Quadratic alternative algebras

Alberto Elduque
Departamento de Matemáticas, Escuela Técnica Superior de Ingenieros Industriales, Universidad de Zaragoza, 50015 Zaragoza, Spain

(Received 28 February 1989; accepted for publication 26 July 1989)

Quadratic alternative algebras are completely classified, up to isomorphism, by means of the associated set of "vectors." They include the quaternion and octonion algebras, but also many other nonassociative algebras used in physics.

I. INTRODUCTION

Quadratic algebras, other than quaternion and octonion algebras, have been recently studied, in connection with some physical questions, by Jantzen,[1] Plebanski and Przanowski,[2,3] Ilamed and Salingaros,[4] Domokos and Kövesi-Domokos,[5,6] Wene,[7–10] and others.

The aim of this paper is to unify and cover most of the "algebraic" results of these authors, providing a description of a broad class of quadratic algebras, namely, all the quadratic alternative algebras.

Any quadratic algebra over a field of characteristic not 2 can be decomposed into a direct sum of the field and an anticommutative algebra equipped with a bilinear form. Hence the properties satisfied by the algebra are closely related to the properties of the anticommutative algebra and the bilinear form. If the quadratic algebra is alternative then the structure of the anticommutative algebra is severely restricted. This will be used to describe these algebras. We shall recover in this way some classical results on nondegenerate quadratic forms permitting composition and results of Kunze and Scheinberg[11] on alternative algebras with scalar involution.

II. QUADRATIC ALGEBRAS

By a quadratic algebra A, over the field F (F will be either $\mathbb{R}$ or $\mathbb{C}$, but everything will be valid over arbitrary fields of characteristic not 2), we shall understand any nonassociative algebra with an identity element 1 and such that 1, a, and a^2 are linearly dependent for any a in A.

Following Osborn,[12] if A is a quadratic algebra over F (multiplication denoted by juxtaposition) and V is the set of "vectors" [that is, the set of the elements x in A such that $x \notin F$ but $x^2 \in F$ (here we identify F with $F1$)], then V is a subspace of A and A decomposes as a direct sum $A = F \oplus V$.

If $x,y \in V$, we write

$$xy = -(x,y) + x \cdot y, \tag{1}$$

where $(x,y) \in F$ and $x \cdot y \in V$. Clearly $(\ ,\ )$ is a bilinear form on V (not necessarily symmetric) and $\cdot$ is an anticommutative product on V. The multiplication in A is then given by

$$(\alpha + x)(\beta + y) = (\alpha\beta - (x,y)) + (\alpha y + \beta x + x \cdot y), \tag{2}$$

for α and β in F and x, y in V.

Notice that we have added a minus sign to the bilinear form $(\ ,\ )$ as it appears in Ref. 12. Usually, for any element a in a quadratic algebra, one writes

$$a^2 - t(a)a + n(a) = 0, \tag{3}$$

with $t(a)$ and $n(a)$ in F.[13]

Now, if $x \in V$, then $x^2 = -(x,x)$, and thus $t(x) = 0$ and $n(x) = (x,x)$; that is, the quadratic form associated to $(\ ,\ )$ coincides with the restriction to V of the quadratic form n. Moreover

$$(\alpha + x)^2 = (\alpha^2 - (x,x)) + 2\alpha x$$

so

$$t(\alpha + x) = 2\alpha,\ n(\alpha + x) = \alpha^2 + (x,x).$$

We list in the next proposition some properties of quadratic algebras that will be useful in the sequel.

Proposition 2.1: Let A be a quadratic algebra over F. Then we have the following.

(i) The map $a \to t(a)1 - a$ is an involution of A if and only if $(\ ,\ )$ is symmetric.

(ii) A is flexible [that is, $(xy)x = x(yx)$, for all x, y in A] if and only if $(\ ,\ )$ is symmetric and invariant [that is, $(x \cdot y, z) = (x, y \cdot z)$, for all x, y, z in V].

(iii) If A is a division algebra then for all x ($\neq 0$) in V we have $(x,x) \neq 0, -1$ and for all linearly independent x, y in V, x, y, and $x \cdot y$ are linearly independent. Moreover, for finite-dimensional algebras these conditions are also sufficient.

(iv) If n is nondegenerate [equivalently, if the symmetric bilinear form on V given by $(x,y) + (y,x)$ is nondegenerate] then either A is simple or A is isomorphic to $F \oplus F$. If A is flexible and simple, then n is nondegenerate [equivalently $(\ ,\ )$ is nondegenerate on V].

(v) A is associative if and only if it is flexible and the identity of the double cross product

$$(x \cdot y) \cdot z = (x,z)y - (y,z)x \tag{4}$$

is satisfied.

(vi) A is alternative [that is, $(xy)y = xy^2$ and $y(yx) = y^2x$, for all x, y in A] if and only if it is flexible and the weak identity of the double cross product

$$(x \cdot y) \cdot y = (x,y)y - (y,y)x \tag{5}$$

is satisfied.

(vii) n admits composition [$n(ab) = n(a)n(b)$] if and only if A is flexible and

$$(x,x)(y,y) = (x,y)^2 + (x \cdot y, x \cdot y). \tag{6}$$

In particular, if A is alternative, then n admits composition.

(viii) A is a form of the algebra of color[5] if and only if A is flexible, n is nondegenerate, for all x, y in V:

$$((x \cdot y) \cdot y) \cdot y = \tfrac{1}{2}(y,y)(x \cdot y), \tag{7}$$

 0022-2488/90/010001-05$03.00

and there are elements $x,y \in V$ such that x,y and $(x\cdot y)\cdot y$ are linearly independent.

Proof: The assertions (ii) and (iii) have been proved by Osborn[12]; (i), (v), and (vi) are straightforward[14,15]; and (viii) has been proved by the author.[16]

With respect to (iv) we can proceed as follows: Let I be a nontrivial ideal of A and let $\alpha + x$ be a nonzero element in I. Since the bilinear form $(u,v)^+ = (u,v) + (v,u)$ is nondegenerate, $(x,y)^+ \neq 0$ for some $y \in V$, and hence if $\alpha = 0$ then

$$0 \neq xy + yx = -(x,y)^+ \in I.$$

Thus we may assume that $1 + x \in I$. Now,

$$(1+x)x = -(x,x) + x \in I,$$

so $1 + (x,x) \in I$, but $I \neq A$, so $(x,x) = -1$. If V has dimension greater than 1, then we take y in V with $(x,y)^+ = 0$, $(y,y) \neq 0$. Then

$$(1+x)y + y(1+x) = 2y \in I$$

and $y^2 = -(y,y) \in I$. Thus $I = A$, a contradiction. If $\dim V = 1$, $V = Fx$, and $(x,x) = -1$, then $F(1+x)$ is an ideal of A and A is isomorphic to $F \oplus F$.

If A is flexible, then $(\ ,\ )$ is symmetric and invariant, so its radical (the set $\{x \in V: (V,x) = 0\}$) is an ideal of A. Hence if A is flexible and simple, then $(\ ,\ )$ is nondegenerate.

For (vii), consider $a = \alpha + x$ and $b = \beta + y$. Then

$$\begin{aligned} n(ab) - n(a)n(b) = &((x,y)^2 + (x\cdot y, x\cdot y) - (x,x)(y,y)) \\ &+ \alpha\beta((y,x) - (x,y)) \\ &+ \alpha((y,x\cdot y) + (x\cdot y,y)) \\ &+ \beta((x,x\cdot y) + (x\cdot y,x)). \end{aligned} \tag{8}$$

Then $n(ab) = n(a)n(b)$, for all a, b in A, if and only if, for all x, y in V, we have

$$(x,x)(y,y) = (x,y)^2 + (x\cdot y, x\cdot y),$$

$$(x,y) = (y,x), \quad (x\cdot y, y) = 0.$$

The last two conditions are equivalent to flexibility. In particular, if A is alternative,

$$\begin{aligned} (x\cdot y, x\cdot y) &= -((x\cdot y)\cdot y, x) \\ &= -((x,y)y - (y,y)x, x), \end{aligned}$$

so

$$(x,x)(y,y) = (x,y)^2 + (x\cdot y, x\cdot y)$$

and n admits composition.

A sufficient condition for the simplicity of A was shown by Sagle.[17] Okubo[18] uses, as in (iv), the nondegeneracy of a bilinear form to obtain simple algebras in a more general setting (see also Ref. 19, Theorem 4.27), but its proof fails for V of dimension 1. A proof for the more general case considered by Okubo may be modeled on the proof above.

If A is a quadratic flexible algebra and N is the radical of $(\ ,\ )$, then the invariance of $(\ ,\ )$ forces N to be an ideal of A. Moreover, Proposition 2.1 (iv) implies that either A/N is simple or isomorphic to $F \oplus F$.

In general, it is not true that if the quadratic algebra A is simple, then n is nondegenerate, as shown by the quadratic algebra determined by $V = \mathrm{sl}(2,F)$ and $(\ ,\ )$ any nonzero degenerate symmetric bilinear form on V.

There are examples of quadratic algebras that are not alternative although the quadratic form n admits composition (n is degenerate in those cases).[20]

Notice that if the dimension of A is finite, say m, and A is alternative, then (vi) shows that

$$\mathrm{Trace}((R_x)^2) = -(m-1)(x,x),$$

so we have that $(\ ,\ )$ equals $-1/(m-1)$ times the Killing form of V [defined as $\{x,y\} = \mathrm{Trace}(R_x R_y)$]. Here R_x denotes the right multiplication in V by the element x.

A Malcev algebra is a nonassociative algebra satisfying the identities $x^2 = 0$ and

$$(xy)(xz) = ((xy)z)x + ((yz)x)x + ((zx)x)y.$$

Any Lie algebra is a Malcev algebra and if A is an alternative algebra, with multiplication denoted by juxtaposition, then the new algebra defined on A by considering the product $[x,y] = xy - yx$ is a Malcev algebra. Thus if A is a quadratic alternative algebra, then V with the multiplication $\cdot$ is a Malcev algebra. Moreover, any central simple non-Lie Malcev algebra appears as the anticommutative algebra V associated to an octonion algebra. Incidentally, notice that (v) [resp. (vi)] proves that any anticommutative algebra equipped with a symmetric invariant bilinear form satisfying (4) [resp. (5)] is a Lie (resp. Malcev) algebra (see, also, Ref. 21, Corollary 2.1).

Over arbitrary fields of characteristic 3, any anticommutative algebra satisfying (5) is a Lie algebra. In any other characteristic, an anticommutative algebra satisfying (5) is a Lie algebra if and only if it satisfies (4), too.

In the next section we shall study the anticommutative algebras over $\mathbb{C}$ that appear as the algebra of vectors of a quadratic alternative algebra. The method of working will use almost exclusively very elementary linear algebra. Finally, in Sec. IV, all quadratic alternative algebras will be described, recovering in a unified way the classical generalized theorem of Hurwitz and the results of Kunze and Scheinberg.[11]

III. ANTICOMMUTATIVE ALGEBRAS VERIFYING (5) OVER C

We shall assume in this section that $F = \mathbb{C}$ (or any algebraically closed field of characteristic not 2) and V is an anticommutative algebra equipped with a symmetric invariant bilinear form so that (5) is satisfied. For any element x in V, R_x will denote the map $y \to y\cdot x$. The linear span of all products of n elements in V (in any order of parentheses) will be denoted by V^n.

Proposition 3.1: The bilinear form $(\ ,\ )$ is trivial if and only if V is a two-engelian Malcev algebra [that is, $(R_x)^2 = 0$, for all $x \in V$]. In this case $V^4 = 0$ and V satisfies (4) if and only if $V^3 = 0$.

Proof: The first and last assertions are obvious. Since $(V,V) = 0$, by linearizing (5) we get

$$(x\cdot y)\cdot z = (z\cdot x)\cdot y \quad \forall x,y,z \in V.$$

Then

$$((x\cdot y)\cdot z)\cdot t = (t\cdot(x\cdot y))\cdot z = (y\cdot(t\cdot x))\cdot z = (z\cdot y)\cdot(t\cdot x),$$

but

$$((x\cdot y)\cdot z)\cdot t = ((y\cdot z)\cdot x)\cdot t = (x\cdot t)\cdot(y\cdot z) = -(z\cdot y)\cdot(t\cdot x)$$

so

$$((x\cdot y)\cdot z)\cdot t = (z\cdot y)\cdot(t\cdot x) = 0$$

and therefore $V^4 = 0$.

The fact that any two-engelian Malcev algebra is nilpotent of class $\leqslant 3$ ($V^4 = 0$) follows also from some results of Filippov.[22]

Corollary 3.2: The radical N of $(\ ,\ )$ in any quadratic alternative algebra A is the only maximal nilpotent ideal of A, it is of class of nilpotency at most 3, and the quotient algebra is either simple or isomorphic to $F\oplus F$.

Let us assume now that $(\ ,\ )\neq 0$; thus there is an element $h\in V$ with $(h,h) = -1$. Then from (5) we get $(R_h)^3 = R_h$, so $V = V_0\oplus V_1\oplus V_{-1}$, where

$$V_i = \{x\in V: xR_h = ix\}.$$

Now if $z\in V_0$, then $0 = (z\cdot h)\cdot h = (z,h)h + z$; thus $V_0 = Fh$. Moreover

$$(V_0, V_1 + V_{-1}) = (V_0, h\cdot(V_1 + V_{-1})) = (V_0\cdot h, V_1 + V_{-1}) = 0,$$

so $(V_0)^\perp = V_1 + V_{-1}$.

If $x,y\in V_1$, then

$$(x,y) = (x\cdot h,y) = (x,h\cdot y) = -(x,y);$$

thus $(V_1,V_1) = 0$ and also $(V_{-1},V_{-1}) = 0$. Besides, $(x\cdot y)\cdot h + x\cdot y = 0$, so $(V_1)^2\subseteq V_{-1}$ and $(V_{-1})^2\subseteq V_1$.

If $x\in V_1$ and $y\in V_{-1}$, then $(x\cdot y)\cdot h + x\cdot y = (x,y)h$ and $(y\cdot x)\cdot h - y\cdot x = (y,x)h$; hence $x\cdot y = (x,y)h$.

Three cases appear.

(A) $N = V_1 + V_{-1}$. In this case $V_1\cdot V_{-1} = 0$ and, for x,y in V_1 and z in V_{-1}, we have $(x\cdot y)\cdot z = 0$ by the linearization of (5). We therefore have $N^3 = 0$. Moreover, if V satisfies (4), then for $x,y\in V_1$, $(x\cdot y)\cdot h = 0$, so $N^2 = 0$.

(B) There exist elements $e\in V_1$ and $f\in V_{-1}$ such that $(e,f) = \frac{1}{2}$, $V_1 = Fe + N\cap V_1$, and $V_{-1} = Ff + N\cap V_{-1}$.

In this case, if $x,y\in N\cap V_1$, $(x\cdot y)\cdot f = 0$ by (5), but

$$0 = ((x\cdot y)\cdot f)\cdot e = -2(e,f)x\cdot y = -x\cdot y.$$

Hence $N^2 = 0$. Moreover, the subalgebra generated by e, f, and h is isomorphic to $\mathrm{sl}(2,F)$ and if $\{u_i: i\in I\}$ is a basis of $N\cap V_1$ and $v_i = u_i\cdot e$, then $\{v_i: i\in I\}$ is a basis of $N\cap V_{-1}$, so $N = \oplus\{Fu_i + Fv_i: i\in I\}$.

There is only one, up to isomorphism, irreducible module for $\mathrm{sl}(2,F)$ considered as a Malcev algebra, that is not a module for $\mathrm{sl}(2,F)$ as a Lie algebra.[23] This module has dimension 2.

Since

$$u_i\cdot f = (u_i,f)h = 0 = v_i\cdot e,\quad u_i\cdot e = v_i,$$

$$v_i\cdot f = (u_i\cdot e)\cdot f = -2(e,f)u_i = -u_i,$$

we check that $Fu_i + Fv_i$ is precisely this non-Lie module [notice that $u_i\cdot(e\cdot f) = u_i$, but $(u_i\cdot e)\cdot f - (u_i\cdot f)\cdot e = -u_i$, so $Fu_i + Fv_i$ is not a Lie module for $\mathrm{sl}(2,F)$].

If V satisfies (4) and $x\in N\cap V_1$, then $(x\cdot e)\cdot h = 0$, so $x\cdot e = 0$, but $0 = (x\cdot e)\cdot f = -(e,f)x$. Hence $N = 0$.

(C) There are elements $e_1,e_2\in V_1$ and $f_1,f_2\in V_{-1}$ with $(e_i,f_j) = \frac{1}{2}\delta_{ij}$ $(i,j = 1,2)$.

In this case we write $e_3 = f_1\cdot f_2$ and $f_3 = e_1\cdot e_2$. The invariance of $(\ ,\ )$ and (5) show that $e_i\cdot e_j = \epsilon_{ijk}f_k$, $f_i\cdot f_j = \epsilon_{ijk}e_k$ and $(e_i,f_j) = \frac{1}{2}\delta_{ij}$ $(i,j = 1,2,3)$, where ϵ_{ijk} is the skew-symmetric tensor of Levi-Civita ($\epsilon_{123} = 1$). Now, if $z\in V_1$ and $(z,f_i) = 0$, $i = 1,2,3$, then

$$e_3\cdot z = (f_1\cdot f_2)\cdot z = -(f_1\cdot z)\cdot f_2 = 0,$$

but

$$0 = (z\cdot e_3)\cdot f_3 = -2(e_3,f_3)z = -z.$$

Therefore, in this case, $N = 0$ and V is, up to isomorphism, the only simple non-Lie Malcev algebra over $\mathbb{C}$.[24]

Summarizing these we get the following theorem.

Theorem 3.3: Let V be an anticommutative algebra over $\mathbb{C}$ equipped with a symmetric invariant bilinear form satisfying (5). Then either (a) V is a two-engelian Malcev algebra; (b) there are vector subspaces M_i, N_i, $i = \pm 1$, and skew-symmetric mappings ϕ_i; $M_i\times M_i\to N_{-i}$ $(i = \pm 1)$ such that $V = \mathbb{C}h\oplus M_1\oplus N_1\oplus M_{-1}\oplus N_{-1}$ and the multiplication in V is given by

$$m_i\cdot h = im_i,\quad n_i\cdot h = in_i,$$
$$n_i\cdot M_j = n_i\cdot N_j = m_i\cdot N_j = M_i\cdot M_{-i} = 0, \tag{9}$$

and

$$m_i\cdot m_i' = \phi_i(m_i,m_i'), \tag{10}$$

for $m_i,m_i'\in M_i$, $n_i\in N_i$, $i,j = \pm 1$; (c) $V = \mathrm{sl}(2,\mathbb{C})\oplus(\oplus\{N_i: i\in I\})$, where I is an indexing set, $N = \oplus\{N_i: i\in I\}$ is Abelian, and each N_i is a non-Lie Malcev irreducible module for $\mathrm{sl}(2,\mathbb{C})$; (d) V is the simple non-Lie Malcev algebra over $\mathbb{C}$.

Conversely, all the algebras listed here satisfy (5). Moreover, V satisfies (4) if and only if either (a′) V is a nilpotent Lie algebra of class $\leqslant 2$; (b′) there are vector subspaces N_i, $i = \pm 1$, such that $V = \mathbb{C}h\oplus N_1\oplus N_{-1}$ and the multiplication is given by $n_i\cdot h = in_i$, $N_i\cdot N_j = 0$, $i,j = \pm 1$; or (c′) V is isomorphic to $\mathrm{sl}(2,\mathbb{C})$.

Notice that in all cases the algebra splits over the ideal N. This could have been easily proved without the reasoning before Theorem 3.3. It should also be remarked that Theorem 3.3 is valid over any field F provided either $(\ ,\ ) = 0$ or there is $h\in V$ with $0\neq -(h,h)\in F^2$.

IV. QUADRATIC ALTERNATIVE ALGEBRAS

In this section A will denote a quadratic alternative algebra over $\mathbb{R}$ or $\mathbb{C}$ (or, with some obvious changes, over any arbitrary field F, char $F\neq 2$). Then $A = F\oplus V$ and the algebra V satisfies (5). If either $F = \mathbb{C}$ or $F = \mathbb{R}$ and V contains an element x such that $(x,x) < 0$, then everything works as in Sec. III and A is completely determined. In the case $F = \mathbb{R}$, by extending scalars up to $\mathbb{C}$ we see that the dimension of the quotient of A, by the radical of $(\ ,\ )$, is 1, 2, 4, or 8. In the latter case, $A\otimes_{\mathbb{R}}\mathbb{C}$ is the split Cayley–Dickson algebra over

$\mathbb{C}$,[15] so A is an octonion algebra and we are done.

So we have to pay attention to the cases in which $F=\mathbb{R}$, the dimension of A/N is 2 or 4 and A does not contain any element $x\in V$ with $(x,x)<0$.

If the dimension of A/N is 2 and e is an element of V with $(e,e)=1$ then $\mathbb{R}+\mathbb{R}e$ is isomorphic to $\mathbb{C}$. We will identify both fields. Let σ be the conjugate operator of $\mathbb{C}$ over $\mathbb{R}$ ($1^\sigma=1$, $e^\sigma=-e$). Then $A=\mathbb{C}\oplus N$ and $N^3=0$.

In this case, for $x,y\in N$,

$$(xy)e=(x\cdot y)\cdot e=(e\cdot x)\cdot y\in N^2,$$

so N and N^2 are vector spaces over $\mathbb{C}$ (action on the right). For $a=\alpha+\beta e$ in $\mathbb{C}$, x in N, we have $ax=xa^\sigma$. Let S be a supplementary subspace to N^2 in N. The multiplication in N determines a skew-symmetric bilinear mapping φ: $S\times S\to N^2$. Now, for $u,v\in S$, $(u\cdot e)\cdot v=-(u\cdot v)\cdot e$, so for $a\in\mathbb{C}$, $\varphi(ua,v)=\varphi(u,v)a^\sigma$.

It is immediate to see that this construction gives an alternative algebra.

Assume now that the dimension of A/N is 4 and that e_1 and e_2 are orthogonal elements in V with $(e_i,e_i)=1, i=1,2$. Then $Q=F1+Fe_1+Fe_2+F(e_1\cdot e_2)$ is the quaternion division algebra, $A=Q\oplus N$, and $N^2=0$. For $u\in N$, $a,b\in Q\cap V$, we have

$$(ua)b=(u\cdot a)\cdot b=-(a\cdot u)\cdot b=(a\cdot b)\cdot u-(a,b)u$$
$$=u(ba),$$

so N is a vector space over the division algebra Q^{op} (the opposite algebra of Q). On the left, if $q\in Q$, $u\in N$, then $qu=uq^\sigma$, where σ is the standard involution in Q. In this way A is completely determined.

Gathering together these results and the ones in Sec. III we obtain the following theorem.

Theorem 4.1: Let A be a nonassociative algebra over F. Then A is a quadratic alternative algebra if and only if one of the following holds.

(i) $A=F\oplus V$, with V a two-engelian Malcev algebra and multiplication given by $(\alpha+v)(\beta+w)=\alpha\beta+v\cdot w$.

(ii) A decomposes as a direct sum of vector spaces:

$$A=F1\oplus Fh\oplus M_1\oplus N_1\oplus M_{-1}\oplus N_{-1}. \tag{11}$$

Also there are skew-symmetric bilinear mappings ϕ_i: $M_i\times M_i\to N_{-i}$ ($i=\pm1$) such that the multiplication in A is determined by

$$h^2=1,\ m_ih=-hm_i=im_i,$$
$$n_ih=-hn_i=in_i,\quad m_im_i'=\phi_i(m_i,m_i'), \tag{12}$$
$$N_iM_j=M_jN_i=N_iN_j=M_iM_{-i}=0,$$

for $m_i,m_i'\in M_i$, $n_i\in N_i$, $i=\pm1$.

(iii) $F=\mathbb{R}$ and A decomposes as a direct sum of vector spaces:

$$A=\mathbb{C}\oplus S\oplus T. \tag{13}$$

Here S and T are vector spaces over $\mathbb{C}$ (action on the right); and there is a skew-symmetric bilinear mapping (over $\mathbb{R}$) φ: $S\times S\to T$, such that $\varphi(s_1k,s_2)=\varphi(s_1,s_2)k^\sigma$, for $s_1,s_2\in S$ and $k\in\mathbb{C}$, where σ is the standard involution in $\mathbb{C}$. The multiplication in A is determined by the multiplication in $\mathbb{C}$, the actions of $\mathbb{C}$ on the right on S and T, and

$$ks=sk^\sigma,\quad kt=tk^\sigma,\quad s_1s_2=\varphi(s_1,s_2),$$
$$ST=TS=T^2=0, \tag{14}$$

for $k\in\mathbb{C}$, $s,s_1,s_2\in S$, and $t\in T$.

(iv) A decomposes as a direct sum:

$$A=M_2(F)\oplus\Big(\bigoplus_{i\in I}F^2\Big). \tag{15}$$

Here I is an indexing set, and the multiplication in A is determined by the usual multiplication in $M_2(F)$, the fact that $\oplus\{F^2:i\in I\}$ is an ideal of A that squares to 0, and, for (α,β) in any copy of F^2, by

$$(\alpha,\beta)\begin{pmatrix}a&b\\c&d\end{pmatrix}=(\alpha a+\beta b,\alpha c+\beta d), \tag{16}$$

that is, the usual matrix multiplication of (α,β) by the transpose of $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$, and

$$\begin{pmatrix}a&b\\c&d\end{pmatrix}(\alpha,\beta)=(\alpha,\beta)\begin{pmatrix}d&-b\\-c&a\end{pmatrix}. \tag{17}$$

(v) $F=\mathbb{R}$ and A decomposes as a direct sum:

$$A=Q\oplus N. \tag{18}$$

Here Q is the quaternion division algebra, N is a linear space over Q^{op} (the action on the right), and the multiplication on A is determined by the multiplication in Q and

$$N^2=0,\quad qn=nq^\sigma, \tag{19}$$

for $n\in N$ and $q\in Q$, where σ is the standard involution in Q.

(vi) A is an octonion algebra.

Moreover, A is a quadratic associative algebra over F if and only if either (i') A is as in (i) but with V a nilpotent Lie algebra of class at most 2; (ii') A is as in (ii) but with $M_1=M_{-1}=0$; (iii') A is as in (iii) but with $S=0$; (iv') A is isomorphic to $M_2(F)$; or (v') A is the real quaternion division algebra.

Proof: Only a few details must be checked. Case (iv) corresponds to the case $V=\mathrm{sl}(2,F)\oplus\{N_i;\ i\in I\}$ of Sec. III. Then we can identify h with $\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$, e with $\left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right)$, f with $\left(\begin{smallmatrix}0&-1\\0&0\end{smallmatrix}\right)$, and, for any i, u_i with (1,0) and v_i with (0,1) in the ith copy of F^2. The rest is straightforward.

Notice that this provides a proof of Hurwitz's Theorem, once it is proved that any composition algebra is alternative (see Ref. 13).

Remark: Quadratic algebras with the corresponding norm form n, possibly degenerate, admitting composition form a broader class of algebras, since this condition affects only A/N, so it allows a great diversity for the radical N. For instance, it is not always true that there is a subalgebra S such that $A=S\oplus N$, as the following example shows:

$$A=F1+Fh+Fe+Ff+Fu,$$
$$V=Fh+Fe+Ff+Fu, \tag{20}$$

with anticommutative product in V (the algebra of vectors) given by

$$e\cdot h=e+u,\quad f\cdot h=-f,\quad e\cdot f=\tfrac{1}{2}h,$$
$$u\cdot h=u,\quad u\cdot e=u\cdot f=0, \tag{21}$$

and symmetric bilinear form (,) given by

$$(u,V) = 0, \quad (h,h) = -1, \quad (e,f) = \tfrac{1}{2}, \tag{22}$$

other pairings being equal to 0.

In this example, $N = Fu$, $A/N \cong M_2(F)$, the bilinear form is the one in $\mathrm{sl}(2,F)$ so that the corresponding norm form admits composition. But this algebra does not split over the ideal N, since it is easily shown that there is no subalgebra T of V with $V = T \oplus N$.

Actually, Proposition 2.1 (vii) tells us that a quadratic algebra admits composition if and only if it is a flexible algebra such that its quotient algebra modulo the radical of (,) is a composition algebra.

V. CONCLUDING REMARKS

Let us see how the algebras described in this paper unify some of the quadratic algebras considered in the references.

Ilamed and Salingaros[4] considered quadratic alternative algebras with dim $V = 3$ and nondegenerate (,). They correspond to the cases (iv) and (v) of Theorem 4.1 with $N = 0$. That is, they obtained, up to isomorphism, the quaternions, $M_2(\mathbb{R})$ and $M_2(\mathbb{C})$.

Jantzen[1] considered finite-dimensional quadratic algebras in which V is a Lie algebra and (,) equals $-1/(m-1)$ times the Killing form of V. we have seen in Sec. II that this condition is necessary for A to be alternative. Then he proved that the quadratic algebra so obtained is associative in the following cases: (i) dim $V = 1$; (ii) $m > 1$ and $x \cdot y = 1/(m-1)$ [Trace(ad(x))y − Trace(ad(y))x]; and (iii) dim $V = 3$ and Trace(ad(x)) $= 0$ for all x.

The first of these cases corresponds to Theorem 4.1 (i') and (ii') with $N_1 = N_{-1} = 0$.

In the second case, either V is Abelian [and we are in the situation of Theorem 4.1 (i')] or there is an element $h \in V$ such that $V = Fh \oplus S$ with Trace(ad(h)) $= -(m-1)$ and $S = \{v \in V$: Trace(ad(v)) $= 0\}$. Hence this latter case corresponds to Theorem 4.1(ii') with the additional condition of N_{-1} being equal to 0.

Finally, in the third case we get that either V is nilpotent of class at most 2—so we are in Theorem 4.1(i')—or V is simple—Theorem 4.1(iv') and (v').

Plebanski and Przanowski[2] get the finite-dimensional associative version of Theorem 4.1 by means of the Bianchi–Behr classification of real Lie algebras. They call the resulting algebras "quaternionlike algebras" and study the relation of these algebras with abstract cross products on vector spaces[3] [recall Proposition 2.1(v)].

ACKNOWLEDGMENT

This work was partially supported by the Dirección General de Investigación Cientifica y Técnica (DGICYT) (PS 87-0054).

[1] R. T. Jantzen, "Generalized quaternions and space-time symmetries," J. Math. Phys. **23**, 1741 (1982).
[2] J. F. Plebanski and M. Przanowski, "Generalizations of the quaternion algebra and Lie algebras," J. Math. Phys. **29**, 529 (1988).
[3] J. F. Plebanski and M. Przanowski, "Notes on a cross product of vectors," J. Math. Phys. **29**, 2334 (1988).
[4] Y. Ilamed and N. Salingaros, "Algebras with three anticommuting elements. I. Spinors and quaternions," J. Math. Phys. **22**, 2091 (1981).
[5] G. Domokos and S. Kövesi-Domokos, "The algebra of color," J. Math. Phys. **19**, 1477 (1978).
[6] G. Domokos and S. Kövesi-Domokos, "Towards an algebraic quantum chromodynamics," Phys. Rev. D **19**, 298 (1979).
[7] G. P. Wene, "An example of a flexible, Jordan-admissible algebra of current use in physics," Hadronic J. **1**, 944 (1978).
[8] G. P. Wene, "A generalization of the construction of Ilamed and Salingaros," J. Math. Phys. **24**, 221 (1983).
[9] G. P. Wene, "Algebras with anticommuting basal elements, space-time symmetries, and quantum theory," J. Math. Phys. **25**, 414 (1984).
[10] G. P. Wene, "The metric of a space and quadratic algebras," J. Math. Phys. **30**, 1034 (1989).
[11] R. A. Kunze and S. Scheinberg, "Alternative algebras having scalar involution," Pac. J. Math. **124**, 159 (1986).
[12] J. M. Osborn, "Quadratic division algebras," Trans. Am. Math. Soc. **105**, 202 (1962).
[13] K. A. Zhevlakov, A. M. Slinko, I. P. Shestakov, and A. I. Shirsov, *Rings That Are Nearly Associative* (Academic, New York, 1982).
[14] A. Micali and A. Paques, "Le groupe des algèbres quadratiques," Commun. Algebra **10**, 1765 (1982).
[15] A. Elduque, "On maximal subalgebras of central simple Malcev algebras," J. Algebra **103**, 216 (1986).
[16] A. Elduque, "On the color algebra," Algebras Groups Geom. **5**, 395 (1988).
[17] A. A. Sagle, "On anticommutative algebras with an invariant form," Can. J. Math. **16**, 370 (1964).
[18] S. Okubo, "Some classes of flexible Lie–Jordan admissible algebras," Hadronic J. **4**, 354 (1981).
[19] H. C. Myung, *Malcev-admissible Algebras* (Birkhäuser, Boston, 1986).
[20] K. McCrimmon, "Nonassociative algebras with scalar involution," Pac. J. Math. **116**, 85 (1985), Example 4.13.
[21] S. Okubo, "Construction of nonassociative algebras from representation modules of simple Lie algebras," Algebras Groups Geom. **3**, 60 (1986), Corollary 2.1.
[22] V. T. Filippov, "Semiprimary Mal'cev algebras of characteristic three," Algebra Logika **14** (1), 100 (1975).
[23] R. Carlsson, "Malcev-Moduln," J. Reine Angew. Math. **281**, 199 (1976).
[24] E. N. Kuzmin, "Mal'cev algebras and their representations," Algebra Logika **7** (4), 48 (1968).

Extended BRST quantization of gauge theories in the generalized canonical formalism

I. A. Batalin
Lebedev Physical Institute, Moscow 117924, USSR

P. M. Lavrov
Tomsk State Pedagogical Institute, Tomsk 634044, USSR

I. V. Tyutin
Moscow Institute of Electronic Engineering, Moscow 109028, USSR

(Received 15 June 1989; accepted for publication 2 August 1989)

The rules of canonical quantization of gauge theories are formulated on the basis of the extended BRST symmetry principle. The existence of solutions of the generating equations of the gauge algebra is proved. Equivalence between the extended BRST quantization and the standard method of generalized canonical quantization is established. Ward identities corresponding to invariance of a theory under the extended BRST symmetry are obtained.

I. INTRODUCTION

The method of generalized canonical quantization is now becoming increasingly popular as the most effective means for solving the problems of quantization of dynamical systems with constraints[1-3] [see, also, the review (Ref. 4)]. This method is based on the idea of a special type global supersymmetry, which in the Hamiltonian formalism is a generalization of the so-called BRST symmetry originally introduced in the Lagrangian formalism in the Abelian gauge theory[5] and in the Yang–Mills theory.[6] According to the now generally accepted terminology, the Lagrangian BRST supersymmetry along with its Hamiltonian analog are united by the general term "BRST symmetry."

In its original version, the BRST symmetry means invariance of a resultant action under global nilpotent transformations of dynamical variables with one single fermion parameter. The original BRST symmetry is, in principle, quite sufficient to construct a correct quantum description of an arbitrary dynamical system with constraints.

It turns out, however, that the requirement of BRST symmetry may be substantially strengthened in such a way that the action be invariant not only under BRST transformations, but at the same time also under so-called anti-BRST transformations.[7,8] The totality of global transformations extended in such a manner introduces instead of the original "restricted BRST algebra" an "extended BRST algebra."

The requirement of an extended BRST symmetry yields two most essential additional advantages. First, the fermion parameters of BRST and anti-BRST transformations, the same as their generators, turn out to form a natural doublet under the global symplectic group Sp(2). Second, the extended BRST symmetry in a number of cases leads directly to a natural geometrical formulation in which BRST and anti-BRST transformations assume the form of supertranslations in configuration and phase superspaces.

The above-mentioned advantages of the extended BRST symmetry have been pointed out earlier by a number of authors,[9-11] who have also discussed the various technical aspects of the corresponding quantization procedure for special classes of dynamical systems.

In the present paper we shall, for the most part, concern ourselves with a consistent formulation of the version of generalized canonical quantization based on the requirement of extended BRST symmetry for dynamical systems with linearly independent constraints. Besides, the most essential points of the formulation proposed are substantiated. First, this is the proof of existence and the description of arbitrariness of solutions of the generating equations and, second, the proof of equivalence of our formulation and the usual version based on the restricted BRST symmetry.

We shall restrict our consideration to the case of first-class constraints. The extension to the case of second-class constraints can be obtained in an obvious manner.

The notation used is as follows. We use the standard definition[12] of the Poisson superbracket in phase space $\Gamma = (P_A, Q^A)$:

$$\{G,F\} = \frac{\delta G}{\delta Q^A}\frac{\delta F}{\delta P_A} - \frac{\delta F}{\delta Q^A}\frac{\delta G}{\delta P_A}(-1)^{\epsilon(G)\epsilon(F)}. \quad (1)$$

By $\epsilon(G)$ we denote Grassmann parity of the quantity G. Derivatives with respect to generalized momenta P_A are always understood as left-hand, and those with respect to generalized coordinates Q^A (unless specified) as right-hand ones. Left-hand derivatives with respect to Q^A have a special sign "l": $\delta_l/\delta Q^A$. The Grassmann parities P_A and Q^A coincide: $\epsilon(P_A) = \epsilon(Q^A) = \epsilon_A$. The superbracket (1) possesses the standard algebraic properties:

$$\begin{aligned} &\{G,F\} = -\{F,G\}(-1)^{\epsilon(G)\epsilon(F)}, \\ &\{F,GH\} = \{F,G\}H + G\{F,H\}(-1)^{\epsilon(F)\epsilon(G)}, \quad (2) \\ &\{\{F,G\},H\}(-1)^{\epsilon(F)\epsilon(H)} + \text{cycl. perm. } (F,G,H) \equiv 0. \end{aligned}$$

The last relation is the Jacobi identity for the superbracket.

Next, the indices of the global symplectic group Sp(2) are denoted by lowercase latin characters $a,b,c,\ldots$ and take on two values $a = 1,2$. The invariant tensor of the group Sp(2) is denoted by

 0022-2488/90/010006-08$03.00

$$\epsilon^{ab}=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},\quad \epsilon^{ac}\epsilon_{cb}=\delta^a_b\,.$$

Symmetrization in Sp(2) indices is denoted by

$$A^{\{ab\}}=A^{ab}+A^{ba}\,.$$

Gauge indices are denoted by lowercase greek characters $\alpha,\beta,\gamma,\ldots$.

The ghost number of the quantity A is denoted by $\mathrm{gh}(A)$. For a new ghost number of the quantity A we use the notation $\mathrm{ngh}(A)$.

II. BRST QUANTIZATION

We consider a dynamical system that in the phase space of initial canonical variables (p_i,q^i), $i=1,2,\ldots,n$, is described by the Hamiltonian $H_0=H_0(p,q)$ and by the set of linearly independent first-class constraints $T_\alpha=T_\alpha(p,q)$, $\alpha=1,2,\ldots,m$, whose Grassmann parities are $\epsilon(H_0)=0$, $\epsilon(T_\alpha)=\epsilon_\alpha$, and the involution relations hold true:

$$\{T_\alpha,T_\beta\}=T_\gamma U^\gamma_{\alpha\beta},\quad \{H_0,T_\alpha\}=T_\beta V^\beta_\alpha\,, \tag{3}$$

where the structure functions $U^\gamma_{\alpha\beta}$ possess the properties of generalized antisymmetry $U^\gamma_{\alpha\beta}=-(-1)^{\epsilon_\alpha\epsilon_\beta}U^\gamma_{\beta\alpha}$. Next, we introduce an extended phase space Γ parametrized by the following set of canonical variables:

$$\begin{aligned}&\Gamma=(P_A,Q^A)=(p_i,q^i;\mathscr{P}_{\alpha a},C^{\alpha a};\lambda_\alpha,\pi^\alpha)\,,\\ &\epsilon(q^i)=\epsilon_i,\quad \epsilon(C^{\alpha a})=\epsilon_\alpha+1,\quad \epsilon(\pi^\alpha)=\epsilon_\alpha\,,\end{aligned} \tag{4}$$

where the ghost momenta $\mathscr{P}_{\alpha a}$ and the coordinates $C^{\alpha a}$ form doublets with respect to the index a under the group Sp(2).

The key role in the procedure of extended BRST quantization is played by the generating functions Ω^a and $\mathscr{H}$. The fermion functions Ω^a are the solutions of Sp(2)-covariant generating equations

$$\{\Omega^a,\Omega^b\}=0\,, \tag{5}$$

which also satisfy the boundary conditions

$$\begin{aligned}&\left.\frac{\delta\Omega^a}{\delta C^{\alpha b}}\right|_{C=\pi=\mathscr{P}=\lambda=0}=T_\alpha\delta^a_b,\\ &\left.\frac{\delta\Omega^a}{\delta\pi^\alpha}\right|_{C=\pi=\lambda=0}=\epsilon^{ab}\mathscr{P}_{\alpha b}\,.\end{aligned} \tag{6}$$

In its turn, the boson function $\mathscr{H}$ satisfies generating equations of the form

$$\{\mathscr{H},\Omega^a\}=0\,, \tag{7}$$

with the boundary condition

$$\mathscr{H}|_{C=\pi=\mathscr{P}=\lambda=0}=H_0\,. \tag{8}$$

The total unitarizing Hamiltonian H is now determined in terms of $\mathscr{H}$ and Ω^a by the formula

$$H=\mathscr{H}+\tfrac{1}{2}\epsilon_{ab}\{\{\Phi,\Omega^b\},\Omega^a\}\,, \tag{9}$$

where Φ is the boson function fixing a concrete choice of admissible gauge. An essential property of the unitarizing Hamiltonian H [(9)] is its invariance

$$\delta H=\{H,\Omega^a\}\mu_a=0\,, \tag{10}$$

under extended BRST transformations of the variables of phase space Γ [see (4)]

$$\delta\Gamma=\{\Gamma,\Omega^a\}\mu_a\,. \tag{11}$$

Here μ_a is a doublet of constant Grassmann parameters of extended BRST symmetry. The invariance (10) obviously follows from Eqs. (5) and (7), as well as from the Jacobi identity (2) for Ω^a.

We now consider from a quantum point of view the standard consequences of invariance of the unitarizing Hamiltonian H under the global transformations (11). To this end we define the vacuum functional Z_Φ in terms of the following functional integral:

$$Z_\Phi=\int D\Gamma\exp\left\{\frac{i}{\hbar}\int dt(P_A\dot{Q}^A-H)\right\}. \tag{12}$$

Then Z_Φ does not, in fact, depend on the choice of the gauge functional Φ. Indeed, one can readily establish that any change of the gauge $\Phi\to\Phi+\Delta\Phi$ in the integral (12) can be compensated by the change of integration variables $\Gamma\to\Gamma+\delta\Gamma$, where

$$\delta\Gamma=(i/2\hbar)\{\Gamma,\Omega^a\}\epsilon_{ab}\{\Omega^b,\Delta\Phi\}\,.$$

Hence $Z_{\Phi+\Delta\Phi}=Z_\Phi$, and therefore the S matrix is gauge invariant in the formalism of the extended BRST quantization.

Another consequence of invariance of the total Hamiltonian, which we discuss here, is the presence of gauge Ward identities. To derive these Ward identities, we consider the generating functional

$$\begin{aligned}&Z(J,\Gamma^*,\overline{\Gamma})\\ &\quad=\int D\Gamma\exp\left[\frac{i}{\hbar}\int dt\Big(P_A\dot{Q}^A-H+J\Gamma\right.\\ &\qquad\left.+\Gamma^*_a\{\Gamma,\Omega^a\}+\frac{1}{2}\overline{\Gamma}\epsilon_{ab}\{\{\Gamma,\Omega^b\},\Omega^a\}\Big)\right].\end{aligned} \tag{13}$$

Using this generating functional, the Green's functions of the theory with the Hamiltonian H [Eq. (5)] are calculated through differentiation with respect to the sources J for $\Gamma^*_a=\overline{\Gamma}=J=0$. In (13) we have introduced additional sources Γ^*_a to the transformations $\{\Gamma,\Omega^a\}$ of extended BRST symmetry and the source $\overline{\Gamma}$ to the generator $\frac{1}{2}\epsilon_{ab}\{\{\Gamma,\Omega^b\},\Omega^a\}$ and have taken into account that by virtue of Eqs. (5) and the Jacobi identity (2) there are no other nontrivial generators. We shall now make the change of variables (11) in the functional integral (13), use the invariance property of H [(10)], and also use the fact that the Berezinian of the change of variables (11) is equal to unity. Then we shall obtain the following Ward identities for the generating functional Z [(13)]:

$$J\frac{\delta Z}{\delta\Gamma^*_a}-\epsilon^{ab}\Gamma^*_b\frac{\delta Z}{\delta\overline{\Gamma}}=0\,. \tag{14}$$

It should be noted that the derivatives with respect to the additional sources Γ^*_a and $\overline{\Gamma}$ in (14) are left-hand ones. The identities (14) can be rewritten for the generating functional of the vertex functions (the effective action). The effective action S is determined in the standard manner as the Legendre transformation of $\ln Z$ with respect to the sources J:

$$S(\langle\Gamma\rangle,\Gamma^*,\overline{\Gamma})=(\hbar/i)\ln Z(J,\Gamma^*,\overline{\Gamma})-J\cdot\langle\Gamma\rangle\,, \tag{15}$$

where the averaged variables $\langle\Gamma\rangle$ are determined as

$$\langle\Gamma\rangle = \frac{\hbar}{i}\frac{\delta \ln Z(J,\Gamma^*,\overline{\Gamma})}{\delta J}. \tag{16}$$

Then (15) and (16) obviously imply

$$\frac{\delta S(\langle\Gamma\rangle,\Gamma^*,\overline{\Gamma})}{\delta\langle\Gamma\rangle} = -J \tag{17}$$

and the identities (14) can be rewritten for S to become

$$\frac{\delta S}{\delta\langle\Gamma\rangle}\cdot\frac{\delta S}{\delta\Gamma^*_a} + \epsilon^{ab}\Gamma^*_b\frac{\delta S}{\delta\overline{\Gamma}} = 0. \tag{18}$$

In Eqs. (16)–(18) the derivatives with respect to J are left-hand and those with respect to $\langle\Gamma\rangle$ are right-hand.

III. THE EXISTENCE THEOREM FOR GENERATING EQUATIONS OF THE GAUGE ALGEBRA

The question of existence of solutions of the generating equations (5) and (7) satisfying the boundary conditions (6) and (7) is the crucial point of the whole scheme of extended BRST quantization. Here we shall prove, using a power series expansion of canonical variables $C^{\alpha a}$ and π^α, the existence of solutions of Eqs. (5) and (7) and describe the arbitrariness present in these solutions. We shall prove the possibility of choosing solutions of generating equations in the form symmetric under the group Sp(2).

In the standard BRST quantization procedure, the initial canonical variables p_i and q^i are associated with the zero ghost number, the ghost variables C^α ($C^\alpha \equiv C^{\alpha 1}$) are associated with $\mathrm{gh}(C^\alpha) = 1$, and the antighost variables $\overline{C}^\alpha$ ($\overline{C}^\alpha \equiv C^{\alpha 2}$) with $\mathrm{gh}(\overline{C}^\alpha) = -1$, etc. The solutions of generating equations are expanded in power series of the components C^α ($\overline{C}^\alpha$ enters only through gauges) associated with a positive ghost number. Therefore, the requirement of conservation of the ghost number in each order determines completely the structure of the expansion. As a result of the fact that $\mathrm{gh}(C^{\alpha 1}C^{\beta 2}) = \mathrm{gh}(\pi^\alpha) = 0$ in the procedure of extended BRST quantization, the structure of the expansion in power series of $C^{\alpha a}$ and π^α is now not fixed by the requirement of ghost number conservation. In view of this, it is convenient to introduce for all variables of the extended phase space (4) the so-called "new ghost number" $\mathrm{ngh}(\Gamma)$ by the rule

$$\begin{aligned}
&\mathrm{ngh}(p_i) = \mathrm{ngh}(q^i) = 0, \quad \mathrm{ngh}(C^{\alpha a}) = 1,\\
&\mathrm{ngh}(\pi^\alpha) = 2, \quad \mathrm{ngh}(\mathscr{P}_{\alpha a}) = -1,\\
&\mathrm{ngh}(\lambda_\alpha) = -2,\\
&\mathrm{ngh}(AB) = \mathrm{ngh}(A) + \mathrm{ngh}(B),
\end{aligned} \tag{19}$$

and to require fulfillment of the conditions

$$\mathrm{ngh}(\mathscr{H}) = 0, \quad \mathrm{ngh}(\Omega^a) = 1. \tag{20}$$

We shall seek solutions of Eqs. (5) and (7) in the form of expansions in power series of $C^{\alpha a}$ and π^α:

$$\Omega^a = \sum_{n=1}^{\infty}\Omega^a_n, \quad \mathrm{ngh}(\Omega^a_n) = 1, \quad \Omega^a_n \sim C^{n-m}\pi^m, \tag{21}$$

$$\mathscr{H} = H_0 + \sum_{n=1}^{\infty}\mathscr{H}_n, \quad \mathrm{ngh}(\mathscr{H}_n) = 0, \quad \mathscr{H}_n \sim C^{n-m}\pi^m. \tag{22}$$

To begin with, we shall prove the existence of solutions of Eqs. (5). In the first-order perturbation series, the solution of Eqs. (5) and (6) has the form

$$\Omega^a_1 = T_\alpha C^{\alpha a} + \epsilon^{ab}\mathscr{P}_{\alpha b}\pi^\alpha. \tag{23}$$

Suppose that we are given quantities Ω^a_n such that Eqs. (5) are satisfied in the nth order. Now we shall find the expression for $\{\Omega^a,\Omega^b\}$ in the $(n+1)$-th order:

$$\{\Omega^a,\Omega^b\}_{n+1} = W^{\{a}\Omega^{b\}}_{n+1} + B^{ab}_{n+1}, \tag{24}$$

where the quantities B^{ab}_{n+1} are constructed from Ω^a_k, $k \leqslant n$, and possess the symmetry properties $B^{ab}_{n+1} = B^{ba}_{n+1}$. The operators W^a in (24) are given by the formula

$$W^a = T_\alpha\frac{\delta}{\delta\mathscr{P}_{\alpha a}} + \epsilon^{ab}\mathscr{P}_{\alpha b}\frac{\delta}{\delta\lambda_\alpha} + (-1)^{\epsilon_\alpha}\epsilon^{ab}\pi^\alpha\frac{\delta_l}{\delta C^{ab}}. \tag{25}$$

One can directly verify that the W^a [(25)] form a set of nilpotent anticommuting operators

$$W^{\{a}W^{b\}} = 0. \tag{26}$$

The quantities B^{ab}_{n+1} satisfy the equations

$$W^a B^{bc}_{n+1} + \text{cycl. perm.} \quad (a,b,c) = 0, \tag{27}$$

which follow, with allowance made for (26), from the Jacobi identities for Ω^a calculated in the $(n+1)$-th order.

All further steps of the proof are based on the following lemma (a similar lemma for the case of a single nilpotent operator W was proved in Ref. 13).

Lemma 1: Any regular solution of the equations

$$W^a X = 0, \tag{28}$$

$$W^{\{a}X^{a_1\cdots a_n\}} = 0, \tag{29}$$

which vanishes when $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$, has the form

$$X = \tfrac{1}{2}\epsilon_{ab}W^aW^bY, \tag{30}$$

$$X^{a_1\cdots a_n} = W^{\{a_1}Y^{a_2\cdots a_n\}}, \tag{31}$$

respectively, where Y, $Y^{a_1\cdots a_{n-1}}$ are some functions of the variables of the extended phase space (4). In Eq. (29) the functions $X^{a_1\cdots a_n}$ are regarded as symmetric under permutation of any indices. The symbol $\{a_1a_2\cdots a_n\}$ in (29) and (31) stands for cyclic permutation in $a_1, a_2,\ldots,a_n$.

To prove the lemma, we shall introduce a set of auxiliary fermion operators Γ_a "conjugated" to W^a by defining their algebra as

$$\Gamma_{\{\alpha}\Gamma_{b\}} = 0, \quad W^a\Gamma_b + \Gamma_bW^a = \delta^a_b N. \tag{32}$$

The solution of Eq. (32) exists, for example,

$$\Gamma_a = \mathscr{P}_{\alpha a}\frac{\delta_l}{\delta T_\alpha} - \epsilon_{ab}\lambda_\alpha\frac{\delta}{\delta\mathscr{P}_{\alpha b}}. \tag{33}$$

Given this, the operator N in (32), scalar under the group Sp(2), takes on the form of "conformal" operator

$$N = T_\alpha\frac{\delta_l}{\delta T_\alpha} + \mathscr{P}_{\alpha a}\frac{\delta}{\delta\mathscr{P}_{\alpha a}} + \lambda_\alpha\frac{\delta}{\delta\lambda_\alpha} \tag{34}$$

commuting with W^a and Γ_a

$$NW^a = W^aN, \quad N\Gamma_a = \Gamma_aN. \tag{35}$$

We shall first consider the solution of Eqs. (28). We shall act upon Eqs. (28) from the left by the operator Γ_b and take into account (32). Then, with allowance made for the fact that on the solutions $N>0$, we have

$$X=(1/N)W^1\Gamma_1 X=(1/N)W^2\Gamma_2 X. \tag{36}$$

From (36) it follows that

$$\begin{aligned} X&=(1/N)W^1\Gamma_1(1/N)W^2\Gamma_2 X\\ &=-W^1W^2((1/N^2)\Gamma_1\Gamma_2 X)\\ &=\tfrac{1}{2}\epsilon_{ab}W^aW^b((1/2N^2)\epsilon^{cd}\Gamma_c\Gamma_d X), \end{aligned} \tag{37}$$

which proves the validity of Lemma 1 concerning solutions of Eqs. (28).

Now let us proceed to the solution of Eqs. (29). To this end, in the algebra of the operators W^a and Γ_a we define the Sp(2) scalar operator $M=\Gamma_a W^a$. One can readily establish the following essential properties of this operator:

$$M^n=(2^{n-1}-1)N^{n-2}M^2-(2^{n-1}-2)N^{n-1}M, \quad n\geqslant 3, \tag{38}$$

$$\Gamma_a M^2=N\Gamma_a M. \tag{39}$$

Let us act on Eqs. (29) from the left by the operators Γ_a and sum over the index "a." We obtain

$$MX^{a_1\cdots a_n}+nNX^{a_1\cdots a_n}-W^{\{a_1}\Gamma_a X^{a_2\cdots a_n\}a}=0. \tag{40}$$

For the functions $X^{a_1\cdots a_n}$ vanishing when $T_\alpha=\mathscr{P}_{\alpha\alpha}=\lambda_\alpha=0$, Eq. (40) implies

$$X^{a_2\cdots a_n}=W^{\{a_1}Y_1^{a_2\cdots a_n\}}+X_1^{a_1\cdots a_n}, \tag{41}$$

where

$$Y_1^{a_1\cdots a_n}=(1/nN)\Gamma_a X^{a_2\cdots a_n a}, \tag{42}$$

$$X_1^{a_1\cdots a_n}=-(1/nN)MX^{a_1\cdots a_n}. \tag{43}$$

The functions $X_1^{a_1\cdots a_n}$ satisfy Eqs. (29) and, therefore, by virtue of (40), admit the following representation:

$$X_1^{a_1\cdots a_n}=W^{\{a_1}Y_2^{a_2\cdots a_n\}}+X_2^{a_1\cdots a_n}, \tag{44}$$

where

$$Y_2^{a_2\cdots a_n}=\frac{1}{nN}\Gamma_a X_1^{a_2\cdots a_n a}=-\frac{1}{n^2N^2}\Gamma_a MX^{a_2\cdots a_n a}, \tag{45}$$

$$X_2^{a_1\cdots a_n}=-\frac{1}{nN}MX_1^{a_1\cdots a_n}=\frac{1}{n^2N^2}M^2X^{a_1\cdots a_n}. \tag{46}$$

In turn, the functions $X_2^{a_1\cdots a_n}$ also satisfy Eqs. (29) and, therefore,

$$X_2^{a_1\cdots a_n}=W^{\{a_1}Y_3^{a_2\cdots a_n\}}+X_3^{a_1\cdots a_n}, \tag{47}$$

$$Y_3^{a_2\cdots a_n}=\frac{1}{nN}\Gamma_a X_2^{a_2\cdots a_n a}=\frac{1}{n^3N^3}\Gamma_a M^2X^{a_2\cdots a_n a}, \tag{48}$$

$$X_3^{a_1\cdots a_n}=-\frac{1}{nN}MX_2^{a_1\cdots a_n}=-\frac{1}{n^3N^3}M^3X^{a_1\cdots a_n}. \tag{49}$$

With allowance made for (38), from (43), (46), and (49) it follows that

$$X_3^{a_1\cdots a_n}=-(3/n)X_2^{a_1\cdots a_n}-(2/n^2)X_1^{a_1\cdots a_n}. \tag{50}$$

Substituting (50) into (47), we find

$$X_2^{a_1\cdots a_n}=\frac{n}{n+3}W^{\{a_1}Y_3^{a_2\cdots a_n\}}-\frac{2}{n(n+3)}X_1^{a_1\cdots a_n}. \tag{51}$$

Then (51) and (44) imply

$$\begin{aligned} X_1^{a_1\cdots a_n}=&\frac{n(n+3)}{(n+1)(n+2)}W^{\{a_1}Y_2^{a_2\cdots a_n\}}\\ &+\frac{n^2}{(n+1)(n+2)}W^{\{a_1}Y_3^{a_2\cdots a_n\}}. \end{aligned} \tag{52}$$

Finally, taking into account (39), we find, from (41), (42), (45), (48), and (52),

$$X^{a_1\cdots a_n}=W^{\{a_1}\Gamma_a\left(\frac{1}{nN}-\frac{M}{n(n+1)N^2}\right)X^{a_2\cdots a_n\}a}. \tag{53}$$

Comparison of (53) with (31) completes the proof of Lemma 1.

It should also be noted that the proof of Lemma 1 that we have carried out here exhibits an important fact concerning solutions of Eqs. (28) and (29). Namely, in case the solutions of Eqs. (28) and (29) are Sp(2) covariant, the quantities Y and $Y^{a_1\cdots a_{n-1}}$ can be chosen in a Sp(2)-covariant form. To prove this assertion, it suffices to refer to the representation of solutions of Eqs. (28) and (29) in the form (37) and (53), respectively.

We now return to the proof of the existence of solutions of Eqs. (5). Since $\mathrm{ngh}(B^{ab}_{n+1})=2$, it follows obviously that, for $n\geqslant 2$, we have $B^{ab}_{n+1}=0$ for $T_\alpha=\mathscr{P}_{\alpha a}=\lambda_\alpha=0$. When $n=1$, the representation

$$\begin{aligned} B_2^{ab}&=-\{T_\alpha,T_\beta\}C^{\beta b}C^{\alpha a}(-1)^{\epsilon_\alpha}\\ &=-T_\gamma U^\gamma_{\alpha\beta}C^{\beta b}C^{\alpha a}(-1)^{\epsilon_\alpha} \end{aligned}$$

holds true, and hence we have $B_2^{ab}=0$ for $T_\alpha=0$. Then by virtue of Lemma 1 there exist functions Y^a_{n+1}, such that

$$B^{ab}_{n+1}=W^{\{a}Y^{b\}}_{n+1}. \tag{54}$$

Next, choosing

$$\Omega^a_{n+1}=-Y^a_{n+1}, \tag{55}$$

we obtain that Eqs. (5) are satisfied already in the $(n+1)$-th order. Applying induction, we conclude the proof of existence of solutions of the generating equations (5) with given boundary conditions (6). Note that, instead of $\Omega^a_{n+1}=-Y^a_{n+1}$, we could take other functions

$$\Omega^a_{n+1}=-Y^a_{n+1}+W^aZ_{n+1} \tag{56}$$

and again the corresponding Ω^a would satisfy Eqs. (5) in the $(n+1)$-th order. On the basis of Lemma 1 proved just now, one can readily show that the terms W^aZ_{n+1} in (56) exhaust [with given boundary conditions (6)] all the arbitrariness in the solution of Eqs. (5) in the $(n+1)$-th order.

The above-mentioned arbitrariness in the solution of Eqs. (5) can be transformed to the form of canonical transformation. Indeed, suppose there exist two solutions Ω^a and Ω^a_0 of Eqs. (5). Suppose next that these solutions coincide up

to nth order inclusive [in the first order they coincide by virtue of the boundary conditions (6)], while in the $(n+1)$-th approximation they are already distinct. We shall write

$$\Omega^a_{n+1} = \Omega^a_{0\,n+1} + \Delta\Omega^a_{n+1}\,.$$

Then the functions $\Delta\Omega^a_{n+1}$ satisfy the equations

$$W^{\{a}\Delta\Omega^{b\}}_{n+1} = 0\,,$$

whose solution has by virtue of Lemma 1 the following form:

$$\Delta\Omega^a_{n+1} = W^a X_{n+1}\,.$$

Now we shall perform a canonical transformation of Ω^a with the generating function X_{n+1}:

$$\Omega'^a = \exp\{-\hat{X}_{n+1}\}\Omega^a \exp\{\hat{X}_{n+1}\}\,,$$

where we have introduced the notation

$$\hat{X}_{n+1} = \frac{\delta X_{n+1}}{\delta P_A}\frac{\delta_l}{\delta Q^A} - \frac{\delta X_{n+1}}{\delta Q^A}\frac{\delta}{\delta P_A}\,.$$

Then in the $(n+1)$-th approximation we have

$$\Omega'^a_k = \Omega^a_k = \Omega^a_{0k}, \quad k\leqslant n\,,$$

$$\Omega'^a_{n+1} = \Omega^a_{n+1} - \{\Omega^a, X_{n+1}\}_{n+1}$$

$$= \Omega^a_{n+1} - W^a X_{n+1} = \Omega^a_{0\,n+1}\,,$$

i.e., solutions of Eqs. (5) already coincide in the $(n+1)$-th order. Then by induction we have that any two solutions of Eqs. (5) with the boundary conditions (6) and with identical new ghost numbers equal to unity are related by a canonical transformation.

The situation is similar to the existence of solutions of equations for the boson generating function (7) with the boundary condition (8). In considering Eqs. (7) we shall think of the functions Ω^a as given solutions of Eqs. (5). In the zero approximation, $\mathscr{H} = H_0$. Next we assume that quantities $\mathscr{H}_n$ are constructed such that Eqs. (7) are satisfied in the nth order. Let us find the expression for $\{\mathscr{H},\Omega^a\}$ in the $(n+1)$-th order:

$$\{\mathscr{H},\Omega^a\}_{n+1} = -W^a\mathscr{H}_{n+1} + D^a_{n+1}\,,$$

where the operators W^a are defined in (25), and the quantities D^a_{n+1} are constructed from $\mathscr{H}_m$ and Ω^a_{m+1}, $m\leqslant n$. The Jacobi identities for the functions $\mathscr{H}$, Ω^a, Ω^b, and Eqs. (5) imply that

$$\{\{\mathscr{H},\Omega^a\},\Omega^b\} + \{\{\mathscr{H},\Omega^b\},\Omega^a\} \equiv 0\,. \tag{57}$$

Considering the identities (57) in the $(n+1)$-th order and taking (26) into account, we obtain

$$W^{\{a}D^{b\}}_{n+1} = 0\,. \tag{58}$$

Since $\mathrm{ngh}(D^a_{n+1}) = 1$ and $n+1\geqslant 1$, it obviously follows that $D^a_{n+1} = 0$ for $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$. By virtue of Lemma 1, the general solution of Eqs. (58) can be written as

$$D^a_{n+1} = W^a X_{n+1}\,.$$

Choosing

$$\mathscr{H}_{n+1} = X_{n+1}\,, \tag{59}$$

we find that Eqs. (7) are already satisfied in the $(n+1)$-th order. Applying induction, we complete the proof of existence of the solution of Eqs. (7) with given boundary condition (8). Note that instead of (59) we could take as $\mathscr{H}_{n+1}$ the function

$$\mathscr{H}_{n+1} = X_{n+1} + \tfrac{1}{2}\epsilon_{ab}W^bW^aY_{n+1} \tag{60}$$

and again the corresponding $\mathscr{H}$ would satisfy Eqs. (7) in the $(n+1)$-th order. On the basis of Lemma 1, one can easily show that the second summand in the right-hand side of (60) exhausts all the arbitrariness in the solution of Eqs. (7) in the $(n+1)$-th order.

The description of arbitrariness in the solution of Eqs. (7) with the boundary condition (8) rests upon the following lemma.

Lemma 2: Any regular solution of equations

$$\{\Omega^a, X\} = 0\,, \tag{61}$$

vanishing when $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$, has the form

$$X = \tfrac{1}{2}\epsilon_{ab}\{\Omega^b,\{\Omega^a, Y\}\}\,, \tag{62}$$

where Y is a certain function of the variables of the extended phase space.

We shall seek the solution of Eqs. (61) in the form of expansion in power series of $C^{\alpha a}$ and π^α:

$$X = \sum_{n=0}^{\infty} X_n, \quad X_n \sim C^{n-m}\pi^m\,.$$

In the zero-th order, Eq. (61) has the form

$$W^a X_0 = 0\,.$$

By virtue of Lemma 1, the solution of this equation,

$$X_0 = \tfrac{1}{2}\epsilon_{ab}W^bW^aY_0 = \tfrac{1}{2}\epsilon_{ab}\{\Omega^b,\{\Omega^a, Y_0\}\} + \cdots\,,$$

with accuracy up to terms of higher order in $C^{\alpha a}$ and Π^α can be represented in the form (62). Next we suppose that up to the nth order the solution X of Eq. (61) can be represented in the form of (62):

$$X = \tfrac{1}{2}\epsilon_{ab}\{\Omega^b,\{\Omega^a,[Y]_n\}\} + X_{(n+1)}\,, \tag{63}$$

with a certain function $[Y]_n$, which is a polynomial of degree n with respect to powers of C and π. In the $(n+1)$-th approximation,

$$X_{n+1} = \tfrac{1}{2}\epsilon_{ab}\{\Omega^b,\{\Omega^a,[Y]_n\}\}_{n+1} + \Delta X_{n+1},$$
$$\Delta X_{n+1} = (X_{(n+1)})_{n+1}\,, \tag{64}$$

where the function ΔX_{n+1} satisfies the equation

$$W^a\Delta X_{n+1} = 0\,,$$

and, therefore,

$$\Delta X_{n+1} = \tfrac{1}{2}\epsilon_{ab}W^bW^aY_{n+1}$$

$$= \tfrac{1}{2}\epsilon_{ab}\{\Omega^b,\{\Omega^a, Y_{n+1}\}\} + \cdots\,, \tag{65}$$

with accuracy up to terms of higher order than $(n+1)$. Taking (63)–(65) into account one can represent the solutions of Eq. (61) in the $(n+1)$-th approximation in the form (62). Now the use of induction completes the proof of Lemma 2.

One can similarly prove the assertion that any solution of the equation

$$\{\Omega^{\{a}, X^{a_1\cdots a_n\}}\} = 0\,,$$

for $X^{a_1\cdots a_n}$ symmetric with respect to the indices a_i and vanishing when $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$, has the form

$$X^{a_1\cdots a_n} = \Omega^{\{a_1} Y^{a_2\cdots a_n\}}.$$

Furthermore, one can show that there exist odd functions Ξ_a and an even function Θ that form, along with Ω^a, the complex

$$\{\Omega^a,\Omega^b\} = \{\Xi_a,\Xi_b\} = \{\Omega^a,\Theta\} = \{\Xi_a,\Theta\} = 0,$$
$$\{\Omega^a,\Xi_b\} = \delta^a_b\,\Theta.$$

Making use of Lemma 2, one can now easily describe the arbitrariness present in the solution of Eqs. (7) with the boundary condition (8). Indeed, suppose there exist two solutions of Eqs. (7), $\mathscr{H}^{(1)}$ and $\mathscr{H}^{(2)}$, with the boundary condition (8). Consider the function $\Delta\mathscr{H} = \mathscr{H}^{(1)} - \mathscr{H}^{(2)}$. This function satisfies the equation

$$\{\Omega^a,\Delta\mathscr{H}\} = 0$$

and possesses the property $\Delta\mathscr{H} = 0$ when $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$. By virtue of Lemma 2,

$$\Delta\mathscr{H} = \tfrac{1}{2}\epsilon_{ab}\{\Omega^b,\{\Omega^a,Y\}\}$$

and the arbitrariness in the solution of Eqs. (7) with the boundary condition (8) corresponds to the change of the gauge in the total unitarizing Hamiltonian.

IV. EQUIVALENCE OF STANDARD AND EXTENDED BRST QUANTIZATIONS

As is well known, in the standard BRST quantization procedure the first step is determination of the fermion $\Omega_{\min}$ and the boson $\mathscr{H}_{\min}$ generating functions using the following equations and boundary conditions:

$$\{\Omega_{\min},\Omega_{\min}\} = 0, \quad \left.\frac{\delta\Omega_{\min}}{\delta C^\alpha}\right|_{c=\bar{\mathscr{P}}=0} = T_\alpha, \tag{66}$$

$$\{\mathscr{H}_{\min},\Omega_{\min}\} = 0, \quad \mathscr{H}_{\min}|_{c=\bar{\mathscr{P}}=0} = H_0, \tag{67}$$

in the minimal sector $\Gamma_{\min} = (p_i,q^i;\bar{\mathscr{P}}_\alpha,C^\alpha)$, where $\bar{\mathscr{P}}_\alpha \equiv \mathscr{P}_{\alpha 1}$, $C^\alpha \equiv C^{\alpha 1}$. Next, making use of $\Omega_{\min}$ and $\mathscr{H}_{\min}$, one constructs the fermion function Ω,

$$\Omega = \Omega_{\min} + \mathscr{P}_\alpha\pi^\alpha \quad (\mathscr{P}_\alpha \equiv \mathscr{P}_{\alpha 2}), \tag{68}$$

and the total Hamiltonian H,

$$H = \mathscr{H}_{\min} + \{\Psi,\Omega\}, \tag{69}$$

where Ψ is a gauge fermion.

Returning to the procedure of extended BRST quantization we note that using a canonical transformation, one can always bring the function Ω^1 (or Ω^2) to the form (68). Indeed, consider the canonical transformation $P_A, Q^A \to P'_A, Q'^A$ with the generating function

$$X(P',Q) = X_0(p',q) + \mathscr{P}'_{\alpha a}(\Lambda^{-1})^\alpha_\beta(p',q)C^{\beta a} + \lambda'_\alpha(\Lambda^{-1})^\alpha_\beta(p',q)\pi^\beta, \tag{70}$$

where $X_0(p',q)$ is the generator of canonical transformation of the initial phase variables under which the constraints T_α take on the form

$$T_\alpha(p(p',q),q) = p'_\beta\Lambda^\beta_\alpha(p',q). \tag{71}$$

Here Λ^β_α is the matrix nondegenerate on the hypersurface $T_\alpha = 0$.

Under the canonical transformations (70),

$$\Omega^a \to \Omega'^a = p'_\alpha C'^{\alpha a} + \epsilon^{ab}\mathscr{P}'_{ab}\pi'^\alpha + \cdots. \tag{72}$$

The functions Ω'^a [(72)] satisfy Eqs. (5) and the boundary conditions

$$\left.\frac{\delta\Omega'^a}{\delta C'^{\alpha b}}\right|_{C'=\pi'=\mathscr{P}'=\lambda'=0} = \delta^a_b\,p'_\alpha,$$
$$\left.\frac{\delta\Omega'^a}{\delta\pi'^\alpha}\right|_{C'=\pi'=\lambda'=0} = \epsilon^{ab}\mathscr{P}'_{\alpha b}. \tag{73}$$

In the class of boundary conditions (73) there exists a solution of Eqs. (5) in the form

$$\tilde{\Omega}'^a = p'_\alpha C'^{\alpha a} + \epsilon^{ab}\mathscr{P}'_{\alpha b}\pi'^\alpha. \tag{74}$$

Any two solutions of Eqs. (5) satisfying one and the same boundary conditions are known to be related through a canonical transformation. Therefore, there exists a canonical transformation $P'_A, Q'^A \to P''_A, Q''^A$, such that

$$\Omega'^a \to \Omega''^a = p''_\alpha C''^{\alpha a} + \epsilon^{ab}\mathscr{P}''_{ab}\pi''^\alpha. \tag{75}$$

We now change from the variables P''_A, Q''^A to $\tilde{P}_A, \tilde{Q}^A$, employing a canonical transformation whose generating function is

$$Y(P'',\tilde{Q}) = X_0(p'',\tilde{q}) + \bar{\mathscr{P}}''_\alpha(\Lambda^{-1})^\alpha_\beta(p'',\tilde{q})\tilde{C}^\beta + \lambda''_\alpha\tilde{\pi}^\alpha + \mathscr{P}''_\alpha\tilde{\bar{C}}^\alpha. \tag{76}$$

Taking into account that under such a transformation the relations (hereafter the tilde is omitted)

$$p''_\alpha C''^{\alpha} = T_\alpha(p,q)C^\alpha + O((C,\bar{\mathscr{P}})^3),$$
$$\mathscr{P}''_\alpha\pi''^\alpha = \mathscr{P}_\alpha\pi^\alpha,$$

hold, we obtain

$$\Omega''^1 \to \Omega^1 = \Omega_{\min}(p,q,\bar{\mathscr{P}},C) + \mathscr{P}_\alpha\pi^\alpha, \tag{77}$$

where $\Omega_{\min}$ satisfies the equation and the boundary conditions (66). Consequently, the function Ω^1 [(77)] is canonically equivalent to the function Ω defined by (68).

Now let us represent the boson generating function $\mathscr{H}$ as

$$\mathscr{H} = \mathscr{H}_1 + \mathscr{H}_2,$$
$$\mathscr{H}_1 = \mathscr{H}|_{\bar{C}=\mathscr{P}=\pi=\lambda=0}, \quad \mathscr{H}_1|_{C=\bar{\mathscr{P}}=0} = H_0. \tag{78}$$

The equation $\{\mathscr{H},\Omega^1\} = 0$ falls into two parts:

$$\{\mathscr{H}_1,\Omega_{\min}\} = 0, \quad \{\mathscr{H}_2,\Omega^1\} = 0. \tag{79}$$

By virtue of (67) and (78), it follows from (79) that in the standard version of BRST quantization, $\mathscr{H}_1$ can be identified with $\mathscr{H}_{\min}$: $\mathscr{H}_1 = \mathscr{H}_{\min}$. Next, taking into account Lemma 2 and the fact that $\mathscr{H}_2 = 0$ when $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$, we conclude from (79) that $\mathscr{H}_2$ can be represented in the form

$$\mathscr{H}_2 = \{X,\Omega^1\},$$

where X is a certain fermion function.

Thus the total Hamiltonian of the extended BRST quantization admits the representation [cf. (69)]

$$H = \mathscr{H}_{\min} + \{X + \{\bar{\Omega},\Phi\},\Omega\}, \quad \bar{\Omega} \equiv \Omega^2.$$

We have proved that the extended BRST quantization in a special basis of canonical variables is a particular case of standard BRST quantization that corresponds to a special choice of the gauge.

It should also be noted that at the same time we have established the following fact: any regular solution of Eqs.

(5) with $\mathrm{ngh}(\Omega^a) = 1$ is canonically equivalent to a linearized solution of the type (75) with the same new ghost number. Indeed, suffice it to note that the generating functions of canonical transformations (70) and (76) conserve the new ghost number.

V. CONCLUDING REMARKS

We have proved the possibility of a consistent version of extended BRST quantization and established equivalence between extended and standard BRST quantizations.

The lemmas proved here also make it possible to describe the arbitrariness in solutions of the generating equations (5) and (7) due to ambiguity in the choice of boundary conditions (6) and (8). It is a well-known fact that the classical dynamics does not change if to the Hamiltonian H_0 we add a linear combination of constraints T_α: $H_0' = H_0 + \Lambda^\alpha T_\alpha$. Consider the solutions of Eqs. (7) that correspond to boundary conditions with H_0 and H_0'. Denote these solutions by $\mathscr{H}$ and $\mathscr{H}'$. Their difference $\Delta\mathscr{H} = \mathscr{H}' - \mathscr{H}$ obviously satisfies the condition $\Delta\mathscr{H} = 0$ for $T_\alpha = \mathscr{P}_{\alpha a} = \lambda_\alpha = 0$ and the equations $\{\Delta\mathscr{H}, \Omega^a\} = 0$. It is immediate from Lemma 2 that

$$\Delta\mathscr{H} = \tfrac{1}{2}\epsilon_{ab}\{\Omega^b, \{\Omega^a, Z\}\},$$

with a certain function Z. Consequently, passing over from the boundary condition with H_0 to $H_0' = H_0 + \Lambda^\alpha T_\alpha$ corresponds to a change of the gauge in the unitarizing Hamiltonian H [(9)].

The classical dynamics is also known to be independent of the choice of linear combinations of constraints T_α. Let us investigate the behavior of solutions of Eqs. (5) in passing over from the constraints T_α to $T'_\alpha = \Sigma^\beta_\alpha T_\beta$, where Σ^β_α is a certain nondegenerate matrix. We have already seen that any solution of Eqs. (5) is canonically equivalent to a linearized solution. Let Ω^a and Ω'^a be solutions of Eqs. (5) that correspond to boundary conditions T_α and T'_α. These solutions are canonically equivalent to one and the same linearized solution and hence are related one to another through a certain canonical transformation.

Thus all the arbitrariness existing in the solutions of Eqs. (5) is described by a canonical transformation, and the arbitrariness in (7) transforms into the change of the gauge in the total unitarizing Hamiltonian.

APPENDIX: THE EXISTENCE THEOREM IN THE STANDARD FORMULATION

We present here, from methodical considerations, the proof of the existence theorem for generating equations of the gauge algebra in the standard (nonextended) BRST quantization procedure. It should be noted that the proof of this theorem is given in the review (Ref. 4), but it is rather cumbersome there. Here we shall present a simpler version of the proof based on a systematic application of the algebraic mechanism the reader may find in Sec. III of this paper.

So, in the minimal sector

$$\Gamma_{\min} = (p_i, q^i; \overline{\mathscr{P}}_\alpha, C^\alpha), \tag{A1}$$

we consider the equations

$$\{\Omega_{\min}, \Omega_{\min}\} = 0, \quad \left.\frac{\delta\Omega_{\min}}{\delta C^\alpha}\right|_{C = \overline{\mathscr{P}} = 0} = T_\alpha, \tag{A2}$$

$$\{\mathscr{H}_{\min}, \Omega_{\min}\} = 0, \quad \mathscr{H}_{\min}|_{C = \overline{\mathscr{P}} = 0} = H_0, \tag{A3}$$

whose solution is sought in the form

$$\Omega_{\min} = \sum_{n=1}^{\infty} \Omega_n, \quad \Omega_n \sim \overline{\mathscr{P}}^{n-1} C^n, \tag{A4}$$

$$\mathscr{H}_{\min} = H_0 + \sum_{n=1}^{\infty} \mathscr{H}_n, \quad \mathscr{H}_n \sim \overline{\mathscr{P}}^n C^n. \tag{A5}$$

The lowest approximation for $\Omega_{\min}$ is

$$\Omega_1 = T_\alpha C^\alpha. \tag{A6}$$

For higher approximations, the first of Eqs. (A2) gives the recurrent relations

$$W\Omega_{n+1} + B_{n+1} = 0, \quad n = 1,2,\ldots, \tag{A7}$$

where

$$W = T_\alpha \frac{\delta}{\delta \overline{\mathscr{P}}_\alpha}, \quad W^2 = 0, \tag{A8}$$

$$B_{n+1} \equiv \tfrac{1}{2}\{[\Omega]_n, [\Omega]_n\}_{n+1}, \quad [\Omega]_n \equiv \sum_{k=1}^{n} \Omega_k, \tag{A9}$$

and $\{\ ,\ \}_n$ implies the C^n approximation of the bracket $\{\ ,\ \}$.

The lowest approximation for $\mathscr{H}_{\min}$ is obviously H_0 and for the approximations following this one the first of Eqs. (A3) yields

$$W\mathscr{H}_{n+1} = D_{n+1}, \quad n = 0,1,\ldots, \tag{A10}$$

where

$$D_{n+1} \equiv \{[\mathscr{H}]_n, [\Omega]_{n+1}\}_{n+1}, \tag{A11}$$

$$[\mathscr{H}]_n \equiv \sum_{k=0}^{n} \mathscr{H}_k. \tag{A12}$$

For our further puposes we need the following lemma: any regular solution of the equation

$$WX = 0, \tag{A13}$$

which vanishes when $T_\alpha = 0$, $\overline{\mathscr{P}}_\alpha = 0$, has the form

$$X = WY. \tag{A14}$$

To prove this lemma, it suffices, as in the main text, to introduce the operator Γ with the following properties:

$$\Gamma^2 = 0, \quad \Gamma W + W\Gamma = N, \tag{A15}$$

where N is the operator positive on the solutions of the above-mentioned class.

The operator Γ corresponding to W from (A8) is chosen to take the form

$$\Gamma = \overline{\mathscr{P}}_\alpha \frac{\delta_l}{\delta T_\alpha} \tag{A16}$$

and then

$$N = T_\alpha \frac{\delta_l}{\delta T_\alpha} + \overline{\mathscr{P}}_\alpha \frac{\delta}{\delta \overline{\mathscr{P}}_\alpha}. \tag{A17}$$

Applying to (A13) the operator Γ from the left and taking into account (A15), we are led to the assertion of the lemma.

Turning again to the proof of the existence theorem, we shall first consider Eqs. (A2). Suppose that these equations

are satisfied to an accuracy of C^n terms, that is, the functions Ω_k, $k = 1,...,n$, exist. Then from the identity

$$\{\{[\Omega]_n,[\Omega]_n\},[\Omega]_n\}\equiv 0 \tag{A18}$$

in the C^{n+1} order we have

$$WB_{n+1} = 0 . \tag{A19}$$

From (A4) and (A9) it follows that

$$B_{n+1}|_{T_\alpha = \bar{\mathscr{P}}_\alpha = 0} = 0 . \tag{A20}$$

Then, by virtue of the lemma, we find, from (A19) and (A20),

$$B_{n+1} = -W\Omega_{n+1} ; \tag{A21}$$

this equality being the *definition* of the function $\Omega_{n+1} \sim \bar{\mathscr{P}}^n C^{n+1}$.

Thus (A21) implies that Eq. (A2) is already satisfied in the C^{n+1} approximation. Induction by n completes the proof of the existence of a solution of Eq. (A2).

We now turn to Eq. (A3). It should only be noted here that when repeating quite similar arguments, rather than the identity (A18) one should use the relation

$$\{\{[\mathscr{H}]_n,[\Omega]_{n+1}\},[\Omega]_{n+1}\}\equiv O(C^{n+2}) ,$$

which is identically fulfilled with respect to $[H]_n$ by virtue of (A2).

[1]E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. B **55**, 224 (1975).
[2]I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).
[3]E. S. Fradkin and T. E. Fradkina, Phys. Lett. B **72**, 343 (1978).
[4]M. Henneaux, Phys. Rev. **126**, 2 (1985).
[5]C. Becchi, A. Rouet, and R. Stora, Commun. Math. Phys. **42**, 127 (1975).
[6]I. V. Tyutin, Preprint of the P. N. Lebedev Institute, N °39, 1975.
[7]G. Curci and R. Ferrari, Nuovo Cimento A **32**, 151 (1976).
[8]I. Ojima, Prog. Theor. Phys. Suppl. **66**, 324 (1979).
[9]N. Nakanishi and I. Ojima, Z. Phys. C **6**, 155 (1980).
[10]L. Alvarez-Gaume and L. Baulieu, Nucl. Phys. B **212**, 255 (1983).
[11]V. P. Spiridonov, Nucl. Phys. B **308**, 527 (1988).
[12]E. S. Fradkin, Proceedings of the X Winter School of Theoretical Physics in Karpacs, N° 207, 1973, p. 93.
[13]I. A. Batalin and G. A. Vilkovisky, J. Math. Phys. **26**, 172 (1985).

Superfield and matrix realizations of highest weight representations for osp(*m*/2*n*)

R. Le Blanc
Rutherford Physics Building, Physics Department, McGill University, Montreal, Quebec H3A 2T8, Canada
D. J. Rowe
Physics Department, University of Toronto, Toronto, Ontario M5S 1A7, Canada

(Received 14 June 1989; accepted for publication 23 August 1989)

A differential representation of the classical Lie superalgebra osp(*m*/2*n*) acting on superfield functions is given. This representation is used to construct matrix representations of the finite-dimensional irreducible representations of the algebra. Inner products on the irreducible spaces are discussed and classes of star and grade-star equivalent representations are identified.

I. INTRODUCTION

It has recently been shown in considerable detail[1–5] how vector coherent state (VCS) theory provides, under certain conditions, means to induce irreducible highest weight representations of a complex Lie algebra or superalgebra $\mathbf{g}$ from highest weight irreps of a subalgebra $\mathbf{n}_0 \subset \mathbf{g}$ of rank $(\mathbf{n}_0) = \text{rank}(\mathbf{g})$.

VCS theory makes use of a Z gradation of $\mathbf{g}$,

$$\mathbf{g} = \mathbf{n}_0 + \sum_{i=1,2,\ldots} \mathbf{n}_{\pm i}\,, \tag{1.1}$$

defined in terms of a grading operator $\hat{Z}$, belonging to a Cartan subalgebra, by the equation

$$\langle \hat{Z}, x\rangle = ix, \quad \forall x \in \mathbf{n}_i\,, \tag{1.2}$$

where the bracket $\langle \cdot,\cdot \rangle$ is the graded Lie product. Such a gradation endows the algebra with a Z graded structure. The zero grade component $\mathbf{n}_0$ is called the *stability subalgebra*.

By definition, a Lie superalgebra, endowed with a Z graded structure, is a vector space $\mathbf{g}$ that (i) is a direct sum of vector subspaces $\mathbf{n}_i$, where the index i takes integer values; and that (ii) has a bilinear product that satisfies

$$\langle x,y\rangle \in \mathbf{n}_{i+j}\,, \tag{1.3a}$$

$$\langle x,y\rangle = -(-1)^{ij}\langle y,x\rangle\,, \tag{1.3b}$$

$$\langle x,\langle y,z\rangle\rangle = \langle\langle x,y\rangle,z\rangle + (-1)^{ij}\langle y,\langle x,z\rangle\rangle\,, \tag{1.3c}$$

for $x \in \mathbf{n}_i$, $y \in \mathbf{n}_j$, and any z in $\mathbf{g}$.

It can be shown that any classical Lie superalgebra can be assigned a convenient Z graded structure, with either $i_{\max} = 1$ or 2. We then have for a classical Lie superalgebra that the even sector

$$\mathbf{g}_{\bar{0}} = \sum_{i \text{ even}} \mathbf{n}_i \tag{1.4a}$$

of the superalgebra is its Lie subalgebra $\mathbf{g}_{\bar{0}} = \mathbf{n}_0$ or $\mathbf{n}_0 + \mathbf{n}_{-2} + \mathbf{n}_{+2}$. The odd sector is

$$\mathbf{g}_{\bar{1}} = \sum_{i \text{ odd}} \mathbf{n}_i = \mathbf{n}_{-1} + \mathbf{n}_{+1}\,. \tag{1.4b}$$

The subspaces

$$\mathbf{n}_+ = \sum_{i>0}^{i_{\max}} \mathbf{n}_i\,,$$

$$\mathbf{n}_- = \sum_{i<0}^{-i_{\max}} \mathbf{n}_i \tag{1.5}$$

are nilpotent subalgebras of raising and lowering operators. Since each level $\mathbf{n}_i$ is invariant under the adjoint action $\text{ad}_{\mathbf{n}_0}$ of the stability algebra $\mathbf{n}_0$, the subalgebras $\mathbf{n}_\pm$ are generally reducible under $\text{ad}_{\mathbf{n}_0}$.

For $i_{\max} = 1$, the subalgebras $\mathbf{n}_\pm$ are necessarily Abelian. The first applications of VCS theory addressed Lie algebra cases with $\mathbf{n}_\pm$ Abelian,[1–3,5] which present some simplifying features. The Lie superalgebra gl(m/n) with $\mathbf{n}_\pm$ Abelian and $\mathbf{g}_{\bar{0}} = \mathbf{n}_0$ has been studied in Ref. 6.

Recent developments[4] have shown that VCS theory applies equally well to Lie algebras with $i_{\max} \geqslant 2$ and $\mathbf{n}_\pm$ non-Abelian. The aim of this paper is to study classical superalgebras for which

$$\mathbf{g}_{\bar{0}} = \mathbf{n}_0 + \mathbf{n}_{-2} + \mathbf{n}_{+2}$$

and for which $\mathbf{n}_\pm$ are non-Abelian and reducible under $\mathbf{n}_0$. More precisely, we consider the osp(m/2n) superalgebras for all m and n. Among the extra complications arising, note that, in contrast to the Abelian superalgebraic case for which the coset representative $\exp X$ of G/N_0, with $X \in \mathbf{n}_+$, is parametrized by Grassmann variables only, the non-Abelian case requires a parametrization by both Bargmann and Grassmann variables since $\mathbf{n}_+$ contains elements from both the odd and even parts of the superalgebra $\mathbf{g}$ (see also Farmer and Jarvis, Ref. 7).

To improve the efficiency of VCS theory as a tool for the explicit construction of irreducible representations of simple or semisimple algebras, we introduce herein some significant developments over previous expositions of the subject.

We give an explicit construction of the irreducible VCS module as a submodule of the vector–Grassmann–Bargmann (VGB) space. This is achieved by expanding VCS states on an *ab initio* orthonormal VGB basis. The expansion operator $\mathcal{O}$, defined by Eq. (6.1), thereby defines a mapping from theVGB space onto the irreducible VCS subspace. Note that previous treatments skipped this expansion and concentrated rather on the construction of matrix representations of the algebra.

We introduce a projection operator in Eq. (6.3) that significantly simplifies the identification of the highest

 0022-2488/90/010014-23$03.00

weight states of the $\mathfrak{g}_{\bar{0}}$ submodules contained in an irreducible VCS module of $\mathfrak{g}$. Previous treatments unnecessarily linked this identification to the independent problem of constructing equivalent Hermitian (star or grade-star) representations. Following the introduction of the projector, the algebraic derivation of relevant recursion formulas [Eqs. (6.2) and (7.13)] is considerably simplified.

The identification of classes of star or grade-star equivalent representations[8] is shown to be quite simple within the present formalism. One has only to identify the circumstances under which the expression $(-1)^{\Phi}\mathcal{O}$, where $(-1)^{\Phi}$ is a phase signature proper to either the star or grade-star case [Eq. (7.14)], defines a positive-definite operator $\mathcal{K}^2$ on the VCS irreducible module.

Also noteworthy are the following points.

We show how VCS theory provides a rationale for Kac's subsidiary conditions[9] for the finite-dimensionality of irreducible modules of the osp$(m/2n)$ algebra (Sec. VI C).

We show how VCS theory predicts the decoupling[10] of $\mathfrak{g}$-invariant submodules for atypical representations (Sec. VI D).

We construct *explicitly* all the irreducible finite-dimensional representations of the superalgebras osp$(1/2n)$ for $n \geqslant 1$ (Secs. VIII and IX B), of osp$(2/2)$ (Sec. IX C) and osp$(3/2)$ (Sec. IX D). Analytical expressions are given for all matrix elements of the representations. We identify all classes of star and grade-star equivalent representations for these algebras and give the (generalized) unitarized forms of the representations with their properties under (generalized) adjoint conjugation.

II. THE osp$(m/2n)$ SUPERALGEBRA

A. The even subalgebra $\mathfrak{g}_{\bar{0}} = \mathrm{so}(m) \oplus \mathrm{sp}(2n)$

The even part $\mathfrak{g}_{\bar{0}}$ of the Lie superalgebra $\mathfrak{g} = \mathrm{osp}(m/2n)$ is the direct sum[9,11] $\mathrm{so}(m) \oplus \mathrm{sp}(2n)$ [and only $\mathrm{sp}(2n)$ if $m = 1$].

1. The so(m) subalgebra

A basis for so(m), $m \geqslant 2$, is given by $\{H_{ab} = -H_{ba};\ 1 \leqslant a < b \leqslant m\}$ with commutation relations

$$\langle H_{ab}, H_{cd} \rangle = \delta_{bc} H_{ad} + \delta_{bd} H_{ca} + \delta_{ad} H_{bc} + \delta_{ac} H_{db} \quad (2.1)$$

(Z_2 graded commutators are denoted $\langle \ldots, \ldots \rangle$ in this manuscript).

It is convenient to introduce a set of Cartan (raising, lowering, and weight) operators for so(m): we set (we use the convention that $1 \leqslant a,b,c,d,\ldots \leqslant m$ while $1 \leqslant i,j,k,l,\ldots \leqslant r$) with $r = [m/2]$

$$\mathscr{A}_{ij} = -\mathscr{A}_{ji} = \tfrac{1}{2}(H_{2i,2j} - H_{2i-1,2j-1} - iH_{2i-1,2j} - iH_{2i,2j-1}), \quad 1 \leqslant i < j \leqslant r, \quad (2.2a)$$

$$\mathscr{B}_{ij} = -\mathscr{B}_{ji} = \tfrac{1}{2}(H_{2i-1,2j-1} - H_{2i,2j} - iH_{2i-1,2j} - iH_{2i,2j-1}), \quad 1 \leqslant i < j \leqslant r, \quad (2.2b)$$

$$\mathscr{C}_{ij} = \tfrac{1}{2}(H_{2i,2j} + H_{2i-1,2j-1} - iH_{2i-1,2j} + iH_{2i,2j-1}), \quad 1 \leqslant i, j \leqslant r, \quad (2.2c)$$

to which we add, if $m = 2r + 1$,

$$\mathscr{D}_i = -(1/\sqrt{2})(H_{2r+1,2i-1} + iH_{2r+1,2i}), \quad 1 \leqslant i \leqslant r, \quad (2.2d)$$

$$\mathscr{E}_i = (1/\sqrt{2})(H_{2r+1,2i-1} - iH_{2r+1,2i}), \quad 1 \leqslant i \leqslant r. \quad (2.2e)$$

From (2.1), we find the following set of commutation relations:

$$\langle \mathscr{C}_{ij}, \mathscr{A}_{kl} \rangle = \delta_{jk}\mathscr{A}_{il} + \delta_{jl}\mathscr{A}_{ki}, \quad (2.3a)$$

$$\langle \mathscr{C}_{ij}, \mathscr{D}_k \rangle = \delta_{jk}\mathscr{D}_i, \quad (2.3b)$$

$$\langle \mathscr{C}_{ij}, \mathscr{C}_{kl} \rangle = \delta_{jk}\mathscr{C}_{il} - \delta_{il}\mathscr{C}_{kj}, \quad (2.3c)$$

$$\langle \mathscr{C}_{ij}, \mathscr{E}_k \rangle = -\delta_{ik}\mathscr{E}_j, \quad (2.3d)$$

$$\langle \mathscr{C}_{ij}, \mathscr{B}_{kl} \rangle = -\delta_{ik}\mathscr{B}_{jl} - \delta_{il}\mathscr{B}_{kj}. \quad (2.3e)$$

From these we conclude that the set $\{\mathscr{C}_{ij},\ 1 \leqslant i,j \leqslant r\}$ generates a u(r) Lie algebra [or rather its complexification, not to be confused with the u$(n) \subset$ sp$(2n)$ algebra to be introduced below] while the sets $\{\mathscr{A}_{ij}\}$, $\{\mathscr{D}_i\}$, $\{\mathscr{E}_i\}$, and $\{\mathscr{B}_{ij}\}$ span, under $\mathrm{ad}_{\mathrm{u}(r)}$, irreducible tensorial sets of rank $\{11\}$, $\{1\}$, $\{-1\}$, and $\{-1-1\}$, respectively.

Similarly, we find, whenever $m = 2r + 1$, that

$$\langle \mathscr{D}_i, \mathscr{B}_{kl} \rangle = \delta_{il}\mathscr{E}_k - \delta_{ik}\mathscr{E}_l, \quad (2.3f)$$

$$\langle \mathscr{E}_i, \mathscr{A}_{kl} \rangle = \delta_{ik}\mathscr{D}_l - \delta_{il}\mathscr{D}_k, \quad (2.3g)$$

from which we conclude that the union of the two u(r) irreducible tensorial sets $\{\mathscr{D}_i\}$ and $\{\mathscr{E}_i\}$ spans an irreducible tensorial set of rank [1] under the adjoint action of the algebra

$$\mathrm{so}(2r) = \mathrm{span}\{\mathscr{C}_{ij},\ 1 \leqslant i,j \leqslant r;\quad \mathscr{A}_{ij}, \mathscr{B}_{ij},\quad 1 \leqslant i < j \leqslant r\},$$

where the so$(2r)$ algebra closes through

$$\langle \mathscr{B}_{ij}, \mathscr{A}_{kl} \rangle = \delta_{il}\mathscr{C}_{kj} - \delta_{ik}\mathscr{C}_{lj} + \delta_{jk}\mathscr{C}_{li} - \delta_{jl}\mathscr{C}_{ki}. \quad (2.3h)$$

Finally, whenever $m = 2r + 1$, we verify that the so(m) algebra closes through

$$\langle \mathscr{D}_i, \mathscr{D}_j \rangle = \mathscr{A}_{ij}, \quad (2.3i)$$

$$\langle \mathscr{E}_i, \mathscr{E}_j \rangle = \mathscr{B}_{ji}, \quad (2.3j)$$

$$\langle \mathscr{D}_i, \mathscr{E}_j \rangle = \mathscr{C}_{ij}. \quad (2.3k)$$

All other so(m) commutators involving elements from the basis (2.2) vanish.

Associated with the generators (2.2) of so(m) are, for $r = [m/2] \geqslant 2$,

the set of positive roots $+(\epsilon_i - \epsilon_j)$
associated with $\mathscr{C}_{ij}$, $1 \leqslant i < j \leqslant r$, (2.4a)

the set of negative roots $-(\epsilon_i - \epsilon_j)$
associated with $\mathscr{C}_{ji}$, $1 \leqslant i < j \leqslant r$, (2.4b)

the set of positive roots $+(\epsilon_i + \epsilon_j)$
associated with $\mathscr{A}_{ij}$, $1 \leqslant i < j \leqslant r$, (2.4c)

the set of negative roots $-(\epsilon_i + \epsilon_j)$
associated with $\mathscr{B}_{ij}$, $1 \leqslant i < j \leqslant r$; (2.4d)

and, for $m = 2r + 1$,

the set of positive roots $+\epsilon_i$
associated with $\mathscr{D}_i$, $1 \leqslant i \leqslant r$, (2.4e)

the set of negative roots $-\epsilon_i$
associated with $\mathscr{E}_i$, $1 \leqslant i \leqslant r$. (2.4f)

For $so(2)\sim u(1)$ ($m=2$), we only have the weight operator H_{12} with an associated null root.

2. The sp(2n) subalgebra

Similarly, a Cartan basis for $sp(2n)\supset u(n)$, $n\geqslant 1$, is given by

$$\{C_{\alpha\beta},A_{\alpha\beta}=A_{\beta\alpha},\quad B_{\alpha\beta}=B_{\beta\alpha};\quad 1\leqslant\alpha,\beta\leqslant n\} \tag{2.5}$$

with commutation relations

$$\langle C_{\alpha\beta},A_{\mu\nu}\rangle=\delta_{\beta\mu}A_{\alpha\nu}+\delta_{\beta\nu}A_{\mu\alpha}\,, \tag{2.6a}$$

$$\langle C_{\alpha\beta},C_{\mu\nu}\rangle=\delta_{\beta\mu}C_{\alpha\nu}-\delta_{\alpha\nu}C_{\mu\beta}\,, \tag{2.6b}$$

$$\langle C_{\alpha\beta},B_{\mu\nu}\rangle=-\delta_{\alpha\mu}B_{\beta\nu}-\delta_{\alpha\nu}B_{\mu\beta}\,, \tag{2.6c}$$

and

$$\langle A_{\alpha\beta},B_{\mu\nu}\rangle=\delta_{\alpha\mu}C_{\beta\nu}+\delta_{\alpha\nu}C_{\beta\mu}+\delta_{\beta\mu}C_{\alpha\nu}+\delta_{\beta\nu}C_{\alpha\mu}\,. \tag{2.6d}$$

All other $sp(2n)$ commutators involving elements from the basis (2.5) vanish. Obviously,

$$\mathbf{n}_0''\equiv\text{span}\{C_{\alpha\beta},\quad 1\leqslant\alpha,\beta\leqslant n\} \tag{2.7a}$$

generates a $u(n)$ Lie algebra (more precisely, its complexification) while the Abelian subalgebras

$$\mathbf{n}_{+2}\equiv\text{span}\{A_{\alpha\beta},\quad 1\leqslant\alpha,\beta\leqslant n\} \tag{2.7b}$$

and

$$\mathbf{n}_{-2}\equiv\text{span}\{B_{\alpha\beta},\quad 1\leqslant\alpha,\beta\leqslant n\} \tag{2.7c}$$

span, under $\text{ad}_{u(n)}$, irreducible tensorial sets of rank $\{2\}$ and $\{-2\}$, respectively.

Associated with the generators (2.5) of $sp(2n)$ are

the set of positive roots $+(\delta_\alpha-\delta_\beta)$

associated with $C_{\alpha\beta}$, $1\leqslant\alpha<\beta\leqslant n$, (2.8a)

the set of negative roots $-(\delta_\alpha-\delta_\beta)$

associated with $C_{\beta\alpha}$, $1\leqslant\alpha<\beta\leqslant n$, (2.8b)

the set of positive roots $+(\delta_\alpha+\delta_\beta)$

associated with $A_{\alpha\beta}$, $1\leqslant\alpha,\beta\leqslant n$, (2.8c)

the set of negative roots $-(\delta_\alpha+\delta_\beta)$

associated with $B_{\alpha\beta}$, $1\leqslant\alpha,\beta\leqslant n$. (2.8d)

3. Cartan subalgebra and Z grading operator

The $sp(2n)$ roots (2.8) and their $so(m)$ counterparts (2.4) belong to the space $\mathbf{h}^*$ dual to the Cartan subalgebra $\mathbf{h}$ of $\mathbf{g}_{\bar 0}$ generated by the set of weight generators

$$\mathbf{h}=\text{span}\{h_i,\quad 1\leqslant i\leqslant r;\quad h_\alpha,\quad 1\leqslant\alpha\leqslant n\}\,, \tag{2.9a}$$

where

$$h_i=\mathscr{C}_{ii}\quad(\text{no sum on } i)\,, \tag{2.9b}$$

$$h_\alpha=C_{\alpha\alpha}\quad(\text{no sum on }\alpha)\,. \tag{2.9c}$$

We therefore have a dual basis $\{\epsilon_i,i=1,\ldots,r;\delta_\alpha,\alpha=1,\ldots,n\}$ with

$$\epsilon_i(h_j)=\delta_{ij},\quad \delta_\alpha(h_j)=0\,,$$
$$\epsilon_i(h_\alpha)=0,\quad \delta_\alpha(h_\beta)=\delta_{\alpha\beta}\,. \tag{2.10}$$

The trace operator

$$\hat Z=\sum_{\beta=1}^{n}h_\beta\in\mathbf{h} \tag{2.11}$$

naturally grades $sp(2n)$ into the three subalgebras $\mathbf{n}_0''$, $\mathbf{n}_{+2}$, and $\mathbf{n}_{-2}$ with Z grade 0, $+2$, and -2, respectively. Since the generators of

$$so(m)\equiv\mathbf{n}_0' \tag{2.12}$$

commute with the generators of $sp(2n)$ in general and with $\hat Z$ in particular, and since all generators of $u(n)\equiv\mathbf{n}_0''$ similarly commute with $\hat Z$, the subalgebra

$$\mathbf{n}_0=\mathbf{n}_0'\oplus\mathbf{n}_0''=so(m)\oplus u(n) \tag{2.13}$$

can be regarded as a stability algebra (null Z grade subalgebra) leaving invariant all subalgebras $\mathbf{n}_{\pm i}$ of given Z grade and, by extension, subspaces of representations of given Z grade. We remark that $\hat Z$ belongs to the center of $\mathbf{n}_0$:

$$\langle\hat Z,x\rangle=0,\quad\forall x\in\mathbf{n}_0\,. \tag{2.14}$$

B. The odd subalgebra $\mathbf{g}_{\bar 1}$ of osp(m/2n)

The odd part $\mathbf{g}_{\bar 1}$ of the Lie superalgebra $\mathbf{g}=osp(m/2n)$, $m\neq 2$, carries an irreducible representation[9,11] $[1]:\langle 1\rangle$ of its even part $\mathbf{g}_{\bar 0}=so(m)\oplus sp(2n)$, where $[1]$ and $\langle 1\rangle$ stand for the fundamental irreps of $so(m)$ and $sp(2n)$, respectively. Under $u(n)$, the $sp(2n)$ irrep $\langle 1\rangle$ splits into the two irreps $\{1\}$ and $\{-1\}$:

$$sp(2n)\downarrow u(n)\ :\ \langle 1\rangle\downarrow\{1\}+\{-1\}\,.$$

[For the case $m=2$, $\mathbf{g}_{\bar 1}$ is reducible and spans the irreducible representations $[\pm 1]$: $\langle 1\rangle$ of $so(2)\oplus sp(2n)\sim u(1)\oplus sp(2n)$.] We therefore introduce the $\mathbf{n}_0=so(m)\oplus u(n)$ tensorial sets

$$\mathbf{n}_{+1}=\text{span}\{D_{a\alpha};\quad 1\leqslant a\leqslant m,\quad 1\leqslant\alpha\leqslant n\} \tag{2.15a}$$

and

$$\mathbf{n}_{-1}=\text{span}\{E_{a\alpha};\quad 1\leqslant a\leqslant m,\quad 1\leqslant\alpha\leqslant n\} \tag{2.15b}$$

spanning the irreducible representations $[1]:\{+1\}$ and $[1]:\{-1\}$ of $\mathbf{n}_0$ (and, for $m=2$, the irreducible representations $[\pm 1]:\{+1\}$ and $[\pm 1]:\{-1\}$ of $\mathbf{n}_0$), respectively. We necessarily have

$$\langle H_{ab},D_{c\mu}\rangle=\delta_{bc}D_{a\mu}-\delta_{ac}D_{b\mu}\,, \tag{2.16a}$$

$$\langle C_{\alpha\beta},D_{c\mu}\rangle=\delta_{\beta\mu}D_{c\alpha}\,, \tag{2.16b}$$

$$\langle H_{ab},E_{c\mu}\rangle=\delta_{bc}E_{a\mu}-\delta_{ac}E_{b\mu}\,, \tag{2.16c}$$

$$\langle C_{\alpha\beta},E_{c\mu}\rangle=-\delta_{\beta\mu}E_{c\alpha}\,, \tag{2.16d}$$

while

$$\langle A_{\alpha\beta},E_{c\mu}\rangle=\delta_{\beta\mu}D_{c\alpha}+\delta_{\alpha\mu}D_{c\beta}\,, \tag{2.16e}$$

$$\langle B_{\alpha\beta},D_{c\mu}\rangle=\delta_{\beta\mu}E_{c\alpha}+\delta_{\alpha\mu}E_{c\beta}\,, \tag{2.16f}$$

$$\langle A_{\alpha\beta},D_{c\mu}\rangle=\langle B_{\alpha\beta},E_{c\mu}\rangle=0\,. \tag{2.16g}$$

Obviously, $\mathbf{n}_{+1}$ and $\mathbf{n}_{-1}$ have Z grade $+1$ and -1, respectively.

In order for

$$\mathbf{g}=\mathbf{g}_{\bar 0}\oplus\mathbf{g}_{\bar 1}=(\mathbf{n}_0\oplus\mathbf{n}_{+2}\oplus\mathbf{n}_{-2})\oplus(\mathbf{n}_{+1}\oplus\mathbf{n}_{-1})$$

to close upon the Lie superalgebra $osp(m/2n)$, one must specify a set of anticommutation relations $\mathbf{g}_{\bar 1}\times\mathbf{g}_{\bar 1}\to\mathbf{g}_{\bar 0}$ compatible with the Jacobi superidentity

$$\langle P,\langle Q,R\rangle\rangle = \langle\langle P,Q\rangle,R\rangle + (-1)^{\zeta(P)\cdot\zeta(Q)}\langle P,\langle Q,R\rangle\rangle\,, \tag{2.17}$$

where $\zeta(P)$ is the Z_2 grade of the operator $P\in\mathbf{g}=\mathrm{osp}(m/2n)$, herein assumed to be consistent with the Z grade of P. We verify that

$$\langle D_{a\alpha},D_{b\beta}\rangle = -\delta_{ab}A_{\alpha\beta}\,, \tag{2.18a}$$

$$\langle E_{a\alpha},E_{b\beta}\rangle = \delta_{ab}B_{\alpha\beta}\,, \tag{2.18b}$$

$$\langle D_{a\alpha},E_{b\beta}\rangle = \delta_{ab}C_{\alpha\beta} - \delta_{\alpha\beta}H_{ab} \tag{2.18c}$$

satisfy (2.17). Equations (2.18) may present a normalization different from those appearing in the literature.

Associated with the generators of $\mathbf{g}_{\bar{1}}$ are, for $r=[m/2]\geqslant 1$,

the set of positive roots $(\pm\epsilon_i+\delta_\alpha)$ associated with

$$\tilde{D}_{\pm i,\alpha} = (1/\sqrt{2})(D_{2i-1,\alpha} \pm iD_{2i,\alpha}),\quad 1\leqslant i\leqslant r,\quad 1\leqslant\alpha\leqslant n\,, \tag{2.19a}$$

the set of negative roots $(\pm\epsilon_i-\delta_\alpha)$ associated with

$$\tilde{E}_{\pm i,\alpha} = (1/\sqrt{2})(E_{2i-1,\alpha} \pm iE_{2i,\alpha}),\quad 1\leqslant i\leqslant r,\quad 1\leqslant\alpha\leqslant n\,; \tag{2.19b}$$

and, if $m=2r+1\geqslant 3$,

the set of positive roots $+\delta_\alpha$ associated with

$$\tilde{D}_{0\alpha} = D_{2r+1,\alpha},\quad 1\leqslant\alpha\leqslant n\,, \tag{2.19c}$$

the set of negative roots $-\delta_\alpha$ associated with

$$\tilde{E}_{0\alpha} = E_{2r+1,\alpha},\quad 1\leqslant\alpha\leqslant n\,. \tag{2.19d}$$

C. Supersymmetric invariant bilinear form on root space

The set of all even roots (2.4) and (2.8) is referred to as $\Delta_{\bar{0}}$ while the set of all odd roots (2.19) is referred to as $\Delta_{\bar{1}}$. We further distinguish[9,11] the sets of roots

$$\overline{\Delta}_{\bar{1}}^{\pm} = \{\pm(+\epsilon_i+\delta_\alpha);\quad \pm(-\epsilon_i+\delta_\alpha);\quad 1\leqslant i\leqslant r,\quad 1\leqslant\alpha\leqslant n\}\,, \tag{2.20a}$$

$$= \varnothing,\quad \text{for } m=1\ (r=0)\,, \tag{2.20b}$$

i.e., the sets of positive and negative odd roots $\vartheta\in\mathbf{h}^*$ such that $2\vartheta\notin\Delta_{\bar{0}}$.

We find[9,11] that half the sum of the even positive roots $\rho_{\bar{0}}$ is given by

$$\rho_{\bar{0}} = \rho_{\bar{0}}^{\mathrm{so}(m)} + \rho_{\bar{0}}^{\mathrm{sp}(2n)}\,, \tag{2.21a}$$

where

$$\rho_{\bar{0}}^{\mathrm{sp}(2n)} = \frac{1}{2}\sum_{\alpha=1}^{n}(2n+2-2\alpha)\delta_\alpha\,, \tag{2.21b}$$

$$\rho_{\bar{0}}^{\mathrm{so}(m)} = 0,\quad m=1,2\,, \tag{2.21c}$$

$$\rho_{\bar{0}}^{\mathrm{so}(m)} = \frac{1}{2}\sum_{i=1}^{r}(2r+1-2i)\epsilon_i,\quad m=2r+1\geqslant 3\,, \tag{2.21d}$$

$$\rho_{\bar{0}}^{\mathrm{so}(m)} = \frac{1}{2}\sum_{i=1}^{r}(2r-2i)\epsilon_i,\quad m=2r\geqslant 4\,, \tag{2.21e}$$

while half the sum of the odd positive roots $(\rho_{\bar{1}})$ is given by

$$\rho_{\bar{1}} = \frac{m}{2}\sum_{\alpha=1}^{n}\delta_\alpha\,, \tag{2.21f}$$

for all m. Finally, we denote by ρ the supersum

$$\rho = \rho_{\bar{0}} - \rho_{\bar{1}}\,. \tag{2.22}$$

A supersymmetric invariant and nondegenerate bilinear form on $\mathbf{h}^*$ is given by[9,11]

$$(\epsilon_i,\epsilon_j) = -\delta_{ij}\,, \tag{2.23a}$$

$$(\epsilon_i,\delta_\alpha) = 0\,, \tag{2.23b}$$

$$(\delta_\alpha,\delta_\beta) = +\delta_{\alpha\beta}\,, \tag{2.23c}$$

implying

$$(\vartheta,\vartheta) = 0,\quad \forall\vartheta\in\overline{\Delta}_{\bar{1}} = \overline{\Delta}_{\bar{1}}^{+} + \overline{\Delta}_{\bar{1}}^{-}\,. \tag{2.24}$$

D. Simple roots for osp $(m/2n)$

We denote by Π a distinguished set[9] of simple roots for $\mathrm{osp}(m/2n)$ containing a single odd root. It is given (i) for $\mathrm{osp}(1/2n)$, by

$$\Pi = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,\delta_n\}\,, \tag{2.25a}$$

with δ_n the odd root; (ii) for $\mathrm{osp}(2/2n)$, by

$$\Pi = \{\epsilon_1-\delta_1,\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,2\delta_n\}\,, \tag{2.25b}$$

with $(\epsilon_1-\delta_1)$ the odd root; (iii) for $\mathrm{osp}(2r+1/2n)$, $r\geqslant 1$, by

$$\Pi = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,\ \delta_n-\epsilon_1,\epsilon_1-\epsilon_2,\ldots,\epsilon_{r-1}-\epsilon_r,\epsilon_r\}\,, \tag{2.25c}$$

with $(\delta_n-\epsilon_1)$ the odd root; and (iv) for $\mathrm{osp}(2r/2n)$, $r\geqslant 2$, by

$$\Pi = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,\ \delta_n-\epsilon_1,\epsilon_1-\epsilon_2,\ldots,\epsilon_{r-1}-\epsilon_r,\epsilon_{r-1}+\epsilon_r\}\,, \tag{2.25d}$$

with $(\delta_n-\epsilon_1)$ to odd root. The distinguished set of simple roots Π minus its odd root plus the even root $2\delta_n$ forms a set of simple roots $\Pi_{\bar{0}}$ for $\mathbf{g}_{\bar{0}}=\mathrm{so}(m)\oplus\mathrm{sp}(2n)$ [except for $\mathrm{osp}(2/2n)$, where $\Pi_{\bar{0}}$ is given by Π minus its odd root]: it is given (i) for $\mathrm{osp}(1/2n)$, by

$$\Pi_{\bar{0}} = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,2\delta_n\}\,; \tag{2.26a}$$

(ii) for $\mathrm{osp}(2/2n)$, by

$$\Pi_{\bar{0}} = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,2\delta_n\}\,; \tag{2.26b}$$

(iii) for $\mathrm{osp}(2r+1/2n)$, $r\geqslant 1$, by

$$\Pi_{\bar{0}} = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,2\delta_n,\ \epsilon_1-\epsilon_2,\ldots,\epsilon_{r-1}-\epsilon_r,\epsilon_r\}\,; \tag{2.26c}$$

and (iv) for $\mathrm{osp}(2r/2n)$, $r\geqslant 2$, by

$$\Pi_{\bar{0}} = \{\delta_1-\delta_2,\delta_2-\delta_3,\ldots,\delta_{n-1}-\delta_n,2\delta_n,\ \epsilon_1-\epsilon_2,\ldots,\epsilon_{r-1}-\epsilon_r,\epsilon_{r-1}+\epsilon_r\}\,. \tag{2.26d}$$

III. GRADED HIGHEST WEIGHT MODULES

A. Graded highest weight modules for osp$(m/2n)$

A Z_2 graded carrier space for a representatioin of $\mathrm{osp}(m/2n)$ is referred to as a $\mathrm{osp}(m/2n)$ module. Let $M(\Lambda^0)$ be a module for a finite-dimensional irreducible representation of $\mathrm{osp}(m/2n)$ with highest weight

$$\Lambda^0 = \sum_{i=1}^{r} \lambda_i^0 \epsilon_i + \sum_{\alpha=1}^{n} \sigma_\alpha^0 \delta_\alpha, \quad r = [m/2], \tag{3.1}$$

and highest weight state $|\Lambda^0\rangle$ such that

$$\begin{aligned}
h_i|\Lambda^0\rangle &= \mathscr{C}_{ii}|\Lambda^0\rangle = \lambda_i^0|\Lambda^0\rangle, && 1\leqslant i\leqslant r \quad \text{(no sum on } i\text{)},\\
\mathscr{C}_{ij}|\Lambda^0\rangle &= 0, && 1\leqslant i<j\leqslant r,\\
\mathscr{D}_i|\Lambda^0\rangle &= 0, && 1\leqslant i\leqslant r \quad (\text{if } m=2r+1),\\
\mathscr{A}_{ij}|\Lambda^0\rangle &= 0, && 1\leqslant i<j\leqslant r;\\
h_\alpha|\Lambda^0\rangle &= C_{\alpha\alpha}|\Lambda^0\rangle = \sigma_\alpha^0|\Lambda^0\rangle, && 1\leqslant\alpha\leqslant n \quad \text{(no sum on } \alpha\text{)},\\
C_{\alpha\beta}|\Lambda^0\rangle &= 0, && 1\leqslant\alpha<\beta\leqslant n,\\
A_{\alpha\beta}|\Lambda^0\rangle &= 0, && 1\leqslant\alpha<\beta\leqslant n,\\
D_{a\alpha}|\Lambda^0\rangle &= 0, && 1\leqslant a\leqslant r, \quad 1\leqslant\alpha\leqslant n.
\end{aligned} \tag{3.2}$$

The Z gradation of $\mathrm{osp}(m/2n)$ naturally imparts a Z gradation on this module consistent with the Z_2 gradatioin.

B. Intrinsic highest Z grade module for $\mathbf{n}_0$

Let $M^{\Lambda^0}(\Lambda^0)$ be the highest Z grade subspace with respect to the Z gradation, i.e., the subspace of weight vectors in $M(\Lambda^0)$ of homogeneous highest Z grade $z^{\max}(\Lambda^0)$,

$$z^{\max}(\Lambda^0) = \sum_{\alpha=1}^{n} \sigma_\alpha^0, \tag{3.3}$$

annihilated by the subalgebra of raising operators $\mathbf{n}_+ = \mathbf{n}_{+1} \oplus \mathbf{n}_{+2}$:

$$M^{(\Lambda^0)}(\Lambda^0) = \left\{ |\eta\rangle \in M(\Lambda^0) \text{ such that } \begin{matrix} \hat{Z}|\eta\rangle = z^{\max}(\Lambda^0)|\eta\rangle \\ X|\eta\rangle = 0, \quad \forall X \in \mathbf{n}_+ = \mathbf{n}_{+1} \oplus \mathbf{n}_{+2} \end{matrix} \right\}. \tag{3.4}$$

This subspace carries an irreducible representation of the stability algebra $\mathbf{n}_0 = \mathrm{so}(m) \oplus \mathrm{u}(n)$ conveniently labeled by its $\mathrm{so}(m)$ and $\mathrm{u}(n)$ highest weights $[\lambda^0]:\{\sigma^0\}$, where $[\lambda^0] = [\lambda_1^0 \lambda_2^0 \cdots \lambda_r^0]$ and $\{\sigma^0\} = \{\sigma_1^0 \sigma_2^0 \cdots \sigma_n^0\}$ refer to $\mathrm{so}(m)$ and $\mathrm{u}(n)$, respectively [see Eq. (3.2)]. This subspace will be referred to as the *intrinsic* $\mathbf{n}_0$ module.[1]

It is assumed herein (although this is not essential for the validity of the present induction formalism) that the intrinsic $\mathbf{n}_0$ module $M^{(\Lambda^0)}(\Lambda^0)$ is finite dimensional and that the labels λ_i^0 and σ_α^0 are real numbers such that

$$\begin{aligned}
&(\sigma_\alpha^0 - \sigma_{\alpha+1}^0) \in Z^+, && 1\leqslant\alpha\leqslant n-1,\\
&\sigma_n^0 \in Z^+,\\
&(\lambda_i^0 - \lambda_{i+1}^0) \in Z^+, && 1\leqslant i\leqslant r-1\\
&2\lambda_r^0 \in Z^+, && m = 2r+1 \geqslant 3,\\
&\lambda_{r-1}^0 + \lambda_r^0 \in Z^+, && m = 2r \geqslant 4,
\end{aligned} \tag{3.5a}$$

where Z^+ is the set of non-negative integers. The subsidiary requirements[9] (see also Sec. VI C)

$$\lambda_{\sigma_n^0+1}^0 = \lambda_{\sigma_n^0+2}^0 = \cdots = \lambda_r^0 = 0 \tag{3.5b}$$

must also be imposed when $m \geqslant 3$ and $\sigma_n^0 < r = [m/2]$.

C. Intrinsic modules for $\mathbf{g}_{\bar{0}}$

Besides the intrinsic highest Z grade $\mathbf{n}_0$ module defined above, there exist other subspaces of importance for the present study. They are defined as the subspaces $M^{(\Lambda)}(\Lambda^0)$,

$$\Lambda = \sum_{i=1}^{r} \lambda_i \epsilon_i + \sum_{\alpha=1}^{n} \sigma_\alpha \delta_\alpha, \quad m = 2r+1 \tag{3.6}$$

[compare to (3.1)], of homogeneous Z grade $z(\Lambda)$,

$$z(\Lambda) = \sum_{\alpha=1}^{n} \sigma_\alpha, \tag{3.7}$$

now annihilated by the subalgebra of $\mathbf{n}_{+2}$ raising operators only,

$$M^{(\Lambda)}(\Lambda^0) = \left\{ |\eta\rangle \in M(\Lambda^0) \text{ such that } \begin{matrix} \hat{Z}|\eta\rangle = z(\Lambda)|\eta\rangle \\ A_{\alpha\beta}|\eta\rangle = 0, \quad \forall A_{\alpha\beta} \in \mathbf{n}_{+2} \end{matrix} \right\}, \tag{3.8}$$

and carrying generally reducible highest weight representations $(\Lambda) = [\lambda]:\{\sigma\}$ of the stability algebra $\mathbf{n}_0 = \mathrm{so}(m) \oplus \mathrm{u}(n)$. A $\mathbf{n}_0$ submodule $M^{(\Lambda)}(\Lambda^0)$ is *isotypic* in the sense that it is a direct sum of a set of equivalent irreducible $\mathbf{n}_0$ submodules. It is an intrinsic module in the same sense as above for a generally reducible $\mathbf{g}_{\bar{0}}$ representation in the decomposition

$$\mathbf{g}\downarrow\mathbf{g}_{\bar{0}} : (\Lambda^0)\downarrow \sum_{\Lambda} m_\Lambda(\Lambda), \quad \mathrm{osp}(m/2n)\downarrow\mathrm{so}(m) \oplus \mathrm{sp}(2n) : [\lambda^0]:\langle\sigma^0\rangle \downarrow \sum_{\lambda,\sigma} m_{\lambda,\sigma}[\lambda]:\langle\sigma\rangle, \tag{3.9}$$

where $m_\Lambda = m_{\lambda,\sigma}$ is the multiplicity of occurrence of a given irrep$(\Lambda) = ([\lambda]:\langle\sigma\rangle)$ of $\mathbf{g}_{\bar{0}} = \mathrm{so}(m) \oplus \mathrm{sp}(2n)$. From a knowledge of the representations of $\mathrm{so}(m) \oplus \mathrm{sp}(2n)$, an irreducible module for a representation (Λ) of $\mathrm{so}(m) \oplus \mathrm{sp}(2n)$ can be induced from an irreducible intrinsic module for $\mathrm{so}(m) \oplus \mathrm{u}(n)$. It follows that the module for an irrep of $\mathrm{osp}(m/2n)$ is uniquely identified by specification of the irreducible components of the various $M^{(\Lambda)}(\Lambda^0)$.

IV. VCS THEORY FOR osp(m/2n)

The VCS theory of $\mathrm{osp}(m/2n)$ is similar to that for $\mathrm{gl}(m/n)$ (Le Blanc and Rowe[6]), the main difference stem-

ming from the fact that the raising nilpotent subalgebra $n_+ = \mathbf{n}_{+1} \oplus \mathbf{n}_{+2}$ is non-Abelian, that it is reducible under $\mathbf{n}_0$, and that it contains part of $\mathbf{g}_{\bar{0}}$, namely, $\mathbf{n}_{+2}$. Consequently, the supercoset G/N_0 must be parametrized by both bosonic (Bargmann) and fermionic (Grassmann) variables (see also Refs. 7 and 10).

A. Embedding of an irreducible module in a vector–Grassmann–Bargmann Hilbert space

An important aspect of VCS theory is the embedding of an irreducible graded highest weight g module $M(\Lambda^0)$ in a vector–Grassmann–Bargmann (VGB) space.[1,4,6] For the classical Lie superalebra $\mathrm{osp}(m/2n)$, the VGB space is the tensor product space $V \otimes \mathrm{Gr} \otimes \mathrm{Bg}$, where we have the following.

(i) V is the intrinsic n_0 module $V = M^{(\Lambda^0)}(\Lambda^0)$ defined by Eq. (3.4). We recall that it carries a unitary irreducible finite-dimensional representation of the stability albegra $\mathbf{n}_0$. Let $\mathscr{B}_V = \{|\eta\rangle\}$ be an orthonormal basis for V with respect to the inner product on V and let $\{\langle\eta|\}$ be a dual basis satisfying $\langle\eta|\eta'\rangle = \delta_{\eta\eta'}$.

(ii) Gr is the space of polynomials in the mn ($= \dim \mathbf{n}_{+1}$) anticommuting [we have assigned a Z_2 grade $\bar{1}$ to the variable $\theta_{a\alpha}$; the commutator in (4.1), interpreted as a graded commutator, thus stands for an anticommutator] Grassmann variables $\{\theta_{a\alpha};\ 1 \leqslant a \leqslant m,\ 1 \leqslant \alpha \leqslant n\}$;

$$\langle \theta_{a\alpha}, \theta_{b\beta} \rangle = 0. \tag{4.1}$$

Equation (4.1) implies

$$(\theta_{a\alpha})^2 = 0; \tag{4.2}$$

the Grassmann space Gr is thus 2^{mn} dimensional. It is isomorphic to the antisymmetric (exterior) tensor algebra over $\mathscr{C}^{mn}$, and has a natural inner product for which the nonzero polynomials

$$\left\{ \prod_{a=1}^{m} \prod_{\alpha=1}^{n} (\theta_{a\alpha})^{n_{a\alpha}} |\eta\rangle;\quad n_{a\alpha} = 0,1;\ |\eta\rangle \in \mathscr{B}_V \right\} \tag{4.3}$$

form an orthonormal basis for the vector–Grassman $\mathrm{VG} = V \otimes \mathrm{Gr}$ space. The space Gr carries an irreducible representation of the Grassman algebra $\mathrm{Gr}(mn)$,

$$\mathrm{Gr}(mn) = \mathrm{span}\left\{ \theta_{a\alpha}, \partial_{a\alpha} \equiv \frac{\partial}{\partial \theta_{a\alpha}}, \mathbf{1};\ 1 \leqslant a \leqslant m,\ 1 \leqslant \alpha \leqslant n \right\}, \tag{4.4}$$

defined by the anticommutation relations

$$\langle \theta_{a\alpha}, \theta_{b\beta} \rangle = 0, \quad \langle \partial_{a\alpha}, \partial_{b\beta} \rangle = 0,$$
$$\langle \partial_{a\alpha}, \theta_{b\beta} \rangle = \delta_{ab}\delta_{\alpha\beta}. \tag{4.5}$$

With respect to the inner product on Gr, we have

$$(\theta_{a\alpha})^\dagger = \partial_{a\alpha}, \quad (\partial_{a\alpha})^\dagger = \theta_{a\alpha}. \tag{4.6}$$

The Grassman variables and their derivatives can be interpreted as fermion annihilation and creation operators; Gr is thus isomorphic to a fermion Fock space. In terms of the inner products on V and Gr, the VG space $V \otimes \mathrm{Gr}$ becomes a Hilbert space $\mathscr{H}_{\mathrm{VG}}$.

(iii) Bg is the space of polynomials in the $\dim n(n+1)/2 = \mathbf{n}_{+2}$ Bargmann (complex) symmetric variables $\{z_{\alpha\beta} = z_{\beta\alpha};\ 1 \leqslant \alpha,\beta \leqslant n\}$. The Bargmann space Bg is infinite dimensional. It is isomorphic to the symmetric tensor algebra over $\mathscr{C}^{n(n+1)/2}$, and has a natural inner product for which the nonzero polynomials

$$\left\{ \prod_{\alpha,\beta,\gamma=1}^{n} \prod_{\alpha=1}^{m} \frac{(z_{\beta\gamma})^{n_{\beta\gamma}}}{\sqrt{(1+\delta_{\beta\gamma})^{M_{\beta\gamma}} n_{\beta\gamma}!}} (\theta_{a\alpha})^{n_{a\alpha}} |\eta\rangle; \right.$$
$$\left. n_{a\alpha} = 0,1, \quad n_{\beta\gamma} = 0,1,2,\ldots;\ |\eta\rangle \in \mathscr{B}_V \right\} \tag{4.7}$$

form an orthonormal basis for the full VGB space. The space Bg carries an irreducible representation of the Heisenberg–Weyl algebra $\mathrm{hw}(n(n+1)/2)$,

$$\mathrm{hw}(n(n+1)/2) = \mathrm{span}\left\{ z_{\alpha\beta}, \nabla_{\alpha\beta} \equiv \frac{\partial}{\partial z_{\alpha\beta}}, \mathbf{1};\ 1 \leqslant \alpha,\beta \leqslant n \right\}, \tag{4.8}$$

defined by the commutation relations

$$\langle z_{\alpha\beta}, z_{\mu\nu} \rangle = 0, \quad \langle \nabla_{\alpha\beta}, \nabla_{\mu\nu} \rangle = 0,$$
$$\langle \nabla_{\alpha\beta}, z_{\mu\nu} \rangle = \delta_{\alpha\mu}\delta_{\beta\nu} + \delta_{\alpha\nu}\delta_{\beta\mu}. \tag{4.9}$$

With respect to the inner product on Bg, we have

$$(z_{i\alpha})^\dagger = \nabla_{i\alpha}, \quad (\nabla_{i\alpha})^\dagger = z_{i\alpha}. \tag{4.10}$$

The Bargmann variables and their derivatives can be interpreted as symmetric boson annihilation and creation operators.[12] In terms of the inner products on V, Gr, and Bg, the VGB space $V \otimes \mathrm{Bg}$ becomes a Hilbert space $\mathscr{H}_{\mathrm{VGB}}$.

The dimension of the VG basis (4.3) is 2^{mn} times the dimension of the intrinsic $\mathbf{n}_0$ module V. Levels can be defined on this basis in terms of the eigenvalue n_θ of the θ-number operator $\hat{N}^{(\theta)} = \theta_{a\alpha}\partial_{a\alpha}$. There are $mn+1$ such levels. The number of VG states on a given level is $\binom{mn}{n_\theta}$ times the dimension of the intrinsic module. The Z grade of a VG state is defined by

$$z = z^{\max} - n_\theta, \tag{4.11}$$

consistent with the definition (2.11) for the grading operator, Eq. (3.4) and Eq. (4.20d) below. Now, if the intrinsic space is assigned a Z_2 grade $\bar{0}$, consistency requires us to identify the even (odd) subspace of $\mathscr{H}_{\mathrm{VB}}$ with the set of all VG states with n_θ even (odd). The Z_2 grade ζ of a state is then given by

$$\zeta = n_\theta \bmod 2. \tag{4.12a}$$

Conversely, if the intrinsic space is assigned a Z_2 grade $\bar{1}$, we identify the even (odd) subspace of $\mathscr{H}_{\mathrm{VB}}$ with the set of all VG states with n_θ odd (even). The Z_2 grade ζ of a state is then given by

$$\zeta = (n_\theta + 1) \bmod 2. \tag{4.12b}$$

The dimension of the even subspace of $\mathscr{H}_{\mathrm{VG}}$ is equal to the dimension of its odd subspace since

$$\sum_{n_\theta \text{ even}} \binom{mn}{n_\theta} = \sum_{n_\theta \text{ odd}} \binom{mn}{n_\theta}.$$

Levels can be also defined on the VGB basis (4.7) in terms of the eigenvalues n_θ and the *even* eigenvalues n_z of the operator $\hat{N}^{(z)} = z_{\alpha\beta}\nabla_{\alpha\beta}$. We define the Z grade of a VGB state by

$$z = z^{\max} - n_\theta - n_z. \tag{4.13}$$

This definition is again consistent with the definition (2.11)

for the grading operator, Eq. (3.4) and Eq. (4.20d) below. Obviously, the Z_2 grade of a VGB state is the Z_2 grade of its underlying VG state:

$$\zeta = (z^{\max} - n_\theta - n_z) \bmod 2 = (z^{\max} - n_\theta) \bmod 2. \tag{4.14}$$

The VCS embedding (this embedding could be considered as a superalgebraic generalization of the development of superfunctions[7,10]) $M(\Lambda^0) \to \mathscr{H}_{\mathrm{VGB}}$ is defined by

$$|\psi\rangle \to \psi(\theta,z) = \sum_\eta |\eta\rangle\langle\eta|\exp \mathscr{T}(\theta,z)|\psi\rangle, \quad |\eta\rangle \in \mathscr{B}_V, \tag{4.15a}$$

where

$$\mathscr{T}(\theta,z) = \theta\cdot D + \frac{1}{2} z\cdot A = \sum_{a=1}^{m}\sum_{\alpha=1}^{n} \theta_{a\alpha} D_{a\alpha} + \frac{1}{2}\sum_{\mu,\nu=1}^{n} z_{\mu\nu} A_{\mu\nu}. \tag{4.15b}$$

This embedding invokes a projection $M(\Lambda^0) \to M^{(\Lambda^0)}(\Lambda^0)$ in which

$$|\bar{\psi}\rangle = \exp \mathscr{T}(\theta,z)|\psi\rangle$$

projects to its highest Z grade component

$$|\bar{\psi}\rangle \to \psi(\theta,z) = \sum_\eta |\eta\rangle\langle\eta|\bar{\psi}\rangle$$

and which, since $M(\Lambda^0)$ is a direct sum of graded subspaces, is well defined without the necessity of assuming that $M(\Lambda^0)$ is a Hilbert space.

B. The VCS expansion for osp(*m*/2*n*)

The VCS realization $\Gamma(X)$ of an arbitrary generator $X \in \mathrm{osp}(m/2n)$ is defined by

$$\begin{aligned}\Gamma(X)\psi(\theta,z) &= \sum_\eta |\eta\rangle\langle\eta|\exp(\mathscr{T})X|\psi\rangle \\ &= \sum_\eta |\eta\rangle\langle\eta|\Big(X + \frac{1}{1!}\langle\mathscr{T},X\rangle \\ &\quad + \frac{1}{2!}\langle\mathscr{T},\langle\mathscr{T},X\rangle\rangle + \cdots\Big)\times\exp(\mathscr{T})|\psi\rangle.\end{aligned} \tag{4.16}$$

Because the variables $\theta_{a\alpha}$ and $z_{\mu\nu}$ belong to algebras that are independent of the superalgebra osp$(m/2n)$, we have that all of the following (graded) commutators vanish:

$$\langle\theta_{a\alpha},X\rangle = \langle z_{\mu\nu},X\rangle = 0, \quad \forall X \in \mathrm{osp}(m/2n). \tag{4.17}$$

For example, the set $\{\theta_{a\alpha}\}$ of Grassmann variables *anticommutes* with the set of *odd* raising and lowering operators $\{D_{a\alpha}\}$ and $\{E_{a\alpha}\}$, and (4.16) should be developed accordingly.

The operator $\Gamma(X)$ can be expressed as a differential operator on the superfield $\psi(\theta,z)$. First note that, since $|\eta\rangle$ belongs to the highest weight $\mathfrak{g}_{\bar{0}}$ submodule $M^{(\Lambda^0)}(\Lambda^0)$,

$$\langle\eta|B_{\alpha\beta}\exp(\mathscr{T})|\psi\rangle = 0, \quad \forall B_{\alpha\beta} \in \mathfrak{n}_{-2}; \tag{4.18}$$

this is easily verified by considering the Z graded structure of the highest weight module $M(\Lambda^0)$ with the understanding that states of different Z grade are orthogonal. We also define

$$\begin{aligned}&\sum_\eta H_{ij}^{(\Lambda^0)}|\eta\rangle\langle\eta|\exp(\mathscr{T})|\psi\rangle \\ &\quad = \sum_\eta |\eta\rangle\langle\eta|H_{ij}\exp(\mathscr{T})|\psi\rangle, \\ &\sum_\eta C_{\alpha\beta}^{(\Lambda^0)}|\eta\rangle\langle\eta|\exp(\mathscr{T})|\psi\rangle \\ &\quad = \sum_\eta |\eta\rangle\langle\eta|C_{\alpha\beta}\exp(\mathscr{T})|\psi\rangle,\end{aligned} \tag{4.19}$$

where $\mathrm{span}\{H_{ij}^{(\Lambda^0)}\} \oplus \mathrm{span}\{C_{\alpha\beta}^{(\Lambda^0)}\}$ is the *intrinsic* $\mathrm{so}(m) \oplus \mathrm{u}(n)$ representation carried by the irreducible highest weight submodule $M^{(\Lambda^0)}(\Lambda^0)$. We find, with the usual convention (we recall that we use the convention that $1 \leqslant a,b,\ldots \leqslant m$, $1 \leqslant \alpha,\beta,\ldots \leqslant n$) concerning the summation of repeated indices, the following VCS expansion for the generators of osp$(m/2n)$:

$$\Gamma(A_{\alpha\beta}) = \nabla_{\alpha\beta}, \tag{4.20a}$$

$$\Gamma(D_{a\alpha}) = \partial_{a\alpha} - \tfrac{1}{2}\theta_{a\mu}\nabla_{\mu\alpha}, \tag{4.20b}$$

$$\Gamma(H_{ab}) = H_{ab}^{(\Lambda^0)} + (\theta_{a\mu}\,\partial_{b\mu} - \theta_{b\mu}\,\partial_{a\mu}), \tag{4.20c}$$

$$\Gamma(C_{\alpha\beta}) = C_{\alpha\beta}^{(\Lambda^0)} - \theta_{c\beta}\,\partial_{c\alpha} - z_{\beta\mu}\nabla_{\mu\alpha}, \tag{4.20d}$$

$$\begin{aligned}\Gamma(E_{a\alpha}) &= \theta_{a\mu}(C_{\mu\alpha}^{(\Lambda^0)} - \tfrac{1}{2} z_{a\sigma}\nabla_{\sigma\mu} - \tfrac{1}{2}\theta_{c\alpha}\,\partial_{c\mu}) \\ &\quad + \theta_{c\alpha}(H_{ac}^{(\Lambda^0)} + \tfrac{1}{2}(\theta_{a\sigma}\,\partial_{c\sigma} - \theta_{c\sigma}\,\partial_{a\sigma})) \\ &\quad + z_{a\sigma}\,\partial_{a\sigma} - \tfrac{1}{4}\theta_{a\xi}\theta_{c\alpha}\theta_{c\sigma}\nabla_{\sigma\xi},\end{aligned} \tag{4.20e}$$

$$\begin{aligned}\Gamma(B_{\alpha\beta}) &= z_{a\sigma}(C_{\sigma\beta}^{(\Lambda^0)} - \theta_{c\beta}\,\partial_{c\sigma} - \tfrac{1}{2} z_{\beta\xi}\nabla_{\xi\sigma}) \\ &\quad - \tfrac{1}{2}\theta_{c\alpha}\theta_{c\sigma}C_{\sigma\beta}^{(\Lambda^0)} + \tfrac{1}{2}\theta_{c\alpha}\theta_{c\sigma}\theta_{d\beta}\,\partial_{d\sigma} \\ &\quad - \tfrac{1}{2}\theta_{d\alpha}\theta_{e\beta}H_{de}^{(\Lambda^0)} + \tfrac{1}{8}\theta_{c\alpha}\theta_{c\mu}\theta_{d\beta}\theta_{d\xi}\nabla_{\xi\mu} + (\alpha\leftrightarrow\beta),\end{aligned} \tag{4.20f}$$

where we have used the identities

$$\nabla_{\alpha\beta}\exp(\mathscr{T}) = A_{\alpha\beta}\exp(\mathscr{T}),$$

$$\partial_{a\alpha}\exp(\mathscr{T}) = \exp(\mathscr{T})(D_{a\alpha} + \tfrac{1}{2}\theta_{a\sigma}A_{\sigma\alpha}).$$

Note that in the VCS realization, the so(m) subalgebra consists of the piecewise sum by component of an intrinsic subalgebra $(H_{ab}^{(\Lambda^0)})$ and a Grassmann realization

$$H_{ab}^{(\theta)} = (\theta_{a\mu}\,\partial_{b\mu} - \theta_{b\mu}\,\partial_{a\mu}) \tag{4.21a}$$

of so(m). Similarly, the u(n) subalgebra consists of the piecewise sum of an intrinsic subalgebra $(C_{\alpha\beta}^{(\Lambda^0)})$, a Grassmann realization

$$C_{\alpha\beta}^{(0)} = (-\theta_{c\beta}\,\partial_{c\alpha}), \tag{4.21b}$$

and a Bargmann realization

$$C_{\alpha\beta}^{(z)} = (-z_{\beta\mu}\nabla_{\mu\alpha}) \tag{4.21c}$$

of u(n). Note also that the VCS expansion for the stability subalgebra $\mathfrak{n}_0 = \mathrm{so}(m) \oplus \mathrm{u}(n)$ is Hermitian with respect to the VGB measure whenever the intrinsic representation is Hermitian (recall that a Hermitian representation of a Lie algebra corresponds to a unitary representation of the Lie group). This is an important characteristic of VCS theory that allows one to construct (see Sec. V) orthonormal VGB bases with good $\mathfrak{n}_0$ transformation properties upon which VCS states can be developed. In contrast, one sees that the VCS expansion for the maximal (here even) subalgebra $\mathfrak{n}_0 \oplus \mathfrak{n}_{+2} \oplus \mathfrak{n}^{-2}$ of osp$(m/2n)$ is not Hermitian with respect

to the VGB measure. This is in spite of the fact that, for positive integral highest weight, it is known to be equivalent to a Hermitian representation. The concepts of adjoint operations and Hermiticity for irreducible representations of osp$(m/2n)$ will be discussed in Sec. VII. For the moment, we simply remark that all the computations carried out below exploit the inherent Hermiticity properties of the VGB bases under $\mathfrak{n}_0$ transformations.

V. VCS GRADED HIGHEST WEIGHT MODULES

The VCS construction naturally exploits the observed fact that the restriction of the VCS representation (4.20) of osp$(m/2n)$ to its stability subalgebra $\mathfrak{n}_0 = \mathrm{so}(m) \oplus \mathrm{u}(n)$ is Hermitian under the VGB inner product. This facilitates the construction of an orthonormal VGB basis that [unlike the basis (4.7)] reduces the stability subalgebra $\mathrm{so}(m) \oplus \mathrm{u}(n)$. The construction is given in Secs. V B and V C below. But first, we must distinguish a basic VCS representation, which is irreducible, from an extended VCS representation, which, generally, is not.

A. Irreducible and extended VCS modules

Starting from an irreducible highest weight representation of a subalgebra, VCS theory induces (finite- or infinite-dimensional) irreducible highest weight representations of the Lie algebra or classical Lie superalgebra under consideration. More specifically, starting from the intrinsic module $M^{(\Lambda^0)}(\Lambda^0)$, the Γ realization (4.20) generates, through the usual laddering down process, the irreducible invariant subspace $M(\Lambda^0)$ of the VGB space. According to its definition, Eq. (4.16), the domain of a VCS operator $\Gamma(X)$ is restricted to this subspace. However, the VCS operators of Eq. (4.20) have a natural extension to the whole VGB space. We call the representation in which the domains of the VCS Γ operators are extended to the whole VGB space the *extended* representation and denote it by $\overline{\Gamma}$.

B. Basis for the VG space

Since

$$\langle \overline{\Gamma}(H_{ab}), \theta_{c\sigma} \rangle = \delta_{bc}\theta_{a\mu} - \delta_{ac}\theta_{b\mu}, \tag{5.1a}$$

$$\langle \overline{\Gamma}(C_{\alpha\beta}), \theta_{a\mu} \rangle = -\delta_{\alpha\mu}\theta_{a\beta}, \tag{5.1b}$$

it follows that the Grassmann variable $\theta_{a\mu}$ transforms as the a component of a so(m) [1] tensor and the μ component of a u(n) $\{-1\}$ tensor. The set $\{\theta_{a\alpha}\}$ thus transforms contragradiently to the set of raising operators $\{D_{a\alpha}\}$, which has a VCS realization involving the set of partial derivatives $\{\partial_{a\mu}\}$ comprising a $[1]:\{-1\}$ so$(m) \oplus$ u(n) tensor.

More generally, a basis of fully *antisymmetric* polynomials of higher rank in the Grassmann variables, orthogonal with respect to the Grassmann inner product, and having good $\mathfrak{n}_0$ transformation properties, can be constructed by considering tensor products of the fundamental Grassmann tensor $\theta^{\{1\}}_{\{-1\}}$. We shall denote these polynomials by

$$\Theta^{[\kappa](m_\kappa)}_{\{-\tau\}(m_{-\tau})}(\theta), \tag{5.2}$$

where (m_κ) and $(m_{-\tau})$ stand for basis labels for the so(m) and u(n) irreps $[\kappa]$ and $\{-\tau\}$, respectively.

The $\mathfrak{n}_0$ ranks of the polynomials that one can construct in this way are given by noting that the polynomials of degree N in the Grassmann variables span a fully antisymmetric irrep $\{1^N\}$ of a u(mn) algebra,

$$\mathrm{u}(mn) = \mathrm{span}\{\theta_{a\alpha}\,\partial_{b\beta},\ 1 \leqslant a,b \leqslant m,\ 1 \leqslant \alpha,\beta \leqslant n\}. \tag{5.3}$$

This fully antisymmetric irrep decomposes, under the restriction u$(mn)\downarrow$u$(m) \oplus$ u(n), into the sum of irreps

$$\mathrm{u}(mn)\downarrow\mathrm{u}(m) \oplus \mathrm{u}(n) : \{1^N\}\downarrow \sum_{\tau} \{\tilde{\tau}\}:\{-\tau\}$$

where the u(m) algebra

$$\mathrm{u}(m) = \mathrm{span}\left\{\mathscr{C}_{ab} = \sum_{\nu=1}^{n} \theta_{a\nu}\,\partial_{b\nu},\ 1 \leqslant a,b \leqslant m\right\} \tag{5.4a}$$

is said to be complementary to the u(n) algebra

$$\mathrm{u}(n) = \mathrm{span}\left\{C_{\mu\nu} = -\sum_{a=1}^{m} \theta_{a\nu}\,\partial_{a\mu},\ 1 \leqslant \mu,\nu \leqslant n\right\} \tag{5.4b}$$

in u(mn), where the partition $\{\tau\}$ has n rows of length τ_α in the range $0 \leqslant \tau_\alpha \leqslant m$, where $\{-\tau\} = \{-\tau_n, -\tau_{n-1}, \ldots, -\tau_1\}$, and where, as a result of the antisymmetry of the polynomial, $\{\tilde{\tau}\}$ is the partition conjugate to $\{\tau\}$ having n columns of respective height τ_α. Necessarily

$$\sum_{a=1}^{m} \tilde{\tau}_a = \sum_{\alpha=1}^{n} \tau_\alpha = N. \tag{5.5}$$

Since the so(m) irreps contained in a u(m) irrep $\{\tilde{\tau}\}$ are given by the decomposition[13]

$$\{\tilde{\tau}\}\downarrow \sum_{\xi \in D} [\tilde{\tau}/\xi], \tag{5.6}$$

where D is the set of partitions having even parts ($\{\xi\} = \{\xi_1\xi_2\ldots\}, \xi_a$ even), we determine that the polynomials of degree N in the Grassmann variables span the so$(m) \oplus$ u(n) irreps given by the decomposition

$$\mathrm{u}(mn)\downarrow\mathrm{so}(m) \oplus \mathrm{u}(n) : \{1^N\}\downarrow \sum_{\tau} \sum_{\xi \in D} [\tilde{\tau}/\xi]:\{-\tau\}. \tag{5.7}$$

The construction of orthonormal bases for the (generally non-multiplicity-free) decomposition

$$\mathrm{u}(m)\downarrow\mathrm{so}(m) : \{\tilde{\tau}\}\downarrow \sum_{\xi \in D} [\tilde{\tau}/\xi] \equiv [\kappa],$$

$$[\kappa] = [\kappa_1\kappa_2\cdots\kappa_r], \quad r = [m/2], \tag{5.8}$$

using VCS techniques is given in Ref. 14.

Now, an orthonormal basis for the VG space, which reduces the stability subalgebra $\mathfrak{n}_0$, is defined by the coupled products

$$\left|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa][\lambda](m_\lambda)}_{\{-\tau\}\{\sigma\}(m_\sigma)}\right\rangle = \left[\Theta^{[\kappa]}_{\{-\tau\}}(\theta) \times \left|{}^{[\lambda^0]}_{\{\sigma^0\}}\right\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}, \tag{5.9}$$

where (i) the polynomials $\Theta(\theta)$ have been defined above; (ii) the kets

$$|\eta\rangle = \left|{}^{[\lambda^0](m_{\lambda^0})}_{\{\sigma^0\}(m_{\sigma^0})}\right\rangle \tag{5.10}$$

span the intrinsic highest Z grade $\mathfrak{n}_0$ module defined in Sec. III A (see also Sec. IV A); and (iii) the square brackets $[\cdots\times\cdots]$ refer to a so$(m) \oplus$ u(n) (upper and lower) cou-

pling of the Θ polynomials with the intrinsic basis. (All couplings in this manuscript should be understood to be going from right to left. Also all multiplicity indices resolving the possible multiplicites arising from the various Kronecker products are implicitly understood but explicitly ignored for the sake of notational simplicity.)

Now, observe that all states of the VG subspace of the VGB space have no dependency on the Bargmann variables and therefore satisfy the equation

$$\overline{\Gamma}(A_{\alpha\beta})\left|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa][\lambda](m_\lambda)}_{\{-\tau\}\{\sigma\}(m_\sigma)}\right\rangle = 0, \quad 1\leqslant\alpha,\beta\leqslant n. \tag{5.11}$$

In accord with Eq. (3.8), we conclude the irreducible $\mathbf{n}_0$ subspaces of the VG space defined by Eq. (5.9) form a basis for the $\mathbf{n}_0$ highest weight (intrinsic) subspaces $M^{(\Lambda)}(\Lambda^0)$. This fact will be highly relevant to the computation of $\mathbf{g}_{\bar{0}}$-reduced matrix elements in Sec. VI.

C. Basis for the VGB space

The construction of an orthonormal basis for the VG space is easily extended to the full VGB space by observing that, since

$$\langle\overline{\Gamma}(H_{ab}),z_{\mu\nu}\rangle = 0, \tag{5.12a}$$

$$\langle\overline{\Gamma}(C_{\alpha\beta}),z_{\mu\nu}\rangle = -\delta_{\alpha\mu}z_{\beta\nu} - \delta_{\alpha\nu}z_{\mu\beta}, \tag{5.12b}$$

the Bargmann variable $z_{\mu\nu}$ transforms as a [0] tensor under $\mathrm{so}(m)$ and as the $(\mu\nu)$ component of a $\mathrm{u}(n)$ $\{-2\}$ tensor. The set $\{z_{\mu\nu}\}$ thus transforms contragradiently to the set of raising operators $\{A_{\mu\nu}\}$, which has a VCS realization given by the set of partial derivatives $\{\nabla_{\mu\nu}\}$; thus it comprises a $[0]:\{2\}$ $\mathrm{so}(m)\oplus\mathrm{u}(n)$ tensor.

A basis of fully *symmetric* polynomials of higher rank in the Bargmann variables, orthogonal with respect to the Bargmann measure, and having good $\mathbf{n}_0$ transformation properties can be constructed[12] by considering tensor products of the fundamental Bargmann tensor $z^{[0]}_{\{-2\}}$. We shall denote these polynomials by

$$Z^{[0]}_{\{-\xi\}(m_{-\xi})}(z), \tag{5.13}$$

where $(m_{-\xi})$ stands for basis labels for the $\mathrm{u}(n)$ irreps $\{-\xi\}$ with $\xi\in D$, the set of even partitions.

An orthonormal basis for the whole VGB space, which reduces the stability subalgebra $\mathbf{n}_0$, is defined by the $\mathrm{u}(n)$ coupling of the basis of Bargmann polynomials (5.13) with the orthonormal VG basis (5.9):

$$\begin{aligned}
&\left|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa][\lambda]}_{\{-\tau\}\{\sigma\}};{}^{[0][\lambda](m_\lambda)}_{\{-\xi\}\{\omega\}(m_\omega)}\right\rangle \\
&\quad = \left[Z^{[0]}_{\{-\xi\}}(z)\times\left|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa][\lambda]}_{\{-\tau\}\{\sigma\}}\right\rangle\right]^{[\lambda](m_\lambda)}_{\{\omega\}(m_\omega)} \\
&\quad = \left[Z^{[0]}_{\{-\xi\}}(z)\times\left[\Theta^{[\kappa]}_{\{-\tau\}}(\theta)\times\left|{}^{[\lambda^0]}_{\{\sigma^0\}}\right\rangle\right]^{[\lambda]}_{\{\sigma\}}\right]^{[\lambda](m_\lambda)}_{\{\omega\}(m_\omega)}.
\end{aligned} \tag{5.14}$$

VI. Γ-MATRIX REPRESENTATIONS FOR osp(*m*/2*n*)

A. Identification of the irreducible submodule of the VGB space

Considerable economy in the expression of the matrix elements of an algebra's matrix representation can be gained by exploiting the Wigner–Eckart theorem. Furthermore, if one assumes a complete knowledge of the subrepresentations of the Lie subalgebra $\mathbf{g}_{\bar{0}}$ contained in a given representation of a Lie superalgebra $\mathbf{g}$, it is then sufficient to determine the $\mathbf{g}_{\bar{0}}$-reduced matrix elements of the odd tensors of $\mathbf{g}$ to obtain a full knowledge of the superstructure of the representations under consideration. We show in this section how such a program is implemented within the VCS framework.

The irreducible $\mathrm{osp}(m/2n)$ module $M(\Lambda^0)$ can be constructed by laddering down from its highest Z grade $\mathbf{n}_0$-invariant intrinsic module $M^{(\Lambda^0)}(\Lambda^0)$. The VCS implementation of this laddering-down process is achieved by performing a $\mathbf{n}_0$ coupling of polynomials $\Theta(\overline{\Gamma}(E))$ in the VCS lowering operators $\overline{\Gamma}(E)$ with the states of $M^{(\Lambda^0)}(\Lambda^0)$

$$\left[\Theta^{[\kappa]}_{\{-\tau\}}(\overline{\Gamma}(E))\times\left|{}^{[\lambda^0]}_{\{\sigma^0\}}\right\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}$$

[compare to (5.9) where we have substituted $\theta\to\overline{\Gamma}(E)$]. In this way, $M(\Lambda^0)$ is generated as a direct sum of $\mathbf{n}_0$-invariant submodules. However, since we assume a complete knowledge of the irreps of $\mathbf{g}_{\bar{0}}$, it is more economical to represent $M(\Lambda^0)$ as a direct sum of $\mathbf{g}_{\bar{0}}$-invariant submodules each of which contains, and is characterized by, an intrinsic ($\mathbf{n}_0$-invariant) highest Z grade subspace $M^{(\Lambda)}(\Lambda^0)$. We therefore seek to identify these intrinsic $\mathbf{n}_0$-invariant subspaces to characterize completely $M(\Lambda^0)$.

