, \quad y^0 = -|\mathbf{y}|, \tag{7}$$

with the vector $\mathbf{y}$ taking all values in $\mathbb{R}^3$. The three-surface that the observer regards as "the present time τ/c" is actually the past light cone (7), not the hyperplane $x^0 = z^0(\tau)$. With increasing τ the events of space-time unfold as a succession of past light cones, and not as a sequence of spacelike hyperplanes $x^0 = \text{const}$.

Motivated by the above picture, a previous paper[2] developed a version of quantum mechanics, first suggested by Dirac,[3] in which an observer's quantum state refers to the past light cone of his current here now. This is in contrast to the conventional idea of a quantum state "at a given time." In this new approach a system containing one charged boson of rest mass m and spin zero is represented by an SO(1,3) scalar wave function $\psi(\mathbf{y},\tau)$. Physically acceptable wave functions are required to be elements of the Hilbert space $\mathscr{H}_y$, defined by the Lorentz invariant scalar product

$$(\psi_1,\psi_2)_y = \int \frac{d^3\mathbf{y}}{y} \psi_1^* \psi_2. \tag{8}$$

If during the interval τ_0 to τ the observer receives no data arising from measurements on the system, then the evolution is assumed to be unitary,

$$\psi(\mathbf{y},\tau) = \exp\{-(i/\hbar)[z^\lambda(\tau) - z^\lambda(\tau_0)]P_\lambda\}\psi(\mathbf{y},\tau_0), \tag{9}$$

where the generator of translations P_λ is an Hermitian four-vector operator defined on a dense subspace of $\mathscr{H}_y$, and satisfies $P^\lambda P_\lambda = m^2c^2$. The eigenfunctions of P^λ are of the form

$$\psi_{\mathbf{k}1}(\mathbf{y}) = (2\pi)^{-2} i \int_0^\infty d\sigma\, \sigma e^{ig(\sigma)}(-k_\lambda y^\lambda)^{i\sigma-1},$$
$$\psi_{\mathbf{k}(-1)}(\mathbf{y}) = [\psi_{\mathbf{k}1}(\mathbf{y})]^*, \tag{10}$$

where $g(\sigma)$ is an arbitrary real function, and k^λ any vector lying on the upper mass shell $k^0 = \varepsilon_k = [\mathbf{k}^2 + (mc/\hbar)^2]^{1/2}$. These functions satisfy completeness and orthogonality,

$$\int \frac{d^3\mathbf{y}}{y} \psi_{\mathbf{k}q}^*(\mathbf{y})\psi_{\mathbf{k}'q'}(\mathbf{y}) = \varepsilon_k \delta(\mathbf{k} - \mathbf{k}')\delta_{qq'},$$
$$\sum_q \int \frac{d^3\mathbf{k}}{\varepsilon_k} \psi_{\mathbf{k}q}(\mathbf{y})\psi_{\mathbf{k}q}^*(\mathbf{y}') = y\delta(\mathbf{y} - \mathbf{y}'), \tag{11}$$

with the charge index q assuming the two values ± 1. One may interpret $\psi_{\mathbf{k}q}(\mathbf{y})$ as an eigenfunction with momentum eigenvalue $\hbar k^\lambda$ and charge eigenvalue q. The momentum operator p^λ and charge operator Q in $\mathscr{H}_y$ are then given by

$$p^\lambda \psi(\mathbf{y}) = \sum_q \int \frac{d^3\mathbf{k}}{\varepsilon_k} \psi_{\mathbf{k}q}(\mathbf{y})\hbar k^\lambda(\psi_{\mathbf{k}q},\psi)_y,$$
$$Q\psi_y = \sum_q \int \frac{d^3\mathbf{k}}{\varepsilon_k} \psi_{\mathbf{k}q}(\mathbf{y})q(\psi_{\mathbf{k}q},\psi)_y. \tag{12}$$

The translation generator P^λ in (9) is identified either with p^λ or with Qp^λ, depending on which of two alternative hypotheses is made concerning the Hilbert space of physical states.[2]

B. Statement of the problem

The aim of this series of papers is to derive complete orthonormal sets that generalize the spin-zero results (10), (11) to other values of spin. To define the problem mathematically we need to recall some results from group representation theory as applied to the Poincaré group.

The functions $\psi_{\mathbf{k}q}(\mathbf{y})$ of (10) *belong to* the mass m, spin-zero unitary representation of the Poincaré group. Recall the following definition introduced by Wigner.[4] Suppose that for each element R of a group $\mathscr{G}$ there exists a representation by an operator P_R on some Hilbert space $\mathscr{H}$. If there is a set of elements $\psi_A \in \mathscr{H}$ such that

$$P_R \psi_A = \sum_B \psi_B D_{BA}(R), \tag{13}$$

where $\{D(R)\}$ is a representation of $\mathscr{G}$, then we say that the set ψ_A *belongs to* $\{D(R)\}$. If the index A takes values in a continuum then the summation in (13) is replaced by integration.

The set $\psi_{\mathbf{k}q}$ furnishes an example of (13), $\mathscr{G}$ being the Poincaré group and $\mathscr{H}$ being $\mathscr{H}_y$. In this case the representation $\{P_R\}$ is defined by the infinitesimal generators p^λ of (12) and the angular momentum operators $j^{\mu\kappa}$ given by

$$\mathbf{J} \equiv (j^{23},j^{31},j^{12}) = -i\hbar\mathbf{y} \times \frac{\partial}{\partial \mathbf{y}},$$
$$\mathbf{K} \equiv (j^{01},j^{02},j^{03}) = i\hbar y \frac{\partial}{\partial \mathbf{y}}. \tag{14}$$

The carrier space for the representation $\{D(R)\}$ is the Hilbert space $\mathscr{H}_k$ consisting of functions $\varphi(\mathbf{k})$ subject to the scalar product

$$(\varphi_1,\varphi_2) = \int \frac{d^3\mathbf{k}}{\varepsilon_k} \varphi_1^* \varphi_2. \tag{15}$$

As representative operators for momentum and angular momentum in $\mathscr{H}_k$ we have

$$D(p^\lambda) = \hbar k^\lambda, \quad D(\mathbf{J}) = -i\hbar\mathbf{k} \times \frac{\partial}{\partial \mathbf{k}},$$
$$D(\mathbf{K}) = -i\hbar\varepsilon_k \frac{\partial}{\partial \mathbf{k}}. \tag{16}$$

The set $\psi_{\mathbf{k}q}$ then satisfies

$$p^\lambda \psi_{\mathbf{k}q} = \hbar k^\lambda \psi_{\mathbf{k}q}, \quad \mathbf{J}\psi_{\mathbf{k}q} = i\hbar\mathbf{k} \times \frac{\partial}{\partial \mathbf{k}} \psi_{\mathbf{k}q},$$
$$\mathbf{K}\psi_{\mathbf{k}q} = i\hbar\varepsilon_k \frac{\partial}{\partial \mathbf{k}} \psi_{\mathbf{k}q}, \tag{17}$$

which corresponds to (13). Note the sign change in the derivative terms between (16) and (17), arising from the fact that (13) involves *right* multiplication by the operator $D(R)$. Since (17) holds for each value of the charge index q, $D(R)$ is actually the direct sum of two copies of the mass m spin-zero representation. [See (3).]

To summarize, what we have in the spin-zero case is a representation in $\mathscr{H}_y$ of Hermitian operators $y^\lambda, p^\lambda, j^{\mu\kappa}$ subject to the commutation relations

$$[y^\kappa, y^\lambda] = 0, \quad [j^{\mu\kappa}, y^\lambda] = i\hbar[g^{\kappa\lambda}y^\iota - g^{\iota\lambda}y^\kappa], \tag{18}$$

$$[p^\kappa, p^\lambda] = 0, \quad [j^{\iota\kappa}, p^\lambda] = i\hbar[g^{\kappa\lambda}p^\iota - g^{\iota\lambda}p^\kappa], \tag{19}$$

$$[j^{\iota\kappa}, j^{\lambda\mu}] = i\hbar[g^{\iota\mu}j^{\kappa\lambda} + g^{\kappa\lambda}j^{\iota\mu} - g^{\iota\lambda}j^{\kappa\mu} - g^{\kappa\mu}j^{\iota\lambda}], \tag{20}$$

and the constraints

$$y_\lambda y^\lambda = 0, \quad p_\lambda p^\lambda = m^2c^2; \tag{21}$$

with $-y^0, p^0$ having only positive eigenvalues. In this representation y^λ is diagonal and p^λ is defined by specifying the complete orthonormal set of eigenfunctions $\psi_{\mathbf{k}q}$ using the ansatz (12). These eigenfunctions belong to the mass m and spin-zero representation of the Poincaré group.

What we seek in this series of papers are representations of the system (18)–(21) that correspond to nonzero values of spin or helicity. Once again the operator y^λ is required to be diagonal, so that the carrier space $\mathscr{H}$ should be $\mathscr{H}_y$ or the direct sum of a finite number of copies of $\mathscr{H}_y$. To define the momentum operator p^λ we then need a set of states $|\mathbf{k}sq\rangle$ that are complete and orthonormal in $\mathscr{H}$ and belong to one of the unitary representations of the Poincaré group. To label these states one would expect the need for a momentum variable $\mathbf{k}$, some spin or helicity index s, and possibly additional labels q.

In the mass zero, helicity s case [see (1)] we require

$$\begin{aligned} \mathbf{J}|\mathbf{k}sq\rangle &= \left(i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}} + \hbar s\mathbf{T}\right)|\mathbf{k}sq\rangle, \\ \mathbf{K}|\mathbf{k}sq\rangle &= \left(i\hbar k\frac{\partial}{\partial\mathbf{k}} - \hbar s\hat{\mathbf{k}}\times\mathbf{T}\right)|\mathbf{k}sq\rangle, \end{aligned} \tag{22}$$

where $s = 0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \ldots$, $k = |\mathbf{k}| = k^0$, and $\hat{\mathbf{k}}$ and $\mathbf{T}$ are given by (2). The vector k^λ is thus future pointing and null, corresponding to zero mass.

In the case of nonzero mass m and spin j, (22) is replaced by

$$\begin{aligned} \mathbf{J}|\mathbf{k}sq\rangle &= i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}}|\mathbf{k}sq\rangle + \sum_{s'}|\mathbf{k}s'q\rangle\hbar\mathbf{S}^j_{s's}, \\ \mathbf{K}|\mathbf{k}sq\rangle &= i\hbar\varepsilon_k\frac{\partial}{\partial\mathbf{k}}|\mathbf{k}sq\rangle - \sum_{s'}|\mathbf{k}s'q\rangle\frac{\hbar\mathbf{k}\times\mathbf{S}^j_{s's}}{\varepsilon_k + mc/\hbar}, \end{aligned} \tag{23}$$

where k^λ now lies on the mass shell $k^0 = \varepsilon_k = (\mathbf{k}^2 + m^2c^2/\hbar^2)^{1/2}$. The spin j takes the values $0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$, with s ranging from $-j$ to $+j$ at integral steps, and $\mathbf{S}^j$ is the $(2j+1)\times(2j+1)$ matrix representation of the SU(2) generators. [See (3).]

If the Hilbert space relevant to (22) or (23) is the direct sum of N copies of $\mathscr{H}_y$ for some integer N, then the states $|\mathbf{k}sq\rangle$ may be represented by a set of N functions of a past pointing null vector y^λ. However it is not clear *a priori* what values of N will allow a solution of (22) or (23). Nor is it clear whether the form (14) for the angular momentum operators $\mathbf{J}$ and $\mathbf{K}$ remains appropriate. It will turn out to be necessary to modify (14) by the addition of helicitylike terms.

From now on in this paper, only mass-zero representations will be considered. The case of nonzero mass will be discussed in subsequent papers of the series.

III. MASS ZERO WITH HELICITY s

A. The spinor Hilbert spaces $\mathscr{H}_\xi$ and $\mathscr{H}_\nu$

Rather than work with the Hilbert spaces $\mathscr{H}_y$ and $\mathscr{H}_k$ it is simpler to introduce spinor Hilbert spaces similar to those used by Dirac[5] in another context.

Any further pointing null vector $k^\lambda = k(1, \sin\theta_k\cos\varphi_k, \sin\theta_k\sin\varphi_k, \cos\theta_k)$ may be written in terms of a contravariant $D^{(1/2)0}$ spinor ξ,

$$k^\lambda = \xi^\dagger\sigma^\lambda\xi, \tag{24}$$

with

$$\xi = \begin{bmatrix}\xi^1 + i\xi^2 \\ \xi^3 + i\xi^4\end{bmatrix} = k^{1/2}\begin{bmatrix}\cos\frac{1}{2}\theta_k \\ e^{i\varphi_k}\sin\frac{1}{2}\theta_k\end{bmatrix}e^{-i\eta_k}. \tag{25}$$

For a given k^λ the phase angle η_k may be chosen arbitrarily in $(0, 2\pi)$. We now introduce a Hilbert space $\mathscr{H}_\xi$, the set of all square integrable functions $\varphi(\xi^1,\xi^2,\xi^3,\xi^4)$ of the real and imaginary parts of the components of ξ, defined by the scalar product

$$\begin{aligned} (\varphi_1,\varphi_2)_\xi &= \int \varphi_1^*\varphi_2\, d\xi^1\, d\xi^2\, d\xi^3\, d\xi^4, \\ &= \frac{1}{8}\int \varphi_1^*\varphi_2\frac{d^3\mathbf{k}}{k}\,d\eta_k. \end{aligned} \tag{26}$$

The angular momentum tensor operator $j^{\iota\kappa}$ that arises from the $D^{(1/2)0}$ transformation law for ξ under SL(2,C) transformations is given by

$$\begin{aligned} (j^{23}_\xi, j^{31}_\xi, j^{12}_\xi) = \mathbf{J}_\xi &= -i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}} + \frac{1}{2}\hbar\mathbf{T}\left(-i\frac{\partial}{\partial\eta_k}\right), \\ (j^{01}_\xi, j^{02}_\xi, j^{03}_\xi) = \mathbf{K}_\xi &= -i\hbar k\frac{\partial}{\partial\mathbf{k}} - \frac{1}{2}\hbar(\hat{\mathbf{k}}\times\mathbf{T})\left(-i\frac{\partial}{\partial\eta_k}\right), \end{aligned} \tag{27}$$

with $\mathbf{T}$ as in (2). We have the commutators

$$[\mathbf{J}_\xi, \xi] = -\tfrac{1}{2}\hbar\sigma\xi, \quad [\mathbf{K}_\xi, \xi] = -(i/2)\hbar\sigma\xi. \tag{28}$$

In the subspace of functions of the form $\varphi(\mathbf{k})\exp(in\eta_k)$ with integral n, the angular momentum operator is equivalent to (1) with helicity $s = \frac{1}{2}n$.

Similarly, we can represent any past pointing null vector $y^\lambda = y(-1, \sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ in terms of a covariant $D^{1/20}$ spinor ν,

$$y^\lambda = -\nu\bar{\sigma}^\lambda\nu^\dagger.$$

[See (4).] Writing

$$\nu = [\nu_1 - i\nu_2, \nu_3 - i\nu_4] = y^{1/2}[e^{i\varphi}\cos\tfrac{1}{2}\theta, \sin\tfrac{1}{2}\theta]e^{i\eta}, \tag{29}$$

we then introduce a Hilbert space $\mathscr{H}_\nu$, the set of all square integrable functions $\psi(\nu_1,\nu_2,\nu_3,\nu_4)$ with scalar product

$$(\psi_1,\psi_2)_\nu = \int \psi_1^* \psi_2 \, d\nu_1 \, d\nu_2 \, d\nu_3 \, d\nu_4,$$
$$= \frac{1}{8} \int \psi_1^* \psi_2 \frac{d^3\mathbf{y}}{y} \, d\eta. \quad (30)$$

The analogs of (27) and (28) are (replacing k^λ by $-y^\lambda$ and η_k by $-\eta$),

$$\mathbf{J}_\nu = -i\hbar \mathbf{y} \times \frac{\partial}{\partial \mathbf{y}} + \frac{1}{2} \hbar \mathbf{W}\left(-i \frac{\partial}{\partial \eta}\right),$$
$$\mathbf{K}_\nu = i\hbar y \frac{\partial}{\partial \mathbf{y}} + \frac{1}{2} \hbar \hat{\mathbf{y}} \times \mathbf{W}\left(-i \frac{\partial}{\partial \eta}\right), \quad (31)$$

with

$$\hat{\mathbf{y}} = \mathbf{y}/y, \quad \mathbf{W} = [\cot(\tfrac{1}{2}\theta)\cos\varphi, \cot(\tfrac{1}{2}\theta)\sin\varphi, -1], \quad (32)$$

and

$$[\mathbf{J}_\nu, \nu] = \tfrac{1}{2}\hbar\nu\boldsymbol{\sigma}, \quad [\mathbf{K}_\nu, \nu] = (i/2)\hbar\nu\boldsymbol{\sigma}. \quad (33)$$

B. Orthonormal states for mass zero and helicity s

Consider now the complete orthonormal set of functions

$$\psi_\xi(\nu) = (2/\pi^2)\exp[i2^{1/2}(\nu\xi + \xi^\dagger\nu^\dagger)],$$
$$= (2/\pi^2)\exp[i2^{3/2}(\nu_1\xi^1 + \nu_2\xi^2 + \nu_3\xi^3 + \nu_4\xi^4)], \quad (34)$$

which span $\mathscr{H}_\nu$. We have

$$\int \psi_\xi^*(\nu)\psi_{\xi'}(\nu)d^4\nu = \prod_{A=1}^{4} \delta(\xi^A - \xi^{A'}),$$
$$\int \psi_\xi(\nu)\psi_\xi^*(\nu')d^4\xi = \prod_{A=1}^{4} \delta(\nu_A - \nu_A'). \quad (35)$$

We can obtain complete and orthonormal functions of $\mathbf{y}$ that belong to a definite helicity by writing (34) in terms of the variables $\mathbf{y},\eta,\mathbf{k},\eta_k$ and then projecting out the appropriate helicity components. From the definitions (25) and (29) one obtains

$$e^{-i(\eta-\eta_k)}\nu\xi$$
$$= (ky)^{1/2}(e^{i\varphi}\cos\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta_k + e^{i\varphi_k}\sin\tfrac{1}{2}\theta\sin\tfrac{1}{2}\theta_k),$$
$$|\nu\xi|^2 = \tfrac{1}{2}\zeta, \quad (36)$$

where

$$\zeta = -k_\lambda y^\lambda = ky + \mathbf{k}\cdot\mathbf{y}. \quad (37)$$

Thus we may write

$$\nu\xi = (\zeta/2)^{1/2}\exp[i\mu(\hat{\mathbf{y}},\hat{\mathbf{k}}) + i(\eta - \eta_k)], \quad (38)$$

with $\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})$ the real function of $\theta,\varphi,\theta_k,\varphi_k$ given by

$$e^{i\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})} = \{2/(1+\hat{\mathbf{k}}\cdot\hat{\mathbf{y}})\}^{1/2}$$
$$\times(e^{i\varphi}\cos\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta_k + e^{i\varphi_k}\sin\tfrac{1}{2}\theta\sin\tfrac{1}{2}\theta_k). \quad (39)$$

Substituting (38) into (34) yields

$$\psi_\nu(\xi) = \frac{2}{\pi^2}\exp\{2i\zeta^{1/2}\cos[\mu(\hat{\mathbf{y}},\hat{\mathbf{k}}) + \eta - \eta_k]\}$$
$$= \frac{4}{\pi}\sum_{n=-\infty}^{\infty} u_{\mathbf{k}n}(\mathbf{y})\exp\left\{in\left(\eta - \eta_k + \frac{1}{2}\pi\right)\right\}, \quad (40)$$

with

$$u_{\mathbf{k}n}(\mathbf{y}) = (1/2\pi)J_n(2\zeta^{1/2})e^{in\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})}. \quad (41)$$

If (40) is inserted into (35) the dependence on the angles η,η_k drops out and we arrive at the relations

$$\int u_{\mathbf{k}n}^*(\mathbf{y})u_{\mathbf{k}'n}(\mathbf{y})\frac{d^3\mathbf{y}}{y} = k\delta(\mathbf{k}-\mathbf{k}'),$$
$$\int u_{\mathbf{k}n}(\mathbf{y})u_{\mathbf{k}n}^*(\mathbf{y}')\frac{d^3\mathbf{k}}{k} = y\delta(\mathbf{y}-\mathbf{y}'). \quad (42)$$

That (41) defines a function belonging to helicity $s = \frac{1}{2}n$ may be seen as follows. A combined SL(2,ℂ) transformation of the spinors ξ and ν leaves $\psi_\xi(\nu)$ of (34) invariant, and hence

$$(j_\xi^{\mu\kappa} + j_\nu^{\mu\kappa})\psi_\xi(\nu) = 0. \quad (43)$$

Applying (27) and (31) to (40) then yields

$$\left(-i\hbar\mathbf{y}\times\frac{\partial}{\partial\mathbf{y}} + \frac{1}{2}\hbar n\mathbf{W}\right)u_{\mathbf{k}n}(\mathbf{y})$$
$$= \left(i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}} + \frac{1}{2}\hbar n\mathbf{T}\right)u_{\mathbf{k}n}(\mathbf{y}),$$
$$\left(i\hbar y\frac{\partial}{\partial\mathbf{y}} + \frac{1}{2}\hbar n\hat{\mathbf{y}}\times\mathbf{W}\right)u_{\mathbf{k}n}(\mathbf{y})$$
$$= \left(i\hbar k\frac{\partial}{\partial\mathbf{k}} - \frac{1}{2}\hbar n\hat{\mathbf{k}}\times\mathbf{T}\right)u_{\mathbf{k}n}(\mathbf{y}). \quad (44)$$

Thus the functions $u_{\mathbf{k}n}(\mathbf{y})$ defined by (39) and (41) satisfy all the requirements, viz. completeness, orthogonality, and definite helicity. Note that the carrier space for the representations of the (nonclosed) algebra (18)–(21) is $\mathscr{H}_y$, but the operators $\mathbf{J}$ and $\mathbf{K}$ of (22) now include helicity terms $\frac{1}{2}\hbar n\mathbf{W}$ and $\frac{1}{2}\hbar n\hat{\mathbf{y}}\times\mathbf{W}$ [see (44)]. The algebra (18)–(21) includes two Poincaré subalgebras $\{y^\lambda, j^{\mu\kappa}\}$ and $\{p^\lambda, j^{\mu\kappa}\}$, and for the representation associated with $u_{\mathbf{k}n}(\mathbf{y})$ both $\hat{\mathbf{y}}\cdot\mathbf{J}/\hbar$ and $(p^0)^{-1}\mathbf{p}\cdot\mathbf{J}/\hbar$ have the eigenvalue $s = \frac{1}{2}n$.

An explicit form can be given for the momentum operator p^λ that has eigenfunctions $u_{\mathbf{k}n}(\mathbf{y})$ and eigenvalues $\hbar k^\lambda$. Consider the Hermitian operator π^λ on $\mathscr{H}_\nu$ defined by

$$\pi^\lambda = -(\hbar/2)\partial_\nu^\dagger\sigma^\lambda\partial_\nu, \quad (45)$$

where ∂_ν is the contravariant $D^{(1/2)0}$ differential operator

$$\partial_\nu = \frac{1}{2}\begin{bmatrix} \dfrac{\partial}{\partial\nu_1} + i\dfrac{\partial}{\partial\nu_2} \\ \dfrac{\partial}{\partial\nu_3} + i\dfrac{\partial}{\partial\nu_4} \end{bmatrix}. \quad (46)$$

Applying π^λ to the functions $\psi_\xi(\nu)$ of (34) yields

$$\pi^\lambda\psi_\xi(\nu) = \hbar k^\lambda\psi_\xi(\nu). \quad (47)$$

Note that the operator

$$S = \tfrac{1}{2}\nu\,\partial_\nu - \tfrac{1}{2}(\nu\,\partial_\nu)^*$$
$$\equiv -\frac{i}{2}\frac{\partial}{\partial\eta}, \quad (48)$$

commutes with π^λ, so that the expansion (40) represents a decomposition of $\psi_\xi(\nu)$ according to the eigenvalues $s = \frac{1}{2}n$ of S. Thus

$$(\pi^\lambda - \hbar k^\lambda)u_{\mathbf{k}n}(\mathbf{y})e^{in\eta} = 0, \quad (S - \tfrac{1}{2}n)u_{\mathbf{k}n}(\mathbf{y})e^{in\eta} = 0. \quad (49)$$

In terms of the coordinates $\mathbf{y},\eta$ the operator π^λ takes the form

$$\pi^\lambda = \pi^\lambda_{(0)} + \pi^\lambda_{(1)} S + \pi^\lambda_{(2)} S^2, \tag{50}$$

with

$$\pi^\lambda_{(0)} = \hbar\left[-\frac{1}{2} y \frac{\partial^2}{\partial \mathbf{y}^2}, \frac{1}{2}\mathbf{y}\frac{\partial^2}{\partial \mathbf{y}^2} - \frac{\partial}{\partial \mathbf{y}}\left(\mathbf{y}\cdot\frac{\partial}{\partial \mathbf{y}}\right)\right],$$

$$\pi^\lambda_{(1)} = \frac{\hbar}{2}\left[\frac{i}{y\sin^2(\frac{1}{2}\theta)}\frac{\partial}{\partial\varphi}, -2i\mathbf{W}\times\frac{\partial}{\partial \mathbf{y}}\right], \tag{51}$$

$$\pi^\lambda_{(2)} = \frac{\hbar}{2y\sin^2(\frac{1}{2}\theta)}[1,0,0,-1].$$

Whence from (49)

$$p^\lambda u_{\mathbf{k}n}(\mathbf{y}) = \hbar k^\lambda u_{\mathbf{k}n}(\mathbf{y}), \tag{52}$$

where

$$p^\lambda = \pi^\lambda_{(0)} + \tfrac{1}{2} n \pi^\lambda_{(1)} + \tfrac{1}{4} n^2 \pi^\lambda_{(2)} \tag{53}$$

is a Hermitian operator in $\mathscr{H}_y$. It is a vector operator satisfying (19), wth $j^{\mu\kappa}$ given by the operators on the left-hand side of (44).

C. Complete orthonormal sets obtained by unitary transformations

We can now derive other complete orthonormal sets by making unitary transformations of the $u_{\mathbf{k}n}(\mathbf{y})$ given by (41). Writing the Bessel function as an integral[6] (41) becomes

$$u_{\mathbf{k}n}(\mathbf{y}) = \frac{e^{in\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})}}{4\pi^2} \times \int_{-\infty}^{\infty} d\sigma\left(\frac{1}{2}n - i\sigma\right)\frac{\Gamma(\frac{1}{2}n - i\sigma)}{\Gamma(\frac{1}{2}n + i\sigma)}\zeta^{i\sigma-1}. \tag{54}$$

We now show that if the phase factor $\Gamma(\frac{1}{2}n - i\sigma)/\Gamma(\frac{1}{2}n + i\sigma)$ in (54) is replaced by a arbitrary phase factor $\exp[ig(\sigma)]$, $g(\sigma)$ real, then one again obtains a complete orthonormal set satisfying (42) and (44). This is most readily seen by noticing that

$$u'_{\mathbf{k}n}(\mathbf{y}) = \frac{e^{in\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})}}{4\pi^2} \times \int_{-\infty}^{\infty} d\sigma\left(\frac{1}{2}n - i\sigma\right)e^{ig(\sigma)}\zeta^{i\sigma-1}, \tag{55}$$

is related to $u_{\mathbf{k}n}(\mathbf{y})$ by the unitary transformation

$$u'_{\mathbf{k}n}(\mathbf{y}) = e^{ig(D)}\frac{\Gamma(\frac{1}{2}n + iD)}{\Gamma(\frac{1}{2}n - iD)}u_{\mathbf{k}n}(\mathbf{y}), \tag{56}$$

where

$$D = -i\left(\mathbf{y}\cdot\frac{\partial}{\partial \mathbf{y}} + 1\right) = -(i/2)\,[\nu\,\partial_\nu + (\nu\,\partial_\nu)^* + 2]. \tag{57}$$

Note that D is a Hermitian operator (in both $\mathscr{H}_y$ and $\mathscr{H}_\nu$) that commutes with $j^{\mu\kappa}_\nu$ and S [see (31) and (48)] and satisfies the eigenvalue equation

$$(D - \sigma)\zeta^{i\sigma-1} = 0. \tag{58}$$

Particular examples of (55) follow:

(1) $\exp[ig(\sigma)] = \pi^{-1/2}\Gamma(\frac{1}{2} - i\sigma) \times[\cosh(\frac{1}{2}\pi\sigma) - i\sinh(\frac{1}{2}\pi\sigma)]$,

$$u'_{\mathbf{k}n}(\mathbf{y}) = \frac{-e^{in\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})}}{(2\pi^3)^{1/2}}\zeta^{(1/2)n} \times\frac{d}{d\zeta}(\zeta^{1/2-(1/2)n}\sin\zeta). \tag{59}$$

(2) $\exp[ig(\sigma)] = \pi^{-1/2}\Gamma(\frac{1}{2} - i\sigma) \times[\cosh(\frac{1}{2}\pi\sigma) + i\sinh(\frac{1}{2}\pi\sigma)]$,

$$u'_{\mathbf{k}n}(\mathbf{y}) = \frac{e^{in\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})}}{(2\pi^3)^{1/2}}\zeta^{(1/2)n} \times\frac{d}{d\zeta}(\zeta^{1/2-(1/2)n}\cos\zeta). \tag{60}$$

(3) $g(\sigma) = 0$,

$$u'_{\mathbf{k}n}(\mathbf{y}) = \frac{-e^{in\mu(\hat{\mathbf{y}},\hat{\mathbf{k}})}}{2\pi}\zeta^{(1/2)n+1}\frac{d}{d\zeta}\delta(\zeta - 1). \tag{61}$$

D. Further complete orthonormal sets

A class of complete orthonormal sets of quite a different type may be derived from unitary transformation of

$$\chi_{\mathbf{k}}(\mathbf{y}) = l^2 y\delta(\mathbf{y} + l^2\mathbf{k}), \tag{62}$$

where l is an arbitrary constant of dimensions length.

We have the relations

$$\int\frac{d^3\mathbf{y}}{y}\chi^*_{\mathbf{k}}(\mathbf{y})\chi_{\mathbf{k}'}(\mathbf{y}) = k\delta(\mathbf{k} - \mathbf{k}'),$$
$$\int\frac{d^3\mathbf{k}}{k}\chi_{\mathbf{k}}(\mathbf{y})\chi^*_{\mathbf{k}}(\mathbf{y}') = y\delta(\mathbf{y} - \mathbf{y}'), \tag{63}$$

$$\left(-i\hbar\mathbf{y}\times\frac{\partial}{\partial\mathbf{y}} - \frac{1}{2}\hbar n\mathbf{W}\right)\chi_{\mathbf{k}}(\mathbf{y}) = \left(i\hbar\mathbf{k}\times\frac{\partial}{\partial\mathbf{k}} + \frac{1}{2}\hbar n\mathbf{T}\right)\chi_{\mathbf{k}}(\mathbf{y}),$$
$$\left(i\hbar y\frac{\partial}{\partial\mathbf{y}} - \frac{1}{2}\hbar n\hat{\mathbf{y}}\times\mathbf{W}\right)\chi_{\mathbf{k}}(\mathbf{y}) = \left(i\hbar k\frac{\partial}{\partial\mathbf{k}} - \frac{1}{2}\hbar n\hat{\mathbf{k}}\times\mathbf{T}\right)\chi_{\mathbf{k}}(\mathbf{y}). \tag{64}$$

Note that (64) holds in a trivial way for *any* value of n, and that the helicity term on the left-hand side has the opposite sign to that in (44). A whole class of orthonormal functions $\chi'_{\mathbf{k}}(\mathbf{y})$ may now be defined by a procedure analogous to that of (56). Let $g(\sigma)$ be an arbitrary real function of a real variable σ, and D the Hermitian operator on $\mathscr{H}_y$ defined by (57). Now define

$$\chi'_{\mathbf{k}}(\mathbf{y}) = e^{ig(D)}\chi_{\mathbf{k}}(\mathbf{y}) = \frac{\delta(\hat{\mathbf{y}}, -\hat{\mathbf{k}})}{2\pi k^2 l^2}\int_{-\infty}^{\infty}d\sigma\left(\frac{y}{kl^2}\right)^{i\sigma-1}e^{ig(\sigma)}. \tag{65}$$

The delta function in (65) is the surface delta function for a unit sphere. Particular examples of (65) follow:

(1) $\exp[ig(\sigma)] = \pi^{-1/2}\Gamma(\frac{1}{2} - i\sigma) \times[\cosh(\frac{1}{2}\pi\sigma) - i\sinh(\frac{1}{2}\pi\sigma)]$,

$$\chi'_{\mathbf{k}}(\mathbf{y}) = \delta(\hat{\mathbf{y}}, -\hat{\mathbf{k}})(2/\pi y k^3 l^2)^{1/2} \sin(y/kl^2), \tag{66}$$

(2) $$\exp[ig(\sigma)] = \Gamma(\tfrac{1}{2} + i\sigma) \times [\cosh(\tfrac{1}{2}\pi\sigma) + i \sinh(\tfrac{1}{2}\pi\sigma)],$$

$$\chi'_{\mathbf{k}}(\mathbf{y}) = \delta(\hat{\mathbf{y}}, -\hat{\mathbf{k}})(2l^2/\pi y^3 k)^{1/2} \sin(kl^2/y). \tag{67}$$

IV. SUMMARY AND CONCLUDING REMARKS

The functions $u_{\mathbf{k}n}(\mathbf{y})$, $u'_{\mathbf{k}n}(\mathbf{y})$, $\chi_{\mathbf{k}}(\mathbf{y})$, $\chi'_{\mathbf{k}}(\mathbf{y})$ given by (54), (55), (62), and (65), respectively, are complete and orthonormal sets of functions of $\mathbf{y}$. Any of these alternative sets can be used as a basis in $\mathscr{H}_y$ the Hilbert space defined by (8). An important property of these sets is that they belong to the unitary irreducible representation of the Poincaré group with rest mass zero and helicity $\frac{1}{2}n$. [See (44), (64).] This suggests the possibility of formulating a past light-cone quantum mechanics of lightlike quanta. Thus a theory of neutrinos and antineutrinos might be based on $u_{\mathbf{k}(\pm 1)}(\mathbf{y})$, and a theory of photons on $u_{\mathbf{k}(\pm 2)}(\mathbf{y})$, just as a theory for spinless bosons of nonzero mass can be based on the $\psi_{\mathbf{k}q}(\mathbf{y})$ of (10). Such theories will be explored in later papers of this series.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste.

[1]E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1948); L. L. Foldy, Phys. Rev. **102**, 568 (1956); J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962).

[2]G. H. Derrick, J. Math. Phys. **28**, 64, 1327 (1987).

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

[4]E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Chap. 12.

[5]P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963).

[6]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Natl. Bur. Stand., Washington, DC, 1965), 9.1.26.

Quantum systems with external electromagnetic fields: The large mass asymptotics

L. Papiez and T. A. Osborn
Department of Physics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
F. H. Molzahn
Department of Physics, University of California, Berkeley, California 94720

(Received 3 June 1987; accepted for publication 30 September 1987)

The large mass asymptotics of the quantum evolution problem for a system of charged particles that mutually interact through scalar fields and couple to an arbitrary time-varying external electromagnetic field is rigorously described. If $K(x,t;y,s;m)$ denotes the coordinate space propagator (time evolution kernel) of this system, the singular perturbation behavior of K as mass $m\to\infty$ is expressed in terms of a gauge invariant asymptotic expansion. In terms of the external fields and interparticle interactions, this expansion provides a nonperturbative approximation for the propagator K that is valid for all particle coordinates x, y and for finite time displacements $t-s$. For the class of analytic scalar and vector fields that are defined as Fourier transforms of time-dependent measures, the existence of this asymptotic series for K in powers of $(m)^{-1}$ is established for both real and complex masses. Explicit bounds for the error term are obtained and a manifestly gauge invariant transport recurrence relation is derived that uniquely determines all the coefficient functions of the asymptotic series. The small time asymptotic expansion of K is shown to be embedded within the large mass expansion.

I. INTRODUCTION

The time-dependent Hamiltonian of an N-body quantum system of spinless nonrelativistic particles, each having mass m and charge q, that mutually interact through scalar fields and couple via the Lorentz force to an external electromagnetic field is given by

$$H(x,p,t,m)=(2m)^{-1}[p-qa(x,t)]^2 + q\phi(x,t)+V(x,t). \tag{1.1}$$

Here (x,t) is the space-time point in the $(d+1)$-dimensional Euclidean space that specifies the generic position of the particles of the system at time t. If the individual particles move in three dimensions then $d=3N$. It is assumed that the time t takes values in the interval $[0,T]$. The symbol p represents the momentum operator $-i\hbar\nabla$ conjugate to x. The vector and scalar potentials that are responsible for the interaction with the external electromagnetic field are denoted by $a: \mathbb{R}^d\times[0,T]\to\mathbb{R}^d$ and $\phi: \mathbb{R}^d\times[0,T]\to\mathbb{R}$, respectively. The mutual interaction of all N particles, and their interaction with other possible forces, is described by potential V.

Consider the propagator (evolution kernel) K for the system (1.1). If $(y,s)\in\mathbb{R}^d\times[0,T]$ is an arbitrary initial space-time point, then the propagator is a distribution-valued solution of the time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}K(x,t;y,s;m)=H(x,-i\hbar\nabla,t,m)K(x,t;y,s;m), \tag{1.2}$$

that satisfies the delta-function initial condition

$$K(x,s;y,s;m)=\delta(x-y). \tag{1.3}$$

The large mass limit is one mechanism through which quantum systems exhibit classical-like behavior. In this paper we obtain a detailed description of the analytic behavior of the propagator K for both real and complex masses. In the case of forward time evolution, the mass parameter m is restricted to take values in the upper-half complex plane. The open (closed) upper-half plane is denoted by $\mathbb{C}_>$ ($\overline{\mathbb{C}}_>$) and $\mathbb{C}_+$ represents the closed upper-half plane with the origin deleted, $\mathbb{C}_+=\overline{\mathbb{C}}_>\setminus\{0\}$. The mass behavior in $\mathbb{C}_+$ is then utilized to find simple nonperturbative approximations for K. There is no loss of generality in the assumption of a common mass m for all N particles since a coordinate scale change transforms a Hamiltonian with different particle masses into one with the form (1.1). Similarly the effect of different charge coupling constants for each particle can be absorbed into the definitions of a and ϕ.

Stated in general terms, the large mass expansion we find takes the following form. The kernel K is shown to admit the factorization

$$K(x,t;y,s;m)=K_0(t-s;x-y;m)F(x,t;y,s;m^{-1}), \quad m\in\mathbb{C}_+, \tag{1.4}$$

where K_0 is the well-known free evolution kernel (i.e., the propagator for the Hamiltonian with $a=0$ and $\phi=V=0$),

$$K_0(t-s;x-y;m)=[m/2\pi i\hbar(t-s)]^{d/2}\times\exp[im(x-y)^2/(2\hbar(t-s))] \tag{1.5}$$

and F turns out to be a smooth bounded function, for each allowed $(x,t;y,s)$ as $|m|\to\infty$ in $\mathbb{C}_+$. This implies that K and K_0 have exactly the same essential singularity at $|m|=\infty$. In addition, the function F has the large $m\in\mathbb{C}_+$ asymptotic expansion

$$F\sim\{\exp[(i\hbar)^{-1}J(x,t;y,s)]\}\{1+m^{-1}T_1(x,t;y,s)$$

0022-2488/88/030642-18$02.50

$$+ m^{-2}T_2(x,t;y,s) + \cdots\}. \tag{1.6}$$

The phase factor J is

$$J(x,t;y,s) = \int_0^1 d\xi\{(t-s)[q\phi + V](w(\xi)) - (x-y)\cdot qa(w(\xi))\}. \tag{1.7}$$

In this expression w is the linear path in $\mathbb{R}^d \times [0,T]$ connecting the initial space-time point (y,s) to the final point (x,t),

$$w(\xi) = w(\xi;x,t;y,s) = (y + \xi(x-y), s + \xi(t-s)), \quad \xi\in[0,1]. \tag{1.8}$$

Clearly, J is real valued and independent of both m and $\hbar$. The functions T_j turn out to be gauge invariant and are uniquely determined as solutions of a transport recurrence relation that is associated with the linear path w. The gauge dependence in the propagator is carried entirely by the mass-independent phase factor J. The appearance of a, ϕ, and V in the exponential factor J illustrates the nonperturbative nature of the approximation (1.4)–(1.7).

Expansion (1.6) was recently derived[1] in a heuristic fashion by implementing a large mass expansion of the higher-order Wentzel–Kramers–Brillouin (WKB) approximation for the propagator K. The objective of this paper is to obtain the large mass asymptotics described in Eqs. (1.4)–(1.7), rigorously, for a sufficiently smooth class of potentials a, ϕ, and V.

The method of solution, devised for this problem, is to employ a constructive representation of the propagator K. For the class of potentials that can be represented as the Fourier transforms of complex-valued time-dependent measures, a convergent infinite series expression[2] for K is known. Section II reviews the status of the operator-valued and kernel-valued solutions of the quantum evolution problem for Hamiltonian (1.1). Those features of the constructive representation of K needed in this investigation are outlined. In Sec. III the factorization property (1.4) is verified. The boundedness of F in the neighborhood of $|m| = \infty$ is established. We prove that m^{-1} is the appropriate small expansion parameter of F. Furthermore, if the asymptotic expansion (1.6) is carried out to an arbitrary order M, we obtain bounds for the remainder term that describes the total error. In Sec. IV the phase factor J is derived by summing all the mass-independent parts of the constructive representation of K. Finally a manifestly gauge invariant recurrence relation is obtained for the coefficient functions T_j, from which T_1 and T_2 are computed. Section V summarizes our conclusions and gives the physical interpretation of representation (1.4)–(1.8) that is applicable if the external fields are solutions of Maxwell's equations.

II. THE COMPLEX MASS PROPAGATOR: DEFINITIONS AND KNOWN RESULTS

In this section the constructive description of the propagator is recounted. Precise definitions of the operator-valued and kernel-valued solutions of the evolution problem are presented. In particular, this section defines the numerous quantities that enter the constructive formulas for K. A class of Fourier image potentials is discussed. For these potentials one can prove that the kernel-valued Dyson series[3,4] gives, via the complex mass extension method,[2] an explicit series representation of the propagator. Finally, the behavior of the propagator with respect to the $U(1)$ gauge group is discussed.

The Hilbert space of square integrable functions on $\mathbb{R}^d$ is indicated by $\mathscr{H} = L^2(\mathbb{R}^d)$. The identity operator on $\mathscr{H}$ will be I, and our notation for the inner product (defined to be antilinear in the left argument) is $\langle\cdot,\cdot\rangle$. In the following analysis certain restrictions are placed on the vector and scalar potentials that ensure that the operator $H(x, -i\hbar\nabla,t,m)$ has a unique closed extension, $H(t,m)$. Furthermore, these restrictions will imply that the (dense) domain of $H(t,m)$ is independent of t, i.e., $D(H(t,m)) = D_0 \subset \mathscr{H}$ for all $t\in[0,T]$. Note that for complex masses, the Hamiltonian operators $H(t,m)$ are not generally self-adjoint unless $\text{Im } m = 0$. Finally the symbols T_Δ and (T_Δ^0) denote the closed (and open) two-dimensional time regions $\{(t,s)\in\mathbb{R}^2\colon 0\leqslant s\leqslant t\leqslant T\}$ and $\{(t,s)\in\mathbb{R}^2\colon 0<s<t<T\}$.

The abstract ($\mathscr{H}$-valued) evolution problem[5] in T_Δ takes the following form. A function $\psi\colon [s,T]\to\mathscr{H}$ is said to be a solution of

$$i\hbar\dot{\psi}(t) = H(t,m)\psi(t), \tag{2.1}$$

if ψ takes values in D_0, possesses a strong derivative $\dot{\psi}$ throughout the interval $[s,T]$, and satisfies (2.1) for all $t\in[s,T]$. Suppose f is an arbitrary function chosen from D_0 and s is the time at which the initial data condition is imposed. The *Cauchy problem in the triangle* T_Δ is the problem of finding, for each fixed $s\in[0,T]$, a solution $\psi(\cdot,s)$ of (2.1) on the interval $[s,T]$ that satisfies the initial condition

$$\psi(s,s) = f. \tag{2.2}$$

Consider the description of the solution to the Cauchy problem in T_Δ in terms of an evolution operator. Let $\mathscr{B}(\mathscr{H})$ be the Banach space (with operator norm $\|\cdot\|$) of all bounded operators mapping $\mathscr{H}$ into $\mathscr{H}$. The evolution map $f\mapsto\psi(t,s)$ defines a linear operator from D_0 into $\mathscr{H}$. The extension of this operator to $\mathscr{H}$ is defined to be the evolution operator $U(t,s;m)$. In greater detail, for each fixed value of $m\in\mathbb{C}_+$, one has the statement.

Definition 1: A two-parameter operator-valued function $U\colon T_\Delta\to\mathscr{B}(\mathscr{H})$ is said to be the *Schrödinger evolution* generated by $\{H(t,m)\colon t\in[0,T]\}$ if the following holds.

(1) For $(t,s)\in T_\Delta$, $U(t,s;m)$ maps the domain D_0 into itself. (2.3a)

(2) U is uniformly bounded in T_Δ and for some positive finite c,

$$\|U(t,s;m)\|\leqslant\exp[c(t-s)], \quad t>s. \tag{2.3b}$$

(3) U is strongly continuous in T_Δ.

(4) The following identities hold in $\mathscr{B}(\mathscr{H})$:

$$U(t,s;m) = U(t,\tau;m)U(\tau,s;m), \quad 0\leqslant s\leqslant\tau\leqslant t\leqslant T, \tag{2.3c}$$

$$U(s,s;m) = I, \quad s\in[0,T]. \tag{2.3d}$$

(5) On the domain D_0, U is strongly continuously differentiable relative to t and s. Furthermore, U satisfies the equations of motion on T_Δ^0,

$$i\hbar\frac{\partial}{\partial t}U(t,s;m)f = H(t,m)U(t,s;m)f, \quad f\in D_0, \tag{2.3e}$$

$$-i\hbar\frac{\partial}{\partial s}U(t,s;m)f=U(t,s;m)H(s,m)f,\quad f\in D_0. \tag{2.3f}$$

Now we introduce the family of potentials that are the Fourier transforms of complex bounded measures on $\mathbb{R}^d$. First, it is notationally simpler to combine the two scalar fields ϕ and V into a single potential, viz.,

$$v(x,t)=q\phi(x,t)+V(x,t). \tag{2.4}$$

At this point the charge coupling constant q will be taken into the definition of a so that qa in (1.1) is replaced by a. The vector and combined scalar potentials are assumed to have the general form,

$$a(x,t)=\int e^{i\alpha\cdot x}\,d\gamma(t), \tag{2.5a}$$

$$v(x,t)=\int e^{i\alpha\cdot x}\,d\nu(t). \tag{2.5b}$$

In these integrals the measures $\gamma(t)$ and $\nu(t)$ are time dependent, while the variable of integration $\alpha\in\mathbb{R}^d$ (the wave vector) is *not* displayed in the measure symbol $d\gamma(t)$ or $d\nu(t)$.

Our measures [$\gamma(t)$ and $\nu(t)$] will be chosen from the Banach spaces $\mathscr{M}(\mathbb{R}^d,\mathbb{C}^r)$, ($r=d$ or 1) of $\mathbb{C}^r$-valued Borel measures γ on $\mathbb{R}^d$, which have complex-valued Fourier images (2.5a) and (2.5b). The space $\mathscr{M}^*(\mathbb{R}^d,\mathbb{C}^r)$ is the subspace of $\mathscr{M}(\mathbb{R}^d,\mathbb{C}^r)$ whose images are real valued. The norm $\|\cdot\|$ for $\mathscr{M}(\mathbb{R}^d,\mathbb{C}^r)$ and $\mathscr{M}^*(\mathbb{R}^d,\mathbb{C}^r)$ is defined using the total variation measure $|\gamma|$, via

$$\|\gamma\|\equiv|\gamma|(\mathbb{R}^d)<\infty. \tag{2.6}$$

The same symbol $\|\cdot\|$ is used as the norm for a variety of different spaces. The context will determine its correct meaning. An additional restriction on the measures is the requirement that they have compact support. Let $S_k\subset\mathbb{R}^d$ be the closed ball of radius k and center at 0. Then $\mathscr{M}^*(S_k,\mathbb{C}^r)$ will denote the Banach subspace of measures in $\mathscr{M}^*(\mathbb{R}^d,\mathbb{C}^r)$ that have their support contained by S_k.

A time-dependent measure is defined by the map

$$\gamma(\cdot):\ [0,T]\to\mathscr{M}^*(\mathbb{R}^d,\mathbb{C}^r).$$

From this point of view, $\gamma(t)$ is a Banach-space-valued function of t. In the space $\mathscr{M}^*(\mathbb{R}^d,\mathbb{C}^r)$ one has the conventional definitions[2] of continuity and differentiability with respect to $\|\cdot\|$. The symbol $\dot{\gamma}(t)$ denotes the derivative of $\gamma(\cdot)$ at t. With this terminology in place we may state the hypothesis on the potentials that is required for the remainder of this paper.

Potential class (A): Let $k<\infty$. The potentials a and v are said to be in the class (A) if a and v are the Fourier images, Eqs. (2.5a) and (2.5b), of time-dependent measures $\gamma(\cdot)$ and $\nu(\cdot)$ satisfying

(1) $\gamma(t)\in\mathscr{M}^*(S_{k/2},\mathbb{C}^d)$, $t\in[0,T]$,

(2) $\nu(t)\in\mathscr{M}^*(S_k,\mathbb{C})$, $t\in[0,T]$,

(3) both $\gamma(\cdot)$ and $\nu(\cdot)$ are continuously differentiable on $[0,T]$.

Hereafter the hypothesis that a and v are in (A) will always be assumed and so will not usually be cited as a part of the various lemmas and theorems. The functions $a(\cdot,t)$ and $v(\cdot,t)$ in class (A) are $\mathbb{R}^d$ and $\mathbb{R}$-valued analytic functions. The requirement that $k<\infty$ means that the electric and magnetic fields have a space frequency cutoff k. The portion of our analysis that forces us to adopt class (A) is the constructive determination of the propagator K. The Schrödinger evolution operators $U(t,s;m)$ are known to exist[5,6] for a much wider class of potentials.

Some useful constants related to a and v that often appear in the subsequent estimates are

$$\nu_T=\sup\|\nu(t)\|,\quad \gamma_T=\sup\|\gamma(t)\|,\quad \dot{\gamma}_T=\sup\|\dot{\gamma}(t)\|, \tag{2.7}$$

where each supremum is taken over $t\in[0,T]$. For a more complete discussion of the measures $\gamma(t)$ and $\nu(t)$ and their properties consult Ref. 2, Sec. II.

Three immediate consequences of hypothesis (A) are (for all $m\in\mathbb{C}_+$, $t\in[0,T]$) (1) that $H(x,-i\hbar\nabla,t,m)$, interpreted as the minimal operator on $C_0^\infty(\mathbb{R}^d)$, has a unique closed extension $H(t,m)$ in $\mathscr{H}$; (2) that the domain of $H(t,m)$ is time independent and is the same domain D_0 as that of the self-adjoint extension of the Laplacian; and (3) that $H(t,m)$ is strongly continuously differentiable in t on D_0. Given the validity of these three properties for the family of Hamiltonian operators $\{H(t,m):\ t\in[0,T]\}$, one can adapt without difficulty the general theory[5,6] of evolution equations in Banach space, with unbounded operator coefficients, to obtain the existence of the complex mass evolution operator satisfying all the properties of Definition 1. For details of the proof see Ref. 2, Theorem 2.

Theorem 1: For each $m\in\mathbb{C}_+$, the family of Hamiltonian operators $\{H(t,m):t\in[0,T]\}$ generates a Schrödinger evolution operator $U(\cdot,\cdot;m):T_\Delta\to\mathscr{B}(\mathscr{H})$.

It is often the case in physical problems that the bounded evolution operators $U(t,s;m)$ turn out to be represented in terms of an integral kernel.[7] For the system (1.1) an appropriate definition of the propagator is as follows.

Definition 2: Fix $m\in\mathbb{C}_+$. A two-parameter family (in T_Δ^0) of functions $K(\cdot,t;\cdot,s;m):\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ that are measurable and locally integrable on $\mathbb{R}^d\times\mathbb{R}^d$ is called the *propagator* for evolution $\{U(t,s;m):\ (t,s)\in T_\Delta^0\}$ if for all $f\in L_0^\infty(\mathbb{R}^d)$,

$$[U(t,s;m)f](x)=\int K(x,t;y,s;m)f(y)dy,\quad \text{a.a. } x. \tag{2.8}$$

The notation $L_0^p(\mathbb{R}^d)$ denotes the L^p functions of compact support. Observe that (2.8) determines $U(t,s;m)f$ for all $f\in\mathscr{H}$. The space $L_0^\infty(\mathbb{R}^d)$ is a dense subset of $\mathscr{H}$. Thus each $f\in\mathscr{H}$ is the strong limit of a sequence $\{f_i\}\subset L_0^\infty(\mathbb{R}^d)$. The value of $U(t,s;m)f$ is then given by

$$U(t,s;m)f=\operatorname*{s-lim}_{i\to\infty}\int K(\cdot,t;y,s;m)f_i(y)dy,$$

which holds for all $f\in\mathscr{H}$.

Definition 2 of the propagator as a type of integral kernel is structured to deal with the difficulties of interpreting the free propagator as a kernel. The function $K_0(t-s;x-y;m)$ is bounded but has no decay as $|x-y|\to\infty$. As a consequence, for an arbitrary wave function $f\in\mathscr{H}$, one does not generally have that $K_0(t-s;x-y;m)f(y)$ is $L^1(\mathbb{R}_y^d)$. This difficulty is circumvented by requiring that f be in $L_0^\infty(\mathbb{R}^d)$. In addition, note that Definition 2 implies that for each $(t,s)\in T_\Delta^0$, the propagator $K(\cdot,t;\cdot s;m)$ is uniquely defined almost everywhere. The point $t=s$ is excluded from the specification of the propagator since, as the case of the free propagator K_0 shows, one cannot expect that K be defined for zero time displacement.

The Dyson expansion[3,4] provides, when successful, a perturbative method for obtaining the operator $U(t,s;m)$. The non-Laplacian portion of Hamiltonian (1.1) has the differential structure

$$\frac{i\hbar}{m}a(x,t)\cdot\nabla+\frac{i\hbar}{2m}(\nabla\cdot a)(x,t)+\frac{1}{2m}a(x,t)^2+v(x,t). \tag{2.9}$$

Using standard techniques for perturbing closed operators,[8] $H(t,m)$ can be shown to be the sum of the closed extension $H_0(m)$ of the free Hamiltonian $-\hbar^2(2m)^{-1}\Delta$ and a perturbing operator $V(t,m)$ associated with (2.9), i.e.,

$$H(t,m)=H_0(m)+V(t,m).$$

The unbounded operator $V(t,m)$ is defined on the domain D_0 [the domain of $H_0(m)$] and is $H_0(m)$ bounded. The formal integral equation equivalent to the equation of motion (2.3e) for $U(t,s;m)$ is

$$U(t,s;m)f=U_0(t,s)f-\frac{i}{\hbar}\int_s^t d\tau \times U_0(t,\tau)V(\tau,m)U(\tau,s;m)f, \tag{2.10}$$

where $U_0(t,s)$ is the evolution operator generated by $H_0(m)$. Iterating (2.10) leads to the formula

$$D_n(t,s;m)f=\left(-\frac{i}{\hbar}\right)^n\int_< d\mathbf{t}_n\, U_0(t,t_n)V(t_n,m)U_0(t_n,t_{n-1}) \times\cdots\times V(t_1,m)U_0(t_1,s)f, \tag{2.11}$$

where $\mathbf{t}_n=(t_1,\ldots,t_n)$ and $<$ is a shorthand notation for the n-dimensional time-ordered domain $\Delta_n(t,s)=\{\mathbf{t}_n\in\mathbb{R}^n: s\leqslant t_1\leqslant t_2\leqslant\cdots\leqslant t_n\leqslant t\}$.

In the circumstances where $V(\cdot,m)$ is uniformly bounded in the interval $[0,T]$ then it is well known (Ref. 5, Chap. II) that the sum of terms (2.11) converges strongly to $U(t,s;m)f$. However, for the problem at hand (with $a\neq 0$), the term $(i\hbar/m)a(x,t)\cdot\nabla$ is unbounded no matter how nicely $a(x,t)$ behaves. This difficulty may be overcome[2] by the use of the complex mass embedding method. In this method one obtains the propagator K for real mass values by continuity from the evolution kernels for complex masses. This technique is similar to Nelson's program[9] of using analytic continuation in mass to define the Feynman path integral.

For $\operatorname{Im} m>0$ an operator characterization of the nth Dyson iterate (2.11) and its summation over n is found in Ref. 2. Let $\mathscr{S}$ be the Schwartz space of complex-valued functions on $\mathbb{R}^d$ of rapid decrease. If $f\in\mathscr{S}$, $m\in\mathbb{C}_>$, then the right-hand side of (2.11) is defined as the n-dimensional strong Riemann integral on $\mathscr{H}$. Thus for each $(t,s;m)\in T_\Delta\times\mathbb{C}_>$ and $n\geqslant 1$ the map $D_n(t,s;m)$: $\mathscr{S}\to\mathscr{H}$ is well defined. In addition, for sufficiently short time displacements, $t-s$, the sum over n of $D_n(t,s;m)f$ converges strongly to $U(t,s;m)f$.

Consider next a kernel representation for $D_n(t,s;m)$, $\operatorname{Im} m>0$. The relevant formulas are built up in terms of integrals of composite measures formed from $\gamma(\cdot)$ and $\nu(\cdot)$. These measures and their combinatorics are defined as follows. First, the measure $\mu(\cdot)$ is given by

$$\mu(t)=(2m)^{-1}\gamma(t)*\gamma(t)+\nu(t).$$

Here $*$ is the scalar convolution of two measures in $\mathscr{M}^*(\mathbb{R}^d,\mathbb{C}^d)$ that constructs a measure in $\mathscr{M}^*(\mathbb{R}^d,\mathbb{C})$. Recall that if $\gamma(t)$ has support in $S_{k/2}$, then $\gamma(t)*\gamma(t)$ has its support contained in S_k. The Fourier transform of $\mu(t)$ is the scalar function $(2m)^{-1}a(x,t)^2+v(x,t)$, which is the derivative-free part of expression (2.9).

It is helpful to introduce the polar factorization of $\gamma(t)$ relative to $|\gamma(t)|$. For each $t\in[0,T]$ there exists a Borel-measurable function $\eta(t,\cdot)$: $\mathbb{R}^d\to\mathbb{C}^d$ whose $\mathbb{C}^d$ Hermitian norm $|\eta(t,\cdot)|=1$. Specifically if B denotes the Borel subsets of $\mathbb{R}^d$, then

$$\int_e d\gamma(t)=\int_e\eta(t,\alpha)d|\gamma(t)| \quad (e\in B).$$

An i-tuple of vectors in $\mathbb{R}^d$ is represented as $\boldsymbol{\alpha}_i=(\alpha_1,\ldots,\alpha_i)$. In terms of the parameters n, i, $\boldsymbol{\alpha}_{i-1}$, t, define the measure

$$\mu_i^n(t,\boldsymbol{\alpha}_{i-1})(e)=\int_e\left(\frac{\alpha}{2}+\sum_{j=1}^{i-1}\alpha_j\right)\cdot\eta(t,\alpha)d|\gamma(t)|, \tag{2.12}$$

where the dot denotes the summation over the components of vectors in $\mathbb{C}^d$. A combination of the previous two measures leads to

$$\rho_i(t)=\mu(t)-(\hbar/m)\mu_i^n(t,\boldsymbol{\alpha}_{i-1}).$$

It follows from its definition that $\rho_i(t)\in\mathscr{M}^*(S_k,\mathbb{C})$.

The measures that appear in the formula for $D_n(t,s;m)$ are constructed from $\rho_i(t_i)$ and $|\gamma(t_j)|$. For $0\leqslant r\leqslant n$, let $J_{n,r}$ denote the collection of all r-element subsets of $\{1,\ldots,n\}$. Thus $J_{n,r}=\{\varnothing\}$ if $r=0$, while if $1\leqslant r\leqslant n$, $J_{n,r}$ contains $\binom{n}{r}$ sets $\mathbf{j}_r=\{j_1,\ldots,j_r\}$, where we may suppose $j_1<j_2<\cdots<j_r$. Each $\mathbf{j}_r\in J_{n,r}$ defines a measure in the n-fold product space $(\mathbb{R}^d\times\cdots\times\mathbb{R}^d, B\times\cdots\times B)$ by

$$\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)=\rho_1(t_1)\times\cdots\times|\gamma(t_{j_1})| \times\cdots\times|\gamma(t_{j_r})|\times\cdots\times\rho_n(t_n). \tag{2.13}$$

The right-hand side of this equality is to be understood in the following way. If $r=0$ the measure involves only products of $\rho_i(t_i)$ for $i=1\sim n$. In the case where $r>0$ and $\mathbf{j}_r=\{j_1,\ldots,j_r\}$ then the j_ith term of the product for the $r=0$ case has element $\rho_{j_i}(t_{j_i})$ replaced with $|\gamma(t_{j_i})|$. Finally we specify the summation convention

$$\sum_{r,\mathbf{j}_r}=\sum_{r=0}^n\sum_{\mathbf{j}_r\in J_{n,r}}$$

and set $c_{n,r}$ to be

$$c_{n,r}=\left(\frac{m}{2\pi i\hbar(t-s)}\right)^{d/2}\left(\frac{-i}{\hbar}\right)^n\left(\frac{\hbar}{im}\right)^r.$$

The integral kernel behavior of D_n is then summarized by the statement (Ref. 2, Lemma 9).

Lemma 1: Suppose that $m\in\mathbb{C}_>$ and $(t,s)\in T_\Delta^0$. Let $d_n(\cdot,t;\cdot,s;m)$: $\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ be the parametric integral $(n\geqslant 1)$,

$$d_n(x,t;y,s;m) = \sum_{r,\mathbf{j}_r} c_{n,r} \int_< d\,\mathbf{t}_n \int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)\mathscr{I}, \tag{2.14a}$$

with the integrand

$$\mathscr{I} = \exp\Big[ix\cdot \sum_{p=1}^{n} \alpha_p - \frac{i\hbar}{2m} \sum_{p',p=1}^{n} (t - t_{p'} \vee t_p)\alpha_{p'}\cdot\alpha_p + \frac{im}{2\hbar(t-s)}(X_n - y)^2\Big]\mathscr{P}, \tag{2.14b}$$

where $t_{p'} \vee t_p = \max\{t_{p'},t_p\}$ and

$$\mathscr{P} = \exp\Big[\frac{-im}{2\hbar(t-s)}(X_n - y)^2\Big]\Big\{\prod_{i=1}^{r}[\eta(t_{j_i},\alpha_{j_i})\cdot\nabla_y]\Big\} \times\exp\Big[\frac{im}{2\hbar(t-s)}(X_n - y)^2\Big], \tag{2.14c}$$

$$X_n = x - \frac{\hbar}{m}\sum_{p=1}^{n}(t - t_p)\alpha_p. \tag{2.14d}$$

Then $d_n(\cdot,t;\cdot s;m)$ is a Carleman[10] kernel for the operator $D_n(t,s;m)$,

$$D_n(t,s;m)f = \int d_n(\cdot,t;y,s;m)f(y)dy, \quad f\in\mathscr{H}. \tag{2.15}$$

Basically this conclusion emerges from the study of iterations of the map $\exp(i\tau_2 H_0(m)/\hbar)V(\tau_1,m)$ $(\tau_1,\tau_2\geqslant 0)$ acting on an element of $\mathscr{S}$. Observe that the function d_n remains well defined for nonzero real values of m. Hereafter d_n will denote the function (2.14a) on the enlarged domain $T^0_\Delta \times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{C}_+$.

The propagator K is obtained as the sum over n of the functions d_n, with $d_0 \equiv K_0$. This fact and the explicit formula (2.14) for d_n is the reason for calling this result a constructive representation. The sum over n has a finite radius of convergence which may restrict the allowed time displacement $t-s$ in T_Δ, but which is independent of the x, y variables. We introduce the convenient convergence parameter

$$T_\gamma = \min\{|m|(2ek\gamma_T)^{-1},T\}. \tag{2.16}$$

Theorem 2: Let $m\in\mathbb{C}_+$ and $(t,s)\in T^0_\Delta$. If $t-s<T_\gamma$ then for each $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$, the (pointwise) sum over n of $d_n(x,t;y,s;m)$ is absolutely convergent and gives an x,y jointly continuous function

$$K(x,t;y,s;m) = \sum_{n=0}^{\infty} d_n(x,t;y,s;m), \tag{2.17}$$

which is the propagator (in the sense of Definition 2) of the Schrödinger evolution operator $U(t,s;m)$.

Proof: For masses that are in $\mathbb{C}_>$ or have positive real values this result is established in Ref. 2 (Proposition 4 and Theorem 3). The proof given there is also applicable if $m<0$. □

The $U(1)$ gauge dependence of time evolution for system (1.1) is well understood.[11,12] The fact that the concept of evolution has been widened here (via Definition 1) to include complex masses in $\mathbb{C}_+$ leaves this situation unchanged since the mass parameter does not appear in a gauge transformation. However, in view of the specific results above several questions relating to gauge invariance arise.

The first question concerns the stability of the potential class (A) under gauge transformations. Is there a natural class of gauge transformations that leave class (A) invariant? An affirmative answer is provided by the following construction. Let λ: $\mathbb{R}^d\times[0,T]\to\mathbb{R}$ denote a gauge potential.

Gauge class $\mathscr{G}$: Let $k<\infty$. The gauge potential λ is said to be in the class $\mathscr{G}(k)$ if λ is the Fourier transform,

$$\lambda(x,t) = \int e^{i\alpha\cdot x}\,d\zeta(t), \tag{2.18}$$

of a time-dependent measure $\zeta(t)$ satisfying

(1) $\zeta(t)\in\mathscr{M}^*(S_{k/2},\mathbb{C})$, $t\in[0,T]$,

(2) $\zeta(\cdot)$ is twice continuously differentiable on $[0,T]$.

Note that $\lambda(x,t)$ defines a t-dependent family of bounded operators on $\mathscr{H}$. Namely let $\Lambda(t)\in\mathscr{B}(\mathscr{H})$ be specified, for each $t\in[0,T]$, by

$$[\Lambda(t)f](x) = \lambda(x,t)f(x), \quad f\in\mathscr{H}.$$

The operator norm of $\Lambda(t)$ obeys $\|\Lambda(t)\|\leqslant\|\zeta(t)\|$. Similarly, the strong t derivative of $\Lambda(t)$ is also a uniformly bounded operator on $[0,T]$ and is given by

$$[\dot\Lambda(t)f](x) = \partial\lambda(x,t)f(x), \quad f\in\mathscr{H}.$$

The symbol $\partial\lambda$ denotes the partial derivative of λ with respect to time. The Abelian $U(1)$ gauge group is conventionally taken to be the family of unitary operators $\{\exp[(q/i\hbar)\times\Lambda(t)]: q\in\mathbb{R}\}$. However in (2.4) and thereafter the charge coupling constant q was incorporated into the definition of a and ϕ. To be notationally consistent, here, it is necessary to set $q=1$ and write the unitary gauge operator as $\exp[(i\hbar)^{-1}\Lambda(t)]$.

It is easy to see that the measure images of the two gauge transformation equations,

$$a(x,t;\Lambda) = a(x,t) + \nabla\lambda(x,t), \tag{2.19a}$$

$$v(x,t;\Lambda) = v(x,t) - \partial\lambda(x,t), \tag{2.19b}$$

take the respective forms

$$\gamma(t;\Lambda) = \gamma(t) + \nabla\zeta(t), \tag{2.20a}$$

$$\nu(t;\Lambda) = \nu(t) - \dot\zeta(t), \tag{2.20b}$$

for all $t\in[0,T]$. Here $\nabla\zeta(t)\in\mathscr{M}^*(S_{k/2},\mathbb{C}^d)$ is defined by

$$\nabla\zeta(t)(e) = \int_e i\alpha\,d\zeta(t) \quad (e\in B).$$

The definition of $\mathscr{G}(k)$ ensures that the right-hand sides of (2.20a) and (2.20b) are, respectively, in $\mathscr{M}^*(S_{k/2},\mathbb{C}^d)$ and $\mathscr{M}^*(S_k,\mathbb{C})$, and satisfy the t-differentiability conditions required in (A).