The identification of the intrinsic $\mathbf{n}_0$-invariant subspaces $M^{(\Lambda)}(\Lambda^0)$ is grately facilitated by the use of projector operators. The introduction of such operators is necessitated by the fact that the subalgebra $\mathbf{n}_{-1}$ is not super-Abelian; anticommutators of generators belonging to $\mathbf{n}_{-1}$ lie in $\mathbf{n}_{-2}$, the set of lowering operators for $\mathrm{sp}(2n)$. Let P denote the projection operator that projects each $\mathbf{g}_{\bar{0}}$ module in the VGB space onto its intrinsic $\mathbf{n}_0$ subspace $M^{(\Lambda)}(\Lambda^0)$. If $|\phi\rangle$ belongs to a $\mathbf{g}_{\bar{0}}$ module of highest weight $(\Lambda)=([\lambda]:\{\sigma\})$, then $P|\phi\rangle\in M^{(\Lambda)}(\Lambda^0)$. Consequently, the intrinsic $\mathbf{n}_0$ submodules $M^{(\Lambda)}(\Lambda^0)$ contained in $M(\Lambda^0)$ are spanned by the states

$$P\left[\Theta^{[\kappa]}_{\{-\tau\}}(\overline{\Gamma}(E))\times\left|{}^{[\lambda^0]}_{\{\sigma^0\}}\right\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}.$$

These states can be expanded on the VG basis (5.9);

$$\begin{aligned}
&P\left[\Theta^{[\kappa]}_{\{-\tau\}}(\overline{\Gamma}(E))\times\left|{}^{[\lambda^0]}_{\{\sigma^0\}}\right\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)} \\
&\quad = \mathscr{O}\left|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa]}_{\{-\tau\}}{}^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\right\rangle \\
&\quad = \sum_{\kappa'\tau'}\left|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\right\rangle\mathscr{O}\left({}^{[\lambda^0]}_{\{\sigma^0\}}{}^{[\lambda]}_{\{\sigma\}}\right)_{[\kappa'][\kappa]}{}_{\{-\tau'\}\{-\tau\}}.
\end{aligned} \tag{6.1}$$

The matrix $\mathscr{O}$ of expansion coefficients is computed using the recursion formula (double bars indicates $\mathbf{n}_0$-reduced matrix elements throughout)

$$\begin{aligned}
&\left\langle{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\middle\|\mathscr{O}\theta\middle\|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa][\lambda]}_{\{-\tau\}\{\sigma\}}\right\rangle \\
&\quad = \left\langle{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\middle\|P\overline{\Gamma}(E)\mathscr{O}\middle\|{}^{[\lambda^0]}_{\{\sigma^0\}};{}^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\right\rangle.
\end{aligned} \tag{6.2}$$

easily obtained from (6.1) using straightforward recoupling techniques.

The projector P can be written

$$P = \mathbf{1} - \sum_y \frac{\overline{\Gamma}(B_{\alpha\beta})\overline{\Gamma}(A_{\alpha\beta})}{\langle y|\overline{\Gamma}(B_{\alpha\beta})\overline{\Gamma}(A_{\alpha\beta})|y\rangle}|y\rangle\langle y| - \cdots \tag{6.3}$$

[the extra terms, of higher order in $\overline{\Gamma}(A)$ and $\overline{\Gamma}(B)$, are not needed here]. When this expression for P is introduced in the recursion formula (6.2a), the latter becomes

$$\langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa'] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \mathcal{O}\theta \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix} \rangle$$
$$= \langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa'] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \overline{\Gamma}^{(0)}(E) \mathcal{O} \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix} \rangle$$
$$- \sum_{\sigma'\tau'\kappa''} \frac{\langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa'] \\ \{-\tau'\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \overline{\Gamma}^{(0)}(B) \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau''\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \rangle}{\langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau''\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-2\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \overline{\Gamma}^{(1)}(B) \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau''\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \rangle} \langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau''\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-2\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \overline{\Gamma}^{(1)}(E) \mathcal{O} \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix} \rangle , \tag{6.2'}$$

where

$$\overline{\Gamma}^{(0)}(E_{a\alpha}) = \theta_{a\mu}(C^{(\Lambda^0)}_{\mu\alpha} - \tfrac{1}{2}\theta_{c\alpha}\,\partial_{c\mu}) + \theta_{c\alpha}(H^{(\Lambda^0)}_{ac} + \tfrac{1}{2}(\theta_{a\sigma}\,\partial_{c\sigma} - \theta_{c\sigma}\,\partial_{a\sigma}), \tag{6.4a}$$

$$\overline{\Gamma}^{(1)}(E_{a\alpha}) = z_{a\sigma}\,\partial_{a\sigma}, \tag{6.4b}$$

$$\overline{\Gamma}^{(0)}(B_{\alpha\beta}) = -\tfrac{1}{2}\theta_{c\alpha}\theta_{c\sigma}C^{(\Lambda^0)}_{\sigma\beta} + \tfrac{1}{2}\theta_{c\alpha}\theta_{c\sigma}\theta_{d\beta}\,\partial_{d\sigma} - \tfrac{1}{2}\theta_{d\alpha}\theta_{e\beta}H^{(\Lambda^0)}_{de} + (\alpha\leftrightarrow\beta), \tag{6.4c}$$

$$\overline{\Gamma}^{(1)}(B_{\alpha\beta}) = z_{a\sigma}(C^{(\Lambda^0)}_{\sigma\beta} - \theta_{c\beta}\,\partial_{c\sigma} - \tfrac{1}{2}z_{\beta\xi}\nabla_{\xi\sigma}) + (\alpha\leftrightarrow\beta). \tag{6.4d}$$

The various partial expansion $\overline{\Gamma}^{(0)}$ and $\overline{\Gamma}^{(1)}$ in (6.4) are extracted from the expansion (4.20) and stand for their $\hat{N}^{(z)}$-invariant [z-number operator, Eq. (6.5i) below] and their $\hat{N}^{(z)}$-increasing components, respectively. It is appropriate to set

$$\mathcal{O}(\begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}) = 1$$

as a starting point for the solution of the recursion formula (6.2) since the states of $M^{(\Lambda^0)}(\Lambda^0)$ are unchanged by the projection operator.

Matrix elements involving $\overline{\Gamma}^{(0)}(E)$ are easily computed by means of the matrix element equality

$$\langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa'] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \overline{\Gamma}^{(0)}(E) \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix} \rangle = \langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa'] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \langle \hat{\Omega}, \theta \rangle \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix} \rangle \tag{6.5a}$$

where

$$\hat{\Omega} = \tfrac{1}{2}\hat{I}_{so(m)} - \tfrac{1}{2}\hat{I}_{u(n)} - \tfrac{1}{4}\hat{I}^{(\theta)}_{so(m)} + \tfrac{1}{4}\hat{I}^{(\theta)}_{u(n)} + \tfrac{1}{4}\hat{I}^{(z)}_{u(n)} - [(m-n-1)/4]\hat{N}^{(\theta)} + [(n+1)/4]\hat{N}^{(z)} \tag{6.5b}$$

is a $\mathfrak{n}_0$-invariant operator with (i) $\hat{I}_{so(m)}$ the $so(m)$ quadratic Casimir operator

$$\hat{I}_{so(m)} = \tfrac{1}{2}\overline{\Gamma}(H_{ab})\overline{\Gamma}(H_{ab}); \tag{6.5c}$$

(ii) $\hat{I}^{(\theta)}_{so(m)}$ the quadratic Casimir operator for the Grassmann realization of so (m) [see Eq. (4.21a)],

$$\hat{I}^{(\theta)}_{so(m)} = \tfrac{1}{2}H^{(\theta)}_{ab}H^{(\theta)}_{ab}, \tag{6.5d}$$

(iii) $\hat{I}_{u(n)}$ the u(n) quadratic Casimir operator

$$\hat{I}_{u(n)} = \overline{\Gamma}(C_{a\beta})\overline{\Gamma}(C_{\beta a}); \tag{6.5e}$$

(iv) $\hat{I}^{(\theta)}_{u(n)}$ the quadratic Casimir operator for the Grassmann realization of u(n) [see Eq. (4.21b)],

$$\hat{I}^{(\theta)}_{u(n)} = C^{(\theta)}_{\alpha\beta}C^{(\theta)}_{\beta\alpha}; \tag{6.5f}$$

(v) $\hat{I}^{(z)}_{u(n)}$ the quadratic Casimir operator for the Bargmann realization of u(n) [see Eq. (4.21c)],

$$\hat{I}_{u(n)} = C^{(z)}_{\alpha\beta}C^{(z)}_{\beta\alpha}; \tag{6.5g}$$

(vi) $\hat{N}^{(\theta)}$ the Grassmann number operator

$$\hat{N}^{(\theta)} = \theta_{a\alpha}\,\partial_{a\alpha}; \tag{6.5h}$$

(vii) $\hat{N}^{(z)}$ the Bargmann number operator

$$\hat{N}^{(z)} = z_{a\beta}\nabla_{\beta a}. \tag{6.5i}$$

Similarly, matrix elements involving $\overline{\Gamma}^{(1)}(B)$ are easily computed by means of the matrix element equality

$$\langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau'\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-2\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \overline{\Gamma}^{(1)}(B) \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau'\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \rangle = \langle \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau'\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-2\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \| \langle \hat{\Omega}, z \rangle \| \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa''] \\ \{-\tau'\} \end{Bmatrix} \begin{Bmatrix} [\lambda'] \\ \{\sigma'\} \end{Bmatrix} \rangle. \tag{6.5j}$$

Finally, matrix elements involving $\overline{\Gamma}^{(1)}(E)$ are easily computed by noting that

$$\overline{\Gamma}^{(1)}(E) = \sqrt{\frac{(n+1)}{2}}\,[z^{[0]}_{\{-2\}} \times \partial^{[1]}_{\{1\}}]^{[1]}_{\{-1\}}. \tag{6.5k}$$

Eigenvalues of the $\mathfrak{n}_0$-invariant operator $\hat{\Omega}$ [Eq. (6.5b)] on the VBG basis (5.14) will be noted

$$\hat{\Omega} | \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-\xi\} \end{Bmatrix} \begin{matrix} [\lambda](m_\lambda) \\ \{\omega\}(m_\omega) \end{matrix} \rangle = \Omega \left(\begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-\xi\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\omega\} \end{Bmatrix} \right) | \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-\xi\} \end{Bmatrix} \begin{matrix} [\lambda](m_\lambda) \\ \{\omega\}(m_\omega) \end{matrix} \rangle, \tag{6.6a}$$

where

$$\Omega(\begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix}; \begin{Bmatrix} [0] \\ \{-\xi\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\omega\} \end{Bmatrix}) = \tfrac{1}{2}I_{so(m)}([\lambda]) - \tfrac{1}{2}I_{u(n)}(\{\omega\}) - \tfrac{1}{4}I_{so(m)}([\kappa]) + \tfrac{1}{4}I_{u(n)}(\{-\tau\}) + \tfrac{1}{4}I_{u(n)}(\{-\xi\}) - \frac{(m-n-1)}{4}\sum_{\alpha=1}^{r}\tau_\alpha + \frac{(n+1)}{4}\sum_{\alpha=1}^{n}\xi_\alpha, \tag{6.6b}$$

or, more simply,

$$\hat{\Omega} | \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix} \rangle = \Omega(\begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{Bmatrix} [\lambda] \\ \{\sigma\} \end{Bmatrix}) | \begin{Bmatrix} [\lambda^0] \\ \{\sigma^0\} \end{Bmatrix}; \begin{Bmatrix} [\kappa] \\ \{-\tau\} \end{Bmatrix} \begin{matrix} [\lambda](m_\lambda) \\ \{\sigma\}(m_\sigma) \end{matrix} \rangle, \tag{6.6c}$$

when $\xi = 0$ ($\{\omega\} = \{\sigma\}$).

B. Γ—matrix representations for osp($m/2n$)

For $m \neq 2$, the odd elements of the osp($m/2n$) superalgebra tranform as the components of a single irreducible tensor of rank $[1]:\langle 1\rangle$ under the Lie subalgebra $\mathbf{g}_{\bar{0}} = \mathrm{so}(m) \oplus \mathrm{sp}(2n)$. This tensor, denoted F, comprises the irreducible $[1]:\{1\}$ and $[1]:\{-1\}$ $\mathbf{n}_0$ tensors D and E, respectively [cf. Eqs. (2.15)]. For $m = 2$, F is the sum of two irreducible tensors of rank $[\pm]:\langle 1\rangle$, which, respectively, comprises the $\mathbf{n}_0$ tensors $[\pm]:\{1\}$ and $[\pm]:\{-1\}$. To calculate $\mathbf{g}_{\bar{0}}$-reduced matrix elements of F in the $\overline{\Gamma}$ representation, one clearly needs a basis for the representation that reduces the $\mathbf{g}_{\bar{0}} \supset \mathbf{n}_0$ subalgebras. Since the basis (5.14) only reduces the $\mathbf{n}_0$ subalgebra, we introduce another basis that also reduces $\mathbf{g}_{\bar{0}}$. The latter basis states can be expected to have complicated expansions in terms of the VGB basis (5.14). However, we shall show that, since we presuppose a knowledge of the $\mathbf{g}_{\bar{0}}$-irreducible representations, the explicit construction of the latter basis can be avoided.

Suppose that a basis for the VGB space that reduces $\mathbf{g}_{\bar{0}} \supset \mathbf{n}_0$ is given by a set of states

$$|\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix};\begin{smallmatrix}[0]\\ \{-\xi\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \{\omega\}(m_\omega)\end{smallmatrix}\right)\rangle. \tag{6.7}$$

Note that, as distinct from the notation used in defining the VGB states that do not reduce $\mathbf{g}_{\bar{0}}$, here we use angle brackets to designate the irreducible representations $\langle\sigma\rangle$ of sp($2n$) contained in the irreducible representation $(\Lambda^0) = ([\lambda^0]:\langle\sigma^0\rangle)$ of osp($m/2n$). (Recall that we restrict consideration to representation of $\mathbf{g}_{\bar{0}}$ equivalent to Hermitian representations of the compact real form of the Lie algebra.) We also assume that the matrix elements for these Hermitian representations are known.[1–5] It will be noted that, since the (extended) VCS representation $\overline{\Gamma}$ is not Hermitian with respect to the VCB inner product, the basis (6.7) is not orthonormal with respect to the VGB inner product. It is, in fact, orthonormal with respect to an alternative inner product, previously referred to as the VCS inner product[2] (see, also, Sec. VII below). Since we assume full knowledge of the $\mathbf{g}_{\bar{0}}$ representations, these questions need not concern us here. Instead, we simply introduce the orthorgonal dual states

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix};\begin{smallmatrix}[0]\\ \{-\xi\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \{\omega\}(m_\omega)\end{smallmatrix}\right)| \tag{6.8}$$

to the basis (6.7).

Observe now that we can, in this $\mathbf{g}_{\bar{0}}$-reducing basis, safely use the Wigner–Eckart theorem to define $\mathbf{g}_{\bar{0}}$-reduced (triple-bar) matrix elements of the $\mathbf{g}_{\bar{0}}$ tensor $\overline{\Gamma}(F)$ in terms of $\mathbf{n}_0$-reduced (double-bar) matrix elements of $\overline{\Gamma}(D)$ and $\overline{\Gamma}(E)$ according to the equalities

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix};\begin{smallmatrix}[0]\\ \{-\xi'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\omega'\}\end{smallmatrix}\right)\|\overline{\Gamma}(D)\|\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix};\begin{smallmatrix}[0]\\ \{-\xi\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\omega\}\end{smallmatrix}\right)\rangle$$
$$= \langle\langle\sigma\rangle\{-\xi\}\{\omega\};[1]\{1\}\|\langle\sigma'\rangle\{-\xi'\}\{\omega'\}\rangle\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}\right)|||\overline{\Gamma}(F)|||\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\right)\rangle \tag{6.9a}$$

and

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix};\begin{smallmatrix}[0]\\ \{-\xi'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\omega'\}\end{smallmatrix}\right)\|\overline{\Gamma}(E)\|\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix};\begin{smallmatrix}[0]\\ \{-\xi\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\omega\}\end{smallmatrix}\right)\rangle$$
$$= \langle\langle\sigma\rangle\{-\xi\}\{\omega\};[1]\{-1\}\|\langle\sigma'\rangle\{-\xi'\}\{\omega'\}\rangle\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}\right)|||\overline{\Gamma}(F)|||\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\right)\rangle, \tag{6.9b}$$

where $\langle\langle\sigma\rangle\{-\xi\}\{\omega\};[1]\{\pm 1\}\|\langle\sigma'\rangle\{-\xi'\}\{\omega'\}\rangle$ are sp($2n$) $\supset u(\mathrm{n})$ reduced Wigner coefficents. From a knowledge of these $\mathbf{g}_{\bar{0}}$-reduced matrix elements and the appropriate Wigner coefficients, one can retrieve all the matrix elements of $\overline{\Gamma}(D)$ and $\overline{\Gamma}(E)$ between states belonging to Hermitian representations of the $\mathbf{g}_{\bar{0}}$ algebra.

To determine the $\mathbf{g}_{\bar{0}}$-reduced matrix elements in the right-hand sides of (6.9), it is sufficient to determine only the $\mathbf{n}_0$-reducing highest weight states $|\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \langle\sigma\rangle(m_\sigma)\end{smallmatrix}\right)\rangle$ and to use the relationships

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}\right)\|\overline{\Gamma}(D)\|\psi^{\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\right)}\rangle$$
$$= \langle\langle\sigma\rangle\{\sigma\};[1]\{1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}\right)|||\overline{\Gamma}(F)|||\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\right)\rangle, \tag{6.10a}$$

when $\sigma' > \sigma$, and

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}\right)\|\overline{\Gamma}(E)\|\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\right)\rangle$$
$$= \langle\langle\sigma\rangle\{\sigma\};[1]\{-1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}\right)|||\overline{\Gamma}(F)|||\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\right)\rangle, \tag{6.10b}$$

when $\sigma' < \sigma$.

Now, since the $\mathbf{n}_0$-reducing states (5.9) provide a natural and convenient basis for the intrinsic modules $M^{(\Lambda)}(\Lambda^0)$ defined by Eq. (3.8), we are allowed the identification

$$|\psi\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \langle\sigma\rangle(m_\sigma)\end{smallmatrix}\right)\rangle = |\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \{\sigma\}(m_\sigma)\end{smallmatrix}\rangle \tag{6.11}$$

between the subset of the highest grade $\mathbf{g}_{\bar{0}}$-reducing basis and the VG basis.

The dual states

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \langle\sigma\rangle(m_\sigma)\end{smallmatrix}\right)|$$

are not immediately identifiable in this simple way, However, for any state $|\Psi\rangle$ in the VGB space, we have the identity

$$\langle\tilde{\psi}\left(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \langle\sigma\rangle(m_\sigma)\end{smallmatrix}\right)|\Psi\rangle = \langle\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda](m_\lambda)\\ \{\sigma\}(m_\sigma)\end{smallmatrix}|P|\Psi\rangle, \tag{6.12}$$

where P is the projection operator Eq. (6.3) that projects any $\mathbf{g}_{\bar{0}}$ state to its highest weight component. We therefore obtain

$$\langle\tilde{\psi}(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix})\|\overline{\Gamma}(X)\|\psi(\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix})\rangle = \langle\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\sigma'\}\end{smallmatrix}\|P\overline{\Gamma}(X)\|\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\sigma\}\end{smallmatrix}\rangle, \tag{6.13}$$

where X stands for either D or E. Note that we suppress for sake of notational simplicity the symbols ψ, $\tilde{\psi}$ in equations like (6.13) [see also Eq. (6.14) below] in the following since the use of angle brackets of label the irreps $[\lambda]:\langle\sigma\rangle$ of $\mathbf{g}_{\bar{0}}$ is sufficient to distinguish the $\mathbf{g}_{\bar{0}}$-reducing basis.

In terms of matrix elements of the operator $\mathscr{O}$ defined by Eq. (6.1), and of the Bargmann tensor θ and its Grassmann conjugate ∂ in the VG basis (5.9), we finally obtain the $\mathbf{g}_{\bar{0}}$-reducing matrix elements

$$\langle\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}|||\overline{\Gamma}(F)|||\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\rangle = \frac{\langle\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\sigma'\}\end{smallmatrix}\|\overline{\Gamma}(D)\|\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\sigma\}\end{smallmatrix}\rangle}{\langle\langle\sigma\rangle\{\sigma\};(1)\{1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle}$$

$$= \frac{\langle\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\sigma'\}\end{smallmatrix}\|\partial\|\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\sigma\}\end{smallmatrix}\rangle}{\langle\langle\sigma\rangle\{\sigma\};(1)\{1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle}, \tag{6.14a}$$

for $\sigma' > \sigma$, while, for $\sigma' < \sigma$, Eqs. (6.1) and (6.2a) give

$$\langle\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \langle\sigma'\rangle\end{smallmatrix}|||\overline{\Gamma}(F)|||\begin{smallmatrix}[\lambda^0]\\ \langle\sigma^0\rangle\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \langle\sigma\rangle\end{smallmatrix}\rangle = \frac{\langle\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\sigma'\}\end{smallmatrix}\|P\overline{\Gamma}(E)\|\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\sigma\}\end{smallmatrix}\rangle}{\langle\langle\sigma\rangle\{\sigma\};(1)\{-1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle}$$

$$= \frac{\langle\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa']\\ \{-\tau'\}\end{smallmatrix}\begin{smallmatrix}[\lambda']\\ \{\sigma'\}\end{smallmatrix}\|\mathscr{O}\theta\mathscr{O}^{-1}\|\begin{smallmatrix}[\lambda^0]\\ \{\sigma^0\}\end{smallmatrix};\begin{smallmatrix}[\kappa]\\ \{-\tau\}\end{smallmatrix}\begin{smallmatrix}[\lambda]\\ \{\sigma\}\end{smallmatrix}\rangle}{\langle\langle\sigma\rangle\{\sigma\};(1)\{-1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle}. \tag{6.14b}$$

These reduced matrix elements obey $\mathbf{g}_{\bar{0}}$-reduced commutation relations, examples of which will be given in Sec. IX.

C. Subsidiary conditions for finite-dimensionality of representations

Necessary and sufficient conditions for the finite-dimensionality of a representation of any semisimple Lie algebra $\mathbf{g}_{\bar{0}}$ are given by[15]

$$(X_{-\alpha})^{c+1}|\Lambda^0\rangle = 0, \quad c = 2(\Lambda^0,\alpha)/(\alpha,\alpha), \tag{6.15}$$

where $|\Lambda^0\rangle$ is the highest weight state with weight Λ^0, and $X_{-\alpha}$ is an element of $\mathbf{g}_{\bar{0}}$ associated with the root $-\alpha$, where $\alpha\in\Pi_{\bar{0}}$ is a positive simple root. As seen in Sec. II D, the set $\Pi_{\bar{0}}$ for $\mathbf{g}_{\bar{0}}$ corresponds to the distinguished set Π of simple root for $\mathrm{osp}(m/2n)$ except for its (even) simple root $2\delta_n$, which is replaced by the single odd root of Π. Thus, except for the special case $\alpha = 2\delta_n$, which, as just argued, is not a simple root of $\mathrm{osp}(m/2n)$, Eq. (6.15) yields the usual conditions [Eq. (3.5a)] for finite-dimensionality of representations of the Lie algebra $\mathbf{g}_{\bar{0}}$. The VCS formalism provides the following rationale for the subsidiary requirements (3.5b) stemming from consideration of the special case $\alpha = 2\delta_n$ in (6.15).

We want to find the conditions for which

$$(X_{-2\delta_n})^{\sigma_n^0+1}|\Lambda^0\rangle = 0,$$

whenever

$$\sigma_n^0 = 2(\Lambda^0,2\delta_n)/(2\delta_n,2\delta_n) \leqslant r = [m/2]. \tag{6.16}$$

For $X_{-2\delta_n} \sim B_{nn}$, we easily verify that

$$(\Gamma(B_{nn}))^k|\Lambda^0\rangle = (ax+b)^k|\Lambda^0\rangle \tag{6.17a}$$

(no sum on n), where

$$a = 2z_{nn},$$
$$x = \sigma_n^0 - \theta_{cn}\partial_{cn} - \tfrac{1}{2}z_{nn}\nabla_{nn}, \tag{6.17b}$$
$$b = -\theta_{dn}\theta_{en}H_{de}^{(\Lambda^0)}.$$

By recursion, we obtain

$$(ax+b)^{l+1}|\Lambda^0\rangle = \left\{\sum_{i=0}^{l} a^{l+1-i}b^i\binom{l+1}{i} \times\left[\prod_{j=0}^{l-i}(\sigma_n^0 - l + j)\right] + b^{l+1}\right\}|\Lambda^0\rangle,$$

which, for $l = \sigma_n^0$, yields the simple expression

$$(\Gamma(B_{nn}))^{\sigma_n^0+1}|\Lambda^0\rangle = (-\theta_{dn}\theta_{en}H_{de}^{(\Lambda^0)})^{\sigma_n^0+1}|\Lambda^0\rangle, \tag{6.18}$$

an expression fully antisymmetric in the Bargmann variables upon expansion. Using a standard Clifford representation for the intrinsic orthogonal algebra $H^{(\Lambda^0)}$, one can ascertain that necessary conditions for the vanishing of the right-hand side of (6.18) are given by the subsidiary conditions (3.5b) (that they are sufficient has been concisely argued by Kac in Ref. 9).

D. VCS expansion for $\mathbf{g}_{\bar{0}}$ highest weight irreps $\Lambda = \Lambda^0 - \beta$, $\beta\in\Delta_{\bar{1}}^+$

It is interesting to look at the VCS expansion of states obtained from the intrinsic highest Z grade module $M^{(\Lambda^0)}(\Lambda^0)$ by lowering once with the operator $\Gamma(E)$ in Eq. (6.1). The possible $\mathbf{g}_{\bar{0}}$ representation labels are then given by $(\Lambda) = (\Lambda^0 - \beta)$, $\beta\in\Delta_{\bar{1}}^+$, and each such $\mathbf{g}_{\bar{0}}$ irrep appears in a multiplicity-free fashion. The recursion formula for the expansion of these states on the VG basis then simplifies to the first term on the right-hand side of Eq. (6.2b). Setting first β to $\beta = \pm\epsilon_i + \delta_\alpha\in\overline{\Delta}_{\bar{1}}^+$ [cf. Eq. (2.20)] and using (6.5) and (6.6), we readily derive [in (6.19), $\Delta(l)$ is, e.g., a partition having null entries everywhere except for unity in the lth entry] that

$$\mathscr{O}\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}\lambda^0\mp\Delta(i)\\ \sigma^0-\Delta(\alpha)\end{Bmatrix}\right)$$
$$=\Omega\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}[1]\\ -1\end{Bmatrix}\begin{Bmatrix}\lambda^0\mp\Delta(i)\\ \sigma^0-\Delta(\alpha)\end{Bmatrix}\right)-\Omega\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}[0]\\ 0\end{Bmatrix}\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\right)$$
$$=(\Lambda^0+\rho,\pm\epsilon_i+\delta_\alpha), \tag{6.19}$$

in terms of the supersum ρ defined by Eq. (2.22), and the invariant bilinear form defined on the root space by Eq. (2.23). For completeness, we also derive for $\beta=\delta_\alpha\notin\bar{\Delta}_{\bar{1}}^+$ that

$$\mathscr{O}\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\begin{Bmatrix}\lambda^0\\ \sigma^0-\Delta(\alpha)\end{Bmatrix}\right)$$
$$=\Omega\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\begin{Bmatrix}[1]\\ -1\end{Bmatrix}\begin{Bmatrix}\lambda^0\\ \sigma^0-\Delta(\alpha)\end{Bmatrix}\right)-\Omega\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\begin{Bmatrix}[0]\\ 0\end{Bmatrix}\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\right)$$
$$=(\Lambda^0+\rho,\delta_\alpha)-\tfrac{1}{2}. \tag{6.20}$$

Equation (6.19) conveniently summarizes for osp$(m/2n)$ the content of Theorem 3 and Lemmas 5 and 6 of Thierry-Mieg,[10] which state that the $\mathbf{g}_{\bar{0}}$-invariant subspaces identified by the highest weights $\Lambda=\Lambda^0-\beta_i$ $\beta_i\in\bar{\Delta}_{\bar{1}}^+$, such that $(\Lambda^0+\rho,\beta_i)=0$, decouple from the highest weight irreducible representation (Λ^0) of $\mathbf{g}_{\bar{0}}=\text{osp}(m/2n)$.

Irreducible highest weight representation (Λ^0) for which

$$(\Lambda^0+\rho,\beta_i)=0,\quad \beta_i\in\bar{\Delta}_{\bar{1}}, \tag{6.21}$$

have been qualified *atypical* by Kac[9]; otherwise, the representations are called *typical*. [Note that if $(\Lambda^0+\rho,\beta_i)=0$, we also have $(\Lambda^0-\beta_i+\rho,\beta_i)=0$ as a consequence of Eq. (2.24).] As discussed in Sec. V A, Γ-matrix representations of typical and atypical representation are always irreducible. The same does not hold true for the extended $\bar{\Gamma}$-matrix representations defined in Sec. V A when the representation under consideration is atypical: the extended representation is then reducible but not fully reducible, i.e., indecomposable. This is exemplified in Sec. IX.

VII. STAR AND GRADE STAR REPRESENTATIONS

One can define two types of adjoint operations[8] for an irreducible representation γ of a classical Lie superalgebra $\mathbf{g}$ on a Hilbert space.

(i) The star adjoint $\gamma^\dagger(X)$ of an operator $\gamma(X)$ for $X\in\mathbf{g}$ is defined by the usual Hermitian adjoint rule

$$\langle x|\gamma^\dagger(X)|y\rangle=\langle\gamma(X)x|y\rangle. \tag{7.1a}$$

A representation γ of a Lie classical superalgebra $\mathbf{g}$ is then said to be a star representation if, for every $X\in\mathbf{g}$, there is some $Z_X\in\mathbf{g}$ for which

$$\gamma^\dagger(X)=\pm\gamma(Z_X),\quad \gamma^\dagger(Z_X)=\pm\gamma(X). \tag{7.1b}$$

A star representation of a classical Lie superalgebra corresponds to a Hermitian representation of a standard Lie algebra.

(ii) A grade-star adjoint is defined by

$$\langle x|\gamma^\ddagger(X)|y\rangle=(-1)^{\zeta(X)\cdot\zeta(y)}\langle\gamma(X)x|y\rangle, \tag{7.2a}$$

where $\zeta(X)$ is the Z_2 grade of the element $X\in\mathbf{g}$ and $\zeta(y)$ is the Z_2 grade of the ket $|y\rangle$. A representation γ of a classical Lie superalgebra $\mathbf{g}$ is then said to be a grade-star representation if, for every $X\in\mathbf{g}$, there is some $Z_X\in\mathbf{g}$ for which

$$\gamma^\ddagger(X)=\pm\gamma(Z_X),\quad \gamma^\ddagger(Z_X)=\mp\gamma(X). \tag{7.2b}$$

Since we are considering finite-dimensional irreducible representations of osp$(m/2n)$, we require the star and grade-star adjoint operations to be compatible with the Hermitian adjoint operation

$$(H_{ij})^\dagger=H_{ji},\quad (C_{\alpha\beta})^\dagger=C_{\beta\alpha},$$
$$(A_{\alpha\beta})^\dagger=B_{\alpha\beta},\quad (B_{\alpha\beta})^\dagger=A_{\alpha\beta}, \tag{7.3}$$

on the compact real form of (the complexification of) $\mathbf{g}_{\bar{0}}$.

There exist for (finite-dimensional representations of) osp $(m/2n)$ two possibilities (up to equivalence) for the grade star $(\ddagger)$ adjoint operation: we have

$$(D_{a\alpha})^\ddagger=\pm E_{a\alpha}, \tag{7.4a}$$
$$(E_{a\alpha})^\ddagger=\mp D_{a\alpha}. \tag{7.4b}$$

There also exist, for the special case $m=2$, two possibilities for the star adjoint operation: we have

$$(D_{1\alpha})^\dagger=\pm iE_{2\alpha},\quad (E_{2\alpha})^\dagger=\mp iD_{1\alpha},$$
$$(E_{1\alpha})^\dagger=\pm iD_{2\alpha},\quad (D_{2\alpha})^\dagger=\mp iE_{1\alpha}. \tag{7.5}$$

The conditions for a star or grade-star representation can be expressed succintly for an arbitrary representaiton γ of osp $(m/2n)$ by the equation

$$\langle x|\gamma^\dagger(X)|y\rangle=\pm(-1)^{\phi(y)\cdot\zeta(X)}\langle x|\gamma(Z_X)|y\rangle. \tag{7.6}$$

If a solution to this equation exists for every $X\in\mathbf{g}$, then the representation γ is a star representation if $\phi(y)=0$ and a grade-star representation if $\phi(y)=\zeta(y)$.

To bring the irreducible VCS representation Γ into a form in which we can apply this criterion, we seek a K mapping from the VGB space to the irreducible subspace such that the given VGB basis maps to a basis that reduces the $\text{osp}(m/2n)\supset\text{so}(m)\oplus\text{sp}(2n)\supset\text{so}(m)\oplus\text{u}(n)$ subalgebra chain.

The similarity transform K can be equivalently defined by

$$K\colon\Gamma(X)\to\gamma(X)=K^{-1}\Gamma(X)K,\quad X\in\text{osp}(m/2n). \tag{7.7}$$

Since the VCS representation Γ is, by construction, Hermitian with respect to the VGB measure on restriction to the stability algebra $\mathbf{n}_0$, it is convenient to require that K commute with the VCS representation of the stability algebra, i.e.,

$$\Gamma(X)K=K\Gamma(X),\quad \forall X\in\mathbf{n}_0. \tag{7.8}$$

The K operator, diagonal in $\mathbf{n}_0$, can be defined in terms of its matrix elements,

$$\left\langle\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}[\kappa]\\ -\tau\end{Bmatrix}\begin{Bmatrix}\lambda'\\ \sigma'\end{Bmatrix};\begin{Bmatrix}[0]\\ -\xi'\end{Bmatrix}\begin{Bmatrix}\lambda'\\ \omega\end{Bmatrix}\middle|K\middle|\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}[\kappa]\\ -\tau\end{Bmatrix}\begin{Bmatrix}\lambda\\ \sigma\end{Bmatrix};\begin{Bmatrix}[0]\\ -\xi\end{Bmatrix}\begin{Bmatrix}\lambda\\ \omega\end{Bmatrix}\right\rangle=0,$$
$$\text{for } \sigma'\geqslant\sigma, \tag{7.9}$$

between VGB $\mathbf{n}_0$ submodules. One need not compute K on the whole VGB space; rather, it is sufficient to know the value of its restriction

$$\mathscr{K}\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\begin{Bmatrix}\lambda\\ \sigma\end{Bmatrix}\right)_{\begin{Bmatrix}[\kappa]\\ -\tau\end{Bmatrix}\begin{Bmatrix}[\kappa']\\ -\tau'\end{Bmatrix}}$$
$$=\left\langle\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}[\kappa]\\ -\tau\end{Bmatrix}\begin{Bmatrix}\lambda\\ \sigma\end{Bmatrix}\middle|K\middle|\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}[\kappa']\\ -\tau'\end{Bmatrix}\begin{Bmatrix}\lambda\\ \sigma\end{Bmatrix}\right\rangle, \tag{7.10}$$

for the $M^{(\Lambda)}(\Lambda^0)$ subspaces defined in Sec. III C. It is convenient to require that the restriction $\mathscr{K}$ be Hermitian. It is also convenient to set

$$\mathscr{K}\left(\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix};\begin{Bmatrix}\lambda^0\\ \sigma^0\end{Bmatrix}\right)=1.$$

Following a projection procedure similar to the one invoked in Sec. VI A, one easily derives from the definition of the γ representation that

$$\begin{aligned}\tilde{P}\left[\Theta^{[\kappa]}_{\{-\tau\}}(\gamma(E))\times|^{[\lambda^0]}_{\{\sigma^0\}}\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}&=\tilde{P}\left[\Theta^{[\kappa]}_{\{-\tau\}}(K^{-1}\Gamma(E)K)\times|^{[\lambda^0]}_{\{\sigma^0\}}\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\\&=\mathscr{K}^{-1}(^{[\lambda^0]}_{\{\sigma^0\}}{}^{[\lambda]}_{\{\sigma\}})P\left[\Theta^{[\kappa]}_{\{-\tau\}}(\Gamma(E))\times|^{[\lambda^0]}_{\{\sigma^0\}}\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)},\end{aligned}\tag{7.11a}$$

where $\tilde{P}=K^{-1}PK$. Similarly, from Eqs. (7.4)–(7.6), one finds that

$$\begin{aligned}&\tilde{P}\left[\Theta^{[\kappa]}_{\{-\tau\}}(\gamma(E))\times|^{[\lambda^0]}_{\{\sigma^0\}}\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\\&\quad=(\pm1)^{n_\theta}(-1)^{\sum_{i=0}^{n_\theta-1}\phi(i)}\\&\qquad\times\tilde{P}(^{[\lambda]}_{\{\sigma\}}\left[\Theta^{[\kappa]}_{\{-\tau\}}(K^{\dagger}\Gamma^{\dagger}(D)K^{-1\dagger})\times|^{[\lambda^0]}_{\{\sigma^0\}}\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\\&\quad=(\pm1)^{n_\theta}(-1)^{\sum_{i=0}^{n_\theta-1}\phi(i)}\\&\qquad\times\mathscr{K}(^{[\lambda^0]}_{\{\sigma^0\}}{}^{[\lambda]}_{\{\sigma\}})|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\rangle,\end{aligned}\tag{7.11b}$$

where $\phi(i)$ either stands for $\phi(i)=0$ for a star equivalent representation, or for the Z_2 grade $\phi(i)=\zeta(\Lambda^0)-i$ of a VG state of rank i in the Grassmann variables for a grade-star equivalent representation and where n_θ is the rank of the polynominal Θ. Equating the right-hand sides of Eqs. (7.11a) and (7.11b), we thus obtain

$$\begin{aligned}&P\left[\Theta^{[\kappa]}_{\{-\tau\}}(\Gamma(E))\times|^{[\lambda^0]}_{\{\sigma^0\}}\rangle\right]^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\\&\quad=(\pm1)^{n_\theta}(-1)^{\sum_{i=0}^{n_\theta-1}\phi(i)}\\&\qquad\times\mathscr{K}^2(^{[\lambda^0]}_{\{\sigma^0\}}{}^{[\lambda]}_{\{\sigma\}})|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda](m_\lambda)}_{\{\sigma\}(m_\sigma)}\rangle\end{aligned}\tag{7.12}$$

from which one could derive the following recursion formula for the square of the Hermitian restriction $\mathscr{K}$ of K [Eq. (7.10)]

$$\begin{aligned}\langle^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda']}_{\{\sigma'\}}\|\mathscr{K}^2\theta\|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\rangle&=\pm(-1)^{\phi(\Lambda)}\\&\times\langle^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\|P\Gamma(E)\mathscr{K}^2\|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\rangle.\end{aligned}\tag{7.13}$$

Actually, it is not necessary to solve (7.13) as (6.1) and (7.12) readily imply that

$$\mathscr{K}^2=(\pm1)^{n_\theta}(-1)^{\sum_{i=0}^{n_\theta-1}\phi(i)}\mathscr{O},\tag{7.14a}$$

which, for star equivalent representations, simplifies to

$$\mathscr{K}^2=(\pm1)^{n_\theta}\mathscr{O},\tag{7.14b}$$

and, for grade-star equivalent representations, to

$$\mathscr{K}^2=(\pm(-1)^{\zeta(\Lambda^0)})^{n_\theta}(-1)^{n_\theta(n_\theta-1)/2}\mathscr{O}.\tag{7.14c}$$

Equation (7.14) provides us with the means to identify the classes of finite-dimensional representations (Λ^0) of osp$(m/2n)$ which can be declared equivalent to star or grade-star representations (this will be demonstrated in the next two sections); simply, it is sufficient to identify the representations for which the right-hand sides of Eqs. (7.14b) and (7.14c) define positive-definite operators $\mathscr{K}^2$ as requested by the definition of the latter.

Finally, we derive from Eq. (6.14) that $\mathfrak{g}_{\bar{0}}$-reduced matrix elements for the γ representations are given, for $\sigma'>\sigma$, by

$$\begin{aligned}\langle^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda']}_{\langle\sigma'\rangle}|||\gamma(F)|||^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\langle\sigma\rangle}\rangle&=\frac{\langle^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\|\gamma(D)\|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\rangle}{\langle\langle\sigma\rangle\{\sigma\};\langle1\rangle\{1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle}\\&=\frac{\langle^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\|\mathscr{K}^{-1}\partial\mathscr{K}\|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\rangle}{\langle\langle\sigma\rangle\{\sigma\};\langle1\rangle\{1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle};\end{aligned}\tag{7.15a}$$

thus, more simply, by

$$\langle^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\langle\sigma'\rangle}|||\gamma(F)|||^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\langle\sigma\rangle}\rangle=\langle^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\langle\sigma'\rangle}|||\mathscr{K}^{-1}\Gamma(F)\mathscr{K}|||^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\langle\sigma\rangle}\rangle.\tag{7.15b}$$

For $\sigma'<\sigma$, the elements are given by

$$\begin{aligned}\langle^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\langle\sigma'\rangle}|||\gamma(F)|||^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\langle\sigma\rangle}\rangle&=\frac{\langle^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\|\tilde{P}\gamma(E)\|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\rangle}{\langle\langle\sigma\rangle\{\sigma\};\langle1\rangle\{-1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle}\\&=\pm(-1)^{\phi(\Lambda)}\frac{\langle^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\{\sigma'\}}\|\mathscr{K}\theta\mathscr{K}^{-1}\|^{[\lambda^0]}_{\{\sigma^0\}};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\{\sigma\}}\rangle}{\langle\langle\sigma\rangle\{\sigma\};\langle1\rangle\{-1\}\|\langle\sigma'\rangle\{\sigma'\}\rangle};\end{aligned}\tag{7.16a}$$

or, more simply, by

$$\langle^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\langle\sigma'\rangle}|||\gamma(F)|||^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\langle\sigma\rangle}\rangle=\pm(-1)^{\phi(\Lambda)}\langle^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa']}_{\{-\tau'\}}{}^{[\lambda']}_{\langle\sigma'\rangle}|||\mathscr{K}^{-1}\Gamma(F)\mathscr{K}|||^{[\lambda^0]}_{\langle\sigma^0\rangle};^{[\kappa]}_{\{-\tau\}}{}^{[\lambda]}_{\langle\sigma\rangle}\rangle.\tag{7.16b}$$

VIII. THE LIE SUPERALGEBRA osp(1/2n)

We consider in this section the superalgebra osp$(1/2n)$. An irrep of osp$(1/2n)$ is finite dimensional if the highest weight

$$\Lambda^0=\sigma^0=\sum_{\alpha=1}^{n}\sigma^0_\alpha\delta_\alpha\tag{8.1a}$$

is such that

$$\sigma^0_i-\sigma^0_{i+1}\in Z^+,\quad\sigma^0_n\in Z^+\tag{8.1b}$$

[cf. Eq. (3.5)]. Since the partitions $\{-\tau\}$ for the Grassmann polynomials $\Theta_{\{-\tau\}}(\theta)$ restrict to the set $\{-1\}^k$, $0\leqslant k\leqslant n$, we conclude that representations of osp$(1/2n)$ are multiplicity-free on restriction to $\mathbf{g}_{\bar{0}}$ [here sp$(2n)$]. This can be easily verified by noting that the u(n) couplings

$$\{\sigma^0\}\times\{-1^k\}\to\{\sigma^0-\Delta(\mathbf{j})\}=\{\sigma^0-\Delta(j_1,j_2,\ldots,j_k)\}=\{\sigma\} \tag{8.2}$$

[where $\Delta(j_1,j_2,\ldots j_k)$, $1\leqslant j_1<j_2<\ldots<j_k\leqslant n$, is a null n vector except for the numerical one is its $(j_1,j_2,\ldots,j_k)$ entries] in Eq. (5.9) are multiplicity free.

The VCS expansion (4.20), with $\theta_\alpha=\theta_{1\alpha}$, simplifies to

$$\Gamma(A_{\alpha\beta})=\nabla_{\alpha\beta}, \tag{8.3a}$$

$$\Gamma(D_\alpha)=\partial_\alpha-\tfrac{1}{2}\theta_\mu\nabla_{\mu\alpha}, \tag{8.3b}$$

$$\Gamma(C_{\alpha\beta})=C_{\alpha\beta}^{(\Lambda^0)}-\theta_\beta\,\partial_\alpha-z_{\beta\mu}\nabla_{\mu\alpha}, \tag{8.3c}$$

$$\Gamma(E_\alpha)=\theta_\mu(C_{\mu\alpha}^{(\Lambda^0)}-\tfrac{1}{2}z_{\alpha\sigma}\nabla_{\sigma\mu}-\tfrac{1}{2}\theta_\alpha\,\partial_\mu)+z_{\alpha\sigma}\,\partial_\sigma, \tag{8.3d}$$

$$\Gamma(B_{\alpha\beta})=z_{\alpha\sigma}(C_{\sigma\beta}^{(\Lambda^0)}-\theta_\beta\,\partial_\sigma-\tfrac{1}{2}z_{\beta\xi}\nabla_{\xi\sigma})-\tfrac{1}{2}\theta_\alpha\,\theta_\sigma C_{\sigma\beta}^{(\Lambda^0)}+(\alpha\leftrightarrow\beta). \tag{8.3e}$$

We find that $\Gamma^{(0)}(B)$, the term quadratic in the Grassmann variables in (8.3e), here can be rewritten more simply as

$$\Gamma^{(0)}(B_{\{-2\}})=\tfrac{1}{2}[\theta_{\{-1\}}\times\langle\hat{I}_{\mathrm{u}(n)}-\hat{I}^{(\theta)}_{\mathrm{u}(n)},\theta_{\{-1\}}\rangle]_{\{-2\}}, \tag{8.4}$$

where the product $[\theta\times\langle\ldots,\theta\rangle]$ stands for a u(n) coupling, and where the quadratic Casimir operators $\hat{I}$ have been defined in Eq. (6.5).

The recursion formula (6.2) for the coefficients of expansion $\mathscr{O}$ for the VCS submodules $M^{(\sigma)}(\sigma^0)$ can be written

$$\begin{aligned}&\langle\{\sigma^0\}\{-1^k\}\{\sigma\}\|\mathscr{O}\theta\|\{\sigma^0\}\{-1^{k-1}\}\{\sigma+\Delta(j_d)\}\rangle\\&\quad=\langle\{\sigma^0\}\{-1^k\}\{\sigma\}\|\Gamma^{(0)}(E)\mathscr{O}\|\{\sigma^0\}\{-1^{k-1}\}\{\sigma+\Delta(j_d)\}\rangle\\&\qquad-\sum_{\substack{j_c=1\\ \neq j_d}}^{n}\frac{\langle\{\sigma^0\}\{-1^k\}\{\sigma\}\|\Gamma^{(0)}(B)\|\{\sigma^0\}\{-1^{k-2}\}\{\sigma+\Delta(j_c,j_d)\}\rangle}{\langle\{\sigma^0\}\{-1^{k-2}\}\{\sigma+\Delta(j_c,j_d)\}\{-2\}\{\sigma\}\|\Gamma^{(1)}(B)\|\{\sigma^0\}\{-1^{k-2}\}\{\sigma+\Delta(j_c,j_d)\}\rangle}\\&\qquad\times\langle\{\sigma^0\}\{-1^{k-2}\}\{\sigma+\Delta(j_c,j_d)\}\{-2\}\{\sigma\}\|\Gamma^{(1)}(E)\mathscr{O}\|\{\sigma^0\}\{-1^{k-1}\}\{\sigma+\Delta(j_d)\}\rangle,\end{aligned} \tag{8.5}$$

which yields

$$\begin{aligned}\mathscr{O}(\{\sigma^0\};\{\sigma\})&=\left\{p^0_{j_dn}-\frac{k-1}{2}+\frac{1}{2}\sum_{\substack{j_c=1\\ \neq j_d}}^{n}\frac{(p^0_{j_cn}-p^0_{j_dn})}{(p^0_{j_cn}+p^0_{j_dn}+1)}\prod_{\substack{l=1\\ \neq c}}^{k}\frac{(p^0_{j_cn}-p^0_{j_ln}+1)}{(p^0_{j_cn}-p^0_{j_ln})}\right\}\mathscr{O}(\{\sigma^0\};\{\sigma+\Delta(j_d)\})\\&=p^0_{j_dn}\prod_{\substack{l=1\\ \neq d}}^{k}\frac{(p^0_{j_dn}+p^0_{j_ln})}{(p^0_{j_dn}+p^0_{j_ln}+1)}\mathscr{O}(\{\sigma^0\};\{\sigma+\Delta(j_d)\}),\end{aligned} \tag{8.6}$$

and which has for a solution

$$\mathscr{O}(\{\sigma^0\};\{\sigma\})=\left[\prod_{l=1}^{k}p^0_{j_ln}\right]\times\left[\prod_{l<m}^{k}\frac{(p^0_{j_ln}+p^0_{j_mn})}{(p^0_{j_ln}+p^0_{j_mn}+1)}\right], \tag{8.7}$$

where p^0_{in} is the so-called partial hook

$$p^0_{in}=p^0_i+n-i.$$

In particular, for $k=1$, we obtain

$$\mathscr{O}(\{\sigma^0\};\{\sigma^0-\Delta(j_1)\})=(\Lambda+\rho,\delta_{j_1})-\tfrac{1}{2}=p_{j_1n} \tag{8.8}$$

[see Eq. (6.20)]. Under the restrictions (8.1), we conclude that these coefficients of expansion are semi-positive-definite. In fact, they are strictly positive definite except for the coefficients associated with the partitions $\{\sigma\}$ defined by (8.2) with $j_k=n$ and $p^0_{nn}=\sigma^0_n=0$.

Equations (8.6) and (8.7) and very strong analytical results since they can allow one to compute in principle the sp$(2n)$-reduced matrix elements (6.14). This would amount to an explicit construction of the representations. In order to carry out this program successfully, one needs only the relevant sp$(2n)\supset$u(n) Wigner coefficients.

Only the grade-star adjoint operations (7.4) can be defined on osp$(1/2n)$ when we restrict our attention to the real compact Lie subalgebra sp$(2n)$ of osp$(1/2n)$. We therefore seek to identify the possible classes of grade-star equivalent representations of osp$(1/2n)$. From Eq. (7.14c), we know that the restriction $\mathscr{K}^2$ is given by

$$\begin{aligned}&\mathscr{K}^2(\{\sigma^0\};\{\sigma\})\\&\quad=(\pm(-1)^{\zeta(\sigma^0)})^k(-1)^{k(k-1)/2}\mathscr{O}(\{\sigma^0\};\{\sigma\}).\end{aligned} \tag{8.9a}$$

But the positive definitiveness of $\mathscr{K}^2$ requires that $\pm(-1)^{\zeta(\sigma^0)}=+1$, for $k=1$. This determines which one of the two possible grade-star operations (7.4a) is compatible with the Z_2 grade of the highest Z (intrinsic) subspace.

Equation (8.9a) then simplifies to

$$\mathscr{K}^2(\{\sigma^0\};\{\sigma\}) = (-1)^{k(k-1)/2}\mathscr{O}(\{\sigma^0\};\{\sigma\}). \tag{8.9b}$$

Since the various quantities $\mathscr{O}(\{\sigma^0\};\{\sigma\})$ are semi-positive-definite, we conclude that they must actually vanish whenever the phase $(-1)^{k(k-1)/2}$ is negative for the representation to be grade-star equivalent. For example, consider the osp(1/4) case: we have, from (8.7) and (8.9b),

$$\begin{aligned}
&\mathscr{K}^2(\{\sigma_1^0\sigma_2^0\};\{\sigma_1^0\sigma_2^0\}) = 1,\\
&\mathscr{K}^2(\{\sigma_1^0\sigma_2^0\};\{\sigma_1^0-1,\sigma_2^0\}) = \sigma_1^0+1,\\
&\mathscr{K}^2(\{\sigma_1^0\sigma_2^0\};\{\sigma_1^0,\sigma_2^0-1\}) = \sigma_2^0,\\
&\mathscr{K}^2(\{\sigma_1^0\sigma_2^0\};\{\sigma_1^0-1,\sigma_2^0-1\})\\
&\quad = -\sigma_2^0(\sigma_1^0+1)\left[\frac{\sigma_1^0+\sigma_2^0+1}{\sigma_1^0+\sigma_2^0+2}\right],
\end{aligned} \tag{8.10}$$

from which we conclude that a representation of osp(1/4) is grade star if and only if the σ_2^0 vanish (see, also, Ref. 16). More generally, only a very restricted set of representations of osp(1/2n) can be realized as grade-star representations for $n>1$. The case $n=1$ is studied in more depth in Sec. IX.

IX. THE LIE SUPERALGEBRA osp(m/2)

In this section, we intend to study the compact classical Lie superalgebra osp(m/2), for $m=1$, 2, and 3. These algebras have been studied to a large extent in the literature. We nevertheless choose, in addition to the computation of new closed analytical results, to rederive some known results[16–18] in order to show that the somewhat disparate approaches found in the literature can be encompassed in the present unifying framework. These low rank subalgebras have been chosen, first and foremost, because the Wigner–Racah calculus for sp(2)$\sim$su(2) is well known; and since Kronecker products in su(2) are multiplicity-free, we are able to carry out explicitly all the computations. Also, in going from $m=1$ to $m=3$, the tensorial structure of the Lie superalgebras osp(m/2) increases in complexity. The case $m=2$ is notable for the fact that the odd subalgebra $\mathbf{g}_{\bar{1}}$ is reducible under $\mathbf{g}_{\bar{0}}$, while the case $m=3$ presents all the complexities of the general problem except for the fact that the $\mathbf{g}_{\bar{0}}=\mathrm{so}(3)\oplus\mathrm{sp}(2)$ algebra is isomorphic to su(2)$\oplus$su(2), thus allowing us once more to give fully analytical results. It should be noted, though, that the corresponding VCS algebraic manipulations do not significantly increase in complexity with increasing m; the relevant information concerning the superstructure of the algebras and their representations is concisely carried by a few $\mathbf{n}_0$- and $\mathbf{g}_{\bar{0}}$-reduced quantities easily computed within the present framework. This results in a large economy for the more general situation, especially as the ranks of the algebras increase.

A. Tensorial structure of osp(1/2), osp(2/2), and osp(3/2)

First, we slightly modify the notation to take advantage of the sp(2)$\sim$su(2) isomorphism and of our knowledge of the Wigner–Racah calculus for su(2). In terms of the basis $\{A_{11}, C_{11}, B_{11}\}$ for sp(2) (Sec. II A 2), we define the sp(2)$\sim$su$_J$(2) angular momentum algebra,

$$J_+ = \tfrac{1}{2}A_{11},\quad J_0 = \tfrac{1}{2}C_{11},\quad J_- = \tfrac{1}{2}B_{11}, \tag{9.1a}$$

with the ususal commutation relations

$$\langle J_0, J_\pm\rangle = \pm J_+,\quad \langle J_+, J_-\rangle = 2J_0. \tag{9.1b}$$

Also, for osp(2/2), the Lie subalgebra so(2)$\sim$u$_B$(1) will be generated by the "baryonic" operator[17]

$$B = \mathscr{C}_{11}/2 = -iH_{12}/2 \tag{9.2}$$

(this redefinition should not give rise to any confusion with the symplectic generator $B_{\alpha\beta}$ in the following), while, for osp(3/2), the so(3) subalgebra will be generated by

$$\begin{aligned}
L_{+1} &= \mathscr{D}_1 = (1/\sqrt{2})(H_{13}+iH_{23}),\\
L_0 &= \mathscr{C}_{11} = -iH_{12},\\
L_{-1} &= \mathscr{E}_1 = (1/\sqrt{2})(H_{13}-iH_{23}).
\end{aligned} \tag{9.3}$$

We also have

$$\begin{aligned}
&\text{osp(1/2):}\quad \mathbf{n}_0 \simeq \mathrm{u}_J(1),\quad \mathbf{g}_{\bar{0}} \simeq \mathrm{su}_J(2),\\
&\text{osp(2/2):}\quad \mathbf{n}_0 \simeq u_B(1)\oplus \mathrm{u}_J(1),\quad \mathbf{g}_{\bar{0}} \simeq \mathrm{u}_B(1)\oplus \mathrm{su}_J(2),\\
&\text{osp(3/2):}\quad \mathbf{n}_0 \simeq \mathrm{so}(3)\oplus \mathrm{u}_J(1),\quad \mathbf{g}_{\bar{0}} \simeq \mathrm{so}(3)\oplus \mathrm{su}_J(2).
\end{aligned} \tag{9.4}$$

The grading operator Z is here given by the weight operator $2J_0$, spanning the u$_J$(1)$\subset$su$_J$(2) subalgebra. We thus have that the tensorial sets $\mathbf{n}_{+1}=\mathrm{span}\{D\}$ and $\mathbf{n}_{-1}=\mathrm{span}\{E\}$ are, respectively, the spin-up ($+\frac{1}{2}$) and spin-down ($-\frac{1}{2}$) components of a (reducible for $m=2$) spinorial ($J=\frac{1}{2}$) tensor $\{F\}=\{D\}\cup\{E\}$. More precisely, we have

$$\begin{aligned}
&\text{osp(1/2):}\quad \mathbf{n}_{+1}=\mathrm{span}\{F_{+1/2}=D_{11}\},\\
&\qquad\qquad \mathbf{n}_{-1}=\mathrm{span}\{F_{-1/2}=E_{11}\};\\
&\text{osp(2/2):}\quad \mathbf{n}_{+1}=\mathrm{span}\{F^{[\pm 1/2]}_{+1/2}=\mp(1/\sqrt{2})(D_{11}\pm iD_{21})\},\\
&\qquad\qquad \mathbf{n}_{-1}=\mathrm{span}\{F^{[\pm 1/2]}_{-1/2}=\mp(1/\sqrt{2})(E_{11}\pm iE_{21})\};\\
&\text{osp(3/2):}\quad \mathbf{n}_{+1}=\mathrm{span}\{F^{[1]m=0}_{+1/2}=D_{31}, F^{[1]m=\pm 1}_{+1/2}=\mp(1/\sqrt{2})(D_{11}\pm iD_{21})\},\\
&\qquad\qquad \mathbf{n}_{-1}=\mathrm{span}\{F^{[1]m=0}_{-1/2}=E_{31}, F^{[1]m=\pm 1}_{-1/2}=\mp(1/\sqrt{2})(E_{11}\pm iE_{21})\};
\end{aligned} \tag{9.5}$$

i.e., $\mathfrak{n}_{+1}$ is n_0-irreducible for osp(1/2); $\mathfrak{n}_0$-reducible for osp(2/2) (with $F^{[+1/2]}_{+1/2}$ a B eigenvector of eigenvalue $b=+\frac{1}{2}$, $F^{[-1/2]}_{+1/2}$ of eigenvalue $b=-\frac{1}{2}$); and $\mathfrak{n}_0$-irreducible for osp(3/2) [with $F^{[1]}_{+1/2}$ an $L=1$ so(3) tensor]; with similar conclusions for $\mathfrak{n}_{-1}$. We also have that $\mathfrak{g}_{\bar{1}}=\mathfrak{n}_{+1}\oplus\mathfrak{n}_{-1}$ is $\mathfrak{g}_{\bar{0}}$-irreducible for osp(1/2) (with F a J spinor); $\mathfrak{g}_{\bar{0}}$-reducible for osp(2/2) (with $F^{[+1/2]}$ a $b=\frac{1}{2}J$ spinor, $F^{[-1/2]}$ a $b=-\frac{1}{2}J$ spinor); and $\mathfrak{g}_{\bar{0}}$-irreducible for osp(3/2) (with $F^{[1]}$ a $L=1$ J spinor).

We shall make the substitutions

$$\nabla=\tfrac{1}{2}\nabla_{11},\quad z=z_{11}, \tag{9.6a}$$

in the VCS expansion (4.20) so that

$$\langle\nabla,z\rangle=1. \tag{9.6b}$$

Also, the $\mathfrak{u}_J(1)$ intrinsic subalgebra

$$J_0^{(\Lambda^0)}=\tfrac{1}{2}C_{11}^{(\Lambda^0)} \tag{9.7a}$$

is defined such that

$$J_0^{(\Lambda^0)}|^{[\lambda^0](m^0_\lambda)}_{j^0}\rangle=j^0|^{[\lambda^0](m_{\lambda^0})}_{j^0}\rangle;\quad 2j^0=\sigma^0_1, \tag{9.7b}$$

where $\{|^{[\lambda^0](m_{\lambda^0})}_{j^0}\rangle\}$ spans a basis for the $M^{(\Lambda^0)}(\Lambda^0)$ $\mathfrak{n}_0$-intrinsic module, and the set $\{(m_{\lambda^0})\}$ spans a basis for the so(m) intrinsic irrep $[\lambda^0]$.

The label $[\lambda^0]$ is redundant for $m=1$. For $m=2$, we set

$$[\lambda^0](m_{\lambda^0})\to[b^0] \tag{9.8a}$$

(irreducible representations $[b^0]$ of $so(2)\sim\mathfrak{u}_B(1)$ are one dimensional) such that the intrinsic operator $B^{(\Lambda^0)}$ is defined by

$$B^{(\Lambda^0)}|^{[b^0]}_{j^0}\rangle=b^0|^{[b^0]}_{j^0}\rangle. \tag{9.8b}$$

For $m=3$, we set

$$[\lambda^0](m^0_\lambda)\to[l^0]m \tag{9.9a}$$

such that the intrinsic angular momentum algebra $\mathbf{L}^{(\Lambda^0)}$ acts on the intrinsic basis according to the transformation law

$$L^{(\Lambda^0)}_{m'}|^{[l^0]m}_{j^0}\rangle=\langle l^0m;1m'|l^0m+m'\rangle\sqrt{l^0(l^0+1)}\;|^{[l^0]m+m'}_{j^0}\rangle, \tag{9.9b}$$

where $\langle l^0m;1m'|l^0m+m'\rangle$ is the usual su(2) coupling (Clebsch–Gordan) coefficient.

2. The Lie superalgebra osp(1/2)

In this section we examine the finite-dimensional representations of the Lie superalgebra osp(1/2) with highest weights

$$\Lambda^0=\sigma^0_1\delta_1=2j^0\delta_1. \tag{9.10}$$

We find, in angular momentum notation, the following commutation relations for osp(1/2):

$$\langle J_\alpha,J_\beta\rangle=\sqrt{2}\langle 1\beta;1\alpha|1\gamma\rangle J_\gamma, \tag{9.11a}$$

$$\langle J_\alpha,F_\mu\rangle=\sqrt{\tfrac{3}{4}}\langle\tfrac{1}{2}\mu;1\alpha|\tfrac{1}{2}\nu\rangle F_\nu, \tag{9.11b}$$

$$\langle F_\alpha,F_\beta\rangle=2\sqrt{2}\langle\tfrac{1}{2}\beta;\tfrac{1}{2}\alpha|1m\rangle J_m. \tag{9.11c}$$

The VCS expansion (4.20) (with $\theta_{11}=\theta$) simplifies to

$$\Gamma(J_+)=\nabla, \tag{9.12a}$$

$$\Gamma(F_{+1/2})=\partial-\theta\nabla, \tag{9.12b}$$

$$\Gamma(J_0)=J_0^{(\Lambda^0)}-\tfrac{1}{2}\theta\partial-z\nabla, \tag{9.12c}$$

$$\Gamma(F_{-1/2})=\theta(2J_0^{(\Lambda^0)}-z\nabla)+z\partial, \tag{9.12d}$$

$$\Gamma(J_-)=z(2J_0^{(\Lambda^0)}-\theta\partial-z\nabla). \tag{9.12e}$$

The VGB basis $|j^0;j\rangle$ for $m_j=j$ states is given by

$$|j^0;j^0\rangle=|j^0\rangle, \tag{9.13a}$$

$$|j^0;j^0-\tfrac{1}{2}\rangle=\theta|j^0\rangle. \tag{9.13b}$$

From (6.1) and (6.2), we find the following values for the coefficients of expansion of the VCS basis on the VG basis:

$$\mathcal{O}(j^0;j^0)=1,$$
$$\mathcal{O}(j^0;j^0-\tfrac{1}{2})=(\Lambda^0+\rho,\delta_1)-\tfrac{1}{2}=2j^0, \tag{9.14}$$

also given by (8.7) with $k=n=1$.

We find, in the ordered VG basis (9.13), the following $\mathfrak{g}_{\bar{0}}$-reduced Γ-matrix representation:

$$\langle j^0;j'|||\Gamma(F)|||j^0;j\rangle=\begin{pmatrix}0 & 1\\ 2j^0[(2j^0+1)/2j^0]^{1/2} & 0\end{pmatrix}. \tag{9.15}$$

We verify that it obeys the reduced-commutator algebra

$$\sum_{j''}U\left(j,\frac{1}{2},j',\frac{1}{2};j''1\right)\times\langle j^0;j'|||F|||j^0;j''\rangle\langle j^0;j''|||F|||j^0;j\rangle=\delta_{jj'}\sqrt{2j(j+1)}, \tag{9.16}$$

easily deduced from Eq. (9.11c).

Only the two grade-star adjoint operations

$$(F_\alpha)^\ddagger=\pm(-1)^{1/2-\alpha}F_{-\alpha} \tag{9.17}$$

are possible for osp(1/2) whenever we restrict our attention to its real compact Lie subalgebra sp(2). Accordingly, we find the following values for the restrictions $\mathscr{K}^2$ [Eq. (7.14c)]:

$$\mathscr{K}^2(j^0;j^0)=1,$$
$$\mathscr{K}^2(j^0;j^0-\tfrac{1}{2})=\pm(-1)^{\xi(j^0)}2j^0. \tag{9.18}$$

The positive-definitiveness of $\mathscr{K}^2$ implies that $\pm(-1)^{\xi(j^0)}=+1$, which determines which one of the two grade-star adjoint operations in (9.17) is compatible with the Z_2 grade of the highest weight state. We conclude that the irrep j^0 of osp(1/2) is a grade-star representation for all $j^0\geqslant 0$.

The $\mathfrak{g}_{\bar{0}}$-reduced γ-matrix representation of the odd tensor of osp(1/2) equivalent to (9.15) is given, in the ordered basis (9.13), by

$$\langle j^0;j'|||\gamma(F)|||j^0;j\rangle=\begin{pmatrix}0 & [\pm(-1)^{\xi(j^0)}(2j^0)]^{1/2}\\ \pm(-1)^{\xi(j^0)}[\pm(-1)^{\xi(j^0)}(2j^0+1)]^{1/2} & 0\end{pmatrix}, \tag{9.19}$$

which also verifies (9.16). These sp(2)-reduced γ-matrix elements obey the grade-star conjugation rule

$$\langle j^0;j'|||\gamma(F)|||j^0;j\rangle = \pm(-1)^{\zeta(j')}(-1)^{j+1/2-j'}[\dim(j)/\dim(j')]^{1/2}\langle j^0;j|||\gamma(F)|||j^0;j'\rangle \tag{9.20}$$

easily derived from Eqs. (7.2) and (9.17). These results reproduce concisely the analysis of Scheunert *et al.*[17]

C. The Lie superalgebra osp(2/2)

In this section we examined the finite-dimensional representations of the Lie superalgebra osp(2/2) with highest weights

$$\Lambda^0 = \lambda_1^0\epsilon_1 + \sigma_1^0\delta_1 = 2b^0\epsilon_1 + 2j^0\delta_1, \tag{9.21}$$

with b^0 here a real number. From Sec. II, we find the following commutation relations in su(2) coupled form:

$$\langle J_\alpha, J_\beta\rangle = \sqrt{2}\langle 1\beta;1\alpha|1\gamma\rangle J_\gamma, \tag{9.22a}$$

$$\langle J_\alpha, F_\mu^{[\pm 1/2]}\rangle = \sqrt{\tfrac{3}{4}}\langle \tfrac{1}{2}\mu;1\alpha|\tfrac{1}{2}\nu\rangle F_\nu^{[\pm 1/2]}, \tag{9.22b}$$

$$\langle B, F_\alpha^{[+1/2]}\rangle = \tfrac{1}{2}F_\alpha^{[+1/2]}, \quad \langle B, F_\alpha^{[-1/2]}\rangle = -\tfrac{1}{2}F_\alpha^{[-1/2]}, \tag{9.22c}$$

$$\langle F_\alpha^{[+1/2]}, F_\beta^{[-1/2]}\rangle = -2\sqrt{2}\langle \tfrac{1}{2}\beta;\tfrac{1}{2}\alpha|1\gamma\rangle J_\gamma - 2\sqrt{2}\langle \tfrac{1}{2}\beta;\tfrac{1}{2}\alpha|00\rangle B. \tag{9.22d}$$

The VCS expansion for the superalgebra is given by

$$\Gamma(J_+) = \nabla, \tag{9.23a}$$

$$\Gamma(F_{+1/2}^{[\pm 1/2]}) = -\partial_\mp - \theta_\pm \nabla, \tag{9.23b}$$

$$\Gamma(B) = B^{(\Lambda^0)} + \tfrac{1}{2}\theta_+\partial_+ - \tfrac{1}{2}\theta_-\partial_-, \tag{9.23c}$$

$$\Gamma(J_0) = J_0^{(\Lambda^0)} - \tfrac{1}{2}\theta_+\partial_+ - \tfrac{1}{2}\theta_-\partial_- - z\nabla, \tag{9.23d}$$

$$\Gamma(F_{-1/2}^{[\pm 1/2]}) = \theta_\pm(2J_0^{(\Lambda^0)} \pm 2B^{(\Lambda^0)} - z\nabla - \theta_\mp\partial_\mp) - z\partial_\mp, \tag{9.23e}$$

$$\Gamma(J_-) = z(2J_0^{(\Lambda^0)} - \theta_+\partial_+ - \theta_-\partial_- - z\nabla) + 2\theta_-\theta_+B^{(\Lambda^0)}. \tag{9.23f}$$

The substitutions

$$\theta_+ \leftrightarrow \theta_-, \quad B^{(\Lambda^0)} \to -B^{(\Lambda^0)}, \tag{9.24a}$$

clearly reflect the superalgebra automorphism

$$F_{\pm 1/2}^{[+1/2]} \leftrightarrow F_{\pm 1/2}^{[-1/2]}, \quad B \to -B, \quad J_\alpha \to J_\alpha. \tag{9.24b}$$

The orthonormal BG basis of $m_j = j$ states is defined by

$$|^{[b^0],[b^0]}_{j^0;j^0}\rangle = |^{[b^0]}_{j^0}\rangle, \tag{9.25a}$$

$$|^{[b^0],[b^0+\frac{1}{2}]}_{j^0;j^0-\frac{1}{2}}\rangle = \theta_+|^{[b^0]}_{j^0}\rangle, \tag{9.25b}$$

$$|^{[b^0],[b^0-\frac{1}{2}]}_{j^0;j^0-\frac{1}{2}}\rangle = \theta_-|^{[b^0]}_{j^0}\rangle, \tag{9.25c}$$

$$|^{[b^0],[b^0]}_{j^0;j^0-1}\rangle = \theta_-\theta_+|^{[b^0]}_{j^0}\rangle. \tag{9.25d}$$

From (6.1) and (6.2), we find the following values for the coefficients of expansion of the VCS basis on the VG basis:

$$\mathcal{O}(^{[b^0],[b^0]}_{j^0;j^0}) = 1, \tag{9.26a}$$

$$\mathcal{O}(^{[b^0],[b^0+\frac{1}{2}]}_{j^0;j^0-\frac{1}{2}}) = (\Lambda^0 + \rho, -\epsilon_1 + \delta_1) = 2(j^0 + b^0), \tag{9.26b}$$

$$\mathcal{O}(^{[b^0],[b^0-\frac{1}{2}]}_{j^0;j^0-\frac{1}{2}}) = (\Lambda^0 + \rho, \epsilon_1 + \delta_1) = 2(j^0 - b^0), \tag{9.26c}$$

$$\mathcal{O}(^{[b^0],[b^0]}_{j^0;j^0-1}) = 2(2j^0 - 1)((b^0)^2 - (j^0)^2)/j^0 \tag{9.26d}$$

(note that $\rho = 0$ here). Vanishing of the coefficients (9.26b) or (9.26c) gives us the atypicality conditions for osp(2/2). Note that Eqs. (9.26) also provide us the branching rule (3.9) for osp(2/2), e.g., it is clear that when $j^0 = b^0$ the states (9.25c) and (9.25d) do not belong to the representation.

We find, in the ordered basis (9.25), the following $\mathfrak{g}_{\bar{0}}$-reduced Γ-matrix representation:

$$\langle^{[b^0],[b']}_{j^0;j'}|||\Gamma(F^{[+1/2]})|||^{[b^0],[b]}_{j^0;j}\rangle = \begin{pmatrix} 0 & 0 & -1 & 0 \\ +2(j^0+b^0)[(2j^0+1)/2j^0]^{1/2} & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2(j^0+b^0)[(2j^0-1)/2j^0]^{1/2} & 0 \end{pmatrix}, \tag{9.27a}$$

$$\langle {}^{[b^0],[b']}_{j^0;j}|||\Gamma(F^{[-1/2]})|||{}^{[b^0],[b]}_{j^0;j}\rangle = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2(b^0-j^0)[(2j^0+1)/2j^0]^{1/2} & 0 & 0 & 1 \\ 0 & -2(b^0-j^0)[(2j^0-1)/2j^0]^{1/2} & 0 & 0 \end{pmatrix}. \tag{9.27b}$$

We verify that it obeys the reduced-commutator algebra

$$\sum_{b'',j''} U\left(j,\frac{1}{2},j,\frac{1}{2};j''0\right)\langle {}^{[b^0],[b']}_{j^0;j}|||F^{[+1/2]}|||{}^{[b^0],[b'']}_{j^0;j''}\rangle\langle {}^{[b^0],[b'']}_{j^0;j''}|||F^{[-1/2]}|||{}^{[b^0],[b]}_{j^0;j}\rangle$$
$$-\sum_{b'',j''} U\left(j,\frac{1}{2},j,\frac{1}{2};j''0\right)\langle {}^{[b^0],[b']}_{j^0;j}|||F^{[-1/2]}|||{}^{[b^0],[b'']}_{j^0;j''}\rangle\langle {}^{[b^0],[b'']}_{j^0;j''}|||F^{[+1/2]}|||{}^{[b^0],[b]}_{j^0;j}\rangle = -\delta_{bb}\cdot\delta_{jj}\,2\sqrt{2b}, \tag{9.28a}$$

$$\sum_{b'',j''} U\left(j,\frac{1}{2},j,\frac{1}{2};j''1\right)\langle {}^{[b^0],[b']}_{j^0;j}|||F^{[+1/2]}|||{}^{[b^0],[b'']}_{j^0;j''}\rangle\langle {}^{[b^0],[b'']}_{j^0;j''}|||F^{[-1/2]}|||{}^{[b^0],[b]}_{j^0;j}\rangle$$
$$+\sum_{b'',j''} U\left(j,\frac{1}{2},j,\frac{1}{2};j''1\right)\langle {}^{[b^0],[b']}_{j^0;j}|||F^{[-1/2]}|||{}^{[b^0],[b'']}_{j^0;j''}\rangle\langle {}^{[b^0],[b'']}_{j^0;j''}|||F^{[+1/2]}|||{}^{[b^0],[b]}_{j^0;j}\rangle = -\delta_{bb}\cdot\delta_{jj}\,2\sqrt{2j(j+1)}, \tag{9.28b}$$

easily deduced from Eq. (9.22d).

The matrix representation (9.27) reproduces very concisely the analysis of Scheunert *et al.*[17] For example, we see that, for $b^0 = j^0$, the (extended) $\bar{\Gamma}$-matrix representation is of the form

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

i.e., it is reducible but not fully reducible with A the left upper 2×2 matrix. Similar conclusions can be drawn for the case $b^0 = -j^0$ to within a reordering of the basis (9.25).

Two star and two grade-star adjoint operations can be defined on osp(2/2) (see Sec. VII) upon restriction of its Lie subalgebra to the compact real form so(2) ⊕ sp(2): from (7.5) and (9.5), we find

$$(F_\alpha^{[+1/2]})^\dagger = \pm(-1)^{1/2-\alpha}F_{-\alpha}^{[-1/2]}, \quad (F_\alpha^{[-1/2]})^\dagger = \mp(-1)^{1/2-\alpha}F_{-\alpha}^{[+1/2]}, \tag{9.29a}$$

for the star adjoint operation and, from (7.4) and (9.5),

$$(F_\alpha^{[+1/2]})^\ddagger = \mp(-1)^{1/2-\alpha}F_{-\alpha}^{[-1/2]}, \quad (F_\alpha^{[-1/2]})^\ddagger = \mp(-1)^{1/2-\alpha}F_{-\alpha}^{[+1/2]}, \tag{9.29b}$$

for the grade-star adjoint operation.

We first seek to identify the star equivalent representations. From (7.14b), we find the following values for the restrictions $\mathscr{K}^2$:

$$\mathscr{K}^2({}^{[b^0],[b^0]}_{j^0;j^0}) = 1, \tag{9.30a}$$

$$\mathscr{K}^2\left({}^{[b^0],[b^0+\frac{1}{2}]}_{j^0;j^0-\frac{1}{2}}\right) = \pm 2(b^0+j^0), \tag{9.30b}$$

$$\mathscr{K}^2\left({}^{[b^0],[b^0-\frac{1}{2}]}_{j^0;j^0-\frac{1}{2}}\right) = \pm 2(b^0-j^0), \tag{9.30c}$$

$$\mathscr{K}^2\left({}^{[b^0],[b^0]}_{j^0;j^0-1}\right) = 2(2j^0-1)((b^0)^2-(j^0)^2)/j^0. \tag{9.30d}$$

Positive definitiveness of $\mathscr{K}^2$ forces us to conclude that we have a star representation whenever $\pm b^0 \geqslant j^0$. In the ordered basis (9.25), we then find, for the star equivalent $\mathfrak{g}_{\bar{0}}$-reduced γ-matrix representation,

$$\langle {}^{[b^0],[b']}_{j^0;j}|||\gamma(F^{[+1/2]})|||{}^{[b^0],[b]}_{j^0;j}\rangle$$
$$= \begin{pmatrix} 0 & 0 & -[\pm 2(b^0-j^0)]^{1/2} & 0 \\ \pm[\pm(b^0+j^0)(2j^0+1)/j^0]^{1/2} & 0 & 0 & -[\pm(b^0-j^0)(2j^0-1)/j^0]^{1/2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mp[\pm 2(b^0+j^0)]^{1/2} & 0 \end{pmatrix}, \tag{9.31a}$$

$$\langle {}^{[b^0],[b']}_{j^0;j}|||\gamma(F^{[-1/2]})|||{}^{[b^0],[b]}_{j^0;j}\rangle$$
$$= \begin{pmatrix} 0 & -[\pm 2(b^0+j^0)]^{1/2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mp[\pm(b^0-j^0)(2j^0+1)/j^0]^{1/2} & 0 & 0 & [\pm(b^0+j^0)(2j^0-1)/j^0]^{1/2} \\ 0 & \mp[\pm 2(b^0-j^0)]^{1/2} & 0 & 0 \end{pmatrix}, \tag{9.31b}$$

which verifies (9.28). These $\mathfrak{g}_{\bar{0}}$-reduced γ-matrix elements (9.31) obey the star adjoint condition

$$\langle {}^{[b^0],[b']}_{j^0;j}|||\gamma(F^{[+1/2]})|||{}^{[b^0],[b]}_{j^0;j}\rangle = \pm(-1)^{j+1/2-j'}[\dim(j)/\dim(j')]^{1/2}\langle {}^{[b^0],[b]}_{j^0;j}|||\gamma(F^{[-1/2]})|||{}^{[b^0],[b']}_{j^0;j}\rangle, \tag{9.32}$$

easily derived from (9.29a).

From (7.14c), we find, for the grade-star adjoint operation, the values for restrictions $\mathcal{K}^2$ of the K operator;

$$\mathcal{K}^2(\begin{smallmatrix}[b^0];[b^0]\\ j^0;j^0\end{smallmatrix}) = 1, \tag{9.33a}$$

$$\mathcal{K}^2(\begin{smallmatrix}[b^0];[b^0+\frac{1}{2}]\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}) = \pm(-1)^{\xi(j^0)}2(b^0+j^0), \tag{9.33b}$$

$$\mathcal{K}^2(\begin{smallmatrix}[b^0];[b^0-\frac{1}{2}]\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}) = \pm(-1)^{\xi(j^0)}2(j^0-b^0), \tag{9.33c}$$

$$\mathcal{K}^2(\begin{smallmatrix}[b^0];[b^0]\\ j^0;j^0-1\end{smallmatrix}) = -2(2j^0-1)((b^0)^2-(j^0)^2)/j^0, \tag{9.33d}$$

from which we conclude that a representation of osp(2/2) is grade-star equivalent if and only if $\pm(-1)^{\xi(j^0)} = +1, j^0 = \frac{1}{2}$, and $-\frac{1}{2} < b^0 < \frac{1}{2}$. The state (9.25d) then has null VCS norm and one must truncate (9.27) to its left upper 3×3 matrix. We then find for the grade-star $\mathbf{g}_{\bar{0}}$-reduced γ-matrix equivalent representation

$$\langle\begin{smallmatrix}[b^0];[b']\\ j^0;j\end{smallmatrix}|||\gamma(F^{[+1/2]})|||\begin{smallmatrix}[b^0];[b]\\ j^0;j\end{smallmatrix}\rangle = \begin{pmatrix} 0 & 0 & -[\pm(-1)^{\xi(j^0)}(1-2b^0)]^{1/2} \\ \pm(-1)^{\xi(j^0)}[\pm(-1)^{\xi(j^0)}2(1+2b^0)]^{1/2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{9.34a}$$

$$\langle\begin{smallmatrix}[b^0];[b']\\ j^0;j\end{smallmatrix}|||\gamma(F^{[-1/2]})|||\begin{smallmatrix}[b^0];[b]\\ j^0;j\end{smallmatrix}\rangle = \begin{pmatrix} 0 & -[\pm(-1)^{\xi(j^0)}(1+2b^0)]^{1/2} & 0 \\ 0 & 0 & 0 \\ \pm(-1)^{\xi(j^0)}[\pm(-1)^{\xi(j^0)}2(1-2b^0)]^{1/2} & 0 & 0 \end{pmatrix}, \tag{9.34b}$$

which verifies the reduced-commutator algebra (9.28). These $\mathbf{g}_{\bar{0}}$-reduced γ-matrix elements obey the grade-star adjoint condition

$$\langle\begin{smallmatrix}[b^0];[b']\\ j^0;j\end{smallmatrix}|||\gamma(F^{[+1/2]})|||\begin{smallmatrix}[b^0];[b']\\ j^0;j\end{smallmatrix}\rangle = \pm(-1)^{\xi(j)}(-1)^{j+1/2-j'}[\dim(j)/\dim(j')]^{1/2}\langle\begin{smallmatrix}[b^0];[b]\\ j^0;j'\end{smallmatrix}|||\gamma(F^{[-1/2]})|||\begin{smallmatrix}[b^0];[b']\\ j^0;j\end{smallmatrix}\rangle, \tag{9.35}$$

easily derived from (9.29b).

D. The Lie superalgebra osp(3/2)

In this section we examined the finite-dimensional representations of the compact Lie superalgebra osp(3/2) with highest weights

$$\Lambda^0 = l^0\epsilon_1 + 2j^0\delta_1. \tag{9.36}$$

From Sec. II, we find the following commutation relations in angular momentum coupled form:

$$\langle J_\alpha, J_\beta\rangle = \sqrt{2}\langle 1\beta;1\alpha|1\gamma\rangle J_\gamma, \tag{9.37a}$$

$$\langle J_\alpha, F^{[1]m}_{(1/2)\mu}\rangle = \sqrt{\tfrac{3}{4}}\langle\tfrac{1}{2}\mu;1\alpha|\tfrac{1}{2}\nu\rangle F^{[1]m}_{(1/2)\nu}, \tag{9.37b}$$

$$\langle L_m, F^{[1]m'}_{(1/2)\mu}\rangle = \sqrt{2}\langle 1m';1m|1m''\rangle F^{[1]m''}_{(1/2)\mu}, \tag{9.37c}$$

$$\langle F^{[1]m}_{(1/2)\mu}, F^{[1]m'}_{(1/2)\nu}\rangle = -2\sqrt{6}\langle 1m';1m|00\rangle\langle\tfrac{1}{2}\nu;\tfrac{1}{2}\mu|1\alpha\rangle J_\alpha + 2\langle 1m';1m|1m''\rangle\langle\tfrac{1}{2}\nu;\tfrac{1}{2}\mu|00\rangle L_{m''}. \tag{9.37d}$$

In angular momentum coupled notation, the VCS expansion (4.20) simplifies to

$$\Gamma(J_+) = \nabla \tag{9.38a}$$

$$\Gamma(F^{[1]m}_{+(1/2)}) = -\partial_{\bar{m}} - \theta_m\nabla, \quad \partial_{\bar{m}} = (-1)^{1-m}\frac{\partial}{\partial\theta_m}, \tag{9.38b}$$

$$\Gamma(L_m) = L^{(\Lambda^0)}_m - \sqrt{2}[\theta\times\partial]^{1m} \equiv L^{(\Lambda^0)}_m + L^{(\theta)}_m, \tag{9.38c}$$

$$\Gamma(J_0) = J^{(\Lambda^0)}_0 - \tfrac{1}{2}\hat{N}^\theta - \hat{N}^z, \quad \hat{N}^\theta = \sum_m \theta_m\partial_m, \quad \hat{N}_z = z\nabla, \tag{9.38d}$$

$$\Gamma(F^{[1]m}_{-(1/2)}) = \theta_m(2J^{(\Lambda^0)}_0 - \tfrac{1}{2}\hat{N}_\theta - \hat{N}_z) - z\partial_{\bar{m}} - \sqrt{2}[\theta\times L^{(\Lambda^0)}]^{1m} + [\theta\times[\theta\times\partial]^1]^{1m}, \tag{9.38e}$$

$$\Gamma(J_-) = z(2J^{(\Lambda^0)}_0 - \hat{N}_\theta - \hat{N}_z) - \sqrt{\tfrac{3}{2}}[[\theta\times\theta]^1\times L^{(\Lambda^0)}]^0, \tag{9.38f}$$

where the square brackets represent so(3) couplings, e.g.,

$$[\theta\times\theta]_{1m} = \sum_{m_1,m_2}\langle 1m_1;1m_2|1m\rangle\theta_{m_2}\theta_{m_1}.$$

The $(2^3 = 8)$ fully antisymmetric orthonormal polynomials $\Theta^{[l_\theta]m_\theta}_{-n_\theta/2}(\theta)$, which can be constructed in terms of the Grassmann variables $\{\theta_{-1},\theta_0,\theta_1\}$, are defined by

$$\begin{aligned}
&\Theta^{[0]0}_0(\theta) = 1, \quad \dim = 1,\\
&\Theta^{[1]m}_{-1/2}(\theta) = \theta_m, \quad \dim = 3,\\
&\Theta^{[1]m}_{-1}(\theta) = -(1/\sqrt{2})[\theta\times\theta]_{1m}, \quad \dim = 3,\\
&\Theta^{[0]0}_{-3/2}(\theta) = -(1/\sqrt{6})[\theta\times[\theta\times\theta]^1]^0, \quad \dim = 1.
\end{aligned} \tag{9.39}$$

The orthonormal BG basis of $m_j = j$ states (with $j = j^0 - n_\theta/2$) is defined by

$$|\begin{smallmatrix}[l^0];[l^0]m\\ j^0;j^0\end{smallmatrix}\rangle = |\begin{smallmatrix}[l^0]m\\ j^0\end{smallmatrix}\rangle, \tag{9.40a}$$

$$|\begin{smallmatrix}[l^0];[l]m\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}\rangle = [\Theta^{[1]}_{-1/2}(\theta)\times|\begin{smallmatrix}[l^0]\\ j^0\end{smallmatrix}\rangle]^{[l]m}, \tag{9.40b}$$

$$|\begin{smallmatrix}[l^0];[l]m\\ j^0;j-1\end{smallmatrix}\rangle = [\Theta^{[1]}_{-1}(\theta)\times|\begin{smallmatrix}[l^0]\\ j^0\end{smallmatrix}\rangle]^{[l]m}, \tag{9.40c}$$

$$|\begin{smallmatrix}[l^0];[l^0]m\\ j^0;j-\frac{3}{2}\end{smallmatrix}\rangle = \Theta^{[0]0}_{-3/2}(\theta)|\begin{smallmatrix}[l^0]m\\ j^0\end{smallmatrix}\rangle. \tag{9.40d}$$

One notes the emergence of the (multiplicity-free) $j[l]$ weight diagram[15,17]

$j\backslash[l]$	$[l^0-1]$	$[l^0]$	$[l^0+1]$	
j^0		●		$(n_\theta=0)$
$j^0-\frac{1}{2}$	●	●	●	$(n_\theta=1)$
j^0-1	●	●	●	$(n_\theta=2)$
$j^0-\frac{3}{2}$		●		$(n_\theta=3)$

(9.41)

from the L couplings in (9.40). Of course, angular momentum forbidden coupled states ($l<0$) should be ignored. Also, states with $j\leqslant 0$ will be truncated by the VCS expansion itself in accordance with the conditions (3.5) [see Eqs. (9.42) below].

From (6.1) and (6.2), we find the following values for the coefficients of expansion of the VCS basis on the VG basis:

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0]\\ j^0;j^0\end{smallmatrix}\right)=1, \tag{9.42a}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0+1]\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}\right)=(\Lambda^0+\rho,-\epsilon_1+\delta_1)=(2j^0+l^0), \tag{9.42b}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0]\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}\right)=(\Lambda^0+\rho,\delta_1)-\tfrac{1}{2}=(2j^0-1), \tag{9.42c}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0-1]\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}\right)=(\Lambda^0+\rho,\epsilon_1+\delta_1)=(2j^0-l^0-1), \tag{9.42d}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0+1]\\ j^0;j^0-1\end{smallmatrix}\right)=(2j^0-1)(2j^0+l^0), \tag{9.42e}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0]\\ j^0;j^0-1\end{smallmatrix}\right)=(2j^0-l^0-1)(2j^0+l^0)(2j^0-1)/(2j^0), \tag{9.42f}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0-1]\\ j^0;j^0-1\end{smallmatrix}\right)=(2j^0-1)(2j^0-l^0-1), \tag{9.42g}$$

$$\mathcal{O}\left(\begin{smallmatrix}[l^0];[l^0]\\ j^0;j^0-\frac{1}{2}\end{smallmatrix}\right)=(2j^0-2)(2j^0-l^0-1)(2j^0+l^0). \tag{9.42h}$$

Equations (9.42) give explicitly the branching rule $\mathrm{osp}(3/2)\downarrow\mathrm{so}(3)\oplus\mathrm{sp}(2)$: the vanishing of any of the above coefficients $\mathcal{O}$ indicates that the corresponding $[l]:j$ multiplet of $\mathrm{so}(3)\oplus\mathrm{sp}(2)$ does not belong to the irreducible representation.

The extended $\overline{\Gamma}$-matrix representation of osp(3/2) in the ordered basis

$j\backslash[l]$	$[l^0-1]$	$[l^0]$	$[l^0+1]$	
j^0		1		$(n_\theta=0)$
$j^0-\frac{1}{2}$	5	2	3	$(n_\theta=1)$
j^0-1	6	7	4	$(n_\theta=2)$
$j^0-\frac{3}{2}$		8		$(n_\theta=3)$

(9.43)

is given by the matrix

TABLE I. $\mathfrak{g}_{\bar{0}}$-reduced Γ submatrices for osp(3/2) [see Sec. IX D, Eq. (9.44)].

$$A=\begin{pmatrix} 0 & 1 & -\left[\frac{2l^0+3}{2l^0+1}\right]^{1/2} & 0 \\ (2j^0-1)\left[\frac{2j^0+1}{2j^0}\right]^{1/2} & 0 & 0 & -\left[\frac{l^0(2l^0+3)}{(l^0+1)(2l^0+1)}\right]^{1/2} \\ (2j^0+l^0)\left[\frac{2j^0+1}{2j^0}\right]^{1/2} & 0 & 0 & -\left[\frac{l^0+2}{l^0+1}\right]^{1/2} \\ 0 & (2j^0+l^0)\left[\frac{2l^0j^0}{(l^0+1)(2j^0-1)}\right]^{1/2} & \left[\frac{2j^0(2j^0-1)(l^0+2)}{(l^0+1)}\right]^{1/2} & 0 \end{pmatrix}$$

$$B=\begin{pmatrix} \left[\frac{2l^0-1}{2l^0-1}\right]^{1/2} & 0 & 0 & 0 \\ 0 & \left[\frac{(l^0+1)(2l^0-1)}{l^0(2l^0+1)}\right]^{1/2} & -\frac{1}{[l^0(l^0+1)]^{1/2}} & 0 \\ 0 & 0 & \left[\frac{l^0}{(l^0+1)}\right]^{1/2} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

$$C=(2j^0-l^0-1)\begin{pmatrix} \left[\frac{2j^0+1}{2j^0}\right]^{1/2} & 0 & 0 & 0 \\ 0 & -\left[\frac{2j^0(l^0-1)}{l^0(2j^0-1)}\right]^{1/2} & 0 & 0 \\ 0 & \frac{-(2j^0+l^0)}{(2j^0(2j^0-1)l^0(l^0+1)^{1/2}} & -\left[\frac{(2j^0-1)l^0(2l^0+3)}{2j^0(l^0+1)(2l^0+1)}\right]^{1/2} & 0 \\ 0 & 0 & 0 & \left[\frac{(2j^0-2)(2l^0+3)}{(2j^0-1)(2l^0+1)}\right]^{1/2} \end{pmatrix}$$

$$D=\begin{pmatrix} 0 & -\left[\frac{(l^0-1)}{l^0}\right]^{1/2} & -\left[\frac{(l^0+1)}{l^0}\right]^{1/2} & 0 \\ \left[\frac{2j^0(2j^0-1)(l^0-1)}{l^0}\right]^{1/2} & 0 & 0 & -1 \\ (2j^0+l^0)\left[\frac{(2j^0-1)(l^0+1)(2l^0-1)}{2l^0j^0(2l^0+1)}\right]^{1/2} & 0 & 0 & -1 \\ 0 & (2j^0+l^0)\left[\frac{(2j^0-2)(2l^0-1)}{(2j^0-1)(2l^0+1)}\right]^{1/2} & -\frac{4j^0(j^0-1)}{(2j^0-1)}\left[\frac{2j^0-1}{2j^0-2}\right]^{1/2} & 0 \end{pmatrix}$$

$$\langle {}^{[l^0];[l']}_{j^0;j'} | | | \bar{\Gamma}(F^{[1]}) | | | {}^{[l^0];[l]}_{j^0;j} \rangle = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{9.44}$$

where A, B, C, and D are 4×4 matrices given in Table I. Under the assumption that $2j^0 \geqslant 0$ and $l^0 \geqslant 0$, the $([l^0]j^0)$ representation of osp(3/2) is found to be atypical only if (9.42d) vanishes. The (extended) $\bar{\Gamma}$-matrix representation (9.44) is then reducible but not fully reducible as the lower left submatrix C vanishes.

One easily verified that the $\mathfrak{g}_{\bar{0}}$-reduced $\bar{\Gamma}$-matrix representation obeys the reduced commutator algebra

$$\sum_{j''j''} U(l\,1l'1;l''0)\,U\left(j\,\frac{1}{2}\,j'\,\frac{1}{2};j''1\right)\langle {}^{[l^0];[l']}_{j^0;j} | | | F^{[1]} | | | {}^{[l^0];[l'']}_{j^0;j''} \rangle \langle {}^{[l^0];[l'']}_{j^0;j''} | | | F^{[1]} | | | {}^{[l^0];[l]}_{j^0;j} \rangle$$
$$= -\delta_{ll'}\delta_{jj'}\sqrt{6j(j+1)}, \tag{9.45a}$$

and

$$\sum_{j''j''} U(l\,1l'1;l''1)\,U\left(j\,\frac{1}{2}\,j'\,\frac{1}{2};j''0\right)\langle {}^{[l^0];[l']}_{j^0;j} | | | F^{[1]} | | | {}^{[l^0];[l'']}_{j^0;j''} \rangle \langle {}^{[l^0];[l'']}_{j^0;j''} | | | F^{[1]} | | | {}^{[l^0];[l]}_{j^0;j} \rangle$$
$$= \delta_{ll'}\delta_{jj'}\sqrt{l(l+1)}, \tag{9.45b}$$

easily obtained from (9.37d).

Only two grade-star adjoint operations can be defined on the superalgebra whenever one restricts one's attention to the real compact Lie subalgebra $so(3) \oplus sp(2)$ of osp(3/2): from (7.4) and (9.5), we find

$$(F_\alpha^{[1]m})^\ddagger = \mp(-1)^{1/2-\alpha}(-1)^{1-m}F_{-\alpha}^{[1]-m}. \tag{9.46}$$

We find the following values for the various restrictions $\mathscr{K}^2$:

$$\mathscr{K}^2({}^{[l^0];[l^0]}_{j^0;j^0}) = 1, \tag{9.47a}$$
$$\mathscr{K}^2({}^{[l^0];[l^0+1]}_{j^0;j^0-\frac{1}{2}}) = \pm(-1)^{\xi(j^0)}(2j^0+l^0), \tag{9.47b}$$
$$\mathscr{K}^2({}^{[l^0];[l^0]}_{j^0;j^0-\frac{1}{2}}) = \pm(-1)^{\xi(j^0)}(2j^0-1), \tag{9.47c}$$
$$\mathscr{K}^2({}^{[l^0];[l^0-1]}_{j^0;j^0-\frac{1}{2}}) = \pm(-1)^{\xi(j^0)}(2j^0-l^0-1), \tag{9.47d}$$
$$\mathscr{K}^2({}^{[l^0];[l^0+1]}_{j^0;j^0-1}) = -(2j^0-1)(2j^0+l^0), \tag{9.47e}$$
$$\mathscr{K}^2({}^{[l^0];[l^0]}_{j^0;j^0-1}) = -(2j^0-l^0-1)(2j^0+l^0)(2j^0-1)/(2j^0), \tag{9.47f}$$
$$\mathscr{K}^2({}^{[l^0];[l^0-1]}_{j^0;j^0-1}) = -(2j^0-1)(2j^0-l^0-1), \tag{9.47g}$$
$$\mathscr{K}^2({}^{[l^0];[l^0]}_{j^0;j^0-\frac{3}{2}}) = \mp(-1)^{\xi(j^0)}(2j^0-2)(2j^0-l^0-1)(2j^0+l^0). \tag{9.47h}$$

With $j^0 \geqslant 0$ and $l^0 \geqslant 0$, we see that the overall sign of the various restrictions depends of the values of the expression $(2j^0 - l^0 - 1)$ and it is easy to conclude that generic representations of osp(3/2) are not grade-star equivalent. The only two exceptions[15,17] are the cases $([l^0]j^0) = ([0]\frac{1}{2})$ and $([\frac{1}{2}]\frac{1}{2})$ [we admit spinor representations of so(3)]. For these cases, we obtain, in the $\{[l\,]j = [l^0]\frac{1}{2}, [l^0+1]0\}$ ordered basis the 2×2 γ-matrix representation

$$\langle {}^{[l^0];[l']}_{j^0;j} | | | \gamma(F^{[1]}) | | | {}^{[l^0];[l]}_{j^0;j} \rangle$$
$$= \begin{pmatrix} 0 & -[\pm(-1)^{\xi(j^0)}(l^0+1)(2l^0+3)/(2l^0+1)]^{1/2} \\ \pm(-1)^{\xi(j^0)}[\pm(-1)^{\xi(j^0)}2(l^0+1)]^{1/2} & 0 \end{pmatrix} \tag{9.48}$$

with grade adjoint conditions

$$\langle {}^{[l^0];[l']}_{j^0;j'} | | | \gamma(F_2^{[1]}) | | | {}^{[l^0];[l]}_{j^0;j} \rangle = \pm(-1)^{\xi(j)}(-1)^{l+1-l'}(-1)^{j+1/2-j'}\left[\frac{\dim[l\,]}{\dim[l']}\frac{\dim(j)}{\dim(j')}\right]^{1/2}$$
$$\times \langle {}^{[l^0];[l]}_{j^0;j} | | | \Gamma(F_2^{[1]}) | | | {}^{[l^0];[l']}_{j^0;j'} \rangle. \tag{9.49}$$

[1] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); Coherent states, contractions and classical limits of the non-compact symplectic groups," in *Proceedings of the XIII International Colloquium on Group Theoretical Methods in Physics*, edited by W. W. Zachary (World Scientific, Singapore, 1984).

[2] D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).

[3] K. T. Hecht, R. Le Blanc, and D. J. Rowe, J. Phys. A: Math. Gen. **20**, 2241 (1987); R. Le Blanc and D. J. Rowe, *ibid.* **19**, 1111 (1986); D. J. Rowe and J. Carvalho, Phys. Lett. B **175**, 243 (1986).

[4] R. Le Blanc and D. J. Rowe, J. Math. Phys. **29**, 758, 767 (1988); D. J. Rowe, R. Le Blanc, and K. T. Hecht, *ibid.* **29**, 287 (1988).

[5] J. Deenen and C. Quesne, J. Math. Phys. **25**, 1638, 2354 (1984); **26**, 2705

(1985); K. T. Hecht and J. P. Elliott, Nucl. Phys. A **438**, 29 (1985); C. Quesne, J. Math. Phys. **27**, 428,869 (1986); K. T. Hecht, "The Vector Coherent State method and its application to problems of higher symmetry." *Lecture Notes in Physics*, Vol. 290 (Springer, Berlin, 1987).
[6]R. Le Blanc and D. J. Rowe, J. Math. Phys. **30**, 1415 (1989).
[7]R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Gen. **16**, 473 (1983).
[8]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[9]V. G. Kac, Lect. Notes Math. **676**, 597 (1977); Commun. Math. Phys. **53**, 31 (1977); Adv. Math. **26**, 8 (1978).
[10]J. Thierry-Mieg, *in Lecture Notes in Physics*, Vol. 201 (Springer, Berlin, 1984), p. 94.
[11]M. Scheunert, "The theory of Lie superalgebras," *Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979); *Proceedings of a Nato Advanced Study Institute on Supersymmetry*, 20–31 August 1984, Bonn, West Germany, edited by K. Dietz, R. Flume, G. v.Gehlen, and V. Rittenberg (Plenum, New York, 1985), p. 421.
[12]R. Le Blanc and D. J. Rowe, J. Phys. A: Math. Gen. **20**, L681 (1987).
[13]R. C. King, J. Phys. A: Math. Gen. **8**, 429 (1975).
[14]R. Le Blanc and D. J. Rowe, J. Phys. A: Math. Gen. **18**, 1891 (1985); K. T. Hecht, R. Le Blanc, and D. J. Rowe, *ibid.* **20**, 257 (1987); D. J. Rowe, R. Le Blanc, and J. Repka, J. Phys. A: Math. Gen. **22**, L309 (1989).
[15]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[16]Han Qi-shi and Sun Hong-zhou, Comm. Theor. Phys. (Bejing, China) **2**, 1137 (1983).
[17]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[18]J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).

On generalization of the Sinai theorem in the random site problem

E. Ahmed
Department of Mathematics, Faculty of Science, U. A. E. University, Al Ain, United Arab Emirates
A. S. Hegazi
Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt

(Received 19 January 1988; accepted for publication 14 June 1989)

The Sinai theorem is generalized to determine the transformations that preserve the critical concentration P_c and the critical exponent β in the random site problem.

The interest in spin glass,[1] polymers, and hopping conduction theory has increased the importance of the random site problem.[2] In this problem sites are distributed randomly. Let ζ_{ij} be some function of the $\mathbf{r}_{ij}$, which is the position vector of the jth site with respect to the ith site. The sites i and j are considered bonded, hence belong to the same cluster, if for some number ζ,

$$\zeta_{ij} \leqslant \zeta. \tag{1}$$

This relation is called the bonding criterion. The percolation threshold ζ_c is the lower bound of ζ that permits the existence of an infinite cluster. The surface defined by

$$\zeta_{ij} \equiv \phi(\mathbf{r}_{ij}) = \zeta,$$

where $\zeta > 0$, ϕ is a homogeneous, increasing, and positive function of the components of $\mathbf{r}_{ij}$, is called the bonding surface. An identical surface Q_ζ is constructed around every site and we increase ζ until $\zeta = \zeta_c$, where an infinite cluster is formed for the first time and the bonding surfaces of two consecutive sites intersect with each other. This is called the overlapping figure construction of the random site problem.

The problem under consideration is what are the transformations that preserve ζ_c? The first answer was obtained in Ref. 2 and is stated in the following theorem.

Theorem 1 (Sinai): If a surface Q'_ζ can be obtained from a surface Q_ζ via a linear transformation of the coordinates

$$x'_k = A_{kl} x_l \tag{2}$$

(we are using the Einstein summation convention), where the constants A_{kl} involve both rotation and dilatation, then ξ_c for the two surfaces are identical.

This result has been generalized Ref. 3 in two dimensions to the following theorem.

Theorem 2: If a curve Q'_ζ can be obtained from the curve Q_ζ via a conformal transformation

$$z' = f(z), \tag{3}$$

where $z' = x' + iy'$, $z = x + iy$, then ζ_c and the critical exponent β are identical for both cases. The critical exponent β is defined via the percolation probability $P(\zeta)$ (the probability that a given site belongs to an infinite cluster) by

$$P(\zeta) \sim (\zeta - \zeta_c)^\beta, \quad \text{as } \zeta \to \zeta_c^+. \tag{4}$$

Now the question is can we generalize Theorem 2 to higher dimension, especially $d = 3$ and 4. This question is related to the existence of conformal mappings and analytic functions in dimension $d > 2$. As for the conformal mappings it is known[4] that conformal algebra exists in any dimension and is given by

$$\begin{aligned}
&[D,P_\mu] = -P_\mu, \quad [D,K_\mu] = K_\mu,\\
&[J_{\mu\nu},P_\lambda] = \eta_{\mu\lambda}P_\nu - \eta_{\nu\lambda}P_\mu,\\
&[J_{\mu\nu},K_\lambda] = \eta_{\mu\lambda}K_\nu - \eta_{\nu\lambda}K_\mu,\\
&[P_\mu,K_\nu] = 2(\eta_{\mu\nu}D - J_{\mu\nu}),\\
&[J_{\mu\nu},J_{\rho\sigma}] = \eta_{\mu\rho}J_{\nu\sigma} - J_{\nu\rho}\eta_{\mu\sigma} - \eta_{\nu\rho}J_{\mu\sigma} + \eta_{\nu\sigma}J_{\mu\rho}.
\end{aligned} \tag{5}$$

Therefore we infer that, at least locally, conformal mapping exists in any dimension.

Analytic fucntions, however, are a different matter. The generalization of the complex variable $z = x + iy$ to four dimensions is the quanternion, which we denote by

$$z \equiv x_0 + i_1 x_1 + i_2 x_2 + i_3 x_3, \tag{6}$$

such that

$$\begin{aligned}
&i_1^2 = i_2^2 = i_3^2 = i_1 i_2 i_3 = -1,\\
&i_\rho i_\sigma = \epsilon_{\rho\sigma\tau} i_\tau, \quad \rho,\sigma,\tau = 1,2,3;
\end{aligned} \tag{7}$$

$\epsilon_{\rho\sigma\tau}$ is a totally skew symmetric symbol, $\epsilon_{123} = 1$. To get $d = 3$ one sets $x_3 = 0$.

Following the concepts of the two-dimensional complex analysis, we introduce the following definitions.

Definition 1: There are three types of conjugations; $z^{(1)}$, $z^{(2)}$, $z^{(3)}$ defined by

$$\begin{aligned}
z^{(1)} &= x_0 - i_1 x_1 + i_2 x_2 + i_3 x_3,\\
z^{(2)} &= x_0 + i_1 x_1 - i_2 x_2 + i_3 x_3,\\
z^{(3)} &= x_0 + i_1 x_1 + i_2 x_2 - i_3 x_3.
\end{aligned} \tag{8}$$

It is clear that they commute with each other.

Definition 2: An analytic function in four dimensions is defined by

$$f(z) = \sum_{n=0}^{\infty} C_n (z)^n, \tag{9}$$

where C_n are constant quaternions.

Now we have the following theorem.

Theorem 3: According to definitions 1 and 2 the only analytic function in three and four dimensions is

$$f(z) = c_0 + c_1 z. \tag{10}$$

Proof: Denoting $\partial / \partial z^{(\rho)}$ by $\partial_{(\rho)}$, $\rho = 1,2,3$, the generalized Cauchy–Riemann conditions necessary for the existence of analytic functions are

$$\partial_{(1)} f(z) = \partial_{(2)} f(z) = \partial_{(3)} f(z) = 0. \tag{11}$$

Assuming

$$f(z) \equiv U_0 + i_1 U_1 + i_2 U_2 + i_3 U_3, \tag{12}$$

Cauchy–Riemann generalized conditions take the form

 0022-2488/90/010037-02$03.00

$$\partial_0 U_0 - \partial_1 U_1 = 0, \quad \partial_0 U_1 + \partial_1 U_0 = 0,$$
$$\partial_0 U_2 - \partial_1 U_3 = 0, \quad \partial_0 U_3 + \partial_1 U_2 = 0, \tag{13}$$
$$\partial_0 U_0 - \partial_2 U_2 = 0, \quad \partial_0 U_1 + \partial_2 U_3 = 0,$$
$$\partial_0 U_2 + \partial_2 U_0 = 0, \quad \partial_0 U_3 - \partial_2 U_1 = 0, \tag{14}$$
$$\partial_0 U_0 - \partial_3 U_3 = 0, \quad \partial_0 U_1 - \partial_3 U_2 = 0,$$
$$\partial_0 U_2 + \partial_3 U_1 = 0, \quad \partial_0 U_3 + \partial_3 U_0 = 0, \tag{15}$$

where

$$\partial_\mu \equiv \partial / \partial x^\mu, \quad \mu = 0,1,2,3.$$

Studying the compatibility conditions of the four dimensional Cauchy–Riemann conditions (13), (14), and (15) it is straightforward, though tedious, to conclude that

$$U_0 = U_0(x_0), \quad U_1 = U_1(x_1),$$
$$U_2 = U_2(x_2), \quad U_3 = U_3(x_3), \tag{16}$$

as well as the condition

$$\partial_\mu^2 U_\mu = 0 \quad (\text{no sum over } \mu). \tag{17}$$

Equation (17) is reminiscent of the relation between analytic functions and harmonic functions in two dimensions, though it is more stringent. The only solution of (16) and (17) that is compatible with (9) is the linear function.

This completes the proof of the theorem.

It is interesting to notice that there is an analogous theorem[5] proved for supermanifolds. It is known that quaternions are closely related to spinors. This also explains why the conformal algebra is infinite only in $d = 2$ since in this case it is related to analytic function.[6]

Therefore, despite losing analytic functions, we still have the conformal transformations. However, recalling that what we really need is the set of transformations that map intersecting surfaces onto intersecting surfaces and vice versa, the requirement of preserving the sense of the angle of intersection is a luxury we can afford to lose. Since 1-1 and onto transformations maps intersecting surfaces onto intersecting surfaces and vice versa, we have the following theorem.

Theorem 4: In any dimension if a surface Q'_ζ is obtained from the surface Q_ζ by a 1-1 and onto transformation then ζ_c and the critical exponent β are identical for both cases.

We give some examples. The transformation

$$x \to ax, \quad y \to y, \quad z \to z \tag{18}$$

maps a sphere onto an ellipse; hence they both have the same ζ_c and β. The transformation

$$\mathbf{r} \to -\mathbf{r} \tag{19}$$

is a reflection through the origin that is not a conformal mapping; however, it preserves ζ_c and β. Finally, one can define a 1-1 mapping from a unit circle onto a square via the argument $\theta(z \equiv ae^{i\theta}$ on the circle); hence according to Theorem 4 both circle and square have identical ζ_c. This result has been confirmed using computer simulation.[7]

[1] D. Mattis *Magnetism* (Springer, Berlin, 1984), Vol. 2.
[2] B. Shklovskii and A. Efros, *Electronic Properties of Doped Semiconductors* (Springer, Berlin, 1984).
[3] E. Ahmed and A. Tawansi, Phys. Lett. A **129**, 151 (1988).
[4] J. van Holten and A. van Proyen, J. Phys. A **15**, 3763 (1982).
[5] S. Matsumoto and K. Kakazu, J. Math. Phys. **27**, 2690 (1986).
[6] D. Friedan, E. Martinec, and S. Shenker, Nucl. Phys. B **271**, 93 (1986).
[7] G. Pike and C. Seager, Phys. Rev. B **10**, 142 (1974).

Wavelet analysis on the circle

Matthias Holschneider
Centre de Physique Théorique,[a] *Centre National de la Recherche Scientifique, Luminy, Case 907, 13288 Marseille Cedex 9, France*

(Received 16 December 1988; accepted for publication 26 July 1989)

The construction of a wavelet analysis over the circle is presented. The spaces of infinitely times differentiable functions, tempered distributions, and square integrable functions over the circle are analyzed by means of the wavelet transform.

I. INTRODUCTION

In this paper we want to show how to analyze fairly arbitrary functions over the circle $\mathbb{T}^1$ with the help of a two-parameter family $g_{\phi,a}$ of functions called wavelets. They are labeled by a position parameter $\phi\in\mathbb{T}^1$ and a scale parameter a, $a>0$. In standard wavelet analysis of functions over the real line $\mathbb{R}$, the family of analyzing wavelets is obtained from a single function by means of dilations and translations (e.g., Ref. 1). On the circle it is difficult to define a good dilation operator, and therefore the wavelets over the circle cannot be obtained by an irreducible representation of the affine group, as was the case in Refs. 1–4.

As proposed before[5] in the case of orthogonal wavelet analysis the wavelets that we will use are obtained from the standard ones $(1/a)g((x-\phi)/a)$ by means of periodization:

$$g_{\phi,a}(x)=\sum_{n\in\mathbb{Z}}\frac{1}{a}g\left(\frac{x-\phi+n}{a}\right),\quad \phi\in\mathbb{T}^1,\quad a\in\mathbb{R}^+. \tag{1.1}$$

This series converges whenever g decays sufficiently fast at infinity. The wavelet transform of a complex-valued function over $\mathbb{T}^1$ is a function over the position–scale space, which is an open, infinite, cylinder $\mathbb{Y}=\mathbb{T}^1\times\mathbb{R}^+$. It is given by the following scalar products:

$$(T_g s)(\phi,a)=(g_{\phi,a},s),\quad (\phi,a)\in\mathbb{Y}. \tag{1.2}$$

The wavelet transform is a sort of mathematical microscope where the position is fixed by the parameter b, the enlargement is $1/a$, and the optic is given by the wavelet itself. We now shall give a precise meaning to all these expressions, and we shall show how to characterize various functional spaces over $\mathbb{T}^1$ with the help of this transform.

II. SOME DEFINITIONS AND EASY PROPERTIES

The space $C^\infty(\mathbb{T}^1)$ is made of complex-valued functions s over $\mathbb{T}^1$ that are arbitrarily many times differentiable. We identify the circle $\mathbb{T}^1$ with the interval $[0,2\pi)$. A topology on $C^\infty(\mathbb{T}^1)$ is given by the following directed family of norms:

$$\|s\|_{C^\infty(\mathbb{T}^1);n}=\sum_{0<p<n}\sup_{\mathbb{T}^1}|\partial^p s|. \tag{2.1}$$

In this topology $C^\infty(\mathbb{T}^1)$ is a Fréchet space, that is, a complete, locally convex, metrizable, linear space. For any function s in $C^\infty(\mathbb{T}^1)$, we can define its Fourier coefficients $(Fs)(n)$, $n\in\mathbb{Z}$:

$$(Fs)(n)=\frac{1}{2\pi}\int_{\mathbb{T}^1}s(x)e^{-inx}\,dx. \tag{2.2}$$

The sequences $(Fs)(n)$ that can appear as the Fourier coefficients of some function s in $C^\infty(\mathbb{T}^1)$ are exactly the sequences that decrease as $|n|$ goes to infinity faster than any power of n. And conversely every such sequence defines a function in $C^\infty(\mathbb{T}^1)$. This sequence space will be called $S(\mathbb{Z})$. A topology on $S(\mathbb{Z})$ is given by the following directed family of norms:

$$\|r\|_{S(\mathbb{Z});n}=\sum_{0<p<n}\sup_{k\in\mathbb{Z}}|k^p r(k)|,\quad n=0,1,\ldots. \tag{2.3}$$

For any sequence r in $S(\mathbb{Z})$ we define the inverse Fourier transform F^{-1}:

$$(F^{-1}r)(x)=\sum_{n\in\mathbb{Z}}r(n)e^{inx}. \tag{2.4}$$

The following well known theorem shows that $C^\infty(\mathbb{T}^1)$ and $S(\mathbb{Z})$ are topologically the same spaces, and that any function in $C^\infty(\mathbb{T}^1)$ can be decomposed into a Fourier series.

Theorem 2.1:
(i) F: $C^\infty(\mathbb{T}^1)\to S(\mathbb{Z})$ is continuous;
(ii) F^{-1}: $S(\mathbb{Z})\to C^\infty(\mathbb{T}^1)$ is continuous,
(iii) F^{-1}: $F=\mathbf{1}_{C^\infty(\mathbb{T}^1)}$, $FF^{-1}=\mathbf{1}_{S(\mathbb{Z})}$.

Proof: [We only shall prove (i) and (ii).] With the help of a partial integration we can write

$$\left|\int_{\mathbb{T}^1}k^p s(x)e^{-ikx}\,dx\right| = \left|\int_{\mathbb{T}^1}\partial_x^p s(x)e^{-ikx}\,dx\right| \leqslant 2\pi\|s\|_{C^\infty(\mathbb{T}^1);p},$$

which proves (i). On the other hand, since any $r\in S(\mathbb{Z})$ is rapidly decreasing, we may exchange the differentiation and the summation:

$$\left|\partial_x^p\sum_{n\in\mathbb{Z}}r(n)e^{inx}\right| = \left|\sum_{n\in\mathbb{Z}}n^p r(n)\frac{1+n^2}{1+n^2}e^{inx}\right|$$

$$\leqslant(\|r\|_{S(\mathbb{Z});p}+\|r\|_{S(\mathbb{Z});p+2})\sum_{n\in\mathbb{Z}}\frac{1}{1+n^2}\leqslant C\|r\|_{S(\mathbb{Z});p+2}.$$

Q.E.D.

[a] Laboratoire Propre LP-7061, Centre Nationale de la Recherche Scientifique.

 0022-2488/90/010039-06$03.00

This proves (ii). Q.E.D.

The functions g that define via (1.1) the wavelets shall all be in the class $S(\mathbb{R})$ of Schwarz, that is, the set of functions decaying at infinity together with all their derivatives faster than any polynomial. A topology on $S(\mathbb{R})$ is given by the following directed family of norms:

$$\|s\|_{S(\mathbb{R});n,\alpha} = \sum_{\substack{0\le p\le n \\ 0\le l\le \alpha}} \sup_{\mathbb{R}} |x^p \partial_x^l s(x)|, \quad n,\alpha = 0,1,\ldots . \tag{2.5}$$

With this topology $S(\mathbb{R})$ is a Fréchet space. On $S(\mathbb{R})$ we define translations and dilations in the usual manner:

$$T^b\colon S(\mathbb{R})\to S(\mathbb{R}), \quad (T^b s)(x) = s(x-b), \quad b\in\mathbb{R}, \tag{2.6}$$

$$D^a\colon S(\mathbb{R})\to S(\mathbb{R}), \quad (D^a s)(x) = (1/a)s(x/a), \quad a>0. \tag{2.7}$$

Obviously these operators are continuous. On $S(\mathbb{R})$ we define the Fourier transform $\mathcal{F}$ and the inverse Fourier transform $\mathcal{F}^{-1}$ as

$$(\mathcal{F}s)(\omega) = \int_{\mathbb{R}} s(x) e^{-i\omega x}\, dx, \tag{2.8}$$

$$(\mathcal{F}^{-1}r)(x) = (2\pi)^{-1} \int_{\mathbb{R}} r(\omega) e^{i\omega x}\, dx. \tag{2.9}$$

We will use the notation $\hat{s}$ for $\mathcal{F}s$. The Fourier transform is a bijective, bicontinuous map.

Theorem 2.2:

(i) $\mathcal{F}\colon S(\mathbb{R})\to S(\mathbb{R})$ is continuous,

(ii) $\mathcal{F}^{-1}\colon S(\mathbb{R})\to S(\mathbb{R})$ is continuous,

(iii) $\mathcal{F}^{-1}\mathcal{F} = \mathcal{F}\mathcal{F}^{-1} = \mathbb{1}_{S(\mathbb{R})}$.

For a proof, see any textbook about functional analysis.

The passage from a function s in $S(\mathbb{R})$ to a function in $C^\infty(\mathbb{T}^1)$ will be done by the periodization operator Π:

$$(\Pi s)(x) = \sum_{n\in\mathbb{Z}} s(x+2\pi n), \quad x\in\mathbb{T}^1. \tag{2.10}$$

Theorem 2.3: $\Pi\colon S(\mathbb{R})\to C^\infty(\mathbb{T}^1)$ is continuous.

We shall prove this theorem in a moment. To any function in $S(\mathbb{R})$ we can associate a sequence in $S(\mathbb{Z})$ with the help of the sampling operator:

$$\Sigma\colon S(\mathbb{R})\to S(\mathbb{Z}), \quad (\Sigma s)(n) = s(n), \quad n = \ldots,-1,0,1,\ldots . \tag{2.11}$$

It obviously is a continuous operator. A natural question is to ask what the Fourier coefficients of periodized function are. The answer is given by the Poisson summation formula, which reads

$$F\Pi = \Sigma\mathcal{F}, \tag{2.12}$$

or, more explicitly ($\hat{s} = \mathcal{F}s$),

$$\sum_{n\in\mathbb{Z}} s(x+2\pi n) = \sum_{n\in\mathbb{Z}} \hat{s}(n) e^{inx}. \tag{2.13}$$

For a proof of this equation, see, e.g., Ref. 6.

Proof of Theorem 2.3: We have $\Pi = F^{-1}\Sigma\mathcal{F}$. All mappings are continuous. Q.E.D.

The space $L^2(\mathbb{T})$ is made of functions s with finite norm

$$\|s\| = \int_{\mathbb{T}^1} |s(\phi)|^2\, d\phi. \tag{2.14}$$

It is a Hilbert space if it is given the following scalar product:

$$(r,s) = \int_{\mathbb{T}^1} \bar{r}(\phi) s(\phi)\, d\phi. \tag{2.15}$$

Clearly $L^2(\mathbb{T}^1)\supset C^\infty(\mathbb{T}^1)$. The Fourier transform extends to a map from $L^2(\mathbb{T})$ to $L^2(\mathbb{Z})$, the Hilbert space of square summable sequences. We may split $L^2(\mathbb{T})$ into the direct sum of $H^2_-(\mathbb{T}^1)$, the space of functions that have only negative frequencies; $H^2_+(\mathbb{T}^1)$, the space of functions that contain only positive frequencies; and $K(\mathbb{T}^1)$, the constant functions:

$$L^2(\mathbb{T}^1) = H^2_+(\mathbb{T}^1)\oplus K(\mathbb{T}^1)\oplus H^2_-(\mathbb{T}^1). \tag{2.16}$$

The corresponding subspaces of $C^\infty(\mathbb{T}^1)$ shall be denoted by $C^\infty_+(\mathbb{T}^1)$ and $C^\infty_-(\mathbb{T}^1)$.

III. THE WAVELET TRANSFORM OF $C^\infty(\mathbb{T}^1)$

In this section we analyze the space functions with the highest possible regularity. It will turn out that this is mirrored in the wavelet transform by a fast decay of the wavelet coefficients as the scale a goes to 0. The regularity of the analyzing wavelet, or, what is the same, the fast decay of the Fourier transform at infinity, in turn gives rise to a fast decay of the wavelet coefficients, as the scale a goes to infinity. Therefore we will be able to characterize this space as the set of functions that are well localized in the scales; that is, every such function has a minimal effective length scale. We shall characterize the range of the transform, and further give an inversion formula.

First we introduce some notations. For any $f\in S(\mathbb{R})$ and any $a>0$, we define $f_a\in C^\infty(\mathbb{T}^1)$ as

$$f_a(x) = (\Pi D^a f)(x) = \sum_{n\in\mathbb{Z}} \frac{1}{a} f\left(\frac{x+2\pi n}{a}\right), \tag{3.1}$$

and $f_{\phi,a}\in C^\infty(\mathbb{T}^1)$ with $\phi\in\mathbb{T}^1$ will stand for

$$f_{\phi,a}(x) = (\Pi T^\phi D^a f)(x) = f_a(x-\phi). \tag{3.2}$$

For reasons that will become clear later on we shall require that all the moments of the wavelet, g, vanish:

$$\int_{\mathbb{R}} x^n g(x)\, dx = 0, \quad n = 0,1,\ldots . \tag{3.3}$$

An equivalent condition is that the Fourier transform $\hat{g} = \mathcal{F}g$ vanishes at the origin in infinite order:

$$\hat{g}(\omega) = O(\omega^n), \quad n = 0,1,\ldots \quad (\omega\to 0). \tag{3.4}$$

The subset of $S(\mathbb{R})$ of functions that satisfy one (and therefore both) of these conditions will be called $S_0(\mathbb{R})$.

Definition 3.1: The wavelet transform T_g of any function $s\in C^\infty(\mathbb{T}^1)$ with respect to a function $g\in S_0(\mathbb{R})$ (called the wavelet) is given by the following scalar products:

$$(T_g s)(\phi,a) = (g_{\phi,a}, s), \quad \phi\in\mathbb{T}^1, \quad a>0. \tag{3.5a}$$

The same expression in Fourier space reads (using the Poisson summation formula)

$$(T_g s)(\phi,a) = \sum_{k\in\mathbb{Z}} (\overline{\mathcal{F}g})(ak) e^{ik\phi} (Fs)(k). \tag{3.5b}$$

The wavelet transform is a function over the position scale space, which in our case is a cylinder $\mathbb{Y} = \mathbb{T}^1\times\mathbb{R}^+$. Obvious-

ly it is a function that is infinitely differentiable. It turns out that T_g is a continuous map from $C^\infty(\mathbb{T}^1)$ into the space of functions $y(\phi,a)$ over $\mathbb{Y}$ that are infinitely differentiable, and that decay for $a\to 0$ and $a\to\infty$ faster than any fractional polynomial in a. We call this space $S(\mathbb{Y})$. The topology of this space is given by a directed family of norms:

$$\|y\|_{S(\mathbb{Y});n,\alpha,\beta} = \sum_{\substack{-n<p<n \\ 0<l<\alpha \\ 0<k<\beta}} \sup_{\mathbb{Y}} |a^p \partial_a^l \partial_p^k y(\rho,a)|,$$

$$n,\alpha,\beta = 0,1,\ldots . \tag{3.6}$$

Then $S(\mathbb{Y})$ is a Fréchet space. We have the following theorem.

Theorem 3.2: For $g\in S_0(\mathbb{R})$, we have T_g: $C^\infty(\mathbb{T}^1)\to S(\mathbb{Y})$ is continuous.

We first shall prove two lemmas.

Lemma 3.3: Let $s\in S_0(\mathbb{R})$, and let s_a be given by (3.1). Then

(i) $\lim_{a\to 0} \|s_a\|_{L^1(\mathbb{T}^1)} = \|s\|_{L^1(\mathbb{R})}$,

(ii) $\forall n = 0,1,\ldots$: $\|s_a\|_{C^\infty(\mathbb{T}^1);n} = O(1/a^m)$ $(a\to\infty)$, for $m = 0,1,\ldots$.

Proof: Assertion (i) follows from the fact that for small a essentially only the term $(n=0)$ in the sum (3.1) remains because of the localization of f. To prove (ii) we expand s_a into a Fourier series using the Poisson summation formula:

$$s_a(x) = \sum_{n\in\mathbb{Z}} \hat{s}(an)e^{inx}.$$

Since $\hat{s}(0) = 0$ and $\hat{s}(\omega) = O(\omega^{p+m+2})$ as ω goes to infinity, we can estimate, for a large enough,

$$|a^m \partial_x^p s_a(x)| \leqslant \sum_{n\neq 0} a^m |n|^p |\hat{s}(a\cdot n)| \leqslant \sum_{n\neq 0} \frac{c}{a^p |n|^{m+2}}.$$

Since the sum remains finite if a goes to infinity we have finished the proof. Q.E.D.

Lemma 3.4: For any $s\in S_0(\mathbb{R})$, there are functions u, v, $w\in S_0(\mathbb{R})$ such that

(i) $\partial_\phi s_{\phi,a}(x) = \partial_x u_{\phi,a}(x)$,

(ii) $\partial_a s_{\phi,a}(x) = \partial_x v_{\phi,a}(x)$,

(iii) $(1/a)s_{\phi,a}(x) = \partial_x w_{\phi,a}(x)$.

Proof: We use again the Poisson summation formula to decompose $s_{\phi,a}$ into a Fourier series ($s = \mathcal{F}s$):

$$s_{\phi,a}(x) = \sum_{n\in\mathbb{Z}} \hat{s}(an)e^{in(x-\phi)}.$$

The following functions are in $S_0(\mathbb{R})$:

(i) $u = -s$,

(ii) $\hat{u}(\omega) = -i\partial_\omega \hat{s}(\omega)$,

(iii) $\hat{w}(\omega) = (i/\omega)\hat{s}(\omega)$.

A direct computation using the Fourier expansion of $s_{\phi,a}$ shows that they satisfy the identities of the lemma.

Proof of Theorem 3.2: First let $p\geqslant 0$. Using Lemma 3.4, for $g\in S_0(\mathbb{R})$ we can find a function $r\in S_0(\mathbb{R})$ such that $\partial_a^l \partial_\phi^k g_{\phi,a}(x) = \partial_x^{l+k} r_{\phi,a}(x)$. With the help of a partial integration we can write

$$|a^p \partial_a^l \partial_\phi^k (T_g s)(\phi,a)|$$

$$= \left| \int_{\mathbb{T}^1} a^p r_{\phi,a}(x) \partial_x^{l+k} s(x)dx \right|$$

$$\leqslant \sup_{a\in\mathbb{R}^+} \|a^p r_a\|_{L^1(\mathbb{T}^1)} \|s\|_{C^\infty(\mathbb{T}^1);l+k}.$$

Lemma 3.3 shows that the sup is finite. Now let $p<0$. Again Lemma 3.4 can be used to find a function $r\in S_0(\mathbb{R})$ such that $a^p\partial_a^l\partial_\phi^k g_{\phi,a}(x) = \partial_x^{|p|+l+k} r_{\phi,a}(x)$. Therefore we can write

$$|a^p \partial_a^l \partial_\phi^k (T_g s)(\phi,a)|$$

$$= \left| \int_{\mathbb{T}^1} r_{\phi,a}(x) \partial_x^{|p|+l+k} s(x)dx \right|$$

$$\leqslant \sup_{a\in\mathbb{R}^+} \|r_a\|_{L^1(\mathbb{T}^1)} \|s\|_{C^\infty(\mathbb{T}^1);|p|+l+k}.$$

Again Lemma 3.3 can be used to conclude. Q.E.D.

Definition 3.5: For any function h in $S(\mathbb{Y})$ the inverse wavelet transform $T_g^{-1}h$ is defined as

$$(T_g^{-1}h)(x) = \int_{\mathbb{Y}} g_{\phi,a}(x)h(\phi,a)\frac{da\,d\phi}{a}.$$

(We shall see in a moment in which sense T_g^{-1} is the inverse of T_g.) This integral is well defined since h is rapidly decreasing as a tends to 0 or infinity. It again turns out to be a continuous map.

Theorem 3.6: For $g\in S_0(\mathbb{R})$, we have T_g^{-1}: $S(\mathbb{Y})\to C^\infty(\mathbb{T}^1)$ is continuous.

Proof: Since h is rapidly decreasing we may exchange the integration and the differentiation and we may write

$$|\partial_x^n (T_g^{-1}h)(x)| = \left| \int_{\mathbb{Y}} \partial_x^n g_{\phi,a}(x)h(\phi,a)\frac{da\,d\phi}{a} \right|$$

$$= \left| \int_{\mathbb{Y}} g_{\phi,a}(x)\frac{1}{a}\partial_x^n h(\phi,a)da\,d\phi \right|$$

$$\leqslant \int_0^\infty \|g_a\|_{L^1(\mathbb{T}^1)}\, da\cdot \|h\|_{S(\mathbb{Y});1,n,0}.$$

Lemma 3.3 assures that the integral is finite. Q.E.D.

We want to establish the relation between T_g and T_g^{-1}. From Definition 3.1 it follows that the positive (negative) frequencies of the wavelet do only interact with the positive (negative) frequencies of the function s. Therefore we will use the splitting (2.16) to separate positive and negative frequencies. Clearly the image of the constant functions is equal to zero and therefore we can only hope to find an inversion formula that holds on the other two parts.

Theorem 3.7: For all $s\in C^\infty_{+(-)}(\mathbb{T}^1)$ we have

$$T_g^{-1}T_g s = c_g^{+(-)}s.$$

The constants c_g^+ and c_g^- are determined by g:

$$c_g^+ = \int_0^\infty \frac{da}{a}|\hat{g}(a)|^2, \quad c_g^- = \int_0^\infty \frac{da}{a}|\hat{g}(-a)|^2.$$

Proof: Let $s\in C^\infty_{+(-)}(\mathbb{T}^1)$ be given. Since all negative (positive) frequencies of s vanish, we may suppose that the Fourier transform of g is symmetric. We call s^ε the function that is obtained when using the inverse transformation with a cutoff at the small scales:

$$s^\varepsilon(x) = \int_\varepsilon^\infty \frac{da}{a} \int_{\mathbb{T}^1} d\phi\, g_{\phi,a}(x)(T_g s)(\phi,a)$$
$$= \int_\varepsilon^\infty \frac{da}{a} \int_{\mathbb{T}^1\times\mathbb{T}^1} d\phi\, dy\, g_{\phi,a}(x)\bar{g}_{\phi,a}(y)s(y).$$

For fixed $\varepsilon > 0$, Lemma 3.3 guarantees the absolute convergence of all integrals, and therefore we could exchange the integration. Integrating first over ϕ and a yields

$$s^\varepsilon(x) = \int_{\mathbb{T}^1} K^\varepsilon(x-y)s(y)dy = (K^\varepsilon * s)(x),$$

and the kernel K^ε is given by

$$K^\varepsilon(u) = \int_\varepsilon^\infty \frac{da}{a} \int_{\mathbb{T}^1} d\phi\, g_a(u-\phi)\bar{g}_a(\phi).$$

Expanding these expressions into a Fourier series we obtain the following equations for the Fourier coefficients (denoted by a tilde):

$$\tilde{K}^\varepsilon(n) = \int_\varepsilon^\infty \frac{da}{a} |\hat{g}(an)|^2 = \int_{|n|\varepsilon}^\infty \frac{da}{a} |\hat{g}(a)|^2 = \hat{K}(\varepsilon n),$$

where we have posed

$$\hat{K}(\omega) = \int_{|\omega|}^\infty \frac{da}{a} |\hat{g}(a)|^2.$$

Here we have used the fact that $\hat{g}$ was chosen symmetric. The Fourier coefficients depend only on εn. Applying the Poisson summation formula (2.12), we obtain

$$K^\varepsilon = \Pi D_\varepsilon K = K_\varepsilon, \quad \text{with } K = \mathcal{F}^{-1}\hat{K}.$$

But $[c_g^{+(-)}]^{-1} K_\varepsilon$ is a summability kernel (e.g., Ref. 6). Therefore we can apply the theorem of the approximation of the identity[6] to conclude. Q.E.D.

We now want to characterize the range of the wavelet transform.

Theorem 3.8: For any $g \in S_0(\mathbb{R})$, the image of $C^\infty(\mathbb{T}^1)$ under T_g is a closed subspace of $S(\mathbb{Y})$. The range of T_g restricted to $C^\infty_{+(-)}(\mathbb{T}^1)$ consists of exactly those functions y in $S(\mathbb{Y})$ that satisfy the following "reproducing kernel" equation:

$$y(\phi,a) = \int_{\mathbb{Y}} p_g(\phi,a;\phi',a')y(\phi',a') \frac{da'\,d\phi'}{a},$$

where the reproducing kernel p_g is given by

$$p_g(\phi,a;\phi',a') = (1/c_g^{+(-)})(g_{\phi,a}, g_{\phi',a'}),$$

or, in Fourier space,

$$p_g(\phi,a;\phi',a') = \frac{1}{c_g^{+(-)}} \sum_{k\in\mathbb{Z},k>(<)0} (\overline{\mathcal{F}g})(ak)(\mathcal{F}g)(a'k)e^{ik(\phi-\phi')}.$$

Proof: As continuous preimage of a closed space under the inverse wavelet transform, the image of the wavelet transform is closed. Now for any $y \in S(\mathbb{Y})$ in the range of T_g on $C^\infty_{+(-)}(\mathbb{T}^1)$ there is a function s in $C^\infty_{+(-)}(\mathbb{T}^1)$ such that $T_g s = y$. From Theorem 3.7 we have $T_g^{-1}T_g s = c_g^{+(-)}s$, and therefore we can write

$$T_g T_g^{-1} y = T_g T_g^{-1} T_g s = c_g^{+(-)} T_g s = c_g^{+(-)} y.$$

On the other hand, if $y \in S(\mathbb{Y})$ satisfies $T_g T_g^{-1} y = c_g^{+(-)} y$, then, for $x = c_g^{+(-)} T_g^{-1} y$, we have $T_g x = y$, which proves that y is in the range of T_g. We now rewrite the above identity more explicitly (we can exchange all integrations, since the integrals are absolutely convergent):

$$\int_{\mathbb{Y}} \frac{da'\,d\phi'}{a'} y(\phi',a') \int_{\mathbb{T}^1} g_{\phi,a}(x)\bar{g}_{\phi',a'}(x)dx$$
$$= c_g^{+(-)} y(\phi,a),$$

which proves the theorem. Q.E.D.

IV. THE WAVELET TRANSFORM OF $\mathcal{D}'(\mathbb{T}^1)$

In this section we will be interested in the wavelet transform of the "functions" over the circle, with very low regularity, that is, the space of distributions $\mathcal{D}'(\mathbb{T}^1)$. The lack of local smoothness is reflected in the wavelet transform by a polynomial growth of the coefficients at small scale. However, the wavelet transformation allows us to represent any distribution by a C^∞ function over the position scale space.

The elements of $\mathcal{D}'(\mathbb{T}^1)$ are continuous linear functionals of $C^\infty(\mathbb{T}^1)$; that is, for any X in $\mathcal{D}'(\mathbb{T}^1)$ there is an integer number n such that, for any $s \in C^\infty(\mathbb{T}^1)$, we have

$$|X(s)| \leqslant C_{te} \|s\|_{C^\infty(\mathbb{T}^1);n}. \tag{4.1}$$

The smallest such n is called the order of the distribution X. A topology in $\mathcal{D}'(\mathbb{T}^1)$ is given by requiring that any sequence X_n in $\mathcal{D}'(\mathbb{T}^1)$ tends to zero if and only if $X_n(s)$ goes to zero for all $s \in C^\infty(\mathbb{T}^1)$. We identify any function s of $C^\infty(\mathbb{T}^1)$ with the distribution $(s,\cdot)$. This embedding is continuous. The space $C^\infty(\mathbb{T}^1)$ is dense in $\mathcal{D}'(\mathbb{T}^1)$, and therefore any distribution can be approximated by functions in $C^\infty(\mathbb{T}^1)$. We may even choose s_n ($s_n \to X$ in the sense of distributions) in such a way that, for all $r \in C^\infty(\mathbb{T}^1)$ and n, we have (m being the order of X)

$$|(s_n,r)| \leqslant C_{te} \|r\|_{C^\infty(\mathbb{T}^1);m}. \tag{4.2}$$

In the same way we denote by $\mathcal{D}'(\mathbb{Y})$ the space of linear continuous functionals over $S(\mathbb{Y})$. For any pair of functions f, h over $\mathbb{Y}$ we define the following "scalar product" $(\cdot,\cdot)_{L^2(\mathbb{Y})}$ whenever the following integral converges absolutely:

$$(f,g)_{L^2(\mathbb{Y})} = \int_{\mathbb{Y}} \bar{f}(\phi,a)g(\phi,a) \frac{da\,d\phi}{a}. \tag{4.3}$$

We now define the wavelet transformation of distributions.

Definition 4.1: T_g: $\mathcal{D}'(\mathbb{T}^1) \to \mathcal{D}'(\mathbb{Y})$ is defined by $X \in \mathcal{D}'(\mathbb{T}^1) \Rightarrow (T_g X)(y) = X(T_g^{-1}y)$, for all $y \in S(\mathbb{Y})$. T_g^{-1}: $\mathcal{D}'(\mathbb{Y}) \to \mathcal{D}'(\mathbb{T}^1)$ is defined by $Y \in \mathcal{D}'(\mathbb{Y}) \Rightarrow (T_g^{-1}Y)(s) = Y(T_g s)$, for all $s \in C^\infty(\mathbb{T}^1)$.

Here, for functions in $C^\infty(\mathbb{T}^1)$ and $S(\mathbb{Y})$, the definitions of Sec. II apply.

These definitions are reasonable, since T_g and T_g^{-1} are continuous maps between $C^\infty(\mathbb{T}^1)$ and $S(\mathbb{Y})$. In the following theorem we show that Definition 4.1 actually extends Definitions 3.1 and 3.5.

Theorem 4.2: T_g and T_g^{-1} are the only possible continuous extensions of T_g restricted to $C^\infty(\mathbb{T}^1)$ and T_g^{-1} restricted to $S(\mathbb{Y})$.

Proof: Let $X_n \in \mathcal{D}'(\mathbf{T}^1)$, $X_n \to 0$ $(n \to \infty)$ in the sense of distributions. For all $y \in S(\mathbf{Y})$, we have

$$(T_g X_n)(y) = X_n(T_g^{-1} y) \to 0 \quad (n \to \infty),$$

which shows that T_g is a continuous map from $\mathcal{D}'(\mathbf{T}^1)$ to $\mathcal{D}'(\mathbf{Y})$. In exactly the same way we can show that T_g^{-1} is continuous, too. Let $s \in C^\infty(\mathbf{T}^1)$. Taking s as a distribution in $\mathcal{D}'(\mathbf{T}^1)$, we write, for any $y \in S(\mathbf{Y})$,

$$(T_g s)(y) = \int_{\mathbf{Y}} \frac{da\, d\phi}{a} \int_{\mathbf{T}^1} dx\, \bar{g}_{\phi,a}(x) s(x) y(\phi,a)$$
$$= (T_g s, y)_{L^2(\mathbf{Y})}.$$

We could exchange the integrations since all integrals converge absolutely. Since y was arbitrary, we have proved that Definition 4.1 coincides with Definition 3.1 in the case where a distribution is a function in $C^\infty(\mathbf{T}^1)$. But $C^\infty(\mathbf{T}^1)$ is dense in $\mathcal{D}'(\mathbf{T}^1)$ and therefore by continuity there is exactly one continuous extension. The proof for T_g^{-1} is the same. Q.E.D.

The next theorem shows that the wavelet transform of any distribution in $\mathcal{D}'(\mathbf{T}^1)$ can be identified with a function in $C^\infty(\mathbf{Y})$, that is, the space of functions over $\mathbf{Y}$ that are infinitely differentiable.

Theorem 4.3: Let $X \in \mathcal{D}'(\mathbf{T}^1)$. For all $y \in S(\mathbf{Y})$, we have

$$(T_g X)(y) = (\mathscr{X}, y)_{L^2(\mathbf{Y})},$$

where the function $\mathscr{X}$ is in $C^\infty(\mathbf{Y})$ and is given by $\mathscr{X}(\phi,a) = X(g_{\phi,a})$.

Proof: We can find a sequence s_n of functions in $C^\infty(\mathbf{T}^1)$ that converges to X in the sense of distributions and that satisifies (4.2). Therefore we may write

$$(T_g X)(y) = \lim_{n \to \infty} (s_n, T_g^{-1} y)_{L^2(\mathbf{T}^1)}$$
$$= \lim_{n \to \infty} \int_{\mathbf{Y}} \overline{(g_{\phi,a}, s_n)}\, y(\phi,s) \frac{da\, d\phi}{a}. \qquad (4.4)$$

We could exchange the integration for fixed n because of the absolute convergence of all the integrals. Clearly $(g_{\phi,a}, s_n)$ tends to $X(g_{\phi,a})$ pointwise, that is, for each $(\phi,a) \in \mathbf{Y}$. But from (4.2) it follows that

$$|(g_{\phi,a}, s_n)| \leqslant C \|g_a\|_{C^\infty(\mathbf{T}^1);m},$$

where m is the order of X. Therefore Lemma 3.3 shows that the integrand in (4.4) is uniformly bounded by a function absolutely integrable over $\mathbf{Y}$. We can apply the theorem of dominated convergence to conclude.

Again it is useful to split the whole space into $\mathcal{D}'_{+(-)}(\mathbf{T}^1)$, the space of distributions that are acting on the positive (negative) frequencies only, and $\mathscr{K}(\mathbf{T}^1)$, the distributions that are multiples of the integral over the circle. Again every distribution can be written in a unique way as the superposition of three distributions, each one belonging to one of these three classes:

$$\mathcal{D}'(\mathbf{T}^1) = \mathcal{D}'_+(\mathbf{T}^1) \oplus \mathscr{K}(\mathbf{T}^1) \oplus \mathcal{D}'_-(\mathbf{T}^1).$$

We now will write a distribution in $\mathcal{D}'(\mathbf{T}^1)$ as a well defined "scalar product" of functions in $C^\infty(\mathbf{Y})$ obtained by an absolutely convergent integral over $\mathbf{Y}$.

Theorem 4.4: Let $s \in C^\infty_{+(-)}(\mathbf{T}^1)$ and let X be a distribution in $\mathcal{D}'_{+(-)}(\mathbf{T}^1)$. Then

$$X(s) = (1/c_g^{+(-)})(\mathscr{X}, T_g s)_{L^2(\mathbf{Y})},$$

and $\mathscr{X} = T_g X$ as given by Theorem 4.3.

Proof: From Theorem 3.7 it follows that

$$(T_g X)(T_g s) = X(T_g^{-1} T_g s) = c_g^{+(-)} X(s).$$

Since $T_g s \in S(\mathbf{Y})$, Theorem 4.3 shows how $(T_g X)(T_g s)$ can be written as a scalar product. Q.E.D.

We now want to characterize the image of $\mathcal{D}'(\mathbf{T}^1)$ under the wavelet transform.

Theorem 4.5: The image of $\mathcal{D}'_{+(-)}(\mathbf{T}^1)$ under T_g are exactly those functions $\mathscr{Y}$ in $C^\infty(\mathbf{Y})$ that satisfy (i) there exists $m \in \mathbf{Z}$ such that

$$\mathscr{Y}(\phi,a) = O(1/a^m) \quad (a \to 0)$$

uniformly in ϕ; (ii) for all $p > 0$, we have

$$\mathscr{Y}(\phi,a) = O(1/a^p) \quad (a \to \infty)$$

uniformly in ϕ; and (iii) $\mathscr{Y}$ satisfies the reproducing kernel equation (pointwise)

$$\mathscr{Y}(\phi,a) = (1/c_g^{+(-)})(p_g(\phi,a;\cdot,\cdot), \mathscr{Y})_{L^2(\mathbf{Y})}.$$

Proof: Let $X \in \mathcal{D}'_{+(-)}(\mathbf{T}^1)$. Then $\mathscr{Y}(\phi,a) = T_g X = X(g_{\phi,a})$ satisfies (i) and (ii) as follows by direct computation from (4.1) and Lemma 3.3. To show (iii), note that, for any $s \in C^\infty(\mathbf{T}^1)$, we can write, with the help of Theorem 2.4,

$$(T_g^{-1} T_g X)(s) = X(T_g^{-1} T_g s) = c_g^{+(-)} X(s).$$

Therefore, we can write, for $\mathscr{Y} = T_g X$,

$$(T_g T_g^{-1} \mathscr{Y})(\phi,a) = c_g^{+(-)} \mathscr{Y}(\phi,a).$$

On the other hand, a direct computation shows that

$$(T_g T_g^{-1} \mathscr{Y})(\phi,a) = c_g^{+(-)} (p_g(\phi,a;\cdot,\cdot), \mathscr{Y})_{L^2(\mathbf{Y})}.$$

The integral on the right-hand side converges absolutely for every $(\phi,a) \in \mathbf{Y}$ due to the rapid decrease of the reproducing kernel p_g at small scales.

Now suppose that $\mathscr{Y}$ is a locally integrable function over $\mathbf{Y}$ that satisfies (i)–(iii). We define $X \in \mathcal{D}'(\mathbf{T}^1)$ by

$$X(s) = (1/c_g^{+(-)})(\mathscr{Y}, T_g s)_{L^2(\mathbf{Y})}.$$

Clearly X is well defined since $T_g s$ is rapidly decreasing (Theorem 3.2), and a direct computation shows that $T_g X = \mathscr{Y}$, thus showing that $\mathscr{Y}$ is in the image of $\mathcal{D}'_{+(-)}(\mathbf{T}^1)$ under T_g. Q.E.D.

V. THE WAVELET TRANSFORM OF $L^2(\mathbf{T}^1)$

In this section we will analyze the Hilbert space of square integrable functions over the circle. As subspace of $\mathcal{D}'(\mathbf{T}^1)$ all theorems of the previous section hold for $L^2(\mathbf{T}^1)$. In particular, the image of $L^2(\mathbf{T}^1)$ are functions in $C^\infty(\mathbf{Y})$ that satisfy the reproducing kernel equation. It turns out that the wavelet transform is an isometry. Its range is a closed subspace of $L^2(\mathbf{Y})$, the Hilbert space over $\mathbf{Y}$ with scalar product (4.3). So the image of $L^2(\mathbf{T}^1)$ under T_g turns out to be a Hilbert space with reproducing kernel.

Theorem 5.1: The operator

$$\left(1/\sqrt{c_g^{+(-)}}\right) T_g : \; H^2_{+(-)}(\mathbf{T}^1) \to L^2(\mathbf{Y})$$

is an isometry. Its adjoint is the only bounded operator

form $L^2(\mathbf{Y})$ to $H^2_{+(-)}(\mathbf{T}^1)$ that coincides with $(1/\sqrt{c_g^{+(-)}})T_g^{-1}$ when it is restricted to $S(\mathbf{Y})$. Let the wavelet g have only positive (negative) frequency contributions. Then the reproducing kernel gives the orthogonal projector $P_g^{+(-)}$ on the image of $H^2_{+(-)}(\mathbf{T}^1)$ under T_g:

$$P_g^{+(-)}: \quad L^2(\mathbf{Y}) \to L^2(\mathbf{Y}),$$

$$(P_g^{+(-)}y)(\phi,a) = (1/c_g^{+(-)})(p_g(\phi,a;\cdot,\cdot),y)_{L^2(\mathbf{Y})}.$$

Proof: It is enough to show the theorem on a dense subset of $H^2_{+(-)}(\mathbf{T}^1)$. Let $s,u \in C^\infty_{+(-)}(\mathbf{T}^1)$. We look upon s as a distribution in $\mathscr{D}'_{+(-)}(\mathbf{T}^1)$. From Theorem 4.4 it follows that

$$(s,u) = (T_g s, T_g u)_{L^2(\mathbf{Y})},$$

and therefore T_g is an isometry. A direct computation shows that the adjoint is as stated in the theorem. The fact that the reproducing kernel equation is an orthogonal projection operator follows from the well known statement about partial isometries. Q.E.D.

[1] A. Grossmann and J. Morlet, in *Mathematics and Physics, Lectures on Recent Results*, edited by L. Streit (World Scientific, Singapore, 1987).

[2] A. Grossmann, J. Morlet, and T. Paul, "Integral transforms associated to squareintegrable representations. I," J. Math. Phys. **26**, 2473 (1985).

[3] A. Grossmann, J. Morlet, and T. Paul, "Integral transforms associated to squareintegrable representations. II," Ann. Inst. H. Poincaré **45**, 293 (1986).

[4] A. Grossmann and R. Murenzi, "Integral transforms associated to squareintegrable representations III," in preparation.

[5] Y. Meyer, "Principe d'incertitude, bases Hilbertiennes et algèbres d'operateurs," Asterisque **145-145**, 209 (1987).

[6] Y. Katznelson, *An Introduction to Harmonic Analysis* (Dover, New York).

Chern–Simons forms on principal superfiber bundles

Claudio Bartocci and Ugo Bruzzo
Dipartimento di Matematica, Università di Genova, Via L. B. Alberti 4, 16132 Genova, Italy
Giovanni Landi[a)]
Department of Physics, Syracuse University, Syracuse, New York 13244-1130

(Received 21 February 1989; accepted for publication 19 July 1989)

A graded Weil homomorphism is defined for principal superfiber bundles and the related transgression (or Chern–Simons) forms are introduced. As an example of the application of these concepts, a "superextension" of the Dirac monopole is discussed.

I. INTRODUCTION

In recent years Chern–Simons forms[1] have found a variety of applications in quantum field theory.[2] Chern–Simons forms are used as topological mass terms in gravity and Yang–Mills theories[3]; moreover, the consistent and covariant anomalies of Yang–Mills-type theories can be interpreted in terms of them[4–6] and they play a fundamental role in the anomaly cancellation mechanism in string theory.[7] Finally, several supergravity models have Lagrangians involving Chern–Simons terms.[8]

Therefore, it seems interesting to generalize concepts involving Chern–Simons forms to the context of supermanifolds and in particular to the case of principal superfiber bundles over supermanifolds [some work in this direction, with a particular choice of the structure (super) group and using local techniques, has already been done[9–11]]. This generalization should be relevant to the study of the anomalies of superstring and supersymmetric field theories; see, e.g., Ref. 11 and the references therein. Some related work (Chern classes for superbundles and cohomological treatment of anomalies of supersymmetric gauge theories) can be found in Refs. 12–14.

In this paper, using rigorous supermanifold theory and relying on a general algebraic description of the Weil homomorphism[15] in the graded setting,[16] we construct a Weil homomorphism for principal superfiber bundles. This is used to construct in terms of curvature forms some invariants associated with the superbundle and in particular to define Chern–Simons forms. Some of the concepts introduced are applied to the study of a "superextension" of the Dirac monopole.[17]

The relationship of the Weil homomorphism to the cohomology of the classifying space for the structure supergroup of the principal superfiber bundle will be discussed in a future paper. This relationship could be of some consequence in physics, in accordance with the work of Bonora *et al.*[18]

This paper is arranged as follows. In Sec. II we review some basic material concerning supermanifolds, including super Lie groups and superfiber bundles. Supermanifolds are intended in the sense of DeWitt and Rogers,[19,20] i.e., they are topological manifolds modeled on superspace whose transition functions fulfill a suitable smoothness condition. We also recall some facts about the cohomology of supermanifolds. In Sec. III the theory of the graded Weil homomorphism is expounded and the transgression (or Chern–Simons) forms associated with a given prinicipal superfiber bundle are introduced. In Sec. IV we define the elementary Ad–invariant polynomials of the general linear superalgebra and construct the related invariants of principal super fiber bundles, when the structure supergroup is the general linear supergroup or one of its subgroups. In Sec. V some of these concepts are exemplified by studying a Grassmann extension of the Dirac monopole.

We shall use some elementary sheaf theory. Indeed, while all information concerning a smooth manifold is encoded in the ring of global functions over the manifold, this is not true for supermanifolds (as it happens in the case of complex manifolds), so that the use of sheaves is mandatory.

II. SUPERMANIFOLDS AND PRINCIPAL SUPERFIBER BUNDLES

A. Algebraic preliminaries

In this paper we shall be constantly concerned with $\mathbf{Z}_2$-graded modules. We shall always say "graded" for "$\mathbf{Z}_2$ graded" and the grading will be denoted as follows: If $M = M_0 \oplus M_1$, $|x| = j$ means $x \in M_j$. If either $x \in M_0$ or $x \in M_1$, x is said to be *homogeneous*. A morphism of graded modules $f: M \to N$ is said to be *even* (resp., *odd*) if $f(M_i) \subset N_i$ [resp., $f(M_i) \subset N_{i+1}$].

We denote by B_L the exterior algebra over $\mathbf{R}^L$, $L < \infty$. Here B_L is naturally graded, $B_L = (B_L)_0 \oplus (B_L)_1$ and is a graded commutative algebra, i.e.,

$$ab \in (B_L)_{|a|+|b|}, \quad ab = (-1)^{|a||b|} ba,$$

if $a, b \in B_L$ are homogeneous. If N_L is the nilpotent ideal of B_L, one has $B_L = \mathbf{R} \oplus N_L$. The relevant projections $\sigma: B_L \to \mathbf{R}$ and $s: B_L \to N_L$ are called *body* and *soul* maps, respectively. The Cartesian product B_L^{m+n} can be made into a graded B_L module by setting

$$B_L^{m+n} = B_L^{m,n} \oplus B_L^{\bar{m},\bar{n}},$$

with $B_L^{m,n} = (B_L)_0^m \times (B_L)_1^n$, $B_L^{\bar{m},\bar{n}} = (B_L)_1^m \times (B_L)_0^n$. A body map $\sigma^{m,n}: B_L^{m,n} \to \mathbf{R}^m$ is defined by letting $\sigma^{m,n}(x^1 \cdots x^m, y^1 \cdots y^n) = (\sigma(x^1) \cdots \sigma(x^m))$.

A graded B_L module is said to be *free of dimension* (m,n) if it is free of rank $m+n$ over B_L and has bases formed by m even and n odd elements. Any left-graded B_L module M can be turned into a right module, and vice versa, by letting

$$xa = (-1)^{|a||x|} ax, \quad \forall \text{ homogeneous } x \in M, \ a \in B_L.$$

[a)] Also at Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Mostra d'Oltremare Pad. 19, I-80125, Napoli, Italy.

 0022-2488/90/010045-10$03.00

Given two graded B_L modules M, N, their tensor product over B_L, calculated by considering M as a right module and N as a left module, can be canonically given a structure of a graded B_L module. We shall always regard $M \otimes_{B_L} N$ as endowed with such a structure (of course, this applies to any graded commutative ring and not only to B_L).

The set of $(m+n)\times(m+n)$ matrices with entries in B_L, denoted by $\mathrm{gl}(m+n)$, is a graded B_L module of dimension $(m^2+n^2, 2mn)$. It can be graded so that its even part, denoted by $\mathrm{gl}(m,n)$, is formed by matrices of the form

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{2.1}$$

where the matrices A and D are $m\times m$ and $n\times n$, respectively, and have entries in $(B_L)_0$; while B and C are $m\times n$ and $n\times m$, respectively, and have entries in $(B_L)_1$. Endowed with the bracket whose action on homogeneous elements is $[X,Y] = XY - (-1)^{|X||Y|}YX$, $\mathrm{gl}(m+n)$ is a graded Lie B_L algebra (i.e., a Lie superalgebra over B_L).

B. Supermanifolds

Following Rogers,[21] we introduce a sheaf of B_L-valued functions on $B_L^{m,n}$ called GH$^\infty$ functions. We denote by $\mathscr{C}[V;B_L]$ the graded algebra of C^∞ B_L-valued functions on $V\subset X$, where X is any manifold, and regard B_L and $B_L^{m,n}$ as topological spaces by endowing them with their vector space topology. Let us fix two positive integers L and L', with $L'\leqslant L$ and, for any open set U in $\mathbb{R}^m$, let us denote by

$$Z_{L',L}: \mathscr{C}[U;B_L] \to \mathscr{C}[(\sigma^{m,0})^{-1}(U);B_L]$$

the morphism of graded $B_{L'}$ algebras defined by

$$Z_{L',L}(f)(x^1\cdots x^n) = \sum_{i_1\cdots i_m=0}^{L} \frac{1}{i_1!\cdots i_m!} (\partial_1^{i_1}\cdots\partial_m^{i_m} f)_{|(\sigma(x^1)\cdots\sigma(x^m))} \times s(x^1)^{i_1}\cdots s(x^m)^{i_m}.$$

One proves that $Z_{L',L}$ is injective; its image will be denoted by $\widehat{\mathscr{G}\mathscr{H}}[(\sigma^{m,0})^{-1}(U)]$ and will be identified with the graded algebra of GH^∞ functions of even variables on $(\sigma^{m,0})^{-1}(U)$. The GH$^\infty$ functions of even and odd variables are naturally defined on the sets $(\sigma^{m,n})^{-1}(U)$, where U is an open set in $\mathbb{R}^m$. The relevant function algebra is denoted by $\mathscr{G}\mathscr{H}[(\sigma^{m,n})^{-1}(U)]$ and its elements have the form

$$F(x^1\cdots x^m, y^1\cdots y^n) = F_0(x^1\cdots x^m) + \sum_{\substack{k=1\cdots n \\ 1\leqslant\alpha_1<\cdots<\alpha_k\leqslant n}} F_{\alpha_1\cdots\alpha_k}(x^1\cdots x^m) \times y^{\alpha_1}\cdots y^{\alpha_k},$$

where $F_{\alpha_1\cdots\alpha_k}\in\widehat{\mathscr{G}\mathscr{H}}[(\sigma^{m,0})^{-1}(U)]$. The first derivatives of F are *uniquely* determined by the development

$$F(x+h,y+k) = F(x,y) + \sum_{i=1}^{m} h^i \frac{\partial F}{\partial x^i}(x,y) + \sum_{\alpha=1}^{n} k^\alpha \frac{\partial F}{\partial y^\alpha}(x,y) + O(h,k)^2,$$

provided that

$$L - L' > n. \tag{2.2}$$

We may as well consider GH^∞ functions on arbitrary open sets $W\subset B_L^{m,n}$: An element of $\mathscr{G}\mathscr{H}[W]$ is a function in $\mathscr{G}\mathscr{H}[(\sigma^{m,n})^{-1}\sigma^{m,n}(W)]$ restricted to W. One can easily check that this defines a sheaf of graded $B_{L'}$ algebras over $B_L^{m,n}$. One has also a sheaf $\widehat{\mathscr{G}\mathscr{H}}$ of germs of GH^∞ functions of even variables and the two sheaves are related by

$$\mathscr{G}\mathscr{H} \simeq \widehat{\mathscr{G}\mathscr{H}} \otimes_{\mathbb{R}} \Lambda[n],$$

where $\Lambda[n]$ is the exterior algebra over $\mathbb{R}$ with n generators.

Setting $L = L'$ one recovers the so-called G^∞ functions originally introduced by Rogers.[20] These are badly behaved since condition (2.2) is violated; the lack of definition of partial derivatives with respect to odd variables implies that the modules of derivations of the local function algebras are not free. However, the sheaf $\mathscr{G}\mathscr{H}$, where L, L', and n are supposed to satisfy condition (2.2), also has some unpleasant features, mainly related with the definition of a tangent space and the relationship between derivations and tangent vectors.[22] Such drawbacks can be eliminated by considering a new, enlarged structure sheaf

$$\mathscr{G} = \mathscr{G}\mathscr{H} \otimes_{B_{L'}} B_L, \tag{2.3}$$

which is a sheaf of graded B_L algebras. The partial derivatives of sections of $\mathscr{G}$ are defined according to the rule

$$\frac{\partial(f\otimes a)}{\partial x^A} = \frac{\partial f}{\partial x^A}\otimes a.$$

The definition of supermanifold we shall adopt is such that the structure sheaf of a supermanifold locally has the form (2.3). A precise definition of the resulting category of supermanifolds, whose objects we call $\mathscr{G}$ supermanifolds, while the morphisms are called $\mathscr{G}$ maps, was given in Refs. 22 and 23. Here we wish only to recall the following.

(i) If U, V are open sets in $B_L^{m,n}$ and $f: U\to V$ is a $\mathscr{G}$ map, then if $f^*(\mathscr{G}|V)$ is a subsheaf of $\mathscr{G}|U$; in other words a $\mathscr{G}$ map pullbacks sections of $\mathscr{G}$ into sections of $\mathscr{G}$.

(ii) If S is an (m,n)-dimensional $\mathscr{G}$ supermanifold, the topological space underlying S is Hausdorff second countable and on S there is an atlas $\mathscr{A} = \{(U_\alpha,\psi_\alpha)\,|\,\psi_\alpha: U_\alpha \to B_L^{m,n}\}$ such that its transition functions are $\mathscr{G}$ maps.

Defining the evaluation morphism $\delta: f\otimes a \mapsto fa$ mapping $\mathscr{G}$ into the sheaf of B_L-valued C^∞ functions on S, the triple $(S,\mathscr{G},\delta)$ is a supermanifold in the sense of Rothstein[24]; moreover, $\delta(\mathscr{G})$ is the sheaf of G^∞ functions on S.

Remarks: (i) The supermanifold S also inherits a structure of ordinary differentiable manifold of dimension $2^{L-1}(m+n)$.

(ii) The module $B_L^{m,n}$ is obviously a supermanifold; also, B_L^{m+n} has a structure of supermanifold of dimension $(m+n,m+n)$ given by the canonical isomorphism of $(B_L)_0$ modules of $B_L^{m+n} \simeq B_L^{m+n,m+n}$.

Henceforth we shall say simply "supermanifold" instead of "$\mathscr{G}$ supermanifold." A class of supermanifolds that is important for physical applications is given by the so-called *De Witt supermanifolds*.[19,20,25] De Witt supermanifolds are defined in terms of a coarse topology on $B_L^{m,n}$, called

the De Witt topology, whose open sets are the counterimages of open sets in $\mathbb{R}^m$ through the body map $\sigma^{m,n}: B_L^{m,n} \to \mathbb{R}^m$. We say that the (m,n) supermanifold is De Witt if it has an atlas such that the images of the coordinate maps are open in the De Witt topology. Loosely speaking, a supermanifold is De Witt if it has an atlas with a "tubelike" cover. It is easily shown[20] that a De Witt (m,n) supermanifold is a locally trivial fiber bundle over an m manifold S_0 with a vector fiber. The manifold S_0 is usually called the *body* of S and the bundle projection $\Phi: S \to S_0$ is given in local bundle coordinates by the body map $\sigma^{m,n}$.

C. Supervector bundles[22,23]

A supervector bundle (SVB) of rank (p,q) over a supermanifold S is a pair (ξ,π), where ξ is a supermanifold, $\pi: \xi \to S$ is a $\mathcal{G}$ map, and there is a cover $\{U_\alpha\}$ of S with $\mathcal{G}$ diffeomorphisms

$$\psi_\alpha: \pi^{-1}(U_\alpha) \to U_\alpha \times B_L^{p+q}$$

such that $pr_1 \circ \psi_\alpha = \pi$. Moreover, the transition functions $g_{\alpha\beta}$ defined as usual, i.e.,

$$g_{\alpha\beta}(x)(v) = pr_2 \circ \psi_\alpha \circ \psi_\beta^{-1}(x,v),$$

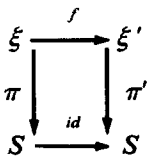
$$\forall x \in U_\alpha \cap U_\beta, \quad v \in B_L^{p+q},$$

are required to be $(B_L)_0$ linear, i.e.,

$$g_{\alpha\beta}: U_\alpha \cap U_\beta \to \mathrm{GL}(p,q),$$

where $\mathrm{GL}(p,q)$ is the group of even automorphisms of the graded B_L module B_L^{p+q}, which is the open subset of $\mathrm{gl}(p,q)$ formed by invertible matrices.

Given two SVB's ξ, ξ' over a supermanifold S, a $\mathcal{G}$ map $f: \xi \to \xi'$ is said to be a *morphism of SVB's* if the diagram



commutes and f induces morphisms of $\mathcal{G}$-modules between the modules of local sections of ξ and ξ'.[23]

D. Graded tangent space

Let S be an (m,n) supermanifold and $\mathcal{G}$ its structure sheaf. Let us denote by $\mathrm{Der}\,\mathcal{G}$ the sheaf of derivations of $\mathcal{G}$[22]; by $\mathrm{Der}^*\,\mathcal{G}$ the dual sheaf; and finally,

$$\Lambda^p = \wedge^p_{\mathcal{G}} \mathrm{Der}^*\,\mathcal{G}.$$

The sections of Λ^p are called differential p-forms. It is easily proved that $\mathrm{Der}\,\mathcal{G}$, $\mathrm{Der}^*\,\mathcal{G}$, and Λ^p are free graded $\mathcal{G}$ modules. Differential one-forms on an open set U are said to be even (or odd) if they are even (or odd) as morphisms $\mathrm{Der}\,\mathcal{G}(U) \to \mathcal{G}(U)$.

We wish to relate $\mathrm{Der}\,\mathcal{G}$ with a notion of graded tangent space. For each $x \in S$, denote by T_xS the space of graded B_L linear maps $X: \mathcal{G}_x \to B_L$ satisfying

$$X(fg) = X(f)\tilde{g} + (-1)^{|X||f|}\tilde{f}X(g),$$

where the overtilde denotes evaluation of germs at x.[22,24] Here T_xS is a free graded B_L module of dimension (m,n) and its even part is canonically isomorphic, both as a real vector space and as a $(B_L)_0$ module, with the ordinary tangent space to S at x.[19,20]

Finally, the disjoint union $\cup_{x \in S} T_xS$ can be given the structure of a rank (m,n) SVB over S, denoted by TS, and one has a natural identification $TS \equiv \mathrm{Der}\,\mathcal{G}$.

Whenever $D \in \mathrm{Der}\,\mathcal{G}(U)$, $\eta \in \mathrm{Der}^*\,\mathcal{G}(U)$, we shall denote the action of η on D by $D \lrcorner \eta$.

E. Supermanifold cohomology[26,27]

A sheaf morphism $d: \mathcal{G} \equiv \Lambda^0 \to \mathrm{Der}^*\,\mathcal{G} \equiv \Lambda^1$, called *exterior differential*, is defined by letting $D \lrcorner df = D(f)\ \forall f \in \mathcal{G}(U)$, $D \in \mathrm{Der}^*\,\mathcal{G}(U)$ and is extended to morphisms $\Lambda^p \to \Lambda^{p+1}$, $p \geqslant 0$, in the usual way, so that $d^2 = 0$. We introduce the differential complex

$$\Lambda^0(S) \xrightarrow{d} \Lambda^1(S) \xrightarrow{d} \Lambda^2(S) \to \cdots$$

and denoted by H^*_{SDR} it cohomology (super de Rham cohomology). Since S has also a structure of ordinary manifold, it is natural to compare $H^*_{\mathrm{SDR}}(S)$ with the ordinary de Rham cohomology of S, $H^*_{\mathrm{DR}}(S)$. Each $\Lambda^p(S)$ into $\mathcal{C}^p(S) \otimes_{\mathbb{R}} B_L$, where $\mathcal{C}^p(S)$ is the vector space of C^∞ differential p forms on S, and one has a corresponding morphism in cohomology,

$$H^p_{\mathrm{SDR}}(S) \to H^p_{\mathrm{DR}}(S) \otimes B_L, \quad p \geqslant 0. \tag{2.4}$$

In general, this morphism is neither injective nor surjective. However, we have the following result:

If S is a De Witt supermanifold, the morphism (2.4) is bijective. Moreover, $H^*_{\mathrm{DR}}(S) \simeq H^*_{\mathrm{DR}}(S_0)$.

These facts rely on the triviality of the Čech cohomology of $\mathcal{G}$ when S is De Witt.[27] Another property of SDR cohomology is that it is not a topological invariant: Two homeomorphic supermanifolds may have different SDR cohomologies.[26] Therefore, SDR cohomology carries different information than de Rham cohomology; in particular, it "feels" the superdifferentiable structure.

F. Super Lie groups[28]

A super Lie group G is an algebraic group also carrying a supermanifold structure such that $(g,h) \mapsto gh^{-1}$ is a $\mathcal{G}$ map. An example of a super Lie group is the group $\mathrm{GL}(p,q)$ introduced in Sec. II C. Let us denote by $\mathcal{G}_G$ the structure sheaf of G. Having introduced the left and right transport operators L_g and R_g as usual, we define the *Lie module* W_G of G as the space of left-invariant global graded derivations of $\mathcal{G}_G$, i.e.,

$$W_G = \{D \in \mathrm{Der}\,\mathcal{G}_G(G) \text{ s.t. } L_g D = D, \quad \forall g \in G\}.$$

Now W_G is endowed with a bracket by setting

$$[D_1, D_2] = D_1 \circ D_2 - (-1)^{|D_1||D_2|} D_2 \circ D_1.$$

The following results are easily proved.[28]

Proposition 2.1: We find that W_G is a free graded B_L module of the same dimension as G. The bracket defined above fulfills the following properties:

(i) $[D_1,D_2] = -(-1)^{|D_1||D_2|}[D_2,D_1]$;

(ii) $[aD_1,D_2] = a[D_1,D_2], \quad \forall a \in B_L$;

(iii) $\sum_{\text{cyclic}} (-1)^{|D_1||D_3|}[[D_1,D_2],D_3] = 0.$ ■

For instance, the graded B_L module $\mathrm{gl}(p,q)$ introduced in Sec. II A is the Lie module of $\mathrm{GL}(p,q)$.

The module W_G free over B_L, for a fixed basis formed by the homogeneous elements $\{D_A, A = 1,...,m+n\}$ of W_G there exist elements C^A_{BE} in B_L such that

$$[D_A,D_B] = C^E_{AB} D_E. \tag{2.5}$$

The grading of C^E_{AB} is the following: $|C^E_{AB}| = |A| + |B| + |E|$. The B_L dual W^*_G of W_G is given by the left-invariant one-forms on G, i.e.,

$$W^*_G = \{\theta \in \Lambda^1_G(G) \text{ s.t. } L^*_g \theta = \theta, \quad \forall g \in G\}.$$

The analogous property of (2.5) on W^*_G is expressed by the Maurer–Cartan equations: If $D_A \lrcorner \theta^B = \delta^B_A$,

$$d\theta^A = \tfrac{1}{2}\theta^B \wedge \theta^E C^A_{EB}.$$

Many other features of ordinary Lie groups have suitable counterparts for super Lie groups. For instance, one can define an exponential map exp: $(W_G)_0 \to G$, which is a $\mathscr{G}$ map and, in a neighborhood of 0 in $(W_G)_0$, is injective.

G. Principal superfiber bundles

Let G be a super Lie group. A PSFB over S with the structure supergroup G is a supermanifold Π carrying a right action of G such that $S = \Pi/G$ and having local trivializations satisfying the usual conditions (obviously, the quotient Π/G must be suitably defined, but this can be done easily). Moreover, the right action of G, the natural projection π: $\Pi \to G$, and the local trivializations are required to be $\mathscr{G}$ maps.

We introduce on Π the *vertical graded tangent bundle* $T^V\Pi$, whose sections are vertical derivations of $\mathscr{G}_\Pi$:

$$\Gamma[U,T^V\Pi] = \{D \in \mathrm{Der}\,\mathscr{G}_\Pi(U) \equiv \Gamma[U,T\Pi] \mid \pi_* D = 0\}.$$

We shall also denote by Λ^q_Π the SVB of differential q forms on Π. Let us consider the exact sequence of SVB's

$$0 \to T^V\Pi \to T\Pi \to \pi^{-1}TS \to 0. \tag{2.6}$$

A connection is an even B_L-linear morphism of SVB's[29]

$$\nabla : T\Pi \to T^V\Pi, \tag{2.7}$$

which splits the sequence (2.6) and is G invariant in the sense that

$$\nabla \circ R_{g*} = R_{g*} \circ \nabla.$$

Therefore, one has a G-invariant splitting $T\Pi \simeq T^V\Pi \oplus \pi^{-1}TS$ or, equivalently, a G-invariant splitting of graded B_L modules

$$T_u\Pi \simeq T^V_u\Pi \oplus \mathrm{Hor}_u\,\Pi, \quad \forall u \in \Pi, \tag{2.8}$$

where $\mathrm{Hor}_u\,\Pi$ is isomorphic with $T_{\pi(u)}S$. Equation (2.8) shows that one can associate with any connection ∇ an even differential one-form ω on Π, with values in W_G, the Lie module of the structure supergroup, satisfying the usual properties of connection forms on principal bundles. To state these properties, we must define the *fundamental vertical derivations* of $\mathscr{G}_\Pi$. Fixing a $u \in \Pi$, one considers the right action of the structure supergroup as a map $f_u : G \to \Pi$, $g \to ug$ and takes the tangent map to f_u at the identity of G, thus obtaining a map $T_e G \simeq W_G \to T_u\Pi$, $D \in W_G \mapsto D^* \in T_u\Pi$. One can check that by varying $u \in \Pi$ one obtains a section of $\mathrm{Der}\,\mathscr{G}_\Pi$ called the *fundamental vertical derivation* D^* associated with D. Then we have

$$\begin{aligned} &R^*_g \omega = (\mathrm{Ad}\, g^{-1})\omega, \quad \forall g \in G; \\ &D^* \lrcorner \omega = D, \quad \forall D \in W_G \end{aligned} \tag{2.9}$$

Conversely, any even W_G-valued differential one-form ω on Π fulfilling conditions (2.9) gives rise to a connection ∇ on Π.

Let $\mathrm{Hor}^1(\Pi)$ be the subbundle of Λ^1_Π whose sections vanish when applied on vertical derivations of $\mathscr{G}_\Pi$ and define $\mathrm{Hor}^q(\Pi) = \wedge^q_{\mathscr{G}_\Pi} \mathrm{Hor}^1(\Pi)$. Then the splitting (2.8) yields a projection $h : \Lambda^q_\Pi \to \mathrm{Hor}^q(\Pi)$. Given any $\eta \in \Gamma[U,\Lambda^q_\Pi]$, where U is an open set in Π, its covariant derivative is defined as

$$D^\nabla \eta = h(d\eta).$$

The covariant derivative Ω of the connection form ω is called the *curvature of* ∇; it fulfills the identities

$$\begin{aligned} &\Omega^A = d\omega^A - \tfrac{1}{2}\omega^B \wedge \omega^E C^A_{EB}, \\ &D^\nabla \Omega = 0 \text{ (Bianchi identity)}. \end{aligned} \tag{2.10}$$

Finally, we note that if η is a horizontal q form on Π of type (Ad, W_G), i.e., $\eta \in \Gamma[\Pi, \mathrm{Hor}^q\Pi \otimes_{B_L} W_G]$ and $R^*_g \eta = (\mathrm{Ad}\, g^{-1})\eta \, \forall g \in G$, then

$$D^\nabla \eta^A = d\eta^A + (-1)^q \omega^E \wedge \eta^B C^A_{BE}. \tag{2.11}$$

Equations (2.10) and (2.11) are conveniently written using the following notation. Let η and τ be a W_G-valued p and q form on Π, respectively, both homogeneous as elements in W_G, and define the W_G-valued $(p+q)$ form $[\eta,\tau]$ by letting, for homogeneous $Y_1 \cdots Y_{p+q} \in \mathrm{Der}\,\mathscr{G}_\Pi(\Pi)$,

$$\begin{aligned} &Y_1 \wedge \cdots \wedge Y_{p+q} \lrcorner [\eta,\tau] \\ &\quad = \frac{1}{p!q!} \sum_\sigma \chi(\sigma)(-1)^{c(\eta,\sigma,p,q) + d(\sigma)} \\ &\quad \times [Y_{\sigma(1)} \wedge \cdots \wedge Y_{\sigma(p)} \lrcorner \eta, \\ &\quad \times Y_{\sigma(p+1)} \wedge \cdots \wedge Y_{\sigma(p+q)} \lrcorner \tau], \end{aligned} \tag{2.12}$$

where $\chi(\sigma)$ is the sign of the permutation σ,

$$c(\eta,\sigma,p,q) = |\eta| \sum_{i=p+1}^{p+q} |Y_{\sigma(i)}|,$$

$d(\sigma)$ = number of minus signs that occur going from the sequence $Y_1 \cdots Y_{p+q}$ to the sequence $Y_{\sigma(1)} \cdots Y_{\sigma(p+q)}$,

and the summation is on all permutations of the first $p+q$ integers.

Then Eqs. (2.10) and (2.11) are written as

$$\Omega = d\omega + \tfrac{1}{2}[\omega,\omega], \quad D^{\nabla}\eta = d\eta + [\omega,\eta].$$

Further results on PSFB's and examples may be found in Refs. 30 and 31.

III. WEIL HOMOMORPHISM AND TRANSGRESSION FORMS

Now we shall use the general algebraic construction developed in Ref. 13 to construct a Weil homomorphism and the secondary Chern–Simons characteristic classes for PSFB's. Throughout this section we assume that we are given a PSFB $\pi:\Pi\to S$ with structure supergroup G; W_G will denote the Lie module of G. Moreover, we assume that on Π there is a connection, that we describe by means of a connection one-form ω with curvature Ω. We would like to stress that in general this is a genuine assumption since, as in the case of holomorphic bundles in the ordinary theory, the lack of a partition of unity prevents one from proving that any PSFB has a connection. Actually, one can define a cohomological invariant (the *Atiyah class* of the bundle) which vanishes if and only if the bundle carries a connection.[31]

A. Differential calculus of W_G-valued forms on Π

The first ingredient we need is the tensor algebra of W_G:

$$W_G^{\bullet} = \bigoplus_{k\in\mathbb{N}} W_G^k, \quad W_G^k = W_G \otimes_{B_L} \cdots \otimes_{B_L} W_G \quad (k \text{ times}, W_G^0 \equiv B_L). \tag{3.1}$$

Here $W_G^{\bullet}$ is a $\mathbb{Z}$-$\mathbb{Z}_2$-graded algebra with $\otimes_{B_L}$ as multiplication. In the following, the subscript B_L will be omitted. Next we take $\Lambda^{p,k} \equiv \Lambda^p(\Pi; W_G^k)$, the space of p forms on Π with values in W_G^k. If $\varphi\in\Lambda^{p,k}$, then for all $Y_i\in \mathrm{Der}\,\mathcal{S}_{\Pi}(\Pi)$,

$$\begin{aligned} &Y_1\wedge\cdots Y_i\wedge Y_{i+1}\wedge\cdots\wedge Y_p \lrcorner \varphi \\ &\quad = -(-1)^{|Y_i||Y_{i+1}|} Y_1\wedge\cdots Y_{i+1} \\ &\qquad \wedge Y_i\wedge\cdots\wedge Y_p \lrcorner \varphi. \end{aligned} \tag{3.2}$$

The vector space $\Lambda^{p,k}$ is graded according to

$$|Y_1\wedge\cdots\wedge Y_p \lrcorner \varphi| = |\varphi| + \sum_i |Y_i|.$$

The exterior differential on Π can be extended to $d:\Lambda^{p,k}\to\Lambda^{p+1,k}$ by letting, for homogeneous $Y_i\in\mathrm{Der}\,\mathcal{S}_{\Pi}(\Pi)$,

$$\begin{aligned} &Y_1\wedge\cdots\wedge Y_{p+1}\lrcorner\varphi \\ &\quad \equiv \sum_{i=1}^{p+1}(-1)^{a(i)} Y_i(Y_1\wedge\cdots\wedge\hat{Y}_i\wedge\cdots\wedge Y_{p+1}\lrcorner\varphi) \\ &\qquad + \sum_{1<i<j<p}(-1)^{b(i,j)}[Y_i,Y_j]\wedge Y_1\wedge\cdots\wedge\hat{Y}_i\wedge\cdots \\ &\qquad \wedge\hat{Y}_j\wedge\cdots\wedge Y_{p+1}\lrcorner\varphi, \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} a(i) &= 1+i+|Y_i|\sum_{h=1}^{i-1}|Y_h|, \\ b(i,j) &= i+j+|Y_i|\sum_{h=1}^{i-1}|Y_h| + |Y_j|\sum_{\substack{h=1\\h\neq i}}^{j-1}|Y_h|, \end{aligned} \tag{3.4}$$

and the caret indicates omission. An exterior product $\wedge$: $\Lambda^{p,k}\times\Lambda^{q,h}\to\Lambda^{p+q,k+h}$ is defined by letting

$$\begin{aligned} &Y_1\wedge\cdots\wedge Y_{p+q}\lrcorner(\varphi\wedge\psi) \\ &\quad = \frac{1}{p!q!}\sum_{\sigma}\chi(\sigma)(-1)^{c(\varphi,\sigma,p,q)+d(\sigma)} \\ &\qquad \times(Y_{\sigma(1)}\wedge\cdots\wedge Y_{\sigma(p)}\lrcorner\varphi) \\ &\qquad \otimes(Y_{\sigma(p+1)}\wedge\cdots\wedge Y_{\sigma(p+q)}\lrcorner\psi), \end{aligned} \tag{3.5}$$

where $\chi(\sigma)$, $c(\varphi,\sigma,p,q)$, and $d(\sigma)$ have the same meaning as in Eq. (2.13).

Finally, let us recall that in Sec. II we defined an operation

$$[\ ,\ ]:\Lambda^{p,1}\times\Lambda^{q,1}\to\Lambda^{p+q,1}. \tag{3.6}$$

Using definitions (3.3)–(3.6) one can prove the following proposition.

Proposition 3.1: If $\varphi\in\Lambda^{p,k}$ and $\psi\in\Lambda^{q,h}$, then

$$d(\varphi\wedge\psi) = d\varphi\wedge\psi + (-1)^p\varphi\wedge d\psi; \tag{3.7}$$

moreover, if $k=h=1$,

$$d([\varphi,\psi]) = [d\varphi,\psi] + (-1)^p[\varphi,d\psi], \tag{3.8}$$

$$[\varphi,\psi] = -(-1)^{pq+|\varphi||\psi|}[\psi,\varphi], \tag{3.9}$$

$$[\varphi,[\varphi,\varphi]] = 0. \tag{3.10}$$

B. Graded Weil homomorphism

Let us now consider the set $I_{\mathrm{Gr}}(W_G^k;B_L)$ of graded algebra morphisms P: $W_G^k\to B_L$, which are graded symmetric and adjoint-invariant, i.e.,

$$\begin{aligned} &P(Z_1\otimes\cdots Z_i\otimes Z_{i+1}\otimes\cdots\otimes Z_k) \\ &\quad = (-1)^{|Z_i||Z_{i+1}|}P(Z_1\otimes\cdots Z_{i+1} \\ &\qquad \otimes Z_i\otimes\cdots\otimes Z_k), \quad \forall Z_i\in W_G, \end{aligned} \tag{3.11}$$

$$P((\mathrm{Ad}\,g)Z_1\otimes\cdots\otimes(\mathrm{Ad}\,g)Z_k) = P(Z_1\otimes\cdots\otimes Z_k), \quad \forall Z_i\in W_G,\ g\in G \tag{3.12}$$

(one could, as well, consider the morphisms $P:W_G^k\to C_L$, where C_L is the complexfication of B_L, as we shall do in Sec. IV). Condition (3.12) implies that

$$\begin{aligned} &\sum_{i=1}^{k}(-1)^{|Z|(|Z_1|+\cdots+|Z_k|)}P(Z_1\otimes\cdots\otimes[Z,Z_i] \\ &\qquad \otimes\cdots\otimes Z_k) = 0, \quad \forall Z,\ Z_i\in W_G. \end{aligned} \tag{3.13}$$

We take $I_{\mathrm{Gr}}(W_G;B_L)\equiv\oplus_{k\in\mathbb{N}}I_{\mathrm{Gr}}(W_G^k;B_L)$ and make it into a graded B_L algebra by defining, for $P\in I_{\mathrm{Gr}}(W_G^k;B_L)$ and $\mathcal{T}\in I_{\mathrm{Gr}}(W_G^h;B_L)$, the product $P\mathcal{T}\in I_{\mathrm{Gr}}(W_G^{k+h};B_L)$ as follows:

$$\begin{aligned} &P\mathcal{T}(Z_1\otimes\cdots\otimes Z_{k+h}) \\ &\quad = \frac{1}{k!h!}\sum_{\sigma}(-1)^{d(\sigma)+|\mathcal{T}|(|Z_{\sigma(1)}|+\cdots+|Z_{\sigma(k)}|)} \\ &\qquad \times P(Z_{\sigma(1)}\otimes\cdots\otimes Z_{\sigma(k)} \\ &\qquad \otimes\mathcal{T}(Z_{\sigma(k+1)}\otimes\cdots\otimes Z_{\sigma(k+h)}), \end{aligned} \tag{3.14}$$

where $d(\sigma)$ is as in Eq. (3.5).

If $\varphi\in\Lambda^{p,k}$ and $P\in I_{\mathrm{Gr}}(W_G^k;B_L)$, by composition we ob-

tain a B_L- valued k form on Π, $P(\varphi) = P \circ \varphi \in \Lambda^p(\Pi)$. One shows easily that

$$dP(\varphi) = P(d\varphi). \tag{3.15}$$

Moreover, property (3.13) implies that given a collection $\{\psi_i \in \Lambda^{p_i,1}, i = 1,...,k\}$ and $\varphi \in \Lambda^{1,1}$,

$$\sum_{i=1}^{k} (-1)^{p_1 + \cdots + p_i + |\varphi|(|\psi_1| + \cdots + |\psi_i|)} \times P(\psi_1 \wedge \cdots \wedge [\psi_i, \varphi] \wedge \cdots \wedge \psi_k) = 0. \tag{3.16}$$

We are now ready to prove the following proposition.

Proposition 3.2: Given a PSFB $\pi: \Pi \to S$ with the structure supergroup G, let Ω be the curvature form of a connection on Π and let $P \in I_{\mathrm{Gr}}(W_G^k; B_L)$. Then the following holds.

(i) The $2k$ form $P(\Omega^k)$, where $\Omega^k = \Omega \wedge \cdots \wedge \Omega$ (k times), projects onto a closed $2k$ form $\overline{P(\Omega^k)} \in \Lambda^{2k}(S)$.

(ii) If $\mathscr{W}(P)$ is the element of the super de Rham cohomology group $H^{2k}_{\mathrm{SDR}}(S)$ defined by $\overline{P(\Omega^k)}$, then $\mathscr{W}(P)$ does not depend on the connection and $\mathscr{W}: I_{\mathrm{Gr}}(W_G; B_L) \to H^*_{\mathrm{SDR}}(S)$ is a graded algebra homomorphism (graded Weil homomorphism).

If we consider the morphisms $P: W_G^k \to C_L$, then $\mathscr{W}(P)$ takes values in $H^{2k}_{\mathrm{SDR}}(S) \otimes C_L$.

Proof of part (i) of Proposition 3.2: Since Ω is horizontal of type Ad, so is Ω^k. As a consequence, $P(\Omega^k)$ is invariant and horizontal. Therefore, there exists a unique $2k$ form $\overline{P(\Omega^k)} \in \Lambda^{2k}(S)$ whose pullback by π is $P(\Omega^k)$. To prove $\overline{dP(\Omega^k)} = 0$ it is sufficient to show $dP(\Omega^k) = 0$. However, using Eqs. (3.15), (3.7), (3.16), and the Bianchi identity, one has

$$dP(\Omega^k) = P(d\Omega^k) = kP(d\Omega \wedge \Omega^{k-1}) = kP([\Omega, \omega] \wedge \Omega^{k-1}) = 0.$$

In order to prove part (ii), we first state, without proof, the following lemma.

Lemma 3.1: Let ω_0 and ω_1 be two connection forms on Π and define $\omega_t = \omega_0 + t\alpha$, $\alpha = \omega_1 - \omega_0$, $0 \leqslant t \leqslant 1$; then the following holds.

(i) We find that α is an even horizontal one-form on Π of type (Ad, W_G).

(ii) We find that ω_t is a one parameter family of connection forms.

(iii) We find that $(d/dt)\Omega_t = d\alpha + [\omega_t, \alpha]$. ■

Another result we shall need is the following proposition.

Proposition 3.3:

$$P(\Omega_1^k) - P(\Omega_0^k) = k\, d \int_0^1 P(\alpha \wedge \Omega_t^{k-1}) dt. \tag{3.17}$$

Proof: By part (iii) of Lemma 3.1,

$$\frac{d}{dt} P(\Omega_t^k) = kP\left(\left(\frac{d}{dt}\Omega_t\right)\right) \wedge \Omega_t^{k-1} = kP(d\alpha \wedge \Omega_t^{k-1}) + kP([\omega_t, \alpha] \wedge \Omega_t^{k-1}).$$

On the other hand, using the Bianchi identity and Eq. (3.16),

$$k\, dP(\alpha \wedge \Omega_t^{k-1}) = kP(d\alpha \wedge \Omega_t^{k-1}) - k(k-1)P(\alpha \wedge d\Omega_t \wedge \Omega_t^{k-2}) = kP(d\alpha \wedge \Omega_t^{k-1}) + kP([\omega_t, \alpha] \wedge \Omega_t^{k-1})$$

and, integrating by parts over t, one obtains Eq. (3.17). ■

Proof of part (ii) of Proposition 3.2: The $(2k-1)$ form $\Phi = k \int_0^1 P(\alpha \wedge \Omega_t^{k-1}) dt$, being horizontal and invariant, projects onto a form $\overline{\Phi} \in \Lambda^{2k-1}(S)$ and Eq. (3.17) in turn projects onto

$$\overline{P(\Omega_1^k)} - \overline{P(\Omega_0^k)} = d\overline{\Phi}. \tag{3.18}$$

Finally, one shows directly that $\mathscr{W}: I_{\mathrm{Gr}}(W_G; B_L) \to H^*_{\mathrm{SDR}}(S)$ is a graded algebra homomorphism. ■

C. Transgression formula

The Weil homomorphism constructed in Sec. II B describes properties of the superbundle structure which do not depend upon the connection. However, the realization of the homomorphism in terms of the curvature Ω of the connection gives rise to new interesting objects. One indeed has the following propositions.

Proposition 3.4: The $2k$ form $P(\Omega^k)$ on Π is exact and

$$P(\Omega^k) = dT P(\omega), \tag{3.19}$$

with

$$TP(\omega) = k \int_0^1 P(\omega \wedge \Psi_t^{k-1}) dt, \quad \Psi_t = t\, d\omega + \frac{1}{2} t^2 [\omega, \omega]. \tag{3.20}$$

■

The proof, similar to that of Proposition 3.3, will be omitted.

We call Eq. (3.19) the *graded transgression formula* and the forms on Π given by $TP(\omega)$ for different P, *graded Chern–Simons forms*. These are invariant under the action of the supergroup G but are not horizontal, so that they do not project onto forms on the base supermanifold S.

IV. ELEMENTARY INVARIANT POLYNOMIALS ON gl($m+n$;C)

Let $C_L = B_L \otimes_{\mathrm{R}} \mathbb{C}$. The set of $(m+n) \times (m+n)$ matrices with entries in C_L, denoted by $\mathrm{gl}(m+n;\mathbb{C})$, is a free graded C_L module; its even part, denoted by $\mathrm{gl}(m,n;\mathbb{C})$, is formed by matrices having the structure given in Eq. (2.1), but with entries in $(C_L)_0$ or $(C_L)_1$. Here $\mathrm{gl}(m+n;\mathbb{C})$ is the Lie module of the super Lie group $\mathrm{GL}(m,n;\mathbb{C})$, which is the open subset of $\mathrm{gl}(m,n;\mathbb{C})$ whose elements are invertible matrices. The adjoint action of $\mathrm{GL}(m,n;\mathbb{C})$ over $\mathrm{gl}(m+n;\mathbb{C})$ is given as in the ordinary case by

$$\mathrm{Ad}_H X = HXH^{-1}, \quad H \in \mathrm{GL}(m,n;\mathbb{C}), \quad X \in \mathrm{gl}(m+n;\mathbb{C}).$$

Finally, we recall that the ordinary concept of trace is here replaced by the *supertrace*, which is defined as follows: If $X = \{X^B_A,\ A,B = 1,...,m+n\}$ is homogeneous [i.e., $X \in \mathrm{gl}(m+n;\mathbb{C})_i$ with $i = 0$ or $i = 1$], one defines

$$\mathrm{Str}\, X = \sum_{A=1}^{m+n} (-1)^{A(i+1)} X^A_A.$$

If X is not homogeneous, then $X = X_0 + X_1$ and $\mathrm{Str}\, X = \mathrm{Str}\,(X_0) + \mathrm{Str}(X_1)$. The supertrace has the following properties:

$$\mathrm{Str}(XY) = (-1)^{|X||Y|}\,\mathrm{Str}(YX),$$
$$\forall \text{ homogeneous } X,Y \in \mathrm{gl}(m+n;\mathbb{C}),$$
$$\mathrm{Str}(HXH^{-1}) = \mathrm{Str}\, X, \quad \forall X \in \mathrm{gl}(m+n;\mathbb{C}),$$
$$\forall H \in \mathrm{GL}(m,n;\mathbb{C}). \tag{4.1}$$

Using the supertrace we can construct functions on $\mathrm{gl}(m+n;\mathbb{C})$, invariant under the adjoint action of $\mathrm{GL}(m,n;\mathbb{C})$, which are polynomials in the entries of their argument. Such functions, which we shall call *elementary invariant polynomials*, are in a sense the "simplest" invariant functions on $\mathrm{gl}(m+n;\mathbb{C})$; however, we do not know whether these functions generate *all* invariant polynomials, as happens in the ordinary case. Indeed, in order to prove this one would need a spectral theory for matrices in $\mathrm{gl}(m+n;\mathbb{C})$, which is not available.

The elementary invariant polynomials are defined by

$$P^k(X) = \sum_{j=1}^{k} a_j (\mathrm{Str}\, X^j)^{k-j}, \tag{4.2}$$

where the coefficients a_j are the same that appear in the ordinary expression of the elementary invariant polynomials on the Lie algebra $\mathrm{gl}(N;\mathbb{C})$ (with $N = m+n$).[32] The Ad invariance of these polynomials is assured by Eq. (4.1). The first few polynomials are the following:

$$P^1(X) = \mathrm{Str}\, X, \quad P^2(X) = \tfrac{1}{2}[(\mathrm{Str}\, X)^2 - \mathrm{Str}\, X^2],$$
$$P^3(X) = \tfrac{1}{6}(\mathrm{Str}\, X)^3 - \tfrac{1}{2}(\mathrm{Str}\, X)^2(\mathrm{Str}\, X) + \tfrac{1}{3}\,\mathrm{Str}\, X^3.$$

Obviously, these polynomials are naturally defined on any subalgebra of $\mathrm{gl}(m+n;\mathbb{C})$. If X is even, i.e., $X \in \mathrm{gl}(m,n;\mathbb{C})$, the polynomials (4.2) can be given the following compact representation:

$$P^k(X) = \frac{1}{k!}\left[\frac{d^k}{dt^k}\,\mathrm{Sdet}(I + tX)\right]_{t=0},$$

where I is the $(m+n)\times(m+n)$ identity matrix, t is a real number, and the *superdeterminant* (or Berezinian) of matrices in $\mathrm{GL}(m,n;\mathbb{C})$ is defined as follows: If X has the form (2.1), with inverse

$$X^{-1} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix},$$

then

$$\mathrm{Sdet}\, X = (\det A)(\det D').$$

Now, let $\pi:\Pi \to S$ be a principal superfiber bundle, with structure supergroup $\mathrm{GL}(m,n;\mathbb{C})$ or one of its subgroups. Given a connection on Π with curvature Ω, we can construct the forms $P^k[(i/2\pi)\Omega]$. These forms can be written in terms of the *polarization* $\tilde{P}^k$ associated with the polynomial P^k. The polarization, which is defined as in the ordinary case,[32] is an element of $I_{\mathrm{Gr}}(W_G^k;C_L)$, with $W_G = \mathrm{gl}(m+n;\mathbb{C})$; one has

$$P^k[(i/2\pi)\Omega] = \tilde{P}^k([(i/2\pi)\Omega]^k),$$

so that the results of Sec. III apply to the forms $P^k[(i/2\pi)\Omega]$. Thus the cohomology classes in $H^*_{\mathrm{SDR}}(S) \otimes C_L$ represented by these forms are invariants associated with the superbundle. We define the k th *Chern class* of Π as

$$c_k(\Pi) = \{\overline{P^k[(i/2\pi)\Omega]}\} \in H^{2k}_{\mathrm{SDR}}(S) \otimes C_L. \tag{4.3}$$

The reader may find in Ref. 12 a more general definition of integer Chern classes of superbundles. To our knowledge it is not possible to prove that the cohomology class in (4.3) is B_L valued, since we are not able to construct on Π some analog of a Hermitian connection in terms of which $c_1(\Pi)$ would turn out to be B_L valued.

In accordance with the discussion of Sec. III C, we can also associate with the bundle Π and the connection ω transgression forms $T_k(\omega)$ such that

$$P^k[(i/2\pi)\Omega] = dT_k(\omega). \tag{4.4}$$

These are the Chern–Simons forms considered in Refs. 8–10.

V. EXAMPLE: THE GRASSMANN DIRAC MONOPOLE

In this section we give a simple example to illustrate some of the ideas developed previously. The example is the Grassmann version of the Hopf fibration as constructed in Ref. 17 in purely algebraic terms.

The total space of the fibration is the (1,2)-dimensional super Lie group UOSP(1,2), which can be realized as follows.[33,34] Let osp(1,2) be the Lie B_L superalgebra of dimension (3,2) with even generators $\{A_i, i = 1,2,3\}$ and odd generators $\{R_\alpha, \alpha = 1,2\}$ given by the matrix representation

$$A_1 = \frac{i}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad A_2 = \frac{i}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix},$$
$$A_3 = \frac{i}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad R_1 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$R_2 = \frac{1}{2}\begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \tag{5.1}$$

Moreover, let L be an even integer, introduce the complexified Grassmann algebra $C_L = B_L \otimes \mathbb{C}$, and consider a graded involution $^\diamond: C_L \to C_L$ verifying

$$(xy)^\diamond = x^\diamond y^\diamond, \quad |x^\diamond| = |x|, \quad x^{\diamond\diamond} = (-1)^{|x|}x$$
$$\forall \text{ homogeneous } x \in C_L.$$

The existence of such a map is assured by the fact that L is even.[34] Now, we introduce the Lie C_L superalgebra

$$W = C_L \otimes_{\mathbb{R}} \mathrm{osp}(1,2)$$

and define uosp(1,2) as the set of $X \in W$ which can be written as

$$X = a^i A_i + \eta R_1 + \eta^\diamond R_2, \quad a_i, \eta \in C_L, \quad a_i^\diamond = a_i.$$

Here uosp(1,2) is a subalgebra of $\mathrm{gl}(3+2)$. The super Lie group UOSP(1,2) is defined as the image of the exponential map exp: $\mathrm{gl}(3,2) \to \mathrm{GL}(3,2)$ restricted to $\mathrm{uosp}(1,2)_0$.

An arbitrary element $s \in \mathrm{UOSP}(1,2)$ can be parametrized as follows:

$$s=\begin{pmatrix} 1+\frac{1}{4}\eta^{\diamond}\eta & -\frac{1}{2}\eta^{\diamond} & \frac{1}{2}\eta \\ -\frac{1}{2}(z_0^{\diamond}\eta+z_1\eta^{\diamond}) & z_0^{\diamond}(1-\frac{1}{8}\eta^{\diamond}\eta) & z_1(1-\frac{1}{8}\eta^{\diamond}\eta) \\ \frac{1}{2}(z_1^{\diamond}\eta-z_0\eta^{\diamond}) & -z_1^{\diamond}(1-\frac{1}{8}\eta^{\diamond}\eta) & z_0(1-\frac{1}{8}\eta^{\diamond}\eta) \end{pmatrix}. \tag{5.2}$$

Here the elements $z_0, z_1 \in (C_L)_0$ satisfy $z_0 z_0^{\diamond} + z_1 z_1^{\diamond} = 1$ and $\eta \in (C_L)_1$.

The structure supergroup of the fibration is $\mathscr{U}(1)$, the Grassmann extension[35] of $U(1)$: It can be realized as

$$\mathscr{U}(1) = \{w \in (C_L)_0 \text{ s.t. } ww^{\diamond} = 1\}. \tag{5.3}$$

Here $\mathscr{U}(1)$ imbeds into UOSP(1,2) by

$$w \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & w & 0 \\ 0 & 0 & w^{\diamond} \end{pmatrix},$$

so that we can think of A_3 as the generator of $\mathscr{U}(1)$, i.e.,

$$\mathscr{U}(1) \simeq \{\exp(\lambda A_3) | \lambda \in (C_L)_0 \text{ s.t. } \lambda = \lambda^{\diamond}\}.$$

By taking the right action of $\mathscr{U}(1)$ on UOSP(1,2) one obtains a principal superfiber bundle

$$\begin{array}{l} \Pi \equiv \mathrm{UOSP}(1,2) \leftarrow \mathscr{U}(1) \\ \quad \pi \downarrow \\ \quad S^{2}_{\cdot} \end{array} \tag{5.4}$$

where $S^2_{\cdot} \equiv \mathrm{UOSP}(1,2)/\mathscr{U}(1)$. The projection π can be given explicitly as

$$\pi(s) = s[(2/i)A_3]s^{\dagger} \equiv x^k[(2/i)A_k] + \xi^{\alpha}(2R_{\alpha}), \tag{5.5}$$

where $s^{\dagger}$ is the adjoint of s.[33] It turns out that the x^k's are "real" even [i.e., $x^k \in (C_L)_0$ and $x^{k\diamond} = x^k$], the ξ^{α}'s are in $(C_L)_1$ and satisfy $\xi^2 = -\xi^{1\diamond}$, and the following constraint holds:

$$(x^1)^2 + (x^2)^2 + (x^3)^2 + 2\xi^1\xi^2 = 1. \tag{5.6}$$

On Π there is a natural connection: take the Maurer–Cartan form on UOSP(1,2),

$$\omega_0 = S^{\dagger}\,ds \equiv \theta^k A_k + \tilde{\theta}^{\alpha} R_{\alpha}; \tag{5.7}$$

then the component of ω_0 along A_3 is a connection form on Π given explicitly by

$$\begin{aligned} \omega &= \theta^3 A_3, \\ \theta^3 &= -2i(1-\tfrac{1}{4}\eta^{\diamond}\eta)(z_0^{\diamond}\,dz_0 + z_1^{\diamond}\,dz_1) \\ &\quad + (i/4)(\eta^{\diamond}\,d\eta + \eta\,d\eta^{\diamond}). \end{aligned} \tag{5.8}$$

As for the curvature, one has

$$\begin{aligned} \Omega &= [-\theta^1 \wedge \theta^2 - (i/2)\tilde{\theta}^1\tilde{\theta}^2]A_3 \\ &= [-2i(1-\tfrac{1}{4}\eta^{\diamond}\eta) \times dz_0^{\diamond} \wedge dz_0 + dz_1^{\diamond} \wedge dz_1) \\ &\quad - (i/2)(\eta^{\diamond}\,d\eta + \eta\,d\eta^{\diamond}) \\ &\quad \wedge (z_0^{\diamond}\,dz_0 + z_1^{\diamond}\,dz_1) + (i/2)d\eta^{\diamond} \wedge d\eta]A_3. \end{aligned} \tag{5.9}$$

The connection ω given by Eq. (5.8) is the Grassmann extension of a Dirac monopole. We shall show that Π is not trivial by checking that its first Chern class, which we shall compute in terms of ω, does not vanish. Let us denote by $\check{H}^*(X,\mathscr{F})$ the Čech cohomology ring of the topological space X with coefficients in the sheaf $\mathscr{F}$.

Proposition 5.1: We find that $S^2_{\cdot}$ is a De Witt supermanifold with body S^2 (the two-dimensional sphere). If Φ: $S^2_{\cdot} \to S^2$ is the bundle projection, $\Phi^*: \check{H}^*(S^2,\mathbf{Z}) \to \check{H}^*(S^2_{\cdot},\mathbf{Z})$ is an isomorphism.

Proof: It is evident that $S^2_{\cdot} \simeq \tilde{S}^2 \times B_L^{0,2}$, where $\tilde{S}^2$ is the unit sphere in $B_L^{3,0}$, i.e.,

$$\begin{aligned} \tilde{S} = \{&(x^1,x^2,x^3) \in B_L^{3,0} \text{ s.t.} \\ &\times (x^1)^2 + (x^2)^2 + (x^3)^2 = 1\}. \end{aligned}$$

Let τ: $\tilde{S}^2 \to S^2$ be the projection, given by $\tau((x^1,x^2,x^3)) = (\sigma(x^1),\sigma(x^2),\sigma(x^3))$, where S^2 is regarded as the unit sphere in $\mathbf{R}^3$. Here $\tilde{S}^2$ can be covered by the two open sets

$$U_+ = \tilde{S}^2 - \tau^{-1}((0,0,1)), \quad U_- = \tilde{S}^2 - \tau^{-1}((0,0,-1)),$$

both of which are homeomorphic to $B_L^{2,0}$ via the maps $\Psi_{\pm}$: $U_{\pm} \to B_L^{2,0}$ defined as follows:

$$\begin{aligned} \Psi_{\pm}(x^1,x^2,x^3) = \Bigg(&\sqrt{\frac{1\pm x^3}{1\mp x^3}}\,\frac{x^1}{\sqrt{(x^1)^2+(x^2)^2}} \\ &\times \sqrt{\frac{1\pm x^3}{1\mp x^3}}\,\frac{x^2}{\sqrt{(x^1)^2+(x^2)^2}}\Bigg). \end{aligned}$$

The transition function $\psi_+ \circ \psi_-^{-1}: B_L^{2,0} - (\sigma^{2,0})^{-1}(0) \to B_L^{2,0} - (\sigma^{2,0})^{-1}(0)$ is the map

$$(z^1,z^2) \mapsto \left(\frac{z^1}{(z^1)^2+(z^2)^2}, \frac{z^2}{(z^2)^2+(z^2)^2}\right),$$

which is a $\mathscr{G}$ map. This proves that $\tilde{S}^2$ is a De Witt supermanifold of dimension (2,0). The second part of Proposition 5.1 is a consequence of the fact that $S^2_{\cdot}$, being De Witt, is a locally trivial bundle over S^2 with vector space fiber. ■

We also introduce the morphism

$$j: \check{H}^2(S^2_{\cdot},\mathbf{Z}) \to H^2_{\mathrm{SDR}}(S^2_{\cdot}), \tag{5.10}$$

which is induced by the morphism $\mathbf{Z} \to B_L$ of constant sheaves over $S^2_{\cdot}$ and by the isomorphism $H^*_{\mathrm{SDR}}(S^2_{\cdot}) \simeq \check{H}^*(S^2_{\cdot},B_L)$. The morphism j is injective since the exact sequence of constant sheaves of $S^2_{\cdot}$,

$$0 \to \mathbf{Z} \to B_L \to \mathscr{U}(1) \times (B_L)_1 \to 0,$$

induces in cohomology a long exact sequence[32] which splits into several pieces, among which one has

$$\begin{aligned} &0 \to \check{H}^2(S^2_{\cdot},\mathbf{Z}) \xrightarrow{j} \check{H}^2(S^2_{\cdot},B_L) \\ &\to \check{H}^2 \times (S^2_{\cdot}, \mathscr{U}(1) \times (B_L)_1) \to 0. \end{aligned}$$

After this preparatory material, we are ready to prove that $c_1(\Pi) \neq 0$. The $\mathscr{U}(1)$ superbundle Π over $S^2_{\cdot}$ can be identified with an element of the cohomology group $\check{H}^1(S^2_{\cdot},\mathscr{F}_1)$, where $\mathscr{F}_1$ is the sheaf of $\mathscr{G}$ maps $S^2_{\cdot} \to \mathscr{U}(1)$. The exactness of the sequence of sheaves over $S^2_{\cdot}$,

$$0 \to \mathbf{Z} \to \mathscr{G}_0 \to \mathscr{F}_1 \to 0,$$

where $\mathscr{G}_0$ is the even part of the structure sheaf of $S^2_{\cdot}$, and

the triviality of the cohomology of $\mathcal{G}_0$, due to the fact that S^2_* is De Witt, give an isomorphism

$$\delta:\check{H}^1(S^2_*,\mathcal{F}_1)\to\check{H}^2(S^2_*,\mathbf{Z})$$

[here we have used the canonical identification of $\mathcal{G}_0$ with the sheaf of germs of $\mathcal{G}$ maps $S^2_*:(B_L)_0$(Ref. 23)]. The element $\delta(\Pi)$ in $\check{H}^2(S^2_*,\mathbf{Z})$ is called the *obstruction class* of Π and has the property[27,36]

$$-j\circ\delta(\Pi)=[(i/2\pi)\Omega]_{\mathrm{SDR}}\equiv c_1(\Pi). \tag{5.11}$$

We wish to show that for the superbundle Π previously constructed, one has

$$\delta(\Pi)=-1 \tag{5.12}$$

[notice that $\check{H}^2(S^2_*,\mathbf{Z})\simeq\check{H}^2(S^2,\mathbf{Z})=\mathbf{Z}$]. Thus we have the following result of the Gauss-Bonnet type:

$$j(1)=c_1(\Pi),$$

which, together with the injectivity of j, proves that Π is not trivial.

Let f: $S^2\to S^2_*$ be a global cross section of the bundle $S^2_*\to S^2$. The pullback $f^{-1}\Pi$ is a principal fiber bundle on S^2, with the structure supergroup $\mathcal{U}(1)$, and there is a bundle map $\bar{f}:f^{-1}\Pi\to\Pi$ such that the diagram

$$\begin{array}{ccc} f^{-1} & \xrightarrow{\bar{f}} & \Pi \\ \downarrow & & \downarrow \pi \\ S^2 & \xrightarrow[f]{} & S^2_* \end{array} \tag{5.13}$$

commutes. Let σ: $\mathcal{U}(1)\to U(1)$ be the restriction of the body map to $\mathcal{U}(1)$. Composing the transition functions of $f^{-1}\Pi$ with σ one obtains a $U(1)$ bundle Q over S^2, which is nothing but a Hopf fibration. Taking the body of the pullback $\bar{f}^*\omega$ of the connection (2.8) one obtains a connection over Q, which is a Dirac monopole of the lowest strength. Therefore, the obstruction class of Q is -1. Then the result (5.12) is a consequence of the following proposition.

Proposition 5.2: The bundles Π, $f^{-1}\Pi$, and Q have the same obstruction class.

Proof: First we prove $\delta(\Pi)=\delta(f^{-1}\Pi)$. This is equivalent to the commutativity of the diagram

$$\begin{array}{ccc} \check{H}^1(S^2_*),\mathcal{F}_1) & \longrightarrow & \check{H}^2(S^2_*,\mathbf{Z}) \\ f^* \downarrow & & \downarrow f^*\equiv\Phi^{*-1}, \\ \check{H}^1(S^2,F_1) & \longrightarrow & \check{H}^2(S^2,\mathbf{Z}) \end{array}$$

where F_1 is the sheaf of smooth maps $S^2\to U(1)$. The equality $f^*\equiv\Phi^{*-1}$ follows from the fact that $f\circ\Phi$ is homotopic to the identity map.

Moreover, the commutativity of the diagram

$$\begin{array}{ccc} \check{H}^1(S^2,\mathcal{F}_1) & \xrightarrow{\partial} & \check{H}^2(S^2,\mathbf{Z}) \\ \sigma \downarrow & & \downarrow id \\ \check{H}^1(S^2,F_1) & \xrightarrow[\partial]{} & \check{H}^2(S^2,\mathbf{Z}) \end{array}$$

implies $\delta(f^{-1}\Pi)=\delta(Q)$. Now, the connecting morphisms ∂: $\check{H}^1(S^2,F_1)\to\check{H}^2(S^2,\mathbf{Z})$ and ∂: $\check{H}^1(S^2,F_1)\to\check{H}^2(S^2,\mathbf{Z})$ have the following form: If $\mathcal{V}=\{V_\alpha\}$ is a cover of S^2 and $g=\{g_{\alpha\beta}\}$ is an element in $\check{H}^1(\mathcal{V},\mathcal{F}_1)$ or in $\check{H}^1(\mathcal{V},F_1)$, then $\partial(g)=h=\{h_{\alpha\beta\gamma}\}$, with

$$h_{\alpha\beta\gamma}=(1/2\pi i)(\log g_{\alpha\beta}+\log g_{\beta\gamma}+\log g_{\gamma\alpha}).$$

Then the equality $\partial\circ\sigma=\partial$ is equivalent to the following algebraic result: If a, b, $c\in(C_L)_0$ with $abc=1$,

$$\log a+\log b+\log c=\log\sigma(a)+\log\sigma(b)+\log\sigma(c).$$

■

ACKNOWLEDGMENTS

We thank Professor G. Marmo for helpful discussions.

Research partially supported by the National Group for Mathematical Physics of the Italian Research Council, by the Italian Ministry for Public Education through the research project "Geometria e Fisica," and by Istituto Nazionale per la Fisica Nucleare.

[1]S. S. Chern and J. Simons, Ann. Math. **99**, 48 (1974).
[2]R. Jackiw, in *Proceedings of the Santa Fe Meeting of the Division of Particles and Fields of the American Physical Society, 1984,* edited by T. Goldman and M. M. Nieto (World Scientific, Singapore, 1985).
[3]S. Deser, R. Jackiw, and S. Templeton, Ann. Phys. **140**, 372 (1982).
[4]B. Zumino, *Relativity, Groups and Topology*, edited by B. S. De Witt and R. Stora (North-Holland, Amsterdam, 1984).
[5]R. Stora, in *Progress in Gauge Field Theory*, NATO ASI Series B, Vol. 115, edited by G. t'Hooft, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer, and R. Stora (Plenum, New York, 1984).
[6]M. Dubois-Violette, M. Talon, and C. M. Viallet, Commun. Math. Phys. **102**, 105 (1985).
[7]M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 117 (1984).
[8]P. Van Nieuwenhuizen, in *Proceedings of the Santa Fe Meeting of the Division of Particles and Fields of the American Physical Society, 1984,* edited by T. Goldman and M. M. Nieto (World Scientific, Singapore, 1985).
[9]A. Ceresole, A. Lerda, and P. Van Nieuwenhuizen, Phys. Rev. D **34**, 1744 (1986).
[10]G. Girardi and R. Grimm, Nucl. Phys. B **292**, 181 (1987).
[11]L. Bonora, P. Pasti, and M. Tonin, Nucl. Phys. B **286**, 150 (1987).
[12]U. Bruzzo and D. Hernández Ruipérez, J. Math. Phys. **30**, 1233 (1989).
[13]U. Bruzzo and G. Landi, Phys. Rev. D **39**, 1174 (1989).
[14]C. Bartocci, U. Bruzzo, and G. Landi, Lett. Math. Phys. **18**, 235 (1989).
[15]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York 1969), Vol. II.
[16]G. Landi and G. Marmo, Phys. Lett. B **192**, 81 (1987).
[17]G. Landi and G. Marmo, Phys. Lett. **193**, 61 (1987).
[18]L. Bonora, P. Cotta-Ramusino, M. Rinaldi, and J. Stasheff, Commun. Math. Phys. **112**, 237 (1987); Commun. Math. Phys. **114**, 381 (1988).
[19]B. De Witt, *Supermanifolds* (Cambridge U.P., London, 1984).
[20]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[21]A. Rogers, Commun. Math. Phys. **105**, 375 (1986).
[22]C. Bartocci and U. Bruzzo, J. Geom. Phys. **4**, 391 (1987).
[23]C. Bartocci, U. Bruzzo, and D. Hernández Ruipérez, J. Geom. Phys., in press (1989); Preprint 105/1989, Dipartimento di Matematica, Universitá di Genova.
[24]M. Rothstein, Trans. AMS **297**, 159 (1986).
[25]M. Batchelor, in *Mathematical Aspects of Superspace*, edited by C. J. S. Clarke, A. Rosenblum, and H.-J. Seifert (Reidel, Dordrecht, 1984); A. Rogers, J. Math. Phys. **26**, 2749 (1985).
[26]C. Bartocci and U. Bruzzo, J. Math. Phys. **28**, 2363 (1987).
[27]C. Bartocci and U. Bruzzo, J. Math. Phys. **29**, 1789 (1988).
[28]A. Rogers, J. Math. Phys. **22**, 939 (1981).
[29]M. F. Atiyah and R. Bott, Philos. Trans. R. Soc. London Ser. A **308**, 523

(1982); A. López Almorox, "Supergauge theories in graded manifolds," in *Differential Geometric Methods in Mathematical Physics*, edited by A. Pérez-Rendón and P. L. García, Lecture Notes in Math. **1251** (Springer, Berlin, 1987).
[30]U. Bruzzo and R. Cianci, Class. Quantum Grav. **1**, 213 (1984).
[31]C. Bartocci and U. Bruzzo, Lett. Math. Phys. **17**, 61 (1989).
[32]P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1978).
[33]F. A. Berezin and V. N. Tolstoy, Commun. Math. Phys. **78**, 409 (1981).
[34]V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 713 (1978).
[35]L. Bonora, P. Pasti, and M. Tonin, J. Math. Phys. **23**, 839 (1982).
[36]C. Bartocci and U. Bruzzo, Lett. Math. Phys. **17**, 263 (1989).

Space-time geometry of relativistic particles

Y. S. Kim
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
E. P. Wigner
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544

(Received 20 July 1989; accepted for publication 6 September 1989)

A three-dimensional space-time geometry of relativistic particles is constructed within the framework of the little groups of the Poincaré group. Since the little group for a massive particle is the three-dimensional rotation group, its relevant geometry is a sphere. For massless particles and massive particles in the infinite-momentum limit, it is shown that the geometry is that of a cylinder and a two-dimensional plane. The geometry of a massive particle continuously becomes that of a massless particle as the momentum/mass becomes large. The geometry of relativistic extended particles is also considered. It is shown that the cylindrical geometry leads to the concept of gauge transformations, while the two-dimensional Euclidean geometry leads to a deeper understanding of the Lorentz condition.

I. INTRODUCTION

The internal space-time symmetries of relativistic particles are governed by the little groups of the Poincaré group.[1,2] The internal space-time symmetry group for massive and massless particles are isomorphic to the three-dimensional rotation group and the two-dimensional Euclidean group, respectively. We have shown in our previous paper[3] that the internal space-time symmetry of massless particles is dictated by the cylindrical group, which is isomorphic to the Euclidean group. The cylindrical axis is parallel to the momentum. For the case of electromagnetic fields satisfying the Lorentz condition, the rotation around the axis corresponds to helicity, while the translation on the surface of the cylinder along the direction of the axis corresponds to a gauge transformation.[4]

The purpose of the present paper is to present a more complete geometrical picture of relativistic particles. Since the little groups for massive and massless particles are three-parameter groups,[1] it is possible to construct a three-dimensional geometry of internal space-time symmetries for all relativistic particles starting from a sphere for a massive particle at rest. It was observed in Ref. 3 that the three-dimensional rotation group can be contracted either to the two-dimensional Euclidean group or to the cylindrical group.[3,5] In the present paper, we point out first that both the cylindrical and Euclidean geometries are needed for the little group for massless particles.[3,6]

We shall then show that the Euclidean geometry leads to a deeper understanding of the Lorentz condition applicable to massless particles and to massive particles in the infinite-momentum limit. It is then shown that the cylindrical symmetry is shared by all those particles, even without the requirement of the Lorentz condition. This means that the concept of gauge transformation can be extended to all massless particles or massive particles with infinite momentum.

Also in this paper, we shall discuss relativistic extended particles often called hadrons. It is not difficult to visualize the symmetry of an extended particle as the three-dimensional rotation group.[7] However, it is not trivial to construct the geometry of a relativistic extended particle or hadron if it moves with a speed close to that of light. We attack this problem by constructing the generators of the little groups in differential form and the wave functions to which these operators are applicable.

In Sec. II, we discuss the three-dimensional rotation group and its contractions to the cylindrical and the two-dimensional Euclidean group. It is shown that both of these contractions can be combined into a single representation. In Sec. III, the generators of the little group are discussed in the light-cone coordinate system. It is shown that these generators are identical to the combined geometry of the cylindrical group and the Euclidean group discussed in Sec. II.

In Sec. IV, we show that the Lorentz condition is not a prerequisite for the cylindrical symmetry and that the Euclidean symmetry replaces the role of the Lorentz condition. In Sec. V, the formalism developed in Secs. II–IV is applied to the space-time geometry of relativistic extended hadrons. It is shown that the relativistic hadron can be described in terms of the parameters of the cylindrical group. Feynman's parton picture is discussed as an illustrative example.

II. THREE-DIMENSIONAL GEOMETRY OF THE LITTLE GROUPS

It is not difficult to construct the geometry of the little group for a massive particle at rest.[1] It is the three-dimensional rotation group whose generators L_i satisfy the commutation relations

$$[L_i,L_j] = i\varepsilon_{ijk}L_k. \tag{2.1}$$

Transformations applicable to the coordinate variables x, y, and z are generated by

$$L_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix},$$
$$L_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \tag{2.2}$$
$$L_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

 0022-2488/90/010055-06$03.00

In our previous paper,[3] we have shown that this group can be contracted either to the cylindrical group or the two-dimensional Euclidean group. In either case, we can start from a sphere. The contraction to the two-dimensional Euclidean group can be achieved through a plane tangent to the sphere at the north pole.[4] The contraction to the cylindrical group corresponds to the cylinder that makes contact with the sphere at the equatorial belt.[3]

The Euclidean group is generated by L_3, P_1, and P_2, where

$$P_1=\begin{pmatrix}0&0&i\\0&0&0\\0&0&0\end{pmatrix},\quad P_2=\begin{pmatrix}0&0&0\\0&0&i\\0&0&0\end{pmatrix},\tag{2.3}$$

and they satisfy the commutation relations

$$[P_1,P_2]=0,\quad [L_3,P_1]=iP_2,\quad [L_3,P_2]=-iP_1.\tag{2.4}$$

The cylindrical group is generated by L_3, Q_1, and Q_2, where

$$Q_1=\begin{pmatrix}0&0&0\\0&0&0\\i&0&0\end{pmatrix},\quad Q_2=\begin{pmatrix}0&0&0\\0&0&0\\0&i&0\end{pmatrix}.\tag{2.5}$$

These generators satisfy the same set of commutation relations as that for L_3, P_1, and P_2 given in Eq. (2.4):

$$[Q_1,Q_2]=0,\quad [L_3,Q_1]=iQ_2,\quad [L_3,Q_2]=-iQ_1.\tag{2.6}$$

We achieve the contractions to the Euclidean and cylindrical groups by taking the large-radius limits of

$$\begin{aligned}&P_1=(1/R)B^{-1}(L_2)B,\quad P_2=-(1/R)B^{-1}(L_1)B,\\&Q_1=-(1/R)B(L_2)B^{-1},\\&Q_2=(1/R)B(L_1)B^{-1},\end{aligned}\tag{2.7}$$

where

$$B(R)=\begin{pmatrix}1&0&0\\0&1&0\\0&0&R\end{pmatrix}.\tag{2.8}$$

The vector spaces to which the above generators are applicable are $(x,y,z/R)$ and (x,y,Rz) for the Euclidean and cylindrical groups, respectively.

In differential forms, the generators of the rotation group can be written as

$$\begin{aligned}&L_1=-i\left(y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right),\quad L_2=-i\left(z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right),\\&L_3=-i\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right),\end{aligned}\tag{2.9}$$

applicable functions of x, y, and z. The $B(R)$ transformation applicable to these operators is

$$B(R)=\exp\left(-\rho z\frac{\partial}{\partial z}\right),\tag{2.10}$$

where $\rho=\ln(R)$. This operator commutes with L_3. The application of this formula to Eq. (2.9) in the large-R limit leads to

$$\begin{aligned}&P_1=-i\frac{\partial}{\partial x},\quad P_2=-i\frac{\partial}{\partial y}.\\&Q_1=-i\left(\frac{x}{R}\right)\frac{\partial}{\partial z},\quad Q_2=-i\left(\frac{y}{R}\right)\frac{\partial}{\partial z}.\end{aligned}\tag{2.11}$$

Since P_1 (P_2) commutes with Q_2 (Q_1), we can consider the following combination of generators:

$$F_1=P_1+Q_1,\quad F_2=P_2+Q_2.\tag{2.12}$$

Then these operators also satisfy the commutation relations:

$$[F_1,F_2]=0,\quad [L_3,F_1]=iF_2,\quad [L_3,F_2]=-iF_1.\tag{2.13}$$

On the other hand, this is not true if we add the three-by-three matrices for P_i and Q_i to construct three-by-three matrices for F_1 and F_2. This is due to the fact that the vector spaces are different for the P_i and Q_i representations. We can accommodate this difference by creating two different z coordinates, one with a contracted z and the other with an expanded z, namely, $(x,y,Rz,z/R)$. Then the generators become

$$P_1=\begin{pmatrix}0&0&0&i\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\quad P_2=\begin{pmatrix}0&0&0&0\\0&0&0&i\\0&0&0&0\\0&0&0&0\end{pmatrix},\tag{2.14}$$

$$Q_1=\begin{pmatrix}0&0&0&0\\0&0&0&0\\i&0&0&0\\0&0&0&0\end{pmatrix},\quad Q_2=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&i&0&0\\0&0&0&0\end{pmatrix}.$$

Then F_1 and F_2 will take the form

$$F_1=\begin{pmatrix}0&0&0&i\\0&0&0&0\\i&0&0&0\\0&0&0&0\end{pmatrix},\quad F_2=\begin{pmatrix}0&0&0&0\\0&0&0&i\\0&i&0&0\\0&0&0&0\end{pmatrix}.\tag{2.15}$$

The contraction and expansion of the z axis are illustrated in Fig. 1.

Next, let us consider the transformation matrix generated by the above matrices. It is easy to visualize the transformations generated by P_i and Q_i. It would be easy to visualize the transformation generated by F_1 and F_2, if P_i commuted with Q_i. However, P_i and Q_i do not commute with each other. Thus the transformation matrix takes a somewhat complicated form:

$$\exp(-i(\xi F_1+\eta F_2))=\begin{pmatrix}1&0&0&\xi\\0&1&0&\eta\\\xi&\eta&1&(\xi^2+\eta^2)/2\\0&0&0&1\end{pmatrix}.\tag{2.16}$$

If we make a similarity transformation on the above form

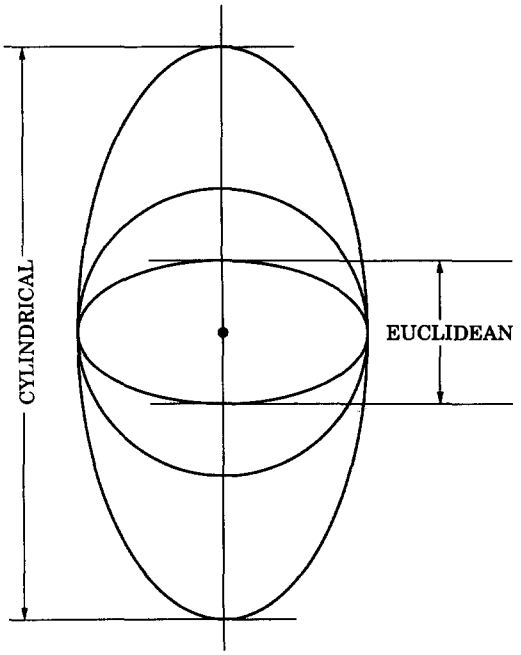


FIG. 1. Cylindrical and Euclidean deformations of the sphere. It is possible to contract the z axis by dividing it by R. This contraction of the z axis leads to the contraction of O(3) to the two-dimensional Euclidean group. If the z axis is multiplied by R, then it becomes expanded. This expansion of the z axis leads to the contraction of O(3) to the cylindrical group. The expanding and contracting z axes are treated as different coordinates, and are called the u and v coordinates, respectively, in Secs. III–V.

using the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}, \tag{2.17}$$

then $\exp(-i(\xi F_1 + \eta F_2))$ of Eq. (2.16) becomes

$$\begin{pmatrix} 1 & 0 & -\xi/\sqrt{2} & \xi/\sqrt{2} \\ 0 & 1 & -\eta/\sqrt{2} & \eta/\sqrt{2} \\ \xi/\sqrt{2} & \eta/\sqrt{2} & 1-(\xi^2+\eta^2)/4 & (\xi^2+\eta^2)/4 \\ \xi/\sqrt{2} & \eta/\sqrt{2} & -(\xi^2+\eta^2)/4 & 1+(\xi^2+\eta^2)/4 \end{pmatrix}. \tag{2.18}$$

This form is readily available in the literature[1,4] as the translationlike transformation matrix for the little group for massless particles. In this section, we have given a geometrical interpretation to this matrix.

III. LITTLE GROUPS IN THE LIGHT-CONE COORDINATE SYSTEM

Let us now study the group of Lorentz transformations using the light-cone coordinate system. If the space-time metric coordinate is specified by (x,y,z,t), then the light-cone coordinate variables are (x,y,u,v) for a particle moving along the z direction, where

$$u = (z+t)/\sqrt{2}, \quad v = (z-t)/\sqrt{2}. \tag{3.1}$$

The generators of Lorentz transformations are then

$$J_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & i \\ 0 & i & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix},$$

$$K_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix},$$

$$J_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & i & -i \\ 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \tag{3.2}$$

$$K_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix},$$

$$J_3 = \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}.$$

For J_1, J_2, and J_3, we can consider the three-by-three matrices consisting of the first three rows and columns. Then they are clearly the generators of the rotation group. The set of three-by-three matrices consisting of the first, second, and fourth rows and columns also constitutes the set of rotation generators.

If a massive particle is at rest, its little group is generated by J_1, J_2, and J_3. For a massless particle, the little group is generated by J_3, N_1, and N_2, where

$$N_1 = (K_1 - J_2), \quad N_2 = (K_2 + J_1), \tag{3.3}$$

which can be written in the matrix form as

$$N_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$N_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.4}$$

These matrices satisfy the commutation relations:

$$[J_3,N_1] = iN_2, \quad [J_3,N_2] = -iN_1, \quad [N_1,N_2] = 0. \tag{3.5}$$

Let us go back to F_1 and F_2 of Eq. (2.15). Indeed, they are proportional to N_1 and N_2, respectively:

$$N_1 = (1/\sqrt{2})F_1, \quad N_2 = (1/\sqrt{2})F_2. \tag{3.6}$$

Since F_1 and F_2 are somewhat simpler than N_1 and N_2, and since the commutation relations of Eq. (3.5) are invariant under multiplication of N_1 and N_2 by constant factors, we shall hereafter use F_1 and F_2 for N_1 and N_2.

In the light-cone coordinate system, the boost matrix

along the z direction takes the form

$$B(R)=\exp(-i\rho K_3)=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&R&0\\0&0&0&1/R\end{pmatrix}, \quad (3.7)$$

with $\rho=\ln(R)$, and $R=((1+\beta)/(1-\beta))^{1/2}$, where β is the velocity parameter of the particle. Under this transformation, x and y coordinates are invariant, and the light-cone variables u and v are transformed as

$$u'=Ru, \quad v'=v/R. \quad (3.8)$$

If we boost J_2 and J_1 and multiply them by $\sqrt{2}/R$, as

$$W_1(R)=-\frac{\sqrt{2}}{R}BJ_2B^{-1}=\begin{pmatrix}0&0&-i/R^2&i\\0&0&0&0\\i&0&0&0\\-i/R^2&0&0&0\end{pmatrix},$$
$$W_2(R)=\frac{\sqrt{2}}{R}BJ_1B^{-1}=\begin{pmatrix}0&0&0&0\\0&0&-i/R^2&i\\0&i&0&0\\0&-i/R^2&0&0\end{pmatrix}, \quad (3.9)$$

then $W_1(R)$ and $W_2(R)$ become F_1 and F_2, respectively, in the large-R limit.

The algebra given in this section is identical with that of Sec. II based on the three-dimensional geometry of a sphere going through a contraction/expansion of the z axis. Therefore, it is possible to give a concrete geometrical picture to the little groups of the Poincaré group governing the internal space-time symmetries of relativistic particles.

The most general form of the transformation matrix is

$$D(\xi,\eta,\alpha)=D(\xi,\eta,0)D(0,0,\alpha), \quad (3.10)$$

where

$$D(\xi,\eta,0)=\exp(-i(\xi F_1+\eta F_2)),$$
$$D(0,0,\alpha)=\exp(-i\alpha J_3).$$

Here, $D(0,0,\alpha)$ represents a rotation around the z axis, and does not need further explanation. In the light-cone coordinate system, $D(\xi,\eta,0)$ takes the form of Eq. (2.16). It is then possible to decompose it into

$$D(\xi,\eta,0)=C(\xi,\eta)E(\xi,\eta)S(\xi,\eta), \quad (3.11)$$

where

$$C(\xi,\eta)=\exp(-i\xi Q_1-i\eta Q_2)=\begin{pmatrix}1&0&0&0\\0&1&0&0\\\xi&\eta&1&0\\0&0&0&1\end{pmatrix}, \quad (3.12)$$

$$E(\xi,\eta)=\exp(-i\xi P_1-i\eta P_2)=\begin{pmatrix}1&0&0&\xi\\0&1&0&\eta\\0&0&1&0\\0&0&0&1\end{pmatrix}, \quad (3.13)$$

$$S(\xi,\eta)=I+\tfrac{1}{2}[C(\xi,\eta),E(\xi,\eta)]=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&(\xi^2+\eta^2)/2\\0&0&0&1\end{pmatrix}. \quad (3.14)$$

The matrix $C(\xi,\eta)$ performs a cylindrical transformation on the first, second, and third components, while $E(\xi,\eta)$ is for a Euclidean transformation on the first, second, and fourth components. The matrix $S(\xi,\eta)$ performs a translation along the third axis and commutes with both $C(\xi,\eta)$ and $E(\xi,\eta)$. As we noted in Ref. 3, both $E(\xi,\eta)$ and $S(\xi,\eta)$ become identity matrices when applied to four-vectors satisfying the Lorentz condition which have a vanishing fourth component.[3]

IV. CYLINDRICAL GROUP AND GAUGE TRANSFORMATIONS

In order to illustrate the transformation property of the vector to which the above matrices are applicable, let us consider a particle represented by a four-vector:

$$A^{\mu}(x)=A^{\mu}e^{i(kz-\omega t)}, \quad (4.1)$$

where $A^{\mu}=(A_1,A_2,A_3,A_0)$. In the light-cone coordinate system,

$$A^{\mu}=(A_1,A_2,A_u,A_v), \quad (4.2)$$

where $A_u=(A_3+A_0)/\sqrt{2}$, and $A_v=(A_3-A_0)/\sqrt{2}$. If it is boosted by the matrix of Eq. (3.7), then

$$A'^{\mu}=(A_1,A_2,RA_u,A_v/R). \quad (4.3)$$

Thus the fourth component will vanish in the large-R limit, while the third component becomes large.

The momentum–energy four-vector is

$$P^{\mu}=(0,0,(k+\omega)/\sqrt{2},(k-\omega)/\sqrt{2}), \quad (4.4)$$

which in the rest frame becomes

$$P^{\mu}=(0,0,m/\sqrt{2},-m/\sqrt{2}), \quad (4.5)$$

where m is the mass. If we boost this four-momentum using the matrix of Eq. (3.7), then

$$P'^{\mu}(0,0,Rm/\sqrt{2},-m/\sqrt{2}R). \quad (4.6)$$

Here again, the fourth component vanishes for large values of R, while the third component becomes large.

Let us go back to $W_1(R)$ and $W_2(R)$ of Eq. (3.9). If $W_1(R)$ is applied to the four-vector A'^{μ}, the result is

$$i((A_u-A_v)/R,0,A_1,-A_1/R^2), \quad (4.7)$$

which becomes $(0,0,-iA_1,0)$. When $W_2(R)$ is applied, the result is $(0,0,-iA_2,0)$. Thus the i/R^2 factors in W_1R^2 and $W_2(R)$ can be dropped in the large-R limit. We can thus safely apply the transformation matrix generated by F_1 and F_2.

Since the fourth component of the vector vanishes or becomes vanishingly small, the application of $S(\xi,\eta)$ of Eq. (3.14) on A'^{μ} and P'^{μ} will produce no effects in the large-R limit. The same is true for $E(\xi,\eta)$ of Eq. (3.13). Thus, among the three factors of the transformation matrix, only the matrix $C(\xi,\eta)$ given in Eq. (3.12) will produce a nontrivial effect. This is the cylindrical transformation discussed in Ref. 3.

During the limiting process, the three-dimensional geometry consisting of the x, y, and v coordinates describes a pancakelike compression of the sphere in which the v coordinate shrinks to zero, as is indicated in Fig. 1. Because of this contraction of the v coordinate, the Euclidean component of the little group disappears. This is the content of the Lorentz condition for massive particles in the infinite-momentum limit. The three-dimensional geometry of the x, y, and u coordinates corresponds to the expanding z coordinate, resulting in the cylindrical symmetry, as is indicated in Fig. 1.

Let us see the effect of $C(\xi,\eta)$ on the four-vector of Eq. (4.3). If we apply $C(\xi,\eta)$ to the four-vector, then

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \xi & \eta & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ RA_u \\ A_v/R \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ RA_u + \xi A_1 + \eta A_2 \\ A_v/R \end{pmatrix}. \tag{4.8}$$

This is not unlike the $D(\xi,\eta)$ transformation applied to the four-vector satisfying the Lorentz condition $A_v = 0$:

$$\begin{pmatrix} 1 & 0 & 0 & \xi \\ 0 & 1 & 0 & \eta \\ \xi & \eta & 1 & (\xi^2+\eta^2)/2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ RA_u \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \xi & \eta & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ RA_u \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} A_1 \\ A_2 \\ RA_u + \xi A_1 + \eta A_2 \\ 0 \end{pmatrix}. \tag{4.9}$$

As we noted at the end of Sec. III, the Lorentz condition eliminates the Euclidean component in the $D(\xi,\eta,0)$ matrix. It is remarkable that Eq. (4.9) is strikingly similar to Eq. (4.8). The cylindrical transformation is quite independent of the fourth component in both cases, and it produces the same result for the first three components. Thus the elimination of the Euclidean component that led to Eq. (4.8) can thus be regarded as an extension of the Lorentz condition to all four-vectors.

V. LITTLE GROUPS FOR RELATIVISTIC EXTENDED PARTICLES

We are now ready to discuss the symmetry property discussed in Sec. III for relativistic extended particles or hadrons. Let us consider a hadron consisting of two quarks bound together by an attractive force such as the harmonic oscillator force. We use four-vectors x_a and x_b to specify space-time positions of the two quarks. Then it is more convenient to use the following variables[8]:

$$X = (x_a + x_b)/2, \quad x = (x_a - x_b)/2\sqrt{2}. \tag{5.1}$$

The four-vector X specifies where the hadron is located in space-time, while the variable x measures the space-time separation between the quarks.

In the light-cone coordinate system, the generators of rotations applicable to functions localized in the four-dimensional space-time of x are

$$J_1 = -\frac{i}{\sqrt{2}}\left(y\left(\frac{\partial}{\partial u}+\frac{\partial}{\partial v}\right) - (u+v)\frac{\partial}{\partial y}\right),$$
$$J_2 = \frac{i}{\sqrt{2}}\left(x\left(\frac{\partial}{\partial u}+\frac{\partial}{\partial v}\right) - (u+v)\frac{\partial}{\partial x}\right), \tag{5.2}$$
$$J_3 = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial z}\right).$$

The boost generators are

$$K_1 = \frac{i}{\sqrt{2}}\left(x\left(\frac{\partial}{\partial u}-\frac{\partial}{\partial v}\right) + (u-v)\frac{\partial}{\partial x}\right),$$
$$K_2 = \frac{i}{\sqrt{2}}\left(y\left(\frac{\partial}{\partial u}-\frac{\partial}{\partial v}\right) + (u-v)\frac{\partial}{\partial y}\right), \tag{5.3}$$
$$K_3 = -i\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right).$$

These generators do not contain the hadronic coordinate variable X, as transformations of the little group do not change the hadronic momentum.

The boost operator along the z direction is

$$B(R) = \exp\left(-\rho\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)\right). \tag{5.4}$$

If this boost is applied to J_2 and J_1, as in the case of Eq. (3.9),

$$W_1(R) = -i\left(x\frac{\partial}{\partial u} - v\frac{\partial}{\partial x} - \left(\frac{1}{R}\right)^2\left(u\frac{\partial}{\partial x} - x\frac{\partial}{\partial v}\right)\right), \tag{5.5}$$
$$W_2(R) = -i\left(y\frac{\partial}{\partial u} - v\frac{\partial}{\partial y} - \left(\frac{1}{R}\right)^2\left(u\frac{\partial}{\partial y} - y\frac{\partial}{\partial v}\right)\right).$$

In the limit of large R, W_1 and W_2 become F_1 and F_2, respectively[9]:

$$F_1 = -i\left(x\frac{\partial}{\partial u} - v\frac{\partial}{\partial x}\right), \quad F_2 = -i\left(y\frac{\partial}{\partial u} - v\frac{\partial}{\partial y}\right). \tag{5.6}$$

The transformation operator is now

$$D(\xi,\eta,0) = \exp\left(-i(\xi x+\eta y)\frac{\partial}{\partial u} - iv\left(\xi\frac{\partial}{\partial x} + \eta\frac{\partial}{\partial y}\right)\right), \tag{5.7}$$

which can be decomposed into

$$D(\xi,\eta,0) = \exp\left(-i(\xi x+\eta y)\frac{\partial}{\partial u}\right) \times \exp\left(-iv\left(\xi\frac{\partial}{\partial x} + \eta\frac{\partial}{\partial y}\right)\right) \times \exp\left(-i\frac{v}{2}(\xi^2+\eta^2)\frac{\partial}{\partial u}\right), \tag{5.8}$$

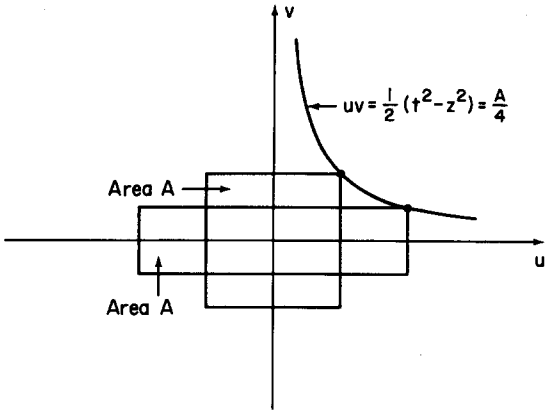


FIG. 2. Lorentz deformation in the uv plane. As the velocity parameter increases, the distribution along the u axis becomes expanded while the v axis becomes contracted, in such a way that the area remains constant. In the infinite-momentum limit, the v distribution becomes like that of $\delta(v)$, while the distribution along the u axis becomes widespread. The translation along the u axis becomes a gauge transformation.

as in the case of Eq. (3.11).

We are applying this operator on functions localized in the four-dimensional space-time. As an illustration, let us consider Dirac's Gaussian form[10]:

$$\psi(x) = (1/\pi)\exp(-(x^2+y^2+z^2+t^2)/2). \quad (5.9)$$

This form is not invariant under Lorentz boosts, but undergoes a Lorentz deformation when the system is boosted.[9,11] If it is boosted along the z direction, the x and y coordinates are not affected. We can therefore delete these transverse variables, and concentrate on the Lorentz deformation property of

$$\psi(z,t) = (1/\pi)^{1/2}\exp(-(u^2+v^2)/2), \quad (5.10)$$

in the uv plane. The light-cone variables u and v are defined in Eq. (3.1), and their Lorentz-transformation property is given in Eq. (3.8). If this function is Lorentz boosted along the z axis,

$$\psi_\beta(z,t) = (1/\pi)^{1/2}\exp[-((u/R)^2+(Rv)^2)/2]. \quad (5.11)$$

The width of this function along the u axis increases as R becomes large, while the distribution along the v axis becomes narrow, as is described in Fig. 2.

This function illustrates the Lorentz-deformation property of functions localized in the uv plane. The width of the v distribution decreases as $1/R$. When the v distribution is very narrow, we can consider the transformation in the subspace where $v=0$. Then the factors

$$\exp\left(-iv\left(\xi\frac{\partial}{\partial x}+\eta\frac{\partial}{\partial y}\right)\right) \text{ and } \exp\left(-i\frac{v}{2}(\xi^2+\eta^2)\frac{\partial}{\partial u}\right)$$

in Eq. (5.8) for $D(\xi,\eta,0)$ can be dropped. As a consequence,

$$D(\xi,\eta,0) = \exp\left(-i(\xi x+\eta y)\frac{\partial}{\partial u}\right). \quad (5.12)$$

This means that the terms $v\,\partial/\partial x$ and $v\,\partial/\partial y$ in Eq. (5.6) can be dropped, and F_1 and F_2 can be written

$$F_1 = -ix\frac{\partial}{\partial u}, \quad F_2 = -iy\frac{\partial}{\partial u}. \quad (5.13)$$

These operators generate translations along the u axis. These operators, together with the rotation generator J_3 of Eq. (5.2), are the generators of the cylindrical group. The differential operators F_1 and F_2 are now the generators of gauge transformations applicable to functions with a narrow distribution in v (Ref. 9).

Here again, a complete description of the little group for massive particles in the infinite-momentum limit requires both the cylindrical and Euclidean components. The Euclidean component can be deleted in the infinite-momentum limit or in the $v=0$ subspace. As we observed at the end of Sec. IV, this is the Lorentz condition applicable to massive particles in the infinite-momentum limit.

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. Wigner, Proc. Natl. Acad. Sci. U.S.A. **34**, 211 (1948).

[2] M. Hamermesh, *Group Theory* (Addison–Wesley, Reading, MA, 1962); E. P. Wigner, in *Theoretical Physics*, edited by A. Salam (IAEA, Vienna, 1962); A. Janner and T. Jenssen, Physica **53**, 1 (1971); **60**, 292 (1972); J. L. Richard, Nuovo Cimento A **8**, 485 (1972); H. P. W. Gottlieb, Proc. R. Soc. London Ser. A **368**, 429 (1979); H. van Dam, Y. J. Ng, and L. C. Biedenharn, Phys. Lett. B **158**, 227 (1985). For a recent textbook on this subject, see Y. S. Kim and M. E. Noz, *Theory and Applications of the Poincaré Group* (Reidel, Dordrecht, 1986). Also, see L. C. Biedenharn, H. W. Braden, P. Truini, and H. van Dam, J. Phys. A **21**, 3593 (1988).

[3] Y. S. Kim and E. P. Wigner, J. Math. Phys. **28**, 1175 (1987). For an earlier discussion of the cylindrical geometry in connection with Maxwell's equations, see L. J. Boya and J. A. de Azcarraga, An. R. Soc. Esp. Fis. Quim. A **63**, 143 (1967).

[4] In the picture of the two-dimensional Euclidean group, the gauge degree of freedom is associated to translations in the two-dimensional Euclidean plane. See S. Weinberg, Phys. Rev. B **134**, 882 (1964); **135**, 1049 (1964); J. Kuperzstych, Nuovo Cimento B **31**, 1 (1976); D. Han and Y. S. Kim, Am. J. Phys. **49**, 348 (1981); J. J. van der Bij, H. van Dam, and Y. J. Ng, Physica A **116**, 307 (1982); D. Han, Y. S. Kim, and D. Son, Phys. Rev. D **31**, 328 (1985).

[5] For earlier efforts to obtain the internal space-time symmetry for massless particles as a limiting case of the O(3)-like symmetry of a massive particle, see E. P. Wigner, Rev. Mod. Phys. **29**, 255 (1957). See, also, D. W. Robinson, Helv. Phys. Acta **35**, 98 (1962); D. Korff, J. Math. Phys. **5**, 869 (1964); S. Weinberg, in *Lectures on Particles and Field Theory*, Brandeis 1964, Vol. 2, edited by S. Deser and K. W. Ford (Prentice–Hall, Englewood Cliffs, NJ, 1965); S. P. Misra and J. Maharana, Phys. Rev. D. **14**, 133 (1976); D. Han, Y. S. Kim, and D. Son, J. Math. Phys. **27**, 2228 (1986).

[6] For the contraction of O(3) to the two-dimensional Euclidean group, see E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. U.S.A. **39**, 510 (1953); J. D. Talman, *Special Functions, A Group Theoretical Approach based on Lectures by E. P. Wigner* (Benjamin, New York, 1968). See also R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications in Physics* (Wiley, New York, 1974); D. Han, Y. S. Kim, and D. Son, Phys. Lett. B **131**, 327 (1983).

[7] E. P. Wigner, *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra*, translated from the German by J. J. Griffin (Academic, New York, 1959).

[8] R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971).

[9] Y. S. Kim, Phys. Rev. Lett. **63**, 348 (1989).

[10] P. A. M. Dirac, Proc. R. Soc. London Ser. A **183**, 284 (1945). For physical implications of this paper, see Y. S. Kim and E. P. Wigner, Phys. Rev. A **36**, 1159 (1988); **39**, 2829 (1989).

[11] Y. S. Kim, M. E. Noz, and S. H. Oh, J. Math. Phys. **20**, 1341 (1979).

Scattering of waves from a random spherical surface—Mie scattering

Hisanao Ogura and Nobuyuki Takahashi
Department of Electronics, Kyoto University, Yoshida, Kyoto 606, Japan

(Received 18 October 1988; accepted for publication 30 August 1989)

The stochastic theory developed by the authors for the scattering from a random planar surface is extended to the case of a random spherical surface, which is assumed to be a homogeneous random field on the sphere, homogeneous with respect to spherical rotations. Based on the group-theoretical analogies between the two, the formulation of the theory is closely connected to the representation theory of the rotation group. The concept of the "stochastic" spherical harmonics associated with the rotation group and their several formulas are introduced and discussed at the beginning. For the plane wave incident on a random spherical surface, the scattered random wave field can be expanded systematically in terms of the stochastic spherical harmonics in much the same way as the nonrandom case, and several formulas are derived for the coherent scattering amplitude, the coherent and incoherent power flows, and the coherent and incoherent scattering cross sections. The power-flow conservation law is cast into the stochastic version of the optical theorem stating that the total scattering cross section consisting of the coherent and incoherent power flow is equal to the imaginary part of the coherent forward-scattering amplitude. Approximate solutions are obtained for the Mie scattering with a slightly random spherical surface where the single scattering approximation is valid due to the absence of a real resonance, as shown in the previous work on the two-dimensional case. Some numerical calculations are made for the coherent and incoherent scattering cross sections.

I. INTRODUCTION

In a series of papers,[1–8] the authors have been studying the scattering of waves from a random planar surface, by means of a stochastic functional approach that is entirely different from the ordinary scattering theories. It is to be stressed that, in addition to the stochastic functional calculus due to Wiener and Ito,[9–11] the formulation of our theory is based on the group-theoretic consideration associated with homogeneity of the random surface, where homogeneity implies the probabilistic invariance under the group of translational motions on the plane, and that this leads to the "stochastic" Floquet theorem[12,1–8] as an irreducible representation for the translation group.

The present paper deals with the problem of scattering from a random spherical object, which is practically related to, for instance, the light scattering by interplanetary dust particles,[13–15] Raman scattering by microcrystalline particles,[16,17] radio-wave scattering by deformed rain drops,[18] diffraction by a rough planetary surface, etc. Although the scattering by a random sphere has been treated in several ways, assuming a suitable model for a random spherical surface,[13–15,19] a theoretical formulation like the well-known scattering theory for a nonrandom sphere[20] has not yet been made for a random sphere. It is perhaps because of the lack of the techniques to handle a random field on the sphere and partly because of the intrincate manipulation of spherical functions in the perturbation calculus. The problem, therefore, attracts our attention not only for practical reasons but also for the theoretical interest aroused by it.

A preliminary study[21] has been made on a simple two-dimensional (2-D) model, that is, the scattering of plane wave from a random cylindrical surface that is homogenous with respect to circular rotations. Since the rotation is only a translational motion along a circle, the irreducible representation of the circular rotation group is again 1D, and the group index is given by a Fourier exponential function, so that the theory can be easily formulated in a manner analogous to the case of random planar surface. The differential cross sections for the coherent and incoherent scattering and a stochastic version of the optical theorem have been obtained. In the case of a planar random surface even if the roughness is negligibly small, the multiple scattering has an important effect causing the so-called anomalous scattering (e.g., scalar wave with Neumann surface,[7] electromagnetic wave with perfectly conducting surface,[3] surface plasmon mode[8]). In the case of random cylindrical surface, although the multiple scattering can be treated by means of Wiener–Hermite expansion in much the same way as the planar case, it is shown that the multiple scattering does not create an appreciable effect in the Mie scattering if the roughness is small enough: This is due to the lack of a real resonance or of a surface mode on the circular surface. Therefore, it is expected for the same reason that, if the roughness is small enough, the Mie scattering from a random spherical surface could be well treated based on the single scattering approximation for the incoherent part and the second-order approximation for the coherent part.

In the present paper, the random surface is assumed to be a homogeneous random field on the sphere, homogeneous with respect to the spherical rotations. The scattered wave field is then regarded as the stochastic functional of the homogeneous random surface, and at the same time it is expected that the random wave field reflects a group-theoretic property associated with the rotational homogeneity. The totality of spherical rotations forms the rotation group that

 0022-2488/90/010061-15$03.00

we denote by G. It is to be noticed that the representation theory of the rotation group[22,23] plays an essential part in our treatment of stochastic functionals associated with the random spherical surface. A great difference from the planar or the circular case lies in the fact that due to the noncommutative property of G the vector space for the irreducible representation is finite dimensional, i.e., $(2l+1)$-D, $l=0,1,2,...$, and its $(2l+1)$-D transformation matrix is given in terms of generalized spherical functions[22]; this is to be contrasted to the 1-D representation in terms of a Fourier exponential function for the translational or the circular motion group. With this difference in mind, we can formulate the scattering theory for a random sphere in a manner analogous to the planar case. Table I shows the analogies between the two formulations for the spherical random surface, which will be helpful in what follows to formulate the scattering theory for a random spherical surface. The scattering from a cylindrical random surface,[21] which stands between the two cases, is theoretically close to the planar case in view of its group property, but physically similar to the spherical case with respect to its scattering characteristics.

A brief summary on the theory of the rotation group and related formulas for vector spherical harmonics are given in the Appendix for reference in the text. Before dealing with the stochastic scattering problem, we need to prepare some mathematical tools in Secs. II–IV, which are indispensable for manipulating the random wave fields arising from the homogeneous random spherical surface. Section II gives a new group-theoretic interpretation for the spectral representation of a homogeneous random field on a sphere. In this connection, we introduce a concept of the stochastic representation of the rotation group in a vector space of random variables: it is a novel way of a group representation, since the representation space has been usually taken as an ordinary vector space or a function space. Section III deals with a more general stochastic representation in the space of random fields by means of the shift transformation associated with the rotational homogeneity. In Sec. IV, a "stochastic spherical harmonic" is defined as a homogeneous vector random field in a $(2l+1)$-D vector space of an irreducible representation. And in terms of stochastic spherical harmonics, a "stochastic solid harmonic" is defined, which satisfies the vector Helmholtz equation and which represents a stochastic spherical wave belonging to the same invariant vector space.

These mathematical definitions and formulas are conveniently applied to the theory of the wave scattering from a random spherical surface in Secs. V–VIII. In Sec. V, the formulation for the scattering problem is given in two steps. First, the theory is developed for the spherical wave incidence where the scattered wave field can be expanded in terms of stochastic solid harmonics belonging to the same invariant vector space with the incident wave: this corresponds to the stochastic Floquet theorem in the case of a planar random surface.[1] Then, in much the same way as the ordinary scattering theory from a sphere,[20] the scattered random wave field for a plane wave incident on the random sphere can be readily obtained by superposition in the form of an expansion in terms of stochastic spherical harmonics (i.e., an irreducible decomposition in the representation theory of the rotation group). Once the stochastic wave field is obtained, the statistical characteristics of the scattered wave can be easily calculated by averaging, which are given in Sec. VI, such as the coherent amplitude, coherent and incoherent power flows, total scattering cross section, optical theorem, and coherent and incoherent scattering distributions. Section VII gives the method of approximate solution for the expansion coefficients by solvintg the boundary condition on the random spherical surface. The calculations are given for the first-order (single-scattering) approximation for the incoherent field and the second-order approximation for the coherent field. Numerical evaluations for several statistical scattering characteristics are finally given in Sec. VIII for the Mie scattering where the first-order approximation is valid for small roughness parameter as shown in the case of a random circular surface. The range of validity is checked using

TABLE I. The analogies between the two formulations for the scattering from a planar random surface and the scattering from a spherical random surface.

	Scattering from a random plane	Scattering from a random sphere
Random surface	1′ homogeneous random field on plane	1 homogeneous random field on sphere
Homogeneity	2′ probabilistic invariance under translations	2 probabilistic invariance under spherical rotations
Group representation	3′ stochastic representation of translation group	3 stochastic representation of the rotation group
Spectral representation	4′ Fourier integral type	4 expansion in spherical harmonics
Analog of Floquet theorem	5′ random field in an irreducible representation space of the translation group	5 random field in an irreducible representation space of the rotation group
Stochastic wave field	6′ integration of stochastic plane wave	6 sum of stochastic solid harmonics
Incident wave	7′ plane wave	7 spherical wave
Synthesized wave field	8′ superposition of plane waves; stochastic radiation field for a point source	8 superposition of spherical waves; stochastic wave field for plane-wave injection

the power conservation formula and the optical theorem.

It is to be remembered that, as often used in quantum mechanics, a group-theoretic treatment based on the symmetry simplifies a great deal of manipulations for calculating the solution, and in the same manner our stochastic group-theoretic treatment based on the rotational homogeneity again considerably simplifies the calculation as well as the description of the solution. Otherwise, even in the single-scattering approximation, the representation and the statistical properties of the scattered wave field are much too complicated to obtain by means of a conventional scattering analysis, because of the coupling of three spherical harmonics arising from the spectral representation of the random surface, the expansions of the Green function and the incident plane wave all in terms of spherical harmonics.

The multiple-scattering correction will be required if we go far beyond the Mie scattering range or treat the case of the short wavelength limit in the diffraction theory. As in the polar or circular rough surface, the multiple scattering can be taken into account, if necessary, by introducing the second Wiener kernel in the Wiener expansion of the stochastic functionals. This involves the irreducible decomposition of the tensorial products describing the second-order Wiener–Hermite functional. To do this, however, implies a lot of complication in addition to the present calculation, although it was trivially easy in the planar case[1,2] or the cylindrical case[21] because of the 1-D representation of the translational or rotational group by means of Fourier exponential functions. Therefore, we avoid this in calculating the approximate solution and limit ourselves to the Mie scattering with small roughness due to the reason mentioned above. The case of multiple scattering will be discussed in a future work based on the present treatment.

II. HOMOGENEOUS RANDOM FIELD ON A SPHERE

For the applications in the following, we briefly summarize the mathematical definitions and formulas with concise description about the derivation and theoretical meanings.

Let $(\Omega,\mathscr{B},\mathbf{P})$ denote the probability space (Ω denotes the sample space, $\mathscr{B}$ the Borel field on Ω, and $\mathbf{P}$ the probability measure), and let $Y(\omega)$ denote a random variable ($\mathscr{B}$ measurable function), ω indicating the probability parameter denoting a sample point in Ω, which will be often supressed for brevity. Let $L^2(\Omega)$ denote the Hilbert space of random variables such that $\langle |Y|^2\rangle < \infty$ with inner product $(Y_1,Y_2)_\Omega = \langle \overline{Y}_1 Y_2\rangle$, $\langle\ \rangle$ denoting the average over Ω.

Let a three-dimensional (3-D) vector $\mathbf{r}$ be denoted by $\mathbf{r}\equiv(r,\theta,\varphi)$ in the spherical coordinates, and a 3-D spherical surface by $S_3=(r,\theta,\varphi)$ with radius $r=1$. The rotational motion g, which describes the transformation $\mathbf{r}\to\mathbf{r}'=g\mathbf{r}$, forms the rotation group G (see the Appendix). Here, D_l denotes a $(2l+1)$-D vector space for an irreducible representation of G with weight l. It is to be kept in mind that the term "representation" is used in two ways in the following; the representation of the rotation group and the representation of a random field, which, however, may not be confusing for us.

Let $f(\mathbf{r})\equiv f(\mathbf{r},\omega)$ represent a q.m. (quadratic mean) continuous random field on S_3. In this section, let $L_f^2(\Omega)$ denote the sub-Hilbert space of random variables which is linearly generated from $f(\mathbf{r})$ (q.m. limit of linear transformation). In the present paper, the probability space $(\Omega,\mathscr{B},\mathbf{P})$ is associated with the homogeneous random field on S_3 describing a random spherical surface, so that a random variable is considered as a stochastic functional of the random spherical surface.

A random field on the sphere whose probability distribution (correlation function) is invariant under arbitrary rotations is said to be homogeneous in the strict (wide) sense. More generally, a homogeneous random field is defined with respect to the invariance under the group of motions in a homogeneous space.[24–26] For our purpose, we assume $f(\mathbf{r},\omega)$ be a homogeneous Gaussian random field on S_3 with mean 0, so that it is homogeneous in either sense. Since the correlation function $R(\mathbf{r}_1,\mathbf{r}_2)=\langle f(\mathbf{r}_1)f(\mathbf{r}_2)\rangle$ is invariant under arbitrary rotations, we have

$$R(\mathbf{r}_1,\mathbf{r}_2)=R(g\mathbf{r}_1,g\mathbf{r}_2),\quad \mathbf{r}_1,\mathbf{r}_2\in S_3,\quad g\in G, \tag{1}$$

from which follows that $R(\mathbf{r}_1,\mathbf{r}_2)$ is a function only of θ, which denotes the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$ [cf. (A19)], so that we write the correlation function as $R(\theta)\equiv R(\mathbf{r}_1,\mathbf{r}_2)$.

We first show three forms for the representation of the homogeneous random field $f(\mathbf{r},\omega)$ on the sphere:

$$f(\mathbf{r},\omega)=f(\mathbf{e}_0,T^{g^{-1}}\omega),\quad \mathbf{r}\equiv g\mathbf{e}_0, \tag{2}$$

$$=\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_l^m(\theta,\varphi)F_l B_l^m(\omega) \tag{3}$$

$$=\sum_{l=0}^{\infty}\sqrt{\frac{2l+1}{4\pi}}\,F_l B_l^0(\mathbf{r},\omega),\quad \mathbf{r}\equiv g\mathbf{e}_0. \tag{4}$$

Equations (2) and (4) will be explained later. Equation (3) is the classical form of the spectral representation[27,28] given in terms of spherical harmonics and the orthogonal random spectrum $F_l B_l^m$, where $B_l^m\equiv B_l^m(\omega)$ denotes a Gaussian random variable with mean 0 having orthogonal property:

$$\langle \overline{B_l^m} B_{l'}^{m'}\rangle = \langle \overline{B_l^m(\mathbf{r})}\, B_{l'}^{m'}(\mathbf{r})\rangle = \delta_{ll'}\cdot\delta_{mm'}, \tag{5}$$

where the middle member in the equality will be explained later. The spectral representation for the correlation function can be easily given using the addition formula for the spherical function:

$$R(\theta)=\frac{1}{4\pi}\sum_{l=0}^{\infty}(2l+1)|F_l|^2 P_l(\cos\theta), \tag{6}$$

where θ is denoting the angle between two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, and we call $|F_l|^2$ the "power spectrum." Particularly, the "white" spectrum $|F_l|^2=1$(const) gives the delta correlation for the white noise on the sphere: $R(\theta)=\delta(\theta),\delta(\theta)$ denoting the delta function on the sphere with the measure $dS\equiv\sin\theta\,d\theta\,d\varphi$.

The homogeneous random field on the sphere, $f(\mathbf{r})$, can be as well regarded as a random field on G by (A2);

$$f(\mathbf{r})\equiv f(g_{\mathbf{r}}\mathbf{e}_0),\quad \mathbf{r}\equiv g_{\mathbf{r}}\mathbf{e}_0,\quad g_{\mathbf{r}}\in G, \tag{7}$$

where $\mathbf{e}_0$ denotes a unit vector along the polar axis, and $g_{\mathbf{r}}$ denotes the rotation that brings $\mathbf{e}_0$ into $\mathbf{r}$. The scalar field (7) is independent of the third Euler angle φ_2 or of the rotation around $\mathbf{r}$.

Let the rotational transformation S^g of the homogeneous random field $f(\mathbf{r})$ be as defined by (A4);

$$S^g: f(\mathbf{r}) \to f(g^{-1}\mathbf{r}), \quad g \in G. \tag{8}$$

Then, the transformation S^g on $f(\mathbf{r})$ induces a transformation U^g on random variables $Y \in L_f^2(\Omega)$,

$$U^g: Y(\omega) \to Y'(\omega) \equiv U^g Y(\omega), \tag{9}$$

which we call the shift transformation. From the invariance (1), U^g is a unitary transformation in L_f^2. The strict sense homogeneity of $f(\mathbf{r})$ implies that the measure-preserving set transformation T^g: $A \to A' = T^g A$ is induced on Ω, such that $\mathbf{P}(T^g A) = \mathbf{P}(A)$, $A, A' \in \mathscr{B}$. For convenience, we write this formally as a point transformation on Ω without loss of rigor[29]:

$$T^g: \omega \to \omega' = T^g \omega, \quad \omega, \omega' \in \Omega, g \in G, \tag{10}$$

and we can write the shift transformation (9) in the following manner:

$$U^g Y(\omega) = Y(\omega') \equiv Y(T^g \omega), \tag{11}$$

which is intuitively understandable because the sample point ω can be looked upon as if it is a coordinate parameter. As easily demonstrated, the transformation T^g, equivalent to U^g, has the group property:

$$T^{g_1} T^{g_2} = T^{g_1 g_2}, \quad [T^g]^{-1} = T^{g^{-1}}, \quad T^e = I, \tag{12}$$

which is also satisfied by U^g, in the same manner as S^g in (A5). Therefore, the group homomorphism $g \to T^g$ (or U^g), $g \in G$, gives the representation of the rotation group in the space $L_f^2(\Omega)$.

Then using the shift operator T^g, we can write the homogeneous random field $f(\mathbf{r},\omega)$ in the form (2):

$$f(\mathbf{r},\omega) = f(\mathbf{e}_0, T^{g^{-1}}\omega) \equiv U^{g^{-1}} f(\mathbf{e}_0,\omega), \quad \mathbf{r} \equiv g\mathbf{e}_0, \tag{13}$$

where g denotes the rotation that brings $\mathbf{e}_0$ into $\mathbf{r}$. Then, we note that the scalar random field $f(\mathbf{r})$ is invariant under rotations around $\mathbf{r}$:

$$U^h f(\mathbf{r}) = f(\mathbf{r}), \quad h \in H_{\mathbf{r}}; U^h f(\mathbf{e}_0) = f(\mathbf{e}_0), \quad h \in H, \tag{14}$$

where H denotes the subgroup of rotations around the "polar axis" $\mathbf{e}_0$; $h\mathbf{e}_0 = \mathbf{e}_0$, $h \in H$, and by $H_{\mathbf{r}}$ the subgroup of rotations around the vector $\mathbf{r} = g\mathbf{e}_0$: $h_{\mathbf{r}}\mathbf{r} = \mathbf{r}$; $h_{\mathbf{r}} = ghg^{-1} \in H_{\mathbf{r}}$, $h \in H$.

By the representation theory of the rotation group,[22,23] the representation space $L_f^2(\Omega)$ for T^g (or U^g) can be decomposed into the sum of irreducible spaces, and correspondingly a vector $f(\mathbf{e}_0)$ in $L_f^2(\Omega)$ can be decomposed into the vectors of orthogonal irreducible spaces. For convenience, we denote by $D_l(\Omega)$, $l = 0,1,2,...$, an irreducible space of the weight-l representation for T^g (U^g).

We can show that the independent Gaussian variable $B_l^m = B_l^m(\omega)$ in the spectral representation (3) is transformed according to the operation $T^{g^{-1}}$ in such a manner as

$$B_l^m(\mathbf{r},\omega) = B_l^m(T^{g^{-1}}\omega) = \sum_{s=-l}^{l} \overline{T_{sm}^l(g)} B_l^s(\omega), \quad \mathbf{r} \equiv g\mathbf{e}_0 \tag{15}$$

and its special case with $m = 0$ is written using (A8),

$$B_l^0(\mathbf{r}) = \sqrt{\frac{4\pi}{2l+1}} \sum_{m=-l}^{l} Y_l^m(\theta,\varphi) B_l^m. \tag{16}$$

This implies that B_l^m, $m = -l,...,l$, is the mth canonical basis relative to the "north pole" $\mathbf{e}_0$ in the $(2l+1)$-D vector space of $D_l(\Omega)$, which is subject to the transformation by means of the unitary matrix $T_{sm}^l(g)$, that the canonical basis B_l^m ought to satisfy the orthogonality (5), and that $B_l^m(\mathbf{r}) \equiv B_l^m(\mathbf{r},\omega)$ in the left-hand member of (15) gives the mth canonical basis relative to $\mathbf{r} = g\mathbf{e}_0$, satisfying the orthogonality (5) as well.

Since by (14) $f(\mathbf{e}_0)$ has only the 0th canonical components relative to $\mathbf{e}_0$, its irreducible decomposition can be written in terms only of B_l^0,

$$f(\mathbf{e}_0) = \sum_{l=0}^{\infty} \sqrt{\frac{2l+1}{4\pi}} F_l B_l^0. \tag{17}$$

If we apply $U^{g^{-1}}$ (or $T^{g^{-1}}$) on both sides, using (13) and (15), we get the expression (4), which is the spectral representation in terms of the moving canonical basis in $D_l(\Omega)$. That each term of (4) or (17) has only 0th canonical component implies according to (A27) that the homogeneous random field $f(\mathbf{r})$ is decomposed into the sum of isotropic l-vector field in $D_l(\Omega)$, which is a simple geometrical interpretation for the spectral representation.

Substituting (16) into (4) we recover the spectral representation (3) in the original form, where the random spectrum $F_l B_l^m$ can be interpreted as the fixed canonical basis of $D_l(\Omega)$ at the north pole $\mathbf{e}_0$. Therefore, the simple spectral representation (4) is a "coordinate-free" representation while the original form (3) is a "coordinate-fixed" representation. Such a group-theoretic or geometric simplification is greatly helpful when we deal with the random fields generated by the original homogeneous random field.

III. SHIFT TRANSFORMATION AND HOMOGENEOUS RANDOM FIELDS ON A SPHERE

In what follows, we deal with a random field on the sphere $\psi(\mathbf{r},\omega)$ as a ($\mathscr{B}_S \times \mathscr{B}$ measurable) function on $S_3 \times \Omega$, or more generally a random field $\psi(g,\omega)$ on G. For the sake of practical applications we introduce the shift transformation D^g operating on random fields using the convenient notation T^g, instead of U^g. Define the operator D^g, $g \in G$, by

$$D^g \psi(\mathbf{r},\omega) = \psi(g^{-1}\mathbf{r}, T^{g^{-1}}\omega), \tag{18}$$

$$D^g \psi(g_0,\omega) = \psi(g^{-1}g_0, T^{g^{-1}}\omega). \tag{19}$$

Writing $\mathbf{r} = g_0\mathbf{e}_0$, (18) is a special case of (19). From (A5) and (12) it easily follows that D^g gives a representation of the rotation group G:

$$D^{g_1} D^{g_2} = D^{g_1 g_2}, \quad [D^g]^{-1} = D^{g^{-1}}, \quad D^e = I. \tag{20}$$

The operator D^g, being a measure transformation on $S_3 \times \Omega$, can be applied to the random measure as well. The shift operator D^g introduced here is an analogy to the shift operator operating on stationary processes.[12]

It should be noticed that the homogeneous random field (2) is invariant under D^g, $g \in G$. More generally, if a random field $X(\mathbf{r},\omega)$ is D^g invariant, that is,

$$D^g X(\mathbf{r},\omega) \equiv X(g^{-1}\mathbf{r}, T^{g^{-1}}\omega) = X(\mathbf{r},\omega), \tag{21}$$

then it is shown that $X(\mathbf{r},\omega)$ is a homogeneous (scalar) random field on S_3 expressible in the same form as (2):

$$X(\mathbf{r},\omega) = X(\mathbf{e}_0, T^{g^{-1}}\omega) \equiv U^{g^{-1}} X(\mathbf{e}_0,\omega), \quad \mathbf{r}\in g\mathbf{e}_0, \tag{22}$$

that is, the value at $\mathbf{r} = g\mathbf{e}_0$ can be obtained by applying $T^{g^{-1}}$ to the value at the north pole, $X(\mathbf{e}_0,\omega)$, which is a scalar quantity invariant under rotations around $\mathbf{e}_0$.

More generally, using a random variable $Y(\omega)$ ($\mathscr{B}$ measurable) we make a random field on G;

$$Y(g,\omega) \equiv Y(T^{g^{-1}}\omega), \quad g\in G, \tag{23}$$

which is easily shown to be D^g invariant and hence is a homogeneous random field on G. A random field on S_3, like (22), is a special case of (23) such that $Y(T^h\omega) = Y(\omega)$, $h\in H$.

Next, we consider an l-vector random field with components,

$$X_l^m(\mathbf{r},\omega), \quad m = -l,...,l, \quad \mathbf{r}\equiv g_r\mathbf{e}_0, \tag{24}$$

which are transformed by D^g in the following manner:

$$D^g X_l^m(\mathbf{r},\omega) = \sum_{s=l}^{l} \overline{T_{sm}^l(g)} X_l^s(\mathbf{r},\omega), \quad m = -l,...,l. \tag{25}$$

Such random field is said to be a homogeneous l-vector random field. One of such l-vector random field can be represented in the following form:

$$X_l^m(\mathbf{r},\omega) = T_{mn}^l(g_r) X_n(\mathbf{r},\omega), \quad m = -l,...,l, \tag{26}$$

where $X_n(\mathbf{r},\omega)$, $n = -l,...,l$, represents a D^g-invariant scalar random field of the form (22). In fact, the factor $T_{mn}^l(g_r)$ in (26) is transformed under D^g or S^g like (25) according to (A16). Since by (A15) $T_{mn}^l(g_r)$ is the mth component of $\mathbf{e}_{(l)n}(\mathbf{r})$ in the fixed canonical basis, (26) can be expressed in the l-vector notation as

$$X_{(l)n}(\mathbf{r},\omega) = \mathbf{e}_{(l)n}(\mathbf{r}) X_n(\mathbf{r},\omega). \tag{27}$$

The l-vector fields, (27) with different n, is linearly independent of (orthogonal to) each other. A homogeneous l-vector random field can be generally written as a linear combination of (27) in n, examples of which will appear in the next section. It is to be noted that the random field of the form (26) or (27) is a rotational counterpart of the stochastic Floquet theorem based on the translational motion.[12]

IV. STOCHASTIC SPHERICAL HARMONICS AND STOCHASTIC SOLID HARMONICS

Let $Z_l^m = Z_l^m(\omega)$ be a fixed canonical basis in $D_l(\omega)$, and $Z_l^m(\mathbf{r}) = Z_l^m(\mathbf{r},\omega)$ be the moving canonical basis relative to $\mathbf{r} = g\mathbf{e}_0$, such that

$$\langle \overline{Z_l^m} Z_{l'}^{m'} \rangle = \langle \overline{Z_l^m(\mathbf{r})} Z_{l'}^{m'}(\mathbf{r}) \rangle = \delta_{ll'}\delta_{mm'}, \tag{28}$$

$$Z_l^m(\mathbf{r}) \equiv Z_l^m(T^{g^{-1}}\omega) = \sum_{s=-l}^{l} \overline{T_{sm}^l(g)} Z_l^s, \quad \mathbf{r}\equiv g\mathbf{e}_0 \tag{29}$$

[cf. (5), (15)]. For comparison and reference it should be noticed that the similar relations (A14) and (A15) do hold for an l-vector canonical basis $\mathbf{e}_{(l)n}$ in D_l and the moving canonical basis $\mathbf{e}_{(l)n}(\mathbf{r})$.

We define the "stochastic l-vector spherical harmonics" associated with Z_l^m by the formula

$$\mathbf{P}_{(l)n}^{l'}(\mathbf{r},\omega) = \sum_{m=-l'}^{l'} \mathbf{P}_{(l)n}^{l'm}(\theta,\varphi) Z_{l'}^m(\omega), \quad \mathbf{r}\equiv(1,\theta,\varphi) \tag{30}$$

$$= \sqrt{\frac{2l'+1}{4\pi}}\, \mathbf{e}_{(l)n}(\mathbf{r}) Z_{l'}^n(\mathbf{r}) \tag{31}$$

$$l,l' = 0,1,2,..., \quad L = \min(l,l'), \quad n = -L,...,L,$$

where $\mathbf{P}_{(l)n}^{l'm}(\theta,\varphi)$ denotes the l-vector spherical harmonic defined by (A29), and (30) is rewritten into (31) using (A29), (29), and (A15). Since, by definition (29), $Z_l^n(\mathbf{r})$ is invariant under $D^{g'}$, (31) is a homogeneous l-vector random field of the form (27). The correlation function can be easily calculated; we omit its details here.

Now we introduce the "stochastic solid harmonics" $\mathbf{J}_{(l)n}^l(\mathbf{r};\omega)$ such that it is a homogeneous l-vector random field on R_3 that satisfies the l-vector Helmholtz equation. We denote the position and the wave vector by $\mathbf{r}\equiv(r,\theta,\varphi)$ and $\mathbf{k} = (k,u,v)$ in the polar coordinate, respectively, and let them stand for the spherical coordinates (θ,φ) and (u,v) as well. Let us define the stochastic solid harmonic by the integral

$$\mathbf{J}_{(l)n}^{l'}(\mathbf{r};\omega) \equiv \frac{1}{4\pi i^{l'-l}} \int_{S_3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{P}_{(l)n}^{l'}(\mathbf{k};\omega)\, dS_{\mathbf{k}} \tag{32}$$

$$l' = 0,1,2,..., \quad n = -L,...,L, \quad L = \min(l,l').$$

That this bears the desired transformation properties under D^g easily follows from that of the integrand, or from the following expressions. Substituting (30) into (32) and using (A35), (A31), (A29), and (31), we obtain several expressions for (32);

$$\mathbf{J}_{(l)n}^{l'}(\mathbf{r};\omega) = \sum_{m=-L}^{L} \mathbf{J}_{(l)n}^{l'm}(kr,\theta,\varphi) Z_{l'}^m \tag{33}$$

$$= \sum_{t=-L}^{L} j_{nt}^{l'l}(kr) \mathbf{P}_{(l)t}^{l'}(\mathbf{r};\omega) \tag{34}$$

$$= \sqrt{\frac{2l'+1}{4\pi}} \sum_{t=-L}^{L} j_{nt}^{l'l}(kr) \mathbf{e}_{(l)t}(\mathbf{r}) Z_{l'}^t(\mathbf{r}). \tag{35}$$

Since $Z_{l'}^t(\mathbf{r})$ is D^g invariant, (35) is the sum of the functions of the form (27). The integral representation (32) is a stochastic analog to (A35), and (34) is another analog to (A31). Furthermore, substituting (34) into the left-hand side of (32), and (31) into the right-hand side, we obtain the formula analogous to (A36):

$$\sum_{t=-L}^{L} j_{nt}^{l'l}(kr) \mathbf{e}_{(l)t}(\mathbf{r}) Z_{l'}^t(\mathbf{r}) = \frac{1}{4\pi i^{l'-l}} \int_{S_3} \mathbf{e}_{(l)n}(\mathbf{k}) Z_{l'}^n(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}\, dS_{\mathbf{k}}, \tag{36}$$

which is the tensor integral representation, where $\mathbf{e}_{l(n)}$ is a canonical l-vector in D_l and $Z_{l'}^n$ a canonical $\overline{l'}$ vector in $D_{l'}(\Omega)$ so that $\mathbf{e}_{(l)n}(\mathbf{k}) Z_{l'}^n(\mathbf{k})$ gives an isotropic $l\times\overline{l'}$-tensor field in $D_l\times D_{l'}(\Omega)$ according to (A28).

It is obvious from (32) or from (33) and (A33) that $\mathbf{J}^{l'}_{(l)n}(\mathbf{r},\omega)$ satisfies the l-vector Helmholtz equation;

$$(\nabla^2 + k^2)\mathbf{J}^{l'}_{(l)n}(\mathbf{r},\omega) = 0 \tag{37}$$

analogous to (A33). The above-mentioned analogies would justify the name of "stochastic l-vector solid harmonics." Similarly, we can define the stochastic solid harmonics $\mathbf{H}^{(1)l'}_{(l)n}(\mathbf{r};\omega)$ by replacing $\mathbf{J}^{l'}_{(l)n}$ by $\mathbf{H}^{(1)l'}_{(l)n}$ and $j^{l'l}_{nt}(kr)$ by $h^{(1)l'l}_{nt}(kr)$ in the right-hand side of (33)–(35), which represent the stochastic outgoing l-vector wave satisfying the Helmholtz wave equation.

When Z^m_l is a linear functional of B^m_l's, then Z^m_l can be replaced by B^m_l on the above formulas. Otherwise it is to be represented as a nonlinear functional in terms of the Wiener–Hermite tensorial expansion, details of which will be omitted here.

V. SCATTERING FROM A SPHERE WITH HOMOGENEOUS RANDOM SURFACE

A. Random spherical surface

Let the random surface on the sphere with radius a be described by

$$r = a + f(\mathbf{r},\omega), \quad \langle f(\mathbf{r},\omega)\rangle = 0, \tag{38}$$

where $f(\mathbf{r},\omega)$ denotes a homogeneous Gaussian random field on S_3 given by the representations (2)–(4) with the correlation function (6). From now on, let $\mathbf{r}$ represent a vector with length $r = |\mathbf{r}|$. The parameter describing the roughness is given by the variance

$$\sigma^2 \equiv \langle f(\mathbf{r},\omega)^2\rangle = \frac{1}{4\pi}\sum_{l=0}^{\infty}(2l+1)|F_l|^2. \tag{39}$$

B. Wave equation and the boundary condition

The stochastic wave field scattered by the random spherical surface has to satisfy the Helmholtz equation and the boundary condition on the random surface (38) as well as the radiation condition;

$$(\nabla^2 + k^2)\psi(\mathbf{r},\omega) = 0, \tag{40}$$

$$\psi(\mathbf{r},\omega) = 0, \quad r = a + f(\mathbf{r}_s,\omega) \quad \text{(Dirichlet)}, \tag{41}$$

$$\frac{\partial\psi(\mathbf{r},\omega)}{\partial n} = 0, \quad r = a + f(\mathbf{r}_s,\omega) \quad \text{(Neumann)}, \tag{42}$$

where ∇^2 denotes 3-D Laplacian, and $\mathbf{r}_s$ a point on S_3 crossed by the vector $\mathbf{r}$. Assuming the roughness is small enough, the boundary condition (41) or (42) on the random surface can be replaced by the following approximate boundary condition on the sphere:

$$\left[\psi + f\frac{\partial\psi}{\partial r}\right]_{r=a} = 0 \quad \text{(Dirichlet)}, \tag{43}$$

$$\left[\frac{\partial\psi}{\partial r} - \nabla f\cdot\nabla\psi + f\frac{\partial^2\psi}{\partial r^2}\right]_{r=a} = 0 \quad \text{(Neumann)}. \tag{44}$$

At this point we note that the random wave field $\psi(\mathbf{r},\omega)$ as a function of ω is regarded as a stochastic functional of the random surface via the boundary condition, and accordingly, all random quantities appearing in the following are considered as generated from the Gaussian variables B^m_l. We also note that the random surface (38), as well as the Laplacian and the boundary condition, is invariant under the transformation D^g as in (21).

C. Spherical-wave expansion of a plane wave

As well known, the plane wave with the wave vector $\mathbf{k} \equiv (k,u,v)$ is expanded in terms of spherical harmonics:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l} i^l j_l(kr)\,\overline{Y^m_l(\theta,\varphi)}\,Y^m_l(u,v) \tag{45}$$

$$= \sum_{l=0}^{\infty} i^l(2l+1)j_l(kr)(\mathbf{e}_{(l)0}(\mathbf{k})\cdot\mathbf{e}_{(l)0}(\mathbf{r})). \tag{46}$$

Here, (46) is a simple form using the inner product: The parenthesis implies the $(2l+1)$-D inner product of two l-vectors [see (A12), (A18)], and $\mathbf{e}_{(l)m}(\mathbf{r})$, $m = -l,\ldots,l$, denote the canonical l-vector basis relative to $\mathbf{r}$ in the space D_l, which is related to the fixed canonical basis $\mathbf{e}_{(l)m}$ in such a way that [see (A14), (A15)]

$$(\mathbf{e}_{(l)m}\cdot\mathbf{e}_{(l)n}) = (\mathbf{e}_{(l)m}(\mathbf{r})\cdot\mathbf{e}_{(l)n}(\mathbf{r})) = \delta_{mn}, \tag{47}$$

$$\mathbf{e}_{(l)m}(\mathbf{r}) = T^l(g_{\mathbf{r}})\mathbf{e}_{(l)m} = \sum_{s=-l}^{l} T^l_{sm}(g_{\mathbf{r}})\mathbf{e}_{(l)s}, \quad m = -l,\ldots,l, \tag{48}$$

where (47) shows the orthogonality and (48) the transformation property under rotation [cf., (5), (15) or (28), (29)].

The expansion (45) or (46) in terms of spherical wave should be compared with the expansion in terms of cylindrical wave in the 2-D problem [Ref. 21, Eq. (24)]. A spherical wave with the quantum number $l = 0,1,2,\ldots$ is regarded as an l-vector wave with $(2l+1)$ components numbered with $m = -l,\ldots,l$ while in 2-D cases the cylindrical wave is a scalar wave with the quantum number $m = 0,1,2,\ldots$, which corresponds to l in the 3-D problem. As the 2-D scattering problem was treated separately for each mth cylindrical wave,[21] in the present 3-D problem each l-vector spherical wave can be dealt with separately since it is transformed within the same vector space under the transformation D^g.

In view of the relation (46), the wave solution for the plane wave incident on a random sphere can be obtained in the following manner: First, we find the l-vector wave solution for the spherical wave incidence, namely, the solution for the primary wave of an isotropic l-vector field $2j_l(kr)\mathbf{e}_{(l)0}(\mathbf{r})$ [cf. (A27)]; second, taking the inner product of the l-vector wave solution with the l-vector $\mathbf{e}_{(l)0}(\mathbf{k})$ and summing the inner products over all l, we obtain the complete wave field for the incidence of a plane wave with the wave vector $\mathbf{k}$.

D. Unperturbed field (primary wave)

In the nonrandom case with $\sigma^2 = 0$ (smooth sphere), the unperturbed wave field for the spherical wave injection can be written in the well-known manner in terms of spherical harmonics and spherical Bessel functions.[20] According to the remark made above, we will write this in the vector notation:

$$\psi^0_{(l)}(\mathbf{r}) \equiv \psi^0_{(l)}(r)\mathbf{e}_{(l)0} \equiv [j_l(kr) + \alpha^0_l h^{(1)}_l(kr)]\mathbf{e}_{(l)0}(\mathbf{r}) \tag{49}$$

$$= \tfrac{1}{2}\left[h_{(l)}^{(2)}(kr) + e^{2i\delta_l} h_l^{(1)}(kr)\right]\mathbf{e}_{(l)0}(\mathbf{r}) \quad (50)$$

$$\alpha_l^0 \equiv ie^{i\delta_l}\sin\delta_l = -j_l(ka)/h_l^{(1)}(ka) \quad \text{(Dirichlet)}, \quad (51)$$

$$= -j_l'(ka)/h_l'^{(1)}(ka) \quad \text{(Neumann)}, \quad (52)$$

where $j_l(z)$ is the spherical Bessel function, $h_l^{(1)}(z)$ the spherical Hankel function of the first kind, and the prime denotes the differentiation; $j_l'(z) = dj_l(z)/dz$. Here, α_l^0 and δ_l are the so-called scattering coefficient and phase shift. The l-vector $\mathbf{e}_{(l)0}(\mathbf{r})$ in (49) is the 0th canonical basis whose $2l+1$ components relative to the fixed basis are the spherical harmonics [see (A15), (A8)]. It should be noticed that under the rotational shift D^g the spherical wave $\psi_{(l)}^0(\mathbf{r})$ given by (49) (independent of ω) is transformed as an isotropic l-vector field in the space D_l [see (A23) and (A27)].

E. Perturbed wave field (secondary wave)

In the case of random surface ($\sigma^2 > 0$), let the total l-vector wave field for the spherical wave injection be written

$$\psi_{(l)}(\mathbf{r},\omega) \equiv \psi_{(l)}^0(\mathbf{r}) + \psi_{(l)}^s(\mathbf{r},\omega), \quad (53)$$

where the second term is the perturbed wave field created by the surface roughness. In view of the fact that the boundary condition on the homogeneous random surface (2), as well as the wave equation (40), is invariant under the transformation D^g, $g \in G$, and that $\psi_{(l)}^0$ is an isotropic l-vector field [cf. (A27)], the perturbed random wave field also should be an l-vector random field subject to the same transformation as $\psi_{(l)}^0$ under D^g. That is, it is a homogeneous l-vector random field transformed under D^g as in (25), and at the same time it should represent an outgoing l-vector wave satisfying the Helmholtz equation. To sum up, the secondary wave should be expanded in terms of stochastic solid harmonics $\mathbf{H}_{(l)n}^{(1)l'}(\mathbf{r},\omega)$, which satisfy the l-vector Helmholtz equation (37) and which are given by Eqs. (33)–(35) where $\mathbf{J}_{(l)n}^{l'm}(kr,\theta,\varphi)$ and $j_{nl}^{l'l}(kr)$ are to be replaced by $\mathbf{H}_{(l)n}^{(1)l'm}(kr,\theta,\varphi)$ and $h_{nl}^{(1)}(kr)$, respectively, so that they express outgoing waves corresponding to the spherical Hankel functions. Such complication did not arise in the 2D problem because the wave components are all scalar quantities and can be readily expressed in terms of familiar functions [cf. Ref. 21, (32)].

Therefore, our expansion can be written

$$\psi_{(l)}^s(\mathbf{r},\omega) = \sum_{l'=0}^{\infty}\sum_{n=-L}^{L}\sqrt{\frac{4\pi}{2l'+1}}A_n^{ll'}\mathbf{H}_{(l)n}^{l'}(k\mathbf{r},\omega), \quad (54)$$

$$= A_l^c h_l^{(1)}(kr)\mathbf{e}_{(l)0}(\mathbf{r}) + \sum_{l'=0}^{\infty}\sum_{m=-L}^{L} H_m^{ll'}(kr) \times \mathbf{e}_{(l)m}(\mathbf{r}) Z_{l'}^m(\mathbf{r},\omega), \quad (55)$$

where $A_n^{ll'}$ denotes the expansion coefficient, $Z_{l'}^m(\mathbf{r})$ are the orthogonal random variables forming the moving canonical basis in $D_l(\Omega)$ satisfying the same relations as (5) and (15) [see (28), (29)], as we have put

$$H_m^{ll'}(kr) \equiv \sum_{n=-L}^{L} h_{mn}^{(1)ll'}(kr) A_n^{ll'}, \quad L = \min(l,l'). \quad (56)$$

Equation (55), a rewritten form of (54) using (35), corresponds to the expansion in terms of the "stochastic spherical harmonics" (31). Concerning the term with $l' = 0$ in (54), we should notice the following. It should be noticed that, although $Z_0(\mathbf{r},\omega) \equiv Z_0(\omega)$ is a D^g-invariant scalar, namely, $Z_0(T^g\omega) = Z_0(\omega)$, it is not necessarily a nonrandom constant. Thus the first term of (56) with coefficient A_l^c denotes the average part extracted from the term with $l' = 0$, and accordingly $\langle Z_{l'}^m(\mathbf{r})\rangle = 0$ in the rest of terms.

The random variable $Z_{l'}^m(\omega)$, which generally is a nonlinear functional of the random surface, could be expanded in the Wiener–Hermite series of $B_{l'}^m$ (Refs. 12 and 21). However, in view of the result in the cylindrical case,[21] we can assume that the multiple scattering due to the surface roughness could be neglected in the Mie scattering range with small roughness, since there is not real reasonance nor surface mode on the sphere. Therefore, in such a case, the incoherent part of the scattered wave u_l^{ic} can be well represented as a linear functional of the random surface or of the variables $B_{l'}^m$. In the expansion of (55), accordingly, we replace $Z_{l'}^m(\mathbf{r},\omega)$ by $B_{l'}^m(\mathbf{r},\omega)$ instead of a nonlinear Wiener expansion.

F. Scattered wave field for spherical wave injection

To summarize the total wave field $\psi_{(l)}$ for the l-vector spherical wave injection can be cast into the coherent (average) part $\psi_{(l)}^c$ plus the incoherent (random) part $\psi_{(l)}^{ic}$ with mean 0;

$$\psi_{(l)}(\mathbf{r},\omega) \equiv \psi_{(l)}^0(\mathbf{r}) + \psi_{(l)}^s(\mathbf{r},\omega) \quad (57)$$

$$= \psi_{(l)}^c(\mathbf{r}) + \psi_{(l)}^{ic}(\mathbf{r},\omega), \quad \langle\psi_{(l)}^{ic}\rangle = 0, \quad (58)$$

where

$$\psi_{(l)}^c(\mathbf{r}) = \left[j_l(kr) + \alpha_l h_l^{(1)}(kr)\right]\mathbf{e}_{(l)0}(\mathbf{r}), \quad (59)$$

$$\alpha_l \equiv \alpha_l^0 + A_l^c, \quad (60)$$

$$\psi_{(l)}^{ic}(\mathbf{r},\omega) = \sum_{l'=0}^{\infty}\sum_{m=-L}^{L} H_m^{ll'}(kr)\mathbf{e}_{(l)m}(\mathbf{r})B_{l'}^m(\mathbf{r},\omega), \quad (61)$$

$\psi_{(l)}^c$ gives the coherent (average) part and $\psi_{(l)}^{ic}$ the incoherent (random) part with mean 0. We call α_l the coherent scattering coefficient.

G. Scattered wave field for plane-wave injection

To obtain the scattered wave field for the incidence of the plane wave (46), according to the procedure given above, we simply take the inner products of the l-vector $\mathbf{e}_{(l)0}(\mathbf{k})$ with the l-vector fields (59) and (61), and sum them up with respect to l: This corresponds to the irreducible decomposition of the rotation group. In what follows, without loss of generality, we take the plane wave progressing along the polar axis, i.e., $\mathbf{k} = k\mathbf{e}_0$, and use the formulas (A17), (A18), and (A8) to calculate the inner products.

Thus, the total wave field for plane-wave incidence with the wave vector $\mathbf{k} \equiv k\mathbf{e}_0$ can be written

$$\psi(\mathbf{r},\omega)\equiv\psi^0(\mathbf{r})+\psi^s(\mathbf{r},\omega) \tag{62}$$

$$=\psi^c(\mathbf{r})+\psi^{ic}(\mathbf{r},\omega),\quad \langle\psi^{ic}\rangle=0, \tag{63}$$

where

$$\psi^c(\mathbf{r})=\sum_{l=0}^{\infty} i^l(2l+1)(\mathbf{e}_{(l)0}(\mathbf{k})\cdot\psi^c_{(l)}(\mathbf{r})) \tag{64}$$

$$=\sum_{l=0}^{\infty} i^l(2l+1)[j_l(kr)+\alpha_l h_l^{(1)}(kr)]P_l(\cos\theta) \tag{65}$$

$$=\sum_{l=0}^{\infty} i^l\frac{(2l+1)}{2}[h_l^{(2)}(kr)+(e^{2i\delta_l}+2A_l^c) \times h_l^{(1)}(kr)]P_l(\cos\theta) \tag{66}$$

$$\psi^{ic}(\mathbf{r},\omega)=\sum_{l=0}^{\infty} i^l(2l+1)(\mathbf{e}_{(l)}(\mathbf{k})\cdot\psi^{ic}_{(l)}(\mathbf{r},\omega)) \tag{67}$$

$$=\sum_{l=0}^{\infty} i^l(2l+1)\sum_{l'=0}^{\infty}\sum_{m=-L}^{L} H_m^{ll'}(kr)T^l_{0m}(g) \times B^m_{l'}(\mathbf{r},\omega), \tag{68}$$

the third Euler angle φ_2 in $T^l_{0m}(g)$ $(\mathbf{r}\equiv r\mathbf{e}_0)$ being canceled by the term $\overline{T^{l'}_{sm}(g)}$ in $B^m_{l'}(\mathbf{r},\omega)$ [see (15)], and $T^l_{0m}(g)$ being given by spherical harmonics [see (A8)].

VI. STATISTICAL CHARACTERISTICS OF THE SCATTERED WAVE FIELD

A. Coherent scattering amplitude

The coherent scattering part $\psi^c_s(\mathbf{r})(=\psi^c(\mathbf{r})-\psi_0(\mathbf{r}))$ arising from the second term of (65) has the asymptotic form

$$\psi^c(\mathbf{r})=\sum_{l=0}^{\infty} i^l(2l+1)\alpha_l h_l^{(1)}(kr)P_l(\cos\theta) \tag{69}$$

$$\sim(e^{ikr}/r)\Phi(\theta),\quad r\to\infty \tag{70}$$

$$\Phi(\theta)\equiv\frac{1}{ki}\sum_{l=0}^{\infty}(2l+1)\alpha_l P_l(\cos\theta), \tag{71}$$

$\Phi(\theta)$ is called the coherent scattering amplitude.

B. Power flow conservation

If we apply the Gauss theorem to a spherical region with radius r surrounding the random sphere, we obtain the conservation law for the total power flow crossing the sphere:

$$0=\frac{r^2}{k}\int_{S_3}\mathrm{Im}\left[\overline{\psi(\mathbf{r},\omega)}\frac{\partial\psi(\mathbf{r},\omega)}{\partial r}\right]dS$$

$$=\frac{r^2}{k}\int_{S_3}\mathrm{Im}\left\langle\overline{\psi(\mathbf{r},\omega)}\frac{\partial\psi(\mathbf{r},\omega)}{\partial r}\right\rangle dS,\quad \text{a.s.} \tag{72}$$

where the total power flow equals 0 with probability 1 (regardless of random surface) so that it must be equal to its average also, which is the last equality. Substituting (63) into (72) we obtain

$$\frac{r^2}{k}\int_{S_3}\mathrm{Im}\left[\overline{\psi^c(\mathbf{r})}\frac{\partial\psi^c(\mathbf{r})}{\partial r}\right]dS$$

$$+\frac{r^2}{k}\int_{S_3}\mathrm{Im}\left\langle\overline{\psi^{ic}(\mathbf{r},\omega)}\frac{\partial\psi^{ic}(\mathbf{r},\omega)}{\partial r}\right\rangle dS=0, \tag{73}$$

where the first term represents the total coherent power flow while the second the total incoherent power flow, which we denote S_c and σ_{ic}, respectively.

C. Total coherent power flow

Since (73) is independent of r, the total coherent power can be evaluated at $r\to\infty$, substituting (65):

$$S_c=\lim_{r\to\infty}\frac{r^2}{k}\int_{S_3}\mathrm{Im}\left[\overline{\psi^c(\mathbf{r})}\frac{\partial\psi^c(\mathbf{r})}{\partial r}\right]dS$$

$$=\frac{\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)[-1+|1+2\alpha_l|^2]$$

$$=\frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)[\mathrm{Re}\,\alpha_l+|2\alpha_l|^2]. \tag{74}$$

The term arising from the second term in the bracket of (74) gives the coherent power flow for the scattered wave which we denote by σ_c.

D. Total incoherent power flow

Substituting (68) into the second terms of (73), and using the orthogonality relations (5) for B^m_l and (A21) for $T^l_{nm}(t)$, we can evaluate the total incoherent power flow at $r\to\infty$:

$$\sigma_{ic}=\lim_{r\to\infty}\frac{r^2}{k}\int_S\mathrm{Im}\left[\overline{\psi^{ic}}\frac{\partial\psi^{ic}}{\partial r}\right]dS$$

$$=4\pi\sum_{l=0}^{\infty}(2l+1)\sum_{l'=0}^{\infty}\sum_{m=-L}^{L}\lim_{r\to\infty} \times\mathrm{Im}\left[\overline{H_m^{ll'}(kr)}\,\dot{H}_m^{ll'}(kr)\right] \tag{75}$$

$$=\frac{4\pi}{k^2}\sum_{l=0}^{\infty}\sum_{l'=0}^{\infty}\sum_{m=-L}^{L}(2l+1)|H_m^{ll'}|^2, \tag{76}$$

where the overdot denotes the differentiation and we have used the asymptotic formulas for $H_m^{ll'}(kr)$ and $h_{mn}^{(1)ll'}(kr)$,

$$H_m^{ll'}(kr)\sim\frac{e^{ikr}}{ikr}H_m^{ll'},\quad h_{mn}^{(1)l'l}(kr)\sim\frac{e^{ikr}}{ikr}h_{mn}^{l'l},\quad r\to\infty, \tag{77}$$

$$H_m^{ll'}\equiv\sum_{n=-L}^{L}h_{mn}^{l'l}A_n^{ll'},\quad L=\min(l,l'), \tag{78}$$

$$h_{mn}^{l'l}\equiv\sum_{L=|l-l|}^{l+l}i^{-l+l'}(-1)^{m+n}(l-ml'm|ll'L0) \times(l-nl'n|ll'L0). \tag{79}$$

E. Total scattering cross section

The total power-flow conservation formula (73) can be rewritten using (74) and (76):

$$\frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)\Big[\mathrm{Re}\,\alpha_l+|2\alpha_l|^2 +\sum_{l'=0}^{\infty}\sum_{m=-L}^{L}|H_m^{ll'}|^2\Big]=0, \tag{80}$$

where the sum corresponding to the second term in the bracket gives the coherent power flow σ_c of the scattered wave and the third term in the bracket corresponds to the

incoherent power flow σ_{ic}. The sum, therefore, is to give the total scattering cross section S, that is,

$$S=\sigma_c+\sigma_{ic}=\frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)\times\left[|2\alpha_l|^2+\sum_{l'=0}^{\infty}\sum_{m=-L}^{L}|H_m^{ll'}|^2\right]. \quad (81)$$

F. Optical theorem

In terms of (71) and (81), the conservation law (80) can be cast into the well-known form of optical theorem:

$$S=(4\pi/k)\,\mathrm{Im}\,\Phi(0), \quad (82)$$

where $\Phi(0)$ is the forward coherent scattering amplitude. This is the stochastic version of the optical theorem for the scattering from a random sphere: The total scattering cross section consisting of the coherent and incoherent power flow is given by the imaginary part of the coherent forward-scattering amplitude.

G. Angular distribution of coherent scattering (differential cross-section for coherent scattering)

In view of (70), the angular distribution $\sigma_c(\theta)$ of the coherent scattering can be given by

$$\sigma_c(\theta)dS=|\Phi(\theta)|dS=\frac{1}{k^2}\left|\sum_{l=0}^{\infty}(2l+1)\alpha_l\times P_l(\cos\theta)\right|^2 dS, \quad (83)$$

$dS=\sin\theta\,d\theta\,d\varphi$ denoting the spherical element.

H. Angular distribution of incoherent scattering (differential cross section for incoherent scattering)

Calculating the average power flow of (68) using (5), (77), and (A8), the angular distribution $\sigma_{ic}(\theta)$ of the incoherent scattering can be given by

$$\sigma_{ic}(\theta)dS=\frac{1}{k^2}\sum_{l'=0}^{\infty}\sum_{m=-l'}^{l'}\left|\sum_{l=0}^{\infty}i^l(2l+1)H_m^{ll'}\times\sqrt{\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)\right|^2 dS, \quad (84)$$

which depends only on θ as expected. Integrating (83) and (84) over dS gives σ_c and σ_{ic}, respectively.

VII. METHOD OF APPROXIMATE SOLUTION FOR THE BOUNDARY CONDITIONS

We deal with the boundary condition on the random surface for the spherical wave incidence and calculate the expansion coefficients A_l^c and $A_n^{ll'}$ to obtain the scattered wave field. As mentioned in Sec. V, the wave field is a homogeneous l-vector random field so that we can calculate only the value at the polar point $\mathbf{r}\equiv r\mathbf{e}_0$, putting $\mathbf{e}_{(l)m}(\mathbf{r})\to\mathbf{e}_{(l)m}$ and $B_{l'}^m(\mathbf{r})\to B_{l'}^m$ in the equation: The value of the field at arbitrary point $\mathbf{r}\equiv r\mathbf{e}_0$ can be obtained from this by means of the transformation (15) and (48). Since the l-vector field has $(2l+1)$ components, the number of equations for the components arising from the boundary condition (43) or (44) increases with increasing l: Such complication does not occur in the 2-D scattering problem since the cylindrical wave field is a scalar quantity under circular rotations.[21]

A. Dirichlet condition

Substituting (57) and (58) into (43), and using the boundary condition for (49), the equation is first separated into the two equations for the coherent (average) part and the incoherent part:

$$\left[\psi_{(l)}^c(\mathbf{r})+\left\langle f(\mathbf{r},\omega)\frac{\partial\psi_{(l)}^{ic}(\mathbf{r},\omega)}{\partial r}\right\rangle\right]_{\mathbf{r}=a\mathbf{e}}=0, \quad (85)$$

$$\left[\psi_{(l)}^{ic}(\mathbf{r},\omega)+f(\mathbf{r},\omega)\frac{\partial\psi_{(l)}^0(\mathbf{r})}{\partial r}\right]_{\mathbf{r}=a\mathbf{e}}=0, \quad (86)$$

where (86) retains only the linear terms with respect to surface roughness, the higher-order terms being neglected according to our assumption, whereas (85) is the second-order equation.

First, substituting (85), (49), and (61) into (86), and putting $G_l\equiv\sqrt{(2l+1)/4\pi}F_l$ for notational brevity, we obtain the equation

$$\sum_{l'=0}^{\infty}\left[\sum_{m=-L}^{L}H_m^{ll'}(ka)\mathbf{e}_{(l)m}B_{l'}^m+k\dot{\psi}_l^0(ka)G_{l'}\mathbf{e}_{(l)0}B_{l'}^0\right]=0, \quad (87)$$

where the overdot denotes differentiation. Using (56) and orthogonality relations (5) and (47), Eq. (87) turns into

$$H_m^{ll'}(ka)\equiv\sum_{n=-L}^{L}h_{mn}^{(1)l'l}(ka)A_n^{ll'}=-\delta_{m0}kG_{l'}\dot{\psi}_l^0(ka),\quad m=-L,\ldots,L, \quad (88)$$

which is a set of linear equations to be solved for $A_n^{ll'}$, $n=-L,\ldots,L$. From $A_n^{ll'}$ we obtain $H_m^{ll'}(kr)$ by (56), which in turn gives us the incoherent part $\psi_{(l)}^c(\mathbf{r},\omega)$ in the single-scattering approximation.

Next, substituting into (85) the expressions (4), (58) and the incoherent part (61) so obtained, we get the perturbed term for the coherent scattering coefficient:

$$A_l^c=-\frac{1}{h_l^{(1)}(ka)}\sum_{l'=0}^{\infty}k\,\overline{G_{l'}}\dot{H}_0^{ll'}(ka), \quad (89)$$

which is the second-order quantity in $G_{l'}$ or $F_{l'}$ in view of (88), and therefore is given in terms of the power spectrum $|F_{l'}|^2$.

B. Neumann condition

In a similar manner, the boundary condition (44) can be separated into the two equations for the coherent and incoherent parts:

$$\left[\frac{\partial\psi_{(l)}^c}{\partial r}-\langle\nabla_s f\cdot\nabla_s\psi_{(l)}^{ic}\rangle+\left\langle f\frac{\partial^2\psi_{(l)}^{ic}}{\partial r^2}\right\rangle\right]_{\mathbf{r}=a\mathbf{e}_0}=0, \quad (90)$$

$$\left[\frac{\partial\psi_{(l)}^{ic}}{\partial r}-\nabla_s f\cdot\nabla_s\psi_{(l)}^0+f\frac{\partial^2\psi_{(l)}^0}{\partial r^2}\right]_{\mathbf{r}=a\mathbf{e}_0}=0, \quad (91)$$

where (91) retains only linear terms in surface roughness. Here, ∇_s denotes the angular part ($m = -1,1$) of the gradient operator $\nabla \equiv (\nabla_-, \partial/\partial r, \nabla_+)$ written in the canonical basis ($m = -1,0,1$), so that ∇_s is expressed as

$$\nabla_s \equiv (\nabla_-, \nabla_+) = (\overline{H_-}, \overline{H_+})/\sqrt{2}r \tag{92}$$

(cf., Ref. 25) where $H_\pm$ ($= -\overline{H_\mp}$) are differential operators[22] such that

$$H_- T^l_{mn} = \beta^l_n T^l_{mn-1}, \quad H_+ T^l_{mn} = \beta^l_{n+1} T^l_{mn+1},$$
$$\beta^l_n \equiv \sqrt{(l+n)(l-n+1)}. \tag{93}$$

Applying this to (15) and (48) yields the following relations:

$$H_- \mathbf{e}_{(l)n} = \beta^l_n \mathbf{e}_{(l)n-1}, \quad H_+ \mathbf{e}_{(l)n} = \beta^l_{n+1} \mathbf{e}_{(l)n+1}, \tag{94}$$

$$\overline{H_-} B^n_l = \beta^l_n B^{n-1}_l, \quad \overline{H_+} B^n_l = \beta^l_{n+1} B^{n+1}_l, \tag{95}$$

r and ω being supressed here. In particular, from (7) and (95), we have

$$\nabla_s f \equiv (\nabla_- f, \nabla_1 f) = \frac{1}{\sqrt{2}r} \sum_{l'=0}^{\infty} G_{l'} (\beta^{l'}_0 B^{-1}_{l'}, \beta^{l'}_1 B^1_{l'}), \tag{96}$$

where we have again put $G_l = \sqrt{(2l+1)/4\pi} F_l$ for notational brevity in the formula.