Time evolution, whether described in terms of $U(t,s;m)$ or its kernel from K, possesses a simple gauge dependence. Let $H(t,m;\Lambda)$ be the Hamiltonian operator determined by (2.9) with $a(x,t;\Lambda)$ and $v(x,t;\Lambda)$ substituting for $a(x,t)$ and $v(x,t)$. Further, let $U(t,s;m;\Lambda)$ be the family of complex mass Schrödinger evolution operators (described in Theorem 1) generated by $\{H(t,m;\Lambda): t\in[0,T]\}$. It follows, without difficulty from (2.3e), that

$$U(t,s;m;\Lambda) = \exp[i\Lambda(t)/\hbar]\,U(t,s;m)\exp[-i\Lambda(s)/\hbar].$$

This is the operator-valued form of the $U(1)$ gauge dependence of the evolution process. In an obvious notation its kernel analog reads

$$K(x,t;y,s;m;\Lambda)=\exp[i\lambda(x,t)/\hbar]K(x,t;y,s;m)\times\exp[-i\lambda(y,s)/\hbar]. \quad (2.21)$$

For sufficiently short time displacements, Theorem 2 guarantees the existence of both propagators in (2.21). However the condition $t-s<T_\gamma$ for Hamiltonian $H(t,m;\Lambda)$ involves $\sup\|\gamma(t,\Lambda)\|$, whereas Hamiltonian $H(t,m)$ uses $\sup\|\gamma(t)\| = \gamma_T$. Thus the convergence criterion (2.16) is not gauge invariant. This circumstance is just an artifact of the estimates of $d_n(x,t;y,s;m)$, used in Ref. 2, to study the convergence properties of the series (2.17). Each term in series (2.17) is highly gauge dependent and it is difficult to bound them in a manner that reflects the simple $U(1)$ gauge dependence of the exact K.

III. LARGE MASS ASYMPTOTIC BEHAVIOR

This section focuses on the mass dependence of the kernels $d_n(x,t;y,s;m)$ and $K(x,t;y,s;m)$. It is verified that the factorization (1.4) is valid and that $K_0(t-s;x-y;m)$ carries all the essential singularity of the propagator K in the inverse mass variable at $m^{-1}=0$. We show that $F(x,t;y,s;m^{-1})$ admits a m^{-1} expansion about the point 0. An explicit bound is obtained for the total truncation error of this large mass expansion.

The basic formula upon which the results of this section rest is the expression (2.14) for d_n. We discuss this formula in detail and show that, in spite of its rather elaborate nature, it has a structure that permits one to find simple estimates. These estimates will suffice to determine the mass dependence of K.

It is useful to employ the variable $u=m^{-1}$. Let $\mathscr{U}$ be the u-complex plane and let $\mathscr{U}_<$ ($\mathscr{U}_{\leqslant}$) represent the lower half-planes $\operatorname{Im} u<0$ ($\operatorname{Im} u\leqslant 0$). Furthermore, denote the open semidisk of radius u_0 by $\mathscr{U}_<(u_0)=\{u\in\mathscr{U}_<:|u|<u_0\}$ and its closure by $\mathscr{U}_{\leqslant}(u_0)$. The large mass limit then corresponds to $u\to 0$ in $\mathscr{U}_{\leqslant}(u_0)$.

To begin, consider the integrand $\mathscr{I}$ defined in (2.14b). The product of gradients ∇_y appearing in $\mathscr{P}$ may be evaluated (Ref. 2, Lemma 8) with the result

$$\mathscr{P}=\sum_{l=0}^{[r/2]}\sum_{\mathbf{q}_r}[-i\hbar u(t-s)]^{l-r}\Phi(\mathbf{q}_r;\boldsymbol{\alpha}_n)\Psi(\mathbf{q}_r;\boldsymbol{\alpha}_n), \quad (3.1a)$$

where Φ and Ψ are the functions

$$\Phi(\mathbf{q}_r;\boldsymbol{\alpha}_n)=\prod_{i=0}^{l-1}\eta(t_{q_{r-2i}},\alpha_{q_{r-2i}})\cdot\eta(t_{q_{r-2i-1}},\alpha_{q_{r-2i-1}}), \quad (3.1b)$$

$$\Psi(\mathbf{q}_r;\boldsymbol{\alpha}_n)=\prod_{i=1}^{r-2l}\eta(t_{q_i},\alpha_{q_i})\cdot\left[y-x+(\hbar u)\sum_{p=1}^{n}(t-t_p)\alpha_p\right]. \quad (3.1c)$$

The summation convention in (3.1a) for $\mathbf{q}_r$ is the following. The symbol $[r/2]$ is the greatest integer less than or equal to $r/2$. Suppose the index set $\mathbf{j}_r=(j_1,j_2,\ldots,j_r)$ is given. For each $0\leqslant l\leqslant[r/2]$, $\mathbf{q}_r$ represents a particular two-stage selection from the set $\mathbf{j}_r$. First choose $r-2l$ elements from $\mathbf{j}_r$, and denote them $q_1,q_2,\ldots,q_{r-2l}$. Next select l pairs from the remaining $2l$ elements in $\mathbf{j}_r$, and denote them by $\{q_{r-2l+1},q_{r-2l+2}\},\ldots,\{q_{r-1},q_r\}$. The summation involving $\mathbf{q}_r$ denotes all $r![2^l(r-2l)!l!]^{-1}$ distinct choices of this type.

It is convenient to abbreviate the space-time arguments of d_n by writing $Q=(x,t;y,s)$. The dependence of Φ and Ψ upon Q and $\mathbf{t}_n$ is suppressed, for reasons of notational economy. Note the simple numerical bounds that Φ and Ψ obey

$$|\Phi|\leqslant 1,\quad |\Psi|\leqslant(Z_n)^{r-2l}, \quad (3.2a)$$

$$Z_n=|y-x|+|u|n\hbar kT. \quad (3.2b)$$

A form of d_n more suitable for estimates results from making the change of variables

$$t_i=s+\xi_i(t-s),\quad \xi_i\in[0,1],\quad i=1\sim n.$$

In the new variables $\boldsymbol{\xi}_n=(\xi_1,\xi_2,\ldots,\xi_n)\in[0,1]^n$ the time-ordered integral in (2.11) becomes the ξ-ordered integration $(0\leqslant\xi_1\leqslant\xi_2\leqslant\cdots\leqslant\xi_n\leqslant 1)$,

$$\int_{\Delta_n(t,s)}d\mathbf{t}_n=(t-s)^n\int_<d\boldsymbol{\xi}_n.$$

Writing $\mathscr{I}$ in the ξ_i variables gives us, after combining the arguments of the exponential in (2.14b),

$$\mathscr{I}=\exp[i(x-y)^2/2\hbar u(t-s)]\exp(ib_n)f_2, \quad (3.3a)$$

where

$$f_2=f_2(\boldsymbol{\xi}_n,\boldsymbol{\alpha}_n,t-s;u)=\exp[-i\hbar u(t-s)a_n/2], \quad (3.3b)$$

$$a_n=a_n(\boldsymbol{\xi}_n,\boldsymbol{\alpha}_n)=\sum_{i,j=1}^{n}g(\xi_i,\xi_j)\alpha_i\cdot\alpha_j, \quad (3.3c)$$

$$b_n=b_n(\boldsymbol{\xi}_n,\boldsymbol{\alpha}_n;x,y)=\sum_{p=1}^{n}\alpha_p\cdot[y+\xi_p(x-y)], \quad (3.3d)$$

$$g(\xi,\xi')=\xi_<(1-\xi_>). \quad (3.3e)$$

Here $\xi_<=\operatorname{Min}\{\xi,\xi'\}$ and $\xi_>=\operatorname{Max}\{\xi,\xi'\}$. The function g is a Green's function for the operator $d^2/d\xi^2$ on the interval $[0,1]$. Observe that the vector $y+\xi_p(x-y)$ in (3.3d) is the space part of $w(\xi_p)$ in (1.8).

A generalized form of the integral (2.14) occurs in much of the subsequent analysis. This generic form results when the factors corresponding to $K_0(t-s,x-y;m)$ are deleted and f_2 is replaced by other related functions of $\boldsymbol{\xi}_n,\boldsymbol{\alpha}_n,t-s;u$. Denoting these functions by $f(\boldsymbol{\xi}_n,\boldsymbol{\alpha}_n,t-s;u)$ define

$$S_n(f,Q;u)=(-1)^n\sum_{r,\mathbf{j}_r}\sum_{l,\mathbf{q}_r}\left(\frac{i}{\hbar}\right)^{n-l}u^l(t-s)^{n-r+l}\times\int_<d\boldsymbol{\xi}_n\int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n(\boldsymbol{\xi}_n))f(\boldsymbol{\xi}_n,\boldsymbol{\alpha}_n,t-s;u)\times\Phi(\mathbf{q}_r;\boldsymbol{\alpha}_n)\Psi(\mathbf{q}_r;\boldsymbol{\alpha}_n)\exp(ib_n), \quad (3.4)$$

where $\{\mathbf{t}_n(\boldsymbol{\xi}_n)\}_i=s+\xi_i(t-s)$, $i=1\sim n$. If we return to the case with $f=f_2$ then (2.14a), (3.1), and (3.3) imply the factorization

$$d_n(Q;m) = K_0(t-s,x-y;m)\tilde{d}_n(Q;m^{-1}), \qquad (3.5a)$$

where

$$\tilde{d}_n(Q;u) = S_n(f_2,Q;u). \qquad (3.5b)$$

Notice that in (3.5a) one must restrict $t>s$ and $m^{-1}\neq 0$ in order to avoid the essential singularity in K_0, while $\tilde{d}_n$ remains well defined by (3.5b) for $t=s$ and $u=0$. Of course (3.5a) gives $\tilde{d}_0 \equiv 1$.

A number of the basic properties of $\tilde{d}_n(Q;u)$ follow immediately from its integral form $S_n(f_2,Q;u)$. Note first that $\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$ has finite total variation in the n-fold product space $(\mathbb{R}^d\times\cdots\times\mathbb{R}^d, B\times\cdots\times B)$. In the notation of Sec. II, Λ^n is a mapping of $J_{n,r}\times\Delta_n(T,0)$ into the Banach space $\mathscr{M}^*([S_k]^n,\mathbb{C})$. As a consequence of (2.7), (2.12), (2.13), and the fact that α_i $(i=1\sim n)$ has compact support S_k, the norm of $\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$ has the estimate

$$\|\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)\| = |\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)|(\mathbb{R}^d)$$
$$\leqslant (\mu_T + |u|\hbar nk\gamma_T)^{n-r}(\gamma_T)^r \qquad (3.6)$$

for all $\mathbf{j}_r\in J_{n,r}$ and $\mathbf{t}_n\in\Delta_n(T,0)$.

Now observe that f_2 and Φ are x,y independent while $\exp(ib_n)$ is a C^∞ function of x,y and that $\Psi(\mathbf{q}_r;\alpha_n)$ is a polynomial of order $r-2l$ in $x-y$. Combining these facts shows that $\tilde{d}_n$ has partial derivatives with respect to x and y to arbitrary order that are continuous on the domain $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}$. Furthermore, $\tilde{d}_n$ has first-order derivatives with respect to t and s that are continuous on the domain $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}$. Finally for fixed $Q\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d$, $\tilde{d}_n$ is an entire function of u. Verification of this last statement follows from an application of Morera's theorem. The measure $\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$ is a polynomial in u and the remaining portion of the integrand is an entire function of u. Integrate $\tilde{d}_n$ over an arbitrary smooth finite length contour in $\mathscr{U}$. In this case the multiple integral is absolutely convergent, and thus Fubini's theorem shows one may interchange the order of the u and the Λ^n integration. Doing the u integration first shows that the complete multiple integral is zero for all contours. Thus $\tilde{d}_n$ is entire. This result is a particular consequence of the fact that the measures $\gamma(t)$ and $\nu(t)$ have compact supports S_k. If the supports for these measures were all of $\mathbb{R}^d$ then the multiple integral would not be absolutely convergent for u in the upper half complex plane and Fubini's theorem would no longer apply.

Next consider F of (1.4). From (3.5a) and (2.17) it is evident that $F(x,t;y,s;m^{-1})$ $[=F(Q;u)]$ is the sum over n of $\tilde{d}_n(Q;u)$. In fact this sum provides a proper definition of F. More specifically the following proposition is found.

Proposition 1: Assume the potentials a and v are in class (A). Let $u_0<(2ekt\gamma_T)^{-1}$.

(a) For each $(Q;u)\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$ the sum over n of $\tilde{d}_n(Q;u)$ is absolutely convergent and provides a pointwise definition of the function F, i.e.,

$$F(Q;u) = \sum_{n=0}^{\infty}\tilde{d}_n(Q;u). \qquad (3.7)$$

(b) The function F has partial derivatives to arbitrary order in x,y that are (jointly) continuous on the domain $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$.

(c) F has first-order partial derivatives with respect to t and s that are continuous functions on $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$.

(d) For each $Q\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d$, $F(Q;\cdot)$ is holomorphic in $\mathscr{U}_<(u_0)$ and continuous in $\mathscr{U}_<(u_0)$.

(e) Let K be the propagator defined in Theorem 2. On the domain $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d\times\{m\in\mathbb{C}_+: |m^{-1}|\leqslant u_0\}$ the function K admits the factorization (1.4).

Proof (sketch): Consider (a) and (d) together. The series (3.7) and (2.17) are the same, modulo the multiplicative function K_0, so the convergence proof (Ref. 2, Lemma 10) with obvious modifications applies to (3.7). A minor difference in the series (3.7) and (2.17) is that the functions $\tilde{d}_n$ are nonsingular at the point $u=0$ and so this point may be added to the domain of convergence of (3.7). The estimates obtained in demonstrating the pointwise convergence of (3.7) also show for each fixed $Q\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d$ that the sum (3.7) is absolutely and uniformly convergent in $\mathscr{U}_<(u_0)$. Since each $\tilde{d}_n(Q;\cdot)$ is holomorphic in $\mathscr{U}_<(u_0)$ and continuous in the compact $\mathscr{U}_<(u_0)$, it follows that $F(Q;\cdot)$ is holomorphic in $\mathscr{U}_<(u_0)$ and continuous in $\mathscr{U}_<(u_0)$.

Examine (b) and (c). Let $\nabla^\gamma_{x,y}$ denote the partial derivative with respect to the variable set $(x_1,\ldots,x_d,y_1,\ldots,y_d)$ that is specified by the multi-index $\gamma=(\gamma_1,\ldots,\gamma_{2d})$. Use the method of estimating $\tilde{d}_n(Q;u)$ found in Eq. (6.25) of Ref. 2 and which is elaborated on in the proof of Theorem 3 below. In this way bounds for $\nabla^\gamma_{x,y}\tilde{d}_n(Q;u)$ are obtained that show the series over n of $\nabla^\gamma_{x,y}\tilde{d}_n$ is absolutely and uniformly convergent for (Q,u) in arbitrary compact subsets of $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$. Thus

$$\nabla^\gamma_{x,y}\sum_{n=0}^{\infty}\tilde{d}_n(Q;u) = \sum_{n=0}^{\infty}\nabla^\gamma_{x,y}\tilde{d}_n(Q;u). \qquad (3.8)$$

Each term in the sum on the right is uniformly continuous in the compact subsets selected above and so this sum defines a continuous function in $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$.

A similar argument verifies that the partial derivative on the set $t\in(s,T)$ [or $s\in(0,t)$] may be interchanged with the sum over n in (3.7). The sum of $\partial\tilde{d}_n(Q;u)/\partial t$ is uniformly convergent for compact subsets of $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$. Thus the partial derivative of F with respect to t exists and is a (jointly) continuous function on the domain $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$. Finally (e) is an immediate consequence of (a) and (3.5a). □

From now on we use the factored form K_0F of Eq. (1.4) as the preferred representation of the propagator K. The behavior of K in the neighborhoods of $t=s$ and $u=0$ is conveniently studied with the representation K_0F because both $t=s$ and $u=0$ are allowed in the domain of F.

Proposition 1 is just one of several ways of summarizing the conclusions that result from analyzing the convergence properties of series (3.7). In the specific form above the value of u_0 was chosen sufficiently small so that the time-displacement condition (2.16) is $T_\gamma=T$, i.e., no restriction beyond the standard requirement $(t,s)\in T_\Delta$. Moreover, even if the physical mass of a system is such that $|m^{-1}|>u_0$, then the results of Proposition 1 apply in an altered form. Let $c>1$ be large enough so that $m^{-1}\in\mathscr{U}_<(cu_0)$. The series

(3.7) remains convergent if the time displacement obeys the limitation $t-s<T/c$.

The holomorphy of $F(Q;u)$ in $\mathscr{U}_<(u_0)$ means that for each Q, $F(Q;u)$ has a convergent Taylor series expansion about every point in $\mathscr{U}_<(u_0)$. Unfortunately this fact is of no help in deriving expansion (1.6) since the point $u=0$ is not in the domain $\mathscr{U}_<(u_0)$. Instead we proceed to derive the small u expansion of $F(Q;u)$ by an appropriate restructuring of series (3.7).

To begin with let us consider the behavior of f_2 in $\mathscr{U}_<(u_0)$ that is critical in the following analysis. Recall that a_n is non-negative and has the simple bound (Ref. 13, Lemma 5),

$$0\leqslant a_n(\xi_n;\alpha_n)\leqslant\frac{n}{4}\sum_{i=1}^{n}\alpha_i^2\leqslant\left(\frac{nk}{2}\right)^2. \tag{3.9}$$

If $u\in\mathscr{U}_<$ the argument of the exponential in f_2 has a nonpositive real part for all ξ_n, α_n and so f_2 admits an M-term asymptotic expansion. Upon setting $c_j=(j!)^{-1}[\hbar(t-s)/2i]^j$ we have for $M\geqslant1$,

$$f_2(\xi_n,\alpha_n,t-s;u)=\sum_{j=0}^{M-1}c_ju^j(a_n)^j+c_Mu^MH_M(\xi_n,\alpha_n,t-s;u), \tag{3.10a}$$

where the remainder H_M has the u-independent bound

$$|H_M|\leqslant(a_n)^M\leqslant(nk/2)^{2M} \tag{3.10b}$$

for all $0\leqslant\xi_1\leqslant\cdots\leqslant\xi_n<1$ and all $\alpha_i\in S_k$.

In addition to the u dependence in f_2 the integrals $S_n(f_2,Q;u)$ acquire u dependence from the measure $\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$ and the function Ψ. A convenient description of the latter u dependence is given by the following lemma.

Lemma 2: Let the symbol $L\equiv\{n,r,\mathbf{j}_r,l,\mathbf{q}_r\}$ represent a set of summation indices that characterize the functions in (3.1a)–(3.1c). Let $(t,s)\in T_\Delta$, $i\geqslant0$, and let $u\in\mathscr{U}$ with $|u|\leqslant u_0$. Denote by $\mathscr{I}(L)$ the following multiple integral that occurs when (3.10a) is substituted into (3.5b):

$$\mathscr{I}(L)\equiv\int_<d\xi_n\int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n(\xi_n))[a_n(\xi_n,\alpha_n)]^i\times\exp[\sqrt{-1}b_n]\Phi(\mathbf{q}_r;\alpha_n)\Psi(\mathbf{q}_r;\alpha_n). \tag{3.11}$$

Then $\mathscr{I}$ is a polynomial in u of the form

$$\mathscr{I}(L)=\sum_{p=0}^{n-2l}u^pA_p(L), \tag{3.12}$$

whose coefficients obey the bound

$$(t-s)^{n-r}(u_0)^p|A_p(L)|\leqslant(t-s)^p\tilde{A}_p(L), \tag{3.13a}$$

$$\tilde{A}_p(L)\equiv T^{n-r-p}(n!)^{-1}(nk/2)^{2i}(\mu_T+\hbar u_0nk\gamma_T)^{n-r}(\gamma_T)^r(Z_n^0)^{r-2l}, \tag{3.13b}$$

$$Z_n^0\equiv|x-y|+u_0n\hbar kT. \tag{3.13c}$$

Proof: The measure Λ^n contains $n-r$ factors of $\rho_i=\mu-u\hbar\mu_i^n$ having u dependence, while Ψ of (3.1c) has $r-2l$ factors each containing one power of u. This leads to (3.12).

Contributions to $A_p(L)$ thus arise by selecting a factors of u from Λ^n, where $a=0,\ldots,\min\{n-r,p\}$. These terms contain a measures $-\hbar\mu_i^n$ of bound $\hbar nk\gamma_T$, $n-r-a$ measures μ_i of bound μ_T, and r measures $|\gamma_i|$ of bound γ_T. The remaining $p-a$ factors of u are selected from Ψ. The coefficient functions of these factors of u in Ψ are bounded by $\hbar n(t-s)k$, while the complementary factors from Ψ are bounded by $|x-y|$.

Noting that $\int_<d\xi_n=(n!)^{-1}$ and upon employing (3.9) these observations give

$$(u_0)^p|A_p(L)|\leqslant\frac{1}{n!}\sum_{a=0}^{n-r}\binom{n-r}{a}(\hbar u_0nk\gamma_T)^a(\mu_T)^{n-r-a}(\gamma_T)^r\left(\frac{nk}{2}\right)^{2i}\times\binom{r-2l}{p-a}(u_0n\hbar k[t-s])^{p-a}|x-y|^{r-2l-(p-a)}, \tag{3.14}$$

where $\binom{r-2l}{p-a}$ is understood to be zero if $a>p$. Upon multiplying (3.14) by $(t-s)^{n-r}$ one can extract $(t-s)^p$ from the sum over a, and replace $(t-s)^{n-r-a}$ by the larger T^{n-r-a}, since $n-r-a\geqslant0$. Next multiply and divide the resulting right-hand side of (3.14) by T^p. The result of these manipulations is that the factor $[t-s]$ in (3.14) has been replaced by T, and a factor $(t-s)^pT^{n-r-p}$ appears outside the summation over a. Finally replace the index $p-a$ with b inside Σ_a and sum over $b=0,\ldots,r-2l$ to obtain (3.13). □

The elementary observations made in (3.10) and Lemma 2 play a key role in the derivation of the small u asymptotic expansion of $F(Q;u)$. It is helpful to first introduce the functions that appear in the remainder term bound of this expansion. Define σ_i: $[0,1)\to\mathbb{R}$ by the convergent series

$$\sigma_i(v)\equiv\sum_{n=1}^{\infty}v^n(n+1)n^{2i}\quad(i\geqslant0), \tag{3.15a}$$

and let I_i: $\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ be

$$I_i(v_1,v_2)=(i!)^{-1}(\hbar k^2/8)^i(v_1)^i\sigma_i(2eu_0\gamma_TkT)\times\exp(c_1''+c_2''v_2), \tag{3.15b}$$

where

$$c_1''=(2u_0\hbar k)^{-1}(\mu_T/\gamma_T+1/2kT),$$
$$c_2''=(2u_0\hbar kT)^{-1}.$$

With these conventions we have the following theorem.

Theorem 3: Let $u_0<(2e\gamma_TkT)^{-1}$. For all integers $M\geqslant1$ and each $Q\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d$, $F(Q;u)$ has the small u asymptotic expansion in $\mathscr{U}_<(u_0)$,

$$F(Q;u)=\sum_{j=0}^{M-1}u^jP_j(Q)+u^ME_M(Q;u). \tag{3.16}$$

Throughout their domains of definition the M complex-valued u-independent coefficient functions P_j: $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$, and the error function E_M: $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)\to\mathbb{C}$ possess continuous partial derivatives up to arbitrary order in (x,y). On the more restricted domains $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d$ and $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)$, P_j and E_M have continuous first-order partial derivatives with respect to t and s.

Furthermore, for each $(Q,u)\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}(u_0)$, P_j and E_M obey the estimates

$$|P_j(Q)|\leqslant\left(\frac{t-s}{u_0T}\right)^j\sum_{i=0}^{j}I_i(u_0T,|x-y|)+\delta_{j,0}, \tag{3.17}$$

$$|E_M(Q;u)| \leqslant \left(\frac{t-s}{u_0 T}\right)^M \sum_{i=0}^{M} I_i(u_0 T,|x-y|). \quad (3.18)$$

In addition, the derivatives of E_M obey the following order estimates, as $t-s \to 0^+$, uniformly for (x,y,u) in compact subsets of $\mathbb{R}^d \times \mathbb{R}^d \times \mathscr{U}_<(u_0)$. For all $|\gamma|>0$ the spatial derivative $\nabla^\gamma_{x,y} E_M = O((t-s)^M)$, for $(t,s) \in T_\Delta$. The time derivatives $(\partial/\partial t)E_M$ and $(\partial/\partial s)E_M$ are $O((t-s)^{M-1})$ for $(t,s) \in T^0_\Delta$. Thus expansion (3.16) may be differentiated to first order in the time variables and to arbitrary order in the space arguments.

Proof: Replace f_2 in $S_n(f_2,Q;u)$ with expansion (3.10a). After summing over n to obtain $F(Q;u)$ one finds

$$F(Q;u) = \sum_{i=0}^{M-1} F_i(Q;u) + \widetilde{F}_M(Q;u), \quad (3.19a)$$

where

$$F_i(Q;u) = c_i u^i \sum_{n=1}^{\infty} S_n((a_n)^i,Q;u) + \delta_{i,0}, \quad (3.19b)$$

and

$$\widetilde{F}_M(Q;u) = c_M u^M \sum_{n=1}^{\infty} S_n(H_M,Q;u). \quad (3.19c)$$

Consider the series that sums to $F_i(Q;u)$, and let $u \in \mathscr{U}$, $|u| \leqslant u_0$ for the moment. The function $S_n((a_n)^i,Q;u)$ is a finite sum of integrals $\mathscr{I}(L)$ that are precisely those characterized in Lemma 2. Hence $u^i S_n((a_n)^i,Q;u)$ is a polynomial in u, and F_i is given by the multiple series [with $c'_i \equiv (-1)^n(\sqrt{-1})^{n-l-i}/(i!2^i)$],

$$F_i(Q;u) - \delta_{i,0} = \sum_L \sum_{p=0}^{n-2l} c'_i \hbar^{i+l-n} \times (t-s)^{i+l+n-r} u^{i+l+p} A_p(L). \quad (3.20)$$

If we bound the coefficients $(t-s)^{n-r} u^p A_p(L)$ using (3.13) then (3.20) has the following majorizing series, which we denote as $\widetilde{I}_i \equiv \widetilde{I}_i(|u|,t-s,|x-y|)$,

$$\widetilde{I}_i = \sum_L \sum_{p=0}^{n-2l} (i!2^i)^{-1} \hbar^{i+l-n} |u|^{i+l} (t-s)^{i+l+p} \widetilde{A}_p(L). \quad (3.21)$$

An important property of this majorizing series is that it is monotone increasing in each of the three variables $|u|$, $t-s$, and $|x-y|$. This is a consequence of the fact that the algebraic powers of these variables are always non-negative. Thus a bound of $\widetilde{I}_i$ found for any particular value of $|u|$, $t-s$, and $|x-y|$ is applicable to all smaller values of these variables.

Note that in terms of the indices in the set L, the bound (3.21) and (3.13) is independent of the more elaborate sets $\mathbf{j}_r$ and $\mathbf{q}_r$. In (3.21) one always has $(t-s)^p T^{-p} \leqslant 1$; the sum over p has a maximum of $n+1$ terms, and there are $r![2^l(r-2l)!l!]^{-1}$ terms in the $\mathbf{q}_r$ sum, thereby

$$\sum_{\mathbf{q}_r} \sum_{p=0}^{n-2l} (t-s)^p \widetilde{A}_p(L) \leqslant \frac{(n+1)r!T^{n-r}}{2^l(r-2l)!l!n!}\left(\frac{nk}{2}\right)^{2i} \times (\mu_T + \hbar u_0 nk\gamma_T)^{n-r}(\gamma_T)^r (Z^0_n)^{r-2l}. \quad (3.22)$$

Multiply (3.22) by $(i!2^i)^{-1}\hbar^{i+l-n}(|u|(t-s))^{i+l}$ and then sum over the additional indices $n,r,\mathbf{j}_r,l$ to obtain

$$\widetilde{I}_i \leqslant \frac{1}{i!}\left(\frac{\hbar|u|(t-s)k^2}{8}\right)^i \sum_{n,r,\mathbf{j}_r} \frac{(n+1)n^{2i}}{n!}\left(\frac{T}{\hbar}\right)^n \times (\mu_T + \hbar u_0 nk\gamma_T)^{n-r}(\gamma_T)^r B_{n,r}, \quad (3.23a)$$

where

$$B_{n,r} \equiv \sum_{l=0}^{[r/2]} \frac{r!}{(r-2l)!l!}\left(\frac{Z^0_n}{T}\right)^{r-l}\left(\frac{\hbar u_0}{2Z^0_n}\right)^l. \quad (3.23b)$$

One completes the bound study of $\widetilde{I}_i$ by using the same pattern of estimates found in (6.22)–(6.25) of Ref. 2. Namely in (3.23b) replace $1/(r-2l)!$ by the larger $n^l/(r-l)!$ and extend the l sum from $[r/2]$ to r. The final estimate for $\widetilde{I}_i$ is

$$\widetilde{I}_i(|u|,t-s,|x-y|) \leqslant ((t-s)/T)^i I_i(u_0 T,|x-y|) < \infty, \quad (3.24)$$

where I_i is the function defined by (3.15b). This bound applies for all $i \geqslant 0$, and requires $2eu_0\gamma_T kT < 1$ for its validity.

This convergence property of the infinite series (3.19b) of polynomials in u implies that $F_i(Q;\cdot)$ is holomorphic on the complex disk $|u| < u_0$. Let $f_{i,j}$ be its Taylor series coefficients at the origin, viz.

$$f_{i,j}(Q) = (j!)^{-1}\left(\frac{d}{du}\right)^j F_i(Q;u)\Big|_{u=0} \quad (j \geqslant 0). \quad (3.25)$$

From (3.20) it is clear that $f_{i,j} = 0$ whenever $j < i$. Then write

$$F_i(Q;u) = \sum_{j=i}^{M-1} u^j f_{i,j}(Q) + u^M \widetilde{F}_{i,M}(Q;u), \quad (3.26a)$$

which defines $\widetilde{F}_{i,M}$ for $0 < |u| \leqslant u_0$. Combining (3.26a) with (3.19a) gives the desired small u expansion (3.16), where

$$P_j(Q) = \sum_{i=0}^{j} f_{i,j}(Q) \quad (j=0,\ldots,M-1), \quad (3.26b)$$

and the remainder term is

$$E_M(Q;u) = \sum_{i=0}^{M-1} \widetilde{F}_{i,M}(Q;u) + u^{-M}\widetilde{F}_M(Q;u). \quad (3.26c)$$

Consider next the bound (3.17). To compute $f_{i,j}$ from (3.25), we may differentiate series (3.20) term by term. In this way one finds that $f_{i,j}$ is the coefficient of the jth power of u in the series (3.20),

$$f_{i,j}(Q) = \sum_L \sum_{p=0}^{n-2l} \delta_{j,i+l+p} c'_i \hbar^{i+l-n} \times (t-s)^{n-r+j-p} A_p(L) + \delta_{i,0}\delta_{j,0}. \quad (3.27)$$

Estimate this sum term by term. First use (3.13a) in order to bound $(t-s)^{n-r} A_p(L)$, next employ the Kronecker delta restriction $j = i+l+p$ to write $T^{-p} = T^{-j+i+l} u_0^{i+l+p-j}$ and finally replace the Kronecker delta with 1. In this way one finds

$$|f_{i,j}(Q)| \leqslant \delta_{i,0}\delta_{j,0} + \left(\frac{t-s}{u_0 T}\right)^j \times \sum_L \sum_{p=0}^{n-2l} \frac{1}{i!2^i} \hbar^{i+l-n} u_0^{i+l} T^{i+l+n-r} \times \frac{(nk/2)^{2i}}{n!} (\mu_T + \hbar u_0 nk\gamma_T)^{n-r}$$

$$\times(\gamma_T)^r(Z_n^0)^{r-2l}. \quad (3.28)$$

Summing inequality (3.28) from $i=0$ to j leads immediately to the bound (3.17).

The error function E_M is a sum of terms $\widetilde{F}_{i,M}$ and $u^{-M}\widetilde{F}_M$. Examine $\widetilde{F}_{i,M}$ first. Comparing (3.20) and (3.26a) shows that

$$\widetilde{F}_{i,M}(Q;u) = \sum_L \sum_{p=0}^{n-2l} \Theta(i+l+p-M) c_i' \hbar^{i+l-n} \times (t-s)^{n-r+i+l} u^{i+l+p-M} A_p(L), \quad (3.29)$$

where Θ is the Heaviside function with value 1 for argument 0. After taking the modulus of each term in (3.29) employ inequality (3.13a) and use the restriction $i+l+p \geqslant M$ to replace $|u|^{i+l+p-M}$ by $u_0^{i+l+p-M}$ and $[(t-s)/T]^{i+l+p}$ by $[(t-s)/T]^M$, then replace the Heaviside function by 1. Now proceed in the same fashion as in the estimate (3.23) to show that

$$|\widetilde{F}_{i,M}(Q;u)| \leqslant ((t-s)/u_0T)^M I_i(u_0T,|x-y|). \quad (3.30)$$

The final function to bound is $u^{-M}\widetilde{F}_M$. Upon using (3.10b) the series (3.19c) for $\widetilde{F}_M$ is majorized term by term by $\widetilde{I}_M(|u|,t-s,|x-y|)$. Since the series $\widetilde{I}_M$ has only positive powers of $|u|$, with the least power equal to M, it follows that

$$|u^{-M}\widetilde{F}_M(Q;u)| \leqslant u_0^{-M}\widetilde{I}_M(u_0,t-s,|x-y|) \leqslant ((t-s)/u_0T)^M I_M(u_0T,|x-y|), \quad (3.31)$$

where the final inequality has utilized (3.24) with $i=M$. Adding inequalities (3.30) and (3.31) establishes (3.18). Observe that P_j is of order $(t-s)^j$ and E_M is of order $(t-s)^M$.

It remains to verify the differentiability claims for P_j, since those for E_m then follow from Proposition 1 and (3.16). The basic idea is to show that the integrals defining $A_p(L)$ have the desired differentiability, and then that series (3.27) has the required absolute and uniform convergence. This lengthy but elementary task can be carried out using similar methods to the one sketched in Proposition 1. The order estimates for $\nabla_{x,y}^{\gamma} E_M$ and the time derivatives of E_M are obtained by a direct (similar to that above) majorization of their defining series. □

Several remarks about Theorem 3 and the analysis leading up to it are warranted. It is natural to question why expansion (3.16) is an asymptotic expansion of arbitrary order M and not a convergent series expansion in u. If one attempts to write (3.16) as a series by setting $M=\infty$ one must replace asymptotic expansion (3.10a) by the related Taylor series for f_2. After making this substitution and using the estimates (3.22)–(3.24), then expansion (3.16) would form a convergent series if the sum over $i \geqslant 0$ of $\widetilde{I}_i$ were finite. However, a little study shows that this sum is divergent. This negative result does not entirely rule out the possibility that a convergent series expansion of $F(Q;u)$ in u exists, but rather indicates the methods found in this section, for bounding the multiple series expansion of $F(Q;u)$, are not precise enough to resolve this issue.

The new feature used in estimating the series (3.7) and its companion series such as (3.29), which was not utilized in Ref. 2, is the observation that the majorizing series like (3.21) are monotone increasing in the variables $|u|$, $t-s$, and $|x-y|$. As an example consider the bound for series (3.7). After employing the inequality $|f_2| \leqslant 1$ and upon using the fact that only non-negative powers of u and $t-s$ occur in the majorizing series we find

$$|F(Q;u)| \leqslant (1-2eu_0\gamma_T kT)^{-1}\exp(c_1''+c_2''|x-y|). \quad (3.32)$$

This bound is an improvement on (6.30) of Ref. 2, in that it is uniform with respect to all compact subsets of $T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d \times \mathscr{U}_<(u_0)$.

IV. GAUGE-INVARIANT RECURRENCE RELATIONS AND EXPANSION COEFFICIENTS

This section completes and consolidates the derivation of the large mass asymptotic expansion (1.6). It is established that the propagator K, as constructed by Theorem 2, is a solution of the time-dependent Schrödinger equation (1.2) interpreted as a classical partial differential equation in the open region $T_\Delta^0 \times \mathbb{R}^d \times \mathbb{R}^d$. If the boundary points $t=s$ are added to T_Δ^0, the propagator determines the fundamental solution of (1.2). The series (3.7), which defines F, is explicitly summed for $u=0$ in order to determine the coefficient function $P_0(Q)$. It turns out that $P_0(Q) = \exp[(i\hbar)^{-1}J(Q)]$, where $J(Q)$ is given by (1.7). By introducing the representation

$$F(Q;u) = T(Q;u)\exp[(i\hbar)^{-1}J(Q)], \quad (4.1)$$

the function F is split into a gauge-dependent part $\exp[(i\hbar)^{-1}J]$ and a gauge-independent function T. The small u expansion for T follows from representation (4.1) together with expansion (3.16) of Theorem 3. The resultant coefficient functions T_j are shown to be determined by a manifestly gauge invariant recurrence relation. Bound estimates for all T_j are found and explicit formulas are given for T_1 and T_2.

First, let us find the specific form of $P_0(Q)$. The asymptotic expansion (3.16) is applicable if $u=0$, i.e.,

$$P_0(Q) = F(Q;0) = \sum_{n=0}^{\infty} \tilde{d}_n(Q;0). \quad (4.2)$$

The following is a consequence of formulas (3.5b) and (4.2).

Lemma 3: For all $Q \in T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d$,

$$P_0(Q) = \exp[(i\hbar)^{-1}J(Q)]. \quad (4.3)$$

Proof: It will be shown that for all integers $n \geqslant 0$,

$$\tilde{d}_n(Q;0) = (1/n!)[(i\hbar)^{-1}J(Q)]^n. \quad (4.4)$$

Given identities (4.2) and (4.4) the result (4.3) follows at once. In order to establish (4.4) use

$$\tilde{d}_n(Q;0) = S_n(f_2,Q;0).$$

Note the various simplifications of (3.4) that occur if $u=0$.

(a) All terms on its right-hand side with $l>0$ vanish since they contain a multiplicative factor u^l and the remaining functions in the integrand have no negative powers of u.

(b) The measure $\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$ of Eq. (2.13) has the $u=0$ form

$$\Lambda^{(n)}(\mathbf{j}_r,\mathbf{t}_n)\Big|_{u=0} \equiv \Lambda_0^{(n)}(\mathbf{j}_r,\mathbf{t}_n) = \nu(t_1)\times\cdots\times|\gamma(t_{j_1})|\cdots\times|\gamma(t_{j_r})|\cdots\times\nu(t_n). \quad (4.5)$$

On the right-hand side of (4.5) all measures $\rho_i(t_i)$, $i\notin\mathbf{j}_r$, in $\Lambda^{(n)}(\mathbf{j}_r,\mathbf{t}_n)$ have been replaced by their $u=0$ values $\nu(t_i)$.

(c) For $u=0$ we have $f_2=1$, and for $l=0$ the function Φ has value 1. Furthermore the sum over $\mathbf{q}_r$ has one term $\mathbf{q}_r = \mathbf{j}_r$ so that

$$\Psi(\mathbf{q}_r;\boldsymbol{\alpha}_n)|_{l=0} = \prod_{i=1}^{r} \eta(t_{j_i},\alpha_{j_i})\cdot(y-x). \quad (4.6)$$

Taken together properties (a)–(c) imply that

$$\tilde{d}_n(Q;0) = (i\hbar)^{-n}\sum_{r=0}^{n}\sum_{\mathbf{j}_r}(t-s)^{n-r}\int_< d\xi_n \times\int d\Lambda_0^n(\mathbf{j}_r,\mathbf{t}_n(\xi_n))\exp(ib_n)\Psi(\mathbf{j}_r;\boldsymbol{\alpha}_n)|_{l=0}.$$

Observe the simple product form now assumed by the r, $\mathbf{j}_r$ sum

$$\sum_{r,\mathbf{j}_r} d\Lambda_0^n(\mathbf{j}_r,\mathbf{t}_n(\xi_n))\Psi(\mathbf{j}_r,\boldsymbol{\alpha}_n)(t-s)^{-r} = \prod_{j=1}^{n}\left[d\nu(t_j(\xi_n)) - \frac{x-y}{t-s}\cdot d\gamma(t_j(\xi_n))\right].$$

Since $\exp(ib_n)$ factors into a product of exponentials each with argument being $i\alpha_j\cdot[y+\xi_j(x-y)]$, the n-fold multiple integral over $\boldsymbol{\alpha}_n$ becomes a product of n integrals, viz.,

$$\tilde{d}_n(Q;0) = (i\hbar)^{-n}\int_< d\xi_n\int\prod_{j=1}^{n}\{[(t-s)d\nu(t_j(\xi_n)) - (x-y)\cdot d\gamma(t_j(\xi_n))]\exp(i\alpha_j\cdot[y+\xi_j(x-y)])\}$$

$$= (i\hbar)^{-n}\int_< d\xi_n\prod_{j=1}^{n}[(t-s)v(w(\xi_j)) - (x-y)\cdot a(w(\xi_j))] \quad (4.7a)$$

$$= \frac{(i\hbar)^{-n}}{n!}\left\{\int_0^1 d\xi\,[(t-s)v(w(\xi)) - (x-y)\cdot a(w(\xi))]\right\}^n. \quad (4.7b)$$

Employing the identities (2.5a) and (2.5b) shows that the second equality follows from the first. The integrand of (4.7a) is invariant with respect to permutations among the set $\{\xi_1,\ldots,\xi_n\}$. Using this symmetry leads to (4.7b), which is just the statement (4.4). □

The next task is to establish that the propagator K is the pointwise solution of (1.2) which in the limit $t\to s^+$ obeys the initial condition (1.3). As a preliminary it is useful to recall the following multidimensional stationary phase asymptotic formula.

Lemma 4: Let $\delta>0$ and let h: $\mathbb{R}^d\times\mathscr{U}_<(\delta)\to\mathbb{C}$ be a function satisfying the following.

(i) There exists a compact set $Y\subset\mathbb{R}^d$, whose interior contains supp $h(\cdot,\lambda)$ for all $\lambda\in\mathscr{U}_<(\delta)$.

(ii) For every d-component multi-index γ of length $|\gamma|\leqslant d$, the partial derivatives $\nabla^\gamma h$: $\mathbb{R}^d\times\mathscr{U}_<(\delta)\to\mathbb{C}$ exist and are continuous.

For $0\neq\lambda\in\mathscr{U}_<(\delta)$ define the complex-valued integral

$$I(\lambda)\equiv\left(\frac{1}{\pi i\lambda}\right)^{d/2}\int_{\mathbb{R}^d}\exp\left[\frac{iy^2}{\lambda}\right]h(y,\lambda)dy.$$

Then

$$\lim_{\lambda\to 0} I(\lambda) = h(0,0). \quad (4.8)$$

Proof: We omit the details that demonstrate this familiar result. For real values of λ the principal term $h(0;0)$ of (4.8) is the same as that found in Theorem 2.2 of Fedoriuk's review article[14] on the stationary phase method and pseudodifferential operators. However, our hypotheses differ from those of Ref. 14. The result above is most easily obtained by treating the multidimensional integral on the left-hand side of (4.8) as d iterated one-dimensional integrals and then applying iteratively Fedoriuk's method for deriving the stationary phase expansion for one-dimensional integrals. The y differentiability of h is necessary since the proof depends on integration by parts with respect to y. □

We denote the nth-order continuously differentiable functions of compact support on $\mathbb{R}^d$ by $C_0^n(\mathbb{R}^d)$.

Proposition 2: Let $u_0<(2ekT\gamma_T)^{-1}$ and $m^{-1}\in\mathscr{U}_<(u_0)\setminus\{0\}$. The propagator K of Theorem 2 satisfies the Schrödinger partial differential equation

$$i\hbar\frac{\partial}{\partial t}K(x,t;y,s;m) = H(x,-i\hbar\nabla,t,m)K(x,t;y,s;m) \quad (4.9a)$$

identically for all $(t,s;x,y)\in T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d$. Furthermore, for all $\phi\in C_0^d(\mathbb{R}^d)$,

$$\lim_{t\to s^+}\int K(x,t;y,s;m)\phi(y)dy = \phi(x),\quad x\in\mathbb{R}^d,\quad s\in[0,T). \quad (4.9b)$$

Proof: It is most convenient to use representation (1.4) for K, wherein F is defined by series (3.7). Fix $s\in[0,T)$. Theorem 3 tells us that within the domain $\mathbb{R}^d\times(s,T)\times\mathbb{R}^d$, $F(x,t;y,s;m^{-1})$ has continuous partial derivatives to first order in t and to second order in x. In view of the explicit form (1.5) for $K_0(t-s;x-y;m)$ it follows that both the left- and right-hand sides of (4.9a) exist and are continuous.

Suppose $\phi\in D_0\cap L_0^2(\mathbb{R}^d)$. Let the function Φ: $\mathbb{R}^d\times(s,T)\to\mathbb{C}$ be the integral

$$\Phi(x,t) = \int K(x,t;y,s;m)\phi(y)dy. \quad (4.10)$$

From Theorems 1 and 2 one has that $\Phi(\cdot,t)\in L^2(\mathbb{R}^d)$ and

$$U(t,s;m)\phi = \Phi(\cdot,t),\quad t\in(s,T). \quad (4.11)$$

We claim that $\Phi(\cdot,t)\in C^2(\mathbb{R}_x^d)$, $t\in(s,T)$; $\Phi(x,\cdot)\in C^1(s,T)$ for each $x\in\mathbb{R}^d$; and furthermore that

$$\nabla_x^\gamma \Phi(x,t) = \int \nabla_x^\gamma K(x,t;y,s;m)\phi(y)dy, \quad |\gamma|\leqslant 2, \qquad (4.12a)$$

$$\frac{\partial}{\partial t}\Phi(x,t) = \int \frac{\partial}{\partial t} K(x,t;y,s;m)\phi(y)dy. \qquad (4.12b)$$

The proof of (4.12) uses the theorem in analysis justifying the interchange of partial derivative of a parameter and integration. Consider (4.12a) first. Under the hypotheses that (i) for a.a. $y\in\mathbb{R}^d$, $K(\cdot,t;y,s;m)\phi(y)$ is $C^2(\mathbb{R}_x^d)$; (ii) for every $x\in\mathbb{R}^d$, and γ such that $|\gamma|\leqslant 2$, $\nabla_x^\gamma K(x,t;\cdot,s;m)\phi(\cdot)$ is $L^1(\mathbb{R}_y^d)$; and (iii) for every compact $X\subset\mathbb{R}^d$ and γ with $|\gamma| = 2$, there exists a function $g_X^\gamma\in L^1(\mathbb{R}_y^d)$ such that for all $(x,y)\in X\times\mathbb{R}^d$, $|\nabla_x^\gamma K(x,t;y,s;m)\phi(y)|\leqslant g_X^\gamma(y)$, then $\Phi(\cdot,t)\in C^2(\mathbb{R}_x^d)$ and (4.12a) is valid for every $x\in\mathbb{R}^d$ and every γ with $|\gamma|\leqslant 2$ (cf. Ref. 15, Appendix B.3). The specific form of K_0 and Proposition 1(b) shows that (i) is valid. Requirements (ii) and (iii) are also a consequence of Proposition 1(b). In particular, note that for $|\gamma|\leqslant 2$, $\nabla_x^\gamma F(x,t;y,s;m^{-1})$ is $C(\mathbb{R}_x^d\times\mathbb{R}_y^d)$, thereby $\nabla_x^\gamma K(x,t;y,s;m)$ has a uniform bound, $A(\gamma,X)<\infty$ on the compact set $X\times\operatorname{supp}\phi$. This fact together with $\phi\in L^2(\mathbb{R}^d)$ suffices to establish (ii). In (iii) an acceptable choice for the x-independent $L^1(\mathbb{R}_y^d)$ majorizing function $g_X^\gamma(y)$ is $A(\gamma,X)|\phi(y)|$. Thus $\Phi(\cdot,t)\in C^2(\mathbb{R}_x^d)$ and (4.12a) holds. A revision of this argument using Proposition 1 (c) and (d) verifies (4.12b) for each $(x,t)\in\mathbb{R}^d\times(s,T)$.

Consider the effect of $H(t,m)$ on $\Phi(\cdot,t)$. By (4.11), (2.3a), and (4.12a) one has that $\Phi(\cdot,t)\in D_0\cap C^2(\mathbb{R}_x^d)$. On functions of this class the action of $H(t,m)$ is the same as that of the classical differential operator $H(x,-i\hbar\nabla,t,m)$, i.e., for almost all x,

$$[H(t,m)\Phi(\cdot,t)](x) = H(x,-i\hbar\nabla,t,m)\Phi(x,t). \qquad (4.13a)$$

Identity (4.12a) shows that

$$[H(t,m)\Phi(\cdot,t)](x) = \int H(x,-i\hbar\nabla,t,m)K(x,t;y,s;m)\phi(y)dy. \qquad (4.13b)$$

A similar reduction is possible for the strong t derivative, $(d/dt)\Phi(\cdot,t)$, which exists by virtue of Definition 1(5) and Theorem 1. Letting $h_n(t)$, $t\in(s,T)$, be the difference quotient

$$h_n(t) = n[\Phi(\cdot,t+n^{-1}) - \Phi(\cdot,t)]$$

for suitably large integers n, then $(d/dt)\Phi(\cdot,t)$ is the strong limit in $\mathscr{H} = L^2(\mathbb{R}^d)$ of $h_n(t)$ as $n\to\infty$. From the $L^2(\mathbb{R}^d)$ convergent sequence $\{h_n(t)\}_n$ we may extract a subsequence $\{h_{n_j}(t)\}_j$ that converges pointwise almost everywhere (Ref. 16, Theorem 3.12). Since $\Phi(x,\cdot)\in C^1(s,T)$, this pointwise limit is just the classical partial derivative $\partial\Phi/\partial t$. In other words,

$$\left[\frac{d}{dt}\Phi(\cdot,t)\right](x) = \frac{\partial\Phi}{\partial t}(x,t) \quad (\text{a.a. } x). \qquad (4.14a)$$

Applying relation (4.12b) gives

$$\left[i\hbar\frac{d}{dt}\Phi(\cdot,t)\right](x) = i\hbar\int\frac{\partial}{\partial t}K(x,t;y,s;m)\phi(y)dy. \qquad (4.14b)$$

The $\mathscr{H}$-valued equation of motion (2.3e) requires that the left-hand sides of (4.13b) and (4.14b) are equal almost everywhere. By continuity it is then true for all x that

$$\int\left\{i\hbar\frac{\partial}{\partial t}K(x,t;y,s;m) - H(x,-i\hbar\nabla,t,m)\times K(x,t;y,s;m)\right\}\phi(y)dy = 0. \qquad (4.15)$$

For fixed x,t,s,m the function in the curly bracket is $C(\mathbb{R}_y^d)$ and thus it is $L^2(Y)$ on any compact subset $Y\subset\mathbb{R}_y^d$. Interpreting (4.15) as an inner product on $L^2(Y)$ shows that (4.9a) holds almost everywhere in $\mathbb{R}_y^d$. The y continuity of all the functions in (4.9a) show it holds everywhere in $T_\Delta^0\times\mathbb{R}^d\times\mathbb{R}^d$.

Finally consider statement (4.9b). Fix any $x\in\mathbb{R}^d$ and $s\in[0,T)$. In the integral appearing in (4.10) change the variable of integration by $y\to y+x$ and define $\lambda = 2\hbar(t-s)/m$. Then

$$\Phi(x,t) = (\pi i\lambda)^{-d/2}\int\exp\left(\frac{iy^2}{\lambda}\right)h(y,\lambda)dy = I(\lambda), \qquad (4.16a)$$

where

$$h(y,\lambda)\equiv F(x,s+|\lambda|\,|m|/(2\hbar);y+x,s;m^{-1})\phi(y+x) \qquad (4.16b)$$

is defined for all $\lambda\in\mathscr{U}_<(\delta)$ with $\delta = 2\hbar(T-s)/|m|$. Now h satisfies the hypothesis of Lemma 4 because (i) we may take Y as the closure of any open ball containing $-x+\operatorname{supp}\phi$, and (ii) the differentiability and continuity are a consequence of (4.16b), Proposition 1(b), and the hypothesis $\phi\in C_0^d(\mathbb{R}^d)$. Now take the limit $t\to s^+$ in (4.16a), which implies $\lambda\to 0$. Applying (4.8) it is seen that

$$\lim_{t\to s^+}\Phi(x,t) = h(0,0) = F(x,s;x,s)\phi(x) = \phi(x),$$

which is just result (4.9b). The last equality used (3.16)–(3.18), (4.3), and (1.7). □

We note that (4.9b) implies, in particular, that the family of linear functionals[17] $K_x^{t,s}$ ($(t,s)\in T_\Delta^0$, $x\in\mathbb{R}^d$) on $\mathscr{D}(\mathbb{R}^d) = C_0^\infty(\mathbb{R}^d)$ defined by

$$\langle K_x^{t,s},\phi\rangle = \int K(x,t;y,s;m)\phi(y)dy$$

satisfies

$$\lim_{t\to s^+}\langle K_x^{t,s},\phi\rangle = \phi(x) = \langle\delta_x,\phi\rangle,$$

where $\phi\in\mathscr{D}(\mathbb{R}^d)$ and δ_x is the Dirac delta function with support at x. This means that $K_x^{t,s}$ is a distribution in $\mathscr{D}'(\mathbb{R}^d)$ [the space dual to $\mathscr{D}(\mathbb{R}^d)$] and converges in the topology of $\mathscr{D}'(\mathbb{R}^d)$ to δ_x as $t\to s^+$. In other words, the propagator K of Theorem 2 constructs the fundamental solution of the Schrödinger equation on the time domain T_Δ with initial condition (1.3).

The function J is real for all arguments, thus $\exp[(i\hbar)^{-1}J(Q)]$ is a complex-valued unimodular function. This lets us use (4.1) as a definition of the function T: $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{U}_<(u_0)\to\mathbb{C}$ in terms of the known function F. Consider the gauge-dependent behavior of T. Let F_Λ, J_Λ, and T_Λ be the functions F, J, and T obtained from (3.7), (1.7), and (4.1) when $a(x,t)$ and $v(x,t)$ are replaced by the gauge transformed potentials $a(x,t;\Lambda)$ and $v(x,t;\Lambda)$. Choose the time displacement small enough so that both propagators $K(Q;m;\Lambda)$ and $K(Q;m)$ are defined by their constructive series (2.17). A simple calculation shows

$$J_\Lambda(Q) = J(Q) - \lambda(x,t) + \lambda(y,s). \qquad (4.17a)$$

Upon writing the gauge transformation identity (2.21) for propagators in terms of J,T and J_Λ, T_Λ it is evident that (4.17a) implies

$$T_\Lambda(Q;u) = T(Q;u), \quad \lambda \in \mathscr{G}(k), \qquad (4.17b)$$

i.e., that $T(Q;u)$ is the same function for all gauges in $\mathscr{G}(k)$. Clearly $J(Q)$ carries all the gauge dependence of the propagator K.

The functions T and F both suffice to completely determine the propagator K and both have well-defined small u expansions. However, the large mass asymptotic expansion of T is of greater interest for physical applications since T and the associated expansion coefficients are gauge invariant.

In order to prepare for the next theorem recall the following formulas from electromagnetism. The vector f: $\mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ will represent the electric force on the system plus the contribution from V,

$$f_i(x,t) = -(\nabla^i v)(x,t) - \partial a_i(x,t). \qquad (4.18a)$$

The magnetic part of the electromagnetic field tensor is given by

$$A_{ij}(x,t) = (\nabla^i a_j)(x,t) - (\nabla^j a_i)(x,t). \qquad (4.18b)$$

In (4.18) the indices i and j denote the Cartesian components of vectors and rank-2 tensors in $\mathbb{R}^d$. The symbol ∇^i is the partial derivative with respect to the ith component of the vector argument $x \in \mathbb{R}^d$. On the other hand, the notation ∇_1 (and Δ_1) describes the gradient (Laplacian) with respect to the first-vector argument of a function, e.g., $(\nabla_1 J)(Q)$ is the x gradient of J. Similarly, ∂ is a partial derivative with respect to a scalar argument, in (4.18a) the time argument.

The functions f_i and A_{ij} are well known to be gauge invariant quantities. In terms of (4.18) and the path $w(\xi)$ of (1.8), we define the function $\bar{f}$: $T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$\bar{f}_i(Q) = \int_0^1 d\xi\, \xi \left[(t-s) f_i(w(\xi)) + \sum_{j=1}^{d} (x-y)_j A_{ij}(w(\xi)) \right]. \qquad (4.19)$$

Note how similar in form (4.19) is to integral (1.7) defining J. The function $\bar{f}/(t-s)$ has the interpretation of a ξ-weighted average Lorentz force experienced by a system of classical particles moving with constant velocity $(x-y)/(t-s)$ from y to x. Because f_i and A_{ij} are gauge invariant so is $\bar{f}_i$. It is helpful to recall a basic identity linking $\bar{f}$ to J and a, namely

$$\bar{f}(Q) = -(\nabla_1 J)(Q) - a(x,t). \qquad (4.20)$$

Equation (4.20) is verified by using definition (1.7) for J and then forming the partial derivative $(\nabla_1 J)(Q)$. An integration by parts with respect to ξ leads to (4.20). For more details see the discussion prior to Eq. (4.11) in Ref. 1. Given that a and v are in class (A), it is evident that $\bar{f}_i$ is differentiable to arbitrary order in x, y and differentiable to first order in t and s on the domain $T_\Delta^0 \times \mathbb{R}^d \times \mathbb{R}^d$.

The principal result of this paper is the following.

Theorem 4: Assume the potentials a and v are in class (A) and that $u_0 < (2ekT\gamma_T)^{-1}$. For all integers $M \geq 1$ the gauge-invariant function T: $T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d \times \mathscr{U}_<(u_0) \to \mathbb{C}$ has the small u asymptotic expansion in $\mathscr{U}_<(u_0)$,

$$T(Q;u) = \sum_{j=0}^{M-1} u^j T_j(Q) + u^M \widetilde{E}_M(Q;u), \qquad (4.21)$$

where the coefficients T_j are defined in terms of P_j by

$$T_j(Q) = \exp[-(i\hbar)^{-1} J(Q)] P_j(Q) \quad (j = 0,\dots,M-1), \qquad (4.22a)$$

and the error function $\widetilde{E}_M$ in terms of E_M by

$$\widetilde{E}_M(Q;u) = \exp[-(i\hbar)^{-1} J(Q)] E_M(Q;u). \qquad (4.22b)$$

The functions T_j, $\widetilde{E}_M$ are the same functions for all gauges $\lambda \in \mathscr{G}(k)$ and have the same differentiability properties as P_j and E_M. The error term satisfies $|\widetilde{E}_M| = |E_M|$ and is bounded in $\mathscr{U}_<(u_0)$ by the u-independent estimate (3.18).

For all $Q \in T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d$, $T_0(Q) = 1$ and the higher-order coefficient functions $T_j(Q)$ are uniquely determined by the manifestly gauge invariant transport recurrence relation

$$\begin{aligned} T_j(Q) = \frac{(t-s)}{2} \int_0^1 d\xi \{ & i\hbar \Delta_1 T_{j-1}(w(\xi);y,s) \\ & - 2\bar{f}(w(\xi);y,s) \cdot \nabla_1 T_{j-1}(w(\xi);y,s) \\ & + [(i\hbar)^{-1} \bar{f}(w(\xi);y,s)^2 \\ & - \nabla_1 \cdot \bar{f}(w(\xi);y,s)] T_{j-1}(w(\xi);y,s) \}. \end{aligned} \qquad (4.23)$$

Proof: Expansion (4.21) is obtained from (3.16) by multiplication with $\exp[-(i\hbar)^{-1}J(Q)]$. Since a and v are in class (A) and J is defined by (1.7) and (1.8) it follows for each $(t,s) \in T_\Delta$ that $\exp[-(i\hbar)^{-1}J(Q)]$ is a $C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ function with partial derivatives in x and y that are (jointly) continuous functions on $T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d$. Furthermore, the phase function J has first-order derivatives with respect to t and s that are continuous on the domain $T_\Delta^0 \times \mathbb{R}^d \times \mathbb{R}^d$. Thus T, T_j, and $\widetilde{E}_M$ have the same differentiability properties as F, P_j, and E_M cited in Proposition 1 and Theorem 3. The coefficients T_j and error term $\widetilde{E}_M$ are gauge invariant because both T and the expansion parameter u are gauge invariant.

Consider the recurrence relation (4.23). For $(Q,u) \in T_\Delta^0 \times \mathbb{R}^d \times \mathbb{R}^d \times [\mathscr{U}_<(u_0) \setminus \{0\}]$ the representation

$$K(Q;u^{-1}) = K_0(t-s;x-y;u^{-1}) \times \exp[(i\hbar)^{-1} J(Q)] T(Q;u) \qquad (4.24)$$

follows from Proposition 1(e) and (4.1). In addition, Proposition 2 states that $K(Q;u^{-1})$ is the pointwise solution of (4.9a). The insertion of (4.24) into (4.9a) determines the partial differential equation satisfied by T,

$$\left[\frac{\partial}{\partial t} + (i\hbar)^{-1} (\partial_t J) \right] T$$

$$= \frac{u}{2}\{(i\hbar)\Delta_1 T - 2\bar{f}\cdot(\nabla_1 T) + [(i\hbar)^{-1}\bar{f}^2 - (\nabla_1\cdot\bar{f})]T\} - \left\{\frac{x-y}{t-s}\cdot\nabla_1 T - (i\hbar)^{-1}\left[\frac{x-y}{t-s}\cdot\bar{f} + v\right]T\right\}. \tag{4.25}$$

In obtaining the right-hand side of (4.25) we have employed the identity (4.20). In (4.25) the omitted arguments of the various functions are $T = T(Q;u)$, $v = v(x,t)$, $f = \bar{f}(Q)$, and $J = J(Q)$.

The next step is to substitute expansion (4.21) into (4.25) and then to equate the common coefficients of the differing powers of the independent functions u^j, $j = 0 \sim M-1$. In carrying out this substitution we rely on the facts that $T_j(Q)$ and $\tilde{E}_M(Q;u)$ have x derivatives of order 2 and t derivatives of order 1 that are continuous throughout the domain $T^0_\Delta \times \mathbb{R}^d \times \mathbb{R}^d \times \mathscr{U}_<(u_0)$. Furthermore, we use the fact that the x and t derivatives of $\tilde{E}_M(Q;u)$ are uniformly bounded in $\mathscr{U}_<(u_0)$ for each $Q \in T^0_\Delta \times \mathbb{R}^d \times \mathbb{R}^d$. The coefficient proportional to $(u)^0$ cancels identically by virtue of the definition of J. This mass-independent term is exceptional in that it is the only part of (4.25) that is gauge dependent.

The coefficient function $T_j(Q)$ is related to $T_{j-1}(Q)$ by the family [in the parameters $(y,s) \in \mathbb{R}^d \times [0,T)$] of partial differential equations

$$\left[(t-s)\frac{\partial}{\partial t} + (x-y)\cdot\nabla_x\right] T_j(x,t;y,s) = (t-s)g_{j-1}(x,t;y,s), \tag{4.26a}$$

where

$$g_{j-1}(x,t;y,s) = \tfrac{1}{2}\{(i\hbar)\Delta_x - 2\bar{f}\cdot\nabla_x - [\nabla_1\cdot\bar{f} - (i\hbar)^{-1}\bar{f}^2]\}T_{j-1}(x,t;y,s). \tag{4.26b}$$

The continuity of $\nabla_x^\gamma T_{j-1}$ for $0 \leqslant |\gamma| \leqslant 3$ on $T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d$ and the continuity of $\nabla_x^\gamma a$ for $0 \leqslant |\gamma| \leqslant 1$ on $\mathbb{R}^d \times [0,T]$ means that g_{j-1} is a continuously differentiable function of $(x,t) \in \mathbb{R}^d \times (s,T)$. If $j=1$, g_0 is determined from the known coefficient T_0.

In studying (4.26a) it is convenient to have a common Euclidean notation for the time and space variables. To this end set $x_{d+1} = t$ and $y_{d+1} = s$, whereby $(x,t) = (x_1,\ldots,x_{d+1})$ and $(y,s) = (y_1,\ldots,y_{d+1})$. In these variables (4.26a) reads

$$\sum_{i=1}^{d+1} (x_i - y_i)\frac{\partial}{\partial x_i} T_j = (x_{d+1} - y_{d+1})g_{j-1}, \quad j \geqslant 1. \tag{4.27a}$$

The (joint) continuity of T_j on the set $T_\Delta \times \mathbb{R}^d \times \mathbb{R}^d$ together with (4.22a) and (3.17) gives us the condition

$$\lim_{(x,t)\to(y,s)} T_j(x,t;y,s) = T_j(y,s;y,s) = 0, \quad j \geqslant 1. \tag{4.27b}$$

For each $j \geqslant 1$, (4.27a) is a linear first-order partial differential equation for T_j in the variables $\{x_i\}_1^{d+1}$, containing the parameters $\{y_i\}_1^{d+1}$. Finding the appropriate solution T_j of (4.26a) subject to the condition (4.27b) is similar to the Cauchy initial value problem[18] for this partial differential equation. We seek a solution in the region $\Omega = \mathbb{R}^d \times [s,T)$ of (x,t) space. Specifically, if the initial value manifold is chosen to be the plane, $\omega(\delta) \equiv \{(v,s+\delta) \in \Omega: v \in \mathbb{R}^d\}$, where $\delta \in (0, T-s)$, then the Cauchy problem is to determine the solution T_j of (4.27a) provided that the values T_j are given on the manifold $\omega(\delta) \subset \Omega$. In the problem faced here the step of assigning values of T_j on $\omega(\delta)$ is replaced by the limiting condition (4.27b).

The solution for this type of initial value problem can be obtained by the method of characteristics. The characteristic curves for (4.27a) are the solutions of the autonomous system of $d+2$ ordinary differential equations,

$$\frac{d}{d\tau} x_i(\tau) = x_i(\tau) - y_i \quad (i = 1 \sim d+1), \tag{4.28a}$$

$$\frac{d}{d\tau} z(\tau) = (x_{d+1}(\tau) - s)g_{j-1}(x_i(\tau);y,s). \tag{4.28b}$$

Because (4.27a) is linear, the equations (4.28a) decouple from (4.28b) and may be solved independently to provide the base characteristics in the space Ω. The initial conditions appropriate for the manifold $\omega(\delta)$ are $x_i(0) = v_i$ $(i = 1 \sim d)$, $x_{d+1}(0) = s+\delta$ and $z(0) = T_j(v,s+\delta;y,s)$. Upon introducing the change of variables $\xi = e^\tau$ the solutions of (4.28) are

$$x_i(\xi,v) = y_i + (v_i - y_i)\xi \quad (i = 1 \sim d), \tag{4.29a}$$

$$x_{d+1}(\xi,v) = s + \delta\xi, \tag{4.29b}$$

$$z(\xi,v) = T_j(v,s+\delta;y,s) + \int_1^\xi \delta g_{j-1}(x_i(\xi',v);y,s)d\xi'. \tag{4.29c}$$

Equations (4.29a) and (4.29b) represent a coordinate transformation $(\xi,v) \to (x_1,\ldots,x_{d+1}) = (x,t)$. The Jacobian of this transformation is $\delta\xi^d$. Thereby it is seen that only the point corresponding to $\xi = 0$ is singular. This point is $(x,t) = (y,s)$. The base characteristics defined by (4.29a) and (4.29b) are straight lines that emanate from the singular point at (y,s).

Given any point $(x,t) \in \Omega$ at which we wish to evaluate the solution T_j to (4.27), we may choose a coordinate $v(x,t)$ on the manifold $\omega(\delta)$ and $\xi(x,t) \in \mathbb{R}$ such that the corresponding base characteristic passes through (x,t), viz., $x_i(\xi(x,t),v(x,t)) = (x,t)$. Specifically, one finds

$$\xi(x,t) = (t-x)\delta^{-1}, \tag{4.29d}$$

$$v(x,t) = y + \delta(t-s)^{-1}(x-y). \tag{4.29e}$$

According to the general method of characteristics, the unique C^1 solution of the initial value problem is given by $T_j(x,t;y,s) = z(\xi(x,t),v(x,t))$. This result becomes, after employing (4.29d) and (4.29e) and then changing the integration variable to $\xi = \delta(t-s)^{-1}\xi'$,

$$T_j(Q) = T_j(y + \delta(t-s)^{-1}(x-y), s+\delta; y,s) + (t-s)\int_{\delta/(t-s)}^1 g_{j-1}(w(\xi;Q);y,s)d\xi. \tag{4.30a}$$

Now let $\delta \to 0^+$. The limiting condition (4.27b) shows that the second factor of T_j in (4.30a) vanishes while the continuity properties of g_{j-1} allow the lower limit of the integral term to be replaced by 0. Thus we obtain, for $t > s$, $j \geqslant 1$,

$$T_j(x,t;y,s) = (t-s)\int_0^1 g_{j-1}(w(\xi;Q);y,s)d\xi. \quad (4.30b)$$

This is just recurrence relation (4.23). The right-hand side of (4.30b) displays the base characteristic transport structure that is commonly found[1,19–21] in the determination of integral recurrence relations for expansion coefficients.