First, substituting (49) and (61) into the incoherent part (91) and using (94) and (96), we get

$$\sum_{l'=0}^{\infty} \Bigg[\sum_{m=-L}^{L} k \dot{H}^{ll'}_n(ka) \mathbf{e}_{(l)m} B^m_{l'} - \frac{\psi^0_l(ka)}{2a^2} G_{l'} (\beta^l_0 \beta^{l'}_0 (\mathbf{e}_{(l)-1} B^{-1}_{l'} + \mathbf{e}_{(l)1} B^1_{l'})) + k^2 \ddot{\psi}^0_l(ka) G_{l'} B^0_l \mathbf{e}_{(l)0} \Bigg] = 0. \tag{97}$$

In the same manner as (87) we obtain

$$\dot{H}^{ll'}_n(ka) \equiv \sum_{m=-L}^{L} \dot{h}^{(1)ll'}_{mn}(ka) A^{ll'}_n = kG_{l'} \Bigg[(\delta_{m-1} + \delta_{m1}) \frac{\beta^l_0 \beta^{l'}_0}{2(ka)^2} \psi^0_l(ka) - \delta_{m0} \ddot{\psi}^0_l(ka) \Bigg], \quad m = -L,\ldots,L, \quad L \equiv \min(l,l'). \tag{98}$$

This is the set of linear equations to be solved for $(2L+1)$ coefficients $A^{ll'}_n$, $n = -L,\ldots,L$, from which we obtain the incoherent field $\psi^{ic}_{(l)}(\mathbf{r},\omega)$ through (56) and (61).

Next, substituting (49), (96), and the incoherent field (61) so obtained into the boundary condition (90), and after some manipulations using (92), (94), and (95), we finally obtain the coherent scattering coefficient in the following form:

$$A^c_l = -\frac{1}{\dot{h}^{(1)}_l(ka)} \sum_{l'=0}^{\infty} k\, \overline{G_{l'}} \Bigg\{ \ddot{H}^{ll'}_0(ka) + \frac{1}{2(ka)^2} \times [\beta^l_0 \beta^{l'}_0 (H^{ll'}_{-1}(ka) + H^{ll'}_1(ka)) - 2(\beta^{l'}_0)^2 H^{ll'}_0(ka)] \Bigg\}, \tag{99}$$

which is the second-order quantity in $G_{l'}$ expressible in terms of the power spectrum $|F_{l'}|^2$.

VIII. NUMERICAL EVALUATIONS FOR SCATTERING CHARACTERISTICS

Once the approximate solutions are obtained for the expansion coefficients $A^{ll'}_n$ and A^c_l as in (88) and (89) for the Dirichelt conditions and (98) and (99) for the Neumann condition, then we can evaluate various scattering characteristics using the formulas (69) to (84) with an appropriate power spectrum for the random spherical surface. As well known in the nonrandom case, the spherical wave expansion of the type (65) and (68) is effective in the Mie scattering range, so that we choose the Mie parameter to be $ka = 1$ or 2 in the numerical calculation.

There is no typical power spectrum known for a random field on the sphere. So, as in the previous example for the circular case,[21] we again assume the power spectrum of a Gaussian form in l for the sake of numerical calculation:

$$|F_l|^2 = (\sigma^2/N) e^{-K^2 l^2/2}, \quad l = 0,1,2,\ldots, \tag{100}$$

where

$$N \equiv \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) e^{-K^2 l^2/2} \tag{101}$$

is the normalization constant given as a Dirichlet series. The correlation function corresponding to (100)

$$R(\theta) = \frac{\sigma^2}{N} \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) e^{-K^2 l^2/2} P_l(\cos\theta), \quad 0 \leqslant \theta \leqslant \pi, \tag{102}$$

can be numerically calculated as a function of θ: Some examples are given in Fig. 1 for $K = 0.2 - 1.0$, where the parameter $K(<\pi)$ can be roughly considered as the correlation distance (rad) on the sphere; as $K \to 0$, $R(\theta)$ looks more like a Gaussian form.

We can check the validity of the approximate solutions by means of the power conservation law (73). It is to be noticed that the conservation formula does hold also for

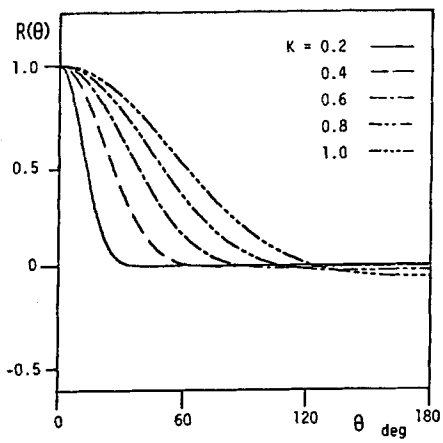


FIG. 1. Correlation function on the sphere (102), plotted as a function of the angular distance. ($\sigma^2 = 1$, $K = 0.2 - 1.0$). $R(0)$ looks more like a Gaussian form as K becomes smaller, K being a correlation distance.

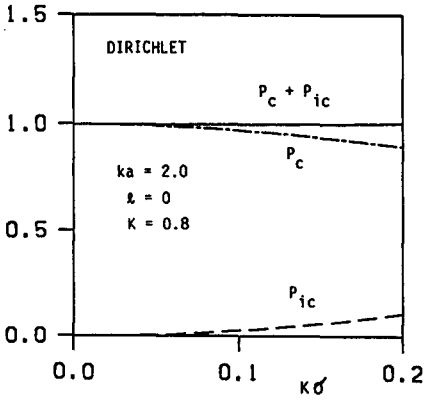


FIG. 2. Power flow conservation for a spherical wave incidence versus the roughness parameter $k\sigma$ (Dirichlet, $l=0$, $ka=2.0$, $K=0.8$). Here, P_c and P_{ic} denote the total coherent power flow and the incoherent power flow, respectively. For a rigorous solution, $P_c + P_{ic} = 1$.

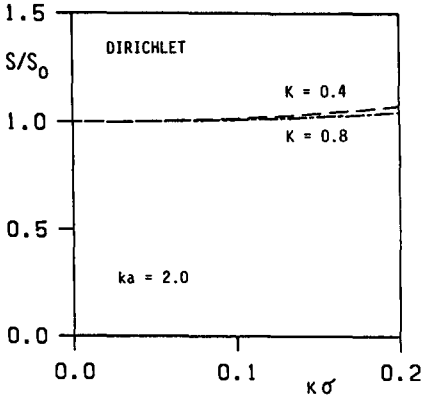


FIG. 4. Optical theorem for a plane-wave incidence (Dirichlet, $ka=2.0$, $K=0.4, 0.8$). Here, S denotes the total cross section calculated by (81) and S_0 the right-hand side of (82) due to the imaginary part of the forward-scattering amplitude. $S/S_0 = 1$ for rigorous solution.

each spherical wave injection. Then the power equality for the l th spherical wave injection can be written

$$1 = P_l^c + P_l^{ic}, \tag{103}$$

where in the present approximation the total coherent power P_l^c and the incoherent power P_l^{ic} can be written, using (74) and (76),

$$P_l^c = |1 + 2(\alpha_l^0 + A_l^c)|^2,$$

$$P_l^{ic} = 4 \sum_{l'=0}^{\infty} \sum_{m=-L}^{L} |H_m^{ll'}|^2. \tag{104}$$

We can use (103) and (104) to check the validity of the single-scattering approximation. An example for $l=0$ is shown in Figs. 2 and 3 for the Dirichlet and Neumann cases, respectively, where $P_l^c + P_l^{ic}$ is plotted against the roughness parameter $k\sigma$: The equality (103) is nearly satisfied within the parameter range $k\sigma < 0.2$ shown in these figures. It is shown by numerical calculations that $1 - P_c$ and P_{ic} rapidly aproach 0 for larger l, so that in the Mie scattering range, the equality (80) consisting of the sum over l does hold for plane-wave incidence. However, to check the power equality in the plane-wave case, we can make use of the optical theorem (82), namely, $S = S_0$, S denoting the total cross section (81) and S_0 the right-hand member of (82) due to the coherent forward-scattering amplitude. Figures 4 and 5 show the ratio S/S_0 plotted against $k\sigma$, which roughly equal 1 within the range shown in the figures.

We are then ready to calculate the angular distributions of coherent scattering (83) for a plane-wave injection, which are shown in Figs. 6 and 7 for the Dirichlet and Neumann case, respectively, with $ka = 1.0$ and 2.0. Correspondingly, the angular distributions of the incoherent scattering calculated by (84) are shown in Figs. 8 and 9, respectively, which show that the incoherent scattering is generally stronger in the backward direction than in the forward in either case. These characteristics shown in Figs. 6–9 are somewhat similar to those for the cylindrical case.[21]

As mentioned at the beginning, the single-scattering approximation is valid in the Mie scattering if $k\sigma$ is small enough; this is due to the absence of a real resonance or a

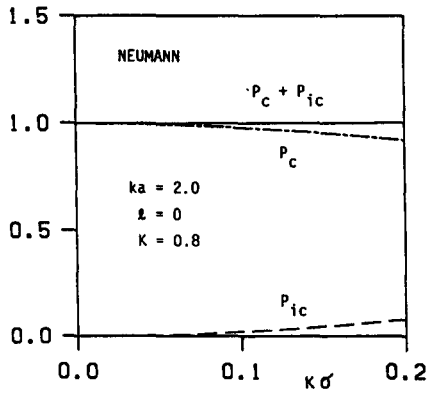


FIG. 3. Power flow conservation for a spherical wave incidence versus the roughness parameter (Neumann, $l=0$, $ka=2.0$, $K=0.8$). Here, P_c and P_{ic} denote the total coherent power flow and the incoherent power flow, respectively. For a rigorous solution, $P_c + P_{ic} = 1$.

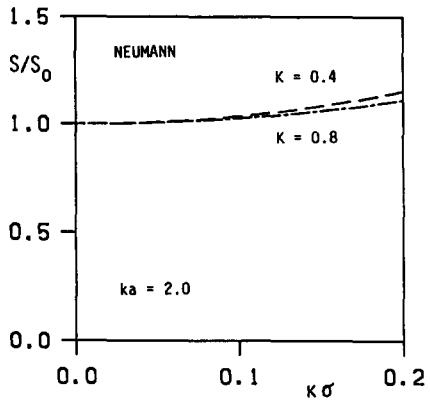


FIG. 5. Optical theorem for a plane-wave incidence (Neumann, $ka=2.0$, $K=0.4, 0.8$). Here, S denotes the total cross section calculated by (81) and S_0 the right-hand side of (82) due to the imaginary part of the forward-scattering amplitude. $S/S_0 = 1$ for rigorous solution.

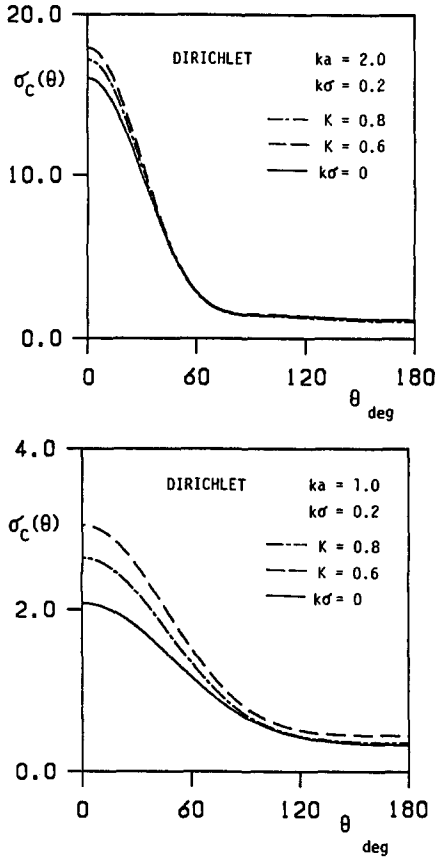


FIG. 6. Angular distribution of the coherent scattering for plane wave incident in the direction $\theta = 0$ (Dirichlet, $ka = 1.0, 2.0$, $k\sigma = 0.2$, $K = 0.4, 0.8$). The solid line shows the case of smooth surface with $k\sigma = 0$.

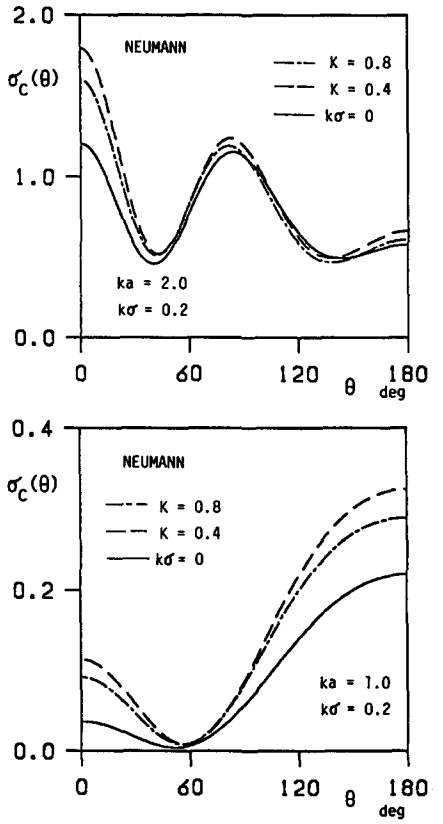


FIG. 7. Angular distribution of the coherent scattering for plane wave incident in the direction $\theta = 0$ (Neumann, $ka = 1.0, 2.0$, $k\sigma = 0.2$, $K = 0.4, 0.8$). The solid line shows the case of smooth surface with $k\sigma = 0$.

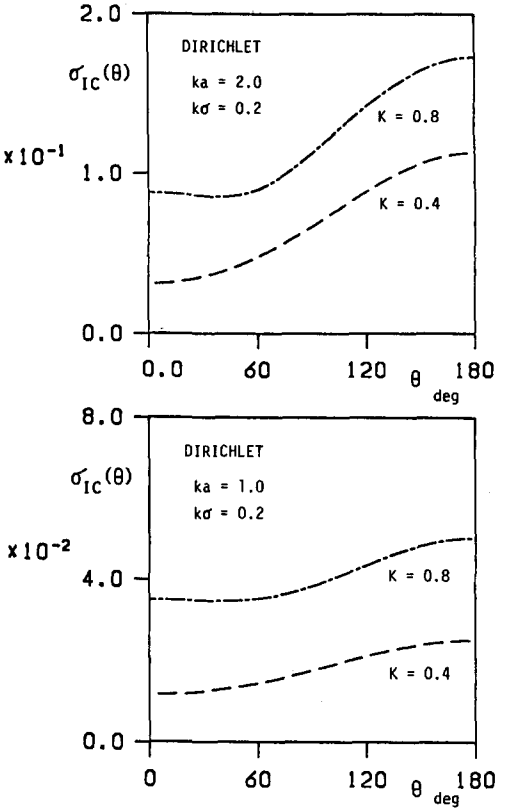


FIG. 8. Angular distribution of the incoherent scattering for a plane-wave incidence (Dirichlet, $ka = 1.0, 2.0$, $k\sigma = 0.2$, $K = 0.4, 0.8$).

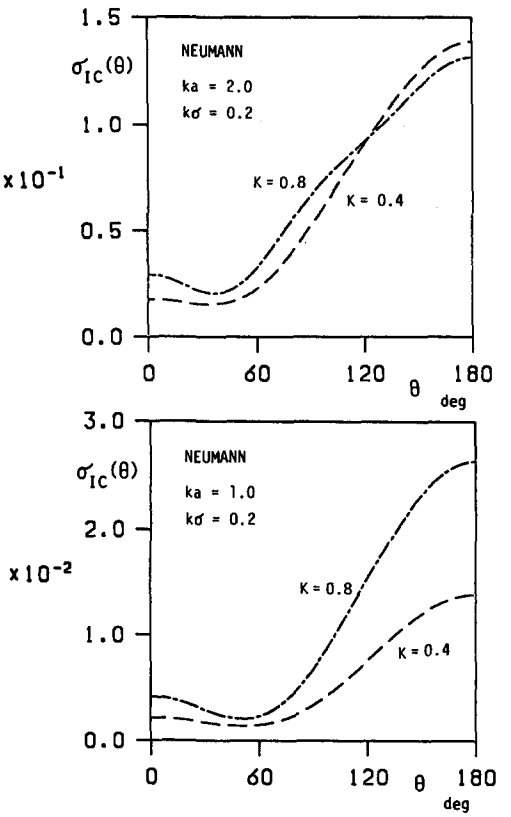


FIG. 9. Angular distribution of the incoherent scattering for a plane-wave incidence (Neumann, $ka = 1.0, 2.0$, $k\sigma = 0.2$, $K = 0.4, 0.8$).

surface mode on a spherical surface. This fact is to be contrasted to the case of planar random surface corresponding to the case $ka \to \infty$, where the multiple scattering has a crucial effect on the scattering characteristics even if the roughness is negligibly small (Neumann surface,[7] perfectly conducting surface,[3] surface plasmon mode[8]). In this regard, when the Mie parameter ka is made much larger beyond the Mie scattering range, the effect of multiple scattering has to be taken into account in the formulas and calculations in a manner similar to Ref. 21. This can be achieved in principle by incorporating the second-order Wiener kernel in the calculation for the approximate solution as remarked at the beginning, which, however, would be much more involved than the present treatment.

ACKNOWLEDGMENT

The authors are indebted to Professor T. Mukai of the Kanazawa Institute of Technology for his suggesting the problem and related references.

APPENDIX: REPRESENTATION OF ROTATION GROUP AND SPHERICAL HARMONICS

Some necessary definitions and notations concerning the representation of the rotation group are briefly summarized for reference and for indicating our choise among various definitions: Several formulas are given in convenient forms for our applications. For details of the theory of the rotation group see Refs. 22 and 23, and also the appendix of Ref. 30.

Rotation group:

A rotation $g \equiv g(\varphi_1,\theta,\varphi_2)$ described by the three Euler angles is defined by the successive rotations in the order: a rotation g_{φ_1} about $\mathbf{e}_z$, g_0 about $\mathbf{e}'_x = g_{\varphi_1}\mathbf{e}_x$, and g_{φ_2} about $\mathbf{e}'_z = g_0\mathbf{e}_z$, $(\mathbf{e}_x,\mathbf{e}_y,\mathbf{e}_z)$ denoting the 3-D unit vectors along x, y, z axes, respectively. The rotations g (identity $\mathbf{e}$, inverse g^{-1}) form the rotation group G, and the rotation of a 3-D vector is represented by the Euler matrix $[g] = [g_{\varphi_1}][g_0][g_{\varphi_2}]$.

Canonical basis:

We call the set of vectors $\mathbf{e}_m$, $m = -1,0,1$, i.e.,

$$\mathbf{e}_0 = \mathbf{e}_z, \quad \mathbf{e}_{\mp} = [\pm\mathbf{e}_x - i\mathbf{e}_y]/\sqrt{2}, \tag{A1}$$

the fixed canonical basis relative to $\mathbf{e}_z$. The Euler matrix is to be represented in the fixed canonical basis. Let $(\mathbf{e}_r,\mathbf{e}_\theta,\mathbf{e}_\varphi)$ be the unit orthogonal vector basis for the polar coordinates $\mathbf{r} \equiv (r,\theta,\varphi)$. The vector $\mathbf{t} \equiv \mathbf{e}_r$ giving a point $\mathbf{t} = (1,\theta,\varphi)$ on the sphere can be written

$$\mathbf{t} \equiv \mathbf{e}_r = g_r\mathbf{e}_z, \quad g_r \equiv g(\varphi + \pi/2,\theta,\varphi_2). \tag{A2}$$

We define the moving canonical basis by the set of vectors, $\mathbf{e}_m(\mathbf{r}) \equiv g_r\mathbf{e}_m$, $m = -1,0,1$, namely,

$$\mathbf{e}_0(\mathbf{r}) \equiv g_r\mathbf{e}_0 = \mathbf{e}_r,$$
$$\mathbf{e}_{\mp 1}(\mathbf{r}) = g_r\mathbf{e}_+ = [\pm\mathbf{e}_\varphi + i\mathbf{e}_\theta]e^{\pm i\varphi_2/\sqrt{2}}, \tag{A3}$$

which is the canonical basis relative to $\mathbf{r}$ or $\mathbf{t}$.

Representation of the rotation group:

For the functions on the sphere $\psi(\mathbf{t})$, $\mathbf{t} \in S_3$, or more generally the functions on G, $\psi(g)$, $g \in G$, we define the transformation S^g by

$$S^g\psi(g_0) \equiv \psi(g^{-1}g_0), \quad g,\ g_0 \in G, \tag{A4}$$

which gives the representation of G:

$$S^{g_1}S^{g_2} = S^{g_1g_2}, \quad [S^g]^{-1} = S^{g^{-1}}, \quad S^e = I. \tag{A5}$$

We denote by D_l the $(2l+1)$-D invariant space of the irreducible representation of weight l, of which the matrix of unitary representation is written

$$\mathbf{T}^l(g) \equiv [T^l_{mn}(\varphi_1,\theta,\varphi_2)], \quad -l \le m,n \le l, \tag{A6}$$
$$T^l_{mn}(\varphi_1,\theta,\varphi_2) \equiv e^{-im\varphi_1}P^l_{mn}(\cos\theta)e^{-in\varphi_2} \tag{A7}$$

where $T^l_{mn}(\mathbf{e}) = \delta_{mn}$ and $T^1_{mn}(g) = g_{mn}$, g_{nm} being the Euler matrix relative to (A1). The matrix representation is referred to the fixed canonical basis in D_l, which is a set of $(2l+1)$ orthogonal vectors of $(2l+1)$ dimension, $\mathbf{e}_{(l)n}$, $n = -l,\ldots,l$: each being the eigenvector with the eigenvalue $e^{-in\varphi_2}$ for the rotation g around $\mathbf{e}_z$. In the present paper, we deal with the representation of the integral weight l: $l = 0,1,2,\ldots$. The matrix element (A7) is called the generalized spherical function of order l (Ref. 22) and in particular for $n = 0$, we have

$$T^l_{m0}(\varphi_1,\theta,\varphi_2) = \sqrt{\frac{4\pi}{(2l+1)}}\, i^m Y^m_l(\theta,\varphi_1), \tag{A8}$$

where $Y^m_l(\theta,\varphi)$ denotes the normalized spherical harmonics:

$$Y^m_l(\theta,\varphi) = (-1)^l\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\,P^m_l(\cos\theta)e^{im\varphi}. \tag{A9}$$

Vector and tensor:

A $(2l+1)$-D vector with components $a_{(l)n}$, $n = -l,\ldots,l$, which is transformed by the matrix $\mathbf{T}^l(g)$ upon rotation g as

$$a'_{(l)m} = \sum_{n=-l}^{l} T^l_{mn}(g)a_{(l)n}, \quad m = -l,\ldots,l \tag{A10}$$

is called an l-vector in D_l; hence, the ordinary 3-D vector transformed by g_{mn} is a 1-vector. Similarly, a $(2l+1)$-D vector transformed by the matrix $\overline{\mathbf{T}^l(g)}$ is called an $\bar{l}$-vector in $\overline{D_l}$ (the overbar implying the complex conjugate). Further, we consider a tensorial quantity in a product space; for instance, a tensor with $(2l'+1)\times(2l+1)$ components $a^{(l')n'}_{(l)n}$ which is transformed under rotation g as

$$a^{(l')n'}_{(l)n} = \sum_{n=-l}^{l}\sum_{n'=-l'}^{l'} \overline{T^{l'}_{mn}(g)}\, T^l_{mn}(g)a^{(l')n'}_{(l)n} \tag{A11}$$

is called for simplicity an $\bar{l}'\times l$ tensor in $\overline{D_{l'}}\times D_l$, where the superscript refers to the component of an $\bar{l}'$-vector in $\overline{D_{l'}}$. The inner product of two l-vectors, $\mathbf{a}_{(l)}$ and $\mathbf{b}_{(l)}$, as well as the contraction of a tensor, can be defined as

$$(\mathbf{a}_{(l)}\cdot\mathbf{b}_{(l)}) = \sum_{m=-l}^{l} \overline{a_{(l)m}}\,b_{(l)m}. \tag{A12}$$

Properties of the matrix:

The unitarity of the representation matrix,

$$\sum_{s=-l}^{l} \overline{T^l_{sm}(g)}\,T^l_{sn}(g) = \delta_{nm}, \tag{A13}$$

can be interpreted as the orthonormal relation of a set of

$(2l+1)$ l-vectors $\mathbf{e}_{(l)n}(\mathbf{r})$ with respect to the inner product; namely

$$(\mathbf{e}_{(l)m}(\mathbf{r})\cdot\mathbf{e}_{(l)n}(\mathbf{r}))=\delta_{mn}, \tag{A14}$$

where we have put

$$\mathbf{e}_{(l)n}(\mathbf{r})\equiv\mathbf{T}^l(g_\mathbf{r})\mathbf{e}_{(l)n}$$
$$=\sum_{s=-l}^{l}T^l_{sn}(g_\mathbf{r})\mathbf{e}_{(l)s},\quad n=-l,...,l, \tag{A15}$$

which is obtained from $\mathbf{e}_{(l)n}$ by rotagion $g_\mathbf{r}$: $T^l_{sn}(g_\mathbf{r})$ in the right-hand side giving the $2l+1$ components in the fixed canonical basis of D_l. The set of $(2l+1)$ vectors (A15) is called the moving canonical basis in D_l relative to $\mathbf{r}$, which is reduced to (A3) for $l=1$. As (A14) shows, the coordinate-free l-vector notation in the left-hand side of (A15) make vector and tensor formulas considerably simpler than the coordinate-fixed notation in the right-hand side.

The multiplicative law of the group representation can be written as

$$T^l_{mn}(g_2g_1)=\sum_{s=-l}^{l}T^l_{ms}(g_2)T^l_{sn}(g_1),$$
$$T^l_{mn}(g_2^{-1}g_1)=\sum_{s=-l}^{l}\overline{T^l_{ms}(g_2)}\,T^l_{sn}(g_1). \tag{A16}$$

The first equality simply shows that $\mathbf{e}_{(l)n}(\mathbf{r})$ is an l-vector, having the property (A10). The second following from the first implies the addition theorem for generalized spherical functions, which can be interpreted as the inner product of two l-vectors:

$$T^l_{mn}(g_2^{-1}g_1)=(\mathbf{e}_{(l)m}(\mathbf{r}_2)\cdot\mathbf{e}_{(l)n}(\mathbf{r}_2)), \tag{A17}$$

where $\mathbf{r}_1\equiv g_1\mathbf{e}_0$, $\mathbf{r}_2\equiv g_2\mathbf{e}_0$; (A17) reduces to(A14) when $g_1=g_2$. In particular, for $m=n=0$, (A17) gives the well-known addition formula for the zonal spherical function;

$$P_l(\cos\theta)=(\mathbf{e}_{(l)0}(\mathbf{r}_2)\cdot\mathbf{e}_{(l)0}(\mathbf{r}_1))$$
$$=\frac{4\pi}{2l+1}\sum_{m=-l}^{l}\overline{Y^m_l(\theta_1,\varphi_1)}\,Y^m_l(\theta_2,\varphi_2) \tag{A18}$$

$$\cos\theta=\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2). \tag{A19}$$

The integration over G has the invariance properties under rotational transformation of the variable:

$$\int_G f(g)dg=\int_G f(gg_0)dg=\int_G f(g_0g)dg=\int_G f(g^{-1})dg,$$
$$f(g)\equiv f(\varphi_1,\theta,\varphi_2),\quad dg=d\varphi_1\sin\theta\,d\theta\,d\varphi_2. \tag{A20}$$

For the function $f(\mathbf{t})$ on S_3 (A20) implies the integral over S_3 multiplied by 2π. The orthogonality and the completeness of the set of generalized spherical functions are written as

$$\int_G\overline{T^{l'}_{m'n'}(g)}\,T^l_{mn}(g)dg=\delta_{ll'}\delta_{mm'}\delta_{nn'}\frac{8\pi^2}{2l+1}, \tag{A21}$$

$$\frac{1}{8\pi^2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{n=-l}^{l}(2l+1)\,\overline{T^l_{mn}(g)}\,T^l_{mn}(g')$$
$$=\delta(g-g'). \tag{A22}$$

Vector and tensor fields:

Upon rotation g, an l-vector field on R_3 is transformed into a new vector field by the formula

$$a'_{(l)m}(\mathbf{r})=\sum_{n=-l}^{l}T^l_{mn}(g)a_{(l)n}(g^{-1}\mathbf{r}),\quad r\in R_3. \tag{A23}$$

In a similar manner, a tensor field, e.g., an $\overline{l'}\times l$-tensor field is transformed according to

$$a'^{(l')m'}_{(l)m}(\mathbf{r})$$
$$=\sum_{n=-l}^{l}\sum_{n'=-l'}^{l'}\overline{T^{l'}_{m'n'}(g)}\,T^l_{mn}(g)a^{(l')n'}_{(l)n}(g^{-1}\mathbf{r}). \tag{A24}$$

Isotropic vector field and isotropic tensor field:

When a vector or a tensor field is invariant under rotations, for instance, when

$$a'_{(l)m}(\mathbf{r})=a_{(l)m}(\mathbf{r}), \tag{A25}$$
$$a'^{(l')m'}_{(l)m}(\mathbf{r})=a^{(l')m'}_{(l)m}(\mathbf{r}), \tag{A26}$$

hold for (A23) and (A24), then the l-vector field or the $\overline{l'}\times l$-tensor field is said to be isotropic. It is easily shown[30] that, when referred to the moving canonical basis, an isotropic l-vector field has only the 0th canonical component depending on $r=|\mathbf{r}|$, that is

$$a_{(l)m}(\mathbf{r})=\delta_{m0}a(r),\quad m=-l,...,l \tag{A27}$$

and that similarly an isotropic $\overline{l'}\times l$-tensor field has the components only for $m'=m$;

$$a^{(l')m'}_{(l)m}(\mathbf{r})=\delta_{mm'}a_m(r),$$
$$m=-l,...,l,\quad m'=-l',...,l'. \tag{A28}$$

For the isotropic field on the sphere $r=1$, the components are constants.

Vector harmonic functions:

For reference in the text we summarize the definitions and formulas concerning the l-vector spherical and solid harmonics that are derived from the representation of the rotation group.[30,31]

Let an l-vector function on S_3 having only nth canonical component be

$$\mathbf{P}^{l'm}_{(l)n}(\theta,\varphi)\equiv\sqrt{\frac{(2l'+1)}{4\pi}}\,\overline{T^{l'}_{mn}(g)}\,\mathbf{e}_{(l)n}(\mathbf{r}), \tag{A29}$$
$$n=-l,...,l,l'=0,1,2,...,m=-l',...,l'.$$

in the coordinate-free notation. The l-vector function (A29) is called the l-vector spherical harmonic and satisfies the orthogonality relation,

$$\int_{S_3}(\mathbf{P}^{l'm'}_{(l)n}\cdot\mathbf{P}^{l''m''}_{(l)n''})\,dS=\delta_{n'n''}\delta_{l'l''}\delta_{m'm''},$$
$$dS\equiv\sin\theta\,d\theta\,d\varphi. \tag{A30}$$

Let l-vector functions on R_3 be defined by

$$\mathbf{J}^{l'm}_{(l)n}(kr,\theta,\varphi)=\sum_{t=-l}^{l}j^{l'l}_{nt}(kr)\mathbf{P}^{l'm}_{(l)t}(\theta,\varphi), \tag{A31}$$

which we call the l-vector solid harmonic, where $j^{l'l}_{nt}(kr)$ is defined by

$$j^{ll'}_{mn}(kr)=\sum_{L=|l-l'|}^{l+l'}i^{L-l+l'}(-1)^{m+n}(l-ml'm|ll'L0)$$
$$\times(l-nl'n|ll'L0)j_L(kr), \tag{A32}$$

$j_L(kr)$ being the spherical Bessel function and $(l-ml'm|ll'L0)$ denoting the Clebsch–Gordan coefficient.[22,30] Here, $j^{ll'}_{mn}(kr)$ defined by (A32) is called the generalized spherical Bessel function, having orthogonality with respect to integration, and is derived from the matrix element of the translation group in R_3.[30] The l-vector solid harmonics are shown to satisfy the l-vector Helmholtz equation,

$$(\nabla^2 + k^2)\mathbf{J}^{l'm}_{(l)n}(kr,\theta,\varphi) = 0\,, \tag{A33}$$

and the orthogonality relation;

$$\frac{2}{\pi}\int_0^\infty \int_{S_3} \left(\mathbf{J}^{l'm'}_{(l)n'}(k'r,\theta,\varphi)\cdot\mathbf{J}^{l''m''}_{(l)n''}(k''r,\theta,\varphi)\right) dS\, r^2\, dr$$
$$= \delta_{n'n''}\delta_{l'l''}\delta_{m'm''}[\delta(k'-k'')/k'^2]\,. \tag{A34}$$

The following vector and tensor integral representations hold for l-vector harmonic functions:

$$\mathbf{J}^{l'm}_{(l)n}(kr,\theta,\varphi)$$
$$= \frac{1}{4\pi i^{l'-l}}\int_{S_3} e^{ik\cdot\mathbf{r}}\mathbf{P}^{l'm}_{(l)n}(u,v)dS\,, \tag{A35}$$

$$\sum_{t=-L}^{L} j^{ll'}_{nt}(kr)\mathbf{e}_{(l)t}(\mathbf{r})\,\overline{\mathbf{e}_{(l')t}(\mathbf{r})}$$
$$= \frac{1}{4\pi i^{l'-l}}\int_{S_3} \mathbf{e}_{(l)n}(\mathbf{k})\,\overline{\mathbf{e}_{(l')n}(\mathbf{k})}e^{i\mathbf{k}\cdot\mathbf{r}}dS$$
$$(L = \min(l,l'))\,, \tag{A36}$$

where $\mathbf{k} = (k,u,v)$ in the polar coordinates and $dS = \sin u\, du\, dv$. These two are equivalent representations with different interpretation; the first is written as the Fourier transform of the l-vector field over a sphere, while the second gives the Fourier transform of an isotropic $l\times\bar{l}'$-tensor field over the sphere.

Analogous to $j^{ll'}_{mn}$ and $\mathbf{J}^{(l')n}_{(l)n}$ given in terms of $j_L(kr)$ we can define $h^{(1)l'm}_{mn}$ and the solid harmonics $H^{(1)l'm}_{(l)n}$ in terms of spherical Hankel function $h^{(1)}_L(kr)$. The l-vector solid harmonic $\mathbf{H}^{(1)l'm}_{(l)n}$ satisfies the Helmholtz equation (A33) also and has similar integral representations. The definitions and formulas for l-vector harmonics are reduced to the vector case for $l=1$ and to the familiar scalar case for $l=0$.

[1]J. Nakayama, H. Ogura, and B. Matsumoto, Radio Sci. **15**, 1049 (1980).
[2]J. Nakayama, H. Ogura, and M. Sakata, J. Math. Phys. **22**, 471 (1981).
[3]J. Nakayama, H. Ogura, and M. Sakata, Radio Sci. **16**, 831 (1981).
[4]J. Nakayama, M. Sakata, and H. Ogura, Radio Sci. **16**, 847 (1981).
[5]J. Nakayama, Radio Sci. **17**, 558 (1982).
[6]H. Ogura and N. Takahashi, J. Opt. Soc. Am. A **2**, 2208 (1985).
[7]J. Nakayama, Ph.D. dissertation, Kyoto University, Kyoto, Japan, 1982.
[8]J. Nakayama, K. Mizutani, H. Ogura, and S. Hayashi, J. Appl. Phys. **56**, 1465 (1984).
[9]N. Wiener, *Nonlinear Problems in Random Theory* (MIT, Cambridge, MA, 1958).
[10]K. Ito, J. Math. Soc. Jpn. **13**, 157 (1951).
[11]K. Ito, Jpn. J. Math. **22**, 63 (1952).
[12]H. Ogura, Phys. Rev. A **11**, 942 (1975).
[13]*Light Scattering by Irregularly Shaped Particles*, edited by D. W. Schuerman (Plenum, New York, 1980).
[14]S. Mukai, T. Mukai, R. H. Giese, K. Weiss, and R. H. Zerull, Moon Planets **26**, 197 (1982).
[15]R. Schiffer, Astron. Astrophys. **148**, 347 (1985).
[16]C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).
[17]S. Hayashi, Jpn. J. Appl. Phys. **23**, 665 (1984).
[18]T. Oguchi, Proc. IEEE **71**, 1029 (1983).
[19]R. Schiffer and K. O. Thielheim, J. Appl. Phys. **57**, 2437 (1985).
[20]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[21]H. Ogura and H. Nakayama, J. Math. Phys. **29**, 851 (1988).
[22]I. M. Gel'fand, R. A. Minlos, and Z. Ya Shapiro, *Representation of the Rotation and Lorentz Groups and their Applications*, English translation (Pergamon, London, 1963).
[23]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*, English translation (American Mathematical Society, Providence, 1968).
[24]A. M. Yaglom, Theory Prob. Appls. **2**, 273 (1957).
[25]A. M. Yaglom, Proc. Fourth Berkeley Symp. Math. Stat. Prob. **2**, 593 (1961).
[26]A. I. Ponomarenko, Sov. Math. Dokl. **14**, 582 (1973).
[27]R. H. Jones, Ann. Math. Statist. **34**, 1 (1963).
[28]M. I. Yadrenko, *Spectral Theory of Random Fields*, English translation (Optimization Software, Inc., New York, 1983).
[29]J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[30]H. Ogura, J. Phys. Soc. Jpn. **25**, 586 (1968).
[31]H. Ogura, J. Phys. Soc. Jpn. **24**, 1370 (1968).

The trivialization of constraints in quantum theory (working in a general gauge/parametrization)

A. Y. Shiekh
International Centre for Theoretical Physics, Miramare, Trieste, Italy and Department of Physics, University of Waterloo, Waterloo, Ontario N2L 5V1, Canada

(Received 30 October 1987; accepted for publication 2 August 1989)

A way is found to trivialize and so deal with constraints in a gauge invariant manner.

I. CONSTRAINTS IN CLASSICAL THEORY

In the Lagrange formulation of mechanics (see Goldstein[1] and Landau and Lifshitz[2]) one has a scalar quantity (the action $\int L\,dt$) from which the classical (Euler–Lagrange) equations of motion

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_n}\right)=\frac{\partial L}{\partial q_n}$$

follow from an extremizing principle where, for the variation, the coordinates are treated as independent of each other

If the motion is constrained to a surface,

$$\phi_m(q,t)=0,$$

then the independence of the coordinates may be restored by supplementing the Lagrangian:

$$L=L_0+\lambda_m\phi_m$$

and treating the Lagrange multipliers λ_m in the manner of additional independent coordinates. Extremization with respect to the λ_m then yields the constraints as equations of motion. If a constraint is nontrivial (that is, alters the motion), then it follows that the associated λ_m is determined.

When moving from the Lagrange to Hamilton formalism,

$$H(p,q,t)\equiv\sum_n p_n\dot{q}_n-L(q,\dot{q},t),\quad p_n\equiv\frac{\partial L}{\partial \dot{q}_n}.$$

In addition to an imposed constraint, a similar but different kind of constraint can arise (Dirac,[3] and references therein). In the Hamilton formulation the equations of motion follow from arbitrary variations of the q_n and p_n leading to

$$\dot{q}_n=\frac{\partial H}{\partial p_n},\quad \dot{p}_n=-\frac{\partial H}{\partial q_n},$$

or, in condensed notation,

$$\dot{g}=\{g,H\}+\frac{\partial g}{\partial t},$$

where the Poisson bracket is defined by

$$\{f,g\}\equiv\frac{\partial f}{\partial q_n}\frac{\partial g}{\partial p_n}-\frac{\partial f}{\partial p_n}\frac{\partial g}{\partial q_n}.$$

If the action has symmetries, then there exist relations

$$\delta_m^{\text{sym}}\int L(q,\dot{q},t)dt=0,$$

which become the constraints (explicitly time-dependent symmetry constraints are ignored for simplicity)

$$\phi_m(p,q)=0$$

in the Hamilton formalism. These constraints that arise directly from the symmetry are called primary. To reinstate the independence of the q_n and p_n the same technique as before may be adopted to yield the supplemented Hamiltonian

$$H=H_0+\lambda_m\phi_m,$$

with the important distinction that these constraints are not externally imposed but arise from invariances of the system. In contrast to previously, where the constraints in general alter the motion and the Lagrange multipliers are determined, here the Lagrange multipliers should be undetermined. However, this is not the end of the story since there are consistency conditions that must be satisfied for the constraints to be maintained for all times; namely,

$$\dot{\phi}_m=\{\phi_m,H\}\approx 0,$$

i.e.,

$$\{\phi_m,H_0\}+\lambda_k\{\phi_m,\phi_k\}\approx 0,$$

where $\approx$ means equality by virtue of the constraints (weakly equal). These consistency conditions might imply further (so-called secondary) constraints (to be attached to the Hamiltonian with associated Lagrange multipliers and which must also undergo the consistency conditions, so possibly leading to further secondary constraints). If all the constraints are first class [commute (in the sense of zero Poisson bracket) with the total Hamiltonian for all λ_m], then the Lagrange multiplies are truly arbitrary (as they must be for a symmetry constraint) and the system is in a consistent condition. However, if the constraints are second class (do not commute), then conditions are imposed upon the Lagrange multipliers in contradiction to their undetermined nature, and the Hamiltonian as it stands is therefore inconsistent. Dirac showed a way to eliminate these second class constraints via the now so-called "Dirac bracket,"[3] which yields a first class (and so consistent) Hamilton formulation of the system. But an alternative approach exists; namely, to apply the Lagrange multiplier conditions to the total Hamiltonian, which then also leads to a first class formulation, albeit different from Dirac's.

The restrictions upon the Lagrange multipliers fall into two distinct classes: (I) $\lambda\neq 0$, where the system is altered by the constraint; or (II) $\lambda=0$, which is a condition selectively removing the offending second class constraint.

Having applied these conditions to obtain a consistent "pilot" Hamiltonian (denoted by a prime)

$$H=H_0'+\lambda_{m'}\phi_{m'},$$

it is easy to see that the now first class $\phi_{m'}$ are generators of

 0022-2488/90/010076-07$03.00

gauge transformations, since, for any $f(q,p)$,

$$\dot{f}=\{f,H\}+\frac{\partial f}{\partial t}=\{f,H_0'\}+\frac{\partial f}{\partial t}+\lambda_{m'}\{f,\phi_{m'}\},$$

which includes an arbitrary part given by

$$\delta^{\text{arb}}(f)=\lambda_{m'}\{f,\phi_{m'}\}\delta t,$$

so that the $\phi_{m'}$ generate arbitrary, nonphysical, changes; each $\lambda_{m'}$ choice corresponding to a particular gauge.

This procedure has a very nice interpretation; namely, since a first class constraint is the generator of a gauge transformation, and a second class constraint is indicative of gauge fixing (leads to a particular choice of λ_m), (this is the reason why second class constraints come in even numbers; since an invariance gives rise to one constraint and its gauge fixing another), the presence of second class constraints in an ungauged system is symptomatic of incompatible symmetries. The automated repair consists of removing the unintentional gauge choice (which does not alter the equations of motion) and restoring compatiblity by extending the system.

In general, a second class formulation, be it so through intentional gauge fixing or incompatible symmetries, is unsuitable for quantizing. Path integral quantization is obstructed because one cannot integrate over Lagrange multipliers that are not undetermined, while operator quantization (which is itself ambiguous[4])

$$\{f,g\}\rightarrow(i/\hbar)[\hat{f},\hat{g}]$$

(where the hat symbolizes an operator), leads to a contradiction, since second class constraints, by definition, do not commute, but are individually zero. Once a first class formulation is achieved the system may be quantized without obstruction, where symmetries lead to another dilemma. Before discussing this new problem and its solution, the technique developed above might be profitably illustrated upon a minimal example.

Consider, therefore, the system characterized by the Lagrangian

$$L=q_2\dot{q}_1.$$

The equations of motion follow as

$$\dot{q}_2=0,\quad \dot{q}_1=0.$$

Now move to the Hamilton formulation

$$p_1\equiv\frac{\partial L}{\partial\dot{q}_1}=q_2,\quad p_2\equiv\frac{\partial L}{\partial\dot{q}_2}=0,$$

which therefore has the primary constraints

$$\phi_1\equiv p_1-q_2=0,\quad \phi_2\equiv p_2=0,$$

which are second class (noncommuting) since $\{\phi_1,\phi_2\}\not\approx 0$. This leads to the pilot Hamiltonian

$$H=\lambda_1(p_1-q_2)+\lambda_2(p_2).$$

The consistency conditions read as

$$\dot{\phi}_1=\{\phi_1,H\}=-\lambda_2=0,\quad \dot{\phi}_2=\{\phi_2,H\}=\lambda_1=0.$$

Applying these yields

$$H=0,$$

which, although leading to a first class system and reproducing the correct equations of motion

$$\dot{p}_1=0,\quad \dot{q}_1=0,$$
$$\dot{p}_2=0,\quad \dot{q}_2=0,$$

would seem not to regenerate the original Lagrangian

$$L=p_1\dot{q}_1+p_2\dot{q}_2-H.$$

However, it should be recalled that the removal of second class constraints indicates that the pilot Hamiltonian possesses extended gauge freedom. This is confirmed by the ability to pick the originally inconsistent constraints as the gauge fixing conditions:

$$p_1=q_2,\quad p_2=0,$$

which then correctly leads back to the original Lagrangian

$$L=q_2\dot{q}_1.$$

To avoid returning to a second class formulation, the second class constraints should be applied as gauge fixing conditions after quantization, where such troubles are avoided.

II. CONSTRAINTS IN QUANTUM THEORY

Having obtained a first class formulation of the classical theory one might quantize it via the Hamilton (q,p) path integral formulation of quantum theory. Formally one would need to deal with objects like

$$\int_{-\infty}^{\infty}\cdots D\lambda\int_{-\infty}^{\infty}\cdots Dq\int_{-\infty}^{\infty}\cdots\frac{Dp}{2\pi\hbar}$$
$$\times\exp\left[\frac{i}{\hbar}\int_{t_a}^{t_b}[p\dot{q}-(H_0+\lambda\phi)]dt\right],$$

where the labels on λ, q, and p are suppressed. This is formal in that the path integral depends upon the finite difference scheme adopted in its time discretization as well as the end point conditions.[5–9] On top of this ambiguity the above integral is divergent, since the infinite number of integrations over λ, while imposing the constraint, overcounts by integrating over equivalent systems (each choice of λ corresponding to a particular gauge). One way of dealing with this is to pick a specific gauge in order to isolate and so discard the overcounting factor[10,11] (this must inevitably involve subtleties related to the obstruction to quantizing a gauged classical Hamiltonian theory; the procedure is less involved in the Lagrange formalism[12]). It would be more esthetically pleasing, however, to identify and factor out the overcounting without recourse to a particular choice of gauge.

Following the philosophy of Omnes,[13] one might consider dealing with a constrained system by transforming it to an equivalent unconstrained one. A direct approach to transforming away the constraint is not viable; for consider a canonical transformation generated by $F(q,P,t)$:

$$K=H+\left(\frac{\partial F}{\partial t}\right)_{qP},\quad p=\left(\frac{\partial F}{\partial q}\right)_{Pt},\quad Q=\left(\frac{\partial F}{\partial P}\right)_{qt}.$$

If one calls upon this transformation to trivialize the constraint $\phi(q,p,t)\approx 0$, then the generating function is determined from

$$\phi\left(q, \frac{\partial F}{\partial q}, t\right) = 0,$$

and so

$F =$ determined function (q,t)

$+$ undetermined function (P,t).

Hence

$p =$ function (q,t), $Q =$ function (P,t),

which does *not* bridge the two spaces!

One might instead follow the Hamilton–Jacobi philosophy[1] and require that the *entire* system with constraint be trivialized. Starting with the general Hamiltonian with constraint

$$H(q,p,t) = H_0(q,p,t) + \lambda(t)\phi(q,p,t),$$

and performing a Hamilton–Jacobi transformation based on $F(q,P,t)$, F is determined (albeit not uniquely) from

$$H_0\left(q, \frac{\partial F}{\partial q}, t\right) + \alpha + \frac{\partial F}{\partial t}(q,P,t) = 0,$$

$$\lambda(t)\phi\left(q, \frac{\partial F}{\partial q}, t\right) = \alpha,$$

which is a generalized Hamilton–Jacobi equation, where α is to become an undertermined (although restricted) function of P. This might be used to advantage on the Hamilton (q,p) path integral. However, up to this point, the discussion has been based on canonical transformations in the context of classical mechanics and formal expressions for the path integral. For this reason we must digress to a discussion of canonical transformations in the context of the path integral.

In classical mechanics a canonical transformation is defined to be one that preserves Hamilton's equations. A quantum canonical transformation might analogously be defined to be one that leads to a path integral representation [6,9,14–16] in the new variables if one existed in the old ones, i.e., formally, with the end points (a,b) in phase space held fixed:

$$\int_{-\infty}^{\infty}\cdots Dq\int_{-\infty}^{\infty}\cdots\frac{Dp}{2\pi\hbar}\exp\left[\frac{i}{\hbar}\int_{t_a}^{t_b}[p\dot{q}-H]dt\right]$$
$$=\exp\left[\frac{i}{\hbar}(\Gamma_b-\Gamma_b)\right]\int_{-\infty}^{\infty}\cdots DQ\int_{-\infty}^{\infty}\cdots\frac{DP}{2\pi\hbar}$$
$$\times\exp\left[\frac{i}{\hbar}\int_{t_a}^{t_b}[P\dot{Q}-K]dt\right]$$

$(\Gamma = F - QP)$.

It has been found[17,18] that this formal statement becomes true for the Stratonovich[19] (midpoint scheme) coherent state path integral

$$\langle p_b,q_b,t_b|p_a q_a t_a\rangle$$

$$\equiv\lim_{N\to\infty}\prod_{j=1}^{N-1}\left(\int_{-\infty}^{\infty}dq(j)\int_{-\infty}^{\infty}\frac{d\bar{p}(j)}{2\pi\hbar}\right)$$

$$\times\exp\left[\frac{i}{\hbar}\sum_{k=1}^{N}(\bar{p}(k)(q(k)-q(k-1))\right.$$

$$\left.-\mathbf{H}(\bar{p}(k),\bar{q}(k),k-1/2)\Delta t)\right],$$

where the Hamiltonian is altered (which may give rise to quantum anomalies that spoil classical symmetries that might be present):

$$\mathbb{H}\equiv H+\frac{\hbar}{4}\left(\frac{\partial^2 H}{\partial p^2}\right)_{p=0},\quad \Delta t\equiv(t_b-t_a)/N,$$

$$\bar{p}(k)\equiv\tfrac{1}{2}(p(k)+p(k-1)),\quad \bar{q}(k)\equiv\tfrac{1}{2}(q(k)+q(k-1)),$$

$$p_b\equiv p(N),\quad p_a\equiv p(0),$$

$$q_b\equiv q(N),\quad q_a\equiv q(0).$$

This understanding gives formal manipulations validity, and it is found that quantum canonical transformations are, in fact, a proper subset of classical canonical transformations[18] (scaling transformations, being normalization ruining, are disallowed).

Coherent states, it should be recalled, are not observable (the reason why the end points could be fixed in phase space without violating Heisenberg's uncertainty principle), and it still remains to convert the coherent state path integral to a physical amplitude. For example, the position to position amplitude

$$\langle x_b,t_b|x_a,t_a\rangle=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dq_b\int_{-\infty}^{\infty}dp_b\,\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dq_a\int_{-\infty}^{\infty}dp_a$$
$$\times\frac{1}{\sqrt{\pi\hbar}}\exp\left[-\frac{1}{2\hbar}((x_b-(q_b+ip_b))^2+p_b^2)\right]\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle\exp\left[-\frac{1}{2\hbar}((x_a-(q_a-ip_a))^2+p_a^2)\right].$$

For the particular case of the trivializing (Hamilton–Jacobi) transformation this reduces to a double integral[18]

$$\langle x_b,t_b|x_a,t_a\rangle=\frac{1}{2(\pi\hbar)^{3/2}}\int_{-\infty}^{\infty}dQ\int_{-\infty}^{\infty}dP\exp\left[-\frac{1}{2\hbar}((x_b-(q_b+ip_b))^2+p_b^2)\right]\exp\left[\frac{i}{\hbar}(\Gamma_b-\Gamma_a)\right]$$
$$\times\exp\left[-\frac{1}{2\hbar}((x_a-(q_a-ip_a))^2+p_a^2)\right],$$

where the Hamilton–Jacobi analysis yields

$$q_a(P,Q,t_a),\quad p_a(P,Q,t_a),\quad \Gamma_a(P,Q,t_a),$$
$$q_b(P,Q,t_b),\quad p_b(P,Q,t_b),\quad \Gamma_b(P,Q,t_b),$$

remembering to employ the altered Hamiltonian. This still leaves a subtlety concerning the sequence of performing the momentum and position integrals, a matter dealt with in the Appendix.

Returning, then, to the constraint case with the now well specified representation of the path integral

$$\int_{-\infty}^{\infty}\cdots D\lambda\int_{-\infty}^{\infty}\cdots Dq\int_{-\infty}^{\infty}\cdots\frac{Dp}{2\pi\hbar}$$
$$\times\exp\left[\frac{i}{\hbar}\int_{t_a}^{t_b}[p\dot{q}-(H_0+\lambda\phi)]dt\right].$$

The above expression, being understood to be a coherent state amplitude, has fixed q and p end points. It is convenient then to also not include the end point λ integrations in this expression and perform all the end point integrations together when converting to a physical amplitude. Trivializing this object through the generalized Hamilton–Jacobi transformation given previously yields:

$$\exp\left[\frac{i}{\hbar}(\Gamma_b-\Gamma_a)\right]\int_{-\infty}^{\infty}\cdots DQ\int_{-\infty}^{\infty}\cdots\frac{DP}{2\pi\hbar}$$
$$\times\exp\left[\frac{i}{\hbar}\int_{t_a}^{t_b}[P\dot{Q}]dt\right]\int_{-\infty}^{\infty}\cdots D\lambda.$$

Here the infinite gauge factor ($\int D\lambda$, not including end points) is isolated and may be factored out (abandoned) and the path integral performed to yield

$$\exp[(i/\hbar)(\Gamma_b-\Gamma_a)]\delta(Q_b-Q_a)$$
$$\times 2\pi\hbar\,\delta(P_b-P_a)\delta(\lambda_b-\lambda_a).$$

The delta function in λ follows from the delta function in Q and P, since once the constraint is enforced at one end, it is then automatically satisfied at the other. It still remains to transform this coherent state amplitude to an amplitude between physical states when the remaining end point λ dependence is integrated out.

III. AN EXPLICIT EXAMPLE

To illustrate the technique one might investigate the reparameterization invariant free massive point particle[20,21] as an example of a system with general covariance freedom. This taking a "hammer to crack a nut" approach has the virtue of illustrating the technique in an uncluttered example, but suffers the penalty of seeming cumbersome.

Starting from the Lagrangian formalism,

$$S=-m\int\sqrt{\frac{\partial q_\mu}{\partial\tau}\frac{\partial q^\mu}{\partial\tau}}\,d\tau.$$

Following Dirac,[3] this has a trivial Hamiltonian if one fails to account for the primary constraint $p_\mu p^\mu=m^2$; hence the Hamilton formalism

$$H=\frac{\lambda(\tau)}{2m}(p_\mu p^\mu-m^2),$$
$$S=\int\left(p_\mu\dot{q}^\mu-\frac{\lambda(\tau)}{2m}(p_\mu p^\mu-m^2)\right)d\tau.$$

In this case the potentially anomalous term to be added to the Hamiltonian when moving to the coherent state path integral (see before) is just a constant and may be neglected. Having only one constraint, this is a first class system and there is no repair to be made.

Analyze this system under the trivializing scheme. A subtlety arises for the Hamilton path integral in that it is, in general, necessary to perform the momentum before the position integrations (see the Appendix). To relax this condition, consider instead the simple harmonic system

$$H=(\lambda/2m)(p_\mu p^\mu+m^2\omega^2q_\mu q^\mu-m^2),$$

in the limit as $\omega\to 0$, where momentum and position are then further placed on an equal footing by performing the trivial counterscaling canonical transformation generated by

$$F=(1/\sqrt{m\omega})q_\mu P^\mu,$$

which yields

$$q_\mu\to q_\mu/\sqrt{m\omega},\quad p_\mu\to p_\mu\sqrt{m\omega},$$

and introduces an overall factor of $\sqrt{m\omega}$ into the position to position amplitude. The system under consideration is then characterized by

$$H=(\lambda/2m)(m\omega p_\mu p^\mu+m\omega q_\mu q^\mu-m^2),$$

and the generalized Hamilton–Jacobi equations determining the generating function then read as

$$\alpha+\frac{\partial F}{\partial\tau}=0,\quad\lambda(\tau)\phi\left(q_\mu\frac{\partial F}{\partial q^\mu},\tau\right)=\alpha,$$

which become

$$F(q_\mu\alpha,\tau)=W(q_\mu\alpha)-\alpha\int d\tau,$$

where W is determined from

$$\frac{\lambda}{2m}\left(m\omega\frac{\partial W}{\partial q_\mu}\frac{\partial W}{\partial q^\mu}+m\omega q_\mu q^\mu-m^2\right)=\alpha,$$

Solving this [Gradshteyn and Ryzhik,[22] p. 86, from 2.271(3)], yields

$$F=\frac{\sqrt{q_\mu q^\mu}}{2}\sqrt{\left(\frac{2\alpha}{\omega\lambda}+\frac{m}{\omega}-q_\mu q^\mu\right)}+\frac{1}{2}\left(\frac{2\alpha}{\omega\lambda}+\frac{m}{\omega}\right)\arcsin\left[\left(\frac{2\alpha}{\omega\lambda}+\frac{m}{\omega}\right)^{-1/2}\sqrt{q_\mu q^\mu}\right]-\alpha\int d\tau.$$

This generates

$$p_\mu=\frac{\partial F}{\partial q^\mu}=\frac{q_\mu}{\sqrt{q_\mu q^\mu}}\sqrt{\left(\frac{2\alpha}{\omega\lambda}+\frac{m}{\omega}-q_\mu q^\mu\right)},$$
$$Q_\mu=\frac{\partial F}{\partial P^\mu}=\frac{\partial F}{\partial\alpha}\frac{\partial\alpha}{\partial P^\mu}=\left(\frac{1}{\omega\lambda}\arcsin\left[\left(\frac{2\alpha}{\omega\lambda}+\frac{m}{\omega}\right)^{-1/2}\sqrt{q_\mu q^\mu}\right]-\int d\tau\right)\frac{\partial\alpha}{\alpha P^\mu}.$$

Unravel to expose q_μ and p_μ:

$$q_\mu = n_\mu \sqrt{\frac{2\alpha}{\omega\lambda} + \frac{m}{\omega}} \sin\left[\omega\lambda\left(Q_\mu \frac{\partial P^\mu}{\partial \alpha} + \int d\tau\right)\right],$$

$$p_\mu = n_\mu \sqrt{\frac{2\alpha}{\omega\lambda} + \frac{m}{\omega}} \cos\left[\omega\lambda\left(Q_\mu \frac{\partial P^\mu}{\partial \alpha} + \int d\tau\right)\right],$$

$$F = \frac{1}{2}\left(\frac{2\alpha}{\omega\lambda} + \frac{m}{\omega}\right)\sin\left[\omega\lambda\left(Q_\mu \frac{dP^\mu}{d\alpha} + \int d\tau\right)\right]\cos\left[\omega\lambda\left(Q_\mu \frac{dP^\mu}{d\alpha} + \int d\tau\right)\right] + \left(\alpha + \frac{m\lambda}{2}\right)Q_\mu \frac{dP^\mu}{d\alpha} + \frac{m\lambda}{2}\int d\tau$$

$$(\Gamma = F - Q_\mu P^\mu),$$

where n_μ is the unit vector in the direction of the corresponding x_μ (since $q_a^\mu = x_a^\mu, q_b^\mu = x_b^\mu$). The phase space integral is thus reduced to one canonical pair. Dropping indices therefore and exploiting the arbitrariness in α to let $\alpha \to \lambda$ ($\omega\alpha - m/2$) leads to

$$q = \sqrt{2\alpha}\sin\left[Q\frac{dP}{d\alpha} + \omega\lambda\int d\tau\right], \quad p = \sqrt{2\alpha}\cos\left[Q\frac{dP}{d\alpha} + \omega\lambda\int d\tau\right],$$

$$F = \alpha\sin\left[Q\frac{dP}{d\alpha} + \omega\lambda\int d\tau\right]\cos\left[Q\frac{dP}{d\alpha} + \omega\lambda\int d\tau\right] + \alpha Q\frac{dP}{d\alpha} + \frac{m\lambda}{2}\int d\tau.$$

To determine the position to position amplitude, recall

$$\langle x_b,\tau_b | x_a,\tau_a\rangle \propto \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} d\lambda \int_{-\infty}^{\infty} dQ \int_{-\infty}^{\infty} dP \frac{1}{\sqrt{\pi\hbar}} \exp\left[-\frac{1}{2\hbar}((x_b - (q_b + ip_b))^2 + p_b^2)\right]\exp\left[\frac{i}{\hbar}(\Gamma_b - \Gamma_a)\right]$$

$$\times \exp\left[-\frac{1}{2\hbar}((x_a - (q_a - ip_a))^2 + p_a^2)\right].$$

Substituting and transforming $P, Q \to \alpha, \theta$, where $\theta \equiv Q\,dP/d\alpha$ (unit Jacobian) proves independence (in general) from the functional form of $\alpha(P)$, α becoming a dummy variable. Further replacing $\alpha = r^2$ leads to

$$\frac{\sqrt{m\omega}}{(\pi\hbar)^{3/2}}\int_{-\infty}^{\infty} d\lambda \int_0^{\infty} r\,dr \int_{-\pi}^{\pi} d\theta \exp[C_2(\theta)r^2 + C_1(\theta)r + C_0(\theta)],$$

where

$$C_2 = -\frac{1}{\hbar}\left(1 - \frac{1}{2}\exp\left[-2i\left(\theta + \omega\lambda\int^{\tau_b} d\tau\right)\right] - \frac{1}{2}\exp\left[2i\left(\theta + \omega\lambda\int^{\tau_a} d\tau\right)\right]\right),$$

$$C_1 = i\frac{\sqrt{2m\omega}}{\hbar}\left(x_b \exp\left[-i\left(\theta + \omega\lambda\int^{\tau_b} d\tau\right)\right] - x_a \exp\left[i\left(\theta + \omega\lambda\int^{\tau_a} d\tau\right)\right]\right),$$

$$C_0 = -m\omega\frac{(x_a^2 + x_b^2)}{2\hbar} + i\frac{m\lambda}{2\hbar}\int_{\tau_a}^{\tau_b} d\tau,$$

having included the $\sqrt{m\omega}$ factors stemming from the first trivial counterscaling canonical transformation.

Translating

$$\theta \to \theta - \frac{\omega\lambda}{2}\int^{\tau_b} d\tau - \frac{\omega\lambda}{2}\int^{\tau_a} d\tau$$

leads to the simplification

$$C_2 = -(1/\hbar)(1 - e^{i\omega\lambda T}\cos(2\theta)), \quad C_1 = i(\sqrt{2m\omega/\hbar})(x_b e^{-i\theta} - x_a e^{i\theta})e^{-i\omega\lambda T/2}, \quad C_0 = -m\omega\frac{(x_a^2 + x_b^2)}{2\hbar} + i\,m\lambda T/2\hbar,$$

where

$$T \equiv \tau_b - \tau_a,$$

Let $\lambda \to \lambda T$ (remembering that $\langle x_b,\tau_b | x_a,\tau_a\rangle = 0$, for $T < 0$), the T dependence then being lost as a result. The parametrization labels are then also dropped [$\langle x_b,\tau_b | x_a,\tau_a\rangle \to K(x_b,x_a)$]. Further transform these "cylindrical polar" coordinates to Cartesian to yield

$$K(x_b,x_a) \propto \frac{\sqrt{m\omega}}{(\pi\hbar)^{3/2}}\int_0^{\infty} d\lambda \exp\left[i\frac{m\lambda}{2\hbar}\right]\exp\left[-\frac{m\omega}{2\hbar}(x_a^2 + x_b^2)\right]$$

$$\times \int_{-\infty}^{\infty} du \exp\left[-\frac{1}{\hbar}((1 - e^{-i\omega\lambda})u^2 + i\sqrt{2m\omega}(x_a - x_b)e^{-i\omega\lambda/2}u)\right]$$

$$\times \int_{-\infty}^{\infty} dv \exp\left[-\frac{1}{\hbar}((1 + e^{-i\omega\lambda})v^2 + \sqrt{2m\omega}(x_a + x_b)e^{-i\omega\lambda/2}v)\right],$$

but

$$\int_{-\infty}^{\infty} \exp[-\alpha x^2+\beta x]dx=\sqrt{\frac{\pi}{\alpha}}\exp\left[\frac{\beta^2}{4\alpha}\right].$$

Hence

$$K(x_b,x_a)\propto\int_0^{\infty}d\lambda\sqrt{\frac{m\omega}{2i\pi\hbar\sin(\omega\lambda)}}\exp\left[i\frac{m}{2\hbar}\left(\lambda+\omega\frac{(x_a^2+x_b^2)\cos(\omega\lambda)-2x_ax_b}{\sin(\omega\lambda)}\right)\right]$$

(the five-dimensional s.h.o. kernel integrated over the "fifth time" λ).

Taking the limit $\omega\to 0$ leads to

$$K(x_b,x_a)\propto\int_0^{\infty}d\lambda\sqrt{\frac{m}{2i\pi\hbar\lambda}}\exp\left[i\frac{m}{2\hbar}\left(\lambda+\frac{(x_a-x_b)^2}{\lambda}\right)\right].$$

Take the four-dimensional Fourier transform

$$\tilde{K}\propto\int_0^{\infty}d\lambda\exp\left[-i\frac{p_\mu p^\mu-m^2}{2m\hbar}\lambda\right]\propto\frac{i}{p_\mu p^\mu-m^2-i\varepsilon},$$

which is the usual Klein–Gordon propagator.

This then verifies that this technique is able to identify and eliminate the symmetry overcounting, while maintaining the symmetry throughout.

ACKNOWLEDGMENTS

Thanks to the theoretical physics group at Imperial College who gently kicked some shape into this work when prematurely presented, as well as the "locals," M. Rakowski and G. Thompson, who each influenced this work in their individual ways.

To the people of the centre itself: Professor A. Salam for his help and hospitality, E. Sezgin for his kindness, and the other whose word processors I held hostage.

This work was supported by the Royal Society (London), the International Centre for Theoretical Physics (Trieste), and the University of Waterloo, Canada.

APPENDIX

In general, the result of performing the phase space integrals depends on the sequence in which they are performed.[23,24] It is understood that if there is any ambiguity, the p integrals are to be performed first.

To understand this ambiguity, study the basic integral

$$I\equiv\int_{-\infty}^{\infty}dq\int_{-\infty}^{\infty}\frac{dp}{2\pi\hbar}\exp\left[\frac{i}{\hbar}(p\Delta q-H(p,q,t)\Delta t)\right]$$

$$=\int_{-\infty}^{\infty}dq\int_{-\infty}^{\infty}\frac{dp}{2\pi}(C_{00}(t)+C_{10}(t)p+C_{01}(t)q+\cdots)$$

$$\times\exp[i(\alpha(t)p^2+\beta(t)pq+\gamma(t)q^2)],$$

having expanded all but the Gaussian terms in the exponent.

The presence of terms of the form p^n, for $n>2$, leads to an unhealthy theory and corresponding divergences in the path integral.[18] Canonical transformations leading to such terms are then implicitly excluded.

The above integrals may be performed using

$$\int_{-\infty}^{\infty}du\exp[-au^2+bu]=\sqrt{\frac{\pi}{a}}\exp\left[\frac{b^2}{4a}\right],\qquad \mathrm{Re}(a)>0,$$

and offspring stemming from taking derivatives with respect to b.

Consider, then, performing the p integrals first. Convergence requires that $\mathrm{Im}(\alpha)>0$ and leads to

$$I=\int_{-\infty}^{\infty}\frac{dq}{2\pi}\left(C_{00}\sqrt{\frac{\pi}{\alpha}}+C_{10}\sqrt{\frac{\pi}{\alpha}}\frac{\beta}{2\alpha}q+C_{01}q+\cdots\right)\times\exp\left[i\left(\gamma-\frac{\beta^2}{4\alpha}\right)q^2\right].$$

The further requirement that this second integral be convergent leads to

$$\mathrm{Im}\,(\gamma-\beta^2/4\alpha)>0$$

and so the overall convergence requirements

$$\gamma_i-[2\alpha_r\beta_r\beta_i-\alpha_i(\beta_r^2-\beta_i^2)]/4|\alpha|^2>0$$

and

$$\alpha_i>0,$$

where $\delta_r\equiv\mathrm{Re}(\delta)$, $\delta_i\equiv\mathrm{Im}(\delta)$.

If the sequence of performing the integrations is reversed (q, then p), the criteria read

$$\alpha_i-[2\gamma_r\beta_r\beta_i-\gamma_i(\beta_r^2-\beta_i^2]/4|\gamma|^2>0$$

and

$$\gamma_i>0.$$

It then follows that the same solution will be obtained regardless of the sequence, if it can be arranged that $\alpha=\gamma$, since otherwise it becomes possible that the first set of convergence criteria are respected while the second are not, and then the two results would, in general, disagree.

This equality can always be achieved via the trivial canonical counterscaling generated by

$$F = (\alpha/\gamma)^{1/4} qP,$$

which yields

$$q \to (\alpha/\gamma)^{1/4} q, \quad p \to (\gamma/\alpha)^{1/4} p.$$

This transformation introduces an overall factor of $(\gamma/\alpha)^{1/4}$ into the position to position amplitude; while generating the new Hamiltonian

$$H((\alpha/\gamma)^{1/4} q, (\gamma/\alpha)^{1/4} p, t),$$

which then permits one the freedom to perform the integrals in any order, i.e., employ variable changes.

This ordering dependence was encountered for the example performed in the main body of the text, and avoided through the above counterscaling.

[1]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd ed.

[2]L. Landau and E. Lifshitz, *Classical Mechanics* (Pergamon, New York, 1976), 3rd ed.

[3]P. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964), and references therein.

[4]P. Chernoff, "Mathematical obstructions to quantization," Hadronic J. **4**, 879 (1981).

[5]J. Klauder, "Path Integrals," Proceedings of the XIX Internationale Universitätswochen für Kernphysik, Schladming, Austria, Acta Phys. Austriaca Supp. **XXII**, 3 (1980).

[6]F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions* (Reidel, Dordrecht, 1982).

[7]I. Mayes and J. Dowker, "Canonical functional integrals in general coordinates," Proc. R. Soc. London Ser. A **327**, 131 (1972).

[8]I. Mayes and J. Dowker, "Hamiltonian orderings and functional integrals," J. Math. Phys. **14**, 434 (1973).

[9]L. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

[10]E. Fradkin and G. Vilkovisky, "Quantization of relativistic systems with constraints," Phys. Lett. B **55**, 224 (1975). See also *Quantum Field Theory and Quantum Statistics, Essays in Honour of the Sixtieth Birthday of E. S. Fradkin*, edited by I. Batalin, C. Isham, and G. Vikovisky (Hilger, London, 1987).

[11]I. Batalin and E. Fradkin, "Operational quantization of dynamical systems subject to second class contraints," Nucl. Phys. B **279**, 514 (1987), and references therein.

[12]L. Faddeev and V. Popov, "Feynman diagrams for the Yang–Mills field," Phys. Lett. B **25**, 29 (1967).

[13]R. Omnes, "A new geometric approach to the relativistic string," Nucl. Phys. B **149**, 269 (1979).

[14]R. Feynman, "Space-time approach to quantum mechanics," Rev. Mod. Phys. **20**, 267 (1948) [reprinted in *Quantum Electrodynamics*, edited by J. Schwinger (Dover, New York, 1958)].

[15]R. Feynman and A. Hibb, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[16]D. Kandaekar and S. Lawande, "Feynman path integrals: Some exact results and applications," Phys. Rep. **137**, 115 (1986).

[17]J. Klauder, "Wiener measure regularization for quantum mechanical path integrals," Schladming Winter School, 1987.

[18]A. Shiekh, "Canonical transformations in quantum mechanics," J. Math. Phys. **29**, 913 (1988).

[19]R. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control* (American Elsevier, New York, 1968).

[20]C. Teitelboim, "Constrained Hamiltonian systems," lectures given at the University of Texas at Austin, 1981.

[21]P. Mannheim, "Klein–Gordon propagator via first quantization," Phys. Lett. B **166**, 191 (1986).

[22]I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980), 4th ed.

[23]C. DeWitt-Morette, A. Maheshwari, and B. Nelson, "Path integration in non-relativistic quantum mechanics," Phys. Rep. **50**, 255 (1979).

[24]S. Coleman, *Aspects of Symmetry* (Cambridge U.P., Cambridge, 1985); "Secret symmetry," originally published in *Laws of Hadronic Matter* (Academic, New York, 1975).

Reduction, quantization, and nonunimodular groups

G. M. Tuynman
CPT-CNRS, Luminy, Case 907, F-13288 Marseille Cedex 9, France

(Received 17 May 1989; accepted for publication 16 August 1989)

It is shown that even in relatively nice cases the naive approach to the quantization of constraints is not correct in general [i.e., the procedure that if $f=0$ is a classical constraint and $\tau(f)$ is the associated quantum operator, then the quantum constraint is $\tau(f)=0$]. An explicit procedure for the quantization of constraints in the case of a configuration space with a symmetry group is provided and proven, where the reduced configuration space is the orbit space. It is not thought that the group acts freely, merely that all isotropy subgroups are conjugated to each other.

I. INTRODUCTION

It is well known that, in general, quantization of constraints poses insurmountable difficulties. However, one does not expect serious problems if at the classical level we have the following situation: (i) a phase space (symplectic manifold) (M,ω) with a symmetry group G (acting as symplectomorphisms) that admits an Ad-equivariant momentum map $J: M\to\mathfrak{g}^*$ ($\mathfrak{g}$ the Lie algebra of G), and (ii) a reduced space (B,ω_{red}) which is obtained by the (classical) constraints $\langle J(\cdot),\xi\rangle = 0\ \forall\xi\in\mathfrak{g}$. In such a situation the reduced phase space (B,ω_{red}) is obtained as the quotient $B=J^{-1}(0)/G$ (see Ref. 1). In this context, each observable $H\in C^\infty(M)$ that is invariant under the group action induces in a natural way an observable $H_{\text{red}}\in C^\infty(B)$ by restricting H to $J^{-1}(0)\subset M$ and observing that this restricted function descends to $J^{-1}(0)/G=B$.

Going to the quantum level we assume to have a Hilbert space $\mathcal{H}_M$ for the classical system (M,ω) and a quantization τ for the constraint functions $\langle J(\cdot),\xi\rangle$ and the (invariant) observable H as skew self-adjoint operators on $\mathcal{H}_M$. One would expect self-adjoint operators, but we have absorbed a factor i in the definition of the operators in order to avoid factors i later on. Now we note that since J is Ad-equivariant, the constraint functions $\langle J(\cdot),\xi\rangle$ are all first-class constraints, so nothing is more natural than to suppose that the Hilbert space $\mathcal{H}_B$ for the classical system (B,ω_{red}) can be identified with

$$\mathcal{H}_{\text{red}}=\{\psi\in\mathcal{H}_M \mid \tau(\langle J(\cdot),\xi\rangle)\psi=0 \quad \forall\xi\in\mathfrak{g}\} \tag{1}$$

and that the quantization $\tau(H_{\text{red}})$ of H_{red} as a skew self-adjoint operator on $\mathcal{H}_B$ coincides under this identification with $\tau(H)$ "restricted to" $\mathcal{H}_{\text{red}}$. Of course, one has to assume that τ maps brackets in $\mathfrak{g}$ to commutators of operators (i.e., $\tau(\langle J(\cdot),[\xi,\eta]\rangle)=[\tau(\langle J(\cdot),\xi\rangle),\tau(\langle J(\cdot),\eta\rangle)]$) in order to get a consistent description of $\mathcal{H}_{\text{red}}$, and one has to assume that $\tau(H)$ commutes with all $\tau(\langle J(\cdot),\xi\rangle)$ in order to get a well defined restriction of $\tau(H)$ to $\mathcal{H}_{\text{red}}$, but those conditions are usually satisfied.

However, in their paper on BRS quantization,[2] Kostant and Sternberg gave implicitly a slightly different definition of $\mathcal{H}_{\text{red}}$ (in the case of a free action of G):

$$\mathcal{H}_{\text{red}}=\{\psi\in\mathcal{H}_M \mid \forall\xi\in\mathfrak{g}:\tau(\langle J(\cdot),\xi\rangle)\psi=-\tfrac{1}{2}\operatorname{tr}(\operatorname{ad}(\xi))\psi\} \tag{2}$$

(see Ref. 3 for an explicit derivation of this formula). Their derivation uses super Poisson algebras, Clifford algebras, and cohomological arguments, but no reference to any specific form of quantization. Now their formula might seem strange, especially if one realizes that skew self-adjoint operators usually do not have real eigenvalues. Nevertheless, as is shown in Ref. 3 and this paper, their formula for $\mathcal{H}_{\text{red}}$ is correct and the naive formula is not.

The outline of this paper is as follows. At the classical level we consider the original phase space M to be the cotangent bundle of a configuration space Q: $M=T^*Q$, and we suppose that there are symmetries at the level of the configuration space, i.e., that a group G acts on Q (but not necessarily freely). The typical example of such a situation is a gauge theory in which Q is the configuration space of all connections on a principal fiber bundle and in which G is the group of all gauge transformations. However, this particular example is not covered by our finite dimensional manifold approach. The reduced space then is the cotangent bundle $B=T^*(Q/G)$, where Q/G is the reduced configuration space. Our quantization of the cotangent bundles T^*Q and $T^*(Q/G)$ consists of taking the Hilbert space of square integrable functions on their configuration spaces Q and Q/G (with respect to a certain measure). Such a quantization of cotangent bundles is natural and can be "derived" by various quantizations, e.g., geometric quantization,[4] or Segal's quantization.[5] Since our constraint operators have in general a continuous spectrum, neither formula (1) nor formula (2) will define a nonzero space $\mathcal{H}_{\text{red}}$. Therefore, in order to obtain meaningful results, we will interpret the Hilbert spaces as the spaces of all C^∞ functions on the configuration spaces, instead of only the square-integrable ones (in the main text we will use a correct language). This problem and the related fact that one has to change the measure when going from the original space $\mathcal{H}_M$ to the subspace $\mathcal{H}_{\text{red}}$ has been noted by several authors (e.g., Refs. 6 and 7). The appropriate setting would probably be a rigged Hilbert space approach, but since we are interested in geometrical properties, we will ignore these questions. It is in this context that we discuss the identification of $\mathcal{H}_B$ with a subspace $\mathcal{H}_{\text{red}}\subset\mathcal{H}_M$ and the relation between $\tau(H_{\text{red}})$ on $\mathcal{H}_B$ and $\tau(H)$ restricted to $\mathcal{H}_{\text{red}}$.

We show that in general there is no intrinsic (i.e., in terms of the group action) description of $\mathcal{H}_B$ as subspace $\mathcal{H}_{\text{red}}\subset\mathcal{H}_M$, but that there always is a (nonunique) identification for which $\tau(H_{\text{red}})$ on $\mathcal{H}_B$ and $\tau(H)$ restricted to

 0022-2488/90/010083-08$03.00

$\mathscr{H}_{\text{red}}$ coincide. In special cases though, there is an intrinsic definition of $\mathscr{H}_B$ as subspace $\mathscr{H}_{\text{red}} \subset \mathscr{H}_M$. First if the action is free, then formula (2) of Kostant and Sternberg applies. A naive explanation in the trivial case $Q = G \times Q/G$ is as follows. Let μ be a measure on Q/G, then the pull back to Q is not a measure. If however s_0 is a measure on G then $\mu \cdot s_0$ is a product measure on Q. The correction term stems from the behavior of s_0 under the group action. As a second special case, suppose G admits a bi-invariant metric, then formula (1) applies. Note that if G admits a bi-invariant metric, then G admits a bi-invariant measure (i.e., G is unimodular) which implies that $\text{tr}(\text{ad}(\xi)) \equiv 0$, so the two cases do not contradict each other.

Since the quantization we use can be derived by geometric quantization, our results generalize a result of Gotay,[8] either in the direction of nonunimodular groups or in the direction of nonfree actions. However, our approach is quite different since we impose the quantization of the reduced space instead of deriving it from the quantization of the original space. In the last section we give two examples: the first one to show that for nonunimodular groups G the term $\frac{1}{2}\text{tr}(\text{ad}(\xi))$ is essential for $\tau(H_{\text{red}})$ and $\tau(H)$ to coincide and the second one to prove that in general no intrinsic description of $\mathscr{H}_B$ as subspace $\mathscr{H}_{\text{red}} \subset \mathscr{H}_M$ can be given.

For the moment it remains a mystery how the algebraic approach of Kostant and Sternberg without any reference to a specific quantization relates to our geometric description which makes in an essential way use of the specific structure of $\mathscr{H}_M$ and $\mathscr{H}_B$. However, even if we are not in a situation covered by this paper, the term $\frac{1}{2}\text{tr}(\text{ad}(\xi))$ proves to be essential.[3]

II. THE CLASSICAL SETTING

Let the original configuration space Q be a manifold and let the connected Lie group G act on Q smoothly and properly. We do not suppose that G acts freely, but we do suppose that all isotropy subgroups $G_q := \{g \in G \mid g(q) = q\}$ are conjugated, say to the subgroup $H \subset G$. It then follows (Ref. 9; exer. 4.1M) that the reduced configuration space $R = Q/G$ (the orbit space) is a manifold and that $p: Q \to R$ is a locally trivial fiber bundle with typical fiber G/H and structure group N/H, where N is the normalizer of H in G, i.e., $N = \{g \in G \mid gHg^{-1} = H\}$.

Since we will need a more precise description of this bundle, we introduce some notations. For $g \in G$ we denote the action of g on $q \in Q$ by $L_g(q)$ to stress that we consider it to be an action on the left. By abuse of notation L_g will also denote the canonical left action of G on G/H defined by $L_g[k] = [gk]$, where brackets denote taking cosets. The group N/H acts canonically on the right on G/H as follows: if $n \in N$, $[n]$ its projection in N/H, then $R_{[n]}[k] = [kn]$. Obviously the left action of G on G/H commutes with the right action of N/H.

From our assumptions it follows that each point $u \in R$ admits a neighborhood U such that $p^{-1}(U) \cong U \times G/H$ where the G action on $p^{-1}(U)$ is just the G action on the second factor G/H. Moreover, given two such trivializing neighborhoods U and U', they are related by transition functions

$$U \times G/H \ni (u,[k]) \mapsto (u, R_{[n](u)}[k]) \in U' \times G/H, \tag{3}$$

where $[n]: U \cap U' \to N/H$ gives the gauge transformations related to these two trivializations.

We now go over to the phase spaces by taking cotangent bundles with their canonical symplectic forms. Here, T^*Q is the original phase space with symplectic form $\omega_Q = d\theta_Q$, θ_Q being the canonical one-form on T^*Q; in the same way $(T^*R, \omega_R = d\theta_R)$ is the phase space of the reduced system. The action of G on Q can be lifted canonically to a symplectic action of G on T^*Q (which even leaves θ_Q invariant) and this symplectic action has a canonically defined Ad-equivariant momentum map $J: T^*Q \to \mathfrak{g}^*$ ($\mathfrak{g}$ the Lie algebra of G) defined by the following prescription. For $\xi \in \mathfrak{g}$ denote by ξ_Q the associated fundamental vector field on Q whose flow is $L_{\exp(-\xi t)}$; for $\alpha \in T_q^*Q$, the value $J(\alpha)$ is given by

$$\langle J(\alpha), \xi \rangle := \alpha(\xi_{Q|q}). \tag{4}$$

By abuse of notation $J(\xi)$ will denote the function on T^*Q defined by $J(\xi)(\alpha) = \langle J(\alpha), \xi \rangle$. If in local coordinates ξ_Q is given by $\Sigma_i \xi^i(q) \partial/(\partial q^i)$ then $J(\xi)$ is given by $\Sigma_i \xi^i(q) p_i$, where the p_i are the associated momentum coordinates.

With our hypotheses it is easy to show that $0 \in \mathfrak{g}^*$ is a weakly regular value for J, that the constraint set $J^{-1}(0) = \{\alpha \in T^*Q \mid \forall \xi \in \mathfrak{g}: J(\xi)(\alpha) = 0\}$ is a submanifold of T^*Q and that the Marsden–Weinstein reduced symplectic manifold $J^{-1}(0)/G$ is symplectomorphic to (T^*R, ω_R). We refer the interested reader to Ref. 10 for the general case of reduction for nonfree actions.

III. INTERMEZZO ON d-DENSITIES

Let Q be a manifold and $F(Q)$ its frame bundle, a principal $\text{GL}(n,\mathbb{R})$ bundle over Q with $n = \dim Q$. The bundle $\Delta^d Q$ is the (real or complex) line bundle over Q associated to $F(Q)$ by the representation $\rho: \text{GL}(n,\mathbb{R}) \to \mathbb{R}^+ \subset \text{Aut}(\mathbb{R})$, $(a_{ij}) \mapsto |\det(a_{ij})|^d$, and a d-density ψ is a global section of $\Delta^d Q$. In other words, a d-density is a map that assigns a (real or complex) number to each $(n+1)$-tuple $(q;(e_1,\ldots,e_n))$ where $q \in Q$ and $(e_1,\ldots,e_n) \equiv (e)$ a frame at (i.e., a basis of) T_qQ. Moreover, if $(f_1,\ldots,f_n)$ is a different frame at T_qQ with $f_i = \Sigma_j a_{ij} e_j$ then

$$\psi(q;(f)) \equiv \psi(q;(ae)) = |\det(a_{ij})|^d \cdot \psi(q;(e)). \tag{5}$$

Since any two frames at T_qQ are related by a matrix a, $\psi(q;\cdot)$ is completely determined by its value on a single frame.

With an appeal to the transformation law of multiple integrals, a one-density is just a measure on Q. Since the pointwise product of a d-density and a d'-density is a $(d+d')$-density, it follows that the product of two $\frac{1}{2}$-densities is a measure on Q which can be integrated over Q, giving rise to the Hilbert space $\mathscr{H}^{1/2}(Q)$ of square-integrable $\frac{1}{2}$-densities on Q. An elementary partition of unity argument shows that $\Delta^d Q$ is a trivial line bundle, hence the choice of a nowhere vanishing d-density ψ_0^d identifies the set $\Omega^d(Q)$ of all smooth d-densities with the set $C^\infty(Q)$ of all smooth functions by $\psi = f\psi_0^d$. In the particular case of $\frac{1}{2}$-densities, the choice of a trivializing section $\psi_0^{1/2}$ identifies $\mathscr{H}^{1/2}(Q)$

with $L^2(q,|\psi_0^{1/2}|^2)$, the space of square-integrable functions with respect to the measure $|\psi_0^{1/2}|^2$.

Now let ϕ be a diffeomorphism of Q and $\psi\in\Omega^d(Q)$, then there exists a natural notion of pull-back $\phi^*\psi$ defined by

$$(\phi^*\psi)(q;(e_1,\ldots,e_n)) := \psi(\phi(q);(\phi_* e_1,\ldots,\phi_* e_n)). \tag{6}$$

It follows that, given a vector field X on Q, there is a natural definition of a Lie derivative $\mathcal{L}_X\psi$ given by

$$\begin{aligned}(\mathcal{L}_X\psi)(q;(e_1,\ldots,e_n)) &:= \frac{d}{dt}\Big|_{t=0}(\phi_t^*\psi)(q;(e_1,\ldots,e_n))\\ &= \frac{d}{dt}\Big|_{t=0}\psi(\phi_t(q);(\phi_{t*}e_1,\ldots,\phi_{t*}e_n)),\end{aligned} \tag{7}$$

where ϕ_t is the flow of the vector field X. It is elementary to verify that $[\mathcal{L}_X,\mathcal{L}_Y]\equiv\mathcal{L}_X\circ\mathcal{L}_Y-\mathcal{L}_Y\circ\mathcal{L}_X=\mathcal{L}_{[X,Y]}$.

IV. QUANTIZATION OF COTANGENT BUNDLES

If the phase space M of a physical system is a cotangent bundle T^*Q, it is quite natural to assume that the Hilbert space $\mathcal{H}_M$ that describes this system in quantum theory is a L^2-space of functions on Q with respect to a certain measure μ on Q. When no natural choices for μ are available, it is more convenient to interpret $\mathcal{H}_M$ as the space $\mathcal{H}^{1/2}(Q)$ of (square-integrable) $\frac{1}{2}$-densities on Q. The main problem of quantization then becomes the problem of finding quantum operators corresponding to classical observables.

For obvious reasons we will restrict our attention to observables H: $T^*Q\to\mathbf{R}$ which are at the most linear in the momentum variables. Let X be a vector field on Q, then there is a naturally defined function H_X: $T^*Q\to\mathbf{R}$ which is linear in the momenta: $H_X(\alpha)=\alpha(X|_q)$ for $\alpha\in T_q^*Q$ (recall that the momentum map J is constructed in this way!). Vice versa, every function on T^*Q that is linear in the momenta is of this form. It follows that we restrict our attention to observables of the form

$$H = H_Q\circ\pi_Q + H_X \tag{8}$$

with H_Q: $Q\to\mathbf{R}$, π_Q: $T^*Q\to Q$ the canonical projection and X a vector field on Q. In local coordinates q on Q with associated momenta this becomes $H(q,p)=H_Q(q)+\Sigma_i\xi^i(q)p_i$ with $X=\xi^i(q)(\partial/\partial q^i)$.

For these observables there is a natural quantization as skew self-adjoint operators on $\mathcal{H}^{1/2}(Q)$:

$$\tau(H_Q\circ\pi_Q+H_X)\psi = iH_Q\cdot\psi+\mathcal{L}_X\psi. \tag{9}$$

This quantization can be derived rigorously by geometric quantization (for $\frac{1}{2}$-densities without additional assumptions; for $\frac{1}{2}$-forms this result is guaranteed if Q is orientable), but can also be found in Segal's approach to quantization. In the case $Q=\mathbf{R}^n$ this quantization procedure is equivalent to the usual prescription of symmetrization:

$$\begin{aligned}&\tau\Big(\sum_i\xi^i(q)p_i\Big)\psi(q)\\ &\quad=\frac{1}{2}\sum_i\Big(\xi^i(q)\frac{\partial\psi}{\partial q^i}+\frac{\partial}{\partial q^i}(\xi^i(q)\psi(q))\Big)\\ &\quad=\sum_i\xi^i(q)\frac{\partial\psi}{\partial q^i}+\frac{1}{2}\Big(\sum_i\frac{\partial\xi^i(q)}{\partial q^i}\Big)\psi(q).\end{aligned} \tag{10}$$

It follows directly from the formula $[\mathcal{L}_X,\mathcal{L}_Y]=\mathcal{L}_{[X,Y]}$ for d-densities that this quantization maps Poisson brackets on commutators, i.e., if H_1 and H_2 are both of the form (8), then $\tau(\{H_1,H_2\}_{\text{P.B.}})=[\tau(H_1),\tau(H_2)]$.

V. QUANTIZATION AND REDUCTION I

In the classical setup we had two phase spaces: the original one T^*Q and the reduced one T^*R that is determined by the constraint functions $J(\xi)$, $\xi\in\mathfrak{g}$. As we have seen in the previous section, the associated Hilbert spaces are derived from $\Omega^{1/2}(Q)$ and $\Omega^{1/2}(R)$. In this section we will try to identify $\Omega^{1/2}(R)$ as a subspace of $\Omega^{1/2}(Q)$ determined by the quantized constraints $\tau(J(\xi))$.

Contrary to all didactical rules we will directly proceed with the correct approach instead of presenting first the natural approach and then noticing that that does not work very well. Our basic object will be $\Omega^{1/2}(R;G/N)$ of $\frac{1}{2}$-densities on R which depend smoothly on a parameter $z\in G/N$. More precisely $\psi\in\Omega^{1/2}(R;G/N)$ is a smooth function of a point $(u,z)\in R\times G/N$ and a frame $(e_1,\ldots,e_n)$ at T_uR with the property that if another frame (f_i) is related to the frame (e_i) by the matrix a_{ij}, i.e., $f_i=\Sigma_j a_{ij}e_j$, then

$$\psi(u,z;(f)) = |\det a_{ij}|^{1/2}\psi(u,z;(e)). \tag{11}$$

As with d-densities, a nowhere vanishing element $\psi_0\in\Omega^{1/2}(R;G/N)$ identifies $\Omega^{1/2}(R;G/N)$ with $C^\infty(R\times G/N)$.

The map $gHg^{-1}\mapsto g \bmod N$ is an isomorphism between the manifold G/N and the set $I(H)=\{gHg^{-1}|g\in G\}$ of all subgroups of G conjugated to H. Moreover, there is a natural map $\mathcal{J}$: $Q\to I(H)\cong G/N$ given by $\mathcal{J}(q)=G_q$. With our assumptions, this is a smooth surjective map given in a local trivialization $U\times G/H$ of Q by $(u,[k])\mapsto k \bmod N$ (see Sec. 7 for a generic example of $\mathcal{J}$). We thus obtain a natural surjective submersion $p\times\mathcal{J}$: $Q\to R\times G/N$, explaining partially our interest in the product $R\times G/N$.

Let s_0 be a nowhere vanishing $\frac{1}{2}$-density on G/H which is invariant under the (right) action of N/H. A partition of unity argument applied to the principal N/H bundle $G/H\to G/N$ shows that such s_0 exists and that two such differ by a nowhere zero function on G/N. Using this s_0 we will define an injective map Φ: $\Omega^{1/2}(R;G/N)\to\Omega^{1/2}(Q)$. For $\psi\in\Omega^{1/2}(R;G/N)$ we define $\Phi(\psi)$ in a local trivialization $U\times G/H$ of Q by the formula:

$$\begin{aligned}&\Phi(\psi)((u,[k]);(f_u,f_{[k]}))\\ &\quad:=\psi(u,k \bmod N;(f_u))\cdot s_0([k];(f_{[k]})),\end{aligned} \tag{12}$$

where (f_u) is a frame at T_uR and $f_{[k]}$ a frame at $T_{[k]}G/H$, hence $(f_u f_{[k]})$ is a frame at $T_{(u,[k])}Q$. Since ψ and s_0 are $\frac{1}{2}$-densities this formula defines a $\frac{1}{2}$-density on $U\times G/H$. This definition of $\Phi(\psi)$ is independent of the local trivialization

because the gauge transformations between two such trivializations are right actions of N/H and s_0 is invariant under this action (i.e., $\forall [n]\in N/H: R_{[n]}{}^* s_0 = s_0$). That Φ is injective is a direct consequence of the fact that s_0 is nowhere vanishing.

In order to describe $\mathrm{im}(\Phi)\subset\Omega^{1/2}(Q)$ intrinsically, we need some preparations. Denote by $\mathfrak{h}$ the Lie algebra of H, then $\forall n\in N$ we have $\mathrm{Ad}(n)\mathfrak{h}\subset\mathfrak{h}$, hence there is an induced action $\overline{\mathrm{Ad}}(n)$ on $\mathfrak{g}/\mathfrak{h}$. If G_q is the isotropy subgroup at q with Lie algebra $\mathfrak{g}_q$ and normalizer N_q, then as above $\forall n\in N_q$ there is an induced action $\overline{\mathrm{Ad}}_q(n)$ on $\mathfrak{g}/\mathfrak{g}_q$. Since G_q is conjugated to H there exists a $g\in G$ with $G_q = gHg^{-1}$. It follows that $\mathfrak{g}_q = \mathrm{Ad}(g)\mathfrak{h}$, $N_q = gNg^{-1}$, and that $\mathrm{Ad}(g)$ induces an isomorphism between $\mathfrak{g}/\mathfrak{h}\to\mathfrak{g}/\mathfrak{g}_q$ which intertwines the actions $\overline{\mathrm{Ad}}(n)$ and $\overline{\mathrm{Ad}}_q(gng^{-1})$. We now recall that for $\xi\in\mathfrak{g}$, ξ_Q denotes the associated fundamental vector field on Q whose flow is $L_{\exp(-\xi t)}$. In the same way $\xi_{G/H}$ is the fundamental vector field on G/H whose flow is $L_{\exp(-\xi t)}$; since the G action commutes with the N/H action on G/H we have $R_{[n]^*}(\xi_{G/H}) = \xi_{G/H}$. On the other hand, for $g\in G$ we have the characteristic property of fundamental vector fields: $L_{g^*}(\xi_{G/H}) = (\mathrm{Ad}(g)\xi)_{G/H}$ and the same on Q: $L_{g^*}(\xi_Q) = (\mathrm{Ad}(g)\xi)_Q$. With these preparations we now can state our main theorem.

Theorem 1: $\hat{\psi}\in\Omega^{1/2}(Q)$ is an element of $\mathrm{im}(\Phi)$ if and only if

$$\forall q\in Q\ \forall n\in N_q:\ (L_n^*\hat{\psi})(q,\cdot) = |\det\ \overline{\mathrm{Ad}}_q(n)|^{1/2}\hat{\psi}(q,\cdot). \tag{13}$$

If $\hat{\psi}\in\mathrm{im}(\Phi)$ then $\Phi^{-1}(\hat{\psi})\in\Omega^{1/2}(R;G/N)$ is given by

$$\Phi^{-1}(\hat{\psi})(u,z;(f_u)) = \hat{\psi}(q;(\hat{f}_u,\xi_Q))/s_0([k];\xi_{G/H}), \tag{14}$$

where $p(q) = u$, $\mathscr{I}(q) = z = k \bmod N$, $\hat{f}_u$ any set of vectors in T_qQ with $p_*\hat{f}_u = f_u$ a frame at T_uR, and $\xi\in\mathfrak{g}$ such that $(\hat{f}_u,\xi_Q)$ becomes a frame at T_qQ.

Proof: We first show that $\Phi(\psi)$ has the above property, using a local trivialization $U\times G/H$ for Q. Now, $N_{(u,[k])} = kNk^{-1}$ so if $n\in N_{(u,[k])}$ then $n_0 = k^{-1}nk\in N$ and $L_n[k] = R_{[n_0]}[k]$. Since G acts transitively on G/H, the map $\mathfrak{g}\to T_{[k]}G/H$, $\xi\mapsto\xi_{G/H}|_{[k]}$ defines an isomorphism $\mathfrak{g}/\mathfrak{g}_{[k]}\cong T_{[k]}G/H$. Hence for each frame $(f_{[k]})$ at $T_{[k]}G/H$ there exist (nonunique) $\xi\in\mathfrak{g}$ such that $(f_{[k]}) = (\xi_{G/H}|_{[k]})$. Now $L_{n^*}(f_{[k]}) = L_{n^*}(\xi_{G/H}) = ((\mathrm{Ad}(n)\xi)_{G/H})$ so by definition of $\overline{\mathrm{Ad}}_q(n)$ the frame $((\mathrm{Ad}(n)\xi)_{G/H})$ is related to the frame $(\xi_{G/H})$ by the matrix $\overline{\mathrm{Ad}}_q(n)$. Finally,

$$\begin{aligned}
&(L_n^*\Phi(\psi))((u,[k]);(f_u,f_{[k]}))\\
&\quad = \Phi(\psi)((u,L_n[k]);(f_u,L_{n^*}(\xi_{G/H})))\\
&\quad = \psi(u;(f_u))\cdot s_0(R_{[n_0]}[k];(\mathrm{Ad}(n)\xi)_{G/H}))\\
&\quad = |\det\overline{\mathrm{Ad}}_{(u,[k])}(n)|^{1/2}\psi(u;(f_u))\\
&\qquad\cdot s_0(R_{[n_0]}[k];(\xi_{G/H}))\\
&\quad = |\det\overline{\mathrm{Ad}}_{(u,[k])}(n)|^{1/2}\Phi(\psi)((u,[k]);(f_u,f_{[k]})),
\end{aligned}$$

which shows that elements in $\mathrm{im}(\Phi)$ have the above property. In order to show the rest of the theorem, it suffices to show that the formula for $\Phi^{-1}(\hat{\psi})$ gives a well-defined element of $\Omega^{1/2}(R;G/N)$ if $\hat{\psi}$ has the above property, i.e., that the right-hand side is independent of the possible choices.

First of all notice that a different choice of (ξ) does not affect the right-hand side because both $\hat{\psi}$ and s_0 transform as $\frac{1}{2}$-densities. In the second place, a different choice for $\hat{f}_u$ is related to the original one by a 2×2 block matrix of the form $((\mathrm{id},0),(?,\mathrm{id}))$ whose determinant is one, so again the right-hand side does not change. If $k'\bmod N = k\bmod N$ then $k' = kn$ with $n\in N$. It follows that $[k'] = R_{[n]}[k]$ and since s_0 is N/H invariant (and because $R_{[n]^*}\xi_{G/H} = \xi_{G/H}$) this implies that the right-hand side is independent of the choice of k. Finally a different choice q' is related to q by $q' = L_nq$ with $n\in N_q$, which implies that $\hat{f}'_u = L_{n^*}\hat{f}_u$ is also a lift of the frame f_u (use $p\circ L_n = p$). We then compute

$$\begin{aligned}
\hat{\psi}(q';(\hat{f}'_u,\xi_Q)) &= \hat{\psi}(L_nq;(\hat{f}'_u,\xi_Q))\\
&= (L_n^*\hat{\psi})(q;(\hat{f}_u,L_{n^{-1}*}(\xi_Q)))\\
&= |\det\overline{\mathrm{Ad}}_q(n)|^{1/2}\hat{\psi}(q;(\hat{f}_u,(\mathrm{Ad}(n^{-1})\xi)_Q))\\
&= \hat{\psi}(q;(\hat{f}_u,\xi_Q)),
\end{aligned}$$

where the last equality follows because the frames $(\hat{f}_u,(\mathrm{Ad}(n^{-1})\xi)_Q)$ and $(\hat{f}_u,\xi_Q)$ are related by the 2×2 block matrix $((\mathrm{id},0),(0,\overline{\mathrm{Ad}}_q(n^{-1})))$. Q.E.D.

As we said before, $\Omega^{1/2}(R;G/N)$ is the space of $\frac{1}{2}$-densities on R which depend smoothly upon a parameter $z\in G/N$. There is, however, a natural injection i: $\Omega^{1/2}(R)\hookrightarrow\Omega^{1/2}(R;G/N)$ describing $\Omega^{1/2}(R)$ as those elements of $\Omega^{1/2}(R;G/N)$ that do not depend on this parameter. One might ask why we did not consider the composite injective map $\Phi\circ i$: $\Omega^{1/2}(R)\hookrightarrow\Omega^{1/2}(Q)$ directly. The reason is the following. We have seen that, although Φ itself depends upon the choice of s_0, $\mathrm{im}(\Phi)$ does not. On the contrary, $\mathrm{im}(\Phi\circ i)$ does depend upon the choice of s_0 and hence cannot, in general, be described intrinsically using only the group action. However, we will describe two different circumstances in which $\mathrm{im}(\Phi\circ i)$ can be described intrinsically: (i) $N = G$ and (ii) if s_0 can be chosen to be G-invariant.

If $N = G$ we can immediately draw two conclusions: (i) the parameter space G/N is a point so i is a bijection and (ii) H is a normal subgroup implying that there is an induced free action of the Lie group G/H on Q. Moreover, one can easily show that $\overline{\mathrm{Ad}}(n) = \mathrm{Ad}([n])$ for $n\in N = G$ and $\mathrm{Ad}([n])$ the adjoint representation of G/H on its Lie algebra $\mathfrak{g}/\mathfrak{h}$. In the following proposition we recover, in our restricted cotangent bundle setting, the result of Kostant and Sternberg as described in the introduction. It is a direct consequence of Theorem 1 and the fact that G is connected.

Proposition 2: If $G = N$, then $\mathrm{im}(\Phi\circ i)\equiv\mathrm{im}(\Phi)$ is characterized by the conditions:

$$\forall g\in G:\ L_g^*\hat{\psi} = |\det\mathrm{Ad}([g])|^{1/2}\hat{\psi} \tag{15}$$

or equivalently by the conditions:

$$\forall\xi\in\mathfrak{g}:\ \mathscr{L}_{\xi_Q}\hat{\psi} = -\tfrac{1}{2}\mathrm{tr}(\mathrm{ad}(\xi \bmod \mathfrak{h}))\cdot\hat{\psi}. \tag{16}$$

The second case we want to investigate is when, among all allowed choices of s_0 there is one which is (also) invariant under the left action of G on G/H; in other words, when there exists a preferred choice for s_0.

Lemma 3: There exists a G and N/H invariant nowhere vanishing $\frac{1}{2}$-density s_0 on G/H if and only if $\forall n\in N$: $|\det\overline{\mathrm{Ad}}(n)| = 1$. If it exists, s_0 is unique up to a nonzero (real or complex) factor.

Proof: Let s_0 be such a $\frac{1}{2}$-density. Reasoning as in the proof of the theorem, we obtain $(L_n^* s_0)([e],\cdot) = |\det \overline{\mathrm{Ad}}_{[e]}(n)|^{1/2} s_0([e],\cdot)$. Since $N = N_{[e]}$ and hence $\overline{\mathrm{Ad}}_{[e]}(n) = \overline{\mathrm{Ad}}(n)$ it now follows from the invariance of s_0 under the left action that $|\det \overline{\mathrm{Ad}}(n)| = 1$.

On the other hand, suppose $\forall n \in N$: $|\det \overline{\mathrm{Ad}}(n)| = 1$, which implies that $\forall n \in N_{[k]}$: $|\det \overline{\mathrm{Ad}}_{[k]}(n)| = 1$ because the isomorphism between $\mathfrak{g}/\mathfrak{h}$ and $\mathfrak{g}/\mathfrak{g}_{[k]}$ induced by $\mathrm{Ad}(k)$ intertwines $\overline{\mathrm{Ad}}(n)$ and $\overline{\mathrm{Ad}}_{[k]}(knk^{-1})$. We now construct s_0 as follows. We first choose a $\frac{1}{2}$-density s_0 at $[e]$, i.e., a map which assigns a number to a frame at $T_{[e]}G/H$ with the correct transformation property under change of frames. We then define s_0 globally by

$$s_0([k];(f_{[k]})) := s_0([e];L_{k^{-1}*}(f_{[k]})).$$

What we have to show is that this is well defined and G and N/H invariant. To show that it is well defined, choose $\xi \in \mathfrak{g}$ such that $(f_{[k]}) = (\xi_{G/H}|_{[k]})$ and choose any $h \in H$. Then $[k] = [kh]$ and we calculate

$$\begin{aligned} &s_0([e];L_{(kh)^{-1}*}(\xi_{G/H})) \\ &\quad = s_0([e];(\mathrm{Ad}(h^{-1})(\mathrm{Ad}(k^{-1})\xi))_{G/H}) \\ &\quad = |\det \overline{\mathrm{Ad}}_{[e]}(h)|^{1/2} \cdot s_0([e];((\mathrm{Ad}(k^{-1})\xi)_{G/H})) \\ &\quad = s_0([e];L_{k^{-1}*}(\xi_{G/H})) \end{aligned}$$

which shows that the definition of s_0 is independent of the choice of k representing $[k]$, i.e., that s_0 is well defined. A similar reasoning, which is left to the reader, shows that this s_0 is invariant under both the G and N/H actions. The uniqueness follows from the fact that the only degree of freedom in the determination of s_0 lies in the choice of the value of s_0 at $[e]$, which is uniquely determined by a nonzero number (nonzero to guarantee that s_0 is nowhere vanishing). Q.E.D.

Proposition 4: If s_0 used in the construction of Φ is (also) G-invariant, then $\Omega^{1/2}(R) \cong \mathrm{im}(\Phi \circ i)$ is characterized by the conditions:

$$\forall g \in G\colon\ L_g^* \hat{\psi} = \hat{\psi} \text{ or equivalently } \forall \xi \in \mathfrak{g}\colon\ \mathscr{L}_{\xi_Q} \hat{\psi} = 0. \tag{17}$$

Proof: According to Lemma 3, the above condition is compatible with, but stronger than the condition in Theorem 1. It follows directly from formula (12) and the G-invariance of s_0 that $\Phi(\psi)$ satisfies the above condition if ψ is independent of the parameter in G/N. On the other hand, if $\hat{\psi}$ satisfies the above condition then it follows from the G-invariance of s_0 and the reconstruction formula (14) that $\Phi^{-1}(\hat{\psi})$ is independent of G/N. Q.E.D.

Remark 5: If G admits a bi-invariant metric, then each homogeneous space G/H admits a G and N/H invariant metric from which one deduces that there exists a nowhere vanishing G and N/H invariant $\frac{1}{2}$-density. It follows that in such a case Proposition 4 provides us with a nice intrinsic description of $\Omega^{1/2}(R)$ as subspace of $\Omega^{1/2}(Q)$.

Remark 6: If we recall that the classical constraint $J(\xi)$ is given by $J(\xi) = H_{\xi_Q}$, then according to Sec. IV $\tau(J(\xi)) = \mathscr{L}_{\xi_Q}$. It follows that propositions 2 and 4 express $\Omega^{1/2}(R)$ as a subspace of $\Omega^{1/2}(Q)$ in terms of the quantized constraints. However, if the fibers G/H are not compact, then the only square integrable element in $\mathrm{im}(\Phi \circ i)$ is identically zero. Hence in such a case this description does not carry over to the Hilbert spaces. This negative result can be stated in several positive ways of which we give one well known example.

Corollary 7: If G is compact, then the unique positive s_0 such that $\int_{G/H} |s_0|^2 = 1$ makes $\Phi \circ i$ an isomorphism between the Hilbert space $\mathscr{H}^{1/2}(R)$ and the Hilbert subspace $\mathscr{H}_{\mathrm{red}} \subset \mathscr{H}^{1/2}(Q)$ defined by the quantized constraint equations $\tau(J(\xi))\psi = 0$.

VI. QUANTIZATION AND REDUCTION II

In this section we will justify our approach of the previous section by showing in corollary nine that quantization and reduction commute for all choices of s_0. Let $H\colon T^*R \to \mathbf{R}$ be an observable which is at the most linear in momenta (Sec. IV), i.e., $H = H_R \circ \pi_R + H_X$, X a vector field on R. Since T^*R is obtained as a Marsden–Weinstein reduction: $T^*R \cong J^{-1}(0)/G$, we obtain a pull-back $\tilde{H}$ of H to $J^{-1}(0) \subset T^*Q$, which is then by definition a G-invariant function on the constraint set $J^{-1}(0)$. We now suppose that there exists a function $\hat{H}\colon T^*Q \to \mathbf{R}$ which is at the most linear in momenta such that $\hat{H}|_{J^{-1}(0)} = \tilde{H}$ and $\hat{H}$ is invariant under the G-action on T^*Q. This implies that $\hat{H} = H_Q \circ \pi_Q + H_{\hat{X}}$ with $H_Q = H_R \circ p$ and $\hat{X}$ a vector field on Q, invariant under the G-action on Q, which projects to X: $p_* \hat{X} = X$. In the classical setting, the fact $\hat{H}|_{J^{-1}(0)} = \tilde{H}$ means that $\hat{H} \in C^\infty(T^*Q)$ represents, after reduction, the observable $H \in C^\infty(T^*R)$. We will now show that our identification of $\Omega^{1/2}(R)$ with a subspace of $\Omega^{1/2}(Q)$ commutes with quantization, i.e., that $\tau(\hat{H}) \circ \Phi \circ i = \Phi \circ i \circ \tau(H)$. Since we will show this by using $\Omega^{1/2}(R;G/N)$, we need to define $\mathscr{L}_X \psi$ for elements $\psi \in \Omega^{1/2}(R;G/N)$ and X a vector field on R. But this we do exactly as for ordinary $\frac{1}{2}$-densities on R, just by neglecting the additional parameter $z \in G/N$. It then follows automatically that $\mathscr{L}_X \circ i = i \circ \mathscr{L}_X$ (slight abuse of notation regarding $\mathscr{L}_X$).

Proposition 8: Let $H \in C^\infty(T^*R)$ and $\hat{H} \in C^\infty(T^*Q)$ be as above, then

$$\tau(\hat{H}) \circ \Phi = \Phi \circ \tau(H). \tag{18}$$

Proof: $\tau(\hat{H})\hat{\psi} = iH_Q \cdot \hat{\psi} + \mathscr{L}_{\hat{X}}\hat{\psi}$ and $\tau(H)\psi = iH_R \cdot \psi + \mathscr{L}_X \psi$. A glance at the defining formula for Φ and the fact that $H_Q = p^* H_R$ shows that $H_Q \cdot \Phi(\psi) = (p^* H_R) \cdot \Phi(\psi) = \Phi(H_R \cdot \psi)$, so it only remains to show that $\mathscr{L}_{\hat{X}} \Phi(\psi) = \Phi(\mathscr{L}_X \cdot \psi)$. Denote by $\hat{\phi}_t$ the flow of $\hat{X}$ and by ϕ_t the flow of X, then $\phi_t \circ p = p \circ \hat{\phi}_t$ (because $p_* \hat{X} = X$), $\mathscr{J} \circ \phi_t = \mathscr{J}$ (because $\hat{\phi}_t$ commutes with the G-action) and for $\xi \in \mathfrak{g}$: $\hat{\phi}_{t*} \xi_Q = \xi_Q$ (again because $\hat{\phi}_t$ commutes with the G-action). Using these facts and the notations (and conditions) of formula 14 in Theorem 1, we get for $\hat{\psi} \in \mathrm{im}(\Phi)$ the identity:

$$\begin{aligned} &\Phi^{-1}(\hat{\psi})(\phi_t(u),z;(\phi_{t*} f_u)) \\ &\quad = \hat{\psi}(\hat{\phi}_t(q);(\hat{\psi}_{t*} \hat{f}_u, \hat{\phi}_{t*} \xi_Q))/s_0([k];(\xi_{G/H})). \end{aligned}$$

By taking derivatives in t we find

$$\begin{aligned} &\mathscr{L}_X(\Phi^{-1}(\hat{\psi}))(u,z;(f_u)) \\ &\quad = (\mathscr{L}_{\hat{X}}\hat{\psi})(q;(\hat{f}_u,\xi_Q))/s_0([k];(\xi_{G/H})) \end{aligned}$$

or in other words $\mathscr{L}_X(\Phi^{-1}(\hat{\psi})) = \Phi^{-1}(\mathscr{L}_{\hat{X}}\hat{\psi})$. Q.E.D.

Corollary 9: For any s_0 used in the construction of Φ we have:

$$\tau(\hat{H})\circ\Phi\circ i = \Phi\circ i\circ\tau(H). \tag{19}$$

VII. TWO EXAMPLES

In this section we give two examples to substantiate our claims that (i) the term $\frac{1}{2}\mathrm{tr}(\mathrm{ad}(\xi))$ in Proposition 2 is essential and (ii) $\mathrm{im}(\Phi\circ i)$ cannot be described intrinsically in the general case.

In our first example we consider $Q=\mathbf{R}^3\ni(x,y,z)$, $G=\{((a,b),(0,1/a))|a\in\mathbf{R}^+,b\in\mathbf{R}\}\subset\mathrm{SL}(2,\mathbf{R})$ and the free action of G on Q given by $L_{(a,b)}(x,y,z)=(x,y+\log a, az+b\exp(\lambda x-y))$, where λ is a fixed real parameter. In this case the quotient R is the real line $R\equiv Q/G=\mathbf{R}$ with projection $p(x,y,z)=(x)$. As basis for the Lie algebra $\mathfrak{g}$ we choose $\xi_1=((-1,0),(0,1))$ and $\xi_2=((0,-1),(0,0))$ whose fundamental vector fields on Q are given by

$$\xi_{1Q}=\partial_y+z\partial_z \text{ and } \xi_{2Q}=\exp(\lambda x-y)\partial_z;$$

the momentum map J: $T^*Q\to\mathfrak{g}^*$ is given by

$$J(\xi_1)(x,y,z,p_x,p_y,p_z)=p_y+zp_z$$

and

$$J(\xi_2)=\exp(\lambda x-y)p_z$$

and the "correction" terms $\frac{1}{2}$tr ad are given by

$$\tfrac{1}{2}\mathrm{tr}(\mathrm{ad}(\xi_1))=-1 \text{ and } \tfrac{1}{2}\mathrm{tr}(\mathrm{ad}(\xi_2))=0.$$

We will identify the $\frac{1}{2}$-density spaces $\Omega^{1/2}(Q \text{ resp. } R)$ with the function spaces $C^\infty(Q \text{ resp. } R)$ using the nowhere vanishing $\frac{1}{2}$-densities $\sqrt{|dx\,dy\,dz|}$, resp. $\sqrt{|dx|}$, where $\sqrt{|dx\,dy\,dz|}(x,y,z;(a_{ij}\partial_j))=|\det a|^{1/2}$ and $\sqrt{|dx|}(x;a\partial_x)=|a|^{1/2}$. The squares of these trivializing $\frac{1}{2}$-densities are the Lebesgue measures on Q resp. R, so they give us the usual interpretation of the associated Hilbert spaces $\mathscr{H}^{1/2}$ (Q resp. $R)\cong L^2(Q \text{ resp. } R, d \text{ Leb})$. In order to compute the quantum operators in terms of functions we need the following formula (which can be obtained by a straightforward computation) for a vector field $\hat{X}$ on Q resp. X on R:

$$\mathscr{L}_{\hat{X}}(|dx\,dy\,dz|^{1/2})=\tfrac{1}{2}\,\mathrm{div}\,(\hat{X})\cdot|dx\,dy\,dz|^{1/2},$$
$$\text{resp. } \mathscr{L}_X(|dx|^{1/2})=\tfrac{1}{2}\,\mathrm{div}\,(X)\cdot|dx|^{1/2}, \tag{20}$$

where the divergence of a vector field $Y=\Sigma_j\eta^j(y)\partial_j$ is defined as $\mathrm{div}(Y)=\Sigma_j(\partial_j\eta^j)$. It follows that the quantized constraints are given by

$$\tau(J(\xi_1))\hat{\psi}=\partial_y\hat{\psi}+z\,\partial_z\hat{\psi}+\tfrac{1}{2}\hat{\psi}$$

and

$$\tau(J(\xi_2))\hat{\psi}=\exp(\lambda x-y)\partial_z\hat{\psi}.$$

Proposition 2 now gives us $\mathrm{im}(\Phi)\equiv\mathrm{im}(\Phi\circ i)\subset\Omega^{1/2}(Q)$ as

$$\mathrm{im}(\Phi\circ i)=\{\hat{\psi}\in C^\infty(Q)|\hat{\psi}(x,y,z)=\exp(y/2)\tilde{\psi}(x),\ \tilde{\psi}\in C^\infty(R)\}.$$

One might be tempted to identify $\hat{\psi}=\exp(y/2)\tilde{\psi}\in\mathrm{im}(\Phi\circ i)$ with $\tilde{\psi}\in\Omega^{1/2}(R)$, but this is incorrect. We must use the identification as given by Φ! Since G acts freely, s_0 is completely determined by its value at the identity and we choose $s_0(e;(\xi_1,\xi_2))=1$. Theorem 1 now gives us the correct identification:

$$\begin{aligned}\Phi^{-1}(\hat{\psi}|dx\,dy\,dz|^{1/2})(x;\partial_x)&\equiv\Phi^{-1}(\exp(y/2)\tilde{\psi}(x)|dx\,dy\,dz|^{1/2})(x;\partial_x)\\&=(\exp(y/2)\tilde{\psi}(x)|dx\,dy\,dz|^{1/2})(x,y,z;(\partial_x,\xi_1,\xi_2))/s_0(e;(\xi_1,\xi_2))\\&=(\exp(y/2)\tilde{\psi}(x)|dx\,dy\,dz|^{1/2})(x,y,z;(\partial_x,\partial_y+z\partial_z,\exp(\lambda x-y)\partial_z))\\&=\exp(y/2)\tilde{\psi}(x)|\exp(\lambda x-y)|^{1/2}\\&\equiv\exp(\lambda x/2)\tilde{\psi}(x)\end{aligned}$$
$$\Leftrightarrow\Phi^{-1}(\exp(y/2)\tilde{\psi}(x))=\exp(\lambda x/2)\tilde{\psi}(x).$$

In other words, the element $\hat{\psi}=\exp(y/2-\lambda x/2)\psi(x)$ has to be identified with $\psi(x)$ by the map Φ.

In order to see what happens with quantized observables, we determine the G-invariant vector fields $\hat{X}$ on Q, which are of the form

$$\hat{X}=f(x)\partial_x+g(x)\partial_y+(h(x)\exp(y)+z(\lambda f(x)-g(x)))\partial_z;$$

they project to vector fields $X=p_*\hat{X}$ on R as $X=f(x)\partial_x$. Again using formula (20) we find for the quantized observables:

$$\begin{aligned}\tau(H_{\hat{X}})&=\tau(fp_z+gp_y+(h\exp(y)+z(\lambda f-g))p_z)\\&=f\partial_x+g\partial_y+(h\exp(y)+z(\lambda f-g))\partial_z\\&\quad+\tfrac{1}{2}((\partial_x f)+\lambda f-g)\end{aligned}$$

$$\tau(H_X)=\tau(fp_x)=f\partial_x+\tfrac{1}{2}(\partial_x f).$$

With these formulas we can check Proposition 7:

$$\begin{aligned}\tau(H_{\hat{X}})\Phi(\psi)&=\tau(H_{\hat{X}})(\exp(y/2-\lambda x/2)\psi(x))\\&=(f(x)\partial_x\psi+\tfrac{1}{2}(\partial_x f)\psi(x))\\&\quad\times\exp(y/2-\lambda x/2)\\&=\Phi(\tau(H_X)\psi).\end{aligned}$$

This calculation shows that *if* we had naively identified $\hat{\psi}=\exp(y/2)\tilde{\psi}(x)\in\mathrm{im}(\Phi\circ i)$ with $\tilde{\psi}\in\Omega^{1/2}(R)$, then we would not have found $\tau(H_{\hat{X}})\circ\Phi'=\Phi'\circ\tau(H_X)$, where Φ' is our naive identification.

To show that we really need the term $\frac{1}{2}\mathrm{tr}(\mathrm{ad}(\xi))$ let us see what happens if we do not use it. The subspace

of $\Omega^{1/2}(Q)$ defined by the equations $\tau(J(\xi_1))\hat{\psi}=0=\tau(J(\xi_2))\hat{\psi}$ is given by

$$\{\hat{\psi}(x,y,z)=\exp(-y/2)\tilde{\psi}(x)\}.$$

Computing $\tau(H_{\hat{\chi}})$ on such an element we find

$$\tau(H_{\hat{\chi}})\exp(-y/2)\tilde{\psi}(x)$$
$$=\exp(-y/2)(f\partial_x\tilde{\psi}-g\tilde{\psi}+\tfrac{1}{2}\lambda f\tilde{\psi}+\tfrac{1}{2}(\partial_x f)\tilde{\psi})$$

which does not resemble $\tau(H_\chi)\tilde{\psi}$. We can get rid of the term $\frac{1}{2}\lambda f\tilde{\psi}$ by absorbing it in $\tilde{\psi}$: $\tilde{\psi}(x)=\exp(-\lambda x/2)\psi(x)$ which gives:

$$\tau(H_{\hat{\chi}})(\exp(-y/2-\lambda x/2)\psi(x))$$
$$=\exp(-y/2-\lambda x/2)((\tau(H_\chi)\psi)-g\psi).$$

However, since g is an arbitrary function, we cannot get rid of it by changing our identification (which should be independent of the observable we want to quantize). It follows that if we wish to have an identification which "commutes" with quantization of observables, we cannot forget the term $\frac{1}{2}\,\mathrm{tr}(\mathrm{ad}(\xi))$.

As the second example we consider $Q=\dot{\mathbb{R}}^2=\mathbb{R}^2\setminus\{(0,0)\}$ and $G=SL(2,\mathbb{R})$ with its natural action on $\dot{\mathbb{R}}^2\subset\mathbb{R}^2$. All isotropy subgroups are conjugated to $H=\{\left(\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}\right)\ b\in\mathbb{R}\}$ whose normalizer is $N=\{\left(\begin{smallmatrix}a & b\\ 0 & 1/a\end{smallmatrix}\right)\mid a\neq 0,\ b\in\mathbb{R}$ which has two components. $N/H\cong\dot{\mathbb{R}}=\mathbb{R}\setminus\{0\}$ with ordinary multiplication as group operation and projection $\left(\begin{smallmatrix}a & b\\ 0 & 1/a\end{smallmatrix}\right)\mapsto a$; the right action of N/H on Q is given by $R_a\left(\begin{smallmatrix}x\\ y\end{smallmatrix}\right)=\left(\begin{smallmatrix}ax\\ ay\end{smallmatrix}\right)$. The space G/N of isotropy groups is S^1 and the isotropy map $\mathcal{I}: Q=G/H\to G/N$ is given by $z\to z/\bar{z}$, where $z=x+iy$ represents $\left(\begin{smallmatrix}x\\ y\end{smallmatrix}\right)\in Q$. Since G acts transitively on Q, the reduced configuration space R is a single point. Hence the question whether $\mathrm{im}(\Phi\circ i)$ can be described intrinsically, using only the left action of G on Q, now boils down to the question whether we can describe a N/H-invariant $\frac{1}{2}$-density s_0 on $Q=B/H$ in terms of the left action of G on G/H.

An elementary calculation shows that for $n=\left(\begin{smallmatrix}a & b\\ 0 & 1/a\end{smallmatrix}\right)\in N$ we have $|\det\overline{\mathrm{Ad}}(n)|=a^{-2}$ which is not identically 1 on N. Consequently there does not exist a G and N/H invariant $\frac{1}{2}$-density on G/H, so Proposition 4 does not apply. The right action of N/H on G/H suggests the use of polar coordinates (r,ϕ) on $G/H=\dot{\mathbb{R}}^2$ ($x=r\cos\phi$, $y=r\sin\phi$). We thus can identify $\frac{1}{2}$-densities s_0 on G/H with functions f according to

$$f(r,\phi)=s_0(r,\phi;(\partial_r,\partial_\phi)),$$

which is equivalent to saying that we use $|dr\,d\phi|^{1/2}$ as trivializing nowhere vanishing $\frac{1}{2}$-density on G/H (see the previous example). The condition that s_0 is N/H-invariant then becomes (for $a>0$):

$$f(r,\phi)=s_0(r,\phi;(\partial_r,\partial_\phi))$$
$$=(R_a^*s_0)(r,\phi;(\partial_r,\partial_\phi))$$
$$=s_0(ar,\phi;(a\partial_r,\partial_\phi))$$
$$=|a|^{1/2}s_0(ar,\phi;(\partial_r,\partial_\phi))$$
$$=a^{1/2}f(ar,\phi)$$
$$\Leftrightarrow f(r,\phi)=r^{-1/2}\cdot f(1,\phi).$$

But, we should not forget the second component of N/H:

$$f(r,\phi)=(R_{-1}^*s_0)(r,\phi;(\partial_r,\partial_\phi))$$
$$=s_0(r,\phi+\pi;(\partial_r,\partial_\phi))$$
$$=f(r,\phi+\pi).$$

It follows that a N/H-invariant $\frac{1}{2}$-density s_0 on G/H is completely determined by a function f_0 on the circle S^1 by the formula:

$$f(r,\phi)=r^{-1/2}\cdot f_0(2\phi)\Leftrightarrow s_0=r^{-1/2}\cdot f_0(2\phi)\,|dr\,d\phi|^{1/2},\tag{21}$$

which is in complete agreement with our observation that two allowed s_0 differ by a function on G/N since in polar coordinates the isotropy map $\mathcal{I}$ is given by $(r,\phi)\mapsto(2\phi)$.

In order to see whether any such s_0 can be described in terms of the left action of G on G/H, let us calculate the fundamental vector fields on Q. We choose as basis for the Lie algebra $\mathfrak{g}$ the matrices $\xi_1=\left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$, $\xi_2=\left(\begin{smallmatrix}0 & -1\\ -1 & 0\end{smallmatrix}\right)$, and $\xi_3=\left(\begin{smallmatrix}-1 & 0\\ 0 & 1\end{smallmatrix}\right)$; for these we find:

$$\xi_{1Q}=\partial_\phi,$$
$$\xi_{2Q}=\sin(2\phi)r\partial_r+\cos(2\phi)\partial_\phi,$$
$$\xi_{3Q}=\cos(2\phi)r\partial_r-\sin(2\phi)\partial_\phi.$$

Using formula (20), we compute the quantized constraints $\tau(J(\xi))$ as operators on functions f [formula (21)]:

$$\tau(J(\xi_1))=\partial_\phi,$$
$$\tau(J(\xi_2))=\sin(2\phi)r\partial_r+\cos(2\phi)\partial_\phi$$
$$+\tfrac{1}{2}(\sin(2\phi)-2\sin(2\phi))$$
$$=\sin(2\phi)(r\partial_r-\tfrac{1}{2})+\cos(2\phi)\partial_\phi,$$
$$\tau(J(\xi_3))=\cos(2\phi)(r\partial_r-\tfrac{1}{2})-\sin(2\phi)\partial_\phi.$$

A simple exercise shows that there is no way to find functions $f(r,\phi)$ of the form $f(r,\phi)=r^{-1/2}\cdot f_0(2\phi)$, using the operators $\tau(J(\xi))$, without introducing "arbitrary" functions on Q. For instance, the "natural" choice $f(r,\phi)=r^{-1/2}$ is determined (up to a multiplicative constant) by the equations $\tau(J(\xi_1))f=0$ and $\tau(J(\xi_2))f=-\sin(2\phi)f$. This example also shows that, although $\mathrm{im}(\Phi)$ can be described in terms of the G-action only, it is not a very easy description because the conditions (in their infinitesimal form PDE's) vary from point to point.

ACKNOWLEDGMENTS

I would like to thank A. Weinstein and R. Montgomery for stimulating discussions and helpful remarks.

This work is supported by the Netherlands Organization for Scientific Research (NWO).

[1] J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[2] B. Kostant and S. Sternberg, Ann. Phys. **176**, 49 (1987).
[3] C. Duval, J. Elhadad, and G. M. Tuynman, Preprint CPT-89/P.2248; CPT-CNRS, Marseille, 1989, 38 pp.; to appear in Comm. Math. Phys.
[4] G. M. Tuynman, *Proceedings Seminar 1983–1985 Mathematical Structures in Field Theories, Vol. 1: Geometric Quantization* (CWI, Amsterdam, 1985), CWI Syllabus 8, 158 pp.

[5]I. E. Segal, J. Math. Phys. **1**, 468 (1960).
[6]K. Kuchar, Phys. Rev. D **34**, 3031, 3044 (1986).
[7]D. McMullan and J. Paterson, J. Math. Phys. **30**, 477 (1989).
[8]M. J. Gotay, J. Math. Phys. **27**, 2051 (1986).
[9]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[10]R. Montgomery, preprint of the Center for Pure and Applied Mathematics, University of California, Berkeley, PAM-143, 1983, 52 pp.

Unitary measurements of discrete quantities in quantum mechanics[a)]

Enrico G. Beltrametti and Gianni Cassinelli
Dipartimento di Fisica, Università di Genova, Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy

Pekka J. Lahti
Department of Physical Sciences, University of Turku, Finland

(Received 8 September 1988; accepted for publication 2 August 1989)

The pure measurements of discrete physical quantities are characterized within quantum theory of measurement and their unitary representations are given. Probabilistic aspects of measurements related to the so-called strong correlation conditions and a probabilistic characterization of the first kind measurements are examined. The problem of the objectification of the measurement result is analyzed in terms of a classical behavior of the measuring apparatus. As a by-product a generalization of the Wigner–Araki–Yanase theorem is given.

I. INTRODUCTION

In this paper, we discuss the quantum mechanical description of the measuring processes of an abstract way and determine the general form of the final state in the combined system of the object system and the measuring apparatus. The inputs for this determination are a probability reproducibility condition and an objectification requirement. The aim of this paper is to clarify the connections between individual physical requirements on the measuring process and the general form of the final state satisfying such requirements.

In this paper we consider only the so-called discrete physical quantities. This introduces drastic technical simplifications. We believe that the complete characterization of measurements of discrete physical quantities obtained here justifies this assumption.

The usual Hilbert space formulation of quantum mechanics will be applied here. (For basic results in functional analysis, see, e.g., Ref. 1.) Let us just fix some notation and terminology.

The description of a physical system $\mathcal{S}$ is based on a (complex, separable) Hilbert space $\mathcal{H}$, with the inner product $\langle\cdot|\cdot\rangle$. We let $\mathcal{L}(\mathcal{H})$ denote the set of bounded linear operators on $\mathcal{H}$; $\mathcal{P}(\mathcal{H})$ denotes the subset of $\mathcal{L}(\mathcal{H})$ consisting of the (orthogonal) projections. Any physical quantity of the system $\mathcal{S}$ is represented as (and identified with) a self-adjoint operator A in $\mathcal{H}$. The spectral measure of A is denoted by $P_A: \mathcal{B}(\mathcal{R}) \to \mathcal{P}(\mathcal{H})$, where $\mathcal{B}(\mathcal{R})$ is the Borel σ algebra of the real line $\mathcal{R}$. Any state of the system $\mathcal{S}$ is represented as (and identified with) an element T of $\mathcal{T}(\mathcal{H})_1^+$ of positive normalized trace class operators on $\mathcal{H}$. The extreme elements of the (convex) set $\mathcal{T}(\mathcal{H})_1^+$ are the one-dimensional projection operators $P[\varphi]$ on $\mathcal{H}$, $\varphi \in \mathcal{H}$, so that they may be identified, modulo a phase factor, with the unit vectors φ of $\mathcal{H}$. We refer to the extreme elements of $\mathcal{T}(\mathcal{H})_1^+$ (or unit vectors of $\mathcal{H}$) as the vector states of $\mathcal{S}$. In the absence of any superselection rule the vector states of $\mathcal{S}$ are exactly its pure states. The probability measure $P_A^T: \mathcal{B}(\mathcal{R}) \to [0,1]$, $X \mapsto P_A^T(X) := \mathrm{tr}(TP_A(X))$ defined by a physical quantity A and a state T is interpreted as the probability distribution of the values of the quantity A in the state T. (For further details of basic Hilbert space quantum mechanics, see, e.g., Ref. 2.)

As usual in the quantum theory of measurement (QTM), we restrict to measurements that preserve the identity of the object system $\mathcal{S}$ and the measuring apparatus $\mathcal{M}$. In this way QTM can be viewed as a part of the theory of compound systems with its own specific questions: which kind of state transformations of the compound system $\mathcal{S}+\mathcal{M}$ may serve as measurements of a quantity $A_{\mathcal{S}}$ of the subsystem $\mathcal{S}$, and which kind of properties of $\mathcal{M}$ allow us to determine in an unambiguous way the value of the measured quantity. In Sec. II we follow this approach, and we give a definition of the concept of measurement. In Sec. III we characterize all pure measurements of discrete physical quantities and give their unitary representations. In Sec. IV we deal with a characterization of measurements leading to strong correlations. In Sec. 5 we study some properties of measurements via the state transformations they induce. In Sec. VI we face the objectification problem by investigating the possibility of determining the value of the measured quantity. The results of the paper lead also to a generalization of the Wigner–Araki–Yanase theorem, which is given in Sec. VII.

II. PREMEASUREMENTS

Let $\mathcal{H}_{\mathcal{M}}$ be the Hilbert space of the measuring apparatus $\mathcal{M}$, and $A_{\mathcal{M}}$ the so-called pointer observable, i.e., a quantity of $\mathcal{M}$ that corresponds to the measured quantity $A_{\mathcal{S}}$ of the object system. (Hereafter every symbol referring to $\mathcal{S}$ or $\mathcal{M}$ will have the corresponding subindex.) Let $T_{\mathcal{S}}$ and $T_{\mathcal{M}}$ be the initial states of $\mathcal{S}$ and $\mathcal{M}$, so that the initial state of $\mathcal{S}+\mathcal{M}$ is uniquely determined as $T_{\mathcal{S}} \otimes T_{\mathcal{M}}$ since we assume that prior to the measurement $\mathcal{S}$ and $\mathcal{M}$ are both dynamically and probabilistically independent of each other. We write $W(T_{\mathcal{S}} \otimes T_{\mathcal{M}})$ for the final state of $\mathcal{S}+\mathcal{M}$; the final states of $\mathcal{S}$ and $\mathcal{M}$ will then be the reduced states $T_{\mathcal{S},W}$ and $T_{\mathcal{M},W}$, respectively. Here, e.g., $T_{\mathcal{S},W}$ is defined through

$$\mathrm{tr}(T_{\mathcal{S},W}P_{\mathcal{S}}) = \mathrm{tr}(W(T_S \otimes T_{\mathcal{M}})P_{\mathcal{S}} \otimes \mathbf{I}_{\mathcal{M}}),$$

[a)] This paper is a revised version of a preprint by the authors circulated under the same title.

 0022-2488/90/010091-08$03.00

where $P_{\mathscr{S}} \in \mathscr{P}(\mathscr{H}_{\mathscr{S}})$, and $\mathbf{I}_{\mathscr{M}}$ is the identity operator on $\mathscr{H}_{\mathscr{M}}$.

In order to qualify the four-tuple $\langle H_{\mathscr{M}}, A_{\mathscr{M}}, T_{\mathscr{M}}, W \rangle$ as a measurement of the quantity $A_{\mathscr{S}}$ it is necessary that it reproduce the basic probabilities $P_{A_{\mathscr{S}}}^{T_{\mathscr{S}}}$ via the pointer observable $A_{\mathscr{M}}$ and the final state $T_{\mathscr{M},W}$ of the measuring apparatus. We take this probability reproducibility condition as the defining property of the notion of premeasurement.

Definition 2.1: With the above notation, a *premeasurement* of $A_{\mathscr{S}}$ on $\mathscr{S}$ is a four-tuple $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, T_{\mathscr{M}}, W \rangle$ for which

$$P_{A_{\mathscr{S}}}^{T_{\mathscr{S}}}(X) = P_{A_{\mathscr{M}}}^{T_{\mathscr{M}},W}(X),$$

for all $X \in \mathscr{B}(\mathscr{R})$, and for any $T_{\mathscr{S}} \in \mathscr{T}(\mathscr{H}_{\mathscr{S}})_1^+$.

The additional requirement that allows us to go from the notion of premeasurement to the one of measurement is the objectification requirement. This is the requirement that a measurement should lead to an unambiguous result, and it is connected with the problem of justifying the subjective ignorance interpretation for the final state $T_{\mathscr{S},W}$ of the object system $\mathscr{S}$ through that of the final state $T_{\mathscr{M},W}$ of the measuring apparatus $\mathscr{M}$. This question is put aside till Sec. VI.

One of the main results of QTM is that for each physical quantity $A_{\mathscr{S}}$ of $\mathscr{S}$ there are premeasurements. An explicit example of such a premeasurement was already given by von Neumann.[3] A result of Ozawa[4] shows that such premeasurements exist also for continuous quantities.

We assume that the state transformation $T_{\mathscr{S}} \otimes T_{\mathscr{M}} \to W(T_{\mathscr{S}} \otimes T_{\mathscr{M}})$ induced by the measurement preserves the convex structure of the set of states. In that case W can be viewed as a trace preserving positive linear map on the state space $T(\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}})$, the Banach space of the trace class operators on $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$. Furthermore we require that the map $T_{\mathscr{S}} \otimes T_{\mathscr{M}} \to W(T_{\mathscr{S}} \otimes T_{\mathscr{M}})$ preserves the extreme points of the set of states of $\mathscr{S} + \mathscr{M}$, i.e., W is a pure map. In that case the premeasurement $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, T_{\mathscr{M}}, W \rangle$ is said to be *pure*. Its form is significantly constrained by the following result due to Davies.[5]

Lemma 2.2: Every pure positive linear map $W: \mathscr{T}(\mathscr{H}) \to \mathscr{T}(\mathscr{H})$ is of one of the following three forms:

(i) $W(T) = BTB^*$,

where $B \in \mathscr{L}(\mathscr{H})$, and B^* is the adjoint of B;

(ii) $W(T) = BT^*B^*$,

where $B: \mathscr{H} \to \mathscr{H}$ is bounded and conjugate linear;

(iii) $W(T) = \mathrm{tr}(TB)P[\xi]$,

where $B \in \mathscr{L}(\mathscr{H})$, $B \geqslant 0$, and $\xi \in \mathscr{H}$.

Because of the linearity and continuity of the theory, and because of the fact that the vector states determine all the states of a physical system, we may assume, without any loss of generality, that the initial states of $\mathscr{S}$ and $\mathscr{M}$ are vector states φ and Φ, i.e., $T_{\mathscr{S}} = P[\varphi]$ and $T_{\mathscr{M}} = P[\Phi]$ for some unit vectors $\varphi \in \mathscr{H}_{\mathscr{S}}$ and $\Phi \in \mathscr{H}_{\mathscr{M}}$. In that case

$$T_{\mathscr{S}} \otimes T_{\mathscr{M}} = P[\varphi] \otimes P[\Phi] = P[\varphi \otimes \Phi]$$

is the initial state of $\mathscr{S} + \mathscr{M}$.

Remark 2.3: Any measurement process $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, T_{\mathscr{M}}, W \rangle$ of a physical quantity $A_{\mathscr{S}}$ determines an *instrument*[5] (i.e., a state transformation valued measure) $\mathscr{I}_{\mathscr{M},W}$: $\mathscr{B}(\mathscr{R}) \to \mathscr{L}(\mathscr{T}(\mathscr{H}_{\mathscr{S}}))$ through the relation

$$\mathrm{tr}(\mathscr{I}_{\mathscr{M},W}(X)(T_{\mathscr{S}})P_{\mathscr{S}}) = \mathrm{tr}(W(T_{\mathscr{S}} \otimes T_{\mathscr{M}})P_{\mathscr{S}} \otimes P_{A_{\mathscr{M}}}(X)),$$

for all $P_{\mathscr{S}} \in \mathscr{P}(\mathscr{H}_{\mathscr{S}})$, $X \in \mathscr{B}(\mathscr{R})$, and $T_{\mathscr{S}} \in \mathscr{T}(\mathscr{H}_{\mathscr{S}})_1^+$. The instrument $\mathscr{I}_{\mathscr{M},W}$ contains all the information on the measurement $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, T_{\mathscr{M}}, W \rangle$ relevant to the object system $\mathscr{S}$. It gives the correct probabilities as

$$\mathrm{tr}(T_{\mathscr{S}} P_{A_{\mathscr{S}}}(X)) = \mathrm{tr}(\mathscr{I}_{\mathscr{M},W}(X)(T_{\mathscr{S}})), \quad \text{for all } X \in \mathscr{B}(\mathscr{R}),\ T_{\mathscr{S}} \in \mathscr{T}(\mathscr{H}_{\mathscr{S}})_1^+,$$

and it gives the (non-normalized) final state $\mathscr{I}_{\mathscr{M},W}(X)(T_{\mathscr{S}})$ of $\mathscr{S}$, with the condition that the measurement yielded a result in X. In particular, $\mathscr{I}_{\mathscr{M},W}(\mathscr{R})(T_{\mathscr{S}}) = T_{\mathscr{S},W}$. The properties of a measurement, like the one of being ideal or of the first kind, can most directly be studied through the properties of their instruments (cf. Sec. V).

III. UNITARY PREMEASUREMENTS OF DISCRETE QUANTITIES

Hereafter we restrict ourselves to discrete physical quantities, namely, to those that are represented by self-adjoint operators with pure point spectrum. The class of measurements of discrete quantities ranges from the yes–no measurements of the elementary quantities associated with the properties (projection operators) of the system to the measurements of discrete quantities with nondegenerate eigenvalues, including the measurements of discrete approximations of continuous quantities. Let us also notice that discrete quantities are characterized as those that admit repeatable measurements.[4]

We shall work out the general form of a pure premeasurement of a discrete quantity. Let $A_{\mathscr{S}}$ (in $\mathscr{H}_{\mathscr{S}}$) be a self-adjoint operator with a pure point spectrum. Let the distinct eigenvalues of $A_{\mathscr{S}}$ be a_i, $i = 1,2,...,N$, where $N \in \mathbf{N}$ (the set of natural numbers), or $N = \infty$. Let $n(i)$ be the degeneracy of the eigenvalue a_i [so that, again, $n(i) \in \mathbf{N}$, or $n(i) = \infty$] and let $\{\varphi_{ij} \mid j = 1,...,n(i)\}$ be an orthonormal basis of the corresponding eigenspace. Then

$$\{\varphi_{ij} \mid i = 1,...,N;\ j = 1,...,n(i)\} \subset \mathscr{H}_{\mathscr{S}}$$

is an orthonormal basis, and $A_{\mathscr{S}}\varphi_{ij} = a_i\varphi_{ij}$, for all $i = 1,...,N$, $j = 1,...,n(i)$. The spectral projections of $A_{\mathscr{S}}$ are

$$P_{A_{\mathscr{S}}}(X) = \sum_{i(a_i \in X)} \sum_j P[\varphi_{ij}].$$

For any $\varphi \in \mathscr{H}_{\mathscr{S}}$,

$$\varphi = \sum_{i,j} c_{ij}\varphi_{ij},$$

with $c_{ij} = \langle \varphi_{ij} \mid \varphi \rangle$, and we have that $\varphi \in \mathrm{dom}(A_{\mathscr{S}})$ if and only if the series

$$\sum_{i,j} |a_i c_{ij}|^2$$

is convergent.

Let $\mathscr{H}_{\mathscr{M}}$ be any (complex) Hilbert space with dimen-

sion N (the number of distinct eigenvalues of $A_{\mathscr{S}}$). Let $\{\Phi_i | i=1,...,N\} \subset \mathscr{H}_{\mathscr{M}}$ be an orthornormal basis. We define $A_{\mathscr{M}}$ as the simple self-adjoint operator

$$\sum_{i=1}^{N} a_i P[\Phi_i]$$

(in $\mathscr{H}_{\mathscr{M}}$) with the spectral measure

$$P_{A_{\mathscr{M}}}: X \to \sum_{i(a_i \in X)} P[\Phi_i]$$

and with the (nondegenerate) eigenvalues a_i, $i=1,...,N$. Let Φ be a fixed unit vector of $\mathscr{H}_{\mathscr{M}}$.

We shall first show that the form (iii) of Lemma 2.2 is not a possible premeasurement map of a nontrivial discrete quantity $A_{\mathscr{S}}$.

Theorem 3.1: If $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, P[\Phi], W \rangle$ is a premeasurement of $A_{\mathscr{S}}$, then W cannot be of the form

$$W(P[\varphi \otimes \Phi]) = \mathrm{tr}(P[\varphi \otimes \Phi]B)P[\xi],$$

with $B \in \mathscr{L}(\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}})$, $B \geqslant 0$, $\xi \in \mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$.

Proof: Assume that W has this form. Then, by Definition 2.1,

$$\langle \varphi | P_{A_{\mathscr{S}}}(X)\varphi \rangle = \langle \varphi \otimes \Phi | B(\varphi \otimes \Phi) \rangle \langle \xi | (\mathbf{I}_{\mathscr{S}} \otimes P_{A_{\mathscr{M}}}(X))\xi \rangle, \quad X \in \mathscr{B}(\mathscr{R}).$$

For $X = \mathscr{R}$, we have

$$1 = \langle \varphi \otimes \Phi | B(\varphi \otimes \Phi) \rangle \langle \xi | \xi \rangle,$$

while for $X = \{a_i\}$ and $\varphi = \varphi_{ij}$ we get

$$1 = \langle \varphi_{i,j} \otimes \Phi | B(\varphi_{i,j} \otimes \Phi) \rangle \langle \xi | (\mathbf{I}_{\mathscr{S}} \otimes P[\Phi_i])\xi \rangle,$$

for any $j = 1,...,n(i)$. We conclude that

$$\langle \xi | \xi \rangle = \langle \xi | (\mathbf{I}_{\mathscr{S}} \otimes P[\Phi_i])\xi \rangle,$$

for any $i = 1,...,N$. As $\Sigma P[\Phi_i] = \mathbf{I}_{\mathscr{M}}$ we also have

$$\langle \xi | \xi \rangle = \sum \langle \xi | (\mathbf{I}_{\mathscr{S}} \otimes P[\Phi_i])\xi \rangle.$$

Hence $\langle \xi | (\mathbf{I}_{\mathscr{S}} \otimes P[\Phi_i])\xi \rangle = 0$, for any $i = 1,...,N$, so that either $\xi = 0$, or $P[\Phi_i] = 0$ for any $i = 1,...,N$, i.e., $A_{\mathscr{M}}$ is constant. □

In view of this result we use the notation $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, B \rangle$ as a premeasurement of $A_{\mathscr{S}}$ whenever $\Phi \in \mathscr{H}_{\mathscr{M}}$ is a unit vector and B is a bounded linear or conjugate linear map on $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$. The probability reproducibility condition of Definition 2.1 then takes the simple form

$$\langle \varphi | P_{A_{\mathscr{S}}}(X)\varphi \rangle = \langle B(\varphi \otimes \Phi) | (\mathbf{I}_{\mathscr{S}} \otimes P_{A_{\mathscr{M}}}(X))B(\varphi \otimes \Phi) \rangle, \tag{1}$$

for all $X \in \mathscr{B}(\mathscr{R})$ and $\varphi \in \mathscr{H}_{\mathscr{S}}$, $\|\varphi\| = 1$.

The next theorem characterizes all pure premeasurements of $A_{\mathscr{S}}$.

Theorem 3.2: A four-tuple $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, B \rangle$ is a pure premeasurement of $A_{\mathscr{S}}$ if and only if B is a continuous linear or conjugate linear extension on $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$ of a map of the form

$$\varphi_{ij} \otimes \Phi \to \psi_{ij} \otimes \Phi_i, \quad i = 1,...,N, \ j = 1,...,n(i), \tag{2}$$

where $\{\psi_{ij}: i = 1,...,N; j = 1,...,n(i)\}$ is any set of unit vectors (in $\mathscr{H}_{\mathscr{S}}$) that are orthogonal with respect to the second index, i.e., for any $i = 1,...,N$, $\langle \psi_{ij} | \psi_{ik} \rangle = \delta_{jk}$, for all $j,k = 1,...,n(i)$. Moreover, B can always be chosen as a unitary or an antiunitary mapping.

Proof: Let $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, B \rangle$ be a pure premeasurement of $A_{\mathscr{S}}$. From Eq. (1) we get $1 = \|B(\varphi \otimes \Phi)\|^2$ for any unit vector $\varphi \in \mathscr{H}_{\mathscr{S}}$. Putting $X = \{a_i\}$ and $\varphi = \varphi_{i,j}$ we obtain

$$1 = \langle B(\varphi_{i,j} \otimes \Phi) | (\mathbf{I}_{\mathscr{S}} \otimes P[\Phi_i])B(\varphi_{i,j} \otimes \Phi) \rangle.$$

As $B(\varphi_{i,j} \otimes \Phi)$ is a unit vector this shows that

$$(\mathbf{I}_{\mathscr{S}} \otimes P[\Phi_i])B(\varphi_{i,j} \otimes \Phi) = B(\varphi_{i,j} \otimes \Phi)$$

so that $B(\varphi_{ij} \otimes \Phi)$ is of the form (2) with

$$\psi_{ij} = \sum_{n,m} \langle \varphi_{nm} \otimes \Phi_i | B(\varphi_{ij} \otimes \Phi) \rangle \varphi_{nm} \otimes \Phi_i.$$

The vectors ψ_{ij} are unit vectors, for

$$\langle \psi_{ij} | \psi_{ij} \rangle = \langle B(\varphi_{i,j} \otimes \Phi) | B(\varphi_{i,j} \otimes \Phi) \rangle = 1.$$

The orthogonality condition $\langle \psi_{ij} | \psi_{il} \rangle = \delta_{jl}$, $j,l = 1,...,n(i)$, is readily obtained from Eq. (1) when applied, e.g., to the vectors $\varphi = (1/\sqrt{2})(\varphi_{ij} + \varphi_{il})$, and $\varphi = (1/\sqrt{2})(\varphi_{ij} + i\varphi_{il})$, respectively, with $X = \{a_i\}$.

Conversely, assume that B is a continuous linear or conjugate linear extension of the map (2). Then a direct computation shows that $\langle H_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, B \rangle$ satisfies the probability reproducibility condition (1).

Since $\{\varphi_{ij} \otimes \Phi\}$ and $\{\psi_{ij} \otimes \Phi_i\}$ are orthonormal sets of $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$ they can be extended to orthonormal bases of $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$. Any bijective mapping U between any two such bases such that

$$U(\varphi_{ij} \otimes \Phi) = \psi_{ij} \otimes \Phi, \quad \text{for all } i \text{ and } j,$$

can be extended uniquely by linearity (conjugate linearity) and continuity to a unitary (antiunitary) operator U satisfying the probability reproducibility condition. □

The nonuniqueness of the operator B in this theorem has no consequence on the physics of the measurement process which is completely determined by the map (2). In view of this fact we consider a fixed unitary premeasurement $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, U \rangle$ of $A_{\mathscr{S}}$.

The final states of $\mathscr{S}$ and $\mathscr{M}$ can be calculated from the final state

$$P[U(\varphi \otimes \Phi)] = P\left[\sum_{i=1}^{N} \sum_{j=1}^{n(i)} c_{ij} \psi_{ij} \otimes \Phi_i\right]$$

of $\mathscr{S} + \mathscr{M}$. We get

$$T_{\mathscr{S},U} = \sum_{i=1}^{N} \sum_{j=1}^{n(i)} \sum_{l=1}^{n(i)} \bar{c}_{ij} c_{il} |\psi_{il}\rangle\langle\psi_{ij}|$$

and

$$T_{\mathscr{M},U} = \sum_{i=1}^{N} \sum_{j=1}^{n(i)} \sum_{k=1}^{N} \sum_{l=1}^{n(k)} \bar{c}_{ij} c_{kl} \langle \psi_{ij} | \psi_{kl} \rangle |\Phi_k\rangle\langle\Phi_i|.$$

We define

$$\gamma_i = N_i^{-1} \sum_{j=1}^{n(i)} c_{ij} \psi_{ij},$$

with

$$N_i^2 = \left\| \sum_j c_{ij} \psi_{ij} \right\|^2 = \sum_j |c_{ij}|^2,$$

whenever $N_i \neq 0$, and $\gamma_i = 0$ otherwise. Then

$$T_{\mathscr{S},U} = \sum_{i=1}^{N} N_i^2 P[\gamma_i] \tag{3}$$

and

$$T_{\mathscr{M},U} = \sum_{i=1}^{N} \sum_{k=1}^{N} N_i N_k \langle \gamma_i | \gamma_k \rangle | \Phi_k \rangle \langle \Phi_i |. \tag{4}$$

A particular choice for the set $\{\psi_{ij}\}$ is $\{\varphi_{ij}\}$. Let V be a unitary operator on $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$ that has the restriction

$$V(\varphi_{ij} \otimes \Phi) = \varphi_{ij} \otimes \Phi_i.$$

The four-tuple $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, V \rangle$ is referred to as the von Neumann–Lüders measurement of $A_{\mathscr{S}}$. The corresponding final states of $\mathscr{S}$ and $\mathscr{M}$ will become

$$T_{\mathscr{S},V} = \sum_{i=1}^{N} N_i^2 P[\omega_i], \tag{5}$$

with

$$\omega_i = N_i^{-1} \sum_{j=1}^{n(i)} c_{ij} \varphi_{ij}, \quad \text{if } N_i \neq 0,$$

while

$$\omega_i = 0, \quad \text{if } N_i = 0,$$

and

$$T_{\mathscr{M},V} = \sum_{i=1}^{N} N_i^2 P[\phi_i]. \tag{6}$$

IV. CORRELATIONS

In addition to the basic probability reproducibility condition (1) there are other probabilistic aspects of measurements especially important for the objectification of the measurement result (Sec. VI). These aspects refer to the type of correlations produced by a measurement.

In the Appendix the general definition of correlation is given; here we recall only that for any $P_{\mathscr{S}} \in \mathscr{P}(\mathscr{H}_{\mathscr{S}})$, $P_{\mathscr{M}} \in \mathscr{P}(\mathscr{H}_{\mathscr{M}})$, and $T \in \mathscr{T}(\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}})_1^+$, the correlation $\rho(P_{\mathscr{S}}, P_{\mathscr{M}}, T)$ can be written as

$$\rho(P_{\mathscr{S}}, P_{\mathscr{M}}, T) = \frac{\operatorname{tr}(TP_{\mathscr{S}} \otimes P_{\mathscr{M}}) - \operatorname{tr}(T_{\mathscr{S}} P_{\mathscr{S}}) \operatorname{tr}(T_{\mathscr{M}} P_{\mathscr{M}})}{\sqrt{\operatorname{tr}(T_{\mathscr{S}} P_{\mathscr{S}})(1 - \operatorname{tr}(T_{\mathscr{S}} P_{\mathscr{S}})) \operatorname{tr}(T_{\mathscr{M}} P_{\mathscr{M}})(1 - \operatorname{tr}(T_{\mathscr{M}} P_{\mathscr{M}}))}},$$

where $T_{\mathscr{S}}$ and $T_{\mathscr{M}}$ are the reduced state of T. As shown in the Appendix,

$$\rho(P_{\mathscr{S}}, P_{\mathscr{M}}, T) = 1 \text{ iff } \operatorname{tr}(T_{\mathscr{S}} P_{\mathscr{S}}) = \operatorname{tr}(T_{\mathscr{M}} P_{\mathscr{M}}).$$

Let $\langle H_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, U \rangle$, $U \in \mathscr{U}(\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}})$, be a unitary premeasurement of $A_{\mathscr{S}}$. The strong correlations

$$\rho(P[\gamma_i], P[\Phi_i], U(\varphi \otimes \Phi)) = 1 \tag{7}$$

and

$$\rho(P_{A_{\mathscr{S}}}(\{a_i\}), P_{A_{\mathscr{M}}}(\{a_i\}), U(\varphi \otimes \Phi)) = 1, \tag{8}$$

$i = 1,...,N$, need not hold, in general, so that they may be stated as further requirements on the premeasurement $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, U \rangle$, and either of them will imply substantial restrictions on U. We shall now study these restrictions.

Assume first that $N_i = 0$, i.e., $\langle \varphi | P_{A_{\mathscr{S}}}(\{a_i\}) \varphi \rangle = 0$, for some $i = 1,...,N$. In that case

$$\gamma_i = \sum_j c_{ij} \psi_{ij} = 0,$$

as $c_{ij} = 0$, for any $j = 1,...,n(i)$, so that neither (7) nor (8) can be required. Assume next that $N_i = 1$, for some $i = 1,...,N$. In that case, $c_{kl} = 0$, for any $k \neq i$, so that the final state $P[U(\varphi \otimes \Phi)]$ of $\mathscr{S} + \mathscr{M}$ is the uncorrelated state $P[\gamma_i] \otimes P[\Phi_i]$. The pointer observable $A_{\mathscr{M}}$ now has the value a_i (with probability equal to 1), though the measured quantity $A_{\mathscr{S}}$ does not need to have the value a_i, i.e., $\langle \gamma_i | P_{A_{\mathscr{S}}}(\{a_i\}) \gamma_i \rangle$ does not need to be 1; nor does $P[\gamma_i]$ need to be equal to $P[\varphi]$.

The above considerations show that the strong correlation conditions (7) and (8) can be required only for those $i = 1,...,N$, and $\varphi \in \mathscr{H}_{\mathscr{S}}$, $\|\varphi\| = 1$, for which $0 \neq N_i \neq 1$.

Theorem 4.1: Let $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, U \rangle$ be a unitary premeasurement of $A_{\mathscr{S}}$. Then

$$\rho(P[\gamma_i], P[\Phi_i], U(\varphi \otimes \Phi)) = 1,$$

for any $i = 1,...,N$, and $\varphi \in \mathscr{H}_{\mathscr{S}}$, $\|\varphi\| = 1$, for which $0 \neq N_i \neq 1$, if and only if

$$\{\psi_{ij} | i = 1,...,N,\ j = 1,...,n(i)\} \subset \mathscr{H}_{\mathscr{S}}$$

is an orthonormal system.

Proof: We have

$$\rho(P[\gamma_i], P[\Phi_i], U(\varphi \otimes \Phi)) = 1$$

if and only if

$$\operatorname{tr}(T_{\mathscr{S},U} P[\gamma_i]) = \operatorname{tr}(T_{\mathscr{M},U} P[\Phi_i]).$$

Moreover,

$$\operatorname{tr}(T_{\mathscr{S},U} P[\gamma_i]) = \sum_{k=1}^{N} N_k^2 \operatorname{tr}(P[\gamma_k] P[\gamma_i]) = N_i^2 + \sum_{k(k \neq i)} N_k^2 |\langle \gamma_k | \gamma_i \rangle|^2,$$

and

$$\operatorname{tr}(T_{\mathscr{M},U} P[\Phi_i]) = \sum_j \bar{c}_{ij} c_{il} \langle \psi_{ij} | \psi_{il} \rangle = \sum_j |c_{ij}|^2 = N_i^2.$$

If

$$\{\psi_{ij} | i = 1,...,N,\ j = 1,...,n(i)\}$$

is an orthonormal system, then

$$\langle \gamma_k | \gamma_i \rangle = N_k^{-1} N_i^{-1} \sum_j \sum_l \bar{c}_{kj} c_{il} \langle \psi_{kj} | \psi_{il} \rangle = 0,$$

for all $k \neq i$, so that

$$\rho(P[\gamma_i], P[\Phi_i], U(\varphi \otimes \Phi)) = 1$$

(whenever $0 \neq N_i \neq 1$). To show the converse, choose, e.g.,

$$\varphi = (1/\sqrt{2})\varphi_{ij} + (1/\sqrt{2})\varphi_{kl}, \quad i \neq k.$$

To obtain

$$\rho(P[\gamma_i],P[\Phi_i],U(\varphi\otimes\Phi))=1,$$

it is then necessary that $|\langle\psi_{ij}|\psi_{kl}\rangle|^2=0$, i.e., $\langle\psi_{ij}|\psi_{kl}\rangle=0$. The same argument can be repeated for any pairs (i,j) and (k,l) such that $i\neq k$. Hence

$$\{\psi_{ij}|i=1,...,N,\quad j=1,...,n(i)\}$$

is an orthonormal system. □

Theorem 4.2: Let $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ be a unitary premeasurement of $A_{\mathscr{S}}$. Then

$$\rho(P_{A_{\mathscr{S}}}(\{a_i\}),P_{A_{\mathscr{M}}}(\{a_i\}),U(\varphi\otimes\Phi))=1,$$

for any $i=1,...,N$, and $\varphi\in\mathscr{H}_{\mathscr{S}}$, $\|\varphi\|=1$, for which $0\neq N_i\neq 1$, if and only if

$$\{\psi_{ij}|i=1,...,N,\quad j=1,...,n(i)\}\subset\mathscr{H}_{\mathscr{S}}$$

is an orthonormal system with $P_{A_{\mathscr{S}}}(\{a_i\})\gamma_i=\gamma_i$, for any $i=1,...,N$.

Proof: As

$$P_{A_{\mathscr{S}}}(\{a_i\})=\sum_j P[\varphi_{ij}]$$

and

$$P_{A_{\mathscr{M}}}(\{a_i\})=P[\Phi_i],$$

we have

$$\mathrm{tr}\left(T_{\mathscr{S},U}\left(\sum_j P[\varphi_{ij}]\right)\right)$$

$$=N_i^2\sum_j|\langle\gamma_i|\varphi_{ij}\rangle|^2+\sum_{k(k\neq i)}N_k^2\sum_{j=1}^{n(i)}|\langle\gamma_k|\varphi_{ij}\rangle|^2,$$

$\mathrm{tr}(T_{\mathscr{M},U}P[\phi_i])=N_i^2$.

Thus

$$\rho(P_{A_{\mathscr{S}}}(\{a_i\}),P_{A_{\mathscr{M}}}(\{a_i\}),U(\varphi\otimes\Phi))=1$$

if and only if

$$\sum_{j=1}^{n(i)}|\langle\gamma_i|\varphi_{ij}\rangle|^2=1\quad\text{and}\quad\langle\gamma_k|\varphi_{ij}\rangle=0,$$

for any $k\neq i$. This is the case if and only if $P_{A_{\mathscr{S}}}(\{a_i\})\gamma_i=\gamma_i$. As this is to hold for any $i=1,...,N$, $\varphi\in\mathscr{H}_{\mathscr{S}}$, $\|\varphi\|=1$, we have that

$$\{\psi_{ij}|i=1,...,N,\quad j=1,...,n(i)\}$$

is an orthonormal system. □

If the premeasurement $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ of $A_{\mathscr{S}}$ produces strong correlations between the possible values of $A_{\mathscr{M}}$ and $A_{\mathscr{S}}$, then it also produces strong correlations between the final "component states" $P[\Phi_i]$ and $P[\gamma_i]$ of $\mathscr{M}$ and $\mathscr{S}$. As is well known, the von Neumann–Lüders measurement $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,V\rangle$ of $A_{\mathscr{S}}$ also has these correlation properties.

V. INSTRUMENTS OF THE UNITARY MEASUREMENTS

The probabilistic aspects of $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ do not exhaust the physics of the measurement process. The measurements that produce strong correlations may differ from each other in the transformations of the states they induce on the object system $\mathscr{S}$. Such differences can most directly be studied by the instruments $\mathscr{I}_{\mathscr{M},U}$ determined by the premeasurements $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ of $A_{\mathscr{S}}$ (cf. Remark 2.3). We shall now consider some properties of the instruments $\mathscr{I}_{\mathscr{M},U}$ that serve to distinguish the von Neumann–Lüders instrument $\mathscr{I}_{\mathscr{M},V}$ from those associated with other unitary measurements.

Let us determine first the form of the von Neumann–Lüders instrument $\mathscr{I}_{\mathscr{M},V}$. Rewriting the final state $T_{\mathscr{S},V}$ of $\mathscr{S}$ as

$$T_{\mathscr{S},V}=\sum_{i=1}^{N}N_i^{-2}P[\omega_i]$$

$$=\sum_i P_{A_{\mathscr{S}}}(\{a_i\})P[\varphi]P_{A_{\mathscr{S}}}(\{a_i\}),$$

one sees that $\mathscr{I}_{\mathscr{M},V}$ has the following form:

$$\mathscr{I}_{\mathscr{M},V}(X)(P[\varphi])=\sum_{i(a_i\in X)}P_{A_{\mathscr{S}}}(\{a_i\})P[\varphi]P_{A_{\mathscr{S}}}(\{a_i\})$$

$$=\sum_{i(a_i\in X)}N_i^2P[\omega_i],\tag{9}$$

for all $X\in\mathscr{B}(\mathscr{R})$, and for any $\varphi\in\mathscr{H}_{\mathscr{S}}$, $\|\varphi\|=1$. This instrument has the well known properties of being *ideal*, i.e.,

$$\text{if } \mathrm{tr}(\mathscr{I}_{\mathscr{M},V}(\{a_i\})(P[\varphi]))=1,\quad\text{for some } i=1,...,N,$$
$$\text{then } \mathscr{I}_{\mathscr{M},V}(\{a_i\})(P[\varphi])=P[\varphi],\tag{10}$$

and of the *first kind*, i.e.,

$$\mathrm{tr}(\mathscr{I}_{\mathscr{M},V}(\{a_i\})(P[\varphi]))$$
$$=\mathrm{tr}(\mathscr{I}_{\mathscr{M},V}(\{a_i\})\circ\mathscr{I}_{\mathscr{M},V}(\mathscr{R})(P[\varphi])),$$
$$\text{for any } \varphi\in\mathscr{H}_{\mathscr{S}},\ \|\varphi\|=1,\text{ and for all } i=1,...,N.\tag{11}$$

Consider next any unitary measurement $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ of $A_{\mathscr{S}}$. The form of $\mathscr{I}_{\mathscr{M},U}$ can again readily be extracted from the final state

$$T_{\mathscr{S},U}=\sum_{i=1}^{N}N_i^2P[\gamma_i]$$

$$=\sum_i\sum_j\sum_l|\psi_{il}\rangle\langle\varphi_{il}|\varphi\rangle\langle\varphi|\varphi_{ij}\rangle\langle\psi_{ij}|,$$

and we get

$$\mathscr{I}_{\mathscr{M},U}(X)(P[\varphi])=\sum_{i(a_i\in X)}\sum_j\sum_l|\psi_{il}\rangle\langle\varphi_{il}|P[\varphi]|\varphi_{ij}\rangle\langle\psi_{ij}|$$

$$=\sum_{(a_i\in X)}N_i^2P[\gamma_i],\tag{12}$$

for any $X\in\mathscr{B}(\mathscr{R})$, and for all $\varphi\in\mathscr{H}_{\mathscr{S}}$, $\|\varphi\|=1$.

The following theorem shows that the premeasurements $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ of $A_{\mathscr{S}}$ that have the strong correlation property (8) are characterized by instruments of the first kind.

Theorem 5.1: Let $\langle\mathscr{H}_{\mathscr{M}},A_{\mathscr{M}},\Phi,U\rangle$ be a unitary premeasurement of $A_{\mathscr{S}}$, and let $\mathscr{I}_{\mathscr{M},U}$ be the corresponding instrument. Then $\mathscr{I}_{\mathscr{M},U}$ is of the first kind if and only if

$$\rho(P_{A_{\mathscr{S}}}(\{a_i\}),P_{A_{\mathscr{M}}}(\{a_i\}),U(\varphi\otimes\Phi))=1,$$

for all $i=1,...,N$, and for any $\varphi\in\mathscr{H}_{\mathscr{S}}$, $\|\varphi\|=1$, for which $0\neq N_i\neq 1$.

Proof: By definition, $\mathcal{I}_{\mathcal{M},U}$ is of the first kind if and only if

$$\mathrm{tr}(I_{\mathcal{M},U}(\{a_i\})(P[\varphi]))$$
$$= \mathrm{tr}(\mathcal{I}_{\mathcal{M},U}(\{a_i\})\circ\mathcal{I}_{\mathcal{M},U}(\mathcal{R})(P[\varphi])),$$

i.e.,

$$\mathrm{tr}(P[\varphi]P_{A_{\mathcal{S}}}(\{a_i\})) = \mathrm{tr}(T_{\mathcal{S},U}P_{A_{\mathcal{S}}}(\{a_i\})),$$

for any $i = 1,...,N$, and $\varphi\in\mathcal{H}_{\mathcal{S}}$, $\|\varphi\| = 1$. But we also have

$$\mathrm{tr}(P[\varphi]P_{A_{\mathcal{S}}}(\{a_i\})) = \mathrm{tr}(T_{\mathcal{M},U}P_{A_{\mathcal{M}}}(\{a_i\})),$$

for any $\varphi\in\mathcal{H}_{\mathcal{S}}$, $\|\varphi\| = 1$, and for all $i = 1,...,N$. This is exactly the case when

$$\rho(P_{A_{\mathcal{S}}}(\{a_i\}),P_{A_{\mathcal{M}}}(\{a_i\}),U(\varphi\otimes\Phi)) = 1,$$

for any $i = 1,...,N$, and $\varphi\in\mathcal{H}_{\mathcal{S}}$, $\|\varphi\| = 1$, for which $0\neq N_i\neq 1$. □

The ideality of an instrument $\mathcal{I}_{\mathcal{M},U}$ leads to a further specification of U.

Theorem 5.2: Let $\langle\mathcal{H}_{\mathcal{M}},A_{\mathcal{M}},\Phi,U\rangle$ be a unitary premeasurement of $A_{\mathcal{S}}$, and let $\mathcal{I}_{\mathcal{M},U}$ be the corresponding instrument. If $\mathcal{I}_{\mathcal{M},U}$ is ideal, then U is of the form $U(w_i\otimes\Phi) = e^{i\alpha_i}V(w_i\otimes\Phi)$, for any $w_i\in P_{A_{\mathcal{S}}}(\{a_i\})(\mathcal{H}_{\mathcal{S}})$, $i = 1,...,N$, and thus $\mathcal{I}_{\mathcal{M},U} = I_{\mathcal{M},V}$.

Proof: Assume first that $\mathcal{I}_{\mathcal{M},U}$ is ideal, i.e., for any $i = 1,...,N$, $\mathcal{I}_{\mathcal{M},U}(\{a_i\})(P[\varphi]) = P[\varphi]$ whenever $\mathrm{tr}(I_{\mathcal{M},U}(\{a_i\})(P[\varphi])) = 1$, $\varphi\in\mathcal{H}_{\mathcal{S}}$, $\|\varphi\| = 1$. Now for any $i = 1,...,N$,

$$\mathrm{tr}(\mathcal{I}_{\mathcal{M},U}(\{a_i\})(P[\varphi_{ij}])) = \langle\varphi_{ij}|P_{A_{\mathcal{S}}}(\{a_i\})\varphi_{ij}\rangle = 1,$$

for all $j = 1,...,n(i)$. Hence $\mathcal{I}_{\mathcal{M},U}(\{a_i\})(P[\varphi_{ij}]) = P[\psi_{ij}] = P[\varphi_{ij}]$, i.e., $\psi_{ij} = e^{i\theta_{ij}}\varphi_{ij}$, $\theta_{ij}\in\mathcal{R}$, for all $i = 1,...,N$, $j = 1,...,n(i)$.

Applying the same argument for the vector states $\varphi = (1/\sqrt{2})(\varphi_{ij}+\varphi_{il})$ we see that, for any $i = 1,...,N$, $e^{i\theta_{ij}} = e^{i\theta_{il}}$, for all $j, l = 1,...,n(i)$. This shows that $U = U'\circ V = U_{\mathcal{S}}\otimes\mathbf{I}_{\mathcal{M}}\circ V$, with $U_{\mathcal{S}}|_{M_i} = c_i\mathbf{I}_{\mathcal{S}}|_{M_i}$, $|c_i| = 1$, where $M_i = P_A(\{a_i\})(\mathcal{H}_{\mathcal{S}})$. Clearly, $\mathcal{I}_{\mathcal{M},U} = \mathcal{I}_{\mathcal{M},V}$. □

VI. OBJECTIFICATION OF THE MEASUREMENT RESULT

It is implicit in the very notion of a measuring apparatus that $A_{\mathcal{M}}$ has an actual value in the final state of $\mathcal{M}$ though this value can be subjectively unknown. The objectivity of $A_{\mathcal{M}}$ in the final state of $\mathcal{M}$ is not coded in the notion of premeasurement $\langle\mathcal{H}_{\mathcal{M}},A_{\mathcal{M}},T_{\mathcal{M}},W\rangle$ but it has to be taken as an additional requirement. The stronger requirement of $A_{\mathcal{S}}$ being objective in the final state of $\mathcal{S}$ can then be deduced from the objectivity of $A_{\mathcal{M}}$ via the strong correlations.

To discuss these highly interpretational items we introduce first some appropriate definitions. We say that a discrete quantity A is *objective* in a vector state φ if $\langle\varphi|P_A(\{a_k\})\varphi\rangle = 1$, for some eigenvalue a_k of A, i.e., if φ is an eigenvector of A. This is the case exactly when A commutes with $P[\varphi]$. If A is objective in the state φ, then its ideal first kind measurement does not change the state of the system since then $\mathcal{I}_{\mathcal{M},V}(\mathcal{R})(P[\varphi]) = P[\varphi]$. We say that a physical quantity is *classical* if it commutes with any other physical quantity of the system (cf., e.g., Ref. 6). The only self-adjoint operators in $\mathcal{H}$ that commute with any other self-adjoint operator are the constants $a\mathbf{I}$, $a\in\mathcal{R}$. Hence the assumption that A is a (nonconstant) classical quantity implies that there is no longer either the one-to-one correspondence between the physical quantities and the self-adjoint operators in $\mathcal{H}$ nor the one between the physical states and the elements of $\mathcal{T}(\mathcal{H})_1^+$. In particular, a physical state corresponds to an equivalence class of $\mathcal{T}(\mathcal{H})_1^+$ and the presence of a one-dimensional projection in this equivalence class does not imply that the state is pure. A vector state is pure if and only if its equivalence class contains only one element.[2,6]

We assume now that the pointer observable $A_{\mathcal{M}}$ is a classical quantity. In that case *the only pure states* of the measuring apparatus $\mathcal{M}$ are the eigenstates $P[\Phi_k]$ of the pointer observable $A_{\mathcal{M}}$. Moreover, the final state (4) of the measuring apparatus $\mathcal{M}$ reduces to (i.e., it is in the same equivalence class as)

$$T_{\mathcal{M},U} = \sum_{i=1}^{N} N_i^2 P[\Phi_i]. \tag{13}$$

We stress that a mixed state never has a unique decomposition into the vector states.[2,7] The assumption that $A_{\mathcal{M}}$ is classical implies that (13) is the only decomposition of $T_{\mathcal{M},U}$ into pure states of $\mathcal{M}$. This also means that the subjective ignorance interpretation can be adopted for the final state $T_{\mathcal{M},U}$ of the measuring apparatus: when $\mathcal{M}$ is in the state $T_{\mathcal{M},U}$, it is actually in one of the pure states $P[\Phi_k]$, the coefficients N_i^2 describing the degree of our knowledge on the actual state of $\mathcal{M}$. The *actual value* of the pointer observable $A_{\mathcal{M}}$ can be read (by an ideal first kind test), without changing the actual state of $\mathcal{M}$. In that state $A_{\mathcal{M}}$ is clearly objective.

The final state of the object system $\mathcal{S}$ is now

$$T_{\mathcal{S},U} = \sum_{i=1}^{N} N_i^2 P[\gamma_i].$$

This state is not directly affected by the assumption that $A_{\mathcal{M}}$ is classical. But, if

$$\rho(P[\gamma_i],P[\Phi_i],U(\varphi\otimes\Phi)) = 1$$

holds true for each $i = 1,...,N$ (for which $0\neq N_i\neq 1$), then the ignorance interpretation of $T_{\mathcal{M},U}$ can be adopted for $T_{\mathcal{S},U}$, too. According to Theorem 4.1 this occurs exactly when $\{\psi_{ij}\}$ is an orthonormal system. Assuming that this is the case, if $P[\Phi_k]$ is the actual final state of $\mathcal{M}$, then $P[\gamma_k]$ is the actual final state of $\mathcal{S}$. Then $\langle\gamma_k|P_{A_{\mathcal{S}}}(\{a_k\})\gamma_k\rangle$ is the probability that in the actual final state $P[\gamma_k]$ of the object system $\mathcal{S}$ the measured quantity $A_{\mathcal{S}}$ has the value a_k. This number needs not to be 1, however. If, in addition, the strong correlation condition (8) is required, then γ_i's are eigenvectors of $A_{\mathcal{S}}$, and $A_{\mathcal{S}}$ is objective in the actual final state of $\mathcal{S}$.

There is, however, a puzzling fact that arises from the present approach, to which we shall now turn.

Consider again a unitary measurement $\langle\mathcal{H}_{\mathcal{M}},A_{\mathcal{M}},\Phi,U\rangle$ of the quantity $A_{\mathcal{S}}$. The pointer observable $A_{\mathcal{M}}$ can be interpreted through $I_{\mathcal{S}}\otimes A_{\mathcal{M}}$ as a quantity of the compound system $\mathcal{S}+\mathcal{M}$. As the von Neumann algebra $\mathcal{L}(\mathcal{H}_{\mathcal{S}}\otimes\mathcal{H}_{\mathcal{M}})$ of the bounded operators on

$\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$ is generated by the operators of the product form $A \otimes B$, $A \in \mathscr{L}(\mathscr{H}_{\mathscr{S}})$, $B \in \mathscr{L}(\mathscr{H}_{\mathscr{M}})$ (see Ref. 8), we see that if $A_{\mathscr{M}}$ is a classical quantity of the measuring apparatus $\mathscr{M}$, then $\mathbf{I}_{\mathscr{S}} \otimes A_{\mathscr{M}}$ is a classical quantity of the compound system $\mathscr{S} + \mathscr{M}$. This fact has two interesting consequences. First, the physical state represented by the projection operator $P[U(\varphi \otimes \Phi)]$ is equivalent to $\Sigma N_i^2 P[\gamma_i] \otimes P[\Phi_i]$. (Recall that this result also can be obtained as a solution of the so-called consistency problem of the measurement theory.[2]) Second, the unitary measurement U, which can be given the form e^{iH}, does not qualify H as a physical quantity (like a Hamiltonian between $\mathscr{S}$ and $\mathscr{M}$) of the compound system. Indeed, if H is a physical quantity and $A_{\mathscr{M}}$ is a classical quantity of $\mathscr{M}$, then

$$P^{P[\Phi]}_{A_{\mathscr{M}}} = P^{P[\varphi\otimes\Phi]}_{I_{\mathscr{S}}\otimes A_{\mathscr{M}}} = P^{P[U(\varphi\otimes\Phi)]}_{I_{\mathscr{S}}\otimes A_{\mathscr{M}}} = P^{T_{\mathscr{M},U}}_{A_{\mathscr{M}}},$$

for any unit vector $\varphi \in \mathscr{H}_{\mathscr{S}}$, as U commutes with $\mathbf{I}_{\mathscr{S}} \otimes A_{\mathscr{M}}$. In particular, this would imply that for any $i = 1,..,N$, and for any vector state φ of $\mathscr{S}$,

$$|\langle \Phi_i | \Phi \rangle|^2 = \sum_j |\langle \varphi_{ij} | \varphi \rangle|^2,$$

and thus

$$P[\Phi] = \sum_i |\langle \Phi_i | \Phi \rangle|^2 P[\Phi_i]$$

$$= \sum_{i,j} |\langle \varphi_{ij} | \varphi \rangle|^2 P[\Phi_i] = T_{\mathscr{M},U}.$$

These results show that the following two assumptions are mutually incompatible: (a) $A_{\mathscr{M}}$ is a classical quantity, and (b) $U = e^{iH}$ qualifies H as a physical quantity.

Such an incompatibility has been independently pointed out by van Fraassen.[9]

We do not go deeper into this issue: we remark only that the inconsistency of assumptions (a) and (b) seems to be in accordance with the ideas that the measurement evolutions could be irreversible and thus nonunitary (see, e.g., Ref. 10), or that, instead of a sharp objectification, only an unsharp objectification might be realized. The latter proposal can be justified, e.g., by the results of Ozawa[4] and by the fact that the so-called realistic measurement of position, momentum, and spin seem to define not sharp but unsharp quantities.[11]

VII. WIGNER–ARAKI–YANASE THEOREM

The quantum theory of measurement shows up some limitations on the measurability of physical quantities. Complementary physical quantities like position and momentum cannot be measured together, i.e., none of the measurements (in the sense of Definition 2.1) of complementary quantities can be combined as a joint measurement of such quantities.[12] Continuous quantities, like position or momentum, do not admit an ideal first kind measurement; in particular, none of the measurements with a completely positive measurement map (like a unitary one) of such quantities can satisfy the (weak) repeatability condition.[4]

Another type of limitation of measurability of physical quantities is said to arise from the existence of conservation laws, like, e.g., the conservation of linear momentum or angular momentum. This type of limitation was discovered by Wigner[13] and it was later cast in the form of a theorem by Araki and Yanase[14] which, with our notations, would read as follows.

Theorem 7.1: Let $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, V \rangle$ be the von Neumann–Lüders measurement of $A_{\mathscr{S}}$. Let $L = L_{\mathscr{S}} \otimes \mathbf{I}_{\mathscr{M}} + \mathbf{I}_{\mathscr{S}} \otimes L_{\mathscr{M}}$ be a (bounded) self-adjoint operator on $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$. Assume that L is a constant of motion of $\mathscr{S} + \mathscr{M}$ with respect to V, i.e., $[L, V] = 0$. Then also $[L_{\mathscr{S}}, A_{\mathscr{S}}] = 0$.

The results of previous sections suggest a formal generalization expressed by the following theorem whose proof includes, as a particular case, that of Theorem 7.1.

Theorem 7.2: Let $\langle \mathscr{H}_{\mathscr{M}}, A_{\mathscr{M}}, \Phi, U \rangle$ be a unitary measurement of $A_{\mathscr{S}}$. Let $L = L_{\mathscr{S}} \otimes \mathbf{I}_{\mathscr{M}} + \mathbf{I}_{\mathscr{S}} \otimes L_{\mathscr{M}}$ be a bounded self-adjoint operator on $\mathscr{H}_{\mathscr{S}} \otimes \mathscr{H}_{\mathscr{M}}$. If L commutes with U, then

$$L_{\mathscr{S}} \text{ commutes with } A_{\mathscr{S}}, \tag{14}$$

provided that either

$$\rho(P[\gamma_i], P[\Phi_i], U(\varphi \otimes \Phi)) = 1, \tag{15}$$

for any $i = 1,...,N$ and $\varphi \in \mathscr{H}_{\mathscr{S}}$, $\|\varphi\| = 1$, for which $0 \neq N_i \neq 1$, or

$$L_{\mathscr{M}} \text{ commutes with } A_{\mathscr{M}}. \tag{16}$$

Proof: As U is unitary and L commutes with U we have, for any pair of indices (i,j) and (k,l),

$$\begin{aligned}\langle \varphi_{ij} \otimes \Phi | L(\varphi_{kl} \otimes \Phi) \rangle &= \langle \varphi_{ij} | L_{\mathscr{S}} \varphi_{kl} \rangle + \delta_{ik}\delta_{jl} \langle \Phi | L_{\mathscr{M}} \Phi \rangle \\ &= \langle \varphi_{ij} \otimes \Phi | U^* U L (\varphi_{kl} \otimes \Phi) \rangle = \langle U(\varphi_{ij} \otimes \Phi) | L U (\varphi_{kl} \otimes \Phi) \rangle \\ &= \langle \psi_{ij} \otimes \Phi_i | L(\psi_{kl} \otimes \Phi_k) \rangle = \langle \psi_{ij} | L_{\mathscr{S}} \psi_{kl} \rangle \delta_{ik} + \langle \psi_{ij} | \psi_{kl} \rangle \langle \Phi_i | L_{\mathscr{M}} \Phi_k \rangle.\end{aligned}$$

If (15) holds, then by Theorem 4.1 $\langle \psi_{ij} | \psi_{kl} \rangle = \delta_{ik}\delta_{jl}$. If (16) holds, then $\langle \Phi_i | L_{\mathscr{M}} \Phi_k \rangle = g(a_k)\delta_{ik}$, for some Borel function g. In both cases we obtain

$$\langle \varphi_{ij} | L_{\mathscr{S}} \varphi_{kl} \rangle = \delta_{ik} \langle \varphi_{ij} | L_{\mathscr{S}} \varphi_{kl} \rangle,$$

for all $i,k = 1,2,...,N$.

Let $P_n = P_{A_{\mathscr{S}}}(\{a_n\})$, $n = 1,2,...,N$. Then

$$\langle \varphi_{ij} | L_{\mathscr{S}} P_n \varphi_{kl} \rangle - \langle \varphi_{ij} | P_n L_{\mathscr{S}} \varphi_{kl} \rangle = (\delta_{nk} - \delta_{ni}) \delta_{ik} \langle \varphi_{ij} | L_{\mathscr{S}} \varphi_{kl} \rangle = 0,$$

for any $i,k,n = 1,2,...,N$, independently of the indices j and l. As $\{\varphi_{ij}\}$ is an orthonormal base of $\mathscr{H}_{\mathscr{S}}$ we thus have $L_{\mathscr{S}} P_n = P_n L_{\mathscr{S}}$, for any $n = 1,2,...,N$, i.e., $L_{\mathscr{S}}$ and $A_{\mathscr{S}}$ commute. □

ACKNOWLEDGMENTS

The main parts of this work were carried out during some visits of P.L. to the University of Genoa. These visits were made possible by the grants of the Department of Phys-

ics of the University of Genoa and of I.N.F.N., which are gratefully acknowledged. This author (P.L.) also thanks his co-workers for the warm hospitality they offered to him during these visits. Finally, he wishes to express his debt to Dr. Paul Busch, Dr. Peter Mittelstaedt, and Dr. Frank Schroeck for useful discussions on the quantum measurement problem.

APPENDIX: PROPERTIES OF CORRELATION

Let A and B be commuting self-adjoint (not necessarily bounded) operators in a (complex, separable) Hilbert space $\mathcal{H}$. Then A and B have a joint projection valued measure, which we denote $P_{A,B}$. We recall only that the spectral measures of A and B are related to $P_{A,B}$ as follows:

$$P_A(X) = P_{A,B}(X \times \mathcal{R}), \quad P_B(Y) = P_{A,B}(\mathcal{R} \times Y),$$

for all $X, Y \in \mathcal{B}(\mathcal{R})$.

Let φ be any unit vector in the common domain of A and B, and consider the probability measures $P_A^{P[\varphi]}$ and $P_B^{P[\varphi]}$ (on $\mathcal{R}$) and $P_{A,B}^{P[\varphi]}$ (on $\mathcal{R}^2$). Then we have the following facts: (a) the function $x \to x^2$ is both $P_A^{P[\varphi]}$- and $P_B^{P[\varphi]}$-integrable; and (b) the function $(x,y) \to xy$ is $P_{A,B}^{P[\varphi]}$-integrable. By (a), the expectations $\mathscr{E}(A,\varphi)$ and $\mathscr{E}(B,\varphi)$, the variances $\mathscr{V}(A,\varphi)$ and $\mathscr{V}(B,\varphi)$, and the standard deviations $\sigma(A,\varphi) = \mathscr{V}(A,\varphi)^{1/2}$ and $\sigma(B,\varphi) = \mathscr{V}(B,\varphi)^{1/2}$ of the probability measures $P_A^{P[\varphi]}$ and $P_B^{P[\varphi]}$ are well defined. Fact (b) assures that we may define the *correlation of A and B in the state* φ, denoted by $\rho(A,B,\varphi)$, as the (normalized) correlation of the probability measure $P_{A,B}^{P[\varphi]}$. Explicitly

$$\rho(A,B,\varphi) = \frac{\int (x - \mathscr{E}(A,\varphi))(y - \mathscr{E}(B,\varphi)) dP_{A,B}^{P[\varphi]}(x,y)}{\sigma(A,\varphi)\sigma(B,\varphi)}.$$

To simplify the notation we define

$$A' = \frac{A - \mathscr{E}(A,\varphi)I}{\sigma(A,\varphi)}, \quad B' = \frac{B - \mathscr{E}(B,\varphi)I}{\sigma(B,\varphi)}.$$

Then, by the properties of the spectral measure $P_{A,B}$, we have

$$\rho(A,B,\varphi) = \langle A'\varphi \,|\, B'\varphi \rangle.$$

Moreover, a quick calculation shows that the following three conditions are equivalent:

$$\rho(A,B,\varphi) = 1,$$

$$\mathscr{V}(A/\sigma(A,\varphi) - B/\sigma(B,\varphi), \varphi) = 0,$$

$$A'\varphi = B'\varphi.$$

Let f be a bounded Borel function on $\mathcal{R}$. If $A'\varphi = B'\varphi$, then $f(A')\varphi = f(B')\varphi$, and, in particular, $P_{A'}^{P[\varphi]} = P_{B'}^{P[\varphi]}$ since A' and B' are commuting self-adjoint operators. The condition $P_{A'}^{P[\varphi]} = P_{B'}^{P[\varphi]}$ is obviously also sufficient for the equality $A'\varphi = B'\varphi$. We may thus conclude that

$$\rho(A,B,\varphi) = 1 \quad \text{iff} \quad P_{A'}^{P[\varphi]} = P_{B'}^{P[\varphi]}. \tag{A1}$$

We remark that (A1) is the Hilbert space counterpart of a classical result in probability theory that can be found, e.g., in Ref. 15.

In case A and B are *commuting projections* in $\mathcal{H}$, (A1) is easily seen to take the form

$$\rho(A,B,\varphi) = 1 \quad \text{iff} \quad \langle \varphi \,|\, A\varphi \rangle = \langle \varphi \,|\, B\varphi \rangle. \tag{A2}$$

In Sec. IV the previous results are used in the particular context of the Hilbert space $\mathcal{H}$ being the tensor product $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{M}}$. If $\mathcal{P}_{\mathcal{S}} \in \mathcal{P}(\mathcal{H}_{\mathcal{S}})$ and $P_{\mathcal{M}} \in \mathcal{P}(\mathcal{H}_{\mathcal{M}})$, then $P_{\mathcal{S}} \otimes I_{\mathcal{M}}$ and $I_{\mathcal{S}} \otimes P_{\mathcal{M}}$ are commuting operators on $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{M}}$ and their correlation is well defined. If $T \in \mathcal{T}(\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{M}})_1^+$, then

$$\mathscr{E}(P_{\mathcal{S}} \otimes I_{\mathcal{M}}, T) = \mathscr{E}(P_{\mathcal{S}}, T_{\mathcal{S}}),$$

$$\mathscr{E}(I_{\mathcal{S}} \otimes P_{\mathcal{M}}, T) = \mathscr{E}(P_{\mathcal{M}}, T_{\mathcal{M}}),$$

$$\mathscr{V}(P_{\mathcal{S}} \otimes I_{\mathcal{M}}, T) = \mathscr{V}(P_{\mathcal{S}}, T_{\mathcal{S}}),$$

$$\mathscr{V}(I_{\mathcal{S}} \otimes P_{\mathcal{M}}, T) = \mathscr{V}(P_{\mathcal{M}}, T_{\mathcal{M}})$$

(where $T_{\mathcal{S}}$ and $T_{\mathcal{M}}$ are the reduced states of T). Thus we obtain the definition of correlation stated in Sec. IV, and (A2) can be written as

$$\rho(P_{\mathcal{S}} \otimes I_{\mathcal{M}}, I_{\mathcal{S}} \otimes P_{\mathcal{M}}, T) = 1 \quad \text{iff} \quad \mathscr{E}(P_{\mathcal{S}}, T_{\mathcal{S}}) = \mathscr{E}(P_{\mathcal{M}}, T_{\mathcal{M}}). \tag{A3}$$

[1] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I: Function Analysis* (Academic, New York, 1972).
[2] E. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1981).
[3] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1955) (original German text in 1932).
[4] M. Ozawa, "Quantum measuring processes of continuous observables," J. Math. Phys. **25**, 79 (1984).
[5] E. Davies, *Quantum Theory of Open Systems* (Academic, New York, 1976).
[6] C. Piron, *Foundations of Quantum Physics* (Benjamin, New York, 1976).
[7] E. Beltrametti, "The non-unique decomposition of mixtures: Some remarks," *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore, 1985), pp. 85–95.
[8] M. Takesaki, *Theory of Operator Algebras* (Springer, Berlin, 1979).
[9] B. van Fraassen (private communication).
[10] *Quantum Theory and Measurement*, edited by J. Wheeler and W. Zurek (Princeton, U.P., Princeton, NJ, 1983).
[11] P. Busch, "Unsharp reality and the question of quantum systems," *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore, 1987), pp. 105–125.
[12] P. Lahti, "Uncertainty and Complementarity, some measurement-theoretical and information-theoretical aspects," *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore, 1987), pp. 181–207.
[13] E. Wigner, "Die Messung Quantenmechanischer Operatoren," Z. Phys. **133**, 101 (1952).
[14] H. Araki and M. Yanase, "Measurements of quantum mechanical operators," Phys. Rev. **120**, 622 (1960).
[15] W. Feller, *An Introduction to Probability Theory* (Wiley, New York, 1950).

Variational and perturbative schemes for a spiked harmonic oscillator

V. C. Aguilera-Navarro
Instituto de Física Teórica—UNESP, 01405 São Paulo, SP, Brazil

G. A. Estévez
Department of Mathematics and Physical Science, Inter American University, San German, Puerto Rico 00753

R. Guardiola
Departamento de Física Moderna, Universidad de Granada, E-18071 Granada, Spain

(Received 20 June 1989; accepted for publication 23 August 1989)

A variational analysis of the spiked harmonic oscillator Hamiltonian operator $-d^2/dx^2 + x^2 + l(l+1)/x^2 + \lambda |x|^{-\alpha}$, where α is a real positive parameter, is reported in this work. The formalism makes use of the functional space spanned by the solutions of the Schrödinger equation for the linear harmonic oscillator Hamiltonian supplemented by a Dirichlet boundary condition, and a standard procedure for diagonalizing symmetric matrices. The eigenvalues obtained by increasing the dimension of the basis set provide accurate approximations for the ground state energy of the model system, valid for positive and relatively large values of the coupling parameter λ. Additionally, a large coupling perturbative expansion is carried out and the contributions up to fourth-order to the ground state energy are explicitly evaluated. Numerical results are compared for the special case $\alpha = 5/2$.

I. INTRODUCTION

A general problem connected with the so-called spiked harmonic oscillator (SHO) Hamiltonian $-d^2/dx^2 + x^2 + l(l+1)/x^2 + \lambda |x|^{-\alpha}$, where α is a positive constant, has been thoroughly studied by Harrell.[1] The name spiked derives from the graphical appearance of the perturbative term $\lambda |x|^{-\alpha}$. The quantity λ is a positive definite parameter and measures the strength of the perturbative potential. The angular momentum term is represented by the expression $l(l+1)/x^2$. The spiked harmonic oscillator problem is of practical importance as it occurs in both chemical and nuclear physics.[1-7] In the elegant work reported in Ref. 1, a modified perturbation series to a finite order is employed to obtain analytical expressions for the eigenenergies of a SHO Hamiltonian for small values of λ, and arbitrary values of the exponent α. In this work, we report attempts to solve the SHO problem, employing a variational procedure and a large coupling perturbative calculation. A short review of the SHO problem is presented in Sec. II. The variational approach is outlined in Sec. III. The large coupling expansion is discussed and developed in Sec. IV, and a summary of the results and conclusions is found in Sec. V.

II. BACKGROUND

To compare our results with those reported by Harrell,[1] we concentrate ourselves on the zero angular momentum case. The Hamiltonian associated with the SHO then reads

$$H(\alpha,\lambda) = \frac{-d^2}{dx^2} + x^2 + \lambda |x|^{-\alpha} \equiv H_0 + \lambda V, \qquad (2.1)$$

where H_0 is formally equal to the simple harmonic oscillator Hamiltonian, and $V = |x|^{-\alpha}$. The sum of H_0 and λV must be understood to be the Friedrichs extension of the quadratic form defined by Eq. (2.1) on the domain of the Schwartz space with the boundary condition $u(0) = 0$, with $u(x)$ denoting a solution of the Schrödinger equation for the simple harmonic oscillator. The latter condition is necessary since not all functions in the domain of H_0 are in the domain of V. Therefore, when $\lambda \to 0$, α fixed, the operator $H(\alpha,\lambda)$ converges to an operator formally equal to $-d^2/dx^2 + x^2$, but supplemented by the Dirichlet boundary condition (DBC) that all functions in its domain vanish at $x = 0$. This operator is H_0. With this definition, the family of operators $H(\alpha,\lambda)$ is both analytic for $\lambda > 0$, and continuous for $\lambda \to 0^+$.

The spectrum of H_0 consists of the (two-fold degenerate) simple harmonic oscillator eigenvalues, whose eigenfunctions vanish at $x = 0$. Since one purpose of the present work is to consider the perturbation expansion of the eigenvalues of $H(\alpha,\lambda)$, all operators will be restricted to the space $L^2[0,\infty]$, with the DBC $u(0) = 0$, to avoid problems stemming from the degeneracy of the spectrum.

As has been pointed out elsewhere, the perturbation V considered in this article is singular.[1,4-6] The series expansion for the eigenvalues $E_n(\alpha,\lambda)$ of $H_0 + \lambda V$, calculated by means of the Rayleigh–Schrödinger procedure, yields divergent expressions for the second- and higher-order perturbative corrections. Harrell[1] has thus utilized an improved perturbation scheme to obtain corrections for order greater than unity for the eigenvalues. For instance, his expression for the SHO ground state energy and for $\alpha = 5/2$ reads,[1]

$$E_0\left(\frac{5}{2},\lambda\right) = 3 + \frac{2\Gamma(1/4)}{\Gamma(1/2)}\lambda + \frac{16}{\Gamma(1/2)}\lambda^2 \ln\lambda + O(\lambda^2). \qquad (2.2)$$

In the expression above, we have corrected for a misprint in the sign of the log term.

The presence in Eq. (2.2) of an explicit term between first and second orders in λ should be mentioned. This equation, valid for small values of the coupling parameter λ, is strikingly similar to the expansion for the ground state ener-

 0022-2488/90/010099-06$03.00

gy per particle for a boson system.[8] An alternative scheme will be employed in the present work to obtain corrections to $E_0(\alpha,\lambda)$. The methodology followed here is well known and is briefly presented in the next sections.

III. THE VARIATIONAL APPROACH

In the variational approach, the first step is to choose a complete set of basis functions. Although in principle one can employ an arbitrary basis of sufficiently smooth functions, in numerical practice the rate of convergence for singular problems depends a great deal on how the basis is chosen. The more clever one is in choosing a set of functions, the faster the convergence of the method. In this article, we take the basis set constructed with the normalized solutions of the Schrödinger equation for the linear harmonic oscillator with a DBC, i.e., the harmonic oscillator eigenfunctions $u(x)$ normalized in the interval $0\leqslant x<\infty$, with the prescription that $u(0)=0$. As is well known, these energy eigenfunctions are essentially the product of Hermite polynomials of odd degree with a Gaussian function. We write

$$\langle x|n\rangle \equiv u_n(x) = A_n e^{-x^2/2} H_{2n+1}(x), \quad n=0,1,2,..., \tag{3.1}$$

A_n being a normalizing factor. The functions $u_n(x)$ define a complete orthonormal set of solutions of H_0 in the interval $0\leqslant x\leqslant\infty$, with

$$A_n^{-2} = 4^n(2n+1)!\Gamma(1/2). \tag{3.2}$$

Now, let $\varphi(x)$ be an eigenfunction of the SHO Hamiltonian (1), and let us expand $\varphi(x)$ in terms of $u_n(x)$, namely,

$$\varphi(x) = \sum_{n=0}^{\infty} a_n u_n(x). \tag{3.3}$$

Next, we want to minimize the eigenenergies of (2.1), with respect to the variational parameters a_n, $n=0,1,...,N-1$, in the finite dimensional subspace spanned by the N functions $u_0,u_1,...,u_{N-1}$. This variational problem is equivalent to diagonalizing the Hamiltonian (2.1) in the chosen basis representation. By varying the dimension N, we get the trend of the method.

All we need is to evaluate the matrix elements of $H(\alpha,\lambda)$ in the basis (3.1). They can be separated into two contributions ($\beta\equiv-\alpha$),

$$H_{m+1,n+1}(\alpha,\lambda) \equiv \langle m|H_0|n\rangle + \lambda\,\langle m|x^\beta|n\rangle, \qquad m,n=0,1,...,N-1. \tag{3.4}$$

Since H_0 is diagonal in the chosen basis, we see that the first term on the right hand side of Eq. (3.4) is simply the expression for the energy eigenvalues of the harmonic oscillator with DBC, that is,

$$\langle m|H_0|n\rangle = (4n+3)\delta_{m,n}, \quad m,n=0,1,2,...,N-1. \tag{3.5}$$

Alternative procedures exist for deriving the matrix elements of the operator x^β appearing in (3.4). A direct way we develop here is to use the following representation for the odd-degree Hermite polynomials,

$$H_{2n+1}(x) = \frac{(-1)^n(2n+1)!\Gamma(1/2)}{n!} \times \sum_{m=0}^{n}(-1)^m\binom{n}{m}\frac{x^{2m+1}}{\Gamma(m+3/2)}. \tag{3.6}$$

The desired result, which involves finite double summations, reads

$$\langle m|x^\beta|n\rangle = T_m T_n \Gamma(3/2) \times \sum_{k=0}^{m}\sum_{l=0}^{n}(-1)^{k+l}\binom{m}{k}\binom{n}{l} \times \frac{\Gamma[k+l+(\beta+3)/2]}{\Gamma(k+3/2)\Gamma(l+3/2)}, \tag{3.7}$$

with

$$T_s \equiv (-1)^s\sqrt{(2s+1)!}/2^s s!. \tag{3.8}$$

Equation (3.7) is simple enough to be used in practical applications. However, it may be still reduced by carrying out explicitly one of the sums, e.g., the sum over l. The relevant terms for this sum are

$$S = \sum_{l=0}^{n}(-1)^l\binom{n}{l}\frac{\Gamma(l+u)}{\Gamma(l+3/2)}, \tag{3.9}$$

where $u=k+(\beta+3)/2$. Now we note that the binomial coefficient can be expressed as

$$\binom{n}{l} = (-1)^l(-n)_l/l!,$$

where $(-n)_l$ is the Pochhammer symbol.[9] Moreover, $\Gamma(l+z)=\Gamma(z)(z)_l$, so that (3.9) is the same as

$$S = \frac{\Gamma(u)}{\Gamma(3/2)}\sum_{l=0}^{n}\frac{1}{l!}\frac{(-n)_l(u)_l}{(3/2)_l}. \tag{3.10}$$

As n is integer, the sum in (3.10) corresponds to the hypergeometric function ${}_2F_1(-n,u;3/2;1)$ which has a very simple known expression.[9] Thus we can write (3.10) as

$$S = \frac{\Gamma[k+(\beta+3)/2]\Gamma(n-k-\beta/2)}{\Gamma(n+3/2)\Gamma(-k-\beta/2)}. \tag{3.11}$$

Finally, the matrix element of interest is given by

$$\langle m|x^\beta|n\rangle = T_m T_n \frac{\Gamma(3/2)}{\Gamma(n+3/2)} \times \sum_{k=0}^{m}(-1)^k\binom{m}{k} \times \frac{\Gamma[k+(\beta+3)/2]\Gamma(n-k-\beta/2)}{\Gamma(k+3/2)\Gamma(-k-\beta/2)}, \tag{3.12}$$

where the sum can be shown to be a polynomial of degree $m+n$ in β. A case of particular interest is the matrix element $\langle 0|x^\beta|n\rangle$ which, after some trivial algebra, reduces to

$$\langle 0|x^\beta|n\rangle = \frac{1}{\sqrt{(2n+1)!}}\frac{\Gamma[(\beta+3)/2]}{\Gamma(3/2)} \times\beta(\beta-2)\cdots(\beta-2n+2), \quad n=1,2,... . \tag{3.13}$$

Equation (3.12) is an exact closed form expression for the matrix elements of the operator x^β in the simple harmon-

ic oscillator representation supplemented by DBC. Explicit forms of the first few matrix elements of x^β are given in the Appendix. The same procedure leads to similar closed form expression for the matrix elements of x^β in the regular harmonic oscillator representation.

IV. LARGE COUPLING EXPANSION

The idea behind the large coupling expansion is to consider the potential

$$V(x) = x^2 + \lambda x^{-\alpha} \tag{4.1}$$

and Taylor expand it around its minimum. Let x_m and V_m be the values of x and $V(x)$, respectively, at the minimum. It is easy to see that

$$x_m = (\lambda\alpha/2)^{1/(\alpha+2)} \tag{4.2}$$

and

$$V_m = \lambda^{2/(\alpha+2)}[(\alpha/2)^{2/(\alpha+2)} + (2/\alpha)^{\alpha/(\alpha+2)}]. \tag{4.3}$$

Let $z = x - x_m$. The expansion of $V(x)$ around $z = 0$ can be written as

$$V(z) = V_m + (\alpha+2)z^2 + \sum_{k=3}^{\infty} (-1)^k \frac{2}{\alpha}(\alpha)_k \mu^{k-2} \frac{z^k}{k!}, \tag{4.4}$$

where $(\alpha)_k$ is the Pochhammer symbol[9] and

$$\mu = (2/\lambda\alpha)^{1/(\alpha+2)}. \tag{4.5}$$

Now let us rewrite the Schrödinger equation as

$$\left[\frac{-d^2}{dz^2} + V(z)\right] y = Ey. \tag{4.6}$$

The zero-order contribution to E is given by

$$E_0 = V_m = (1 + 2/\alpha)\mu^{-2} \tag{4.7}$$

and the next contribution comes from a harmonic oscillator in z, characterized by the energy parameter

$$\omega = \sqrt{a+2}. \tag{4.8}$$

Thus, in the large coupling expansion method we have

$$E(\mu) = (1 + 2/\alpha)\mu^{-2} + (\alpha+2)^{1/2} + \text{higher-order terms in } \mu. \tag{4.9}$$

For the particular case $\alpha = 5/2$, we have

$$E(\lambda) = \frac{9}{5}\left(\frac{5}{4}\right)^{4/9}\lambda^{4/9} + \left(\frac{9}{2}\right)^{1/2} + \cdots. \tag{4.10}$$

Numerical results for several values of λ are displayed in Table I.

The higher-order contributions can be obtained through a special perturbation expansion as shown in the following discussion. Consider

$$H = H_0 + \sum_{n=1}^{\infty} \mu^n H_n, \tag{4.11}$$

with,

$$H_0 = \frac{-d^2}{dz^2} + \omega^2 z^2 + V_m, \tag{4.12}$$

where, as before, $z = x - x_m$, and V_m and ω are given by (4.3) and (4.8), respectively. In (4.11), H_n is given by

$$H_{m-2} = (-1)^m (2/\alpha)(\alpha)_m (z^m/m!) \tag{4.13}$$

with μ, defined in (4.5), playing the role of a coupling constant.

Now, as usual, set

$$E(\mu) = E_0 + \mu E_1 + \mu^2 E_2 + \cdots \tag{4.14}$$

TABLE I. Ground state energy eigenvalues of the spiked harmonic oscillator for $\alpha = 5/2$. The superscripts in the energy E denote the dimension of the harmonic oscillator basis set (supplemented by Dirichlet boundary condition) employed for diagonalizing the matrix of the energy operator defined by Eq. (3.4). Also shown are the energies obtained from Eq. (4.10) and from fourth-order large coupling perturbative calculation Eq. (4.33). For comparison, the values obtained from Ref. 1 and from numerical integration of the corresponding Schrödinger equation, labeled "Exact," are also tabulated. All energies are displayed in arbitrary units.

λ	$E^{(1)}$	$E^{(2)}$	$E^{(10)}$	$E^{(20)}$	Large coupling expansion	4th order perturbation	Ref. 1	Exact
0.001	3.004 091	3.004 086	3.004 078	3.004 075	···	···	3.004 028	3.004 022
0.005	3.020 455	3.020 346	3.020 148	3.020 071	···	···	3.019 259	3.019 142
0.01	3.040 910	3.040 475	3.039 701	3.039 409	···	···	3.036 753	3.036 729
0.05	3.204 553	3.193 800	3.177 840	3.172 753	···	···	3.136 946	3.152 429
0.1	3.409 106	3.366 866	3.316 061	3.302 485	2.835 650	···	3.201 251	3.266 873
0.5	5.045 531	4.216 199	3.919 691	3.882 167	3.581 992	3.860 533	3.481 265	3.848 553
1	7.091 062	4.688 097	4.354 247	4.329 449	4.108 987	4.323 602	···	4.317 311
5	23.455 313	6.304 223	6.297 319	6.296 712	6.185 725	6.297 553	···	6.296 472
10	43.910 626	7.951 033	7.735 637	7.735 136	7.652 122	7.735 582	···	7.735 111
100	412.106 269	36.802 319	17.541 891	17.541 890	17.511 104	17.541 916	···	17.541 889
1000	4094.062 688	324.897 482	44.967 048	44.955 485	44.944 307	44.955 486	···	44.955 485

$$\psi(\mu) = \psi_0 + \mu\psi_1 + \mu^2\psi_2 + \cdots . \tag{4.15}$$

From

$$H\psi(\mu) = E_\mu \psi(\mu) \tag{4.16}$$

and collecting terms in powers of μ, we get

$$\sum_{m=0}^{n} H_m \psi_{n-m} = \sum_{m=0}^{n} E_m \Psi_{n-m}, \quad n = 0,1,2\ldots . \tag{4.17}$$

To solve this hierarchy of equations, let us project on the complete basis $|N\rangle$ generated by the harmonic oscillator eigenvalue problem

$$H_0|N\rangle = \left(\frac{-d^2}{dz^2} + \omega^2 z^2\right)|N\rangle$$

$$= e_N|N\rangle, \quad N = 0,1,2,\ldots . \tag{4.18}$$

The solution to (4.18) is well known and given by

$$|N\rangle = \left(\frac{1}{N!2^N}\right)^{1/2}\left(\frac{\omega}{\pi}\right)^{1/4} e^{-\omega z^2/2} H_N(\sqrt{\omega}z), \tag{4.19}$$

with

$$e_N = 2\omega(N + 1/2), \tag{4.20}$$

where H_N is a Hermite polynomial normalized in the interval $-\infty \leqslant z \leqslant \infty$. Moreover, Eqs. (4.17) are to be supplemented by the following orthogonality conditions

$$\langle\psi_0|\psi_0\rangle = 1 \text{ and } \langle\psi_0|\psi_n\rangle = 0, \quad \text{for } n > 0. \tag{4.21}$$

The energy expression is obtained by projecting (4.17) on $|\psi_0> \equiv |0>$. In doing so, we get

$$E_n = \sum_{m=0}^{n-1} \langle 0|H_{n-m}|\psi_m\rangle, \quad n = 1,2,\ldots . \tag{4.22}$$

From parity considerations, it can be seen that E_1 (as well as any E_{2n+1}) vanishes. So, the first high-order contribution comes from E_2 and is given by

$$E_2 = \langle 0|H_2|0\rangle + \langle 0|H_1|\psi_1\rangle \tag{4.23}$$

with H_1 and H_2 given by (4.13). The first contribution to E_2 is easily calculated. To evaluate the second one, we need to express $|\psi_1\rangle$ in terms of the basis functions $|N\rangle$. To do this, let us first project (4.17) on $|N\rangle$. From

$$\sum_{m=0}^{n} \langle N|H_m - E_m|\psi_{n-m}\rangle = 0, \tag{4.24}$$

it is straightforward to obtain that ($N \neq 0, n = 1,2,\ldots$)

$$\langle N|\psi_n\rangle = \frac{1}{2N\omega} \sum_{m=1}^{N} \sum_{p} \langle N|E_m - H_m|p\rangle\langle p|\psi_{n-m}\rangle, \tag{4.25}$$

where p runs over the complete set (4.19). For $n = 1$, we get

$$\langle N|\psi_1\rangle = \frac{(\alpha+1)(\alpha+2)}{6N\omega}\langle N|z^3|0\rangle, \tag{4.26}$$

which tells us that the only possibilities for N are $N = 1$ and $N = 3$, as in (4.26) we have harmonic oscillator matrix elements. These can be easily obtained through direct calculations. The results are

$$\langle 0|z^3|1\rangle = (9/8\omega^3)^{1/2}, \tag{4.27}$$

$$\langle 0|z^3|3\rangle = (6/8\omega^3)^{1/2}. \tag{4.28}$$

Thus, from (4.26) and the results above, we obtain

$$|\psi_1\rangle = \frac{(\alpha+1)(\alpha+2)}{18\omega(8\omega^3)^{1/2}}[9|1\rangle + \sqrt{6}|3\rangle]. \tag{4.29}$$

Going back to (4.23), we see that to get the contribution E_2 we need the following harmonic oscillator matrix elements

$$\langle 0|H_2|0\rangle = (\alpha+1)(\alpha+2)(\alpha+3)/16\omega^2, \tag{4.30a}$$

$$\langle 0|H_1|1\rangle = -(\alpha+1)(\alpha+2)/(8\omega^3)^{1/2}, \tag{4.30b}$$

$$\langle 0|H_1|3\rangle = -\sqrt{6}(\alpha+1)(\alpha+2)/3(8\omega^3)^{1/2}. \tag{4.30c}$$

Thus, using (4.8) we get the total second-order contribution

$$\mu^2 E_2 = (2/\lambda\alpha)^{2/(\alpha+2)}(\alpha+1)(8-\alpha)/72. \tag{4.31}$$

The calculation of the next correction is quite involved, because of the long emerging expression to be manipulated. We have used a (deceptively simple) algorithm using the symbolic manipulation package SMP[10] to get the next E_4 correction. We ran out of memory when trying to evaluate E_6. The contribution is given by

$$\mu^4 E_4 = \frac{(\alpha+1)(\alpha-2)(\alpha^2-\alpha-74)}{1728(2+\alpha)^{1/2}}\left(\frac{2}{\lambda\alpha}\right)^{4/(\alpha+2)}, \tag{4.32}$$

and putting everything together we finally obtain, for $\alpha = 5/2$,

$$E(\lambda) = \frac{9}{5}\left(\frac{5\lambda}{4}\right)^{4/9} + \left(\frac{9}{2}\right)^{1/2} + \frac{77}{288}\left(\frac{4}{5\lambda}\right)^{4/9} - \frac{1967}{27648}\left(\frac{2}{9}\right)^{1/2}\left(\frac{4}{5\lambda}\right)^{8/9} + \cdots . \tag{4.33}$$

Numerical results for several values of λ are displayed in Table I.

To go further we must abandon algebraic methods and use some kind of seminumerical algorithm. The hierarchy of equations (4.22) and (4.25) can be quite easily coded in FORTRAN, in order to obtain the value of the amplitudes $\langle N|\psi_n\rangle$ for a given value of α and subsequently the corresponding energy corrections E_{n+1} in a chained way. Note that in order to obtain E_{n+1} we require all wavefunctions $\psi_0,\psi_1,\ldots,\psi_n$, but to obtain $\langle N|\psi_{n+1}\rangle$ we need the value of E_{n+1}. The calculation does not involve any approximation, aside from rounding errors. The values of the first twenty coefficients of the expansion corresponding to $\alpha = 5/2$ are displayed in Table II. The notation is such that the energy $E(\lambda)$ is given by

TABLE II. Coefficients for the large coupling perturbation expansion equation (4.34), for $\alpha = 5/2$.

n	E_n
2	0.267 361 111
4	−0.033 537 785
6	−0.017 395 489
8	0.011 679 410
10	0.001 109 577
12	−0.006 825 514
14	0.002 542 598
16	0.008 178 874
18	−0.011 093 142
20	−0.014 427 769

$$E(\lambda)=\frac{9}{5}\left(\frac{5\lambda}{4}\right)^{4/9}+\left(\frac{9}{2}\right)^{1/2}+\sum_{\substack{n=2\\(\text{even})}} E_n\left(\frac{4}{5\lambda}\right)^{2n/9}. \tag{4.34}$$

Here, E_n behaves quite erratically with n, and probably this is not a well behaved series expansion. When λ is large enough, say larger than 2, one can expect to get very precise results from the expansion. However, $\lambda = 4/5$ is certainly outside the convergence radius, if such radius exists.

One can try to understand this strange behavior. The large coupling perturbation method expands both interaction and wavefunction around the classical minimum of the potential, extending the new coordinate $z = x - x_m$ to the full real axis. Certainly the region $(-\infty,0]$ is spurious because of the mere statement of the problem. When λ is large, the minimum is placed at a large value of x, so that the harmonic oscillator wavefunctions centered at x_m will not penetrate too much into the forbidden region. Obviously this is not the case for small λ. Probably the proper way of controlling this unwanted characteristic is to change variables to a new coordinate extending along the full real axis and carry out afterwards the large coupling perturbative expansion.

V. NUMERICAL RESULTS AND CONCLUSION

Since it is of some interest to have an idea of the relative size of the ground state spiked harmonic oscillator energies as the potential parameter increases, we have numerically integrated the Schrödinger equation. The exponent α was fixed at 5/2 to compare with Harrell's result Eq. (2.2). The energies so obtained are displayed in Table I under the entry "Exact."

From the computational point of view it resulted in a quite complex problem. First of all, because of the singular character of the potential near the origin, we could not use a small error integration formula, like, e.g., Numerov's method.[11] Instead, we had to use the lowest order approximation to the second derivative

$$D^2 = \delta^2/h^2,$$

where δ^2 represents the second-order centered differences, and h is the mesh spacing. Nevertheless, the results could be improved and tested by means of the Richardson extrapolation algorithm. To give an idea of the numerical difficulties, let us mention that to obtain the energy for $\lambda = 0.001$ with six decimal places, we had to use a mesh with 80 000 points.

From Eqs. (3.4), (3.5), and (3.7) and the results given in the Appendix, it can be readily seen that the first variational approximation (subspace of dimension 1) to the ground state eigenenergy of the SHO is

$$E^{(1)} = H_{11} = \langle 0|H|0\rangle = 3 + \lambda\Gamma\left[\frac{(\beta+3)}{2}\right]\Big/\Gamma(3/2), \tag{5.1}$$

which coincides with the $0(\lambda)$ correction of Harrell, Eq. (2.2).

When $N = 2$ the diagonalization can also be easily performed analytically, via the secular equation approach, and we have

$$E^{(2)} = \frac{1}{2}\left[10 + \frac{53}{24}\lambda\gamma \pm \left(16 + \frac{5}{3}\lambda\gamma + \frac{2425}{576}\lambda^2\gamma^2\right)^{1/2}\right]. \tag{5.2}$$

From this expression, we notice that when the spikelike perturbation vanishes ($\lambda = 0$), the two eigenvalues become 3 and 7, as expected.

For higher values of N we have to resort to numerical diagonalization procedures and we employed the known Jacobi method. Table I shows the convergence of the results for the ground state energy of the SHO, for selected values of λ, as the dimension of the basis set is increased.

To analyze the results of the variational calculation it is convenient to distinguish among three cases, corresponding to a) large values of the coupling constant ($\lambda \geqslant 10$); b) small values of λ (< 0.1); and c) medium values ($\lambda \simeq 1$).

For large λ the calculations involving few basis states are definitely poor, but very good results are obtained when the basis space is enlarged (dimension 10 or more). In this region of λ, the variational method converges quite rapidly to the exact results.

In the case of very small λ, one can get the wrong impression that the variational method behaves properly. Just to clarify the previous statement let us consider the number 3.004 075 which appears in the column labeled $E^{(20)}$ corresponding to $\lambda = 0.001$, in Table I. That number is actually $3 + 4.091\lambda - 16\lambda^2$, where the first term is the unperturbed energy, the second is the first-order perturbation correction, Eq. (2.2), and the third is the contribution of all remaining nineteen states. The coefficient of λ^2 varies very slowly with the number of states of the basis: adding more figures, it changes from -14.97 for $N = 10$ to -16.16 for $N = 20$. Certainly, the energy eigenvalue decreases when increasing the number of basis states, but very slowly. This behavior is a direct consequence of the abnormal properties of the perturbation, as pointed out by Harrell in Ref. 1.

It is interesting to study this point more closely. Assume a very small λ and solve the matrix eigenvalue problem by expanding the determinant in powers of λ up to and including λ^2. As it is well known, we end up with the perturbation-like formula

$$E = E_0 + \langle 0|x^\beta|0\rangle\lambda - \lambda^2\sum_{n\neq 0}^{N}\frac{|\langle 0|x^\beta|n\rangle|^2}{E_n - E_0}, \tag{5.3}$$

with the difference that the sum in the λ^2 correction is limited to the chosen number of basis states. Analyzing this sum, we will understand the slow variation of the matrix eigenvalue problem solutions.

Using (3.13) for $\beta = -5/2$ there results

$$\sum_{n\neq 0}\frac{|\langle 0|x^\beta|n\rangle|^2}{E_n - E_0} = \gamma\sum_{n\neq 0}\frac{1}{4n}\frac{(2+\frac{1}{2})^2}{2\cdot 3}\frac{(4+\frac{1}{2})^2}{4\cdot 5}\cdots\frac{(2n+\frac{1}{2})^2}{2n(2n+1)}, \tag{5.4}$$

where $\gamma = \Gamma[(\beta+3)/2]/\Gamma(3/2)$ and the factors $(2n+1)!$ and $\beta(\beta-2)\cdots(\beta-2n+2)$ have been ordered in a special form. The ratio of the $(n+1)$th and nth terms of the sum in (5.4) is

$$\frac{\langle 0|x^\beta|n+1\rangle^2/(E_{n+1}-E_0)}{\langle 0|x^\beta|n\rangle^2/(E_n-E_0)} = \frac{n}{n+1}\frac{(2n+2+\frac{1}{2})^2}{2(n+1)(2n+3)} \underset{n\to\infty}{\to} 1,$$

so that successive contributions to the coefficient of λ^2 are of the same size. Moreover, each of the grouped terms in (5.4) is bigger than 1, i.e.,

$$(2k+1/2)/2k(2k+1) > 1$$

and each term of (5.4), apart from a global constant, is bigger than the corresponding term of the harmonic series $\Sigma(1/n)$ which is known to be divergent.

Thus, the standard perturbation theory, Eq. (5.3) for $N\to\infty$, makes no sense, and our variational method will give a ground state energy which converges very slowly when the number of basis states increases. Note finally that Eq. (5.3) is no longer an upper bound formula, and it is only valid for N sufficiently small.

The variational results are somewhat poor for small λ and quite good for large λ. At small λ the non-power series expansion of Harrell[1] is appropriate, as well as at large λ our large coupling perturbation expansion gives a proper description of the ground state energy.

Finally we have the region of intermediate $\lambda(\simeq 1)$. In this region, the best method is the variational one. By using sufficiently large basis one could obtain the correct value of the energy, but no definite statements about the speed of the convergence can be drawn from our results.

In conclusion, it seems that the appropriate method to deal with this class of potentials is to use a nonpower series expansion for small coupling constant, and a large coupling perturbative expansion for large λ. Both expansions should be an appropriate extension of the presently known forms. Moreover, it would also be of interest to find a connection of both expansions for intermediate values of the coupling constant.

ACKNOWLEDGMENTS

One of the authors (G. A. E.) wants to dedicate his contribution to this work to Professor Kerson Huang with admiration and affection on the occasion of his sixtieth birthday. G. A. E. also wishes to express his sincere thanks to R. G. Barrera, B. Bhuiyan, E. A. Castro, F. M. Fernandez, J. Giraldo, Wilfrido Solano, and T. Kato for their kind encouragement in so many ways. The authors acknowledge Professor M. de Llano for suggesting the theme, for useful discussions, and for hospitality at the Physics Department of the North Dakota State University, Fargo, ND, where the authors met and started the present paper.

One of us (R. G.) acknowledges the financial support of the Comision Interministerial de Ciencia y Tecnología (Spain) under contract number 969/87.

APPENDIX

A program has been written in BASIC for symbolically handling the evaluation of the matrix elements of the operator x^β. A listing of the program can be obtained from the authors on request. The first ten matrix elements of x^β run as follows; where

$$x^\beta_{mn} \equiv \langle m-1|x^\beta|n-1\rangle,$$

$$x^\beta_{11} \equiv \gamma = \Gamma\left(\frac{\beta+3}{2}\right)\Big/\Gamma(3/2),$$

$$x^\beta_{12} = \beta\gamma/\sqrt{3!},$$

$$x^\beta_{13} = \beta(\beta-2)\gamma/\sqrt{5!},$$

$$x^\beta_{14} = \beta(\beta-2)(\beta-4)\gamma/\sqrt{7!},$$

$$x^\beta_{22} = (\beta^2+2\beta+6)\gamma/3!,$$

$$x^\beta_{23} = \beta(\beta^2+2\beta+12)\gamma/\sqrt{3!5!},$$

$$x^\beta_{24} = \beta(\beta^3+14\beta-36)\gamma/\sqrt{3!7!},$$

$$x^\beta_{33} = (\beta^4+4\beta^3+36\beta^2+64\beta+120)\gamma/5!,$$

$$x^\beta_{34} = \beta(\beta^4+4\beta^3+56\beta^2+104\beta+360)\gamma/\sqrt{5!7!},$$

$$x^\beta_{44} = (\beta^6+6\beta^5+106\beta^4+454\beta^3 + 1660\beta^2+3968\beta+5040)\gamma/7!.$$

[1] E. M. Harrell, Ann. Phys. (NY) **105**, 379 (1977) and references therein.
[2] T. Kato, "Perturbation Theory for Linear Operators," in *Die Grundlehren der Matematischen Wissenschaften* (Springer, New York, 1966), Vol. 132.
[3] F. Rellich, *Perturbation Theory of Eigenvalue Problems* (Gordon and Breach, New York, 1969).
[4] L. C. Detwiler and J. R. Klauder, Phys. Rev. D **11**, 1436 (1975).
[5] M. de Llano, Rev. Mex. Fis. **27**, 243 (1981).
[6] J. R. Klauder, Science **199**, 735 (1978).
[7] R. M. Spector, J. Math. Phys. **8**, 2357 (1967).
[8] T. T. Wu, Phys. Rev. **115**, 1390 (1959). See also K. Huang, *Statistical Mechanics* (Wiley, New York, 1987), 2nd ed., p. 332.
[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[10] SMP is a licensed symbolic manipulation language of INFERENCE Company.
[11] E. Buendía and R. Guardiola, J. Comput. Phys. **60**, 561 (1985).

Some remarks on the Feynman–Kac formula

Zdzisław Brzeźniak
Institute of Mathematics, Jagellonian University, Reymonta 4, 30-059 Kraków, Poland

(Received 28 March 1989; accepted for publication 21 June 1989)

A simple necessary condition for the existence of the representation of solutions of partial differential equations is found. This condition is applied to obtain the known results on the Schrödinger equation and the Dirac system in a unified way. Applications for further equations are also possible (Weyl's equations are discussed).

I. INTRODUCTION

The aim of this paper is to find a necessary condition for the existence of the Feynman–Kac representation for the solutions of evolution equations or systems of equations in terms of their fundamental solutions; the condition does not seem to be far from sufficient (see Remark 1). This global condition can be applied to many systems of equations, including Schrödinger and Dirac. Thus we are able to obtain in a simple way the result of Zastawniak.[1] Moreover, that condition can be expressed in a very simple way: in order that the Feynman–Kac measure does exist, the fundamental solution $K(t)$, being *a priori* a distribution, should be a finite (matrix valued) Borel measure.

In recent years a growing interest in finding representations of solutions to some evolutionary systems by the so-called Feynman–Kac formula in terms of some vector valued measure, which we shall call the Feynman–Kac measure, has been observed. Such a representation for a two space-time dimensional Dirac system was found by several authors in different ways (see, for example, Ichinose,[2] Zastawniak,[3,4] Ichinose and Tamura,[5] and Blanchard *et al.*[6]). However, until recently such a representation for other systems (including the four space-time dimensional Dirac system) had been unknown. It was observed by Ichinose[2] that the existence of the Feynman–Kac measure for the two space-time dimensional Dirac system does not seem to generalize easily to the four space-time dimensional Dirac system for the lack of an L^∞ estimate. It was Zastawniak[1] who first proved that such an L^∞ estimate does not exist and therefore the Feynman–Kac measure does not exist either.

For a long time it has been well known that solutions to parabolic equations can be represented in terms of the Feynman–Kac formula (see Kac[7] and Reed and Simon[8]). It is also known that one cannot represent solutions to the Schrödinger equation in terms of that formula; this fact was first observed by Cameron.[8,9] As was observed by Zastawniak,[1] the corresponding proofs of nonexistence for the Schrödinger equation and the Dirac system are completely different, although both rely on showing the nonexistence of L^∞ estimates for solutions to the corresponding Cauchy problems.

Our approach unifies in a certain sense the Schrödinger equation and the Dirac system. It also explains the difficulties with constructing the Feynman–Kac measure, and it may prove useful in the future.

II. THE MAIN RESULT

Let us consider on $\mathbf{R}^n$ a system of m differential operators with constant coefficients acting on $\mathbf{C}^m$ valued functions. We denote that system by $A(\partial_x)$. Let us consider the following Cauchy problem:

$$\frac{\partial u(t,x)}{\partial t} = A(\partial_x)u(t,x), \quad t>0, \quad x\in\mathbf{R}^n, \qquad u(0,x) = u_0(x), \quad x\in\mathbf{R}^n. \tag{2.1}$$

We assume that there exists a fundamental solution of (2.1), i.e., a $\mathscr{D}'(\mathbf{R}^n,\mathbf{C}^{m\times m})$ valued function $\{K(t)\}_{t>0}$ such that for any $u_0\in\mathscr{D}(\mathbf{R}^n,\mathbf{C}^m)$ the unique solution u to problem (2.1) is given by

$$u(t) = K(t)*u_0. \tag{2.2}$$

Definition 1: If, for any $t>0$, there exists a measure space (Ω,Σ,μ), where Ω consists of $\mathbf{R}^n$ valued paths $\gamma(s)$, $0\le s\le t$, $\gamma(0)=0$, while μ is a $\mathbf{C}^{m\times m}$ valued measure, such that for any $u_0\in\mathscr{D}(\mathbf{R}^n,\mathbf{C}^m)$ the unique solution u to (2.1) can be represented by the Feynman–Kac formula

$$u(t,x) = \int_\Omega d\mu(\gamma)u_0(\gamma(t)+x), \tag{2.3}$$

then we say that problem (2.1) possesses a Feynman–Kac measure.

Proposition 1: Assume that problem (2.1) possesses a Feynman–Kac measure. Then for any $t>0$ there exists $C>0$ such that, for any $u_0\in\mathscr{D}(\mathbf{R}^n,\mathbf{C}^m)$,

$$|u(t,\cdot)|_\infty \le C|u_0|_\infty. \tag{2.4}$$

Proposition 1 can be easily deduced from the fact that the variation $\|\mu\|$ of the measure μ is finite (see, also, Ref. 10).

Now we are ready to state our main result.

Theorem 1: Assume that the Cauchy problem (2.1) possesses a fundamental solution $\{K(t)\}_{t>0}$, $K(t)\in\mathscr{D}'(\mathbf{R}^n,\mathbf{C}^{m\times m})$, for $t>0$. Assume that problem (2.1) possesses a Feynman–Kac measure. Then $K(t)$ is a finite $\mathbf{C}^{m\times m}$ valued Borel measure on $\mathbf{R}^n$.

Theorem 1 follows immediately from Proposition 1 the following propositions.

Proposition 2: If a distribution $T\in\mathscr{D}'(\mathbf{R}^n,\mathbf{C}^{m\times m})$ satisfies $|T*\phi|_\infty \le C|\phi|_\infty$ for some constant $C>0$ and any $\phi\in\mathscr{D}(\mathbf{R}^n,\mathbf{C}^m)$, then T is a finite Borel measure.

 0022-2488/90/010105-03$03.00

Proof of Proposition 2: For $\phi \in \mathcal{D}(\mathbf{R}^n, \mathbf{C}^m)$, $T*\phi$ is a $\mathcal{C}^\infty$ function defined by $(T*\phi)(x) = T(\tau_x, \check{\phi})$, where $(\tau_x \psi)(y) = \psi(y-x)$, $\check{\psi}(y) = \psi(-y)$. Hence

$$|T(\phi)|_\infty = |T(\tau_0 \check{\phi})|_\infty = |(T*\check{\phi})(0)|_\infty \leq C|\phi|_\infty . \quad (2.5)$$

We conclude the proof by applying the Riesz Representation Theorem.

Let us also observe that Proposition 2 can be also derived from Proposition 5.20 in Ref. 11.

Remark 1: In the future the author will investigate the problem of sufficiency of the condition in Theorem 1, i.e., the existence of the Feynman–Kac measure under the assumption that the fundamental solution is a finite (matrix valued) measure, of course, with some additional technical assumptions. This should be a straightforward corollary from the Kolmogorov–Jessen Extension Theorem.

Also some applications to hyperbolic systems will be considered. In that case we may consider problem (2.1) with the operator $A(\partial_x)$ perturbed by some "nice" potential function $V(x)$, i.e.,

$$\begin{aligned} &\frac{\partial u(t,x)}{\partial t} = A(\partial_x)u(t,x) + V(x)u(t,x), \\ &\qquad t>0, \quad x\in\mathbf{R}^n, \\ &u(0,x) = u_0(x), \quad x\in\mathbf{R}^n. \end{aligned} \quad (2.6)$$

Then we are interested in representing a solution u to (2.6) in the following form:

$$u(t,x) = \int_\Omega d\mu(\gamma) u_0(\gamma(t)+x) \times \exp\left\{\int_0^t V(\gamma(s)+x)ds\right\}. \quad (2.7)$$

The precise meaning of the above expression will be given.

The problem of continuity of paths should be discussed also. For example, it is known[2–4] that almost all paths of the underlying Feynman–Kac space for the two space-time dimensional Dirac system are continous (in fact, they are zigzag functions).

Remark 2: Once one knows that the fundamental solution $K(t)$ for the system (2.1) is a finite (matrix valued) Borel measure, then, under some additional assumptions, one can infer that the same is true for the following system:

$$\begin{aligned} &\frac{\partial u(t,x)}{\partial t} = A(\partial_x)u(t,x) + B\,u(t,x), \\ &\qquad t>0, \quad x\in\mathbf{R}^n, \\ &\qquad u(0,x) = u_0(x), \quad x\in\mathbf{R}^n, \end{aligned} \quad (2.8)$$

where B is a linear differential operator satisfying a certain growth condition with respect to A (for example, one can take as B, $u \to miKu$, where $m\in\mathbf{R}$, $K\in\mathbf{C}^{m\times m}$ are fixed, $i^2 = -1$). Indeed, if problem (2.1) generates a $\mathcal{C}_0$ semigroup on $X := C_0(\mathbf{R}^n, \mathbf{C}^m)$, then under the assumptions concerning the operators A and B (these assumptions are satisfied when B is of the form miK as above) we get, by standard theory on $\mathcal{C}_0$ semigroups (see Ref. 12, Chap. III), that problem (2.8) generates a $\mathcal{C}_0$ semigroup on X. Thus by Theorem 1 we conclude that the fundamental solution for (2.8) is a finite (matrix valued) Borel measure.

III. SOME EXAMPLES AND APPLICATIONS

Example 1: Consider the Cauchy problem for the following first-order equation:

$$\begin{aligned} &\frac{\partial u(t,x)}{\partial t} + \frac{\partial u(t,x)}{\partial x} = 0, \quad t>0, \quad x\in\mathbf{R}, \\ &u(0,x) = u_0(x), \quad x\in\mathbf{R}. \end{aligned} \quad (3.1)$$

As the fundamental solution is given by $K(t) = \delta_t$, it is not difficult to show that the Feynman–Kac measure does exist.

Example 2: For the two space-time dimensional Dirac system

$$\begin{aligned} &\frac{\partial u(t,x)}{\partial t} + A\frac{\partial u(t,x)}{\partial x} = miKu(t,x), \quad t>0, \quad x\in\mathbf{R}, \\ &u(0,x) = u_0(x), \quad x\in\mathbf{R}, \end{aligned} \quad (3.2)$$

where the 2×2 matrices are given by

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad K = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

the fundamental solution $E(t)$ is given by

$$E(t) = \begin{bmatrix} \dfrac{\partial D(t)}{\partial t} - \dfrac{\partial D(t)}{\partial x}, & imD(t) \\ imD(t), & \dfrac{\partial D(t)}{\partial t} + \dfrac{\partial D(t)}{\partial x} \end{bmatrix}.$$

Here $D(t)$ is the fundamental solution to the Klein–Gordon equation. It is given by the following formula:

$$D(t,x) = \tfrac{1}{2} J_0(m\sqrt{t^2 - x^2})1_{(-t,t)}(x).$$

Here $J_0(z)$ is the Bessel function of zeroth order:

$$J_n(z) = (\pi^{1/2}\Gamma(n+\tfrac{1}{2}))^{-1}(z/2)^n \times \int_{-1}^{1} e^{i\xi z}(1-\xi^2)^{n-1/2}\,d\xi, \quad z\in\mathbf{R}, \quad n\in\mathbf{N}.$$

By direct calculations we find that both $\partial D(t)/\partial t$ and $\partial D(t)/\partial x$ are finite Borel measures, and so is $E(t)$. Therefore (see Remark 1) we get Ichinose's result on the existence of Feynman–Kac measure (cf. Ref. 2).

Remark 3: Using the ideas described in Remark 2 one can also derive the existence of the Feynman–Kac measure for the system (3.2).

Example 3: The fundamental solution $K(t)$ for the Schrödinger equation in $\mathbf{R}^n \times [0,\infty)$ has a density

$$K(t,x) = (4\pi it)^{-n/2}\exp(i|x|^2/4t).$$

Thus, although $K(t)$ is a locally bounded Borel measure, it is not a finite Borel measure and therefore the Feynman–Kac measure for the Schrödinger equation does not exist.

Example 4: Consider the four space-time dimensional Dirac system (in Weyl representation)

$$\begin{aligned} &\frac{\partial u(t,x)}{\partial t} - \sum_{j=1}^{a} A_j \frac{\partial u(t,x)}{\partial x_j} - miKu(t,x) = 0, \\ &t>0, \quad x\in\mathbf{R}^3, \\ &u(0,x) = u_c(x), \quad x\in\mathbf{R}^3, \end{aligned} \quad (3.3)$$

where the 4×4 matrices A_j, K are given by

$$K=\begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix},\quad A_j=\begin{bmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{bmatrix},$$

$$\sigma_1=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\quad \sigma_2=i\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\quad \sigma_3=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and $m\geqslant 0$ denotes the mass. In order to write down the fundamental solution of (3.3) let us introduce the distribution $D(t)$ defined by the following formula:

$$D(t)(\phi)=\frac{t}{4\pi}\int_{|y|=1}\phi(ty)d\sigma(y) - \frac{mt^2}{4\pi}\int_{|y|<1}\phi(ty)\frac{J_1(mt\sqrt{1-|y|^2})}{\sqrt{1-|y|^2}}\,dy,$$

where $\phi\in\mathscr{D}(\mathbb{R}^3)$, $d\sigma(y)$ denotes the Lebesque measure on the unit sphere in $\mathbb{R}^3$, and J_1 is the Bessel function of the first order. Then the fundamental solution $E(t)$ for the Dirac system (2.3) is as follows:

$$E(t)=\frac{\partial D(t)}{\partial t}I+miK\,D(t)+\sum_{j=1}^{3}A_j^*\frac{\partial D(t)}{\partial x_j}$$

$$=\frac{\partial D(t)}{\partial t}+miK\,D(t)+\sum_{j=1}^{3}(-1)^jA_j\frac{\partial D(t)}{\partial x_j}.$$

Observe that a matrix valued distribution $E(t)$ is a finite measure iff all the entries of the matrix $E(t)$ are finite complex Borel measures. Since the entries on the main diagonal of the matrices $A_j\,\partial D(t)/\partial x_j$ vanish (due to the form of the matrices A_j) and $miKD(t)$ is obviously a finite measure, in order to show that $E(t)$ is not a measure it is enough to show that neither is $\partial D(t)/\partial t$. For $\phi\in\mathscr{D}(\mathbb{R}^3,\mathbb{C})$ we have the explicit formula

$$\frac{\partial D(t)}{\partial t}(\phi)=\frac{1}{4\pi}\int_{|y|=1}\phi(ty)d\sigma(y)-\frac{mt^2}{4\pi}\int_{|y|<1}\langle\nabla\phi(ty),y\rangle\frac{J_1(mt\sqrt{1-|y|^2})}{\sqrt{1-|y|^2}}$$

$$-\frac{mt}{4\pi}\int_{|y|<1}\phi(ty)\left\{2\frac{J_1(mt\sqrt{1-|y|^2})}{\sqrt{1-|y|^2}}-tmJ_1'(mt\sqrt{1-|y|^2})\right\}dy+\frac{t}{4\pi}\int_{|y|=1}\langle\nabla\phi(ty),y\rangle d\sigma(y).$$

Since the last term in the above expression, contrary to the others, is not a measure, the proof that $\partial D(t)/\partial t$ is not a measure is completed. Therefore, in view of Theorem 1, the Feynman–Kac measure for the four space-time dimensional Dirac system does not exist (see, also, Zastawniak).[1]

Example 5: Now we consider the so-called Weyl equation, i.e.,

$$\frac{\partial u(t,x)}{\partial t}+\sigma\cdot\nabla\,u(t,x)=0,\quad t>0,\quad x\in\mathbb{R}^n,$$

$$u(0,x)=u_0(x),\quad x\in\mathbb{R}^n, \tag{3.4}$$

where σ is a 2×2 matrix with complex entries satisfying $\sigma^2=I$, ∇ stands for the gradient operator, and the unknown function u takes values in $\mathbb{C}^2$.

The form of the fundamental solution $E(t)$ for (3.4) depends on n, the dimension of the space $\mathbb{R}^n$. Indeed,

$$E(t)=\frac{\partial D(t)}{\partial t}I-(\sigma\cdot\nabla)D(t),$$

where

$$D(t,x)=\tfrac{1}{2}1_{(-t,t)}(x),\quad \text{for } n=1,$$

$$D(t,x)=\frac{1}{2\pi}\frac{1}{\sqrt{t^2-|x|^2}}1_{(-t,t)}(|x|),\quad \text{for } n=2,$$

$$D(t)=\frac{1}{4\pi t}\text{ times the Lebesque measure supported on } S(0,t),\text{ the sphere of radius } t,\text{ for } n=3.$$

Once again, by direct computation, we infer that $E(t)$ is a finite Borel measure for $n=1,2$, while $E(t)$ is not a measure (thus not a finite measure) for $n=3$. Hence the Feynman–Kac measure should exist for $n=1,2$ and does not exist for $n=3$.

[1] T. Zastawniak, "The nonexistence of the path-space measure for the Dirac equation in four space-time dimensions," J. Math. Phys. **30**, 1354 (1989).
[2] T. Ichinose, "Path integrals for system of the first order," Duke Math. J. **51**, 1 (1984).
[3] T. Zastawniak, Ph.D. thesis, Jagellonian University, Kraków, 1987.
[4] T. Zastawniak, "Path integrals for the telegrapher's and Dirac equations; the analytic family of measures and underlying Poisson process," Bull. Pol. Acad. Sci. Math. **37** (1989).
[5] T. Ichinose and H. Tamura, J. Math. Phys. **25**, 1810 (1984); Suppl. Prog. Theor. Phys. **92**, 144 (1987); J. Math. Phys. **29**, 103 (1988).
[6] Ph. Blanchard, Ph. Combe, M. Siruge, and M. Siruge-Collin, "Probabilistic solution of the Dirac equation," "Path integral representation in then presence of an electromagnetic field," BiBoS, Univ. Bielefeld preprints Nos. 44 and 66, 1985.
[7] M. Kac, "On some connections between probability theory and differential and integral equations." in *Proceedings of the Second Berkeley Symposium* (Univ. California Press, Berkeley, 1951). pp. 189–215.
[8] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. 2 (Academic, New York, 1975).
[9] R. Cameron, "The Ilstow and Feynman integrals," J. Anal. Math. **10**, 287 (1962/63).
[10] W. Rudin, *Real and Complex Analysis* (McGraw–Hill, New York, 1975), 2nd ed.
[11] G. B. Folland, *Lectures on Partial Differential Equations*, Tata Institute of Fundamental Research, Bombay, 1983 (Springer, Berlin, 1983).
[12] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations* (Springer, Berlin, 1983).

Singular anharmonicities and the analytic continued fractions. II. The potentials $V(r)=ar^2+br^{-4}+cr^{-6}$

M. Znojil
Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež, Czechoslovakia

(Received 15 November 1988; accepted for publication 23 August 1989)

The $c=0$ results of Paper I [J. Math. Phys. 30, 23 (1989)] are extended. In spite of the presence of an additional coupling constant, the Laurent series solutions of the Schrödinger equation that are obtained remain similar to Mathieu functions. Indeed, the recurrences for coefficients preserve their three-term character, their analytic continued fraction solutions still converge, etc. The formulas become even slightly simpler for $c\neq 0$ due to a certain symmetry of the equations to be solved. An acceleration of convergence is better understood and a few numerical illustrations of efficiency are also delivered.

I. INTRODUCTION

For the class of potentials

$$V(r)=ar^2+br^{-4}+cr^{-6}, \quad a>0, \quad c>0, \tag{1.1}$$

the radial Schrödinger bound-state problem

$$\left[-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+V(r)\right]\psi(r)=E\psi(r), \qquad l=0,1,..., \quad \psi\in L_2(0,\infty), \tag{1.2}$$

may easily be solved numerically, by some standard (say, Runge–Kutta[1]) method. Indeed, in both the threshold and asymptotic domains, the required logarithmic derivatives have a simple WKB form:

$$\psi'(r_0)/\psi(r_0)=2\lambda r_0^{-3}+O(r_0^{-1}), \quad \lambda=\tfrac{1}{2}\sqrt{c}>0, \quad r_0\ll 1, \tag{1.3}$$

and

$$\psi'(r_\infty)/\psi(r_\infty)=-2\mu r_\infty+O(r_\infty^{-1}) \quad \mu=\tfrac{1}{2}\sqrt{a}>0, \quad r_\infty\gg 1, \tag{1.4}$$

respectively. Nevertheless, a few extremely interesting features of the interaction (1.1) emerge after its deeper nonnumerical analysis.

(A) The finite polynomial interpolation between $r=0$ (1.3) and $r=\infty$ (1.4), namely,

$$\psi(r)=r^{\varkappa}\exp(-\mu r^2-\lambda r^{-2})\sum_{m=-M}^{N}h_m r^{2m}, \tag{1.5}$$

may coincide with the exact bound states for certain values of the couplings.[2]

(B) In the infinite "polynomial-approximation" $N\to\infty$ and $M\to\infty$ limit, the Laurent-series *Ansatz* (1.5) remains useful: It converts our second-order differential equation (1.2) into its second-order difference-equation counterpart:

$$\begin{aligned}
&A_{n+1}h_n+B_{n+1}h_{n+1}+C_{n+1}h_{n+2}=0,\\
&A_{n+1}=2\mu(4n+2\varkappa+1)-E,\\
&B_{n+1}=-(2n+\varkappa+1)(2n+\varkappa+2)+l(l+1)+8\mu\lambda,\\
&C_{n+1}=-2\lambda(4n+2\varkappa+5)+b,\\
&n=...,-1,0,1,... .
\end{aligned} \tag{1.6}$$

This motivated the present paper: In the spirit of its preceding part I (Ref. 3) [devoted to the $c=0$ special case of (1.1) where only property (B) holds], we shall show how our transition to (1.6) simplifies the original problem (1.2). In Sec. II, we relate the "redundant" $n\to\infty$ boundary conditions to the "Floquet" parameter[4] $\varkappa$. In the improved notation of Sec. III, we then define the continued fraction solutions to (1.6) and describe a new type of acceleration of their convergence. Finally, we illustrate and summarize the resulting bound-state prescription in Secs. V and VI, respectively.

II. THE DIFFERENCE SCHRÖDINGER EQUATION

A. The $r\to 1/r$ symmetries

The form of our differential equation (1.2) remains unchanged after the reflection of coordinates $r\to\hat{r}$, $\psi\to\hat{\psi}$, i.e., say,

$$\hat{r}=1/r, \quad \hat{\psi}(\hat{r})=\psi(r)/r, \tag{2.1a}$$

which implies merely the change of notation

$$\hat{l}=l, \quad \hat{a}=c, \quad \hat{b}=-E, \quad \hat{c}=a, \quad \hat{E}=-b. \tag{2.1b}$$

The related interchange of asymptotics (1.3) and (1.4) may be characterized by an introduction of the "capped" symbols

$$\hat{\mu}=\lambda, \quad \hat{\lambda}=\mu. \tag{2.1c}$$

Moreover, we may put $\hat{n}=-n$, denote

$$\hat{A}_{\hat{n}}=C_n, \quad \hat{B}_{\hat{n}}=B_n, \quad \hat{C}_{\hat{n}}=A_n, \quad \hat{\varkappa}=1-\varkappa, \quad \hat{h}_{\hat{n}}=h_n, \tag{2.1d}$$

and extend the same formal symmetry to our difference Schrödinger equation (1.6). As a consequence, a transition $n\to -n$ may formally be interpreted as mere "capping" transformation of the parameters.

B. The $|n|\to\infty$ asymptotics

The leading-order asymptotical form of (1.6) reads

$$8\mu n h_n-4n^2h_{n+1}-8\lambda n h_{n+2}=0, \quad |n|\gg 1, \tag{2.2}$$

and admits the two independent "Jost" solutions $h_n^{(i)}$, $i=1,2$, with

$$h_{n+1}^{(1)}/h_n^{(1)}=2\mu/n+O(n^{-2}) \tag{2.3}$$

and

 0022-2488/90/010108-05$03.00

$$h^{(2)}_{n+2}/h^{(2)}_{n+1} + \cdots - n/2\lambda + O(1), \quad |n| \gg 1. \tag{2.4}$$

Obviously, the "large" $C(n^{|n|})$ components contradict the $M, N \to \infty$ convergence of our Laurent series (1.5), so that we must demand

$$h_n \approx h^{(1)}_n, \quad n \gg 1, \tag{2.5}$$

and

$$h_{-m} \approx h^{(2)}_{-m}, \quad m \gg 1. \tag{2.6}$$

The latter two boundary conditions do not replace the physical requirements (1.3) and (1.4): In light of the Floquet theory,[4] they merely fix the values of the "additional" free parameters $\varkappa$.

III. SCHRÖDINGER EQUATION AS RECURRENCES

A. The improved asymptotics

An insertion of the asymptotics (2.3) and (2.4) in the difference Schrödinger equation (1.6) or (2.2) recovers its asymptotical degeneracy to the two-term recurrences. Thus, for the positive or negative large n, respectively, the third or first item in (1.6) or (2.2) represents just a small correction of the order $O(n^{-2})$.

In the former case, the *Ansatz*

$$h_{n+1} = \frac{(2\mu)^n \Gamma(n+\alpha+1)}{\Gamma(n+\beta+1)\Gamma(n+\gamma+1)} p_n, \tag{3.1}$$

with

$$\alpha = \tfrac{1}{2}\varkappa + \tfrac{1}{4} - E/8\mu, \quad \beta,\gamma = \tfrac{1}{2}\varkappa + \tfrac{3}{4} \pm \tfrac{1}{2}U,$$
$$U^2 = (l+\tfrac{1}{2})^2 + 8\mu\lambda, \tag{3.2}$$

improves the first boundary condition (2.5) by a few further corrections. It also simplifies the difference Schrödinger equation which, in the new notation, acquires the form

$$p_{n-1} - p_n = \varphi_n p_{n+1},$$
$$\varphi_n = 4\mu\lambda \frac{(n+\alpha+1)(n+\delta)}{(n+\beta)(n+\gamma)(n+\beta+1)(n+\gamma+1)},$$
$$\delta = \tfrac{1}{2}\varkappa + \tfrac{5}{4} - b/8\lambda. \tag{3.3}$$

Similarly, the second boundary condition (2.6) finds an adequate gamma-function representation in the alternative *Ansatz*

$$h_{-m-1} = \frac{(2\lambda)^m \Gamma(m+s+1)}{\Gamma(m+v+1)\Gamma(m+w+1)} q_m, \tag{3.4}$$

with the parameters

$$s = -\tfrac{1}{2}\varkappa + \tfrac{3}{4} + b/8\lambda, \quad w,v = -\tfrac{1}{2}\varkappa + \tfrac{5}{4} + \tfrac{1}{2}U. \tag{3.5}$$

Obviously, the related new difference Schrödinger equation

$$q_{m-1} - q_m = \psi_m q_{m+1},$$
$$\psi_m = 4\mu\lambda \frac{(m+s+1)(m+\epsilon)}{(m+v)(m+w)(m+v+1)(m+w+1)},$$
$$\epsilon = -\tfrac{1}{2}\varkappa + \tfrac{7}{4} + E/8\mu, \tag{3.6}$$

remains related to (3.3) by the capping transformations (2.1), with

$$\hat{p}_n = q_n, \quad \hat{q}_m = p_m, \quad \hat{\varphi}_n = \psi_n, \quad \hat{\psi}_m = \varphi_m, \tag{3.7}$$

and

$$\hat{\alpha} = s, \quad \hat{\beta} = v, \quad \hat{\gamma} = w, \quad \hat{\delta} = \epsilon. \tag{3.8}$$

Thus, without any loss of generality, we may work with one of the two limits $|n| \to \infty$ only.

B. The initial values

Strictly speaking, the boundary conditions (2.5) and (2.6) need not necessarily be postulated in a capping-symmetric manner. Indeed, both these respective requirements are easily seen to represent just the conditions $p_n \approx$ const and $q_m \approx$ const for n and $m \gg 1$. In light of Eqs. (3.3) and (3.6) we may replace them immediately by an alternative requirement

$$p_{N+1} = 0, \quad q_{M+1} = 0, \quad N \to \infty, \quad M \to \infty. \tag{3.9}$$

Equation (3.9) admits arbitrary normalizations (say, $p^{(1)}_N = 1$ and $q^{(2)}_M = 1$) and we may then use Eqs. (3.3) and (3.6), respectively as recurrences. In such an interpretation (see, also, Paper I), the respective alternatives (3.3) (with $n \geqslant 0$) and (3.6) (with $m \geqslant 0$) are equivalent to the $n \geqslant 0$ and $n < -2$ rows of our original Schrödinger difference equation (1.6), and define the "asymptotically correct" (equivalently, "Jost-like") sequences $p^{(1)}_{N-1}, p^{(1)}_{N-2}, \ldots, p^{(1)}_{-1}$ and $q^{(2)}_{M-1}, q^{(2)}_{M-2}, \ldots, q^{(2)}_{-1}$, respectively. The remaining $n = -1$ row,

$$A_0 h^{(2)}_{-1}/h^{(2)}_0 + B_0 + C_0 h^{(1)}_1/h^{(1)}_0 = 0, \tag{3.10}$$

may be employed as a matching condition for the two recurrences. Computationally, the latter equation restricts only our freedom in choosing parameters: In what follows, we shall use it as a definition of $\varkappa$'s for each choice of energy E.

IV. THE ANALYTIC CONTINUED FRACTIONS AND AN ACCELERATION OF THEIR CONVERGENCE

We may introduce the quantities

$$f_n = p_n/p_{n-1}, \quad g_m = q_m/q_{m-1}, \tag{4.1}$$

and rewrite our Schrödinger equation [recurrences (3.3) and (3.6)] in the form

$$f_n = 1/(1+\varphi_n f_{n+1}), \quad g_m = 1/(1+\psi_m g_{m+1}) \tag{4.2}$$

In combination with (3.9), they represent the standard analytic continued fractions[5-7]

$$f_n = \cfrac{1}{1+\varphi_n \cfrac{1}{1+\varphi_{n+1}\cfrac{1}{1+\cdots}}},$$
$$g_m = \cfrac{1}{1+\psi_m \cfrac{1}{1+\psi_{m+1}\cfrac{1}{1+\cdots}}}. \tag{4.3}$$

With the finite truncation parameters and initial values

$$f_{N+1} = 0, \quad g_{M+1} = 0, \tag{4.4}$$

our recurrences (4.2) still define good approximants since the corresponding coefficients are asymptotically very small,

$$\varphi_n = O(1/n^2), \quad \psi_m = O(1/m^2). \tag{4.5}$$

Numerically, it is therefore reasonable to define our solutions (3.1) and (3.4) as products, with

$$p_n = f_n f_{n-1} \cdots f_{n_0+1} p_{n_0}, \quad q_m = g_m g_{m-1} \cdots g_{m_0+1} q_{m_0}, \tag{4.6}$$

where p_{n_0} and q_{m_0} represent the two alternative normalizations.

In general, the continued fractional convergence may be accelerated by means of the fixed-point technique as described in Paper I. An alternative rearrangement will be described now—it seems to be much more efficient for the particular form of our recurrences (4.2) [with the small parameters (4.5)]. For the sake of brevity, we shall recall the capping symmetry (3.7) and restrict our attention to f_n's in what follows.

Starting from the slightly modified form of Eq. (3.3),

$$\begin{aligned} &p_{n+1} = \chi_{n+1}^{(0)} p_n - \varphi_{n+1}^{(0)} p_{n+2}, \\ &\chi_{n+1}^{(0)} = 1, \quad \varphi_{n+1}^{(0)} = \varphi_{n+1}, \end{aligned} \tag{4.7}$$

we shall eliminate p_{n+1} from Eq. (3.3). We get the $j=1$ formula

$$p_n = \chi_n^{(j)} p_{n-1} - \varphi_n^{(j)} p_{n+j+1}, \tag{4.8}$$

where

$$\begin{aligned} \chi_n^{(1)} &= \frac{\chi_n^{(0)}}{1+\varphi_n^{(0)}\chi_{n+1}^{(0)}} = \frac{1}{1+\varphi_n}, \\ \varphi_n^{(1)} &= -\frac{\varphi_n^{(0)}\varphi_{n+1}^{(0)}}{1+\varphi_n^{(0)}\chi_{n+1}^{(0)}} = -\frac{\varphi_n\varphi_{n+1}}{1+\varphi_n}. \end{aligned} \tag{4.9}$$

Next, we modify the preceding $j=0$ and $j=1$ formulas,

$$\begin{aligned} &p_{n+2} = \chi_{n+2}^{(0)} p_{n+1} - \varphi_{n+2}^{(0)} p_{n+3}, \\ &p_{n+1} = \chi_{n+1}^{(1)} p_n - \varphi_{n+1}^{(1)} p_{n+3}, \end{aligned} \tag{4.7'}$$

eliminate p_{n+1} and p_{n+2} from (4.8) with $j=1$, and get the explicit $j=2$ form of the latter prescription, with

$$\begin{aligned} \chi_n^{(2)} &= \frac{\chi_n^{(1)}}{1+\varphi_n^{(1)}\chi_{n+2}^{(0)}\chi_{n+1}^{(1)}} = \frac{1+\varphi_{n+1}}{1+\varphi_n+\varphi_{n+1}}, \\ \varphi_n^{(2)} &= -\varphi_n^{(1)} \frac{\varphi_{n+2}^{(0)} + \chi_{n+2}^{(0)}\varphi_{n+1}^{(1)}}{1+\varphi_n^{(1)}\chi_{n+2}^{(0)}\chi_{n+1}^{(1)}} = \frac{\varphi_n\varphi_{n+1}\varphi_{n+2}}{1+\varphi_n+\varphi_{n+1}}. \end{aligned} \tag{4.9'}$$

For the subsequent j's, the coefficients may also be found in the same manner.

In combination with definition (4.1), our relation (4.8) implies that

$$f_n = \chi_n^{(j)} / (1 + \varphi_n^{(j)} f_{n+1} f_{n+2} \cdots f_{n+j+1}). \tag{4.10}$$

These recurrences have the structure of the "extended" continued fractions[8] and degenerate back to the ordinary continued fractions and Eq. (4.2) for $j=0$. Since $\varphi_n^{(j)} = O(n^{-2j-2})$ is extremely small for large n's, even the drastic truncations of (4.10), e.g.,

$$\begin{aligned} f_n &= \chi_n^{(j)}(1 + O(1/n^{2j+2})) \\ &= \chi_n^{(j)} / (1 + \varphi_n^{(j)} \chi_{n+1}^{(j)} \cdots \chi_{n+j+1}^{(j)})(1 + O(1/n^{4j+4})), \end{aligned} \tag{4.11}$$

etc., represent the extremely efficient approximants.

The first nontrivial formulas (4.11) may already be derived from the explicit equation (4.10),

$$f_n = 1/(1 + \varphi_n - \varphi_n\varphi_{n+1} f_{n+1} f_{n+2}). \tag{4.12}$$

Its simplicity is, surprisingly, preserved also by the higher-order formulas

$$f_n = \frac{Q_{n+1}^{(j-1)}}{\left(Q_n^{(j)} - \prod_{m=n}^{n+j} (-\varphi_m f_{m+1})\right)}, \tag{4.13}$$

where

$$Q_n^{(-1)} = Q_n^{(0)} = 1 \tag{4.14}$$

and

$$Q_n^{(j+1)} = Q_{n+1}^{(j)} + \varphi_n Q_{n+2}^{(j-1)}, \quad j = 0,1,..., \tag{4.15}$$

in general. The proof by induction is a simple consequence of the identity $(a+b)^{-1} = a^{-1} + a^{-1}b(a+b)^{-1}$. A rearrangement

$$Q_n^{(j)} = 1 + \varphi_n + \sum_{l=0}^{j-2} \varphi_{n+l+1} Q_n^{(l)}, \quad j = 2,3,..., \tag{4.16}$$

of (4.15) quickly gives the final algorithm for each reasonable j.

V. THE PHYSICAL BOUND STATES

A. An imposition of boundary conditions

In a way paralleling Paper I, we may pick up two roots $\varkappa_1(E)$ and $\varkappa_2(E)$ of the "Hill determinant" [Eq. (3.10)] or of its equivalent continued fractional (or extended continued fractional) form

$$1 + \psi_{-1} g_0 + \varphi_{-1} f_0 = 0. \tag{5.1}$$

The corresponding pair of the Laurent-series solutions $\psi^{(i)}(r) = \psi(r,\varkappa_i)$, $i=1,2$, may be then used in a superposition

$$\psi(r) = d_1\psi^{(1)}(r) + d_2\psi^{(2)}(r) \tag{5.2}$$

and defines the general solution of our differential equation (1.2) of the second order. We have to avoid the cases where the two functions $\psi^{(i)}(r)$ are linearly dependent (e.g., for $\varkappa_1 = \varkappa_2 +$ an even integer).

In accord with the standard oscillation theorems,[4] the number of nodes (zeros) of $\psi(r)$ may only increase with the increasing energy E. As a consequence (see, also, Paper I), all the binding energies may be characterized by an asymptotic emergence of a node in $\psi(r)$,

$$\psi(r_0) = 0, \quad \psi(r_\infty) = 0, \quad r_0 \ll 1, \quad r_\infty \gg 1. \tag{5.3}$$

In the limit $r_0 \to$ and $r_\infty \to \infty$, such a pair of equations will specify the exact binding energies and coefficients d_i in (5.2) in principle. Thus an insertion of (5.2) gives

$$\begin{aligned} &d_1\psi^{(1)}(r_0) + d_2\psi^{(2)}(r_0) = 0, \\ &d_1\psi^{(1)}(r_\infty) + d_2\psi^{(2)}(r_\infty) = 0, \end{aligned} \tag{5.4}$$

and the related secular equation

$$\det\begin{pmatrix} \psi^{(1)}(r_0) & \psi^{(2)}(r_0) \\ \psi^{(1)}(r_\infty) & \psi^{(2)}(r_\infty) \end{pmatrix} = 0 \tag{5.5}$$

defines the spectrum of energies.

In place of the simple-minded requirement (5.3), we may also employ our knowledge of the more precise asymp-

totics (1.3) and (1.4). This leads to a pair of equations similar to (5.4),

$$d_1 A_{10} + d_2 A_{20} = 0,$$
$$d_1 A_{1\infty} + d_2 A_{2\infty} = 0, \tag{5.6}$$

where

$$A_{ij} = \sum_{m=-\infty}^{\infty} h_m(\varkappa = \varkappa_i) r^{2m+\varkappa_i}\cdot(2m + \varkappa_i + \Delta_j),$$
$$i = 1,2, \quad j = 0,\infty, \tag{5.7}$$

and the constants Δ_j represent the explicit second-order WKB corrections

$$\Delta_0 = -\tfrac{3}{2} - b/4\lambda, \quad \Delta_\infty = \tfrac{1}{2} - E/4\mu. \tag{5.8}$$

The energies may also be computed as roots of the modified secular equation

$$\det \begin{pmatrix} A_{10} & A_{20} \\ A_{1\infty} & A_{2\infty} \end{pmatrix} = 0 \tag{5.9}$$

in principle.

B. Numerical examples

For a given set of couplings in our Schrödinger equation [(1.1) and (1.2)], we may employ Eq. (5.1) as an implicit definition of functions $\varkappa_1(E)$ and $\varkappa_2(E)$, and determine numerically the binding-energy root from Eq. (5.5). The former procedure resembles a search for solutions of quadratic equations: We have chosen our examples in such a manner that the two roots happen to be complex conjugated,

$$\varkappa_{1,2} = \mathrm{Re}\,\varkappa \pm i\omega(E).$$

We have also scaled $r \to \mathrm{const} \times r$ in such a way that $a = 1$. Then, we have employed the purely numerical Runge–Kutta[1] algorithm (cf., also Paper I) and determined the "reference" exact energies E_{exact}. Their sample is listed in Table I for $c = 0.8$ and for a few different values of b. For our choice of the matching points (quite close to the semiclassical turning points: $r_0^{(\mathrm{WKB})} \doteq 0.74$ for $E = 4.8$, and $r_0^{(\mathrm{WKB})} \doteq 0.66$ for $E = 9.44$, etc.), the ground-state precision is already quite satisfactory. For the first excited state, high precision is not yet achieved.

The first comparison of roots E of Eq. (5.5) with E_{exact} is displayed in Table II. Again, very good precision is achieved for ground states, and good results also appear in the excited states. An exception (the last excited-state column) may immediately be understood after an inspection of Table III where the corresponding $\varkappa$'s are listed: Close to the real axis, the values of $\omega(E)$ remain still strongly sensitive to the variation of r_0 and r_∞. This represents one of the immanent limitations of the method—in a way analogous to the theory of Mathieu functions,[9] a modification of the *Ansatz* is needed at the singularity $\omega(E) \to 0$. Because of the worsening of convergence as observed above, such a situation is less interesting in the present context; vice versa, the increase of $|\omega(E)|$ becomes reflected by an improvement of precision—this may be illustrated by a comparison of Tables II and III in the ground-state cases (A).

A priori, an improvement of the boundary conditions [equal to a transition to the secular equation (5.9)] should also improve the rate of convergence. In practice, it is not always so—this problem appears even in implementations of the Runge–Kutta method sometimes. Here, an explanation is quite easy—the differentiation worsens the rate of convergence of the expansions (5.7). As a consequence of the related rounding errors, Eq. (5.9) is not always superior to Eq. (5.5) in the numerical sense. A typical illustration of this ambivalence is presented in Table IV. Again, the position of singularities is relevant—in accord with Table V, the ground-state root $\varkappa_1(E)$ moves towards the real axis during an increase of the coupling b from the value 0.8. In fact, the complex conjugate roots meet [reach $\omega(E) = 0$] somewhere in between the values of $b = 0.904$ and 0.908. Table V illustrates also the further movement of $\varkappa_1(E)$—presumably, it reaches the value $\varkappa_1 = 1$ at $b = 1$.

TABLE I. The convergence of energies with $r_0 \to 0$ and $r_\infty \to \infty$ (the standard Runge–Kutta method, $a = 1$, $c = 0.8$): (A) ground states, (B) the first excited states.

	b		1.00	1.02	1.04	1.06
	r_0	r_∞	E_{exact}			
(A)	0.3	3.5	4.876 70	4.883 68	4.890 63	4.897 54
	0.25	4.0	4.875 98	4.882 95	4.889 89	4.896 80
	0.2	4.5	4.875 93	4.882 94	4.889 86	4.896 77
(B)	0.3	3.5	9.441 4	9.448 7	9.456 0	9.463 1
	0.25	4.0	9.378 7	9.385 7	9.392 6	9.399 5
	0.2	4.5	9.373 6	9.380 6	9.387 5	9.394 4

TABLE II. A comparison of present results [E from Eq. (5.5)] with E_{exact} of Table I. (A) $E - E_{\mathrm{exact}}$, ground states. (B) $E - E_{\mathrm{exact}}$, the first excited states.

	b		1.00	1.02	1.04	1.06
	r_0	r_∞	$E - E_{\mathrm{exact}}$			
(A)	0.3	3.5	0.007 30	0.007 34	0.007 39	0.007 43
	0.25	4.0	0.000 370	0.000 373	0.000 376	0.000 379
	0.2	4.5	0.000 016	0.000 016	0.000 016	0.000 016
(B)	0.3	3.5	0.063	0.520	0.185	0.021
	0.25	4.0	0.023	0.023	0.023	0.023
	0.2	4.5	0.001 3	0.001 3	−0.000 5	−0.064

TABLE III. Parameters $\varkappa_{1,2} = 0.5 \pm i\omega$ [= roots of Eq. (5.1)] pertaining to Table II.

	b		1.00	1.02	1.04	1.06
	r_0	r_∞	ω			
(A)	0.30	3.5	0.660 42	0.678 22	0.695 15	0.711 30
	0.25	4.0	0.660 58	0.678 35	0.695 24	0.711 35
	0.20	4.5	0.660 59	0.678 35	0.695 25	0.711 35
(B)	0.30	3.5	0.568	0.651	0.490	0.305
	0.25	4.0	0.534	0.469	0.383	0.253
	0.20	4.5	0.525	0.458	0.367	0.137

TABLE IV. A comparison of results for the simple and WKB boundary conditions [Eq. (5.1) for $x_{1,2} = 1.5 \pm i\omega(E)$ and the respective Eqs. (5.5) and (5.9) for energies]. (A) The ground state with $a = c = 1$ and $b = 0.8$. (B) The first excited state with $a = c = 1$ and $b = 1$.

	Boundary conditions		Simple		WKB	
	r_0	r_∞	$\omega(E)$	$E - E_{\text{exact}}$	$\omega(E)$	$E - E_{\text{exact}}$
(A)	0.3	3.5	0.444 35	0.007 75	0.459	0.044
	0.25	4.0	0.441 05	0.000 44	0.452	0.029
	0.2	4.5	0.440 88	0.000 14	0.447	0.016
(B)	0.3	3.5	1.170 2	0.200	1.142 49	0.0641
	0.25	4.0	1.144 5	0.258	1.141 25	0.0049
	0.2	4.5	1.141 3	0.002	1.141 17	0.0004

VI. CONCLUDING REMARKS

In a way that resembles the use of the exact Mathieu solutions for potentials $V \sim 1/r^4$ (see, e.g., Ref. 10 or the review by Newton[11]), we have described here the quasi-exact construction of bound states in the potential (1.1).

A core of our construction lies in the use of continued fractions. In this respect, our method and, in particular, the related new technique of acceleration of their convergence, might prove relevant in a broader methodical context. In conclusion, let us mention at least the following three possibilities.

(a) In a number of papers, we may find an analytic continued fraction rearrangement of the three-term recurrences that resemble our Eq. (1.6) even by their power-series origin. In the underlying spectrum of technical problems, we may distinguish between the mathematically "safe" variational tractability of the potentials derived from

$$r^2 + fr^2/(1 + gr^2) \tag{6.1}$$

(see e.g., Ref. 12), and the straightforward (so-called "Hill-determinant") treatments of the forces of the type

$$ar^2 + br^4 + cr^6 \tag{6.2}$$

(see, e.g., Ref. 13). Both these techniques lie very close to our present construction: The acceleration of the continued-fractional convergence plays an important role in it.

(b) Formally, a very broad class of applications of continued fractions has been inspired by the Lanczos ideas[14] related to the theory of moments.[5-7] For a sketchy illustration, we may recall, e.g., the diagonalization of the so-called chain models in the condensed-matter physics[15] or the extensive Lanczos-type computations in nuclear physics.[16] Also, we may mention here the closely related rearrangements of the divergent Born series[17] or the whole rich domain of the nonequilibrium statistics,[18] etc. Often, an acceleration of convergence remains an open problem in this context.[19]

(c) Whenever we interpret continued fractions as a special case of the Padé resummation,[20] we immediately get in close contact with the perturbation theory.[21] In this setting, an acceleration of convergence is of extreme importance[22] and, presumably, new inspiration could stem from the present results. Unfortunately, this question already lies too far beyond the scope of the present paper.

TABLE V. The first ground-state root $x_1(E)$ of Eq. (5.1). An example of its transition in the complex plane ($a = c = 1$).

r_0	0.30		0.25	
r_∞	3.50		4.00	
b	Re x_1	Im x_1	Re x_1	Im x_1
1.00	1.0005	0.0	1.0000	0.0
0.98	1.0705	0.0	1.0699	0.0
0.94	1.2259	0.0	1.2251	0.0
0.92	1.3331	0.0	1.3319	0.0
0.908	1.4639	0.0	1.4590	0.0
0.904	1.5000	0.1035	1.5000	0.0843
0.90	1.5000	0.1383	1.5000	0.1247
0.88	1.5000	0.2438	1.5000	0.2367
0.80	1.5000	0.4443	1.5000	0.4410
0.60	1.5000	0.6686	1.5000	0.6669
0.20	−0.5000	−0.8731	−0.5000	−0.8721

ACKNOWLEDGMENTS

I appreciate the inspiring discussions with Professor A. Vančura and Dr. R. Horsley.

The partial support by the "Freundeskreis der Universität Kaiserslautern" during the numerical computations is also much appreciated.

[1] F. S. Acton, *Numerical Methods That Work* (Harper and Row, New York, 1970).
[2] M. Znojil, J. Phys. A: Math. Gen. **15**, 2111 (1982).
[3] M. Znojil, J. Math. Phys. **30**, 23 (1989).
[4] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[5] N. I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, London, 1965).
[6] O. Perron, *Die Lehre von der Kettenbruechen* (Teubner, Stuttgart, 1957).
[7] W. B. Jones and W. J. Thron, *Continued Fractions: Analytic Theory and Applications*, Vol. 11 of *Encyclopedia of Mathematics and its Applications*, edited by G.-C. Rota (Addison–Wesley, Reading, MA, 1980).
[8] M. Znojil, J. Math. Phys. **24**, 1136 (1983).
[9] H. Bateman and A. Erdelyi, *Higher Transcendental Functions III* (McGraw-Hill, New York, 1955).
[10] E. Vogt and G. H. Wanier, Phys. Rev. **95**, 1190 (1954).
[11] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982).
[12] M. Znojil, J. Phys. A: Math. Gen. **16**, 293 (1983); Phys. Lett. A **94**, 120 (1983).
[13] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978); D. P. Datta and S. Mukherjee, J. Phys. A: Math. Gen. **13**, 3161 (1980); M. Znojil, Phys. Rev. D **26**, 3750 (1982); D. Masson, J. Math. Phys. **24**, 2074 (1983); M. Tater, J. Phys. A: Math. Gen. **20**, 2483 (1987); M. A. Shifman, "New findings in Quantum Mechanics," preprint CERN-TH.5265/88 (unpublished).
[14] C. Lanczos, J. Res. Natl. Bur. Stand. **49**, 33 (1952).
[15] R. Haydock, in *The Recurrent Method and Its Applications*, edited by D. G. Pettifor and D. L. Wearie (Springer, Berlin, 1985).
[16] R. R. Whitehead, in *Theory and Applications of Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1983).
[17] M. Znojil, Phys. Rev. A **34**, 2697 (1986).
[18] M. H. Lee, Phys. Rev. B **26**, 1072 (1982); J. Math. Phys. **24**, 2512 (1983).
[19] M. Znojil, J. Math. Phys. **29**, 139 (1988).
[20] G. A. Baker, Jr. and P. Graves-Morris, *Pade Approximants I, II*, Vols. 13 and 14 of *Encyclopedia of Mathematics and its Applications*, edited by G.-C. Rota (Addison–Wesley, Reading, MA, 1980).
[21] J. Cizek and E. R. Vrscay, Int. J. Quantum Chem. **21**, 27 (1982); P.-O. Loewdin, *ibid.* **21**, 69 (1982).
[22] E. R. Vrscay, Phys. Rev. A **31**, 2054 (1985), and references therein.

Quantization of bi-Hamiltonian systems

D. J. Kaup
Department of Mathematics, Clarkson University, Potsdam, New York 13676
Peter J. Olver
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 27 July 1989; accepted for publication 6 September 1989)

One of the distinguishing features of soliton equations is the fact that they can be written in Hamiltonian form in more than one way. Here we compare the different quantized versions of the soliton equations arising in the AKNS inverse scattering scheme. It is found that, when expressed in terms of the scattering data, both quantized versions are essentially identical.

In 1975, one of the present authors[1] showed how to obtain the quantized levels of the nonlinear Schrödinger equation using the action-angle variables (canonical coordinates) of the AKNS scattering data. The symplectic form used to effect the reduction to canonical coordinates was based on the standard Hamiltonian structure for the nonlinear Schrödinger equation. The method used was a nonlinear generalization of one of the standard methods for the second quantization of the electromagnetic field. As presented in the textbook by Schiff,[2] one takes the classical electromagnetic field and decomposes it into normal modes (Fourier components). The key idea in this approach is that the classical electromagnetic Hamiltonian will decompose into a sum of noninteracting classical Hamiltonians, each of which has just two degrees of freedom and is easily quantized by itself. This method of quantization bypasses all the inherent difficulties of fully quantizing the system, including the factor-ordering problem, defining the quantum field operators for the fundamental fields, etc.[3] It is fundamentally based on the symmetries of the classical system, and reduces the problem to one of quantizing noninteracting particles.[4] In this way, the original difficult second quantization problem is reduced to a simpler set of noninteracting problems. The advantage of this simpler solution is tremendous when one considers the information that one can glean from it. First, one can obtain the spacings of the energy levels. One also discovers which quantum variables will commute, and which modes will have a particle-like behavior. Of course, for a full quantum theory, one still has to deal with a number of remaining difficult problems, including finding a consistent factor-ordering for the quantum operators, evaluating matrix elements, etc. Unfortunately, the solution to this larger quantization problem may well be multivalued.[3] However, in the meantime, one has been able to immediately isolate the above mentioned important features of second quantization, and, very importantly, those quantities which would have the same common solution for every possible consistent second quantization. Thus, any difficulty which would be found at this level would also be present in *any* quantum field theory. And a study by this method can provide valuable insight into the structure of the more thorny parts of the second-quantization problem.

The symplectic form used in Ref. 1 to effect the reduction to canonical coordinates was based on the first Hamiltonian structure for the nonlinear Schrödinger equation. In 1978, Magri[5] showed how many soliton equations, including the nonlinear Schrödinger equation, could be written as bi-Hamiltonian systems, meaning that they have two distinct, but compatible, Hamiltonian structures. Indeed, his fundamental result showed that, subject to some technical hypotheses[5,6] any bi-Hamiltonian system is completely integrable in the sense that it has infinitely many conservation laws in involution and corresponding commuting Hamiltonian flows.

From the viewpoint of quantum mechanics, the existence of more than one Hamiltonian structure for a given classical mechanical system raises the possibility of there existing more than one quantized version of this system, even at the level of quantization considered in Ref. 1. The resulting ambiguity in the quantization procedure raises serious physical doubts as to the mathematical framework of quantization. However, the main result to be proven here is that, for AKNS soliton equations,[7] both quantized versions are essentially the same. We demonstrate that, in terms of the respective canonical coordinates on the scattering data, the two Hamiltonians have identical expressions, and hence identical quantum versions. Indeed, we conjecture that this phenomenon is true in general: *quantization does not depend on the underlying Hamiltonian structure.* (The results of Dodonov *et al.*,[8] in which an ambiguity in the quantization procedure for certain finite-dimensional bi-Hamiltonian systems is supposedly demonstrated, are erroneous, since they fail to incorporate the important topological properties of phase space properly in their picture. Indeed, their ambiguity is just a version of the ambiguity inherent in the quantization of two-dimensional Hamiltonian systems, which we discuss in detail below.) Moreover, we will see that for the other members of the associated hierarchy of soliton equations the only difference in the quantum versions is in the choice of weighting factor for the quantum operators corresponding to the continuous spectrum, the weight being determined by the classical dispersion relation, and the replacement of the bound state Hamiltonians. Thus, the effect of quantizing different members of the soliton hierarchy will only be significant for the bound states/solitons.

Our presentation relies heavily on the notation and results in earlier papers by Kaup and Newell[1,9,10] on the closure of the squared eigenfunctions for the AKNS scattering

 0022-2488/90/010113-05$03.00

problem. The key to our result is the well-known fact that the recursion operator, which is built out of the two Hamiltonian operators for the system[5,6] is essentially the squared eigenfunction operator. Since variations in the potential for the AKNS scattering problem are expressed in terms of the squared eigenfunctions, the second symplectic form can be simply written down in explicit form. In terms of the scattering data, it differs from the first symplectic form only by a weighting factor in the continuous spectrum, and a change in the discrete components. However, the corresponding difference in weighting factors for the two Hamiltonians exactly cancels out the weighting factor for the two symplectic forms, while the discrete components reduce simply to the quantization of a two-dimensional Hamiltonian system, based on different symplectic structures. Thus, the entire quantum ambiguity reduces to the simple matter of an ambiguity in the quantization of two-dimensional Hamiltonian systems, a problem that is easily handled.

Our notation is as follows. Hamilton's equations are

$$\partial_t Q^\alpha = J^{\alpha\beta}\,\partial_\beta H\,, \tag{1}$$

where $Q = \{Q^\alpha\}$ are the dynamical variables (the p's and the q's), $J = [J^{\alpha\beta}]$ is the Hamiltonian operator, which determines the underlying Hamiltonian structure of the phase space, and H is the Hamiltonian function or density. For instance, for a harmonic oscillator, one would take

$$Q = \begin{pmatrix} q \\ p \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \text{and} \quad H = \tfrac{1}{2}(p^2 + q^2).$$

When Q is a function of a continuous variable, the sum over the dummy indices in (1) is understood to include the appropriate integration, and the partial derivative is understood to be a functional derivative instead. The Poisson bracket determined by such a Hamiltonian operator has the form

$$\{F, G\} = (\partial_\alpha F) J^{\alpha\beta}\,\partial_\beta G, \tag{2}$$

which requires the symplectic two-form to be

$$\Omega = \tfrac{1}{2}\, dQ^\alpha \wedge J^{-1}_{\alpha\beta}\, dQ^\beta\,. \tag{3}$$

For the harmonic oscillator, this reduces to the familiar canonical form

$$\Omega = dp \wedge dq\,. \tag{4}$$

Therefore, the operator J needs to be skew adjoint, and satisfy the additional condition that the Poisson bracket (2) satisfy the Jacobi identity, which is equivalent to the requirement that the two-form Ω can be closed.[6]

Before presenting the main results, we discuss a simple but crucial fact that any two-dimensional Hamiltonian system has a unique quantized version, even though it has many different Hamiltonian structures. In terms of the standard Hamiltonian structure prescribed by the canonical two-form (4), Hamilton's equations take the classical form[11]

$$p_t = -\frac{\partial H}{\partial q}, \quad q_t = \frac{\partial H}{\partial p}\,. \tag{5}$$

In $\mathbb{R}^2$, any nonzero two-form $\lambda(p, q)\,dp \wedge dq$ is always closed, and hence determines a Hamiltonian operator

$$\tilde{J} = \begin{pmatrix} 0 & -\dfrac{1}{\lambda} \\ \dfrac{1}{\lambda} & 0 \end{pmatrix}.$$

It is easy to see that (5) can be written in Hamiltonian form using this second Hamiltonian structure if and only if λ is a function of the Hamiltonian H. In this case, the new Hamiltonian function is

$$H_2(p,q) = \Phi[H(p,q)],$$

where $\Phi(\xi)$ is any nonvanishing scalar function, and

$$\Omega_2 = \Phi'[H(p,q)]\,dp \wedge dq \tag{6}$$

is the second symplectic form. Re-expressing Ω_2 in canonical form will lead to new canonical variables $\tilde{p}$, $\tilde{q}$, and an ostensibly different quantized version. However, provided this transformation does not affect the phase space topology, it is not hard to see that these two quantized versions will end up being identical, at least in the semi-classical limit, and so there is no ambiguity in the (semi-classical) quantization of two-dimensional Hamiltonian systems.

We now turn to our problem at hand. For simplicity, we will consider the general nonlinear Schrödinger equation

$$iq_t = -q_{xx} + 2rq^2, \tag{7a}$$

$$ir_t = r_{xx} + 2qr^2, \tag{7b}$$

in detail. However, our arguments will work equally well for any other soliton equation associated with the AKNS spectral problem[7]; see the remarks at the end of the paper. For $r = \pm q^*$, (7) reduces to the single equation

$$iq_t = -q_{xx} \pm 2(q^* q)\, q\,, \tag{8}$$

which is the form of the nonlinear Schrödinger equation in which all physical constants, e.g., $\hbar$, m, etc., have been set equal to 1. According to Magri,[5] the nonlinear Schrödinger equation can be written as a bi-Hamiltonian system

$$\Psi_t = J_1\,\partial H_1 = J_2\,\partial H_2\,. \tag{9}$$

The first Hamiltonian can be identified with the (signed) energy

$$H_1 = \pm E = \int_{-\infty}^{\infty} (q_x r_x + q^2 r^2)\,dx\,, \tag{10}$$

while the second Hamiltonian is the field momentum

$$H_2 = P = i\int_{-\infty}^{\infty} (rq_x - qr_x)\,dx\,. \tag{11}$$

The two Hamiltonian operators are given by

$$J_1 = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{12}$$

$$J_2 = \frac{1}{2}\,\sigma_1 \partial_x + \begin{pmatrix} q\displaystyle\int_{-\infty}^{x} q & -q\displaystyle\int_{-\infty}^{x} r \\ -r\displaystyle\int_{-\infty}^{x} q & r\displaystyle\int_{-\infty}^{x} r \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{13}$$

(In our notation,[6] we have omitted the delta functions used by some authors.) Moreover, these Hamiltonian structures are compatible, in the sense that any linear combination $c_1 J_1 + c_2 J_2$ is also Hamiltonian. Therefore, according to the theorem of Magri the operator

$$R = J_2 \cdot J_1^{-1} \tag{14}$$

is a recursion operator for the general nonlinear Schrödinger equation, leading to an infinite hierarchy of mutually commuting bi-Hamiltonian flows.

To determine the two quantized versions of the nonlinear Schrödinger equation, we need to introduce canonical coordinates and momenta, which will be found among the scattering data for the associated eigenvalue problem. We begin by recalling how this was done in Ref. 1 for the first symplectic form. The general nonlinear Schrödinger equation can be solved using the AKNS eigenvalue problem[7]

$$v_{1,x} + i\zeta v_1 = q v_2, \quad v_{2,x} - i\zeta v_2 = r v_1 . \tag{15}$$

We let

$$\phi = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

be the solution to (15) satisfying the boundary conditions

$$\phi \to e^{-i\zeta x}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad x \to -\infty, \quad \phi \to \begin{pmatrix} a(\zeta)e^{-i\zeta x} \\ b(\zeta)e^{i\zeta x} \end{pmatrix}, \quad x \to \infty,$$

for $\operatorname{Im}\zeta > 0$. Similarly, let

$$\bar{\phi} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix}$$

be the solution to (15) satisfying the boundary conditions

$$\bar{\phi} \to e^{-i\zeta x}\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad x \to -\infty,$$

$$\bar{\phi} \to \begin{pmatrix} \bar{b}(\zeta)e^{-i\zeta x} \\ -\bar{a}(\zeta)e^{i\zeta x} \end{pmatrix}, \quad x \to \infty ,$$

for $\operatorname{Im}\zeta < 0$. This serves to define the scattering coefficients $a, b, \bar{a}, \bar{b}$, which also satisfy

$$\bar{a}(\zeta)a(\zeta) + \bar{b}(\zeta)b(\zeta) = 1 . \tag{16}$$

The ratio $\rho(\xi) = b(\xi)/a(\xi)$, ξ real, serves to define the continuous spectrum of the scattering data for (16). The zeros of $a(\zeta)$ in the upper half plane correspond to the bound states, and are denoted as $\zeta_j = \xi_j + i\eta_j, j = 1,\ldots,N$. Finally let b_j denote the value of b at ζ_j, and let ρ_j denote the residue of ρ at the pole ζ_j. Similar quantities are defined for the eigenvalues $\bar{\zeta}_j$.

In Ref. 1 it was shown how to express the first symplectic two-form in terms of the scattering data in the case $r = \pm q^*$. Tracing through the calculation there in the more general case, we find that

$$\begin{aligned} \Omega_1 &= i\int_{-\infty}^{\infty} \{\delta q \wedge \delta r\} dx \\ &= \frac{i}{\pi}\int_{-\infty}^{\infty} \{\delta \log b(\xi) \wedge \delta \log[\bar{a}(\xi)a(\xi)]\} d\xi \\ &\quad - 2\sum_{j=1}^{N} (\delta\zeta_j \wedge \delta \log b_j + \delta\bar{\zeta}_j \wedge \delta \log \bar{b}_j) , \end{aligned} \tag{17}$$

where the last sum is absent if $r = +q^*$, since there are no bound states. When $r = \pm q^*$, then $\bar{a}(\xi) = a(\xi)^*$, and $\bar{b}(\xi) = \mp b(\xi)^*$. In this case one can choose canonically conjugate variables by letting

$$A_j = 4\eta_j, \quad p_j = -4\xi_j, \quad p(\xi) = -(i/\pi)\log|a(\xi)| ,$$

represent the momenta (p's), and letting

$$B_j = \arg b_j, \quad q_j = \log|b_j|, \quad q(\xi) = \arg b(\xi)$$

represent the conjugate coordinates (q's) for the system. The first Hamiltonian functional is then expressed as

$$\begin{aligned} H_1 = \pm E &= \frac{4}{\pi}\int_{-\infty}^{\infty} \xi^2 \log(|a(\xi)|) d\xi \\ &\quad - \frac{8i}{3}\sum_{j=1}^{N} (\bar{\zeta}_j^3 - \zeta_j^3) . \end{aligned} \tag{18}$$

From this expression, the quantized form follows directly as in Ref. 1.

For the second symplectic form, we first recognize that by (12), (13) and Ref. 7,

$$J_2 = L^A J_1 = L^A \sigma_2 , \tag{19}$$

where L^A is the recursion operator for the squared eigenfunctions. Recall that the *squared eigenfunctions* corresponding to (15) are the functions

$$\Psi(\zeta,x) = \begin{pmatrix} v_1(\zeta,x)^2 \\ v_2(\zeta,x)^2 \end{pmatrix}.$$

We define the corresponding quantities Ψ_j for the bound states ζ_j similarly. The key result[10] is that the recursion operator L^A, given in (19), has the squared eigenfunctions as eigenstates:

$$L^A\Psi = \zeta\Psi, \quad L^A\Psi_j = \zeta_j\Psi_j . \tag{20}$$

Thus we can compute the second symplectic form

$$\Omega_2 = \tfrac{1}{2}\langle \delta V^A | \wedge \sigma_2 (L^A)^{-1} | \delta V \rangle .$$

Now, according to (B3) of Ref. 10,

$$\begin{aligned} \delta V &= \frac{1}{\pi}\int_{-\infty}^{\infty} [\delta\rho(\xi)\Psi(\xi) - \delta\bar{\rho}(\xi)\bar{\Psi}(\xi)] d\xi \\ &\quad - 2i\sum_{j=1}^{N} (\delta\rho_j \Psi_j + \rho_j \delta\zeta_j \chi_j + \delta\bar{\rho}_j \bar{\Psi}_j + \bar{\rho}_j \delta\bar{\zeta}_j \bar{\chi}_j). \end{aligned}$$

Therefore, using (20),

$$\begin{aligned} (L^A)^{-1}\delta V &= \frac{1}{\pi}\int_{-\infty}^{\infty} [\delta\rho(\xi)(L^A)^{-1}\Psi(\xi) - \delta\bar{\rho}(\xi)(L^A)^{-1}\bar{\Psi}(\xi)] d\xi \\ &\quad - 2i\sum_{j=1}^{N} (\delta\rho_j (L^A)^{-1}\Psi_j + \rho_j \delta\zeta_j (L^A)^{-1}\chi_j + \delta\bar{\rho}_j (L^A)^{-1}\bar{\Psi}_j + \bar{\rho}_j \delta\bar{\zeta}_j (L^A)^{-1}\bar{\chi}_j) \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \left(\frac{\delta\rho(\xi)\Psi(\xi)}{\xi + i\epsilon} - \frac{\delta\bar{\rho}(\xi)\bar{\Psi}(\xi)}{\xi - i\epsilon} \right) d\xi - 2i\sum_{j=1}^{N} \left(\delta\left(\frac{\rho_j}{\zeta_j}\right)\Psi_j + \frac{\rho_j}{\zeta_j}\delta\zeta_j\chi_j + \delta\left(\frac{\bar{\rho}_j}{\bar{\zeta}_j}\right)\bar{\Psi}_j + \frac{\bar{\rho}_j}{\bar{\zeta}_j}\delta\bar{\zeta}_j\bar{\chi}_j \right), \end{aligned}$$

where we have moved the integral over the continuous spectrum off the real axis to avoid the singularity at $\zeta = 0$. Therefore the only difference between the computation of Ω_1 and the new symplectic form Ω_2 are the weighting factors $1/\xi$ in the continuous spectrum, and $1/\zeta_j$ in the discrete spectrum. A similar calculation as was used to produce (17) now gives

$$\Omega_2 = \frac{i}{\pi}\int_{-\infty}^{\infty} \{\delta \log[\bar{a}(\xi)a(\xi)] \wedge \delta \arg b(\xi)\}\frac{d\xi}{\xi} + \frac{1}{2}\bar{b}(0)b(0)\delta \log \frac{\bar{a}(0)}{a(0)} \wedge \delta \log \frac{\bar{b}(0)}{b(0)} - 2\sum_{j=1}^{N} \{\delta \log \zeta_j \wedge \delta \log b_j + \delta \log \bar{\zeta}_j \wedge \delta \log \bar{b}_j\}, \tag{21}$$

where the two complex integrals have combined to give the principal value in the leading term, and extra discrete term comes from the associated residues at the pole $\zeta = 0$. When $r = \pm q^*$, canonically conjugate variables are provided by the momenta

$$\hat{A}_j = 4 \arg \zeta_j, \quad \hat{p}_j = -4 \log|\zeta_j|,$$

$$\hat{p}(\xi) = -(i/\pi\xi) \log|a(\xi)|,$$

and the conjugate coordinates

$$\hat{B}_j = \arg b_j, \quad \hat{q}_j = \log|b_j|, \quad \hat{q}(\xi) = \arg b(\xi),$$

provided $\xi \neq 0$. In addition, the point $\xi = 0$ appears separately as the extra residue term in the expression for Ω_2, so this particular mode survives the principal value cancellation in a new discrete form. However, there is no simple formula for the relevant canonical variables there. Also, in the case $r = \pm q^*$, this term vanishes because $\bar{a}(0) = a(0)$, and so this extra complication does not arise. All the other modes for the continuous spectrum are related according to the simple reweighting

$$p(\xi) = \xi\hat{p}(\xi). \tag{22}$$

For the second Hamiltonian structure, the Hamiltonian functional giving the nonlinear Schrödinger equation is the momentum (11). According to the calculations in Ref. 1, it can be expressed in terms of the scattering data as

$$H_2 = P = \frac{4}{\pi}\int_{-\infty}^{\infty} \xi \log|a(\xi)|d\xi - 4i\sum_{j=1}^{N} (\bar{\zeta}_j^2 - \zeta_j^2). \tag{23}$$

Comparing with (18), we see that, in terms of the respective canonical variables, the continuous spectrum contribution is exactly the same weighted sum of the continuous canonical momentum variable associated with the respective symplectic two forms:

$$H_1: \ \frac{4}{\pi}\int_{-\infty}^{\infty} \xi^2 p(\xi)d\xi \quad \text{versus}$$

$$H_2: \ \frac{4}{\pi}\int_{-\infty}^{\infty} \xi p(\xi)d\xi = \frac{4}{\pi}\int_{-\infty}^{\infty} \xi^2 \hat{p}(\xi)d\xi.$$

Therefore, the continuous modes have identical quantizations. (The singular point $\xi = 0$ plays no role as both Hamiltonians make no contribution to this mode.) As for the bound states, we are reduced to the case of a collection of integrable two-dimensional Hamiltonian systems with different Hamiltonian structures. For the original symplectic form Ω_1, the Hamiltonian system corresponding to the discrete eigenvalue ζ_j has the form

$$(\log b_j)_t = -\frac{1}{2}\frac{\partial H_1}{\partial \zeta_j} = 4i\zeta_j^2, \quad (\zeta_j)_t = \frac{1}{2}\frac{\partial H_1}{\partial \log b_j} = 0,$$

and similarly for the eigenvalues $\bar{\zeta}_j$. (We are just reproducing the classical calculation of the evolution of the discrete scattering data for soliton equations.) For the second symplectic form Ω_2, the Hamiltonian system corresponding to the discrete eigenvalue ζ_j now takes the form

$$(\log b_j)_t = -\frac{1}{2}\frac{\partial H_1}{\partial \log \zeta_j} = 4i\zeta_j^2,$$

$$(\log \zeta_j)_t = \frac{1}{2}\frac{\partial H_1}{\partial \log b_j} = 0,$$

and similarly for the eigenvalues $\bar{\zeta}_j$. Thus, these two dimensional Hamiltonian systems are identical, even though they use two different Hamiltonian structures:

$$-2\delta\zeta_j \wedge \delta \log b_j \quad \text{versus} \quad -2\delta \log \zeta_j \wedge \delta \log b_j.$$

However, as we remarked above, we take as fundamental the fact that a two-dimensional Hamiltonian system has a unique quantization, even though it has many different Hamiltonian structures. Therefore the bound states for the nonlinear Schrödinger equation also have identical quantizations. We conclude that both Hamiltonians lead to the same quantized verison of the nonlinear Schrödinger equation.

As a final remark, we recall that the other soliton equations appearing in the AKNS scheme can be written in the form

$$\begin{pmatrix} q \\ r \end{pmatrix}_t = \Omega(L^A)\begin{pmatrix} q \\ r \end{pmatrix}_x,$$

where $\Omega(\xi)$ determines the linear dispersion relation.[7] These can all be written in bi-Hamiltonian form using the same two Hamiltonian structures as above. An identical calculation, which we omit for the sake of brevity, will show that the two quantized versions of any member of these AKNS hierarchies will lead to the same quantum version. Moreover, it is not hard to see that the only difference between the quantized versions of two different members of the same soliton hierarchy is in the weighting factor $\Omega(\xi)$ for the modes corresponding to the continuous spectrum [with appropriate discrete contributions at the points where $\Omega(\xi) = 0$] and replacement of the discrete Hamiltonians by $\Omega(\zeta_j)$ and $\Omega(\bar{\zeta}_j)$, respectively. Thus the only distinction between the various quantized versions of a soliton hierarchy is in the weighting assigned to the continuous modes, and the replacement of the Hamiltonian governing the evolution of the bound states. Finally, we note that the same considerations will apply to other soliton equations, such as the Korteweg–de Vries equation, as the key fact that the recursion operator is the squared eigenfunction operator remains valid.

ACKNOWLEDGMENTS

We gratefully acknowledge the support and hospitality of the Institute for Mathematics and Its Applications

(I.M.A.) during the fall progam on Nonlinear Waves-Solitions in 1988.

One of us (D.J.K) is supported in part by NSF Grant DMS 88-03471 and AFOSR Grant 86-0277. One of us (P.J.O.) is supported in part by NSF Grant DMS 86-02004.

[1]D. J. Kaup, J. Math. Phys. **16**, 2036 (1975).
[2]L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1955), Chap. 14.
[3]E. Gutkin, Phys. Rep. **167**, 1 (1988).
[4]L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1955), Chaps. 2–6.
[5]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[6]P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer-Verlag, Berlin, 1986).
[7]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[8]V. V. Dodonov, V. I. Man'ko, and V. D. Skarzhinsky, Nuovo Cimento **B69**, 185 (1982).
[9]D. J. Kaup, J. Math. Anal. Appl. **54**, 849 (1976).
[10]D. J. Kaup and A. C. Newell, Adv. Math. **31**, 67 (1979).
[11]H. Goldstein, *Classical Mechanics* (Addison-Wesley, New York, 1959).

The Yang–Baxter equations and differential identities

Fu-Cho Pu
International Centre for Theoretical Physics, 34100 Trieste, Italy

D. H. Sattinger
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 17 May 1989; accepted for publication 6 September 1989)

The solution of the Yang–Baxter equation for integrable systems is shown to be equivalent to the existence of a differential identity. Quantum integration formulas for the calculation of commutators of monodromy matrices are given. Based on the integration formulas and the systematic use of differential identities, the Yang–Baxter equations for the nonlinear Schrödinger model for the quantum case of both bosons and fermions are derived. The case for discrete models is also included. The parallelism between the classical and quantum case and the classical limiting process from the latter to the former are discussed.

I. INTRODUCTION

The calculation of Poisson brackets for the scattering data for the KdV equation by the CISM (classical inverse scattering method) was first carried out by Zakharov and Faddeev.[1] The calculations depended on a set of differential identities and on the evaluation of certain limits in the sense of distributions. After the formulation of the QISM (quantum inverse scattering method) by Faddeev and his collaborators,[2] in which the solution of the Yang–Baxter equation is the important step of the problem, the CISM was recast in terms of the solution of a classical Yang–Baxter equation.[3] In this paper, we first repeat the calculation of Poisson brackets for monodromy matrices of the NS (nonlinear Schrödinger) model by the use of an integration formula[3] and differential identity, and then generalize the formula to the quantum cases to calculate commutators and obtain the Yang–Baxter equations of the NS model for both bosons and fermions by carrying out the integration with the use of corresponding differential identities. We present a simple proof that the Yang–Baxter equation is equivalent to a differential identity. These differential identities can be chosen to be the same as those that have been used to derive Yang–Baxter equations by our generalized formulas. Further, the differential identity is shown to be equivalent to the local Yang–Baxter equation. We do not claim that the establishment of differential identities is always simpler than the algebraic solutions of local Yang–Baxter equations but aim to clarify the relations between them in various contexts. Finally, we show how the classical results are obtained by the classical limit from the quantum ones. The corresponding results for discrete models are similar where the differential identities are replaced by difference identities. We discuss the discrete model of the Heisenberg chain as an illustration.

II. POISSON BRACKETS FOR CLASSICAL SYSTEMS

In this section we recapitulate the calculation of the Poisson brackets by the use of differential identities. We use the following integration formula for the Poisson bracket of monodromy matrices (see Ref. 3, p. 192):

$$\{T(x,y|\lambda)\otimes T(x,y|\mu)\} = \int_y^x \int_y^x dz\,dz'\; T(x,z|\lambda)\otimes T(x,z'|\mu) \times\{U(z,\lambda)\otimes U(z',\mu)\}T(z,y|\lambda)\otimes T(z',y|\mu), \quad (2.1)$$

where $T(x,y|\lambda)$ is the monodromy matrix defined by

$$\frac{\partial}{\partial x}T(x,y|\lambda) = U(x,\lambda)T(x,y|\lambda), \qquad T(x,x|\lambda) = 1. \quad (2.2)$$

For the NS model, the Hamiltonian takes the form

$$H = \int dx\left(\frac{\partial u^*(x)}{\partial x}\frac{\partial u(x)}{\partial x} + cu^*(x)u^*(x)u(x)u(x)\right), \quad (2.3)$$

where c, the coupling constant, is positive for repulsive interaction, and

$$U(x,\lambda) = (i/2)\lambda\sigma_3 + i\sqrt{c}u(x)\sigma_+ - i\sqrt{c}u^*(x)\sigma_-. \quad (2.4)$$

Here $\sigma_\pm = \frac{1}{2}(\sigma_1 \pm i\sigma_2)$, where $\sigma_i (i = 1,2,3)$ are the Pauli matrices.

Let F and G be (2×2) matrix valued functionals of u and u^*. The Poisson bracket of F and G is (see Ref. 3, p. 187)

$$\{F\otimes G\} = i\int_{-\infty}^{\infty}\left(\frac{\delta F}{\delta u(x)}\otimes\frac{\delta G}{\delta u^*(x)} - \frac{\delta F}{\delta u^*(x)}\otimes\frac{\delta G}{\delta u(x)}\right)dx.$$

Considering $U(x,\lambda)$ as a functional of $u(z)$ and $u^*(z)$, we have

$$\frac{\delta U(x,\lambda)}{\delta u(z)} = i\sqrt{c}\delta(x-z)\sigma_+, \qquad \frac{\delta U(x,\lambda)}{\delta u^*(z)} = -i\sqrt{c}\delta(x-z)\sigma_-.$$

Hence

 0022-2488/90/010118-09$03.00

$$\{U(x,\lambda)\otimes U(y,\mu)\} = i\int dz\Big(\frac{\delta U(x,\lambda)}{\delta u(z)} \otimes \frac{\delta U(y,\mu)}{\delta u^*(z)} - \frac{\delta U(x,\lambda)}{\delta u^*(z)}\otimes\frac{\delta U(y,\mu)}{\delta u(z)}\Big)$$
$$= ic\delta(x-y)(\sigma_+\otimes\sigma_- - \sigma_-\otimes\sigma_+)$$
$$= (ic/2)(\sigma_3\otimes I - I\otimes\sigma_3)P\delta(x-y), \tag{2.5}$$

where the identity

$$(\sigma_+\otimes\sigma_- - \sigma_-\otimes\sigma_+) = \tfrac{1}{2}(\sigma_3\otimes I - I\otimes\sigma_3)P \tag{2.6}$$

has been used. In (2.6) the symbol P stands for the permutation matrix, equal to

$$P = \frac{1}{2}\Big(I\otimes I + \sum_{i=1}^{3}\sigma_i\otimes\sigma_i\Big), \tag{2.7}$$

where I is the 2×2 identity matrix. The matrix P acts on $V\otimes V$ by $(\xi\otimes\eta)P = \eta\otimes\xi$ and on matrices by $P(A\otimes B) = (B\otimes A)P$. Clearly, $P^2 = I$ and

$$[(A\otimes B)P]_{ij,kl} = (A\otimes B)_{ij,lk},$$
$$[P(A\otimes B)]_{ij,kl} = (A\otimes B)_{ji,kl}.$$

Substituting (2.5) into (2.1), we obtain

$$\{T(x,y|\lambda)\otimes T(x,y|\mu)\}$$
$$= \frac{ic}{2}\int_y^x dz\big(T(x,z|\lambda)\sigma_3 T(z,y|\mu)\otimes T(x,z|\mu)T(z,y|\lambda)$$
$$- T(x,z|\lambda)T(z,y|\mu)\otimes T(x,z|\mu)\sigma_3 T(z,y|\lambda)\big)P.$$

Performing the integration with the help of the differential identity

$$\frac{\partial}{\partial z}\big(T(x,z|\lambda)T(z,y|\mu)\otimes T(x,z|\mu)T(z,y|\lambda)\big)$$
$$= -(i/2)(\lambda-\mu)$$
$$\times\big(T(x,z|\lambda)\sigma_3 T(z,y|\mu)\otimes T(x,z|\mu)T(z,y|\lambda)$$
$$- T(x,z|\lambda)T(z,y|\mu)\otimes T(x,z|\mu)\sigma_3 T(z,y|\lambda)\big), \tag{2.8}$$

we finally obtain[3]

$$\{T(x,y|\lambda)\otimes T(x,y|\mu)\} = [r(\lambda,\mu), T(x,y|\lambda)\otimes T(x,y|\mu)],$$

where

$$r(\lambda,\mu) = -cP/(\lambda-\mu). \tag{2.9}$$

III. THE QUANTUM CASE

In Sec. II, we showed that the classical Yang–Baxter equation for the monodromy matrix can be obtained from the differential equation for the monodromy matrix by the use of a certain differential identity. This is also true for the quantum case. We demonstrate this point for the NS model for bosons. The Hamiltonian in this case is again (2.3), but now u, u^* are operators acting on Fock space satisfying the commutation relations

$$[u(x),u^*(y)] = \delta(x-y),$$
$$[u(x),u(y)] = [u^*(x),u^*(y)] = 0. \tag{3.1}$$

The monodromy matrix $T(x,y|\lambda)$ satisfies the forward and backward differential equations

$$\frac{\partial}{\partial x}T(x,y|\lambda) = :U(x,\lambda)T(x,y|\lambda):,$$
$$\frac{\partial}{\partial y}T(x,y|\lambda) = -:T(x,y|\lambda)U(y,\lambda):, \tag{3.2}$$
$$T(x,x|\lambda) = 1,$$

where $U(x,\lambda)$ is given by (2.4).

The notation :A: means that an operator A, depending on the creation and annihilation operators u^* and u, is to be written in *normal order*, meaning that the destruction operators u are to appear all the way to the right. For example, u^*u^2 is in normal order, while $u\,u^*u$ is not. The reasons for writing expressions in normal order are subtle. (In particular, we found that the order in which the operators appear in an expression can have an effect on the asymptotic behavior of its expectations as $x\to\pm\infty$.)

The following commutation relations hold[4,5] ($y<x$):

$$[u(z),T(x,y|\lambda)] = [u^*(z),T(x,y|\lambda)] = 0, \quad z\notin[y,x],$$
$$[u(x),T(x,y|\lambda)] = -i(\sqrt{c}/2)\sigma_- T(x,y|\lambda),$$
$$[u(y),T(x,y|\lambda)] = -i(\sqrt{c}/2)T(x,y|\lambda)\sigma_-,$$
$$[u^*(x),T(x,y|\lambda)] = -i(\sqrt{c}/2)\sigma_+ T(x,y|\lambda),$$
$$[u^*(y),T(x,y|\lambda)] = -i(\sqrt{c}/2)T(x,y|\lambda)\sigma_+,$$
$$[u(z),T(x,y,\lambda)] = -i\sqrt{c}T(x,z|\lambda)\sigma_- T(z,y|\lambda), \quad y<z<x. \tag{3.3}$$

With the help of (3.3), Eqs. (3.2) can be rewritten in *standard order* as

$$\frac{\partial}{\partial x}T(x,y|\lambda) = V(x,\lambda)T(x,y|\lambda),$$
$$\frac{\partial}{\partial y}T(x,y|\lambda) = -T(x,y|\lambda)V(y,\lambda),$$

where

$$V(x,\lambda) = U(x,\lambda) - (c/2)\sigma_+\sigma_-. \tag{3.4}$$

We now generalize (2.1) to the quantum case. First, we define

$$[T(x,y|\lambda)\mathbin{\dot\otimes} T(x,y|\mu)]$$

to be a matrix with elements

$$[T(x,y\lambda)\mathbin{\dot\otimes} T(x,y|\mu)]_{ij,kl} = [T_{ik}(x,y|\lambda), T_{jl}(x,y|\mu)]. \tag{3.5}$$

It is easy to see that

$$[T(x,y|\lambda)\mathbin{\dot\otimes} T(x,y|\mu)]$$
$$= T(x,y|\lambda)\otimes T(x,y|\mu) - PT(x,y|\mu)\otimes T(x,y|\lambda)P. \tag{3.6}$$

We state our generalized formula as the following theorem.