All the functions of ξ in the integral (4.23) are gauge invariant and thereby the recurrence relation is gauge invariant. For $j=0$, Lemma 3 shows that $T_0(Q)=1$, thus (4.23) recursively determines all $T_j(Q), j\geqslant 1$. □

The recurrence relations (4.23) allow one to obtain numerical bounds on the expansion coefficients T_j that are superior to those that result from their constructive series definition. Inspection of (4.23) reveals that, in order to estimate T_j, one also requires bounds on $\bar{f}$, T_{j-1}, and their first few spatial derivatives. Hence the recursive process requires estimates of arbitrary partial derivatives of $\bar{f}$ and T_j with respect to x, from the outset.

Bounds for $\bar{f}$, defined by the average (4.19), are based on the following simple bounds obeyed by the potentials. For example, using bounds (2.7) one finds (with γ a d-component multi-index)

$$|\nabla^\gamma v(x,t)| = \left|\int i^{|\gamma|}\alpha^\gamma e^{i\alpha\cdot x}\,d\nu(\tau)\right| \leqslant k^{|\gamma|}\nu_T,$$

and similarly

$$|\nabla^\gamma a_i(x,t)| \leqslant (\tfrac{1}{2}k)^{|\gamma|}\gamma_T, \quad |\nabla^\gamma \partial a_i(x,t)| \leqslant (\tfrac{1}{2}k)^{|\gamma|}\dot{\gamma}_T.$$

Using these simple facts it is not difficult to obtain the following bound for the components of $\bar{f}$. Setting Z to be

$$Z \equiv Z(|x-y|,t-s) = \sqrt{d}\,\gamma_T(1+k\,|x-y|) + (t-s)(k\nu_T + \dot{\gamma}_T),$$

then

$$|\nabla_1^\gamma \bar{f}_i(Q)| \leqslant k^{|\gamma|} Z(|x-y|,t-s).$$

Note that Z is γ independent and is a monotonically increasing function of its arguments. The final results, stated below, use the additional quantities

$$L \equiv \max\{Z,\hbar k\}, \quad \lambda_j \equiv (1+3^{j-1})^2.$$

Corollary 1: For $j\geqslant 1$ and any multi-index γ the coefficient functions T_j have the estimate

$$|\nabla_1^\gamma T_j(Q)| \leqslant k^{|\gamma|}[\tfrac{1}{2}(t-s)d\,]^j Z L^{j-1} \times [3^{|\gamma|}(Z/\hbar + \lambda_j k)]^j. \quad (4.31)$$

Proof (sketch): Estimate (4.31) is verified by induction on j. It is simple to verify (4.31) for $j=1$, and it may be extended to larger j by using (4.23). □

Solving the recurrence relation (4.23) for the coefficients T_1 and T_2 give us the following explicit formulas. First it is convenient to set

$$\Omega_i(z,\tau) \equiv \Omega_i(z,\tau;Q) = f_i(z,\tau) + (t-s)^{-1}(x-y)_j A_{ij}(z,\tau),$$

where $i = 1\sim d$ and (hereafter) the repeated index j is summed from 1 to d. Then

$$T_1(Q) = (2i\hbar)^{-1}(t-s)^3\int_{I^2} d\xi_1\,d\xi_2\,g(\xi_1,\xi_2)\Omega_j(w(\xi_1))\Omega_j(w(\xi_2)) - (2)^{-1}(t-s)^2\int_I d\xi_1(\xi_1-\xi_1^2)\left[\nabla\cdot f(w(\xi_1)) + \left(\frac{x-y}{t-s}\right)_j \nabla^i A_{ij}(w(\xi_1))\right]. \quad (4.32)$$

Here g is the one-dimensional Green's function defined in (3.3e), and $I^n = [0,1]^n$. The function T_2 is more elaborate. One finds

$$T_2(Q) = \frac{1}{2}T_1(Q)^2 + \sum_{l=0}^{2}(i\hbar)^{l-1}G_l(Q), \quad (4.33a)$$

where the functions G_l are given by

$$G_0(Q) = -\frac{1}{2}(t-s)^5\int_{I^3} d\xi_1\,d\xi_2\,d\xi_3\,g(\xi_1,\xi_2)\Omega_i(w(\xi_1))\Omega_j(w(\xi_3)) \times [g(\xi_2,\xi_3)\nabla^j\Omega_i(w(\xi_3)) + (t-s)^{-1}\partial_1 g(\xi_2,\xi_3)A_{ij}(w(\xi_2))], \quad (4.33b)$$

$$\begin{aligned} G_1(Q) = \frac{1}{4}(t-s)^4\int_{I^2} d\xi_1\,d\xi_2 g(\xi_1,\xi_2)\{&g(\xi_1,\xi_2)[\nabla^i\Omega_j(w(\xi_1)) \\ &+ (\xi_1(t-s))^{-1}F_{ji}(w(\xi_1))][\nabla^i\Omega_j(w(\xi_2)) + (\xi_2(t-s))^{-1}A_{ji}(w(\xi_2))] \\ &+ \xi_2^2\xi_>^{-1}(1-\xi_>)\Omega_j(w(\xi_1))[\Delta\Omega_j(w(\xi_2)) + 2(\xi_2(t-s))^{-1}\nabla^i A_{ji}(w(\xi_2))] \\ &+ \xi_2\xi_>^{-1}[2\xi_> - (\xi_> + 1)\xi_2]\Omega_j(w(\xi_1))[\nabla^j\nabla^i\Omega_i(w(\xi_2)) + (\xi_2(t-s))^{-1}\nabla^i A_{ij}((\xi_2))]\}, \end{aligned} \quad (4.33c)$$

$$G_2(Q) = -(8)^{-1}(t-s)^3\int_I d\xi_1[\xi_1(1-\xi_1)]^2[\Delta(\nabla\cdot\Omega)](w(\xi_1)). \quad (4.33d)$$

A few comments on the technical aspects of computing T_j using the recurrence relations are helpful. When T_{j-1} and its derivatives are substituted into the right-hand side of (4.23), one may employ the composition law for the linear path w,

$$w(\xi;w(\lambda;Q);y,s)=w(\xi\lambda;Q).$$

Then one can change the integration variables λ in the integrals arising from T_{j-1} to $\gamma=\xi\lambda$. It is then possible to perform explicitly the single integral over ξ that occurs in (4.23). In this way some of the Green's functions $g(\xi_i,\xi_j)$ appearing in (4.32) and (4.33) arise, cf. Ref. 1, p. 1704.

V. CONCLUSIONS

The two previous sections demonstrate that for each space-time coordinate $Q=(x,t;y,s)$, the propagator K admits the multiple factorization

$$K(Q;m)=K_0(t-s;x-y;m)\times\exp[(i\hbar)^{-1}J(Q)]T(Q;m^{-1}). \quad (5.1)$$

The $u=m^{-1}$ singular behavior of K as well as its gauge dependence are completely characterized by representation (5.1). Within the closed semidisk $\mathscr{U}_<(u_0)$ and for each value of Q with $t>s$, the propagator $K(Q;u^{-1})$ has only one singular point, namely an essential singularity at $u=0$. This singularity is entirely carried by the free propagator $K_0(t-s;x-y;m)$. The mass-independent unimodular phase factor $\exp[(i\hbar)^{-1}J(Q)]$ carries all the gauge dependence of $K(Q;m)$. As a result the function $T(Q;u)$ is gauge independent and sufficiently smooth in $\mathscr{U}_<(u_0)$ to be described by the asymptotic expansion of Theorem 4. Since u is the multiplier of the highest-order differential operator (the Laplacian) in the differential equation (4.9a) obeyed by $K(Q;u^{-1})$, the asymptotic expansion (4.21) together with (5.1) provide a detailed characterization of the singular perturbation behavior as $u\to0$ of the fundamental solution of the time-dependent Schrödinger equation (2.1).

The asymptotic expansion of K via (5.1) and (4.21) has a number of attractive features. It constructs an approximation for the propagator that is nonperturbative in the sense that it has contributions from the Dyson series to all orders, cf. (4.2). In addition the values of the potentials a, ϕ, and V appear only in $J(Q)$; all expansion coefficients $T_j(Q)$ are functions of derivatives of the fields. The expansion is valid for all $Q\in T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d$ to an arbitrary order M. The expansion has uniquely determined gauge invariant coefficients and error term both of which are uniformly bounded for Q taking values in compact subsets of $T_\Delta\times\mathbb{R}^d\times\mathbb{R}^d$. The expansion is robust (stable) in the sense that identity (4.21) may be differentiated to first order in t or s and to arbitrary order in x and y. The resultant identities are also valid asymptotic expansions with an error term whose order estimate remains $O(|u|^M)$. This stability feature of the expansion means that one has all the ingredients to determine from (4.21) the time evolving expectation values of operators that are representable as sums of partial derivatives with locally integrable coefficients. This class of operators includes most observables (self-adjoint operators) of interest in quantum mechanics.

Another interesting property of (4.21) is that contained within the small u expansion is a type of small $t-s$ expansion. Estimates (3.17) and (3.18) together with (4.22) or (4.30b) allow one to define

$$T_j(Q)=(t-s)^jT_j^*(Q), \quad (5.2a)$$

$$\widetilde{E}_M(Q;u)=(t-s)^ME_M^*(Q;u), \quad (5.2b)$$

where $T_j^*(Q)$ has (t,s) independent bounds, i.e., $T_j^*(Q)=O((t-s)^0)$ $(j\geqslant0)$. In this notation (4.21) assumes the form

$$T(Q;u)=\sum_{j=0}^{M-1}[u(t-s)]^jT_j^*(Q)+[u(t-s)]^ME_M^*(Q;u). \quad (5.2c)$$

Expansion (5.2c) is not a conventional double asymptotic expansion in the variables u and $t-s$ since there is a residual t and s dependence in T_j^*. However, in the case where a and v have higher-order derivatives in t it is not a difficult calculation to expand $T_j^*(Q)$ in powers of $t-s$ in order to find the short-time asymptotics of $T(Q;u)$ implied by (5.2c). In examining the small $t-s$ implications of representation (5.1) we have emphasized the behavior of $T(Q;u)$. It is also possible to expand the phase factor $\exp[(i\hbar)^{-1}J(Q)]$ for small $t-s$ but doing so will break the gauge invariance of the representation. A recent overview of the widely studied small time expansion of quantum propagators may be found in the review[22] of Fulling.

A particular advantage of having solved the evolution problem for complex mass is that the representation (5.1) is also capable of describing the equilibrium statistical physics of our N-body system. First observe that the constructive series (2.17) for the propagator K remains valid if the potential class (A) is enlarged to allow complex valued scalar potentials, i.e., the measure $\nu(t)$ for $v(x,t)$ is in $\mathscr{M}(S_k,\mathbb{C})$ rather than $\mathscr{M}^*(S_k,\mathbb{C})$. Suppose the interactions a, ϕ and V are static, and replace $q\phi+V$ in the Hamiltonian (1.1) by $i^{-1}(q\phi+V)$. If $m_0>0$ is the physical mass of a particle and after setting $t-s=\hbar\beta$, $s=0$, and $m^{-1}=u=(im_0)^{-1}\in\mathscr{U}_<(u_0)$ one finds that the Schrödinger equation (4.9a) becomes the Bloch equation which describes the equilibrium behavior of the N-body system interacting with static fields and having inverse temperature β. In this case the expansion (4.21) provides us with the large mass asymptotics of the fundamental solution of

$$-\frac{\partial}{\partial\beta}K(x,\hbar\beta;y,0;(im_0)^{-1})=\left[\frac{1}{2m_0}(-i\hbar\nabla_x-qa(x))^2+q\phi(x)+V(x)\right]\times K(x,\hbar\beta;y,0;(im_0)^{-1}). \quad (5.3)$$

Consider in detail the physical meaning of the phase J. The construction of the propagator in Theorem 2 requires only that the potentials a, ϕ, and V be in class (A). However, the circumstances of greatest interest in physics occur when the external fields arise as solutions of Maxwell's equations. Let index $i=1\sim N$ label particles whose coordinate positions are specified by $\mathbf{r}_i\in\mathbb{R}^3$, i.e., $x=(\mathbf{r}_1,...,\mathbf{r}_N)$. Each particle i interacts with a four-potential $\{\mathbf{a}_i(\mathbf{r}_i,t),\phi_i(\mathbf{r}_i,t)\}$.

All N four-potentials determine the total fields appearing in Hamiltonian (1.1) via

$$a(x,t) = (\mathbf{a}_1(\mathbf{r}_1,t),\ldots,\mathbf{a}_N(\mathbf{r}_N,t)), \tag{5.4a}$$

$$\phi(x,t) = \sum_{i=1}^{N} \phi_i(\mathbf{r}_i,t). \tag{5.4b}$$

Now assume that $\{\mathbf{a}_i,\phi_i\}_1^N$ are solutions of Maxwell's equations. Structurally J is similar to the well-known Dirac magnetic phase factor[23] that plays the central role in the Aharonov–Bohm effect.[24] If $\Gamma(\mathbf{r}',\mathbf{r})$ is a smooth directed path in $\mathbb{R}^3$ from initial point $\mathbf{r}'$ to final point $\mathbf{r}$ and $\mathbf{a}(\mathbf{r}'')$ is a static vector potential then the one-body Dirac phase factor is

$$\exp\left\{-\frac{q}{i\hbar}\int_{\Gamma(\mathbf{r}',\mathbf{r})} \mathbf{a}(\mathbf{r}'')\cdot d\mathbf{r}''\right\}. \tag{5.5}$$

In comparing J to the Dirac phase it is helpful to split J into two parts. Let J_V denote the contribution to J that is proportional to V, and write

$$J(Q) = J_L(Q) + J_V(Q). \tag{5.6}$$

Then J_L is defined solely in terms of the electromagnetic potentials. Let $\{\mathbf{w}_i(\xi),\tau(\xi)\}$ be the projection of path $w(\xi;x,t;y,s)$ onto the space-time coordinates of the ith particle. If $y = (\mathbf{r}_1',\ldots,\mathbf{r}_N')$ then $\mathbf{w}_i(\xi) = \mathbf{r}_i' + \xi(\mathbf{r}_i - \mathbf{r}_i')$ and $\tau = s + \xi(t-s)$ so

$$J_L(Q) = q\sum_{i=1}^{N}\int_0^1 [(t-s)\phi_i(\mathbf{w}_i(\xi),\tau(\xi)) - (\mathbf{r}_i - \mathbf{r}_i')\cdot\mathbf{a}_i(\mathbf{w}_i(\xi),\tau(\xi))]\,d\xi. \tag{5.7}$$

For each value of $\xi\in[0,1]$ the integrand above is a Lorentz scalar formed by the product of two Lorentz four-vectors. Inspecting (5.5) and (5.7) shows that J_L is an extended version of the Dirac phase in which time-dependent potentials are allowed and in which the scalar fields $\{\phi_i\}$ are adjoined in such a way that the phase J_L is a Lorentz scalar. The static Dirac path $\Gamma(\mathbf{r}',\mathbf{r}'')$ is extended to the linear space-time path $\{\mathbf{w}_i(\xi),\tau(\xi)\}$.

It may appear unexpected that Lorentz invariant features appear in a problem whose particle dynamics are strictly nonrelativistic. But of course phase J_L is an average of the electromagnetic potentials $\{\mathbf{a}_i,\phi_i\}$ with respect to the path $\{\mathbf{w}_i(\xi),\tau(\xi)\}$. The path integrals of the form (5.7) would define a Lorentz invariant for any smooth space-time path connecting y,s to x,t. The residual effect of our constructive solution is that it selects the particular path in J_L to be $w(\xi;Q)$.

As is well known, quantum systems exhibit semiclassical behavior if the particle mass is large. The semiclassical aspect of the representation (5.1) and its companion asymptotic expansion (4.21) is reflected in the geometrical character of the transport averages over the linear path $w(\xi;Q)$ that enter the phase factor $J(Q)$ and the expansion coefficients $T_j(Q)$. Let $\tau = s + \xi(t-s)$ be the running time variable, then (1.8) leads to

$$x(\tau) = y + [(\tau - s)/(t-s)](x-y). \tag{5.8}$$

Path (5.8) is the geodesic for the free evolution problem having initial point (y,s) and end point (x,t). Furthermore note that the one-dimensional Green's function $g(\xi,\xi')$ in (3.3e) and (4.32) is also a manifestation of this underlying two-point classical boundary value problem. These semiclassical properties have emerged directly from the exact Dyson series description of quantum evolution without the need to resort to any semiclassical ansatz about the analytic character of K near $u=0$.

It is interesting to contrast the large mass expansion of Theorem 4 with available results for the WKB approximation for the same dynamical system. A major difference between the $\hbar\to 0$ asymptotics of the propagator and the $u = m^{-1}\to 0$ asymptotics is that $\hbar$ appears only in the quantum evolution problem (1.2) whereas the variable mass parameter m enters both the quantum evolution and the companion classical evolution problems. In Ref. 1 a detailed comparison of the $\hbar\to 0$ and $u\to 0$ limits was used to formally obtain expansion (1.6) from the higher-order WKB representation of K. Such an approach is instructive in how the formula (1.6) emerges from the classical trajectories having two fixed end points that enter the WKB approximation but it suffers from the drawback that it is difficult to make rigorous. Here we have not had to make any ansatz concerning single valuedness of the action, the absence of caustics or the type of singular behavior the propagator has in the limit $\hbar\to 0$. We note finally that expansion (4.21) is relatively easy to use in the calculation of observables since the phase factor J and expansion coefficients T_i are explicitly given expressions of the fields whereas the analogous calculation in the WKB approximation for K requires one to first solve the difficult two point boundary value problems for the Hamiltonian dynamical equations in order to obtain the action function and expansion coefficients.

In the physics literature the factorization (5.1) has been postulated a number of times. The first detailed account is apparently that given by Valatin.[25] For a related representation of the propagator for the Dirac equation see Refs. 26 and 27. Some attempts[28,29] have been made to use the path phase factors like $J(Q)$ in the description of the time-dependent wave function $\Phi(x,t)$, cf. (4.10). However this is unnatural (and not very successful) since there is no geometrically distinguished initial point y,s for a wave function as there is in the case of the propagator K.

In the special case where the interaction is static and the evolution problem is that for the Bloch equation the large mass expansion is known to be equivalent to the Wigner–Kirkwood[30,31] approximation. Extensive discussions of the large mass expansions for the Bloch equation can be found in Refs. 32–34. A Wigner function analog of (5.1) is treated in Refs. 35 and 36.

In the theory of stochastic processes[37] results parallel to ours have been obtained. If $i\hbar$ is replaced by 1 throughout (1.2) a parabolic differential equation results. This equation can be investigated by the constructive Dyson series method in the same manner as the Schrödinger equation but it also can be studied, unlike the Schrödinger equation, through the specific methods of stochastic differential equations. The propagator of this parabolic equation has the interpretation of the transition probability density for the diffusion process starting form position y at time s and ending at x at time t. A singular perturbation problem that includes the $m \to \infty$ limit for the parabolic equation is investigated by Kifer.[38] Upon specializing Kifer's asymptotic expansion [Ref. 38, Eq. (4)] from a Riemannian to a Euclidean manifold, it can be shown equivalent to expansion (4.21).

ACKNOWLEDGMENTS

The authors have benefited from discussions of various portions of this material with Professor Balram Bhakar, Professor Roderick Wong, and Professor Joe Williams.

This work has been supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). F.H.M. gratefully acknowledges a postdoctoral fellowship awarded by NSERC and thanks Professor Robert Littlejohn's group at Berkeley for their hospitality.

[1]T. A. Osborn and F. H. Molzahn, Phys. Rev. A **34**, 1696 (1986).
[2]T. A. Osborn, L. Papiez, and R. Corns, J. Math. Phys. **28**, 103 (1987).
[3]F. J. Dyson, Phys. Rev. **75**, 1736 (1949).
[4]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II, Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[5]S. G. Krein, *Linear Differential Equations in Banach Space* (Am. Math. Soc., Providence, RI, 1971).
[6]H. Tanabe, *Equations of Evolution* (Pitman, London, 1979), Chap. 4.
[7]B. Simon, Bull. Am. Math. Soc. **7**, 447 (1982).
[8]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1984), 2nd ed.
[9]E. Nelson, J. Math. Phys. **5**, 332 (1964).
[10]P. R. Halmos and V. S. Sunder, *Bounded Integral Operators on L^2 Spaces* (Springer, Berlin, 1978).
[11]K.-H. Yang, Ann. Phys. (NY) **101**, 62 (1976). D. H. Kobe and K-H. Yang, Phys. Rev. A **32**, 952 (1985).
[12]K. Moriyasu, *An Elementary Primer for Gauge Theory* (World Scientific, Singapore, 1983).
[13]T. A. Osborn and Y. Fujiwara, J. Math. Phys. **24**, 1093 (1983).
[14]M. V. Fedoriuk, Usp. Mat. Nauk **26**:3, 67 (1971); Russ. Math. Surv. **26**:3, 65 (1971).
[15]B. C. Carlson, *Special Functions of Applied Mathematics* (Academic, New York, 1977).
[16]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1974).
[17]W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973), Chap. 6.
[18]F. John, *Partial Differential Equations* (Springer, New York, 1982), 4th ed.
[19]V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Dordrecht, 1981).
[20]V. Guillemin and S. Sternberg, *Geometrical Asymptotics* (Am. Math. Soc., Providence, RI, 1977).
[21]D. Fujiwara, J. Anal. Math. **35**, 41 (1979).
[22]S. A. Fulling, SIAM J. Math. Anal. **13**, 891 (1982).
[23]P. A. Dirac, Proc. R. Soc. London Ser. A **133**, 60 (1931).
[24]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[25]J. G. Valatin, Proc. R. Soc. London Ser. A **222**, 93 (1954).
[26]J. Schwinger, Phys. Rev. **82**, 664 (1951).
[27]C. Itzykson and J-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980), pp. 100–104.
[28]S. Mandelstam, Ann. Phys. (NY) **19**, 1 (1962).
[29]J. L. Friar and S. Fallieros, Phys. Rev. C **34**, 2029 (1986).
[30]E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[31]J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[32]E. Tirapegui, F. Langouche, and D. Roekaerts, Phys. Rev. A **27**, 2649 (1983).
[33]D. Bollé and D. Roekaerts, Phys. Rev. A **30**, 2024 (1984); **31**, 1667 (1985); **33**, 1427(E) (1986).
[34]V. V. Dodonov, V. I. Man'ko, and D. L. Ossipov, Physica A **132**, 269 (1985).
[35]F. J. Narcowich and S. A. Fulling, *Wigner Distribution Functions, Seminars in Mathematical Physics* (Texas A&M Univ., College Station, TX, 1986), No. 1.
[36]O. T. Serimaa, J. Javanainen, and S. Varro, Phys. Rev. A **33**, 2913 (1986).
[37]M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1984), Chap. 9.
[38]Yu. I. Kifer, Theor. Prob. Appl. **21**, 513 (1976).

Finitely many δ interactions with supports on concentric spheres

J. Shabani[a)]
International Centre for Theoretical Physics, Trieste, Italy and Institut de Physique Theorique, Université de Louvain, Louvain la Neuve, Belgium

(Received 18 March 1987; accepted for publication 24 June 1987)

Using the theory of self-adjoint extensions of symmetric operators the precise mathematical definition of the quantum Hamiltonian describing a finite number of δ interactions with supports on concentric spheres is given. Its resolvent is also derived, its spectral properties are described, and it is shown how this Hamiltonian can be obtained as a norm resolvent limit of a family of local scaled short-range Hamiltonians.

I. INTRODUCTION

Recently Antoine *et al.*[1] performed a rigorous and systematic study of the quantum Hamiltonian describing a δ interaction with support on a sphere in arbitrary dimensions $n \geqslant 2$.

In this paper we obtain a generalization of some of the results of Ref. 1 by considering a finite number N of δ interactions with supports on concentric spheres of radii $0 < R_1 < R_2 < \cdots < R_N$. In fact, using the techniques of Ref. 2 one can treat the case $N = \infty$.

In Sec. II we employ the theory of self-adjoint extensions of symmetric operators in order to give the precise mathematical meaning of the formal expression

$$H = -\Delta + \sum_{j=1}^{N} \alpha_j\, \delta(|x| - R_j), \tag{1.1}$$

where

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$$

is the Laplacian.

We show that (1.1) corresponds to the self-adjoint operator $H_{\{\alpha_j\},\{R\}}$ given by Eq. (2.12), i.e.,

$$H_{\{\alpha_j\},\{R\}} = \bigoplus_{l=0}^{\infty} U^{-1} h_{l,\{\alpha_j\},\{R\}}\, U \otimes 1.$$

In this section we also derive the resolvent of $H_{\{\alpha_j\},\{R\}}$.

Section III is devoted to the description of spectral properties of $h_{l,\{\alpha_j\},\{R\}}$ and finally in Sec. IV we show how $h_{l,\{\alpha_j\},\{R\}}$ can be obtained as a norm resolvent limit of a family of local scaled short-range Hamiltonians.

II. DEFINITION OF THE HAMILTONIAN

In this section we give (in dimension $n = 3$) the precise mathematical formulation of the quantum Hamiltonian describing N δ interactions with supports on concentric spheres of radii $0 < R_1 < R_2 \cdots < R_N$ formally given by

$$H = -\Delta + \sum_{j=1}^{N} \alpha_j\, \delta(|x| - R_j). \tag{2.1}$$

Consider in $L^2(\mathbb{R}^3)$ the closed, non-negative minimal operator

$$\dot{H} = \overline{-\Delta|_{C_0^\infty(\mathbb{R}^3 \setminus \bigcup_{j=1}^{N} \partial \overline{K(0,R_j)})}}, \tag{2.2}$$

where $\overline{K(0,R_j)}$ is a closed ball of radius R_j centered at the origin in $\mathbb{R}^3$.

Following, e.g., Ref. 3, p. 160, one can decompose $L^2(\mathbb{R}^3)$ with respect to angular momenta

$$L^2(\mathbb{R}^3) = L^2((0,\infty); r^2\,dr) \otimes L^2(S^2) \tag{2.3}$$

(S^2 is the unit sphere in $\mathbb{R}^3$), and introduce the unitary transformation

$$U: \begin{cases} L^2((0,\infty); r^2\,dr) \to L^2((0,\infty)), \\ f \to (Uf)(r) = rf(r), \quad r > 0, \end{cases} \tag{2.4}$$

in order to obtain the following decomposition of $L^2(\mathbb{R}^3)$:

$$L^2(\mathbb{R}^3) = \bigoplus_{l=0}^{\infty} U^{-1} L^2((0,\infty); dr) \otimes [Y_l^m], \quad l \in \mathbb{N}_0, \quad -l \leqslant m \leqslant l, \tag{2.5}$$

where $[Y_l^m]$ represents the linear span of the spherical harmonics.

With respect to the decomposition (2.5) $\dot{H}$ reads

$$\dot{H} = \bigoplus_{l=0}^{\infty} U^{-1} \dot{h}_{l,\{R\}}\, U \otimes 1, \tag{2.6}$$

where

$$\dot{h}_{l,\{R\}} = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2},$$

$$\begin{aligned} \mathscr{D}(\dot{h}_{l,\{R\}}) = \{ & f \in L^2((0,\infty)) \,|\, f, f' \in AC_{\text{loc}}((0,\infty)); \\ & f(0_+) = 0 \text{ if } l = 0;\ f(R_{j\pm}) = 0; \\ & -f'' + l(l+1) r^{-2} f \in L^2((0,\infty))\}, \\ & l \in \mathbb{N}_0, \quad 1 \leqslant j \leqslant N, \quad \{R\} = \{R_1, \ldots, R_N\}. \end{aligned} \tag{2.7}$$

Here $AC_{\text{loc}}((0,\infty))$ stands for the set of locally absolutely continuous functions on $(0,\infty)$, and

$$f(x_\pm) = \lim_{\varepsilon \to 0_+} f(x \pm \varepsilon).$$

The adjoint $\dot{H}^*$ of $\dot{H}$ is given by

$$\dot{H}^* = \bigoplus_{l=0}^{\infty} U^{-1} \dot{h}^*_{l,\{R\}}\, U \otimes 1, \tag{2.8}$$

[a)] On leave of absence from the Department of Mathematics, University of Burundi, BP 2700 Bujumbura, Burundi.

 0022-2488/88/030660-05$02.50

$$\dot h^*_{l,\{R\}} = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2},$$

$$\mathscr{D}(\dot h^*_{l,\{R\}}) = \{f \in L^2((0,\infty)) | f, f' \in AC_{loc}(0,\infty)\setminus\{R\});$$
$$f(0_+) = 0 \text{ if } l = 0;$$
$$f(R_{j+}) = f(R_{j-}) \equiv f(R_j);$$
$$-f'' + l(l+1)r^{-2} f \in L^2((0,\infty))\},$$
$$l \in \mathbb{N}_0, \quad 1 \leqslant j \leqslant N. \tag{2.9}$$

A straightforward calculation shows that the equation

$$\dot h^*_{l,\{R\}} \phi_l(k) = k^2 \phi_l(k), \quad \phi_l(k) \in \mathscr{D}(\dot h^*_{l,\{R\}}),$$
$$\operatorname{Im} k > 0,$$

has the solutions

$$\phi_{l,j}(k,r) = \begin{cases} (i\pi/2)\, R_j^{1/2} H^{(1)}_{l+1/2}(kR_j) r^{1/2} J_{l+1/2}(kr), & r \leqslant R_j, \\ (i\pi/2)\, R_j^{1/2} J_{l+1/2}(kR_j) r^{1/2} H^{(1)}_{l+1/2}(kr), & r \geqslant R_j, \quad \operatorname{Im} k > 0, \quad 1 \leqslant j \leqslant N, \end{cases} \tag{2.10}$$

where $J_\nu(z)$ and $H^{(1)}_\nu(z)$ are, respectively, Bessel and Hankel functions of order ν.[4] Thus the deficiency indices of $\dot h_{l,\{R\}}$ are (N,N) [we write $\operatorname{def}(\dot h_{l,\{R\}}) = (N,N)$], and consequently all self-adjoint (s.a.) extensions of $\dot h_{l,\{R\}}$ are given by an N^2-parameter family of s.a. operators.

In this paper we consider a special N-parameter family of s.a. extensions of $\dot h_{l,\{R\}}$ corresponding to the formal expression (2.1). The construction of the general (N^2-parameter) family of s.a. extensions of $\dot h_{l,\{R\}}$ will be reported elsewhere.[5]

We introduce in $L^2((0,\infty))$ the following family of closed extensions of $\dot h_{l,\{R\}}$:

$$h_{l,\{\alpha_l\},\{R\}} = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2},$$
$$\mathscr{D}(h_{l,\{\alpha_l\},\{R\}}) = \{f \in L^2((0,\infty)) | f, f' \in AC_{loc}((0,\infty)\setminus\{R\});$$
$$f(0_+) = 0 \quad \text{if} \quad l = 0; \quad f(R_{j-}) = f(R_{j+}) \equiv f(R_j),$$
$$f'(R_{j+}) - f'(R_{j-}) = \alpha_{jl} f(R_j), \quad -f'' + l(l+1)r^{-2} f \in L^2((0,\infty))\},$$
$$\{\alpha_l\} = \{\alpha_{1l}, \ldots, \alpha_{Nl}\}, \quad -\infty < \alpha_{jl} \leqslant \infty, \quad l \in \mathbb{N}_0, \quad 1 \leqslant j \leqslant N. \tag{2.11}$$

A simple integration by parts shows that $h_{l,\{\alpha_l\},\{R\}}$ is symmetric. Moreover, since $\operatorname{def}(\dot h_{l,\{R\}}) = (N,N)$ and the N boundary conditions in Eq. (2.11) are symmetric and linearly independent it follows from Ref. 6, Theorem XII, 4.30 that $h_{l,\{\alpha_l\},\{R\}}$ is self-adjoint.

The case $\alpha_{j_0 l} = \infty$ for some $j_0 \in \{1,\ldots,N\}$ in Eq. (2.11) describes a Dirichlet boundary condition at R_{j_0} while $\alpha_{jl} = 0$ for all $j = 1,\ldots, N$ (i.e., $\{\alpha_l\} = 0$) coincides with the free kinetic energy Hamiltonian $h_{l,0}$ for fixed angular momentum l.

By definition the operator $h_{\{\alpha_l\},\{R\}}$ defined in $L^2(\mathbb{R}^3)$ by

$$H_{\{\alpha_l\},\{R\}} = \bigoplus_{l=0}^{\infty} U^{-1} h_{l,\{\alpha_l\},\{R\}}\, U \otimes 1 \tag{2.12}$$

describes N δ interactions with supports on concentric spheres of radii $0 < R_1 < \cdots < R_N$. Actually $H_{\{\alpha_l\},\{R\}}$ provides a slight generalization of (2.1) since α_j: $1 \leqslant j \leqslant N$ may depend on $l \in \mathbb{N}_0$. If $\{\alpha_l\} = \infty$, then $H_{\infty,\{R\}}$ coincide with the Laplacian with Dirichlet boundary conditions at $\partial \overline{K(0,R_j)}$, $1 \leqslant j \leqslant N$. The case $\{\alpha_l\} = 0$ yields the free Hamiltonian

$$H_0 = -\Delta, \quad \mathscr{D}(H_0) = H^{2,2}(\mathbb{R}^3),$$

where $H^{2,2}(\Omega)$ is the Sobolev space of indices (2,2).[7]

The resolvent of $h_{l,\{\alpha_l\},\{R\}}$ is given by the following theorem.

Theorem 2.1: If $\alpha_{jl} \neq 0, j = 1,\ldots,N$, then the resolvent of $h_{l,\{\alpha_l\},\{R\}}$ is given by

$$(h_{l,\{\alpha_l\},\{R\}} - k^2)^{-1}$$
$$= (h_{l,0} - k^2)^{-1} + \sum_{j,j'=1}^{N} \mu_{jj'}(k) (\phi_{l,j'}(-\bar k), \cdot) \phi_{l,j}(k),$$
$$k^2 \in \rho(h_{l,\{\alpha_l\},\{R\}}), \quad \operatorname{Im} k > 0, \quad l \in \mathbb{N}_0 \tag{2.13}$$

[$\rho(\cdot)$ the resolvent set], where

$$[\mu(k)]^{-1}_{jj'} = -[\alpha_{jl}\,\delta_{jj'} + g_{l,k}(R_j, R_{j'})]^N_{jj'=1} \tag{2.14}$$

with

$$g_{l,k} = (h_{l,0} - k^2)^{-1}, \quad \operatorname{Im} k > 0, \tag{2.15}$$

the free resolvent with integral kernel;

$$g_{l,k}(r,r') = \begin{cases} (i\pi/2) r^{1/2} H^{(1)}_{l+1/2}(kr) r'^{1/2} J_{l+1/2}(kr'), & r' \leqslant r, \\ (i\pi/2) r'^{1/2} H^{(1)}_{l+1/2}(kr') r^{1/2} J_{l+1/2}(kr), & r' \geqslant r, \quad \operatorname{Im} k \geqslant 0. \end{cases} \tag{2.16}$$

Proof: Equation (2.13) except for the factors $\mu_{jj'}(k)$ follows from Krein's formula.[8] In order to determine the factors $\mu_{jj'}(k)$ we proceed as follows.

Let $g_l \in L^2((0,\infty))$ and define

$$\chi_l(k,r) = ((h_{l,\{\alpha_l\},\{R\}} - k^2)^{-1} g_l)(r), \tag{2.17}$$

where k is chosen in such a way that $\det \mu(k) \neq 0$. Since $\chi_l \in \mathscr{D}(h_{l,\{\alpha_l\},\{R\}})$, it follows from Eq. (2.11) that χ_l satisfies the following boundary conditions:

$$\chi_l \in AC_{loc}((0,\infty)), \tag{2.18}$$

$$\chi_l'(R_{j+}) - \chi_l'(R_{j-}) = \alpha_{jl}\,\chi_l(R_j), \tag{2.19}$$

$$((h_{l,\{\alpha_l\},\{R\}} - k^2)\,\chi_l)(r) = -\chi_l''(k,r) + (l(l+1)/r^2)\,\chi_l(k,r) - k^2\chi_l(k,r) = g_l(r), \quad r>0, \quad r\neq R_j, \quad 1\leqslant j\leqslant N. \tag{2.20}$$

The verification of the boundary conditions (2.18)–(2.20) gives the factors $\mu_{jj'}(k)$. The resolvent of $H_{\{\alpha_l\},\{R\}}$ may easily be obtained using Eqs. (2.12) and (2.13). We get

$$(H_{\{\alpha_l\},\{R\}} - k^2)^{-1} = (H_0 - k^2)^{-1} + \bigoplus_{l=0}^{\infty} \bigoplus_{m=-l}^{l} \sum_{j,j'=1}^{N} \mu_{jj'}(k) \times(|\cdot|^{-1}\phi_{l,j'}(-\bar{k})\,Y_l^m,\cdot) \times|\cdot|^{-1}\phi_{l,j}(k)\,Y_l^m, \quad k^2\in\rho(H_{\{\alpha_l\},\{R\}}), \quad \operatorname{Im} k>0, \quad l\in\mathbb{N}_0. \tag{2.21}$$

III. SPECTRAL PROPERTIES OF $h_{l,\{\alpha_l\},\{R\}}$

Spectral properties of $h_{l,\{\alpha_l\},\{R\}}$ are contained in the following theorem [$\sigma(\cdot)$, $\sigma_{\rm ess}(\cdot)$, $\sigma_{\rm ac}(\cdot)$, $\sigma_{\rm sc}(\cdot)$, $\sigma_{\rm p}(\cdot)$ denote the spectrum, essential spectrum, absolutely continuous spectrum, singularly continuous spectrum, and point spectrum, respectively.[9]]

Theorem 3.1: Assume $\alpha_{jl}\neq 0$, $1\leqslant j\leqslant N$. If all $\alpha_{jl}\neq\infty$, then $h_{l,\{\alpha_l\},\{R\}}$ has at most N eigenvalues which are all negative and simple. If $\alpha_{jl}=\infty$ for at least one $j\in\{1,...,N\}$, then $h_{l,\{\alpha_l\},\{R\}}$ has at most N negative eigenvalues (counting multiplicity) and infinitely many non-negative eigenvalues accumulating at ∞. The remaining part of the spectrum is purely absolutely continuous and covers the non-negative real axis

$$\sigma_{\rm ess}(h_{l,\{\alpha_l\},\{R\}}) = \sigma_{\rm ac}(h_{l,\{\alpha_l\},\{R\}}) = [0,\infty), \qquad \sigma_{\rm sc}(h_{l,\{\alpha_l\},\{R\}}) = \phi, \quad -\infty<\alpha_{jl}\leqslant\infty, \quad 1\leqslant j\leqslant N. \tag{3.1}$$

Proof: Since $\dot{h}_{l,\{R\}}\geqslant 0$ and $\operatorname{def}(\dot{h}_{l,\{R\}}) = (N,N)$ it follows from Ref. 10, p. 246, that $h_{l,\{\alpha_l\},\{R\}}$ has at most N negative eigenvalues counting multiplicity. Suppose $0<R_1<R_2<\cdots R_N$. If $|\alpha_{jl}|<\infty$, $1\leqslant j\leqslant N$, then following, e.g., Ref. 11, one can define

$$\psi_l(k,r) = \begin{cases} b r^{1/2} J_{l+1/2}(kr), & 0<r\leqslant R_1, \\ a_{m+1} r^{1/2} H^{(1)}_{l+1/2}(kr) + b_{m+1} r^{1/2} J_{l+1/2}(kr), & R_m\leqslant r\leqslant R_{m+1}, \quad 1\leqslant m\leqslant N-1, \\ a_{N+1} r^{1/2} H^{(1)}_{l+1/2}(kr) + b_{N+1} r^{1/2} J_{l+1/2}(kr), & r\geqslant R_N, \quad \operatorname{Im} k>0, \quad k\neq 0, \end{cases} \tag{3.2}$$

where a_{m+1} and b_{m+1} are unique nontrivial solutions of the system

$$\begin{aligned} & a_{m+1} H^{(1)}_{l+1/2}(kR_m) + b_{m+1} J_{l+1/2}(kR_m) = a_m H^{(1)}_{l+1/2}(kR_m) + b_m J_{l+1/2}(kR_m), \\ & a_{m+1} [r^{1/2} H^{(1)}_{l+1/2}(kr)]'_{r=R_m} + b_{m+1} [r^{1/2} J_{l+1/2}(kr)]'_{r=R_m} \\ & \quad - a_m [r^{1/2} H^{(1)}_{l+1/2}(kr)]'_{r=R_m} - b_m [r^{1/2} J_{l+1/2}(kr)]'_{r=R_m} \\ & \quad = \alpha_{ml} [a_{m+1} R_m^{1/2} H^{(1)}_{l+1/2}(kR_m) + b_{m+1} R_m^{1/2} J_{l+1/2}(kR_m)], \quad a_1=0, \quad b_1=b. \end{aligned} \tag{3.3}$$

A straightforward computation shows that the function $\psi_l(k,r)$ satisfies the boundary conditions

$$\psi_l(k,R_{j+}) = \psi_l(k,R_{j-}), \tag{3.4}$$

$$\psi_l'(k,R_{j+}) - \psi_l'(k,R_{j-}) = \alpha_{jl}\,\psi_l(k,R_j), \quad j=1,...,N. \tag{3.5}$$

Furthermore, the uniqueness of the coefficients a_{m+1} and b_{m+1}, $1\leqslant m\leqslant N$ implies that $\psi_l(k,r)$ is the unique solution of the differential equation

$$-\frac{d^2}{dr^2}\psi_l(k,r) + \frac{l(l+1)}{r^2}\psi_l(k,r) = k^2\psi_l(k,r), \quad r>0, \quad r\neq R_j, \quad j=1,...,N, \tag{3.6}$$

satisfying the boundary conditions (3.4) and (3.5). If $k^2>0$, then $\psi_l(k,r)\in L^2((0,\infty))$ if and only if $a_{N+1}=b_{N+1}=0$, i.e., $\psi_l(k,r)=0$. Since the same argument may be used for $k=0$, we conclude that

$$\sigma_{\rm p}(h_{l,\{\alpha_l\},\{R\}}) \subset (-\infty,0).$$

Suppose now that $k^2<0$. The simplicity of this eigenvalue follows from the uniqueness of $\psi_l(k,r)$. (We observe that $k^2<0$ is an eigenvalue of $h_{l,\{\alpha_l\},\{R\}}$ if and only if $b_{N+1}=0$.)

Consider now the case when exactly one $\alpha_{jl}=\infty$, e.g., $\alpha_{j_0l}=\infty$ and $N\geqslant 2$. The boundary condition at $r=R_{j_0}$ becomes a Dirichlet boundary condition and therefore divides $(0,\infty)$ into two independent intervals $(0,R_{j_0})$ and (R_{j_0},∞). The operator $h_{l,\{\alpha_l\},\{R\}}$ with $\alpha_{j_0l}=\infty$ is then a direct sum,

$$h_{l,\{\alpha_l\},\{R\}} = h^{(1)}_{l,\{\alpha_l\},\{R\}} \oplus h^{(2)}_{l,\{\alpha_l\},\{R\}}$$

acting in $L^2((0,\infty)) = L^2((0,R_{j_0}))\oplus L^2((R_{j_0},\infty))$ (and satisfying a Dirichlet boundary condition at $r=R_{j_0}$). Hence $h^{(1)}_{l,\{\alpha_l\},\{R\}}$ in $L^2((0,R_{j_0}))$ has a pure point spectrum in $(0,\infty)$ accumulating at ∞. The relation (3.1) follows from Weyl's theorem (Ref.12, p. 112) and the absence of singularly continuous spectrum follows, e.g., from Ref. 13, Lemma 2.4.

IV. APPROXIMATION OF $h_{l,\{\alpha_l\},\{R\}}$ BY A FAMILY OF LOCAL SCALED SHORT-RANGE HAMILTONIANS

In this section we show how $h_{l,\{\alpha_l\},\{R\}}$ can be obtained as a limit of a sequence of local scaled short-range Hamiltonians. Let λ_{jl}: $[0,\infty)\to\mathbb{R}$, $l\in\mathbb{N}_0$ be analytic near the origin with $\lambda_{jl}(0_+) = 0$ and denote by U_ϵ the unitary dilation group in $L^2((0,\infty))$, given by

$$(U_\epsilon f)(r) = \epsilon^{-1/2} f(r/\epsilon), \quad \epsilon > 0, \quad f \in L^2((0,\infty)). \tag{4.1}$$

For all $j = 1,\ldots,N$, let $V_j: \mathbb{R} \to \mathbb{R}$ be measurable, $V_j(r) \equiv 0$ for $r<0$, $V_j \in L^1((R,\infty))$, and define

$$v_j(r) = |V_j(r)|^{1/2}, \quad u_j(r) = |V_j(r)|^{1/2} \operatorname{sgn}[V_j(r)]. \tag{4.2}$$

Next we introduce

$$\widetilde{B}_{l,\epsilon}(k): L^2((0,\infty))^N \to L^2((0,\infty))^N,$$

$$[\widetilde{B}_{l,\epsilon}(k)(g_1,\ldots,g_N)]_j = \sum_{j'=1}^{N} \widetilde{B}_{l,\epsilon,jj'}(k)\, g_{j'}, \tag{4.3}$$

$$g_j \in L^2((0,\infty)),$$

where

$$\widetilde{B}_{l,\epsilon,jj'}(k) = \lambda_{jl}(\epsilon)\tilde{u}_j\, g_{l,k}\, \tilde{v}_{j'}, \quad \epsilon>0, \quad \operatorname{Im} k>0, \quad j,j' = 1,\ldots,N, \tag{4.4}$$

with

$$\tilde{u}_j(r) = u_j(r - (1/\epsilon)R_j), \quad \tilde{v}_j(r) = v_j(r - (1/\epsilon)R_j), \quad \epsilon>0, \quad j = 1,\ldots,N, \tag{4.5}$$

and $g_{l,k}$ given by Eq. (2.15).

Following, e.g., Ref. 11, one can show that $\widetilde{B}_{l,\epsilon,jj'}(k)$, $j,j' = 1,\ldots,N$, extend to Hilbert–Schmidt operators for $\operatorname{Im} k \geqslant 0$, $k \neq 0$.

Let us define the form sum[14] in $L^2((0,\infty))$:

$$h_l(\epsilon) = h_{l,0} \dotplus \sum_{j=1}^{N} \lambda_{jl}(\epsilon) V_j\Big(\cdot + \frac{1}{\epsilon} R_j\Big), \quad \epsilon>0, \tag{4.6}$$

with resolvent given by

$$(h_l(\epsilon) - k^2)^{-1} = g_{l,k} - \sum_{j,j'=1}^{N} (g_{l,k}\, \tilde{v}_j)\, [1 + \widetilde{B}_{l,\epsilon}(k)]^{-1} (\tilde{u}_{j'}, g_{l,k}),$$
$$\epsilon>0, \quad k^2 \in \rho(h_l(\epsilon)), \quad \operatorname{Im} k>0. \tag{4.7}$$

Next we define the Hamiltonian $h_{l,\epsilon}$ in $L^2((0,\infty))$:

$$h_{l,\epsilon} = \epsilon^{-2} U_\epsilon\, h_l(\epsilon)\, U_\epsilon^{-1} = h_{l,0} \dotplus \epsilon^{-2} \sum_{j=1}^{N} \lambda_{jl}(\epsilon) V_j\Big(\frac{(\cdot - R_j)}{\epsilon}\Big). \tag{4.8}$$

The scaling behavior

$$U_\epsilon\, g_{l,k}\, U_\epsilon^{-1} = \epsilon^{-2} g_{l,\epsilon^{-1}k}, \quad \operatorname{Im} k>0, \quad \epsilon>0, \tag{4.9}$$

and a translation $r \to r + \epsilon^{-1} R_j$, $\epsilon>0$, $j = 1,\ldots,N$, then yields

$$(h_{l,\epsilon} - k^2)^{-1} = \epsilon^2 U_\epsilon\, [h_l(\epsilon) - (\epsilon k)^2]^{-1} U_\epsilon^{-1}$$
$$= g_{l,k} - \epsilon^{-1} \sum_{j,j'=1}^{N} A_{l,\epsilon,j}(k)\, [1 + B_{l,\epsilon}(k)]_{jj'}^{-1} \times \lambda_{j'l}(\epsilon) C_{l,\epsilon,j'}(k), \quad \epsilon>0, \quad k^2 \in \rho(h_{l,\epsilon}), \quad \operatorname{Im} k>0, \tag{4.10}$$

where the Hilbert–Schmidt operators $A_{l,\epsilon,j}(k)$, $B_{l,\epsilon,jj'}(k)$, and $C_{l,\epsilon,j}(k)$ are defined through their integral kernels

$$A_{l,\epsilon,j}(k,r,r') = g_{l,k}(r, \epsilon r' + R_j) v_j(r'), \quad \operatorname{Im} k>0, \tag{4.11}$$

$$B_{l,\epsilon,jj'}(k,r,r') = \epsilon^{-1} \lambda_{jl}(\epsilon) u_j(r) \times g_{l,k}(\epsilon r + R_j, \epsilon r' + R_{j'}) v_{j'}(r'), \quad \operatorname{Im} k \geqslant 0, \tag{4.12}$$

$$C_{l,\epsilon,j}(k,r,r') = u_j(r) g_{l,k}(\epsilon r + R_j, r'), \quad \operatorname{Im} k>0. \tag{4.13}$$

Next we define the rank 1 operators $A_{l,j}(k)$, $B_{l,jj'}(k)$, and $C_{l,j}(k)$ via their integral kernels

$$A_{l,j}(k,r,r') = g_{l,k}(r, R_j) v_j(r'), \quad \operatorname{Im} k>0, \tag{4.14}$$

$$B_{l,jj'}(k,r,r') = \lambda'_{jl}(0) g_{l,k}(R_j, R_{j'}) u_j(r) v_{j'}(r'), \quad \operatorname{Im} k \geqslant 0, \quad k \neq 0, \tag{4.15}$$

$$C_{l,j}(k,r,r') = u_j(r) g_{l,k}(R_j, r'), \quad \operatorname{Im} k>0. \tag{4.16}$$

Lemma 4.1: For fixed k, $\operatorname{Im} k>0$; and for all $j = 1,\ldots,N$, $A_{l,\epsilon,j}(k)$, $B_{l,\epsilon,jj'}(k)$, and $C_{l,\epsilon,j}(k)$ converge in Hilbert–Schmidt norm to $A_{l,j}(k)$, $B_{l,jj'}(k)$, and $C_{l,j}(k)$, respectively, as $\epsilon \to 0_+$.

Proof: Using dominated convergence, one can easily show that

$$\operatorname*{w-lim}_{\epsilon\to 0_+} A_{l,\epsilon,j}(k) = A_{l,j}(k),$$
$$\operatorname*{w-lim}_{\epsilon\to 0_+} B_{l,\epsilon,jj'}(k) = B_{l,jj'}(k), \tag{4.17}$$
$$\operatorname*{w-lim}_{\epsilon\to 0_+} C_{l,\epsilon,j}(k) = C_{l,j}(k).$$

By Theorem 2.21 of Ref. 15 it suffices to prove

$$\lim_{\epsilon\to 0_+} \|A_{l,\epsilon,j}(k)\|_2 = \|A_{l,j}(k)\|_2,$$
$$\lim_{\epsilon\to 0_+} \|B_{l,\epsilon,jj'}\|_2 = \|B_{l,jj'}(k)\|_2, \tag{4.18}$$
$$\lim_{\epsilon\to 0_+} \|C_{l,\epsilon,j}(k)\|_2 = \|C_{l,j}\|_2,$$

which can be easily done, again using dominated convergence. □

Now we can state the main result of this section.

Theorem 4.2: For all $j = 1,\ldots,N$, let $V_j: \mathbb{R} \to \mathbb{R}$ be measurable, $V_j(r) \equiv 0$ for $r<0$, and $V_j \in L^1((R,\infty))$. Then $h_{l,\epsilon}$ converges in norm resolvent sense to $h_{l,\{\alpha_j\},\{R\}}$ as $\epsilon \to 0_+$, i.e., if $k^2 \in \rho(h_{l,\{\alpha_j\},\{R\}})$ then $k^2 \in \rho(h_{l,\epsilon})$ for ϵ small enough and

$$\operatorname*{n-lim}_{\epsilon\to 0_+} (h_{l,\epsilon} - k^2)^{-1} = (h_{l,\{\alpha_j\},\{R\}} - k^2)^{-1}, \tag{4.19}$$

where

$$\alpha_{jl} = \lambda'_{jl}(0) \int_R^\infty dr\, V_j(r), \quad l \in \mathbb{N}_0. \tag{4.20}$$

Proof: By (4.10) and Lemma 4.1 we obtain

$$\text{n-lim}_{\epsilon\to 0_+} (h_{l,\epsilon} - k^2)^{-1} = g_{l,k} - \sum_{j,j'=1}^{N} A_{l,j}(k) \times [1 + B_l(k)]^{-1}_{jj'} \lambda'_{j'}(0) C_{j'}(k), \quad k^2 \in \mathbb{C}\setminus\mathbb{R}, \quad \text{Im}\, k > 0, \tag{4.21}$$

where $B_l(k)$ is defined by

$$B_l(k):\ L^2((0,\infty))^N \to L^2((0,\infty))^N, \quad \text{Im}\, k \geqslant 0, \quad k \neq 0,$$

$$[B_l(k)(g_1,...,g_N)]_j = \sum_{j'=1}^{N} B_{l,jj'}(k) g_{j'}, \quad g_j \in L^2((0,\infty)), \quad 1 \leqslant j \leqslant N. \tag{4.22}$$

But

$$B_{l,jj'}(k) = \lambda'_{jl}(0) g_{l,k}(R_j, R_{j'}) (v_j, \cdot) u_j \tag{4.23}$$

implies

$$[1 + B_l(k)]^{-1}_{jj'} = 1\delta_{jj'} - \lambda'_{jl}(0) \sum_{m=1}^{N} g_{l,k}(R_j, R_m) \times [\hat{\mu}(k)]^{-1}_{mj'} (v_{j'}, \cdot) u_j, \tag{4.24}$$

where

$$\hat{\mu}(k) = [\delta_{jj'} + \lambda'_{jl}(0)(v_j, u_j) g_{l,k}(R_j, R_{j'})]^{N}_{j,j'=1}, \quad \text{Im}\, k > 0. \tag{4.25}$$

If $\lambda'_{jl}(0)(v_j, u_j) \neq 0$ for all $j = 1,...,N$ then a comparison with Eq. (2.14) shows that

$$[\hat{\mu}(k)]^{-1}_{jj'} \lambda'_{j'l}(0)(v_{j'}, u_{j'}) = -[\mu(k)]^{-1}_{jj'}, \quad \alpha_{jl} = \lambda'_{jl}(0)(v_j, u_j), \quad j, j' = 1,...,N, \tag{4.26}$$

which by (2.13) completes the proof after inserting (4.26), (4.14), and (4.16) into (4.21). ■

Formulas (4.21), (4.24), and (4.26) show that bound states (resp. resonances) of $h_{l,\{\alpha_l\},\{R\}}$ are given by zeros of the Fredholm determinant $\det[1 + B_l(k)]$ in the upper (resp. lower) k-half plane.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. He would also like to thank Professors S. Albeverio, Ph. Blanchard, and L. Streit for their invitation to the BiBoS Research Centre, University of Bielefeld, and Professor J. P. Antoine and Dr. F. Gesztesy for the joy of collaboration and for many interesting discussions.

This paper was partially supported by the Administration Générale de la Coopération au Développement, Belgium.

[1] J. P. Antoine, F. Gesztesy, and J. Shabani, J. Phys. A: Math. Gen. **20**, 3687 (1987).

[2] F. Gesztesy and W. Kirsch, J. Reine Angew. Math. **362**, 28 (1985).

[3] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, II—Fourier Analysis, Self-adjointness* (Academic, New York, 1975).

[4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Publisher, New York, 1972).

[5] L. Dabrowski and J. Shabani, in preparation.

[6] N. Dunford and J. T. Schwarz, *Linear Operators Part II—Spectral Theory* (Interscience, New York, 1963).

[7] R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).

[8] N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Pitman, Boston, 1981), Vol. 2.

[9] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, I—Functional Analysis* (Academic, New York, 1972).

[10] J. Weidmann, *Linear Operators in Hilbert Space* (Springer, New York, 1980).

[11] S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and H. Holden, "Solvable models in quantum mechanics," in *Texts and Monographs in Physics* (Springer, Berlin, 1987).

[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, IV—Analysis of Operators* (Academic, New York, 1978).

[13] S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and W. Kirch, J. Oper. Theory **12**, 101 (1984).

[14] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, NJ, 1971).

[15] B. Simon, *Trace Ideals and their Application* (Cambridge U. P., Cambridge, 1979).

The Pauli matrices in n dimensions and finest gradings of simple Lie algebras of type A_{n-1}

J. Patera and H. Zassenhaus
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada and Department of Mathematics, Ohio State University, Columbus, Ohio 43210

(Received 4 August 1987; accepted for publication 16 September 1987)

Properties of the Lie algebra gl(n,C) are described for a basis which is a generalization of the 2×2 Pauli matrices. The 3×3 case is described in detail. The remarkable properties of that basis are the grading of the Lie algebra it offers (each grading subspace is one dimensional) and the matrix group it generates [it is a finite group with the center of SL(n,C) as its commutator group].

I. INTRODUCTION

The purpose of this paper is to exploit recent results in mathematics[1,2] in order to generalize the 2×2 Pauli matrices to the $n\times n$ case. The generalization is unique up to normalization and change of basis. For $n = 3$ it is very different from the familiar generalization of the generators of the Lie algebra su(2) to su(3), known as the Gell-Mann matrices.[3]

We start by asking the following question: What is the most important property of the Pauli matrices? A definitive answer to this question cannot be given since "importance" is relative to the purpose one may have in mind, and because the familiar case of 2×2 Pauli matrices is too small in size to really appreciate the analogous properties for larger values of n. However, it is well known that the 2×2 Pauli matrices have other nontrivial properties besides spanning the Lie algebras su(2) and sl(2,C) (real and complex parameters, respectively). We list their properties in Sec. II. The generalization of the Pauli matrices is thus related to what one considers to be the defining important properties of these matrices.

In this paper we adopt the following point of view: The first one of the defining properties of what will henceforth be called the generalization of the Pauli matrices and denoted by $\mathscr{P}_n$ is that they provide a finest grading of the Lie algebra gl(n,C). The role $\mathscr{P}_n$ plays in grading gl(n,C) has two aspects: The adjoint action of $\mathscr{P}_n$ on gl(n,C) provides the grading group, and the generators of the graded gl(n,C) are found among the elements of $\mathscr{P}_n$.

The second defining requirement is that the set of $n\times n$ matrices $\mathscr{P}_n$ generates a subgroup of SL(n,C) with the center of SL(n,C) as its commutator subgroup. It simply means that the group commutator of $\mathscr{P}_n$ must be as large as possible given its role in the grading of gl(n,C). Throughout this paper we try to emphasize those basic properties of $\mathscr{P}_n$ that, in our opinion, should find a reflection in any lasting application of the results in physics.

Until now the role of the gradings in physical applications of Lie algebras and their representations were rarely noticed or emphasized except perhaps for the $\mathbb{Z}_2$ gradings underlying the classification of real forms of simple Lie algebras, the structure of superalgebras, and the Wigner–Inönü contractions of Lie algebras. Also the affinization $\widehat{A}$ of finite simple Lie algebra A involves an infinite $\mathbb{Z}$ grading of the algebra $\widehat{A}$. Implicitly another type of grading underlies the Cartan or root space decomposition of simple Lie algebras (finite and Kac–Moody ones).

The role of gradings of a Lie algebra in physics cannot be overestimated. In conventional terms it means the existence of preferred bases of the Lie algebra which admit additive quantum numbers. Naturally one wants to know all such bases and all nonequivalent choices available in a given situation. Moreover, such bases "force their way" into physics even if one is not set up to study them. Thus the matrices **A** and **D** below which generate $\mathscr{P}_{2n+1}$ are encountered in physics literature.[4]

In general terms a grading of a Lie algebra L means that L can be written as a direct sum of linear subspaces,

$$L = X_a \oplus X_b \oplus X_c \oplus \cdots, \quad a,b,c,...\in S, \tag{1.1}$$

labeled by a set S of finite sequences of integers or integers to a module $a = \{a_1,a_2,...,a_m\}$, $b = \{b_1,b_2,...,b_m\}$. The set S may be finite or infinite, there may be more than one integer labeling each subspace, etc.; the subspaces are supposed to be not zero, often even of dimension greater than 1. The decomposition (1.1) of L is called a grading provided the nonzero commutation relations of L have the following form:

$$[x_a, y_b] = z_{a+b}, \tag{1.2}$$

for any x_a, y_b of L for which $a,b\in S$, $x_a\in X_a$, $y_b\in X_b$, $[x_a, y_b] \neq 0$ so that $a+b\in S$, $z_{a+b}\in X_{a+b}$. Note that the m-tuple $a+b$ is formed componentwise and it must also be a part of the labeling set S. Practically grading L means to find generators of L and a labeling set S such that (1.2) is satisfied.

In the case of a $\mathbb{Z}_2$ grading the decomposition (1.1) contains exactly two subspaces labeled by integers mod 2. Such gradings most often can be refined to gradings with more than two components, they are coarse gradings. Of interest to us here are the fine gradings, where the sum in (1.1) contains as many subspaces as possible given the requirements of (1.2), among which are the finest gradings in case all subspaces X_i in (1.1) are one-dimensional. The finest gradings of A_n algebras are described here for the first time although we exploit results of Refs. 1 and 2.

Furthermore, it may be possible to grade simultaneously the Lie algebra and its representations, decomposing a representation space V of L into a direct sum of subspaces

 0022-2488/88/030665-09$02.50

$$V = V_d \oplus V_e \oplus V_f \oplus \cdots, \quad d,e,f \in S, \tag{1.3}$$

with the property

$$x_a | y_d \rangle \in V_{a+d}, \quad a,d,a+d \in S; \ | y_d \rangle \in V_d. \tag{1.4}$$

The relations (1.3), (1.4) contain (1.1), (1.2) as the particular case of the adjoint representation of L.

In quantum mechanics the labels of the set S are the admissible additive quantum numbers, which are eigenvalues of a chosen set of mutually commuting diagonable operators. In the case of a semisimple or reductive Lie algebra L, or of the Kac–Moody algebras, the traditional choice of the "diagonal" operators are the generators (i.e., a basis) of a Cartan subalgebra $\mathfrak{h} = \{h_1,h_2,...,h_r\}$. The remaining generators of the Lie algebra are then taken to be the eigenvectors of $\mathfrak{h}$. This is the traditional scenario which leads to the shift-up and shift-down generators similar to L_+ and L_- generators of the angular momentum theory. If the rank of L is r, then each label has r components. Such a label is called a weight of L; in the case of the Lie algebra these weights are the roots of the algebra, and the decomposition (1.1) of L is a grading called either the root space decomposition or the Cartan decomposition of L. Such a grading is fine but not the finest since $\dim \mathfrak{h} = r > 1$ for all but the 2×2 case. Note how restrictive the grading concept is in comparison with arbitrary decompositions of a Lie algebra into linear subspaces [cf. the matrices (2.2) below], that is, most decompositions do not admit a labeling of the generators with the property (1.2).

Our construction departs from the traditional approach by the observation that the 2×2 Pauli matrices generate a very particular maximal nilpotent subgroup $\mathscr{P}_2$ of SU(2), the quaternion group of order 2^3. This group is non-Abelian and therefore it is not a subgroup of the maximal torus of SL(2,$\mathbb{C}$). However, its adjoint action on the Lie algebra sl(2,$\mathbb{C}$) is Abelian and hence in many standard situations it can be used *instead* of the maximal torus.

Since the features of the general case appear already in the lowest case, $n = 3$, we describe them in detail for the 3×3 example in Secs. III and IV often leaving to the reader the verification of the properties by straightforward computation. In Sec. V the general $(2n+1)\times(2n+1)$ case is dealt with because it is somewhat simpler than the even size generalization presented in the last section. The 4×4 example is also briefly considered there.

The matrices $\mathscr{P}_n$ of any degree n provide a finest grading of A_{n-1}. But not every finest grading of A_{n-1} is conjugate under SL(n,$\mathbb{C}$) to the grading provided by the group $\mathscr{P}_n$. The general theory of Ref. 2 provides the answer that all finest gradings of A_{n-1} (with the exception of some low rank cases) are obtained upon using the Kronecker product groups

$$\mathscr{P}_{m_1} \otimes \mathscr{P}_{m_2} \otimes \cdots \otimes \mathscr{P}_{m_j}, \quad m_1 m_2 \cdots m_j = n.$$

An appropriate name for these matrices would be generalized Dirac matrices since the ordinary Dirac matrices correspond to $\mathscr{P}_2 \otimes \mathscr{P}_2$.

II. PROPERTIES OF THE PAULI MATRICES

The set of matrices

$$\sigma_0 = N' \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = N \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sigma_2 = N \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = N \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \tag{2.1}$$

with any complex nonzero normalization constants N, N' we shall call the Pauli matrices. Sometimes it is convenient to admit also the value $N' = 0$ and thus consider σ_1, σ_2, and σ_3 as the Pauli matrices without the identity matrix σ_0. In physics the most common normalization is $N' = 1$, $N = -i$, which makes all four matrices Hermitian.

A well known 3×3 analog of (1.1) are the Gell-Mann matrices,[3]

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{2.2}$$
$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$
$$\lambda_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrices (2.2) generalize (2.1) in that their $\mathbb{R}$- or $\mathbb{C}$-linear combinations span the Lie algebras u(3) and gl(3,$\mathbb{C}$), respectively, just as the Pauli matrices span the Lie algebras u(2) and gl(2,$\mathbb{C}$). However, the Pauli matrices have other remarkable properties not shared by (2.2). They are as follows.

(1) With $N = 1$, $N' = 1$ the Pauli matrices (2.1) (equipped with matrix multiplication) generate the maximal nilpotent subgroup $\mathscr{P}_2$ of SL(2,$\mathbb{C}$), a group of order 2^3. Explicitly the group $\mathscr{P}_2$ consists of the following elements:

$$\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \pm \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \pm \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \tag{2.3}$$

Note the coincidence of the centers of the groups $\mathscr{P}_2$, SL(2,$\mathbb{C}$), and SU(2). All the matrices (2.3), except the multiples of identity, belong to the same conjugacy class of SL(2,$\mathbb{C}$) elements of order 4 denoted[5] by [11]. It is the unique class of the lowest order regular elements. Note also that the Hermitian normalization of (2.1) would generate a finite group which is quite *different* from $\mathscr{P}_2$.

(2) The adjoint action of the Pauli matrices on themselves is diagonal and does not depend on $N \neq 0$, $N' \neq 0$:

$$\sigma_j \sigma_k \sigma_j^{-1} = \begin{cases} \sigma_k, & \text{if } j = k \text{ or } k = 0 \text{ or } j = 0, \\ -\sigma_k, & \text{if } 0 \neq j \neq k \neq 0. \end{cases} \tag{2.4}$$

Existence of the group $\mathscr{P}_n$ satisfying (2.4) and the irreducibility of $\mathscr{P}_n$ are the requirements defining the generalization of Pauli matrices in this paper.

Among the interesting consequences of (1) and (2) let us point out the following.

(3) With $N = i/2$ the commutation relations of (2.1) have integer structure constants. The normalization of σ_0 is irrelevant for this property since σ_0 commutes with all the others.

(4) Introducing the following notations for the generators of sl(2,$\mathbb{C}$):

$$\sigma_1 = (1,0), \quad \sigma_2 = (1,1), \quad \sigma_3 = (0,1),$$

the grading of the algebra is made obvious:

$$[(p,q)(p',q')] = \text{const}(p+p', q+q'), \tag{2.5}$$

where $p,q,p',q',p+p',q+q'$ are integers mod 2.

Let us note the following properties which find some reflection in the generalization.

(5) The Lie algebra su(2) [or sl(2,$\mathbb{C}$)] decomposes into a sum of one-dimensional real (or complex) subspaces generated by $\sigma_1,\sigma_2,\sigma_3$ each of which is a Cartan subalgebra. For $n=2$ this means that $\sigma_1,\sigma_2,\sigma_3$ are diagonable.

(6) The three Cartan subalgebras are pairwise orthogonal,

$$\text{tr}(\sigma_j\sigma_k) = 2N\delta_{jk} \quad (j,k = 1,2,3). \tag{2.6}$$

The properties listed above are not independent of each other. The general theory can be found in Ref. 2.

III. THE GENERALIZATION

We will repeatedly use in Secs. III and IV the constants $\omega = e^{2\pi i/3}$ and $\xi = e^{2\pi i/6}$ and the obvious identities they satisfy.