Theorem 3.1:

$$[T(x,y|\lambda) \mathbin{\dot{\otimes}} T(x,y|\mu)]$$
$$= \int_y^x \int_y^x dz\, dz'\, T(x,z|\lambda) \otimes T(x,z'|\mu)$$
$$\times [V(z,\lambda) \mathbin{\dot{\otimes}} V(z',\mu)]$$
$$\times PT(z',y|\mu) \otimes T(z,y|\lambda)P, \tag{3.7}$$

where

$$\frac{\partial}{\partial x} T(x,y|\lambda) = :U(x,\lambda)T(x,y|\lambda): = V(x,\lambda)T(x,y|\lambda).$$

Proof: The $(\alpha\beta,\gamma\delta)$ component of the right-hand side of (3.7) is

$$\int_y^x \int_y^x dz\, dz'\, T_{\alpha\rho}(x,z|\lambda) T_{\beta\sigma}(x,z'|\mu)$$
$$\times (V_{\rho\vartheta}(z,\lambda) V_{\sigma\phi}(z',\mu) - V_{\sigma\phi}(z',\mu)$$
$$\times V_{\rho\vartheta}(z,\lambda)) T_{\phi\delta}(z',y|\mu) T_{\vartheta\gamma}(z,y|\lambda),$$

where, as usual, repeated indices mean summation. Using the forward and backward equations in component form

$$V_{\sigma\phi}(z',\mu) T_{\phi\delta}(z',y|\mu) = \frac{\partial}{2z'} T_{\sigma\delta}(z',y|\mu),$$
$$- T_{\beta\sigma}(x,z'|\mu) V_{\sigma\phi}(z',\mu) = \frac{\partial}{\partial z'} T_{\beta\phi}(x,z'|\mu),$$

we can write this as

$$\int_y^x dz\, T_{\alpha\rho}(x,z|\lambda) \left[\int_y^x dz' \frac{\partial}{\partial z'} (T_{\beta\sigma}(x,z'|\mu) \right.$$
$$\left. \times V_{\rho\vartheta}(z,\lambda) T_{\sigma\delta}(z',y|\mu)) \right] T_{\vartheta\gamma}(z,y|\lambda).$$

Integrating with respect to z' and using the fact that $T_{\beta\sigma}(x,x|\mu) = \delta_{\beta\sigma}$, for example, we get

$$\int_y^x dz [T_{\alpha\rho}(x,z|\lambda) V_{\rho\vartheta}(z,\lambda) T_{\beta\delta}(x,y|\mu) T_{\vartheta\delta}(z,y|\lambda)$$
$$- T_{\alpha\rho}(x,z|\lambda) T_{\beta\delta}(x,y|\mu) V_{\rho\vartheta}(z,\lambda) T_{\vartheta\gamma}(z,y|\lambda)].$$

Similarly, by the use of

$$T_{\alpha\rho}(x,z|x) V_{\rho\vartheta}(z,\lambda) = -\frac{\partial}{\partial z} T_{\alpha\vartheta}(x,z|\lambda),$$
$$V_{\rho\vartheta}(z,\lambda) T_{\vartheta\gamma}(z,y|\lambda) = \frac{\partial}{\partial z} T_{\rho\gamma}(z,y|\lambda),$$

the integration of z is carried out. Finally, we obtain

$$- T_{\beta\delta}(x,y|\mu) T_{\alpha\gamma}(x,y|\lambda) + T_{\alpha\gamma}(x,y|\lambda) T_{\beta\delta}(x,y|\mu),$$

which is just $[T(x,y|\lambda) \otimes T(x,y|\mu)]_{\alpha\beta,\gamma\delta}$; thus, by definition, the theorem is proved.

By a straightforward calculation, we have

$$[V(z,\lambda) \mathbin{\dot{\otimes}} V(z',\mu)]$$
$$= [U(z,\lambda) \mathbin{\dot{\otimes}} U(z',\mu)]$$
$$= c(\sigma_+ \otimes \sigma_- - \sigma_- \otimes \sigma_+)\delta(z-z').$$

Substituting this into Theorem 3.1 and using (2.6), we obtain

$$[T(x,y|\lambda) \mathbin{\dot{\otimes}} T(x,y|\mu)]$$
$$= \frac{c}{2} \int_y^x dz\, T(x,z|\lambda) \otimes T(x,z|\mu)(\sigma_3 \otimes I - I \otimes \sigma_3)$$
$$\times T(z,y|\mu) \otimes T(z,y|\lambda)P.$$

Since $[T(x,z|\mu), T(z,y|\mu)] = 0$, the right-hand side can be written as

$$\frac{c}{2} \int_y^x dz (T(x,z|\lambda)\sigma_3 T(z,y|\mu) \otimes T(x,z|\mu) T(z,y|\lambda)$$
$$- T(x,z|\lambda) T(z,y|\mu) \otimes T(x,z|\mu)\sigma_3 T(z,y|\lambda))P.$$

Because of (3.4), the differential identity (2.8) remains valid in this case for bosons. So the integral is calculated using (2.8). In this way, we obtain

$$[T(x,y|\lambda) \mathbin{\dot{\otimes}} T(x,y|\mu)]$$
$$= [ic/(\lambda-\mu)](T(x,y|\mu) \otimes T(x,y|\lambda)$$
$$- T(x,y|\lambda) \otimes T(x,y|\mu))P. \tag{3.8}$$

Using (3.6), Eq. (3.8) may be rewritten in the form

$$R(\lambda,\mu) T(x,y|\lambda) \otimes T(x,y|\mu)$$
$$= T(x,y|\mu) \otimes T(x,y|\lambda) R(\lambda,\mu), \tag{3.9}$$

where

$$R(\lambda,\mu) = [-ic/(\lambda-\mu-ic)]I$$
$$+ [(\lambda-\mu)/(\lambda-\mu-ic)]P.$$

In fact, from (3.2) and (3.8), we have

$$T(x,y|\lambda) \otimes T(x,y|\mu) - PT(x,y|\mu) \otimes T(x,y|\lambda)P$$
$$= [ic/(\lambda-\mu)](T(x,y|\mu) \otimes T(x,y|\lambda)$$
$$- T(x,y|\lambda) \otimes T(x,y|\mu))P.$$

Multiplying on the right by P we obtain

$$T(x,y|\lambda) \otimes T(x,y|\mu)P - PT(x,y|\mu) \otimes T(x,y|\lambda)$$
$$= [ic/(\lambda-\mu)](T(x,y|\mu) \otimes T(x,y|\lambda)$$
$$- T(x,y|\lambda) \otimes T(x,y|\mu));$$

hence

$$[P + ic/(\lambda-\mu)] T(x,y|\mu) \otimes T(x,y|\lambda)$$
$$= T(x,y|\lambda) \otimes T(x,y|\mu)[P + ic/(\lambda-\mu)].$$

Equation (3.9) is now obtained by interchanging λ and μ and multiplying by the factor $(\lambda-\mu)/(\lambda-\mu-ic)$. This is the result of Refs. 2 and 4.

For the NS model of fermions,[5] the Hamiltonian takes the form

$$H = \int dx \left(\frac{\partial u_i^*}{\partial x} \frac{\partial u_i}{\partial x} + c :u_i^* u_i u_j^* u_j: \right). \tag{3.10}$$

Here we again consider only the case of repulsive interaction $c > 0$. The dummy indices i,j are now summed over 1,2; u_i, u_i^* act on Fock space and satisfy the anticommutation relations

$$[u_i(x),u_j^*(y)]_+ = \delta_{ij}\delta(x-y),$$
$$[u_i(x),u_j(y)]_+ = [u_i^*(x),u_j^*(y)]_+ = 0. \tag{3.11}$$

The monodromy matrix $T(x,y|\lambda)$ is a 3×3 matrix defined by

$$\frac{\partial}{\partial x}T(x,y|\lambda) = :U(x,\lambda)T(x,y|\lambda):,$$
$$T(x,x|\lambda) = 1, \tag{3.12}$$

where

$$U(x,\lambda) = (i\lambda/2)J + i\sqrt{c}u_j(x)e_{j3} - i\sqrt{c}u_j^*(x)e_{3j}, \tag{3.13}$$

$J = \mathrm{diag}(1,1,-1)$, and e_{jk} is a 3×3 matrix defined by $(e_{jk})_{lm} = \delta_{jl}\delta_{km}$.

For fermion fields, the normal product is

$$:u_i(x)u_j^*(y): = -u_j^*(y)u_i(x),$$

and for the monodromy matrix we get as a result[6]

$$:u_j(x)T(x,y|\lambda): = JT(x,y|\lambda)Ju_j(x).$$

Therefore Eq. (3.12) can be written

$$\frac{\partial T}{\partial x}(x,y|\lambda) = \left(\frac{i}{2}\lambda J - i\sqrt{c}u_j^*(x)e_{3j}\right)T(x,y|\lambda) + i\sqrt{c}e_{j3}JT(x,y|\lambda)Ju_j(x).$$

The following relations hold[6] ($y<x$), for $j=1,2$:

$$u_j(z)T(x,y|\lambda) = JT(x,y|\lambda)Ju_j(z), \quad z\notin[x,y],$$
$$T(x,y|\lambda)u_j^*(z) = u_j^*(z)JT(x,y|\lambda)J, \quad z\notin[x,y],$$
$$u_j(x)T(x,y|\lambda) = JT(x,y|\lambda)Ju_j(x) - (i\sqrt{c}/2)e_{3j}T(x,y|\lambda),$$
$$u_j(y)T(x,y|\lambda) = JT(x,y|\lambda)Ju_j(y) - (i\sqrt{c}/2)JT(x,y|\lambda)Je_{3j},$$
$$T(x,y|\lambda)u_j^*(x) = u_j^*(x)JT(x,y|\lambda)J + (i\sqrt{c}/2)JT(x,y|\lambda)Je_{j3},$$
$$T(x,y|\lambda)u_j^*(y) = u_j^*(y)JT(x,y|\lambda)J + (i\sqrt{c}/2)T(x,y|\lambda)e_{j3}. \tag{3.14}$$

From (3.12) and the third equation in (3.14), we find

$$\frac{\partial}{\partial x}T(x,y|\lambda) = V(x,\lambda)T(x,y|\lambda)$$

where

$$V(x,\lambda) = U(x,\lambda) - (c/2)e_{j3}e_{3j}. \tag{3.15}$$

Similarly,

$$\frac{\partial}{\partial y}T(x,y|\lambda) = -T(x,y|\lambda)V(y,\lambda).$$

In dealing with problems involving fermions, it is necessary to introduce *supermatrices*.[5] For a supermatrix we assign a parity $p(i)$ of the ith row (column) $=0$ when the parity is even, and $p(i)=1$ when the parity is odd. The supertensor product of two supermatrices A and B is defined by

$$(A \underset{s}{\otimes} B)_{ij,kl} = A_{ik}B_{jl}(-1)^{p(j)(p(i)+p(k))}.$$

We define $\left[A \underset{s}{\otimes} B\right]$ to be the matrix with elements given by

$$\left[A \underset{s}{\otimes} B\right]_{ij,kl} = A_{ik}B_{jl}(-1)^{p(j)(p(i)+p(k))} - B_{jl}A_{ik}(-1)^{p(l)(p(i)+p(k))}.$$

It is easy to see that

$$\left[A \underset{s}{\otimes} B\right] = A \underset{s}{\otimes} B - P_s B \underset{s}{\otimes} A P_s, \tag{3.16}$$

where $P_s = \Sigma e_{ij}\otimes e_{ji}(-1)^{p(i)p(j)}$ is the superpermutation operator. In analogy with Theorem 3.1, we have the following theorem.

Theorem 3.2:

$$\left[T(x,y|\lambda) \underset{s}{\otimes} T(x,y|\mu)\right] = \int_y^x\int_y^x dz\,dz'\; T(x,z|\lambda) \underset{s}{\otimes} T(x,z'|\mu) \times \left[U(z,\lambda) \underset{s}{\otimes} U(z',\mu)\right] P_s T(z',y|\mu) \underset{s}{\otimes} T(z,y|\lambda)P_s. \tag{3.17}$$

Proof: The proof is similar to that of Theorem 2.1 with consideration of the additional signs due to the parity.

We now return to the NS model for fermions, described by the Hamiltonian (3.10). For the present problem the parity should be chosen as

$$p(i) = \begin{cases} 0, & i=1,2, \\ 1, & i=3. \end{cases}$$

It follows from a straightforward calculation that

$$\left[V(z,\lambda) \underset{s}{\otimes} V(z',\mu)\right] = \left[U(z,\lambda) \underset{s}{\otimes} U(z',\mu)\right] = c\left(e_{3j} \underset{s}{\otimes} e_{j3} + e_{j3} \underset{s}{\otimes} e_{3j}\right)\delta(z-z').$$

Substituting this into (3.17) and taking into account the identity

$$(e_{3j} \underset{s}{\otimes} e_{j3} + e_{j3} \underset{s}{\otimes} e_{3j})P_s = \frac{1}{2}(I \underset{s}{\otimes} J - J \underset{s}{\otimes} I),$$

we obtain

$$\left[T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\right]=\frac{c}{2}\int_y^x dz\,T(x,z|\lambda)\underset{s}{\otimes}T(x,z|\mu)(I\underset{s}{\otimes}J-J\underset{s}{\otimes}I)T(z,y|\mu)\underset{s}{\otimes}T(z,y|\lambda)P_s. \tag{3.18}$$

For nonoverlapping intervals (x,z) and (z,y) $(x<z<y)$, $\left[T(x,z|\mu)\underset{s}{\otimes}T(z,y|\mu)\right]=0$ so (3.18) becomes

$$\frac{c}{2}\int_y^x dz\Big(T(x,z|\lambda)T(z,y|\mu)\underset{s}{\otimes}T(x,z|\mu)JT(z,y|\lambda)-T(x,z|\lambda)JT(z,y|\mu)\underset{s}{\otimes}T(x,z|\mu)T(z,y|\lambda)\Big)P_s. \tag{3.19}$$

Making use of the differential identity

$$\frac{\partial}{\partial z}\Big(T(x,z|\lambda)T(z,y|\mu)\underset{s}{\otimes}T(x,z|\mu)T(z,y|\lambda)\Big)$$
$$=\frac{i}{2}(\lambda-\mu)\Big(T(x,z|\lambda)T(z,y|\mu)\underset{s}{\otimes}T(x,z|\mu)JT(z,y|\lambda)-T(x,z|\lambda)JT(z,y|\mu)\underset{s}{\otimes}T(x,z|\mu)T(z,y|\lambda)\Big) \tag{3.20}$$

[(3.20) follows from (3.15)], the integral of (3.19) is reduced to

$$[-c/(\lambda-\mu)]\Big(T(x,y|\mu)\underset{s}{\otimes}T(x,y|\lambda)-T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\Big)P_s.$$

Thus, we have shown that

$$\left[T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\right]=[-ic/(\lambda-\mu)]\Big(T(x,y|\mu)\underset{s}{\otimes}T(x,y|\lambda)-T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\Big)P_s. \tag{3.21}$$

According to (3.16), the left-hand side of (3.21) is

$$T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)-P_sT(x,y|\mu)\underset{s}{\otimes}T(x,y|\lambda)P_s.$$

Therefore, it is easy to see that (3.17) contains the Yang–Baxter equation

$$R(\lambda,\mu)T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)=T(x,y|\mu)\underset{s}{\otimes}T(x,y|\lambda)R(\lambda,\mu),$$

where

$$R(\lambda,\mu)=[ic/(\lambda-\mu+ic)]I+[(\lambda-\mu)/(\lambda-\mu+ic)]P_s. \tag{3.22}$$

Here I denotes the 3×3 identity matrix. This is the result obtained in Ref. 5.

We point out that both Theorems 3.1 and 3.2 are valid in general for the quantum case, but that Theorem 3.1 is useful for the boson case, while Theorem 3.2 is useful for the fermion case. Further, Theorem 3.1 can be considered as a particular case of Theorem 3.2 when $p(i)=0$, for all i.

For a discrete model on a one-dimensional lattice, define

$$T(i,j|\lambda)=L_{i-1}(\lambda)\cdots L_j(\lambda),\quad j\leqslant i-1,\quad T(i,i|\lambda)=1, \tag{3.23}$$

where $L_j(\lambda)$ is the associated linear operator of the given model. If we define the difference operator Δ_i by $\Delta_i f(i)=f(i+1)-f(i)$, then the forward and backward difference equations are, respectively,

$$\Delta_i T(i,j|\lambda)=U_i(\lambda)T(i,j|\lambda),\quad \Delta_j T(i,j|\lambda)=-T(i,j+1)U_j(\lambda),\quad \text{where } U_i(\lambda)=L_i(\lambda)-1. \tag{3.24}$$

The analogs of Theorems 3.1 and 3.2 are as follows.

Theorem 3.3:

$$\left[T(i,j|\lambda)\underset{s}{\otimes}T(i,j|\mu)\right]=\sum_{k=j}^{i-1}\sum_{k'=j}^{i-1}T(i,k+1|\lambda)\otimes T(i,k'+1|\mu)\left[L_k(\lambda)\underset{s}{\otimes}L_{k'}(\mu)\right]P\,T(k',j|\mu)\otimes T(k,j|\lambda)P. \tag{3.25}$$

Theorem 3.4:

$$\left[T(i,j|\lambda)\underset{s}{\otimes}T(i,j|\mu)\right]=\sum_{k=j}^{i-1}\sum_{k'=j}^{i-1}T(i,k+1|\lambda)\underset{s}{\otimes}T(i,k'+1|\mu)\left[L_k(\lambda)\underset{s}{\otimes}L_{k'}(\mu)\right]P_sT(k',j|\mu)\underset{s}{\otimes}T(k,j|\lambda)P_s. \tag{3.26}$$

Since Theorem 3.3 can be considered as a particular case of Theorem 3.4 when all parities are even, it is sufficient to give the proof for Theorem 3.4.

Take the $(\alpha\beta,\gamma\delta)$ component of the right-hand side of (3.26):

$$(\text{rhs})_{\alpha\beta,\gamma\delta} = \sum_{k=j}^{i-1}\sum_{k'=j}^{i-1} T_{\alpha\rho}(i,k+1|\lambda)T_{\beta\sigma}(i,k'+1|\mu)(-1)^{p(\beta)(p(\alpha)+p(\rho))}$$
$$\times\{L_k(\lambda)_{\rho\vartheta}L_{k'}(\mu)_{\sigma\phi}(-1)^{p(\sigma)(p(\rho)+p(\vartheta))} - L_{k'}(\mu)_{\sigma\phi}L_k(\lambda)_{\rho\vartheta}(-1)^{p(\phi)(p(\rho)+p(\vartheta))}\}$$
$$\times T_{\phi\delta}(k',j|\mu)T_{\vartheta\gamma}(k,j|\lambda)(-1)^{p(\delta)(p(\vartheta)+p(\gamma))}.$$

Notice that

$$T_{\beta\sigma}(i,k'+1|\mu)L_{k'}(\mu)_{\sigma\phi} = T_{\beta\phi}(i,k'|\mu), \quad L_{k'}(\mu)_{\sigma\phi}T_{\phi\delta}(k',j|\mu) = T_{\sigma\delta}(k'+1,j\mu), \tag{3.27}$$

and perform the summation of k' by the use of

$$\sum_{k'=j}^{i-1}(f(k'+1)-f(k')) = f(i)-f(j). \tag{3.28}$$

We obtain

$$(\text{rhs})_{\alpha\beta,\gamma\delta} = \sum_{k=j}^{i-1}\{T_{\alpha\rho}(i,k+1|\lambda)L_k(\lambda)_{\rho\vartheta}T_{\beta\delta}(i,j|\mu)T_{\vartheta\gamma}(k,j|\lambda)(-1)^{p(\beta)(p(\alpha)+p(\vartheta))+p(\delta)(p(\vartheta)+p(\gamma))}$$
$$- T_{\alpha\rho}(i,k+1|\lambda)T_{\beta\delta}(i,j|\mu)L_k(\lambda)_{\rho\vartheta}T_{\vartheta\gamma}(k,j|\lambda)(-1)^{p(\delta)(p(\gamma)+p(\rho))+p(\beta)(p(\alpha)+p(\rho))}\}.$$

Repeating the process for the summation on k, we arrive at

$$(\text{rhs})_{\alpha\beta,\gamma\delta} = T_{\alpha\gamma}(i,j|\lambda)T_{\beta\delta}(i,j|\mu)(-1)^{p(\beta)(p(\alpha)+p(\gamma))} - T_{\beta\gamma}(i,j|\mu)T_{\alpha\gamma}(i,j|\lambda)(-1)^{p(\delta)(p(\alpha))}$$
$$= \left[T(i,j|\lambda)\underset{s}{\otimes}T(i,j|\mu)\right]_{\alpha\beta,\gamma\delta}$$
$$= (\text{lhs})_{\alpha\beta,\gamma\delta}.$$

As an example we consider the Heisenberg spin chain, for which $L_k(\lambda)$ is a 2×2 matrix given by[2]

$$L_k(\lambda) = \lambda I + i\sigma_\alpha S_k^\alpha,$$

where the S_k^α are spin operators satisfying the commutation relations

$$[S_k^\alpha, S_{k'}^\beta] = i\delta_{kk'}\epsilon_{\alpha\beta\gamma}S_k^\gamma.$$

By a straightforward calculation, we obtain

$$\left[L_k(\lambda)\underset{s}{\otimes}L_{k'}(\mu)\right] = -i\delta_{kk'}\epsilon_{\alpha\beta\gamma}S_k^\gamma\sigma^\alpha\otimes\sigma^\beta. \tag{3.29}$$

Substituting (3.29) into (3.25), we have

$$\left[T(i,j|\lambda)\underset{s}{\otimes}T(i,j|\mu)\right] = -i\sum_{k=j}^{i-1}T(i,k+1|\lambda)\otimes T(i,k+1|\mu)\epsilon_{\alpha\beta\gamma}\sigma^\alpha\otimes\sigma^\beta PS_k^\gamma T(k,j|\mu)\otimes T(k,j|\lambda)P. \tag{3.30}$$

From the algebraic identity

$$\epsilon_{\alpha\beta\gamma}\sigma^\alpha\otimes\sigma^\beta P = i(\sigma^\gamma\otimes I - I\otimes\sigma^\gamma),$$

we obtain

$$\epsilon_{\alpha\beta\gamma}\sigma^\alpha\otimes\sigma^\beta S_k^\gamma P = L_k(\lambda)\otimes I - I\otimes L_k(\lambda). \tag{3.31}$$

Using (3.31) and the fact that $L_k(\lambda)$ commutes with $T(i,k+1|\mu)$ and $T(k,j|\mu)$, the right-hand side of (3.30) can be written as

$$-i\sum_{k=j}^{i-1}\{T(i,k+1|\lambda)L_k(\lambda)T(k,j|\mu)\otimes T(i,k+1|\mu)T(k,j|\lambda) - T(i,k+1|\lambda)T(k,j|\mu)\otimes T(i,k+1|\mu)L_k(\lambda)T(k,j|\lambda)\}P$$
$$= -i\sum_{k=j}^{i-1}\{T(i,k|\lambda)T(k,j|\mu)\otimes T(i,k+1|\mu)T(\kappa,j|\lambda) - T(i,k+1|\lambda)T(k,j|\mu)\otimes T(i,k+1|\mu)T(k+1,j|\lambda)\}P. \tag{3.32}$$

The next step is to use the difference identity

$$\Delta_k\{T(i,k|\lambda)T(k,j|\mu)\otimes T(i,k|\mu)T(k,j|\lambda)\}$$
$$= \{(\lambda-\mu)(T(i,k|\lambda)T(k,j|\mu)\otimes T(i,k+1|\mu)T(k,j|\lambda) - T(i,k+1|\lambda)T(k,j|\mu)\otimes T(i,k+1|\mu)T(k,+1,j|\lambda)\}$$

to carry out the summation. Finally we obtain

$$\left[T(i,j|\lambda)\otimes T(k,j|\mu)\right] = [-i/(\lambda-\mu)](T(i,j|\mu)\otimes T(i,j|\lambda) - T(i,j|\lambda)\otimes T(i,j|\mu))P,$$

which is equivalent to

$$R(\lambda,\mu)T(i,j|\lambda)\otimes T(i,j|\mu) = T(i,j|\mu)\otimes T(i,j|\lambda)R(\lambda,\mu),$$

with

$$R(\lambda,\mu) = [i/(\lambda-\mu+i)]I + [(\lambda-\mu)/(\lambda-\mu+i)]P.$$

This is the result obtained in Ref. 2.

IV. CONCLUDING REMARKS

In the preceding section, we derived the Yang–Baxter equations on the basis of integration (summation) formulas and with the help of certain differential (difference) identities. In fact, we may show that the Yang–Baxter equation is equivalent to a differential identity.

Theorem 4.1: The Yang–Baxter equation

$$R(\lambda,\mu)T(x,y|\lambda)\otimes T(x,y/\mu) = T(x,y|\mu)\otimes T(x,y|\lambda)R(\lambda,\mu) \tag{4.1}$$

is equivalent to the differential identity

$$\frac{\partial}{\partial z}(T(x,z|\mu)\otimes T(x,z|\lambda)R(\lambda,\mu)T(z,y|\lambda)\otimes T(z,y|\mu)) = 0. \tag{4.2}$$

Proof: Integrating this identity over the interval $y\leqslant z\leqslant x$, we get the Yang–Baxter equation. Conversely, if the Yang–Baxter equation holds for all x and y, then it certainly holds for $y = z$. Multiplying the Yang–Baxter equation for x,z on the right by $T(z,y|\lambda)\otimes T(z,y|\mu)$, we get

$$R(\lambda,\mu)(T(x,z|\lambda)\otimes T(x,z|\mu))(T(z,y|\lambda)\otimes T(z,y|\mu)) = T(x,z|\mu)\otimes T(x,z|\lambda)R(\lambda,\mu)T(z,y|\lambda)\otimes T(z,y|\mu).$$

Since $T(x,z|\mu)$ and $T(z,y|\lambda)$ depend on field operators on nonoverlapping intervals, and since the field operators commute on nonoverlapping intervals,

$$(T(x,z|\lambda)\otimes T(x,z|\mu))(T(z,y|\lambda)\otimes T(z,y|\mu)) = T(x,y|\lambda)\otimes T(x,y|\mu).$$

Therefore the left side above is independent of z, and the differentiation of this identity with respect to z proves the other half of Theorem 4.1.

The differential identity (4.2) for the NS model for bosons can be written

$$\frac{\partial}{\partial z}(PT(x,z|\mu)\otimes T(x,z|\lambda)P\,T(z,y|\lambda)\otimes T(z,y|\mu)) = \frac{ic}{\lambda-\mu}P\frac{\partial}{\partial z}(T(x,z|\mu)T(z,y|\lambda)\otimes T(x,z|\lambda)T(z,y|\mu)). \tag{4.3}$$

On the other hand, we have[4]

$$\frac{\partial}{\partial z}T(z,y|\lambda)\otimes T(z,y|\mu) = :\Gamma(z|\lambda,\mu)T(z,y|\lambda)\otimes T(z,y|\mu):,$$

$$\frac{\partial}{\partial z}T(x,z|\mu)\otimes T(x,z|\lambda) = -:T(x,z|\mu)\otimes T(x,z|\lambda)\Gamma(z|\mu,\lambda):,$$

where

$$\Gamma(z|\lambda,\mu) = U(z,\lambda)\otimes I + I\otimes U(z,\mu) + c\sigma_+\otimes\sigma_-.$$

The term $c\sigma_+\otimes\sigma_-$ is the quantum correction due to the noncommutativity of the field operators.

Using (2.6), Eq. (4.3) may be reduced to (2.8).

In general, if we write the forward differential equation for the tensor product of two monodromy matrices in normal order

$$\frac{\partial}{\partial z}(T(z,y|\lambda)\otimes T(z,y|\mu)) = :\Gamma(z|\lambda,\mu)T(z,y|\lambda)\otimes T(z,y|\mu):, \tag{4.4}$$

then the backward differential equation in normal order is

$$\frac{\partial}{\partial z}(T(x,z|\lambda)\otimes T(x,z|\mu)) = -:T(x,z|\lambda)\otimes T(x,z|\mu)\Gamma(z|\lambda,\mu):. \tag{4.5}$$

We have the following theorem which was proved in Ref. 4 by another method.

Theorem 4.2: The Yang–Baxter equation is equivalent to

$$R(\lambda,\mu)\Gamma(z|\lambda,\mu) = \Gamma(z|\mu,\lambda)R(\lambda,\mu). \tag{4.6}$$

Proof: From (4.2) and (4.4), we have

$$\frac{\partial}{\partial z}(T(x,z|\mu) \otimes T(x,z|\lambda)R(\lambda,\mu)T(z,y|\lambda) \otimes T(z,y|\mu))$$
$$= :T(x,z|\mu) \otimes T(x,z|\lambda)(-\Gamma(z|\mu,\lambda)R(\lambda,\mu)$$
$$+ R(\lambda,\mu)\Gamma(z|\lambda,\mu))T(z,y|\lambda) \otimes T(z,y|\mu):= 0.$$

The theorem follows from the conclusion of Theorem 4.1.

Similarly, if we write the differential equations for the tensor product of monodromy matrices in standard order,

$$\frac{\partial}{\partial z}(T(z,y|\lambda) \otimes T(z,y|\mu))$$
$$= \Lambda(z|\lambda,\mu)T(z,y|\lambda) \otimes T(z,y|\mu),$$
$$\frac{\partial}{\partial z}(T(x,z|\lambda) \otimes T(x,z|\mu)) \tag{4.7}$$
$$= -T(x,z|\lambda) \otimes T(x,z|\mu)\Lambda(z|\lambda,\mu),$$

then the Yang–Baxter equation is equivalent to

$$R(\lambda,\mu)\Lambda(z|\lambda,\mu) = \Lambda(z|\mu,\lambda)R(\lambda,\mu). \tag{4.8}$$

Analogously, for supermatrices with assigned parities, we have the following theorem.

Theorem 4.3: The Yang–Baxter equation

$$R(\lambda,\mu)T(x,y|\lambda) \underset{s}{\otimes} T(x,y|\mu)$$
$$= T(x,y|\mu) \underset{s}{\otimes} T(x,y|\lambda)R(\lambda,\mu) \tag{4.9}$$

is equivalent to the differential identity

$$\frac{\partial}{\partial z}(T(x,z|\mu) \underset{s}{\otimes} T(x,z|\lambda)R(\lambda,\mu)T(z,y|\lambda)$$
$$\underset{s}{\otimes} T(z,y|\mu)) = 0; \tag{4.10}$$

and the latter is equivalent to

$$R(\lambda,\mu)\Gamma(z|\lambda,\mu) = \Gamma(z|\mu,\lambda)R(\lambda,\mu) \tag{4.11}$$

or

$$R(\lambda,\mu)\Lambda(z|\lambda,\mu) = \Lambda(z|\mu,\lambda)R(\lambda,\mu) \tag{4.12}$$

provided that

$$\frac{\partial}{\partial z}\left(T(z,y|\lambda) \underset{s}{\otimes} T(z,y|\mu)\right)$$
$$= :\Gamma(z|\lambda,\mu)T(z,y|\lambda) \underset{s}{\otimes} T(z,y|\mu):$$
$$= \Lambda(z|\lambda,\mu)T(z,y|\lambda) \underset{s}{\otimes} T(z,y|\mu). \tag{4.13}$$

On account of (3.22), the differential identity (4.10) for the NS model of fermions with repulsive interaction can be written as

$$\frac{\partial}{\partial z}\left(P_s T(x,z|\mu) \underset{s}{\otimes} T(x,z|\lambda)P_s T(z,y|\lambda) \underset{s}{\otimes} T(z,y|\mu)\right)$$
$$= \frac{-ic}{\lambda-\mu}P_s \frac{\partial}{\partial z}\Big(T(x,z|\mu)T(z,y|\lambda)$$
$$\underset{s}{\otimes} T(x,z|\lambda)T(z,y|\mu)\Big). \tag{4.14}$$

From Ref. 6,

$$\frac{\partial}{\partial z}\left(T(z,y|\lambda) \underset{s}{\otimes} T(z,y|\mu)\right)$$
$$= :\Big(U(z,\lambda) \underset{s}{\otimes} I + I \underset{s}{\otimes} U(z,\mu) + ce_{j3} \underset{s}{\otimes} e_{3j}\Big)$$
$$\times T(z,y|\lambda) \underset{s}{\otimes} T(z,y|\mu):,$$
$$\frac{\partial}{\partial z}\Big(T(x,z|\mu) \underset{s}{\otimes} T(x,z|\lambda)\Big)$$
$$= -:T(x,z|\mu) \underset{s}{\otimes} T(x,z|\lambda)\Big(U(z,\mu) \underset{s}{\otimes} I$$
$$+ I \underset{s}{\otimes} U(z,\lambda) + ce_{j3} \underset{s}{\otimes} e_{3j}\Big):,$$

the differentiation on the left side of (4.14) can be completed. Then, (4.14) turns out to be just the differential identity (3.20).

For the discrete model on a lattice the corresponding theorem is as follows.

Theorem 4.4: The Yang–Baxter equation

$$R(\lambda,\mu)T(i,j|\lambda) \otimes T(i,j|\mu) = T(i,j\,|\mu) \otimes T(i,J\,|\lambda)R(\lambda,\mu) \tag{4.15}$$

is equivalent to the difference identity

$$\Delta_k(T(i,k\,|\mu) \otimes T(i,k\,|\lambda)R(\lambda,\mu)T(k,j|\lambda)$$
$$\otimes T(k,j|\mu)) = 0. \tag{4.16}$$

In the classical theory of inverse scattering transform the following theorem is true.

Theorem 4.5: The classical Yang–Baxter equation

$$\Big\{T(x,y|\lambda) \underset{s}{\otimes} T(x,y|\mu)\Big\}$$
$$= [\Gamma(\lambda,\mu), T(x,y|\lambda) \otimes T(x,y|\mu)] \tag{4.17}$$

is equivalent to

$$\frac{\partial}{\partial z}\Big(\Big\{T(x,z|\lambda) \underset{s}{\otimes} T(x,z|\mu)\Big\}T(z,y|\lambda) \otimes T(z,y|\mu)$$
$$+ T(x,z|\lambda) \otimes T(x,z|\mu)r(\lambda,\mu)T(z,y|\lambda)$$
$$\otimes T(z,y|\mu)\Big) = 0. \tag{4.18}$$

Proof: Integrating (4.18) from y to x we obtain (4.17). To prove the converse, set y equal to z in (4.17), multiply on the right by $T(z,y|\lambda) \otimes T(z,y|\mu)$, and differentiate with respect to z; this yields (4.18).

If the mondromy matrix $T(x,y|\lambda)$ is a functional of $u(x)$ and $u^*(x)$, as in the case of the NS model, and the Poisson bracket can be put in the form (cf. Ref. 3)

$$= \Big\{T(x,y|\lambda) \underset{s}{\otimes} T(x,y|\mu)\Big\}$$
$$= i\int_y^x \Big(\frac{\delta T(x,y|\lambda)}{\delta u(z)} \otimes \frac{\delta T(x,y|\mu)}{\delta u^*(z)}$$
$$- \frac{\delta T(x,y|\lambda)}{\delta u^*(z)} \otimes \frac{\delta T(x,y|\mu)}{\delta u(z)}\Big)dz,$$

then

$$\frac{\partial}{\partial z}\left(\left\{T(x,z|\lambda)\underset{\cdot}{\otimes}T(x,z|\mu)\right\}T(z,y|\lambda)\otimes T(z,y|\mu)\right)$$
$$=-i\left(\frac{\delta T(x,y|\lambda)}{\delta u(z)}\otimes\frac{\delta T(x,y|\mu)}{\delta u^*(z)}-\frac{\delta T(x,y|\lambda)}{\delta u^*(z)}\otimes\frac{\delta T(x,y|\mu)}{\delta u(z)}\right).$$

This follows from the fact that

$$\frac{\delta T(z,y|\lambda)}{\delta u(w)}=0,\quad \text{when } w\notin[y,z].$$

Therefore, (4.18) takes the form

$$\frac{\partial}{\partial z}(T(x,y|\lambda)\otimes T(x,y|\mu)r(\lambda,\mu)T(z,y|\lambda)\otimes T(z,y|\mu))$$
$$=i\left(\frac{\delta T(x,y|\lambda)}{\delta u(z)}\otimes\frac{\delta T(x,y|\mu)}{\delta u^*(z)}-\frac{\delta T(x,y|\lambda)}{\delta u^*(z)}\otimes\frac{\delta T(x,y|\mu)}{\delta u(z)}\right). \tag{4.19}$$

For the NS model with repulsive interaction we recall

$$r(\lambda,\mu)=-\frac{c}{\lambda-\mu}P,$$
$$\frac{\delta T(x,y|\lambda)}{\delta u(z)}=T(x,z|\lambda)\sigma_+T(z,y|\lambda),$$
$$\frac{\delta T(x,y|\lambda)}{\delta u^*(z)}=T(x,z|\lambda)\sigma_-T(z,y|\lambda).$$

Then (4.19) is just the same differential identity (2.8).

Finally, we mention that all formulas obtained in the classical theory can be derived from those in quantum theory by taking the classical limit $\hbar\to 0$,

$$\frac{-1}{i\hbar}\left[T(x,y|\lambda)\underset{\cdot}{\otimes}T(x,y|\mu)\right]\to\left\{T(x,y|\lambda)\underset{\cdot}{\otimes}T(x,y|\mu)\right\}. \tag{4.20}$$

For example, (2.1) can be obtained from Theorem 3.1 in this way. Likewise, by taking the limit $\hbar\to 0$,

$$\frac{-1}{i\hbar}\left[T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\right]\to\left\{T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\right\}, \tag{4.21}$$

we obtain, from (3.17),

$$\left\{T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\right\}$$
$$=\int_y^x\int_y^x dz\,dz'\,T(x,z|\lambda)\underset{s}{\otimes}T(x,z'|\mu)\times\left\{U(x,\lambda)\underset{s}{\otimes}U(z',\mu)\right\}T(z,y|\lambda)\underset{s}{\otimes}T(z',y|\mu) \tag{4.22}$$

where

$$\left\{T(x,y|\lambda)\underset{s}{\otimes}T(x,y|\mu)\right\}_{ij,kl}$$
$$=i\int_y^x dz\left(\frac{\delta T_{ik}(x,y|\lambda)}{\delta u(z)}\frac{\delta T_{jl}(x,y|\mu)}{\delta u^*(z)}(-1)^{p(j)(p(i)+p(k))}-\frac{\delta T_{jl}(x,y|\mu)}{\delta u(z)}\frac{\delta T_{ik}(x,y|\lambda)}{\delta u^*(z)}(-1)^{p(l)(p(i)+p(k))}\right).$$

This equation can also be derived directly by taking into account the parity of monodromy matrices in classical theory. The parities of monodromy matrices are defined in the same way for both the quantum and classical cases. In the classical case corresponding to fermion fields, the field operators are anticommutative and form a Grassmann algebra. The monodromy matrices are then supermatrices.

The anticommutators for fermion fields, with Planck's constant written explicitly, are

$$u_i(x)u_j^*(y)+u_j^*(y)u_i(x)=\hbar\delta_{ij}\delta(x-y),$$
$$u_i(x)u_j(y)+u_j(y)u_i(x)=0.$$

In the classical limit $\hbar\to 0$, we have

$$u_i(x)u_j^*(y)+u_j^*(y)u_i(x)=0,$$
$$u_i(x)u_j(y)+u_j(y)u_i(x)=0.$$

Thus in the classical limit the fermion fields are anticommutative, and the monodromy operators are supermatrices.

The differential identity (4.18) equivalent to the classical Yang–Baxter equation (4.17) can be obtained from the differential identity (4.10) equivalent to the quantum Yang–Baxter equation (4.9) by putting[2]

$$R(\lambda,\mu)=P(1+i\hbar r(\lambda,\mu)) \tag{4.23}$$

and taking the limit according to (4.20).

ACKNOWLEDGMENTS

One of the authors (F.-C. Pu) would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the ICTP, Trieste; Professor Avner Friedman for hospitality at the Institute for Mathematics and its Applications, Minneapolis, Minnesota; and Professor Larry McLerran of the Theoretical Physics Institute at the University of Minnesota.

[1]V. E. Zakharov and L. D. Faddeev, Funct. Anal. Appl. **5**, 280 (1971).
[2]L. D. Faddeev, Sov. Sci. Rev. Math. Phys. **CI**, 107 (1981); P. P. Kulish and E. K. Sklyanin *Lecture Notes in Physics*, Vol. 151 (Springer, Berlin, 1982).
[3]L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, New York, 1987).
[4]E. K. Sklyanin, LOMI Proceedings 55, 1980 (in Russian).
[5]F. C. Pu and B. H. Zhao, Phys. Rev. D **30**, 2253 (1984).
[6]F. C. Pu, Y. Z. Wu, and B. H. Zhao, J. Phys. A **20**, 1173 (1987).

Harmonic analysis of directing fields

Robert Alan Coleman and Herbert Korte
Interdisciplinary Studies in Mathematical Physics and Philosophy of Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

(Received 9 November 1988; accepted for publication 26 July 1989)

The structure of a directing field is determined by the projective structure of space-time and by various tensor (force) fields. Given a sufficient variety of such directing fields, which can be measured directly given only the ability to track material bodies with respect to an arbitrary coordinate system, it is shown how the projective and tensor fields involved can be determined (and hence measured). This method employs the technique of harmonic analysis on the forward unit hyperboloid. For the important and physically relevant case of an electromagnetic directing field, the projective structure and the electromagnetic field tensor can be determined using only *one* class of charged monopoles characterized by a given charge-to-mass ratio. The method also provides a new empirical criterion for determining whether or not a directing field is geodesic.

I. ACCELERATION AND DIRECTING FIELDS

Let M be an n-dimensional, C^∞ manifold. A *curve* element of order k at $p\in M$ is an equivalence class $j_0^k\gamma$ of curves through p that have the same Taylor expansion with respect to some (and hence every) coordinate chart $(U,x)_p$ up to and including order k at $0\in\mathbb{R}$. A *path element* of order k at $p\in M$ is an equivalence class of paths $j_p^k\xi$ consisting of all paths corresponding to curves in $j_0^k\gamma$, where $\gamma\in\xi$.

A second-order curve element $j_0^2\gamma$ has local coordinates γ_1^i and γ_2^i, called n-velocity and n-acceleration, respectively, and given by

$$\gamma_1^i=\frac{d}{d\lambda}x^i\circ\gamma(0),\quad \gamma_2^i=\frac{d^2}{d\lambda^2}x^i\circ\gamma(0). \tag{1}$$

A second-order path element $j_p^2\xi$ has local coordinates ξ_1^α and ξ_2^α, called $(n-1)$-velocity and $(n-1)$-acceleration, respectively, and given by

$$\xi_1^\alpha=\left.\frac{dx^\alpha\circ\gamma}{dx^n\circ\gamma}\right|_p,\quad \xi_2^\alpha=\left.\frac{d^2x^\alpha\circ\gamma}{(dx^n\circ\gamma)^2}\right|_p. \tag{2}$$

Under a change of coordinate chart from $(U,x)_p$ to $(\bar U,\bar x)_p$, the coordinates of $j_0^2\gamma$ transform according to

$$\bar\gamma_1^i=\bar X_j^i\gamma_1^j,\quad \bar\gamma_2^i=\bar X_j^i\gamma_2^j+\bar X_{jk}^i\gamma_1^j\gamma_1^k, \tag{3}$$

and the coordinates of $j_p^2\xi$ transform according to

$$\bar\xi_1^\alpha=(\bar X_n^\alpha+\bar X_\beta^\alpha\xi_1^\beta)/(\bar X_n^n+\bar X_\gamma^n\xi_1^\gamma) \tag{4}$$

and

$$\bar\xi_2^\alpha=\frac{\bar X_\rho^\alpha\xi_2^\rho+\bar X_{\rho\sigma}^\alpha\xi_1^\rho\xi_1^\sigma+2\bar X_{n\rho}^\alpha\xi_1^\rho+\bar X_{nn}^\alpha}{(\bar X_n^n+\bar X_\gamma^n\xi_1^\gamma)^2}-\frac{\bar\xi_1^\alpha(\bar X_\rho^n\xi_2^\rho+\bar X_{\rho\sigma}^n\xi_1^\rho\xi_1^\sigma+2\bar X_{n\rho}^n\xi_1^\rho+\bar X_{nn}^n)}{(\bar X_n^n+\bar X_\gamma^n\xi_1^\gamma)^2}, \tag{5}$$

where $\bar X=\bar x\circ x^{-1}$.

Denote by $\mathscr{L}_1^1(M)$ and $\mathscr{L}_1^2(M)$ the bundles of first- and second-order curve elements and by $\mathscr{D}^1(M)$ and $\mathscr{D}^2(M)$ the bundles of first- and second-order path elements. In each case, the bundle of second-order elements can be regarded as a bundle over the corresponding bundle of first-order elements.

An acceleration field is a cross section A: $L_1^1(M)\to L_1^2(M)$. Such a field is described in terms of local coordinates by functions $A_2^i(x^i,\gamma_2^i)$, which transform under a change of coordinates according to

$$\bar A_2^i(\bar x^i,\bar\gamma_1^i)=\bar X_j^i(x^i)A_2^j(x^i,\gamma_1^i)+\bar X_{jk}^i(x^i)\gamma_1^j\gamma_1^k. \tag{6}$$

An acceleration field is called geodesic iff for every $p\in M$, there is some coordinate chart $(\bar U,\bar x)_p$ such that the functions $\bar A_2^i(\bar x^i,\bar\gamma_1^i)$ vanish at p. A geodesic acceleration field is denoted by Γ and has the special functional form

$$\Gamma_2^i(x^i,\gamma_1^i)=-\Gamma_{jk}^i(x^i)\gamma_1^j\gamma_1^k. \tag{7}$$

A directing field is a cross section Ξ: $\mathbb{D}^1(M)\to\mathbb{D}^2(M)$. Such a field is described in terms of local coordinates by functions $\Xi_2^\alpha(x^i,\xi_1^\alpha)$, which transform under a change of coordinate chart according to

$$\bar\Xi_2^\alpha(\bar x^i,\bar\xi_1^\alpha)=[\bar X_\rho^\alpha(x^i)\Xi_2^\rho(x^i,\xi_1^\alpha)+\bar X_{\rho\sigma}^\alpha(x^i)\xi_1^\rho\xi_1^\sigma+2\bar X_{n\rho}^\alpha(x^i)\xi_1^\rho+\bar X_{nn}^\alpha(x^i)]/(\bar X_n^n(x^i)+\bar X_\gamma^n(x^i)\xi_1^\gamma)^2$$
$$-\bar\xi_1^\alpha[\bar X_\rho^n(x^i)\Xi_2^\rho(x^i,\xi_1^\alpha)+\bar X_{\rho\sigma}^n(x^i)\xi_1^\rho\xi_1^\sigma+2\bar X_{n\rho}^n(x^i)\xi_1^\rho+\bar X_{nn}^n(x^i)]/(\bar X_n^n(x^i)+\bar X_\gamma^n(x^i)\xi_1^\gamma)^2. \tag{8}$$

A directing field is called geodesic iff for every $p\in M$, there is some coordinate chart $(\bar U,\bar x)_p$ such that the functions $\bar\Xi_2^\alpha(\bar x^i,\bar\xi_1^\alpha)$ vanish at p. A geodesic directing field is denoted by Π and has the special functional form

$$\Pi_2^\alpha(x^i,\xi_1^\alpha)=\xi_1^\alpha(\Pi_{\rho\sigma}^n(x^i)\xi_1^\rho\xi_1^\sigma+2\Pi_{n\rho}^n(x^i)\xi_1^\rho+\Pi_{nn}^n(x^i))-(\Pi_{\rho\sigma}^\alpha(x^i)\xi_1^\rho\xi_1^\sigma+2\Pi_{n\rho}^\alpha(x^i)\xi_1^\rho+\Pi_{nn}^\alpha(x^i)), \tag{9}$$

 0022-2488/90/010127-04$03.00

where the projective coefficients $\Pi^i_{jk}(x^i)$ are traceless so that $\Pi^n_{nn}(x^i)$ and $\Pi^n_{np}(x^i)$ may be eliminated from (9).

An acceleration field A determines a directing field Ξ iff A is of the form (see Ref. 1, Theorem 3.1)

$$A^i_2(x^i,\gamma^i_1) = B(x^i,\gamma^i_1)\gamma^i_1 + C^i(x^i,\gamma^i_1), \tag{10}$$

where

$$C^i(x^i,\lambda\gamma^i_1) = \lambda^2 C^i(x^i,\gamma^i_1). \tag{11}$$

II. ANALYSIS OF DIRECTING FIELDS

The directing fields $\Xi^\alpha_2(x^i,\xi^\alpha_1)$ considered in this section are all those of the form

$$\Xi^\alpha_2(\xi^\alpha_1) = [\gamma^n_1 A^\alpha_2(\gamma^i_1) - \gamma^\alpha_1 A^n_2(\gamma^i_1)]/(\gamma^n_1)^3, \tag{12}$$

where $A^i_2(\gamma^i_1)$ is any member of the following multiple-parameter family:

$$\begin{aligned} A^i_2(\gamma^i_1) = {} & \mu_1(g_{rs}\gamma^r_1\gamma^s_1)^{1/2}T^i_{j_1}\gamma^{j_1}_1 - \Pi^i_{j_1j_2}\gamma^{j_1}_1\gamma^{j_2}_1 \\ & + \mu_3(g_{rs}\gamma^r_1\gamma^s_1)^{-1/2}T^i_{j_1j_2j_3}\gamma^{j_1}_1\gamma^{j_2}_1\gamma^{j_3}_1 + \cdots \\ & + \mu_k(g_{rs}\gamma^r_1\gamma^s_1)^{(k-2)/2}T^i_{j_1\cdots j_k}\gamma^{j_1}_1\cdots\gamma^{j_k}_1, \end{aligned} \tag{13}$$

where g_{jk} is the space-time metric tensor, the $\Pi^i_{j_1j_2}$ are the projective coefficients, and the

$$T_{ij_1\cdots j_r} = g_{ij}T^j_{j_1\cdots j_r} \tag{14}$$

are tensors antisymmetrized on the first two indices and then symmetrized on the last r indices so that

$$g_{ij}\gamma^i_1 T^j_{j_1\cdots j_k} = 0, \tag{15}$$

and

$$A^\alpha_{2,\alpha} = 0. \tag{16}$$

The fact that the fields $A^i_2(\gamma^i_1)$ are homogeneous of degree 2 in the variables γ^i_1 yields the relation

$$A^i_2(\gamma^i_1) = \gamma^r_1 A^i_{2,r}(\gamma^i_1). \tag{17}$$

The relations (12), (16), and (17) may be used to express the fields $A^i_2(\gamma^i_1)$ in terms of the fields $\Xi^\alpha_2(\xi^\alpha_1)$. The result is

$$\begin{aligned} A^\alpha_2(\gamma^i_1) &= (\gamma^n_1)^2\Xi^\alpha_2(\xi^\alpha_1) - [\gamma^n_1\gamma^\alpha_1/(n+2)]\Xi^\beta_{2,\beta}(\xi^\alpha_1), \\ A^n_2(\gamma^i_1) &= -[(\gamma^n_1)^2/(n+2)]\Xi^\beta_{2,\beta}(\xi^\alpha_1). \end{aligned} \tag{18}$$

Elsewhere[2] we have shown that the fields $\Xi^\alpha_2(\xi^\alpha_1)$ can be measured directly in a noncircular manner given access only to the differential topology of space-time, that is, the ability to track material bodies with respect to an arbitrary coordinate system. It follows that the fields $A^i_2(\gamma^i_1)$ are also measurable.

Although the full space-time metric cannot be measured in a direct manner, the conformal structure of space-time can be directly measured. The remaining unknown scale factor can be determined provided that it is possible to measure the projective structure of space-time.[2,3] In this section we show that the projective structure and the tensor fields that occur in the fields (13) can be determined provided that the (measurable) directing fields corresponding to the fields (13) are known. Our procedure employs harmonic analysis on the forward unit hyperboloid.

We denote the coefficients that determine the conformal structure locally by $\mathbf{g}_{ij}$, which for definiteness we assume to be normalized to satisfy

$$\det(\mathbf{g}_{ij}) = -1. \tag{19}$$

Define

$$\hat{\gamma}^i_1 = \gamma^i_1/\sqrt{\mathbf{g}_{rs}\gamma^r_1\gamma^s_1}. \tag{20}$$

Since the terms on the right-hand side of Eq. (13) are homogeneous of degree 2 in the n-velocities γ^i_1, division by $\mathbf{g}_{ij}\gamma^i_1\gamma^j_1$ yields

$$\begin{aligned} A^i_2(\hat{\gamma}^i_1) = {} & \mu_1\mathscr{T}^i_{j_1}\hat{\gamma}^{j_1}_1 - \Pi^i_{j_1j_2}\hat{\gamma}^{j_1}_1\hat{\gamma}^{j_2}_1 \\ & + \mu_3\mathscr{T}^i_{j_1j_2j_3}\hat{\gamma}^{j_1}_1\hat{\gamma}^{j_2}_1\hat{\gamma}^{j_3}_1 + \cdots \\ & + \mu_k\mathscr{T}^i_{j_1\cdots j_k}\hat{\gamma}^{j_1}_1\cdots\hat{\gamma}^{j_k}_1, \end{aligned} \tag{21}$$

where the unknown conformal factor has been absorbed into the coefficients $\mathscr{T}$. It is worth noting that these functions defined on the forward unit mass shell satisfy the *Lorentz-finite* property; that is, the set of functions (for any k and for any coefficients $\mathscr{T}$ and Π)

$$\{A^i_2(\Lambda_j^{-1i}\hat{\gamma}^j_1)\,|\,\Lambda^i_j\in SO(1,n-1)\} \tag{22}$$

is a finite-dimensional vector space with respect to pointwise addition and scalar multiplication. Helgason[4,5] has shown that such functions must be the restriction to the forward unit hyperboloid of a polynomial on $\mathbb{R}^n$.

The analysis of the fields is carried out on a pointwise basis as far as the space-time coordinates x^i are concerned. Let y^i be coordinates for the interior of the forward light cone at the given space-time point so that

$$\mathbf{g}_{ij}y^iy^j > 0. \tag{23}$$

At the given point, the microinvariance group of the conformal structure acts on the interior of the forward light cone. The generators for the action of the Lorentz subgroup are

$$J^{rs} = y^r\mathbf{g}^{st}\frac{\partial}{\partial y^t} - y^s\mathbf{g}^{rt}\frac{\partial}{\partial y^t}. \tag{24}$$

Set

$$\hat{y}^i = y^i/\sqrt{\mathbf{g}_{rs}y^ry^s}. \tag{25}$$

Denote by $[\hat{y}^{i_1}\hat{y}^{i_2}\cdots\hat{y}^{i_r}]$ the product $\hat{y}^{i_1}\hat{y}^{i_2}\cdots\hat{y}^{i_r}$ with all traces removed so that

$$\begin{aligned} & \hat{y}^i = [\hat{y}^i], \quad \hat{y}^{i_1}\hat{y}^{i_2} = [\hat{y}^{i_1}\hat{y}^{i_2}] + \tfrac{1}{2}\mathbf{g}^{i_1i_2}, \\ & \hat{y}^{i_1}\hat{y}^{i_2}\hat{y}^{i_3} = [\hat{y}^{i_1}\hat{y}^{i_2}\hat{y}^{i_3}] + [1/(n+2)](\mathbf{g}^{i_1i_2}\hat{y}^{i_3} + \mathbf{g}^{i_2i_3}\hat{y}^{i_1} + \mathbf{g}^{i_3i_1}\hat{y}^{i_2}), \end{aligned} \tag{26}$$

and

$$\begin{aligned} \hat{y}^{i_1}\hat{y}^{i_2}\hat{y}^{i_3}\hat{y}^{i_4} = {} & [\hat{y}^{i_1}\hat{y}^{i_2}\hat{y}^{i_3}\hat{y}^{i_4}] + [1/n(n+2)](\mathbf{g}^{i_1i_2}\mathbf{g}^{i_3i_4} + \mathbf{g}^{i_1i_3}\mathbf{g}^{i_2i_4} + \mathbf{g}^{i_1i_4}\mathbf{g}^{i_2i_3}) \\ & + [1/(n+4)](\mathbf{g}^{i_1i_2}[\hat{y}^{i_3}\hat{y}^{i_4}] + \mathbf{g}^{i_1i_3}[\hat{y}^{i_2}\hat{y}^{i_4}] + \mathbf{g}^{i_1i_4}[\hat{y}^{i_2}\hat{y}^{i_3}] + \mathbf{g}^{i_2i_3}[\hat{y}^{i_1}\hat{y}^{i_4}] + \mathbf{g}^{i_2i_4}[\hat{y}^{i_1}\hat{y}^{i_3}] + \mathbf{g}^{i_3i_4}[\hat{y}^{i_1}\hat{y}^{i_2}]). \end{aligned} \tag{27}$$

The Laplace–Beltrami operator on the forward unit hyperboloid is given by

$$L = \tfrac{1}{2}\mathbf{g}_{ra}\mathbf{g}_{sb}J^{rs}J^{ab}. \tag{28}$$

Note that this operator does not depend on the arbitrary normalization of $\mathbf{g}$. The functions $[\hat{y}^{i}\hat{y}^{i_2}\cdots\hat{y}^{i_r}]$ satisfy

$$-L\,[\hat{y}^{i}\hat{y}^{i_2}\cdots\hat{y}^{i_r}] = (r^2+(n-2)r)[\hat{y}^{i}\hat{y}^{i_2}\cdots\hat{y}^{i_r}]\,. \tag{29}$$

By expressing the field (21) in terms of the quantities $[\hat{y}^{i}\hat{y}^{i_2}\cdots\hat{y}^{i_r}]$, one obtains

$$A^{i}_{2}(\hat{y}^{i}) = S^{i} + S^{i}_{j_1}[\hat{y}^{j_1}] + S^{i}_{j_1 j_2}[y^{j_1}y^{j_2}] + \cdots + S^{i}_{j_1\cdots j_k}[y^{j_1}\cdots y^{j_k}]. \tag{30}$$

The first task is to determine the coefficients $S^{i}_{j_1\cdots j_r}$. For convenience, write A for $A^{i}_{2}(\hat{y}^{i})$, L_r for the operator $-L-(r^2+(n-2)r)$, and S_r for $S^{i}_{j_1\cdots j_r}[y^{j_1}\cdots y^{j_r}]$. The operator L_r multiplies the corresponding term S_r by zero and any other term, say S_s by the nonzero integer $(s^2+(n-2)s)-(r^2+(n-2)r)$. Hence the field A has the form

$$A = S_0 + S_1 + \cdots + S_k \tag{31}$$

for some integer k, just in case

$$L_k L_{k-1}\cdots L_1 L_0 A = 0, \tag{32}$$

but

$$L_{k-1}L_{k-2}\cdots L_1 L_0 A \neq 0. \tag{33}$$

In fact, one has

$$L_{k-1}L_{k-2}\cdots L_1 L_0 A = \left(\prod_{r=0}^{k-1}[(k^2+(n-2)k)-(r^2+(n-2)r)]\right)S_k, \tag{34}$$

from which one can obtain S_k. Moreover,

$$L_{k-2}L_{k-3}\cdots L_1 L_0 A = \left(\prod_{r=0}^{k-2}[((k-1)^2+(n-2)(k-1)) - (r^2+(n-2)r)]\right)S_{k-1} + \left(\prod_{r=0}^{k-2}[(k^2+(n-2)k)-(r^2+(n-2)r)]\right)S_k. \tag{35}$$

Since S_k is known, one can subtract the second term and hence determine S_{k-1}. Clearly, this process can be used to determine all of the coefficients S_r.

In general, the coefficients $S^{i}_{j_1 j_2\cdots j_r}$ are mixtures of the coefficients $\mathcal{T}^{i}_{j_1 j_2\cdots j_r}$ and $\Pi^{i}_{j_1 j_2}$ that one desires to determine, and the procedure used to determine them can, in principle, be quite complex. Fortunately, the analysis of directing fields for neutral monopoles and electrically charged monopoles is relatively straightforward. We therefore consider these physically relevant and important cases before considering the complications that arise in a case involving terms up to fourth order. For an electromagnetic directing field, only the first two terms of (21) occur, and μ_1 is the electromagnetic charge-to-mass ratio. One obtains

$$\begin{aligned} S^{i} &= (1/n)\mathbf{g}^{k_1k_2}\Pi^{i}_{k_1k_2}, \\ S^{i}_{j_1} &= \mu_1\mathcal{T}^{i}_{j_1}, \\ S^{i}_{j_1j_2} &= \Pi^{i}_{j_1j_2} - (1/n)\mathbf{g}_{j_1j_2}\mathbf{g}^{k_1k_2}\Pi^{i}_{k_1k_2}. \end{aligned} \tag{36}$$

The projective structure is easily recovered from the first and third of these equations. One obtains

$$\Pi^{i}_{j_1j_2} = S^{i}_{j_1j_2} + \mathbf{g}_{j_1j_2}S^{i}. \tag{37}$$

It is important to note that the projective structure so determined does not depend on the arbitrary normalization of $\mathbf{g}_{j_1j_2}$. The projective structure $\Pi^{i}_{j_1j_2}$ and the conformal structure $\mathbf{g}_{j_1j_2}$ together uniquely determine[2,3] the affine structure $\Gamma^{i}_{j_1j_2}$ and hence a Weyl structure[3,6] on space-time provided that the projective and conformal structures satisfy a second-order compatibility constraint. Moreover, provided that $\Gamma_k = \Gamma^{r}_{rk}$ satisfies

$$\Gamma_{kj} - \Gamma_{j,k} = 0, \tag{38}$$

the conformal structure reduces to a Riemann structure. Thus the scale factor $\lambda(x^{i})$, where

$$g_{j_1j_2}(x^{i}) = \lambda(x^{i})\mathbf{g}_{j_1j_2}(x^{i}), \tag{39}$$

can be determined by parallel transport of the scale factor chosen arbitrarily at any one point, and the electromagnetic field tensor $T^{i}_{j_1}$ is given by

$$\mu_1 T^{i}_{j_1} = \sqrt{\lambda(x^{i})}S^{i}_{j_1} = \mu_1\sqrt{\lambda(x^{i})}\mathcal{T}^{i}_{j_1}, \tag{40}$$

where the charge-to-mass ratio μ_1 must be chosen arbitrarily for at least one class of charged particles in order to determine the scale for charge space.

Remark 1: In all extant work[3,7–10] on the constructive axiomatics of the general theory of relativity, it was necessary to specify that the test particles used to determine the projective structure of space-time were not only monopole but also *neutral*. In a previous paper,[11] we constructed a method of uniquely decomposing a directing field of the type (13) into its projective and force components. This method has the virtue of permitting the determination of the projective coefficients Π^{i}_{jk} given only *one* class of monopoles (13) regardless of the complexity of the force component; however, it requires that the $\mathbf{g}_{ij}$ be measured with sufficient precision to permit the computation of the conformal connection coefficients K^{i}_{jk} from them.

In contrast, in order to determine the Π^{i}_{jk}, the harmonic method discussed in this paper requires only the $\mathbf{g}_{ij}$ and not the K^{i}_{jk}. The K^{i}_{jk} are only required for the determination of the scale factor $\lambda(x^{i})$ and hence the metric g_{ij}. On the other hand, as is shown below, many classes of monopoles with a suitable variety of charge-to-mass ratios may in principle be required to determine the Π^{i}_{jk} In practice, however, this is not really a problem because only the electromagnetic case is physically relevant and in this case only one electromagnetic charge-to-mass ratio is needed. In addition, the general techniques of harmonic analysis are widely known, a fact that makes the harmonic method more transparent.

Remark 2: Although the projective structure (geodesic directing field) that governs the motions of neutral massive monopoles is a special case of the electromagnetic directing field discussed above, it is particularly important because of the central role it plays in the formulation of the Law of

Inertia.[1] In previous work[2] (see, also Ref. 12), we provided an empirical criterion for determining whether or not a directing field is or is not geodesic (projective); namely, a directing field $\Xi_2^\alpha(x^i,\xi_1^\alpha)$ is geodesic iff it is cubic in its $(n-1)$-velocities ξ_1^α. The analysis presented above shows that an alternative criterion is the following.

An Empirical Criterion for Geodesicity: A directing field $\Xi_2^\alpha(\xi_1^\alpha)$ is geodesic iff the corresponding field (21) satisfies

$$L_2L_1L_0A^i_2(\hat{\gamma}^i_1)=0 \tag{41}$$

and

$$L_1L_0A^i_2(\hat{\gamma}^i_1)=(2n-1)L_0A^i_2(\hat{\gamma}^i_1). \tag{42}$$

We conclude with the analysis of a somewhat more complicated possibility, namely, the field (21) with $k=4$. The $S^i_{j_1\cdots j_r}$ are given by

$$\begin{aligned}
S^i &= (1/n)\mathbf{g}^{k_1k_2}\Pi^i_{k_1k_2}+[3\mu_4/n(n+2)]\mathbf{g}^{k_1k_2}\mathbf{g}^{k_3k_4}\mathcal{T}^i_{k_1k_2k_3k_4},\\
S^i_{j_1} &= \mu_1\mathcal{T}^i_{j_1}+[3\mu_3/(n+2)]\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1},\\
S^i_{j_1j_2} &= (\Pi^i_{j_1j_2}-(1/n)\mathbf{g}_{j_1j_2}\mathbf{g}^{k_1k_2}\Pi^i_{k_1k_2})+[6\mu_4/(n+4)](\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1j_2}-(1/n)\mathbf{g}_{j_1j_2}\mathbf{g}^{k_1k_2}\mathbf{g}^{k_3k_4}\mathcal{T}^i_{k_1k_2k_3k_4}),\\
S^i_{j_1j_2j_3} &= \mu_3[\mathcal{T}^i_{j_1j_2j_3}-[1/(n+2)](\mathbf{g}_{j_1j_2}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_3}+\mathbf{g}_{j_2j_3}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1}+\mathbf{g}_{j_3j_1}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_2})],
\end{aligned} \tag{43}$$

and

$$\begin{aligned}
S^i_{j_1j_2j_3j_4} = \mu_4\{&\mathcal{T}^i_{j_1j_2j_3j_4}-[1/(n+4)][\mathbf{g}_{j_1j_2}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_3j_4}+\mathbf{g}_{j_1j_3}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_2j_4}+\mathbf{g}_{j_1j_4}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_2j_3}\\
&+\mathbf{g}_{j_2j_3}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1j_4}+\mathbf{g}_{j_2j_4}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1j_3}+\mathbf{g}_{j_3j_4}\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1j_2}\\
&-[1/(n+2)](\mathbf{g}_{j_1j_2}\mathbf{g}_{j_3j_4}+\mathbf{g}_{j_1j_3}\mathbf{g}_{j_2j_4}+\mathbf{g}_{j_1j_4}\mathbf{g}_{j_2j_3})\mathbf{g}^{k_1k_2}\mathbf{g}^{k_3k_4}\mathcal{T}^i_{k_1k_2k_3k_4}]\}.
\end{aligned} \tag{44}$$

It is clear that one must know the $S^i_{j_1j_2\cdots j_r}$ for a sufficient variety of μ_r. Moreover, one complete set of the μ_r must be fixed arbitrarily in order to determine the scale for each of the charge-to-mass ratios. For one field for which $\mu_3\neq0$ and $\mu_4\neq0$, define the charge-to-mass scales by setting $\mu_3=1$ and $\mu_4=1$. One can then easily determine $\tilde{\mu}_3$ and $\tilde{\mu}_4$ from the ratio of $\tilde{S}^i_{j_1j_2j_3}$ and $S^i_{j_1j_2j_3}$ and the ratio of $\tilde{S}^i_{j_1j_2j_3j_4}$ and $S^i_{j_1j_2j_3j_4}$, respectively. Then from $\tilde{S}^i_{j_1j_2}$ and $S^i_{j_1j_2}$ and from $\tilde{S}^i$ and S^i the projective and nonprojective terms can be recovered. Thus the projective structure can be reconstructed in the same way as in the case of the electromagnetic directing field, and the metric tensor and hence the scale factor $\lambda(x^i)$ can be determined from the projective and conformal structures as before. In addition, the nonprojective terms of S^i, $S^i_{j_1j_2j_3j_4}$ suffice for the reconstruction of $\mathcal{T}^i_{j_1j_2j_3j_4}$ and hence of $T^i_{j_1j_2j_3j_4}$. Finally, given

$$S^i_{j_1}=\mu_1\mathcal{T}^i_{j_1}+[3\mu_3/(n+2)]\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1}, \tag{45}$$

$$\hat{S}^i_{j_1}=\hat{\mu}_1\mathcal{T}^i_{j_1}+[3\hat{\mu}_3/(n+2)]\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1}, \tag{46}$$

one can obtain

$$\hat{\mu}_3S^i_{j_1}-\mu_3\hat{S}^i_{j_1}=(\hat{\mu}_3\mu_1-\mu_3\hat{\mu}_1)\mathcal{T}^i_{j_1}. \tag{47}$$

One could determine the scale for the μ_1 charge-to-mass ratio by setting the coefficient $(\hat{\mu}_3\mu_1-\mu_3\hat{\mu}_1)$ equal to 1 and hence determine $\mathcal{T}^i_{j_1}$; however, μ_1 and $\hat{\mu}_1$ are still not known and there does not seem to be any straightforward way to determine the field $\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1}$. However, if the family of directing fields contained a pair of fields with $\mu_1'=0, \mu_3'\neq0$, $\mu_1''=0$, and $\mu_3''\neq0$, then they could be identified by the fact that the combination corresponding to (47) would vanish (assuming it is already known that $\mathcal{T}^i_{j_1}$ does not vanish). From these fields, $\mathbf{g}^{k_1k_2}\mathcal{T}^i_{k_1k_2j_1}$ can be determined and hence $\mathcal{T}^i_{j_1j_2j_3}$ and $T^i_{j_1j_2j_3}$ can be reconstructed.

ACKNOWLEDGMENT

This work was supported in part by the Natural Science and Engineering Research Council of Canada.

[1] R. A. Coleman and H. Korte, J. Math. Phys. **25**, 3513 (1984).
[2] R. A. Coleman and H. Korte, J. Math. Phys. **21**, 1340 (1980); **23**, 345 (E) (1982).
[3] J. Ehlers, F. A. E. Pirani, and A. Schild, *The Geometry of Free Fall and Light Propagation in General Relativity* (Clarendon, Oxford, 1972).
[4] S. Helgason, *Groups and Geometric Analysis* (Academic, New York, 1984). See pp. 72 and 73. Also, see pp. 16–20 for a discussion of the harmonic analysis in the Euclidean case.
[5] S. Helgason, Acta Math. **109**, 241 (1963).
[6] R. A. Coleman and H. Korte, J. Math. Phys. **22**, 2598 (1981).
[7] F. A. E. Pirani, Symp. Math. **XII**, 67 (1973).
[8] N. M. J. Woodhouse, J. Math. Phys. **14**, 495 (1973).
[9] R. F. Marzke and J. A. Wheeler, *Gravitation as Geometry in Gravitation and Relativity*, edited by H. Chiu and W. F. Hoffmann (Benjamin, New York, 1964), pp. 40–64.
[10] W. Kundt and B. Hoffmann, *Determination of Gravitational Standard Time in Recent Developments in General Relativity* (Pergamon, New York, 1962), pp. 303–306.
[11] R. A. Coleman and H. Korte, J. Math. Phys. **28**, 1492 (1987).
[12] J. Ehlers and E. Köhler, J. Math. Phys. **18**, 2014 (1977).

On Chandrasekhar's perturbation analysis. I. The superfluity of the N.P. system

S. B. Edgar
Department of Mathematics, Linköping University, S-581 83 Linköping, Sweden

(Received 13 April 1989; accepted for publication 30 August 1989)

Chandrasekhar has developed a method of analyzing first-order perturbations about some known metrics using the N.P. system of equations. In this paper it is shown that some of the intriguing aspects that have been noted in his method—the superfluity of the N.P. system, and the existence of very complicated integral identities—are not peculiar to this particular type of perturbation analysis; rather the underlying principles are fundamental properties associated with the differential structure of the N.P. system. Specifically the three different subsystems used in the three space-times where Chandrasekhar's method has been applied, are confirmed directly as sufficient subsystems for extracting all information from the complete N.P. system, for the respective situations in which they have been used.

I. INTRODUCTION

In his work on gravitational perturbations about a Kerr black hole, Chandrasekhar[1–5] draws attention to the need for further investigation of a number of related features that emerge in his analysis. These include

(i) the superfluity of the N.P. system of equations,[6] which permits the complete solution of this system to be obtained by solving explicitly only a particular subsystem (which shall be called subsystem C_1) of the whole system;

(ii) the existence of integral identities (and, in particular, the depth of their integral nature) which are obtained in the process of calculating the solution of subsystem C_1; and

(iii) the question whether a similar type of analysis would supply new integrability identities for existing functions of mathematical physics, when they occur in Einstein's equations.

Chandrasekhar considers a Kerr black hole being perturbed so that the N.P. quantities associated with the Kerr metric change by the first order of smallness and with a particular t and φ dependence. He then solves the subsystem C_1—to this level of accuracy and with this restricted t and φ dependence—to obtain explicit expressions for all the first-order changes.

The various specializations being made, the approximate nature of the analysis, the very long and complicated calculations involved, mean that the features described above emerge in a puzzling manner; it is not immediately apparent whether they are features of this particular approach to this particualr problem, or whether they are of a more general and fundamental nature.

The purpose of this paper is to examine the first feature noted above—although the results obtained will have implications in the greater understanding of the related features. When the differential structure of the N.P. system is considered, together with the relationship of this system to Einstein's equations, it becomes clear that the superfluity of the N.P. system and the existence of associated integral identities are fundamental features—peculiar neither to Chandrasekhar's approach, nor even to this type of perturbation analysis.

The results of this paper follow from work by Papapetrou[7,8] on the structure of the system of equations used in tetrad formalisms. Since these results do not seem to be well known, they are summarized in the following section; and in Sec. III some general implications are deduced regarding existence of sufficient subsystems of the N.P. system.

In Sec. IV it is shown that in general subsystem C_1 is not sufficient to guarantee exact solutions of the whole N.P. system. Of course Chandrasekhar's analysis seeks to find approximate solutions with a restricted type of t and φ behavior; and it does not automatically follow that if a particular subsystem is insufficient (or sufficient) in general to guarantee the exact solution, that it is insufficient (or sufficient) to guarantee an approximate solution of a particular restricted type. In fact the insistance that solutions must have a particular type of t and φ dependence is a crucial restriction, which ensures that subsystem C_1 is sufficient in these circumstances. So it is proved in Sec. V that subsystem C_1 is a sufficient subsystem to guarantee restricted approximate solutions of the type under consideration.

It should be noted that Chandrasekhar did not show directly that subsystem C_1 was sufficient, and some of the N.P. equations were left unsolved in his analysis—although the nature of his solution gives evidence that all of the information in the N.P. system has indeed been extracted using subsystem C_1. However, in one of his original papers[2] Chandrasekhar had prematurely argued that all of the information had been extracted—using an even smaller subsystem than subsystem C_1—only to find later that this was not the case, and additional N.P. equations had to be solved explicitly.[3,4]

The result obtained in this paper therefore confirms unambiguously that Chandrasekhar's analysis, using subsystem C_1 has indeed extracted all possible information from the N.P. system for the type of restricted and approximate solutions under consideration. It is also noted that the significance of subsystem C_1 is limited to this particular approach to this particular problem. However, in any attempt to find exact or approximate solutions to Einstein's equations there will usually be one or more subsystems that suggest themselves in a natural way; the sufficiency of these

 0022-2488/90/010131-06$03.00

subsystems can be tested by Papapetrou's identities.[9–11]

Recently Chandrasekhar and Xanthopoulus[12] have carried out the same type of perturbation analysis on region I and region II of Bell–Szekeres space-time.[13–15] In each region only a subsystem of the complete N.P. system is solved explicitly (subsystems C_2 and C_3, respectively). Although there is no doubt that all of the information has been extracted, it is instructive to use Papapetrou's identities to confirm that the subsystems C_2 and C_3 are indeed sufficient in their respective situations.

This is shown in Sec. VI; and once again it is noted that the usefulness of both subsystems seems to be restricted to these particular situations.

Since subsystems C_1, C_2, and C_3 are all different, some overall insight into the redundancy of the N.P. system and the role of integrability conditions has been gained, and this is summarized in Sec. VII.

II. PAPAPETROU'S IDENTITIES

The basic system of equations for the tetrad formalism can be presented as

$$X_{mnp} = 0, \tag{2.1}$$

$$Y_{mnpq} = 0, \tag{2.2}$$

$$V_{sm[npq]} = 0, \tag{2.3}$$

where

$$X_{mnp} \equiv 2\gamma_{[m|n|p]} - 2Z_{[m}{}^{a}{}_{:p]}Z_{na}, \tag{2.4}$$

$$Y_{mnpq} \equiv R_{mnpq} - 2\gamma_{mn[p:q]} - 2\gamma_{sm[q}\gamma^{s}{}_{|n|p]} - 2\gamma_{mn}{}^{s}\gamma_{s[pq]}, \tag{2.5}$$

$$V_{smnpq} \equiv R_{smnp:q} - 2R_{smtn}\gamma^{t}{}_{sn} + R_{pqtm}\gamma^{t}{}_{sn} - R_{pqts}\gamma^{t}{}_{mn}. \tag{2.6}$$

Following Papapetrou,[8] labels $X_{mnp}, Y_{mnpq}, V_{smnpq}$ have been introduced for the different equations; the remainder of the notation is standard.

The redundancy is then given explicitly by the following three sets of identities:

$$X_{[m}{}^{s}{}_{n:p]} - X_{r}{}^{s}{}_{[m}X_{n}{}^{r}{}_{p]} + X_{[m}{}^{r}{}_{n}\{\gamma_{|r|}{}^{s}{}_{p]} - \gamma_{p]}{}^{s}{}_{r}\} + 2X_{r}{}^{s}{}_{[m}\gamma_{n}{}^{r}{}_{p]} + Y_{[m}{}^{s}{}_{np]} = 0, \tag{2.7}$$

$$V_{mn[pqr]} - Y_{mn[pq:r]} + \gamma_{mn[p:|s|}X_{q}{}^{s}{}_{r]} + 2\gamma^{s}{}_{[pq}Y_{|mns|r]} - \gamma^{s}{}_{m[p}Y_{|sn|qr]} + \gamma^{s}{}_{n[p}Y_{|sm|qr]} + \gamma_{mn}{}^{s}Y_{s[pqr]} = 0, \tag{2.8}$$

$$\eta^{pqrs}\{V_{mnpqr:s} + 3\gamma^{t}{}_{rs}V_{mn[pqt]} - 2\gamma^{t}{}_{[m|p}V_{t|n]qrs} - R_{mntr}Y^{t}{}_{pqs} + R_{qrt[n}Y^{t}{}_{m]ps} + \tfrac{1}{2}R_{mnpq:t}X_{r}{}^{t}{}_{s}\} = 0. \tag{2.9}$$

Also used in the tetrad formalism are the commutator equations

$$\{\nabla_{[m}\nabla_{n]} + \gamma_{[m}{}^{p}{}_{n]}\nabla_{p}\} = 0. \tag{2.10}$$

The N.P. formalism for general relativity[6] is derived from the above by

(i) choosing the four tetrad vectors to be two real and two complex null vectors, so that there are 12 complex spin coefficients and 12 independent complex components of the Riemann tensor;

(ii) using Einstein's equations to replace the Ricci tensor components in (2.2) and (2.3) by the appropriate energy-momentum tensor components; and

(iii) writing out each individual equation from the four sets (2.1)–(2.3) and (2.10) using a different symbol for each spin-coefficient, differential operator, and independent Weyl tensor and energy-momentum tensor component.

Perhaps two points over which there seems to be some uncertainty in the literature[16,17] need to be emphasized:

(a) The four sets of equations (2.1)–(2.3) and (2.10) with the substitutions noted above are completely equivalent to Einstein's equations[7] (although this N.P. system is obviously larger with considerable redundancy).

(b) The structure equations defining the spin coefficients (2.1) and the commutator equations (2.10) are exactly equivalent, so only one of these sets needs to be included in the formal N.P. system.[7]

In the remainder of this paper, "the complete N.P. system of equations" will refer to the three sets of equations (2.2), (2.3), and (2.10) written out individually subject to the substitutions noted above. [These are, respectively, Eqs. (310), (321, 322), and (303, 304) on pages 45–51 of Ref. 5.]

The usefulness of Papapetrou's explicit determination of the redundancy is that one is able to pick out from the complete system of equations a particular subsystem and to test directly whether it is sufficient to ensure that the complete system is satisfied. A simple example will illustrate. Suppose one solves only the Ricci equations (2.2) and the commutator equations (2.10), which in the above notation are equivalent to

$$X_{mnp} = 0, \quad \text{for all } m,n,p, \tag{2.11}$$

$$Y_{mnpq} = 0, \quad \text{for all } m,n,p,q. \tag{2.12}$$

By virtue of identity (2.8),

$$V_{sm[npq]} = 0, \quad \text{for all } m,n,p,q,s, \tag{2.13}$$

and so the complete N.P. system is satisfied. This is of course just the well-known result that Eqs. (2.13) (the Bianchi "identities") are identically satisfied, provided that the two sets of structure equations (2.11) and (2.12) are satisfied. However, it is easy to note that this subsystem is unnecessarily large. An appropriately chosen subset of the Ricci equations (2.12) together with all of (2.11) would suffice, since Eqs. (2.11) substituted in identity (2.7) yield

$$Y_{[m}{}^{s}{}_{np]} = 0. \tag{2.14}$$

In practice, in many calculations the Bianchi equations (2.3) are the easiest equations to manipulate, and so it is often preferable to choose them as part of the basic subsystem, together with the commutators (2.10); then the identities (2.7)–(2.9) can be used to determine just how few of the equations from the remaining set of Ricci equations (2.2) are really needed to be added to (2.3) and (2.10) to ensure a sufficient subsystem.

Since Chandrasekhar's method involves a subsystem consisting of all Eqs. (2.3) and (2.10), and some of Eqs. (2.2), Papapetrou's identities enable an analysis of this system to be made comparatively easily.

III. SUBSYSTEMS OF N.P. SYSTEM

In this section some general results for arbitrary spaces are deduced from Papapetrou's identities (2.7)–(2.9).

The first point to note is that the commutator equations (2.10) [or equivalently the structure equations (2.1)] imply from (2.7) that

$$Y_{[m}{}^{s}{}_{np]} = 0. \tag{3.1}$$

These are the six complex and four real "elimination realtions" quoted by Chandrasekhar.[18] Therefore, provided that Eqs. (2.10) are satisfied, Y_{mnpq} has exactly the same symmetry properties as the Riemann tensor R_{mnpq}. So it will be convenient to use a similar N.P.-type notation,

$$\begin{aligned}
&\tilde{\Psi}_0 = -\underset{0}{Y}_{1313}, \quad \tilde{\Psi}_3 = -\underset{0}{Y}_{1241},\\
&\tilde{\Psi}_1 = -\underset{0}{Y}_{1213}, \quad \tilde{\Psi}_4 = -\underset{0}{Y}_{4242},\\
&\tilde{\Psi}_2 = -\underset{0}{Y}_{1342},\\
&\tilde{\Phi}_{00} = -\tfrac{1}{2}Y_{11}, \quad \tilde{\Phi}_{21} = -\tfrac{1}{2}Y_{24},\\
&\tilde{\Phi}_{11} = -\tfrac{1}{2}(Y_{12}+Y_{34}), \quad \tilde{\Phi}_{02} = -\tfrac{1}{2}Y_{33},\\
&\tilde{\Phi}_{01} = -\tfrac{1}{2}Y_{13}, \quad \tilde{\Phi}_{22} = -\tfrac{1}{2}Y_{22},\\
&\tilde{\Phi}_{12} = -\tfrac{1}{2}Y_{23}, \quad \tilde{\Phi}_{20} = -\tfrac{1}{2}Y_{44},\\
&\tilde{\Phi}_{10} = -\tfrac{1}{2}Y_{14}, \quad \tilde{\Lambda} = Y/24,
\end{aligned} \tag{3.2, 3.3}$$

where $\underset{0}{Y}_{mnpq}$, Y_{mp} relate to Y_{mnpq} in exactly the same way as the Weyl tensor C_{mnpq} and Ricci tensor R_{mp}, respectively, relate to the Riemann tensor R_{mnpq}.

Next it is noted that when the Bianchi equations (2.3) are added to the commutator equations (2.10), then (2.8) becomes

$$Y_{mn[pq;r]} - 2\gamma^{s}{}_{[pq}Y_{|mns|r]} + \gamma^{s}{}_{m[p}Y_{|sn|qr]} - \gamma^{s}{}_{n[p}Y_{|sm|qr]} = 0. \tag{3.4}$$

So Y_{mnpq} also obeys the same differential equations as R_{smnp}. This means that each equation of (3.4) can be written out explicitly in the N.P.-type notation simply by substituting the various $\tilde{\Psi},\tilde{\Phi}$ quantities defined in (3.2) and (3.3) for their corresponding Riemann tensor counterparts in the usual N.P. version of the Bianchi equations—Eqs. (321), (321′), and (322) on pp. 49–51 of Ref. 5.

In addition, (2.3) and (2.10) imply from (2.9)

$$R_{mntr}Y^{t}{}_{pqs} - R_{qrt[n}Y^{t}{}_{m]ps} = 0, \tag{3.5}$$

which written out explicitly gives

$$\begin{aligned}
&\Psi_0\tilde{\Psi}_3 - 3\Psi_1\tilde{\Psi}_2 + 3\Psi_2\tilde{\Psi}_1 - \Psi_3\tilde{\Psi}_0 = 0,\\
&\Psi_1\tilde{\Psi}_4 - 3\Psi_2\tilde{\Psi}_3 + 3\Psi_3\tilde{\Psi}_2 - \Psi_4\tilde{\Psi}_1 = 0,\\
&\Psi_0\tilde{\Psi}_4 - 2\Psi_1\tilde{\Psi}_3 + 2\Psi_3\tilde{\Psi}_1 - \Psi_4\tilde{\Psi}_0 = 0.
\end{aligned} \tag{3.6}$$

So provided (2.3) and (2.10) are satisfied, the eighteen complex Ricci equations (2.2) are subject to nine complex and four real algebraic identities (3.1) and (3.6) together with nine complex and two real integrability conditions (3.4). Clearly the two sets of equation (2.3) and (2.10) need only a small number of equations from (2.2) to ensure that the complete N.P. system is satisfied. There are many variations of sufficient subsystems—the type of analysis being carried out and/or the tetrad formalism being used will suggest the appropriate subsystem to be tested for sufficiency.[9–11]

Although there are many choices of sufficient subsystems, there is one equation from the complete system that must always be included in the subsystem to ensure sufficiency in general cases. Note that if all of the equations in the N.P. system (2.1), (2.2), and (2.10) are satisfied, with the exception of the one equation

$$\tilde{\Lambda} = 0, \tag{3.7}$$

then the identities (2.7)–(2.9) reduce to the four equations

$$\tilde{\Lambda}_{,p} = 0, \tag{3.8}$$

which in general have a nonzero (although constant) solution. Hence any subsystem of equations that does not include Eq. (3.7) explicitly or implicitly will not be a sufficient subsystem for arbitrary spaces.

IV. SUBSYSTEM C_1

In this section Chandrasekhar's subsystem C_1 is considered explicitly. This subsystem consists of all the commutator equations (2.10), all the Bianchi equations (2.3), and six Ricci equations from (2.2) [given by Eqs. (310), members a, b, g, j, n, p, on pp. 46–47 of Ref. 5]. In the notation of the last sections, subsystem C_1 consists of

$$X_{mnp} = 0, \tag{4.1}$$

$$\tilde{\Psi}_0 = 0 = \tilde{\Psi}_4, \tag{4.2a}$$

$$\tilde{\Phi}_{00} = \tilde{\Phi}_{22} = 0 = \tilde{\Phi}_{02} = \tilde{\Phi}_{20}, \tag{4.2b}$$

$$V_{sm[npq]} = 0. \tag{4.3}$$

Since this system does not contain Eq. (3.7), it cannot be a sufficient subsystem for arbitrary spaces.

Of course Chandrasekhar was not seeking arbitrary solutions, but rather solutions to a first level of approximation whose perturbations have a t and φ dependence given by

$$e^{i\sigma^{+}t + m\varphi}, \tag{4.4}$$

where σ^{+} is a real positive constant and m is an integer (positive, negative, or zero). The $X_{mnp}, Y_{mnpq}, V_{smnpq}$ labels for the equations will therefore now only be considered to this level of approximation, and so can be written

$$X_{mnp} \equiv X^{(0)}_{mnp} + X^{(1)}_{mnp}, \tag{4.5a}$$

$$Y_{mnpq} \equiv Y^{(0)}_{mnpq} + Y^{(1)}_{mnpq}, \tag{4.5b}$$

$$V_{smnpq} \equiv V^{(0)}_{smnpq} + V^{(1)}_{smnpq}, \tag{4.5c}$$

Since the equations are satisfied to zeroth order, all (0) marked terms are identically zero. The perturbation quantities, marked with (1), will all have the t and φ dependence given by (4.4), and the usual conventions of dropping the (1) superscript and suppressing the factor (4.4) on such quantities will now be followed.

Subsystem C_1, for the type of analysis carried out by Chandrasekhar, is therefore given by (4.1)–(4.3), where the various quantities are now considered as perturbed quantities of first order with an implicit factor (4.4).

With this interpretation of (4.1)–(4.3), the argument in the last paragraph of the last section is no longer valid. Since $\tilde{\Lambda}$ must have a factor (4.4), the only solution to (3.8) is

$$\tilde{\Lambda} = 0. \tag{4.6}$$

Hence the argument in that paragraph cannot be used to rule out subsystem C_1 as a sufficient subsystem for the restricted type of approximate solution sought by Chandrasekhar. So subsystem C_1 will now be considered explicitly for such situations.

V. SUFFICIENCY OF SUBSYSTEM C_1

The equations (4.1)–(4.3) comprising subsystem C are now substituted into the identities (2.7)–(2.9), and the resulting equations are solved to show that the only soluton is that all the remaining $\tilde{\Phi}$ quantities be zero, i.e., the complete N.P. system is satisfied.

When (4.1) and (4.3) are substituted into identity (2.8) as in the last section, it reduces to (3.6); the additional substitution of (4.2), together with the consideration of terms up to first order only, reduces (3.6) to

$$\tilde{\Psi}_1 = 0, \tag{5.1a}$$

$$\tilde{\Psi}_3 = 0. \tag{5.1b}$$

When (4.1)–(4.3), together with (5.1), are substituted into identity (2.8), the 11 equations can be written out explicitly as

$$D\tilde{\Phi}_{01} - 2(\epsilon + \rho^*)\tilde{\Phi}_{01} = 0, \tag{5.2a}$$

$$\delta\tilde{\Phi}_{01} + 2(\pi^* - \beta)\tilde{\Phi}_{01} = 0, \tag{5.2b}$$

$$\Delta\tilde{\Phi}_{21} + 2(\mu^* + \gamma)\tilde{\Phi}_{21} = 0, \tag{5.2c}$$

$$\delta^*\tilde{\Phi}_{21} + 2(\alpha - \tau^*)\tilde{\Phi}_{21} = 0, \tag{5.2d}$$

$$D\tilde{\Psi}_2 + 2D\tilde{\Lambda} - \delta^*\tilde{\Phi}_{01} = 3\rho\tilde{\Psi}_2 - 2(\alpha + \tau^*)\tilde{\Phi}_{01} + 2\rho\tilde{\Phi}_{11} - 2\tau\tilde{\Phi}_{10}, \tag{5.3a}$$

$$\Delta\tilde{\Psi}_2 + 2\Delta\tilde{\Lambda} - \delta\tilde{\Phi}_{21} = -3\mu\tilde{\Psi}_2 + 2(\pi^* + \beta)\tilde{\Phi}_{21} - 2\mu\tilde{\Phi}_{11} - 2\pi\tilde{\Phi}_{12}, \tag{5.3b}$$

$$\delta\tilde{\Psi}_2 + 2\delta\tilde{\Lambda} + \Delta\tilde{\Phi}_{01} = 3\tau\tilde{\Psi}_2 - 2(\mu^* - \gamma)\tilde{\Phi}_{01} + 2\rho\tilde{\Phi}_{12} - 2\tau\tilde{\Phi}_{11}, \tag{5.3c}$$

$$\delta^*\tilde{\Psi}_2 + 2\delta^*\tilde{\Lambda} + D\tilde{\Phi}_{21} = -3\pi\tilde{\Psi}_2 + 2(\rho^* - \epsilon)\tilde{\Phi}_{21} - 2\mu\tilde{\Phi}_{10} + 2\pi\tilde{\Phi}_{11}, \tag{5.3d}$$

$$D(\tilde{\Psi}_2 + \tilde{\Psi}_2^* - \tilde{\Phi}_{11} + \tilde{\Lambda}) = 3\rho\tilde{\Psi}_2 + 3\rho^*\tilde{\Psi}_2^* - (2\tau^* + \pi)\tilde{\Phi}_{01} - (2\tau + \pi^*)\tilde{\Phi}_{10}, \tag{5.4a}$$

$$\Delta(\tilde{\Psi}_2 + \tilde{\Psi}_2^* - \tilde{\Phi}_{11} + \tilde{\Lambda}) = -3\mu\tilde{\Psi}_2 - 3\mu^*\tilde{\Psi}_2^* + (\tau^* + 2\pi)\tilde{\Phi}_{12} + (\tau + 2\pi^*)\tilde{\Phi}_{21}, \tag{5.4b}$$

$$\delta(\tilde{\Psi}_2 + \tilde{\Psi}_2^* + \tilde{\Phi}_{11} + \tilde{\Lambda}) = 3\tau\tilde{\Psi}_2 - 3\pi^*\tilde{\Psi}_2^* - (2\mu^* - \mu)\tilde{\Phi}_{01} + (2\rho - \rho^*)\tilde{\Phi}_{12}. \tag{5.4c}$$

The notation and conventions of Ref. 5 are being used—in particular, all spin coefficients and operators are zeroth order (Kerr), since the field quantities are all first order. Note that, with respect to the original N.P. notation, (5.4a) and (5.4b) are real.

There are a number of possible integrability conditions for this system (five complex and two real), but the only nontrivial ones are the two real ones from the set of equations (5.4); these conditions are quite complicated in the general case.

When the $\tilde{\Phi}$ quantities in (5.2)–(5.4) are restricted to having the factor (4.4), then the zeroth-order operators acting on such quantities have the form

$$D = \partial_r + iK/\Delta, \quad \Delta = -(\Delta/2\rho^2)(\partial_r - iK/\Delta),$$
$$\delta = (1/\bar{\rho}\sqrt{2})(\partial_\theta - Q), \quad \delta^* = (1/\bar{\rho}^*\sqrt{2})(\partial_\theta + Q), \tag{5.5}$$

where

$$K = (r^2 + a^2)\sigma^+ + am, \quad Q = a\sigma^+ \sin\theta + m\csc\theta,$$
$$\bar{\rho} = r + ia\cos\theta, \quad \bar{\rho}^* = r - ia\cos\theta, \tag{5.6}$$
$$\rho^2 = r^2 + a^2\cos^2\theta, \quad \Delta = r^2 - 2Mr + a^2.$$

When these operators are substituted into Eqs. (5.2)–(5.4), they become a system of differential equations in the two coordinates r and θ. The two nontrivial integrability conditons have a simpler form; in addition, two real and two complex algebraic compatibility conditions arise when

(i) ∂_r terms are eliminated by combining (5.4a) and (5.4b);

(ii) ∂_θ terms are eliminated by combining (5.4c) and (5.4c*);

(iii) ∂_r and ∂_θ terms are eliminated by combining (5.2b) and (5.2d) with (5.3c) and (5.3d); and

(iv) ∂_r and ∂_θ terms are eliminated by combining (5.2a) and (5.2c) with (5.3c) and (5.3d).

So there is an algebraic system of six equations (four real and two complex) in five unknowns ($\tilde{\Lambda}, \tilde{\Phi}_{11}$ real and $\tilde{\Psi}_2, \tilde{\Phi}_{01}, \tilde{\Phi}_{21}$ complex). This is equivalent to a homogeneous system of eight real algebraic equations in eight real unknowns, which will have only the trivial solution if all the equations are linearly independent. To show that this system is indeed linearly independent in the Kerr background it is clearly sufficient (and much easier) to show that it is linearly independent in a special case of the Kerr background, i.e., the Schwarzschild background.

When the algebraic and integrability conditions for the system (5.2)–(5.4) are specialized to a Schwarzschild background they simplify as follows.

The two real integrability conditions found from (5.4a) and (5.4b), and (5.4c) and (5.4c*), respectively, are

$$K(\tilde{\Psi}_2 + \tilde{\Psi}_2^*) = 0, \tag{5.7a}$$

$$K(\tilde{\Psi}_2 - \tilde{\Psi}_2^*) = 0. \tag{5.7b}$$

The compatibility conditions for (5.4a) and (5.4b), and (5.4c) and (5.4c*), respectively, are

$$K(\tilde{\Psi}_2 + \tilde{\Psi}_2^* - \tilde{\Phi}_{11} + \tilde{\Lambda}) = 0, \tag{5.7c}$$

$$\sqrt{2}Q(\tilde{\Psi}_2 + \tilde{\Psi}_2^* + \tilde{\Phi}_{11} + \tilde{\Lambda}) + ((\Delta/2\rho^2)\tilde{\Phi}_{01} + \tilde{\Phi}_{21}) - ((\Delta/2\rho^2)\tilde{\Phi}_{10} + \tilde{\Phi}_{12}) = 0. \tag{5.7d}$$

The compatibility conditions from (5.2b), (5.2d),

(5.3a), (5.3b), and (5.2a), (5.2c), (5.3c), and (5.3d), respectively, are

$$\sqrt{2}Q(\tilde{\Psi}_2+2\tilde{\Lambda})-2\Delta\left(\frac{\Delta}{2\rho^2}\tilde{\Phi}_{01}+\tilde{\Phi}_{21}\right)+2rM\left(\frac{\Delta}{2\rho^2}\tilde{\Phi}_{10}+\tilde{\Phi}_{12}\right)-2riK\left(\frac{\Delta}{2\rho^2}\tilde{\Phi}_{01}-\tilde{\Phi}_{21}\right)=0. \tag{5.7e}$$

$$\frac{iK}{\rho\sqrt{2}}\left(\tilde{\Psi}_2+2\tilde{\Lambda}\right)-\cot\theta\left(\frac{\Delta}{2\rho^2}\tilde{\Phi}_{01}+\tilde{\Phi}_{21}\right)-Q\left(\frac{\Delta}{2\rho^2}\tilde{\Phi}_{10}-\tilde{\Phi}_{12}\right)=0. \tag{5.7f}$$

It is easy to see that the only solution to the system (5.7) is the trivial solution, and this must, therefore, also be the only solution in the Kerr background. So it has been shown that sybsystem C_1 ensures that all other N.P. equations are satisfied,

$$\tilde{\Phi}_{11}=\tilde{\Phi}_{21}=\tilde{\Phi}_{01}=0=\tilde{\Psi}_2=\tilde{\Lambda}, \tag{5.8}$$

for the restricted approximate analysis under consideration.

VI. SUFFICIENCY OF SUBSYSTEMS C_2 AND C_3

In this section the subsystems considered in the perturbation analysis[12] of Bell–Szekeres space-time[13] are examined and shown to be sufficient.

The Bell–Szekeres space-time is conformally flat,

$$\Psi_0=\Psi_1=\Psi_2=\Psi_3=\Psi_4=0. \tag{6.1}$$

In the coordinate and tetrad systems used in region I the only nonzero spin coefficients are

$$\epsilon=-\gamma=\cot\varphi/2\sqrt{2} \text{ and } \alpha=-\beta=\cot\theta/2\sqrt{2}, \tag{6.2}$$

and the only nonzero Maxwell scalar is

$$\phi_1=1/\sqrt{2}. \tag{6.3}$$

In the usual way first-order changes are considered for all the tetrad vectors, spin coefficients, and Weyl and Maxwell tensor components; the same conventions and notation are used as in the case for perturbations about Kerr.

The subsystem C_2 consists of all the nontrivial Maxwell equations together with the following N.P. equations:

$$X_{mnp}=0, \quad \text{for all } m,n,p,q,s, \tag{6.4}$$

$$Y_{1314}=Y_{1313}=Y_{1312}=0=Y_{2441}=Y_{2431}=Y_{2421}, \tag{6.5a}$$

$$Y_{2442}=Y_{3143}=Y_{2443}=0=Y_{2423}=Y_{1332}=Y_{1324}, \tag{6.5b}$$

$$V_{sm[npq]}=0, \quad \text{for all } m,n,p,q,s, \tag{6.6}$$

This means that the only equations of the N.P. system not considered explicitly are the six equations

$$Y_{3414}-Y_{1213}=Y_{1213}-Y_{3413}=Y_{1232}-Y_{3432}=Y_{1242}-Y_{3442}=0, \tag{6.7a}$$

$$Y_{1212}-Y_{3412}=Y_{1234}-Y_{3434}=0. \tag{6.7b}$$

But when Eqs. (6.4)–(6.6) are substituted into the first of Papapetrou's identities (2.7), it is clear that the first four of these remaining equations (6.7a) are also satisfied by virtue of the elimination relations. So in the N.P.-type notation used in Sec. III the only equations still outstanding are

$$\tilde{\Psi}_2+\tilde{\Phi}_{11}-\tilde{\Lambda}=0, \tag{6.8a}$$

$$-\tilde{\Psi}_2+\tilde{\Phi}_{11}+\tilde{\Lambda}=0, \tag{6.8b}$$

where it is known from (6.5) that

$$\tilde{\Psi}_2+2\tilde{\Lambda}=0. \tag{6.9}$$

Since the background metric is conformally flat, there is no contribution from identity (2.8); the only constraints coming from identity (2.9) [using (6.9)] are

$$D(\tilde{\Phi}_{11}+3\tilde{\Lambda})=0,\quad \Delta(\tilde{\Phi}_{11}+3\tilde{\Lambda})=0,\quad \delta(\tilde{\Phi}_{11}-3\tilde{\Lambda})=0. \tag{6.10}$$

It is obvious that in general these four equations are not sufficient to ensure $\tilde{\Phi}_{11}$, $\tilde{\Lambda}$ identically zero. However, since these have an x^1 and x^2 dependence given by

$$e^{i(k_1x^1+k_2x^2)}, \tag{6.11}$$

and on such quantities the operators have the forms

$$D=+(1/\sqrt{2})(\partial_\varphi-ik_2\csc\varphi), \tag{6.12a}$$

$$\Delta=+(1/\sqrt{2})(\partial_\varphi+ik_2\csc\varphi), \tag{6.12b}$$

$$\delta=-(1/\sqrt{2})(\partial_\theta-k_1\csc\theta), \tag{6.12c}$$

$$\delta^*=-(1/\sqrt{2})(\partial_\theta+k_1\csc\theta), \tag{6.12d}$$

it is easily seen that the only solution to (6.10) in these circumstances is

$$\tilde{\Phi}_{11}=0=\tilde{\Lambda}. \tag{6.13}$$

So it has been shown that C_2 is a sufficient subsystem of the N.P. system for region I of Bell–Szekeres space-time in the chosen coordinate and tetrad system.

In region II of Bell–Szekeres space-time the subsystem C_3 used consists of all the nontrivial Maxwell equations and all of the N.P. equations except the three Ricci equations,

$$Y_{2431}=0, \tag{6.14a}$$

$$Y_{2443}=0, \tag{6.14b}$$

$$Y_{2423}=0. \tag{6.14c}$$

But the first two of these equations are identically satisfied because of identity (2.7), leaving only the outstanding equation (6.13c) subject to two constraints from identity (2.9):

$$D\tilde{\Phi}_{22}=0, \tag{6.15a}$$

$$\delta^*\tilde{\Phi}_{22}=0. \tag{6.15b}$$

Once again these equations, in general, are not sufficient to ensure $\tilde{\Phi}_{22}$ identically zero, but for the type of coordinate dependence being considered it is obvious that the only possible solution is

$$\tilde{\Phi}_{22}=0. \tag{6.16}$$

So it has been shown that C_3 is a sufficient subsystem of the N.P. system for region II of Bell–Szekeres space-time in the chosen coordinate and tetrad system.

VII. SUMMARY AND DISCUSSION

It has been pointed out that because of the inherent redundancy within the complete N.P. system of equations

many possible choices of sufficient subsystems can be found; the particular choice will depend on the type of analysis being undertaken. Papapetrou's indentities[8] provide a means of confirming directly the sufficiency of any subsystem.

In particular, it has been shown that subsystems C_1,C_2,C_3 are sufficient subsystems for the restricted approximate analysis in which they have been used. One cannot in general expect a subsystem that is sufficient for one particular analysis to be sufficient in an (even slightly) different analysis. For each particular analysis the potential sufficient subsystem can be confirmed directly from Papapetrou's identities.

Papapetrou's identities enable one to see exactly where and why the redundancy arises. These identities in general are quite complicated in the relations they show between the three different sets of equations from the N.P. system; but their structure is considerably simpler when all of the commutator equations

$$X_{mnp}=0, \quad \text{for all } m,n,p, \tag{7.1}$$

are included within the sufficient subsystem. In this case the first identity (2.7) reduces to

$$Y_{m[npq]}=0, \tag{7.2}$$

the elimination relations which link together some of the Ricci equations. If all of the Bianchi equations

$$V_{sm[npq]}=0, \quad \text{for all } m,n,p,q,s, \tag{7.3}$$

are also included within the sufficient subsystem, then three more simple algebraic relations (3.6) linking the Ricci equations are given by identity (2.9).

However, identity (2.8) provides a more complicated differential link between the Ricci equations. As has been seen in Sec. V, not only does this identity give rise there to algebraic compatibility conditions, but more generally, since it is a set of differential equations, it itself has its own nontrivial integrability conditions, which give rise to yet another set of identities between the Ricci equations. In Sec. V these compatibility conditions and higher integrability conditions are written out for the simple Schwarzschild background metric; however, if these conditions had been written in the Kerr background metric, they would have given a much more complicated set of realtions linking the Ricci equations. When the Ricci equations themselves are written out explicitly in terms of the Teukolsky functions, from Chandrasekhar's analysis, and these complicated expressions are substituted into the already complicated identity (2.8) and its associated compatibility and higher integrability conditions, then it is clear that very complicated identities for the Teukolsky functions will result. The identity (2.8) therefore provides a source for at least some of the deep integral relations found by Chandrasekhar in his perturbation analysis of the Kerr metric.

So the occurrence of deep integral relations is also something fundamental to the N.P. formalism, linked to its inherent redundancy; such relations can be expected, even predicted, in any analysis that relies on solving a sufficient subsystem of the N.P. system. Of course additional identities will occur in particular cases, especially when specific choices of tetrad and coordinate system are made. A fuller discussion of the integral relations associated with the perturbations of the Kerr metric will be presented elsewhere.

It is obvious that subsystems C_2 and C_3 have more equations than subsystem C_1. The Bell–Szekeres space-time is conformally flat, and in both regions the choice of tetrad and coordinate system reduce almost all zeroth-order spin coefficients to zero; this means that a greater number of Papapetrou's identities are trivially satisfied for these subsystems compared to the situation for subsystem C_1, which is associated with the less specialized Kerr space-time. It is clear that the more specialized and symmetric the space-time, the "weaker" will be Papapetrou's identities, and so more equations will be needed to constitute a sufficient subsystem.

Finally it is noted that for subsystems C_2 and C_3 only the "already linearized equations" were needed to ensure a sufficient subsystem; but subsystem C_1 required additional (more complicated) equations alongside its already linearized ones. The most efficient way to supplement the already linearized equations will be considered in a separate paper.

[1]S. Chandrasekhar, Proc. R. Soc. London Ser. A **358**, 421 (1978).
[2]S. Chandrasekhar, Proc. R. Soc. London Ser. A **358**, 441 (1978).
[3]S. Chandrasekhar, Proc. R. Soc. London Ser. A **365**, 425 (1979).
[4]S. Chandrasekhar, Proc. R. Soc. London Ser. A **372**, 475 (1980).
[5]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[6]E. Newman and R. Penrose, J. Math. Phys. **17**, 1226 (1962).
[7]A. Papapetrou, C. R. Acad. Sci. **272**, 1537 (1971a).
[8]A. Papapetrou, C. R. Acad. Sci. **272**, 1613 (1971b).
[9]A. Papapetrou, Ann. Inst. H. Poincaré **13**, 271 (1970).
[10]S. B. Edgar, Gen. Relativ. Gravit. **12**, 347 (1980).
[11]S. B. Edgar, Int. J. Theor. Phys. **18**, 251 (1979).
[12]S. Chandrasekhar and B. C. Xanthopoulus, Proc. R. Soc. London Ser. A **420**, 93 (1988).
[13]P. Bell and P. Szekeres, Gen. Relativ. Gravit. **5**, 275 (1974).
[14]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 223 (1985).
[15]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **410**, 311 (1987).
[16]F. A. E. Pirani, *Lectures on General Relativity, Brandeis Summer Institute* (Prentice–Hall, Englewood Cliffs, NJ, 1964), p. 349.
[17]See Ref. 5, p. 39.
[18]See Ref. 5, pp. 47–48.

Ricci and contracted Ricci collineations of the Robertson–Walker space-time

L. Núñez
Laboratorio de Física Teórica, Departamento de Física, Universidad de Los Andes, Mérida, 5101, Venezuela

U. Percoco and V. M. Villalba[a)]
Centro de Física, Instituto Venezolano de Investigaciones Científicas, Apartado 21827, Caracas 1020A, Venezuela

(Received 28 December 1988; accepted for publication 6 September 1989)

Ricci collineations and contracted Ricci collineations of the Robertson–Walker metric, associated with a vector field of the form $\xi = (\xi^0(t,r),\xi^1(t,r),0,0)$ are presented.

I. INTRODUCTION

In addition to isometries, space-times may admit other symmetries that do not leave the metric tensor invariant.[1] These collineations have been classified by Katzin *et al.*[1,2] and their relation to constants of motion has been indicated.[2–4]

Robertson–Walker space-times, which are described by the metric[5]

$$ds^2 = dt^2 - R^2(t)\left(\frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\right), \tag{1.1}$$

with $k = 0, \pm 1$, provide physically relevant examples to study these geometrical properties.

It has been recently shown that these space-times admit affine collineations, i.e., symmetries that leave invariant the equation of the geodesic. In particular, Bedran and Lesche[6] have given an example of homothetic motion corresponding to $R(t) = at + b$. Another homothetic collineation has been found by Maartens,[7] who has also pointed out the existence of a proper affine collineation associated with the static case $\dot{R} = 0$. Collinson[8] has proved the uniqueness of this proper affine collineation.

A less restrictive class of symmetries corresponds to Ricci collineations[1,2] (RC)

$$\mathcal{L}_\xi R_{\mu\lambda} = \xi^\nu \nabla_\nu R_{\mu\lambda} + R_{\mu\nu}\nabla_\lambda \xi^\nu + R_{\nu\lambda}\nabla_\mu \xi^\nu = 0, \tag{1.2}$$

and to the family of contracted Ricci collineations (FCRC)[9]

$$g^{\alpha\beta}\mathcal{L}_\xi R_{\alpha\beta} = 0. \tag{1.3}$$

Green *et al.*[10] have provided an example of both types of symmetries for Robertson–Walker space-times. These authors have confined their study to symmetries generated by a vector field of the form

$$\xi = (\xi^0(t,r,\theta,\phi),0,0,0).$$

In this paper we further investigate the symmetry properties of the Robertson–Walker space-times by considering Ricci collineations and the family of contracted Ricci collineations associated to a vector field ξ of the form

$$\xi = (\xi^0(t,r), \xi^1(t,r),0,0). \tag{1.4}$$

The paper is organized as follows: in Sec. II we study Ricci collineations while the contracted Ricci collineations are dealt with in Sec. III.

II. RICCI COLLINEATIONS

The nonvanishing components of the Ricci tensor, corresponding to the metric (1.1), read

$$R_{00} = -3\ddot{R}/R, \tag{2.1}$$

$$R_{ii} = -g_{ii}\,(\Delta/R^2), \tag{2.2}$$

where a dot indicates derivative with respect to time. Also, in (2.2), where $i = 1,2,3$, there is no sum over repeated indices and we have defined

$$\Delta = 2k + 2\dot{R}^2 + \ddot{R}R. \tag{2.3}$$

In the next section, in order to single out some of the proper FCRC, it is convenient to study RC first. Substituting (1.4) in Eq. (1.2) we obtain the following equations:

$$\xi^0 R_{11,0} + \xi^1 R_{11,1} + 2R_{11}\xi^1{}_{,1} = 0, \tag{2.4}$$

$$\xi^0 R_{00,0} + 2R_{00}\xi^0{}_{,0} = 0, \tag{2.5}$$

$$R_{00}\xi^0{}_{,1} + R_{11}\xi^1{}_{,0} = 0, \tag{2.6}$$

$$\xi^0 R_{22,0} + \xi^1 R_{22,1} = 0, \tag{2.7}$$

$$\xi^0 R_{33,0} + \xi^1 R_{33,1} = 0. \tag{2.8}$$

Equation (2.7) is equivalent to (2.8), therefore we are left with four independent equations. Integrating, we have

$$\xi^0 = c(1-kr^2)^{1/2}|R_{00}|^{-1/2}, \tag{2.9}$$

$$\xi^1 = g(t)\,r[1-kr^2]^{1/2}, \tag{2.10}$$

where

$$g(t) = -(c/|R_{00}|^{1/2})\,(\dot{\Delta}/2\Delta), \tag{2.11}$$

where c is a constant.

Equations (2.4)–(2.7) also provide an independent expression for $\dot{g}(t)$:

$$\dot{g}(t) = c\,k(|R_{00}|^{1/2}/\Delta). \tag{2.12}$$

From (2.11) and (2.12) an integrability condition for (1.2) emerges:

$$\tfrac{1}{2}\dot{R}_{00}\Delta\dot{\Delta} - R_{00}(\Delta\ddot{\Delta} - \dot{\Delta}^2) = 2k\Delta R_{00}^2. \tag{2.13}$$

Equations (2.9)–(2.13) have been obtained assuming that R_{00} and Δ do not vanish. Nevertheless there may exist

[a)] Postal address: Apartado 4282, Caracas 1010A, Venezuela.

 0022-2488/90/010137-03$03.00

solutions that correspond to the vanishing of any of the above quantities.

It is therefore advantageous to explore the above possibilities.

(i) $\Delta = 0$, $R_{00} \neq 0$. The vanishing of Δ implies $R_{ii} = 0$ which represents a general relativistic fluid with the stiff matter equation of state, that is, $\rho = p$.[11–13]

The equation $\Delta = 0$ can be integrated leading to

$$t = 3\int \frac{R^2\, dr}{(c - 9kR^4)^{1/2}}, \tag{2.14}$$

where c is a positive constant. Three cases emerge as a consequence of this integral.

(ia) $k = 0$. This gives

$$R = c^{1/6} t^{1/3}. \tag{2.15}$$

(ib) $k = 1$. In this case

$$t = (c/36)^{1/4}[2E(\alpha,1/\sqrt{2}) - F(\alpha,1/\sqrt{2})], \tag{2.16}$$

$$\alpha = \sin^{-1}[\sqrt{2}\sin(z/2)],$$

where $F(\alpha,1/\sqrt{2})$ and $E(\alpha,1/\sqrt{2})$ are elliptic integrals of the first and second kind, respectively. The above result is valid for

$$0 < z < \pi/2, \quad R^2 = (c^{1/2}/3)\cos z, \tag{2.17}$$

which implies

$$0 < R^2 < c^{1/2}/3. \tag{2.18}$$

This solution represents a closed oscillating universe.

(ic) $k = -1$. This case leads to

$$t = \left(\frac{c}{144}\right)^{1/4}\left[F(\alpha,r) - 2E(\alpha,r) + 2\frac{\sqrt{\sinh z(1+\sinh^2 z)}}{1+\sinh z}\right], \tag{2.19}$$

where

$$\alpha = \cos^{-1}\frac{1-\sinh z}{1+\sinh z}, \quad r = \frac{1}{\sqrt{2}},$$

with

$$z > 0, \quad R^2 = (c^{1/2}/3)\sinh z.$$

In order to obtain the collineations associated to these space-times, we go back to Eqs. (2.4)–(2.7) obtaining

$$R_{00}\xi^0{}_{,1} = 0, \tag{2.20}$$

$$\xi^0 R_{00,0} + 2R_{00}\xi^0{}_{,0} = 0. \tag{2.21}$$

In these equations ξ^1 emerges as an arbitrary function of t and r. The component ξ^0, on the other hand, is determined by (2.20) and (2.21) which indicates that $R(t)$ is not further restricted and ξ^0 is forced to depend on the time only, as follows:

$$\xi^0 = c/|R_{00}|^{1/2}, \tag{2.22}$$

where c is a constant.

(ii) $\Delta = \text{const} \neq 0$, $R_{00} = 0$. This represents a matter satisfying the equation of state $\rho + 3p = 0$. Also, it implies that the curvature tensor has no components normal to the homogeneous hypersurfaces.[11,12,19]

$$R(t) = at + b, \quad a,b = \text{const}. \tag{2.23}$$

As a result of (2.4)–(2.7) we obtain that ξ^1 vanishes and ξ^0 remains an underdetermined function of t and r.

When $\Delta R_{00} \neq 0$ we can obtain the RC and the corresponding space-times for the case $k = 0$. In these cases the integrability condition (2.13) can be integrated once yielding

$$g(t) = -cR_{00}^{-1/2}(\dot{\Delta}/2\Delta) = \text{const}. \tag{2.24}$$

Then the collineation vector takes the form

$$\xi^0 = c/|R_{00}|^{1/2}, \tag{2.25}$$

$$\xi^1 = g(t)r. \tag{2.26}$$

A one-parameter family of space-times, consistent with (2.24), corresponds to

$$R(t) = \beta e^{\alpha t}. \tag{2.27}$$

This represents an empty de Sitter space-time. Since this is an Einstein space, all the collineations are motions.[1]

The solution of Eq. (2.24) given above is obviously not unique. Another solution is

$$R(t) = \beta t^{\alpha}, \quad \alpha \neq 1. \tag{2.28}$$

This choice of $R(t)$ produces a different Ricci collineation. However the RC associated to (2.28) is not a proper Ricci collineation; it corresponds to the homothetic motion found by Maartens.[7]

Another particular solution to (2.24) is $\dot{\Delta} = 0$. In this case ξ^1 vanishes while ξ^0 is still given by (2.25). This corresponds to the RC found by Green *et al.*[10]

III. FAMILY OF CONTRACTED RICCI COLLINEATIONS

Contracted Ricci collineations are defined by

$$g^{\alpha\beta}\mathcal{L}_{\xi}R_{\alpha\beta} = 0, \tag{3.1}$$

which for the Robertson–Walker metric (1.1) takes the form

$$\xi^0\left(R_{00,0} - \frac{3\dot{\Delta}}{R^2}\right) + 2R_{00}\xi^0_{,0} - \xi^1\left(\frac{2kr}{1-kr^2}\frac{\Delta}{R^2} + \frac{4\Delta}{rR^2}\right) - \frac{2\Delta}{R^2}\xi^1_{,1} = 0. \tag{3.2}$$

A first example of proper FCRC is obtained by setting $\Delta = 0$ in the above equation. Following Sec. II, we have integrated the equation $\Delta = 0$, finding $R(t)$. In the present case ξ^1 is undetermined and Eq. (3.2) becomes

$$2R_{00}\xi^0{}_{,0} + R_{00,0}\xi^0 = 0. \tag{3.3}$$

This equation is the same as (2.5) and can be integrated demanding that R_{00} be different from zero. Thus

$$\xi^0(t,r) = h(r)/|R_{00}|^{1/2}, \tag{3.4}$$

where $h(r)$ is arbitrary.

Now, for the case $\Delta \neq 0$ we introduce solutions of the form

$$\xi^1 = u(t)\, v(r) \tag{3.5}$$

and

$$\xi^0 = f(t)\, h(r) \tag{3.6}$$

into Eq. (3.2). Then, it is possible to separate variables, obtaining

$$v'(r) + (kr/(1-kr^2) + 2/r)v(r) + ch(r), \tag{3.7}$$

$$2R_{00}\dot{f}(t) + (R_{00,0} - 3\dot{\Delta}/R^2)f(t) = c(2\Delta/R^2)\,u(t), \tag{3.8}$$

where c is a constant.

These equations enable us to find two of the four functions needed to determine ξ^0 and ξ^1. Now, in order to leave $u(t)$ and $g(r)$ arbitrary, we set $c=0$. Integration of these equations leads to the proper family of contracted Ricci collineations described by

$$\xi^1(t,r) = u(t)|1-kr^2|^{1/2}/r^2 \tag{3.9}$$

and

$$\xi^0(t,r) = h(r)/\ddot{R}\,R^2, \tag{3.10}$$

where it is assumed that $R_{00}\neq 0$. As an example we consider the collineations associated to

$$R(t) = \beta e^{\alpha t}. \tag{3.11}$$

This particular form of $R(t)$ has already been considered in Sec. II. Nevertheless, since now k can be different from zero, (3.11) does not represent an empty de Sitter space. It turns out that ξ^1 is given by (3.9) while ξ^0 takes the form

$$\xi^0(t,r) = h(r)e^{-3\alpha t}. \tag{3.12}$$

Let us consider now $R_{00}=0$. Then

$$R(t) = \alpha t + \delta, \quad \alpha,\delta = \text{const.} \tag{3.13}$$

In this case, (3.9) and (3.10) are not valid, going back to Eqs. (3.7) and (3.8), with $c=0$, we find that ξ^0 is undetermined, while ξ^1 keeps the form (3.9).

ACKNOWLEDGMENTS

We wish to express our gratitude to the referee for many important suggestions.

This work was partially supported by the Consejo de Desarrollo Científico y Technológico de la Universidad de Los Andes (CDCHT-ULA) under project 379-89-05. One of the authors (VMV) acknowledges CONICIT for financial support.

[1] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1960).
[2] G. H. Katzin and J. Levine, Colloq. Math. (Poland) **XXVI**, 21 (1972).
[3] G. H. Katzin and J. Levine, J. Math. Phys. **22**, 1878 (1981).
[4] S. Hojman, L. Núñez, A. Patiño, and H. Rago, J. Math. Phys. **27**, 281 (1985).
[5] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1975).
[6] M. L. Bedran and B. Lesche, J. Math. Phys. **27**, 2360 (1986).
[7] R. Maartens, J. Math. Phys. **28**, 2051 (1987).
[8] C. D. Collinson, J. Math. Phys. **29**, 1972 (1988).
[9] W. R. Davis, L. H. Green, and L. K. Norris, Nuovo Cimento B **34**, 256 (1976).
[10] L. H. Green, L. K. Norris, D. R. Oliver Jr., and W. R. Davis, Gen. Rel. Grav. **8**, 731 (1977).
[11] H. Stephani, *General Relativity* (Cambridge U. P., London, 1982).
[12] D. R. Oliver Jr. and W. R. Davis, Ann. Inst. Henri Poincaré **30**, 339 (1979).
[13] D. R. Oliver Jr. and W. R. Davis, Gen. Rel. Grav. **8**, 905 (1977).

General exact solution for homogeneous time-dependent self-gravitating perfect fluids

Patricio Gaete
Centro Brasileiro de Pesquisas Fisicas, CNPq/CBPF, rua Xavier Sigaud 150, 22290 Rio de Janeiro, Brazil
Roberto Hojman
Departamento de Fisica, Facultad de Ciencias, Universidad de Chile, casilla 653, Santiago, Chile and Centro de Fisica El Trauco, Fanor Velasco 41 D, Santiago, Chile

(Received 25 August 1988; accepted for publication 23 August 1989)

A procedure to obtain the general exact solution of Einstein equations for a self-gravitating spherically symmetric static perfect fluid obeying an arbitrary equation of state is applied to time-dependent Kantowski–Sachs line elements (with spherical, planar, and hyperbolic symmetry). As in the static case, the solution is generated by an arbitrary function of the independent variable and its first derivative. To illustrate the results, the whole family of (plane-symmetric) solutions with a "gamma-law" equation of state is explicitly obtained in terms of simple known functions. It is also shown that, while in the static plane-symmetric line element, every metric is in one to one correspondence with a "partner metric" (both originated from the same generatrix function); in this case every generatrix function uniquely determines one metric.

I. INTRODUCTION

In this paper we extend a procedure originally conceived to find the general static solution of self-gravitating perfect fluids[1,2] to the case of time-dependent distributions of matter.

The approach consists of looking at the differential equations for the metric coefficients without appealing *a priori* to any equation of state for the self-gravitating perfect fluid. This allows the introduction of an arbitrary function G in terms of which it is possible to determine all the relevant unknown functions. Thus the explicit form of the equation of state becomes fixed only after the integration has been carried out.

It is not claimed that the method necessarily provides a useful tool to find new solutions. Rather, it should be understood as a device to gain extra insight about the structure of the solutions of Einstein equations or as a possible alternative way to classify their solutions.

Nonetheless, because of its physical importance and to gain some acquaintance with the method, an example is exhibited.

Apart from the general form of the solution, two items regarding the features of the used mechanism deserve special mention.

First, it should be remembered that if two solutions have the same equation of state, they may correspond to the same solution written in different coordinates. However, as will be seen, solutions labeled by a different function G truly correspond to different solutions as they do induce different equations of state.

Second, in the static case with planar symmetry the solutions appeared in couples: every solution induced a partner that in turn generated back its own seed. However, in this case an important deviation from such a situation occurs. The particular combination of factors in the quadrature is such that the integrals, although similar, are simpler than in the static case and the problem is reduced to a single integral depending only on t and G (and not on its first derivative).

In Sec. II, we write the Einstein field equations for spherical, plane, and hyperbolic symmetry when the metric coefficients and the matter parameters exhibit time dependence only. Then, in Sec. III, we mimic the method already used in the static case to get the general solution of the equations. In Sec. IV, the plane-symmetric case is studied in more detail and the most general plane-symmetric metric obeying a gamma-law equation of state is obtained using the proposed scheme. In Sec. V, we present our conclusions.

II. FIELD EQUATIONS

Consider the Kantowski–Sachs-type line element

$$ds^2 = dt^2 - A(t)dr^2 - B(t)[d\theta^2 + \Sigma(\theta,K)d\varphi^2], \quad (2.1)$$

where the metric coefficients A and B depend only on t and

$$\Sigma(\theta,K) = \begin{cases} \sin^2\theta, & K=1, \\ \theta, & K=0, \\ \sinh^2\theta, & K=-1. \end{cases} \quad (2.2)$$

The parameter K says whether the matter distribution has spherical $(K=1)$, plane $(K=0)$, or hyperbolic $(K=-1)$ symmetry.

A slightly different form of the metric is obtained if $B(t)$ (instead of t itself) is used as an independent variable:

$$ds^2 = g^2(t)dt^2 - h^2(t)dr^2 - t^2[d\theta + \Sigma(\sigma,K)d\varphi^2] \quad (2.3)$$

(the new time variable has been renamed t). Even if the following analysis is valid—strictly speaking—just in the coordinates defined by Eq. (2.3), only slight modifications have to be introduced to take into account other coordinates.

If the gravitational field is generated by a perfect fluid, the associated energy–momentum tensor reads

$$T_{\mu\nu} = (p+\rho)u_\mu u_\nu - pg_{\mu\nu}, \quad (2.4)$$

 0022-2488/90/010140-04$03.00

where p is the matter pressure, ρ is the energy density, and

$$u^\mu = (1/g)\delta^\mu_0 \tag{2.5}$$

is the matter four-velocity.

The Einstein field equations

$$G^\mu{}_\nu = T^\mu{}_\nu \tag{2.6}$$

are

$$\frac{K}{t^2} + \frac{2}{tg^2}\left(\frac{\dot h}{h} + \frac{1}{2t}\right) = \rho, \tag{2.7a}$$

$$-\frac{K}{t^2} + \frac{2}{tg^2}\left(\frac{\dot g}{g} - \frac{1}{2t}\right) = p, \tag{2.7b}$$

$$-\frac{1}{g^2}\left[\frac{\ddot h}{h} - \frac{\dot h \dot g}{hg} + \frac{1}{t}\left(\frac{\dot h}{h} - \frac{\dot g}{g}\right)\right] = p, \tag{2.7c}$$

whereas the hydrodynamic equation

$$T^\mu{}_{\nu;\mu} = 0 \tag{2.8}$$

is

$$\dot\rho = -(p+\rho)(\dot h/h + 2/t), \tag{2.9}$$

and, as usual, it can be also deduced from the Bianchi identities for the curvature tensor constructed from the line element (2.1).

The system of equations (2.7) serves—in principle—to determine both the metric coefficients (g^2 and h^2) and the thermodynamic variables (p and ρ).

However, unless an equation of state

$$p = p(\rho) \tag{2.10}$$

linking p and ρ is introduced, the problem remains undetermined (only three equations for four unknowns).

Solving (2.7) together with (2.9) can become extremely difficult. Usually, when handling this kind of system, a functional dependence of one of the unknown functions (say ρ) on the independent variable (t, in the present case) is given by hand.

In such a way, the three remaining differential field equations can be solved for g^2, h^2, and p and an equation of state is determined *a posteriori*.[3] For instance, take ρ as given. Then,

$$12m\dot m - 2t(\rho t^2 - 4K)\dot m + 2t^2(3\rho - 2\dot\rho t)m + t^3 + t^3(4K\rho - \rho^2 t^2 - 2Kt\dot\rho) = 0, \tag{2.11}$$

where $\dot m = pt^2/2$.

Nonetheless, there are few choices of ρ leading to an exactly solvable differential equation for m.

III. THE METHOD

Surprisingly, there exists another prescription by means of which it is possible to get the general exact solution of the problem in terms of quadratures.

In fact, Eq. (2.7b) can be rewritten as

$$\frac{1}{t^2}\frac{d}{dt}\left[t\left(K + \frac{1}{g^2}\right)\right] = -p \tag{3.1}$$

and then integrated, giving

$$1/g^2 = -[K + 2m(t)/t], \tag{3.2}$$

where

$$\frac{dm(t)}{dt} = \frac{1}{2}pt^2. \tag{3.3}$$

If Eq. (3.2) and $\dot h/h$ from Eq. (2.9) are introduced into Eq. (2.7a), one gets

$$(2m + Kt)\left(\frac{3}{t^3} + \frac{2}{t^2}\frac{\dot\rho}{p+\rho}\right) = \rho - \frac{K}{t^2}. \tag{3.4}$$

Now define

$$G(t) \equiv [2m(t) + Kt]/[\rho(t) - K/t^2]. \tag{3.5}$$

Then, Eq. (3.4) becomes

$$\dot\rho - \frac{(t^3 - 3G)(t^2 + \dot G)}{G(t^3 + 3G)}\rho + \frac{K(t^3 - 3G)(t^3 + t\dot G - 2G)}{t^3 G(t^3 + 3G)} = 0. \tag{3.6}$$

The last equation is first order and linear in ρ and can be integrated at once if G is a given function of t. In such a case

$$\rho(t) = e^{I(t)}\left[\rho_0 + K\int dt \times \frac{(3G - t^3)(t^3 + t\dot G - 2G)}{t^3 G(t^3 + G)} e^{-I(t)}\right], \tag{3.7}$$

where ρ_0 is an integration constant and

$$I(t) = \int dt \frac{(t^3 - 3G)(t^2 + \dot G)}{G(t^3 + 3G)}. \tag{3.8}$$

Moreover, from Eq. (3.3) and the definition (3.5) of G,

$$p(t) = (1/t^2)[\rho\dot G + G\dot\rho - (K/t^3)(t^3 + t\dot G - 2G)]. \tag{3.9}$$

Also,

$$g^2 = [(1/t)(K/t^2 - \rho)G]^{-1} \tag{3.10}$$

and

$$h^2 = (h_0^2/t)e^{-J(t)} \tag{3.11}$$

where

$$J(t) = \int dt \frac{t^2}{G} \tag{3.12}$$

and h_0 is an integration constant.

Considering G as a given function of t is in a sense equivalent to postulating an equation of state. In fact, once a choice for G in terms of t is made, Eq. (3.5) links p and m. Such a relationship cannot be directly understood as an equation of state because it lacks the invariance property under arbitrary coordinate transformation it should have. (In any case, it would be more satisfactory to have a relationship involving p and ρ rather than p and m.)

IV. AN EXAMPLE: PLANE SYMMETRY

Consider now the $K = 0$ case, which describes plane-symmetric Bianchi type I models.

Then, from Eq. (3.7),

$$\rho = \rho_0 e^{I(t)}, \tag{4.1}$$

where I is given by Eq. (3.8).

A little algebra shows that

$$I(t) = \ln[G/(t^3 + 3G)^2] + J(t). \tag{4.2}$$

Then, from Eq. (4.1),

$$\rho = \rho_0[G/(t^3 + 3G)^2]e^{J(t)}. \tag{4.3}$$

Also, from Eqs. (3.6) and (3.9) (for $K = 0$),

$$p = [(t^3 + 2t\dot{G} - 3G)/(t^3 + 3G)]\rho \tag{4.4}$$

and, from Eq. (3.10),

$$g^2 = -t/\rho G. \tag{4.5}$$

To illustrate the above formulas let us find—following Ref. 4—the plane-symmetric line element generated when the space is filled with a perfect fluid obeying a "gamma-law" equation of state:

$$p = (\gamma - 1)\rho, \tag{4.6}$$

for any γ such that $1 < \gamma < 2$.

Substituting Eq. (4.6) into Eq. (4.4),

$$2t\dot{G} - 3\gamma G + (2 - \gamma)t^3 = 0, \tag{4.7}$$

which is solved by

$$G = \begin{cases} -\frac{1}{3}t^3 + G_0 t^{3\gamma/2}, & \gamma \neq 2, \\ G_0 t^3, & \gamma = 2, \end{cases} \tag{4.8}$$

where G_0 is an integration constant.

Inserting (4.8) into Eq. (3.12),

$$\exp J(t) = \begin{cases} t^{3\gamma/(2-\gamma)} G^{2/(\gamma-2)}, & \gamma \neq 2, \\ t^{G_0}, & \gamma = 2. \end{cases} \tag{4.9}$$

Then, from Eqs. (3.11), (4.3), (4.6), and (4.5) one gets, respectively,

$$h^2(t) = \begin{cases} h_0^2 t^{2(\gamma+1)/(\gamma-2)} G^{2/(2-\gamma)}, & \gamma \neq 2, \\ h_0^2 t^{-(1+G_0)}, & \gamma = 2; \end{cases} \tag{4.10}$$

$$(\gamma - 1)\rho(t) = p(t) = \begin{cases} [\rho_0(\gamma - 1)/9G_0^2] t^{3\gamma(\gamma-1)/(2-\gamma)} G^{\gamma/(\gamma-2)}, & \gamma \neq 2, \\ [\rho_0 G_0(\gamma - 1)/(1 + 3G_0)^2] t^{G_0 - 3}, & \gamma = 2; \end{cases} \tag{4.11}$$

$$g^2(t) = \begin{cases} -(9G_0^2/\rho_0) t^{(2+2\gamma-3\gamma^2)/(2-\gamma)}, & \gamma \neq 2, \\ -[(1 + 3G_0)^2/\rho_0 G_0] t^{-G_0}, & \gamma = 2. \end{cases} \tag{4.12}$$

V. CONCLUSIONS

As has been stated elsewhere,[1,2] expressions (3.7)–(3.11) constitute the whole set of solutions of Eqs. (2.7). No spurious solutions have been introduced anywhere, as can be proved by direct substitution of the solutions into the field equations. On the other hand, any solution can be accommodated in the present scheme; for example, consider a metric obtained using another technique. Then, by inverting relation (3.9) the generatrix function can be determined in terms of the metric coefficient h and its first derivative:

$$G = -ht^3/h_0(h + 2\dot{h}t). \tag{5.1}$$

Moreover, the solutions obtained are not merely coordinate transformations, because every solution does produce a different equation of state.

The case when $\rho = K/t^2$ should be considered separately, as the definition of G loses its meaning [see Eq. (3.5)]. From Eq. (2.7a), h can be readily integrated giving

$$h = h_0/\sqrt{t}. \tag{5.2}$$

Equation (2.9) can be algebraically solved for $p(t)$:

$$p = \tfrac{1}{3}K/t^2 = \tfrac{1}{3}\rho. \tag{5.3}$$

Finally, from Eqs. (3.2) and (3.3) it is found that

$$g^2 = -(\tfrac{4}{3}K + 2m_0/t)^{-1}, \tag{5.4}$$

m_0 being an integration constant.

Thus Eq. (3.4) still holds identically: the right-hand side and the second factor of the left-hand side vanish.

As we have pointed out in the Introduction, one drawback of the present method is that the realistic matter content cannot be predicted—in general—from an *a priori* choice of the function G (e.g., a choice of G does not guarantee the positiveness of p, and the equation of state remains unknown until the integrals are carried out).

It is interesting that in the present context, some restrictions have to be imposed on the relevant parameters of the solutions if we want them to possess physical meaning. In fact, if t is the time (and that happens to be when the metrics are asymptotically flat) then it ranges from $-\infty$ to $+\infty$. Then, by looking at expression (4.8) for the generatrix function, it is seen that there are certain forbidden values of γ in order to maintain G (and the physical quantities derived from it) real.

In order to establish a comparison, let us summarize the main results obtained in the static case with plane symmetry.

The line element to be determined there is

$$ds^2 = g^2(x)dt^2 - h^2(x)dx - x^2(dy^2 + dz^2). \tag{5.5}$$

If G is defined as

$$G \equiv m/p, \tag{5.6}$$

then

$$p(x) = p_0 \exp\left[\int \frac{(x^2 + G')(x^3 + G)}{G(x^3 - G)}\, dx\right]$$
$$= p_0[G/(X^3 - G)^2]e^{J(x)}e^{8H(x)}, \tag{5.7}$$

$$\rho(x) = (1/x^2)(pG' + p'G)$$
$$= [x^3 + 2xG' + G/(x^3 - G)]p(x), \tag{5.8}$$

$$g^2(x) = g_0^2(e^{-J(x)}/x), \tag{5.9}$$

$$h^2(x) = -x/p(x)G, \tag{5.10}$$

where

$$J(x) = \int dx \frac{x^2}{G} \tag{5.11}$$

and

$$H(x) = \int dx \frac{x^2}{x^3 - G}. \tag{5.12}$$

Now, suppose that the integrals J and H can be completely carried out in terms of known functions. A companion metric can be determined at once. In fact, consider

$$G^* = x^3 - G. \tag{5.13}$$

Then, $J^* = H$ and $H^* = J$. [Notice that $(G^*)^* = G$.] Consequently, a star solution can be written as

$$p^*(x) = p_0[(x^3 - G)/G^2]e^{H(x)}e^{8J(x)}, \tag{5.14}$$

$$\rho^*(x) = [(7x^3 - 2xG' - G)/G]p^*(x), \tag{5.15}$$

$$(g^*(x))^2 = g_0^2(e^{-H(x)}/x), \tag{5.16}$$

$$(h^*(x))^2 = -x/p^*(x)(x^3 - G). \tag{5.17}$$

However, in the present case no other solution can be found with the above prescription: every solution is uniquely determined by G and does not induce any other solution.

The physical meaning of the generatrix function G remains unknown to our knowledge. Some attempts to extract additional information about it are presently underway,[5] especially regarding the stability of the solution (3.7)–(3.10) under perturbations by a scalar field. In particular, the exact analytical solution of the Klein–Gordon equation in a class of background metrics presented in Ref. 5 has been found.

In any case, we claim that the method applied in the present paper possesses a very attractive feature when extracting analytical information from the field equations: it allows us to handle the whole family of solutions on the same footing.

ACKNOWLEDGMENTS

We are indebted to Jorge Zanelli for enlightening discussions, encouragement, and careful reading of the manuscript. Thanks are also due to Nelson Zamorano for calling our attention to Kantowski–Sachs metrics, and to Leda Peña for her valuable comments.

This work was partially supported by Grant No. 89-958 of Fondo Nacional de Desarrollo Científico y Tecnológico, Chile.

[1] R. Hojman, and J. Santamarina, J. Math. Phys. **25**, 1973 (1984).
[2] S. Berger, R. Hojman, and J. Santamarina, J. Math. Phys. **28**, 2949 (1987).
[3] D. Kramer, H. Stephani, M. MacCallum, and E. Hertl, *Exact Solutions of Einstein's Equations* (Cambridge U.P., Cambridge, 1980).
[4] C. Collins, J. Math. Phys. **26**, 2268 (1985).
[5] R. Hojman, L. Peña, and N. Zamorano, "Exact relativistic models for spiral galaxies," in preparation.

Matching *pp* waves to the Kerr metric

C. Hoenselaers and F. J. Ernst[a)]
Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 8046 Garching, Federal Republic of Germany

(Received 3 July 1989; accepted for publication 23 August 1989)

Alternative pairs of impulsive *pp* waves (accompanied by shocks) that interact to produce a Kerr geometry are compared. In one case, discovered by Chandrasekhar and Xanthopoulos [Proc. R. Soc. London Ser. A **408**, 175 (1986)], analytic extension across the horizon revealed a timelike singularity analogous to the ring singularity of the Kerr metric. In another, presented here for the first time, analytic extension across the horizon reveals instead an asymptotically flat exterior Kerr geometry. The *pp* wave pulse shapes that result in the formation of these two different Kerr interaction regions are displayed graphically.

I. INTRODUCTION

Many examples of colliding plane-fronted impulsive gravitational waves have been constructed. However, we believe the solution we shall present in this paper is the first that is completely free of curvature singularities. It does have, of course, a pair of unavoidable singularities of a topological nature. Unlike a solution[1] of Chandrasekhar and Xanthopoulos (CX), to which it is closely related, the new solution admits a C^1 extension to Minkowski space after the passage of either incident plane wave. Like the CX solution, it has in the interaction region a null surface that acts as a horizon. Analytic extension across this null surface reveals, however, an asymptotically flat metric indistinguishable from the *exterior* Kerr metric.

In Boyer–Lindquist coordinates the Kerr metric may be expressed in the form

$$\begin{aligned} ds^2 = {} & \Sigma(dr^2/\Delta + d\vartheta^2) \\ & + \Sigma^{-1}\{\sin^2\vartheta\,[(r^2+a^2)d\varphi - a\,dt\,]^2 \\ & - \Delta[a\sin^2\vartheta\,d\varphi - dt\,]^2\}, \end{aligned}$$

where

$$\Sigma = r^2 + a^2\cos^2\vartheta, \qquad \Delta = r^2 + a^2 - 2\,m\,r.$$

Usually one considers the case in which $r > m + (m^2 - a^2)^{1/2}$. However, we shall be using the above metric for values of r such that

$$m + (m^2 - a^2)^{1/2}|\cos\vartheta\,| < r < m + (m^2 - a^2)^{1/2}.$$

In this selection we differ from Chandrasekhar and Xanthopoulos, who employed values of r such that

$$m - (m^2 - a^2)^{1/2} < r < m + (m^2 - a^2)^{1/2}|\cos\vartheta\,|.$$

To facilitate the conversion to null coordinate we introduce $x = p^{-1}(r/m - 1)$, $y = \cos\vartheta$, $x_1 = \phi$, and $x_2 = p\,t$, and set $a = m\,q$ and $m = 1$. Then the metric assumes the form

$$\begin{aligned} ds^2 = {} & \Sigma\big(-dx^2/(1-x^2) + dy^2/(1-y^2)\big) \\ & + \Sigma^{-1}\{(1-y^2)[[(px+1)^2+q^2]dx_1 - p^{-1}q\,dx_2]^2 \\ & + (1-x^2)\,[pq(1-y^2)dx_1 - dx_2]^2\}, \end{aligned} \tag{1}$$

where $\varepsilon = (px+1)^2 + q^2y^2$. The CX metric may be obtained from (1) simply by replacing p by $-p$.

In terms of the x-y coordinates, the horizon is located at $x = 1$, where y runs from -1 to $+1$. For $0 \leqslant x < 1$, y is allowed to assume values that run from $-x$ to x. At $y = \pm x$ occur null surfaces at which we shall match the Kerr metric to plane-fronted gravitational wave metrics. This is accomplished in the usual way. Null coordinates u and v are defined by

$$\begin{aligned} x &= u(1-v^2)^{1/2} + v(1-u^2)^{1/2}, \\ y &= u(1-v^2)^{1/2} - v(1-u^2)^{1/2}. \end{aligned} \tag{2}$$

The joinings then take place at $u = 0, 1 > v \geqslant 0$, and at $v = 0$, $1 > u \geqslant 0$, where Heaviside step functions accomplish the C^0 extensions to Petrov type N solutions of the vacuum field equations.

II. THE INCIDENT PLANE WAVES

In one of the plane wave regions the metric depends only upon the null coordinate u. Here the metric may be expressed in the form

$$\begin{aligned} ds^2 = {} & -4\Sigma(1-u^2)^{-1/2}\,du\,dv \\ & + \Sigma^{-1}(1-u^2)\{[[(pu+1)^2+q^2]\,dx_1 \\ & - p^{-1}q\,dx_2]^2 + [pq(1-u^2)\,dx_1 - dx_2]^2\}, \end{aligned} \tag{3}$$

where $\Sigma = (pu+1)^2 + q^2u^2$.

The joining of this metric to Minkowski space at $u = 0$, $v < 0$, is accomplished in the usual way, simply by substituting $u = 0$ in the above metric. The resulting joining provides a C^0 extension, associated with which there is a δ function discontinuity as well as a step discontinuity in the Weyl tensor.

In order to discuss the limit $u \to 1$ we shall substitute $u = \cos\alpha$ and treat α as small. We find that

$$ds^2 = 4\Sigma\,d\alpha\,dv + \Sigma^{-1}\alpha^2\{(\Sigma\,dx_1 - p^{-1}q\,dx_2)^2 + dx_2^2\}, \tag{4}$$

plus terms of order α^2, where $\Sigma = 2(1+p)$. This is clearly flat space, with the x's adapted to null rotations. Hence we have a C^1 extension across the hypersurface $u = 1$ to flat space, associated with which there is a step discontinuity but no δ function in the Weyl tensor. However, it should be noted that the only geodesics that can reach this null surface are

a) Present address: Department of Mathematics, Clarkson University, Potsdam, New York 13676.

 0022-2488/90/010144-03$03.00

those for which the first integrals associated with the two Killing vectors vanish, for the α^2 terms in the metric appear like an angular momentum barrier in the geodesic equation. It is therefore a moot point as to whether the extension to flat space is particularly significant.

The traditional Kerr–Schild form for *pp* waves is

$$ds^2 = -2\,dU\,dV + dY^T dY + 2Y^T F Y\,dU^2, \tag{5}$$

where $F(U)$ is an arbitrary symmetric trace-free matrix and $Y^T = (y_1, y_2)$. To transform such a metric to u, v, x_1, x_2 coordinates, we substitute

$$U = U(u), \qquad V = v + \tfrac{1}{4} Y^T (A^T A)^{\cdot} Y, \qquad Y = A(u)X, \tag{6}$$

where $A(u)$ is nonsingular and a dot denotes the derivative with respect to U. The metric will assume the form

$$ds^2 = -2\,dU\,dv + dX^T\,h\,dX, \qquad h = A^T A, \tag{7}$$

if A satisfies

$$A^T \dot{A} - \dot{A}^T A = 0, \tag{8}$$

$$\ddot{A} - 2FA = 0. \tag{9}$$

The first of the conditions follows from the symmetry of F, while its being trace-free yields

$$\ddot{A}^T \epsilon A + A^T \epsilon \ddot{A} = 0. \tag{10}$$

Let us now consider the transformation from the point of view of one who wishes to transform a *pp* wave expressed in the form (7) to the Kerr–Schild form (5). First of all, it should be remarked that a metric of form (7) automatically satisfies all the vacuum field equations save one, and that one turns out to be precisely Eq. (10). Equation (9) becomes the definition of the matrix F, which by (10) is trace-free. What, however, is the status of Eq. (8)?

It will be noted that for given $h = A^T A$, the matrix A is determined only up to $A \to RA$, where $R^T R = I$. In order to arrive at the Kerr–Schild form (5) one must impose Eq. (8), which can always be satisfied by choosing R appropriately.

Turning our attention now to the specific *pp* wave (3), we may express A in the form

$$A = \frac{\sqrt{\rho}}{\Sigma} \begin{pmatrix} u+p & q \\ -q & u+p \end{pmatrix} \begin{pmatrix} (pu+1)^2 + q^2 & -p^{-1}q \\ -pq(1-u^2) & 1 \end{pmatrix}, \tag{11}$$

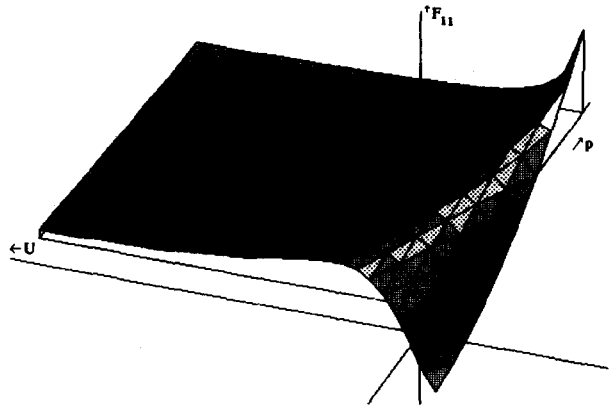


FIG. 1. The F_{11} pulse profile for various values of p for our solution.

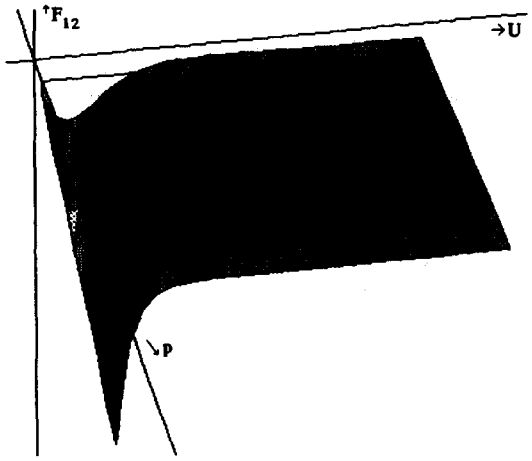


FIG. 2. The F_{12} pulse profile for various values of p for our solutions.

where $\rho = 1 - u^2$. Moreover, the transformation $U(u)$ is determined by

$$U = 2\int \frac{u^2 + 2pu + 1}{\sqrt{1-u^2}}\,du$$

$$= 3\vartheta - \tfrac{1}{2}\sin(2\vartheta) + 4p(1 - \cos(\vartheta)), \quad u = \sin(\vartheta). \tag{12}$$

Using Eqs. (11) and (12) a tedious, although straightforward, calculation yields

$$F_{11} = -(3p^2/8\Sigma^5)\{pu^5 + 5u^4 + 10pu^3 + 10u^2 + 5pu + 1 - 2q^2(2pu^5 + 5u^4 + 1)\},$$

$$F_{12} = -(3p^2 q/8\Sigma^5)\{(3 - 4q^2)u^5 + 10pu^4 + 10u^3 - 5u - 2p\}. \tag{13}$$

These expressions are valid for $0 < u < 1$; at $u = 0$ there is not only a step function but also a δ function discontinuity.

In the accompanying figures (see Figs. 1–4), we have attempted to display the nature of the functions F_{11} and F_{12} both for our solution ($p \geqslant 0$) and for the CX solution ($p \leqslant 0$). The variable u in each case runs from 0 to 1, while U runs

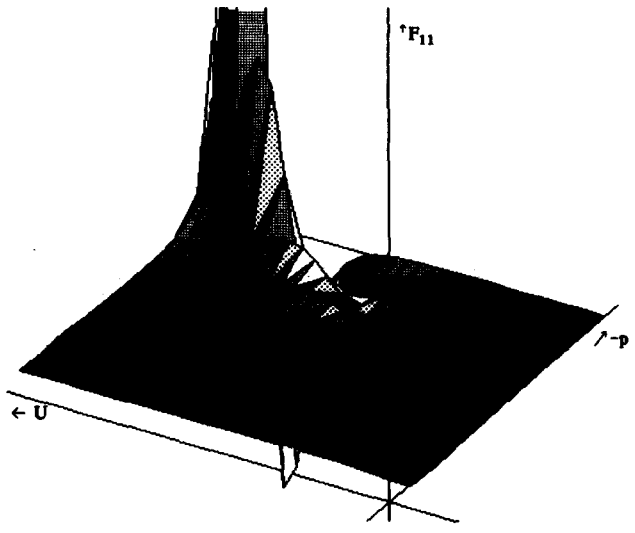


FIG. 3. The F_{11} pulse profile for various values of p for the CX solution.

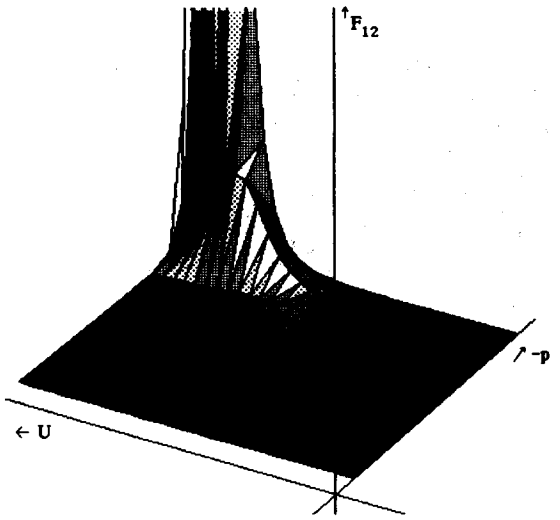


FIG. 4. The F_{12} pulse profile for various values of p for the CX solution.

from 0 to $3\pi/2 \pm 4p$, respectively. The figures are rotated in such a way as best to display the interesting features. Moreover, the u-independent coefficients, including the minus signs, have been suppressed. The horizontal axes of the figures are the p axis and U axis, or more precisely, $U/(3\pi/2 \pm 4p)$.

In the collinear case, i.e., $q = 0, p = \pm 1$, we get $F_{12} = 0$ and $F_{11} \sim (1 \pm u)^{-5}$, which clearly shows the divergence that occurs in the CX shock wave as $u \to 1$, and which is absent from our shock wave. Of course, both pulses have a δ function at $u = 0$, which we have not shown in the sketches.

III. THE INTERACTION REGION AND THE CRITICAL POINTS

The metric (1) describes the interaction of the two incident pp waves. When $\rho = 1 - u^2 - v^2 \to 0$, one reaches the boundary of our chart. However, a method of Carter[2] may be used to construct the unique analytic extension of this metric across the null surface $\rho = 0$. The final result is that the geometry on the other side of the horizon is described by a metric of the same form as (1) except with $x > 1$. It is just the Kerr metric outside the horizon at $\mathscr{I}^-$. At the corresponding horizon at $\mathscr{I}^+$, one may extend the metric analytically again, and, if one wishes, one may join part of the region within that horizon to a pair of outgoing pp waves.

Nowhere in the extended space-time we have been discussing does the curvature become infinite. Nevertheless, there do exist in this solution as well as in the CX solution unavoidable topological singularities at $u = 0$, $v = 1$ and $u = 1$, $v = 0$. This may be seen by approaching either of these points from the appropriate pp wave region and the interaction region. In the former case, one has Minkowski space in the form (4), where the ignorable coordinates x_1 and x_2 are adapted to null rotations, while in the latter case, one approaches Minkowski space in a form in which the ignorable coordinates x_1 and x_2 are adapted to a boost and a rotation. In both cases, the singular point is at the origin ($\alpha = 0$), where it is not possible to define the manifold of tangent vectors.

ACKNOWLEDGMENTS

The authors would like to thank Dr. B. Schmidt for his help in clarifying the topological nature of the singularity that colliding wave metrics invariably have. One of the autors (FJE) acknowledges the support of the National Science Foundation through Grants No. PHY-86-05958 and PHY-89-08231.