Consider the following 27 matrices:

$$A_k = \omega^k \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad A_k^- = \omega^{-k} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$B_k = \omega^k \begin{pmatrix} 0 & \omega & 0 \\ 0 & 0 & \omega^2 \\ 1 & 0 & 0 \end{pmatrix}, \quad B_k^- = \omega^{-k} \begin{pmatrix} 0 & 0 & \omega \\ 1 & 0 & 0 \\ 0 & \omega^2 & 0 \end{pmatrix},$$

$$C_k = \omega^k \begin{pmatrix} 0 & \omega^2 & 0 \\ 0 & 0 & \omega \\ 1 & 0 & 0 \end{pmatrix}, \quad C_k^- = \omega^{-k} \begin{pmatrix} 0 & 0 & \omega^2 \\ 1 & 0 & 0 \\ 0 & \omega & 0 \end{pmatrix},$$

$$D_k = \omega^k \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}, \quad D_k^- = \omega^{-k} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix},$$

$$I_k = \omega^k \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{with } k \text{ an integer mod 3.} \tag{3.1}$$

The set of matrices (3.1) is the 3×3 analog $\mathscr{P}_3$ of the group $\mathscr{P}_2$ of (2.3). Under matrix multiplication they form a subgroup of SL(3,$\mathbb{C}$) of order 3^3 whose center, $\{I_k, k\equiv 0, \pm 1 \bmod 3\}$, coincides with the center of both SL(3,$\mathbb{C}$) and SU(3). All but elements of the center belong to the unique SL(3,$\mathbb{C}$) conjugacy class [111] of lowest order regular elements.[5,6]

Any linearly independent subset of (3.1) is a basis of the Lie algebra gl(3,$\mathbb{C}$). Our choice of the sl(3,$\mathbb{C}$) linear generators will be (dropping the subscripts and writing the generators in bold characters)

$$\begin{aligned}
\mathbf{A} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, & \mathbf{B} &= \begin{pmatrix} 0 & \omega & 0 \\ 0 & 0 & \omega^2 \\ 1 & 0 & 0 \end{pmatrix}, \\
\mathbf{C} &= \begin{pmatrix} 0 & \omega^2 & 0 \\ 0 & 0 & \omega \\ 1 & 0 & 0 \end{pmatrix}, & \mathbf{D} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}, \\
\mathbf{A}^- &= \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, & \mathbf{B}^- &= \begin{pmatrix} 0 & 0 & \omega \\ 1 & 0 & 0 \\ 0 & \omega^2 & 0 \end{pmatrix}, \\
\mathbf{C}^- &= \begin{pmatrix} 0 & 0 & \omega^2 \\ 1 & 0 & 0 \\ 0 & \omega & 0 \end{pmatrix}, & \mathbf{D}^- &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix},
\end{aligned} \tag{3.2}$$

The Lie algebra gl(3,$\mathbb{C}$) is generated by (3.2) and by the identity matrix **I**. Note that the matrices **B** and $\mathbf{B}^-$, **C** and $\mathbf{C}^-$ are not inverse to each other, their products are multiples of the identity. Such a choice makes them a particular case of (5.6) below.

It would be possible from now on to consider only Hermitian (or anti-Hermitian) linear combinations of the generators (3.2), but this would reveal little of the general structure and introduces many cumbersome complications (as happens in the angular momentum theory) although it could prove useful in some applications, for instance where the pairwise orthogonality of the generators with respect to the Killing form is required. The Hermitian version of (3.2) is thus

$$\mathbf{A}_+ = \mathbf{A} + \mathbf{A}^- = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{B}_+ = \mathbf{B} + \omega^2\mathbf{B}^- = \begin{pmatrix} 0 & \omega & 1 \\ \omega^2 & 0 & \omega^2 \\ 1 & \omega & 0 \end{pmatrix},$$

$$\mathbf{A}_- = i(\mathbf{A} - \mathbf{A}^-) = \begin{pmatrix} 0 & i & -i \\ -i & 0 & i \\ i & -i & 0 \end{pmatrix}, \quad \mathbf{B}_- = i(\mathbf{B} - \omega^2\mathbf{B}^-) = \begin{pmatrix} 0 & -i\omega & i \\ i\omega^2 & 0 & -i\omega^2 \\ -i & i\omega & 0 \end{pmatrix},$$

$$\mathbf{D}_+ = \mathbf{D} + \mathbf{D} = \begin{pmatrix} 2 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \quad \mathbf{C}_+ = \mathbf{C} + \omega\mathbf{C}^- = \begin{pmatrix} 0 & \omega^2 & 1 \\ \omega & 0 & \omega \\ 1 & \omega^2 & 0 \end{pmatrix},$$
$$\mathbf{D}_- = i(\mathbf{D} - \mathbf{D}^-) = \sqrt{2}\begin{pmatrix} 0 & & \\ & 1 & \\ & & -1 \end{pmatrix}, \quad \mathbf{C}_- = i(\mathbf{C} - \omega\mathbf{C}^-) = \begin{pmatrix} 0 & -i\omega^2 & i \\ i\omega & 0 & -i\omega \\ -i & i\omega^2 & 0 \end{pmatrix}. \tag{3.2'}$$

The matrices (3.1), besides spanning the Lie algebra gl(3,ℂ) (under matrix commutation), form at the same time a finite subgroup of SL(3,ℂ) (under matrix multiplication). Thus any matrix (3.1) can be interpreted as a group element or a Lie algebra element. The two interpretations differ by the implied composition law: commutation and linear combinations for the Lie algebra, and matrix multiplication for the group.

The generators (3.2) make obvious a decomposition of the Lie algebra sl(3,ℂ) into a sum of four two-dimensional subspaces

$$\mathrm{sl}(3,\mathbb{C}) = \mathfrak{h}_A + \mathfrak{h}_B + \mathfrak{h}_C + \mathfrak{h}_D\,, \tag{3.3}$$

where the subspaces are spanned by two commuting generators,

$$\mathfrak{h}_A = \{\mathbf{A},\mathbf{A}^-\}, \quad \mathfrak{h}_B = \{\mathbf{B},\mathbf{B}^-\},$$
$$\mathfrak{h}_C = \{\mathbf{C},\mathbf{C}^-\}, \quad \mathfrak{h}_D = \{\mathbf{D},\mathbf{D}^-\}\,. \tag{3.4}$$

Hence each of the four subspaces is a Cartan subalgebra of sl(3,ℂ) and, taking suitable linear combinations of generators, also of the su(3). Furthermore, one easily verifies the pairwise orthogonality of the subspaces $\mathfrak{h}_A,\mathfrak{h}_B,\mathfrak{h}_C,\mathfrak{h}_D$ with respect to the Killing form,

$$\operatorname{tr} XY = 0, \quad \text{for } X\in\mathfrak{h}_X,\ Y\in\mathfrak{h}_Y,\ X \neq Y\,. \tag{3.5}$$

The commutation relations of the generators (3.2) are summarized in Table I. The nonzero structure constants are cyclotomic integers of the form

$$\xi^k + \xi^{k+1}\,. \tag{3.6}$$

Finally observe that the property (2.4) of the Pauli matrices also generalizes to higher ranks. Namely,

$$X_k Y_{k'} X_k^{-1} = \omega^j Y_{k'} \tag{3.7a}$$

or equivalently

$$X_k Y_{k'} = \omega^j Y_{k'} X_k \tag{3.7b}$$

and also

$$X_k Y_{k'} X_k^{-1} Y_{k'}^{-1} = I_j \tag{3.7c}$$

for any $X_k, Y_{k'} \in \mathscr{P}_3$. The factor ω^j is given in Table I as the power of ξ^{2j} in the structure constant in $[X,Y] = (\xi^{2j} + \xi^{2j+1})Z$.

The finite group $\mathscr{P}_3$ of the matrices (3.1) is obviously non-Abelian. Hence it is not a subgroup of the maximal torus of SL(3,ℂ). Nevertheless its action (3.7a) on the generators of sl(3,ℂ) is Abelian. As a result of that it can be used instead of the maximal torus in many ways.

IV. SOME FURTHER PROPERTIES

A. The cyclotomic quarks and antiquarks

In (3.2) we have a new basis of sl(3,ℂ) with unique properties. Now let us consider the elementary representation theory in terms of the new basis.

The natural (quark) representation of the generators of sl(3,ℂ) coincides with (3.2). Let us choose the basis vectors (quarks) of the three-dimensional representation space as the eigenvectors of the generator **D**, label them by the power $p \pmod 6$ of ξ in the eigenvalue ξ^p of **D**, and call them the cyclotomic quarks (most of the relevant numbers related to them in the representation theory are cyclotomic integers). Thus we have the quarks

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad |4\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.1}$$

defined by

$$\mathbf{D}|p\rangle = \xi^p|p\rangle, \quad p = \text{even integer mod } 6\,. \tag{4.2}$$

One verifies directly that

$$\mathbf{D}|p\rangle = \xi^p|p\rangle, \quad \mathbf{D}^-|p\rangle = \xi^{-p}|p\rangle\,,$$
$$\mathbf{A}|p\rangle = |p-2\rangle, \quad \mathbf{A}^-|p\rangle = |p+2\rangle\,,$$

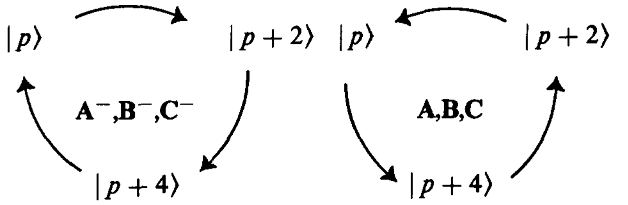
TABLE I. The commutation relations of the sl(3,ℂ) generators (3.2). The 0 blocks on the diagonal indicate the presence of the generators of four Cartan subalgebras in our basis. Only the upper part of the table is shown.

	A	$\mathbf{A}^-$	**B**	$\mathbf{B}^-$	**C**	$\mathbf{C}^-$	**D**	$\mathbf{D}^-$
A	0	0	$(1+\xi)\mathbf{C}^-$	$(1+\xi^5)\mathbf{D}^-$	$(1+\xi^5)\mathbf{B}^-$	$(1+\xi)\mathbf{D}^-$	$(1+\xi)\mathbf{B}^-$	$(1+\xi^5)\mathbf{C}$
$\mathbf{A}^-$	0	0	$(1+\xi^5)\mathbf{D}$	$(1+\xi)\mathbf{C}$	$(1+\xi)\mathbf{D}$	$(1+\xi^5)\mathbf{B}$	$(1+\xi^5)\mathbf{C}^-$	$(1+\xi)\mathbf{B}^-$
B			0	0	$(\xi^4+\xi^5)\mathbf{A}^-$	$(\xi^2+\xi)\mathbf{D}$	$(1+\xi)\mathbf{C}$	$(1+\xi^5)\mathbf{A}$
$\mathbf{B}^-$			0	0	$(\xi^2+\xi)\mathbf{D}$	$(\xi^4+\xi^5)\mathbf{A}$	$(1+\xi^5)\mathbf{A}^-$	$(1+\xi)\mathbf{C}^-$
C					0	0	$(1+\xi)\mathbf{A}$	$(1+\xi^5)\mathbf{B}$
$\mathbf{C}^-$					0	0	$(1+\xi^5)\mathbf{B}^-$	$(1+\xi)\mathbf{A}^-$
D							0	0
$\mathbf{D}^-$							0	0

$$\mathbf{B}|p\rangle = \xi^{p}|p-2\rangle,\quad \mathbf{B}^{-}|p\rangle = \xi^{-p}|p+2\rangle,$$
$$\mathbf{C}|p\rangle = \xi^{-p}|p-2\rangle,\quad \mathbf{C}^{-}|p\rangle = \xi^{p}|p+2\rangle. \tag{4.3}$$

The relations (4.3) are rewritten in (4.15) below in a compact form using different notation for the generators.

Note the "rotating" action of **A**, **B**, **C** and that of $\mathbf{A}^-$, $\mathbf{B}^-$, $\mathbf{C}^-$ on the quarks and the fact that during commutation the rotations add up. Neither of the generators is a "shift-up" or "shift-down" operator. Symbolically one has



(4.4)

In order to consider other representations, say the antiquark one, it is helpful to distinguish between the abstract generators, which we denote by $A, B, C, D, A^-, B^-, C^-, D^-$, and their representations. Thus the matrices (3.2) stand for the sl(3,C) generators in the quark representation q. The abstract generators in the antiquark representation $\bar{q}$ are represented by matrices which are the negative transpose of those of (3.2). In particular,

$$q(D) = \mathbf{D},\quad \bar{q}(D) = -\mathbf{D}^T = -\mathbf{D}. \tag{4.5}$$

Therefore the antiquarks $|p\rangle$ are

$$|1\rangle,\quad |3\rangle,\quad |5\rangle, \tag{4.6}$$

defined by

$$-\mathbf{D}|p\rangle = \xi^{p}|p\rangle,\quad p = \text{odd integer mod } 6. \tag{4.7}$$

Thus to every quark $|p\rangle$ there corresponds an antiquark $|p+3\rangle$. Transformation properties of the antiquarks analogous to (4.3) are given in (4.15c).

B. The finest grading of sl(3,C)

Before proceeding further in this direction, it is useful to consider a grading of the Lie algebra sl(3,C) unique to our basis.

The rotating action of the generators on the quarks allows one to decompose gl(3,C) into three subspaces L_d, $d = 0, \pm 1$ mod 3, spanned by the generators

$$L_1 = \{\mathbf{A}^-,\mathbf{B}^-,\mathbf{C}^-\},\quad L_0 = \{\mathbf{D},\mathbf{D}^-,\mathbf{I}\},$$
$$L_{-1} = \{\mathbf{A},\mathbf{B},\mathbf{C}\}, \tag{4.8}$$

with the grading property

$$[L_r, L_s] \subseteq L_{r+s \bmod 3}. \tag{4.9}$$

The subspaces (4.8) can be defined as eigenspaces of the adjoint action (3.7a) of the generator **D**,

$$L_d = \{X \,|\, \mathbf{D}X\mathbf{D}^{-1} = \omega^d X\}. \tag{4.10}$$

Thus (4.10) allows one to label each gl(3,C) generator by an integer d which can take three values. However, (3.7a) is valid not only for **D** but for any generator (3.2). Therefore we can use any other generator of sl(3,C), or all of them simultaneously, and label the generators by up to eight three-valued integers. In order to label completely all the generators without redundancy of notation, it suffices to use any two of them which do not commute. Choosing in addition to **D**, for instance, the generator **A**, and using its eigenvalues to label the generators, we end up with a new notation for the generators,

$$\begin{aligned} &\mathbf{A} = (0,-1), && \mathbf{A}^- = (0,1,), && \mathbf{D} = (1,0,),\\ &\mathbf{B} = (1,-1,), && \mathbf{B}^- = (-1,1,), && \mathbf{D}^- = (-1,0,),\\ &\mathbf{C} = (-1,-1,), && \mathbf{C}^- = (1,1,), && \mathbf{I} = (0,0,), \end{aligned} \tag{4.11}$$

where the first label refers to **A** and the second to **D** [cf. (4.10)]. The Abelian property of the adjoint action (3.7a) assures the grading structure of the commutation relations

$$[(k,j),(k',j')] = \text{const}(k+k',j+j') \bmod 3 \tag{4.12}$$

of gl(3,C) with the structure constants given as before in Table I. Since no two generators are labeled by the same symbol in (4.11), the grading (4.12) of gl(3,C) cannot be further refined. We say that it is fine. Note that (4.9) is a coarsening of (4.12) obtained when one ignores the first label. The decomposition (3.3), however, is not a grading. Moreover, since the subspaces $\{(i,j)\}$ generated by each (i,j) are one dimensional the grading is finest. We then have the fine decomposition of the Lie algebra gl(3,C) into a sum of one-dimensional subspaces:

$$\text{gl}(3,\mathbb{C}) = \sum_{i,j=-1}^{1} \{(i,j)\}. \tag{4.13}$$

Finally, note that the grading (4.11) allows us to write the commutation table (Table I) in a compact form. Namely,

$$[(k,j),(k',j')] = (\omega^{kj'} - \omega^{k'j})(k+k',j+j') \quad \text{mod } 3, \tag{4.14}$$

and that the transformation properties (4.3) of quarks by the generators (r,s) of (4.11) including the matrix elements can be written in a simple form:

$$(r,s)|p\rangle = \xi^{rp}|p+2s\rangle. \tag{4.15a}$$

In (3.2) and (4.11) we have identified the abstract generators (r,s) with their matrix (quark) representation $q(r,s)$. Without such convention the relations (4.15a) should have been written as

$$q(r,s)|p\rangle = \xi^{rp}|p+2s\rangle \quad (r,s \text{ mod } 3;\ p \text{ even mod } 6). \tag{4.15b}$$

The corresponding relations in the antiquark representation $\bar{q}(r,s)$ of the generators are then

$$\bar{q}(r,s)|p\rangle = -\xi^{(p-3)r}|p-2s\rangle \quad (r,s \text{ mod } 3;\ p \text{ odd mod } 6). \tag{4.15c}$$

C. The gl(2,C) and o(3,C) subalgebras of sl(3,C)

There are two maximal subalgebras of gl(3,C) which are often used. Let us now write their generators in our basis of gl(3,C).

First note that the 3×3 matrices E_{ij}, $i,j = 1,2,3$, with 1 at the intersection of the ith row and jth column and 0 elsewhere, can be written as follows:

$$E_{ii} = \frac{1}{3}\sum_{m=1}^{3} \omega^{(1-i)m}\mathbf{D}^m,\quad E_{ik} = E_{ii}\mathbf{A}^{k-i}. \tag{4.16}$$

The subalgebras are now generated for instance by

$$\text{gl}(2,\mathbb{C})\text{: } E_{32}=E_{33}\mathbf{A}^-,\quad E_{23}=E_{22}\mathbf{A},$$
$$E_{22}-E_{33}=\tfrac{1}{3}(\xi^4+\xi^5)(\mathbf{D}-\mathbf{D}^-)\,, \tag{4.17}$$
$$2E_{11}-E_{22}-E_{33}=\mathbf{D}+\mathbf{D}^-\,;$$

$$\text{o}(3,\mathbb{C})\text{: } E_{12}+E_{23}=\tfrac{1}{3}(2+\xi^{-1}\mathbf{D}+\xi\mathbf{D}^-)\mathbf{A}\,,$$
$$E_{21}+E_{32}=-\tfrac{1}{3}(2+\mathbf{D}+\mathbf{D}^{-1})\mathbf{A}^{-1}\,, \tag{4.18}$$
$$E_{11}-E_{33}=\tfrac{1}{3}((1+\xi^{-1})\mathbf{D}+(1+\xi)\mathbf{D}^{-1})\,.$$

D. The Weyl group and the weight lattice

Among the most important tools of the general representation theory is the Weyl group W and the weight lattices and weight systems of representations. We finish this section by pointing them out in the new basis.

The sl(3,$\mathbb{C}$) weight lattice is usually given as the integer span of the two fundamental weights,

$$Q=\mathbb{Z}\nu_1+\mathbb{Z}\nu_2\,. \tag{4.19}$$

Here $\mathbb{Z}$ denotes any integer. In our notations the fundamental weights are written as the highest weights of the quarks and antiquarks,

$$\nu_1=1 \quad\text{and}\quad \nu_2=\xi\,. \tag{4.20}$$

Hence the weight lattice Q consists of all the points

$$Q=\mathbb{Z}+\mathbb{Z}\xi=\mathbb{Z}+\mathbb{Z}\omega\,. \tag{4.21}$$

The Weyl group action in Q is generated by two reflections,

$$\begin{aligned} r_1(a+b\xi)&=r_1(a+b+b\omega)\\ &=-a+(a+b)\xi=b+(a+b)\omega\,,\\ r_2(a+b\xi)&=r_2(a+b+b\omega)\\ &=a+b-b\xi=a-b\omega\,. \end{aligned} \tag{4.22}$$

In particular all quark sl(3,$\mathbb{C}$) quantum numbers (weights) are found on the same Weyl group orbit,

$$\begin{aligned} &\xi^0=1\leftrightarrow 0\,,\quad \xi^2=r_1\xi^0=-1+\xi\leftrightarrow 2\,,\\ &\xi^4=r_2r_1\xi^0=-\xi\leftrightarrow 4\,. \end{aligned} \tag{4.23}$$

Similarly one finds the antiquarks on another orbit,

$$\begin{aligned} &\xi\leftrightarrow 1\,,\quad \xi^3=r_1r_2\xi=-1\leftrightarrow 3\,,\\ &\xi^5=r_2\xi=1-\xi\leftrightarrow 5\,. \end{aligned} \tag{4.24}$$

The standard representation theory can be developed in terms of this basis, irreducible representations are constructed in tensor products of the quark and antiquark ones, etc.

V. THE GENERAL CASE OF gl(2n+1,$\mathbb{C}$)

The properties of gl(3,$\mathbb{C}$) described in Secs. III and IV are particular cases of those which will be described here. Similar properties of gl(2n,$\mathbb{C}$) also exist; however, some modification is necessary there. They are described in Sec. VI.

The finite group $\mathscr{P}_{2n+1}$ represented as a group of $(2n+1)\times(2n+1)$ matrices of determinant 1 is generated by the cyclic permutation matrix

$$A=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0\\ & & 1 & 0 & 0\\ \vdots & & & \ddots & \vdots\\ 0 & & & & 1\\ 1 & 0 & & \cdots & 0 \end{pmatrix},\quad A^{2n+1}=I\,, \tag{5.1}$$

and by the diagonal matrix

$$D=\text{diag}\,\{1,\zeta,\zeta^2,...,\zeta^{2n}\},\quad \zeta=e^{2\pi i/2n+1}\,. \tag{5.2}$$

The group consists of $(2n+1)^3$ matrices given by

$$K_{kad}=\zeta^kA^aD^d,\quad k,a,d\in\mathbb{Z}_{2n+1}\,, \tag{5.3a}$$

which means that k,a,d assume integral values mod$(2n+1)$. Equivalently we could have chosen

$$K'_{kad}=\zeta^kD^dA^a,\quad k,a,d\in\mathbb{Z}_{2n+1}\,, \tag{5.3b}$$

instead of (5.3a). The transfer between the two conventions is made as follows. Because

$$ADA^{-1}=\zeta D\Leftrightarrow DAD^{-1}=\zeta^{-1}A\,, \tag{5.4}$$

one has

$$A^aD^dA^{-a}=\zeta^{ad}D^d\Leftrightarrow D^dA^aD^{-d}=\zeta^{-ad}A^a \tag{5.5}$$

and therefore

$$K'_{kad}=\zeta^{ad}K_{kad}\,. \tag{5.6}$$

Rewriting (5.5) in terms of K_{kad}, we establish easily the crucial property of the group $\mathscr{P}_{2n+1}$ which generalizes (2.4) and (3.7). Namely,

$$K_{kad}K_{k'a'd'}(K_{kad})^{-1}=\zeta^{ad'-a'd}K_{k'a'd'}\,,\qquad a,a',d,d'\in\mathbb{Z}_{2n+1}\,. \tag{5.7}$$

Linear combinations of the matrices (5.3) with complex coefficients span the Lie algebra gl$(2n+1,\mathbb{C})$. A suitable set of generators can be chosen, for example, by putting $k=0$ in (5.3a). To be specific we choose the generators

$$\mathbf{K}_{ad}=K_{0ad},\quad -n\leqslant a,d\leqslant n\,. \tag{5.8}$$

In particular, the one-dimensional center of gl$(2n+1,\mathbb{C})$ is generated by the identity matrix

$$\mathbf{K}_{00}=K_{000}\,;$$

the matrices A and D are also among the generators

$$A=\mathbf{K}_{10}=K_{010}\,,\quad D=\mathbf{K}_{01}=K_{001}\,.$$

Moreover, the subgroup of SL$(2n+1,\mathbb{C})$ generated by **A**,**D** has as its commutator subgroup the whole center of SL$(2n+1,\mathbb{C})$.

When it is possible to decompose sl$(2n+1,\mathbb{C})$ into the algebraic sum

$$\text{sl}(2n+1,\mathbb{C})=\mathfrak{h}+\sum_{d=-n}^{n}\mathfrak{h}_d \tag{5.9}$$

of $2n+2$ Cartan subalgebras? It can be done, according to a conjecture in Ref. 1, if and only if $2n+1$ is a prime power. If $2n+1$ is a prime number then we find the following solution for which we conjecture that our solution is the only one that can be refined to a finest grading:

$$\mathfrak{h}_d=\{(\mathbf{K}_{1d})^a,\ 1\leqslant a\leqslant 2n\}\,,$$

while $\mathfrak{h}$ is the Cartan subalgebra of diagonal matrices,

$$\mathfrak{h}=\{K_{0d},1\leqslant d\leqslant 2n\}=\{D,D^2,D^3,...,D^{2n}\}\,. \tag{5.10}$$

The property (5.7) specialized for the generators, i.e., $k=k'=0$, allows one to label the generators by the eigenvalues of other generators acting as in (5.7). Thus the $(2n+1)^2$ basis elements of $\mathrm{gl}(2n+1,\mathbb{C})$ can each be labeled by $(2n+1)^2$ eigenvalues. Avoiding redundancy of notation, it suffices to use eigenvalues of any two generators which generate $\mathscr{P}_{2n+1}$ upon multiplication. Our choice of labeling generators from now on is A and D.

A generator $\mathbf{K}_{ad}$ is labeled by the eigenvalues of the transformations

$$A\,\mathbf{K}_{ad}A^{-1}=\zeta^{d}\mathbf{K}_{ad},\quad D\,\mathbf{K}_{ad}D^{-1}=\zeta^{-a}\mathbf{K}_{ad}\,. \tag{5.11}$$

For simplicity of notation we write

$$\mathbf{K}_{ad}=(d,-a)\,. \tag{5.12}$$

Here $-a$ and d are integers $\mathrm{mod}(2n+1)$. Note that each generator of $\mathrm{gl}(2n+1,\mathbb{C})$ is labeled by a distinct pair $(d,-a)$. The identity $\mathbf{K}_{00}$ is labeled by (0,0).

Consider the commutation relations

$$[\mathbf{K}_{pq},\mathbf{K}_{p'q'}]=[(q,-p),(q',-p')]$$
$$=(q,-p)(q',-p')-(q',-p')(q,-p)\,.$$

Since

$$(q,-p)(q',-p')=A^{p}D^{q}A^{p'}D^{q'}$$
$$=A^{p+p'}A^{-p'}D^{q}A^{p'}D^{q'}$$
$$=\zeta^{-p'q}A^{p+p'}D^{q+q'}\,,$$

all the commutation relations of our generators of $\mathrm{gl}(2n+1,\mathbb{C})$ can be written in the explicit form

$$[A^{a}D^{d},A^{a'}D^{d'}]=(\zeta^{-a'd}-\zeta^{-ad'})A^{a+a'}D^{d+d'}\,, \tag{5.13a}$$

$$[(a,b)(a',b')]=(\zeta^{a'b}-\zeta^{ab'})(a+a',b+b')\,, \tag{5.13b}$$

where the addition of the generator labels a,b,a',b' is understood $\mathrm{mod}(2n+1)$. The finest grading of $\mathrm{gl}(2n+1,\mathbb{C})$ realized by our basis (5.10) is made obvious in (5.11). Note that (5.11) is valid also for $\mathrm{sl}(2n+1,\mathbb{C})$ which requires the exclusion of (0,0) from the set of generators of the algebra.

There are $2n$ Casimir operators of $\mathrm{sl}(2n+1,\mathbb{C})$. In our basis they are written in an obvious way. Indeed,

$$C^{(2)}=\sum_{\substack{p_1+p_2=0\\ q_1+q_2=0}}(p_1,q_1)(p_2,q_2)\,;$$

$$C^{(3)}=\sum_{\substack{p_1+p_2+p_3=0\\ q_1+q_2+q_3=0}}(p_1,q_1)(p_2,q_2)(p_3,q_3)\,;$$

$$\vdots$$

$$C^{(2n+1)}=\sum_{\substack{p_1+\cdots+p_{2n+1}=0\\ q_1+\cdots+q_{2n+1}=0}}(p_1,q_1)(p_2,q_2)\times\cdots\times(p_{2n+1},q_{2n+1})\,. \tag{5.14}$$

It is understood that only the generators of $\mathrm{sl}(2n+1,\mathbb{C})$ do appear in (5.14), i.e., (0,0) is excluded.

Finally observe that also the relations (4.16) generalize in an obvious way:

$$E_{ii}=\frac{1}{2n+1}\sum_{m=1}^{2n+1}\zeta^{(1-i)m}D^{m},\qquad E_{ik}=E_{ii}A^{k-i},\quad 1\leqslant i,k\leqslant 2n+1\,. \tag{5.15}$$

VI. THE GENERAL CASE OF gl(2*n*,ℂ)

The development in this case follows the same line as in Sec. V. Differences occur in two ways[2]: the generating matrices A and D have to be modified in order to assure that their determinant is 1, and the orthogonal decomposition of $\mathrm{gl}(2n,\mathbb{C})$ into Cartan subalgebras holds only for $n=1$.

The group $\mathscr{P}_{2n}$ of $2n\times 2n$ matrices of determinant 1 is generated by

$$A=\begin{pmatrix}0&1&0&\cdots&0\\ &&1&0&0\\ \vdots&&&\ddots&\vdots\\ 0&&&&1\\ -1&0&&\cdots&0\end{pmatrix},\quad A^{2n}=-I\,, \tag{6.1}$$

and by the diagonal matrix

$$D=\mathrm{diag}\,\{\eta,\eta^3,\ldots,\eta^{4n-1}\},\quad \eta=e^{2\pi i/4n}\,,\qquad D^{2n}=-I\,. \tag{6.2}$$

Similarly as before $\mathscr{P}_{2n}$ consists of $(2n)^3$ matrices given by

$$K_{kad}=\eta^{2k}A^{a}D^{d},\quad k,a,d \text{ integers mod } 2n\,. \tag{6.3}$$

The property (5.7) of the group $\mathscr{P}_{2n}$, which lies at the origin of our interest in it, is written as

$$K_{kad}K_{k'a'd'}(K_{kad})^{-1}=\eta^{2(ad'-a'd)}K_{k'a'd'},\quad j\in\mathbb{Z}_{2n}\,. \tag{6.4}$$

It is verified directly using (6.3) and the relations

$$ADA^{-1}=\eta^2D,\quad DAD^{-1}=\eta^{-2}A\,. \tag{6.5}$$

Choosing the labeling elements A,D and using the notations

$$\mathbf{K}_{ad}=K_{0ad},\quad a,d \text{ integers mod } 2n \tag{6.6}$$

for the basis of $\mathrm{gl}(2n,\mathbb{C})$, we have

$$A=\mathbf{K}_{10}=(0,-1),\quad D=\mathbf{K}_{01}=(1,0)\,,\qquad A^{a}=(\mathbf{K}_{10})^{a}=(0,-a),\quad D^{d}=(K_{01})^{d}=(d,0)\,. \tag{6.7}$$

The subgroup of $\mathrm{SL}(2n,\mathbb{C})$ generated by A,D has as its commutator subgroup again the whole center of $\mathrm{SL}(2n,\mathbb{C})$.

A generator $\mathbf{K}_{ad}$ is labeled by the eigenvalues of the transformations

$$A\,\mathbf{K}_{ad}A^{-1}=\eta^{2d}\mathbf{K}_{ad},\quad D\,\mathbf{K}_{ad}D^{-1}=\eta^{-2a}\mathbf{K}_{ad}\,. \tag{6.8}$$

For simplicity of notation we write $\mathbf{K}_{ad}=(d,-a)$ [cf. (5.12)] rather than $\mathbf{K}_{ad}=(\eta^{2d},\eta^{-2a})$.

Then the commutation relations of our basis of $\mathrm{gl}(2n,\mathbb{C})$ are given by (5.13a) where $\zeta=e^{2\pi i/2n}$. In the case of (5.13b) one should remember that now, because of the identity $X^{m}=-X^{m+2n}$ for $X=A$ and D, we have

$$[(p,q),(p',q')]=\epsilon(\eta^{2p'q}-\eta^{2pq'})(p+p',q+q')\,. \tag{6.9}$$

Here $\epsilon=-1$ if either $0\leqslant q+q'<2n\leqslant q+q'<4n$ or $0\leqslant q+q'<2n\leqslant q+q'<4n$, and $\epsilon=1$ otherwise. The $2n-1$ Casimir operators of $\mathrm{sl}(2n,\mathbb{C})$ have the structure given by (5.12). Also the relations (5.15) hold practically without change taking into account that A is of order $4n$ in the present case,

$$E_{ii}=\frac{1}{2n}\sum_{m=1}^{2n}\zeta^{(1-2i)m}D^{m},\qquad E_{ik}=E_{ii}A^{k-i},\quad 1\leqslant i,k\leqslant 2n. \tag{6.10}$$

Finally let us briefly consider the generalized Pauli matrices of degree 4. The subgroup $\mathscr{P}_4$ of GL(4,C) is of order 4^3. It is faithfully represented by the following 16 matrices each multiplied by ± 1 and $\pm i$:

$$(1,0)=D=\eta\begin{pmatrix}1&&&\\&i&&\\&&-1&\\&&&-i\end{pmatrix},\quad (2,0)=D^2=\begin{pmatrix}i&&&\\&-i&&\\&&i&\\&&&-i\end{pmatrix},\quad (3,0)=D^3=\eta\begin{pmatrix}i&&&\\&1&&\\&&-i&\\&&&-1\end{pmatrix},$$

$$(0,3)=A=\begin{pmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\-1&0&0&0\end{pmatrix},\quad (0,2)=A^2=\begin{pmatrix}0&0&1&0\\0&0&0&1\\-1&0&0&0\\0&-1&0&0\end{pmatrix},\quad (0,1)=A^3=\begin{pmatrix}0&0&0&1\\-1&0&0&0\\0&-1&0&0\\0&0&-1&0\end{pmatrix},$$

$$(1,3)=AD=\eta\begin{pmatrix}0&i&0&0\\0&0&-1&0\\0&0&0&-i\\-1&0&0&0\end{pmatrix},\quad (2,3)=AD^2=\begin{pmatrix}0&-i&0&0\\0&0&i&0\\0&0&0&-i\\-i&0&0&0\end{pmatrix},$$

$$(3,3)=AD^3=\eta\begin{pmatrix}0&1&0&0\\0&0&-i&0\\0&0&0&-1\\-i&0&0&0\end{pmatrix},$$

$$(1,2)=A^2D=\eta\begin{pmatrix}0&0&-1&0\\0&0&0&-i\\-1&0&0&0\\0&-i&0&0\end{pmatrix},\quad (2,2)=A^2D^2=\begin{pmatrix}0&0&i&0\\0&0&0&-i\\-i&0&0&0\\0&i&0&0\end{pmatrix},$$

$$(3,2)=A^2D^3=\eta\begin{pmatrix}0&0&-i&0\\0&0&0&-1\\-i&0&0&0\\0&-1&0&0\end{pmatrix},$$

$$(1,1)=A^3D=\eta\begin{pmatrix}0&0&0&i\\-1&0&0&0\\0&-i&0&0\\0&0&1&0\end{pmatrix},\quad (2,1)=A^3D^2=\begin{pmatrix}0&0&0&-i\\-i&0&0&0\\0&i&0&0\\0&0&-i&0\end{pmatrix},$$

$$(3,1)=A^3D^3=\eta\begin{pmatrix}0&0&0&-1\\-i&0&0&0\\0&-1&0&0\\0&0&i&0\end{pmatrix},$$

$$(0,0)=I=\begin{pmatrix}1&&&\\&1&&\\&&1&\\&&&1\end{pmatrix}.$$

Here $\eta=\exp(2\pi i/8)$. The 16 matrices above are linearly independent and all but the identity are traceless. Equipped with the commutation relations they generate the Lie algebra gl(4,C). One can also verify that among them one does not find the Dirac matrices given relatively to an uncommon basis. Similarly they do not belong to the symplectic or orthogonal subgroups of GL(4,C).

ACKNOWLEDGMENTS

This work was supported in part by the National Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

The authors are grateful for helpful remarks and comments of Dr. A. J. Coleman, Dr. C. Cummins, Dr. I. Kaplansky, Dr. F. W. Lemire, Dr. R. V. Moody, and Dr. R. T. Sharp. The hospitality of the Aspen Center for Physics where part of the work was done is also appreciated.

[1]A. I. Kostrikin, I. A. Kostrikin, and V. A. Ufnarovskii, "Orthogonal decompositions of simple Lie algebras (type $A(n)$)," Proc. Steklov Inst. Math. **158**, 105 (1981); "Orthogonal decompositions of simple Lie algebras," Sov. Math. Dokl. **24**, 292 (1981); "Multiplicative decompositions of simple Lie algebras," *ibid.* **25**, 2327 (1982).
[2]J. Patera and H. Zassenhaus, "On Lie gradings I," preprint CRM-1459, Université de Montréal, April, 1987.
[3]M. Gell-Mann, "Symmetries of baryons and mesons," Phys. Rev. **125**, 1067 (1962).
[4]G. V. Gehlen, V. Rittenberg, and H. Ruegg, "Conformal invariance and finite-dimensional quantum chains," J. Phys. A: Math. Gen. **19**, 107 (1985); D. S. Freed and C. Vafa, "Global anomalies on orbifolds," Commun. Math. Phys. **110**, 1067 (1987).
[5]R. V. Moody, J. Patera, and R. T. Sharp, "Character generators for elements of finite order in simple Lie groups A_1, A_2, A_3, B_2, and G_2," J. Math. Phys. **24**, 2387 (1983).
[6]R. V. Moody and J. Patera, "Characters of elements of finite order in simple Lie groups," SIAM J. Algebraic Discrete Meth. **5**, 359 (1984).

Some new stationary axisymmetric asymptotically flat space-times obtained from Painlevé transcendents

Sotirios Persides
University of Chicago, Chicago, Illinois 60637 and Department of Astronomy, University of Thessaloniki, Thessaloniki, Greece
Basilis C. Xanthopoulos
University of Chicago, Chicago, Illinois 60637 and Department of Physics, University of Crete and Research Center of Crete, Iraklion, Greece

(Received 18 November 1986; accepted for publication 30 July 1987)

For the stationary axisymmetric Einstein vacuum equations in cylindrical coordinates we find that both the Ernst equation and the two real equations which alternatively describe the stationary axisymmetric problem separate, leading to Painlevé transcendents. The boundary and asymptotic behaviors of the resulting space-times are investigated in both cases. Two families of solutions are determined which, away from the symmetry axis, become asymptotically flat. The analysis provides an example to the conjecture that the Painlevé property implies integrability.

I. INTRODUCTION

The Einstein vacuum equations with one timelike and one spacelike commuting Killing fields (the stationary axisymmetric problem) reduce either to the complex Ernst equation (for the squared norm Ψ and the twist potential Φ of one of the Killing fields) or to two real Ernst-type equations (for suitable combinations of the metric coefficients χ, ω). In the present paper we investigate the separable solutions that these two sets of equations admit in cylindrical coordinates. Although the equations are nonlinear we find that the (axial) z dependence of the solutions can be exponential or sinusoidal and that the radial amplitudes satisfy (ordinary) differential equations of the Painlevé type V or III. We investigate the boundary (near the axis $\rho = 0$) and the asymptotic (for $\rho \to \infty$) behavior of the resulting space-times and we determine two families for which the metric becomes, away from the axis, asymptotically flat.

That the complex Ernst equation separates in cylindrical coordinates and the separation leads to Painlevé equations is known even more generally than the vacuum[1–4] case, for the Einstein–Maxwell equations,[5] and the Einstein equations coupled with any number of U(1) gauge fields.[6,7] It seems to have escaped notice, however, that the system of two real, Ernst-type equations separate as well. We should clarify at this point that the real Ernst-type equations arise only in the studies of the vacuum Einstein equations; they do not arise in the Einstein–Maxwell theory.

All previous investigations have essentially confined themselves to the study of the Ernst equation. None of them has investigated the boundary and asymptotic behaviors of the resulting solutions, nor the integration of the remaining two Einstein equations, leading to the determination of the conformal two-dimensional geometry "orthogonal" to the two Killing fields. These two problems are addressed systematically in the present investigation. In fact the reduction of the separated equations to Painlevé ones is such (the required transformations are not analytic) that the asymptotic behaviors can be obtained with difficulty from the behaviors of the Painlevé transcendents, if the latter were known. So, in the Appendix, we determine the behaviors of the solutions of the separated system of equations before their reduction to Painlevé equations.

There is a long standing conjecture[8,9] that an ordinary differential equation is integrable when it possesses the Painlevé property, meaning that all movable singularities are simple poles. A partial differential equation is integrable[10] when the ordinary differential equations obtained by an exact reduction of the partial equations possess the Painlevé property. The stationary axisymmetric problem is completely integrable: it possesses an infinite number of conserved currents[11,12] and it has been integrated by the inverse scattering method.[13,14] That the (complex) Ernst equation reduces to one of the Painlevé equations (which are actually characterized by the Painlevé property) is one of the few standard examples of the conjecture. The conclusion of the present paper that the real equations of the stationary axisymmetric problem also possess the Painlevé property provides additional new evidence in favor of the conjecture.

II. THE FORMALISM

For stationary axisymmetric space-times[15] in the Papapetrou[16] gauge the metric is of the form

$$(ds)^2 = \rho[\chi(dt)^2 - (1/\chi)(d\varphi - \omega dt)^2] - e^{2\mu}[(d\rho)^2 + (dz)^2], \tag{2.1}$$

where $(\partial/\partial t)$ and $(\partial/\partial\varphi)$ are the two Killing fields. Setting

$$\chi + \omega = (1+F)/(1-F), \quad \chi - \omega = (1+G)/(1-G), \tag{2.2}$$

the vacuum Einstein equations become

$$(1-FG)D^2F = -2G(DF)^2,$$
$$(1-FG)D^2G = -2F(DG)^2, \tag{2.3}$$

$$\mu_{,z} = \frac{\rho}{2\chi^2}(\chi_{,\rho}\chi_{,z} - \omega_{,\rho}\omega_{,z}) = \frac{\rho(F_{,\rho}G_{,z} + F_{,z}G_{,\rho})}{(1-FG)^2}, \tag{2.4a}$$

 0022-2488/88/030674-07$02.50

$$\mu_{,\rho} = -\frac{1}{4\rho} + \frac{\rho}{4\chi^2}(\chi_{,\rho}^2 - \chi_{,z}^2 - \omega_{,\rho}^2 + \omega_{,z}^2)$$
$$= -\frac{1}{4\rho} + \frac{\rho(F_{,\rho}G_{,\rho} - F_{,z}G_{,z})}{(1-FG)^2}, \tag{2.4b}$$

where D is the gradient operator in three-dimensional flat space but acting on scalar fields with azimuthal symmetry; thus, in particular,

$$D^2H = H_{,\rho\rho} + (1/\rho)H_{,\rho} + H_{,zz},$$
$$(DH)(D\widetilde{H}) = H_{,\rho}\widetilde{H}_{,\rho} + H_{,z}\widetilde{H}_{,z}, \tag{2.5}$$

for any two scalar fields H and $\widetilde{H}$.

Alternatively, if instead of Eqs. (2.2) we set

$$\Psi = \rho/\chi, \quad \Phi_{,\rho} = (\rho/\chi^2)\omega_{,z}, \quad \Phi_{,z} = -(\rho/\chi^2)\omega_{,\rho}, \tag{2.6}$$

and introduce the complex Ernst potentials $\mathbf{Z}$ and $\mathbf{E}$ by

$$\Psi + i\Phi = \mathbf{Z} = (1+\mathbf{E})/(1-\mathbf{E}) \tag{2.7}$$

the vacuum Einstein equations (2.3) and (2.4) become, respectively,

$$(\mathbf{E}\mathbf{E}^* - 1)D^2\mathbf{E} = 2\mathbf{E}^*(D\mathbf{E})^2, \tag{2.8}$$

$$(\mu + \tfrac{1}{2}\ln\Psi)_{,z} = (\rho/2\Psi^2)(\Psi_{,\rho}\Psi_{,z} + \Phi_{,\rho}\Phi_{,z})$$
$$= \frac{\rho(\mathbf{E}_{,\rho}\mathbf{E}_{,z}^* + \mathbf{E}_{,\rho}^*\mathbf{E}_{,z})}{(\mathbf{E}\mathbf{E}^* - 1)^2}, \tag{2.9a}$$

$$(\mu + \tfrac{1}{2}\ln\Psi)_{,\rho} = (\rho/4\Psi^2)(\Psi_{,\rho}^2 - \Psi_{,z}^2 + \Phi_{,\rho}^2 - \Phi_{,z}^2)$$
$$= \frac{\rho(\mathbf{E}_{,\rho}\mathbf{E}_{,\rho}^* - \mathbf{E}_{,z}\mathbf{E}_{,z}^*)}{(\mathbf{E}\mathbf{E}^* - 1)^2}. \tag{2.9b}$$

Equations (2.3) and (2.8) are the two real and the one complex Ernst equations, respectively, associated with the metric functions (χ,ω) and the norm and the twist potential (Ψ,Φ) of the azimuthal Killing field $(\partial/\partial\varphi)$.

Finally we mention that setting

$$\chi = \tilde{\chi}/(\tilde{\chi}^2 - \tilde{\omega}^2), \quad \omega = \tilde{\omega}/(\tilde{\chi}^2 - \tilde{\omega}^2), \tag{2.10}$$

and repeating all previous steps with the variables $(\tilde{\chi},\tilde{\omega})$ one obtains the Ernst potential and the Ernst equation associated with the Killing field $(\partial/\partial t)$.

More generally, setting

$$\tilde{\chi} = \frac{\chi}{\pi_1},$$
$$\tilde{\omega} = \frac{(\theta + \omega - \theta\omega)(1 - \theta - \theta\omega) + \theta(1-\theta)\chi^2}{(1 - 2\theta + 2\theta^2)\pi_1}, \tag{2.11}$$
$$\pi_1 = (1 - \theta - \theta\omega)^2 - \theta^2\chi^2, \quad 0 \leqslant \theta \leqslant 1,$$

with inverse

$$\chi = (1/\pi_2)(1 - 2\theta + 2\theta^2)^2\tilde{\chi},$$
$$\omega = -(1/\pi_2)\{[1 - \theta + \theta(1 - 2\theta + 2\theta^2)\tilde{\omega}]$$
$$\times[\theta - (1-\theta)(1 - 2\theta + 2\theta^2)\tilde{\omega}]$$
$$+ \theta(1-\theta)(1 - 2\theta + 2\theta^2)^2\tilde{\chi}^2\}, \tag{2.12}$$
$$\pi_2 = [1 - \theta + \theta(1 - 2\theta + 2\theta^2)\tilde{\omega}]^2$$
$$- \theta^2(1 - 2\theta + 2\theta^2)^2\tilde{\chi}^2,$$

and repeating the previous steps with the variables $(\tilde{\chi},\tilde{\omega})$ one would obtain the Ernst potential and the Ernst equation associated with the Killing field $(\partial/\partial\tilde{\varphi}) = (1-\theta)(\partial/\partial\varphi) + \theta(\partial/\partial t)$, i.e., an arbitrary linear combination of $(\partial/\partial\varphi)$ and $(\partial/\partial t)$. It should be noted, however, that the transformation (2.11) for $\theta = 1$ should be followed by the "trivial" transformation $(\chi,\omega) \to (-\chi, +\omega)$, which also preserves Eqs. (2.2) and (2.3), in order to reduce to Eqs. (2.10).

Since the boundary (near the axis) and the asymptotic (at infinity) behaviors of the Killing fields should be different for the azimuthal and the time-translational Killing fields, any solution of the Ernst equation should be investigated as representing either the untilded or the tilded quantities. Note, however, that μ remains invariant in passing from the untilded to the tilded variables.

For a solution to be physically interesting it should describe a reasonable space-time at least in some region, namely, either near the axis $\rho \to 0+$ or at infinity $\rho \to \infty$. Such solutions should have the following behavior.

(i) Near the axis

$$\chi \simeq c_1\rho^{-1}, \quad \omega \simeq c_2\rho^{-1}, \quad \mu \simeq \text{const}, \tag{2.13}$$

where $c_1 > 0$ and $c_2 \geqslant 0$ are constants.

(ii) Asymptotically (away from the axis)

$$\chi \simeq c_3\rho^{-1}, \quad \omega = o(\rho^{-2}), \quad \mu \simeq \text{const}, \tag{2.14}$$

where $c_3 > 0$ is a constant.

These conditions should be checked for both sets χ, ω, μ as well as $\tilde{\chi}, \tilde{\omega}, \tilde{\mu}$.

III. SEPARABLE SOLUTIONS OF THE REAL EQUATIONS

It does not seem to have been noticed that the two real equations (2.3) admit separable solutions in cylindrical coordinates (ρ, z). Thus setting

$$F = e^{\alpha z}f(\rho), \quad G = e^{-\alpha z}g(\rho), \quad \alpha = \text{const}, \tag{3.1}$$

and changing to dimensionless coordinates $x = \alpha\rho$, $\zeta = \alpha z$, Eqs. (2.4) reduce to

$$(fg - 1)(f'' + (1/x)f' + f) = 2g(f'^2 + f^2), \tag{3.2a}$$
$$(fg - 1)(g'' + (1/x)g' + g) = 2f(g'^2 + g^2), \tag{3.2b}$$

where the prime denotes differentiation with respect to x.

Equations (3.2) admit the integral

$$x(fg' - f'g)/(fg - 1)^2 = k = \text{const}. \tag{3.3}$$

Moreover, with the substitutions (3.1), Eqs. (2.4) reduce considerably; they read

$$\mu_{,\zeta} = \frac{x(fg' - f'g)}{(fg-1)^2} = k,$$
$$\mu_{,x} = -\frac{1}{4x} + \frac{x(f'g' + fg)}{(fg-1)^2}. \tag{3.4}$$

From Eqs. (2.2), (3.1), and (3.4) we find that the resulting solution is determined from

$$\chi = \frac{1 - fg}{(1 - e^{\zeta}f)(1 - e^{-\zeta}g)},$$
$$\omega = \frac{e^{\zeta}f - e^{-\zeta}g}{(1 - e^{\zeta}f)(1 - e^{-\zeta}g)}, \quad \mu = k\zeta + \mu_1(x), \tag{3.5}$$

where

$$\frac{d\mu_1(x)}{dx} = -\frac{1}{4x} + \frac{x(fg+f'g')}{(fg-1)^2}. \quad (3.6)$$

The expressions (3.5) and (3.6), with f and g satisfying Eqs. (3.2), determine a *four-parameter family of solutions*. Because of the presence of the term $-\frac{1}{4}\ln x$ in the expression for μ, arising from the integration of Eq. (3.6), the resulting space-time is not regular near the axis.

Asymptotically $(x \to \infty)$ we find, after an elaborate analysis which is demonstrated in the Appendix, that the solutions of Eqs. (3.2) behave like

$$f = (a_1 e^{i\varphi} + a_1^* e^{-i\varphi})x^{-1/2} + O(x^{-3/2}), \quad (3.7a)$$

$$g = k_1(a_1 e^{i\varphi} + a_1^* e^{-i\varphi})x^{-1/2} + O(x^{-3/2}), \quad (3.7b)$$

where

$$\varphi = x + (4k_1 a_1 a_1^*)\ln x, \quad (3.8)$$

and a_1 and k_1 are complex and real constants, respectively. From the expressions (3.5)–(3.8) we find that

$$\chi = 1 + O(x^{-1/2}), \quad \omega = O(x^{-1/2}), \quad \mu = k\zeta + 4k_1 a_1 a_1^* x + O(\ln x). \quad (3.9)$$

The resulting space-time is not asymptotically flat.

In the presence of the integral (3.3) one would expect that the system of equations (3.2) would be essentially one of the third order. We shall now show, instead, that it can be reduced to a second-order Painlevé equation. For the reduction we set

$$f = Pe^{Q}, \quad g = \epsilon P e^{-Q}, \quad (3.10)$$

where $\epsilon = +1$ or -1 depending on whether $fg \geqslant 0$ or $fg < 0$, respectively. Note that this transformation is locally one-to-one and invertible, but it is not analytic. The integral (3.3) then becomes

$$xP^2Q'/(\epsilon - P^2)^2 = -\epsilon k/2. \quad (3.11)$$

Using Eq. (3.11) to eliminate Q we find that P satisfies the second-order equation

$$(P^2-\epsilon)\left(P'' + \frac{1}{x}P'\right) - 2PP'^2 - P(P^2+\epsilon)\left[1 + \frac{k^2(P^2-\epsilon)^4}{4x^2P^4}\right] = 0. \quad (3.12)$$

Then setting

$$P = \eta(\epsilon W)^{-1/2}, \quad (3.13)$$

where $\eta = +1$ or -1 depending on whether $P \geqslant 0$ or $P < 0$, respectively, we obtain

$$W'' - \left(\frac{1}{2W} + \frac{1}{W-1}\right)W'^2 + \frac{1}{x}W' - \beta\frac{(W+1)(W-1)^3}{x^2W} - \delta\frac{W(W+1)}{W-1} = 0, \quad (3.14)$$

with $\beta = k^2/2$ and $\delta = 2$. This is a particular case of the Painlevé equation[17] of type V. Note that Eq. (3.14) is independent of ϵ and η, i.e., independent of the signs of f, g, and P.

Much simpler is the case when $k = 0$ in Eq. (3.3) (instead of a constant). Then $f/g = \text{const}$ and without loss of generality we can choose $f = g = y$ (say), since the constant of the ratio f/g can be absorbed in a shift of the origin of the z coordinate. Equations (3.2) then reduce to the single equation

$$(y^2-1)(y'' + (1/x)y' + y) = 2y(y'^2 + y^2). \quad (3.15)$$

Now y is smooth near the axis, exhibiting the behavior

$$y = a_0 + [a_0(a_0^2+1)/4(a_0^2-1)]x^2 + O(x^4), \quad (3.16)$$

where a_0 is a real constant different from ± 1. Hence we get

$$\begin{aligned}\chi &= \frac{1-y^2}{(1-e^{\zeta}y)(1-e^{-\zeta}y)} = \frac{1-a_0^2}{(1-a_0e^{\zeta})(1-a_0e^{-\zeta})} + O(x^2),\\ \omega &= \frac{2y\sinh\zeta}{(1-e^{\zeta}y)(1-e^{-\zeta}y)} = \frac{2a_0\sinh\zeta}{(1-a_0e^{\zeta})(1-a_0e^{-\zeta})} + O(x^2),\\ \mu &= k\zeta - \tfrac{1}{4}\ln x + O(x^2),\end{aligned} \quad (3.17)$$

and the space-time is not regular on the axis $x = 0$.

The general asymptotic expansion of y is given in the Appendix. As a special case we find, with $a_1 = (1-i)/2$, that

$$y = (\cos\varphi + \sin\varphi)x^{-1/2} + O(x^{-3/2}), \quad \varphi = x + 2\ln x, \quad (3.18)$$

from which we can find the behavior of the metric coefficients

$$\begin{aligned}\chi &= 1 + (2\cosh\zeta/\sqrt{x})(\cos\varphi + \sin\varphi) + O(x^{-1}),\\ \omega &= (2\sinh\zeta/\sqrt{x})(\cos\varphi + \sin\varphi) + O(x^{-1}),\\ \mu &= 2x + O(\ln x).\end{aligned} \quad (3.19)$$

The space-time is not asymptotically flat.

Finally we note that by the substitution $y = (w-1)/(w+1)$ Eq. (3.21) reduces to

$$w'' - \frac{w'^2}{w} + \frac{1}{x}w' + \frac{\delta}{4}\left(w^3 - \frac{1}{w}\right) = 0, \quad (3.20)$$

with $\delta = 1$, which is a special case of the Painlevé equation of type III.

IV. SEPARABLE SOLUTIONS OF THE COMPLEX EQUATION

The complex Ernst equation (2.8) also admits separable solutions.[2,3] Substituting

$$\mathbf{E} = e^{i\zeta}\mathbf{H}(x), \quad \zeta = \alpha z, \quad x = \alpha\rho, \quad \alpha = \text{real const} \quad (4.1)$$

into Eq. (2.9) we obtain

$$(\mathbf{HH}^* - 1)(\mathbf{H}'' + (1/x)\mathbf{H}' - \mathbf{H}) = 2\mathbf{H}^*(\mathbf{H}'^2 - \mathbf{H}^2). \quad (4.2)$$

This equation admits the integral (as a similar equation in Ref. 4),

$$\frac{x(\mathbf{HH}^{*\prime} - \mathbf{H}'\mathbf{H}^*)}{(\mathbf{HH}^* - 1)^2} = i\lambda, \quad \lambda = \text{real const} \quad (4.3)$$

as it can be shown by direct differentiation.

As in Sec. III, Eq. (4.2), essentially a system of two ordinary differential equations, can be reduced to a single *second-order equation*. Setting

$$\mathbf{H} = \mathbf{H}_0(x)e^{i\theta(x)}, \quad \mathbf{H}_0 = |\mathbf{H}|, \quad \theta(x) = \text{real}, \tag{4.4}$$

the integral (4.3) reads

$$x\mathbf{H}_0^2\theta'(x)/(\mathbf{H}_0^2 - 1)^2 = -\lambda/2, \tag{4.5}$$

while Eq. (4.2) reduces to

$$(\mathbf{H}_0^2 - 1)(\mathbf{H}_0'' + (1/x)\mathbf{H}_0') - 2\mathbf{H}_0\mathbf{H}_0'^2 + \mathbf{H}_0(\mathbf{H}_0^2 + 1) \times [1 + \lambda^2(\mathbf{H}_0^2 - 1)^4/4x^2\mathbf{H}_0^4] = 0. \tag{4.6}$$

This is similar to Eq. (3.12) and with the substitution $\mathbf{H}_0 = W^{-1/2}$ reduces to Eq. (3.14) and $\beta = -\lambda^2/2$ and $\delta = -2$, i.e., a Painlevé equation of type V. The reduction of the (complex) Ernst equation to the Painlevé equation of type V was first noticed by Marek,[1] and later, independently, by Léauté and Marcilhacy.[2]

Using Eqs. (2.6) and (2.7) it is straightforward to express the metric coefficients (χ,ω) in terms of any solution of Eqs. (4.2). We find

$$\chi = \rho/\Psi = [\rho/(1 - \mathbf{H}\mathbf{H}^*)] \times (\mathbf{H}\mathbf{H}^* - \mathbf{H}e^{i\zeta} - \mathbf{H}^*e^{-i\zeta} + 1), \tag{4.7}$$

and

$$\omega_{,\rho} = -[x/(\mathbf{H}\mathbf{H}^* - 1)^2][\mathbf{H}(\mathbf{H}\mathbf{H}^* + 1)e^{i\zeta} + \mathbf{H}^*(\mathbf{H}\mathbf{H}^* + 1)e^{-i\zeta} - 4\mathbf{H}\mathbf{H}^*], \tag{4.8a}$$

$$\omega_{,z} = -[ix/(\mathbf{H}\mathbf{H}^* - 1)^2][(\mathbf{H}' - \mathbf{H}^2\mathbf{H}^{*\prime})e^{i\zeta} - (\mathbf{H}^{*\prime} - \mathbf{H}^{*2}\mathbf{H}')e^{-i\zeta}] + 2\lambda. \tag{4.8b}$$

The remarkable thing is that Eqs. (4.8) can be integrated for ω, using only the knowledge that $\mathbf{H}$ satisfies Eq. (4.2). We find that

$$\omega = -[x/\alpha(\mathbf{H}\mathbf{H}^* - 1)^2][(\mathbf{H}' - \mathbf{H}^2\mathbf{H}^{*\prime})e^{i\zeta} + (\mathbf{H}^{*\prime} - \mathbf{H}^{*2}\mathbf{H}')e^{-i\zeta}] + 2\lambda z + \omega_1(x), \tag{4.9}$$

where

$$\frac{d\omega_1(x)}{dx} = \frac{4x\mathbf{H}\mathbf{H}^*}{\alpha(\mathbf{H}\mathbf{H}^* - 1)^2}. \tag{4.10}$$

The last metric coefficient μ is obtained from Eqs. (2.9). We find

$$(\mu + \tfrac{1}{2}\ln\Psi)_{,\zeta} = -\lambda, \quad (\mu + \tfrac{1}{2}\ln\Psi)_{,x} = x(\mathbf{H}'\mathbf{H}^{*\prime} - \mathbf{H}\mathbf{H}^*)/(\mathbf{H}\mathbf{H}^* - 1)^2, \tag{4.11}$$

from which we obtain

$$\mu = -\alpha\lambda z - \tfrac{1}{2}\ln(\rho/\chi) + \mu_1(\rho), \tag{4.12}$$

where

$$\frac{d\mu_1}{dx} = \frac{x(\mathbf{H}'\mathbf{H}^{*\prime} - \mathbf{H}\mathbf{H}^*)}{(\mathbf{H}\mathbf{H}^* - 1)^2}. \tag{4.13}$$

A. Solutions smooth on the axis? No!

When $\mathbf{H}$ and $\mathbf{H}'$ are finite at $x = 0$, from Eq. (4.3) we obtain $\lambda = 0$ and therefore we should have $\mathbf{H}\mathbf{H}^{*\prime} = \mathbf{H}'\mathbf{H}^*$, for every x. It is then easy to show that $\mathbf{H}$ equals to a real function (of x) times a complex constant which can be absorbed into the factor $e^{i\alpha z}$ by a linear transformation of z. Thus when $\lambda = 0$ we can assume, without any loss of generality, that $\mathbf{H}$ is real, satisfying the equation

$$(\mathbf{H}^2 - 1)(\mathbf{H}'' + (1/x)\mathbf{H}' - \mathbf{H}) = 2\mathbf{H}(\mathbf{H}'^2 - \mathbf{H}^2). \tag{4.14}$$

Note that Eqs. (3.15) and (4.14) differ only in the signs of the last terms in the two sides. By the transformation $\mathbf{H} = (w - 1)/(w + 1)$ Eq. (4.14) transforms to Eq. (3.20) with $\delta = -1$, which is a Painlevé equation of type III.

In terms of any real $\mathbf{H}$ we obtain from Eqs. (4.7), (4.9), (4.10), (4.12), and (4.13) that

$$\chi = \rho(\mathbf{H}^2 + 1 - 2\mathbf{H}\cos\zeta)/(1 - \mathbf{H}^2), \tag{4.15}$$

$$\omega = \frac{2x\mathbf{H}'\cos\zeta}{\alpha(\mathbf{H}^2 - 1)} + \omega_1, \quad \frac{d\omega_1}{dx} = \frac{4x\mathbf{H}^2}{\alpha(\mathbf{H}^2 - 1)^2}, \tag{4.16}$$

and

$$\mu = \frac{1}{2}\ln\left(\frac{\mathbf{H}^2 + 1 - 2\mathbf{H}\cos\zeta}{1 - \mathbf{H}^2}\right) + \mu_1(x), \quad \frac{d\mu_1}{dx} = \frac{x(\mathbf{H}'^2 - \mathbf{H}^2)}{(\mathbf{H}^2 - 1)^2}. \tag{4.17}$$

Since the condition $\lambda = 0$ is necessary but not sufficient for a solution to be well behaved near the axis, we shall investigate the behavior of the solution (4.15)–(4.17) as well as of the solution $(\tilde{\chi},\tilde{\omega},\mu)$ obtained from Eqs. (2.10) or (2.11).

Near the axis Eq. (4.14) admits the smooth solution

$$\mathbf{H} = b_0 + [b_0(b_0^2 + 1)/4(1 - b_0^2)]x^2 + O(x^4), \tag{4.18}$$

where $b_0 \neq \pm 1$ is an arbitrary real constant. Using this expansion we obtain

$$\chi = [(b_0^2 + 1 - 2b_0\cos\zeta)/\alpha(1 - b_0^2)]x + O(x^3), \tag{4.19}$$

$$\omega = -[b_0x^2/\alpha(b_0 + 1)^2] + O(x^4), \tag{4.20}$$

$$\mu = \tfrac{1}{2}\ln[(b_0^2 + 1 - 2b_0\cos\zeta)/(1 - b_0^2)] + O(x^2). \tag{4.21}$$

From Eqs. (2.10), (4.19), and (4.20) we find that

$$\tilde{\chi} = -\alpha(1 - b_0^2)/(b_0^2 + 1 - 2b_0\cos\zeta)x + O(x),$$
$$\tilde{\omega} = \alpha b_0(1 - b_0)^2/(b_0^2 + 1 - 2b_0\cos\zeta)^2 + O(x^2), \tag{4.22}$$

when we interchange the two Killing fields. Similarly from Eqs. (2.11) we find that

$$\tilde{\chi} = [(b_0^2 + 1 - 2b_0\cos\zeta)/\alpha(1 - b_0^2)(1 - \theta)^2]x + O(x^3), \tag{4.23}$$
$$\tilde{\omega} = \theta/(\theta - 1) + O(x^2),$$

when we consider a mixture of the two Killing fields, for $0 < \theta < 1$. None of the resulting space-times behaves well near the axis.

B. Solutions leading to asymptotically flat space-times

For large x we shall consider space-times arising from solutions of Eq. (4.2) for which $\mathbf{H}$ and $\mathbf{H}'$ are finite as $x \to +\infty$. From the analysis in the Appendix we find that these solutions of Eq. (4.2) behave like $e^{-x}x^{-1/2}$ as $x \to +\infty$. Applying the integral (4.3) asymptotically we find that $\lambda = 0$. Thus we shall have $\mathbf{H}\mathbf{H}^{*\prime} = \mathbf{H}'\mathbf{H}^*$ everywhere and, as in Sec. IV A, without any loss of generality we can assume that $\mathbf{H}$ *is real*.

We shall consider solutions of Eq. (4.14) but referring to the Ernst equation for the tilded potentials

$$\tilde{\Psi} + i\tilde{\Phi} = (1+\tilde{\mathbf{E}})/(1-\tilde{\mathbf{E}}), \tag{4.24}$$

corresponding to the interchange of the Killing fields $(\partial/\partial t)$ and $(\partial/\partial\varphi)$.

The expressions (4.15)–(4.17) are now applicable if we replace (χ,ω) by $(\tilde{\chi},\tilde{\omega})$. Using the asymptotic behaviors (A11) and (A12) for a real parameter A we find

$$\tilde{\chi} = \rho - A\alpha^{-1}\sqrt{2\pi}x^{1/2}e^{-x}\cos\zeta + O(x^{-1/2}e^{-x}), \tag{4.25}$$

$$\tilde{\omega} = A\alpha^{-1}\sqrt{2\pi}x^{1/2}e^{-x}\cos\zeta + O(x^{-1/2}e^{-x}), \tag{4.26}$$

$$\mu = -A\sqrt{\pi/2x}e^{-x}\cos\zeta + O(x^{-3/2}e^{-x}). \tag{4.27}$$

Then from Eqs. (2.10) we obtain

$$\chi = 1/\rho + \alpha A\sqrt{2\pi}x^{-3/2}e^{-x}\cos\zeta + O(x^{-5/2}e^{-x}), \tag{4.28}$$

$$\omega = \alpha A\sqrt{2\pi}\,x^{-3/2}e^{-x}\cos\zeta + O(x^{-5/2}e^{-x}). \tag{4.29}$$

For large ρ the corresponding metric (2.1) tends to

$$(ds)^2 = (dt)^2 - (d\rho)^2 - \rho^2(d\varphi)^2 - (dz)^2, \tag{4.30}$$

i.e., *to a flat metric*. It should be pointed out, however, that for large z but near the axis the metric does not tend to a flat metric (it depends sinusoidally on z).

For completeness we also investigate the possibility of using the solution of Eq. (4.14) but now referring to the Ernst equation for the potential

$$\Psi_\theta + i\Phi_\theta = (1+\mathbf{E}_\theta)/(1-\mathbf{E}_\theta), \tag{4.31}$$

corresponding to the arbitrary mixing of the two Killing fields. Now χ and ω are obtained from Eqs. (2.12), where $(\tilde{\chi},\tilde{\omega})$ are given again by Eqs. (4.25) and (4.26). As we have mentioned in Sec. II, μ will remain the same and it will behave as in Eq. (4.27). We find that

$$\pi_2 = -\theta^2(1-2\theta+2\theta^2)^2\rho^2 + (1-\theta)^2 + 2A\alpha^{-2}\theta^2(1-2\theta+2\theta^2)^2\sqrt{2\pi}x^{3/2}e^{-x}\cos\zeta + O(x^{1/2}e^{-x}), \tag{4.32a}$$

$$\chi = -(1/\theta^2\rho)[1 + ((1-\theta)^2/\theta^2(1-2\theta+2\theta^2)^2\rho^2) + A\sqrt{2\pi}x^{-1/2}e^{-x}\cos\zeta + O(x^{-3/2}e^{-x})], \tag{4.32b}$$

$$\omega = [(1-\theta)/\theta][1 + (\theta^2(1-2\theta+2\theta^2)\rho^2)^{-1} + O(x^{-4})]. \tag{4.32c}$$

For large ρ the metric (2.1) tends to

$$(ds)^2 = -\theta^{-2}(dt)^2 + \rho^2[\theta(d\varphi) + (\theta-1)(dt)]^2 - (d\rho)^2 - (dz)^2, \tag{4.33}$$

i.e., to a locally flat but physically unacceptable metric.

V. AN EHLERS TRANSFORMATION

The simplest expression of the Ehlers transformation asserts that whenever $\mathbf{Z} = \Psi + i\Phi$ is a solution of the Ernst equation, so does

$$\bar{\mathbf{Z}} = \frac{1}{\mathbf{Z}+ic} \quad \text{or} \quad \bar{\Psi} = \frac{\Psi}{\Psi^2 + (\Phi+c)^2}, \quad \bar{\Phi} = -\frac{\Phi+c}{\Psi^2+(\Phi+c)^2}, \tag{5.1}$$

where c is a real constant.

We apply the Ehlers transformation to the solution of Sec. IV B for the tilded potentials $\tilde{\Psi}$ and $\tilde{\Phi}$, i.e., the squared norm and the twist potential of the Killing field $(\partial/\partial t)$. Since

$$\tilde{\Psi} = (1-\mathbf{H}^2)/(1+\mathbf{H}^2 - 2\mathbf{H}\cos\zeta), \quad \tilde{\Phi} = 2H\sin\zeta/(1+H^2-2H\cos\zeta), \tag{5.2}$$

where $\mathbf{H}$ is any (real) solution of Eq. (4.14), we find that

$$\bar{\tilde{\Psi}} = \frac{(1-\mathbf{H}^2)(\mathbf{H}^2+1-2\mathbf{H}\cos\zeta)}{(1-\mathbf{H}^2)^2 + [c(\mathbf{H}^2+1-2\mathbf{H}\cos\zeta) + 2\mathbf{H}\sin\zeta]^2}, \tag{5.3a}$$

$$\bar{\tilde{\Phi}} = -\frac{[c(\mathbf{H}^2+1-2\mathbf{H}\cos\zeta)+2\mathbf{H}\sin\zeta](\mathbf{H}^2+1-2\mathbf{H}\cos\zeta)}{(1-\mathbf{H}^2)^2 + [c(\mathbf{H}^2+1-2\mathbf{H}\cos\zeta)+2\mathbf{H}\sin\zeta]^2}. \tag{5.3b}$$

Then $\bar{\tilde{\chi}}$ is readily obtained from $\bar{\tilde{\chi}} = \rho/\bar{\tilde{\Psi}}$.

Using the asymptotic expansion (A11) and (A12) we find that

$$\bar{\tilde{\chi}} = \rho(1+c^2) + A\alpha^{-1}[2c\sin\zeta + (1-c^2)\cos\zeta] \times e^{-x}\sqrt{2\pi x} + O(x^{-1/2}e^{-x}), \tag{5.4}$$

$$\bar{\tilde{\Psi}} = 1/(1+c^2) - [A/(1+c^2)^2][2c\sin\zeta + (1-c^2)\cos\zeta]e^{-x}\sqrt{2\pi/x} + O(x^{-3/2}e^{-x}), \tag{5.5}$$

$$\bar{\tilde{\Phi}} = -c/(1+c^2) - [A/(1+c^2)^2][(1-c^2)\sin\zeta - 2c\cos\zeta]e^{-x}\sqrt{2\pi/x} + O(x^{-3/2}e^{-x}), \tag{5.6}$$

and that

$$\bar{\tilde{\Psi}}_{,\rho} = -\bar{\tilde{\Phi}}_{,z} + O(x^{-3/2}e^{-x}), \quad \bar{\tilde{\Psi}}_{,z} = \bar{\tilde{\Phi}}_{,\rho} + O(x^{-3/2}e^{-x}). \tag{5.7}$$

Also

$$\bar{\tilde{\omega}} = -(A/\alpha)[2c\sin\zeta + (1-c^2)\cos\zeta] \times e^{-x}\sqrt{2\pi x} + O(x^{-1/2}e^{-x}). \tag{5.8}$$

Moreover from Eqs. (2.9) we find that

$$\bar{\mu} = \tfrac{1}{2}\ln(1+c^2) + [A/(1+c^2)] \times [2c\sin\zeta + (1-c^2)\cos\zeta] \times e^{-x}\sqrt{\pi/2x} + O(x^{-3/2}e^{-x}). \tag{5.9}$$

Therefore for $\rho \to +\infty$ the metric tends to

$$(ds^2) = (dt)^2/(1+c^2) - (1+c^2) \times [(d\rho)^2 + \rho^2(d\varphi)^2 + (dz)^2], \quad (5.10)$$

which is *flat*! As in the previous case, local flatness fails near the axis of symmetry. This should have been anticipated, since an asymptotically flat solution should be matched to an interior solution, i.e., a solution of the inhomogeneous Einstein equations that are appropriate for a region that includes the axis.

VI. CONCLUSIONS

We have investigated the separable solutions of the stationary axisymmetric problem in cylindrical coordinates. For all these solutions we have obtained explicit expressions for all of the metric functions and we have determined the behavior of these solutions near the axis and asymptotically. For one of these solutions we found that, for large distances away from the axis, the metric tends to a flat metric. A second space-time with the same property has been obtained by applying an Ehlers transformation.

The separated equations always admit a first integral; and the radial functions reduce to certain Painlevé equations of type III or V, depending on whether the integral vanishes or not. Since the Painlevé transcendents are determined uniquely from the studied functions, we have, in fact, also determined the near the axis behavior and the asymptotic expansions of these (particular) Painlevé transcendents.

The solutions (4.27)–(4.29) and (5.4), (5.8), and (5.9) go to a flat space-time exponentially as $\rho \to \infty$. Furthermore, they are both periodic in the z direction with period $z_0 = 2\pi/\alpha$ [see Eqs. (4.15)–(4.17)] while the first corrections of the asymptotic expansions change signs when ζ takes values in the intervals $(-\pi/2, \pi/2)$ and $(\pi/2, 3\pi/2)$. If we were to allow ourselves to speculate, we could probably say that such gravitational fields may be generated by a matter distribution which is periodic along the z axis and whose rotation reverses itself every $\Delta z = \pi/\alpha$. In addition we could say that the resulting curvature is very strongly localized around the axis of rotation. It should be pointed out, however, that any rigorous interpretation of the solutions would require considerations of the inhomogeneous Einstein equations around the axis. And that any interior solution obtained should be joined, with a C^2 matching, with the asymptotically flat solutions of the present paper. Clearly, this project is beyond the scope of the present investigation.

ACKNOWLEDGMENTS

We would like to thank the Relativity Group of the University of Chicago, where part of this work was done, for the hospitality shown to us. In particular, we would like to thank Professor S. Chandrasekhar for communicating to us his results on cylindrically symmetric gravitational waves from which we were motivated to undertake the present investigation.

APPENDIX: ASYMPTOTIC BEHAVIOR

We determine the asymptotic ($x \to \infty$) behavior of the solutions of the system (3.2). It is convenient to set

$$f = x^{1/2} f_1, \quad g = x^{1/2} g_1; \quad (A1)$$

then the system becomes

$$f_1'' + f_1 = -(2/x) f_1' - 2x g_1 (f_1'^2 + f_1^2) + x f_1 g_1 (f_1'' + f_1) - (1/4x^2) f_1 - (1/4x) f_1^2 g_1, \quad (A2a)$$

$$g_1'' + g_1 = -(2/x) g_1' - 2x f_1 (g_1'^2 + g_1^2) + x f_1 g_1 (g_1'' + g_1) - (1/4x^2) g_1 - (1/4x) g_1^2 f_1 . \quad (A2b)$$

We seek an asymptotic expansion of f_1 and g_1 that would satisfy Eqs. (A2) to leading order. After a lot of trials we find that

$$f_1 = (a_1 e^{i\varphi} + a_2 e^{-i\omega}) x^{-1} + O(x^{-2}),$$
$$g_1 = (b_1 e^{-i\varphi} + b_2 e^{i\omega}) x^{-1} + O(x^{-2}), \quad (A3)$$

where

$$\varphi = x + \alpha \ln x, \quad \omega = x + \beta \ln x, \quad (A4)$$

and a_1, a_2, b_1, b_2, α, and β are free parameters, satisfy Eqs. (A2) to $O(x^{-1})$. In fact, the only nontrivial step is the verification that

$$f_1'' + f_1 = O(x^{-2}), \quad g_1'' + g_1 = O(x^{-2}). \quad (A5)$$

Obviously, the expansions (A3) have too many free parameters, a freedom necessary for the expansion to be continued to higher orders.

To restrict the parameters we consider the asymptotic expansion of the solution of Eqs. (A2) to $O(x^{-2})$. Thus we assume that

$$f_1 = (a_1 e^{i\varphi} + a_2 e^{-i\omega}) x^{-1} + (A_1 e^{i\varphi} + A_2 e^{-i\omega}) x^{-2} + O(x^{-3}), \quad (A6a)$$

$$g_1 = (b_1 e^{-i\varphi} + b_2 e^{i\omega}) x^{-1} + (B_1 e^{-i\varphi} + B_2 e^{i\omega}) x^{-2} + O(x^{-3}), \quad (A6b)$$

and we demand that the system of Eqs. (A2) is satisfied to $O(x^{-2})$ as well. We find that the conditions to $O(x^{-2})$ impose restrictions on the parameters a_1, a_2, b_1, b_2, α, and β of the $O(x^{-1})$ expansion, which do not involve the parameters A_1, A_2, B_1, and B_2. Using

$$f_1'' + f_1 = -2[(\alpha + i) a_1 e^{i\varphi} + (\beta - i) a_2 e^{-i\omega}] x^{-2} + O(x^{-3}), \quad (A7a)$$

$$g_1'' + g_1 = -2[(\alpha - i) b_1 e^{-i\varphi} + (\beta + i) b_2 e^{i\omega}] x^{-2} + O(x^{-3}), \quad (A7b)$$

we obtain the conditions

$$a_1(\alpha - 4a_2 b_2) = 0, \quad a_2(\beta - 4a_1 b_1) = 0 \text{ from Eq. (A2a)}$$

and

$$b_1(\alpha - 4a_2 b_2) = 0, \quad b_2(\beta - 4a_1 b_1) = 0 \text{ from Eq. (A2b)}.$$

The further requirement that the solutions (A6) are real and nontrivial to order $O(x^{-1})$ give that

$$a_2 = a_1^*, \quad b_1 = k_1 a_1^*, \quad b_2 = b_1^*, \quad \alpha = \beta = 4k_1 a_1 a_1^*, \tag{A8}$$

where a_1 and k_1 are arbitrary complex and real constants, respectively.

For $k_1 = 1$, $f_1 = g_1$ and the method gives the asymptotic behavior of the solution of Eq. (3.15). We find that for real y the asymptotic expansion is

$$\begin{aligned} y &= (a_1 e^{i\varphi} + a_1^* e^{-i\varphi})x^{-1/2} + O(x^{-3/2}), \\ \varphi &= x + 4a_1 a_1^* \ln x. \end{aligned} \tag{A9}$$

We determine now the asymptotic behavior of $\mathbf{H}$ that satisfies Eq. (4.2) and goes to 0 at infinity. Since the nonlinear terms will go to 0 faster than the linear terms, the leading term of the expansion is determined by the linear equation

$$\mathbf{H}'' + (1/x)\mathbf{H}' - \mathbf{H} = 0. \tag{A10}$$

This is the differential equation for the modified Bessel function.[18] The solution that goes to 0 at infinity is

$$\mathbf{K}_0 = e^{-x}\sqrt{\pi/2x}\,[1 + O(x^{-1})]. \tag{A11}$$

Hence the (complex) solution of Eq. (4.2) that goes to zero at infinity is

$$\mathbf{H} = A\,\mathbf{K}_0(x) + o(e^{-3x}), \tag{A12}$$

where A is an arbitrary complex constant and the complex terms $o(e^{-3x})$ are obtained from the nonlinear terms of Eq. (4.2).

[1] J. J. Marek, Proc. Cambridge Philos. Soc. **64**, 167 (1968).
[2] B. Léauté and G. Marcilhacy, Phys. Lett. A **87**, 159 (1982).
[3] R. K. Dodd and H. C. Morris, Proc. R. Irish Acad. Sec. A **83**, 95 (1983).
[4] S. Chandrasekhar, Proc. R. Soc. London Ser. A **408**, 209 (1986).
[5] G. Marcilhacy, Phys. Lett. A **73**, 157 (1979).
[6] M. Gürses, Phys. Rev. D **30**, 486 (1984).
[7] B. Léauté and G. Marcilhacy, J. Math. Phys. **27**, 703 (1986).
[8] P. Painlevé, Acta Math. **25**, 1 (1902).
[9] B. Gambier, Acta Math. **33**, 1 (1910).
[10] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **19**, 715 (1980).
[11] R. Geroch, J. Math. Phys. **13**, 394 (1972).
[12] W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 2037 (1978).
[13] V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).
[14] D. Kramer and G. Neugebauer, J. Phys. A **14**, L333 (1981).
[15] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[16] A. Papapetrou, Ann. Phys. (Leipzig) **12**, 309 (1953).
[17] E. L. Ince, *Ordinary Differential Equations* (Longmans, Green, London, 1927).
[18] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), p. 375.

Colliding gravitational plane waves with noncollinear polarization. III

Frederick J. Ernst
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676
Alberto García D.
Departamento de Física, CINVESTAV del Instituto Politécnico Nacional, Apdo. 14-740, 07000 México, D. F., Mexico
Isidore Hauser
1615 Cottonwood Drive, #8, Louisville, Colorado 80027

(Received 1 September 1987; accepted for publication 28 October 1987)

The Hauser–Ernst homogeneous Hilbert problem (HHP) approach, formerly used in connection with the derivation of stationary axisymmetric fields, is here adapted to the derivation of colliding gravitational plane wave solutions of the vacuum Einstein equations. Proceeding from Kasner metrics, and using a double-Harrison transformation, the HHP approach yields a three-parameter generalization of a two-parameter family of colliding wave solutions found recently by Ferrari, Ibañez, and Bruni. In the present paper we provide the details concerning the derivation of this new family of solutions, and we set the stage for future applications of the HHP approach in connection with gravitational waves.

I. INTRODUCTION

A. General problem being considered

In this series of papers[1,2] we are concerned with the interaction of two gravitational plane waves moving in opposite directions. In particular, we have focused attention upon plane waves that do not interact at all until a certain moment called the "moment of collision." At the leading edge of each of the Petrov type N plane waves, we permit the existence of a Dirac delta function behavior of the curvature tensor, i.e., an impulse, and we permit the existence of a jump discontinuity of the curvature tensor, interpreted as a gravitational shock wave.

It is convenient to describe the colliding wave solution in terms of four space-time regions, separated from one another by null surfaces. In region I the metric is simply that of Minkowski space. In the adjacent regions II and III, the metric is, generally, a Petrov type N plane wave solution of the field equations, although in the particular case of the Nutku–Halil solution[3] the metric is flat in the interior of regions II and III. Finally, in another region, region IV, separated from regions II and III by null surfaces, the interaction of the plane waves takes place. As a result of the interaction of the gravitational plane waves, curvature singularities may prevent region IV from being extended indefinitely.

In fact, the solutions of the colliding plane wave problem that have been found thus far were found by working backwards. First, one finds a solution of the vacuum field equations in region IV, the region of interaction. Then one attempts to join this solution to appropriately chosen plane wave solutions in regions II and III, and to join the latter to Minkowski space in region I.

In Paper II we identified a simple condition that the solution has to satisfy in region IV in order that it be joinable to appropriately chosen plane wave solutions in regions II and III. This *colliding wave condition* plays a role roughly analogous to the *asymptotic flatness condition* usually imposed in connection with stationary axisymmetric fields.

B. Drawing upon experience with stationary axisymmetric fields

The search for an effective way to construct *all* asymptotically flat stationary axisymmetric vacuum fields began with a speculation of Geroch[4] that perhaps all such solutions could be derived from a single solution, e.g., Minkowski space, through the action of a group, the free product of two $SL(2,\mathbb{R})$ groups.

The first really useful realization of the Lie algebra of the Geroch group was formulated by Kinnersley and Chitre,[5] who displayed the action of the infinitesimal elements of the group in terms of an infinite hierarchy of potentials. Kinnersley and Chitre, as well as Hoenselaers, Kinnersley, and Xanthopoulos,[6] exploited this formalism to derive new asymptotically flat stationary axisymmetric solutions, and also to demonstrate that certain famous solutions, such as the Kerr solution, could be regenerated using these techniques.

Two of the present authors, Hauser and Ernst,[7] introduced a realization of the finite elements of the Geroch group, and ultimately showed[8] that Kinnersley–Chitre transformations could be carried out by solving an appropriate homogeneous Hilbert problem (HHP). In particular, they employed the HHP approach in order to prove[9,10] the Geroch conjecture.

In Paper II we described, within the context of colliding gravitational plane waves, how the Geroch group arises as the free product of two $SL(2,\mathbb{R})$ groups. We also drew attention to the utility of augmenting the Geroch group with a Kramer–Neugebauer involution.[11] In the present paper we shall replace the rather formal realization of the group described in Paper II by a realization better suited to the generation of new solutions from old ones. We shall formulate an HHP that is particularly adapted to the problem at hand, and we shall show how we used it in order to derive a three-parameter generalization of a two-parameter family of solutions discovered recently by Ferrari, Ibañez, and Bruni.[12] It

 0022-2488/88/030681-09$02.50

should be mentioned that these authors used a formalism that involves solving a different Riemann–Hilbert problem[13] from the one that we solved.

II. THE *H* AND *F* POTENTIALS

A. A class of vacuum space-times for which a linear problem is known to exist

During the last decade much progress has been made handling not only the Einstein field equations but also other nonlinear systems of partial differential equations. Invariably, the key to success is the reduction of the nonlinear problem to a linear one, to which traditional methods may be applied. In general relativity this proved to be possible under certain restrictive assumptions, the principal one being the existence of two commuting Killing vector fields. No one has yet had any success when less symmetry is assumed. Moreover, even when two commuting Killing vector fields are assumed, additional assumptions prove to be necessary if a reduction to a linear problem is to be achieved. Fortunately, the vacuum and electrovac cases fall within the province of problems that can be handled in this way, at least in the absence of a cosmological constant.

We shall be concerned here with vacuum space-times for which the line element may be expressed in the form

$$\sum_{a,b=1}^{2} g_{ab}(u,v)dx^a\, dx^b + 2g_{uv}(u,v)du\, dv. \tag{1}$$

Here $\mathbf{X}_1 := \partial/\partial x^1$ and $\mathbf{X}_2 := \partial/\partial x^2$ are Killing vectors,

$$g_{ab} := \mathbf{X}_a \cdot \mathbf{X}_b \tag{2}$$

has signature $+\ +$, $g_{uv} < 0$, and

$$\rho := \sqrt{g_{11}g_{22} - (g_{12})^2} > 0 \tag{3}$$

over the domain of the chart which consists of all (x^1,x^2,u,v) such that $(x^1,x^2) \in \mathbb{R}^2$, and (u,v) is a member of a connected open subset of R^2. This class of vacuum space-times contains the Kasner solutions as well as the set CW_1 of vacuum metrics which we defined in Paper II. It is sufficiently broad to cover all conceivable vacuum space-times that we are likely to consider in the current sequence of papers.

B. The *H* potential

When one considers vacuum space-times that possess two commuting Killing vector fields, it is useful to introduce a 2×2 matrix generalization H of the Ernst potential $\mathscr{E}$.[14] It should be mentioned that the H potential was originally introduced by Kinnersley[15] in quite a different way from the way we shall now employ. Moreover, throughout the following discussion, the reader should bear in mind that we invariably suppress the wedge symbol $\wedge$ in exterior products of differential forms.

1. Definition of the H potential

We begin with the fact that the Lie derivative of a p-form Y with respect to a vector field $\mathbf{X}$ is expressible in the form

$$\mathscr{L}_{\mathbf{X}} Y = (-1)^p[\mathbf{X}\, dY - d(\mathbf{X}Y)]. \tag{4}$$

When, as in this equation, we write a vector field immediately to the *left* of a differential form, we intend that the differential form should be evaluated as a linear functional acting upon the vector field. Thus, for example, we have

$$\mathbf{X}_\mu\, dx^\nu = \delta^\nu_\mu. \tag{5}$$

More generally, if u is a p-form, v is a one-form, and $\mathbf{w}$ is a vector field, then such contractions are to be evaluated using the relation

$$\mathbf{w}(uv) = u(\mathbf{w}v) - (\mathbf{w}u)v. \tag{6}$$

Suppose now that $\mathbf{X}_1$ and $\mathbf{X}_2$ are commuting Killing vector fields. It then follows that

$$\mathscr{L}_{\mathbf{X}_a} X_b = 0 \quad (a,b = 1,2), \tag{7}$$

where X_b is the *covector* of $\mathbf{X}_b$. Since

$$\mathbf{X}_a X_b = g_{ab} \quad (a,b = 1,2), \tag{8}$$

it follows from Eq. (4) that

$$dg_{ab} = \mathbf{X}_a dX_b. \tag{9}$$

The expression dX_b is a two-form, which can be separated into *self-dual* and *anti-self-dual* parts. Assuming that X_b is real, we may express dX_b in the form

$$-2dX_b = W_b + W_b^*, \tag{10}$$

and we may identify W_b as a self-dual two-form, W_b^* as an anti-self-dual two-form. Hence Eq. (9) may be reexpressed as

$$dg_{ab} = -\mathrm{Re}(\mathbf{X}_a W_b). \tag{11}$$

Now, observe that the Lie derivative of the two-form W_b with respect to the Killing vector field $\mathbf{X}_a$ must also vanish. It follows from Eq. (4) that

$$d(\mathbf{X}_a W_b) = \mathbf{X}_a\, dW_b. \tag{12}$$

However, in the case of a vacuum space-time, it can be shown that

$$dW_b = 0. \tag{13}$$

Therefore, there exists a complex potential H_{ab} such that

$$\mathbf{X}_a W_b = dH_{ab}. \tag{14}$$

Because of Eq. (11), the constants of integration may be chosen so that

$$-\mathrm{Re}(H_{ab}) = g_{ab} \quad (a,b = 1,2). \tag{15}$$

2. The self-duality relation

As a result of the self-dual nature of W, the H potential satisfies a relation which we like to call the "self-duality relation." This relation may be expressed in the form [Ref. 8, Eq. (31)]

$$\tfrac{1}{2}(H + H^\dagger)\Omega\, dH = (z - \rho *)dH, \tag{16}$$

where $H^\dagger$ is the Hermitian conjugate of H,

$$\Omega = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \tag{17}$$

the real field z is defined by

$$2iz := H_{12} - H_{21}, \tag{18}$$

and $*$ is a *two-dimensional duality operator* such that

$$*du = +du, \quad *dv = -dv. \tag{19}$$

3. The z and ρ fields

The real field z defined by Eq. (18) is intimately related to the field ρ defined by Eq. (3). In fact [Ref. 8, Eqs. (26) and (30)],

$$*d\rho = -dz, \quad *dz = -d\rho. \tag{20}$$

Of course, one can use the fields z and ρ, which are like the Weyl canonical fields in the stationary axisymmetric field problem, as coordinates, but it should be kept in mind that $-\rho$ then plays the role of a time coordinate. Generally, however, we prefer to use coordinates u and v such that, in region IV,

$$z = u^2 - v^2, \quad \rho = 1 - u^2 - v^2, \tag{21}$$

and we shall do so throughout this paper.

4. Relationship to Ernst potentials

Both H_{11} and H_{22} can be shown to satisfy the Ernst equation.[14] We shall arbitrarily choose to denote H_{22} by $\mathscr{E}$. As we saw in Paper II, it is convenient to follow Chandrasekhar and Ferrari,[16] and introduce a second Ernst potential E which is directly related to the metric tensor components by

$$E := (g_{22})^{-1}(\rho + i g_{12}). \tag{22}$$

In practice, the H potential can be computed from $\mathscr{E} = f + i\chi$ or from $E = F + i\omega$ by employing three equations derived from the real or imaginary parts of three components of Eq. (16):

$$f^{-1}\, d\chi = -F^{-1} * d\omega, \tag{23}$$

$$d(\operatorname{Im} H_{12}) = F * df + \omega\, d\chi, \tag{24}$$

$$d(\operatorname{Im} H_{11}) = -2\rho * d\omega + 2 g_{12} * dF - f^{-1} g_{11}\, d\chi. \tag{25}$$

It should be noted that

$$f = -g_{22}, \quad fF = -\rho. \tag{26}$$

Thus, for example, one may first evaluate $E = F + i\omega$ and f directly from the metric, then evaluate the *twist potential* χ by integrating Eq. (23). This provides

$$H_{22} = \mathscr{E} = f + i\chi.$$

The real parts of the other components of the H matrix are determined by Eq. (15), while the imaginary parts are determined using Eqs. (18), (24), and (25).

5. The H potential of the Kasner metrics

Let us denote by h the 2×2 matrix whose elements are g_{ab} $(a,b = 1,2)$. In the case of the Kasner metrics, we have

$$h^K = \begin{pmatrix} \rho^{1+n} & 0 \\ 0 & \rho^{1-n} \end{pmatrix}, \tag{27}$$

where we shall refer to the exponent n as the *Kasner parameter*. Using Eqs. (23)–(25) we can easily establish that

$$H^K = \begin{pmatrix} -\rho^{1+n} & i(1-n)z \\ -i(1+n)z & -\rho^{1-n} \end{pmatrix}. \tag{28}$$

6. The class CW_1 of colliding wave solutions

In Paper II, Eq. (2.30), we identified a simple criterion for a solution of the vacuum field equations to be designated as a *colliding wave solution*. It is a necessary and sufficient condition that each of the numbers

$$k := |E_v(0,0)/2F(0,0)|^2, \tag{29}$$

$$l := |E_u(0,0)/2F(0,0)|^2, \tag{30}$$

equals 1. Here subscripts are used to denote partial derivatives with respect to u and v, and the fields are evaluated at $u = v = 0$.

C. The F potential

1. Definition of the F potential

The F potential is a 2×2 matrix field that depends not only upon the nonignorable space-time coordinates, but also upon a complex parameter t analogous to the "spectral parameter" of which others often speak. The F potential itself plays the role assumed by Lax pairs in analyses of other exactly soluble systems, and it is the generator of the infinite hierarchy of complex potentials of Kinnersley.[15] From our perspective, however, the analytic properties of the F potential, regarded as a function of t, assume the greatest importance.

For a given H potential, the F potential is defined to be any 2×2 matrix solution of the equations

$$dF(t) = \Gamma(t)\Omega F(t), \tag{31}$$

$$F(0) = \Omega, \tag{32}$$

$$\dot{F}(0) = H, \tag{33}$$

where

$$\Gamma(t) := t[1 - 2t(z - \rho *)]^{-1}\, dH \tag{34}$$

is a 2×2 matrix of one-forms (which can be computed from the H potential), and where $\dot{F}(t)$ denotes the partial derivative of $F(t)$ with respect to t. For fixed (z,ρ), $F(t)$ is required to be holomorphic in a neighborhood of $t = 0$. Moreover, we require that $F(t)$ be chosen so that, for fixed (z,ρ), it is holomorphic on the t plane minus two cuts that lie on the real axis intervals $t \geqslant \frac{1}{2}$ or $t \leqslant -\frac{1}{2}$, and join ∞ to the branch points[8] at $t = 1/[2(z \mp \rho)]$. Even with this gauge restriction, note that $F(t)$ remains arbitrary up to multiplication on its right by any space-time independent 2×2 matrix $U(t)$ that is an entire function of t and satisfies $U(0) = I$ and $\dot{U}(0) = 0$.

It is important to have a clear understanding of the cuts of $F(t)$. Since, in region IV of the space-times we shall consider, z and ρ are given by Eqs. (21), we conclude that

$$z + \rho = 1 - 2v^2 \quad \text{and} \quad z - \rho = 2u^2 - 1 \tag{35}$$

both lie in the interval $[-1, +1]$. Therefore, both branch points lie outside the interval $(-\frac{1}{2}, +\frac{1}{2})$.

When $|z| < \rho$,

$$1/[2(z-\rho)] \leqslant -\tfrac{1}{2}, \quad 1/[2(z+\rho)] \geqslant +\tfrac{1}{2}. \tag{36}$$

In particular, the origin lies within the gap between the two branch points. We shall introduce a pair of cuts along the real axis of the t plane, one extending leftward to infinity from the branch point at $t = 1/[2(z-\rho)]$, and the other extending rightward to infinity from the branch point at $t = 1/[2(z+\rho)]$. See Fig. 1.

When $z > \rho$, the branch point $1/[2(z+\rho)]$ lies at or to the right of $t = +\frac{1}{2}$, and the other branch point is further to the right. The cuts are then chosen as shown in Fig. 2. When

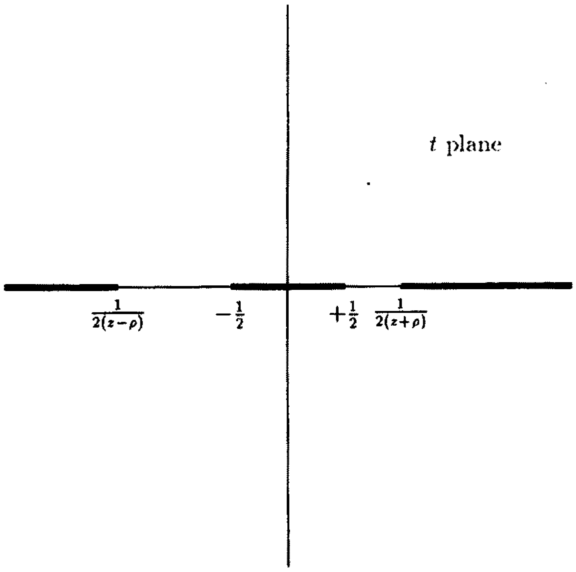


FIG. 1. Cuts associated with $u(t)$ and $F(t)$ when $|z| < \rho$.

$z = \rho$, the layout is similar except that the cut from $1/[2(z-\rho)]$ to ∞ reduces to the singlet set $\{\infty\}$.

When $z < \rho$, the branch point $1/[2(z-\rho)]$ lies at or to the left of $t = -\frac{1}{2}$, and the other branch point is further to the left. The cuts are then as in Fig. 3. When $z = -\rho$, the cut from $1/[2(z+\rho)]$ to ∞ is the singlet set $\{\infty\}$.

2. The F potential of the Kasner metrics

In the case of the Kasner metrics, a solution of Eqs. (31)–(33) may be constructed from the well-known F potential of Minkowski space by using a general expression[5] for the F potential corresponding to those space-times for which

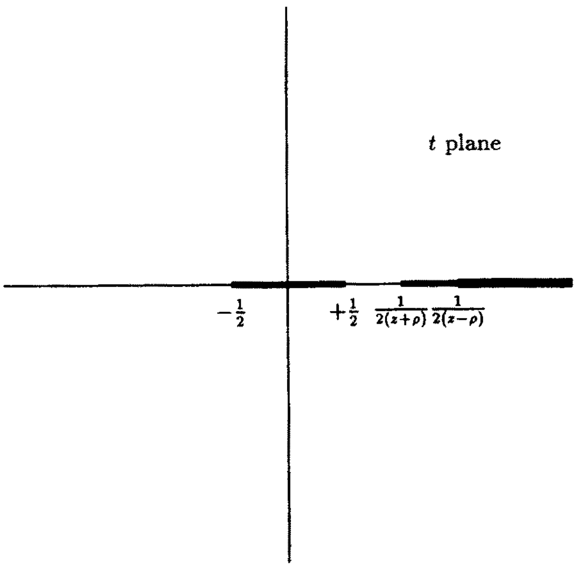


FIG. 2. Cuts associated with $u(t)$ and $F(t)$ when $z > \rho$.

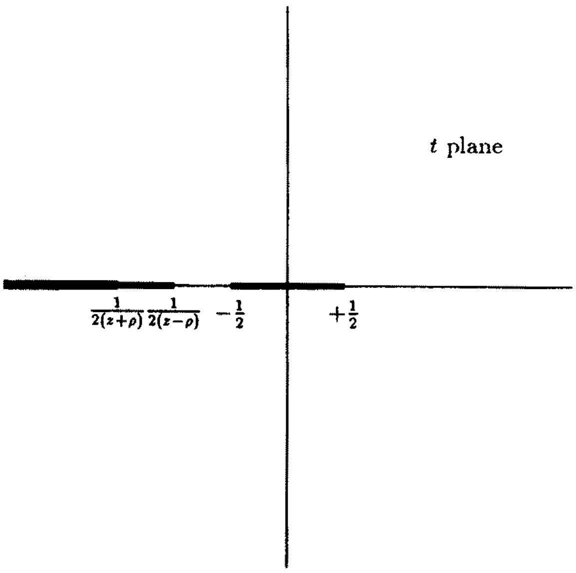


FIG. 3. Cuts associated with $u(t)$ and $F(t)$ when $z < -\rho$.

$g_{12} = 0$ in the line element (1). The result is given by $F^K(t)$

$$= \frac{1}{\lambda(t)} \begin{pmatrix} -t(\rho^2/\Lambda)^{(1+n)/2} & i\Lambda^{(1+n)/2} \\ -i\Lambda^{(1-n)/2} & -t(\rho^2/\Lambda)^{(1-n)/2} \end{pmatrix}, \tag{37}$$

where

$$\lambda(t) := \sqrt{(1-2tz)^2 - (2t\rho)^2}, \tag{38}$$

$$\Lambda(t) := \tfrac{1}{2}[1 - 2tz + \lambda(t)]. \tag{39}$$

The field $\lambda(t)$ plays a fundamental role in the general theory of the F potential. For example, one can prove[8] that a space-time-independent factor $U(t)$ which multiplies $F(t)$ on its right can always be chosen so that

$$\begin{aligned} &\det F(t) = -1/\lambda(t), \\ &[F(t^*)]^* = [-1/\lambda(t)][I - t(H + H^T)\Omega]F(t), \end{aligned} \tag{40}$$

which can readily be shown to hold for the F^K given by Eq. (37) and which are imposed as a matter of convention on all F potentials. Note that this still leaves $F(t)$ arbitrary up to multiplication by a $U(t)$ that, in addition to the properties already given, has unit determinant and is real for real t.

The field $\lambda(t)$ has branch points at $t = 1/[2(z \mp \rho)]$. We introduce the cuts defined in Sec. II C 1 and select that branch of $\lambda(t)$ for which $\lambda(0) = 1$, whereupon $F^K(t)$ is holomorphic in the cut t plane and satisfies all stipulated equations. Observe that, for any point t_0 on the cuts that $t_0 \neq 1/[2(z \mp \rho)]$ and $t_0 \neq \infty$, the limits of $F^K(t)$ as $t \to t_0$, either from above or from below the cuts, exist and are finite. Moreover, the branch points at $1/[2(z \mp \rho)]$ are of index $-\frac{1}{2}$, and $t = \infty$ is generally a pole or branch point whose order or index, as the case may be, is n dependent. This type of singularity structure is typical not only of the Kasner F potentials, but also of the colliding wave F potentials that we shall encounter later.

III. A HOMOGENEOUS HILBERT PROBLEM FOR EFFECTING KINNERSLEY–CHITRE TRANSFORMATIONS

In Paper II we introduced the Geroch group as the *free product* of two SL(2,ℝ) groups, one of which induced rational linear transformations of the E potential, and one of which induced rational linear transformations of the $\mathscr{E}$ potential. We shall in this section introduce a homogeneous Hilbert problem, the solution of which permits one to effect any given Kinnersley–Chitre transformation. As our first application of the HHP approach, we shall reproduce the two SL(2,ℝ) groups of transformations we considered in Paper II. Then we shall formulate and give the solution of the HHP that arises when one considers a double-Harrison transformation.

The realization of the Geroch group we shall use here is the multiplicative group of all space-time-independent, 2×2 matrices $u(t)$ that (i) are real in the sense that $u^* = u$, where

$$u^*(t) := [u(t^*)]^*; \tag{41}$$

(ii) have determinants equal to 1, and (iii) are each holomorphic in a neighborhood of ∞ except perhaps at ∞ itself. This group will be denoted by $K[\mathrm{SL}(2,\mathbb{R})]$. We shall also have occasion to use the group $K[\mathrm{SL}(2,\mathbb{C})]$, which is defined in the same way except that the reality condition on $u(t)$ is dropped.

For any given $u(t)$ in $K[\mathrm{SL}(2,\mathbb{R})]$, consider the F potential of some space-time that you would like to transform, and restrict attention to those space-time points for which $u(t)$ is holomorphic everywhere on the cuts of $F(t)$, except perhaps at $t=\infty$. (By definition, every cut is understood to include its end points.) Our homogeneous Hilbert problem involves identifying a matrix field $F'(t)$, holomorphic in the same cut plane as $F(t)$, and a matrix field $X_-(t)$, holomorphic on both cuts of $F(t)$, and satisfying

$$F'(t)u(t)F(t)^{-1} = X_-(t), \tag{42}$$

$$F'(0) = \Omega. \tag{43}$$

One may establish[8] that $F'(t)$ is the F potential of a space-time, and that $H' := \dot{F}'(0)$ is the H potential of that space-time. The metric is computed using $h' := -\operatorname{Re} H'$.

A. Example: Rational linear transformations of the E potential

It is very easy to identify a $u(t)$ that generates rational linear transformations of the E potential; namely, any t-independent $u(t)$. If we express such a u matrix in the form

$$u(t) = \Omega^{-1} w \Omega, \tag{44}$$

where

$$w = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \tag{45}$$

and $\alpha\delta - \beta\gamma = 1$, then the solution of the HHP (42) is quite obvious; namely,

$$X_-(t) = w, \tag{46}$$

$$F'(t) = wF(t)w^T. \tag{47}$$

It follows immediately that

$$H' = wHw^T, \tag{48}$$

$$h' = whw^T. \tag{49}$$

As we already pointed out in Paper II, the transformation (49) induces a rational linear transformation,

$$E' = i(\alpha E + i\beta)/(\gamma E + i\delta), \tag{50}$$

of the E potential.

B. Example: Rational linear transformations of the $\mathscr{E}$ potential

Consider a $u(t)$ of the form

$$u(t) = \begin{pmatrix} \alpha & \beta t \\ \gamma t^{-1} & \delta \end{pmatrix}, \tag{51}$$

where $\alpha\delta - \beta\gamma = 1$. By expanding each t-dependent matrix in Eq. (42) in a neighborhood of $t = 0$ one easily determines the form of $X_-(t)$ everywhere. This, of course, also determines $F'(t)$. The details of this calculation have been given elsewhere.[17] In particular, it has been shown that the transformation (51) induces the rational linear transformation

$$\mathscr{E}' = i(\alpha\mathscr{E} + i\beta)/(\gamma\mathscr{E} + i\delta) \tag{52}$$

of the $\mathscr{E}$ potential.

C. Example: The double-Harrison transformation

When the so-called double-Harrison transformation[18] was applied for the first time[19] to derive a colliding wave solution from a Kasner metric, some of the subtleties associated with the transformation were not well understood. There seemed, in particular, to be a mysterious minus sign in the transformed metric h'. We now understand this phenomenon much better.

For the case $F(t) = F^K(t)$, we should like to consider a double-Harrison transformation[18] of the form

$$\mathscr{F}'(t)u(t)[F(t)]^{-1} = \mathrm{X}_-(t), \tag{53}$$

$$u(t) = e^{j\eta(t)}, \tag{54}$$

$$j^2 = I, \tag{55}$$

$$\operatorname{Tr} j = 0, \tag{56}$$

$$\eta(t) = \tfrac{1}{2}\ln[(1+2t)/(1-2t)], \tag{57}$$

where the cut for $\eta(t)$ is chosen to be the straight line segment which joins its branch points $\pm\frac{1}{2}$, and where we select that branch for which $\eta(\infty) = -i\pi/2$. Also, we restrict (z,ρ) to values for which

$$1/[2(z-\rho)] \neq -\tfrac{1}{2}$$

and

$$1/[2(z+\rho)] \neq +\tfrac{1}{2},$$

i.e., according to Eq. (35), $u\neq 0$ and $v\neq 0$. The solution, whose derivation will be given in Sec. III C 1, is as follows:

$$\mathscr{F}'(t) = e^{A\eta(t)}F(t)e^{-j\eta(t)}, \tag{58}$$

$$\mathrm{X}_-(t) = e^{A\eta(t)}, \tag{59}$$

$$A := [(M+M^T)\Omega]/[(M-M^T)\Omega], \tag{60}$$

$$M := F(-\tfrac{1}{2})\,\tfrac{1}{2}(I+j)\Omega[F(+\tfrac{1}{2})]^T. \tag{61}$$

In particular, the transformed H potential is given by

$$\mathscr{H}' = H + 2A\Omega, \tag{62}$$

except for an inconsequential gauge transformation.

The catch is that $u(t)$ does not satisfy the reality condition. To understand this, note that, on the real axis to the right of $+\frac{1}{2}$ and to the left of $-\frac{1}{2}$, the imaginary part of $\eta(t)$ is $-i\pi/2$. Therefore, the provisional $u(t)$ given by Eq. (54) is imaginary! It is a member of $K[\mathrm{SL}(2,\mathbb{C})]$, but not of $K[\mathrm{SL}(2,\mathbb{R})]$.

Nevertheless, we knew that this $K[\mathrm{SL}(2,\mathbb{C})]$ transformation did, except for the mysterious sign change in h', yield the Nutku–Halil colliding wave solution when it was applied to the isotropic Kasner metric.[19] It turns out that what one should really identify as a double-Harrison transformation is

$$u(t) = e^{j\eta(t)} i\sigma_3, \tag{63}$$

where j and $\eta(t)$ are as before, and

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is a Pauli spin matrix. The revised u-matrix (63), unlike the original one (54), does satisfy the reality condition. Furthermore, the following theorem explains both why the original $K[\mathrm{SL}(2,\mathbb{C})]$ transformation worked and why there was a change in sign of h'.

Theorem 1: Suppose $u(t)$ is any given member of $K[\mathrm{SL}(2,\mathbb{C})]$, $F(t)$ is the Kasner F potential of Eq. (37), and $F'(t)$ is the solution of the HHP

$$F'(t)u(t)F(t)^{-1} = X_-(t).$$

Then

$$\mathscr{F}'(t) := -[F'(t^*)]^*$$

and

$$\mathrm{X}_-(t) := X_-(t^*)^* i\sigma_3$$

satisfy the HHP

$$\mathscr{F}'(t)[-u^*(t)i\sigma_3]F(t)^{-1} = \mathrm{X}_-(t).$$

The theorem follows easily from the relation

$$[F^K(t^*)]^* = \sigma_3 F^K(t)\sigma_3, \tag{64}$$

which is implied by Eq. (37).

Identifying $u(t)$ with the expression in Eq. (63), we see that $-u^*(t)i\sigma_3$ turns out to be nothing but the expression in Eq. (54). Moreover, if we define $\mathscr{H} := \mathscr{F}(0)$, then

$$\mathscr{H} = -H^*. \tag{65}$$

Thus, the metric is given by

$$h = -\mathrm{Re}\, H = +\mathrm{Re}\, \mathscr{H}. \tag{66}$$

D. Derivation: The double-Harrison transformation

Our derivation will be divided into phases (A), (B) and (C). Phases (A) and (B) will detail how we arrived at the solution (58)–(61) by a mixture of educated guessing and deduction. Phase (C) will contain the actual proof that Eqs. (58)–(61) constitute a solution of the HHP (53). In this connection, there is a general theorem[8] that guarantees the uniqueness of the solution if it exists.

(A) We shall first make an educated guess about the t dependence of $\mathrm{X}_-(t)$.

From Eqs. (54)–(57) and the fact that $(I \pm j)/2$ are projection matrices whose product is zero, it follows that

$$u(t) = \frac{1}{2}(I+j)\sqrt{\frac{1+2t}{1-2t}} + \frac{1}{2}(I-j)\sqrt{\frac{1-2t}{1+2t}}, \tag{67}$$

whereupon the analytic properties required of the various factors in the HHP (53) inform us that the only t-plane singularities of $X_-(t)$ are branch points of index $\frac{1}{2}$ or $-\frac{1}{2}$ at $t = \pm\frac{1}{2}$. This suggests that perhaps

$$\mathrm{X}_-(t) = A_1\sqrt{\frac{1+2t}{1-2t}} + A_2\sqrt{\frac{1-2t}{1+2t}}, \tag{68}$$

where A_1 and A_2 are t-independent matrix fields. Now, $u(0) = \pm I$ is implied by Eq. (67), and $\mathrm{X}_-(0) = \pm I$ is then implied by Eqs. (32) and the HHP (53). Therefore, from Eq. (68),

$$A_1 + A_2 = I.$$

It follows that there is a t-independent matrix field A such that

$$\mathrm{X}_-(t) = \frac{1}{2}(I+A)\sqrt{\frac{1+2t}{1-2t}} + \frac{1}{2}(I-A)\sqrt{\frac{1-2t}{1+2t}}. \tag{69}$$

A remains to be determined.

(B) We shall next grant that $\mathrm{X}_-(t)$ has the form (69), and show that the HHP then implies that A is given by Eqs. (60) and (61), and that $\mathrm{X}_-(t)$ and $\mathscr{F}'(t)$ are given by Eqs. (59) and (58), respectively.

Upon expressing the HHP (53) in the alternative form

$$\mathscr{F}'(t) = \mathrm{X}_-(t)F(t)u(t)^{-1}, \tag{70}$$

and substituting (67) and (69) into the above, we obtain

$$\begin{aligned}\mathscr{F}'(t) = {} & \tfrac{1}{4}(I+A)F(t)(I+j) \\ & + \tfrac{1}{4}(I-A)F(t)(I-j) \\ & + \tfrac{1}{4}(I+A)F(t)(I-j)\frac{1+2t}{1-2t} \\ & + \tfrac{1}{4}(I-A)F(t)(I+j)\frac{1-2t}{1+2t}.\end{aligned} \tag{71}$$

Multiplication of the above by $(1-2t)(1+2t)$, followed by setting $t = \frac{1}{2}$ or, alternatively, $t = -\frac{1}{2}$, yields the equations

$$(I+A)F(\tfrac{1}{2})(I-j) = (I-A)F(-\tfrac{1}{2})(I+j) = 0 \tag{72}$$

or, equivalently,

$$AF(\tfrac{1}{2})\tfrac{1}{2}(I-j) = -F(\tfrac{1}{2})\tfrac{1}{2}(I-j), \tag{73}$$

$$AF(-\tfrac{1}{2})\tfrac{1}{2}(I+j) = F(-\tfrac{1}{2})\tfrac{1}{2}(I+j). \tag{74}$$

Note that any traceless 2×2 matrix multiplied by Ω is symmetric. Therefore, if M is defined as in Eq. (61),

$$M^T = -F(\tfrac{1}{2})\tfrac{1}{2}(I-j)\Omega F^T(-\tfrac{1}{2}). \tag{75}$$

Upon multiplying Eq. (73) on the right by $\Omega F^T(-\frac{1}{2})$ and Eq. (74) on the right by $\Omega F^T(\frac{1}{2})$, we obtain

$$AM^T = -M^T, \quad AM = M, \tag{76}$$

which are equivalent to Eqs. (73) and (74), respectively.

Now, Eqs. (75) and (76) imply that

$$(M - M^T)\Omega \neq 0.$$

Moreover, any antisymmetric 2×2 matrix times Ω is a multiple of the unit matrix. Hence Eqs. (76) yield that expression for A given by Eqs. (60) and (61).

Equations (76) also imply that the two eigenvalues of A are ± 1. Therefore,

$$\operatorname{tr} A = 0, \quad A^2 = I, \tag{77}$$

from which we see that $X_-(t)$, as given by Eq. (69), is expressible in the neat form (59). The expression (58) for $\mathscr{F}'(t)$ then follows from Eqs. (54), (59), and (70).

Finally, we complete this phase of the derivation by noting that Eqs. (32), (33), and (71) imply

$$\mathscr{H}' = \dot{\mathscr{F}}'(0) = H + 2A\Omega - 2\Omega j. \tag{78}$$

For any given metric, the H potential is arbitrary up to addition of any constant matrix that is both imaginary and symmetric. Therefore, in the above equation, we can drop the term $-2\Omega j$ with impunity. The result is that expression for $\mathscr{H}'$ which is given by Eq. (62).

(C) We next prove that Eqs. (58)–(61) constitute a solution of the HHP (53).

First, note that the expressions (58) and (59) for $\mathscr{F}'(t)$ and $X_-(t)$ satisfy the HHP identically when substituted into it. Second, $X_-(t)$ as given by Eq. (59) is clearly holomorphic on the cuts of $F(t)$, and $\mathscr{F}'(t)$ as given by Eq. (58) is clearly holomorphic at $t = 0$ and satisfies $\mathscr{F}'(0) = \Omega$. To complete the proof, it remains only to show that $\mathscr{F}'(t)$, as given by Eq. (58), is holomorphic on the same cut plane as $F(t)$. We shall employ Eq. (72), which is equivalent to Eq. (58).

Consider the definitions (60) of A and (61) of M. Equation (61) implies that M has rank equal to 1. Therefore,

$$M\Omega M^T\Omega = M^T\Omega M\Omega = (\det M)\, I = 0,$$

which imply Eqs. (76), which, in turn, imply Eqs. (77). Recall that Eqs. (73) and (74) are equivalent to Eqs. (76), which are, in turn, equivalent to Eqs. (72).

Equations (59) and (77) imply that the expression (69) for $X_-(t)$ holds. Substitution of (69) and (67) into Eq. (58) then yields the expression (72) for $\mathscr{F}'(t)$.

Let us next apply Eqs. (72) to the right side of the expression (72) for $\mathscr{F}'(t)$, whereupon it is seen that the only t-plane singularities of this expression lie on the cuts of $F(t)$. That completes the proof.

As regards prior efforts on the material that we have covered in Secs. III C and III C 1, the solution of a similar but different HHP has been given by Hauser.[20] Part of the derivation in Sec. III C is patterned after a derivation of that solution due to Hauser and Ernst.[17]

IV. APPLICATION OF DOUBLE-HARRISON TO THE KASNER METRICS

A. The solution for the output H potential

We begin by evaluating $F^K(t)$ at $t = \pm\frac{1}{2}$. It is convenient to write

$$F^K(t) = F_0^K(t)S, \tag{79}$$

where

$$S := \begin{pmatrix} 2^{n/2} & 0 \\ 0 & 2^{-n/2} \end{pmatrix}. \tag{80}$$

Then

$$M = F_0^K(-\tfrac{1}{2})M_0[F_0^K(+\tfrac{1}{2})]^T, \tag{81}$$

where

$$M_0 := S\tfrac{1}{2}(I + j)\Omega S^T \tag{82}$$

is a constant matrix, which may be parametrized as follows:

$$M_0 = \frac{i}{2q}\begin{pmatrix} -p-p' & q+q' \\ -q+q' & -p+p' \end{pmatrix}, \tag{83}$$

where

$$p = \cos\nu, \quad q = \sin\nu, \quad p' = \cos\nu', \quad q' = \sin\nu'. \tag{84}$$

The values of $F_0^K(t)$ at $t = \pm\frac{1}{2}$ are best expressed in terms of (x,y) coordinates, where

$$z = xy, \quad \rho = XY, \quad X := \sqrt{1-x^2}, \quad Y := \sqrt{1-y^2}. \tag{85}$$

This we do by identifying

$$\lambda(+\tfrac{1}{2}) = x - y, \tag{86}$$

$$\lambda(-\tfrac{1}{2}) = x + y, \tag{87}$$

$$\Lambda(+\tfrac{1}{2}) = \tfrac{1}{2}(1+x)(1-y), \tag{88}$$

$$\Lambda(-\tfrac{1}{2}) = \tfrac{1}{2}(1+x)(1+y), \tag{89}$$

$$\rho^2/\Lambda(+\tfrac{1}{2}) = 2(1-x)(1+y), \tag{90}$$

$$\rho^2/\Lambda(-\tfrac{1}{2}) = 2(1-x)(1-y). \tag{91}$$

We then find that

$$[F_0^K(+\tfrac{1}{2})]^T = \frac{1}{\sqrt{2}(x-y)}\begin{pmatrix} -[(1-x)(1+y)]^{(1+n)/2} & -i[(1+x)(1-y)]^{(1-n)/2} \\ i[(1+x)(1-y)]^{(1+n)/2} & -[(1-x)(1+y)]^{(1-n)/2} \end{pmatrix} \tag{92}$$

and

$$F_0^K(-\tfrac{1}{2}) = \frac{1}{\sqrt{2}(x+y)}\begin{pmatrix} [(1-x)(1-y)]^{(1+n)/2} & i[(1+x)(1+y)]^{(1+n)/2} \\ -i[(1+x)(1+y)]^{(1-n)/2} & [(1-x)(1-y)]^{(1-n)/2} \end{pmatrix}. \tag{93}$$

It is then simple to evaluate M using Eq. (81) and to compute the output H potential using Eq. (62).

If one introduces the notation

$$T(n,\nu,\nu') := \frac{1}{2} X \left[(p+p')\left(\frac{1-x}{1+x}\right)^{n/2} + (p-p')\left(\frac{1+x}{1-x}\right)^{n/2} \right] + \frac{i}{2} Y \left[(q+q')\left(\frac{1-y}{1+y}\right)^{n/2} + (q-q')\left(\frac{1+y}{1-y}\right)^{n/2} \right], \quad (94)$$

then the output H potential may be expressed in the following form:

$$\mathscr{H}_{11}(n,\nu,\nu') = \rho^{1+n} \frac{T(n+2,\nu,\nu')}{T(n,\nu,\nu')}, \quad (95)$$

$$\mathscr{H}_{22}(n,\nu,\nu') = \rho^{1-n} \frac{T(n-2,\nu,\nu')}{T(n,\nu,\nu')}, \quad (96)$$

$$\mathscr{H}_{12}(n,\nu,\nu') = i(1-n)z - 2i \times \frac{uVT(n,\nu',\nu) - vUT(n,\nu',\nu)^*}{T(n,\nu,\nu')}. \quad (97)$$

As regards the above solution for the output H potential, recall that the points $(u,0)$ and $(0,v)$ in region IV were necessarily avoided in the HHP (53). It is important to note that, in spite of this avoidance, the final solution can be analytically continued in the (u,v) plane so that the extended domain covers the points $(u,0)$ and $(0,v)$.

B. The metric components g_{ab} and the E potential of the solution

Using the identity

$$|T(n,\nu,\nu')|^2 = \mathrm{Re}[T(n-1,\nu',\nu)T(n+1,\nu',\nu)^*], \quad (98)$$

one may express the metric components $g_{ab} = \mathrm{Re}\,\mathscr{H}_{ab}$ in the following way:

$$g_{11} = \rho^{1+n} \frac{|T(n+1,\nu',\nu)|^2}{|T(n,\nu,\nu')|^2}, \quad (99)$$

$$g_{22} = \rho^{1-n} \frac{|T(n-1,\nu',\nu)|^2}{|T(n,\nu,\nu')|^2}, \quad (100)$$

$$g_{12} = -\rho \frac{\mathrm{Im}[T(n-1,\nu',\nu)T(n+1,\nu',\nu)^*]}{|T(n,\nu,\nu')|^2}. \quad (101)$$

From these expressions it is clear, furthermore, that the complex E potential,

$$E := (\rho + ig_{12})/g_{22}, \quad (102)$$

is given by

$$E(n,\nu,\nu') = \rho^n \frac{T(n+1,\nu',\nu)}{T(n-1,\nu',\nu)}. \quad (103)$$

C. The computation of g_{uv} (i.e., of N), A, and B

The field equations which govern g_{uv} are

$$2u\Gamma_u = 1 - \rho\,|E_u/2F|^2, \quad (104)$$

$$2v\Gamma_v = 1 - \rho\,|E_v/2F|^2, \quad (105)$$

where

$$e^{2\Gamma} := -\sqrt{\rho}\, g_{uv}. \quad (106)$$

The solution may be constructed from E, which was given in Eq. (103). The result may be expressed as

$$e^{2\Gamma} = N/UV, \quad (107)$$

$$N = \rho^{n^2/2} |T(n,\nu,\nu')|^2, \quad (108)$$

or, if one prefers,

$$E = A/B, \quad (109)$$

$$N = \mathrm{Re}(AB^*), \quad (110)$$

$$A = \rho^{[(n+1)^2-1]/4}\, T(n+1,\nu',\nu), \quad (111)$$

$$B = \rho^{[(n-1)^2-1]/4}\, T(n-1,\nu',\nu). \quad (112)$$

It should be noted that the resulting three-parameter family of solutions of the vacuum Einstein equations belongs to the set CW_1 that we introduced in Paper II. This follows from the fact that the constants k and l defined by Eqs. (29) and (30) both evaluate to 1.

V. PERSPECTIVES

In this paper we have exploited a double-Harrison transformation to generate from the F potential of the Kasner metrics a three-parameter generalization of the two-parameter family of colliding wave solutions discovered by Ferrari, Ibañez, and Bruni. It should be noted that the solution (58)–(61) of the associated HHP (53)–(57) holds even when $F \neq F^K$. However, Theorem 1, which was used in order to cope with the fact that $u(t)$ did not satisfy the reality condition, must be replaced by the following more generally applicable theorem.

Theorem 2: Let F' be the solution of the HHP

$$F' e^{j\eta} i\sigma_3 F^{-1} = X_-,$$

and $\mathscr{F}'$ be the solution of the HHP

$$\mathscr{F}' e^{j'\eta} F^{-1} = \mathrm{X}_-,$$

where $j' := \sigma_3 j \sigma_3$. Then

$$\mathscr{F}' = -\sigma_3 F' \sigma_3, \quad \mathrm{X}_- = i\sigma_3 X_-.$$

The derivation of new colliding wave solutions through the application of the double-Harrison transformation to other input F potentials will be the subject of a future paper.

ACKNOWLEDGMENTS

The authors would like to thank Professor Ferrari, Professor Ibañez, and Professor Bruni for kindly sending them a preprint of their article.[12]

One of the authors (FJE) of the present paper was supported in part by Grants No. PHY-8605958 and No. PHY-8306684 from the National Science Foundation.

[1]F. J. Ernst, Alberto García D., and I. Hauser, J. Math. Phys. **28**, 2155 (1987). Henceforth cited as Paper I.
[2]F. J. Ernst, Alberto García D., and I. Hauser, J. Math. Phys. **28**, 2951 (1987). Henceforth cited as Paper II.
[3]Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).

[4]R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).
[5]W. Kinnersley and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926, 2037 (1978).
[6]C. Hoenselaers, W. Kinnersley, and B. C. Xanthopoulos, J. Math. Phys. **20**, 2530 (1979).
[7]I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 362 (1979).
[8]I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980).
[9]I. Hauser and F. J. Ernst, J. Math. Phys. **22**, 1051 (1981).
[10]I. Hauser and F. J. Ernst, "A new proof of an old conjecture," in *Gravitation and Geometry*, edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987), pp. 165–214.
[11]D. Kramer and G. Neugebauer, Commun. Math. Phys. **10**, 132 (1968).
[12]V. Ferrari, J. Ibañez, and M. Bruni, Phys. Rev. D **36**, 1053 (1987).
[13]V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).
[14]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[15]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977).
[16]S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984).
[17]I. Hauser and F. J. Ernst, "The Riemann–Hilbert approach to the axial Einstein equations," in *Solitons in General Relativity*, edited by H. C. Morris and R. Dodd (Plenum, New York, to be published).
[18]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978).
[19]F. J. Ernst, "Derivation of Nutku–Halil colliding plane wave solution from isotropic Kasner metric using double-Harrison transformation," in *Gravitational Collapse and Relativity*, edited by H. Sato and T. Nakamura (World Scientific, Singapore, 1987), pp. 141–149.
[20]I. Hauser, "On the homogeneous Hilbert problem for effecting Kinnersley–Chitre transformations," in *Lecture Notes in Physics, Vol. 205, Solutions of Einstein's Equations: Techniques and Results*, edited by C. Hoenselaers and W. Dietz (Springer, Berlin, 1984), pp. 128–175.

On the collision of planar impulsive gravitational waves

A. H. Taub
Department of Mathematics, University of California, Berkeley, California 94720

(Received 17 September 1987; accepted for publication 21 October 1987)

Exact continuous solutions to the Einstein field equations are determined under the assumption that the subspaces of space-time spanned by the variables X^A $(A = 1,2)$ admit the three-dimensional group of motions of a two-plane, and the Ricci tensor R_{AB} $(A,B = 1,2)$ vanishes. The space-time is assumed to contain two colliding planar impulsive gravitational waves. Each wave may be followed by a distribution of null dust. It is shown that the Cauchy data on a spacelike three-surface does not lead to a unique solution of the Einstein field equations unless additional requirements are imposed on the stress-energy tensor in the region of interaction of the waves.

I. INTRODUCTION

Chandrasekhar and Xanthopolous have shown that there is an ambiguity in the evolution of a space-time containing two colliding-plane impulsive gravitational waves whose leading edges are followed by distributions of null dust. This result follows from their two papers[1,2] in which they report two different exact solutions of the Einstein field equations in space-times admitting two commuting spacelike Killing vectors in which two such gravitational waves collide. In the first paper the region of interaction of the waves is shown to contain a perfect fluid with pressure equal to the energy density. In the second one this region is shown to be filled with a mixture of two noninteracting null dusts moving in opposite directions.

It is purpose of this paper to discuss plane-symmetric space-times containing planar colliding impulsive gravitational waves. That is, space-times admitting three spacelike Killing vectors that generate the group of motions of a two-dimensional plane will be treated. In such a space-time, coordinates may be introduced (cf. Ref. 3) in which the line element contains only two functions of two coordinates. The Einstein field equations are simpler than those solved in Refs. 1 and 2, and classes of solutions of these equations are readily obtained. The nature of the ambiguity in the evolution of such a space-time containing two colliding planar waves with or without trailing distributions of null dust can be determined.

The discussion given below will be modeled after that given in Refs. 1 and 2. Namely, we shall discuss solutions of the Einstein field equations in the region of interaction of the impulsive planar gravitational waves and extend these solutions to the regions of space-time prior to the instant of collision by requiring that the metric tensor be continuous across the null hypersurfaces describing the boundary between the various regions. The derivatives of the metric tensor need not be continuous and the curvature tensors of the space-times discussed will be distribution valued. The formalism developed in Ref. 4 will be used in the sequel.

It has been pointed out in Ref. 3 that coordinates in a plane-symmetric space-time may be chosen so that the line element may be written as

$$ds^2 = g_{ij}\,dx^i\,dx^j - g_{AB}\,dx^A\,dx^B\,, \tag{1.1}$$

where $i,j = 0,3$, $A,B = (1,2)$,

$$g_{ij} = e^{\omega}\eta_{ij} = e^{\omega}(\delta_{ij} - 2\delta_i^3\delta_j^3), \quad g_{AB} = e^{\mu}\delta_{AB}. \tag{1.2}$$

Equation (1.1) is equivalent to

$$ds^2 = e^{\omega}\,du\,dv - e^{\mu}(dx^2 + dy^2)\,, \tag{1.3}$$

with

$$u = t - z, \quad v = t + z\,. \tag{1.4}$$

The functions ω and μ depend on only the variables u and v. The nonvanishing components of the Ricci tensor are given by

$$R_{uu} = \mu_{,uu} + \tfrac{1}{2}\mu^2_{,u} - \mu_{,u}\omega_{,u}\,, \tag{1.5a}$$

$$R_{uv} = \mu_{,uv} + \tfrac{1}{2}\mu_{,u}\mu_{,v} + \omega_{,uv}\,, \tag{1.5b}$$

$$R_{vv} = \mu_{,vv} + \tfrac{1}{2}\mu^2_{,v} - \mu_{,v}\omega_{,v}\,, \tag{1.5c}$$

$$R_{AB} = 2g_{AB}e^{-\omega}(\mu_{,uv} + \mu_{,u}\mu_{,v}) = 2e^{-\omega}(e^{\mu})_{,uv}\delta_{AB}\,. \tag{1.5d}$$

The scalar curvature

$$R = 4e^{-\omega}(2\mu_{,uv} + \tfrac{3}{2}\mu_{,u}\mu_{,v} + \omega_{,uv}) \tag{1.6}$$

and the nonvanishing components of the Einstein tensor

$$G_{\alpha\beta} = R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R$$

are

$$G_{uu} = R_{uu}\,, \tag{1.7a}$$

$$G_{uv} = -(\mu_{,uv} + \mu_{,u}\mu_{,v})\,, \tag{1.7b}$$

$$G_{vv} = R_{vv}\,, \tag{1.7c}$$

$$G_{AB} = -g^{uv}R_{uv}g_{AB}\,, \tag{1.7d}$$

where we have used the notation

$$f_{,u} = \frac{\partial f}{\partial u}, \quad f_{,v} = \frac{\partial f}{\partial v},$$

$$f_{,uu} = \frac{\partial^2 f}{\partial u^2}, \quad f_{,uv} = \frac{\partial^2 f}{\partial u\,\partial v}, \quad f_{,vv} = \frac{\partial^2 f}{\partial v^2}.$$

The indices u and v are related to indices α $(\alpha = 0,3,1.2)$ by the equations

$$V_u = V_\alpha X^\alpha_{,u}, \quad V_v = V_\alpha X^\alpha_{,u}\,.$$

 0022-2488/88/030690-06$02.50

It may be shown that the nonvanishing components of the Reimann curvature tensor are

$$R_{ABCD} = e^{-\omega}\mu_{,u}\mu_{,v}(g_{AD}g_{BC} - g_{AC}g_{BD}) , \quad (1.8a)$$

$$R_{uABu} = \tfrac{1}{2}R_{uu}g_{AB} , \quad (1.8b)$$

$$R_{uABv} = \tfrac{1}{2}(R_{uv} - \omega_{,uv})g_{AB} , \quad (1.8c)$$

$$R_{vABv} = \tfrac{1}{2}R_{vv}g_{AB} , \quad (1.8d)$$

$$R_{uvuv} = \tfrac{1}{2}e^{\omega}\omega_{,uv} . \quad (1.8e)$$

II. $R_{AB}=0$ AND $\mu_{,u}\mu_{,v}\neq 0$

It follows from these conditions and Eqs. (1.5d) that

$$\mu = 1 + U(u) + V(v) , \quad (2.1)$$

where $U(u)$ and $V(v)$ are arbitrary nonconstant functions of their arguments. Let

$$\omega = \Omega - \tfrac{1}{2}\mu + \ln(U'V') - \ln(AB) , \quad (2.2)$$

where the prime denotes the derivative of a function with respect to its argument and

$$A = U'(0), \quad B = V'(0) . \quad (2.3)$$

The line element given by Eq. (1.3) is then given by

$$ds^2 = \pm dU\,dV e^{\Omega}/AB(1 + U + V)^{1/2} - (1 + U + V)(dx^2 + dy^2) , \quad (2.4)$$

where the sign of the first term is to be chosen so that the coefficient of $dU\,dV$ is positive.

It is no restriction to set

$$U + V = Au + Bv . \quad (2.5)$$

In addition we may take

$$Au + Bv = \pm k(v \pm u) , \quad (2.6)$$

where $k = +|AB|^{1/2}$, the sign of u is the same as the sign of AB and the sign of k is the sign of A when AB is positive and the sign of B when AB is negative.

Equation (2.4) then becomes

$$dS^2 = e^{\Omega}\,du\,dv/(1 + Au + Bv)^{1/2} - (1 + Au + Bv)(dx^2 + dy^2) , \quad (2.7)$$

where there are four nonequivalent choices for $Au + Bv$ given by Eq. (2.6) (only two if we admit the transformation $u \to -u, v \to -v$).

It follows from Eqs. (2.1) and (2.6) that Eqs. (1.5a)–(1.5c) become

$$(1 + Au + Bv)R_{uu} = -A\Omega_{,u} , \quad (2.8a)$$

$$R_{uv} = \Omega_{,uv} , \quad (2.8b)$$

$$(1 + Au + Bv)R_{vv} = -B\Omega_{,v} . \quad (2.8c)$$

The integrability conditions for these equations are

$$-R_{uv} = ((1 + Au + Bv)A^{-1}R_{uu})_{,v} = ((1 + Au + Bv)B^{-1}R_{vv})_{,u} . \quad (2.9)$$

These equations are equivalent to the Bianchi identities. When they are satisfied we have

$$\Omega = -\int (1 + Au + Bv)(A^{-1}R_{uu}\,du + B^{-1}R_{vv}\,dv) . \quad (2.10)$$

The line integral in this equation may be taken along an arbitrary curve in the u-v plane joining an arbitrary point (ξ,η) to the point (u,v).

In the subsequent discussion in this section we shall be mainly concerned with the region of the u-v plane where $u \geqslant 0$ and $v \geqslant 0$. The line integral in equation will be evaluated on the curve consisting of the interval of the v (u) axis from (0,0) to $(0,v)$ [$(u,0)$] and the line parallel to the u (v) axis joining $(0,v)$ [$(u,0)$] to (u,v).

We may also write

$$\Omega = \Phi + \ln(\alpha(u)\beta(v)) , \quad (2.11)$$

where

$$A \ln \alpha(u) = \int_0^u (1 + A\xi)R_{uu}(\xi,0)d\xi , \quad (2.12a)$$

$$B \ln \beta(v) = \int_0^v (1 + B\eta)R_{vv}(0,\eta)d\eta , \quad (2.12b)$$

$$\Phi = \int_0^u \int_0^v R_{uv}\,du\,dv . \quad (2.12c)$$

Equation (2.7) may then be written as

$$ds^2 = \alpha(u)\beta(v)e^{\Phi}\,du\,dv/(1 + Au + Bv)^{1/2} - (1 + Au + Bv)(dx^2 + dy^2) . \quad (2.13)$$

Note that the line element given by Eq. (2.7) [or (2.13)] is completely determined in the region $u > 0$, $v > 0$ when R_{uv} is given as a function of these variables in this region and $R_{uu}(u,0)$ $(u>0)$ and $R_{vv}(0,v)$ $(v>0)$ are known.