[1]S.Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **408**, 175 (1986).
[2]B. Carter, Phys. Rev. **174**, 1559 (1968).

A Kerr object embedded in a gravitational field. II

K. D. Krori and R. Bhattacharjee
Mathematical Physics Forum, Cotton College, Gauhati-781001, India

(Received 28 February 1989; accepted for publication 26 July 1989)

Starting with a Weyl solution representing an axially symmetric gravitational field as the seed metric, a generalized Kerr solution is generated by the Belinskii–Zakharov technique [Sov. Phys. JETP **48**, 985 (1978); **50**, 1 (1980)]. This solution may be interpreted as a Kerr metric embedded in a gravitational field.

I. INTRODUCTION

Using a procedure developed by Ernst,[1] Kerns and Wild[2] constructed an axially symmetric solution of the vacuum Einstein field equations for a Schwarzschild source embedded in a gravitational field. Such a gravitational field may be produced by surrounding matter.[3] Kerns and Wild pointed out at the end of their paper that astrophysically it would be more interesting to derive a solution embedding a Kerr source in a gravitational field. With this idea in view, we started an investigation and succeeded in deriving a solution for $a^2 > m^2$ (a is the angular momentum parameter, m is the mass).[4] In this paper, we present a solution of a more general type using the Belinskii–Zakharov technique.[5,6] If the gravitational field parameter vanishes, we obtain the Kerr solution. If, however, the Kerr (angular momentum) parameter vanishes, we obtain a generalized Schwarzschild solution.[2] It is hoped that astrophysically this generalized Kerr solution will be more interesting than our earlier solution.[4]

II. A GENERALIZED KERR SOLUTION

Let us consider the Weyl line element for an axially symmetric gravitational field[1]:

$$ds^2 = e^{2\gamma - 2\chi}(d\rho^2 + dz^2) + \rho^2 e^{-2\chi}\, d\Phi^2 - e^{2\chi}\, dt^2 \quad (1)$$

$$\equiv f(\rho,z)(d\rho^2 + dz^2) + g_{ab}\, dx^a\, dx^b, \quad (2)$$

where

$$\gamma = -\tfrac{1}{2}k^2\rho^2, \quad \chi = kz, \quad (3)$$

a and b take the values 0 and 1 corresponding to t and Φ, respectively, and k is the gravitational field parameter.

Taking (1) as the seed metric and applying the Belinskii–Zakharov technique[5,6] we obtain the following physical (ph) solution:

$$(g_{00})_{\rm ph} = -\frac{e^{2k\sigma xy}}{(px+r)^2 + q^2\Omega^2}\left[p^2x^2 + q^2 - r^2 - \frac{q^2}{4}(1-y^2)\{e^{2k\sigma(x-1)} + e^{-2k\sigma(x-1)}\}^2\right], \quad (4)$$

$$(g_{01})_{\rm ph} = -\frac{\sigma q}{2[(px+r)^2 + q^2\Omega^2]}\,[(p+r)(x+1)\{(1-y)(x+y)e^{-2k\sigma(x-1)} + (1+y)(x-y)e^{2k\sigma(x-1)}\}$$
$$-(r-p)(x-1)\{(1+y)(x+y)e^{2k\sigma(x-1)} + (1-y)(x-y)e^{-2k\alpha(x-1)}\}], \quad (5)$$

$$(g_{11})_{\rm ph} = \frac{a^2e^{-2k\sigma xy}}{2[(px+r)^2 + q^2\Omega^2]}\left[2(x-1)^2(1-y^2)(px+r)^2 + 2x(x+1)^2(1-y^2)(p+r)^2\right.$$
$$\left.-q^2(x^2-1)(1-y^2)(2x-1-y^2) - \frac{q^2}{2}(x^2-1)\{(1+y^2)^4e^{4k\sigma(x-1)} + (1-y^2)^4e^{-4k\sigma(x-1)}\}\right], \quad (6)$$

$$f_{\rm ph} = e^{-k^2\sigma^2(x^2-1)(1-y^2) - 2k\sigma xy}[(px+r)^2 + q^2\Omega^2], \quad (7)$$

where

$$p = e^{2k\sigma y}\cos^2\eta - e^{-2k\sigma y}\sin^2\eta, \quad r = e^{2k\sigma y}\cos^2\eta + e^{-2k\sigma y}\sin^2\eta, \quad q = 2\sin\eta\cos\eta, \quad (8)$$
$$\eta = \text{const}, \quad 2\Omega = (1+y)e^{2k\sigma(x-1)} - (1-y)e^{-2k\sigma(x-1)},$$

and x,y are related to ρ, z as

$$z = \sigma xy, \quad (9)$$

$$\rho^2 = \sigma^2(x^2-1)(1-y^2), \quad (10)$$

σ being a constant.

Using the transformation

$$q = t + (2q/p)\Phi, \quad (11)$$

one can write the new metric in the form

$$ds^2 = e^{-k^2\sigma^2(x^2-1)(1-y^2) - 2k\sigma xy}[(px+r)^2 + q^2\Omega^2]\left[\frac{dx^2}{x^2-1} + \frac{dy^2}{1-y^2}\right] + G_{11}\, d\Phi^2 + G_{01}\, d\Phi\, d\tau + G_{00}\, d\tau^2, \quad (12)$$

 0022-2488/90/010147-02$03.00

where

$$G_{00} = -\frac{e^{2k\sigma xy}}{(px+r)^2+q^2\Omega^2}\left[p^2x^2+q^2-r^2-\frac{q^2}{4}(1-y^2)\{e^{2k\sigma(x-1)}+e^{-2k\sigma(x-1)}\}^2\right], \tag{13}$$

$$G_{01} = -\frac{1}{(px+r)^2+q^2\Omega^2}\Bigg(\frac{\sigma q}{2}[(p+r)(x+1)\{(1-y)(x+y)e^{-2k\sigma(x-1)}+(1+y)(x-y)e^{2k\sigma(x-1)}\}$$
$$-(r-p)(x-1)\{(1+y)(x+y)e^{2k\sigma(x-1)}+(1-y)(x-y)e^{-2k\sigma(x-1)}\}]-\frac{2q}{p}e^{2k\sigma xy}\Big[p^2x^2+q^2-r^2$$
$$-\frac{q^2}{4}(1-y^2)(e^{2k\sigma(x-1)}+e^{-2k\sigma(x-1)})^2\Big]\Bigg), \tag{14}$$

$$G_{11} = \frac{1}{(px+r)^2+q^2\Omega^2}\Bigg(-\frac{4q^2}{p^2}e^{2k\sigma xy}\Big[p^2x^2+q^2-r^2-\frac{q^2}{4}(1-y^2)\{e^{2k\sigma(x-1)}+e^{-2k\sigma(x-1)}\}^2\Big]$$
$$+\frac{2\sigma q^2}{p}[(p+r)(x+1)\{(1-y)(x+y)e^{-2k\sigma(x-1)}+(1+y)(x-y)e^{2k\sigma(x-1)}\}$$
$$-(r-p)(x-1)\{(1+y)(x+y)e^{2k\sigma(x-1)}+(1-y)(x-y)e^{-2k\sigma(x-1)}\}]$$
$$+\frac{\sigma^2}{2}e^{-2k\sigma xy}\Big[2(x-1)^2(1-y^2)(px+r)^2+2x(x+1)^2(1-y^2)(p+r)^2$$
$$-q^2(x^2-1)(1-y^2)(2x-1-y^2)-\frac{q^2}{2}(x^2-1)\{(1+y)^4e^{4k\sigma(x-1)}+(1-y)^4e^{-4k\sigma(x-1)}\}\Big]\Bigg). \tag{15}$$

III. DISCUSSION

Putting $k=0$, one gets $p=\cos 2\eta$, $q=\sin 2\eta$, and, substituting $mp=\sigma$ and $q/p=a$, the new metric (12) reduces to

$$ds^2 = [(px+1)^2+q^2y^2]\left[\frac{dx^2}{x^2-1}+\frac{dy^2}{1-y^2}\right]+\frac{1-y^2}{p^2}\left[(px+1)^2+q^2+\frac{2q^2(px+1)(1-y^2)}{(px+1)^2+q^2y^2}\right]d\Phi^2$$
$$-\frac{2q}{p}\frac{(px+1)(1-y^2)}{(px+1)^2+q^2y^2}d\Phi\,d\tau-\frac{p^2x^2+q^2y^2-1}{(px+1)^2+q^2y^2}d\tau^2, \tag{16}$$

which is the Kerr metric.

On the other hand, if q vanishes, then the new metric (12) reduces to the following form:

$$ds^2 = (x+1)^2e^{-2k\sigma xy-k^2\sigma^2(x^2-1)(1-y^2)+4k\sigma y}\left[\frac{dx^2}{x^2-1}+\frac{dy^2}{1-y^2}\right]+(x+1)^2(1-y^2)e^{-2k\sigma xy}d\Phi^2-\frac{x-1}{x+1}e^{2k\sigma xy}d\tau^2. \tag{17}$$

This is a generalized Schwarzschild solution derived by Kerns and Wild.[2] It represents a Schwarzschild source embedded in a gravitational field. This leads to the interpretation of (12) as a solution for Kerr source embedded in a gravitational field. Kerns and Wild apparently wanted to have a solution of this type. Herein lies the importance of our work.

ACKNOWLEDGMENTS

The authors express their profound gratitude to the Government of Assam for all facilities provided at Cotton College, Gauhati-781001 (India) to carry out the investigations reported in this paper.

[1] F. J. Ernst, J. Math. Phys. **19**, 1986 (1978).
[2] R. M. Kerns and W. J. Wild, Gen. Relat. Gravit. **14**, 1 (1982).
[3] A. Tomimatsu, Phys. Lett. A **103**, 374 (1984).
[4] K. D. Krori and R. Bhattacharjee, J. Math. Phys. **27**, 1056 (1986).
[5] V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).
[6] V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **50**, 1 (1980).

A Marderlike solution in higher dimensions[a)]

K. D. Krori, P. Borgohain, and Kanika Das
Mathematical Physics Forum, Cotton College, Gauhati-781001, India

(Received 3 May 1988; accepted for publication 19 July 1989)

A vacuum solution in higher dimensions that corresponds to the exterior Marder solution in four-dimensional general relativity is presented.

I. INTRODUCTION

Higher-dimensional physics is of great importance today as a result of recent developments in superstring theory.[1-3] Yoshimura and Koikawa have recently presented some spherically symmetric solutions in higher dimensions.[4-6] Myers and Perry[7] and Krori *et al.*[8] have derived Schwarzschild-like exterior and interior solutions, respectively. Here a vacuum solution in higher dimensions has been worked out and can be considered as an extension of the Marder solution[9] to higher dimensions: It reduces to the Marder solution when the number of dimensions becomes four.

II. SOLUTION

We consider the line element in the form

$$ds^2 = e^{2\alpha}dt^2 - e^{2\beta}(dr^2 + dz^2) - r^2e^{2\gamma}d\Omega^2, \tag{1}$$

where α, β, γ are functions of r only and $d\Omega^2$ is the line element on a unit $(N-3)$ sphere.

Using exterior calculus, we find the following Ricci tensors:

$$R_{00} = e^{-2\beta}[\alpha'' + \alpha'^2 + (N-3)(\alpha'/r + \alpha'\gamma')], \tag{2}$$

$$R_{11} = e^{-2\beta}[-\alpha'' - \beta'' - \alpha'^2 + \alpha'\beta' + (N-3) \times(-2\gamma'/r + \beta'/r - \gamma'^2 + \beta'\gamma' - \gamma'')], \tag{3}$$

$$R_{22} = e^{-2\beta}[-\alpha'\beta' - \beta'' - (N-3)(\beta'/r + \beta'\gamma')], \tag{4}$$

$$R_{33} = R_{44} = \cdots = R_{(N-1)(N-1)} = e^{-2\beta}[-\alpha'/r - \alpha'\gamma' - 2\gamma'/r - \gamma'' - \gamma'^2 - (N-4)(1/r + \gamma')^2]. \tag{5}$$

We obtain the following field equations by making $R_{\mu\nu} = 0$:

$$\alpha'' + \alpha'^2 + (N-3)(\alpha'/r + \alpha'\gamma') = 0, \tag{6}$$

$$-\alpha'' - \beta'' - \alpha'^2 + \alpha'\beta' + (N-3)(-2\gamma'/r + \beta'/r - \gamma'^2 + \beta'\gamma' - \gamma'') = 0, \tag{7}$$

$$\alpha'\beta' + \beta'' + (N-3)(\beta'/r + \beta'\gamma') = 0, \tag{8}$$

$$\alpha'/r + \alpha'\gamma' + 2\gamma'/r + \gamma'' + \gamma'^2 + (N-4)(1/r + \gamma')^2 = 0. \tag{9}$$

From Eq. (6) we obtain

$$\gamma' = (-\alpha'' - \alpha'^2)/(N-3)\alpha' - 1/r. \tag{10}$$

Hence

$$\gamma'' = (-\alpha'\alpha''' - 2\alpha'^2\alpha'' + \alpha'' + \alpha'^2\alpha'')/(N-3)\alpha' + 1/r^2. \tag{11}$$

Using (10) and (11) we obtain, from Eq. (9),

$$\alpha'' - C\alpha'^2 = 0, \tag{12}$$

where C is an integration constant. Solving Eq. (12) we obtain

$$\alpha = A - B\log(r + C_1), \tag{13}$$

where A, B, and C_1 are integration constants. From Eq. (10) by using (13) we obtain

$$\gamma = [(1+B)/(N-3)]\log(r + C_1) - \log r + C_2, \tag{14}$$

where C_2 is a constant. From Eq. (8) by using (13) and (14) we obtain

$$\beta = -\tfrac{1}{2}[(1 - B^2) - (1+B)^2/(N-3)] \times \log(r + C_1) + C_3, \tag{15}$$

where C_3 is a constant.

It can be shown that solutions (13)–(15) correspond to the N-dimensional extension of a Marderlike solution. Exact N-dimensional Marderlike solutions can be obtained by adjusting the constants in the following manner:

$$A = 0, \quad B = -C, \quad C_1 = C_2 = 0, \quad C_3 = \log K, \tag{16}$$

where C and K are constants.

Thus we obtain

$$\alpha = C\log r, \tag{17}$$

$$\gamma = [(1-C)/(N-3)]\log r - \log r, \tag{18}$$

$$\beta = -\tfrac{1}{2}[1 - C^2 - (1-C)^2/(N-3)]\log r + \log K. \tag{19}$$

The line element (1) becomes

$$ds^2 = r^{2C}dt^2 - K^2r^{-[1 - C^2 - (1-C)^2/(N-3)]}(dr^2 + dz^2) + r^{2(1-C)/(N-3)}d\Omega^2. \tag{20}$$

When $N = 4$ (i.e., the number of dimensions is four) we obtain the Marder metric[9]

$$ds^2 = r^{2C}dt^2 - K^2r^{-2C(1-C)}(dr^2 + dz^2) + r^{2(1-C)}d\phi^2. \tag{21}$$

III. LIMITING CONDITIONS

We now consider properties of the general solution represented by (13)–(17) and of the Marderlike solution (20) in the limits $r \to 0$ and $r \to \infty$. The metric coefficients for the general solution are

$$e^{2\alpha} = e^{2A}(r + C_1)^{-2B}, \tag{22}$$

$$e^{2\beta} = e^{2C_3}(r + C_1)^{-P}, \tag{23}$$

where

$$P = 1 - B^2 - (1+B)^2/(N-3), \tag{24}$$

$$r^2e^{2\gamma} = e^{2C_2}[(r + C_1)]^{2(1+B)/(N-3)}. \tag{25}$$

[a)] This paper is dedicated in honor of Professor S. D. Chatterjee for his contributions to nuclear physics.

 0022-2488/90/010149-02$03.00

A. $r \to 0$

In the limit $r \to 0$, the metric coefficients of the general solution have the values

$$\begin{aligned} &e^{2\alpha} \to e^{2A} C_1^{-2B}, \\ &e^{2\beta} \to e^{2C_3} C_1^{-P}, \\ &r^2 e^{2\gamma} \to e^{2C_2} C_1^{2(1+B)/(N-3)}. \end{aligned} \tag{26}$$

Thus the general solution is free from any singularity in the limit $r \to 0$. However, the situation is different for the particular Marderlike solution with values of the integration constants given by (16). In this case,

$$\begin{aligned} &e^{2\alpha} \to 0, \\ &e^{2\beta} \to 0, \quad \text{if} \quad P<0, \quad \text{i.e.,} \quad C>1, \\ &e^{2\beta} \to \infty, \quad \text{if} \quad P>0, \quad \text{i.e.,} \quad C<1, \\ &e^{2\beta} = K^2, \quad \text{if} \quad P=0, \quad \text{i.e.,} \quad C=1, \\ &r^2 e^{2\gamma} \to \infty, \quad \text{if} \quad C>1, \\ &r^2 e^{2\gamma} \to 0, \quad \text{if} \quad C<1, \\ &r^2 e^{2\gamma} = 1, \quad \text{if} \quad C=1. \end{aligned} \tag{27}$$

B. $r \to \infty$

In the limit $r \to \infty$ the metric coefficients of the general solution have the values

$$\begin{aligned} &e^{2\alpha} \to \infty, \quad \text{if} \quad B<0, \\ &e^{2\alpha} \to 0, \quad \text{if} \quad B>0, \\ &e^{2\beta} \to \infty, \quad \text{if} \quad P<0, \quad \text{i.e.,} \quad B<-1 \text{ or } B>(N-4)/(N-2), \\ &e^{2\beta} \to 0, \quad \text{if} \quad P>0, \quad \text{i.e.,} \quad -1<B<(N-4)/(N-2), \\ &e^{2\beta} = e^{2C_3}, \quad \text{if} \quad P=0, \quad \text{i.e.,} \quad B=-1 \text{ or } B=(N-4)/(N-2), \\ &r^2 e^{2\gamma} \to \infty, \quad \text{if} \quad B>-1, \\ &r^2 e^{2\gamma} \to 0, \quad \text{if} \quad B<-1, \\ &r^2 e^{2\gamma} = e^{2C_2}, \quad \text{if} \quad B=-1. \end{aligned} \tag{28}$$

Thus the general solution has singularities in the limit $r \to \infty$. For the Marderlike solution we have, in this limit,

$$\begin{aligned} &e^{2\alpha} \to \infty, \\ &e^{2\beta} \to \infty, \quad \text{if} \quad C>1, \\ &e^{2\beta} \to 0, \quad \text{if} \quad C<1, \\ &e^{2\beta} = k^2, \quad \text{if} \quad C=1, \\ &r^2 e^{2\gamma} \to \infty, \quad \text{if} \quad C<1, \\ &r^2 e^{2\gamma} \to 0, \quad \text{if} \quad C>1, \\ &r^2 e^{2\gamma} = 1, \quad \text{if} \quad C=1. \end{aligned} \tag{29}$$

Singularities occur in this case as well.

ACKNOWLEDGMENT

The authors express their profound gratitude to the Government of Assam, Dispur, for all facilities provided at Cotton College, Gauhati-781001, India, in order to carry out the investigations reported in this paper.

[1] M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 177 (1984).
[2] E. Witten, Phys. Lett. B **149**, 351 (1984).
[3] D J. Gross, J. Harvey, E. Martinu, and R. Rohm, Phys. Rev. Lett. **52**, 502 (1985).
[4] T. Koikawa and M. Yoshimura, Prog. Theor. Phys. **75**, 977 (1986).
[5] M. Yoshimura, Phys. Rev. D **34**, 1021 (1986).
[6] T. Koikawa, Phys. Lett. A **117**, 279 (1986).
[7] R. C. Myers and M. J. Perry, Ann. Phys. (NY) **172**, 304 (1986).
[8] K. D. Krori, P. Borgohain, and K. Das, Phys. Lett. A **321**, 132 (1988).
[9] L. Marder, Proc. R. Soc. London Ser. A **244**, 524 (1958).

Conformally symmetric static fluid spheres

R. Maartens and M. S. Maharaj
Centre for Nonlinear Studies and Department of Computational and Applied Mathematics, University of Witwatersrand, P.O. Wits 2050, South Africa

(Received 17 May 1989; accepted for publication 23 August 1989)

Solutions of the Einstein–Maxwell equations for static spheres of charged imperfect fluids are investigated, where the space-time geometry is assumed to admit a conformal symmetry. Previous work is generalized by considering a *nonstatic* conformal symmetry. This allows the possibility of solutions that are nonsingular at the center, unlike the previous solutions based on a static conformal symmetry. Two such regular solutions are presented for charged spheres. The further generalization necessary to find stable exact stellar models with a conformal symmetry is indicated.

I. INTRODUCTION

Despite the power of computer solutions, exact solutions are still of importance in general relativity, because they allow a "global" understanding not tied to a specific choice of parameters and initial conditons. In relativistic astrophysics, exact static fluid solutions are important for developing stellar models. The most general mathematical case in this class of solutions is the charged, imperfect fluid sphere.

Exact solutions of the field equations for static spheres may be found in an *ad hoc* fashion by specifying one or more of the geometric and matter variables, and using the field equations to determine the remaining variables. However, such *ad hoc* solutions are extremely unlikely to produce physically acceptable models. What is needed is a systematic method of searching for exact solutions. One such method has been initiated by Herrera and co-workers (see Refs. 1–3 and references cited therein). In this approach, the fluid space-time is assumed to have, in addition to its static and spherical symmetry, a conformal symmetry. If the vector field ξ is the generator of this conformal symmetry, then the space-time metric $\mathbf{g}$ is mapped conformally onto itself along the trajectories of ξ:

$$\mathcal{L}_\xi \mathbf{g} = \psi \mathbf{g}, \tag{1}$$

where $\mathcal{L}$ is the Lie derivative operator,[4] and ψ is the conformal factor.

The condition (1) is essentially geometric. But it does have two physical points of support. First, it is a generalization (when ψ is not constant) of self-similarity in hydrodynamics. Second, it is a generalization of the property of the incompressible Schwarzschild interior solution, which has 11 independent conformal symmetries in addition to the four Killing symmetries, since it is conformally flat.[4] The Schwarzschild interior solution is perhaps the most realistic known exact static stellar solution.[4] Condition (1) therefore offers the possibility of finding exact solutions that generalize the Schwarzschild interior solution, and are more physically interesting models.

However, as we shall show, the possibility of finding acceptable solutions via condition (1) depends crucially on the form of the conformal Killing vector field ξ. In fact the form of ξ assumed by Herrera and co-workers[1,2] is too restrictive to allow this possibility. A wide range of solutions is presented in Refs. 1 and 2, but all of these solutions are singular at the stellar center. Although one can argue, as in Refs. 1 and 2, for a regular core matched to the solution that satisfies condition (1), this seems a little artificial and perhaps against the intention underlying assumption (1). In any case, our aim in this paper is to investigate the possibility of generalizing the Schwarzschild interior solution in a nonsingular way, using condition (1).

In Sec. II we give the field equations for the general case of charged imperfect fluid spheres. We show in Sec. III why the solutions of Herrera and co-workers[1,2] are necessarily singular at the center. Essentially, this occurs because Herrera and co-workers assume that ξ is not only spherically symmetric, but also static. It follows that the first step in aiming for regular solutions is to relax the assumption that ξ be static. [Note that in the limiting case of flat space-time (regular vacuum solution), none of the nonisometric conformal Killing vectors is static.] Two cases then arise: either ψ is static or nonstatic. The first case is pursued in this paper. The second case, the most general possibility for spherically symmetric ξ, is under investigation.[5]

In Sec. III we show that the case of nonstatic ξ with static ψ admits the possibility of solutions regular at the stellar center. We find a necessary condition on ψ for regularity. Then we show that the new class of solutions contains flat space-time, but not the Schwarzschild interior solution (for which ψ is nonstatic[5]).

In Sec. IV we consider some new solutions for the generalized nonstatic ξ. We show that the upper limit on the mass–radius ratio may exceed that established in Refs. 1 and 2 on the basis of a static ξ. We briefly describe the direct generalization of the solutions of Herrera and Ponce de Leon[2] (which contain the solutions of Ref. 1 as the uncharged special case). In particular, we find a regular perfect fluid solution with uniform charge. We also find a regular charged imperfect fluid solution by giving ψ its simplest polynomial form satisfying the regularity condition. Unfortunately, both of these solutions have unstable features.

In fact, we show in Sec. IV that instability arising from negative pressures is a problem with all the generalized regular solutions. Furthermore, we show that there are no regular uncharged perfect fluid spheres. Either electric repulsion or pressure anisotropy is necessary to maintain a sphere (even if unstable) within this class of solutions. We also

 0022-2488/90/010151-05$03.00

show that there are no regular charged perfect fluid solutions that are self-similar or that satisfy the equation of state $p=(\gamma-1)\mu$, $1\leqslant\gamma\leqslant 2$.

These results show that while we are able to remove the central singularity by generalizing the conformal symmetry of Herrera and co-workers,[1,2] our models have other drawbacks. In particular, we cannot find regular uncharged perfect fluid solutions. This motivates a further generalization of the conformal symmetry.[5]

II. FIELD EQUATIONS

(We use the notation and conventions of Ref. 4, with Einstein's gravitational constant equal to 1.) Consider a charged nonconducting imperfect fluid without heat flow. (For a discussion of the physical relevance of such fluids in astrophyics, see Refs. 2 and 3 and references cited therein.) The total energy–momentum tensor is

$$T_{ab}=M_{ab}+E_{ab}, \tag{2}$$

where the matter contribution is

$$M_{ab}=\mu u_a u_b+ph_{ab}+\pi_{ab} \tag{3}$$

(with $h_{ab}=g_{ab}+u_a u_b$) and the electromagnetic contribution is

$$E_{ab}=F_{ac}F_b{}^c-\tfrac{1}{4}g_{ab}F_{cd}F^{cd}. \tag{4}$$

The Einstein–Maxwell field equations are

$$R_{ab}-\tfrac{1}{2}Rg_{ab}=T_{ab}, \tag{5}$$

$$F_{[ab;c]}=0, \quad F^{ab}{}_{;b}=\epsilon u^a. \tag{6}$$

For a static spherically symmetric fluid, we can choose coordinates $x^i=(x^0,x^\alpha)=(t,r,\theta,\phi)$ such that the metric is

$$g_{ij}=\mathrm{diag}(-e^{\nu(r)},e^{\lambda(r)},r^2,r^2\sin^2\theta). \tag{7}$$

By symmetry, the fluid four-velocity $\mathbf{u}$, energy density μ, isotropic pressure p, and stress tensor π take the form,[3,6]

$$u^i=e^{-\nu/2}\delta^i_0, \quad \mu=\mu(r), \quad p=\tfrac{1}{3}[p_R(r)+2p_T(r)],$$

$$\pi_{ij}=(p_R-p_T)(n_in_j-\tfrac{1}{3}h_{ij}), \quad n^i=e^{-\lambda/2}\delta^i_1, \tag{8}$$

where $\mathbf{n}$ is a unit radial vector, p_R is the radial pressure, and p_T is the tangential pressure. Symmetry also implies that we can give the Maxwell field the form[6]

$$\mathbf{F}=f(r)dt\wedge dr, \tag{9}$$

so that $d\mathbf{F}=0$ and the first set of Maxwell's equations (6) is satisfied. The second set determines the charge density $\epsilon(r)$ in terms of f and g_{ij}. The form (9) shows that the magnetic field vanishes in the fluid rest frame, and the electric field $E_a=F_{ab}u^b$ has the form

$$E_i=E(r)n_i. \tag{10}$$

Putting together (2)–(10), we get the field equations[2,6]

$$\mu+\tfrac{1}{2}E^2=r^{-2}+r^{-2}e^{-\lambda}(r\lambda'-1), \tag{11}$$

$$p_R-\tfrac{1}{2}E^2=-r^{-2}+r^{-2}e^{-\lambda}(r\nu'+1), \tag{12}$$

$$p_T+\tfrac{1}{2}E^2=\tfrac{1}{4}e^{-\lambda}[2\nu''+r^{-1}(\nu'-\lambda')(r\nu'+2)]. \tag{13}$$

By (4)–(10),

$$E^{ab}{}_{;b}=-F^{ab}F_b{}^c{}_{;c}=-\epsilon En^a,$$

$$\epsilon=-u_aF^{ab}{}_{;b}=e^{-\lambda/2}r^{-2}(r^2E)'. \tag{14}$$

Thus the contracted Bianchi identity reduces to $\epsilon E^a=M^{ab}{}_{;b}$, which gives, by (3), (7), (8), and (14),

$$2r^{-2}E(r^2E)'=(\mu+p_R)\nu'+2p_R'+4r^{-1}(p_R-p_T). \tag{15}$$

The field equations (11)–(13) imply (15); alternatively, Eq. (15) may replace one of the field equations.

Finally, we consider the boundary conditions at the radius $r=R$ of the fluid sphere. The metric and energy–momentum tensor must match the Reissner–Nordström exterior space-time. This gives[2]

$$e^{\nu(R)}=e^{-\lambda(R)}=1-2M/R+Q^2/2R^2,$$

$$p_R(R)=0, \quad E(R)=Q/R^2, \tag{16}$$

where M and Q represent the total mass and charge of the sphere.

III. THE CONFORMAL SYMMETRY

The field equations (11)–(13) are underdetermined. One or more (depending on whether $E\neq 0$ and $p_R-p_T\neq 0$) functional relations must be specified in order to solve the equations. One method, which avoids *ad hoc* specification, is to assume that the fluid space-time is mapped conformally onto itself along the direction ξ, so that, by (1),

$$g_{ij,k}\xi^k+g_{kj}\xi^k{}_{,i}+g_{ik}\xi^k{}_{,j}=\psi g_{ij}. \tag{17}$$

Herrera and co-workers[1,2] assume that

$$\xi=\alpha(r)\partial_t+\beta(r)\partial_r. \tag{18}$$

(In fact, it is erroneously asserted in Ref. 1 that this form follows from the static spherical symmetry of $\mathbf{g}$: Minkowski space-time provides an immediate counterexample.) Using (7) and (18) in (17), we get[1,2]

$$\alpha=A, \quad \beta=\tfrac{1}{2}Bre^{-\lambda/2}, \quad \psi=Be^{-\lambda/2}, \quad e^\nu=C^2r^2, \tag{19}$$

where A, B, and C are constants: A may be set to zero since $A\,\partial_t$ is a Killing vector; B may be set to 1 by a rescaling $\xi\to B^{-1}\xi$, $\psi\to B^{-1}\psi$, which leaves (17) invariant. Thus the assumptions (17) and (18) for the geometry (7) determine the metric component e^ν explicitly, and fix α, β, and ψ.

The form (18) is the most general ξ invariant under the Killing symmetries of $\mathbf{g}$, i.e.,

$$[\partial_t,\xi]=0=[\mathbf{X}_\alpha,\xi],$$

where $\{\mathbf{X}_\alpha\}$ generates SO(3). Thus a nonisometric conformal Killing vector that is static and spherically symmetric is necessarily orthogonal to ∂_t. There are no such vectors in Minkowski space-time: thus assumption (18) rules out the limiting case of a regular vacuum solution. This suggests that (18) may lead to solutions that are singular. The world line $\{r=0\}$ is a timelike geodesic, by spherical symmetry. The regularity of space-time along a geodesic imposes stringent conditions on the limiting behavior of g_{ij}, obtained by expanding about the central geodesic. If t measures proper time along $\{r=0\}$, this gives[7]

$$e^\nu=1+O(l^2), \quad r=l+O(l^3),$$

near $l=0$, where l is proper radial distance orthogonal to $\{r=0\}$ ($dl=e^{\lambda/2}\,dr$). By a rescaling: $t\to e^{-a/2}t$, $\nu\to\nu+a$, which leaves the metric invariant, we get the more general

necessary condition for a nonsingular stellar center:

$$e^{\nu} = e^{a} + O(r^2), \quad e^{\lambda} = 1 + O(r^2) \quad \text{near } r = 0. \quad (20)$$

Then (19) shows that all the solutions obtained via (18) are necessarily singular.

Thus a search for regular solutions leads us to weaken the static symmetry of (18), i.e., we take

$$\xi = \alpha(t,r)\partial_t + \beta(t,r)\partial_r \quad (21)$$

(which generalizes the isotropic conformal vector $t\,\partial_t + r\,\partial_r$ of Minkowski space-time). Furthermore we assume that the conformal factor is static:

$$\psi = \psi(r). \quad (22)$$

[Note that (22) follows from (18), but not necessarily from (21).] The general case $\psi = \psi(t,r)$ is under investigation.[5] Note also that (22) in (17) implies that the deformation of $\mathbf{g}$ under ξ is static and spherically symmetric, like $\mathbf{g}$ itself.

Using (21) and (22), (17) gives β, ψ as in (18), and

$$\alpha = A + \frac{1}{2}kt, \quad e^{\nu} = C^2 r^2 \exp\left(-2kB^{-1}\int r^{-1}e^{\lambda/2}\,dr\right),$$

where k is constant. As before, we can set $A = 0$ and $B = 1$ without loss of generality. Thus we have

$$\xi = \tfrac{1}{2}kt\,\partial_t + \tfrac{1}{2}\psi(r)r\,\partial_r, \qquad e^{\lambda} = \psi^{-2}, \quad e^{\nu} = C^2 r^2 \exp\left(-2k\int \frac{dr}{r\psi}\right). \quad (23)$$

The solutions of Refs. 1 and 2 belong to the class $k = 0$. The vacuum solution is $k = 1 = \psi$. The self-similar Tolman solutions[8]

$$ds^2 = -r^{4(\gamma-1)/\gamma}\,dt^2 + b^2\,dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2)$$

are given by (23) with

$$k = (2-\gamma)/b\gamma, \quad \psi = b^{-1}, \quad C = 1.$$

However, the Schwarzschild interior solution[4]

$$e^{-\lambda} = 1 - Ar^2, \quad 2e^{\nu/2} = 3(1 - AR^2)^{1/2} - (1 - Ar^2)^{1/2} \quad (24)$$

is not contained by (23): if we take $\psi = \pm(1 - Ar^2)^{1/2}$, then

$$e^{-\lambda} = 1 - Ar^2, \quad e^{\nu/2} = Cr[2r^{-1}(1 + (1 - Ar^2)^{1/2})]^{\pm k}, \quad (25)$$

which cannot regain (24) for any C,k. It can be shown[5] that (24) requires $\psi = \psi(t,r)$.

With (23), the field equations (11)–(13) become

$$\mu = \tfrac{1}{2}r^{-2}(1 - k^2) - 3r^{-1}\psi\psi' + \Delta, \quad (26)$$

$$p_R = \tfrac{1}{2}r^{-2}(k^2 - 1 + 4\psi^2 - 4k\psi) + r^{-1}\psi\psi' - \Delta, \quad (27)$$

$$E^2 = r^{-2}(k^2 + 1 - 2\psi^2) + 2r^{-1}\psi\psi' - 2\Delta, \quad (28)$$

where, following Herrera and Ponce de Leon,[2] we define

$$\Delta = \tfrac{1}{2}(p_T - p_R) \quad (29)$$

as the measure of pressure anisotropy. Exact solution of the Einstein–Maxwell equations in the general case ($E\Delta \neq 0$) thus requires a choice of $\psi(r)$ and of an equation of state $f(\mu, p_R, \Delta) = 0$ (or equivalent choice).

Comparing (20) and (23), we get a necessary condition for regularity at the stellar center:

$$\psi = 1 + O(r^2) \quad \text{near } r = 0, \quad k = 1. \quad (30)$$

If (30) is satisfied, then the field equations (26)–(28) show that the dynamical variables μ, p_R, Δ, and E are all bounded at $r = 0$, provided any one of them is bounded (which happens automatically in the special case $E\Delta = 0$ of an uncharged or perfect fluid).

Finally, we note that each dynamical tensor is mapped conformally onto itself by (21): by (8)–(10) and (23), we have

$$[\xi,\mathbf{u}] = -\tfrac{1}{2}\psi\mathbf{u}, \quad [\xi,\mathbf{n}] = -\tfrac{1}{2}\psi\mathbf{n},$$
$$\mathcal{L}_\xi \boldsymbol{\pi} = (1 + r\Delta'/2\Delta)\psi\boldsymbol{\pi},$$
$$\mathcal{L}_\xi \mathbf{F} = \tfrac{1}{2}(r\psi f'/f + r\psi' + \psi + k)\mathbf{F},$$
$$[\xi,\mathbf{E}] = \tfrac{1}{2}\psi(rE'/E - 1)\mathbf{E}.$$

(Note that none of the above holds in general for conformal motions.[3]) The conformal Killing vector ξ forms a five dimensional Lie algebra with the Killing vectors $\{\mathbf{X}_0 = \partial_t, \mathbf{X}_\alpha\}$:

$$[\mathbf{X}_0,\xi] = \tfrac{1}{2}k\,\mathbf{X}_0, \quad [\mathbf{X}_\alpha,\xi] = 0,$$
$$[\mathbf{X}_0,\mathbf{X}_\alpha] = 0, \quad [\mathbf{X}_\alpha,\mathbf{X}_\beta] = \epsilon^{\gamma}{}_{\alpha\beta}\mathbf{X}_\gamma.$$

IV. NEW SOLUTIONS

We begin by deriving the mass–charge–radius relations for solutions satisfying (23). By (16), (23), (27), and (28) we have

$$Q^2/2R^2 = 1 - 3\psi(R)^2 + 2k\psi(R), \qquad M/R = 1 - 2\psi(R)^2 + k\psi(R). \quad (31)$$

Then, since $Q^2 \geqslant 0$, $M \geqslant 0$, (31) implies

$$\max\{a_-,b_-\} \leqslant \psi(R) \leqslant \min\{a_+,b_+\}, \quad (32)$$

where $3a_\pm = k \pm (k^2+3)^{1/2}$, $4b_\pm = k \pm (k^2+8)^{1/2}$. Eliminating $\psi(R)^2$ in (31),

$$3M = R + Q^2/R - kR\psi(R). \quad (33)$$

For $k = 0$, we regain the results of Herrera and Ponce de Leon.[2] In particular, we see that for $k \neq 0$ it is no longer necessarily true that charge increases the mass. Furthermore, for uncharged spheres with conformal symmetry, $M/R = \frac{1}{3}$ is not an upper limit, as stated in Refs. 1 and 9. [This assertion is based on the assumption that (18) is the most general form of conformal symmetry.] In fact, by (31) and (33), for uncharged spheres the upper limit is

$$M/R = \tfrac{1}{3} - \tfrac{1}{9}k^2 + \tfrac{1}{9}|k|(k^2+3)^{1/2}.$$

Thus the limit on M/R is independent of the pressure anisotropy, and depends only on the conformal symmetry parameter k. In particular, M/R can exceed $\frac{1}{3}$, and approach $\frac{1}{2}$ arbitrarily closely for large enough $|k|$. For regular uncharged spheres ($k = 1$), the maximum is $M/R = \frac{4}{9}$, equal to the perfect fluid limit.[9]

The condition for the existence of a horizon, $Q^2/2 \leqslant M^2$, is, by (31),

$$\psi(R) \leqslant \tfrac{1}{2}(k-1) \quad \text{or} \quad \psi(R) \geqslant \tfrac{1}{2}(k+1), \quad (34)$$

and then the horizon radius is

$$R_* = M + (M^2 - \tfrac{1}{2}Q^2)^{1/2}. \quad (35)$$

By (31), it follows that $R_* \leqslant R$, for all k and all $\psi(R)$ satisfying (32) and (34) (which ensure the existence of R_*).

We now briefly describe the generalization of some of the singular solutions of Ref. 2. The generalized self-similar perfect fluid solutions are all singular, except for the vacuum limit. They are given by $\Delta = 0 = \psi'$ in (26)–(29) and in (23):

$$\mu = \tfrac{1}{2}(1-k^2)r^{-2}, \quad p = \tfrac{1}{2}(k^2 - 1 + 4\psi^2 - 4k\psi)r^{-2},$$
$$E^2 = (k^2 + 1 - 2\psi^2)r^{-2}, \quad e^{\lambda} = \psi^{-2}, \quad e^{\nu} = C^2 r^{2(1-k/\psi)}.$$

This is a charged generalization of the Tolman solution.[8]

The generalization of the charged dust solution ($\Delta = 0 = p$) is also singular for all k:

$$\mu = \tfrac{1}{2}(1 - k^2 - 6r\psi\psi')r^{-2}, \quad E^2 = (1 - 3\psi^2 + 2k\psi)r^{-2},$$

where ψ satisfies

$$2r\psi\psi' + 4\psi^2 - 4k\psi + k^2 - 1 = 0,$$

which is an Abel-type equation unless $k = 0$, when[2] $\psi^2 = \frac{1}{4} + Ar^{-4}$, or $k = 1$, when

$$\psi = 1 + Ar^{-2} = e^{-\lambda/2}, \quad e^{\nu} = C^2 r^2 (A + r^2)^{-1},$$

by (23).

The generalized solution for perfect fluids with uniform charge density is obtained by solving (28) for ψ with $\Delta = 0$, $E = Br$, and then substituting into (26) and (27):

$$\psi^2 = \tfrac{1}{2}(1 + k^2) - Ar^2 + \tfrac{1}{2}B^2 r^4 = e^{-\lambda},$$
$$\mu = \tfrac{1}{2}(1 - k^2)r^{-2} + 3A - 3B^2 r^2,$$
$$p = \tfrac{1}{2}(1 + 3k^2)r^{-2} - 3A + 2B^2 r^2 - 2k\psi r^{-2},$$

with e^{ν} given by (23). When $k = 1$, this shows that the regularity condition (30) is satisfied if $\psi \geqslant 0$, and further that μ, p, and E are bounded at $r = 0$:

$$\psi = (1 - Ar^2 + \tfrac{1}{2}B^2 r^4)^{1/2} = e^{-\lambda/2},$$
$$e^{\nu} = C^2 r^2 \exp\left(-2\int \frac{dr}{r\psi}\right),$$
$$\mu = 3A - 3B^2 r^2, \quad E = Br,$$
$$p = 2[1 - (1 - Ar^2 + \tfrac{1}{2}B^2 r^4)^{1/2}]r^{-2} - 3A + 2B^2 r^2. \tag{36}$$

Thus $p(0) = -2A$, and the radius of the sphere is given by

$$4B^4 R^6 - 12AB^2 R^4 + (9A^2 + 6B^2)R^2 - 8A = 0,$$

where $A > 0$ by (36), since $\mu > 0$. Thus the cubic in R^2 always has a positive root. It is possible to choose A and B such that μ remains positive throughout the sphere. The limiting case, when μ and p vanish simultaneously at $r = R$, is given by $B = A/\sqrt{2}$ and $R = (2/A)^{1/2}$. Then $\psi(R) = 0$, so that by (31) and (35), the stellar surface is at the horizon ($R = R_*$), with $M = R$ and $Q = \sqrt{2}R$. This limiting case is clearly unstable. Unfortunately, all solutions (36) have unstable features: both p and $dp/d\mu$ are negative, at least near the center. Electric repulsion is holding the matter apart, but this is not stable. We do not claim that the model is realistic, but it is a regular generalization of the singular solution in Ref. 2.

In fact, the solution (36) indicates a general property of the class of solutions that obey the regularity condition (30): all regular fluid spheres have nonpositive pressure at the center. To see this, we begin with the necessary conditions for regularity: μ, p_R, Δ, E bounded at $r = 0$; and, by (30), $\psi = 1 - Ar^2 + O(r^3)$ near $r = 0$ and $k = 1$. Then, by (26)–(28),

$$\mu(0) = 6A + \Delta(0), \quad p_R(0) = -4A - \Delta(0),$$
$$E(0)^2 = -2\Delta(0). \tag{37}$$

Since $\mu, E^2 \geqslant 0$ we have $A \geqslant 0$ and $\Delta(0) \leqslant 0$. Furthermore, the isotropic pressure p, given by (8), satisfies

$$p(0) = \tfrac{1}{3}(\Delta(0) - 12A)$$

by (29) and (37). Thus, except for the special case $A = 0 = \Delta(0)$, the pressure p is negative near the center. In the special case, $p(0) = 0$. Neither of these cases is ruled out: negative pressures could occur in nonequilibrium metastable states,[3] and zero central pressure is not impossible. However, we do not claim that $p(0) \leqslant 0$ is realistic for a static stellar model.

Apart from the central-pressure feature of the regular solutions with the generalized conformal symmetry (21) and (22), there are further general results that follow from (26)–(30).

(a) There are no regular uncharged perfect fluid spheres. Putting $\Delta = 0 = E$ into (26)–(28), we find that $\psi = +(1 - Ar^2)^{1/2}$, giving the solution (25) with $+k = +1$. Furthermore, we get $\mu = 3A > 0$ and

$$p = -A - 2r^{-2}[(1 - Ar^2)^{1/2} - (1 - Ar^2)],$$

which is negative for all r. Thus there is no zero-pressure surface.

(b) There are no regular incompressible perfect fluid spheres. This follows since $\Delta = 0 = \mu'$ implies $E = 0$ by (26) and (28). Then result (a) applies.

(c) There are no regular perfect fluid spheres with equation of state $p = (\gamma - 1)\mu$, $1 \leqslant \gamma \leqslant 2$. If we put $\Delta = 0$, $p = (\gamma - 1)\mu$ into (26)–(28) with $k = 1$, we get

$$\psi = 1 - Ar^{2/(2-3\gamma)},$$

which is singular at $r = 0$.

Finally, we present a solution for a charged and imperfect fluid sphere. We start by taking the simplest regular polynomial form for the conformal factor: $\psi = 1 - Ar^2$. Then we choose a linear pressure anisotropy $\Delta(r)$. By (37), $-6A \leqslant \Delta(0) \leqslant -4A$ ensures that $\mu(0), p_R(0) \geqslant 0$. By (27), $\Delta(R) = 4A(AR^2 - 1)$ ensures $p_R(R) = 0$. Then (26) gives $AR^2 \leqslant 1$ as the condition for $\mu(R) \geqslant 0$. We choose

$$A = \tfrac{1}{10}R^{-2}, \quad 50R^2\Delta = 7x - 25,$$

where $x = r/R$. Then (26)–(29) give

$$50R^2\mu = 5 + 7x - 3x^2,$$
$$50R^2 p_R = 5 - 7x + 2x^2,$$
$$50R^2 p_T = -45 + 7x + 2x^2, \tag{38}$$
$$25R^2 E^2 = 25 - 35x + 5x^2.$$

By (31), the charge and mass are given by

$$Q = \tfrac{3}{10}\sqrt{26}R, \quad M = \tfrac{27}{25}R.$$

Thus the sphere is charge dominated, and has no horizon, by (34). The radial pressure is positive in the interior, decreasing monotonically to zero. However, the tangential pressure

is negative throughout the interior. By (23), the metric components inside the sphere are [using (16)]

$$e^{\lambda} = (1 - \tfrac{1}{10}x^2)^{-2}, \quad e^{\nu} = \tfrac{9}{10}(1 - \tfrac{1}{10}x^2),$$

and the conformal Killing vector is

$$\xi = \tfrac{1}{2}[t\,\partial_t + (1 - \tfrac{1}{10}x^2)x\,\partial_x].$$

The generalization of the solutions of Herrera and Ponce de Leon[2] and the new regular solutions (36) and (38) show the importance of the choice of conformal symmetry vector ξ. The solutions of Ref. 2 are all forced to blow up at the center. While we are able to overcome this singularity via the generalization (21) of ξ, our solutions still suffer some serious drawbacks, particularly the problem of negative pressures. An indication of these drawbacks is the fact that the Schwarzschild interior solution is not contained in our class of solutions. The static nature of the conformal factor in (22) is the root of these limitations. When ψ is allowed to be nonstatic, a new range of possibilities is opened up.[5] [Note that Ponce de Leon[10] obtains regular stable static solutions with anisotropic pressure via a different approach: he assumes that the sphere is conformally flat. Thus there are 11 independent vector fields satisfying (1), some of which may obey (21) with $\psi = \psi(t,r)$.]

[1] L. Herrera, J. Jimenez, L. Leal, J. Ponce de Leon, M. Esculpi, and V. Galina, J. Math. Phys. **25**, 3274 (1984).
[2] L. Herrera and J. Ponce de Leon, J. Math. Phys. **26**, 2302 (1985).
[3] R. Maartens, D. P. Mason, and M. Tsamparlis, J. Math. Phys. **27**, 2987 (1986).
[4] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[5] R. Maartens, in preparation.
[6] S. D. Maharaj and R. Maartens, J. Math. Phys. **27**, 2514 (1986); Gen. Relativ. Gravit. **21**, 899 (1989).
[7] G. F. R. Ellis, R. Maartens, and S. D. Nel, Mon. Not. R. Astron. Soc. **184**, 439 (1978).
[8] J. Wainwright, in *Galaxies, Axisymmetric Systems and Relativity*, edited by M. A. H. MacCallum (Cambridge U.P., Cambridge, 1985).
[9] J. Ponce de Leon, Phys. Rev. D **37**, 309 (1988).
[10] J. Ponce de Leon, J. Math. Phys. **28**, 1114 (1987).

Stochastic quantization of para-Fermi fields

Janaki Balakrishnan, S. N. Biswas, Ashok Goyal, and S. K. Soni
Department of Physics, University of Delhi, Delhi-110007, India

(Received 3 December 1987; accepted for publication 31 May 1989)

The application of the method of stochastic quantization originally attributed to Parisi and Wu has been extended to spinor fields obeying para-Fermi statistics. The connection between Euclidean and stochastic field theories is established in the conventional manner by proving the equivalence between a Langevin equation satisfied by para-Grassmann fields and a Fokker–Planck equation, the Hamiltonian of which has been constructed using para-Grassmann variables analogous to its construction from Grassmann variables in the Fermi case. As an example, a two-point Green function is calculated for any arbitrary value of order p of para-Fermi statistics, barring the pathological case $p = 2$ which has been mentioned briefly.

I. INTRODUCTION

The stochastic quantization of Parisi and Wu[1] provides a viable alternative to path-integral quantization. It is particularly useful in quantizing gauge field theories, as no gauge fixing and associated Faddeev–Popov ghosts are required. In essence, the vacuum expectation value of the time-ordered product of field operators $\langle 0|T(\phi(x_1)\phi(x_2)\cdots\phi(x_n))|0\rangle$ as given by a Euclidean path integral is identical with the steady state equal time ($= t$) stochastic average $\langle \phi_\eta(x_1,t)\phi_\eta(x_2,t)\cdots\phi_\eta(x_n,t)\rangle_\eta$ of the product of stochastic fields. In the approach, t denotes the fictitious time coordinate which is introduced in addition to the usual space-time variable x and for the dynamical evolution of the stochastic field $\phi_\eta(x,t)$ with respect to the new time variable t, one postulates the following Langevin equation LE:

$$\frac{\partial}{\partial t}\phi_\eta(x,t) = -\frac{\delta S[\phi]}{\delta\phi_\eta(x,t)} + \eta(x,t), \tag{1.1}$$

where η is a Gaussian white noise with standard properties,

$$\langle \eta(x,t)\rangle_\eta = 0,$$
$$\langle \eta(x,t)\eta(x',t')\rangle_\eta = 2\delta(x-x')\delta(t-t'), \tag{1.2}$$

and η average is performed with respect to a Gaussian distribution

$$\langle F[\eta]\rangle_\eta = \frac{\int D\eta\, F[\eta]\exp(-\frac{1}{4}\int dx\, dt\, \eta^2)}{\int D\eta \exp(-\frac{1}{4}\int dx\, dt\, \eta^2)} \tag{1.3}$$

for any arbitrary function $F[\eta]$ of η.

This equivalence between D-dimensional Euclidean field theories and the steady-state of the $(D+1)$-dimensional stochastic process is known as the stochastic quantization of Parisi and Wu. The equivalence between Euclidean field theories and stochastic processes can be seen through the well-known relation between a Langevin equation and a Fokker–Planck equation (FPE). Extension of this stochastic quantization method to spinor fields has been considered by Fukai *et al.*[2] and Sakita.[3] Here the path integral expression for an n-point function is given by

$$\frac{\int D\bar\psi\, D\psi(\psi(x_1)\cdots\bar\psi(x_n))\exp(-S[\psi,\bar\psi])}{\int D\bar\psi\, D\psi \exp(-S[\psi,\bar\psi])},$$

where ψ and $\bar\psi$ are independent Grassmann variables and the action is taken to be a bilinear expression $S = \int d^4x\, \bar\psi K\psi$; K may contain not only derivative operators, but also external fields, and K need not be Hermitian in general. An appropriate Fokker–Planck Hamiltonian can be chosen as

$$H_{\rm FP} = \int dx \left[\frac{\delta}{\delta\psi}\left(\frac{\delta}{\delta\bar\psi} + \frac{\delta S}{\delta\bar\psi}\right) - \frac{\delta}{\delta\bar\psi}\left(\frac{\delta}{\delta\psi} + \frac{\delta S}{\delta\psi}\right)\right] \tag{1.4}$$

and the corresponding Langevin equations are

$$\frac{\partial}{\partial t}\psi(x,t) = -K\psi(x,t) + \eta(x,t) = -\frac{\delta S}{\delta\bar\psi} + \eta(x,t) \tag{1.5}$$

and

$$\frac{\partial\bar\psi}{\partial t}(x,t) = -K^T\bar\psi(x,t) + \bar\eta(x,t) = +\frac{\delta S}{\delta\psi} + \bar\eta(x,t). \tag{1.6}$$

Note that η and $\bar\eta$ are Gaussian noise functions obeying

$$\langle \eta_\alpha(x,t)\bar\eta_\beta(x',t')\rangle_\eta = -\langle \bar\eta_\beta(x',t')\eta_\alpha(x,t)\rangle_\eta$$
$$= 2\delta_{\alpha\beta}\delta^4(x-x')\delta(t-t').$$

The η average here means

$$\langle F(\eta)\rangle_\eta = \frac{\int D\bar\eta\, D\eta F(\eta)\exp(-\frac{1}{2}\int dx\, dt\, \bar\eta(x,t)\eta(x,t))}{\int D\bar\eta\, D\eta \exp(-\frac{1}{2}\int dx\, dt\, \bar\eta(x,t)\eta(x,t))}.$$

If the $H_{\rm FP}$ defined above does not have positive semidefinite eigenvalues, then one can modify the Hamiltonian in an appropriate manner. Application of this extension has already been considered in different processes. We in the present paper would like to consider a straightforward extension of Parisi and Wu stochastic quantization for spinor fields obeying generalized statistics, known in the literature as para-Fermi statistics, originally attributed to Green.[4]

The method of canonical quantization for para-Fermi fields is well-known.[5] It is based on the use of trilinear commutation relations satisfied by the field operators $\psi(x)$ and $\bar\psi(x)$,

$$[\psi(x),[\psi^\dagger(y),\psi^\dagger(z)]_-]_- = 2\delta^4(x-y)\psi^\dagger(z) - 2\delta^4(x-z)\psi^\dagger(y), \tag{1.7a}$$

 0022-2488/90/010156-08$03.00

$$[\psi(x),[\psi^\dagger(y),\psi(z)]_-]_- = 2\delta^4(x-y)\psi(z), \tag{1.7b}$$

$$[\psi(x),[\psi(y),\psi(z)]_-]_- = 0. \tag{1.7c}$$

The point is that the Heisenberg equations of motion for the field operators lead to the desired field equations not only if ψ, $\psi^\dagger$ satisfy the standard Fermi bilinear quantization rules but also if they satisfy the trilinear commutation rules (1.7). If one expands ψ, $\psi^\dagger$ in terms of annihilation and creation operators $b_\gamma, b_\gamma^\dagger$, the relevant equations (1.7a)–(1.7c) assume the following equivalent form:

$$[b_\alpha,[b_\beta^\dagger,b_\gamma^\dagger]_-]_- = 2\delta_{\alpha\beta}b_\gamma^\dagger - 2\delta_{\alpha\gamma}b_\beta^\dagger, \tag{1.8a}$$

$$[b_\alpha,[b_\beta^\dagger,b_\gamma]_-]_- = 2\delta_{\alpha\beta}b_\gamma, \tag{1.8b}$$

$$[b_\alpha,[b_\beta,b_\gamma]_-]_- = 0 \tag{1.8c}$$

In order to define the theory completely one postulates a vacuum state which satisfies the following properties:

$$b_\alpha^\dagger|0\rangle = 0; \quad b_\beta b_\alpha^\dagger|0\rangle = p\delta_{\alpha\beta}|0\rangle. \tag{1.9}$$

The positive integer p normally known as the order of para-Fermi statistics is the maximum number of parafermions that a totally symmetric state can accommodate. What is most important for what follows is that à la Green,[4] a para-Fermi operator of order p can be uniquely represented by the sum

$$b_\alpha = \sum_{i=1}^{p} b_\alpha^i, \tag{1.10}$$

where the so-called Green components $b_\alpha^i, b_\alpha^{i\dagger}$ satisfy (anti)-commutation rules of the anomalous type

$$[b_\alpha^i,b_\beta^j]_+ = [b_\alpha^i,b_\beta^{j\dagger}]_+ - \delta_{\alpha\beta} = 0, \tag{1.11a}$$

$$[b_\alpha^i,b_\beta^j]_- = [b_\alpha^i,b_\beta^{j\dagger}]_- = 0; \quad \text{for } i\neq j. \tag{1.11b}$$

The path-integral quantization of para-Fermi fields[6] is not as well known as their canonical quantization just mentioned. The former involves integration over para-Grassmann fields. They are defined in Sec. II where we show the equivalence of the n-point function of Euclidean field theories with the steady state limit of the equal time correlation functions computed from the following LE's [see (2.23) and (2.24)],

$$\frac{\partial}{\partial t}\psi(x,t) = -\frac{\delta S}{\delta\bar\psi(x,t)} + \eta(x,t), \tag{1.12}$$

$$\frac{\partial}{\partial t}\bar\psi(x,t) = +\frac{\partial S}{\partial\psi(x,t)} + \bar\eta(x,t). \tag{1.13}$$

Here $\psi(x,t)$ and $\bar\psi(x,t)$ are independent stochastic para-Grassmann fields. The statistical properties of the para-Grassmann noise functions η and $\bar\eta$ are given by

$$\langle\eta_\alpha(x,t)\bar\eta_\beta(x',t')\rangle_{\eta,\bar\eta} = -\langle\bar\eta_\beta(x',t')\eta_\alpha(x,t)\rangle_{\eta,\bar\eta} = 2p\delta_{\alpha\beta}\delta^4(x-x')\delta(t-t') \tag{1.14}$$

[see (2.25a)–(2.25e)].

Construction of action S, with a bilinear form like that of ordinary Fermi fields, has been discussed at length by Kamefuchi and Ohnuki.[6] They have shown that consistent with weak locality, an action of the bilinear form $S = \frac{1}{2}\int dx[\bar\psi,K\psi]_-$, where K is a linear operator, can always be written down for order of statistics $p\geqslant 3$. For $p=2$ the action contains additional terms such as $\int dx\,\kappa[\bar\psi,\psi]_+$, where κ is a nonvanishing real parameter. The case $p=2$ has been shown by Kamefuchi and Ohnuki to be equivalent to two ordinary Fermi fields with different masses through Klein transformation. We shall restrict ourselves to $p\neq 2$.

In Sec. II we set out with the aforementioned form of the action. We assert a Langevin equation and derive the Fokker–Planck equation which corresponds to it, and examine the spectrum of the Fokker–Planck Hamiltonian thus derived to prove the equivalence between Euclidean field theories and stochastic processes. The construction of this Fokker–Planck Hamiltonian in terms of para-Grassmann variables proceeds parallel to its construction in the Fermi case using Grassmann variables. In Sec. III the formalism is applied to calculate the two-point Green's function for two para-Fermi fields and show the distinction of the Green's function for para-Fermi fields from that for ordinary Fermi fields. Our results agree with those obtained earlier through canonical quantization. In Sec. IV we discuss another application of the formalism and obtain the normal and the anomalous Ward identities for a para-Fermi field, directly from the Langevin equation of motion. The Appendix is devoted to a discussion of the averages defined over para-Grassmann variables and a derivation of the analog of Novikov's theorem for such variables.

II. QUANTIZATION OF PARA-FERMI FIELDS

In this section we are mainly concerned with developing a stochastic quantization scheme for para-Fermi fields. This is done by an extension of these methods which were applied earlier to Fermi fields by Fukai *et al.*[2] For the derivation given below, we try to generalize the work of Chaturvedi, Kapoor, and Srinivasan.[7]

As before, our starting point is the functional formulation of quantum field theory. In this approach, the averages of products of the para-Fermi fields with respect to a weight factor $\exp(-S[\psi,\bar\psi])$ yield the Green's functions $\langle\psi(x_1)\psi(x_2)\cdots\bar\psi(x_n)\rangle$ of the field theory. Here $S[\psi,\bar\psi]$ is the action, which we take to be bilinear in the fields

$$S[\psi,\bar\psi] = \int dx\,\tfrac{1}{2}[\bar\psi,K\psi]_-. \tag{2.1}$$

Unlike the fermion case, an parafermion action cannot be cast in this form with the help of auxiliary scalar fields. But for the application we discuss in Sec. III, the bilinear terms suffice. Unless the order p of parastatistics equals two, Kamefuchi and Ohnuki[6] argue that the most general bilinear in parafields consistent with the requirement of weak locality has the form (2.1). K in (2.1) may in general contain scalar fields, derivative operators, and γ matrices. The peculiar case of order two will be discussed thoroughly elsewhere. The fields, $\psi(x)$ and $\bar\psi(x)$, neither commute nor anticommute. Rather they constitute p Green components,

$$\psi(x) = \sum_{a=1}^{p}\psi^a(x), \quad \bar\psi(x) = \sum_{a=1}^{p}\bar\psi^a(x), \tag{2.2}$$

which satisfy the following anomalous (anti-) commutation relations

$$[\psi^a(x),\psi^a(x')]_+ = [\psi^a(x),\bar{\psi}^a(x')]_+ = [\bar{\psi}^a(x),\bar{\psi}^a(x')]_+ = 0, \tag{2.3a}$$

$$[\psi^a(x),\psi^b(x')]_- = [\psi^a(x),\bar{\psi}^b(x')]_- = [\bar{\psi}^a(x),\bar{\psi}^b(x')]_- = 0, \quad a\neq b. \tag{2.3b}$$

Such fields as ψ, $\bar{\psi}$, known as para-Grassmann fields, are formally defined in terms of an infinite number of para-Grassmann numbers in the following manner:

$$\psi(x) = \sum_{i=1}^{\infty} \phi_i(x)\alpha_i, \quad \bar{\psi}(x) = \sum_{i=1}^{\infty} \phi_i^*(x)\bar{\alpha}_i, \tag{2.4}$$

where $\phi_i(x)$ is a complete set of orthonormal functions. Note that α_i and $\bar{\alpha}_i$ are independent para-Grassmann numbers. Such numbers are in turn decomposable into p Green components

$$\alpha_i = \sum_{a=1}^{p} \alpha_i^a, \quad \bar{\alpha}_i = \sum_{a=1}^{p} \bar{\alpha}_i^a$$

which satisfy

$$[\alpha_i^a,\alpha_j^a]_+ = [\alpha_i^a,\bar{\alpha}_j^a]_+ = [\bar{\alpha}_i^a,\bar{\alpha}_j^a]_+ = 0, \tag{2.5a}$$

$$[\alpha_i^a,\alpha_j^b]_- = [\alpha_i^a,\bar{\alpha}_j^b]_- = [\bar{\alpha}_i^a,\bar{\alpha}_j^b]_- = 0, \quad a\neq b. \tag{2.5b}$$

It is evident from (2.5a) that for a given value of the Green index the numbers $\alpha_i^a,\bar{\alpha}_i^a$ are ordinary Grassmann numbers. The functional integration measure in the prescription for obtaining Green's functions,

$$\langle\psi(x_1)\cdots\bar{\psi}(x_n)\rangle = \frac{\int D\psi\, D\bar{\psi}\, \psi(x_1)\cdots\bar{\psi}(x_n)\exp(-S[\psi,\bar{\psi}])}{\int D\psi\, D\bar{\psi}\exp(-S[\psi,\bar{\psi}])}, \tag{2.6}$$

must be understood as integrations over α_i^a and $\bar{\alpha}_i^a$,

$$\int D\psi\, D\bar{\psi} = \int \prod_{j=1}^{\infty}\prod_{a=1}^{p} d\bar{\alpha}_j^a d\alpha_j^a. \tag{2.7}$$

In terms of α_i^a and $\bar{\alpha}_i^a$ the action can also be written as

$$S = \sum_{a=1}^{p}\int dx\,\bar{\psi}^a K\psi^a = \sum_{a=1}^{p}\sum_{i,j=1}^{\infty}\bar{\alpha}_i^a K_{ij}\alpha_j^a, \tag{2.8}$$

where

$$K_{ij} = \int dx\,\phi_i^* K\phi_j. \tag{2.9}$$

Note that in the double sum over Green indices implied in S, the off-diagonal terms have dropped out, leaving only a direct sum in Green space, because of (2.1).

Computation of Green's functions for the para-Fermi fields thus essentially amounts to computation of averages of products of para-Grassmann numbers α_i and $\bar{\alpha}_i$ with respect to the weight factor

$$\exp\left(-\sum_{a=1}^{p}\sum_{ij=1}^{\infty}\bar{\alpha}_i^a K_{ij}\bar{\alpha}_j^a\right). \tag{2.10}$$

For the stochastic quantization of para-Fermi fields, our aim as mentioned in the Introduction is to introduce a fictitious time t and a probability distribution $P(\alpha,\bar{\alpha},t)$ whose evolution in the fictitious time is such that in the steady state limit, i.e., in the limit $t\to\infty$, $P(\alpha,\bar{\alpha},t)$ relaxes to the weight factor (2.10). The evolution equation, called a Fokker–Planck equation [even though in classical stochastic processes there exists no analog of $P(\alpha,\bar{\alpha},t)$ defined over para-Grassmann variables], is written as

$$\frac{\partial}{\partial t}P(\alpha,\bar{\alpha},t) = -H_{\rm FP}P(\alpha,\bar{\alpha},t). \tag{2.11}$$

A possible choice for the operator $H_{\rm FP}$ is the following:

$$H_{\rm FP} = \sum_{a=1}^{p}\sum_{i=1}^{\infty}\left[\frac{\partial}{\partial\alpha_i^a}\left(\frac{\partial}{\partial\bar{\alpha}_i^a}+\sum_{j=1}^{\infty}K_{ij}\alpha_j^a\right)\times\frac{-\partial}{\partial\bar{\alpha}_i^a}\left(\frac{\partial}{\partial\alpha_i^a}-\sum_{j=1}^{\infty}\bar{\alpha}_j^a K_{ji}\right)\right]. \tag{2.12}$$

In view of the algebra (2.5), it can be easily checked that (2.10) is a stationary solution of (2.11), i.e., it is an eigenfunction of $H_{\rm FP}$ with zero eigenvalue. In terms of the eigenfunctions χ_n and the corresponding eigenvalues λ_n of the operator $H_{\rm FP}$, the general solution of (2.11) can be given by the expansion

$$P(\alpha,\bar{\alpha},t) = \sum_n \chi_n \exp(-\lambda_n t). \tag{2.13}$$

To examine the spectrum of $H_{\rm FP}$, we regard $P(\alpha,\bar{\alpha},t)$ as a representative of an abstract vector $|P(t)\rangle$ in the coherent state representation.[6] This amounts to casting (2.11) into the following form:

$$\frac{\partial}{\partial t}|P(t)\rangle = -H_{\rm FP}|P(t)\rangle, \tag{2.14}$$

where the operator $H_{\rm FP}$ in (2.14) is obtained from the operator $H_{\rm FP}$ in (2.12) by making the following replacements:

$$\frac{\partial}{\partial\alpha_i},\frac{\partial}{\partial\bar{\alpha}_i},\frac{\partial}{\partial\alpha_i^a},\frac{\partial}{\partial\bar{\alpha}_i^a}\to A_i^\dagger,B_i^\dagger,A_i^{a\dagger},B_i^{a\dagger}, \quad \alpha_i,\bar{\alpha}_i,\alpha_i^a,\bar{\alpha}_i^a\to A_i,B_i,A_i^a,B_i^a, \tag{2.15}$$

and is consequently given by

$$H_{\rm FP} = \sum_{i=1}^{\infty}\sum_{a=1}^{p}\left[A_i^{a\dagger}\left(B_i^{a\dagger}+\sum_{j=1}^{\infty}K_{ij}A_j^a\right) - B_i^{a\dagger}\left(A_i^{a\dagger}-\sum_{j=1}^{\infty}B_j^a K_{ji}\right)\right]. \tag{2.16}$$

Here in (2.15) $A_i^\dagger,B_i^\dagger$ are independent parafermion creation operators and A_i,B_i are the corresponding annihilation operators. They satisfy the well-known trilinear relations and the superscript a on them labels their p Green components which satisfy the following well-known anomalous algebra:

$$[A_i^a,A_j^a]_+ = [A_i^a,B_j^a]_+ = [A_i^a,B_j^{\dagger a}]_+ = [A_i^a,A_j^{\dagger a}]_+ - \delta_{ij} = 0, \quad [A_i^a,A_j^b]_- = [A_i^a,B_j^b] = [A_i^a,B_j^{\dagger b}]_- = [A_i^a,A_j^{\dagger b}]_- = 0, \quad a\neq b, \text{ etc.} \tag{2.17}$$

Now we consider the following similarity transformation:

$$H_{\rm FP} = \exp\Big(\sum_{i,j,a} A_i^{a\dagger} K_{ij}^{-1} B_j^a\Big) \times H_{\rm FP} \exp\Big(-\sum_{i,j,a} A_i^{a\dagger} K_{ij} B_j^a\Big). \tag{2.18}$$

Under this transformation,

$$\begin{aligned} &A_i^{a\dagger} \to A_i^{a\dagger}, \quad B_i^{a\dagger} \to B_i^{a\dagger}, \\ &A_i^a \to A_i^a - \sum_j K_{ij}^{-1} B_j^{a\dagger}, \\ &B_i^a \to B_i^a + \sum_j K_{ij}^{-1} A_j^a, \end{aligned} \tag{2.19}$$

so that $\hat{H}_{\rm FP}$ is given by

$$\hat{H}_{\rm FP} = \sum_{i,j,a} (A_i^{a\dagger} K_{ij} A_j^a + B_i^{\dagger a} K_{ji} B_j^a). \tag{2.20}$$

From the structure of $\hat{H}_{\rm FP}$ in (2.20), it follows that if K is a positive definite operator, then so is $\hat{H}_{\rm FP}$ and hence $H_{\rm FP}$. This means that every solution like (2.13) of (2.11) in the limit $t \to \infty$ approaches the eigenfunction (2.10) of $\hat{H}_{\rm FP}$ corresponding to the zero eigenvalue, provided that this eigenvalue is nondegenerate and that there is a gap in the spectrum of H above the zero eigenvalue. Actually, to guarantee the relaxation of $P(\alpha,\bar{\alpha},t)$ to (2.10) regardless of the nature of K, it is possible to modify (2.12) without altering the stationary solution of (2.11) in such a way that the resulting $\hat{H}_{\rm FP}$ is positive definite. Instead of (2.12) we may consider the following form for $H_{\rm FP}$:

$$H_{\rm FP} = \sum_{l,m,a} \left[\frac{\partial}{\partial \alpha_l^a} G_{ml}^* \left(\frac{\partial}{\partial \bar{\alpha}_m^a} + \sum_j K_{mj} \alpha_j^a\right) - \frac{\partial}{\partial \bar{\alpha}_l^a} G_{lm}^* \left(\frac{\partial}{\partial \alpha_m^a} - \sum_j K_{jm} \bar{\alpha}_j^a\right)\right], \tag{2.21}$$

where G is arbitrary. This equation still has (2.10) as its zero eigenmode. However, the corresponding $\hat{H}_{\rm FP}$ is modified and becomes

$$\hat{H}_{\rm FP} = \sum_{a=1}^{p} \sum_{i,j=1}^{\infty} (B_i^{a\dagger} (GK^\dagger)_{ij}^* B_j^a + A_i^{a\dagger} (G^\dagger K)_{ij} A_j^a). \tag{2.22}$$

If we now set $G = K$, then $\hat{H}_{\rm FP}$ becomes manifestly positive definite regardless of the nature of K. In fact, to guarantee the desired relaxation, it is enough to require that eigenvalues of $GK^\dagger$ have positive real parts. This observation proves to be quite convenient for perturbative calculations with stochastically quantized parafermion fields.

Instead of working with the FPE (2.11), for actual computations, it proves much more convenient to work with an equivalent formulation based on the Langevin equation (LE). This is a stochastic differential equation governing the time evolution of ψ and $\bar{\psi}$ in place of an evolution equation for $P[\psi,\bar{\psi},t]$. It can be shown that the single time averages $\langle \psi(x_1,t) \cdots \bar{\psi}(x_n,t) \rangle_p$ computed using the FPE are identical to the equal time averages obtained using the LE. We begin with an assertion that the LE's,

$$\frac{\partial}{\partial t} \psi^a(x,t) = \frac{-\delta S}{\delta \bar{\psi}^a(x,t)} + \eta^a(x,t), \tag{2.23}$$

$$\frac{\partial}{\partial t} \bar{\psi}^a(x,t) = +\frac{\delta S}{\delta \psi^a(x,t)} + \bar{\eta}^a(x,t), \tag{2.24}$$

are equivalent to the FPE (2.11) with $H_{\rm FP}$ given by (2.12). Note that on summing over Green indices in (2.23) and (2.24) we get (1.12) and (1.13). Here $\psi^a(x,t)$ and $\bar{\psi}^a(x,t)$ are treated as independent stochastic fields, and the noise sources $\eta^a(x,t)$ and $\bar{\eta}^a(x,t)$ satisfy the following stochastic properties:

$$\langle \eta^a(x,t) \rangle = \langle \bar{\eta}^a(x,t) \rangle = 0, \tag{2.25a}$$

$$\langle \eta^a(x,t) \bar{\eta}^a(y,t') \rangle = -\langle \bar{\eta}^a(y,t') \eta^a(x,t) \rangle = 2\delta(x-y)\delta(t-t'), \tag{2.25b}$$

$$\langle \eta^a(x,t) \bar{\eta}^b(y,t') \rangle = +\langle \bar{\eta}^b(y,t') \eta^a(x,t) \rangle = 0, \quad a \neq b, \tag{2.25c}$$

$$\langle \eta^a(x,t) \eta^a(y,t') \rangle = -\langle \eta^a(y,t') \eta^a(x,t) \rangle = 0, \tag{2.25d}$$

$$\langle \eta^a(x,t) \eta^b(y,t') \rangle = +\langle \eta^b(y,t') \eta^a(x,t) \rangle = 0,$$

$a \neq b$, etc. (2.25e)

These properties are summarized by the following distribution for the noise sources:

$$\exp\Big(-\frac{1}{2} \sum_{a=1}^{p} \int dx\, dt\, \bar{\eta}^a(x,t) \eta^a(x,t)\Big). \tag{2.26}$$

We are now ready to prove the assertion made above regarding the equivalence between the LE and the FPE. Consider an arbitrary functional F of ψ and $\bar{\psi}$. Then in the Langevin approach,

$$\left\langle \frac{\partial F}{\partial t} \right\rangle_{\eta,\bar{\eta}} = \sum_{a=1}^{p} \int dx \left(\left\langle \frac{\partial \psi^a(x,t)}{\partial t} \frac{\partial F}{\partial \psi^a(x,t)} \right\rangle_{\eta,\bar{\eta}} + \left\langle \frac{\partial \bar{\psi}^a(x,t)}{\partial t} \frac{\delta F}{\delta \bar{\psi}^a(x,t)} \right\rangle_{\eta,\bar{\eta}} \right). \tag{2.27}$$

On using (2.23) and (2.24) we get

$$\left\langle \frac{\partial F}{\partial t} \right\rangle_{\eta,\bar{\eta}} = \sum_{a=1}^{p} \int dx \left(\left\langle -\frac{\delta S}{\delta \bar{\psi}^a(x,t)} \frac{\delta F}{\delta \psi^a(x,t)} \right\rangle_{\eta,\bar{\eta}} + \left\langle \frac{\delta S}{\delta \psi^a(x,t)} \frac{\delta F}{\delta \bar{\psi}^a(x,t)} \right\rangle_{\eta,\bar{\eta}} + \left\langle \eta^a(x,t) \frac{\delta F}{\delta \psi^a(x,t)} \right\rangle_{\eta,\bar{\eta}} + \left\langle \bar{\eta}^a(x,t) \frac{\delta F}{\delta \bar{\psi}^a(x,t)} \right\rangle_{\eta,\bar{\eta}} \right). \tag{2.28}$$

With the help of the distribution of η and $\bar{\eta}$ as given by (2.26) one can prove an analog of Novikov's theorem for these para-Grassmann noise sources (see Appendix A),

$$\left\langle \eta^a(x,t) \frac{\delta F}{\delta \psi^a(x,t)} \right\rangle_{\eta,\bar{\eta}} = 2 \int dx' \sum_{b=1}^{p} \left\langle \frac{\delta \bar{\psi}^b(x',t)}{\delta \bar{\eta}^a(x,t)} \frac{\delta^2 F}{\delta \bar{\psi}^b(x',t) \delta \psi^a(x,t)} \right\rangle_{\eta,\bar{\eta}} \tag{2.29}$$

$$\left\langle \bar{\eta}^a(x,t)\frac{\delta F}{\delta\bar{\psi}^a(x,t)}\right\rangle_{\eta\bar{\eta}} = -2\int dx' \sum_{b=1}^{p}\left\langle \frac{\delta\psi^b(x',t)}{\delta\eta^a(x,t)}\frac{\delta^2 F}{\delta\psi^b(x',t)\delta\bar{\psi}^a(x,t)}\right\rangle_{\eta\bar{\eta}}. \tag{2.30}$$

One can show from LE's (2.23) and (2.24) that

$$\frac{\delta\bar{\psi}^b(x',t)}{\delta\bar{\eta}^a(x,t)} = \frac{\delta\psi^b(x',t)}{\delta\eta^a(x,t)} = \delta_{ab}\delta(x-x')\Theta(0) = (\tfrac{1}{2})\delta_{ab}\delta(x-x'). \tag{2.31}$$

The step function in (2.30) arises from the integration over t of LE's after being differentiated w.r.t. the noise sources. Substituting from (2.31) into (2.29) and (2.30) and using the expressions thus obtained in (2.28), we get

$$\left\langle\frac{\partial F}{\partial t}\right\rangle_{\eta\bar{\eta}} = \sum_{a=1}^{p}\int dx\left(\left\langle -\frac{\delta S}{\delta\bar{\psi}^a(x,t)}\frac{\delta F}{\delta\psi^a(x,t)}\right\rangle_{\eta\bar{\eta}} + \left\langle\frac{\delta S}{\delta\psi^a(x,t)}\frac{\delta F}{\delta\bar{\psi}^a(x,t)}\right\rangle_{\eta\bar{\eta}} + \left\langle\frac{\delta^2 F}{\delta\bar{\psi}^a(x,t)\delta\psi^a(x,t)}\right\rangle_{\eta\bar{\eta}} - \left\langle\frac{\delta^2 F}{\delta\psi^a(x,t)\delta\bar{\psi}^a(x,t)}\right\rangle_{\eta\bar{\eta}}\right). \tag{2.32}$$

The connection with the Fokker–Planck approach is made by identifying the averages in the two approaches,

$$\langle F[\psi,\bar{\psi}]\rangle_{\eta\bar{\eta}} = \int D\psi\, D\bar{\psi}\, F[\psi,\bar{\psi}]P[\psi,\bar{\psi},t] \equiv \langle F[\psi,\bar{\psi}]\rangle_P, \tag{2.33}$$

so that (2.32) also holds for P averages,

$$\int D\psi\, D\bar{\psi}\, F[\psi,\bar{\psi}]\frac{\partial}{\partial t}P(\psi,\bar{\psi},t) = \sum_{a=1}^{p}\int dx\left(-\left\langle\frac{\delta S}{\delta\bar{\psi}^a(x)}\frac{\delta F}{\delta\psi^a(x)}\right\rangle_P + \left\langle\frac{\delta S}{\delta\psi^a(x)}\frac{\delta F}{\delta\bar{\psi}^a(x)}\right\rangle_P + \left\langle\frac{\delta^2 F}{\delta\bar{\psi}^a(x)\delta\psi^a(x)}\right\rangle_P - \left\langle\frac{\delta^2 F}{\delta\psi^a(x)\delta\bar{\psi}^a(x)}\right\rangle_P\right). \tag{2.34}$$

Next our aim is to transfer the functional derivative(s) on F onto P in each of the four terms on the right-hand side of (2.34). This is done by integration by parts, keeping in mind the anticommuting nature of $\psi^a(x)$ and $\bar{\psi}^a(x)$. The result is

$$\int D\psi\, D\bar{\psi}\, F[\psi,\bar{\psi}]\frac{\partial P}{\partial t} = \int D\psi\, D\bar{\psi}\, F \times\left(-\frac{\delta}{\delta\psi^a}\left(\frac{\delta S}{\delta\bar{\psi}^a}P\right)+\frac{\delta}{\delta\bar{\psi}^a}\left(\frac{\delta S}{\delta\psi^a}P\right) - \frac{\delta^2 P}{\delta\psi^a\delta\bar{\psi}^a}+\frac{\delta^2 P}{\delta\bar{\psi}^a\delta\psi^a}\right). \tag{2.35}$$

Since F is arbitrary, we obtain the following equation for P:

$$\frac{\partial}{\partial t}P(\psi,\bar{\psi},t) = -\sum_{a=1}^{p}\int dx\left[\frac{\delta}{\delta\psi^a(x)}\left(\frac{\delta}{\delta\bar{\psi}^a(x)}+\frac{\delta S}{\delta\bar{\psi}^a(x)}\right) - \frac{\delta}{\delta\bar{\psi}^a(x)}\left(\frac{\delta}{\delta\psi^a(x)}+\frac{\delta S}{\delta\psi^a(x)}\right)\right]P. \tag{2.36}$$

This reduces to

$$\frac{\partial}{\partial t}P(\psi,\bar{\psi},t) = -\sum_{a=1}^{p}\int dx\left[\frac{\delta}{\delta\psi^a(x)}\left(\frac{\delta}{\delta\bar{\psi}^a(x)}+K\psi^a(x)\right) - \frac{\delta}{\delta\bar{\psi}^a}(x)\left(\frac{\delta}{\delta\psi^a(x)}-K^T\bar{\psi}^a(x)\right)\right]P \tag{2.37}$$

on using the first equation in (2.8). Noting that the definitions (2.4) for $\bar{\psi}^a(x)$ and $\psi^a(x)$ imply the following definitions for the functional derivatives with respect to $\psi^a(x)$ and $\bar{\psi}^a(x)$,

$$\frac{\delta}{\delta\psi^a(x)} = \sum_{i=1}^{\infty}\phi_i^*(x)\frac{\partial}{\partial\alpha_i^a};\quad \frac{\delta}{\delta\bar{\psi}^a(x)} = \sum_{i=1}^{\infty}\phi_i(x)\frac{\partial}{\partial\bar{\alpha}_i^a}, \tag{2.38}$$

we can translate (2.11) and (2.12) into the following functional equation for P,

$$\frac{\partial}{\partial t}P[\psi,\bar{\psi},t] = -H_{\rm FP}P[\psi,\bar{\psi},t] = -\sum_{a=1}^{p}\int dx\left[\frac{\delta}{\delta\psi^a(x)}\left(\frac{\delta}{\delta\psi^a(x)}+K\psi^a(x)\right) - \frac{\delta}{\delta\bar{\psi}^a(x)}\left(\frac{\delta}{\delta\psi^a(x)}-K^T\bar{\psi}^a(x)\right)\right]P \tag{2.39}$$

which is identical with (2.37).

For completeness it is added that proceeding in exactly the same manner as above one can show that the LE's,

$$\frac{\partial}{\partial t}\psi^a(x,t) = -G^\dagger K\psi^a(x,t)+G^\dagger\eta^a(x,t), \tag{2.40}$$

$$\frac{\partial}{\partial t}\bar{\psi}^a(x,t) = -(KG^\dagger)^T\bar{\psi}^a(x,t)+\bar{\eta}^a(x,t), \tag{2.41}$$

with η and $\bar{\eta}$ having the properties (2.25) and (2.26), are equivalent to the FPE (2.11) with $H_{\rm FP}$ (2.21). As explained before, one may set $G=K$. Finally, in the event G is a constant, i.e., it does not contain any fields which are also stochastically quantized, then we may alternatively write (2.40) and (2.41) in the following form,

$$\frac{\partial}{\partial t}\psi^a(x,t) = G^\dagger K\psi^a(x,t)+\theta^a(x,t), \tag{2.42}$$

$$\frac{\partial}{\partial t}\bar{\psi}^a(x,t) = -(KG^\dagger)^T\bar{\psi}^a(x,t)+\bar{\theta}^a(x,t), \tag{2.43}$$

with

$$\begin{aligned}\langle\theta^a(x,t)\bar{\theta}^a(x',t')\rangle &= -\langle\bar{\theta}^a(x',t')\theta^a(x,t)\rangle\\ &= 2G^\dagger\delta(x-x')\delta(t-t'),\\ \langle\theta^a(x,t)\bar{\theta}^b(x',t')\rangle &= +\langle\bar{\theta}^b(x',t')\theta^a(x,t)\rangle\\ &= 0,\quad a\neq b,\ \text{etc.}\end{aligned} \tag{2.44}$$

Let us now briefly recapitulate the discussion in this section. Here we have considered two FPE's, and subsequently derived the corresponding LE's. They both have $\exp(-\int dx\, \frac{1}{2}[\bar{\psi},K\psi]_-)$ as their stationary solution. For this to be the unique steady state solution, the corresponding $H_{\rm FP}$ must have the following attributes:

(i) Its lowest eigenvalue must be zero and the other eigenvalues must have a positive real part.

(ii) The nonzero eigenvalue must be nondegenerate.

We have been concerned mainly with how (i) could always be ensured. The second condition is in general difficult to establish.

III. FREE PARAFERMIONS

As an application of the extension of the Parisi–Wu formalism for parafermions developed in the last section, we consider free parafermions.

The Euclidean action for free parafermions is given by

$$S=\int d^4x\frac{1}{2}[\bar{\psi}(x),(\gamma_\mu\partial_\mu+m)\psi(x)]$$

$$=\sum_{a=1}^{p}\int d^4x\,\bar{\psi}^a(x)(\gamma_\mu\partial_\mu+m)\psi^a(x), \tag{3.1}$$

where γ_μ are Hermitian matrices obeying the following anticommutation relations:

$$[\gamma_\mu\gamma_\nu]_+=2\delta_{\mu\nu}. \tag{3.2}$$

We first consider the LE's (2.23) and (2.24), which in this case with $K=(\gamma_\mu\partial_\mu+m)$ become

$$\frac{\partial}{\partial t}\psi^a_\alpha(x,t)=-(\gamma_\mu\partial_\mu+m)_{\alpha\epsilon}\psi^a_\epsilon(x,t)+\eta^a_\alpha(x,t), \tag{3.3}$$

$$\frac{\partial}{\partial t}\bar{\psi}^a_\beta(x,t)=-\bar{\psi}^a_\delta(-\gamma_\mu\partial_\mu+m)_{\delta\beta}+\bar{\eta}^a_\beta(x,t), \tag{3.4}$$

with

$$\langle\eta^a_\alpha(x,t)\bar{\eta}^a_\beta(x',t')\rangle=-\langle\bar{\eta}^a_\beta(x',t')\eta^a_\alpha(x,t)\rangle$$
$$=2\delta_{\alpha\beta}\delta(x-x')\delta(t-t'), \tag{3.5a}$$
$$\langle\eta^a_\alpha(x,t)\bar{\eta}^b_\beta(x',t')\rangle=\langle\bar{\eta}^b_\beta(x',t')\eta^a_\alpha(x,t)\rangle$$
$$=0,\quad \text{for } a\neq b. \tag{3.5b}$$

Fourier transforming (3.3) and (3.4) w.r.t. x,

$$\frac{\partial}{\partial t}\psi^a_\alpha(k,t)=-(i\not{k}+m)_{\alpha\epsilon}\psi^a_\epsilon(k,t)+\eta^a_\alpha(k,t), \tag{3.6}$$

$$\frac{\partial}{\partial t}\bar{\psi}^a_\beta(k,t)=-(-i\not{k}+m)_{\delta\beta}\bar{\psi}^a_\delta(k,t)+\bar{\eta}^a_\beta(k,t), \tag{3.7}$$

and solving them with the initial conditions

$$\psi_\alpha(k,0)=\bar{\psi}_\alpha(k,0)=0, \tag{3.8}$$

we get

$$\psi^a_\alpha(k,t)=\int_0^t dt_1[\exp-(i\not{k}+m)(t-t_1)]_{\alpha\epsilon}\eta^a_\epsilon(k,t_1), \tag{3.9}$$

$$\bar{\psi}^b_\beta(k',t')=\int_0^{t'}dt_2\,\bar{\eta}^b_\delta(k',t_2)[\exp-(i\not{k}'+m)(t'-t_2)]_{\delta\beta}. \tag{3.10}$$

Using (3.5) we can easily compute the equal time correlation function $\langle\psi^a_\alpha(k,t)\bar{\psi}^b_\beta(k',t)\rangle_{\eta\bar{\eta}}$,

$$\langle\psi^a_\alpha(k,t)\bar{\psi}^b_\beta(k',t)\rangle=\frac{\delta(k+k')}{(i\not{k}+m)}\delta_{ab}(1-\exp(-2(i\not{k}+m)t)). \tag{3.11}$$

We now want to take the limit $t\to\infty$. If m is nonzero, this limit exists because the eigenvalues of the matrix $(i\not{k}+m)$, which are $m\pm i\sqrt{k^2}$, have positive real parts. Hence so long as $m\neq0$, the limit $t\to\infty$ of the two-point correlation function $\langle\psi_\alpha(k,t)\bar{\psi}_\beta(k',t)\rangle_{\eta\bar{\eta}}$ exists and is given by

$$\lim_{t\to\infty}\langle\psi_\alpha(k,t)\bar{\psi}_\beta(k',t)\rangle_{\eta\bar{\eta}}$$
$$=\sum_{a=1}^{p}\sum_{b=1}^{p}\lim_{t\to\infty}\langle\psi^a_\alpha(k,t)\bar{\psi}^b_\beta(k',t)\rangle$$
$$=p\delta(k+k')(i\not{k}+m)^{-1}. \tag{3.12}$$

To deal with the massless case we consider the other set of LE's based on (2.42) and (2.43) with $G=K$. When these LE's are Fourier transformed w.r.t. x and solved with the same initial conditions as (3.8), we obtain the following expression for the equal-time correlation function,

$$\langle\psi^a_\alpha(k,t)\bar{\psi}^b_\beta(k',t)\rangle_{\eta\bar{\eta}}=\frac{\delta(k+k')}{k^2+m^2}(-i\not{k}+m)_{\alpha\beta}\times\{1-\exp[-2(k^2+m^2)t]\}. \tag{3.13}$$

Clearly the limit $t\to\infty$ exists whether m is zero or not, and we obtain

$$\langle\psi_\alpha(k,t)\bar{\psi}_\beta(k',t)\rangle=\frac{p\delta(k+k')(-i\not{k}+m)_{\alpha\beta}}{(k^2+m^2)}. \tag{3.14}$$

This result agrees with that obtained from canonical quantization of parafermions.[5]

We note that we would have obtained the same results for the propagator if we had used the regularized Langevin equations (discussed in the next section).

IV. THE AXIAL-VECTOR CURRENT ANOMALY AND THE VECTOR-CURRENT CONSERVATION FROM THE LANGEVIN EQUATIONS

As another interesting application of the formalism given above, we discuss in this section how one can obtain the normal and the anomalous ward identities for a para-Fermi field, directly from the Langevin equations of motion without even having to explicitly find their solutions. Following this equation of motion approach for anomalies, we recover the standard result for ordinary fermions when the order of the statistics equals unity. Following Bern *et al.*,[8] we use the covariant derivative regularization scheme in which the noise structure of the Langevin equations is generalized in a covariant way to

$$\frac{\partial\psi^a}{\partial\tau}(x,\tau) = [i\gamma_\mu(\vec{\partial}_\mu - igA_\mu) - m]\psi^a(x,\tau) + \int d^4y\, R(x,y)\eta^a(y,\tau), \tag{4.1}$$

$$\frac{\partial\bar{\psi}^a}{\partial\tau}(x,\tau) = \bar{\psi}^a(x,\tau)[i\gamma_\mu(-\overleftarrow{\partial}_\mu - igA_\mu) - m] + \int d^4y\, \bar{\eta}^a(y,\tau)R(x,y). \tag{4.2}$$

Here we have set $G=1$ for the sake of simplicity. The fermionic regulator $R(x,y)$ is a function of the covariant fermionic Laplacian $\not{D}^2$,

$$\not{D}^2(x,y) = \not{D}_x^2\delta(x-y).$$

The operator

$$R(x,y) = (\exp[-(\not{D}/\Lambda)^2])(x,y). \tag{4.3}$$

Note that R approaches unity as the cutoff Λ approaches infinity.

Multiplying Eq. (4.1) fromt he left-hand side by $\bar{\psi}^a\gamma_5$ and Eq. (4.2) from the right by $\gamma_5\psi^a$, adding the two, and taking the noise averages on both sides, we get

$$\partial_\tau \sum_a \langle\bar{\psi}^a\gamma_5\psi^a\rangle = i\partial_\mu \sum_a \langle\bar{\psi}^a\gamma_5\gamma_\mu\psi^a\rangle - 2m\sum_a\langle\bar{\psi}^a\gamma_5\psi^a\rangle + \sum_a\langle\bar{\psi}^a\gamma_5 R\eta^a\rangle + \sum_a\langle\bar{\eta}^a R\gamma_5\psi^a\rangle. \tag{4.4}$$

Again, multiplying Eq. (4.1) from the left by $\bar{\psi}$ and Eq. (4.2) from the right by ψ, subtracting the two, and noise averaging both sides, we obtain

$$\sum_a\langle\bar{\psi}^a\partial_\tau\psi^a\rangle - \sum_a\langle(\partial_\tau\bar{\psi}^a)\psi^a\rangle = i\partial_\mu\sum_a\langle\bar{\psi}^a\gamma_\mu\psi^a\rangle + \sum_a\langle\bar{\psi}^a R\eta^a\rangle - \sum_a\langle\bar{\eta}^a R\psi^a\rangle. \tag{4.5}$$

In the steady-state limit $\tau\to\infty$ we expect to arrive at the regulated Ward identities, because then the stochastic averages $\Sigma_a\langle\bar{\psi}^a\gamma_5\gamma_\mu\psi^a\rangle$ and $\Sigma_a\langle\bar{\psi}^a\gamma_\mu\psi^a\rangle$ reduce to the axial-vector current and vector current, respectively, in quantum field theory.

To evaluate the left-hand side of Eq. (4.4), consider a function $F(\psi,\bar{\psi})$ which is any arbitrary bilinear in ψ and $\bar{\psi}$,

$$\lim_{\tau\to\infty}\frac{\partial}{\partial\tau}\langle F\rangle = \lim_{\tau\to\infty}\int d\psi\, d\bar{\psi}\, F\frac{\partial}{\partial\tau}P[\psi,\bar{\psi},\tau] = -\lim_{\tau\to\infty}\int d\psi\, d\bar{\psi}\, FH_{\rm FP}P[\psi,\bar{\psi},\tau] = 0, \tag{4.6}$$

where the last equality follows because the probability distribution function relaxes to the ground state with zero eigenvalue in the limit $\tau\to\infty$. The time derivative of $\Sigma_a\langle\bar{\psi}^a\gamma_5\psi^a\rangle$ therefore vanishes.

The left-hand side of Eq. (4.5) is, from first principles,

$$\sum_a\langle\bar{\psi}^a\partial_\tau\psi^a\rangle - \sum_a\langle(\partial_\tau\bar{\psi}^a)\psi^a\rangle = \lim_{\delta\tau\to0}\delta\tau^{-1}\sum_a[\langle\bar{\psi}^a(x,\tau)\psi^a(x,\tau+\delta\tau)\rangle - \langle\bar{\psi}^a(x,\tau)\psi^a(x,\tau)\rangle - \langle\bar{\psi}^a(x,\tau+\delta\tau)\psi^a(x,\tau)\rangle + \langle\bar{\psi}^a(x,\tau)\psi^a(x,\tau)\rangle], \tag{4.7}$$

which vanishes because $\langle\psi(x,\tau)\bar{\psi}(x,\tau')\rangle$ is symmetric under the interchange of the arguments τ and τ'. To evaluate the last two terms on the right-hand sides of Eqs. (4.4) and (4.5), we make use of Novikov's theorem and obtain

$$\sum_a\left\langle\int d^4y\,\bar{\psi}^a(x,\tau)\gamma_5 R(x,y)\eta^a(x,\tau)\right\rangle = \sum_a\left\langle\int d^4y\,\bar{\eta}^a(x,\tau)R(x,y)\gamma_5\psi^a(x,\tau)\right\rangle = -\sum_{a=1}^{p}\left\langle x\left|\int d^4y\,\mathrm{Tr}\,\gamma_5 R^2(x,y)\right|x\right\rangle. \tag{4.8}$$

Finally, following Fujikawa[9] and taking the limit $\Lambda\to\infty$, Eq. (4.8) reduces to

$$\frac{-p}{16\pi^2}\mathrm{Tr}\,(*F^{\mu\nu}F_{\mu\nu}).$$

In addition, we have

$$\sum_a\left\langle\int d^4y\,\bar{\psi}^a(x,\tau)R(x,y)\eta^a(x,\tau)\right\rangle = \sum_a\left\langle\int d^4y\,\bar{\eta}^a(x,\tau)R(x,y)\psi^a(x,\tau)\right\rangle = -p\left\langle x\left|\int d^4y\,\mathrm{Tr}\,R(x,y)\right|x\right\rangle. \tag{4.9}$$

Thus we are led to the Ward identities in the quantum field theory

$$\sum_a\partial_\mu(\bar{\psi}^a\gamma_\mu\gamma_5\psi^a) = 2mi\sum_a(\bar{\psi}^a\gamma_5\psi^a) - \frac{ip}{8\pi^2}\mathrm{Tr}(*F^{\mu\nu}F_{\mu\nu}) \tag{4.10}$$

and

$$\sum_a\partial_\mu(\bar{\psi}^a\gamma_\mu\psi^a) = 0. \tag{4.11}$$

The procedure we have adopted is quite similar in spirit to the derivation of the chiral anomaly given by Namiki *et al.*[10]

APPENDIX

We wish to prove the analog of Novikov's theorem for para-Grassmann variables. The proof depends on a probability distribution,

$$P(\bar{\alpha},\alpha)\propto\exp\left(-\sum_{i,j}\sum_{a=1}^{p}\alpha_i^a A_{ij}\alpha_j^a\right), \tag{A1}$$

over the Green components α_i^a and $\bar{\alpha}_i^a$ of independent para-Grassman variables $\bar{\alpha}_i$ and α_i. For a given value of the Green

index, these Green components anticommute, but for different Green indices they commute. The average of a function F of α_i and $\bar{\alpha}_i$ is defined as

$$\langle F\rangle = \int \prod_a \prod_i d\bar{\alpha}_i^a \, d\alpha_i^a \, FP(\alpha,\bar{\alpha})$$

$$= \int \prod d\bar{\alpha} \, d\alpha \, FP(\alpha,\bar{\alpha}). \tag{A2}$$

We want to prove the relations

$$\langle \alpha_i^a F\rangle = \sum_k \langle \alpha_i^a \bar{\alpha}_k^a\rangle \left\langle \frac{\partial F}{\partial \bar{\alpha}_k^a}\right\rangle, \tag{A3}$$

$$\langle \bar{\alpha}_i^a F\rangle = -\sum_k \left\langle \frac{\partial F}{\partial \alpha_k^a}\right\rangle \langle \alpha_k^a \bar{\alpha}_i^a\rangle. \tag{A4}$$

Noting that $\langle \alpha_i^a \bar{\alpha}_k^a\rangle = A^{-1}_{ik}$, we may rewrite (A3) and (A4) as

$$\sum_k A_{ik}\langle \alpha_k^a F\rangle = \left\langle \frac{\partial F}{\partial \bar{\alpha}_i^a}\right\rangle, \tag{A5}$$

$$\sum_k \langle \bar{\alpha}_k^a F\rangle A_{ki} = -\left\langle \frac{\partial F}{\partial \alpha_i^a}\right\rangle, \tag{A6}$$

Consider the lhs of (A5),

$$\sum_k A_{ik}\langle \alpha_k^a F\rangle = \int \prod d\bar{\alpha} \, d\alpha \, A_{ik}\alpha_k^a FP(\alpha,\bar{\alpha})$$

$$= \int \prod d\bar{\alpha} \, d\alpha \, P(F) A_{ik}\alpha_k^a P(\alpha,\bar{\alpha})$$

$$= -\int \prod d\bar{\alpha} \, d\alpha \, P(F) \frac{\partial}{\partial \bar{\alpha}_i^a} P(\alpha,\bar{\alpha}), \tag{A7}$$

where $P(F)$ is obtained from F on changing the signs of α_i^a's and $\bar{\alpha}_i^a$'s in F. Carrying out an integration by parts in the next step, we get

$$\sum_k A_{ik}\langle \alpha_k^a F\rangle = +\int \prod d\bar{\alpha} \, d\alpha \frac{\partial F}{\partial \bar{\alpha}_i^a} P(\alpha,\bar{\alpha}), \tag{A8}$$

which is (A5) or (A3). Likewise one can prove (A6) or (A4). Equations (A3) and (A4) are the desired analogs of Novikov's theorem for the noise sources used in Sec. II.

ACKNOWLEDGMENT

Two of us (J.B. and S.K.S.) gratefully acknowledge financial support from the Council of Scientific and Industrial Research, New Delhi.

[1] G. Parisi and Y. S. Wu, Sci. Sin. **24**, 483 (1981).
[2] T. Fukai, H. Nakazato, I. Ohba, K. Okano, and Y. Yamanaka, Prog. Theor. Phys. **69**, 1600 (1983).
[3] B. Sakita, "Stochastic Quantization," in *Proceedings of the 7th Johns Hopkins Workshop* (World Scientific, Singapore, 1983).
[4] H. S. Green, Phys. Rev. **90**, 270 (1953).
[5] I. McCarthy, Proc. Cambridge Philos. Soc. **51**, 131 (1955).
[6] S. Kamefuchi and Y. Ohnuki, *Quantum Field Theory and Parastatistics* (University of Tokyo Press, Tokyo, 1982).
[7] S. Chaturvedi, A. K. Kapoor, and V. Srinivasan, "Stochastic Quantization Scheme of Parisi–Wu," HUTP/85-3.
[8] Z. Bern, H. S. Chan, and M. B. Halpern, Z. Phys. C **33**, 77 (1986).
[9] K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1979); Phys. Rev. D **21**, 2848 (1980).
[10] M. Namiki, I. Ohba, S. Tanaka, and D. Yanga, Phys. Lett. B **194**, 530 (1987).

State function and spectral properties of a two-component interacting Fermi gas

J. B. McGuire
Department of Physics, Florida Atlantic University, Boca Raton, Florida 33431

(Received 15 August 1989; accepted for publication 11 October 1989)

Rules are derived for the computation of the spectrum and state function of a system of N fermions in one dimension interacting with a delta function potential. The system is composed of fermions of two types of internal symmetry. The rules are valid for an arbitrary number of either symmetry type and any value of the delta function interaction strength. The state functions satisfy periodic boundary conditions, and therefore describe the interacting system at nonzero density.

I. INTRODUCTION

I address the problem of N fermions of two distinguishable types of internal symmetry (here called u and d), in one dimension, interacting with a delta function potential. There are to be K spin-up u particles and M spin-down d particles, $M \leqslant K$ and $K + M = N$. This problem is of the exactly solvable type, for it is algebraically factorized by an ansatz of no diffraction, often identified with the Bethe ansatz. Gaudin[1] is a major resource in tracing the history of this problem, and in explaining the algebraic relation of this to other problems, e.g., the Hubbard model.

Some time ago substantial effort was directed toward computation of the spectrum in the bulk limit for various values of M and all values of the interaction strength. Flicker and Lieb,[2] Gaudin,[3] Yang,[4] and Sutherland[5] conclude, with reluctance, that the excitation spectrum is not an analytic function of the delta function coupling constant at zero interaction strength. This conclusion is drawn with reluctance because zero interaction strength separates the attractive from the repulsive interaction, and the attractive case promises the most interesting interpretation due to the presence of bound states. The integral equation methods developed by these investigators gave all of the information about the spectrum for a repulsive interaction, but only very limited information about the system for an attractive interaction.

Interest has reawakened recently as a result of the complete characterization of the dynamics of the $M = 1$ case.[6,7] The excitation spectrum is indeed nonanalytic at zero interaction strength, but, through the state function, the nonanalyticity is understood as arising from the presence or absence of a pair. There is a pairing phase transition if the interaction is attractive.[8] The phase transition arises because the state function influences equilibrium, and thus the equilibrium properties are not a function of the spectrum alone. These results have fortified the opinion that a solution to the attractive M-down problem would contain important information about the superconducting or pair superfluid state of more realistic models. From the $M = 1$ case it is clear that understanding the attractive M-down equilibrium problem requires computation of the state function as well as the spectrum.

This work will follow closely the model of $M = 1$, which differs from the work of other investigators in its emphasis on the state function. The fundamental assumption will be the assumption of no diffraction, which is an assumption of properties of the state function. The internal consistency of this assumption requires that the free parameters of the stationary state wavefunctions satisfy a certain dynamical algebra. The invariants of this algebra are integrals of the motion, completely determined by the algebra and boundary conditions. These integrals of motion determine the constants of motion. The constants of the motion include the energy, the total momentum and, as will be shown in subsequent work, a paticular marginal probability distribution.

II. DERIVATION OF THE SPECTRAL LAW

Here I derive the spectral law and the state function for the M-down problem. Since an algebraic condition of no diffraction is satisfied (a stronger condition than the Bethe ansatz), I assume—without loss of generality—a state function of a very particular form.

(1s) The state functions of a complete set of stationary states may be expressed as linear combinations of $N!$ plane waves (with differing coefficients) in an N-dimensional state space, partitioned into $N!$ regions that corespond to orderings of the particles along the one spatial dimension.

(2s) The $N!$ different plane waves are permutations of N *distinct* integrals of motion k_i among N particle coordinates.

The delta function dynamics imply an algebraic rule for computing the coefficients of the plane waves in any region, given the coefficients in an adjacent region. I will call this rule D, to be defined explicitly later. The consistency of the assumption of no diffraction is expressed as a condition on the algebra of rule D.

(A) Given the value of the coefficients in any one region it does not matter what sequence of particle permutations are used to carry the state function to any other region—the answer for the state function is the same.

The condition of no diffraction, then, is a condition on the algebra of the delta function dynamics. Any faithful representation of the algebra of that dynamics will satisfy condition (A). Here I choose a particular representation of the state function that satisfies (1s) and (2s) and a particular representation of the dynamical algebra which manifestly satisfies (A). Subsequently I will assure that the dynamical algebra is consistent with rule D.

 0022-2488/90/010164-06$03.00

By virtue of the Pauli principle and conditions (1s) and (2s) above, I may write a state function valid in $M!K!$ contiguous regions of the state space. It is of a Slater determinant form. Let $x_1,...,x_M$ be the coordinates of the spin-down fermions, and $y_1,...,y_K$ be the coordinates of the spin-up fermions. In all regions of the state space such that $x_1,...,x_M < y_1,...,y_K$, the function

$$\Psi_1 = \mathrm{Det}\begin{Bmatrix} \alpha_1 e^{ik_1x_1} & \alpha_2 e^{ik_2x_1} & \cdots & \alpha_N e^{ik_Nx_1} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_1 e^{ik_1x_M} & \alpha_2 e^{ik_2x_M} & \cdots & \alpha_N e^{ik_Nx_M} \\ \beta_1 e^{ik_1y_1} & \beta_2 e^{ik_2y_1} & \cdots & \beta_N e^{ik_Ny_1} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_1 e^{ik_1y_K} & \beta_2 e^{ik_2y_K} & \cdots & \beta_N e^{ik_Ny_K} \end{Bmatrix}, \quad (1)$$

satisfies the Pauli principle and the state function constraints (1s) and (2s); $\Psi_1 = 0$ if any pair of k's are the same, thus all k's must be distinct. Here, Ψ_1 is of the symmetry class appropriate to region 1; $\Psi_1 = 0$ if any two x's or y's are equal. The α's and β's are independent of the x_i, y_i. Their dependence of the k_i must be chosen to satisfy rule D of the delta function dynamics, which intervenes at every boundary of this region.

Again, consistent with condition (A) and the Pauli principle, the form of the solution in an adjacent region is

$$\Psi_1 = \mathrm{Det}\begin{Bmatrix} \alpha_1 e^{ik_1x_1} & \alpha_2 e^{ik_2x_1} & \cdots & \alpha_N e^{ik_Nx_1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_1\alpha_1 e^{ik_1x_M} & \mu_2\alpha_2 e^{ik_2x_M} & \cdots & \mu_N\alpha_N e^{ik_Nx_M} \\ \nu_1\beta_1 e^{ik_1y_1} & \nu_2\beta_2 e^{ik_2y_1} & \cdots & \nu_N\beta_N e^{ik_Ny_1} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_1 e^{ik_1y_K} & \beta_2 e^{ik_2y_K} & \cdots & \beta_N e^{ik_Ny_K} \end{Bmatrix}, \quad (2)$$

where I have taken the adjacent region to be $x_1,...,x_{M-1} < y_1 < x_M < y_2,...,y_K$. Condition (A) of the algebra of no diffraction is satisfied if the algebra of this two-body problem self-consistently determines all other regions. In particular the constraint (A) on the algebra is satisfied if, in the region M, $y_1 < x_1,...,x_M < y_2,...,y_K$, the state function is

$$\Psi_M = \mathrm{Det}\begin{Bmatrix} \mu_1\alpha_1 e^{ik_1x_1} & \mu_2\alpha_2 e^{ik_2x_1} & \cdots & \mu_N\alpha_N e^{ik_Nx_1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_1\alpha_1 e^{ik_1x_M} & \mu_2\alpha_2 e^{ik_2x_M} & \cdots & \mu_N\alpha_N e^{ik_Nx_M} \\ (\nu_1)^M\beta_1 e^{ik_1y_1} & (\nu_2)^M\beta_2 e^{ik_2y_1} & \cdots & (\nu_N)^M\beta_N e^{ik_Ny_1} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_1 e^{ik_1y_K} & \beta_2 e^{ik_2y_K} & \cdots & \beta_N e^{ik_Ny_K} \end{Bmatrix}. \quad (3)$$

The idea here is that each time a particular u passes from right to left through a particular d, the d-particle amplitudes gain one power of μ_i and the u-particle amplitudes gain one power of ν_i. The effect of a u–d passage on the state function is independent of history, and automatically produces a state function of the correct symmetry class. Therefore, this dynamics of u–d passage satisfies condition (A).

At the common boundary of regions 1 and 2, where $x_M = y_1$, delta function potential two-sided boundary conditions that constitute rule D are to be satisfied. I impose these conditions by factoring out the x_M, y_1 dependence from (2). This is achieved by expanding the respective determinants by 2×2 subdeterminant minors along the rows M and $M+1$.

Continuity of the wavefunction gives, for all i, j

$$\alpha_i\beta_j - \alpha_j\beta_i = \mu_i\nu_j\alpha_i\beta_j - \mu_j\alpha_j\beta_i. \quad (4)$$

The derivative discontinuity condition:

the discontinuity of the normal derivative

$$= -g(\text{value of the wavefunction on the boundary}),$$

gives, for all i and j

$$2i(k_i - k_j)(\alpha_i\beta_j + \alpha_j\beta_i - \mu_i\nu_j\alpha_i\beta_j - \mu_j\nu_i\alpha_j\beta_i) = -g(\alpha_i\beta_j - \alpha_j\beta_i + \mu_i\nu_j\alpha_i\beta_j - \mu_j\nu_i\alpha_j\beta_i), \quad (5)$$

where g is the strength constant of the delta function potential (as chosen here $g > 0$ is an attractive interaction). Some manipulation gives the homogeneous equations

$$\begin{Bmatrix} 1 + 1/s - \mu_i\nu_j & -1/s \\ 1/s & 1 - 1/s - \mu_j\nu_i \end{Bmatrix} \begin{Bmatrix} \alpha_i\beta_j \\ \alpha_j\beta_i \end{Bmatrix} = 0, \quad (6)$$

where $s \equiv 2i(k_i - k_j)/g$. Putting the determinant of the system zero gives

$$s = 2i(k_i - k_j)/g = 1/(1 - \mu_i\nu_j) - 1/(1 - \mu_j\nu_i),$$

or, for all i, j

$$\frac{2ik_i}{g} + \frac{1}{(1 - \mu_j\nu_i)} = \frac{2ik_j}{g} + \frac{1}{(1 - \mu_i\nu_j)}. \quad (7)$$

Equation (7) is the algebraic condition for the consistency of rule D and the algebra of the condition of no diffraction.

Consistency with periodic boundary conditions may only be achieved if the regional state function $\Psi_M(y_1)$ falls in the same symmetry class as $\Psi_1(y_1 + L)$, because the consequence of the dynamics is the translation of a u particle from a region where all of the u's are together to a region where all of the u's are together.[9] Thus, from (3),

$$(\nu_i)^M = e^{ik_iL}$$

$$\nu_i = e^{ik_iL/M} e^{2im_i\pi/M}. \quad (8)$$

Periodic boundary conditions dictate, for all regions J

$$\Psi_J(x_1,...,x_M,y_1,...,y_K) = \Psi_J(x_1+L,...,x_M+L,y_1+L,...,y_K+L).$$

From the determinental form of Ψ_J it is easy to show

$$\Psi_J(x_1,+L,...,x_M+L,y_1+L,...,y_K+L) = e^{i\Sigma k_i L}\Psi_J(x_1,...,x_M,y_1,...,y_K),$$

and thus the condition of periodicity is

$$\Sigma k_i L = 2m\pi, \tag{9}$$

for m any integer. Consistency is achieved if I can find μ's and which allow a self-consistent solution of (7)–(9).

The form of the μ's that admit this consistency is not unique. I will not attempt to provide a deductive path to a suitable form. The following choice is reasonable, given the result for $M=1$, and the investigations in Refs. 2–5. It will be demonstrated to admit the desired consistency. Let

$$\mu_s = e^{2im_s\pi/M} \tag{10}$$

where m_s = any integer. There are M distinct values for μ_i, corresponding to m_s modulo M. These values may be thought of as naming distinct families of the μ_i, where the values of μ_i are the same for all values of i, which I call the generic recurrence index. Each of the ν_i also falls into one of the M families. A substitution in (8)

$$\nu_i = e^{ik_iL/M}e^{2im_i\pi/M} = \mu_i e^{ik_iL/M} \tag{11}$$

establishes ν_i as a function of k_i, as well as a particular one of M values of μ, so the ν's differ from the μ's in that their values are different for different values of the generic recurrence index.

I substitute (10) and (11) into (7) and, after some algebra, obtain for all i and j,

$$\frac{2k_i}{g} + \operatorname{ctn}\left(\frac{k_iL}{2M}+\frac{m_i\pi}{M}+\frac{m_j\pi}{M}\right) = \frac{2k_j}{g} + \operatorname{ctn}\left(\frac{k_jL}{2M}+\frac{m_j\pi}{M}+\frac{m_i\pi}{M}\right), \tag{12}$$

the set of transcendental equations to be solved for the k_i. These transcendental equations are a generalization of the single transcendental equation of Refs. 6 and 7, where $M=1$. Every *distinct* set of N k_i which satisfy these equations, and for which the constraint of periodicity (9) is satisfied, are the integrals of motion of a stationary state of the system.

If $4\pi(m_i+m_j)/gL$ is added to both sides of (12) the transcendental equation becomes

$$\frac{2(k_i+2m_i\pi/L+2m_j\pi/L)}{g} + \operatorname{ctn}\left(\frac{k_iL}{2M}+\frac{m_i\pi}{M}+\frac{m_j\pi}{M}\right) = \frac{2(k_j+2m_i\pi/L+2m_j\pi/L)}{g} + \operatorname{ctn}\left(\frac{k_jL}{2M}+\frac{m_j\pi}{M}+\frac{m_i\pi}{M}\right). \tag{13}$$

Let $m_{ij} = m_i + m_j$ modulo M, and

$$z_i = (L/2M)(k_i^{m_{is}} + 2m_{is}\pi/L), \tag{14}$$

where the distinct $k_i^{m_{is}}$ depend upon s and the generic z_i do not. This change of variable substituted in (13) gives

$$\operatorname{ctn} z_i = 4\Lambda(c-z_i), \tag{15}$$

where $\Lambda = M/gL$ and the constant c is independent of i. This is the generic form of the transcendental equation, as opposed to the distinct form (12). The z's are the generic integrals of the motion. There are N generic integrals of the motion and N_D distinct generic integrals of the motion where $1 \leqslant N_D \leqslant N$. Note that the generic and distinct transcendental equations are identical for $M=1$, as in Refs. 6 and 7.

The connection between the generiz z's and the distinct k's is, from (14)

$$k_i^{m_{is}} = (2/L)(Mz_i - m_{is}\pi), \tag{16}$$

where m_{is} = any integer modulo M, except, to keep k's distinct, $m_{is} \neq m_{it}$ if $s \neq t$.

III. STATEMENT OF THE SPECTRAL LAW

The spectral law for the determination of the distinct integrals of motion, comprised of (9), (15), and (16), is now internally consistent and complete. Let the index i label a branch of the cotangent in the generic transcendental equation. Let the positive integer n_i, $0 \leqslant n_i \leqslant M$, be the occupancy number of the generic branch i. It follows that

$$\sum_i n_i = N. \tag{17}$$

The graphical methods of Refs. 6 and 7 are adequate for explicit visualization of the solutions to the generic transcendental equation (15) for a fixed value of c. The solutions to (15), $z_i(c)$, are continuous monotonic functions of c. Substituting (16) into (9)

$$2\sum_i\sum_s (Mz_i(c) - m_{is}\pi) = 2m\pi,$$

which gives the generic constraint

$$\sum_i n_i z_i(c) = \frac{n\pi}{M}, \tag{18}$$

for n an integer. Each value of c for which this relation is satisfied is an allowed value of c, for which the generic integrals of motion are $z_i(c)$.

IV. THE CONSTANTS OF THE MOTION, STATE FUNCTION, AND NORMALIZATION

The distinct integrals of motion are determined from (16). They depend upon the generic integrals of motion as well as the N quantum numbers m_{is}. The constants of the motion depend upon the distinct integrals of the motion. The total momentum and energy are given by

$$k(c) = \frac{2}{L}\sum_i\sum_{s=0}^{n_i-1}(Mz_i(c) - m_{is}\pi),$$

$$E(c) = \frac{2}{L^2}\sum_i\sum_{s=0}^{n_i-1}(Mz_i(c) - m_{is}\pi)^2. \tag{19}$$

As in the $M=1$ case, the spectral constant c is to be eliminated in favor of k in the expression for the energy. This gives rise to an energy band whose energy is $E_b(k)$, where the

band quantum number is determined by the quantum numbers n_i and m_{is}.

The state function is given by the spatial union of the state functions in the various regions. To make this state function explicit I specify the α_i's and β_i's in region 1. The equation of continuity (4) gives

$$\alpha_i/\beta_i(1-\mu_j\nu_i) = \text{const, independent of } i,j.$$

For simplicity, let the constant $=1$, and substitute the values for the μ and ν

$$\alpha_i = \beta_i(1-e^{ik_iL/M}e^{2i\pi m_{ij}/M}).$$

To be completely explicit I make another arbitrary choice, that $\beta_i = 1$ for all i. This corresponds to assigning a particular amplitude and phase to the state function in region 1. With this choice the complete state function in region 1 is given by specifying the α's in region 1

$$\alpha_i = (1-e^{2iz_i}) = -2ie^{iz_i}\sin z_i. \tag{20}$$

Note that these amplitudes are functions only of the generic integrals of motion.

The state function in region 1 is the determinant of an $N\times N$ matrix whose form is

$$\Psi_1 = \mathrm{Det}\begin{Bmatrix} \alpha_i e^{ik_ix_1} \\ \cdots \\ \cdots \\ e^{ik_jy_1} \\ \cdots \\ \cdots \end{Bmatrix}, \tag{21}$$

where it is understood that the α_i are generic and the k_i are distinct. Now consider the state function in an adjacent region, where the d particle labeled by x_1 and the u particle labeled by y_1 interchange position. Expand the determinant (21) by 2×2 subdeterminant minors along the two rows shown, to factor the x_1,y_1 dependence. According to the self-consistent version of rule D, each 2×2 subminor is transformed by

$$\mathrm{Det}\begin{Bmatrix} \alpha_i e^{ik_ix_1} & \alpha_j e^{ik_jx_1} \\ e^{ik_jy_1} & e^{ik_jy_1} \end{Bmatrix}$$

$$\to \mathrm{Det}\begin{Bmatrix} \mu_i\alpha_i e^{ik_ix_1} & \mu_j\alpha_j e^{ik_jx_1} \\ \mu_i e^{ik_i(y_1+L/M)} & \mu_j e^{ik_j(y_1+L/M)} \end{Bmatrix}. \tag{22}$$

Since this is all of the x_1,y_1 dependence, and the μ's are unimodular, the effect on $\Psi^*\Psi$ is

$$\Psi_2{}^*\Psi_2(x_1,y_1) = \Psi_1{}^*\Psi_1(x_1,y_1+L/M), \tag{23}$$

that is $\Psi^*\Psi$ in any region is generated by a discrete translation of $\Psi^*\Psi$ in any other region.

The complete stationary state wavefunction is the spatial union of the regional state functions, which I write

$$\Psi = \{\cup\Psi_J\}.$$

This notation means that the various regions of state space are labeled with an index J, and the total state function is the nonanalytic function Ψ_1 in region 1, Ψ_2 in region 2, etc. In the same notation

$$\Psi^*\Psi = \{\cup\Psi_J{}^*\Psi_J\}.$$

From (23), $\Psi_J{}^*\Psi_J$ may be written as a set of finite translations of $\Psi_1{}^*\Psi_1$:

$$\Psi_J{}^*\Psi_J(X,y_1,y_2,\ldots,y_K) = \Psi_1{}^*\Psi_1\left(X,y_1+n_1(J)\frac{L}{M}, y_2+n_2(J)\frac{L}{M},\ldots,y_K+n_K(J)\frac{L}{M}\right), \tag{24}$$

where the $n_i(J)$ are integers which have the meaning that they are the number of u steps into the d sea required to connect region 1 to region J. Since $\Psi^*\Psi$ is composed of the union of all $\Psi_J{}^*\Psi_J$, integrals of $\Psi^*\Psi$ may be computed as integrals of a single regional state function over volume L. For example, the unnormalized bivariate u–d probability density in a stationary state is

$$\rho(z,\zeta,x,y) = \int_0^L dX\cdots\int_0^L dY\,\Psi_1(x,y,X,Y)\Psi_1^*(x,y,X,Y), \tag{25}$$

where x is a particular d coordinate, y is a particular u coordinate, and X,Y are, respectively, the set of complementary d and u coordinates. The procedure for evaluating these integrals as functions of the integrals of motion closely follows Refs. 6 and 7 and is outlined in the Appendix.

The normalized u–d relative coordinate marginal probability distribution in a stationary state is

$$\omega(z,r) = 1+\left[\frac{1}{N}-\frac{Q^\dagger Q}{n}\right]\frac{1}{D},$$

where

$$n(z) = \frac{1}{N}\sum_1^N \frac{\alpha_i\alpha_i^\dagger}{(1-\Lambda\alpha_i\alpha_i^\dagger)}$$

$$Q(r) = \frac{1}{N}\sum_1^N \frac{\alpha_i e^{-ik_ir}}{(1-\Lambda\alpha_i\alpha_i^\dagger)} \tag{26}$$

$$D = 1-1/N+\Lambda n.$$

V. CONCLUDING REMARKS

The computation of the integrals of motion is remarkably similar to $M=1$. The generic transcendental equation and constraint of periodicity are identical in form to the corresponding equations for $M=1$. The computation of the distinct integrals from the generic is very simple. It is only necessary to identify with each generic integral a set of equally spaced momentum states. The effect of this seemingly trivial generalization of the spectral law is profound (particularly in the attractive case), and will be dealt with elsewhere.

I do not continue here with the important application of the bulk limit $N=2M$, which motivated this work, because that is only a single application of the ideas expressed here. Further applications include the possibility of removing some of the artificial barriers in other solvable model problems (e.g., negative coupling in the one dimensional Hubbard and Heisenberg models), and extension of this algebraic approach to other quantum fields.

APPENDIX: INTEGRALS OVER REGIONAL STATE FUNCTIONS

Consider integrals of the form

$$I(z,\zeta,x,y) = \int_0^L dX \cdots \int_0^L dY\, \Delta(x,y,X,Y)\Delta^\dagger(x,y,X,Y), \qquad \text{(A1)}$$

where Δ is the determinental form of the state function in region 1:

$$\Delta(\Delta,X,Y) = \mathrm{Det}\begin{Bmatrix} \alpha(z_i)e^{ik_i x_1} \\ \cdots \\ \cdots \\ e^{ik_i y_1} \\ \cdots \\ \cdots \end{Bmatrix} = \mathrm{Det}\begin{Bmatrix} \alpha_i e^{ik_i x_1} \\ \cdots \\ \cdots \\ e^{ik_i y_1} \\ \cdots \\ \cdots \end{Bmatrix}.$$

Let $z_i^* = \zeta_i$, and correspondingly $k_i^* = \kappa_i$, so that the adjoint determinant is written

$$\Delta^\dagger(z,X,Y) = \mathrm{Det}\begin{Bmatrix} \alpha^*(\zeta_i)e^{-i\kappa_i x_1} \\ \cdots \\ \cdots \\ e^{i\kappa_i y_1} \\ \cdots \\ \cdots \end{Bmatrix} = \mathrm{Det}\begin{Bmatrix} \alpha_i^\dagger e^{-i\kappa_i x_1} \\ \cdots \\ \cdots \\ e^{i\kappa_i y_1} \\ \cdots \\ \cdots \end{Bmatrix}.$$

Note the integral identity:

$$\int_{-L/2}^{L/2} dX\, e^{i(k_{-1}-\kappa_j)x} = L\left[\delta_{ij}(1-\Lambda\alpha_i\alpha_i^\dagger) + \frac{4\sin z_i \sin\zeta_i \sin(k_i-\kappa_j)L/2}{gL\sin(k_i-\kappa_j)L/2M}\right]. \qquad \text{(A2)}$$

where I have chosen $k_i = \kappa_j \Leftrightarrow i = j$. This identity is established by doing the integral

$$\int_{-L/2}^{L/2} dx\, e^{i(k_i-\kappa_j)x} = L\begin{cases} \dfrac{2\sin(L/2)(k_i-\kappa_j)}{L(k_i-\kappa_j)}, & k_i \neq \kappa_j, \\ 1, & k_i = \kappa_j \end{cases}$$

and noting that

$$\frac{2}{g}\sin z_i \sin\zeta_j = \frac{\alpha_i\alpha_j^\dagger}{2g} e^{-i(z_i-\zeta_j)} = \frac{\sin(L/2M)(k_i-\kappa_j)}{(k_i-\kappa_j)}$$

is an alternative form of the transcendental equation (12). The second term of (A2) does not contribute in the bulk limit; if $k_i L$ and $\kappa_j L$ are distinct and separated by an integral multiple of 2π it vanishes identically otherwise I shall always be concerned with the bulk limit $gL \to \infty$ where it is negligible; if $k_i L \equiv \kappa_j L$, the corresponding 2×2 determinant vanishes identically.

The integration over y_1 removes all dependence on y_1. Thus

$$I(z,\zeta,X,Y') = \sum_i (1-\Lambda\alpha_i\alpha_i^\dagger)\Delta^\dagger\Delta(z',X,Y'),$$

where $\Delta(z',X,Y')$ is the determinant of an $N-1$ by $N-1$ matrix. Here, z' means leave out all of the z's in the sums, Y' means leave out all of the y's in the integrals. Continue the process until all but one of the y's are integrated out:

$$I(z,\zeta,X,y) = \sum_i \cdots \sum_j \prod_t (1-\Lambda\alpha_t\alpha_t^\dagger)\Delta^\dagger\Delta(z',X,Y'),$$

where Δ is the determinant of an $M+1$ by $M+1$ matrix:

$$\Delta(z,X,y) = \mathrm{Det}\begin{Bmatrix} \alpha(z_i)e^{ik_i x_1} \\ \cdots \\ \cdots \\ e^{ik_i y} \end{Bmatrix}.$$

Now, integrate the x dependences by the same method—expand by minors to factor x_i dependence of the integrals using the integral identity. The only difference is that the typical term is multiplied by $\alpha_i\alpha_i^\dagger$. Integrate over x's until only one x remains:

$$I(z,\zeta,x,y) = \sum_i \sum_j \sum_{N-2} F_{ij} \prod_{N-2} (1-\Lambda\alpha_t\alpha_t^\dagger)\Delta^\dagger\Delta,$$

where

$$F_{ij} = \sum \prod_{t\neq i,j}^{M-1} \alpha_t\alpha_t^\dagger,$$

meaning sum over all possible ways to choose a product of $M-1$ factors $\alpha_t\alpha_t^\dagger$, no index repeated, index $\neq i$ or j. In Refs. 6 and 7 it is observed that $\alpha_t\alpha_t^\dagger$, is the weight, or relative probability, that any given d particle is identified with the integral of the motion z_t. Here, F_{ij} is the weight, or relative probability, that $M-1$ balls (d particles) put in N distinguishable urns (indices) such that no more than one ball is in any urn and no ball is in urn i or j. The events "no ball in i" and "no ball in j" are independent, and the weight of their conjunction is the product of their individual weights:

$$F_{ij} = F_i F_j,$$

and

$$F_i = 1 - \left[\alpha_i\alpha_i^\dagger \Big/ \sum_{i=1}^N \alpha_i\alpha_i^\dagger\right]^{M-1}.$$

Thus the i dependence of F_i is of order $(1/N)^M$, and therefore negligible.

Completing the product over the $N-2$ other factors the result is

$$I(z,\zeta,x,y) = \sum_i \sum_j \frac{\Delta(x,y)\Delta^\dagger(x,y)}{(1-\Lambda\alpha_i\alpha_i^\dagger)(1-\Lambda\alpha_j\alpha_j^\dagger)},$$

where

$$\Delta = \mathrm{Det}\begin{Bmatrix} \alpha_i e^{ik_i x} & \alpha_j e^{ik_j x} \\ e^{ik_i y} & e^{ik_j y} \end{Bmatrix}.$$

Expanding the determinants

$$DD^\dagger = \alpha_i\alpha_i^\dagger + \alpha_j\alpha_j^\dagger - \alpha_i\alpha_j^\dagger e^{i(k_j-\kappa_i)r} - \alpha_j\alpha_i^\dagger e^{i(k_i-\kappa_j)r},$$

where $r = x - y$, the u–d relative coordinate. Substituting and rearranging terms

$$I = n(1 + \Lambda n) - Q^{\dagger}Q(r),$$

where the functions n and Q are

$$n(z) = \frac{1}{N}\sum_{1}^{N} \frac{\alpha_i \alpha_i^{\dagger}}{(1 - \Lambda \alpha_i \alpha_i^{\dagger})},$$

$$Q(r) = \frac{1}{N}\sum_{1}^{N} \frac{\alpha_i e^{-ik_i r}}{(1 - \Lambda \alpha_i \alpha_i^{\dagger})}.$$

The normalization integral may be computed using the integral identity (A2):

$$\int_0^L Q^{\dagger}Q\, dr - \frac{nL}{N}.$$

In the bulk limit, therefore, the normalized u–d marginal probability distribution is

$$\omega(z,r) = 1 + \left[\frac{1}{N} - \frac{Q^{\dagger}Q}{n}\right]\frac{1}{D},$$

where $D = 1 - 1/N + \Lambda n$.

[1]M. Gaudin, *La Fonction d'Onde de Bethe* (Masson, Paris, 1983).
[2]M. Flicker and E. Lieb, Phys. Rev. **161**, 179 (1967).
[3]M. Gaudin, "Etude d'un modele a une dimension pour un systeme de fermions en interaction," Ph.D. thesis, Univ. of Paris (1967), CEA Rep. 5-3569 Saclay (1968).
[4]C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[5]B. Sutherland, Phys. Rev. Lett. **20**, 98 (1968).
[6]J. B. McGuire, J. Math. Phys. **6**, 432 (1965); *ibid.* **7**, 123 (1966).
[7]J. B. McGuire, Phys. Rev. A **15**(1) (1990).
[8]J. B. McGuire (unpublished).

[9]Notice that, for this choice of state function, it is too restrictive to require equality of the two regional state functions.

On the zeta-function regularization of a two-dimensional series of Epstein–Hurwitz type

E. Elizalde
Department of Structure and Constituents of Matter, Faculty of Physics, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

(Received 27 September 1988; accepted for publication 2 August 1989)

As a further step in the general program of zeta-function regularization of multiseries expressions, some original formulas are provided for the analytic continuation, to any value of s, of two-dimensional series of Epstein–Hurwitz type, namely, $\Sigma_{n_1,n_2=0}^{\infty}[a_1(n_1+c_1)^2+a_2(n_2+c_2)^2]^{-s}$, where the a_j are positive reals and the c_j are not simultaneously nonpositive integers. They come out from a generalization to Hurwitz functions of the zeta-function regularization theorem of the author and Romeo [Phys. Rev. D **40**, 436 (1989)] for ordinary zeta functions. For $s=-k$,0,2, with $k=1,2,3,...$, the final results are, in fact, expressed in terms of Hurwitz zeta functions only. For general s they also involve Bessel functions. A partial numerical investigation of the different terms of the exact, algebraic equations is also carried out. As a by-product, the series $\Sigma_{n=0}^{\infty}\exp[-a(n+c)^2]$, $a,c>0$, is conveniently calculated in terms of them.

I. INTRODUCTION

For a few years now, the study of quantum field theories in partially compactified space-time manifolds has acquired increasing importance in several domains of quantum physics. Let me just mention the issues of dimensional reduction and spontaneous compactification, and the multiple questions associated with the study of quantum field theories in the presence of boundaries (like the Casimir effect) and on curved space-time (manifolds with curvature and nontrivial topology), a step towards quantum gravity.

There are many interesting calculations in these theories that can be carried out exactly—and in a very elegant way from the mathematical point of view—by the zeta-function regularization method. In particular, if all the eigenvalues of the Hamiltonian are known, then, very commonly, one is led in this method to the computation of expressions of the general form

$$\sum_{n_1,\ldots,n_N=0}^{\infty}\left[\sum_{j=1}^{N}a_j(n_j+c_j)^{\alpha_j}\right]^{-s},\quad a_j,\alpha_j>0. \tag{1}$$

As such a multiseries, this expression only makes sense for $\mathrm{Re}(s)$ big enough, and an analytic (usually meromorphic) continuation to other values of s is in order. In the zeta-function method, this is provided by the Riemann and Hurwitz (also called Riemann generalized) zeta functions.

However, for an expression as general as (1) this program has proved to be extremely difficult (not to say impossible until now) to carry out. The simplest case is obtained when (1) corresponds to the Hamiltonian zeta function

$$\zeta(s)\equiv\sum_i E_i^{-s} \tag{2}$$

(E_i are the eigenvalues of H) of a system of N noninteracting harmonic oscillators. In this case, $a_j=1$, $j=1,2,...,N$, and the a_j are the eigenfrequencies ω_j.[1] Another important case shows up in the partial toroidal compactification (space-time $T^p\times\mathbb{R}^{q+1}$). Then $\alpha_j=2$ and, usually, $c_j=0,\ \pm\frac{1}{2}$ (Ref. 2). This leads typically to Epstein zeta functions

$$\begin{aligned}Z_N(s)&=\sum_{n_1,\ldots,n_N=-\infty}^{\infty}(n_1^2+\cdots+n_N^2)^{-s},\\ Y_N(s)&=\sum_{n_1,\ldots,n_N=-\infty}^{\infty}{}'\\ &\quad\times\left[\left(n_1+\frac{1}{2}\right)^2+\cdots+\left(n_N+\frac{1}{2}\right)^2\right]^{-s},\end{aligned} \tag{3}$$

(the prime prescribes omission of the term with $n_1=n_2=\cdots=n_N=0$). Other powers α_j appear when one deals with the spherical compactification (space-time $S^p\times\mathbb{R}^{q+1}$). Moreover, as string theory seems to indicate, nothing precludes the possibility of having to consider other compactification manifolds, leading to very general values for the α_j. In this work, however, we shall only deal with the particular case $\alpha_j=2$, $j=1,2,...,N$, leaving more general situations for subsequent study.

The aim of the paper is to derive some new and useful expressions for the analytic continuation of two-dimensional sums of the types just mentioned. My results will come from a rigorous generalization of the zeta-function regularization theorem,[1,3,4] which is carried out in Sec. II, Eq. (7), by obtaining the appropriate counterterm (9). From it, basic expressions for zeta-functions regularization—Eqs. (22), (30), and (32) of Secs. III, IV, and V, respectively—will follow. They will give rise to the general equation (34) of Sec. V, which provides the analytical extension to any complex value of s of two-dimensional sums of the type mentioned in the Abstract, and also to the interesting particular formulas (35)–(38). Finally, in Sec. VI a recurrent procedure to extend these expressions to arbitrary-N multiseries as (1) will be sketched [Eq. (39)].

II. THE CASE $\alpha_j=2$: STATEMENT OF THE MATHEMATICAL PROBLEM

The apparently simple case $\alpha_j=2$ carries enough complication that it deserves a complete study on its own. On the

 0022-2488/90/010170-05$03.00

other hand, at least formally, the general case is actually very similar to this one (the main difference being the transition from the cases $\alpha_j < 2$ to the cases $\alpha_j \geqslant 2$, as will be explained later. Thus I shall restrict myself to the expressions

$$E_N(s;a_1,\ldots,a_N;c_1,\ldots,c_N) \equiv \sum_{n_1,\ldots,n_N=0}^{\infty} \left[\sum_{j=1}^{N} a_j(n_j+c_j)^2\right]^{-s}, \tag{4}$$

where it is understood that all $a_j > 0$ and that not all of the c_j are nonpositive integers. Actually, only the particular situation with $N = 2$ will be worked out in detail. Let me emphasize the fact of the presence in (4) of general nonzero a_j's and c_j's. The only precedents in the literature (to my knowledge) of this kind of evaluations are restricted to very few special cases other than $a_1 = a_2 = \cdots = a_N$ and $c_1 = c_2 = \cdots = c_N = 1$.[2,5] Maybe the most famous expression in this context is the celebrated result of Hardy,[6] which can be obtained as a particular case of our final formulas in Ref. 4.

In Ref. 4, together with Romeo we began an investigation of the general expression (4), limiting ourselves to the simplest case $c_j = 1, j = 1,2,\ldots,N$. It is not that immediate to extend the results there to the present situation, as we shall see.

A basic point in the zeta-function regularization procedure is the interchange of the order of the summations of infinite series in expressions like

$$S_c^{(\alpha)}(s) \equiv \sum_{m=0}^{\infty} (m+c)^{-s-1} \sum_{a=0}^{\infty} \frac{(-1)^a}{a!} (m+c)^{\alpha a}. \tag{5}$$

In the case $c = 1$ and $\alpha < 2$, the correct additional contribution coming from this commutation of sums was obtained by Weldon.[3] Actually, he claimed that his result was valid for any $\alpha \in \mathbb{N}$. This has turned out to be not right, as rigorously shown in Ref. 4 where the correct supplementary contributions for $\alpha \geqslant 2$ (always with $c = 1$) have been obtained.

III. THE FUNDAMENTAL FORMULA FOR ZETA-FUNCTION REGULARIZATION WHEN $\alpha_j = 2$

I shall now proceed with the calculation of (5). It can be written as

$$S_c^{(\alpha)}(s) = \sum_{m=0}^{\infty} (m+c)^{-s-1} \oint_C \frac{da}{2\pi i} (m+c)^{-\alpha a} \Gamma(a), \tag{6}$$

where C is the contour ($C = L + K$) consisting of the straight line (L), $\mathrm{Re}(a) = a_0$, $0 < a_0 < 1$, and of a curved part (K), which is the semicircumference at infinity on the left of this line. For $\mathrm{Re}(s)$ big enough, we obtain

$$S_c^{(\alpha)}(s) = \sum_{a=0}^{\infty} \frac{(-1)^a}{a!} \zeta(s+1-\alpha a,c) + \begin{cases} \dfrac{1}{\alpha}\Gamma\left(-\dfrac{s}{\alpha}\right) + \Delta_c^{(\alpha)}(s), & \dfrac{s}{\alpha} \notin \mathbb{N}, \\ (-1)^{s/\alpha}\left[\dfrac{\gamma}{\Gamma(s/\alpha+1)} - \dfrac{1}{\gamma\Gamma'(s/\alpha+1)}\right] + \Delta_c^{(\alpha)}(s), & \dfrac{s}{\alpha} \in \mathbb{N}, \end{cases} \tag{7}$$

where $\zeta(z,c)$ is Hurwitz's (or Riemann's generalized) zeta function

$$\zeta(z,c) = \sum_{n=0}^{\infty} (n+c)^{-s} \tag{8}$$

and $\Delta_c^{(\alpha)}(s)$ is the following integral over the curved part K of the contour C:

$$\Delta_c^{(\alpha)}(s) \equiv \int_K \frac{da}{2\pi i} \zeta(s+1+\alpha a,c)\Gamma(a). \tag{9}$$

The preceding expressions, Eqs. (7) and (9), constitute the more basic result in this paper. They can be viewed as a generalization of the zeta-function regularization theorem obtained in Ref. 4. There the case of the ordinary Riemann zeta function (i.e., $c = 1$) was studied and a detailed discussion on the nature of the term (9) for $c = 1$ (including numerical computations for different values of s) was provided. It turns out that, for arbitrary positive c, the present term (9) can be related to the one in Ref. 4; in fact, it is numerically comprised between two expressions both obtained from the case $c = 1$ by suppressing a finite number of contributions, namely, the first $[c-1]$ and $[c]$, respectively (here square brackets mean integer part). As the notation (i.e., the *delta*) already suggests, this term (9) always turns out to be a correction to the first, leading terms. It is also clear from the above discussion that the most interesting new case with respect to the one dealt with in Ref. 4 appears now when $0 < c < 1$, and this is precisely the specific situation that I will consider below.

In order to be able to provide an expression for the integral (9) in terms of more elementary functions, I shall restrict myself to the case $\alpha = 2$. Use will be made of the well-known Hurwitz formula,[7] valid (in particular) for $\mathrm{Re}\, z < 0$ and $0 < c \leqslant 1$,

$$\zeta(z,c) = 2(2\pi)^{z-1}\Gamma(z-1)\sum_{n=1}^{\infty} n^{z-1}\sin\left(2\pi nc + \frac{\pi z}{2}\right). \tag{10}$$

The behavior of the lhs for $|z| \to \infty$ with $\mathrm{Re}\, z < 0$ is

$$\zeta(z,c) \sim 2(2\pi)^{z-1}\Gamma(z-1)\sin(2\pi c + \pi z/2), \tag{11}$$

while, for $c = 1$, we obtain

$$\zeta(z) = \zeta(z,1) \sim 2(2\pi)^{z-1}\Gamma(z-1)\sin(\pi z/2). \tag{12}$$

From the last two expressions, we get, for $0 < c \leqslant 1$,

$$\lim_{\substack{|z|\to\infty \\ \mathrm{Re}(z)<0}} \frac{\zeta(z,c)}{\zeta(z)} = \sin(2\pi c)\cot\left(\frac{\pi z}{2}\right) + \cos(2\pi c). \tag{13}$$

Now, making use of the identity, valid also for $\mathrm{Re}(z)<0$ (Ref. 7),

$$\Gamma\left(\frac{1-z}{2}\right)\zeta(1-z)=\int_0^\infty dt\, t^{-(z+1)/2}S(t), \tag{14}$$

with

$$S(t)\equiv\sum_{n=1}^{\infty}e^{-n^2t}, \tag{15}$$

we obtain for the analytic continuation of (9) (with $\alpha=2$) to $s=-1$:

$$\begin{aligned}\Delta_c^{(2)}(-1)&=\int_K\frac{da}{2\pi i}\zeta(2a,c)\Gamma(a)\\&=\int_K\frac{da}{2\pi i}[\sin(2\pi c)\cot(\pi a)\\&\quad+\cos(2\pi c)]\zeta(2a)\Gamma(a).\end{aligned} \tag{16}$$

After making use of Eq. (14), the second term on the rhs can be integrated immediately. Writing it now at the beginning of the second member, we get

$$\begin{aligned}\Delta_c^{(2)}(-1)&=-\sqrt{\pi}S(\pi^2)\cos(2\pi c)\\&\quad+\sin(2\pi c)\int_K\frac{da}{2\pi i}\int_0^\infty dt\\&\quad\times t^{-a-1/2}\cot(\pi a)S(\pi^2t).\end{aligned} \tag{17}$$

The last integral in (17) turns out to be zero. In fact,

$$\begin{aligned}&\int_K da\, it^{-a}\frac{e^{i\pi a}+e^{-i\pi a}}{e^{i\pi a}-e^{-i\pi a}}\\&=\int_{\pi/2}^{\pi}d\theta\,\mathrm{Re}^{i\theta}t^{-R(\cos\theta+i\sin\theta)}\\&\quad-\int_{\pi}^{3\pi/2}d\theta\,\mathrm{Re}^{i\theta}t^{-R(\cos\theta+i\sin\theta)}\\&=-i\frac{t^{iR}+t^{-iR}}{\mathrm{Ln}\,t}=-2i\frac{\cos(R\,\mathrm{Ln}\,t)}{\mathrm{Ln}\,t},\end{aligned} \tag{18}$$

from which it follows that

$$\begin{aligned}&\lim_{R\to\infty}\int_0^\infty dt\,S(\pi^2t)\frac{\cos(R\,\mathrm{Ln}\,t)}{\mathrm{Ln}\,t}\\&=\mathrm{Re}\left[\int_{-\infty}^{\infty}du\,S(\pi^2e^{u/R})\frac{e^{iu}}{u}\right]=0.\end{aligned} \tag{19}$$

We are left with just

$$\Delta_c^{(2)}(-1)=-\sqrt{\pi}\cos(2\pi c)S(\pi^2). \tag{20}$$

Summing up, I have proved that

$$S_c\equiv S_c^{(2)}(-1)=\sum_{m=0}^{\infty}e^{-(m+c)^2} \tag{21}$$

can be expressed in terms of Hurwitz zeta functions, as

$$\begin{aligned}S_c&=\sum_{m=0}^{\infty}\frac{(-1)^m}{m!}\zeta(-2m,c)\\&\quad+\frac{\sqrt{\pi}}{2}+\sqrt{\pi}\cos(2\pi c)S(\pi^2),\end{aligned} \tag{22}$$

with (the standard, related to Jacobi's theta function) $S(t)$ being given by Eq. (15). Equation (22) is another meaningful result of this paper. It is *exact* and holds for *any* value of c.

IV. BEHAVIOR OF THE ZETA-REGULARIZED FUNDAMENTAL SERIES

Let me now investigate the behavior of the different series in (22). Depending on the value of c, the series of Hurwitz functions can be convergent (even finite) or asymptotic. The other two series, i.e., those implicit in the definition of the S functions, are quickly convergent (the one on the rhs much more quickly than the one on the lhs). In fact, to be clearer, let us check some specific cases.

(i) In the particular case $c=1$ we recover the known equality[4,7]

$$S(1)=(\sqrt{\pi}-1)/2+\sqrt{\pi}S(\pi^2). \tag{23}$$

(ii) For $c=\frac{1}{2}$ we have $\zeta(-2m,\frac{1}{2})=0$, $k=0,1,2,\ldots$, and

$$\begin{aligned}&\sum_{m=0}^{\infty}\exp\left[-\left(m+\frac{1}{2}\right)^2\right]\\&=\frac{\sqrt{\pi}}{2}-\sqrt{\pi}\sum_{m=1}^{\infty}\exp(-m^2\pi^2).\end{aligned} \tag{24}$$

The rhs of (24) permits us to obtain the value of the series on the lhs with 10^{-10} accuracy, *with just two terms*

$$\begin{aligned}&\sum_{m=0}^{\infty}\exp\left[-\left(m+\frac{1}{2}\right)^2\right]\\&=\frac{\sqrt{\pi}}{2}-\sqrt{\pi}e^{-\pi2}+\mathcal{O}(10^{-10}).\end{aligned} \tag{25}$$

(iii) For $c=0$ we get an equality equivalent to (23),

$$\sum_{m=0}^{\infty}e^{-m^2}=\frac{1}{2}+\frac{\sqrt{\pi}}{2}+\sqrt{\pi}S(\pi^2). \tag{26}$$

Actually, it is an immediate consequence of the properties of the series in (22) that the equalities one obtains for $c+1$ and for $c-1$ are each equivalent to the corresponding one for c. Therefore, only the equalities (22) corresponding to c, $0<c\leqslant1$, provide interesting (independent) relations.

(iv) For $c=\frac{1}{4}$, we get

$$\begin{aligned}&\sum_{m=0}^{\infty}\exp\left[-\left(m+\frac{1}{4}\right)^2\right]\\&=\frac{\sqrt{\pi}}{2}+\sum_{m=0}^{\infty}\frac{(-1)^m}{m!}\zeta\left(-2m,\frac{1}{4}\right).\end{aligned} \tag{27}$$

The series of Hurwitz functions on the rhs is now asymptotic. It stabilizes between the eighth and the twelfth summands and it provides a best value (with $\simeq10^{-7}$ accuracy) exactly when we add its ten first terms.

(v) For $c=\frac{1}{3}$ and $c=\frac{1}{6}$ we obtain, respectively,

$$\begin{aligned}&\sum_{m=0}^{\infty}\exp\left[-\left(m+\frac{1}{3j}\right)^2\right]\\&=\frac{\sqrt{\pi}}{2}+\sum_{m=0}^{\infty}\frac{(-1)^m}{m!}\zeta\left(-2m,\frac{1}{3j}\right)\\&\quad+(-1)^j\frac{\sqrt{\pi}}{2}\sum_{m=1}^{\infty}\exp(-m^2\pi^2),\quad j=1,2.\end{aligned} \tag{28}$$

In these cases, contributions from the two series in the rhs must be taken into account. The first of them is asymptotic [as in (iv)] and has exactly *the same* characteristics as the one in (27), both for $j=1,2$. The second series is extremely

rapidly convergent (much more than the series on the lhs). These characteristics are maintained over the full range $0<c\leqslant 1$ (but for the very special values $c=\frac{1}{2},1$ considered above).

One may ask what is gained with these asymptotic expressions. The answer has already been given before, in Eqs. (7) and (9), which extend these equalities by analytic continuation to *any* value of s, and not simply to the case $s=-1$ exemplified here. I will be more precise in what follows. Before being so, let me present two more examples of interesting, original relations that come from Eq. (22):

$$\sum_{m=-\infty}^{\infty}\exp[-(m+c)^2]=\sqrt{\pi}+2\sqrt{\pi}\cos(2\pi c)S(\pi^2), \tag{29}$$

$$\begin{aligned}\sum_{m=0}^{\infty} m\exp[-(m+c)^2] &= \frac{1}{2}+\sum_{m=0}^{\infty}\frac{(-1)^m}{m!}\\ &\times[\zeta(-2m-1,c)-c\zeta(-2m,c)]-\frac{\sqrt{\pi}}{2}c\\ &+\sqrt{\pi}[\pi\sin(2\pi c)-c\cos(2\pi c)]S(\pi^2).\end{aligned} \tag{30}$$

V. THE GENERAL EXPRESSION FOR $N=2$

The calculation of the general expression (4) will be now illustrated, for the sake of clarity, in the simpler case $N=2$. By using the Mellin transform, we write

$$\begin{aligned}&E_2(s;a_1,a_2;c_1,c_2)\\ &=\sum_{n_1,n_2=0}^{\infty}[a_1(n_1+c_1)^2+a_2(n_2+c_2)^2]^{-s}\\ &=\frac{1}{\Gamma(s)}\sum_{n_1,n_2=0}^{\infty}\int_0^{\infty}dt\,t^{s-1}\\ &\quad\times\exp\{-t[a_1(n_1+c_1)^2+a_2(n_2+c_2)^2]\}.\end{aligned} \tag{31}$$

We shall need the following generalization of Eq. (22)—again a basic outcome of my regularization theorem (7)–(9) and obtained in the same way—

$$\begin{aligned}\sum_{m=0}^{\infty}\exp[-a(m+c)^2] &= \sum_{m=0}^{\infty}\frac{(-1)^m}{m!}a^m\zeta(-2m,c)\\ &+\frac{1}{2}\sqrt{\frac{\pi}{a}}+\sqrt{\frac{\pi}{a}}\cos(2\pi c)S\left(\frac{\pi^2}{a^2}\right).\end{aligned} \tag{32}$$

Substituting (32) into (31), we get

$$\begin{aligned}&E_2(s;a_1,a_2;c_1,c_2)\\ &=\frac{1}{\Gamma(s)}\sum_{m=0}^{\infty}\frac{(-1)^m}{m!}a_1^m\zeta(-2m,c_1)\sum_{n_2=0}^{\infty}\int_0^{\infty}dt\,t^{s+m-1}\exp[-ta_2(n_2+c_2)^2]\\ &+\frac{1}{2}\sqrt{\frac{\pi}{a_1}}\frac{1}{\Gamma(s)}\sum_{n_2=0}^{\infty}\int_0^{\infty}dt\,t^{s-3/2}\exp[-ta_2(n_2+c_2)^2]\\ &+\sqrt{\frac{\pi}{a_1}}\frac{\cos(2\pi c_1)}{\Gamma(s)}\sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}\int_0^{\infty}dt\,t^{s-3/2}\exp\left[-\frac{\pi^2n_1^2}{a_1t}-ta_2(n_2+c_2)^2\right].\end{aligned} \tag{33}$$

This gives

$$\begin{aligned}&E_2(s;a_1,a_2;c_1,c_2)\\ &=\frac{a_2^{-s}}{\Gamma(s)}\sum_{m=0}^{\infty}\frac{(-1)^m\Gamma(s+m)}{m!}\left(\frac{a_1}{a_2}\right)^m\zeta(-2m,c_1)\zeta(2s+2m,c_2)+\frac{a_2^{-2}}{2}\left(\frac{\pi a_1}{a_2}\right)^{1/2}\frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)}\zeta(2s-1,c_2)\\ &+\frac{2\pi^s}{\Gamma(s)}\cos(2\pi c_1)a_1^{-s/2-1/4}a_2^{-s/4+1/4}\sum_{n_1=1}^{\infty}\sum_{n_2=0}^{\infty}n_1^{s-1/2}(n_2+c_2)^{-s+1/2}K_{s-1/2}\left[2\pi\sqrt{\frac{a_2}{a_1}}n_1(n_2+c_2)\right],\end{aligned} \tag{34}$$

where K_ν is the modified Bessel function of the second kind. Equation (34) constitutes the general analytic continuation formula for two-dimensional series I was looking for. As is apparent, it involves Bessel functions as well as Hurwitz functions. However, the following particular cases look especially simple.