If $\Omega(u,v)$ is a C^2 function of u and v in the region $u \geqslant 0$, $v \geqslant 0$, Eq. (2.7) determines the line element of space-time and Eqs. (2.8) determine the Ricci tensor in this region. Note that when $\Omega(u,v)$ is required to be continuous across $u = 0$ $(v = 0)$, then $\Omega_{,v}$ $(\Omega_{,u})$ is continuous across this null hypersurface.

III. PARTICULAR SOLUTIONS

Case (a): When

$$R_{AB} = R_{uu} = R_{vv} = 0 , \quad (3.1)$$

it follows from Eq. (2.9) that

$$R_{uv} = \Omega_{,uv} = 0 . \quad (3.2)$$

That is, when Eqs. (3.1) hold, the space-time must satisfy the Einstein vacuum field equations. Equations (2.8) then imply that Ω is a constant which may be taken to be zero. Equations (2.6) and (2.7) then give two nonequivalent metrics that satisfy the vacuum field equations. When the plus sign in the coefficient of u is used in Eq. (2.6), the metric is one of the Kasner vacuum solutions given in Ref. 3. When the minus sign is used, it becomes the stationary plane-symmetric vacuum solution given in that reference. The metrics given by Eqs. (2.6) and (2.7) are also given by Dray and 't Hooft[5] and said by them to be "sometimes referred to as Robinson's nullicle."

Case (b): $R_{AB} = R_{uv} = \Omega_{,uv} = 0$. The only nonvanishing components of the Ricci tensor are

$$R_{uu} = -\epsilon_1 \quad \text{and} \quad R_{vv} = -\epsilon_2 \quad (3.3)$$

and the Einstein tensor is given by

$$-G_{\alpha\beta}=\epsilon_1 l_\alpha l_\beta+\epsilon_2 n_\alpha n_\beta\ , \tag{3.4}$$

where

$$l_\alpha=u_{,\alpha}\quad\text{and}\quad n_\alpha=v_{,\alpha}\ , \tag{3.5}$$

with

$$g^{\alpha\beta}l_\alpha l_\beta=g^{\alpha\beta}n_\alpha n_\beta=0\ .$$

Hence

$$l_{\alpha;\beta}l^\beta=l_{\beta;\alpha}l^\beta=0\ ,\quad n_{\alpha;\beta}l^\beta=n_{\beta;\alpha}n^\beta=0\ ,$$

where the semicolon denotes the covariant derivative.

The Bianchi identities

$$G^{\alpha\beta}{}_{;\beta}=0$$

then imply that

$$(\epsilon_1 l^\alpha)_{;\alpha}l^\beta+(\epsilon_2 n^\alpha)_{;\alpha}n^\beta=0\ .$$

Since l^α and n^α are linearly independent we must have

$$(\epsilon_1 l^\alpha)_{;\alpha}=(\epsilon_2 n^\alpha)_{;\alpha}=0\ .$$

That is, the source of a gravitational field given by Eq. (3.4) is that of two noninteracting null fluids. For such a source the line element is given by Eqs. (2.7) and (2.10). It is also given by Eq. (2.13) with $\Phi=0$, and α and β may be determined from the values of Ω on the u and v axes. The latter is the line element used by Dray and 't Hooft[5] in their discussion of colliding planar shells of matter, i.e., colliding impulsive gravitational waves.

Case (c): It is a consequence of Eqs. (2.9) that if $R_{AB}=0$ and

$$R_{uu}R_{vv}-(R_{uv})^2=0\ ,$$

i.e., $R_{\alpha\beta}$ is of rank 1 so that

$$R_{\alpha\beta}=-\tau_\alpha\tau_\beta\ , \tag{3.6}$$

then there exists a function $\sigma(u,v)$ such that

$$\tau_\alpha=\sigma_{,\alpha} \tag{3.7}$$

In other words, if Eqs. (2.9) and (3.6) hold, then the stress-energy tensor that describes the source of the gravitational field is that due to the gradient of a massless scalar field or equivalently of a perfect fluid with pressure equal to energy density. The planar space-times satisfying the Einstein field equations for such a source were discussed by Tabensky and Taub.[6] The proof that Eq. (3.7) follows from Eqs. (3.6) and (2.9) is immediate, since if we substitute for R_{uu} and R_{vv} from Eq. (3.6) into (2.9) we obtain

$$2(1+Au+Bv)\tau_{u,v}=2(1+Au+Bv)\tau_{v,u}=(A\tau_v+B\tau_u)\ . \tag{3.8}$$

It follows from these equations that

$$\tau_{u,v}=\tau_{v,u}$$

and hence Eq. (3.7) obtains.

Equation (3.8) then becomes

$$2(1+Au+Bv)\sigma_{,uv}=-(A\sigma_{,v}+B\sigma_{,u})\ . \tag{3.9}$$

For every solution of this partial differential equation for σ there is a space-time whose source is given by the stress-energy tensor determined by the gradient of σ (i.e., a perfect fluid with energy density equal to pressure). The line element of this space-time is given by Eq. (2.7).

The line element used in Ref. 6 may be obtained from that given in Eq. (2.7) with plus signs used in Eq. (2.6) by the simple transformation

$$u=\bar u-(2k)^{-1},\quad v=\bar v-(2k)^{-1}\ ,$$

followed by a rescaling of the coordinates. Then Eq. (3.9) becomes

$$2(u+v)\sigma_{,uv}=-(\sigma_{,u}+\sigma_{,v})\ . \tag{3.10}$$

Solutions of this equations are determined from the values of σ on the null hypersurfaces $u=0$ and $v=0$. Given such a solution we may determine $\Omega(u,v)$ from Eq. (2.10) or (2.11) by using the equations

$$R_{\alpha\beta}=-\sigma_{,\alpha}\sigma_{,\beta}\ .$$

Note that the particular solutions discussed above may be characterized as follows: case (a), $\Omega=0$; case (b), $\Omega_{,uv}=0$, $\Omega_{,u}\neq0$, and $\Omega_{,v}\neq0$; and case (c), Ω given by Eq. (2.10) or (2.11) where

$$R_{\alpha\beta}=-\sigma_{,\alpha}\sigma_{,\beta} \tag{3.11}$$

with σ given as a solution of Eq. (3.9). Different solutions may be obtained from different specifications of the function $\Omega(u,v)$ via different solutions of the Bianchi identities, Eqs. (2.9), and the field equations, (2.8).

IV. $R_{AB}=0$ AND $\mu_{,u}\mu_{,v}=0$

These conditions imply that the functions $U(u)$, $V(v)$ of Eq. (2.1) are such that either one or both are constant. In case μ is a constant it follows from Eqs. (1.5) that the only nonvanishing component of the Ricci tensor is R_{uv}, when $\omega_{,uv}\neq0$. In this case the Einstein tensor is of the form

$$G_{\alpha\beta}=G_{AB}\delta^A{}_\alpha\delta^B{}_\beta=-g_{AB}g^{uv}R_{uv}\delta^A{}_\alpha\delta^B{}_\beta\ .$$

That is, the Einstein tensor has no nonvanishing timelike proper value and hence cannot be equated to a physically plausible stress-energy tensor.

Thus for physical reasons we must set $\omega_{,uv}=0$ when $\mu_{,u}=\mu_{,v}=0$. In this case the Riemannian curvature tensor vanishes and the space-time is flat.

The situation that obtains when $\mu_{,v}=0$ may be derived from that which holds when $\mu_{,u}=0$ by replacing the variable v and $V(v)$ by u and $U(u)$, respectively. When $\mu_{,u}=0$, $U(u)$ is a constant and with no loss of generality, we may take it to vanish. Equation (2.1) becomes

$$\mu=\ln(1+V(v))\ . \tag{4.1}$$

If we now define

$$\omega=\Omega-\tfrac{1}{2}\mu+\ln(V'/B)\ , \tag{4.2}$$

the line element given by Eq. (1.3) becomes

$$ds^2=\pm e^\Omega\,dV\,du/B(1+V)^{1/2}-(1+V)(dx^2+dy^2)\ . \tag{4.3}$$

It is no restriction to set

$$V=Bv\ , \tag{4.4}$$

and Eq. (4.3) becomes

$$ds^2=e^\Omega\,du\,dv/(1+Bv)^{1/2}-(1+Bv)(dx^2+dy^2)\ . \tag{4.5}$$

Thus when $\mu_{,u}=0$ we may assume that the line element obtained from (2.6) by setting $A = 0$ holds. It further follows from (4.1), (4.4), and (4.5) that

$$R_{uu} = 0\,, \tag{4.6a}$$

$$R_{uv} = \Omega_{,uv}\,, \tag{4.6b}$$

$$(1+Bv)R_{vv} = -B\Omega_{v} \tag{4.6c}$$

[Eqs. (2.8) with $A = 0$]. Hence we must have

$$((1+Bv)R_{vv})_{,u} = -BR_{uv} \tag{4.7}$$

as an integrability condition (Bianchi identity) of Eqs. (4.6b) and (4.6c). It then follows that if $R_{vv} = 0$ we must have $R_{uv} = 0$. That is, the Ricci tensor vanishes and the Riemannian curvature tensor also vanishes. Thus such a space time is flat.

Equation (4.6c) may be integrated to give

$$\Omega = -\int_0^v (1+B\eta)R_{vv}\,d\eta + \ln f(u)\,.$$

It is no restriction to take $f(u) = 1$, for by the transformation

$$\bar{u} = \int f(u)du\,,$$

the term $f(u)du$ in the line element (4.5) (with the bars omitted) becomes $d\bar{u}$. Thus we have as the line element of the space-time the expression

$$ds^2 = e^{\Omega}\,du\,dv/(1+Bv)^{1/2} - (1+Bv)(dx^2+dy^2)\,, \tag{4.8}$$

with

$$B\Omega = -\int_0^v (1+B\eta)R_{vv}\,d\eta\,. \tag{4.9}$$

When $\Omega_{,u} = 0$ and thus $R_{uv} = 0$ in addition to $R_{AB} = R_{uu} = 0$, the Einstein tensor of the space-time is

$$G_{\alpha\beta} = R_{vv}n_\alpha n_\beta\,,$$

where n_α is the null vector defined by the second of Eqs. (3.5). That is, the stress-energy tensor of such a space-time is that of a null fluid with energy density proportional to $-R_{vv}$, and with four-velocity n^α.

The Bianchi identities

$$G^{\alpha\beta}{}_{;\beta} = 0$$

ensure that

$$(n^\alpha R_{vv})_{;\alpha} = 0;$$

that is, R_{vv} is conserved under the motion of the null fluid.

As was pointed out earlier, when $\Omega = 0$, then $R_{vv} = 0$, in addition to the assumptions made above, the space-time is flat and the line element (4.9) becomes

$$ds^2 = du\,dv/(1+Bv)^{1/2} - (1+Bv)(dx^2+dy^2)\,, \tag{4.10}$$

or

$$ds^2 = d\bar{u}\,d\bar{v} - (1+B\bar{v}/2)^2(dx^2+dy^2)\,,$$

when $u = \bar{u}$ and

$$1 + Bv = (1+B\bar{v}/2)^2\,.$$

When $\mu_{,v} = 0$ we may use the results of Sec. II with $B = 0$ and find that $R_{vv} = 0$. In addition, equations similar to Eqs. (4.6), (4.8), and (4.9) must hold. That is

$$ds^2 = e^{\Omega}\,du\,dv/(1+Au)^{1/2} - (1+Au)(dx^2+dy^2)\,, \tag{4.11}$$

$$-A\Omega = \int_0^u (1+A\xi)R_{uu}\,d\xi\,. \tag{4.12}$$

Further, we must have

$$-AR_{uv} = ((1+Au)R_{uu})_{,v}\,, \tag{4.13}$$

the equations obtained from Eq. (2.9) when $B = 0$. The analogs to Eq. (4.6) are

$$(1+Au)R_{uu} = -A\Omega_{,u}\,, \tag{4.14a}$$

$$R_{uv} = \Omega_{,uv}\,, \tag{4.14b}$$

$$R_{vv} = 0\,. \tag{4.14c}$$

When $R_{uu} = R_{vv} = R_{AB} = 0$, then $R_{uv} = 0$, and as follows from Eqs. (1.8) the Riemann curvature tensor vanishes. That is, the space-time is flat. When $\mu_{,v} = R_{AB} = R_{uv} = 0$, the only possible nonvanishing component of the Ricci tensor is R_{uu} and hence the Einstein tensor is given by

$$G_{\alpha\beta} = R_{uu}l_\alpha l_\beta\,,$$

where l_α is the null vector defined by the first of Eqs. (3.5). That is, the source of the gravitational field is a null fluid. In addition we must have

$$(R_{uu}l^\alpha)_{;\alpha} = 0\,.$$

V. EXTENSION OF SOLUTIONS

In this section we shall assume that in a region of space-time we may introduce coordinates u, v, x, y which are such that in region I, where $u>0$ and $v>0$, the line element is given by Eqs. (2.6) and (2.7), where $\Omega(u,v)$ is a known C^2 function of u and v. That is, $R_{AB} = 0$ and R_{uu}, R_{uv}, and R_{vv} are determined by Eqs. (2.8). We shall then extend such a solution across the null hypersurfaces $u = 0$ and $v = 0$ by assuming that the metric tensor is continuous across these hypersurfaces but has discontinuous derivatives across them. Such space-times were discussed in Ref. 4 and were shown to have distribution-valued curvature tensors, i.e., curvature tensors that contain Dirac delta functions whose coefficients depended on the values of the discontinuities of the first derivatives of the metric tensor.

The method of that paper enables one to determine the distribution-valued Ricci and Riemann curvature tensors of the resulting space-time. These tensors may also be calculated by using the equations of Sec. I with

$$\mu = \mu^D = \ln(1 + \mathscr{A}(u) + \mathscr{B}(v))\,, \tag{5.1a}$$

$$\omega = \omega^D = \Omega^D - \tfrac{1}{4}\mu^D\,, \tag{5.1b}$$

where

$$\mathscr{A}(a) = Au\theta(u)\,, \tag{5.1c}$$

$$\mathscr{B}(v) = Bv\theta(v)\,, \tag{5.1d}$$

$$\Omega^D = (\Omega^{\rm I} - \Omega^{\rm II} - \Omega^{\rm III} + \Omega^{\rm IV})\theta(u)\theta(v) + \theta(u)(\Omega^{\rm II} - \Omega^{\rm IV}) + \theta(v)(\Omega^{\rm III} - \Omega^{\rm IV}) + \Omega^{\rm IV}\,, \tag{5.1e}$$

with

$$\Omega^{IV} = \Omega(0,0) = 0 , \quad (5.1f)$$

and $\theta(\phi)$ is defined by the equations

$$\theta(\phi) = \begin{cases} 1, & \phi > 0, \\ \frac{1}{2}, & \phi = 0, \\ 0, & \phi < 0. \end{cases} \quad (5.1g)$$

Also $\theta(\phi)$ may be taken to be the Heaviside function that is unity for positive and zero values of the argument and otherwise zero. Equations (5.1a) and (5.1b) define μ and ω in terms of their values in region I by replacing the variables u and v by $u\theta(u)$ and $v\theta(v)$, respectively (cf. Penrose[7]).

Thus in region I (where $u > 0$, $v > 0$),

$$\mu = \mu^{I} = \ln(1 + Au + Bv) , \quad (5.2a)$$

$$\omega = \omega^{I} = \Omega^{I}(u,v) - \tfrac{1}{2}\mu^{I} . \quad (5.2b)$$

In region II (where $u > 0$, $v < 0$)

$$\mu = \mu^{II} = \ln(1 + Au) \quad (5.3a)$$

$$\omega = \omega^{II} = \Omega^{II} - \tfrac{1}{2}\mu^{II} = \Omega(u,0) - \tfrac{1}{2}\mu^{II} . \quad (5.3b)$$

In region III (where $u < 0$, $v > 0$),

$$\mu = \mu^{III} = \ln(1 + Av) , \quad (5.4a)$$

$$\omega = \omega^{III} = \Omega^{III} - \tfrac{1}{2}\mu^{III} = \Omega(0,v) - \tfrac{1}{2}\mu^{III} , \quad (5.4b)$$

and in region IV (where $u < 0$, $v < 0$),

$$\mu^{IV} = \omega^{IV} = \Omega^{IV} = 0 . \quad (5.5)$$

On the hypersurface $\mu = 0$, it follows from Eq. (5.1e) that

$$\Omega^{D} = \theta(v)\Omega^{III} + (1 - \theta(v))\Omega^{IV} , \quad (5.6a)$$

and on $v = 0$,

$$\Omega^{D} = \theta(u)\Omega^{II} + (1 - \theta(u))\Omega^{IV} . \quad (5.6b)$$

Since we are taking $\Omega(0,0) = 0$, Eqs. (5.6a) and (5.6b) contain only the first terms in each equation irrespective of whether $\theta(\phi)$ is the Heaviside function or is defined by Eq. (5.2).

It follows from Eq. (5.1a) that

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))\mu^{D}_{,u} = A\theta(u) , \quad (5.7a)$$

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))\mu^{D}_{,v} = B\theta(v) , \quad (5.7b)$$

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))^2\mu^{D}_{,uu} = (1 + \mathscr{A}(u) + \mathscr{B}(v))A\delta(u) - A^2\theta^2(u), \quad (5.7c)$$

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))^2\mu^{D}_{,vv} = (1 + \mathscr{A}(u) + \mathscr{B}(v))B\delta(v) - B^2\theta^2(v) , \quad (5.7d)$$

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))^2\mu^{D}_{,uv} = -AB\theta(u)\theta(v) , \quad (5.7e)$$

where $\delta(u)$ is the Dirac delta function.

From Eq. (5.1e) we have

$$\Omega^{D}_{,u} = \theta(u)(\theta(v)\Omega^{I}_{,u} + (1 - \theta(v)))\Omega^{II}_{,u} , \quad (5.8a)$$

$$\Omega^{D}_{,v} = \theta(v)(\theta(u)\Omega^{I}_{,v} + (1 - \theta(u)))\Omega^{III}_{,v}, \quad (5.8b)$$

$$\Omega^{D}_{,uv} = \theta(u)\theta(v)\Omega^{I}_{,uv} . \quad (5.8c)$$

On substituting Eqs. (5.1) into Eqs. (1.5) one

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))Q_{uu} = A\delta(u) - A\theta^2(u)(\theta(v)\Omega^{I}_{,u} + (1 - \theta(v))\Omega^{II}_{,u}) , \quad (5.9a)$$

$$Q_{uv} = \Omega^{I}_{,uv}\theta(u)\theta(v) , \quad (5.9b)$$

$$(1 + \mathscr{A}(u) + \mathscr{B}(v))Q_{vv} = B\delta(v) - B\theta^2(v)(\theta(u)\Omega^{I}_{,v} + (1 - \theta(u))\Omega^{III}_{,v}) , \quad (5.9c)$$

where the $Q_{\alpha\beta}$ are the distribution-valued components of the Ricci tensor. It follows from the results in Ref. 4 that $R^{J}{}_{\alpha\beta}$ ($J =$ I,II,III,IV), the components of the Ricci tensor in region J, are given by $Q(u,v)$ with (u,v) in region J.

Hence from Eqs. (5.9) we have

$$(1 + Au + Bv)R^{I}_{uu} = -A\Omega^{I}_{,u} , \quad (5.10a)$$

$$(1 + Au)R^{II}_{uu} = -A\Omega^{II}_{,u} , \quad (5.10b)$$

$$R^{III}{}_{uu} = R^{IV}{}_{uu} = 0 , \quad (5.10c)$$

$$R^{I}{}_{uv} = \Omega^{I}{}_{,uv} , \quad (5.11a)$$

$$R^{II}{}_{uv} = R^{III}{}_{uv} = R^{IV}{}_{uv} = 0 , \quad (5.11b)$$

$$(1 + Au + Bv)R^{I}_{vv} = -B\Omega^{I}{}_{,v} , \quad (5.12a)$$

$$(1 + Bv)R^{III}{}_{vv} = -B\Omega^{III}{}_{,v} , \quad (5.12b)$$

$$R^{II}{}_{vv} = R^{IV}{}_{vv} = 0 . \quad (5.12c)$$

Note that

$$\lim_{v\to 0} R^{I}{}_{uu} = R^{II}{}_{uu} \quad (5.13a)$$

and

$$\lim_{u\to 0} R^{I}{}_{vv} = R^{III}{}_{vv} . \quad (5.13b)$$

It further follows from Eqs. (5.9) that on $u = 0$,

$$(1 + \mathscr{B}(v))Q_{uu} = A\delta(u) , \quad (5.14a)$$

and on $v = 0$,

$$(1 + \mathscr{A}(u))Q_{vv} = B\delta(v) . \quad (5.14b)$$

From the discussion given in Ref. 4 it may be concluded that the singular hypersurfaces $u = 0$ and $v = 0$ are planar shells of null matter whose stress-energy tensors are given by

$$-\kappa\tau_{\alpha\beta} = A(1 + \mathscr{B}(v))^{-1}l_\alpha l_\beta, \quad (5.15a)$$

for $u = 0$, and

$$-\kappa\tau_{\alpha\beta} = B(1 + \mathscr{A}(u))^{-1}n_\alpha n_\beta , \quad (5.15b)$$

for $v = 0$, where κ is the Einstein gravitational constant (cf. Appendix 2 of Ref. 5). Requiring that the energy density of the shells be positive forces,

$$A < 0, \quad B < 0.$$

In view of Eq. (2.6) we may take

$$A = B = -k . \quad (5.16)$$

There is an essential singularity in region I, even when $\Omega(u,v,) = 0$, along the spacelike hypersurface Σ:

$$k(u + v) = 1 . \quad (5.17)$$

Thus given a function $\Omega(u,v)$ in region I*, region I bounded by the hypersurface Σ, we may extend the space-time with metric tensor (2.7) from this region across the boundaries $u = 0$ and $v = 0$ to regions II, III, and IV. The resulting space-time has the metric tensor

$$ds^2 = \frac{\exp(\Omega^D)du\,dv}{\{1-k(u\theta(u)+v\theta(v))\}^{1/2}} - \{1-k(u\theta(u)+v\theta(v))\}(dx^2+dy^2),$$

where Ω^D is given by Eq. (5.1e).

The space-time contains two colliding planar shells of matter, the hypersurfaces $u=0$ and $v=0$, and the components of its Ricci tensor contain delta functions with support on these hypersurfaces. These null hypersurfaces may be interpreted as planar shells of null matter[5] or as the leading edges of planar impulsive gravitational waves.[7] This latter interpretation is that used by Penrose[7] especially when $\Omega^I=0$. In that case all components of the Ricci tensor vanish everywhere except on the hypersurfaces $u=0$ and $v=0$—the wave fronts—and all regions but region I* are flat. In the latter region, the region of interaction of the colliding waves (planar shells of matter), the metric describes the stationary plane-symmetric vacuum solution given in Ref. 3 and discussed in Ref. 5. We shall interpret the region I* as the region of interaction of the impulsive gravitational waves with wave fronts $u=0$ and $v=0$ in all cases irrespective of the value of Ω^I.

VI. CONCLUSIONS

As has been pointed out above, given $\Omega^I=0$ the line element (5.21) describes the evolution of a space-time in which two planar gravitational waves with wave fronts $u=0$ and $v=0$, colliding in two surfaces $u=v=0$, interact in the region I* to produce a Kasner plane-symmetric vacuum solution of the Einstein field equations. However, if we are given that two planar impulsive gravitational waves with wave fronts $u=0$ and $v=0$ propagate in a flat space-time and collide at $u=v=0$, the nature of the region I* of space-time is not uniquely determined. The ambiguity in the outcome of such a collision results from the fact that, from the data given above and Eqs. (5.10b), (5.12b), and (5.13), one can only conclude that $\Omega^I(u,v)$ is such that $\Omega^I(u,0)=\Omega^I(0,v)=0$ but otherwise arbitrary.

When $\Omega^I_{,uv}=0$ and $\Omega^I_{,u}\Omega^I_{,v}\neq 0$, the nature of region I* is described by case (b) of Sec. III, and the wave fronts $u=0$ and $v=0$ are followed by distributions of null dust. If, however, one assumes that in addition to $\Omega^I_{,uv}=0$, the regions II and III are vacuous (the assumption made throughout Ref. 5), then it follows from Eqs. (5.13) and the requirement that $\Omega^I(0,0)=0$ that $\Omega^I=0$.

When $\Omega^I_{,uv}=-\sigma_{,u}\sigma_{,v}\neq 0$ inside region I*, where σ is a nonconstant solution of Eq. (3.9), the nature of this region is described by case (c) of Sec. III. If, in addition, one requires that regions II and III be vacuous equations, (5.15) imply that σ is constant on $u=0$ and $v=0$. Then Eq. (3.9) in turn implies that σ is constant throughout region I bounded by Σ. Thus the metric for case (c) cannot be extended to regions II and III unless the leading edges of the wave fronts $u=0$ and $v=0$ are followed by distributions of null dust.

Suppose that regions II and III of a planar space-time are occupied by two planar impulsive gravitational waves with wave fronts $u=0$ and $v=0$ each followed by a distribution of null dust, and that these waves collide at $u=0$, $v=0$. The field equations (5.10)–(5.12) and Eqs. (5.13), together with the condition $\Omega(0,0)=0$, only determine $\Omega^I(u,0)$ and $\Omega^I(0,v)$, and $\Omega^I(u,v)$ is arbitrary for $u>0$ and $v>0$.

If, however, one requires that $R^{\,I}_{\;uv}=0$ in addition to the above requirements in regions II, III, and IV, then $\Omega^I(u,v)$ is uniquely determined because of Eqs. (5.13). If, instead of imposing this condition on $R^{\,I}_{\;\alpha\beta}$, one requires that $R^{\,I}_{\;\alpha\beta}$ be of rank 1 with $R^{\,I}_{\;AB}=0$, then $\Omega^I(u,v)$ is again uniquely determined from Eqs. (3.11), (3.9), and (5.13).

Thus the evolution of a planar space-time in which two planar impulsive gravitational waves collide is not uniquely determined by the Einstein field equations after the collision. That is, solutions of the generalized Einstein field equations—equations in which the curvature tensor is distribution valued—are not unique if only the initial values of the metric tensor and its derivatives are prescribed on a spacelike hypersurface. In other words, in such a case the Cauchy problem does not have a unique solution.

For planar symmetric space-times in which two planar impulsive gravitational waves collide, uniqueness can be restored by imposing various conditions on the Ricci tensor (the stress-energy tensor) in the region of interaction of the waves. Different requirements on $R^{\,I}_{\;\alpha\beta}$ is determined from the values of $R^{\,II}_{\;\alpha\beta}$ and $R^{\,III}_{\;\alpha\beta}$ on the singular hypersurfaces $u=0$ and $v=0$, respectively.

[1] S. Chandrasekhar and B. C. Xanthopolous, Proc. R. Soc. London Ser. A **402**, 37 (1985).
[2] S. Chandrasekhar and B. C. Xanthopolous, Proc. R. Soc. London Ser. A. **403**, 189 (1986).
[3] A. H. Taub, Ann. Math. **53**, 472 (1951).
[4] A. H. Taub, J. Math. Phys. **21**, 1423 (1980).
[5] T. Dray and G. 't Hooft, Class. Quantum Grav. **3**, 825 (1986).
[6] R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[7] R. Penrose, *General Relativity*, edited by L. O'Raifeartaigh (Clarendon, Oxford, 1972), pp. 101–118.

Multiplication formulas of orthogonal polynomials of boson field operators: Derivation based on the generalized phase-space method

H. M. Ito
Seismology and Volcanology Division, Meteorological Research Institute, Tsukuba, Ibaraki 305, Japan

(Received 3 July 1986; accepted for publication 21 October 1987)

The products of several orthogonal polynomials of boson field operators, the quantum mechanical version of multiple Wiener integrals, are expressed as linear combinations of the polynomials. The expression is obtained by making use of the correspondence rules of boson operators and complex numbers.

I. INTRODUCTION

A broad class of stochastic processes and random fields are decomposed into direct sums of orthogonal polynomials of a Gaussian white noise. This is called the Wiener–Itô decomposition[1–4] or, more familiarly in physics, the Wiener–Hermite expansion.[5,6] This classical result affords a quantum mechanical interpretation[4–8] by introducing annihilation and creation operators $(a(t), a^\dagger(t))(t\in R^d)$ for a boson free field in the Fock representation. The nth degree Wick products of the form

$$g_n = :Q(t_1)Q(t_2)\cdots Q(t_n):,\quad t_1,t_2,\ldots,t_n\in R^d, \tag{1.1}$$

of commuting operators

$$Q(t) = a(t) + a^\dagger(t),\quad t\in R^d, \tag{1.2}$$

acting on the vacuum state $|0\rangle$, generate n-particle subspace $\{g_n|0\rangle\}$. In the case of $d=1$, this Fock structure on the canonical commutation relation is only represented by the Wiener–Itô decomposition of square integrable random variables.[4] The Wiener–Hermite expansion refers to the cases of $d\geqslant 1$ with the same statement.[5,6]

In view of the nonlinear problems in the Wick polynomial expansions of field operators, attention was naturally directed to the multiplication formulas of Wick polynomials. Jaffe[9] gave a general formula for products of an arbitrary number of g_n's of the above type (1.1) with (1.2). The notion of Wick products itself was extended by Segal[10] to free fields in (possibly) non-Fock states with similar multiplication formulas. The results were generalized, in a context of quantum stochastic differential equations, by Nakazawa[11] to g_n's formed with noncommuting operators,

$$Q(t) = \xi(t)a(t) + \xi^*(t)a^\dagger(t),\quad t\in R^d, \tag{1.3}$$

depending on a complex function ξ.

This paper presents multiplication formulas that express products of g_n's as their linear combinations for the Fock case. The subject is in the domain of Appendix A of Jaffe,[9] but our generalization is the extension to the case (1.3). This is again in the domain of Nakazawa,[11] yet our result subsumes (3.9) of Ref. 11 for two g_n's as a special case. The main contribution of the present analysis will be the clarification of interrelations of the subject with the generalized phase-space method of Agarwal and Wolf.[12] Our method of derivation seems to be valuable because of its being lucid and systematic, thus facilitating practical calculations. This will be seen by an example whose result would have otherwise been difficult to obtain.

Starting from the multiplication formulas for functions of c-numbers,[12] we obtain those for functions of q-numbers (boson operators) with the aid of the correspondence rules between the two kinds of functions. Section II begins with the case of one degree of freedom, which is essential to our approach. As a by-product, formulas for the products of two and three Hermite polynomials are readily reproduced. We then proceed to finitely many degrees of freedom in Sec. III, and extension to boson fields will be carried out in Sec. IV using these results. Remarks will follow in Sec. V, giving a powerful application of the obtained three-term formula.

II. ONE DEGREE OF FREEDOM

Let a, $a^\dagger$ be boson annihilation and creation operators obeying a commutation relation

$$[a,a^\dagger] = 1\,. \tag{2.1}$$

Let $|\alpha\rangle$ denote the coherent state[13] defined by

$$|\alpha\rangle = \exp(\alpha a^\dagger - \alpha^* a)|0\rangle\,, \tag{2.2}$$

which is an eigenstate of the annihilation operator a,

$$a|\alpha\rangle = \alpha|\alpha\rangle\,. \tag{2.3}$$

Let us introduce a linear map $\mathscr{T}$, transforming functions of complex variables α, α^* to those of a, $a^\dagger$ by

$$\mathscr{T}F(\alpha,\alpha^*) = :F(a,a^\dagger):. \tag{2.4}$$

Here $:\cdot:$ denotes the Wick product arranging a, $a^\dagger$ in the normal order without using the commutation relation (2.1). In the normal order, the annihilation operator a is put on the right-hand side of the creation operator $a^\dagger$. Note a simple relation,

$$\mathscr{T}(\langle\alpha|F(a,a^\dagger)|\alpha\rangle) = F(a,a^\dagger)\,. \tag{2.5}$$

This is established as follows: rewrite $F(a,a^\dagger)$ in the normal order using (2.1), where the relation (2.5) clearly holds because of (2.3).

Let $F_1,F_2,\ldots,F_I$ be possibly noncommuting functions of a, $a^\dagger$. A multiplication formula for $\langle\alpha|F_1F_2\cdots F_I|\alpha\rangle$ is given by Agarwal and Wolf[12] from the viewpoint of the generalized phase-space method:

$$\langle\alpha|F_1F_2\cdots F_I|\alpha\rangle = \prod_{j=2}^{I}\prod_{i=1}^{j-1}\exp\left(\frac{\partial^2}{\partial\alpha_i\,\partial\alpha_j^*}\right) \times\prod_{k=1}^{I}\langle\alpha_k|F_k|\alpha_k\rangle\Big|_{\alpha_k=\alpha,\text{c.c.}}\,, \tag{2.6}$$

 0022-2488/88/030696-06$02.50

where $\alpha_k = \alpha$,c.c. indicates $\alpha_1 = \alpha_2 = \cdots = \alpha_I = \alpha$ and $\alpha_1^* = \alpha_2^* = \cdots = \alpha_I^* = \alpha^*$.

For a fixed complex ξ, we define Wick polynomials g_n's by

$$g_n(\xi) = \begin{cases} 1 & (n=0)\,, \\ :(\xi a + \xi^* a^\dagger)^n: & (n \geq 1)\,. \end{cases} \tag{2.7}$$

First, let us derive a multiplication formula for a product of two Wick polynomials. An application of (2.6) to $g_l(\xi) g_m(\xi)$ with $\alpha_1 = \alpha$, $\alpha_2 = \beta$ yields

$$\langle \alpha | g_l g_m | \alpha \rangle = \exp\left(\frac{\partial^2}{\partial \alpha\, \partial \beta^*}\right) \langle \alpha | g_l | \alpha \rangle \langle \beta | g_m | \beta \rangle \Big|_{\beta = \alpha, \text{c.c.}} . \tag{2.8}$$

Using the property [Eq. (2.3)] of the coherent state and the definition of the Wick product $:\cdot:$, we note

$$\langle \alpha | g_k | \alpha \rangle = (\xi \alpha + \xi^* \alpha^*)^k, \quad k = l,m\,. \tag{2.9}$$

Substituting (2.9) into (2.8) and expanding the exponential function of $\partial^2/\partial\alpha\,\partial\beta^*$, we obtain

$$\begin{aligned} \langle \alpha | g_l g_m | \alpha \rangle &= \exp\left(\frac{\partial^2}{\partial \alpha\, \partial \beta^*}\right) (\xi\alpha + \xi^*\alpha^*)^l \\ &\quad \times (\xi\beta + \xi^*\beta^*)^m \Big|_{\beta = \alpha, \text{c.c.}} \\ &= \sum_{i=0}^{\infty} (i!)^{-1} \left(\frac{\partial^2}{\partial \alpha\, \partial \beta^*}\right)^i (\xi\alpha + \xi^*\alpha^*)^l \\ &\quad \times (\xi\beta + \xi^*\beta^*)^m \Big|_{\beta = \alpha, \text{c.c.}} \\ &= \sum_{i=0}^{l \wedge m} A(l,m;i) |\xi|^{2i} (\xi\alpha + \xi^*\alpha^*)^{l+m-2i} \\ &= \sum_{i=0}^{l \wedge m} A(l,m;i) |\xi|^{2i} \\ &\quad \times \langle \alpha | :(\xi a + \xi^* a^\dagger)^{l+m-2i}: | \alpha \rangle\,, \end{aligned} \tag{2.10}$$

where

$$A(l,m;i) = l!m!/i!(l-i)!(m-i)!\,, \tag{2.11}$$

and the symbol $l \wedge m$ represents a minimum of l and m. The operation of $\mathscr{T}$ on (2.10) yields the multiplication formula

$$g_l(\xi) g_m(\xi) = \sum_{i=0}^{l \wedge m} A(l,m;i) |\xi|^{2i} g_{l+m-2i}(\xi)\,, \tag{2.12}$$

with the aid of the property (2.5).

Multiplication formulas for the product of I Wick polynomials are derived similarly. However, we will discuss only the case $I = 3$:

$$g_l(\xi) g_m(\xi) g_n(\xi) = \sum_{i,j,k} B(l,m,n;i,j,k) |\xi|^{2(i+j+k)} \times g_{l+m+n-2(i+j+k)}(\xi)\,. \tag{2.13}$$

Here,

$$B(l,m,n;i,j,k) = \frac{l!m!n!}{i!j!k!(l-i-j)!(m-j-k)!(n-k-i)!}\,, \tag{2.14}$$

and the summation in (2.13) is carried out over the range

$$0 \leq i+j \leq l, \quad 0 \leq j+k \leq m, \quad 0 \leq k+i \leq n, \quad 0 \leq i,j,k\,. \tag{2.15}$$

Derivation of (2.13) is as follows. The formula (2.6) in the present case becomes

$$\begin{aligned} \langle \alpha | g_l g_m g_n | \alpha \rangle &= \exp\left(\frac{\partial^2}{\partial\alpha\,\partial\beta^*} + \frac{\partial^2}{\partial\alpha\,\partial\gamma^*} + \frac{\partial^2}{\partial\beta\,\partial\gamma^*}\right) \\ &\quad \times \langle \alpha | g_l | \alpha \rangle \langle \beta | g_m | \beta \rangle \langle \gamma | g_n | \gamma \rangle \Big|_{\beta = \gamma = \alpha, \text{c.c.}} . \end{aligned}$$

Expanding the exponential function as

$$\sum_{i,j,k} (i!j!k!)^{-1} \left(\frac{\partial^2}{\partial\alpha\,\partial\gamma^*}\right)^i \left(\frac{\partial^2}{\partial\alpha\,\partial\beta^*}\right)^j \left(\frac{\partial^2}{\partial\beta\,\partial\gamma^*}\right)^k,$$

we have

$$\langle \alpha | g_l g_m g_n | \alpha \rangle = \sum_{i,j,k} B(l,m,n;i,j,k) |\xi|^{2(i+j+k)} \times (\xi\alpha + \xi^*\alpha^*)^{l+m+n-2(i+j+k)}\,.$$

Thus the operation of $\mathscr{T}$ and the use of (2.5) result in Eq. (2.13).

The above results immediately give multiplication formulas for the products of two and three Hermite polynomials. For

$$a = \frac{x}{2} + \frac{d}{dx}, \quad a^\dagger = \frac{x}{2} - \frac{d}{dx}, \tag{2.16}$$

the Hermite polynomial $H_n(x)$, defined by

$$H_n(x) = (-1)^n \exp\left(\frac{x^2}{2}\right) \frac{d^n}{dx^n} \exp\left(-\frac{x^2}{2}\right), \tag{2.17}$$

is expressed as

$$H_n(x) = g_n(1)\,. \tag{2.18}$$

The relation (2.18) is shown by induction; the case $n = 0,1$ is obvious. Suppose (2.18) holds for $n = k$. Then

$$\begin{aligned} :(a + a^\dagger)^{k+1}: &= a^\dagger :(a + a^\dagger)^k: + :(a + a^\dagger)^k: a \\ &= \left(\frac{x}{2} - \frac{d}{dx}\right) H_k(x) + H_k(x) \left(\frac{x}{2} + \frac{d}{dx}\right) \\ &= x H_k(x) - \frac{dH_k(x)}{dx} \end{aligned} \tag{2.19}$$

which is equal to $H_{k+1}(x)$ since

$$\frac{dH_k(x)}{dx} = k H_{k-1}(x),$$

$$H_{k+1}(x) = x H_k(x) - k H_{k-1}(x)\,.$$

The multiplication formula (2.12) is now reduced to the well-known formula

$$H_l(x) H_m(x) = \sum_{i=0}^{l \wedge m} A(l,m;i) H_{l+m-2i}(x)\,, \tag{2.20}$$

and (2.13) to the formula

$$H_l(x)H_m(x)H_n(x) = \sum_{i,j,k} B(l,m,n;i,j,k)H_{l+m+n-2(i+j+k)}(x)\,, \quad (2.21)$$

where summation ranges over (2.15).

III. MANY DEGREES OF FREEDOM

The results in Sec. II are now straightforwardly extended to many degrees of freedom. Let $\underline{a} = (a(1), a(2),...,a(J))$, $\underline{a}^\dagger = (a^\dagger(1),a^\dagger(2),...,a^\dagger(J))$ be boson annihilation and creation operators obeying the commutation relations

$$[a(s),a^\dagger(t)] = \delta_{st}\,,$$
$$[a(s),a(t)] = [a^\dagger(s),a^\dagger(t)] = 0, \quad s,t\in \mathbf{J} \equiv (1,2,...,J)\,. \quad (3.1)$$

In this section all quantities with underlines such as $\underline{\alpha}$ represent J-dimensional complex vectors; we write, for example, $\underline{\alpha} = (\alpha(1),\alpha(2),...,\alpha(J))$. By $|\underline{\alpha}\rangle$ we denote a direct product of the coherent states $|\alpha(t)\rangle_t$ of tth boson operators $(a(t),a^\dagger(t))$. Let us extend the linear map $\mathscr{T}$ of (2.4) in such a way so as to transform the functions of the complex variables $\underline{\alpha}$, $\underline{\alpha}^*$ to those of $\underline{a}$, $\underline{a}^\dagger$ by

$$\mathscr{T}F(\underline{\alpha},\underline{\alpha}^*) = :F(\underline{a},\underline{a}^\dagger):\,. \quad (3.2)$$

Here $:\cdot:$ denotes the Wick product making all the annihilation operators $a(t)$ appear on the right-hand side of the creation operators $a^\dagger(t)$ without using the commutation relations (3.1). Again, a relation

$$\mathscr{T}(\langle\underline{\alpha}|F(\underline{a},\underline{a}^\dagger)|\underline{\alpha}\rangle) = F(\underline{a},\underline{a}^\dagger) \quad (3.3)$$

holds.

Let $F_1,F_2,...,F_I$ be possibly noncommuting functions of $\underline{a}$, $\underline{a}^\dagger$. Let us start with a multiplication formula for $\langle\underline{\alpha}|F_1F_2\cdots F_I|\underline{\alpha}\rangle$ given by an extension of (2.6)[12]:

$$\langle\underline{\alpha}|F_1F_2\cdots F_I|\underline{\alpha}\rangle = \prod_{j=2}^{I}\prod_{i=1}^{j-1}\Lambda(\underline{\alpha}_i,\underline{\alpha}_j) \times \prod_{k=1}^{I}\langle\underline{\alpha}_k|F_k|\underline{\alpha}_k\rangle\Big|_{\underline{\alpha}_k=\underline{\alpha},\text{c.c.}}\,. \quad (3.4)$$

Here the differential operator $\Lambda(\underline{\alpha},\underline{\beta})$ is defined by

$$\Lambda(\underline{\alpha},\underline{\beta}) = \exp\left(\sum_{s=1}^{J}\lambda(\underline{\alpha},\underline{\beta};s)\right), \quad (3.5)$$

with

$$\lambda(\underline{\alpha},\underline{\beta};s) = \frac{\partial^2}{\partial\alpha(s)\,\partial\beta^*(s)}\,, \quad s\in\mathbf{J}\,. \quad (3.6)$$

In (3.4), the expression $\underline{\alpha}_k = \underline{\alpha}$,c.c. means $\underline{\alpha}_1 = \underline{\alpha}_2 = \cdots = \underline{\alpha}_I = \underline{\alpha}$ and $\underline{\alpha}_1^* = \underline{\alpha}_2^* = \cdots = \underline{\alpha}_I^* = \underline{\alpha}^*$.

For $t\in\mathbf{J}$, define

$$Q(\underline{\zeta},\underline{a};t) = \zeta(t)a(t) + \zeta^*(t)a^\dagger(t)\,, \quad (3.7)$$

and consider the Wick polynomials

$$g_n(\underline{\zeta}_1\underline{\zeta}_2\cdots\underline{\zeta}_n;t_1,t_2\cdots t_n) = \begin{cases}1 & (n=0)\,,\\ :Q(\underline{\zeta}_1,\underline{a};t_1)Q(\underline{\zeta}_2,\underline{a};t_2)\cdots Q(\underline{\zeta}_n,\underline{a};t_n): & (n\geqslant 1)\,.\end{cases} \quad (3.8)$$

A multiplication formula for g_lg_m now becomes

$$g_lg_m = \sum_{i=0}^{l\wedge m}\sum^{(1)}(i!)^{-1}\theta_1 g_{l+m-2i}\,, \quad (3.9)$$

with

$$\theta_1 = \prod_{p=1}^{i}\delta(t_{\rho(p)},t'_{\sigma(p)})\xi_{\rho(p)}(t_{\rho(p)})\eta^*_{\sigma(p)}(t'_{\sigma(p)})\,, \quad (3.10)$$

and g_l, g_m, g_n are written explicitly as

$$g_l = g_l(\underline{\xi}_1\underline{\xi}_2\cdots\underline{\xi}_l;t_1t_2\cdots t_l)\,, \quad (3.11)$$

$$g_m = g_m(\underline{\eta}_1\underline{\eta}_2\cdots\underline{\eta}_m;t'_1t'_2\cdots t'_m)\,, \quad (3.12)$$

$$g_{l+m-2i} = g_{l+m-2i}(\underline{\xi}_1\cdots[\underline{\xi}_\rho]\cdots\underline{\xi}_l\underline{\eta}_1\cdots[\underline{\eta}_\sigma]\cdots\underline{\eta}_m;t_1\cdots[t_\rho]\cdots t_lt'_1\cdots[t'_\sigma]\cdots t'_m)\,. \quad (3.13)$$

Here $(\rho(1),\rho(2),...,\rho(i))$ and $(\sigma(1),\sigma(2),...,\sigma(i))$ are taken from $(1,2,...,l)$ and $(1,2,...,m)$, respectively, and the summation $\Sigma^{(1)}$ is carried out over all such $l!m!/\{(l-i)!(m-i)!\}$ combinations. The symbol $\delta(\cdot,\cdot)$ is Kronecker's delta, and $[\underline{\xi}_\rho]$, $[\underline{\eta}_\sigma]$, $[t_\rho]$, and $[t'_\sigma]$ indicate the exclusions of $\underline{\xi}_{\rho(1)},...,\underline{\xi}_{\rho(i)}$, $\underline{\eta}_{\sigma(1)},...,\underline{\eta}_{\sigma(i)}$, $t_{\rho(1)},...,t_{\rho(i)}$, and $t'_{\sigma(1)},...,t'_{\sigma(i)}$, respectively.

Let us derive (3.9) by using (3.4) with $I=2$ and $\underline{\alpha}_1=\underline{\alpha}$, $\underline{\alpha}_2=\underline{\beta}$. Expanding the exponential operator Λ, we obtain

$$\langle\underline{\alpha}|g_lg_m|\underline{\alpha}\rangle = \Lambda(\underline{\alpha},\underline{\beta})\langle\underline{\alpha}|g_l|\underline{\alpha}\rangle\langle\underline{\beta}|g_m|\underline{\beta}\rangle\Big|_{\underline{\beta}=\underline{\alpha},\text{c.c.}}$$
$$= \sum_{i=0}^{\infty}\sum_{s_1,s_2,...,s_i=1}^{J}(i!)^{-1}\prod_{p=1}^{i}\lambda(\underline{\alpha},\underline{\beta};s_p) \times \Pi_1\Pi_2\Big|_{\underline{\beta}=\underline{\alpha},\text{c.c.}}\,, \quad (3.14)$$

where

$$\Pi_1 = \prod_{x=1}^{l}Q(\underline{\xi}_x,\underline{\alpha};t_x)\,, \quad (3.15)$$

$$\Pi_2 = \prod_{y=1}^{m}Q(\underline{\eta}_y,\underline{\beta};t'_y)\,. \quad (3.16)$$

The operation of $\lambda(\underline{\alpha},\underline{\beta};s_p)$ transforms (3.14) into

$$\sum_{i=0}^{l\wedge m}\sum_{s_1,s_2,...,s_i=1}^{J}(i!)^{-1}\sum^{(2)}\Pi_4\prod_{p=1}^{i}\xi_{\rho(p)}(t_{\rho(p)})\delta(s_p,t_{\rho(p)}) \times \sum^{(3)}\Pi_5\prod_{p=1}^{i}\eta^*_{\sigma(p)}(t'_{\sigma(p)})\delta(s_p,t'_{\sigma(p)})\Big|_{\underline{\beta}=\underline{\alpha},\text{c.c.}}$$
$$= \sum_{i=0}^{l\wedge m}(i!)^{-1}\sum^{(1)}\theta_1\Pi_4\Pi_5\Big|_{\underline{\beta}=\underline{\alpha},\text{c.c.}} \quad (3.17)$$

Here the summation $\Sigma^{(2)}$ is carried out over all $l!/(l-i)!$ choices of $\rho(1),\rho(2),...,\rho(i)$ from $\{1,2,...,l\}$, and $\Sigma^{(3)}$ is done over all $m!/(m-i)!$ choices of $\sigma(1),\sigma(2),...,\sigma(i)$ from $\{1,2,...,m\}$. The factors Π_4,Π_5 are defined by omitting $x=\rho(1),\rho(2),...,\rho(i)$ and $y=\sigma(1),\sigma(2),...,\sigma(i)$ in (3.15) and (3.16), respectively. In the last expression of (3.17),

$$\Pi_4\Pi_5\Big|_{\underline{\beta}=\underline{\alpha},\text{c.c.}} = \langle\underline{\alpha}|g_{l+m-2i}|\underline{\alpha}\rangle\,,$$

which follows from the property of the coherent state and the definition of g_n's (3.8). By operating $\mathscr{T}$ on (3.17) and using the property (3.3), we readily obtain (3.9).

Here we remark that the relation (3.9) implies the orthogonality of Wick polynomials g_n in the sense

$$\langle \underline{0}|g_l g_m|\underline{0}\rangle = 0 \quad (l \neq m)\,, \tag{3.18}$$

since $\langle \underline{0}|g_n|\underline{0}\rangle = 0$ if $n \neq 0$.

A multiplication formula for the product of the three polynomials g_l, g_m, and

$$g_n = g_n(\zeta_1\zeta_2\cdots\zeta_n;t''_1t''_2,\dots,t''_n) \tag{3.19}$$

is derived similarly:

$$g_l g_m g_n = \sum_{i,j,k} (i!j!k!)^{-1}\Sigma^{(4)}\theta_2\theta_3\theta_4\, g_{l+m+n-2(i+j+k)}\,. \tag{3.20}$$

Here the summation with respect to i,j,k is carried out over the range (2.15), and the summation $\Sigma^{(4)}$ is done over all $l!m!n!/\{(l-i-j)!(m-j-k)!(n-k-i)!\}$ combinations of choosing $(\rho(1),\rho(2),\dots,\rho(i+j))$, $(\sigma(1),\sigma(2),\dots,\sigma(j+k))$, and $(\tau(1),\tau(2),\dots,\tau(k+i))$ from $(1,2,\dots,l)$, $(1,2,\dots,m)$, and $(1,2,\dots,n)$, respectively. The symbols θ_2, θ_3, and θ_4 are defined by

$$\theta_2 = \prod_{p=1}^{i} \delta(t''_{\tau(k+p)},t_{\rho(p)})\zeta^*_{\tau(k+p)}(t''_{\tau(k+p)})\xi_{\rho(p)}(t_{\rho(p)})\,,$$

$$\theta_3 = \prod_{q=1}^{j} \delta(t_{\rho(i+q)},t'_{\sigma(q)})\xi_{\rho(i+q)}(t_{\rho(i+q)})\eta^*_{\sigma(q)}(t'_{\sigma(q)})\,,$$

$$\theta_4 = \prod_{r=1}^{k} \delta(t'_{\sigma(j+r)},t''_{\tau(r)})\eta_{\sigma(j+r)}(t'_{\sigma(j+r)})\zeta^*_{\tau(r)}(t''_{\tau(r)})\,, \tag{3.21}$$

and the argument of $g_{l+m+n-2(i+j+k)}$ is explicitly written as

$$(\xi_1\cdots[\xi_\rho]\cdots\xi_l\eta_1\cdots[\eta_\sigma]\cdots\eta_m\zeta_1\cdots[\zeta_\tau]\cdots\zeta_n;t_1\cdots[t_\rho]\cdots t_l t'_1\cdots[t'_\sigma]\cdots t'_m t''_1\cdots[t''_\tau]\cdots t''_n)\,,$$

where $[\xi_\rho]$, $[\eta_\sigma]$, $[\zeta_\tau]$, $[t_\rho]$, $[t'_\sigma]$, and $[t''_\tau]$ exclude $\xi_{\rho(1)},\dots,\xi_{\rho(i+j)}$, $\eta_{\sigma(1)},\dots,\eta_{\sigma(j+k)}$, $\zeta_{\tau(1)},\dots,\zeta_{\tau(k+i)}$, $t_{\rho(1)},\dots,t_{\rho(i+j)}$, $t'_{\sigma(1)},\dots,t'_{\sigma(j+k)}$, and $t''_{\tau(1)},\dots,t''_{\tau(k+i)}$, respectively.

The derivation of (3.20) is as follows. First, use (3.4) with $I=3$ and $\underline{\alpha}_1=\underline{\alpha}$, $\underline{\alpha}_2=\underline{\beta}$, and $\underline{\alpha}_3=\underline{\gamma}$. Then,

$$\begin{aligned}\langle\underline{\alpha}|g_lg_mg_m|\underline{\alpha}\rangle &= \Lambda(\underline{\alpha},\underline{\beta})\Lambda(\underline{\alpha},\underline{\gamma})\Lambda(\underline{\beta},\underline{\gamma})\langle\underline{\alpha}|g_l|\underline{\alpha}\rangle\langle\underline{\beta}|g_m|\underline{\beta}\rangle\langle\underline{\gamma}|g_n|\underline{\gamma}\rangle\Big|_{\underline{\gamma}=\underline{\beta}=\underline{\alpha},\text{c.c.}}\\ &= \sum_{i,j,k}(i!j!k!)^{-1}\sum_{s_1\cdots s_p s'_1\cdots s'_j,s''_1\cdots s''_k}\prod_{p=1}^{i}\lambda(\underline{\alpha},\underline{\gamma};s_p)\prod_{q=1}^{j}\lambda(\underline{\alpha},\underline{\beta};s'_q)\prod_{r=1}^{k}\lambda(\underline{\beta},\underline{\gamma};s''_r)\Pi_1\Pi_2\Pi_3\Big|_{\underline{\gamma}=\underline{\beta}=\underline{\alpha},\text{c.c.}}\\ &= \sum_{i,j,k}(i!j!k!)^{-1}\,\Sigma^{(4)}\theta_2\theta_3\theta_4\Pi_6\Pi_7\Pi_8\Big|_{\underline{\gamma}=\underline{\beta}=\underline{\alpha},\text{c.c.}}\,.\end{aligned}$$

Here Π_1,Π_2 are given by (3.15) and (3.16), and Π_3 by

$$\Pi_3 = \prod_{z=1}^{n} Q(\zeta_z,\underline{\gamma};t''_z)\,. \tag{3.22}$$

The factors Π_6, Π_7, and Π_8 are given by omitting $x=\rho(1),\dots,\rho(i+j)$, $y=\sigma(1),\dots,\sigma(j+k)$, and $z=\tau(1),\dots,\tau(k+i)$ in (3.15), (3.16), and (3.22), respectively. Again the operation of $\mathscr{T}$ and the use of relations

$$\Pi_6\Pi_7\Pi_8\Big|_{\underline{\gamma}=\underline{\beta}=\underline{\alpha},\text{c.c.}} = \langle\underline{\alpha}|g_{l+m+n-2(i+j+k)}|\underline{\alpha}\rangle$$

and (3.3) yield (3.20).

IV. ORTHOGONAL POLYNOMIAL FUNCTIONALS

Let $(a(t),a^\dagger(t))$ $(t\in R^d)$ be boson field operators satisfying

$$[a(s),a^\dagger(t)] = \delta(t-s)\,,$$
$$[a(s),a(t)] = [a^\dagger(s),a^\dagger(t)] = 0,\quad s,t\in R^d\,. \tag{4.1}$$

Let $\xi_i(t)$, $\eta_i(t)$, $\zeta_i(t) \in L_2(R^d)$ $(i=1,2,\dots)$ be the space of square integrable complex functions on R^d. The purpose of this section is to derive the multiplication formulas for orthogonal functionals defined by

$$\vec{G}_n(\zeta_1\zeta_2\cdots\zeta_n) = \int_{R^d}dt_1\cdots\int_{R^d}dt_n\, g_n(\zeta_1\zeta_2\cdots\zeta_n;t_1t_2\cdots t_n)\,, \tag{4.2}$$

and their linear combination G_n [see Eq. (4.12)]. In (4.2), g_n is defined by

$$g_n(\zeta_1\zeta_2\cdots\zeta_n;t_1t_2\cdots t_n) = \begin{cases}1 & (n=0)\,,\\ :Q(\zeta_1,a;t_1)Q(\zeta_2,a;t_2)\cdots Q(\zeta_n,a;t_n): & (n\geq 1)\,,\end{cases} \tag{4.3}$$

with

$$Q(\zeta,a;t) = \zeta(t)a(t) + \zeta^*(t)a^\dagger(t)\,, \tag{4.4}$$

analogous to (3.7) and (3.8). The discussion in Sec. III is formally extended to the present case; as is usually done, we regard, in (3.8), $\underline{\zeta}_1,\underline{\zeta}_2,\dots,\underline{\zeta}_n$ as the infinite-dimensional vectors $\zeta_1,\zeta_2,\dots,\zeta_n$ and $t_1,t_2,\dots,t_n$ as continuous variables. Then we obtain the multiplication formulas (3.9) and (3.20) with the replacement of Kronecker's delta by the delta function. A multiplication formula for $\vec{G}_l\vec{G}_m$ now reads

$$\vec{G}_l\vec{G}_m = \sum_{i=1}^{l\wedge m}\sum^{(1)}(i!)^{-1}\prod_{p=1}^{i}\langle\xi_{\rho(p)},\eta_{\sigma(p)}\rangle\vec{G}_{l+m-2i}\,, \tag{4.5}$$

with

$$\vec{G}_l = \vec{G}_l(\xi_1\xi_2\cdots\xi_l)\,, \tag{4.6}$$

$$\vec{G}_m = \vec{G}_m(\eta_1\eta_2\cdots\eta_m)\,, \tag{4.7}$$

$$\vec{G}_{l+m-2i} = \vec{G}_{l+m-2i}(\xi_1\cdots[\xi_\rho]\cdots\xi_l\eta_1\cdots[\eta_\sigma]\cdots\eta_m)\,. \tag{4.8}$$

Here $(\rho(1),\rho(2),...,\rho(i))$ and $(\sigma(1),\sigma(2),...,\sigma(i))$ are taken from $(1,2,...,l)$ and $(1,2,...,m)$, respectively, and the summation $\Sigma^{(1)}$ is carried out over all such $l!m!/\{(l-i)!(m-i)!\}$ combinations. The terms $[\xi_\rho]$ and $[\eta_\sigma]$ indicate the exclusion of $\xi_{\rho(1)}, \xi_{\rho(2)},...,\xi_{\rho(i)}$ and $\eta_{\sigma(1)},\eta_{\sigma(2)},...,\eta_{\sigma(i)}$, respectively. The inner product $\langle\cdot,\cdot\rangle$ in $L_2(R^d)$ is defined by

$$\langle\xi_i,\eta_j\rangle = \int_{R^d}\xi_i(t)\eta_j^*(t)\,dt. \tag{4.9}$$

Since $\vec{G}_n(\zeta_1\zeta_2\cdots\zeta_n)$ is invariant under the permutation of $\{1,2,...,n\}$, it is naturally extended to ζ, where it has the form

$$\zeta = \sum_{w=1}^{W}\gamma_w\mathscr{S}\{\zeta_1^{(w)}\cdots\zeta_n^{(w)}\}. \tag{4.10}$$

Here $\mathscr{S}$ is the symmetrizing operator defined by

$$\mathscr{S}\{\zeta_1\cdots\zeta_n\} = (n!)^{-1}\sum_\pi \zeta_{\pi(1)}\otimes\cdots\otimes\zeta_{\pi(n)} \quad (n=1,2,...), \tag{4.11}$$

with π running over all permutations of $\{1,2,...,n\}$. We define $G_n(\zeta)$ by

$$G_n(\zeta) = \sum_{w=1}^{W}\gamma_w\vec{G}_n(\zeta_1^{(w)}\cdots\zeta_n^{(w)}) \tag{4.12}$$

for ζ of (4.10), and similarly $G_l(\xi)$, $G_m(\eta)$ for ξ, η given by

$$\xi = \sum_{u=1}^{U}\alpha_u\mathscr{S}\{\xi_1^{(u)}\cdots\xi_l^{(u)}\}, \tag{4.13}$$

$$\eta = \sum_{v=1}^{V}\beta_v\mathscr{S}\{\eta_1^{(v)}\cdots\eta_m^{(v)}\}. \tag{4.14}$$

A multiplication formula for $G_l(\xi)G_m(\eta)$ is now written in a form similar to (2.20),

$$G_l(\xi)G_m(\eta) = \sum_{i=1}^{l\wedge m}A(l,m;i)G_{l+m-2i}([\xi,\eta;i]), \tag{4.15}$$

where $[\xi,\eta;i]$ is defined as

$$[\xi,\eta;i] = \frac{1}{l!}\sum_\lambda\frac{1}{m!}\sum_\mu\sum_{u=1}^{U}\sum_{v=1}^{V}\alpha_u\beta_v\prod_{p=1}^{i}\langle\xi_{\lambda(p)}^{(u)},\eta_{\mu(p)}^{(v)}\rangle \times\mathscr{S}\{\xi_{\lambda(i+1)}^{(u)}\cdots\xi_{\lambda(l)}^{(u)}\eta_{\mu(i+1)}^{(v)}\cdots\eta_{\mu(m)}^{(v)}\}. \tag{4.16}$$

In (4.16), λ and μ run over all permutations of $\{1,2,...,l\}$ and $\{1,2,...,m\}$, respectively, and the coefficient $A(l,m;i)$ in (4.15) is defined in (2.11). Nakazawa[11] writes (4.15) in a more compact form,

$$G_l(\xi)G_m(\eta) = \sum_{i=1}^{l\wedge m}G_{l+m-2i}(\xi{:}i{:}\eta), \tag{4.17}$$

by introducing the i contraction $\xi{:}i{:}\eta$[14] given by

$$\xi{:}i{:}\eta = A(l,m;i)[\xi,\eta;i]. \tag{4.18}$$

The formula (4.15) is derived as follows: we operate $(l!)^{-1}\Sigma_\lambda(m!)^{-1}\Sigma_\mu$ to (4.5) and rearrange it using the invariance of $\vec{G}_l$, $\vec{G}_m$, $\vec{G}_{l+m-2i}$ under the permutations of their arguments. This rearranged formula and the definition of G_l, G_m, G_{l+m-2i}, and $[\xi,\eta;i]$ immediately yield (4.15).

A similar argument can be used to derive a multiplication formula for the three functionals $G_l(\xi)$, $G_m(\eta)$, and $G_n(\zeta)$ as

$$G_l(\xi)G_m(\eta)G_n(\zeta) = \sum_{i,j,k}B(l,m,n;i,j,k)G_{l+m+n-2(i+j+k)}([\xi\eta\zeta;ijk]). \tag{4.19}$$

Here the summation is over the range (2.15) and

$$[\xi\eta\zeta;ijk] = \frac{1}{l!}\sum_\lambda\frac{1}{m!}\sum_\mu\frac{1}{n!}\sum_\nu\sum_{u=1}^{U}\sum_{v=1}^{V}\sum_{w=1}^{W}\alpha_u\beta_v\gamma_w\Theta_2\Theta_3\Theta_4 \times\mathscr{S}\{\xi_{\lambda(i+j+1)}^{(u)}\cdots\xi_{\lambda(l)}^{(u)}\eta_{\mu(j+k+1)}^{(v)} \times\cdots\eta_{\mu(m)}^{(v)}\zeta_{\nu(k+i+1)}^{(w)}\cdots\zeta_{\nu(n)}^{(w)}\}, \tag{4.20}$$

with

$$\Theta_2 = \prod_{p=1}^{i}\langle\xi_{\lambda(p)}^{(u)},\zeta_{\nu(k+p)}^{(w)}\rangle, \tag{4.21}$$

$$\Theta_3 = \prod_{q=1}^{j}\langle\xi_{\lambda(i+q)}^{(u)},\eta_{\mu(q)}^{(v)}\rangle, \tag{4.22}$$

$$\Theta_4 = \prod_{r=1}^{k}\langle\eta_{\mu(j+r)}^{(v)},\zeta_{\nu(r)}^{(w)}\rangle. \tag{4.23}$$

The summations in (4.20) w.r.t. λ,μ,ν are carried out over all permutations of $\{1,2,...,l\}$, $\{1,2,...,m\}$, $\{1,2,...,n\}$, respectively.

V. REMARKS

The identity of the structure of expectation values of Wiener–Hermite functionals and Wick products of boson Fock fields was realized early by Imamura *et al.*[5] Any (vectorial or tensorial) Gaussian white noise on $t\in R^d$ may be constructed linearly on a scalar Gaussian white noise $f(s)$, $s\in R$, with t also depending linearly on s. The Wiener–Itô decomposition[2,3] of square integrable random variables associated with $f(s)$ is known[4,7,8,15] to give a representation of Fock structures of free fields on s, with annihilation and creation operators possibly defined by[4,8,15]

$$a(s) = \frac{f(s)}{2}+\frac{\delta}{\delta f(s)}, \quad a^\dagger(s) = \frac{f(s)}{2}-\frac{\delta}{\delta f(s)}, \tag{5.1}$$

together with vacuum expectation values realized as expectations w.r.t. the probability measure induced by f. These facts imply that Wiener–Hermite functionals or multiple Wiener integrals are special facets of orthogonal polynomials of commuting operators for real $\zeta_i^{(w)}$'s in (4.10), yet the whole structure is embraced in the original multiple Wiener integrals of Itô[2] and Wiener.[3]

Let us introduce a suggestive notation for (4.12), valid for real ζ of (4.10),

$$G_n(\zeta) = \int_{R^d}dt_1\cdots\int_{R^d}dt_n \times\zeta(t_1,...,t_n)H_n(t_1,...,t_n), \tag{5.2}$$

$$H_n(t_1,...,t_n) = g_n(1,1,...,1;t_1,...,t_n). \tag{5.3}$$

We also use an abbreviated notation for (5.2) and (5.3),

$$G_n(\zeta) = \zeta(1,2,\ldots,n)H_n(1,2,\ldots,n)\,, \tag{5.4}$$

where the integration over R^d is implied by repeated indices $1^{(\prime)},2^{(\prime)},\ldots,n^{(\prime)}$ that stand for $t_1^{(\prime)},t_2^{(\prime)},\ldots,t_n^{(\prime)}$. The formulas (4.15) and (4.19) for real ξ, η, and ζ (or, for commuting operators) are now written simply as follows:

$$\xi(1,\ldots,l)H_l(1,\ldots,l)\times\eta(1',\ldots,m')H_m(1',\ldots,m')$$
$$= \sum_{i=0}^{l\wedge m} A(l,m;i)\mathscr{S}[\xi(1,\ldots,(l-i),1^{\#},\ldots,i^{\#})\eta(1',\ldots,(m-i)',1^{\#},\ldots,i^{\#})]H_{l+m-2i}(1,\ldots,(l-i),1',\ldots,(m-i)')\,, \tag{5.5}$$

$$\xi(1,\ldots,l)H_l(1,\ldots,l)\times\eta(1',\ldots,m')H_m(1',\ldots,m')\times\zeta(1'',\ldots,n'')H_n(1'',\ldots,n'')$$
$$= \sum_{i,j,k} B(l,m,n;i,j,k)\mathscr{S}[\xi(1,\ldots,(l-i-j),1^{\#},\ldots,i^{\#},1^{*},\ldots,j^{*})\eta(1',\ldots,(m-j-k)',1^{*},\ldots,j^{*},1^{b},\ldots,k^{b})$$
$$\times\zeta(1'',\ldots,(n-k-i)'',1^{b},\ldots,k^{b},1^{\#},\ldots,i^{\#})]$$
$$\times H_{l+m+n-2(i+j+k)}(1,\ldots,(l-i-j),1',\ldots,(m-j-k)',1'',\ldots,(n-k-i)'')\,. \tag{5.6}$$

Here $\mathscr{S}$ indicates the symmetrization of the unrepeated arguments, and the summation in (5.6) ranges over (2.15).

We now consider a non-Gaussian random variable with real K_1 and K_3:

$$X = K_1(1)H_1(1) + K_3(123)H_3(123)\,. \tag{5.7}$$

We calculate the *kernels*, L_1, L_3 of X^3, which are of significance in stochastic problems,

$$X^3 = L_1(1)H_1(1) + L_3(123)H_3(123) + \text{higher order terms.} \tag{5.8}$$

Repeated applications of (5.5) will give the desired result,[16] but the procedure is extremely troublesome. It is far more practical to use our (5.6), the result being

$$\begin{aligned}L_1(1) &= 3(Y+6Z)K_1(1)\\ &\quad+18K_3(1pq)K_1(p)K_1(q)\\ &\quad+108K_3(1pq)K_3(pqr)K_1(r)\\ &\quad+324K_3(1pq)K_3(prs)K_3(qrs)\,,\end{aligned} \tag{5.9}$$

$$\begin{aligned}L_3(123) &= K_1(1)K_1(2)K_1(3) + 3(Y+6Z)K_3(123)\\ &\quad+18\mathscr{S}\{K_1(1)K_3(23p)K_1(p)\}\\ &\quad+54\mathscr{S}\{K_1(1)K_3(2pq)K_3(3pq)\}\\ &\quad+108\mathscr{S}\{K_3(12p)K_3(3pq)K_1(q)\}\\ &\quad+216K_3(1pq)K_3(2qr)K_3(3pr)\\ &\quad+324\mathscr{S}\{K_3(12p)K_3(3qr)K_3(pqr)\}\,,\end{aligned} \tag{5.10}$$

$$Y = K_1(1)K_1(1),\quad Z = K_3(123)K_3(123)\,. \tag{5.11}$$

The rhs of (5.10) clarifies that the factor 126 in the expression of $L_3(123)$ in Ref. 16 should correctly be read as 216.

ACKNOWLEDGMENT

I heartily thank Dr. Hiroshi Nakazawa for illuminating discussions on this work.

[1] R. H. Cameron and W. T. Martin, Ann. Math. **48**, 385 (1947).
[2] K. Itô, J. Math. Soc. Jpn. **3**, 157 (1951).
[3] N. Wiener, *Nonlinear Problems in Random Theory* (M.I.T., Cambridge, MA; Wiley, New York, 1958).
[4] T. Hida, *Brownian Motion* (Springer, Berlin, 1980), Chaps. 4 and 5.
[5] T. Imamura, W. C. Meecham, and A. Siegel, J. Math. Phys. **6**, 695 (1965).
[6] T. Imamura, *Mathematics of Random Fields* (Iwanami, Tokyo, 1976) (in Japanese).
[7] B. Simon, *The $P(\phi)_2$ Euclidean Quantum Field Theory* (Princeton U.P., Princeton, NJ, 1974), Chap. I.
[8] T. Hida, *Quantum Fields—Algebras, Processes*, edited by L. Streit, Proceedings of the Symposium "Bielefeld Encounter in Physics and Mathematics II" (Springer, Berlin, 1980).
[9] A. Jaffe, Ann. Phys. (NY) **32**, 127 (1965), Appendix A.
[10] I. E. Segal, J. Funct. Anal. **6**, 29 (1970), Corollary 2.1.
[11] H. Nakazawa, J. Math. Anal. Appl. **III**, 353 (1985).
[12] G. S. Agarwal and E. Wolf, Phys. Rev. D **2**, 2161 (1970).
[13] R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1966).
[14] For $\Sigma_{j=1}^{r}$ on the right-hand side of (3.3) of Ref. 11 read $\Pi_{j=1}^{r}$.
[15] I. Kubo and S. Takenaka, Proc. Jpn. Acad. Ser. A **56**, 411 (1980).
[16] H. Nakazawa, Prog. Theor. Phys. **56**, 1411 (1976).

The covariant effective action in QED. One-loop magnetic moment

A. A. Ostrovsky
State Committee of Standards, Leninskii Prospect 9, Moscow 117049, Union of Soviet Socialist Republics
G. A. Vilkovisky
Lebedev Physical Institute, Academy of Sciences, Leninskii Prospect 53, Moscow 117924, Union of Soviet Socialist Republics

(Received 28 July 1987; accepted for publication 30 September 1987)

The paper starts the program of rewriting quantum electrodynamics in terms of the manifestly covariant and covariantly computed effective action. A general method for obtaining nonlocal terms in the effective action is proposed and the term responsible for the one-loop magnetic moment of the electron is worked out. In contrast to the usual calculation based on Green's functions, the present calculation nowhere encounters the infrared divergences (including renormalization and restriction to the physical mass shell). Comparison with the method of Green's functions shows the inadequacy of the latter.

I. INTRODUCTION

All effects of a given quantum field theory are contained in its effective action and there should be a straightforward way to compute relevant terms in the effective action without recourse to the standard technique of Green's functions, perturbation theory, etc. This is especially important for gauge theories because in this case the Green's functions are inadequate objects whereas the effective action is a manifestly covariant functional and there should be manifestly covariant methods for its computation. In addition, the effective action technique, when sufficiently developed, should save much computational work because it deals only with diagrams without external lines, and this reduces considerably the number of diagrams contributing to a given effect.

In the present paper we start the program of rewriting quantum electrodynamics in terms of the manifestly covariant (and covariantly computed) effective action. The effect of QED that seems most attractive from this point of view is the anomalous magnetic moment of the electron. This is because, first, the magnetic moment has a clear interpretation in terms of the effective action and, second, already the present-day experimental data require its computation with four-loop accuracy.[1–4] If the effective action techniques have any computational advantages, it is here that they must prove their worth.

One more reason why the problem of the anomalous magnetic moment looks tempting is that *apparently*, for this problem, one needs only the *local* term in the effective Lagrangian, of the form

$$\bar{\psi}\gamma^{\mu}\gamma^{\nu}F_{\mu\nu}\psi\,, \tag{1.1}$$

and, therefore, one may hope that the elaborate Schwinger–DeWitt technique[5,6] will be applicable. This simplicity is, however, illusory and, on closer examination, the problem (even to lowest order) requires a qualitative improvement of the existing covariant methods for the calculation of the effective action. Indeed, if using the Schwinger–DeWitt technique one expands the one-loop effective action of QED in inverse powers of the electron mass, the coefficient of the term (1.1) will prove to be exactly zero. In fact, other terms of this local expansion, containing derivatives of ψ,

$$(1/m)\bar{\psi}\gamma^{\mu}\gamma^{\nu}F_{\mu\nu}\nabla\!\!\!/\psi,\quad (1/m^2)\bar{\psi}\gamma^{\mu}\gamma^{\nu}F_{\mu\nu}\nabla\!\!\!/^2\psi,\ \dots \tag{1.2}$$

all contribute and, upon using the mass-shell equation for ψ, all take the form (1.1). To be more correct, for the electron magnetic moment we need the term in the effective action, which is quadratic in ψ and linear in $F_{\mu\nu}$, and while $F_{\mu\nu}$ may be regarded as constant, ψ is arbitrary. This term [which is a sum of all terms (1.2)] is nonlocal. However, when the corresponding effective equation for ψ is solved iteratively, by expanding in powers of the fine-structure constant, then at each iteration order the quadratic in the ψ term of the renormalized effective action takes the *local* form

$$W_R^{\bar{\psi}\psi} = -\int dx\,\bar{\psi}\Big(\nabla\!\!\!/ + m + i\,\frac{g-2}{2}\,\frac{e}{4m}\,\gamma^{\mu}\gamma^{\nu}F_{\mu\nu}\Big)\psi + O(\partial F) + O(F^2) + O(e^{2n})\,, \tag{1.3}$$

where g is the gyromagnetic ratio, $O(\partial F)$ denotes terms with derivatives of $F_{\mu\nu}$, $O(F^2)$ terms higher order in $F_{\mu\nu}$, and $O(e^{2n})$ terms of nth and higher orders in the fine-structure constant.

Thus, at least in the framework of the usual effective action (as distinct from the *unique* effective action,[7] see the discussion in the concluding section), the anomalous magnetic moment is a nonlocal effect. In the present paper we propose a covariant method for computing nonlocal terms in the effective action and work out the relevant term in the one-loop effective action of QED [Eq. (5.8)]. At the heart of the method lies a Gaussian integral with noncommuting sources. Its calculation is discussed in the Appendix. (Another approach to the effective action in quantum electrodynamics, which also leads to integrals with noncommuting parameters, is proposed in Ref. 8.)

In the language of the Green's functions, our final result corresponds to the calculation of the one-loop mass operator and the three-point vertex with the electron lines off shell. However, the reward for covariance is that we nowhere (including the renormalization and restriction to the physical mass shell) encounter the infrared divergences and nowhere need to introduce the photon mass. The comparison with the textbook calculation, carried out in the concluding section, shows clearly the inadequacy of the Green's function method.

II. DIAGONALIZATION OF THE QUANTUM LAGRANGIAN

We start with the QED Lagrangian in a covariant (mean field[9]) gauge

$$\mathscr{L} = -\tfrac{1}{4}\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu} - \bar{\boldsymbol{\psi}}[\gamma^\mu(\partial_\mu - ie\mathbf{A}_\mu) + mI]\boldsymbol{\psi} - \tfrac{1}{2}(\partial_\mu \mathbf{A}^\mu - \partial_\mu\langle \mathbf{A}^\mu\rangle)^2, \quad (2.1)$$

where

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu, \quad (2.2)$$

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}I, \quad \operatorname{tr} g^{\mu\nu} = 2\omega, \quad (2.3)$$

and $g^{\mu\nu}$ is the positive-definite metric of flat 2ω-dimensional space. [In QED the use of mean-field gauges is not crucial. The loop part of the effective action for (2.1) coincides with that for the usual Lorentz gauge, which is covariant by itself. The addition containing $\langle \mathbf{A}\rangle$ affects only the tree term of the effective action making it covariant.] The parameter ω will be used to regularize ultraviolet divergences in proper-time integrals.

Next we introduce the mean fields

$$A_\mu = \langle \mathbf{A}_\mu\rangle, \quad \psi = \langle \boldsymbol{\psi}\rangle, \quad \bar\psi = \langle \bar{\boldsymbol{\psi}}\rangle, \quad (2.4)$$

and define

$$a_\mu = \mathbf{A}_\mu - A_\mu, \quad \eta = \boldsymbol{\psi} - \psi, \quad \bar\eta = \bar{\boldsymbol{\psi}} - \bar\psi. \quad (2.5)$$

The $a_\mu, \eta, \bar\eta$ will be regarded as independent integration variables in the functional integral. By expanding the Lagrangian in powers of the quantum fields we obtain (the term $\mathscr{L}_1$, linear in the quantum fields, always cancels in the equation for the effective action, see, e.g., Ref. 7)

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1 + \mathscr{L}_2 + \mathscr{L}_3, \quad (2.6)$$

$$\mathscr{L}_0 = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \bar\psi[\gamma^\mu(\partial_\mu - ieA_\mu) + mI]\psi, \quad (2.7)$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \quad (2.8)$$

$$\mathscr{L}_2 = \tfrac{1}{2}a_\mu(g^{\mu\nu}\partial^2)a_\nu - \bar\eta[\gamma^\mu(\partial_\mu - ieA_\mu) + mI]\eta + iea_\mu\bar\eta\gamma^\mu\psi + iea_\mu\bar\psi\gamma^\mu\eta, \quad (2.9)$$

$$\mathscr{L}_3 = iea_\mu\bar\eta\gamma^\mu\eta. \quad (2.10)$$

Let us introduce the notation for the Dirac operator in an external field and its (Euclidean) Green's function,

$$\mathscr{D} = \gamma^\mu(\partial_\mu - ieA_\mu) + mI, \quad (2.11)$$

$$\mathscr{D}\mathscr{D}^{-1} = \mathscr{D}^{-1}\mathscr{D} = I, \quad (2.12)$$

$$\mathscr{D}^{-1}J(x) = -\int G(x,y)J(y)dy, \quad (2.13)$$

where all operators are understood as acting to the right on a spinor.

A convenient diagrammatic technique for the effective action in QED emerges if we diagonalize the Lagrangian of quantum fields by making the shift

$$\eta(x) = \xi(x) - ie\int G(x,y)\gamma^\mu\psi(y)a_\mu(y)dy, \quad (2.14a)$$

$$\bar\eta(y) = \bar\xi(y) - ie\int a_\mu(x)\bar\psi(x)\gamma^\mu G(x,y)dx, \quad (2.14b)$$

where now ξ and $\bar\xi$ will be the independent variables in the functional integral. The Jacobian of the replacement (2.14) equals unity. In terms of the new set of quantum fields $a_\mu, \xi, \bar\xi$ we obtain

$$\mathscr{L}_2 = \tfrac{1}{2}a_\mu Q^{\mu\nu}a_\nu - \bar\xi\mathscr{D}\xi, \quad (2.15)$$

$$\mathscr{L}_3 = iea_\mu\bar\xi\gamma^\mu\xi - e^2a_\mu\bar\xi\gamma^\mu\mathscr{D}^{-1}\gamma^\nu\psi a_\nu - e^2a_\mu\bar\psi\gamma^\mu\mathscr{D}^{-1}\gamma^\nu\xi a_\nu - ie^3a_\mu\bar\psi\gamma^\mu\mathscr{D}^{-1}a_\alpha\gamma^\alpha\mathscr{D}^{-1}\gamma^\nu\psi a_\nu, \quad (2.16)$$

where $Q^{\mu\nu}$ is the following vector-field operator:

$$Q^{\mu\nu}\delta(x,y) = g^{\mu\nu}\partial^2\delta(x,y) + e^2\bar\psi(x)\gamma^\mu G(x,y)\gamma^\nu\psi(y) + e^2\bar\psi(y)\gamma^\nu G(y,x)\gamma^\mu\psi(x). \quad (2.17)$$

Its Green's function plays the role of the photon propagator in the resultant diagrammatic technique.

For the one-loop effective action we need only $\mathscr{L}_2$:

$$W_{\text{one-loop}}[A,\psi,\bar\psi] = \operatorname{Tr}\ln(\mathscr{D}\delta(x,y)) - \tfrac{1}{2}\operatorname{Tr}\ln(Q^{\mu\nu}\delta(x,y)). \quad (2.18)$$

The mean fields ψ, $\bar\psi$ enter $W_{\text{one-loop}}$ through the operator $Q^{\mu\nu}$ given by (2.17). By expanding (2.18) in powers of $\psi,\bar\psi$ we obtain

$$W_{\text{one-loop}}[A,\psi,\bar\psi] = \operatorname{Tr}\ln(\mathscr{D}\delta(x,y)) - \tfrac{1}{2}\operatorname{Tr}\ln(g^{\mu\nu}\partial^2\delta(x,y)) - \tfrac{1}{2}\operatorname{Tr}(M^\mu{}_\nu(x,y)) + O((\bar\psi\psi)^2), \quad (2.19)$$

where

$$M^\mu{}_\nu(x,y) = -e^2\bar\psi(x)\gamma^\mu G(x,y)\gamma^\lambda\psi(y)(-g_{\lambda\nu}/\partial_y^2) - (-g_{\nu\lambda}/\partial_y^2)e^2\bar\psi(y)\gamma^\lambda G(y,x)\gamma^\mu\psi(x). \quad (2.20)$$

Thus the term quadratic in ψ is

$$W^{\bar\psi\psi}_{\text{one-loop}} = -\tfrac{1}{2}\operatorname{Tr}(M^\mu{}_\nu(x,y)). \quad (2.21)$$

This expression is graphically shown in Fig. 1.

III. THE PROPAGATORS

If the effective action is to be expanded in powers of ψ, $\bar\psi$, then the photon propagator in the diagrammatic technique will become the free Green's function

$$-\frac{g_{\mu\nu}}{\partial^2}\delta(x,y) = \int_0^\infty \frac{d\tau}{(4\pi\tau)^\omega}\, e^{-\sigma(x,y)/2\tau}g_{\mu\nu}, \quad (3.1)$$

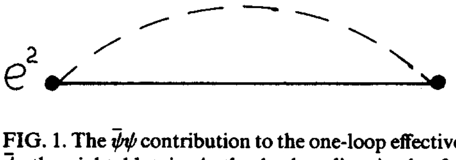


FIG. 1. The $\bar\psi\psi$ contribution to the one-loop effective action. The left blot is $\bar\psi$, the right blot is ψ, the broken line is the free photon propagator $(-g_{\mu\nu}/\partial^2)\delta(x,y)$, the full line is the electron propagator $G(x,y)$ in an external (mean) electromagnetic field, and each vertex is γ^μ.

where $\sigma(x,y)$ is the world function,[5]

$$\sigma = \tfrac{1}{2}\sigma_\mu \sigma^\mu, \quad \sigma_\mu \equiv \partial_\mu \sigma(x,y) . \tag{3.2}$$

For our purposes, the mean electromagnetic field $F_{\mu\nu}$, unlike ψ, may be regarded as constant. This means that the effective action will be expanded in powers of $F_{\mu\nu}$ *and its derivatives*. Therefore, for the calculation of the electron propagator in an external field $G(x,y)$, one may apply the Schwinger–DeWitt technique.[5,6] Bearing in mind renormalization, we shall work with accuracy $O(F^3)$ although in the present paper all quantities will be needed only with accuracy $O(F^2) + O(\partial F)$.

Thus we introduce the covariant derivative ∇_μ which acts on a spinor as

$$\nabla_\mu \psi = (\partial_\mu - ieA_\mu)\psi \tag{3.3a}$$

and on a conjugate spinor as

$$\nabla_\mu \bar\psi = (\partial_\mu + ieA_\mu)\bar\psi . \tag{3.3b}$$

Matrices in the space of spinors will be denoted by letters with a hat. From (3.3a) and (3.3b), the action of the covariant derivative on a matrix is

$$\nabla_\mu \hat X = \partial_\mu \hat X . \tag{3.3c}$$

The quantity $\hat{\mathscr{R}}_{\mu\nu}$, which figures in the general formalism[6]

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\psi = \hat{\mathscr{R}}_{\mu\nu}\psi , \tag{3.4}$$

is in the present case

$$\hat{\mathscr{R}}_{\mu\nu} = -ieF_{\mu\nu} I . \tag{3.5}$$

By using the squaring procedure one obtains

$$G(x,y) = -(\gamma^\mu \nabla_\mu - mI)(H - m^2 I)^{-1}\delta(x,y) , \tag{3.6}$$

$$H = g^{\mu\nu}\nabla_\mu \nabla_\nu + \hat P , \tag{3.7}$$

where we have reduced the squared operator to the canonical form[6] with

$$\hat P = \tfrac{1}{2}\gamma^\mu \gamma^\nu \hat{\mathscr{R}}_{\mu\nu} = -(ie/2)F_{\mu\nu}\gamma^\mu\gamma^\nu . \tag{3.8}$$

Next one writes

$$-(H - m^2 I)^{-1} = \int_0^\infty ds\, e^{-sm^2} e^{sH} , \tag{3.9}$$

$$e^{sH}\delta(x,y) = \frac{1}{(4\pi s)^\omega} e^{-\sigma(x,y)/2s} \sum_{n=0}^{\infty} s^n \hat a_n(x,y) , \tag{3.10}$$

where $\hat a_n(x,y)$ are the DeWitt coefficients[5,6] which in the present case behave like spinors at the point x and conjugate spinors at the point y. Thus

$$G(x,y) = \int_0^\infty \frac{ds}{(4\pi s)^\omega} e^{-sm^2} e^{-(1/4s) g_{\mu\nu}\sigma^\mu\sigma^\nu} \times \left[-\frac{\sigma_\nu \gamma^\nu \hat a_0(x,y)}{2s} + \sum_{n=0}^{\infty} s^n \hat b_n(x,y) \right], \tag{3.11}$$

where

$$\hat b_n(x,y) = (\gamma^\mu \nabla_\mu - mI)\hat a_n(x,y) - \tfrac{1}{2}\sigma_\nu \gamma^\nu \hat a_{n+1}(x,y) \tag{3.12}$$

and $\sigma_\nu = \partial^x_\nu \sigma(x,y)$ is a vector at the point x and a scalar at the point y. The two-point functions (3.12) should now be expanded in the covariant Taylor series in powers of σ^ν (Ref. 6) with coefficients at the point x. These coefficients are built of the coincidence limits of $\hat a_n(x,y)$ and their covariant derivatives, which are tabulated in Ref. 6 in a universal form.

Let us represent the Green's function (3.11) in the form

$$G(x,y) = K(x,y)\hat a_0(x,y) \tag{3.13}$$

by factoring out the zeroth-order DeWitt coefficient $\hat a_0(x,y)$. The $\hat a_0(x,y)$ is the parallel displacement propagator along the geodesic,[5,6] whose explicit form is never needed in loop calculations (we shall see below how this comes about, see also Ref. 6). For $K(x,y)$ one may write down the expansion in σ^ν to the given accuracy

$$K(x,y) = \int_0^\infty \frac{ds}{(4\pi s)^\omega} e^{-sm^2} e^{-(1/4s) g_{\mu\nu}\sigma^\mu\sigma^\nu} \times (\hat Z_0(\sigma) + \hat Z_1(\sigma) + \hat Z_2(\sigma)) + O(F^3) + O(\partial F) , \tag{3.14}$$

where $\hat Z_0(\sigma)$, $\hat Z_1(\sigma)$, and $\hat Z_2(\sigma)$ are polynomials in σ^ν of zeroth, first, and second order in $F_{\mu\nu}$, respectively. All that remains is to use the table of universal coincidence limits in Ref. 6 to obtain

$$\hat Z_0(\sigma) = -mI - (1/2s)\gamma^\nu \sigma_\nu , \tag{3.15}$$

$$\hat Z_1(\sigma) = (iem/2)F_{\mu\nu}\gamma^\mu\gamma^\nu s + (ie/2)F_{\mu\nu}(\gamma^\mu \delta^\nu_\lambda + \tfrac{1}{2}\gamma_\lambda \gamma^\mu \gamma^\nu)\sigma^\lambda , \tag{3.16}$$

$$\begin{aligned}\hat Z_2(\sigma) = {}& (e^2 m/4)\left[\tfrac{1}{2}(F_{\mu\nu}\gamma^\mu\gamma^\nu)^2 + \tfrac{1}{3}F_{\mu\nu}F^{\mu\nu} I\right]s^2 \\ & + (e^2/2)\gamma^\beta \left[\tfrac{1}{8}(F_{\mu\nu}\gamma^\mu\gamma^\nu)^2 g_{\alpha\beta} + \tfrac{1}{2}(F_{\mu\nu}\gamma^\mu\gamma^\nu)F_{\beta\alpha} \right. \\ & \left. + \tfrac{1}{12}F_{\mu\nu}F^{\mu\nu} I g_{\beta\alpha} - \tfrac{1}{3}F_{(\beta\nu}F_{\alpha)\cdot}{}^{\nu} I\right] s\sigma^\alpha \\ & + (e^2 m/12)F_{\alpha\mu}F_{\beta\cdot}{}^{\mu} I s\sigma^\alpha\sigma^\beta \\ & + (e^2/24)F_{\alpha\mu}F_{\beta\cdot}{}^{\mu}\gamma_\nu \sigma^\alpha\sigma^\beta\sigma^\nu .\end{aligned} \tag{3.17}$$

IV. CALCULATION OF THE NONLOCAL EFFECTIVE ACTION

From (2.21), (2.20), and (3.1) we have

$$W^{\bar\psi\psi}_{\text{one-loop}} = e^2 \int dx \int_0^\infty \frac{d\tau}{(4\pi\tau)^\omega} \times \int dy\, e^{-\sigma(x,y)/2\tau}\bar\psi(x)\gamma^\nu G(x,y)\gamma_\nu \psi(y) . \tag{4.1}$$

Here $\int dx$ may be regarded as an "external" integral making the effective action of the effective Lagrangian, and $\int dy$ as a loop integral. The general procedure[6] is now to make the replacement of variables in the loop integral

$$y^\mu \to \sigma^\mu = g^{\mu\nu}(x)\partial^x_\nu \sigma(x,y) \tag{4.2}$$

and use the expansion (3.13), (3.15) for the Green's function $G(x,y)$. In the absence of gravity, the Jacobian of the replacement (4.2) is

$$\left| \frac{\partial \sigma^\mu}{\partial y^\nu} \right| = 1 \tag{4.3}$$

and one obtains

$$W^{\bar\psi\psi}_{\text{one-loop}} = e^2 \int dx \iint_0^\infty \frac{ds\, d\tau}{(4\pi s)^\omega (4\pi\tau)^\omega} e^{-sm^2} \times \int\left(\prod_{\beta=1}^{2\omega} d\sigma^\beta\right) e^{-[(s+\tau)/4s\tau] g_{\mu\nu}\sigma^\mu\sigma^\nu} \times \bar\psi(x)\gamma^\alpha[\hat Z_0(\sigma) + \hat Z_1(\sigma) + O(F^2) + O(\partial F)]\gamma_\alpha \hat a_0(x,y)\psi(y) , \tag{4.4}$$

where use has been made of the fact that $\hat a_0(x,y)$ commutes with γ_ν. [In QED, $\hat a_0(x,y)$ is proportional to the unit matrix,

$$\hat a_0(x,y) = I \exp\left(ie \int_C A_\mu\, dx^\mu\right)$$

where the contour C is the geodesic connecting y with x and directed from y to x.]

In the framework of the Schwinger–DeWitt technique one would now expand also $\psi(y)$ in the covariant Taylor series[6]

$$\hat a_0(x,y)\psi(y) = \sum_{n=0}^\infty \frac{(-1)^n}{n!} \sigma^{\mu_1} \cdots \sigma^{\mu_n} \nabla_{\mu_1} \cdots \nabla_{\mu_n} \psi(x) \tag{4.5}$$

[it is here that the $\hat a_0(x,y)$ completes its job and leaves the stage] and in this way reduce the loop integral to the Gaussian integral over σ^μ. However, this is just what we cannot do in the present case because *all* terms of the expansion (4.5) contribute to the electron magnetic moment. Therefore we represent the expansion (4.5) in the form

$$\hat a_0(x,y)\psi(y) = e^{-\sigma^\mu \nabla_\mu} \psi(x) \tag{4.6}$$

and in this form substitute it in (4.4). As a result the loop integral reduces to Gaussian integrals with noncommuting sources (this method, originally used for the present calculation, was generalized in Refs. 10 and 11 where a nonlocal expansion in powers of a universal set of curvatures is obtained for the generic one-loop effective action:)

$$\frac{1}{(4\pi u)^\omega} \int\left(\prod_{\beta=1}^{2\omega} d\sigma^\beta\right) \sigma^{\alpha_1} \cdots \sigma^{\alpha_n} e^{-(1/4u) g_{\mu\nu}\sigma^\mu\sigma^\nu} e^{-\sigma^\alpha \nabla_\alpha} ,$$
$$[\nabla_\alpha, \nabla_\beta] = -ieF_{\alpha\beta} I, \quad u = s\tau/(s+\tau) . \tag{4.7}$$

The calculation of these integrals is discussed in the Appendix. Since the polynomials $\hat Z_0(\sigma)$ and $\hat Z_1(\sigma)$, entering (4.4) and given by the expressions (3.15) and (3.16), are linear in σ^μ, we need only the first two Gaussian moments, for which we obtain

$$\frac{1}{(4\pi u)^\omega} \int\left(\prod_{\beta=1}^{2\omega} d\sigma^\beta\right) e^{-(1/4u) g_{\mu\nu}\sigma^\mu\sigma^\nu} e^{-\sigma^\alpha \nabla_\alpha} = e^{u\nabla_\alpha \nabla^\alpha} + O(F^2) + O(\partial F) , \tag{4.8}$$

$$\frac{1}{(4\pi u)^\omega} \int\left(\prod_{\beta=1}^{2\omega} d\sigma^\beta\right) \sigma^\lambda e^{-(1/4u) g_{\mu\nu}\sigma^\mu\sigma^\nu} e^{-\sigma^\alpha \nabla_\alpha} = -2u(g^{\lambda\nu} + ieuF^{\lambda\nu})\nabla_\nu e^{u\nabla_\alpha\nabla^\alpha} + O(F^2) + O(\partial F) . \tag{4.9}$$

After the explicit expressions (3.15) and (3.16) have been used and the Gaussian integral over σ^μ done with the aid of (4.8) and (4.9), we obtain

$$W^{\bar\psi\psi}_{\text{one-loop}} = e^2 \int dx \iint_0^\infty \frac{ds\, d\tau}{(4\pi)^\omega (s+\tau)^\omega} e^{-sm^2} \bar\psi(x)\gamma^\nu\Big[-mI + \frac{\tau}{s+\tau}\gamma_\mu\left(\nabla^\mu + ie\frac{s\tau}{s+\tau}F^{\mu\beta}\nabla_\beta\right) + \frac{iem}{2}F_{\mu\beta}\gamma^\mu\gamma^\beta s$$
$$- ie\frac{s\tau}{s+\tau}F_{\mu\beta}(\gamma^\mu\delta^\beta_\lambda + \tfrac{1}{2}\gamma_\lambda\gamma^\mu\gamma^\beta)\nabla^\lambda\Big]\gamma_\nu e^{[s\tau/(s+\tau)]\nabla_\alpha\nabla^\alpha}\psi(x) + O(F^2) + O(\partial F) , \tag{4.10}$$

where all operators act in the indicated order to the right on the spinor $\psi(x)$.

V. EXPLICIT FORM OF THE EFFECTIVE ACTION AND THE ANOMALOUS MAGNETIC MOMENT

It is convenient to express the exponentiated operator in (4.10) in terms of the squared Dirac operator (3.7),

$$e^{u\nabla_\alpha\nabla^\alpha} = (1 + u\,\tfrac{1}{2}\, e\sigma F)e^{uH} + O(F^2) + O(\partial F) , \tag{5.1}$$

$$H = \nabla_\alpha\nabla^\alpha - \tfrac{1}{2}\, e\sigma F , \tag{5.2}$$

where

$$\sigma F \equiv \sigma^{\mu\nu}F_{\mu\nu}, \quad \sigma^{\mu\nu} = (i/2)[\gamma^\mu, \gamma^\nu] . \tag{5.3}$$

In the remaining terms of (4.10) one may use the operator identity

$$4i\gamma_\mu F^{\mu\nu}\nabla_\nu = (\sigma F)\gamma^\mu\nabla_\mu - \gamma^\mu\nabla_\mu(\sigma F) + O(\partial F) , \tag{5.4}$$

which guarantees that the covariant derivatives will act on ψ or $\bar\psi$ only in the Dirac combination $\gamma^\mu\nabla_\mu$ (and this is true for higher loop orders as well).

The proper-time integrals over s and τ remain to be computed. For this purpose we make the standard replacement of variables

$$s = \lambda z, \quad \tau = \lambda(1-z), \quad 0 \leqslant \lambda < \infty, \quad 0 \leqslant z \leqslant 1, \quad \left|\frac{\partial(s,\tau)}{\partial(\lambda,z)}\right| = \lambda , \tag{5.5}$$

which brings expression (4.10) to the form

$$W^{\bar\psi\psi}_{\text{one-loop}} = \frac{e^2}{(4\pi)^\omega}\int dx\,\bar\psi(x)\int_0^1 dz\int_0^\infty \frac{d\lambda}{\lambda^{\omega-1}}[2(2-\omega)mz$$
$$-m(1+z)+2(2-\omega)(\gamma^\mu\nabla_\mu+mI)(1-z)-2(\gamma^\mu\nabla_\mu+mI)(1-z)$$
$$-2(2-\omega)m\,\tfrac{1}{2}e(\sigma F)\lambda z^3-2m\,\tfrac{1}{2}e(\sigma F)\lambda z^2(1-z)$$
$$-\{\gamma^\mu\nabla_\mu+mI,\tfrac{1}{2}e\sigma F\}\lambda z(1-z)(2-z)]e^{-\lambda z[m^2-(1-z)H]}\psi(x)+O(F^2)+O(\partial F) \tag{5.6}$$

where { , } denotes the anticommutator.

The integrals over the parameters are elementary if the integration over λ is done first. The only ultraviolet divergent integral is of the form

$$\int_0^\infty \frac{d\lambda}{\lambda^{\omega-1}}e^{-\lambda z[m^2-(1-z)H]} \tag{5.7}$$

and is calculated at $\omega\to 2$ by the usual trick of integration by parts (see, e.g., Ref. 6). Note that, for small $F_{\mu\nu}$, the coefficient of λ in the exponent of (5.7) is negative definite. There is, therefore, no problem with convergence of the proper-time integrals at the upper limit, which in massless theories is inherent in the local Schwinger–DeWitt technique (see the discussion in Refs. 7 and 10).

After the proper-time integrals have been computed, in the limit $\omega\to 2$ we obtain the final explicit expression for the effective action:

$$W^{\bar\psi\psi}_{\text{one-loop}} = \frac{e^2}{(4\pi)^2}\int dx\,\bar\psi(x)\Big[m\Big(-\frac{3}{2-\omega}+3\ln m^2+3\mathbf{C}-4$$
$$-3\ln 4\pi+\frac{m^2-H}{H}+\frac{m^2-H}{H}\Big(2-\frac{m^2-H}{H}\Big)\ln\frac{m^2}{m^2-H}\Big)$$
$$+(\gamma^\mu\nabla_\mu+m)\Big(-\frac{1}{2-\omega}+\ln m^2+\mathbf{C}-2-\ln 4\pi-\frac{m^2-H}{H}+\frac{m^2-H}{H}\Big(2+\frac{m^2-H}{H}\Big)\ln\frac{m^2}{m^2-H}\Big)$$
$$+\frac{e}{2m}\sigma F\Big(-\Big(1+\frac{m^2-H}{H}\Big)\Big(1+2\frac{m^2-H}{H}\Big)+2\frac{m^2-H}{H}\Big(1+\frac{m^2-H}{H}\Big)^2\ln\frac{m^2}{m^2-H}\Big)$$
$$+\Big\{\gamma^\mu\nabla_\mu+m,\frac{e}{2m^2}\sigma F\Big\}\Big(2\Big(1+\frac{m^2-H}{H}\Big)$$
$$-\Big(1+\frac{m^2-H}{H}\Big)^2\Big(2+\frac{m^2-H}{H}\Big)\ln\frac{m^2}{m^2-H}\Big)+O(2-\omega)\Big]\psi(x)+O(F^2)+O(\partial F)\,, \tag{5.8}$$

where $\mathbf{C}$ is the Euler constant. The renormalization in (5.8) boils down to deleting the terms

$$\frac{e^2}{(4\pi)^2}\int dx\,\bar\psi(x)\Big[m\Big(-\frac{3}{2-\omega}+3\ln m^2$$
$$+3\mathbf{C}-4-3\ln 4\pi\Big)$$
$$+(\gamma^\mu\nabla_\mu+m)\Big(-\frac{1}{2-\omega}+\ln m^2$$
$$+\mathbf{C}-2-\ln 4\pi\Big)\Big]\psi(x)=-W^{\bar\psi\psi}_{\text{counter}}\,, \tag{5.9}$$

proportional to the terms of the classical action

$$W^{\bar\psi\psi}_{\text{tree}}=-\int dx\,\bar\psi(x)(\gamma^\mu\nabla_\mu+m)\psi(x) \tag{5.10}$$

[see Eq. (2.7)]. The renormalized effective action is

$$W^{\bar\psi\psi}_R=W^{\bar\psi\psi}_{\text{tree}}+W^{\bar\psi\psi}_{\text{one-loop}}+W^{\bar\psi\psi}_{\text{counter}}+O(e^4)\,. \tag{5.11}$$

If the effective equations for ψ and $\bar\psi$

$$\frac{\delta W^{\bar\psi\psi}_R}{\delta\bar\psi(x)}=(\gamma^\mu\nabla_\mu+m)\psi(x)+O(e^2)=0\,, \tag{5.12a}$$

$$\frac{\delta W^{\bar\psi\psi}_R}{\delta\psi(x)}=\bar\psi(x)(\gamma^\mu\overleftarrow{\nabla}_\mu-m)+O(e^2)=0 \tag{5.12b}$$

are solved by iteration and the solution is inserted back into the action, then, since we have

$$m^2-H=-(\gamma^\mu\nabla_\mu-m)(\gamma^\mu\nabla_\mu+m)\,, \tag{5.13}$$

it follows that all nonlocal terms in (5.8) vanish. Note that all logarithms are suppressed and no infrared divergences arise. Thus we obtain

$$W^{\bar\psi\psi}_R\Big|_{\text{mass shell}}$$
$$=-\int dx\,\bar\psi(x)\Big(\gamma^\mu\nabla_\mu+m+\frac{e^2}{(4\pi)^2}\frac{e}{2m}\sigma F\Big)$$
$$\times\psi(x)+O(F^2)+O(\partial F)+O(e^4)\,, \tag{5.14}$$

with Schwinger's value

$$(g-2)/2=e^2/8\pi^2+O(e^4) \tag{5.15}$$

for the anomalous magnetic moment.

VI. DISCUSSION

The reader has of course noticed that the present calculation is reminiscent of Schwinger's calculation in Ref. 12 (but pushed to its logical extreme) and differs drastically from the usual textbook calculation. An important difference is that neither in computing the effective action nor in conducting the renormalization nor in restricting the result to the physical mass shell did we encounter the infrared divergences. In particular, the spinor-field renormalization constant is

$$(Z_2^{-1}-1) = \frac{e^2}{(4\pi)^2}\left(\frac{1}{2-\omega} - \ln m^2 - \mathbf{C} + 2 + \ln 4\pi\right) + O(e^4) \tag{6.1}$$

from (5.9) and is infrared finite. [The Ward identity $Z_1 = Z_2$ is, of course, trivially satisfied in (5.9) because the counterterm is covariant.]

To carry out the comparison with the method of Green's functions, let us expand the renormalized effective action (5.11) in powers of the electromagnetic potential A_μ and keep only terms linear in A_μ. The result can be written in the form

$$W_R^{\bar\psi\psi} = -\int dx\,\bar\psi[\partial\!\!\!/ + m - ieA\!\!\!/ + \Sigma_R(\partial\!\!\!/) - ieA\Gamma_R(\partial\!\!\!/)]\psi + O(\cdots) \tag{6.2}$$

in terms of the renormalized mass operator $\Sigma_R(\partial\!\!\!/)$ and vertex function $\Gamma_R(\partial\!\!\!/)$ [$A\Gamma_R(\partial\!\!\!/)$ is the notation for the contraction of the vertex function with $A_\mu(x)$]. The $\Sigma_R(\partial\!\!\!/)$ and $\Gamma_R(\partial\!\!\!/)$ can be read off from (5.8)–(5.11).

Expression (5.14) is obtained by using, in the quantum terms, the mass-shell equation (5.12):

$$(\partial\!\!\!/ + m - ieA\!\!\!/)\psi + O(e^2) = 0 \tag{6.3}$$

and similarly for $\bar\psi$. The expansion in A_μ is then equivalent to the expansion of $\Sigma_R(\partial\!\!\!/)$ and $\Gamma_R(\partial\!\!\!/)$ at the point $\partial\!\!\!/ = -m$:

$$\int dx\,\bar\psi(x)[-ieA\Gamma_R(\partial\!\!\!/)]\psi = \int dx\,\bar\psi[-ieA\Gamma_R(-m)]\psi + O(A^2) + O(e^4), \tag{6.4}$$

$$\int dx\,\bar\psi[\Sigma_R(\partial\!\!\!/)]\psi = \int dx\,\bar\psi[\Sigma_R(-m) + \Sigma_R'(-m)ieA\!\!\!/\,]\psi + O(A^2) + O(e^4). \tag{6.5}$$

But they cannot be expanded! If, nevertheless, we *do* expand, the result will be infrared divergent:

$$\Sigma_R(-m) = 0, \tag{6.6}$$

$$\Sigma_R'(-m) = (e^2/(4\pi)^2)(2 - 4\ln(m^2/0)), \tag{6.7}$$

$$-ieA\Gamma_R(-m) = [e^2/(4\pi)^2][(e/2m)\sigma F - (2 - 4\ln(m^2/0))ieA\!\!\!/\,]. \tag{6.8}$$

Fortunately, the contributions of Green's functions have no meaning separately. The noncovariant infrared-divergent pieces add together to yield the finite covariant result

$$\Sigma_R(-m) + \Sigma_R'(-m)ieA\!\!\!/ - ieA\Gamma_R(-m) = [e^2/(4\pi)^2](e/2m)\sigma F. \tag{6.9}$$

However, what is done in textbooks is not even this. For Green's functions, the restriction to the physical mass shell (6.3) cannot be formulated consistently [note the term with Σ_R' in (6.5)], and, instead of (6.3), the free equation

$$(\partial\!\!\!/ + m)\psi = 0 \tag{6.10}$$

is used. With this mass-shell equation the quantum corrections to the Dirac operator take the form

$$\Sigma_R(-m) - ieA\Gamma_R(-m), \tag{6.11}$$

the contribution from Σ_R' is absent, and the infrared divergence of the vertex function remains uncompensated. The textbook procedure is then to redefine the counterterm,

$$(Z_2^{-1}-1)_{\text{textbook}} = (Z_2^{-1}-1) + [e^2/(4\pi)^2](2 - 4\ln(m^2/0)) \tag{6.12}$$

and correspondingly the mass operator, vertex function, and effective action

$$\Sigma_R(\partial\!\!\!/)\big|_{\text{textbook}} = \Sigma_R(\partial\!\!\!/) - (\partial\!\!\!/ + m)\times\frac{e^2}{(4\pi)^2}(2 - 4\ln(m^2/0)) = \Sigma_R(\partial\!\!\!/) - \Sigma_R'(-m)(\partial\!\!\!/ + m), \tag{6.13}$$

$$A\Gamma_R(\partial\!\!\!/)\big|_{\text{textbook}} = A\Gamma_R(\partial\!\!\!/) - A\!\!\!/\,\frac{e^2}{(4\pi)^2}(2 - 4\ln(m^2/0)), \tag{6.14}$$

$$W_R^{\bar\psi\psi}\big|_{\text{textbook}} = W_R^{\bar\psi\psi} + \frac{e^2}{(4\pi)^2}\left(2 - 4\ln\frac{m^2}{0}\right)\int dx\bar\psi(\nabla\!\!\!\!/ + m)\psi; \tag{6.15}$$

thereby making these good quantities explicitly infrared divergent off shell! At this price, the quantum corrections to the Dirac operator, when restricted to (6.10), take the required form. Note that the structure of the coefficient $\ln(m^2/0)$ in (6.15) clearly indicates that a piece of a nonlocal term has been erroneously included in the counter term.

Equation (6.10) may be understood as a leading approximation in expanding the solution of the effective equations in A_μ. To be more correct, the leading approximation is then

$$(\partial\!\!\!/ + m + \Sigma_R(\partial\!\!\!/))\psi = 0, \tag{6.16}$$

but condition (6.6) is satisfied, and one may think that (6.10) is the only solution of (6.16). It is in this case that the quantum corrections are of the form (6.11). However, for perturbation theory, the correct expansion is an expansion in the charge, and this leads to (6.9) not (6.11). Physically, too, one considers either the problem of an electron in a constant external field or the scattering problem. In the former case the electron field never satisfies Eq. (6.10), while in the latter case the electron is free in the *in* and *out* states (if such states exist), but at any rate the external field cannot be regarded as constant in time.

To return to more interesting matters, an important question is whether the iterative solution (in e^2) is the only solution of the effective equation for ψ (for $F_{\mu\nu}$ constant and small). If it is, then the nonlocality of the corresponding term in the effective action is just an off-shell artifact. In the present paper we did not consider the modification [7,13] of the effective action, which guarantees its gauge independence and parametrization independence. It would be interesting to see if this modification can change the situation off shell, discussed above. An even more unsatisfactory feature of the off-shell result (5.8) is the presence of the term with the anticommutator. For the iterative solution, this term vanishes in the effective action but does not vanish in the effective equations. This is precisely the kind of problem which the unique effective action deals with.[7]

ACKNOWLEDGMENT

The authors are grateful to R. N. Faustov for his advice to use the methods of effective action in the problem of the electron magnetic moment and pointing out the importance of this problem for practical applications.

APPENDIX: THE GAUSSIAN INTEGRAL WITH NONCOMMUTING SOURCES

The Gaussian moments (4.7) can be obtained by differentiating the integral

$$J(\varepsilon) = \frac{1}{(4\pi u)^{\omega}} \int \left(\prod_{\beta=1}^{2\omega} d\sigma^{\beta} \right) e^{-(1/4u) g_{\mu\nu} \sigma^{\mu} \sigma^{\nu}} e^{\sigma^{\alpha}(\varepsilon_{\alpha} - \nabla_{\alpha})} \tag{A1}$$

with respect to the numerical parameters ε_{α}. Here ∇_{α} are, generally, abstract operators whose commutator is known,

$$[\nabla_{\mu}, \nabla_{\nu}] = \hat{\mathscr{R}}_{\mu\nu} . \tag{A2}$$

In general, calculation of the integral (A1) is an outstanding problem, but for our present purposes we do not need much. We shall confine ourselves to the approximation where

$$[\nabla_{\alpha}, \hat{\mathscr{R}}_{\mu\nu}] = 0 \tag{A3}$$

and keep only terms linear in $\hat{\mathscr{R}}_{\mu\nu}$ (with flat-space $g_{\mu\nu}$).

The operator exponential function in (A1) is defined by its power series expansion which may be integrated term by term:

$$\begin{aligned} J(\varepsilon) &= \sum_{n=0}^{\infty} \frac{1}{n!} (\varepsilon - \nabla)_{\alpha_1} \cdots (\varepsilon - \nabla)_{\alpha_n} \frac{1}{(4\pi u)^{\omega}} \\ &\quad \times \int \left(\prod_{\beta=1}^{2\omega} d\sigma^{\beta} \right) e^{-(1/4u) g_{\mu\nu} \sigma^{\mu} \sigma^{\nu}} \sigma^{\alpha_1} \cdots \sigma^{\alpha_n} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(2u)^k}{(2k)!} g^{\alpha_1 \cdots \alpha_{2k}} (\varepsilon - \nabla)_{\alpha_1} \cdots (\varepsilon - \nabla)_{\alpha_{2k}} . \end{aligned} \tag{A4}$$

Here $g^{\alpha_1 \cdots \alpha_{2k}}$ is the completely symmetric tensor defined by the recursion relations

$$g^{\alpha_1 \cdots \alpha_{2k}} = \sum_{j=2}^{2k} g^{\alpha_1 \alpha_j} g^{\alpha_2 \cdots \alpha_{j-1} \alpha_{j+1} \cdots \alpha_{2k}}, \quad g^{\alpha_1 \alpha_2} g_{\alpha_2 \alpha'_1} = \delta^{\alpha_1}_{\alpha'_1} . \tag{A5}$$

Next, in each term of the sum (A4), the multipliers should be commuted in such a way as to form the operator $(\nabla_{\alpha} \nabla^{\alpha})^k$. Using (A3) we find that up to terms $O(\hat{\mathscr{R}}^2)$ the result has the same form as in the case of commuting sources,

$$\begin{aligned} J(\varepsilon) &= \sum_{k=0}^{\infty} \frac{u^k}{k!} [(\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha}]^k + O(\hat{\mathscr{R}}^2) \\ &\quad + O([\nabla, \hat{\mathscr{R}}]) \\ &= e^{u(\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha}} + O(\hat{\mathscr{R}}^2) + O([\nabla, \hat{\mathscr{R}}]) . \end{aligned} \tag{A6}$$

The differentiation of (A6) with respect to ε_{μ} again encounters the problem of noncommutativity. We proceed by using the power series expansion of (A6):

$$\begin{aligned} \frac{\partial}{\partial \varepsilon_{\mu}} e^{u(\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha}} &= \frac{\partial}{\partial \varepsilon_{\mu}} \sum_{n=0}^{\infty} \frac{u^n}{n!} [(\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha}]^n \\ &= \sum_{n=0}^{\infty} \frac{u^n}{n!} \{ 2n (\varepsilon - \nabla)^{\mu} ((\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha})^{n-1} \\ &\quad + n(n-1) [(\varepsilon - \nabla)_{\nu} (\varepsilon - \nabla)^{\nu}, (\varepsilon - \nabla)^{\mu}] ((\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha})^{n-2} \} \\ &\quad + O(\hat{\mathscr{R}}^2) + O([\nabla, \hat{\mathscr{R}}]) . \end{aligned} \tag{A7}$$

Since

$$\begin{aligned} &[(\varepsilon - \nabla)_{\nu} (\varepsilon - \nabla)^{\nu}, (\varepsilon - \nabla)^{\mu}] \\ &\quad = -2 \hat{\mathscr{R}}^{\mu\nu} (\varepsilon - \nabla)_{\nu} + O([\nabla, \hat{\mathscr{R}}]) , \end{aligned} \tag{A8}$$

the final result is

$$\begin{aligned} &\frac{\partial}{\partial \varepsilon_{\mu}} e^{u(\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha}} \\ &\quad = 2u (g^{\mu\nu} - u \hat{\mathscr{R}}^{\mu\nu}) (\varepsilon - \nabla)_{\nu} e^{u(\varepsilon - \nabla)_{\alpha} (\varepsilon - \nabla)^{\alpha}} \\ &\qquad + O(\hat{\mathscr{R}}^2) + O([\nabla, \hat{\mathscr{R}}]) . \end{aligned} \tag{A9}$$

Thus the first Gaussian moment is already modified with the commutator term. Repeated use of (A9) makes it possible to obtain Gaussian moments of any order.

[1]T. Kinoshita and J. R. Sapirstein, in *Atomic Physics 9*, edited by R. S. Van Dyck, Jr., and E. N. Fortson (World Scientific, Singapore, 1984), p. 38.

[2]R. N. Faustov, in *Correlation and Relativistic Effects in Atoms and Ions* (Academy of Sciences, Moscow, 1986), p. 54.

[3]M. A. Samuel, Phys. Rev. Lett. **57**, 3133 (1986).

[4]T. Kinoshita and W. B. Lindquist, Phys. Rev. D **27**, 867, 877, 886 (1983).

[5]B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).

[6]A. O. Barvinsky and G. A. Vilkovisky, Phys. Rep. **119**, 1 (1985).

[7]G. A. Vilkovisky, in *Quantum Theory of Gravity*, edited by S. M. Christensen (Hilger, Bristol, 1984); Nucl. Phys. B **234**, 125 (1984).

[8]Yu. A. Golfand, in *Quantum Field Theory and Quantum Statistics*, edited by I. A. Batalin, C. J. Isham, and G. A. Vilkovisky (Hilger, Bristol, 1987), Vol. 1.

[9]B. S. DeWitt, in *Quantum Gravity 2*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford U. P., Oxford, 1981).

[10]A. O. Barvinsky and G. A. Vilkovisky, Nucl. Phys. B **282**, 163 (1987).

[11]A. O. Barvinsky and G. A. Vilkovisky, to be published in Nucl. Phys. B.

[12]J. S. Schwinger, Phys. Rev. **82**, 664 (1951).

[13]B. S. DeWitt, in *Quantum Field Theory and Quantum Statistics*, edited by I. A. Batalin, C. J. Isham, and G. A. Vilkovisky (Hilger, Bristol, 1987), Vol. 1.

Spin-polarized Thomas–Fermi theory

Jerome A. Goldstein
Department of Mathematics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118

Gisèle Ruiz Rieder
Department of Mathematics, Louisiana State University, Baton Rouge, Louisiana 70803

(Received 22 September 1987; accepted for publication 4 November 1987)

Of concern is a rigorous Thomas–Fermi theory of electron densities for spin-polarized quantum-mechanical systems. The number $N_\uparrow$, $N_\downarrow$ of spin-up and spin-down electrons are specified in advance, and one seeks to minimize the energy functional $E(\rho_\uparrow, \rho_\downarrow) = c_1 \int_{\mathbb{R}^3} (\rho_\uparrow(x)^{5/3} + \rho_\downarrow(x)^{5/3}) dx + c_2 \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} [\rho(x)\rho(y)/|x-y|] dx\,dy + \int_{\mathbb{R}^3} V(x)\rho(x) dx$, where c_1, c_2 are given positive constants, $\rho_\uparrow$ and $\rho_\downarrow$ are non-negative functions, $\rho = \rho_\uparrow + \rho_\downarrow$ is the total electron density, $\int_{\mathbb{R}^3} \rho_\uparrow(x) dx = N_\uparrow$, $\int_{\mathbb{R}^3} \rho_\downarrow(x) dx = N_\downarrow$, and V is a given potential. These results are analogous to the classical rigorous (spin-unpolarized) Thomas–Fermi theory developed by Lieb and Simon [Phys. Rev. Lett. **33**, 681 (1973)] and by Bénilan and Brezis ("The Thomas–Fermi problem," in preparation).

I. INTRODUCTION

Consider a quantum-mechanical system having N electrons with Z_j protons at a nucleus in a fixed location R_j in $\mathbb{R}^3$, for $j = 1,...,M$. In Thomas–Fermi theory one studies the ground-state electron density for such a system. An *N electron density* is a non-negative integrable function ρ on $\mathbb{R}^3$ satisfying

$$\int_{\mathbb{R}^3} \rho(x) dx = \|\rho\|_{L^1} = N.$$

The *ground-state density* ρ_0 is the N electron density that minimizes the total energy of the system, when viewed as a functional of the density. The ground-state density satisfies the property implied by its name, i.e., if Λ is any Borel set in $\mathbb{R}^3$, then $\int_\Lambda \rho_0(x) dx$ is the expected number of electrons to be found in Λ at any instant of time (when the system is in its ground state).

The simplest expression for the energy as a functional of the density goes back to Thomas[1] and Fermi[2] in the early days of quantum mechanics (1927). The resulting Thomas–Fermi (ground-state) energy and density have certain nice properties. For instance, a scaling argument shows that the energy is exact as $Z = \Sigma_{j=1}^M Z_j \to \infty$ (cf. Ref. 3). Thomas–Fermi theory is useful in calculating properties that depend on the "average electron," such as total, kinetic, and exchange energies. On the other hand, the theory is less effective for calculating properties depending on valence shell electrons such as molecular bonding energies.

Thomas–Fermi theory is traditionally a spin-unpolarized theory in which half of the electrons are spin up and half are spin down. A spin-polarized theory is one in which there is an excess of spin-up (or spin-down) electrons. Several physicists and chemists working in density-functional theories have discovered that spin-polarized theories can lead to better approximations of molecular bonding energies, kinetic energies, and other numbers of interest. (Compare, e.g., Ref. 4.)

In high magnetic fields, at high temperatures, and in certain other circumstances, the ground state of the system is known experimentally to be spin polarized. It is not unusual to see the electron configuration for a nitrogen atom in its ground state depicted as

2*p* $\uparrow \uparrow \uparrow$
2*s* $\uparrow\downarrow$
1*s* $\uparrow\downarrow$.

The 1*s* and 2*s* orbitals are filled with one spin-up electron and one spin-down electron, while the three 2*p* orbitals contain a single spin-up electron apiece. Gadiyak and Lozovik[5] and Pathak,[6] in his unpublished thesis, have obtained results in formal spin-polarized Thomas–Fermi theory, but not in a rigorous mathematical context. Lieb and Simon [Ref. 3(b), p. 34] indicated the possibility that their rigorous spin-unpolarized theory could be extended to the spin-polarized case.

Our purpose here is to put spin-polarized Thomas–Fermi theory on a rigorous mathematical foundation. Many simplifying assumptions are present in our model. We treat the usual Thomas–Fermi model but we specify both the member of spin-up and spin-down electrons in advance. (Thus temperature, exchange terms, gradient expansions, and relativistic corrections are ignored.) Nevertheless, even in this simple case, two new mathematical complications arise. First of all, the Euler–Lagrange equation for our minimization problem reduces to a *system* of nonlinear elliptic partial differential equations rather than a single equation. Second, in spin-unpolarized Thomas–Fermi theory the monotonicity properties of the relationship between the chemical potential and the electron number leads easily to certain conclusions. In the spin-unpolarized case, the function alluded to above maps a subset of $\mathbb{R}$ into $\mathbb{R}$. However, the analog of this function in the spin-polarized case maps a subset of the plane $\mathbb{R}^2$ into $\mathbb{R}^2$. Herein lies a key difference in the analysis of the spin-polarized and spin-unpolarized uses. The key steps are more difficult in the spin-polarized case, and, in particular, it is harder to determine the range of this function.

 0022-2488/88/030709-08$02.50

This paper is organized as follows. In Sec. II we formulate and discuss the minimization problem for the ground-state densities and the corresponding Euler–Lagrange problem. In Sec. III we explain how to find the ground-state densities by solving partial differential equations. The technical section, Sec. IV, is devoted to proofs. Novel results on compact support for a maximal system of electrons are obtained. In Sec. V it is shown that a (generalized) atom has radially nonincreasing ground-state densities for both the spin-up and spin-down electrons.

II. THE MINIMIZATION PROBLEM

Consider a system with $N_\uparrow$ electrons of the spin-up variety and $N_\downarrow$ spin-down electrons. We emphasize that $N_\uparrow$ and $N_\downarrow$ are specified in advance. Let $N = N_\uparrow + N_\downarrow$ be the total number of electrons for the system. Let $\rho_\uparrow$ and $\rho_\downarrow$ be the corresponding densities. From now on we replace the subscripts $\uparrow$, $\downarrow$ by 1, 2 for typographical convenience. Thus, for any Borel set Λ in $\mathbb{R}^3$, $\int_\Lambda \rho_i(x)dx$ is the expected number of electrons with spin i in Λ (at any instant of time), and

$$\int_{\mathbb{R}^3} \rho_i(x)dx = N_i$$

for $i = 1,2$. Let

$$L^1_+ = \{\rho \in L^1(\mathbb{R}^3):\ \rho \geqslant 0\},$$

$$L^1_+[N] = \left\{\rho \in L^1_+ :\ \int_{\mathbb{R}^3} \rho(x)dx = N\right\}.$$

Consider the *energy functional* $\mathscr{E}$ defined by

$$\mathscr{E}(\rho_1,\rho_2) = \int_{\mathbb{R}^3} (J(\rho_1) + J(\rho_2))(x)dx + \int_{\mathbb{R}^3} V(x)\rho(x)dx + \frac{c_{ee}}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy. \qquad (1)$$

Here $\rho = \rho_1 + \rho_2$ is the *total electron density*, and $\mathscr{E}$ is defined on the largest subset of $L^1_+ \times L^1_+$ such that each term on the right-hand side of (1) makes sense. We now describe the three integral terms in (1).

The kinetic energy term involves a convex function J: $[0,\infty) \to [0,\infty)$ satisfying

$$J(0) = J'(0) = 0, \quad J'' \geqslant 0, \quad J > 0 \quad \text{on } (0,\infty). \qquad \text{(J)}$$

The usual Thomas–Fermi kinetic energy approximation, in atomic units with $h = 1$, for the spin-polarized case is given by

$$J(r) = \tfrac{3}{10}(6\pi^2)^{2/3} r^{5/3}.$$

This is derived formally in Refs. 5 and 6 and incorporates the Fermi statistics of the electrons.

The second term in (1) represents the electron-nuclear attraction and is the only term in (1) that corresponds exactly to its wave function analog. For a discussion connecting wave function theory with Thomas–Fermi theory, see Ref. 7. In order to allow the potential V to be as general as possible we make at this point only the minimal assumption

$$V \in L^1_{\text{loc}}(\mathbb{R}^3) \quad \text{and} \quad V < 0 \quad \text{on a set of positive measure.} \qquad \text{(V)}$$

This (V) is a *necessary* condition for the existence of a Thomas–Fermi ground-state density in both the spin-polarized and usual (or spin-unpolarized) theories (cf. Ref. 8). The most important special cases are the *molecular Coulomb potential*

$$V(x) = -\sum_{j=1}^{M} \frac{Z_j}{|x - R_j|} \qquad (2)$$

and the *atomic Coulomb potential*

$$V(x) = -Z/|x|. \qquad (3)$$

Here Z and Z_j are given positive numbers.

The final term in the definition of the energy functional $\mathscr{E}$ corresponds to classical electron–electron repulsion. The customary choice of the constant c_{ee} is 1; the Fermi–Amaldi[9] choice of $c_{ee} = (N-1)/N$ vanishes when there is only one electron and hence no electron–electron repulsion, and c_{ee} is approximately 1 for large N. As we shall see later, negative ions do not exist in spin-polarized Thomas–Fermi theory (with $c_{ee} = 1$), but singly negative ions exist under the Fermi–Amaldi hypothesis.

The problem of finding the ground-state energy and densities in spin-polarized Thomas–Fermi theory is stated as follows.

Minimization problem: Assume (J), (V), with $\mathscr{E}$ given by (1). Find $(\bar\rho_1,\bar\rho_2) \in \mathscr{D}(N_1,N_2)$ such that

$$\mathscr{E}(\bar\rho_1,\bar\rho_2) = \min\{\mathscr{E}(\rho_1,\rho_2):\ (\rho_1,\rho_2) \in \mathscr{D}(N_1,N_2)\}, \qquad (4)$$

where

$$\mathscr{D}(N_1,N_2) = \{(\rho_1,\rho_2):\ \rho_i \in L^1_+[N_i],\ J(\rho_i) \in L^1_+,\ V\rho_i \in L^1(\mathbb{R}^3),\ (x,y) \to \rho(x)\rho(y)|x-y|^{-1} \in L^1(\mathbb{R}^3 \times \mathbb{R}^3),\ \text{for } i = 1,2\}.$$

Here N_1 and N_2 are given positive numbers.

Recall that $\rho = \rho_1 + \rho_2$ is the total electron density. One can easily show that the functional $\mathscr{E}$ on the convex set $\mathscr{D}(N_1,N_2)$ is strictly convex. Thus, if a minimizing $(\bar\rho_1,\bar\rho_2)$ exists, it is unique. But the domain $\mathscr{D}(N_1,N_2)$ of the minimization problem incorporates the constraints $\rho_i \geqslant 0$ and $\int_{\mathbb{R}^3} \rho_i(x)dx = N_i$ for $i = 1,2$. The integral constraints suggest the introduction of Lagrange multipliers λ_1, λ_2, and the minimization of the functional

$$E(\rho_1,\rho_2) = \mathscr{E}(\rho_1,\rho_2) + \sum_{i=1}^{2} \lambda_i\left(\int_{\mathbb{R}^3} \rho_i(x)dx - N_i\right).$$

We proceed formally, ignoring the constraints $\rho_i \geqslant 0$. The Euler–Lagrange equations take the form

$$\frac{\partial E}{\partial \rho_i}(\bar\rho_1,\bar\rho_2) = \frac{\partial E}{\partial \rho_2}(\bar\rho_1,\bar\rho_2) = 0.$$

The computation of $\partial E/\partial \rho_i$ is analogous to that in the spin-unpolarized case (cf. Refs. 7, 8, 10), but there is an extra complication in the electron–electron repulsion term. Define the convolution operator B by $B = (4\pi|\cdot|)^{-1}*$, i.e.,

$$Bf(x) = \frac{1}{4\pi}\int_{\mathbb{R}^3} f(y)|x-y|^{-1}\,dy.$$

It is well-known from Newtonian potential theory that B is the inverse of the negative Laplacian acting on functions on $\mathbb{R}^3$, i.e., $B = (-\Delta)^{-1}$. The electron–electron repulsion term thus becomes

$$2\pi c_{ee}\int_{\mathbb{R}^3}\rho B\rho\,dx$$

$$= 2\pi c_{ee}\int_{\mathbb{R}^3}[\rho_1 B\rho_1 + 2\rho_2 B\rho_1 + \rho_2 B\rho_2]dx,$$

since B is (formally) self-adjoint. Consequently,

$$\frac{\partial}{\partial\rho_2}\left[2\pi c_{ee}\int_{\mathbb{R}^3}\rho B\rho\,dx\right] = 2\pi c_{ee}[2B\rho_1 + 2B\rho_2]$$

$$= 4\pi c_{ee}B\rho,$$

and similarly

$$\frac{\partial}{\partial\rho_1}\left[2\pi c_{ee}\int\rho B\rho\,dx\right] = 4\pi c_{ee}B\rho.$$

This leads us to the Euler–Lagrange problem associated with the minimization problem [cf. (4)]; a precise statement follows.

Euler–Lagrange problem: Assume (J), (V), and let

$$\widetilde{\mathscr{D}}(N_1,N_2) = L^1_+[N_1]\times L^1_+[N_2].$$

Find $(\bar\rho_1,\bar\rho_2,\lambda_1,\lambda_2)\in\widetilde{\mathscr{D}}(N_1,N_2)\times\mathbb{R}^2$ such that, a.e.,

$$\begin{aligned} &J'(\bar\rho_1) + V + kB(\bar\rho_1+\bar\rho_2) + \lambda_1\in\mathscr{O}(\bar\rho_1),\\ &J'(\bar\rho_2) + V + kB(\bar\rho_1+\bar\rho_2) + \lambda_2\in\mathscr{O}(\bar\rho_2),\end{aligned}\tag{5}$$

where $k = 4\pi c_{ee}$ and for $i = 1,2$,

$$\mathscr{O}(\bar\rho_i) = \begin{cases}\{0\}, & \text{on } \{x:\ \rho_i(x)>0\},\\ [0,\infty), & \text{on } \{x:\ \rho_i(x)=0\}.\end{cases}$$

The Lagrange multipliers λ_1, λ_2 are the *electronegativities*, while their negative $-\lambda_1$, $-\lambda_2$ are the *chemical potentials*. The notation $g\in\mathscr{O}(\bar\rho_i)$ combines an equation (on $[\bar\rho_i>0]$) together with an inequality (on $[\bar\rho_1=0]$). The inequalities arise as a consequence of the constraint that the densities are non-negative. Since the domain of admissible densities (ρ_1,ρ_2) is larger in the Euler–Lagrange problem than in the minimization problem, we expect solutions of latter to satisfy the former, but not necessarily conversely. The precise relationship between the two problems is as follows.

Theorem 1: Assume (J), (V). If $(\bar\rho_1,\bar\rho_2)$ solves the minimization problem, then there exists a unique pair $(\lambda_1,\lambda_2)\in\mathbb{R}^2$ such that $(\bar\rho_1,\bar\rho_2,\lambda_1,\lambda_2)$ solves the Euler–Lagrange problem. Conversely let $(\bar\rho_1,\bar\rho_2,\lambda_1,\lambda_2)$ solve the Euler–Lagrange problem. If there is a real constant M such that

$$x\to J^*((M-V(x))_+\in L^1(\mathbb{R}^3),$$

then $(\bar\rho_1,\bar\rho_2)$ solves the minimization problem.

Here $a_+ = \max\{a,0\}$, and J^* is the convex conjugate function (or conjugate of J) defined by, for $t\geqslant 0$,

$$J^*(t) = \sup\{ts - J(s):\ s\geqslant 0\}.$$

In particular, if $J(s) = cs^p$ for some $p\in(1,\infty)$ and $c>0$, then $J^*(t) = (cp)^{p/q}t^q$ for $t\geqslant 0$, where $p^{-1}+q^{-1}=1$. The assumption that $J^*((M-V)_+)$ is integrable ensures that $\inf\mathscr{E}(\rho_1,\rho_2) > -\infty$.

Theorem 1 is proved by making obvious modifications of the proof given in Ref. 10, which is the spin-unpolarized case. The ideas were sketched earlier by Brezis,[8] and the full details were given in Ref. 10. □

We remark that letting $|x|\to\infty$ in (5) shows that $\lambda_1\geqslant 0$, $\lambda_2\geqslant 0$ provided that $V(x)\to 0$ as $|x|\to\infty$, either in the usual sense or in a weak sense [e.g., $V\in\text{Weak } L^p(x:\ |x|>R)$ for some $R>0$, $p<\infty$. These spaces will be defined in the next section]. In this case, in the statement of Theorem 1, we may replace $(\lambda_1,\lambda_2)\in\mathbb{R}^2$ by $(\lambda_1,\lambda_2)\in[0,\infty)^2$. Furthermore, $N_1\geqslant N_2$, $N_1 = N_2$ turn out to be equivalent to, respectively, $\lambda_2\geqslant\lambda_1$, $\lambda_2=\lambda_1$, and in the latter case (i.e., $N_1 = N_2$) we have the usual (spin-unpolarized) Thomas–Fermi theory.

III. SOLUTION OF THE EULER–LAGRANGE PROBLEM VIA NONLINEAR ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

Set

$$u_i = -\tfrac{1}{2}V - kB\bar\rho_i,$$

for $i = 1,2$ and define $\Gamma:[0,\infty)\to[0,\infty)$ by

$$\Gamma(s) = \begin{cases}(J')^{-1}(s), & \text{for } s\in(0,\infty),\\ 0, & \text{for } s\in(-\infty,0].\end{cases}\tag{6}$$

The Euler–Lagrange problem (5) (which consists of two equations and two inequalities) reduces to

$$J'(\bar\rho_i)\in u_i + u_j - \lambda_i + \mathscr{O}(\bar\rho_i),$$

for $i = 1,2$ and $j\neq i$ (with $j\in\{1,2\}$). Applying Γ yields the system of two *equations*

$$\bar\rho_i = \Gamma(u_i + u_j - \lambda_i).\tag{7}$$

Since we require that $\bar\rho_i\in L^1_+[N_i]$, integrating (7) yields

$$\int_{\mathbb{R}^3}\Gamma(u_i(x) + u_j(x) - \lambda_i)dx = N_i,\tag{8}$$

for $i,j\in\{1,2\}$ with $i\neq j$. Applying $-\Delta$ to u_i and using (7) and (8) leads to a coupled system of elliptic equations. This version of the spin-polarized Thomas–Fermi problem can be stated precisely as follows.

Nonlinear elliptic problem (first version): Find u_1, $u_2\in\text{Weak } L^3(\mathbb{R}^3)$ and $(\lambda_1,\lambda_2)\in\mathbb{R}^2$ such that

$$\begin{aligned}-\Delta u_1 + k\Gamma(u_1+u_2-\lambda_1) &= \tfrac{1}{2}\Delta V,\\ -\Delta u_2 + k\Gamma(u_1+u_2-\lambda_2) &= \tfrac{1}{2}\Delta V,\end{aligned}\tag{9}$$

in the sense of distributions, and

$$\begin{aligned}N_1 &= \int_{\mathbb{R}^3}\Gamma(u_1(x)+u_2(x)-\lambda_1)dx,\\ N_2 &= \int_{\mathbb{R}^3}\Gamma(u_1(x)+u_2(x)-\lambda_2)dx.\end{aligned}\tag{10}$$

In studying the nonlinear elliptic problem we shall make a stronger assumption on V, namely, that (V) holds and $\Delta V\in\mathscr{M}(\mathbb{R}^3) + L^1(\mathbb{R}^3)$, i.e., ΔV is the sum of a finite signed measure and an integrable function. The results that follow are especially clean when ΔV is non-negative.

The spaces Weak $L^p(\mathbb{R}^3)$ are the weak L^p spaces or Marcinkiewicz spaces. A measurable function f on $\mathbb{R}^3$ is in Weak $L^p(\mathbb{R}^3)$ iff $|||f|||_p<\infty$, where

$$|||f|||_p = \inf\left\{K:\ \int_A |f(x)|^p dx \leqslant K\left[\int_A dx\right]^{1/q},\right.$$
$$\left.\text{for all bounded Borel sets } A\subset\mathbb{R}^3\right\};$$

here $p<\infty$ and $p^{-1}+q^{-1}=1$. Some basic facts concerning these spaces are collected in the following lemma.[11]

Lemma 1 (cf. Ref. 10): (i) Let $1\leqslant r<p<\infty$. Then Weak $L^p(\mathbb{R}^3)\subset L^r_{\text{loc}}(\mathbb{R}^3)$ with continuous injection; and $u\in$Weak $L^p(\mathbb{R}^3)$ implies $|u|^r\in$Weak $L^{p/r}(\mathbb{R}^3)$.

(ii) The function $x\to|x|^{-\alpha}$ belongs to Weak $L^{3/\alpha}(\mathbb{R}^3)$, for $0<\alpha<3$.

(iii) If $E\in$Weak $L^p(\mathbb{R}^3)$, $1<p<\infty$, and $f\in L^1(\mathbb{R}^3)$, then $E*f\in$Weak $L^p(\mathbb{R}^3)$ and

$|||E*f|||_p\leqslant|||E|||_p\|f\|_{\lambda^1}$.

(iv) Write "$g(\infty)=0$" iff for all $\epsilon\geqslant0$ there is a Borel set A_ϵ in $\mathbb{R}^3$ of finite measure such that $|g(x)|<\epsilon$ for all $x\notin A_\epsilon$. If $g\in$Weak $L^p(\mathbb{R}^3)$, $1<p<\infty$, then "$g(\infty)=0$".

Coupled systems of nonlinear elliptic equations are in general difficult to handle. Our particular system has the apparent additional complications caused by the necessity of working in a spaces of densities [in $L^1(\mathbb{R}^3)$] and by the presence of bounded signed measures that arise from ΔV. [Recall that $-\Delta(|x|^{-1})=4\pi\delta_0$, where δ_0 is the Dirac point mass at the origin in $\mathbb{R}^3$.] However, our system of equations is greatly simplified by the introduction of a new variable.

Let

$$w=u_1+u_2.$$

Then the pair of equations (9) can be added to give the single equation

$$-\Delta w+\sum_{i=1}^{2}\Gamma(w-\lambda_i)=\Delta V. \tag{11}$$

This suggests an alternate version of our partial differential equation (PDE) problem.

Nonlinear elliptic problem (second version): Find $w\in$Weak $L^3(\mathbb{R}^3)$ and $(\lambda_1,\lambda_2)\in\mathbb{R}^2$ such that (11) holds in the sense of distributions and

$$N_i=\int_{\mathbb{R}^3}\Gamma(w(x)-\lambda_i)dx,\quad i=1,2.$$

The two versions of the nonlinear elliptic problem are equivalent. The second is simpler in that it only involves one PDE, and the densities can be found directly from $\bar\rho_i=\Gamma(w-\lambda_i)$. Of course, u_i can be found by solving

$$-\Delta u_i=f_i$$

[see (9)], where $f_i=\frac{1}{2}\Delta V-k\Gamma(w-\lambda_i)$ is known once w is known.

Prior to solving the Euler–Lagrange problem by means of the nonlinear elliptic problem (second version), we make additional assumptions on J and V.

Hypothesis 1: Let (V) and (J) hold. Suppose further that "$V(\infty)=0$", $0\leqslant\Delta V\in L^1(\mathbb{R}^3)$, and

$$\int_{|x|>1}\Gamma(c|x|^{-1})dx=\infty, \tag{12}$$

for some $c>0$.

Hypothesis 2: Let (V) and (J) hold. Suppose further that "$V(\infty)=0$", $0\leqslant\Delta V\in\mathscr{M}(\mathbb{R}^3)$, (12) holds, and $x\to\Gamma(|x|^{-1})$ is integrable in a neighborhood of $x=0$.

Recall that Γ is defined by (6). Note that either Hypothesis 1 or Hypothesis 2 implies $0<\int_{\mathbb{R}^3}\Delta V<\infty$.

Lemma 2: Assume either Hypothesis 1 or Hypothesis 2. Then the nonlinear elliptic problem (either version) is equivalent to the Euler–Lagrange problem.

Proof: The equivalence of the first version of the nonlinear elliptic problem and the Euler–Lagrange problem is essentially the same as in the spin-unpolarized case.[9,10] Any solution of the first version of the nonlinear elliptic problem gives a solution of the second version when we set $w=u_1+u_2$. Conversely, passing from the second version to the first involves solving

$$-\Delta u_i=\tfrac{1}{2}\Delta V-k\Gamma(w-\lambda_i)$$

as was discussed following the statement of the second version of the nonlinear elliptic problem. □

We now solve the second version of the nonlinear elliptic problem and thereby obtain the desired solution of the Euler–Lagrange problem (and the minimization problem as well in many cases).

Theorem 2: Assume either Hypothesis 1 or Hypothesis 2. Let

$$N_0=\frac{1}{4\pi c_{ee}}\int_{\mathbb{R}^3}\Delta V. \tag{13}$$

Then (recall $N=N_1+N_2$) the Euler–Lagrange problem has a unique solution whenever $0<N\leqslant N_0$ and no solution when $N>N_0$. If $V(x)\to0$ as $|x|\to\infty$ and if $0<N<N_0$, then the solution densities $\bar\rho_1$, $\bar\rho_2$ have compact support. Moreover, if $N_1>N_2$, then $\bar\rho_2$ has compact support, even if $N=N_0$.

This theorem will be proved in the next section. First some remarks are in order.

According to Theorem 2, $N=N_0$ is the maximum number of electrons our quantum-mechanical system can have (in spin-polarized Thomas–Fermi theory), and this N is determined by V and c_{ee} via (13). If V is the molecular Coulomb potential (2), then

$$\frac{1}{4\pi}\int_{\mathbb{R}^3}\Delta V=\sum_{j=1}^{M}\int Z_j\delta_{R_j}=Z,$$

where $Z=\Sigma_{j=1}^{M}Z_j$ is the total number of protons in the molecule. The same is true for atoms. Thus for $N=N_0$, (13) gives

$$N=c_{ee}^{-1}Z.$$

When $c_{ee}=1$, it follows that $N\leqslant Z$, and therefore no negative ions exist (in this theory). However, in the Fermi–Amaldi case of $c_{ee}=(N-1)/N$, the maximum value of N satisfies (13), i.e.,

$$N=N(N-1)^{-1}Z,$$

or $N=Z+1$. Thus singly negative ions exist (but not doubly negative ions). On the other hand, in Thomas–Fermi theory, neither N nor Z need be integral.

For $J(r)=cr^p$ for $r\geqslant0$ with $c>0$ and $1<p<\infty$, condition (12) is equivalent to $p\geqslant\frac{4}{3}$. If $\Delta V\in\mathscr{M}(\mathbb{R}^3)$ rather than $L^1(\mathbb{R}^3)$, then (12) plus the final condition of Hypothesis 2 is equivalent to $p > \frac{4}{3}$. When V is the molecular Coulomb potential (2), then the condition in the converse of Theorem 1 holds iff $p > \frac{3}{2}$. Thus when $p > \frac{3}{2}$, we can solve the minimization problem. However, for $\frac{4}{3} < p \leqslant \frac{3}{2}$, we have a solution of the Euler–Lagrange problem but no solution of the minimization problem [since in this case $\inf \mathscr{E}(\rho_1,\rho_2) = -\infty$]. It is worth emphasizing that the physical case of $p = \frac{5}{3}$ falls comfortably into the acceptable range for both problems.

IV. PROOF OF THEOREM 2

First fix $(\lambda_1,\lambda_2)\in[0,\infty)^2$. Without loss of generality we may assume $\lambda_2 \geqslant \lambda_1$. For $\xi\in\mathbb{R}$ set

$$\beta(\xi) = \sum_{j=1}^{2} k\Gamma(\xi - \lambda_j) \tag{14}$$

with Γ defined by (6) and, as before, $k = 4\pi c_{ee}$. We then solve, for $\lambda = (\lambda_1,\lambda_2)$,

$$-\Delta w_\lambda + \beta(w_\lambda) = f, \quad \beta(w_\lambda)\in L^1(\mathbb{R}^3), \tag{15}$$

for $f = \Delta V$ together with the condition that "$w_\lambda(\infty) = 0$". The following result of Benilan *et al.*[11] is the right tool for this problem.

Proposition 1: Let β: $\mathbb{R}\to\mathbb{R}$ be continuous, nondecreasing, and satisfy $\beta(0) = 0$.

(i) If $f\in L^1(\mathbb{R}^3)$, then (15) has a unique weak solution w_λ in Weak $L^3(\mathbb{R}^3)$.

(ii) If $x\to\beta(\pm|x|^{-1})$ is integrable in a neighborhood of the origin, then for every $f\in\mathscr{M}(\mathbb{R}^3)$, problem (15) has a unique solution w_λ in Weak $L^3(\mathbb{R}^3)$.

Lemma 3: The function β: $\mathbb{R}\to[0,\infty)$ defined by (14) is continuous, nondecreasing, and satisfies $\beta(0) = 0$.

This lemma follows easily from the hypotheses (J) on J, the definition of Γ, and the fact that $\lambda_1,\lambda_2 \geqslant 0$. □

Thus for each fixed $\lambda = (\lambda_1,\lambda_2)$, Proposition 1 (i) [resp. Proposition 1 (ii)] guarantees the existence of a unique solution w_λ in Weak $L^3(\mathbb{R}^3)$ of (11) under the assumption Hypothesis 1 [resp. Hypothesis 2].

In both cases, $\beta(w_\lambda)\in L^1(\mathbb{R}^3)$.

For $i = 1,2$ set

$$N_i(\lambda) = \int_{\mathbb{R}^3} \Gamma(w_\lambda(x) - \lambda_i)dx,$$
$$\mathbf{N}(\lambda) = (N_1(\lambda),N_2(\lambda)), \tag{16}$$
$$N(\lambda) = N_1(\lambda) + N_2(\lambda).$$

Proposition 2: The function $\mathbf{N}(\cdot)$: $[0,\infty)^2\to[0,\infty)^2$ is continuous. For $i,j\in\{1,2\}$ with $i\neq j$, $N_i(\lambda_1,\lambda_2)$ is a nonincreasing function of λ_i and a nondecreasing function of λ_j. Both $N_1(\lambda)$ and $N_2(\lambda)$ are strictly decreasing on lines of slope 1 that pass through the positive λ_2 axis [i.e., if $\lambda_1 < \mu_1$, $\lambda_2 < \mu_2$, and $\mu_2 - \mu_1 = \lambda_2 - \lambda_1$, then $N_1(\lambda) > N_1(\mu)$ and $N_2(\lambda) > N_2(\mu)$]. Moreover, $N_1(0,\lambda_2) > 0$, $N_2(0,\lambda_2) > 0$, for all $\lambda_2 \geqslant 0$; and $N(\lambda_1,\lambda_2) < N(0,\lambda_2)$ whenever $\lambda_1 > 0$, $\lambda_2 > 0$. Finally, for $i\in\{1,2\}$, $\lim_{\lambda_i\to\infty} N_i(\lambda_1,\lambda_2) = 0$.

In proving this result we may, without loss of generality, restrict ourselves to the infinite triangle $\{(\lambda_1,\lambda_2)\in[0,\infty)^2$: $\lambda_2 \geqslant \lambda_1\}$ rather than the quarter plane $(\lambda_1,\lambda_2)\in[0,\infty)^2$.

Proof: The proof will be broken into several pieces.

Monotonicity: Let $\lambda = (\lambda_1,\lambda_2)$ and $\mu = (\mu_1,\mu_2)$ be in $[0,\infty)^2$ and satisfy $\lambda_1 < \mu_1$ and $\lambda_1 - \lambda_2 \leqslant \mu_1 - \mu_2$. Set

$$\widetilde{w}_\alpha = w_\alpha - \alpha_1,$$

for $\alpha = \lambda,\mu$. Then $\widetilde{w}_\alpha$, by Proposition 1 and Lemma 3, is the unique solution of

$$-\Delta\widetilde{w}_\alpha + \beta_\alpha(\widetilde{w}_\alpha + \alpha_1) = \Delta V, \quad \text{“}\widetilde{w}_\alpha(\infty) + \alpha_1 = 0\text{”}, \tag{17_α}$$

where β_α is given by (14) for $\alpha = \lambda,\mu$. Subtracting (17_λ) from (17_μ) yields

$$-\Delta(\widetilde{w}_\mu - \widetilde{w}_\lambda) + \beta_\mu(\widetilde{w}_\mu + \mu_1) - \beta_\lambda(\widetilde{w}_\lambda + \lambda_1) = 0.$$

Multiply this equation by $(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+$; using (14) and integrating over the ball $B_R = \{x\in\mathbb{R}^3\colon |x| < R\}$ gives

$$\int_{B_R} -\Delta(\widetilde{w}_\mu - \widetilde{w}_\lambda)(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+ + k\int_{B_R}[\Gamma(\widetilde{w}_\mu) + \Gamma(\widetilde{w}_\mu + \mu_1 - \mu_2) - \Gamma(\widetilde{w}_\lambda) - \Gamma(\widetilde{w}_\lambda + \lambda_1 - \lambda_2)](\widetilde{w}_\mu - \widetilde{w}_\lambda)_+ = 0. \tag{18}$$

Let

$$E = \{x\in\mathbb{R}^3\colon \widetilde{w}_\mu(x) > \widetilde{w}_\lambda(x)\}.$$

Since $(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+ = 0$ on $\mathbb{R}^3\setminus E$, the integrals in (18) may be taken over $E\cap B_R$ rather than B_R. Also, on E, $\Gamma(\widetilde{w}_\mu) - \Gamma(\widetilde{w}_\lambda) \geqslant 0$ and $\Gamma(\widetilde{w}_\mu + \mu_1 - \mu_2) - \Gamma(\widetilde{w}_\lambda + \lambda_1 - \lambda_2) \geqslant 0$ since Γ is a nondecreasing function. It follows that, by the divergence theorem,

$$\begin{aligned}0 &\geqslant \int_{E\cap B_R} -\Delta(\widetilde{w}_\mu - \widetilde{w}_\lambda)(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+ \\ &= -\int_{E\cap\partial B_R}\left[\frac{\partial}{\partial r}(\widetilde{w}_\mu - \widetilde{w}_\lambda)\right](\widetilde{w}_\mu - \widetilde{w}_\lambda)_+\,dS + \int_{E\cap B_R}|\nabla(\widetilde{w}_\mu - \widetilde{w}_\lambda)|^2 \\ &= -\frac{1}{2}\int_{E\cap\partial B_R}\frac{\partial}{\partial r}(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+^2\,dS + \int_{E\cap B_R}|\nabla(\widetilde{w}_\mu - \widetilde{w}_\lambda)|^2.\end{aligned} \tag{19}$$

But, by Fubini's theorem,

$$\int_R^{R+k}\int_{\partial B_R\cap E}\frac{\partial}{\partial r}(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+^2\,dS\,dr = \int_{\partial B_R\cap E}[(\widetilde{w}_\mu - \widetilde{w}_\lambda)_+^2(R+k) - (\widetilde{w}_\mu - \widetilde{w}_\lambda)_+^2 R\,]dS\to 0,$$

as $R\to\infty$ for each $k\in\mathbb{R}$ since $\lambda_1 < \mu_1$ and "$\widetilde{w}_\alpha(\infty) + \alpha_1 = 0$" for $\alpha = \lambda,\mu$. Thus setting $R\to\infty$ in (19) allows us to deduce

$$0 \geqslant \int_E |\nabla(w_\mu - w_\lambda)|^2.$$

Consequently E is a Lebesgue null set and $\widetilde{w}_\mu \leqslant \widetilde{w}_\lambda$ a.e. (whenever $\lambda_1 < \mu_1$ and $\lambda_1 - \lambda_2 \leqslant \mu_1 - \mu_2$). It follows that

$$\Gamma(w_\mu - \mu_1) \leqslant \Gamma(w_\lambda - \lambda_1) \quad \text{a.e.}, \tag{20}$$

whence

$$N_1(\mu) \leqslant N_1(\lambda). \tag{21_1}$$

A similar argument establishes

$$N_2(\mu)\leqslant N_2(\lambda) \qquad (21_2)$$

provided $\lambda_2<\mu_2$ and $\lambda_1-\lambda_2\geqslant\mu_1-\mu_2$. In particular, it follows that $N_1(\lambda_1,\lambda_2)$ is nonincreasing in λ_1 and $N_2(\lambda_1,\lambda_2)$ is nonincreasing in λ_2 on the sets specified.

Now suppose $\lambda_1<\mu_1$ and $\lambda_2=\mu_2$. Then

$$-\Delta w_\mu+k\Gamma(w_\mu-\mu_1)+k\Gamma(w_\lambda-\lambda_2)=\Delta V,$$
$$-\Delta w_\lambda+k\Gamma(w_\lambda-\lambda_1)+k\Gamma(w_\lambda-\lambda_2)=\Delta V.$$

Subtracting the former from latter gives

$$-\Delta(w_\lambda-w_\mu)+k\Gamma(w_\lambda-\lambda_1)+k\Gamma(w_\mu-\mu_1)$$
$$-k\Gamma(w_\lambda-\lambda_2)+k\Gamma(w_\mu-\lambda_2)=0.$$

Multiply by $(w_\lambda-w_\mu)_+$, integrate over B_R, and use (20) to obtain

$$\int_{B_R}-\Delta(w_\lambda-w_\mu)(w_\lambda-w_\mu)_+ + k\int_{B_R}(\Gamma(w_\lambda-\lambda_2)-\Gamma(w_\mu-\lambda_2))(w_\lambda-w_\mu)_+\leqslant 0. \qquad (22)$$

Let

$$F=\{x\in\mathbb{R}^3:\ w_\lambda(x)>w_\mu(x)\}.$$

Both integrals in (22) may be taken over $F\cap B_R$ rather than B_R since the integrands vanish on $\mathbb{R}^3\backslash F$. Also, on F,

$$\Gamma(w_\lambda-\lambda_1)\geqslant\Gamma(w_\mu-\lambda_2)$$

since Γ is nondecreasing. Thus

$$\int_{F\cap B_R}-\Delta(w_\lambda-w_\mu)(w_\lambda-w_\mu)_+\leqslant 0.$$

As in our previous calculation, we use the divergence theorem, let $R\to\infty$, and employ "$(w_\lambda-w_\mu)(\infty)=0$" to conclude that

$$\int_F|\nabla(w_\lambda-w_\mu)|^2=0.$$

Thus F is a Lebesgue null set and $w_\lambda\leqslant w_\mu$ a.e. Hence

$$\Gamma(w_\lambda-\lambda_1)\leqslant\Gamma(w_\mu-\lambda_2)$$

and so

$$N_2(\lambda)\leqslant N_2(\mu).$$

The very same argument shows that $\lambda_1=\mu_1$, $\lambda_2<\mu_2$ implies $N_1(\lambda)\leqslant N_1(\mu)$. Thus $N_i(\lambda_1,\lambda_2)$ is nonincreasing in λ_i and nondecreasing in λ_j for $j\neq i$.

Suppose now that $\lambda_2\geqslant\lambda_1>0$ and set $\mu_2=\lambda_2-\lambda_1$, $\mu_1=0$. Then the points $\lambda=(\lambda_1,\lambda_2)$ and $\mu=(0,\mu_2)$ lie on a line of slope 1 that intersects the positive vertical (or λ_2) axis. The preceding arguments show $N_1(\lambda_1,\lambda_2)\leqslant N_2(0,\mu_2)$ and $N_1(\lambda_1,\lambda_2)\leqslant N_2(0,\mu_2)$.

We have verified the monotonicity assertions in their weak form. The strict monotonicity results will be proved presently.

Continuity: Fix $\lambda_2\geqslant 0$. Let $\{\lambda_1^n\}_n$ be a sequence in $[0,\infty)$ with $\lambda_1^n\downarrow\lambda_1$, and let $\lambda^n=(\lambda_1^n,\lambda_2)$. By the monotonicity of $w_{\lambda^n}-\lambda_1^n$, w_{λ^n} converges to a function v almost everywhere and in the sense of distributions. But v is clearly a distributional solution of the same equation as w_λ $[\lambda=(\lambda_1,\lambda_2)]$ and $v\in$weak $L^3(\mathbb{R}^3)$. Thus by uniqueness, $v=w_\lambda$. The preceding argument implies $\Gamma(w_{\lambda^n}-\lambda_1^n)\leqslant\Gamma(w_\lambda-\lambda_1)$; hence by Lebesgue's monotone convergence theorem, $N_1(\lambda^n)\to N_1(\lambda)$.

Now suppose $\lambda_1^n\uparrow\lambda_1$. Again we have $w_{\lambda^n}\to w_\lambda$ and $\Gamma(w_{\lambda^n}-\lambda_1^n)\to\Gamma(w_\lambda-\lambda_1)$. An application of Lebesgue's dominated convergence theorem then gives $N_1(\lambda^n)\to N_1(\lambda)$. It follows that $N_1(\lambda_1,\lambda_2)$ is continuous in λ_1. In both cases, i.e., as λ_1^n approaches λ_1 from either side, $w_{\lambda^n}\to w_\lambda$ holds, and so $w_{\lambda^n}-\lambda_2\to w_\lambda-\lambda_2$. Applying Γ and integrating shows that $N_2(\lambda_1,\lambda_2)$ is continuous in λ_1. An analogous argument with λ_1 fixed shows that $N_1(\lambda_1,\lambda_2)$ and $N_2(\lambda_1,\lambda_2)$ are continuous in λ_2.

We next show that

$$\lim_{\lambda_i\to\infty}N_i(\lambda)=0,$$

for $i=1,2$. Since $-\Delta w_\lambda\leqslant\Delta V$ and $w_\lambda,V\in$Weak $L^3(\mathbb{R}^3)$, (a suitable version of) the maximum principle gives $w_\lambda\leqslant -V$ a.e. (cf. Ref. 11). Consequently, $\Gamma(w_\lambda-\lambda_i)\leqslant\Gamma(-V-\lambda_i)$. Using the fact that "$V(\infty)=0$" and the definition of Γ, it follows that $\Gamma(-V-\lambda_i)\to 0$ as $\lambda_i\to\infty$. Applying the dominated convergence theorem gives the desired result.

Recall our assumption that $\lambda_2\geqslant\lambda_1$. Then $N_2(\lambda_1,\lambda_2)\to 0$ as $\lambda_2\to\infty$, even if λ_1 is fixed. But our condition that $\lambda_2\geqslant\lambda_1$ was inessential and made for convenience only. Thus $N_1(\lambda_1,\lambda_2)\to 0$ as $\lambda_1\to\infty$, whether or not $\lambda_2\geqslant\lambda_1$; in particular, λ_2 can be fixed in this argument.

Strict monotonicity: Assume $\lambda_2\geqslant\lambda_1$, $0<\lambda_1<\mu_1$, $\lambda_2<\mu_2$, $\lambda_1-\lambda_2=\mu_1-\mu_2$, and $N_i(\lambda)=N_i(\mu)$, for $i=1,2$. Assume further that $N_2(\lambda)>0$. We seek a contradiction. The inequalities (20), (21) then become equalities, so $-\Delta w_\mu=-\Delta w_\lambda$, "$w_\mu(\infty)=0$", "$w_\lambda(\infty)=0$". It follows that $w_\lambda=w_\mu$ a.e. Since

$$N_1(\lambda)\geqslant N_2(\lambda)>0,$$

the sets

$$Q_i=\{x\in\mathbb{R}^3:\ \Gamma(w_\lambda(x)-\lambda_i)>0\}$$

have positive Lebesgue measure. But Γ is strictly increasing on $(0,\infty)$, whence

$$\Gamma(w_\mu(x)-\mu_i)=\Gamma(w_\lambda(x)-\lambda_i)$$

for $x\in Q_i$ iff $w_\mu(x)-\mu_i=w_\lambda(x)-\lambda_i$. Thus $\mu_i=\lambda_i$, a contradiction. Thus both $N_1(\mu)=N_1(\lambda)$ and $N_2(\mu)=N_2(\lambda)$ cannot hold.

So we suppose $N_1(\lambda)>N_1(\mu)$ and $N_2(\lambda)=N_2(\mu)$, and we seek a contradiction. [The case of $N_1(\lambda)=N_1(\mu)$, $N_2(\lambda)>N_2(\mu)$ is similar.] Since $N_2(\lambda)=N_2(\mu)$ we must have $w_\lambda-\lambda_2=w_\mu-\mu_2$ a.e. Then $w_\lambda=w_\mu+\lambda_2-\mu_2=w_\mu+\lambda_1-\mu_1$, whence $w_\lambda-\lambda_1=w_\mu-\mu_1$ a.e. But then $\Gamma(w_\lambda-\lambda_1)=\Gamma(w_\mu-\mu_1)$ a.e., which implies $N_1(\lambda)=N_1(\mu)$, a contradiction.

Next we show that

$$N_2(\lambda_1,0)>0,\quad N_1(0,\lambda_2)>0,$$

for $0\leqslant\lambda_1,\lambda_2<\infty$. As the two proofs are essentially the same, we show the latter. Let w_λ be the solution of (15) and "$w_\lambda(\infty)=0$", where $\lambda=(0,\lambda_2)$. Assume $N_1(0,\lambda_2)=0$; we seek a contradiction. Then

$$\beta(w_\lambda) = k\Gamma(w_\lambda) + k\Gamma(w_\lambda - \lambda_2) = 0 \text{ a.e.},$$

whence $w_\lambda \leqslant 0$ a.e. But $-\Delta(w_\lambda + V) = 0$, "$w_\lambda(\infty) = 0$", "$V(\infty) = 0$". By the maximum principle it follows that $w_\lambda = -V$ a.e. Thus $V \geqslant 0$ a.e., which contradicts assumption (V) [which states that $\{x\in\mathbb{R}^3: V(x) < 0\}$ has positive measure].

This completes the proof of Proposition 2. □

Lemma 4: Let $N_0 = N(0,\lambda_2)$ where $\lambda_2 \geqslant 0$ is fixed. Then (13) holds, and N_0 is independent of λ_2.

Proof: We give the proof for the case that $\Delta V \in L^1(\mathbb{R}^3)$. [The more general case of $\Delta V \in \mathscr{M}(\mathbb{R}^3)$ requires only minor modifications.] By Proposition 1 (i), $w_\lambda \in \text{Weak } L^3(\mathbb{R}^3)$ and $\Delta w_\lambda \in L^1(\mathbb{R}^3)$, where w_λ is the solution of (15) and "$w_\lambda(\infty) = 0$" for $\lambda = (0,\lambda_2)$. It follows that

$$w_\lambda = (4\pi|x|)^{-1} * (-\Delta w_\lambda)$$

(cf. the appendix in Ref. 11). Consequently w_λ is asymptotic to $c/|x|$ as $|x| \to \infty$, where $c = (4\pi)^{-1} \int_{\mathbb{R}^3} (-\Delta w_\lambda)$. Assume $\int_{\mathbb{R}^3} \Delta w_\lambda < 0$, or, equivalently, $c > 0$. Then $\Gamma(w_\lambda) \in L^1(\mathbb{R}^3)$, which implies $\Gamma(c|x|^{-1})$ is integrable outside some ball; this contradicts (12).

It follows that $\int_{\mathbb{R}^3} \Delta w_\lambda \geqslant 0$, and so

$$\int_{\mathbb{R}^3} \beta(w_\lambda) \geqslant \int_{\mathbb{R}^3} \Delta V.$$

Therefore it only remains to show that $\int_{\mathbb{R}^3} \beta(w_\lambda) \leqslant \int_{\mathbb{R}^3} \Delta V$. Recall that $\int_{\mathbb{R}^3} (-\Delta w_\lambda)(\text{sgn } w_\lambda)_+ \geqslant 0$ [since $-\Delta$ is accretive on $L^1(\mathbb{R}^3)$]. Multiplying $-\Delta w_\lambda + \beta(w_\lambda) = \Delta V$ by $(\text{sgn } w_\lambda)_+$, integrating over $\mathbb{R}^3$, and noting that $\beta(w_\lambda) = 0$ whenever $w_\lambda \leqslant 0$, we obtain

$$\int_{\mathbb{R}^3} \beta(w_\lambda(x))dx = \int_{\mathbb{R}^3} \beta(w_\lambda(x))(\text{sgn}(w_\lambda(x)))_+ \, dx$$

$$\leqslant \int_{[w_\lambda > 0]} \Delta V \leqslant \int_{\mathbb{R}^3} \Delta V$$

by our assumption that $\Delta V \geqslant 0$. □

View $\mathbf{N} = (N_1,N_2)$ as a map from the triangle $T_0 = \{(\lambda_1,\lambda_2): 0 \leqslant \lambda_1 \leqslant \lambda_2 < \infty\}$ to $[0,\infty)^2$. The monotonicity properties of Proposition 2 (and the proof of Proposition 2) show that $\mathbf{N}$ is injective on T_0. We next show that the image of T_0 under $\mathbf{N}$ is the triangle T_1 pictured in Fig. 1.

Lemma 5: The image of T_0 under $\mathbf{N}$ is the triangle $T_1 = \{(N_1,N_2): 0 < N_2 \leqslant N_1, N_1 + N_2 \leqslant N_0\}$.

Proof: First we prove that the interior of T_1 is in the range of T_0. We do this by contradiction. To that end suppose there is a point $\mathbf{N}^* = (N_1^*,N_2^*)$ in the interior of T_1 and in the boundary of the image of $\mathbf{N}$. Choose a sequence $\{\mathbf{N}^n\}$ in the range of $\mathbf{N}$ that converges to $\mathbf{N}^*$ as $n \to \infty$. Since $\mathbf{N}^*$ is in the interior of T_1, it follows that $N_2^* > 0$, $N_1^* > N_2^*$, and $N_1^* + N_2^* < N_0$. Choose λ^n in T_0 such that $\mathbf{N}(\lambda^n) = N^n$ for each $n \geqslant 1$. Choose a subsequence of $\{\lambda^n\}$, which we denote by $\{\lambda^n = (\lambda_1^n,\lambda_2^n)\}$, such that $\lambda_1^n \to \lambda_1^* \in [0,\infty]$, $\lambda_2^* \to \lambda_2^* \in [0,\infty]$, and $\lambda_2^* \geqslant \lambda_1^*$. If $\lambda_1^* < \lambda_2^* < \infty$, we obtain a contradiction from the continuity of $\mathbf{N}$ (by Proposition 2). If $\lambda_1^* = 0$ and $\lambda_2^* < \infty$, then Lemma 4 implies $N_1^* + N_2^* = N_0$, which is again a contradiction. Next, $\lambda_1^* = \lambda_2^* \in [0,\infty)$ corresponds to spin-unpolarized Thomas–Fermi theory in which case $N_1^* = N_2^*$, again a contradiction. Finally, if $\lambda_2^* = \infty$, then by (16), $N_2^* = 0$. Thus all cases have been handled, and we have obtained a contradiction. Thus the image of $\mathbf{N}$ contains the interior of T_1.

The line segment $\{(N_1,N_2)\in T_1: N_1 = N_2\}$ is the image under $\mathbf{N}$ of $\{(\lambda_1,\lambda_2)\in T_0: \lambda_1 = \lambda_2\}$ by standard spin-unpolarized Thomas–Fermi theory. The line segment $\{(N_1,N_2)\in T_1: N_1 + N_2 = N_0\}$ is the image under $\mathbf{N}$ of $\{(\lambda_1,\lambda_2)\in T_0: \lambda_1 = 0\}$ by Lemma 4 and Proposition 2. The line segment $\{(N_1,N_2)\in T_1: N_2 = 0\}$ is the image of $\{(\lambda_1,\lambda_2)\in \overline{T}_0: \lambda_2 = \infty\}$ under the (extension by continuity of the) map $\mathbf{N}$. Lemma 5 now follows. □

We can now complete the proof of Theorem 2. Most of it follows from Proposition 2 and Lemmas 4 and 5. Next we show that if $V(x) \to 0$ as $|x| \to \infty$, and $0 < N_1 + N_2 < N_0$, then the solution densities $\bar{\rho}_1$ and $\bar{\rho}_2$ have compact support. Note that $0 < N_1 + N_2 < N_0$ iff $\lambda_1 > 0$, $\lambda_2 > 0$, where $N_i = N_i(\lambda_1,\lambda_2) = N_i(\lambda)$, for $i = 1,2$. However, $-\Delta(w_\lambda + V) \leqslant 0$ and "$w_\lambda(\infty) = 0$", "$V(\infty) = 0$", whence $w_\lambda \leqslant -V$ a.e. by the maximum principle. Thus $w_\lambda - \lambda_i \leqslant -V - \lambda_i$, and since $V(x) \to 0$ as $|x| \to \infty$, there exists an $R > 0$ such that $w_\lambda(x) - \lambda_i \leqslant 0$ for a.e. x with $|x| > R$. This implies that

$$\bar{\rho}_i(x) = \Gamma(w_\lambda(x) - \lambda_i) = 0$$

for a.e. x with $|x| > R$ and $i = 1,2$. Thus $\bar{\rho}_1$ and $\bar{\rho}_2$ have compact support. Finally we note that when $N_1 + N_2 = N_0$, $N_1 > N_2$ iff $\lambda_1 = 0 < \lambda_2$. Again invoking the maximum principle we see that $w_\lambda \leqslant -V$ and so $w_\lambda - \lambda_2 \leqslant -V - \lambda_2$ (a.e.). The argument of the preceding paragraph shows that $\bar{\rho}_2$ has compact support.

The proof is finally complete. □

The last assertion of Theorem 2 helps to justify the figure in the Introduction depicting the electronic configuration of a neutal nitrogen atom ($c_{ee} = 1$). The density $\bar{\rho}_2$ of spin-down electrons has compact support while the density $\bar{\rho}_1$ is supported on $\mathbb{R}^3$. Thus the spin-down electrons are more tightly bound to the nucleus.

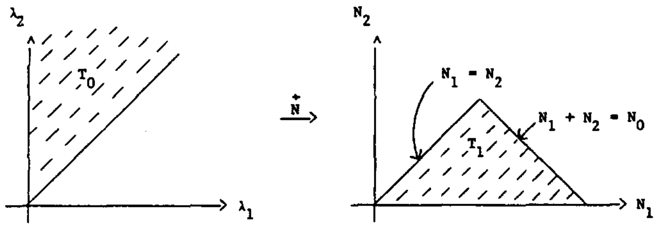


FIG. 1. The image of T_0 under N.

V. MONOTONICITY OF ATOMIC DENSITIES

Let us now consider a generalized atom with Z protons fixed at the origin. We shall show that if ΔV is a radial decreasing function, then the unique solution densities (ρ_1,ρ_2) of the Euler–Lagrange problem are also radial decreasing functions whenever $0 < N \leqslant N_0$. The uniqueness of the solution readily implies that both must be radial functions. The problem is thus to show that ρ_1 and ρ_2 are decreasing. Our result to this effect (Theorem 3 below) is based upon the following result of Gallouët and Morel.[12]

Proposition 3: Let β: $\mathbb{R}\rightarrow\mathbb{R}$ be a continuous nondecreasing function which is increasing on $(0,\infty)$ and satisfies $\beta(0)=0$.

(i) Let $f\in L^1(\mathbb{R}^3)$ and let u be the unique solution of

$$-\Delta u+\beta(u)=f,$$
$$\text{“}u(\infty)=0\text{”},\quad \beta(u)\in L^1(\mathbb{R}^3). \tag{23}$$

If f is radially nonincreasing (resp. decreasing), then u is radially nonincreasing (resp. decreasing).

(ii) Assume further that $x\rightarrow\beta(\pm|x|^{-1})$ is integrable on a neighborhood of the origin in $\mathbb{R}^3$. If $f=a\delta_0+g$ for some $a\geqslant 0$ and $g\in L^1(\mathbb{R}^3)$, where g is radially nonincreasing (resp. decreasing), then the unique solution u of (23) is also radially nonincreasing (resp. decreasing).

Theorem 3: Assume Hypothesis 1 and suppose ΔV is a radially nonincreasing (resp. increasing) function. Then for all $0<N\leqslant N_0$, the solution densities ρ_1,ρ_2 are radially nonincreasing (resp. decreasing) on the sets where they are positive. Furthermore, if Hypothesis 1 is replaced by Hypothesis 2, the same conclusions hold when $\Delta V=a\delta_0+g$, where a is a non-negative constant and $g\in L^1(\mathbb{R}^3)$ is radially nonincreasing (resp. decreasing).

Proof: Assume Hypothesis 1. Then given positive numbers N_1, N_2 with $N_1+N_2\leqslant N_0$, there exists $\lambda=(\lambda_1,\lambda_2)$ in $[0,\infty)^2$ such that the unique densities ρ_1,ρ_2 that solve the Euler–Lagrange problem are given by

$$\rho_i=\Gamma(w_\lambda+\lambda_i),\quad i=1,2,$$

where w_λ is the unique solution of

$$-\Delta w_\lambda+\beta(w_\lambda)=\Delta V,$$
$$\text{“}w_\lambda(\infty)=0\text{”},\quad \beta(w_\lambda)\in L^1(\mathbb{R}^3),$$

where $\beta(w_\lambda)=k\Gamma(w_\lambda-\lambda_1)+k\Gamma(w_\lambda-\lambda_2)$ [see (14)]. Lemma 3 shows that β satisfies the hypotheses of Proposition 3. Since ΔV is radially nonincreasing, so is w_λ by Proposition 3 (i). The strictly increasing assertion of Proposition 3 (i) implies that ρ_i is radially decreasing whenever it is positive. When Hypothesis 2 holds, the above argument goes through by appealing to Proposition 3 (ii). □

ACKNOWLEDGMENTS

We are very grateful to Mel Levy, Rajeev Pathak, and John Perdew for their patient explanations and generous help in our efforts to understand spin-polarized Thomas–Fermi theory.

The first named author gratefully acknowledges the partial support of NSF Grant No. DMS-8620148. The second named author gratefully acknowledges the partial support of the Louisiana Education Quality Support Fund, Contract No. 86-LBR-016-04.

[1]L. H. Thomas, Proc. Cambridge Philos. Soc. **23**, 542 (1927).

[2]E. Fermi, Rend. Acad. Naz. **6**, 602 (1927).

[3](a) E. H. Lieb and B. Simon, Phys. Rev. Lett. **33**, 681 (1973); (b) Adv. Math. **23**, 22 (1977); (c) E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981).

[4]G. L. Oliver and J. P. Perdew, Phys. Rev. A **20**, 397 (1979); O. Gunnarsson, B. I. Lundqvist, and J. W. Wilkins, Phys. Rev. B **10**, 1319 (1974); S. Nordholm, J. Chem. Phys. **86**, 363 (1987); J. P. Perdew, M. Levy, and G. S. Painter, "Chemical bond as a test of density-gradient expansions for kinetic and exchange energies," to appear.

[5]G. V. Gadiyak and Yu. E. Lozovik, J. Phys. B **13**, 1531 (1980).

[6]R. Pathak, Ph.D. thesis, University of Poona, 1982.

[7]J. A. Goldstein and G. R. Rieder, in *Differential Equations in Banach Spaces*, edited by A. Favini and E. Obrecht (Springer, Berlin, 1986), p. 110.

[8]Ph. Bénilan and H. Brezis, "The Thomas–Fermi problem," in preparation; H. Brezis, in *Contemporary Developments in Continuum Mechanics and Partial Differential Equations*, edited by G. M. de la Penha and L. A. Medeiros (North-Holland, Amsterdam, 1978), p. 81; H. Brezis, in *Variational Inequalities and Complementarity Problems: Theory and Applications*, edited by R. W. Cottle, F. Giannessi, and J. L. Lions (Wiley, New York, 1980), p. 53.

[9]E. Fermi and E. Amaldi, Mem. Accad. Ital. **6**, 119 (1934).

[10]G. R. Rieder, Ph.D. thesis, Tulane University, 1986; and an article to appear.

[11]Ph. Bénilan, H. Brezis, and M. G. Crandall, Ann. Scuola Norm. Sup. Pisa **2**, 523 (1975).

[12] Th. Gallouët and J.-M. Morel, Nonlin. Anal. TMA **7**, 971 (1983).

Moment problem formulation of the simplified ideal magnetohydrodynamics ballooning equation

Carlos R. Handy, Daniel Bessis,[a] and Robert M. Williams
Department of Physics, Atlanta University, Atlanta, Georgia 30314

(Received 16 September 1986; accepted for publication 4 November 1987)

A fundamentally new method for determining the eigenvalues of linear differential operators is presented. The method involves the application of moment analysis and affords a fast and precise numerical algorithm for eigenvalue computation, particularly in the intermediate and strong coupling regimes. The most remarkable feature of this approach is that it provides exponentially converging lower and upper bounds to the eigenvalues. The effectiveness of this method is demonstrated by applying it to an important magnetohydrodynamics problem recently studied by Paris, Auby, and Dagazian [J. Math. Phys. 27, 2188 (1986)]. Through the very precise lower and upper bounds obtained, this approach gives full support to their analysis.

I. INTRODUCTION

In a recent work, Paris, Auby, and Dagazian[1] presented a thorough analysis of the simplified ideal magnetohydrodynamics (MHD) ballooning equation given below, based on the earlier work by Antonsen, Ferreira, and Ramos[2]:

$$\frac{d}{dx}\left[(1+x^2)\frac{dy}{dx}\right] - \left[\lambda+\gamma^2(1+x^2)-\frac{\mu^2}{1+x^2}\right]y=0. \qquad (1.1)$$

We will analyze this λ-eigenvalue equation through a fundamentally new approach utilizing a moments equation derivable from Eq. (1.1), together with non-negativity properties of the solutions to Eq. (1.1). In this manner we are able to transform the above system into a true moment problem.[3] Through the use of well-known theorems arising from the traditional "moment problem," it will be seen that a highly effective, simple, and precise numerical algorithm for determining the λ-eigenvalues can be realized. This type of analysis has appeared elsewhere[4–7]; however, the special nature of the present MHD system requires some unprecedented reformulations quite different from those to be found in the cited references.

There are three principal reasons for applying our moment formulaton to Eq. (1.1). First, since our method yields very narrow lower and upper bounds to the eigenvalues, we can unequivocably confirm the results of Paris, Auby, and Dagazian. Second, this approach is simple and readily implementable numerically. Third, few researchers are aware of the generality of this technique. Its dissemination in the context of a physically important problem motivates this work in part.

For simplicity we limit our presentation to the generation of the two lowest $\Lambda(=-\lambda-\frac{1}{4})$-eigenvalues.

[a] On leave of absence from Service de Physique Theorique, CEN-Saclay, France.

II. A SPECIAL TRANSFORMATION

Although Eq. (1.1) is nonsingular along the x axis, it has regular singular points in the complex-x plane at $x=\pm i$. In addition, Eq. (1.1) is defined on the interval $(a=-\infty, b=+\infty)$. At the end points, the physical solution must exhibit "rapid" decrease to zero. In general, on the basis of accumulated empirical data,[4–7] the implementation of a moments analysis appears to be numerically more effective in a representation space in which the number of non-end-point singular points is reduced or completely eliminated. Thus, with respect to Eq. (1.1), consider the transformation

$$z=x/(1+x^2)^{1/2}, \qquad (2.1)$$

or

$$x=z/(1-z^2)^{1/2}. \qquad (2.2)$$

Note that the transformation is invertible and that end points map onto end points. Using

$$\frac{d}{dx}=(1-z^2)^{3/2}\frac{d}{dz},$$

one can transform the original MHD problem to

$$(1-z^2)^3\frac{d^2y}{dz^2}-z(1-z^2)^2\frac{dy}{dz} - [\lambda(1-z^2)+\gamma^2-\mu^2(1-z^2)^2]y=0. \qquad (2.3)$$

Clearly, the new problem is defined on the interval $(a=-1, b=+1)$. Irregular singular points appear at the end points $a=-1$ and $b=1$ (while the singular points at $\pm i$ have been mapped to $\pm i\infty$). As will be seen below, because the physical solution decreases to zero "rapidly" at the transformed end points, the new irregular singular points at $z=\pm 1$ will not affect the exponential convergence of the moments problem analysis.

A simple asymptotic analysis of Eq. (1.1) shows us that the physical solutions must behave as

 0022-2488/88/030717-04$02.50

$$y(x) \underset{|x|\to\infty}{\to} \begin{cases} \exp(-\gamma|x|), & \gamma>0, \\ |x|^{-1/2\pm\sqrt{1/4+\lambda}}, & \gamma=0. \end{cases} \tag{2.4}$$

Accordingly,

$$y(z) \underset{|z|\to 1}{\to} \begin{cases} \exp[-\gamma|z|/(1-z^2)^{1/2}], & \gamma>0, \\ (1-z^2)^{1/4\pm(1/2)\sqrt{1/4+\lambda}}, & \gamma=0. \end{cases} \tag{2.5}$$

Note that from the asymptotic relations, no positive solution can exist for $\lambda < -\frac{1}{4}$.

The moments of the *physical* solutions, $m(p)$, must exist and be finite.

$$m(p) = \int_{-1}^{1} dz\, z^p y(z). \tag{2.6}$$

A recursion relation for these moments follows from Eq. (2.3). In particular, upon multiplying Eq. (2.3) by z^p and integrating by parts over the domain $(-1,1)$, one finds for all γ^2,

$$m(p+4) = \{[-2\mu^2+\lambda+(p+3)(3p+10)]m(p+2) + [-\lambda-\gamma^2+\mu^2-(p+1)(3p+5)]m(p)+p(p-1)m(p-2)\}[(p+5)^2-\mu^2]^{-1}. \tag{2.7}$$

No end point contributions from $y(\pm 1)$, $y'(\pm 1)$ appear because $y(z)$'s end point behavior insures that expressions of the form $z^p(1-z^2)^q y(z)$ and $z^p(1-z^2)^q y'(z)$ $(q\geqslant 1)$ vanish at $z=\pm 1$. Note that this holds for all γ^2 values, including $\gamma=0$!

III. MAKING USE OF THE POSITIVITY PROPERTIES OF THE SOLUTIONS TO THE SCHRÖDINGER EQUATION

In the work of Paris, Auby, and Dagazian,[1] it is shown how Eq. (1.1) and Eq. (2.3) can be transformed into a Schrödinger equation system given by

$$-\frac{d^2\Psi}{d\zeta^2} + q(\zeta)\Psi = \Lambda\Psi, \tag{3.1}$$

where

$$x = \sinh(\zeta), \quad \Psi(\zeta) = (\cosh^{1/2}\zeta) y(x),$$

$$q(\zeta) = \gamma^2\cosh^2\zeta - (\mu^2-\tfrac{1}{4})\operatorname{sech}^2\zeta,$$

and $\Lambda = -\lambda - \frac{1}{4}$. It is known that for such systems, the lowest Λ-eigenvalue corresponds to a positive $(\Psi>0)$ solution.[4] Accordingly, one also has $y_0(z) = S_0(z) > 0$. In addition, because of parity invariance, the next "excited" state (or next higher Λ-eigenvalue) must correspond to a solution with only one zero situated at the origin, $y_1(z) = zS_1(z)$. Because of parity invariance, it also follows $S_i(z)$ $(i=0,1)$ are symmetric in z. This latter observation leads to further simplification of Eq. (2.7). Through a change of variables, one has

$$m_0(2p) = \int_{-1}^{1} dz\, z^{2p} S_0(z), \tag{3.2}$$

$$= \int_0^1 dw\, w^p S_0(w^{1/2})/w^{1/2}, \quad z^2 = w.$$

$$= u_0(p). \tag{3.3}$$

Note that for the above, all the odd-order moments are zero.

For the first excited state we have

$$m_1(2p+1) = \int_{-1}^{1} dz\, z^{2p+1} z S_1(z) \tag{3.4}$$

$$= \int_0^1 dw\, w^p w^{1/2} S_1(w^{1/2})$$

$$= u_1(p). \tag{3.5}$$

Again, note that the even-order moments are zero for $m_1(p)$.

In terms of the $u_i(p)$, $i=0,1$, the moment recursion relation for the "ground" and "first" excited states becomes

$$u_i(p+2) = \left[\begin{Bmatrix} -2\mu^2+\lambda \\ +(2p+3+i)(6p+10+3i) \end{Bmatrix} u_i(p+1) + \begin{Bmatrix} -\lambda-\gamma^2+\mu^2 \\ -(2p+1+i)(6p+5+3i) \end{Bmatrix} u_i(p) + 2p(2p-1+2i)u_i(p-1)\right] \times [(2p+5+i)^2-\mu^2]^{-1}, \tag{3.6}$$

$$u_i(0) = 1. \tag{3.7}$$

Equation (3.7) follows from the arbitrariness of normalization and $S_i(z)$'s positivity. Note that the u-moments are moments of positive function measures $(S_0/w^{1/2}, w^{1/2}S_1)$.

It will be noted that Eq. (3.6) defines a finite difference equation. Once $u_i(1)$ and λ are specified (for fixed μ^2,γ^2 values), all the moments are determined. This is called a "1-missing moment problem."[4] Thus for a 1-missing moment problem, $u_i(1)$ is not known as a function of λ. However, unlike other systems we have examined[4–7] for particular choices of μ^2, Eq. (3.6) actually defines a 0-missing moment problem (where only λ needs to be determined).

Consider the μ^2 values for which the denominator in Eq. (3.6) can vanish. Let $\mu = 2q+5+i$, for some integer q and a chosen value for i (0 or 1). It is known that $S_i(w^{1/2})$ exists and has finite nonzero moments. Thus we must have that the numerator expression for $u_i(q+2)$ be zero. Hence if $\mu = 2q+5+i$, then

$$0 = [-2\mu^2+\lambda+(2q+3+i)(6q+10+3i)]u_i(q+1) + [-\lambda-\gamma^2+\mu^2-(2q+1+i)(6q+5+3i)] \times u_i(q) + 2q(2q-1+2i)u_i(q-1).$$

Only if the above is satisfied will we have $0<u_i(2q+i)<\infty$! Thus an additional constraint on the moments is defined. For given λ, $u_i(1)$ is fixed, and the problem becomes a 0-missing moment problem up to moment order $u_i(q+1)$!

As an example, let $\mu = 5$ $(q = 0, i = 0)$. Then

$$u(1) = (\lambda + \gamma^2 - 20)/(\lambda - 20). \tag{3.8}$$

A second application of the above philosophy leads to an important result. From Eq. (3.8) we see that if a ground state is to exist for $\lambda = 20$, then the finiteness and positivity of $u(1)$ require $\lambda + \gamma^2 - 20 = 0$, or $\gamma^2 = 0$. This is the same result quoted by Paris, Auby, and Dagazian.[1] It should be noted that this is a special case of the general eigenvalue sequence for $\gamma^2 = 0$,

$$\lambda = (\mu - m)(\mu - m - 1), \tag{3.9}$$

where m is a non-negative integer (notice $\lambda \geqslant -\frac{1}{4}$ always). The complete solution in the $\gamma^2 = 0$ case is known (i.e., Refs. 1 and 2) since, for $\gamma^2 = 0$, the relevant equation can be reduced to a form of the hypergeometric equation.

IV. RELEVANT THEOREMS

Handy and Bessis[4] have shown that use of the Hamburger moment theorem,[3] specified below, leads to a rapid algorithm for calculating eigenvalues. Because the system in question, Eq. (2.3), is defined on a finite interval, the results in Ref. 4 need to be appropriately generalized.

We state the Hamburger moment theorem[3]: The necessary and the sufficient conditions for a given set of moments $u(p)$, $p \geqslant 0$, to be the moments of a non-negative function measure, $f(x)$, defined on $(-\infty, +\infty)$, are

$$D_n[u(p)] = \text{Det}\begin{pmatrix} u(0) & u(1) & \cdots & u(n) \\ u(1) & u(2) & \cdots & u(n+1) \\ \vdots & & & \\ u(n) & u(n+1) & \cdots & u(2n) \end{pmatrix} > 0, \quad \text{for all } n \geqslant 0. \tag{4.1}$$

The above are called Hankel–Hadamard determinants.

If a function $f(w)$ is defined on a finite interval $[a,b]$, then the necessary and sufficient conditions for it to be non-negative on $[a,b]$ are obtainable as follows. Let $f_*(w)$ be defined so that

$$f_*(w) = \begin{cases} \text{arbitrary, for } w \notin [a, b], \text{ so long as} \int dx\, x^p f_*(x) \text{ exists,} \\ f(w), \quad \text{if } w \in [a,b]. \end{cases} \tag{4.2}$$

The necessary and sufficient conditions for $f_*(w)$ to be non-negative on $(-\infty, \infty)$, and zero on the complement of $[a,b]$, are that the functions $f_*(w)$, $(w-a)f_*(w)$, and $(b-w)f_*(w)$ *all* be non-negative for $w \in (-\infty, +\infty)$. This is immediately clear. Thus we can say that the necessary and sufficient conditions for $f(w)$ to be non-negative on $[a,b]$ are

$$\left(u(p) = \int_a^b dw\, w^p f(w)\right),$$

$$D_n[u(p)] > 0, \quad D_n[u(p+1) - au(p)] > 0,$$

and

$$D_n[bu(p) - u(p+1)] > 0, \quad \text{for all } n \geqslant 0. \tag{4.3}$$

V. DESCRIPTION OF THE ALGORITHM

All the basic components of the general moments approach have been defined. Thus, for either the ground or first excited state ($i = 0,1$, respectively), one chooses fixed values for μ^2 and γ^2. From moment recursion relations in Eq. (3.6) one can readily generate the first P moments, $u_i(p)$ $(1 \leqslant p \leqslant P)$. They will be polynomial functions of λ and $u_i(1) = u$. Accordingly, the Hankel–Hadamard determinants in Eq. (4.3) will also become polynomials in λ and u. Note that from the w-integrations in Eq. (3.2) and Eq. (3.4), the appropriate choices for the a and b parameters are $a = 0$ and $b = 1$. Thus we have

$$D_n[u_i(p)] = \sum_{k=0}^{(n+1-\delta_{0,n})} C_{i,n}^{(k)}(\lambda) u^k, \tag{5.1}$$

$$D_n[u_i(p+1)] = \sum_{k=0}^{n+1} \tilde{C}_{i,n}^{(k)}(\lambda) u^k, \tag{5.2}$$

$$D_n[u_i(p) - u_i(p+1)] = \sum_{k=0}^{n+1} \tilde{\tilde{C}}_{i,n}^{(k)}(\lambda) u^k. \tag{5.3}$$

The specific numerical algorithm proceeds as follows. An arbitrary λ interval is specified $[\alpha,\beta]$. A sufficiently narrow partitioning is defined. At each λ point the polynomial determinants defined above are determined. That is, the C coefficients are numerically determined. The location of the real u roots are determined. In this manner one can assess if any u-space subregions exist satisfying the Hankel–Hadamard inequalities of Eq. (4.3) (only those determinants involving moments of order at most P are considered). If such u-space subregions exist, then the associated λ-partition point is a *possible* physical value. If no u-space subregions exist, then for that specific λ-partition value one can say that it is not a possible physical value. In this manner both lower and upper bounds to $\lambda_{\text{physical}}$ are determined. The results are given in the various tables.

VI. CONCLUSION

We have presented a simple and numerically effective technique for eigenvalue determination of linear differential systems. Our results confirm the analysis of Paris, Auby, and

TABLE I. Bounds for the ground state eigenvalue.

μ^2	γ^2	Max. order of moments used, P	Lower λ bound	Upper λ bound
0	50	3	− 58.2	− 56.9
		4	− 57.90	− 57.27
		5	− 57.84	− 57.71
		6	− 57.831	− 57.792
		7	− 57.805	− 57.797
		8	− 57.800 3	− 57.797 8
		9	− 57.800 0	− 57.798 1
0	5	9	− 7.942	− 7.923
		12	− 7.929 5	− 7.926 4
		15	− 7.929 3	− 7.928 5
		18	− 7.929 18	− 7.929 02
0	1	12	− 2.73	− 2.65
		18	− 2.660	− 2.652
		24	− 2.654 48	− 2.653 85
		28	− 2.654 36	− 2.653 98
0	0.5	12	− 2.0	− 1.7
		18	− 1.87	− 1.83
		24	− 1.848	− 1.839

Dagazian[1] for those specific μ^2, γ^2 values quoted in the tables. These suggest that the overall analysis of Paris, Auby, and Dagazian is reliable. Our approach yields unequivocal narrow bounds for $\lambda_{\text{physical}}$. As noted elsewhere,[4–7] a moments analysis is specially designed to handle intermediate and strong coupling problems. This is readily apparent from Tables I–III. The larger γ^2 is, the faster is the rate of convergence of our bounds. Also note that in our method, which parameters are fixed and which varied is inconsequential. We have adopted the point of view of Paris, Auby, and Dagazian in treating λ as the undetermined parameter. We could have just as easily switched things around and kept λ fixed, while varying γ^2.

TABLE II. Bounds for the ground state eigenvalue.

μ^2	γ^2	Max. order of moments used, P	Lower λ bound	Upper λ bound
1	50	3	− 57.3	− 55.9
		7	− 56.87	− 56.85
		9	− 56.859 2	− 56.856 8
1	5	10	− 7.08	− 7.06
		15	− 7.070	− 7.068
		18	− 7.069 4	− 7.069 2
1	1	10	− 2.1	− 1.8
		12	− 1.94	− 1.87
		15	− 1.893	− 1.875
		20	− 1.880	− 1.876
		24	− 1.877 8	− 1.877 5
1	0.5	5	− 2.4	6.2
		10	− 1.4	− 0.6
		15	− 1.16	− 1.09
		20	− 1.116	− 1.102
		24	− 1.107	− 1.104

TABLE III. Bounds for the first excited state eigenvalue.

μ^2	γ^2	Max. order of moments used, P	Lower λ bound	Upper λ bound
0	50	3	− 75.0	− 71.1
		7	− 72.90	− 72.84
		9	− 72.858	− 72.854
0	5	12	− 13.21	− 13.18
		17	− 13.204 3	− 13.202 7
0	1	15	− 5.46	− 5.32
		18	− 5.38	− 5.33
		22	− 5.350	− 5.340
0	0.5	15	− 4.2	− 3.8
		24	− 3.92	− 3.88
1	50	3	− 74.0	− 70.0
		7	− 72.06	− 72.01
		9	− 72.017	− 72.013
1	5	13	− 12.542	− 12.530
		17	− 12.541 4	− 12.540 0
1	1	10	− 5.5	− 4.7
		18	− 4.86	− 4.82
		22	− 4.836	− 4.827
1	0.5	15	− 3.8	− 3.4
		24	− 3.46	− 3.43

During the space of time since the original communication of this work, important developments have transpired which further the implementation of the Hankel–Hadamard moments approach. In particular, it is possible to develop an equivalent linear formulation of the nonlinear Hankel–Hadamard theory.[8,9] This linearization allows us to use linear programming methods to solve missing moment problems of any order. Such methods can be used in the present case.

ACKNOWLEDGMENTS

C. R. Handy was supported by Grant No. DOE DE-AC05-840R21400. R. M. Williams' research was performed under appointment to the Historically Black Colleges and Universities Nuclear Energy Training program administered by Oak Ridge Associated Universities for the U. S. Department of Energy. We gratefully acknowledge the support of the cited agencies. In addition, we appreciate the recommendations suggested by the reviewer.

[1]R. B. Paris, N. Auby, and R. Y. Dagazian, J. Math. Phys. **27**, 2188 (1986).
[2]T. M. Antonsen, A. Ferreira, and J. J. Ramos, Plasma Phys. **24**, 197 (1982).
[3]J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (Am. Math. Soc., Providence, RI, 1963).
[4]C. R. Handy, and D. J. Bessis, Phys. Rev. Lett. **55**, 931 (1985).
[5]C. R. Handy, J. Phys. A: Math. Gen. **18**, 3593 (1985).
[6]D. J. Bessis and C. R. Handy, Sanibel Symposium on Quantum Chemistry 20, Int. J. Quantum Chem. (1986).
[7]D. J. Bessis, E. R. Vrscay, and C. R. Handy, J. Phys. A Gen. Phys. **20**, 419 (1986).
[8]C. R. Handy, D. J. Bessis, and T. D. Morley, to be published in Phys. Rev. A.
[9]C. R. Handy, D. J. Bessis, G. Siginmondi, and T. D. Morley, to be published in Phys. Rev. Lett.

Erratum: The diffraction of waves by a penetrable ribbon [J. Math. Phys. 4, 65 (1963)]

C. Yeh and C. S. Kim
Electrical Engineering Department, University of California at Los Angeles, California 90024

(Received 9 December 1987; accepted for publication 16 December 1987)

The problem of the scattering of electromagnetic waves by an elliptical dielectric cylinder was formulated and solved in the original paper. Numerical examples were also presented there. Recently, we discovered a typographical error in the computer program affecting the presented numerical results. The purpose of this erratum is to provide the corrected numerical results. They are shown in Figs. 2–5 and in Table I.

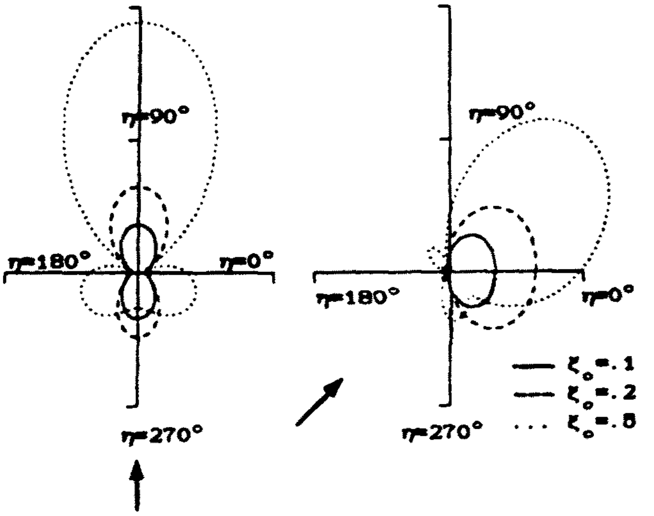


FIG. 2. Polar diagrams for waves ($|H^s_\eta|$) scattered by a dielectric ribbon with $k_0^2q^2 = 10$. The incident electric vector is polarized in the axial direction. (Arrows indicate the direction of incident waves.)

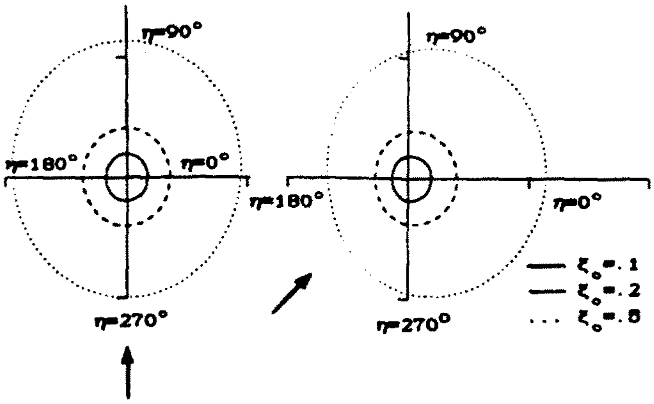


FIG. 3. Polar diagrams for waves ($|H^s_\eta|$) scattered by a dielectric ribbon with $k_0^2q^2 = 1.0$. The incident electric vector is polarized in the axial direction. (Arrows indicate the direction of incident waves.)

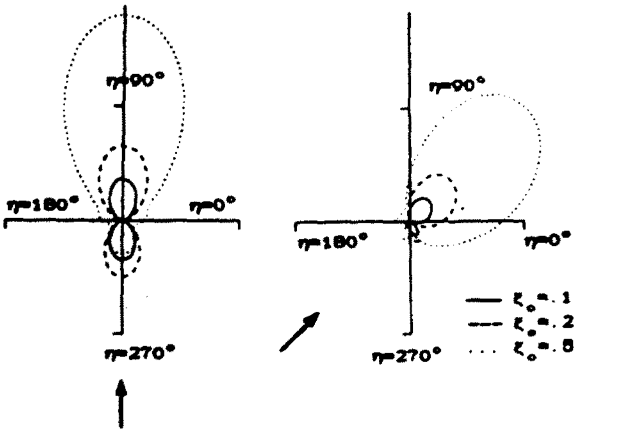


FIG. 4. Polar diagrams for waves ($|E^s_\eta|$) scattered by a dielectric ribbon with $k_0^2q^2 = 10$. The incident magnetic vector is polarized in the axial direction. (Arrows indicate the direction of incident waves.)

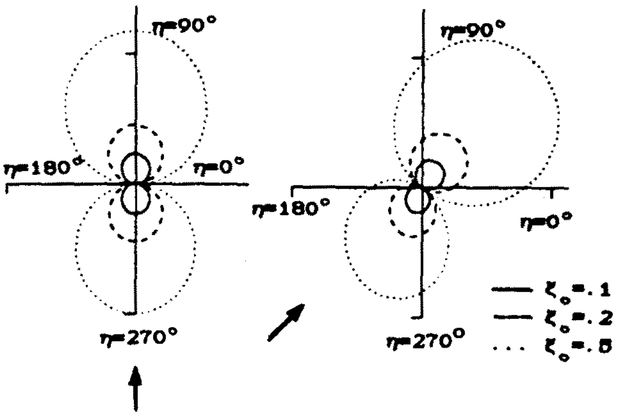


FIG. 5. Polar diagrams for waves ($|E^s_\eta|$) scattered by a dielectric ribbon with $k_0^2q^2 = 1.0$. The incident magnetic vector is polarized in the axial direction. (Arrows indicate the direction of incident waves.)

 0022-2488/88/030721-02$02.50

TABLE I. The rate of convergence for $\xi_0 = 0.2$, $k_0 q = (10)^{1/2}$, and $\theta = 90^{\circ}$.

n	A_{2n} $m=2$	A_{2n} $m=3$	A_{2n} $m=4$	A_{2n+1} $m=2$	A_{2n+1} $m=3$	A_{2n+1} $m=4$
0	-0.151	-0.167	-0.166	-0.254	-0.261	-0.261
	$+0.328i$	$+0.333i$	$+0.333i$	$+0.318i$	$+0.326i$	$+0.326i$
1	$+0.137\times10^{-1}$	$+0.678\times10^{-1}$	$+0.675\times10^{-1}$	$+0.400\times10^{-1}$	$+0.408\times10^{-1}$	$+0.408\times10^{-1}$
	$+0.212i$	$+0.126i$	$+0.127i$	$-0.142\times10^{-1}i$	$-0.205\times10^{-1}i$	$-0.205\times10^{-1}i$
2		-0.107×10^{-2}	-0.105×10^{-2}		-0.307×10^{-3}	-0.307×10^{-3}
		$-0.109\times10^{-1}i$	$-0.111\times10^{-1}i$		$-0.261\times10^{-3}i$	$-0.265\times10^{-3}i$
3			$+0.330\times10^{-5}$			$+0.443\times10^{-6}$
			$+0.271\times10^{-4}i$			$+0.791\times10^{-7}i$

n	B_{2n+2} $m=2$	B_{2n+2} $m=3$	B_{2n+2} $m=4$	B_{2n+1} $m=2$	B_{2n+1} $m=3$	B_{2n+1} $m=4$
0	-0.405×10^{-3}	-0.405×10^{-3}	-0.405×10^{-3}	-0.375×10^{-2}	-0.375×10^{-2}	-0.375×10^{-2}
	$+0.192\times10^{-1}i$	$+0.192\times10^{-1}i$	$+0.192\times10^{-1}i$	$+0.593\times10^{-1}i$	$+0.594\times10^{-1}i$	$+0.594\times10^{-1}i$
1	$+0.152\times10^{-4}$	$+0.153\times10^{-4}$	$+0.153\times10^{-4}$	$+0.373\times10^{-3}$	$+0.375\times10^{-3}$	$+0.375\times10^{-3}$
	$-0.440\times10^{-3}i$	$-0.469\times10^{-3}i$	$-0.469\times10^{-3}i$	$-0.216\times10^{-2}i$	$-0.255\times10^{-2}i$	$-0.255\times10^{-2}i$
2		-0.421×10^{-7}	-0.421×10^{-7}		-0.365×10^{-5}	-0.365×10^{-5}
		$-0.172\times10^{-5}i$	$-0.175\times10^{-5}i$		$-0.427\times10^{-4}i$	$-0.438\times10^{-4}i$
3			$+0.266\times10^{-10}$			$+0.534\times10^{-8}$
			$-0.105\times10^{-8}i$			$-0.571\times10^{-7}i$