For $s=-k$, $k=0,1,2,\ldots$, one obtains

$$\begin{aligned}&E_2(-k;a_1,a_2;c_1,c_2)\\ &=\frac{a_2^k}{\Gamma(-k)}\sum_{m=0}^{k}\frac{(-1)^m\Gamma(m-k)}{m!}\left(\frac{a_1}{a_2}\right)^m\zeta(-2m,c_1)\zeta(2(m-k),c_2)\\ &=a_2^k\left(\frac{1}{2}-c_1\right)\zeta(-2k,c_2)+a_2^k\sum_{m=1}^{\infty}\frac{k(k-1)\cdots(k-m+1)}{m!}\left(\frac{a_1}{a_2}\right)^m\zeta(-2m,c_1)\zeta(2(m-k),c_2).\end{aligned} \tag{35}$$

In particular, for $s=0$,

$$E_2(0;a_1,a_2;c_1,c_2)=(c_1-\tfrac{1}{2})(c_2-\tfrac{1}{2}), \tag{36}$$

and, for $s=-1$,

$$E_2(-1;a_1,a_2;c_1,c_2)$$
$$=a_2(\tfrac{1}{2}-c_1)\zeta(-2,c_2)+a_1(\tfrac{1}{2}-c_2)\zeta(-2,c_1)$$
$$=\tfrac{1}{3}(c_1-\tfrac{1}{2})(c_2-\tfrac{1}{2})[a_1c_1(1-c_1)+a_2c_2(1-c_2)]. \tag{37}$$

For $s=2$, we obtain

$$E_2(2;a_1,a_2;c_1,c_2)$$
$$=\frac{1}{a_2^2}\sum_{m=0}^{\infty}(-1)^m(m+1)\left(\frac{a_1}{a_2}\right)^m\zeta(-2m,c_1)\zeta(2m+4,c_2)+\frac{\pi}{4a_2}\frac{1}{\sqrt{a_1a_2}}\zeta(3,c_2)$$
$$+\frac{\pi^2\cos(2\pi c_1)}{a_1a_2}\sum_{m=0}^{\infty}\left\{(n+c_2)^{-2}\left[\exp\left(2\pi\sqrt{\frac{a_2}{a_1}}(n+c_2)\right)-1\right]^{-2}\right.$$
$$\left.+\left[(n+c_2)^{-2}+\sqrt{\frac{a_1}{a_2}}\frac{(n+c_2)^{-3}}{2\pi}\right]\left[\exp\left(2\pi\sqrt{\frac{a_2}{a_1}}(n+c_2)\right)-1\right]^{-1}\right\}. \tag{38}$$

The first two terms on the rhs yield properly the result of the zeta-function regularization method (naive commutation of the series summations plus Weldon's additional contribution[3]). They produce the desired expression of (4) in terms of zeta functions. The last term in (38) generalizes to arbitrary $c_1,c_2>0$ the supplementary corrections detected in Ref. 4 for $c_1=c_2=1$ and which had been loosely forgotten in Ref. 3. In spite of the imposing aspect of this last term, its contribution is actually very small, and the series in n is very quickly convergent (only the first couple of summands need to be taken into account in practice). For an arbitrary value of s, one must use the general expression (34).

VI. A GENERAL EXPRESSION FOR ARBITRARY *N*

The preceding calculations can be generalized to multiple sums (4) with arbitary N. The fundamental formula (32) introduced into the Mellin transform [as in (31)–(33)] allows us to proceed recurrently. One obtains the (exact) equation

$$E_N(s;a_1,\ldots,a_N;c_1,\ldots,c_N)$$
$$=\frac{1}{\Gamma(s)}\sum_{m=0}^{\infty}\frac{(-1)^m}{m!}a_1^m\zeta(-2m,c_1)\Gamma(s+m)E_{N-1}(s+m;a_2,\ldots,a_N;c_2,\ldots,c_N)$$
$$+\frac{1}{2}\sqrt{\frac{\pi}{a_1}}\frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)}E_{N-1}\left(s-\frac{1}{2};a_2,\ldots,a_N;c_2,\ldots,c_N\right)$$
$$+\sqrt{\frac{\pi}{a_1}}\frac{\cos(2\pi c_1)}{\Gamma(s)}\sum_{n_1=1}^{\infty}\sum_{n_2,\ldots,n_N=0}^{\infty}\int_0^{\infty}dt\,t^{s-3/2}\exp\left[-\frac{\pi^2n_1^2}{a_1t}-t\sum_{j=2}^{N}a_j(n_j+c_j)^2\right]. \tag{39}$$

Notice, once more, that the last term is a small correction to the first two, so that, in practice Eq. (39) can be viewed as a recursive formula with a small correction term Δ (the last one) that can be estimated numerically. This is also discussed in Ref. 4 (for the particular case $c_1=\cdots=c_N=1$) in greater detail.

The application of the formulas derived in this paper to the direct evaluation (exact, or at worst, six to seven decimal places precise) of the Casimir effect, by just summing over modes (provided that they are known exactly) and by zeta-regularizing the resulting expressions, will be developed in a separate publication.

ACKNOWLEDGMENTS

The constructive comments of the referee of a first version of this paper are gratefully acknowledged.

This work was partially supported by Comisión Asesora de Investigación Científica y Técnica (CAICYT, Spain), research project AE 87-0016-3.

[1] A. Actor, J. Phys. A **20**, 927 (1987).
[2] J. Ambjørn and S. Wolfram, Ann. Phys. (N.Y.) **147**, 1 (1983); A. Actor Class. Quantum Grav. **5**, 1415 (1988).
[3] H. A. Weldon, Nucl. Phys. B **270**, (FS16), 79 (1986).
[4] E. Elizalde and A. Romeo, Phys. Rev. D **40**, 436 (1989).
[5] I. J. Zucker, J. Phys. A **7**, 1568 (1974); **8**, 1734 (1975); I. J. Zucker and M. M. Robertson, *ibid.* **8**, 874 (1975).
[6] G. H. Hardy, Mess. Math. **49**, 85 (1919).
[7] *Higher Transcendental Functions, Vol. I*, edited by A. Erdélyi (McGraw-Hill, New York, 1953); *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).

Basic equations of three-dimensional radiative transfer

Alan P. Wang
Department of Mathematics, Arizona State University, Tempe, Arizona 85287

(Received 3 March 1989; accepted for publication 6 September 1989)

Chandrasekhar developed one-dimensional mathematical models of radiative transfer in 1949 [*Radiative Transfer* (Oxford U.P., New York, 1950)]. This paper is a systematic extension of Chandrasekhar's work to three dimensions, including discussions of specular and diffused parts, reciprocity, solutions, and approximation.

I. INTRODUCTION

Lord Rayleigh,[1] in 1899, investigated the problem of specifying the radiative field in a light scattering atmosphere. In 1905, Schuster[2] explained the absorption and emission lines in stellar sectra and, in 1906, Schwarzschild[3] introduced the concept of radiative equilibrium in stellar atmospheres. Until 1949, the subject of radiative transfer had been principally investigated astrophysically rather than mathematically.

Then Chandrasekhar published *Radiative Transfer*.[4] He systematically presented radiative transfer as a branch of mathematical physics. Developing the general "principle of invariance," he constructed a complete one-dimensional mathematical model for radiative transfer in plane-parallel atmospheres. Since then there have been many interesting developments in this area.

Ueno[5] and Bellman and Kalaba[6] described their method of "invariant imbedding." At the same time, 1956, Redheffer developed the closely related transmission-line approach.[7] Other important contributions were made by Preisendorfer,[8] Paszkowski,[9] Wing,[10] Ueno and Wang,[11] Wang,[12] and many others. The mathematical models associated with the above were all based on one-dimensional time-independent radiative transfer.

Recently, investigators have been forced to study three-dimensional radiative transfer: for example, the study of radiative transfer in a free non-plane-parallel atmosphere, or in a plane-parallel atmosphere with nonuniform ground reflection. Models in these three-dimensional cases have immediate applications in Earth image processing and target identification.[13–16]

The purpose of this paper is to construct, in a compact and geometric way, a complete set of mathematical models for three-dimensional radiative transfer using integral operators and scattering theory.[7] With our interpretation of such operators, results may be applied to three-dimensional problems in neutron transport, probability and circuit theory, etc. A following paper extends three-dimensional radiative transfer to a time-dependent case.

Symbols used in this paper can be found in Table I.

II. BASIC EQUATIONS

Splitting the radiative intensity at an incremental element of a transfer medium into input and output, we find its equation of state, from which we derive, in linear operator form, the fundamental equation governing the variation of intensity in the medium. The intensity $I(p,u)$ is separated into two parts, I^+ and I^-, i.e., $I^+(p,u) = I(p;+u)$, with $0<\theta<\pi/2$, and $I^-(p,u) = I(p;-u)$, with $\pi/2<\theta<\pi$. As $I^+(p,u)$ and $I^-(p,u)$ travels from p to $p' = p + \Delta p$ (see Fig. 1), there are differences in energies arising from the trans-

TABLE I. Symbols used in this paper.

$p=(x,y,z)$	the position, i.e., a point in three-dimensional Euclidean space
$u=(\mu,\phi)$	a directional vector; $\mu=\cos\theta$, θ is the polar angle, and ϕ is the azimuthal angle
$\alpha(p)$	the position-dependent attenuation coefficient
$\sigma(p)$	the position-dependent scattering coefficient
$I(p,u)$	radiant intensity at position p in the direction u and in an element of solid angle
$f(p,u,u')$	the position-dependent phase function for an angle between u and u'
$\tilde{t},\tilde{\tau}$	transmission operator, see Sec. II
$\tilde{\rho},\tilde{r}$	reflection operator, see Sec. II
S	scattering matrix, see (2.4)
T,Π	specular part of transmission operator
R,P	diffused part of reflection operator
t,τ	diffused part of transmission operator
r,ρ	diffused part of reflection operator
A,B,C,D	generators associated with specular part of operators, see (2.5)
a,b,c,d	generators associated with diffuse part of operators, see (2.5)
K	ground reflection operator
*	superscript *, see (3.16)
Q	overall reflection operator, see (4.1)
R_m	see (4.1)
$\|\ \|$, $\|\|\ \|\|$	norm and sup norm, see Sec. V.

 0022-2488/90/010175-07$03.00

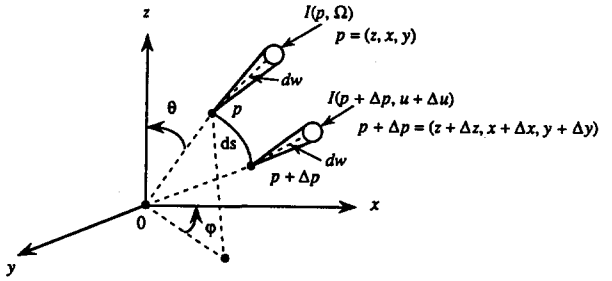


FIG. 1. Radiation fields at points p and $p+\Delta p$.

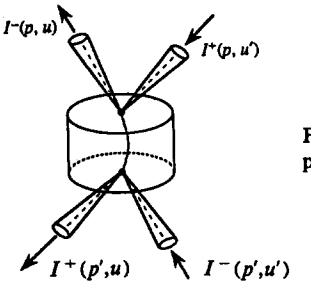


FIG. 2. Reflection and transmission of pencils of radiation.

mission and reflection in the frequency interval and element of solid angle considered. Denoting transmission by $\tilde{t}$ and $\tilde{\tau}$ and reflection by $\tilde{r}$ and $\tilde{\rho}$, their relations are given by

$$\begin{pmatrix} I^+(p',u) \\ I^-(p,u) \end{pmatrix} = \begin{pmatrix} \tilde{t}(p,p',u,u') & \tilde{\rho}(p,p',u,u') \\ \tilde{r}(p,p',u,u') & \tilde{\tau}(p,p',u,u') \end{pmatrix} \begin{pmatrix} I^+(p,u') \\ I^-(p',u') \end{pmatrix}, \tag{2.1}$$

that is, the input–output pairs in direction u' are mapped to the input–output pairs in direction u (see Fig. 2), where, using $\tilde{t}$ to denote the operator as well as its kernel,

$$\tilde{t}(p,p',u,u')\cdot I^+(p,u') = \int \tilde{t}(p,p,u,u')I^+(p,u')du',$$

likewise for the other operators, $\tilde{\rho}$, $\tilde{r}$, and $\tilde{\tau}$. We denote the generators associated with $\tilde{t}$, $\tilde{\tau}$, $\tilde{\rho}$, and $\tilde{r}$, respectively, by $\tilde{b}$, $\tilde{d}$, $\tilde{a}$, and $\tilde{c}$. Then taking the limit as $\Delta p \to 0$, we obtain the *equation of state*

$$\frac{1}{v}\frac{d}{ds}\begin{pmatrix} I^+(p,u) \\ I^-(p,u) \end{pmatrix} = \begin{pmatrix} \tilde{b}(p,u,u') & \tilde{a}(p,u,u') \\ -\tilde{c}(p,u,u') & -\tilde{d}(p,u,u') \end{pmatrix} \cdot \begin{pmatrix} I^+(p,u') \\ I^-(p,u') \end{pmatrix}, \tag{2.2}$$

where d/ds is the directional derivative in the direction of R,

$$\frac{d}{ds} I^{\pm}(p,u) = \nabla I^{\pm}(p,u)\cdot R. \tag{2.3}$$

The advantage of the present formulation is that the apparently preferred variable, s, in (2.2) can be adjusted to the application, e.g., Sec. III, without the transverse or radial symmetry assumption of one-dimensional radiative transfer. If z is the preferred variable, we have transverse or plane-parallel symmetry and (2.3) becomes

$$\frac{d}{ds} I^+(p,u) = \frac{d}{dz} I^+(z,u).$$

Similarly, if the medium is radially symmetric, the preferred variable is $r = (x^2+y^2+z^2)^{1/2}$ with an equation similar to (2.3), cf. Wang.[17] Thus previous treatments are restricted cases of the present treatments.

It is known,[4] in the treatment of radiative transfer, that it is convenient to distinguish between the reduced or specular radiation, which penetrates to a point in a medium without being scattered, and the diffuse radiation field, which has been multiply scattered. The appropriate representation for a scattering matrix

$$S(p,p',u,u') = \begin{pmatrix} \tilde{t} & \tilde{\rho} \\ \tilde{r} & \tilde{\tau} \end{pmatrix}$$

is a linear combination of its *specular* and the *diffused parts*

$$S(p,p',u,u') = \begin{pmatrix} T(u')\delta(u-u') & P(u')\delta(u+u') \\ R(u')\delta(u+u') & \Pi(u')\delta(u-u') \end{pmatrix} + \begin{pmatrix} t(u,u') & \rho(u,u') \\ r(u,u') & \tau(u,u') \end{pmatrix}, \tag{2.4}$$

where δ is the Dirac distribution and the dependence on p and p' are understood.

The associated generator also allows for both specular and diffuse parts,

$$\begin{pmatrix} \tilde{b}(u,u') & \tilde{a}(u,u') \\ \tilde{c}(u,u') & \tilde{d}(u,u') \end{pmatrix} = \begin{pmatrix} B(u')\delta(u-u') & A(u')\delta(u+u') \\ C(u')\delta(u+u') & D(u')\delta(u-u') \end{pmatrix} + \begin{pmatrix} b(u,u') & a(u,u') \\ c(u,u') & d(u,u') \end{pmatrix}, \tag{2.5}$$

suppressing the variable p for convenience.

Set $I^+(p,u') = 0$ in (2.1); then

$$\frac{d}{ds}\tilde{\rho}(u,u') = B(u)\tilde{\rho}(u,u') + \tilde{\rho}(u,u')D(u) + A(u')\delta(u+u') + a(u,u') + \int b(u,\tilde{u})\tilde{\rho}(\tilde{u},u')d\tilde{u} + \int \tilde{\rho}(u,\tilde{u})d(\tilde{u},u')d\tilde{u} + \int \tilde{\rho}(u,\tilde{u})C(-\tilde{u})\tilde{\rho}(-\tilde{u},u)d\tilde{u} + \int\int \tilde{\rho}(u,\tilde{u})c(\tilde{u},\tilde{\tilde{u}})\tilde{\rho}(\tilde{\tilde{u}},u')d\tilde{u}\,d\tilde{\tilde{u}}. \tag{2.6}$$

Substituting

$$\tilde{\rho}(u,u') = P(u')\delta(u+u') + \rho(u,u'),$$

as in (2.4), into Eq. (2.6) and grouping all terms involving $\delta(u+u')$, we obtain the equation containing only the *specular part* of the *reflection operator* $\tilde{\rho}$,

$$\frac{d}{ds}P(u) = A(u) + B(-u)P(u) + P(u)D(u) + P(u)C(-u)P(u). \quad (2.7)$$

The remaining part of (2.6), i.e., the *diffused part*, is represented as

$$\frac{d}{ds}\rho(u,u') = \bar{a}(u,u') + \bar{B}(u)\rho(u,u') + \rho(u,u') + \int \bar{b}(u,\tilde{u})\rho(\tilde{u},u')d\tilde{u} + \int \rho(u,\tilde{u})\bar{d}(\tilde{u},u')d\tilde{u} + \int\int \rho(u,\tilde{u})\bar{c}(\tilde{u},\tilde{\tilde{u}})\rho(\tilde{\tilde{u}},u')d\tilde{u}\,d\tilde{\tilde{u}}, \quad (2.8)$$

where the coefficients are defined as follows:

$$\bar{a}(u,u') = a(u,u') + b(u,-u')P(u') + P(u)d(-u,u') + P(-u)c(-u,-u')P(u'),$$
$$\bar{B}(u) = B(u) + P(-u)C(u),$$
$$\bar{D}(u') = D(u') + C(-u')P(u'),$$
$$\bar{b}(u,\tilde{u}) = b(u,\tilde{u}) + P(-u)c(-u,\tilde{u}),$$
$$\bar{d}(\tilde{u},u') = d(\tilde{u},u') + c(\tilde{u},-u')P(u'),$$
$$\bar{c}(\tilde{u},\tilde{\tilde{u}}) = c(\tilde{u},\tilde{\tilde{u}}) + C(\tilde{\tilde{u}})\delta(\tilde{u}+\tilde{\tilde{u}}).$$

The specular part [Eq. (2.7)] is a one-dimensional one while the diffuse part [(2.8)] is three dimensional and involves the specular part.

Considering $I^-(p',u') = 0$ in (2.1), taking the directional derivative, and separating the $\delta(u-u')$ terms, we have the *specular part* of the transmission,

$$\frac{d}{ds}T(u) = [B(u) + P(u)C(-u)]T(u), \quad (2.9)$$

while the *diffused part* is

$$\frac{d}{ds}t(u,u') = \bar{\bar{b}}(u,u') + \bar{B}(u)t(u,u') + \int \bar{b}(u,\tilde{u})t(\tilde{u},u')d\tilde{u} + \int\int \rho(u,\tilde{u})\bar{c}(\tilde{u},\tilde{\tilde{u}})t(\tilde{\tilde{u}},u')d\tilde{u}\,d\tilde{\tilde{u}}, \quad (2.10)$$

where the new coefficient

$$\bar{\bar{b}}(u,u') = \bar{b}(u,u')T(u') + \int \rho(u,\tilde{u})\bar{c}(\tilde{u},u')d\tilde{u}\,T(u').$$

Similarly, the specular part of $\tilde{r}$ satisfies the equation

$$\frac{d}{ds}R(u) = \Pi(u)C(u)T(u), \quad (2.11)$$

and the diffuse part of $\tilde{r}$ satisfies

$$\frac{d}{ds}r(u,u') = \bar{\bar{c}}(u,u') + \int \tau(u,\tilde{u})\bar{c}(\tilde{u},u')T(u')d\tilde{u} + \int\int \tau(u,\tilde{u})\bar{c}(\tilde{u},\tilde{\tilde{u}})t(\tilde{\tilde{u}},u')d\tilde{u}\,d\tilde{\tilde{u}}, \quad (2.12)$$

with

$$\bar{\bar{c}}(u,u') = \Pi(u)c(u,u')T(u') + \Pi(u)\int \bar{c}(u,\tilde{u})t(\tilde{u},u)d\tilde{u}.$$

Finally, the specular part and the diffuse part of the transmission $\tilde{\tau}$ are

$$\frac{d}{ds}\Pi(u) = \Pi(u)[D(u) + C(-u)P(u)] \quad (2.13)$$

and

$$\frac{d}{ds}\tau(u,u') = \bar{\bar{d}}(u,u') + \tau(u,u')\bar{D}(u') + \int \tau(u,\tilde{u})\bar{d}(\tilde{u},u')d\tilde{u} + \int\int \tau(u,\tilde{u})\bar{c}(u,\tilde{\tilde{u}})\rho(\tilde{\tilde{u}},u')d\tilde{u}\,d\tilde{\tilde{u}}, \quad (2.14)$$

with

$$\bar{\bar{d}}(u,u') = \Pi(u)\bar{d}(u,u') + \Pi(u)\int \bar{c}(u,\tilde{u})\rho(\tilde{u},u')d\tilde{u}.$$

With the basic equations completed we note that the derivation is closely related to Redheffer's transmission line theory, and the concept to Chandrasekhar's principle of invariance. Also, the three-dimensional model is motivated by Ueno. In the case where the operator ρ is independent of (x,y), Eqs. (2.7) and (2.8) reduce to those obtained by Redheffer. When $|p-p'| \to \infty$, so that

$$S(p,p',\Omega,\Omega') \to S_\infty(\Omega,\Omega')$$

and

$$\frac{d}{ds}S(p,p',\Omega,\Omega') = 0,$$

then the above differential-integral equations (2.6)–(2.14) all become integral equations. Solutions of such equations are of great interest in radiative transfer; indeed, they are related to the "law of darkening" in astrophysics.

III. RADIATIVE TRANSFER

In this section we consider the three-dimensional radiative transfer model consisting of an atmosphere extended from optical thickness $z=0$ to $z=z_1$. At the top $z=z_1$, it is uniformly and monodirectionally illuminated by parallel rays of solar radiation of constant net flux, πF, per unit area normal to the incident direction. At the bottom $z=0$, it is bounded by a flat reflecting surface. The upward intensity of radiation emergent in the direction from the level z, $0 \leqslant z \leqslant z_1$, at the horizontal rectangular coordinates (x,y), is $I^-(p,u)$ and the downward intensity is $I^+(p,u)$. In this case,

$$\frac{d}{ds}I^{\pm}(p,u) = \left(\frac{\partial I^{\pm}}{\partial z} + \tan\theta\sin\phi\frac{\partial I^{\pm}}{\partial x} + \tan\theta\cos\phi\frac{\partial I^{\pm}}{\partial y}\right), \quad (3.1)$$

since $R = (1, \tan\theta \sin\phi, \tan\theta \cos\phi)$ with $u = (\theta,\phi)$.

Letting $p \to p'$ as in Fig. 2, the integral equation generators $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, and $\tilde{d}$ of the previous section are determined by the physical properties of a thin increment of the medium in this case. It can be shown[11] that

$$\begin{pmatrix} b & a \\ c & d \end{pmatrix} = \frac{\sigma(z)}{4\pi}\begin{pmatrix} f(z,u,u') & f(z,u,-u') \\ f(z,-u,u') & f(z,-u,-u') \end{pmatrix}, \quad (3.2)$$

where $\sigma(z)$ is the position-dependent scattering coefficient and $f(z,u,u')$ is the position-dependent phase function for the angle between u and u'. It is also well known[14] that specular radiation, in this case, occurs only in transmission; therefore $A = C = 0$. Thus the equation for P in (2.7) reduces to the linear equation

$$\frac{d}{ds}P(u) = B(-u)P(u) + P(u)D(u). \quad (3.3)$$

Because there is no specular reflection radiation when the optical thickness is zero, the proper initial condition in (3.3) is zero. Hence the unique solution of (3.3) for radiative transfer is

$$P(u) = 0. \quad (3.4)$$

It follows [(2.8)–(2.10)] that $\bar{a}$, $\bar{b}$, $\bar{c}$, $\bar{d}$, $\bar{B}$, and $\bar{D}$ are a, b, c, d, B, and D, respectively. In radiative transfer the reduced or specular transmission radiation field is equal to $e^{-\alpha(z)/\mu}$, the amount of incident flux at level z. With $u = (\mu,\phi)$, this yields the values

$$B(u) = -\alpha(z)/\mu \quad \text{and} \quad D(u') = -\alpha(z)/\mu. \quad (3.5)$$

Combining (3.4) and (3.5) with the remarks following (3.4), Eq. (2.8) has the integral form

$$\left[\frac{d}{ds} + \alpha(p)\left(\frac{1}{\mu} + \frac{1}{\mu'}\right)\right]\rho(u,u') = \sigma(z)\left[f(p,u,-u') + \frac{1}{4\pi}\int \rho(u,\tilde{u})\, f(p,-\tilde{u},-u')\,\frac{d\tilde{u}}{\tilde{\mu}} + \frac{1}{4\pi}\int f(p,u,\tilde{\tilde{u}})\rho(\tilde{\tilde{u}},u')\frac{d\tilde{\tilde{u}}}{\tilde{\mu}} + \frac{1}{16\pi^2}\int\int \rho(u,\underline{u})f(p,-\tilde{\tilde{u}},\tilde{u})\rho(\tilde{u},u')\,\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\,\frac{d\tilde{u}}{\tilde{\mu}}\right], \quad (3.6)$$

with initial condition

$$\rho(u,u') = 0, \quad \text{for } p = (0,x,y). \quad (3.7)$$

Equation (3.6) is identical to the result of Ueno and Wang.[11]

Applying the above conditions to Eq. (2.9), we have

$$\frac{d}{ds}T(u) = -\frac{\alpha(p)}{\mu}T(u), \quad (3.8)$$

with initial condition $T(u) = \delta(u - u')$, where δ is the Dirac distribution. The unique solution for (3.8) is

$$T(u) = \exp(-\alpha(p)/\mu). \quad (3.9)$$

The corresponding equation (2.10) for radiative transfer now has the form

$$\left[\frac{d}{ds} + \frac{\alpha(p)}{\mu'}\right]t(u,u') = \lambda(p)\left[\exp\left(-\frac{\alpha(p)}{\mu}\right)f(p,u,u') + \frac{1}{4\pi}\int\left\{\exp\left(-\frac{\alpha(p)}{\mu}\right)\rho(u,\tilde{u})f(p,-\tilde{u},u') + f(p,\tilde{u},u')t(\tilde{u},u')\right\}\frac{d\tilde{u}}{\tilde{\mu}} + \frac{1}{16\pi^2}\int\int \rho(u,\tilde{\tilde{u}})f(p,-\tilde{\tilde{u}},\tilde{u})t(\tilde{u},u')\,\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\,\frac{d\tilde{u}}{\tilde{\mu}}\right], \quad (3.10)$$

with initial condition

$$t(u,u') = 0, \quad \text{when } p = (0,x,y). \quad (3.11)$$

Assuming independence of ρ, α, and λ on (x,y), the atmosphere is homogeneous in the (x,y) but not in the z direction, with anisotropic scattering in a "free space," i.e., $K = 0$, or in a "reflective space," with ground reflection uniform in (x,y). In this case, Eqs. (3.6) and (3.9) reduce to the results of Ueno[5,18] and Busbridge.[19]

As for Eqs. (2.13) and (2.14), it is observed that $D = B$ and $A = C = 0$. We have

$$\Pi(u) = T(u) \quad (3.12)$$

and

$$\left[\frac{d}{ds} + \frac{\alpha(p)}{\mu}\right]\tau(u,u') = \lambda(p)\left[\exp\left(-\frac{\alpha(p)}{\mu'}\right)f(p,u,u') + \frac{1}{4\pi}\int\left\{\exp\left(-\frac{\alpha(p)}{\mu'}\right)f(p,-u,\tilde{u})\rho(p,\tilde{u},u') + \tau(u,\tilde{u})f(p,-\tilde{u},-u')\right\}\frac{d\tilde{u}}{\tilde{\mu}} + \frac{1}{16\pi^2}\int\int \tau(p,u,\tilde{\tilde{u}})f(p,-\tilde{\tilde{u}},\tilde{u})\rho(p,\tilde{u},u')\,\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\,\frac{d\tilde{u}}{\tilde{\mu}}\right], \quad (3.13)$$

with initial condition

$$\tau(p,u,u') = 0, \quad \text{when } p = (0,x,y). \quad (3.14)$$

For Eq. (2.11), $R(u) = 0$, since $C(u) = 0$. Equation (2.12) for radiative transfer reduces to

$$\left[\frac{d}{ds}+\alpha(p)\left(\frac{1}{\mu}+\frac{1}{\mu'}\right)\right]r(p,u,u')$$
$$=\exp\left(-\frac{\alpha(p)}{\mu}\right)\frac{1}{4\pi}\int\tau(u,\tilde{u})f(p,-\tilde{u},u')\frac{d\tilde{u}}{\tilde{\mu}}$$
$$+\exp\left(-\frac{\alpha(p)}{\mu}\right)\frac{1}{4\pi}\int f(p,-u,\tilde{\tilde{u}})t(\tilde{\tilde{u}},u')\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}+\frac{1}{16\pi^2}\int\int\tau(u,\tilde{\tilde{u}})f(p,-\tilde{\tilde{u}},\tilde{u})t(\tilde{u},u')\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\frac{d\tilde{u}}{\tilde{\mu}},\tag{3.15}$$

also with the initial condition

$$r(u,u')=0,\quad \text{when } p=(0,x,y).$$

IV. RECIPROCITY RELATIONS

Suppose the phase function depends only on the cosine of the angle between u and u', i.e.,

$$f(z,u,u')=f(z,\cos\psi),$$

where

$$\cos\psi=\mu\mu'+(1-\mu^2)^{1/2}(1-\mu')^{1/2}\cos(\phi'-\phi).$$

Thus

$$f^*(z,u,u')=f(z,u,u'),\quad f^*(z,u,-u')=f(z,u',-u),\tag{4.1}$$

where f^* is obtained from f by transposing u and u',

$$f^*(z,u,u')=f(z,u',u).\tag{4.2}$$

Let ρ be a solution of (3.6); then interchanging u and u', and using Eq. (4.1),

$$\left[\frac{d}{ds}+\alpha(p)\left(\frac{1}{\mu}+\frac{1}{\mu'}\right)\right]\rho^*(u,u')$$
$$=\sigma(z)\left[f(p,u',-u)+\frac{1}{4\pi}\int\rho(u',\tilde{u})f(p,-\tilde{u},-u)\frac{d\tilde{u}}{\tilde{\mu}}+\frac{1}{4\pi}\int f(p,u',\tilde{\tilde{u}})\rho(\tilde{\tilde{u}},u)\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\right.$$
$$\left.+\frac{1}{16\pi^2}\int\int\rho(u',\tilde{\tilde{u}})f(p,-\tilde{\tilde{u}},\tilde{u})\rho(\tilde{u},u)\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\frac{d\tilde{u}}{\tilde{\mu}}\right]$$
$$=\sigma(z)\left[f(p,u,-u')+\frac{1}{4\pi}\int f(p,u,\tilde{u})\rho(\tilde{u},\mu')\frac{d\tilde{u}}{\tilde{\mu}}+\frac{1}{4\pi}\int\rho^*(u,\tilde{\tilde{u}})(u,\tilde{\tilde{u}})\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}+\frac{1}{16\pi^2}\int\int\rho^*(u,\tilde{u})\right.$$
$$\left.\times f(p,-\tilde{u},-\tilde{\tilde{u}})\rho^*(\tilde{\tilde{u}},u')\frac{d\tilde{\tilde{u}}}{\tilde{\tilde{\mu}}}\frac{d\tilde{u}}{\tilde{\mu}}\right].$$

This Riccati equation has a unique solution[17,20] thus

$$\rho(u,u')=\rho^*(u,u').\tag{4.3}$$

Likewise, by (3.10), (3.13), (3.15), and (4.1), it can be shown that

$$t(u,u')=\tau^*(u,u')\tag{4.4}$$

and

$$(u,u')=r^*(u,u'),\tag{4.5}$$

the *reciprocity relations* for three-dimensional radiative transfer.

Conversely, we assume that the reciprocity relations (4.3)–(4.5) hold. Then

$$\frac{d}{ds}t^*(u,u')=\frac{d}{ds}\tau(u,u').$$

Using (3.2) and letting $z\to(0,x,y)$, we have

$$f(z,u,u')=f(z,-u,-u').\tag{4.6}$$

Similarly, since

$$\frac{d}{ds}r(u,u')=\frac{d}{ds}r^*(u,u')$$

and

$$\frac{d}{ds}\rho(u,u')=\frac{d}{ds}\rho^*(u,u'),$$

we obtain

$$f(z,u,u')=f(z,u',u)\tag{4.7}$$

and

$$f(z,u,-u')=f(z,-u,u').\tag{4.8}$$

Equations (4.6)–(4.8) imply (4.1). Thus Eq. (4.1) holds if and only if the reciprocity relations are true.

As for isotropic scattering,

$$t(u,u')=t^*(u,u'),\tag{4.9}$$

and the above reciprocity relations reduces to the *Holmboltz Principle.*

V. SOLUTIONS AND APPROXIMATION

The basic problem, associated with the physical model in Fig. 3, is to determine the upward radiation field $I^+(z,u)$ at the top of the atmosphere in the direction u due to the solar or other incident radiation, $I_i(z,-u')$, in the direction u'. Let

$$I^+(z,u) = Q(u,u')I_i(z,u'), \tag{5.1}$$

where the linear operator Q is called the *overall reflection* operator, because it consists of the reflection by the atmospheric layer and the multiple reflection between atmosphere and the ground. Letting $\tilde{\rho}$ and K denote the upward reflection operators of the atmosphere and ground, respectively, it is known[7,12] that

$$Q_m = \rho + \tilde{t}R_m(K)\tilde{\tau}, \tag{5.2}$$

with

$$R_m(K) = K\sum_{n=0}^{m}(rK)^n,$$

where $\tilde{r}$ is the downward reflection operator of the atmosphere, and

$$Q(u,u') = \lim_{m\to\infty} Q_m(u,u'). \tag{5.3}$$

Here, uniform convergence is assumed. It is clear from the definition that Q_m approximates the overall reflection by taking the mth order of multiple scattering between the atmosphere and ground. Then Q_m is called mth order overall scattering or the mth-order approximation of Q.

If Eqs. (3.6), (3.10), (3.13), and (3.15) can be solved for ρ, t, τ, and r, then Eq. (5.3) has the form

$$Q_m = \rho + \exp(-\alpha(p)/\mu)[2R_m + \tilde{t}R_m + R_m\tilde{\tau}] + \tilde{t}R_m\tilde{\tau}, \tag{5.4}$$

where, by (3.12), $\tilde{t} = t + T$, $\tilde{\tau} = \tilde{\tau} + \Pi$, and $T = \Pi$, with

$$p = (z_1,x,y).$$

The right side of Eq. (5.4) has the following physical interpretation: first term, the diffuse reflection due to the atmosphere only, i.e., $K = 0$; second term, combinations of specular and diffusion due to the existence of the reflector K; and third term, pure diffusion reflection due to the existence of the ground reflector K.

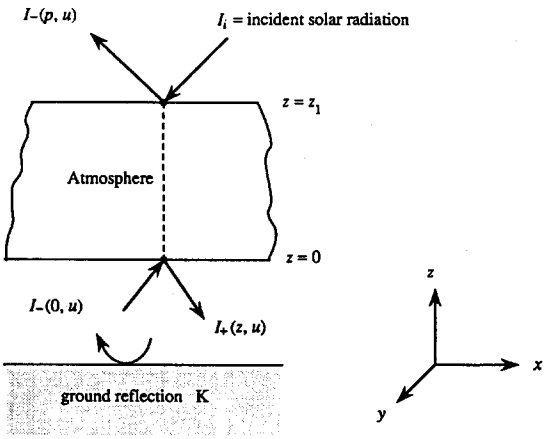


FIG. 3. A three-dimensional model.

Thus Q_m captures the intuitive physical notion of mth-order approximation. Furthermore, Q_m can be easily "updated" from m to $m+1$. Indeed,

$$Q_{m+1} = Q_m + \Delta Q_m, \tag{5.5}$$

where

$$\Delta Q_m = \exp(-\alpha(p)/\mu)[2K(rK)^{m+1} + \tilde{t}K(rK)^{m+1} + K(rK)^{m+1}\tilde{\tau}] + \tilde{t}K(rK)^{m+1}\tilde{\tau}. \tag{5.6}$$

We recall that operators in (5.6) are in integral form. For convergence, the norm

$$|f(p,u,u')I(u')| = \int\int f(p,u,u')I(u')\,du\,du'$$

and the sup norm

$$\|f\| = \sup_p |f(p,u,u')I(u')|$$

are used. Then

$$\|\Delta Q_m\| \leqslant \|rK\|^{m+1}\|K\|\{2 + \|\tilde{t}\| + \|\tilde{\tau}\| + \|\tilde{t}\|\,\|\tilde{\tau}\|\} \tag{5.7}$$

If the system is *dissipative*, i.e.,

$$\|S\| < 1 \quad \text{and} \quad \|K\| \leqslant 1,$$

then

$$\|\Delta Q_m\| \leqslant 5\|rK\|^{m+1}. \tag{5.8}$$

Since $\|rK\| < 1$, $\|\Delta Q_m\| \to 0$, and (5.8) gives the $(m+1)$th-order of convergence.

To compute Q_m and Q_{m+1} in (5.4) and (5.5) it is necessary to obtain the solutions of Sec. III for ρ, $\tilde{t}$, $\tilde{\tau}$, and r, which in this setting is more cumbersome than in the Chandrasekhar problem. A detailed discussion of these solutions is beyond the scope and intention of this paper. However, there are several techniques that can be used: (i) the successive order-of-scattering method of Bellman *et al.*[21]; (ii) the extended adding procedure of Pom and Ueno[22]; and (iii) the modified WKB method of Duddley and Wang.[23]

[1]Lord Raleigh, *Scientific Papers* (Cambridge, England, 1899), pp. 87, 104, 518.
[2]A. Schuster, Astrophys. J. **21**, 1 (1905).
[3]K. Schwarzschild, Göttinger Nachrichten, p. 41 (1906).
[4]S. Chandrasekhar, in *Radiative Transfer* (Oxford, U.P., New York, 1950).
[5]S. Ueno, "On the principle of invariance in a semi-infinite inhomogeneous atmosphere," Prog. Theor. Phys. **24**, 734 (1960).
[6]R. Bellman and R. Kalaba, "On the principle of invariant imbedding and propagation through inhomogeneous media," Proc. Natl. Acad. Sci. U.S.A. **42**, 629 (1956).
[7]R. Redheffer, "On the relation of transmission line theory to scattering and transfer," J. Math. Phys. (Cambridge, Mass) **41**, 1 (1962).
[8]R. Preisendorfer, "Invariant imbedding relation for the principles of invariance," Proc. Natl. Acad. Sci. USA **44**, 320 (1958).
[9]S. Paszkowski, "Sur un probleme du calcul de probabilité (II)," Studia Math. **15**, 273 (1956).
[10]G. M. Wing, Syllabus for a course on the theory of particle transport, Sandia Corp., 1961.
[11]S. Ueno and A. P. Wang, "Scattering and transmission functions of radiation by finite atmospheres with reflecting surfaces," Bull. Am. Astron. Soc. **6**, 385 (1974).
[12]A. P. Wang, "Invariant imbedding and scattering process," J. Math.

Anal. Appl. **17**, 48 (1967).
[13]S. Ueno and S. Mukai, "Atmospheric effects on remotely sensed data from space," in *Proceedings of the IFAC Symposium on Environmental Systems Planning, Design and Control*, edited by H. Akashi (Pergamon, Oxford, 1977), pp. 423–428.
[14]A. P. Odell and J. A. Weinman, "The effect of atmospheric haze on images of the Earth's surface," J. Geophys. Res. **80**, 5035 (1975).
[15]S. Ueno and Y. Terashita, "The atmospheric scattering effects on the remotely sensed Earth's imagery," in *Evaluation and Control of Environmental Systems*, edited by G. Sawaragi (Kyoto, 1979).
[16]A. P. Wang, "Correction of atmospheric effects in remote sensing," J. Math. Modeling **9**(2), 117 (1987).
[17]A. P. Wang, "Linear operator and transfer of radiation with spherical symmetry," J. Math. Phys. **14**, 855 (1973).
[18]S. Ueno, "The probabilistic method for problem of radiative transfer." Astrophys. J. **132**, 729 (1960).
[19]I. W. Busbridge, "The mathematics of radiative transfer," in *Cambridge Tract*, No. 5 (Cambridge, U.P., London, 1960).
[20]W. T. Reid, "Solution of a Riccati matrix differential as functions of initial values," J. Math. Mech. **8**(2), 221 (1959).
[21]R. Bellman, A. L. Fymat, S. Ueno, and R. Vasudevan, "Invariant imbedding and radiation dosimetry: IX. Inverse problem of determining a plane source in a finite isotropically scattering target slab," Math. Biosci. **20**, 315 (1974).
[22]P. T. Poon and S. Ueno, "Scattering matrix and doubling equations for the scattering and transmission functions," J. Atmos. Sci. **30**, 954 (1973).
[23]D. Duddley and A. P. Wang, "Operator theory on WKB and Bremmer series," J. Math. Phys. **18**, 47 (1977).

On the Levinson theorem for Dirac operators

Martin Klaus
Mathematics Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 17 August 1988; accepted for publication 30 August 1989)

For the Dirac equation with potential $V(r)$ obeying $\int_0^\infty (1+r)|V(r)|dr<\infty$ we prove a relativistic version of Levinson's theorem that relates the number of bound states in the spectral gap $[-m,m]$ to the variation of an appropriate phase along the continuous part of the spectrum. In the process, the asymptotic properties of the Jost function as $E\to\pm m$ are analyzed in detail. The connection with the nonrelativistic version of Levinson's theorem is also established.

I. INTRODUCTION

In this paper, we consider the Dirac equation for a particle moving in a central electrostatic potential $V(r)$. Separation of variables leads to the following systems of equations

$$H_\kappa(c)\psi = c\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\psi' + \begin{pmatrix}mc^2+V(r) & \kappa/r\\ \kappa/r & -mc^2+V(r)\end{pmatrix}\psi = E\psi,\ \psi=\begin{pmatrix}\psi_1(r)\\ \psi_2(r)\end{pmatrix} \tag{1.1}$$

on $0<r<\infty$. Here, m is the mass of the particle, c is the velocity of light, E is the energy (in units where $\hbar=1$), and κ is a nonzero integer. We assume that $V(r)$ satisfies

$$\int_0^\infty (1+r)|V(r)|dr<\infty. \tag{1.2}$$

This condition guarantees that the differential operator H is limit point at zero[1] (it is always limit point at infinity[2]) so that H_κ can be viewed as a self-adjoint operator in the Hilbert space of vector-valued functions ψ satisfying $\int_0^\infty(|\psi_1|^2+|\psi_2|^2)dr<\infty$. The spectrum of H_κ is absolutely continuous on $(-\infty,-mc^2]\cup[mc^2,\infty)$ and consists of at most finitely many (simple) eigenvalues in the gap $[-mc^2,mc^2]$.

There is a deep connection between the continuous part and the discrete part of the spectrum. In the Schrödinger case, this is the content of Levinson's theorem.[3] Here we study its relativistic analog. In order to facilitate the comparison with other authors, we make the substitution $\phi=\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)\psi$, which converts (1.1) into

$$\phi' = \begin{pmatrix}\kappa/r & mc+c^{-1}V(r)-c^{-1}E\\ mc-c^{-1}V(r)+c^{-1}E & -\kappa/r\end{pmatrix}\phi. \tag{1.3}$$

Henceforth, we will only consider $\kappa\geqslant 1$ which causes no loss of generality since on interchanging the components of ϕ the problem corresponding to κ, V, E is equivalent to that corresponding to $-\kappa,-V,-E$. We now also set $c=1$ in this section and in Sec. II. Under assumption (1.2), Eq. (1.3) has a solution called the regular solution, which satisfies

$$\lim_{r\to 0} r^{-\kappa}\varphi_\kappa(E,r) = \begin{pmatrix}1/(2\kappa-1)!!\\ 0\end{pmatrix}. \tag{1.4}$$

As $r\to\infty$ this solution behaves like

$$\varphi_\kappa(E,r) = k^{-\kappa}|F_\kappa(E)| \times\begin{pmatrix}\cos(kr-\kappa\pi/2-\delta_\kappa(E))\\ k/(E-m)\sin(kr-\kappa\pi/2-\delta_\kappa(E))\end{pmatrix} + o(1). \tag{1.5}$$

The parameter $k=\sqrt{E^2-m^2}$ is defined by choosing a branch of k such that $k>0$ for $E>m$ and $\operatorname{Im} k\geqslant 0$ for $\operatorname{Im} E\geqslant 0$. Then $k<0$ corresponds to $E<-m$. This choice is different from that in Ref. 4 where $\operatorname{Re} E\geqslant 0$ corresponds to $\operatorname{Im}\kappa\leqslant 0$. Also, due to different conventions, our phase $\delta_\kappa(E)$ differs in sign from that in Refs. 4 or 5 for $E>m$ but agrees with it for $E<m$. Conceptually, the basic parameter for us is E and not k. The function $F_\kappa(E)$ is the analog of the Jost function in the Schrödinger case. It can be written as

$$F_\kappa(E) = |F_\kappa(E)|e^{i\delta_\kappa(E)} = 1+\int_0^\infty (\varphi_\kappa(E,t))^T V(t) f_\kappa^0(E,t)dt, \tag{1.6}$$

where

$$f_\kappa^0(E,r) = k^\kappa\begin{pmatrix}[k^2r/(E+m)]h_{\kappa-1}(kr)\\ krh_\kappa(kr)\end{pmatrix}. \tag{1.7}$$

Here, $h_\kappa(kr)=n_\kappa(kr)+ij_\kappa(kr)$, where n_κ and j_κ denote spherical Bessel functions.[4] Also, T denotes the ordinary transpose and a superscript 0 indicates a solution of the unperturbed ($V=0$) problem. We recall that the zeros of $F_\kappa(E)$ are all simple, lie in the interval $[-m,m]$, and correspond to eigenvalues of H [see Ref. 4, Sec. 2, where $h_\kappa(k)$ corresponds to $F_\kappa(E)$]. The only exception occurs at $E=-m$ for $\kappa=1$ when, if $F_1(-m)=0$, the solution $\varphi_1(-m,r)$ is bounded but not square integrable at infinity [see (2.9),(2.11) below]. Then we say $E=-m$ is a half-bound state.

Theorem (1.1): Let $V(r)$ obey (1.2). Let $N_\kappa(\kappa\geqslant 1)$ denote the number of eigenvalues of H_κ in $[-m,m]$. Then

$$N_\kappa = (1/\pi)(\delta_\kappa(-m)-\delta_\kappa(m)),\quad \kappa\geqslant 2, \tag{1.8}$$

$$N_\kappa = (1/\pi)(\delta_\kappa(-m)-\delta_\kappa(m))+(1/2\pi)\Delta,\quad \kappa=1, \tag{1.9}$$

where

$$\Delta=\begin{cases}0, & E=-m \text{ is not a half-bound state,}\\ -\pi, & E=-m \text{ is a half-bound state.}\end{cases} \tag{1.10}$$

 0022-2488/90/010182-09$03.00

It is important to add that the difference $\delta_\kappa(-m) - \delta_\kappa(m)$ may be viewed as the change of phase as we go, through *real* values, from $E=m$ to $+\infty$ and then from $E=-\infty$ to $E=-m$. Relativistic versions of the Levinson theorem have been studied before, by Barthélémy[4] and more recently by Ni[6] and Ma and Ni[5] and also in Ref. 7. However, in Ref. 5, the authors point out that the results of Refs. 4 and 6 are not correct in general and they go on to derive a correct form of Levinson's theorem in the case where $V(r)$ has compact support. We will comment on the fallacy of Ref. 4 after the proof of Theorem (1.1). The hypothesis (1.2) is weaker than that in Ref. 4 where $\int_0^\infty |V(r)|r^n\,dr<\infty$ for $n=0$, 1, or 2 was assumed and for Levinson's theorem $n=2$ (and $n=0$) were absolutely essential. The condition (1.2) is optimal for large r as far as moment-type conditions go since if $V\sim cr^{-2}$ as $r\to\infty$ then Levinson's theorem must be modified.[8] This remark should answer to some extent a question in Ref. 5 concerning the proper assumptions on V which will insure that Levinson's theorem holds. Concerning the behavior as $r\to 0$, condition (1.2) excludes a r^{-1} singularity. It is conceivable to us that the methods used in this paper can be extended to include such a behavior. However, then the r^{-1} singularity cannot be treated perturbatively, which leads to some complications at the level of the unperturbed problem.

In Ref. 7, Dirac systems containing, in place of κ/r, a coefficient $p(r)$ such that $\int_0^\infty (1+r)|p(r)|dr<\infty$ were considered. In that case, $p(r)$ can be included in the perturbation which leads to some simplifications in the analysis. Conceivably, such a term $p(r)$ could be added to κ/r in (1.1) without essentially altering the analysis but we will not do so here. The Levinson theorem for the Schrödinger equation under assumption (1.2) was studied in Ref. 9. Although both Refs. 9 and 7 have provided us with some guidance for the present paper, we have encountered some unexpected complications in the case where $F_\kappa(\pm m)=0$.

There exist several methods for proving the Levinson theorem in the relativistic and nonrelativistic case. In the relativistic case, the Green's function method was used in Ref. 5 and an approach based on the Sturm–Liouville theorem was used in Refs. 10 and 8. This latter approach was also used in the nonrelativistic case in Refs. 11 and 12. We follow Levinson's original proof for the Schrödinger equation[3] which is based on a detailed study of the asymptotic properties of the Jost function whereby the main effort goes into analyzing the case where $F_\kappa(\pm m)=0$. As in Ref. 9 but in contrast to Ref. 4 we do not work with the Jost *solution* at all, only with the regular solution $\varphi_\kappa(E,r)$ since the latter is better behaved as $E\to\pm m$ than the former [compare also Ref. 9, Corollary (3.31)].

Theorem (1.1) is proved in Sec. II. In Sec. III we discuss the nonrelativistic limit $c\to\infty$.

II. PROPERTIES OF $F_\kappa(E)$ AND PROOF OF THEOREM (1.1)

The solution $\varphi_\kappa(E,r)$ defined by (1.4) satisfies[4]

$$\varphi_\kappa(E,r)=\varphi_\kappa^0(E,r)+\int_0^r K(E;r,t)V(t)\varphi_\kappa(E,t)dt, \tag{2.1}$$

where

$$K(E;r,t)=\varphi_\kappa^0(E,r)(\psi_\kappa^0(E,t))^T-\psi_\kappa^0(E,r)(\varphi_\kappa^0(E,t))^T, \tag{2.2}$$

$$\varphi_\kappa^0(E,r)=k^{-\kappa}\begin{pmatrix} krj_{\kappa-1}(kr) \\ [k/(E-m)]krj_\kappa(kr)\end{pmatrix}, \tag{2.3}$$

$$\psi_\kappa^0(E,r)=k^{\kappa}\begin{pmatrix} (E-m)rn_{\kappa-1}(kr) \\ krn_\kappa(kr)\end{pmatrix}. \tag{2.4}$$

We first collect some results concerning the solutions at $E=\pm m$ which will be used later on. From standard asymptotic analysis based on (2.1) it follows that

$$\varphi_\kappa(m,r)=\frac{1}{(2\kappa-1)!!}\times\begin{pmatrix} F_\kappa(m)r^\kappa+o(r^\kappa) \\ [2m/(2\kappa+1)]F_\kappa(m)r^{\kappa+1}+o(r^{\kappa+1})\end{pmatrix}, \quad r\to\infty, \tag{2.5}$$

$$\varphi_\kappa(-m,r)=\frac{1}{(2\kappa-1)!!}\begin{pmatrix} F_\kappa(-m)r^\kappa+o(r^\kappa) \\ o(r^{\kappa-1})\end{pmatrix}, \quad r\to\infty, \tag{2.6}$$

respectively, where [by (1.6)]

$$F_\kappa(m)=1+(2\kappa-1)!!\int_0^\infty \varphi_{\kappa,2}(m,t)V(t)t^{-\kappa}dt, \tag{2.7}$$

$$F_\kappa(-m)=1-2m(2\kappa-3)!!\int_0^\infty \varphi_{\kappa,1}(-m,t)\times V(t)t^{-\kappa+1}dt+(2\kappa-1)!!\times\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t^{-\kappa}dt. \tag{2.8}$$

The reason for having a term $o(r^{\kappa-1})$ in the second component of (2.6) is that $r^{-\kappa}\int_0^r\varphi_{\kappa,1}(-m,t)V(t)t^\kappa dt = o(r^{\kappa-1})$ which follows from the behavior of $\varphi_{\kappa,1}(-m,t)$ as $t\to\infty$ and (1.2). Besides the solution φ_κ, Eq. (1.3) has, for $E=m$, a second solution $\tilde\varphi$ satisfying

$$\tilde\varphi_\kappa(m,r)=\begin{pmatrix} o(r^{-\kappa-1}) \\ r^{-\kappa}+o(r^{-\kappa})\end{pmatrix}, \quad r\to\infty. \tag{2.9}$$

Similarly, for $E=-m$, we have

$$\tilde\varphi_\kappa(-m,r)=\begin{pmatrix} [2m/(1-2\kappa)]r^{1-\kappa}+o(r^{1-\kappa}) \\ r^{-\kappa}+o(r^{-\kappa})\end{pmatrix}, \quad r\to\infty. \tag{2.10}$$

Here, if we replace the o terms by zero we get exact solutions of the unperturbed ($V=0$) problem which are bounded at infinity. By considering Wronskians, we see that φ_κ and $\tilde\varphi_\kappa$ are linearly dependent if and only if $F_\kappa(\pm m)=0$. If that happens we set

$$\varphi_\kappa(\pm m,r)=-A_\kappa(\pm m)\tilde\varphi_\kappa(\pm m,r) \tag{2.11}$$

and deduce from (2.1) the representations

$$A_\kappa(m)=\int_0^\infty \varphi_{\kappa,1}(m,t)V(t)t^\kappa dt+\frac{2m}{2\kappa+1}\times\int_0^\infty \varphi_{\kappa,2}(m,t)V(t)t^{\kappa+1}dt, \tag{2.12}$$

$$A_\kappa(-m)=\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t^\kappa\,dt. \tag{2.13}$$

If $E=m$ and $F_\kappa(m)=0$, then, of course, there exists a second solution χ_κ of (1.3) such that $W(\varphi_\kappa,\chi_\kappa)=\varphi_\kappa\chi_{\kappa,2}-\chi_{\kappa,1}\varphi_{\kappa,2}=1$. It satisfies

$$\chi_\kappa(m,r)=\frac{1}{A_\kappa(m)}\begin{pmatrix} r^\kappa+o(r^\kappa)\\ [2m/(2\kappa+1)]r^{\kappa+1}+o(r^{\kappa+1})\end{pmatrix},\quad r\to\infty \tag{2.14}$$

and

$$\chi_\kappa(m,r)=\begin{pmatrix} o(r^{-\kappa})\\ (2\kappa-1)!!r^{-\kappa}+o(r^{-\kappa})\end{pmatrix},\quad r\to 0. \tag{2.15}$$

Similarly, at $E=-m$, if $F_\kappa(-m)=0$, then a second solution χ_κ exists such that [again $W(\varphi_\kappa,\chi_\kappa)=1$]

$$\chi_\kappa(-m,r)=\frac{1}{A_\kappa(-m)}\begin{pmatrix} r^\kappa+o(r^\kappa)\\ o(r^{\kappa-1})\end{pmatrix},\quad r\to\infty, \tag{2.16}$$

$$\chi_\kappa(-m,r)=\begin{pmatrix} o(r^{-\kappa})\\ (2\kappa-1)!!r^{-\kappa}+o(r^{-\kappa})\end{pmatrix},\quad r\to 0. \tag{2.17}$$

The solution χ_κ will be needed later. At the heart of our method is the following lemma whose proof we defer to the Appendix.

Lemma (2.1): Let $V(r)$ obey (1.2). Fix $\delta>0$.

(i) If $F_\kappa(m)=0$, $\kappa\geqslant 1$, then

$$|\varphi_{\kappa,j}(E,r)-\varphi_{\kappa,j}(m,r)|\leqslant Ck^2[r/(1+kr)]^{\kappa+1},\quad j=1,2, \tag{2.18}$$

for $E\in[m,m+\delta]$ where C depends on δ but not on k and r.

(ii) If $F_\kappa(-m)=0$ and $\kappa\geqslant 2$, then

$$|\varphi_{\kappa,1}(E,r)-\varphi_{\kappa,1}(-m,r)|\leqslant Ck^2[r/(1+|k|r)]^\kappa, \tag{2.19}$$

$$|\varphi_{\kappa,2}(E,r)-\varphi_{\kappa,2}(-m,r)|\leqslant Ck^2[(r/(1+|k|r))^{\kappa+1}+(r/(1+|k|r))^\kappa], \tag{2.20}$$

while if $\kappa=1$, then

$$|\varphi_{\kappa,j}(E,r)-\varphi_{\kappa,j}(-m,r)|\leqslant C[(|k|r/(1+|k|r))^2+k^2r/(1+|k|r)],\quad j=1,2, \tag{2.21}$$

for $E\in[-m-\delta,-m]$.

The pertinent properties of the Jost function are summarized in the next theorem. We denote the L^2 norm of a vector function by $\|\ \|$.

Theorem (2.2): Let $V(r)$ obey (1.2), then

(i) $F_\kappa(E)$ is analytic for $\operatorname{Im}E>0$ and has an analytic continuation into the half-plane $\operatorname{Im}E<0$. Moreover, the extended function $F_\kappa(E)$ assumes continuous boundary values as E approaches the real axis from either above or below.

(ii) As $|E|\to\infty$, $\operatorname{Im}E\geqslant 0$,

$$F_\kappa(E)\to e^{i\int_0^\infty V(t)dt}. \tag{2.22}$$

(iii) If $F_\kappa(m)=0$, then

$$F_\kappa(E)=c_\kappa k^2+o(k^2), \tag{2.23}$$

$$c_\kappa=[(2\kappa-1)!!/2mA_\kappa(m)]\|\varphi_\kappa(m,\cdot)\|^2, \tag{2.24}$$

as $E\to m$ uniformly in $0\leqslant\arg(E-m)\leqslant 2\pi$.

(iv) If $F_\kappa(-m)=0$ and $\kappa\geqslant 2$, then

$$F_\kappa(E)=d_\kappa k^2+o(k^2), \tag{2.25}$$

$$d_\kappa=-[(2\kappa-1)!!/2mA_\kappa(-m)]\|\varphi_\kappa(m,\cdot)\|^2, \tag{2.26}$$

while if $\kappa=1$, then

$$F_\kappa(E)=d_\kappa k+o(k), \tag{2.27}$$

$$d_\kappa=2miA_\kappa(-m), \tag{2.28}$$

as $E\to -m$ uniformly in $-\pi\leqslant\arg(E+m)\leqslant\pi$.

Proof: (i) It follows from Ref. 4 that there is a constant C such that for all E with $\operatorname{Im}E\geqslant 0$:

$$|\varphi_{\kappa,1}(E,r)|\leqslant Ce^{(\operatorname{Im}k)r}[r/(1+|k|r)]^\kappa \tag{2.29}$$

and

$$|\varphi_{\kappa,2}(E,r)|\leqslant C|E+m|e^{(\operatorname{Im}k)r}[r/(1+|k|r)]^{\kappa+1}.$$

Also,

$$|f^0_{\kappa,1}(E,r)|\leqslant C(k^2/|E+m|)e^{-(\operatorname{Im}k)r}\times[(1+|k|r)/r]^{\kappa-1}, \tag{2.30}$$

$$|f^0_{\kappa,2}(E,r)|\leqslant Ce^{-(\operatorname{Im}k)r}[(1+|k|r)/r]^\kappa. \tag{2.31}$$

Hence,

$$|(\varphi_\kappa(E,r))^TV(r)f^0_\kappa(E,r)|\leqslant C|V(r)|(1+r),\quad |E\pm m|\leqslant\delta, \tag{2.32}$$

$$|(\varphi_\kappa(E,r))^TV(r)f^0_\kappa(E,r)|\leqslant C|V(r)|,\quad |E\pm m|\geqslant\delta, \tag{2.33}$$

for any $\delta>0$ with an appropriate constant C. Since $\varphi_\kappa(E,r)$ is an entire function of E and $f^0_\kappa(E,r)$ is analytic for $\operatorname{Im}E>0$ and continuous for $\operatorname{Im}E\geqslant 0$ the bounds (2.32), (2.33) insure that $F_\kappa(E)$ has the asserted analyticity and continuity properties. We get an analytic continuation into the lower half plane by the Schwarz reflection principle because $F_\kappa(E)$ is real for $-m<E<m$.

(ii) Since the construction of $\varphi_\kappa(E,r)$ only requires knowledge of V on $[0,r]$ the large E behavior of $\varphi_\kappa(E,r)$ can be inferred from Ref. 4,

$$\varphi_\kappa(E,r)=k^{-\kappa}\begin{pmatrix}\cos(kr-\kappa\pi/2-\int_0^r V(t)dt)\\ \sin(kr-\kappa\pi/2-\int_0^r V(t)dt)\end{pmatrix}+o(k^{-\kappa}), \tag{2.34}$$

and also

$$f^0_\kappa(E,r)=k^\kappa e^{i(kr-\kappa\pi/2)}\begin{pmatrix} i\\ 1\end{pmatrix}+o(k^\kappa), \tag{2.35}$$

as $|E|\to\infty$ on $\{E:\operatorname{Im}E\geqslant 0\}$. Owing to (2.29), (2.30), and (2.31) we may insert (2.34) and (2.35) in (1.6) and apply the Lebesgue Dominated Convergence Theorem. Then (2.22) follows.

(iii) Suppose $E>m$. We break the right-hand side of (1.6) into three parts, $F_\kappa(E)=I_1+I_2+I_3$, using $F_\kappa(m)=0$, where

$$I_1=\int_0^\infty(\varphi_\kappa(m,t))^TV(t)[f^0_\kappa(E,t)-f^0_\kappa(m,t)]dt, \tag{2.36}$$

$$I_2=\int_0^\infty[(\varphi_\kappa(E,t))^T-(\varphi_\kappa(m,t))^T]V(t)f^0_\kappa(m,t)dt, \tag{2.37}$$

$$I_3 = \int_0^\infty [(\varphi_\kappa(E,t))^T - (\varphi_\kappa(m,t))^T] V(t) \times [f_\kappa^0(E,t) - f_\kappa^0(m,t)]\,dt. \tag{2.38}$$

Consider I_1 first. By (1.4), (2.9), and (2.11)

$$|\varphi_{\kappa,j}(m,r)| \leqslant C[r^\kappa/(1+r)^{2\kappa}], \quad j=1,2. \tag{2.39}$$

Moreover,

$$f_\kappa^0(E,r) = \begin{pmatrix} 0 \\ (2\kappa-1)!!r^{-\kappa} \end{pmatrix} + k^2 \begin{pmatrix} (2m)^{-1}(2m-3)!!r^{-\kappa+1} \\ (\tfrac{1}{2})(2\kappa-3)!!r^{-\kappa+2} \end{pmatrix} + o(k^2) \tag{2.40}$$

and

$$|f_{\kappa,2}^0(E,r) - f_{\kappa,2}^0(m,r)| \leqslant Ck^2[r^{2-\kappa}/(1+kr)^{2-\kappa}]. \tag{2.41}$$

From (2.39), (2.40), (2.41), and (2.30) we find

$$I_1 = a_\kappa k^2 + o(k^2) \quad \text{as } k \to 0 (E \downarrow m), \tag{2.42}$$

where

$$a_\kappa = \frac{(2\kappa-3)!!}{2m}\left(\int_0^\infty \varphi_{\kappa,1}(m,t) V(t) t^{-\kappa+1}\,dt + m\int_0^\infty \varphi_{\kappa,2}(m,t) V(t) t^{-\kappa+2}\,dt \right) \tag{2.43}$$

[with the convention $(-1)!! = 1$].

Now consider I_2. If we expand

$$\varphi_\kappa(E,r) = \varphi_\kappa(m,r) + (E-m)u_\kappa(m,r) + O(k^4) \tag{2.44}$$

and use (2.18), then by dominated convergence

$$I_2 = b_\kappa k^2 + o(k^2) \text{ as } k \to 0, \tag{2.45}$$

where

$$b_\kappa = \frac{(2\kappa-1)!!}{2m}\int_0^\infty (u_{\kappa,2}(m,t))^T V(t) t^{-\kappa}\,dt. \tag{2.46}$$

For later use we need to know the asymptotic behavior of $u_\kappa(m,r)$ as $r \to 0$ and $r \to \infty$. Equation (1.3) for $E = m$ has the fundamental matrix

$$R_\kappa(r) = \begin{pmatrix} \varphi_{\kappa,1}(m,r) & \chi_{\kappa,1}(m,r) \\ \varphi_{\kappa,2}(m,r) & \chi_{\kappa,2}(m,r) \end{pmatrix}, \tag{2.47}$$

where χ_κ is the solution introduced before, see (2.14), (2.15). Applying the variation of parameters formula to (1.3), viewing $(E-m)\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ as the nonhomogeneous term we find

$$u_\kappa(m,r) = \int_0^r R_\kappa(r) R_\kappa^{-1}(t) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \varphi_\kappa(m,t)\,dt. \tag{2.48}$$

Therefore, by (1.4), (2.9), (2.11), (2.14), and (2.15) we can say that

$$u_\kappa(m,r) = \begin{pmatrix} o(r^{\kappa+1}) \\ r^{\kappa+1}/(2\kappa+1)!! \end{pmatrix}, \quad r \to 0, \tag{2.49}$$

$$u_\kappa(m,r) = \|\varphi_\kappa(m,\cdot)\|^2 \chi_\kappa(m,r)(1+o(1)), \quad r \to \infty. \tag{2.50}$$

Now, by (2.30), (2.41), (2.18), and (1.2) we have

$$|I_3| \leqslant Ck^2 \int_0^\infty \left(\frac{kt}{1+kt}\right)^2 t\,|V(t)|\,dt = o(k^2). \tag{2.51}$$

Therefore, $F_\kappa(E) = (a_\kappa + b_\kappa)k^2 + o(k^2)$ and it only remains to reduce $c_\kappa = a_\kappa + b_\kappa$ to the expression (2.24). This can be accomplished by means of the following identities:

$$\int_0^\infty u_{\kappa,2}(m,t) V(t) t^{-\kappa}\,dt = \int_0^\infty t^{-\kappa}\varphi_{\kappa,2}(m,t)\,dt + A_\kappa^{-1}(m)\|\varphi_\kappa(m,\cdot)\|^2, \tag{2.52}$$

$$(2\kappa-1)\int_0^\infty \varphi_{\kappa,2}(m,t) t^{-\kappa}\,dt = \int_0^\infty \varphi_{\kappa,1}(m,t)(2m - V(t)) t^{-\kappa+1}\,dt, \tag{2.53}$$

$$-2\int_0^\infty \varphi_{\kappa,1}(m,t) t^{-\kappa+1}\,dt = \int_0^\infty \varphi_{\kappa,2}(m,t) V(t) \times t^{-\kappa+2}\,dt. \tag{2.54}$$

To prove (2.52) we use the equation

$$u_\kappa'(m,r) = \begin{pmatrix} \kappa/r & V(r) \\ 2m - V(r) & -\kappa/r \end{pmatrix} u_\kappa(m,r) + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \varphi_\kappa(m,r). \tag{2.55}$$

We multiply the equation $u_{\kappa,1}'(m,t) = (\kappa/t)u_{\kappa,1}(m,t) + V(t)u_{\kappa,2}(m,t) - \varphi_{\kappa,2}(m,t)$ by $t^{-\kappa}$ and integrate by parts. The boundary term is $u_{\kappa,1}(m,t)t^{-\kappa}\big|_0^\infty = A_\kappa^{-1}(m)\|\varphi_\kappa(m,\cdot)\|^2$ by virtue of (2.14), (2.49), and (2.50). Thus (2.52) follows. To prove (2.53) and (2.54) we use (1.3) and multiply the individual component equations by appropriate powers of t and integrate by parts. Now (2.52)–(2.54) are easily combined with (2.43) and (2.46) which yields (2.24). Since $F_\kappa(\bar{E}) = \overline{F_\kappa(E)}$, Eq. (2.23) also holds if $E \downarrow m$ along the lower edge of the cut $[m,\infty]$. Therefore, by a variant of the Phragmén–Lindelöf theorem[13] (the same argument was used in Ref. 7) we obtain the desired uniformity in $\arg(E-m)$. Part (iii) is proved.

(iv) The proof is, of course, similar to that of (iii), but the case $\kappa = 1$ requires special attention. Also, the estimates are more tedious because the two components of $\varphi_\kappa(-m,r)$ must be controlled by separate bounds, namely,

$$|\varphi_{\kappa,1}(-m,r)| \leqslant C[r^\kappa/(1+r)^{2\kappa-1}], \tag{2.56}$$

$$|\varphi_{\kappa,2}(-m,r)| \leqslant C[r^\kappa/(1+r)^{2\kappa}], \tag{2.57}$$

and similarly for the difference $\varphi_\kappa(E,r) - \varphi_\kappa(-m,r)$ according to Lemma (2.1). Assume $\kappa \geqslant 2$ first. Then

$$f_\kappa^0(E,r) = \begin{pmatrix} -2m(2\kappa-3)!!r^{-\kappa+1} \\ (2\kappa-1)!!r^{-\kappa} \end{pmatrix} + \frac{(2\kappa-3)!!}{2}k^2 \times \begin{pmatrix} -r^{-\kappa+1}/m - [2m/(2\kappa-3)]r^{-\kappa+3} \\ r^{-\kappa+3} \end{pmatrix} + o(k^2) \tag{2.58}$$

and

$$|f^0_{\kappa,1}(E,r) - f^0_{\kappa,1}(-m,r)| \leqslant Ck^2[r^{1-\kappa}/(1+|k|r)^{1-\kappa}](1+[r/(1+|k|r)]^2), \tag{2.59}$$

$$|f^0_{\kappa,2}(E,r) - f^0_{\kappa,2}(-m,r)| \leqslant C(kr)^2(1+|k|r)^2. \tag{2.60}$$

So, if we split $F_\kappa(E)$ in analogy to (2.36)–(2.38) with respect to E near $-m$, we have that

$$I_1 = \alpha_\kappa k^2 + o(k^2), \quad k \to 0 (E \uparrow -m), \tag{2.61}$$

where

$$\alpha_\kappa = -(2\kappa-5)!!m\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t^{-\kappa+3}\,dt + \frac{(2\kappa-3)!!}{2}\int_0^\infty \varphi_{\kappa,2}(-m,t)V(t)t^{-\kappa+2}\,dt - \frac{(2\kappa-3)!!}{2m}\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t^{-\kappa+1}\,dt \tag{2.62}$$

and

$$I_2 = \beta_\kappa k^2 + o(k^2), \quad k \to 0, \tag{2.63}$$

where

$$\beta_\kappa = (2\kappa-3)!!\int_0^\infty u_{\kappa,1}(-m,t)V(t)t^{-\kappa+1}\,dt - \frac{(2\kappa-1)!!}{2m}\int_0^\infty u_{\kappa,2}(-m,t)V(t)t^{-\kappa}\,dt. \tag{2.64}$$

Here, $u_\kappa(-m,r)$ obeys $\varphi_\kappa(E,r) = \varphi_\kappa(-m,r) + (E-m)u_\kappa(-m,r) + O(k^4)$. By using dominated convergence $I_3 = o(k^2)$ so that $F_\kappa(E) = d_\kappa k^2 + o(k^2)$ with $d_\kappa = \alpha_\kappa + \beta_\kappa$ and we must reduce this coefficient to the form (2.26). To this end we use the following identities, the proof of which is similar to that of (2.52)–(2.54) and is therefore omitted:

$$\int_0^\infty u_{\kappa,2}(-m,t)V(t)t^{-\kappa}\,dt = \int_0^\infty \varphi_{\kappa,2}(-m,t)t^{-\kappa}\,dt - 2m\int_0^\infty u_{\kappa,2}(-m,t)t^{-\kappa}\,dt + A_\kappa^{-1}(-m)\|\varphi_\kappa(-m,\cdot)\|^2, \tag{2.65}$$

$$\int_0^\infty u_{\kappa,1}(-m,t)V(t)t^{-\kappa+1}\,dt = \int_0^\infty \varphi_{\kappa,1}(-m,t)t^{-\kappa+1}\,dt - (2\kappa-1)\times\int_0^\infty u_{\kappa,2}(-m,t)t^{-\kappa}\,dt, \tag{2.66}$$

$$\int_0^\infty \varphi_{\kappa,2}(-m,t)t^{-\kappa}\,dt = \frac{-1}{2\kappa-1}\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t^{-\kappa+1}\,dt, \tag{2.67}$$

$$-2\int_0^\infty \varphi_{\kappa,1}(-m,t)t^{-\kappa+1}\,dt = 2m\int_0^\infty \varphi_{\kappa,2}(-m,t)t^{-\kappa+2}\,dt + \int_0^\infty \varphi_{\kappa,2}(-m,t)V(t)t^{-\kappa+2}, \tag{2.68}$$

$$(2\kappa-3\int_0^\infty \varphi_{\kappa,2}(-m,t)t^{-\kappa+2}\,dt = -\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t^{-\kappa+3}\,dt. \tag{2.69}$$

Finally, if $\kappa = 1$, then I_2 and I_3 both are $o(k)$ on account of (2.21). Since $f^0_{\kappa,1}(E,r) - f^0_{\kappa,1}(-m,r) = -2mikr + O(k^2)$ and $f^0_{\kappa,2}(E,r) - f^0_{\kappa,2}(-m,r) = O(k^2)$ we get

$$F_\kappa(E) = -2mik\int_0^\infty \varphi_{\kappa,1}(-m,t)V(t)t\,dt + o(k) = -2mikA_\kappa(-m) + o(k). \tag{2.70}$$

The uniform validity in $\arg(E+m)$ of (2.25) and (2.27) again follows from a Phragmén–Lindelöf type argument. This completes the proof of Theorem (2.2).

Proof of Theorem (1.1): As in Ref. 7, we choose a contour in the closed upper half plane consisting of a semicircle of radius R, two line segments $[-R,-m-\epsilon]$ and $[m+\epsilon,R]$ and two semicircles about $\pm m$ of radius ϵ. Then we extend the contour into the lower half plane by reflection and assign a counterclockwise orientation. For ϵ sufficiently small, all zeros of $F_\kappa(E)$ except possibly those at $\pm m$ lie inside the contour. By the argument principle, the change in $\delta_\kappa(E)$ on this contour equals $2\pi\tilde{N}_\kappa$, where $\tilde{N}_\kappa$ denotes the number of eigenvalues that lie in $(-m,m)$. Since $F_\kappa(\bar{E}) = \overline{F_\kappa(E)}$ the change in $\delta_\kappa(E)$ on the top half of the contour is the same as that on the bottom half. For $\kappa \geqslant 2$, the change on the small circles centered at $\pm m$, respectively, approaches, as $\epsilon \to 0$, the value

$$\eta_\pm = \begin{cases} -2\pi, & \text{if } F_\kappa(\pm m) = 0, \\ 0, & \text{if } F_\kappa(\pm m) \neq 0. \end{cases} \tag{2.71}$$

Thus

$$\tilde{N}_\kappa = (1/\pi)(\delta_\kappa(-m) - \delta_\kappa(m)) + (1/2\pi)(\eta_+ + \eta_-). \tag{2.72}$$

Since N_κ also counts the eigenvalues at $\pm m$ if there are any we get (1.8) from (2.72) by dropping the term $(1/2\pi)(\eta_+ + \eta_-)$. Equation (2.71) also holds when $\kappa = 1$ with respect to $E = m$. If $\kappa = 1$ and $F_\kappa(-m) = 0$ then the change on the small circle centered at $-m$ is $-\pi$ if there is a half-bound state at $-m$ and 0 otherwise. This establishes (1.8) and (1.9) with respect to the above contour where now $\epsilon = 0$ but R is still finite. Of course, we can let $R \to \infty$ by using (2.22) so that $\delta_\kappa(-m) - \delta_\kappa(m)$ can be viewed as the change of phase over the continuous spectrum of H. Theorem (1.1) is proved.

The version of Levinson's theorem in Ref. 14 can easily be seen to agree with ours because $\delta_\kappa(\pm m) = 0 (\mathrm{mod}\,\pi)$ except for $\delta_1(-m)$ which equals $\pi/2 (\mathrm{mod}\,\pi)$ when $F_1(-m) = 0$. In connection with Ref. 4 we recall that there the concern was to find a relationship between the phase and the number of eigenvalues in $[0,m]$ and $[-m,0]$, respectively. Let the former be denoted by N_κ^+, the latter by N_κ^-. Suppose $E = 0$ is not an eigenvalue. Then again by a contour argument (take a contour which lies in $\{E: \mathrm{Re}\, E \geqslant 0\}$ such that it coincides with our previous contour for $\mathrm{Re}\, E > 0$ and

consists of a vertical segment joining iR to $-iR$) we have that

$$N_\kappa^+ = (1/\pi)(\delta_\kappa(0) - \delta_\kappa(m)), \quad \kappa \geqslant 1. \tag{2.73}$$

A similar formula holds for N_κ^- if we replace $\delta_\kappa(m)$ by $\delta_\kappa(0)$ in (1.8) and (1.9). The discrepancy with Ref. 4 is that the term $\delta_\kappa(0)$ is missing from the formula corresponding to (2.73). In Ref. 4 (p. 146) the phase changes over two line segments along the imaginary axis were said to cancel, but in our setting these segments correspond precisely to the segments from iR to 0 and from 0 to $-iR$, so the phase changes add, giving rise to the term $\pi^{-1}\delta_\kappa(0)$.

III. NONRELATIVISTIC LIMIT

The Jost function associated with (1.1) when c is no longer equal to one can be obtained from (1.6) by making the replacements $E \to c^{-1}E$, $m \to mc$, $V \to c^{-1}V$ [cf. (1.3)] so that

$$F_\kappa(E,c) = 1 + c^{-1}\int_0^\infty (\varphi_\kappa(E,c,t))^T \times V(t)\, f_\kappa^0(E,c,t)\, dt, \tag{3.1}$$

where

$$f_\kappa^0(E,c,t) = k_c^\kappa \begin{pmatrix} [k_c ct/(E+mc^2)]h_{\kappa-1}(k_c t) \\ k_c t h_\kappa(k_c t) \end{pmatrix}, \tag{3.2}$$

with $k_c = c^{-1}\sqrt{E^2 - m^2c^4}$, and where we have modified our notation in an obvious manner in order to exhibit the c dependence. We are interested in the nonrelativistic limit $c \to \infty$ of $F_\kappa(E,c)$ and its phase $\delta_\kappa(E,c)$ because by taking this limit we should be able to connect the relativistic Levinson theorem with the nonrelativistic one. Recall that if $c \to \infty$, then the Dirac equation goes over into a Schrödinger equation in a sense that has been made precise by several authors, see Hunziker,[14] Gesztesy *et al.*[15] (and the references quoted therein). The main goal of these papers was to develop the perturbation theory of eigenvalues and eigenfunctions in powers of c^{-1}. Some aspects of the scattering theory (convergence of wave operators) in the nonrelativistic limit were studied by Yajima.[16] These authors admit general, not necessarily spherically symmetric potentials. The only paper we are aware of which specifically considers the spherically symmetric case in a rigorous way is the old paper by Titchmarsh.[17] There it is shown that the solution $\varphi_\kappa(E,c,r)$ has a convergent expansion in powers of c^{-1} although under the strong restriction that V is a bounded function. But it has been pointed out in Ref. 17 and is not hard to verify that locally the integrability of V is the only requirement for the results of Ref. 17 to go through. In order to formulate our results we need some notation. Put $F_\kappa^+(e,c) = F_\kappa(E,c)$ if $E = mc^2 + e$ and $F_\kappa^-(e,c) = F_\kappa(E,c)$ if $E = -mc^2 - e$ where in both cases $e \geqslant 0$. Let L_κ^+ denote the Schrödinger operators

$$L_\kappa^\pm y = -(1/2m)y'' + [\kappa(\kappa \pm 1)/2mr^2]y \pm Vy = ey \tag{3.3}$$

[with $y(0) = 0$ when $\kappa = 1$] and let $\tilde{F}_\kappa^\pm(e)$ denote the corresponding Jost functions.[9] Also, put $\delta_\kappa^\pm(e,c) = \arg F_\kappa^\pm(e,c)$ and $\tilde{\delta}_\kappa^\pm(e) = \arg \tilde{F}_\kappa^\pm(e)$. Then we have

Theorem (3.1): (i) As $c \to \infty$, $F_\kappa^+(e,c) \to \tilde{F}_\kappa^+(e)$ and $F_\kappa^-(e,c) \to \overline{\tilde{F}_\kappa^-(e)}$ uniformly on $e \geqslant 0$.

(ii) If $\tilde{F}_\kappa^+(0) \neq 0$, then $\delta_\kappa^+(e,c) \to \tilde{\delta}_\kappa^+(e)$ uniformly on $e \geqslant 0$ as $c \to \infty$, while if $\tilde{F}_\kappa^+(0) = 0$, then $\delta_\kappa^+(e,c) \to \tilde{\delta}_\kappa^+(e)$ uniformly on $e \geqslant \epsilon > 0$ for any ϵ. Analogous statements hold for $F_\kappa^-(e,c)$ with the difference that $\delta_\kappa^-(e,c) \to -\tilde{\delta}_\kappa^-(e)$ [by (i)].

(iii) Let $n_\kappa^\pm$ denote the number of negative eigenvalues of $L_\kappa^\pm$ and let $N_\kappa(c)$ be the number of eigenvalues of $H_\kappa(c)$ in $[-mc^2, mc^2]$. Suppose $\tilde{F}_\kappa^\pm(0) \neq 0$. Then $N_\kappa(c) = n_\kappa^+ + n_\kappa^-$ for c sufficiently large.

Proof: We omit the suffix κ from the solutions φ_κ and f_κ^0 for this proof. Consider $F_\kappa^+(e,c)$. By (2.29) and (2.30) we have

$$\begin{aligned} c^{-1}&|\varphi_1(E,c,t)V(t)f_1^0(E,c,t)| \\ &\leqslant C\,[k_c^2/(e+2mc^2)]\,[|V(t)|t/(1+k_c t)] \\ &\leqslant Cc^{-1}|V(t)|, \end{aligned} \tag{3.4}$$

so this contribution to (3.1) vanishes as $c \to \infty$ uniformly on $e \geqslant 0$ [here we are also using that the constants C in the estimates (2.29), (2.30) can be chosen to be independent of c; this follows from their derivation in Ref. 4]. Regarding the second component we note the bound

$$\begin{aligned} c^{-1}|\varphi_2(E,c,t)V(t)f_2^0(E,c,t)| &\leqslant C\,[(e+2mc^2)/c^2] \\ &\quad \times [|V(t)|t/(1+k_c t)] \\ &\leqslant C(e/c^2 + 2m)|V(t)|t. \end{aligned} \tag{3.5}$$

This shows that the theorem on dominated convergence is applicable to (3.1). Alternatively, the middle term in (3.4) can be estimated by

$$C(1/c + \sqrt{2m/e})|V(t)|, \tag{3.6}$$

(3.5) and (3.6) together imply that in order to prove $F_\kappa^+(e,c) \to \tilde{F}_\kappa^+(e)$ uniformly on $e \geqslant 0$ it suffices to prove

$$\left|c^{-1}\int_0^R \varphi_2(E,c,t)V(t)f_2^0(E,c,t)dt - \int_0^R \tilde{\varphi}_2(e,t)V(t)\tilde{f}_2^0(e,t)dt\right| \to 0 \tag{3.7}$$

uniformly on every bounded interval $0 \leqslant e \leqslant e_0$. Here, $\tilde{\varphi}_2(E,t) = \lim_{c\to\infty} c^{-1}\varphi_2(E,c,t)$ and $\tilde{f}_2^0(e,t) = \lim_{c\to\infty} f_2^0(e,c,t)$. This is so because the difference $|c^{-1}\int_R^\infty \cdots - \int_R^\infty \cdots|$ can be made arbitrarily small uniformly in e by choosing R sufficiently large and letting $c \to \infty$ [use (3.5) for $e \in [0,1]$ and (3.6) for $e \in (1,\infty)$]. Another appeal to (3.6) then shows that the difference in (3.7) can be made arbitrarily small uniformly in e for $e \geqslant e_0$ by choosing e_0 large enough and taking $c \to \infty$. To prove (3.7) for a finite energy interval we estimate separately the integrals

$$\int_0^R (c^{-1}\varphi_2(E,c,t) - \tilde{\varphi}_2(e,t))V(t)f_2^0(E,c,t)dt \tag{3.8}$$

and

$$\int_0^R \tilde{\varphi}_2(e,t)V(t)(f_2^0(E,c,t)-\tilde{f}_2^0(e,t))dt. \tag{3.9}$$

Since t is restricted to a finite interval we can use the methods of Ref. 17 (and also Ref. 1) to show that $|c^{-1}\varphi_2(E,c,t)-\tilde{\varphi}_2(e,t)|\leqslant Cc^{-2}t^{\kappa}$. Since the techniques are standard we omit the details. Inserting this estimate along with (2.31) in (3.8) shows that the integral is $O(c^{-2})$ uniformly on $[0,e_0]$. Moreover, it follows that $\tilde{\varphi}_2(e,r)$ is a solution of $L_\kappa^+ y = ey$ such that $r^{-\kappa-1}\tilde{\varphi}_2(e,r)\to 2m/(2\kappa+1)!!$ as $r\to 0$. Considering (3.9) we have to estimate the difference $f_2^0(E,c,t)-\tilde{f}_2^0(c,t)=k_c^{\kappa}\hat{h}_\kappa(k_c t)-\tilde{k}^{\kappa}\hat{h}_\kappa(\tilde{k}t)=(k_c^{\kappa}-\tilde{k}^{\kappa})\hat{h}_\kappa(k_c t)+\tilde{k}^{\kappa}(\hat{h}_\kappa(k_c t)-\hat{h}_\kappa(\tilde{k}t))$ where $\hat{h}_\kappa(kt)=kth_\kappa(kt)$ and $\tilde{k}=\lim_{c\to\infty}k_c=\sqrt{2me}$. Clearly, $k_c^{\kappa}-\tilde{k}^{\kappa}=O(c^{-2})$ uniformly on $[0,e_0]$. Furthermore, $|\hat{h}_\kappa(k_c t)-\hat{h}_\kappa(\tilde{k}t)|\leqslant(k_c-\tilde{k})t\sup|\hat{h}'_\kappa(\xi)|$ where the sup is over $\xi\in[\tilde{k}t,k_c t]$. Now $\hat{h}'_\kappa(\xi)=[(\kappa+1)/\xi]\hat{h}_\kappa(\xi)-\hat{h}_{\kappa+1}(\xi)$ and $|\hat{h}_\kappa(\xi)|\leqslant C\xi^{-\kappa}(1+\xi)^{\kappa}$ so that

$$\tilde{k}^{\kappa}|\hat{h}_\kappa(k_c t)-\hat{h}_\kappa(\tilde{k}t)|\leqslant C(k_c-\tilde{k})(\tilde{k}/k_c)^{\kappa}t((1+k_c t)/t)^{\kappa+1}. \tag{3.10}$$

Since $|\tilde{\varphi}_2(e,t)|\leqslant Ct^{\kappa+1}$ we see that the integral (3.9) is $O(c^{-2})$. This proves the uniform convergence of $F^+(e,c)$ to $\tilde{F}_\kappa^+(e)=1+\int_0^\infty\tilde{\varphi}_2(e,t)V(t)\tilde{f}_2^0(e,t)dt$. The identification of $\tilde{F}_\kappa^+(e)$ as the Jost function for L_κ^+ is a straightforward computation using the relations $n_\kappa(kr)=(-1)^{\kappa}\times(\pi/2kr)^{1/2}J_{-\kappa-1/2}(kr)$ and $j_\kappa(kr)=(\pi/2kr)^{1/2}\times J_{\kappa+1/2}(kr)$ [and keeping in mind that $\tilde{\varphi}_2(e,r)$ is $(2\kappa+1)!!/2m$ times the standard solution y of $L_\kappa^+ y=ey$ with $r^{-\kappa-1}y\to 1$ as $r\to 0$]. In a similar manner, one proves the statements about $F_\kappa^-(e,c)$. One has to remember that since $E<-mc^2$ then $k_c<0$ and hence one also has to use relations like $J_\nu(k_c r)=e^{i\pi\nu}J_\nu(|k_c|r)$ in the process. Thus part (i) is proved.

The statements in (ii) immediately follow from the uniform convergence of the Jost functions and the fact that $F_\kappa^{\pm}(e,c)$ does not vanish for $e>0$, and also not for $e=0$ if c is sufficiently large and $\tilde{F}_\kappa^{\pm}(0)\neq 0$.

The assertion in (iii) is a consequence of Theorem (1.1) since

$$N_\kappa(c)=(1/\pi)(\delta_\kappa^+(\infty,c)-\delta_\kappa^+(0,c))+(1/\pi)(\delta_\kappa^-(0,c)-\delta_\kappa^-(\infty,c)) \tag{3.11}$$

so that on letting $c\to\infty$ the first term tends to $(1/\pi)(\tilde{\delta}_\kappa^+(\infty)-\tilde{\delta}_\kappa^+(0))=n_\kappa^+$ by the nonrelativistic Levinson theorem[3] (remember our phase convention) and the second term tends to $(1/\pi)(\tilde{\delta}_\kappa^-(\infty)-\tilde{\delta}_\kappa^-(0))=n_\kappa^-$ [since $F_\kappa^-(e,c)\to\tilde{F}_\kappa^-(e)$]. Theorem (3.1) is thus proved.

In closing, we remark that the weaker statement $N_\kappa(c)\geqslant n_\kappa^+ + n_\kappa^-$ for c sufficiently large can also be deduced directly from the results in Refs. 15 and 14 where it is shown (under somewhat different conditions on V) that

$$(H_\kappa(c)-m^2c-z)^{-1}\to\begin{pmatrix}(L_\kappa^+-z)^{-1} & 0\\ 0 & 0\end{pmatrix}$$

and

$$(-H_\kappa(c)-m^2c-z)^{-1}\to\begin{pmatrix}0 & 0\\ 0 & (L_\kappa^- -z)^{-1}\end{pmatrix}$$

in norm as $c\to\infty$ ($\operatorname{Im} z\neq 0$). By some additional arguments one can also obtain equality [i.e., $N_\kappa(c)=n_\kappa^+ + n_\kappa^-$] and one can extend the result to nonspherically symmetric operators [the condition $\tilde{F}_\kappa^{\pm}(0)\neq 0$ then becomes a condition on the absence of zero-energy resonances, resp., zero-energy bound states, for the Schrödinger operators $-(2m)^{-1}\Delta\pm V$]. Moreover, note that if $\kappa=1$ and $\tilde{F}_1^-(0)=0$ then $\tilde{\delta}_1^-(e)\to(\pi/2)(\operatorname{mod}\pi)$ as $e\to 0$ (Ref. 9, Theorem 4.1, case $l=0$). In general, one will have $F_1^-(0,c)\neq 0$ for c large enough and so, since $F_1^-(0,c)$ is real, $\delta_1^-(0,c)=0(\operatorname{mod}\pi)$. Thus in this case $\delta_1^-(0,c)$ does not converge to $-\tilde{\delta}_1^-(0)$ as $c\to\infty$ explaining the restriction $e\geqslant\epsilon>0$ in (ii).

ACKNOWLEDGMENT

I thank the referee for some useful suggestions one of which motivated us to add Sec. III to the paper.

APPENDIX: PROOF OF LEMMA (2.1)

The method of proof is similar to that used in the Schrödinger case to prove a corresponding result (see Ref. 9, Appendix) but as already mentioned, there are several complications which need to be dealt with carefully. We give a detailed proof of part (i) and then only indicate the changes that are needed for part (ii).

(i) Let $\Delta\varphi_\kappa(E,r)=\varphi_\kappa(E,r)-\varphi_\kappa(m,r)$ and define $\Delta\varphi_\kappa^0(E,r)$ and $\Delta\psi_\kappa^0(E,r)$ analogously.

Then using (2.2) we may write

$$\begin{aligned}\Delta\varphi_\kappa(E,r)=\Delta\varphi_\kappa^0(E,r)&+\Delta\varphi_\kappa^0(E,r)\int_0^r(\psi_\kappa^0(m,t))^T V(r)\varphi_\kappa(m,t)dt+\varphi_\kappa^0(E,r)\int_0^r(\Delta\psi_\kappa^0(E,t))^T V(t)\varphi_\kappa(m,t)dt\\&-\Delta\psi_\kappa^0(E,r)\int_0^r(\varphi_\kappa^0(E,t))^T V(t)\varphi_\kappa(m,t)dt-\psi_\kappa^0(m,r)\int_0^r(\Delta\varphi_\kappa^0(E,t))^T V(t)\varphi_\kappa(m,t)dt\\&+\int_0^r[\varphi_\kappa^0(E,t)(\psi_\kappa^0(E,t))^T-\psi_\kappa^0(E,r)(\varphi_\kappa^0(E,t))^T]V(t)\Delta\varphi_\kappa(E,t)dt.\end{aligned} \tag{A1}$$

We denote the six terms on the right-hand side by A_1 through A_6. The idea is to estimate these terms so that Gronwall's inequality can be used at the end. Suppose now that $E\in[m,m+\delta]$ for some $\delta>0$. We use C to denote a generic constant which depends on δ and κ but not on k and r. We will frequently use the following bounds.

$$|krj_\kappa(kr)|\leqslant C[kr/(1+kr)]^{\kappa+1}, \tag{A2}$$

$$|krn_\kappa(kr)|\leqslant C[(1+kr)/kr]^{\kappa}, \tag{A3}$$

$$|krj_\kappa(kr) - (kr)^{\kappa+1}/(2\kappa+1)!!|$$
$$\leqslant C(kr)^{\kappa+3}/(1+kr)^2, \tag{A4}$$
$$|krn_\kappa(kr) - (kr)^{-\kappa}(2\kappa-1)!!|$$
$$\leqslant C\,[kr/(1+kr)]^{2-\kappa}. \tag{A5}$$

In the following if we have a vector $f = \binom{f_1}{f_2}$ and estimates $|f_1| \leqslant a_1, |f_2| \leqslant a_2$, then we use the notation $|f| < \binom{a_1}{a_2}$ to denote this fact. We also set $L(kr) = kr/(1+kr)$:

$$|\varphi_\kappa^0(E,r)| \leqslant Ck^{-\kappa}L^\kappa(kr)\begin{pmatrix}1\\ r/(1+kr)\end{pmatrix}, \tag{A6}$$

$$|\psi_\kappa^0(E,r)| \leqslant Ck^\kappa L^{-\kappa}(kr)\begin{pmatrix}1\\1\end{pmatrix}, \tag{A7}$$

$$|\Delta\varphi_\kappa^0(E,r)| \leqslant Ck^2\frac{r^{\kappa+1}(1+r)}{(1+kr)^2}\begin{pmatrix}1\\1\end{pmatrix}, \tag{A8}$$

$$|\Delta\psi_\kappa^0(E,r)| \leqslant Ck^{1+\kappa}L^{1-\kappa}(kr)\begin{pmatrix}1\\ r/(1+kr)\end{pmatrix}. \tag{A9}$$

Combining A_1 and A_2, using $\psi_\kappa^0(m,r) = (2\kappa-1)!!\binom{0}{r^{-\kappa}}$ and that the right-hand side of (2.7) is zero we get

$$A_1 + A_2 = -\Delta\varphi_\kappa^0(E,r)\int_r^\infty (\psi_\kappa^0(m,t))^T V(t)\varphi_\kappa(m,t)dt, \tag{A10}$$

so that by elementary estimates

$$|A_1 + A_2| \leqslant Ck^{1-\kappa}L^{\kappa+1}(kr)\int_r^\infty \frac{|V(t)|}{(1+t)^\kappa}dt\begin{pmatrix}1\\1\end{pmatrix}. \tag{A11}$$

Estimating the third term in (A1) yields

$$|A_3| \leqslant Ck^{1-\kappa}L^{\kappa+1}(kr)\int_0^r \frac{|V(t)|}{(1+t)^{\kappa-1}}dt\begin{pmatrix}1\\1\end{pmatrix}. \tag{A12}$$

Similarly, for A_4 and A_5 we get

$$|A_4| \leqslant Ck^{1-\kappa}L^{\kappa+1}(kr)\int_0^r \frac{|V(t)|}{(1+t)^{\kappa-1}}dt\begin{pmatrix}1\\1\end{pmatrix}, \tag{A13}$$

$$|A_5| \leqslant Ck^{1-\kappa}L^{\kappa+1}(kr)\int_0^r \frac{|V(t)|}{(1+t)^{\kappa}}dt\begin{pmatrix}1\\1\end{pmatrix}. \tag{A14}$$

The entries of the matrix $\varphi_\kappa^0(E,r)(\psi_\kappa^0(E,t))^T - \psi_\kappa^0(E,r)(\varphi_\kappa^0(E,t))^T$ are each bounded in magnitude by

$$CL^{\kappa+1}(kr)L^{-\kappa-1}(kt)(1+t). \tag{A15}$$

So if we set

$$u(E,r) = (|\Delta\varphi_{\kappa,1}(E,r)| + |\Delta\varphi_{\kappa,2}(E,r)|)L^{-\kappa-1}(kr)k^{\kappa-1} \tag{A16}$$

and combine (A11)–(A15), then we arrive at the inequality

$$u(E,r) \leqslant C + C\int_0^r |V(t)|(1+t)u(E,t)dt. \tag{A17}$$

Hence by Gronwall's inequality $u(E,r) \leqslant C$ which is equivalent to (2.18). Part (i) is proved.

(ii) Here, $E \in [-m-\delta, -m], k < 0$. It turns out that the quantity $k^2L^{\kappa+1}(|k|r)$ is not sufficient to control the difference $\Delta\varphi_\kappa(E,r) = \varphi_\kappa(E,r) - \varphi_\kappa(-m,r)$, we must also use $k^{2-\kappa}L^\kappa(|k|r)$. So we introduce

$$h_\kappa(r) = k^{1-\kappa}L^{\kappa+1}(|k|r) + k^{2-\kappa}L^\kappa(|k|r). \tag{A18}$$

Proceeding as in (i) we can then show that

$$|\Delta\varphi_{\kappa,1}(E,r)|$$
$$\leqslant Ck^{2-\kappa}L^\kappa(|k|r) + Ck^{-1}L^\kappa(|k|r)\int_0^r L^{1-\kappa}(|k|t)$$
$$\times|V(t)|\,|\Delta\varphi_{\kappa,1}(E,t)|dt + CL^\kappa(|k|r)$$
$$\times\int_0^r L^{-\kappa}(|k|t)|V(t)||\Delta\varphi_{\kappa,2}(E,t)|dt, \tag{A19}$$

for $\kappa \geqslant 2$, and

$$|\Delta\varphi_{\kappa,1}(E,r)| \leqslant CL^2(|k|r) + Ck^{-1}L(|k|r)$$
$$\times\int_0^r |V(t)||\Delta\varphi_{\kappa,1}(E,t)|dt + CL(|k|r)$$
$$\times\int_\delta^r L^{-1}(|k|t)|V(t)||\Delta\varphi_{\kappa,2}(E,t)|dt, \tag{A20}$$

for $\kappa = 1$. For all $\kappa \geqslant 1$ we have

$$|\Delta\varphi_{\kappa,2}(E,r)| \leqslant Ch_\kappa(r) + CL^\kappa(|k|r)$$
$$\times\int_0^r L^{-\kappa}(|k|t)|V(t)||\Delta\varphi_{\kappa,1}(E,t)|dt$$
$$+ CkL^{\kappa+1}(|k|r)$$
$$\times\int_0^r L^{-\kappa}(|k|t)|V(t)||\Delta\varphi_{\kappa,2}(E,t)|dt. \tag{A21}$$

Now when $\kappa = 1$ we set $u(E,r) = (|\Delta\varphi_{\kappa,1}(E,r)| + |\Delta\varphi_{\kappa,2}(E,r)|)/h_\kappa(r)$ and when $\kappa \geqslant 2$ we set $u(E,r) = |\Delta\varphi_{\kappa,1}(E,r)|/(k^{2-\kappa}L^\kappa(|k|r)) + |\Delta\varphi_{\kappa,2}(E,r)|/h_\kappa(r)$. Then $u(E,r)$ is seen to obey an inequality of the form (A17) and hence (2.19), (2.20), and (2.21) follow immediately. This concludes the proof of Lemma (2.1).

[1] E. C. Titchmarsh, "On the nature of the spectrum in problems of relativistic quantum mechanics," Q. J. Math., Oxford **12**, 227 (1961).
[2] J. Weidmann, "Oszillationsmethoden für Systeme gewöhnlicher Differentialgleichungen," Math. Z. **119**, 349 (1971).
[3] N. Levinson, "On the uniqueness of the potential in a Schrödinger equation for a given asymptotic phase," Mat.-Fys. Medd. **25**, 1 (1949).
[4] M. C. Barthélémy, "Contribution à l'étude de la diffusion par un potentiel central dans la théorie de l'électron de Dirac II," Ann. Inst. Henri Poincaré A **7**, 115 (1967).
[5] Z.-Q. Ma and G. Ni, "Levinson theorem for Dirac particles," Phys. Rev. D **31**, 1482 (1985).
[6] G. Ni, "The Levinson theorem and its generalization in relativistic quantum mechanics," Phys. Energ. Fortis **3**, 432 (1979).
[7] M. Klaus, K. Shaw, and D. Hinton "Levinson's theorem and Titchmarsh-Weyl theory for Dirac systems," Proc. R. Soc. Edinburgh **109A**, 173 (1988).
[8] Z.-Q. Ma, "Levinson's theorem for Dirac particles with a long-range potential," Phys. Rev. D **32**, 2213 (1985).
[9] M. Klaus, "Exact behavior of Jost functions at low energy," J. Math. Phys. **29**, 148 (1988).
[10] Z.-Q. Ma, "Levinson's theorem for Dirac particles moving in a background magnetic monopole field," Phys. Rev. D **32**, 2203 (1985).
[11] Z. R. Iwinski, L. Rosenberg, and L. Spruch, "Nodal structure of zero-energy wave functions: New approach to Levinson's theorem," Phys. Rev. A **31**, 1229 (1985).
[12] Z.-Q. Ma, "Proof of the Levinson theorem by the Sturm–Liouville theorem," J. Math. Phys. **26**, 1995 (1985).
[13] R. P. Boas, *Entire Functions* (Academic, New York, 1954).

[14]W. Hunziker, "On the nonrelativistic limit of the Dirac theory," Comm. Math. Phys. **40**, 215 (1975).
[15]F. Gesztesy, H. Grosse, and B. Thaller, "A rigorous approach to relativistic corrections of bound state energies for spin-1/2 particles," Ann. Inst. Henri Poincaré **40**, 159 (1984).
[16]K. Yajima, "Non relativistic limit of the Dirac theory, scattering theory," J. Fac. Sci. Univ. Tokyo, Sec. IA **23**, 517 (1976).
[17]E. C. Titchmarsh, On the relation between the eigenvalues in relativistic and nonrelativistic quantum mechanics, Proc. R. Soc. London Ser. A **266**, 33 (1962).

The Dirac equation in Robertson–Walker spaces: A class of solutions

M. Kovalyov and M. Légaré
Department of Mathematics, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

(Received 18 November 1988; accepted for publication 16 August 1989)

Exact solutions of the Dirac equation in open and closed Robertson–Walker spaces are presented. A set of massive solutions is given for static metrics. In the case of nontrivial and arbitrary expansion factors, massless solutions are obtained via a conformal transformation. The set of massless solutions in open Robertson–Walker spaces is shown to be complete.

I. INTRODUCTION

Dirac spinor fields in the background of a gravitational field have been the subject of many investigations. In the last two decades, a number of such studies have been devoted to the determination of the renormalized vacuum expectation value of the energy-momentum tensor and the problem of creation of particles in expanding universes.[1–9] As a starting point, a complete set of solutions to the generalized Dirac equation is desirable. For the flat and closed Robertson–Walker (R–W) spaces, sets of exact solutions to the massless and massive Dirac equations have already been found.[1,2,4,9–12] In the following, we present a complete set of masless solutions in open R–W spaces using the Poincaré (or upper half-space) model of the three-dimensional hyperbolic space (H^3_k). Even if massless solutions in R–W spaces generate a conformally trivial case, preventing any creation of particles with expansion, these solutions can still be used to evaluate back reaction effects of quantum spinor fields on the gravitational field.

The sketch of the paper is as follows. In Sec. II, we briefly introduce the generally covariant formulation of the Dirac equation and we present its explicit form for the three classes of R–W metrics (open, flat, and closed). Section III is devoted to the description of the upper half-space model of the three-dimensional hyperbolic space (H^3_k) and to a discussion of its isometry group. In Sec. IV, the Dirac equation in the space $\mathbb{R}\times H^3_k$ with static metric is solved under specific requirements for the general massive case. Then a set of solutions is generated by the action of isometries of H^3_k and massless solutions to the Dirac equation in open R–W spaces are obtained by application of a conformal transformation to the massless spinor determined for the static metric. In Sec. V, the spinor solutions found in $\mathbb{R}\times H^3_k$ are rewritten in terms of spherical coordinates. As a by-product, a transformation effected on the curvature parameter of these solutions will give rise to solutions in closed R–W spaces. We show that the massless solutions in open R–W spaces form a complete set in Sec. VI. Finally, a summary of the results and possible future developments are given in the last section.

II. THE DIRAC EQUATION IN ROBERTSON–WALKER SPACES

The covariant formulation of the Dirac equation in curved spaces is presented in Lichnerowicz[13] and Choquet-Bruhat *et al.*[14] First, we summarize some of the definitions and notations. Let M be a four-dimensional manifold endowed with a hyperbolic metric g of signature $(+---)$. A Dirac spinor field ψ on M is a (C^∞) section of the vector bundle associated to the spin bundle corresponding to (M,g) via the $D^{(1/2,0)}\oplus D^{(0,1/2)}$ representation of SL(2,C). For each class of Robertson–Walker spaces, a spin bundle exists and thus the spinor fields are globally well defined. Moreover, the Levi-Civita connection associated to g (with coefficients denoted by $\Gamma^\mu_{\nu\lambda}$) on the frame bundle over M determines a connection on the spin bundle (spin connection), which then defines a covariant derivative of the spinor field ψ:

$$\nabla_\mu\psi=(\partial_\mu+\Sigma_\mu)\psi, \tag{2.1}$$

where Σ_μ, $\mu=0,1,2,3$, stand for the spin connection coefficients. They take values in a $D^{(1/2,0)}\oplus D^{(0,1/2)}$ representation of the Lie algebra of SL(2,C) and satisfy the following equation:

$$\frac{\partial\tilde\gamma^\mu}{\partial x^\nu}+\Gamma^\mu_{\nu\lambda}\,\tilde\gamma^\lambda(x)+[\Sigma_\nu,\tilde\gamma^\mu(x)]=0. \tag{2.2}$$

The $\tilde\gamma$'s appearing in Eq. (2.2) are constrained by

$$\tilde\gamma^\mu\tilde\gamma^\nu+\tilde\gamma^\nu\tilde\gamma^\mu=2g^{\mu\nu}\mathbf{1}_4. \tag{2.3}$$

With a choice of orthonormal frames $\{e_\alpha\,|\,\alpha=0,1,2,3\}$ on M:

$$e_\alpha=e^\mu_\alpha(x)\frac{\partial}{\partial x^\mu}, \tag{2.4}$$

where $\alpha=0,1,2,3$, the standard Dirac matrices are retrieved:

$$\gamma_\alpha=e^\mu_\alpha(x)\tilde\gamma_\mu(x). \tag{2.5}$$

If $\eta_{\alpha\beta}$ denotes the Minkowski metric, it follows from (2.5) that the γ's obey the relation defining a Clifford algebra on Minkowski space:

$$\gamma_\alpha\gamma_\beta+\gamma_\beta\gamma_\alpha=2\eta_{\alpha\beta}\mathbf{1}_4. \tag{2.6}$$

The covariant form of the Dirac equation in curved space is then written as

$$(i\tilde\gamma^\mu(x)\nabla_\mu-m)\psi=0. \tag{2.7}$$

For instance, let us consider the R–W spaces with their metric expressed in terms of spherical coordinates[15]:

$$\begin{aligned}g&=g_{\mu\nu}dx^\mu\otimes dx^\nu\\&=R^2(t)(dt^2-dr^2-f^2(r)(d\theta^2+\sin^2\theta\,d\phi^2)),\end{aligned} \tag{2.8}$$

 0022-2488/90/010191-08$03.00

where $R(t)$ is the expansion factor, and

$$f(r)=\begin{cases}(\sinh kr)/k, & \text{with } 0\leqslant r<\infty, \quad \text{for the open case with curvature equal to } -k^2,\\ r, & \text{with } 0\leqslant r<\infty, \quad \text{for the flat case,}\\ (\sin kr)/k, & \text{with } 0\leqslant r\leqslant \pi/k, \quad \text{for the closed case with curvature equal to } k^2.\end{cases}$$

If we choose the following set of orthonormal coframes $\{\theta^\alpha\}$:

$$\theta^0=R(t)dt,\quad \theta^1=R(t)dr,$$
$$\theta^2=R(t)f(r)d\theta,\quad \theta^3=R(t)f(r)\sin\theta\, d\phi, \tag{2.9}$$

with

$$e_\alpha(\theta^\beta)=\delta_\alpha^\beta, \tag{2.10}$$

then the spin connection coefficients are given by

$$\Sigma_0=0,$$
$$\Sigma_1=-\tfrac{1}{4}(R'/R)[\gamma^0,\gamma^1],$$
$$\Sigma_2=-\tfrac{1}{4}((R'/R)f[\gamma^0,\gamma^2]+f'[\gamma^1,\gamma^2]),$$
$$\Sigma_3=-\tfrac{1}{4}((R'/R)f\sin\theta[\gamma^0,\gamma^3]-f'\sin\theta[\gamma^1,\gamma^3]$$
$$-\cos\theta[\gamma^2,\gamma^3]), \tag{2.11}$$

where the prime indicates the differentiation of the function with respect to its argument. Substituting this into equation (2.7), we obtain the generalized form of the Dirac equation for the R–W spaces:

$$\left[\gamma^0\left(\frac{\partial}{\partial t}+\frac{3R'}{2R}\right)+imR\,\mathbf{1}_4+\gamma^1\left(\frac{\partial}{\partial r}+\frac{f'}{f}\right)\right.$$
$$\left.+\frac{\gamma^2}{f}\left(\frac{\partial}{\partial\theta}+\frac{\cot\theta}{2}\right)+\frac{\gamma^3}{f\sin\theta}\frac{\partial}{\partial\phi}\right]\psi=0. \tag{2.12}$$

In the massless case, a reduction of the Dirac equations in R–W spaces to the Dirac equations in spaces with static ($R(t)=1$) metric can be achieved with the conformal mapping[16]:

$$g'=R^{-2}(t)g, \tag{2.13a}$$

and

$$\psi'=R^{3/2}(t)\psi. \tag{2.13b}$$

As a result, a solution to the massless Dirac equation in $\mathbb{R}\times H_k^3$ is also a solution to the massless Dirac equation in open R–W spaces up to the above-mentioned conformal factor. Before showing solutions to this reduced equation, we review in the next section some properties of H_k^3 that will be useful.

III. THE THREE-DIMENSIONAL HYPERBOLIC SPACE

We will work with two models of the hyperbolic manifold (H_k^3). One of them, which is more natural, is the geodesic model. In this model, H_k^3 can be viewed as the set of all triplets (r,θ,ϕ) with range $0\leqslant r<\infty$, $0\leqslant\theta\leqslant\pi$, and $0\leqslant\phi<2\pi$, called spherical coordinates. In these coordinates the metric tensor has the form

$$h=dr^2+(\sinh kr/k)^2(d\theta^2+\sin^2\theta\, d\phi^2), \tag{3.1}$$

where k is a positive constant.

In the second model, H_k^3 is described by the upper half-space of the three-dimensional Euclidean space with coordinates (x^1,x^2,y), where $y>0$, endowed with the metric

$$h=((dx^1)^2+(dx^2)^2+dy^2)/k^2y^2. \tag{3.2}$$

Both models are related to each other by the following transformation:

$$x^1=\frac{\sin\theta\cos\phi}{\coth(kr)-\cos\theta},$$
$$x^2=\frac{\sin\theta\sin\phi}{\coth(kr)-\cos\theta},$$

and

$$y=(\cosh(kr)-\sinh(kr)\cos\theta)^{-1}, \tag{3.3}$$

where the origin of the spherical coordinates is mapped to the point (0,0,1) of the upper half-space model.

The group of isometries of H_k^3 is PSL(2,C). It can be realized using quaternionic notation for the upper half-space coordinates (x^1,x^2,y) (Ref. 17):

$$q=x^1\cdot 1+x^2\cdot i+y\cdot j,$$
$$q'=x'^1\cdot 1+x'^2\cdot i+y'\cdot j. \tag{3.4}$$

The action of PSL(2,C) on q corresponds to a fractional linear transformation,

$$q'=(Aq+B)(Cq+D)^{-1}, \tag{3.5}$$

where the matrix $\left[\begin{smallmatrix}A & B\\ C & D\end{smallmatrix}\right]$ belongs to PSL(2,C).

Let us restrict ourselves to the subgroup leaving invariant the point (0,0,1) in the upper half-space model. One verifies that this subgroup of PSL (2,C), which is SO(3), preserves the origin in the geodesic model. It can be described as the SU(2) subgroup of SL(2,C) quotiented by its center, parametrized by

$$\begin{bmatrix} ae^{i\chi} & (e_1+ie_2)b\\ (-e_1+ie_2)b & ae^{-i\chi}\end{bmatrix}\in SU(2), \tag{3.6}$$

where $a^2+b^2=1$, $e_1^2+e_2^2=1$, and $0\leqslant\chi<2\pi$.

Its action on a point of H_k^3 can be written explicitly in terms of the variables x^1, x^2, and y:

$$x'^1=\frac{1}{\Delta}\left[x^1-2(e_1x^1+e_2x^2)e_1+\frac{a}{b}e_1\right]-\frac{a}{b}e_1,$$
$$x'^2=\frac{1}{\Delta}\left[x^2-2(e_1x^1+e_2x^2)e_2+\frac{a}{b}e_2\right]-\frac{a}{b}e_2,$$
$$y'=y/\Delta, \tag{3.7a}$$

where

$$\Delta=(bx^1-ae_1)^2+(bx^2-ae_2)^2+b^2y^2. \tag{3.7b}$$

A further geometrical meaning can be attributed to the transformation (3.7) if we look at the H_k^3 analogs of the Euclidean planes the so-called horospheres, which are defined by the equation

$$(b'y-(c/b'))^2+(b'x^1-a'e_1')^2 + (b'x^2-a'e_2')^2=(c/b')^2, \tag{3.8}$$

where $c>0$, $a'^2+b'^2=1$, and $e_1'^2+e_2'^2=1$.

If we fix a', b', e_1', and e_2', we obtain a one-parameter family of surfaces which we call a family of parallel horospheres with direction $((a'/b')e_1',(a'/b')\,e_2')$. When $b'\neq 0$, this family is composed of the spheres of radius c/b' tangent to the plane $y=0$ at the point $((a'/b')e_1',(a'/b')\,e_2',0)$. If $b'=0$, we have the family of parallel horosheres with direction (∞,∞) which consists of the planes $y=1/2c$ and can be thought of as a family of spheres of infinite radius tangent to the plane $y=0$ at the point $(\infty,\infty,0)$. It can be shown that the transformation (3.7) maps the set of parallel horospheres with direction $((a/b)e_1,(a/b)e_2)$ onto the set of parallel horospheres with direction (∞,∞).

Moreover, for each set of parallel horospheres with direction $\nu=((a/b)e_1,(a/b)e_2)$, we can define a family of geodesics with direction ν by the equations

$$\frac{1}{\Delta}\left[x^1-2(e_1x^1+e_2x^2)e_1+\frac{a}{b}e_1\right]-\frac{a}{b}e_1=C \tag{3.9a}$$

and

$$\frac{1}{\Delta}\left[x^2-2(e_1x^1+e_2x^2)e_2+\frac{a}{b}e_2\right]-\frac{a}{b}e_2=D, \tag{3.9b}$$

where C and D are constants.

The transformation (3.7) also maps the family of geodesics with direction ν onto the family of geodesics with direction (∞,∞).

IV. SOLUTIONS IN $\mathbb{R}\times H_k^3$

In this section, we find explicit solutions to the massive Dirac equations in the space $\mathbb{R}\times H_k^3$ using the upper half-space model of H_k^3. In this model, the nonzero coefficients of the Levi-Civita connection are

$$\begin{aligned}&\Gamma_{13}^1=\Gamma_{23}^2=\Gamma_{33}^3=-1/y,\\ &\Gamma_{11}^3=\Gamma_{22}^3=1/y.\end{aligned} \tag{4.1}$$

We choose the following set of orthonormal frames and coframes on $\mathbb{R}\times H_k^3$:

$$\tilde\theta^0=dt,\quad \tilde\theta^1=\frac{dx^1}{ky},\quad \tilde\theta^2=\frac{dx^2}{ky},\quad \tilde\theta^3=\frac{dy}{ky}, \tag{4.2}$$

with

$$\tilde e_\alpha(\tilde\theta^\beta)=\delta_\alpha^\beta;\quad \alpha,\beta=0,1,2,3. \tag{4.3}$$

From (2.2), we derive the spin connection coefficients in terms of the Dirac matrices $\{\gamma^\alpha\}$:

$$\begin{aligned}&\Sigma_0=\Sigma_3=0,\\ &\Sigma_1=-(1/2y)\gamma^1\gamma^3,\\ &\Sigma_2=-(1/2y)\gamma^2\gamma^3.\end{aligned} \tag{4.4}$$

The generalized Dirac equation (2.7) in $\mathbb{R}\times H_k^3$ can then be written as

$$\left\{\gamma^0\frac{\partial}{\partial t}+ky\gamma^1\frac{\partial}{\partial x^1}+ky\gamma^2\frac{\partial}{\partial x^2} + ky\gamma^3\frac{\partial}{\partial y}-k\gamma^3+im\mathbf{1}_4\right\}\psi=0, \tag{4.5}$$

where, for explicit calculations, we will use the representation of the Dirac matrices given below:

$$\begin{aligned}&\gamma^0=\begin{bmatrix}O_2 & \sigma_2\\ \sigma_2 & O_2\end{bmatrix}\quad \gamma_1=\begin{bmatrix}O_2 & -i\sigma_3\\ -i\sigma_3 & O_2\end{bmatrix};\\ &\gamma^2=\begin{bmatrix}O_2 & -\mathbf{1}_2\\ \mathbf{1}_2 & O_2\end{bmatrix};\quad \gamma^3=\begin{bmatrix}O_2 & i\sigma_1\\ i\sigma_1 & O_2\end{bmatrix},\end{aligned} \tag{4.6}$$

with the σ_i's ($i=1,2,3$) standing for the Pauli matrices.

Let us determine spinor solutions propagating along the y axis. If they stay constant on the horospheres with direction (∞,∞) (corresponding to the y axis), then they can be expressed as

$$\psi=e^{iwt}\phi(y), \tag{4.7}$$

where $\phi(y)$ satisfies the equation

$$ky\gamma^3\frac{\partial\phi}{\partial y}+(iw\gamma^0-k\gamma^3+im\mathbf{1}_4)\phi=0. \tag{4.8}$$

A general solution to (4.8) is

$$\phi(y)=\sum_{a=1}^{4}c_a y^{\alpha_a}v_a, \tag{4.9}$$

where the c_a's are complex constants;

$$\alpha_1=\alpha_2=1+(\sqrt{m^2-w^2})/k,\quad \alpha_3=\alpha_4=1-(\sqrt{m^2-w^2})/k,$$

$$v_1=\begin{bmatrix}0\\ \sqrt{m^2-w^2}-m+iw\\ \sqrt{m^2-w^2}-m-iw\\ 0\end{bmatrix},\quad v_2=\begin{bmatrix}\sqrt{m^2-w^2}-m-iw\\ 0\\ 0\\ \sqrt{m^2-w^2}-m+iw\end{bmatrix},$$

$$v_3=\begin{bmatrix}0\\ -\sqrt{m^2-w^2}-m+iw\\ -\sqrt{m^2-w^2}-m-iw\\ 0\end{bmatrix},\quad v_4=\begin{bmatrix}-\sqrt{m^2-w^2}-m-iw\\ 0\\ 0\\ -\sqrt{m^2-w^2}-m+iw\end{bmatrix}, \tag{4.10}$$

are, respectively, the eigenvalues and eigenvectors of the matrix:

$$\mathbf{1}_4+(iw/k)\gamma^3\gamma^0+(im/k)\gamma^3. \tag{4.11}$$

Thus any solution of (4.5) with direction (∞,∞) can be put in the form

$$\psi(t,y) = e^{i\omega t} \sum_{a=1}^{4} c_a y^{\alpha_a} v_a. \tag{4.12}$$

Let us mention that the plane wave spinor solutions in Minkowski space traveling along the y axis are recovered in the zero limit of the curvature parameter k with the substitution $y + (1/k)$ for y.

In order to obtain spinor solutions moving in the direction $((a/b)e_1,(a/b)e_2)$, we apply the transformation (3.7) that maps the family of horospheres with direction $((a/b)e_1,(a/b)e_2)$ onto the family of horospheres with direction (∞,∞). Since these transformations leave the metric invariant, the solutions to (4.5) in the direction $((a/b)e_1,(a/b)e_2)$ are

$$\Psi(t,x^1,x^2,y) = S^{\dagger}(a,b,e_1,e_2)\psi(t,y'), \tag{4.13}$$

where S represents the $D^{(1/2,0)} \oplus D^{(0,1/2)}$ SL(2,C) representation of the rotation $\mathscr{R}$ of the orthonormal frames induced by (3.7), that is

$$S\gamma^i S^{\dagger} = (\mathscr{R}^T)^i{}_j \gamma^j \quad (i,j = 1,2,3), \tag{4.14}$$

with

$$\mathscr{R} = (y/y')J, \tag{4.15a}$$

$$[\mathscr{R}]^i_j = \begin{bmatrix} 1 - \dfrac{2}{|z|^2+y^2}[(\operatorname{Im} e^* z)^2 + e_1^2 y^2], & \dfrac{-2}{|z|^2+y^2}[e_1e_2((z^2)^2 - (z^1)^2 + y^2) + z^1z^2(e_2^2 - e_1^2)], & \dfrac{-2yz^1}{|z|^2+y^2}, \\ \dfrac{-2}{|z|^2+y^2}[e_1e_2((z^2)^2 - (z^1)^2 + y^2) + (e_1^2 - e_2^2)z^1z^2], & 1 - \dfrac{2}{|z|^2+y^2}[(\operatorname{Im} e^* z)^2 + e_2^2 y^2], & \dfrac{-2yz^2}{|z|^2+y^2}, \\ \dfrac{-2y}{|z|^2+y^2}[(e_2^2 - e_1^2)z^1 - 2e_1e_2z^2], & \dfrac{-2y}{|z|^2+y^2}[(e_1^2 - e_2^2)z^2 - 2e_1e_2z^1], & 1 - \dfrac{2y^2}{|z|^2+y^2} \end{bmatrix}, \tag{4.15b}$$

where $\mathscr{R}^T$ is the transpose of the rotation matrix $\mathscr{R}$, J is the Jacobian of the transformation (3.7), $z^1 \equiv x^1 - (a/b)e_1$, $z^2 \equiv x^2 - (a/b)e_2$, $z \equiv z^1 + iz^2$, and $e \equiv e_1 + ie_2$. Explicitly[18]:

$$S(a,b,e_1,e_2) = \begin{bmatrix} u & O_2 \\ O_2 & u^* \end{bmatrix}, \tag{4.16}$$

with

$$u = \frac{1}{\sqrt{|z|^2+y^2}} \begin{bmatrix} z^*e, & -ye^* \\ ye, & ze^* \end{bmatrix} \in \mathrm{SU}(2). \tag{4.17}$$

Hence, the spinor solutions with direction $((a/b)e_1, (a/b)e_2)$ to the Dirac equation in $\mathbb{R}\times H^3_k$ have the following form:

$$\Psi(t,x^1,x^2,y) = e^{i\omega t} \sum_{a=1}^{4} c_a \left[\frac{y}{\Delta}\right]^{\alpha_a} \begin{bmatrix} u^{\dagger} & O_2 \\ O_2 & u^T \end{bmatrix} v_a. \tag{4.18}$$

From this set of solutions, we obtain massless solutions to the Dirac equation in open R–W spaces by letting $m = 0$ and by applying a conformal transformation (2.13):

$$\Psi^{\mathrm{R-W}}_{m=0} = R^{-3/2}(t)\Psi_{m=0}, \tag{4.19}$$

where $\Psi^{\mathrm{R-W}}_{m=0}$ and $\Psi_{m=0}$ denote, respectively, massless solutions in R–W space and the solutions (4.18) with $m = 0$.

We note that the limit of the curvature parameter k to zero in (4.18) leads to a set of plane wave spinor solutions in Minkowski space. In order to perform this limit, it is more appropriate to go back to the "Ball model" of H^3_k (Ref. 17).

V. SOLUTIONS IN SPHERICAL COORDINATES

For completeness, we present the spinor solutions found for the open R–W spaces in terms of the usual spherical coordinates. A simple modification to these solutions will allow us to introduce a set of solutions to the Dirac equation in closed R–W spaces. We recall that the spherical coordinates are related to the upper half-space coordinates by Eq. (3.3). However, the transformation of the spinor fields requires the SO(3,1) transformation $(\tilde{\Lambda})$ between the two sets of orthonormal coframes:

$$\tilde{\theta}^{\alpha} = \tilde{\Lambda}^{\alpha}_{\beta}\theta^{\beta}, \tag{5.1}$$

with $\tilde{\Lambda}$ given by

$$[\tilde{\Lambda}]^{\alpha}_{\beta} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\eta & -\sin\eta & 0 \\ 0 & \cos\phi\sin\eta & \cos\phi\cos\eta & -\sin\phi \\ 0 & \sin\phi\sin\eta & \sin\phi\cos\eta & \cos\phi \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{(-2\pi/3\sqrt{3})(L_1+L_2+L_3)}e^{\phi L_1}e^{\eta L_3} & & \\ 0 & & & \\ 0 & & & \end{bmatrix}, \tag{5.2}$$

$$\sin\eta = \frac{\sin\theta}{[\cosh kr - \sinh kr\cos\theta]},$$

$$\cos\eta = \frac{\cosh kr\cos\theta - \sinh kr}{[\cosh kr - \sinh kr\cos\theta]}, \tag{5.3}$$

and L_i stands for the generator of rotations around the ith axis ($i=1,2,3$). The generators satisfy the commutation rules of the SO(3) algebra: $[L_i,L_j]=\epsilon_{ijk}L_k$ ($i,j,k=1,2,3$), and $\epsilon_{123}=1$.

With respect to our choice of Dirac matrices (4.6), the $D^{(1/2,0)}\oplus D^{(0,1/2)}$ SL(2,C) representation of $\tilde{\Lambda}$ can be written as

$$\tilde{S}(\tilde{\Lambda}) = \frac{1}{2}\begin{bmatrix} (\mathbf{1}_2 + i(\sigma_1+\sigma_2+\sigma_3))v & O_2 \\ O_2 & (\mathbf{1}_2 - i(\sigma_1-\sigma_2+\sigma_3))v^* \end{bmatrix}, \tag{5.4}$$

where

$$v = e^{-i(\phi/2)\sigma_1}e^{-i(\eta/2)\sigma_3} = \frac{1}{(\cosh kr - \cos\theta\sinh kr)^{1/2}} \times \begin{bmatrix} \left(\cos\frac{\theta}{2} - ie^{kr/2}\sin\frac{\theta}{2}\right)\cos\frac{\phi}{2}, & \left(-i\cos\frac{\theta}{2} + e^{kr/2}\sin\frac{\theta}{2}\right)\sin\frac{\phi}{2}, \\ \left(-i\cos\frac{\theta}{2} - e^{-kr/2}\sin\frac{\theta}{2}\right)\sin\frac{\phi}{2}, & \left(\cos\frac{\theta}{2} + ie^{kr/2}\sin\frac{\theta}{2}\right)\cos\frac{\phi}{2} \end{bmatrix}. \tag{5.5}$$

Consequently the spinor field, defined in terms of the upper half-space coordinates, will undergo the following transformation:

$$\Psi(t,r,\theta,\phi) = \tilde{S}(\tilde{\Lambda})^{\dagger}\Psi(t,x^1,x^2,y). \tag{5.6}$$

It follows that Eq. (5.6) will give rise to solutions of the Dirac equation in spherical coordinates if $\Psi(t,x^1,x^2,y)$ has the form (4.18). The resulting spinor solutions are

$$\Psi(t,r,\theta,\phi) = \frac{e^{i\omega t}}{2}\sum_{a=1}^{4} c_a\left[\frac{1}{\Delta(\cosh kr - \sinh kr\cos\theta)}\right]^{\alpha_a} \times \begin{bmatrix} v^{\dagger}(\mathbf{1}_2 - i(\sigma_1-\sigma_2+\sigma_3))u^{\dagger} & O_2 \\ O_2 & v^{T}(\mathbf{1}_2 + i(\sigma_1+\sigma_2+\sigma_3))u^{T} \end{bmatrix} v_a, \tag{5.7}$$

where Δ and u are expressed in terms of spherical coordinates via Eq. (3.3). With the conformal mapping (2.13b) of (5.7), we get the spherical coordinate representation of massless spinor solutions in open R–W spaces, that is

$$\Psi^{\mathrm{R-W}}_{m=0}(t,r,\theta,\phi) = R^{-3/2}(t)\Psi_{m=0}(t,r,\theta,\phi). \tag{5.8}$$

Let us note that our choice of parameters a, b, e_1 and e_2 for the solutions in the upper half-space model does not lead to simple expressions when transcribed in spherical coordinates. However, a more suitable parametrization is derived if, first, the spinor solution in the direction $\nu=(\infty,\infty)$ is mapped to spherical coordinates by (5.6), and then the spinor is transformed by the SO(3) isometries.

Finally, we observe that changing the parameter k by ik and restricting the variable r to the interval $[0,(\pi/k)]$ in the open R–W metric, we retrieve exactly the closed R–W class. Applying this transformation to (5.8), we get a set of spinor solutions to the Dirac equation in closed R–W spaces. These solutions are well defined everywhere on S^3, except at the poles $r=0,\ \pi/k$. Let us mention that the massless solutions found in Refs. 2, 10, and 11, which can be expressed as Jacobi polynomials in "cos θ" or "cos r," differ from the above.

VI. COMPLETENESS OF THE SET OF MASSLESS SOLUTIONS

In the following, a sketch of the proof of completeness of the massless solutions in open R–W spaces is presented. It is understood that the set of solutions (4.19) with $m=0$ is complete if any solution to the massless equation (4.5) can be expressed as a linear combination of them. For simplicity, we consider the case when the curvature parameter $k=1$. We also ignore the expansion factor $R(t)$, since the conformal map (2.13) preserves the completeness.

As a generalization to hyperbolic space of the spatial Fourier transform for massless spinors, we define

$$\hat{\psi}(t,\beta,\omega) = \frac{1}{16\pi^3}\int_{H^3}\begin{bmatrix} iw(y/\Delta)^{1+iw} & 0 & 0 & 0 \\ 0 & -iw(y/\Delta)^{1-iw} & 0 & 0 \\ 0 & 0 & iw(y/\Delta)^{1+iw} & 0 \\ 0 & 0 & 0 & -iw(y/\Delta)^{1-iw} \end{bmatrix} \times \begin{bmatrix} u(x^1,x^2,y;\beta) & O_2 \\ O_2 & u^*(x^1,x^2,y;\beta) \end{bmatrix}\psi(t,x^1,x^2,y)\frac{dx^1\,dx^2\,dy}{y^3}, \tag{6.1}$$

with the inversion formula

$$\psi(t,x^1,x^2,y)=\int_{\mathbf{R}}\int_{\partial H^3}\begin{bmatrix}u^\dagger(x^1,x^2,y;\beta) & O_2\\ O_2 & u^T(x^1,x^2,y;\beta)\end{bmatrix}\times\begin{bmatrix}-iw(y/\Delta)^{1-iw} & 0 & 0 & 0\\ 0 & iw(y/\Delta)^{1+iw} & 0 & 0\\ 0 & 0 & -iw(y/\Delta)^{1-iw} & 0\\ 0 & 0 & 0 & -iw(y/\Delta)^{1+iw}\end{bmatrix}\hat{\psi}(t,\beta,w)\,\frac{d\beta_1\,d\beta_2\,dw}{(1+|\beta|^2)^2}, \tag{6.2}$$

where $\beta\equiv(a/b)\,e\in\partial H^3$ (boundary of H^3), u is given by (4.17), and $\Delta=(|x-\beta|^2+y^2)/(1+|\beta|^2)$(3.7b), with $x\equiv x^1+ix^2$.

In this representation, we can write the general solution to Eq. (4.5) with zero mass as

$$\psi(t,x^1,x^2,y)=\int_{\mathbf{R}}\int_{\partial H^3}e^{iwt}\begin{bmatrix}u^\dagger(x^1,x^2,y;\beta) & O_2\\ O_2 & u^T(x^1,x^2,y;\beta)\end{bmatrix}\times\begin{bmatrix}-iw(y/\Delta)^{1-iw} & 0 & 0 & 0\\ 0 & iw(y/\Delta)^{1+iw} & 0 & 0\\ 0 & 0 & -iw(y/\Delta)^{1-iw} & 0\\ 0 & 0 & 0 & iw(y/\Delta)^{1+iw}\end{bmatrix}\hat{\psi}(0,\beta,w)\,\frac{d\beta_1\,d\beta_2\,dw}{(1+|\beta|^2)^2}, \tag{6.3}$$

where $\hat{\psi}(0,\beta,w)$ stand for the "Fourier transform" of the initial data.

The consistence of equations (6.1) and (6.2) follows from the formula

$$\begin{aligned}&\int_{-\infty}^{+\infty}\int_{\partial H^3}u^\dagger(x^{1\prime},x^{2\prime},y',\beta)\\&\quad\times\begin{bmatrix}(y'/\Delta')^{1-iw}(y/\Delta)^{1+iw} & 0\\ 0 & (y'/\Delta')^{1+iw}(y/\Delta)^{1-iw}\end{bmatrix}\\&\quad\times u(x^1,x^2,y;\beta)\,\frac{w^2\,dw\,d\beta_1\,d\beta_2}{(1+|\beta|^2)^2}\\&=16\pi^3y^3\delta(x-x',y-y')\mathbf{1}_2.\end{aligned} \tag{6.4}$$

In order to verify this identity, we first split the w dependent matrix in its w even and w odd parts, that is

$$\begin{aligned}&\begin{bmatrix}(y'/\Delta')^{1-iw}(y/\Delta)^{1+iw} & 0\\ 0 & (y'/\Delta')^{1+iw}(y/\Delta)^{1-iw}\end{bmatrix}\\&=\frac{1}{2}\left[\left(\frac{y'}{\Delta'}\right)^{1-iw}\left(\frac{y}{\Delta}\right)^{1+iw}+\left(\frac{y'}{\Delta'}\right)^{1+iw}\left(\frac{y}{\Delta}\right)^{1-iw}\right]\mathbf{1}_2\\&\quad+\frac{1}{2}\left[\left(\frac{y'}{\Delta'}\right)^{1-iw}\left(\frac{y}{\Delta}\right)^{1+iw}\right.\\&\quad\left.-\left(\frac{y'}{\Delta'}\right)^{1+iw}\left(\frac{y}{\Delta}\right)^{1-iw}\right]\sigma_3.\end{aligned} \tag{6.5}$$

Substituting (6.5) in (6.4), the w odd part vanishes once integrated over w, leaving only the w even contribution. The residual integral can be reduced to the form

$$\begin{aligned}&\int_{-\infty}^{+\infty}\int_{\partial H^3}u^\dagger(x^{1\prime},x^{2\prime},y';\beta)\,u(x^1,x^2,y;\beta)\\&\quad\times\left[\left(\frac{y'}{\Delta'}\right)^{1-iw}\left(\frac{y}{\Delta}\right)^{1+iw}\right]\frac{w^2\,dw\,d\beta_1\,d\beta_2}{(1+|\beta|^2)^2}\\&=16\pi^3\delta(x-x',\,y-y')\mathbf{1}_2.\end{aligned} \tag{6.6}$$

Since the identity (6.6) is invariant with respect to the action of any isometry of H^3 on $(x^{1\prime},x^{2\prime},y')$ and (x^1,x^2,y), we can put without loss of generality: $x^1=x^2=x^{1\prime}=x^{2\prime}=0$ and $y'=1$. In other words, the group of H^3 isometries can be used to map $(x^{1\prime},x^{2\prime},y')$ to $(0,0,1)$ and (x^1,x^2,y) to a point on the positive y-axis, relabelled $(0,0,y)$. As a consequence, (6.6) is simplified to

$$\begin{aligned}&\int_{-\infty}^{+\infty}\int_{\partial H^3}u^\dagger(0,0,1;\beta)\,u(0,0,y;\beta)\\&\quad\times\left[\frac{y(1+|\beta|^2)}{y^2+|\beta|^2}\right]^{1+iw}\frac{w^2\,dw\,d\beta_1\,d\beta_2}{(1+|\beta|^2)^2}\\&=16\pi^3y^3\delta(x,y-1)\mathbf{1}_2.\end{aligned} \tag{6.7}$$

We obtain from (4.17) that

$$\begin{aligned}&u^\dagger(0,0,1;\beta)\,u(0,0,y;\beta)\\&=\frac{1}{\sqrt{(|\beta|^2+1)(|\beta|^2+y^2)}}\\&\quad\times\begin{bmatrix}|\beta|^2+y & -\beta^*(1-y)\\ \beta(1-y) & |\beta|^2+y\end{bmatrix}.\end{aligned} \tag{6.8}$$

The integration over the parameters β_1 and β_2 can now be carried out. Inserting (6.8) in (6.7) and passing to polar coordinates: $\beta=\beta_1+i\beta_2=|\beta|\,e^{i\gamma}$, where $0\leq\gamma<2\pi$, we find that the off-diagonal integrals over the angular variable γ vanish while the diagonal contribution becomes

$$2\pi\int_{-\infty}^{+\infty} dw\, w^2 \int_0^{\infty} \frac{d|\beta|}{(1+|\beta|^2)^2} \frac{|\beta|(|\beta|^2+y)}{\sqrt{(|\beta|^2+1)(|\beta|^2+y^2)}}$$

$$\times\left[\frac{y(1+|\beta|^2)}{y^2+|\beta|^2}\right]^{1+iw} \mathbf{1}_2$$

$$= 16\pi^3 y^3 \delta(x,y-1)\mathbf{1}_2. \tag{6.9}$$

We evaluate the second integral on the left-hand side of (6.9) using the following substitution:

$$|\beta| = \sqrt{(1+t)/(1-t)}, \tag{6.10}$$

where $-1 \leqslant t \leqslant 1$.

After integration, Eq. (6.9) is transformed to

$$\frac{2\pi y^{3/2} I(y)}{(1+y)^2(1-y)} = 16\pi^3 y^3 \delta(x,y-1), \tag{6.11}$$

where

$$I(y) = \int_{-\infty}^{+\infty} dw\, w^2 \,\mathrm{Re}\left\{\frac{y^{-1/2+iw} + y^{1/2-iw}}{1/2 - iw}\right\}. \tag{6.12}$$

The change of variables, $y = e^{\xi}$, where $\xi \in \mathbb{R}$, allows us to rewrite (6.11) as

$$(-1/4 \sinh\xi \cosh(\xi/2))\, I(\xi) = 8\pi^2 y^3 \delta(x,y-1), \tag{6.13}$$

where

$$I(\xi) = \int_{-\infty}^{+\infty} \left(\frac{e^{i\xi(w+(i/2))} - e^{-i\xi(w+(i/2))}}{-i(w+(i/2))}\right) w^2\, dw. \tag{6.14}$$

Differentiating $I(\xi)$, we find that

$$I'(\xi) = 4\pi \cosh(\xi/2)\delta''(\xi). \tag{6.15}$$

A symbolic expression for $I(\xi)$ can be derived with help of the following equalities[19]:

$$\cosh(\xi/2)\delta''(\xi) = \delta''(\xi) \tag{6.16}$$

and

$$\cosh(\xi/2)\delta'(\xi) = \delta'(\xi). \tag{6.17}$$

In fact, we get from Eqs. (6.15) and (6.16):

$$I'(\xi) = 4\pi\delta''(\xi). \tag{6.18}$$

Using (6.17), the integration of (6.18) gives rise to

$$I(\xi) = 4\pi \cosh(\xi/2)\delta'(\xi), \tag{6.19}$$

where the constant of integration vanishes since $I(\xi)$ and $\delta'(\xi)$ are both odd distributions.

Finally, inserting this last result in the identity (6.13), we arrive to

$$-\delta'(\xi)/\sinh\xi = 8\pi y^3 \delta(x,y-1). \tag{6.20}$$

The validity of this identity is shown below. It enables us to conclude that (6.4) is verified, and consequently, that (6.1) and (6.2) are consistent.

In order to justify (6.20), we use spherical coordinates. First, we note that $|\xi|$ can be identified with the variable r in (3.3), which is the geodesic distance between (0,0,1) and (x^1,x^2,y). The identity (6.20) is then confirmed by verification of the next formula for any test function f, expressed in both the upper half-space coordinates (x^1,x^2,y) and the spherical coordinates (r,θ,ϕ):

$$-8\pi\int y^3\delta(x,y-1) f(x^1,x^2,y)\frac{dx^1\,dx^2\,dy}{y^3}$$

$$= 2\int_0^{\infty} \sinh^2 r\, dr \int \sin^2\theta\, d\theta\, d\phi\, f(r,\theta,\phi)\,\frac{\delta'(r)}{\sinh r}. \tag{6.21}$$

VII. SUMMARY

We were concerned in this paper with Dirac fields in the background of R–W metrics. In the case of open R–W spaces, a set of explicit massless solutions to the Dirac equation has been found. Each solution propagates along the direction defined by a set of parallel horospheres, the analogs of the planes in flat space. For this purpose, we have used the upper half-space representation of the three-dimensional spatial submanifold. In the limit of zero curvature, one can show that spinor plane-wave solutions in Minkowski space are retrieved. We also expressed the open R–W spinor solutions in terms of spherical coordinates. It follows that the substitution of the parameter k by its imaginary form and the restriction of the domain of the spatial variable r lead to closed R–W spinor solutions. Finally, it has been verified that the set of spinor massless solutions presented for the open class is complete, a property which is certainly very important for quantum field theoretical considerations.

Let us recall that only massless solutions have been determined for metrics with nonconstant and arbitrary expansion factor $R(t)$, since a conformal map of the static metric was carried out. It would be interesting in a future investigation to exhibit massive spinor solutions in R–W spaces with nontrivial expansion factors, as worked out for the flat case in Ref. 12.

ACKNOWLEDGMENT

It is a pleasure to thank Dr. J. So for his assistance.

This research was supported by grants from the Natural Science and Engineering Research Council of Canada.

[1] L. Parker, Phys. Rev. D **3**, 346 (1971).
[2] L. H. Ford, Phys. Rev. D **14**, 3304 (1976).
[3] G. Schäfer and H. Dehnen, Astron. Astrophys. **54**, 823 (1977).
[4] J. Audretsch and G. Schäfer, J. Phys. A: Math. Gen. **11**, 1583 (1978).
[5] A. A. Grib, S. G. Mamayev, and V. M. Mostepanenko, Fortsch. Phys. **28**, 173 (1980).
[6] M. Castagnino, L. Chimento, D. D. Harari, and C. A. Núñez, J. Math. Phys. **25**, 360 (1984).
[7] A. H. Najmi and A. C. Ottewill, Phys. Rev. D **30**, 2573 (1984).
[8] L. P. Chimento and M. S. Mollerach, Phys. Rev. D **34**, 3689 (1986).
[9] C. J. Isham and J. E. Nelson, Phys. Rev. D **10**, 3226 (1974).
[10] E. Schrödinger, Comment. Pontif. Acad. Sci. **2**, 321 (1938).
[11] E. Schrödinger, Proc. R. Ir. Acad. Ser. A **46**, 25 (1940).
[12] A. O. Barut and I. H. Duru, Phys. Rev. D **36**, 3705 (1987).
[13] A. Lichnerowicz, Bull. Soc. Math. France **92**, 11 (1964).
[14] Y. Choquet-Bruhat, C. De Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, New York, 1982).
[15] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[16] Y. Choquet-Bruhat and D. Christodoulou, Ann. Sci. École Norm. Sup., 4ème série, t. 14, 481 (1981).
[17] A. F. Beardon, "The Geometry of Discrete Groups," *Graduate Texts in Mathematics*, Vol. **91** (Springer–Verlag, New York, 1983).

[18]The spinor transformation S has been determined from the normal eigenvector $(\vec{n})$ corresponding to the eigenvalue 1 and the two other eigenvalues $(e^{\pm i\alpha})$ of $\mathscr{R}$. The canonical homomorphism between SU(2) and SO(3) is then used to arrive at $u(=\cos(\alpha/2)-\vec{in}\cdot\vec{\sigma}\sin(\alpha/2))$.

[19]In the formulas (6.16) and (6.17), the absolute value of the variable ξ will later be interpreted as the spherical coordinate r introduced in Sec. III. If we consider test functions $f(\xi)$, (6.16) and (6.17) are proved by showing that the following relations are satisfied, respectively,

$$\int_{-\infty}^{+\infty} f(\xi)\cosh\frac{\xi}{2}\,\delta''(\xi)\sinh^2\xi\,d\xi=\int_{-\infty}^{+\infty} f(\xi)\delta''(\xi)\sinh^2\xi\,d\xi$$

and

$$\int_{-\infty}^{+\infty} f(\xi)\left(\frac{\delta'(\xi)}{\sinh\xi\cosh(\xi/2)}\right)\sinh^2\xi d\xi=\int_{-\infty}^{+\infty} f(\xi)\left(\frac{\delta'(\xi)}{\sinh\xi}\right)\sinh^2\xi d\xi.$$

An extensions theory setting for scattering by breathing bag

Yu. A. Kuperin, K. A. Makarov,[a] and B. S. Pavlov
Leningrad State University, Leningrad, Union of Soviet Socialist Republics

(Received 5 October 1988; accepted for publication 21 June 1989)

A model of the scattering of a structureless pointlike particle on a spherical bag with an internal structure, imitating "quark" degrees of freedom, is considered. It is assumed that the bag is a dynamical object and its radius plays the role of additional dynamical variable. The energy of the collision is distributed among the quark excitations and the vibration degrees of freedom of the bag surface. In the frame of the theory of extensions the Hamiltonian of the coupled bag-quark system interacting with the pointlike particle is constructed. The formal multichannel S matrix of the problem is obtained.

I. INTRODUCTION

Let us consider two conservative quantum systems Q_1 and Q_2 with Hamiltonians H_1 and H_2 acting in Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$, respectively. We assume that the system Q_1 consists of two noninteracting subsystems Q^b and Q^q. Let the Hamiltonian H_1 of the system Q_1 be a direct sum of the corresponding Hamiltonians H^b and H^q, which determine independent evolutions of the subsystems: $H_1 = H^b \oplus H^q$ in the Hilbert space $\mathscr{H}_1 = \mathscr{H}^b \oplus \mathscr{H}^q$. The total Hamiltonian H_0 of the system $Q_1 \cup Q_2$ is represented as a tensor sum of the operators H_i, $i = 1,2$,

$$H_0 = H_1 \otimes I_2 + I_1 \otimes H_2. \tag{1}$$

Because of the decomposition $\mathscr{H}_1 = \mathscr{H}^b \oplus \mathscr{H}^q$, the relation (1) can be written in the form

$$\begin{aligned} H_0 &= (H^b \otimes I_2 + I_b \otimes H_2) \\ &\quad \oplus (H^q \otimes I_2 + I_q \otimes H_2) \\ &\equiv L_1 \oplus L_2. \end{aligned} \tag{2}$$

One could switch on an interaction between the subsystem in the composite quantum system $Q_1 \cup Q_2$, as was done in Ref. 1. The method of the paper[1] deals with the situation in which the total Hilbert space is a direct sum of its subspaces. An interaction arises as a result of construction of a set of self-adjoint extensions of some symmetric operator. This symmetric operator is the result of the restriction of the direct sum of initial Hamiltonians $L_1 \oplus L_2$. This set, in particular, contains the initial Hamiltonian of the composite system. Any other self-adjoint extension can be interpreted as the total Hamiltonian determining the coupled dynamics of the subsystems and the interaction between them.

In this paper we use the techniques of Refs. 1–4 for the operators L_i, $i = 1,2$, of special structure (2), i.e., in the case when the total energy of the subsystems Q^b, Q^q, and Q_2 is the sum of the subsystem energies. We consider the scattering of a pointlike particle on a dynamical quark bag[5] (DQB) as an illustration of the scheme described above.

In this interpretation the operator H^b determines the dynamics of the DQB surface, H^q corresponds to the internal (quark) degrees of freedom, and H_2 is the operator of kinetic energy of the pointlike particle. We construct an interaction between the abovementioned degrees of freedom on the base of the extensions theory.[1–4] It will be shown that this approach leads to a nontrivial scattering matrix.

II. MODEL

Let us assume that the dynamics of the bag surface is given by the self-adjoint operator acting in the Hilbert space $\mathscr{H}^b = L_2(0,\infty)$[5],

$$H^b\chi = \left(-\frac{1}{2M}\frac{d^2}{dR^2} + V(R)\right)\chi, \tag{3}$$

with boundary condition

$$\chi(0) = 0. \tag{4}$$

To simulate the breathing character of the system one should choose the potential $V(R)$ with confinement: $V(R) \to \infty$ as $R \to \infty$.

In this case the Hamiltonian H^b has the pure discrete spectrum $\sigma(H^b) = \{\epsilon_n\}_{n=0}^{\infty}$ and its eigenfunctions form the basis in the Hilbert space $\mathscr{H}^b$.

The dynamics of quark degrees of freedom inside the DQB is determined by a self-adjoint operator H^q, acting in an abstract Hilbert space $\mathscr{H}^b$. The nature of the Hamiltonian H^b can be treated in different ways. It might be understood as the few-body Schrödinger operator with confining potentials (if one needs a nonrelativistic description); or as some relativistic Hamiltonian. Since there is no adequate mathematical description of multiquark dynamics at distances of the order of the confinement, we use the information about the structure of the spectrum of the Hamiltonian H^q only, omitting concrete details of the evolution of internal (quark) degrees of freedom. In this case the bound states of H^q can be treated as quark excitations of the DQB.

Finally, the role of the operator H_2 is played by the Hamiltonian of the free particle

$$H_2 u = -\frac{1}{2m}\frac{d^2}{dr^2}u, \quad r \geqslant 0, \tag{5}$$

$$u(0) = 0, \tag{6}$$

in the Hilbert space $L_2(0,\infty)$. Let us assume that the interaction between the particle and DQB switches on if and only if the coordinates R and r are equal, and the pointlike particle does not penetrate into the bag. As a consequence, the configuration space $\mathscr{R}^2_+$ is divided by the line $y = \alpha x$ into two sectors $V^{\pm} = \{y \gtrless \alpha x\}$, where $\alpha = (m/M)^{1/2}$,

[a] Present address: Freie Universität Berlin, Berlin (West), Germany.

 0022-2488/90/010199-03$03.00

$x = \sqrt{2M}R, y = \sqrt{2mr}$, and the scattering should be considered in "physical" sector V^+ only.

The condition of nonpenetrability into the bag leads to the following rather simple model of an interaction between the quantum systems Q^b and Q_2. One needs to solve the Schrödinger equation in the physical sector V^+:

$$\left(-\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}+V(x)-E\right)\psi(x,y)=0, \tag{7}$$

with the boundary conditions

$$\psi(x,y)|_{y=\alpha x}=0, \quad \psi(x,0)=0. \tag{8}$$

To include the interactions between the systems Q^b, Q^q, and Q_2 in accordance with Refs. 1–4, let us restrict the domain of the operator L_1 to the set of all smooth functions in the sector V^+, which vanish at the vicinity of the line $y=\alpha x$. The symmetric operator L_{10}, restricted in such a way, has infinite deficiency indices. The boundary form[1-4] of its adjoint operator L_{10}^* can be calculated as follows:

$$\langle L_{10}^*\psi,\varphi\rangle-\langle\psi,L_{10}^*\varphi\rangle=\int_\gamma(\partial_n\psi\,\bar{\varphi}-\partial_n\bar{\varphi}\,\psi)d\gamma, \tag{9}$$

where ∂_n denotes the normal derivative of the functions ψ,φ on the curve γ. [Of course, the expression for the boundary form (9) on the domain $\mathscr{D}(L_{10}^*)$ needs some regularization. But for further purposes it is sufficient to consider the boundary form (9) only on the class of smooth functions, which do not satisfy any boundary condition on the line γ. In this case one does not need any regularization and (9) holds in the ordinary sense.]

The next step of the scheme[1-4] is the restriction of the operator L_2 to some symmetric operator L_{20}. Since the operator $L_2=H_q+H_2$ acts in the tensor product of the Hilbert spaces $\mathscr{H}^q\otimes\mathscr{H}_2$ we shall restrict the domain of the operator H^q only, so that the fixed elements θ from $\mathscr{H}^q$ be the deficiency element for the restriction H_0^q,

$$\mathscr{D}(H_0^q)=\{(H^q-iI)^{-1}\psi,\psi\in\mathscr{H}^q\ominus\theta\}. \tag{10}$$

We call the coefficients $\epsilon_n^\pm$ of the decomposition

$$u=u_0+\epsilon_n^+H^q(H^q-iI)^{-1}\theta+\epsilon_n^-(H^q-iI)\theta,$$
$$u_0\in\mathscr{D}(H_0^q), \quad u\in\mathscr{D}(H_0^{q*}) \tag{11}$$

of arbitrary element u from the domain of the adjoint operator H_0^{q*}, the boundary values. In a general case, the boundary values $\epsilon^\pm$ of the elements u from $\mathscr{D}(L_{20}^*)$ become the functions of the variable y and the boundary form of the operator L_{20}^* is given by

$$\langle L_{20}^*u,v\rangle-\langle u,L_{20}^*v\rangle=\int_{R^+}\left[\epsilon_u^-(y)\,\overline{\epsilon_v^+(y)}-\epsilon_v^-(y)\,\overline{\epsilon_u^+(y)}\right]dy, \tag{12}$$

In order to construct a self-adjoint extension H of the operator $L_{10}\oplus L_{20}$, in accordance with our general method, one should impose on the line γ such boundary conditions that nullify the sum of the boundary forms (9) and (12). One of the simplest possibilities to do it is to study the following boundary conditions, mixing all the channels together:

$$\partial_n\psi|_\gamma=\beta\epsilon^-(y), \quad \epsilon^+(y)=\beta\psi|_\gamma, \quad \beta\in R^1. \tag{13}$$

III. SCATTERING PROBLEM

The spectral analysis of the total two-channel Schrödinger equation

$$(H-E)\psi=0 \tag{14}$$

can be reduced to some boundary-value problem in the sector V^+. More precisely, it can be shown[4] that the boundary values $\epsilon^\pm$ on the solutions of the problem (14) are connected with the relation

$$\epsilon^-(y)=Q\left(E+\frac{d^2}{dy^2}\right)\epsilon^+(y). \tag{15}$$

Here $Q(E+d^2/dy^2)$ is the integral operator acting in the space $L^2(R_+^1)$ with the kernel $Q(y,y',z)$,

$$Q(y,y',z)=\frac{1}{2\pi}\oint_\Gamma\Delta(\zeta)r_0(y-y',z-\zeta)d\zeta, \tag{16}$$

where $r_0(u)=(H_2-z)^{-1}$ is the resolvent of the operator $-d^2/dy^2$ with boundary conditions $\psi|_{y=0}=0$, $\Delta(\zeta)$ is the Schwartz integral of the spectral measure of the operator H^q, and the counter Γ encircles the spectrum of the Hamiltonian H^q in the complex plane (ζ).

Equality (15) reduces the problem (14) to the search of the components of the wave function Ψ in the space $\mathscr{H}^q\otimes\mathscr{H}_2$ as the solutions of the following boundary value problem in the sector V^+:

$$\left(-\frac{\partial^2}{dx^2}-\frac{\partial^2}{dy^2}+V(x)-E\right)\psi=0, \tag{17}$$

with energy-dependent boundary conditions

$$\partial_n\psi|_\gamma=\beta^2Q\left(E+\frac{d^2}{dy^2}\right)\psi|_\gamma, \tag{18}$$

$$\psi|_{x=0}=0, \tag{19}$$

and appropriate asymptotic conditions at infinity. All the information about internal degrees of freedom penetrates into the boundary-value problem (17)–(19) through the Schwartz integral $\Delta(\zeta)$ of the Hamiltonian H^q,

$$\Delta(\zeta)=\int\frac{\lambda_\zeta+1}{\lambda-\zeta}d\langle E_\lambda\theta,\theta\rangle, \tag{20}$$

$$H^q=\int\lambda\,dE_\lambda^q. \tag{21}$$

Here $\theta\in\mathscr{H}^q$ is a vector parameter of the theory [see (10)]. The function $\Delta(\zeta)$ is an analytical function with a positive imaginary part in the upper half-plane of the complex variable ζ. If the Hamiltonian H^q is a finite-dimensional self-adjoint operator with a simple spectrum, λ_s are its eigenvalues and E_s are the corresponding orthogonal eigenprojectors, then $\Delta(\zeta)$ is a rational function,

$$\Delta(\zeta)=\sum_s\frac{\zeta\lambda_s+1}{\lambda_s-\zeta}\langle E_s\theta,\theta\rangle_{\mathscr{H}^q}. \tag{22}$$

The formal solution[6] of the problem (17)–(19) can be represented in the form

$$\psi_m(x,y)=\sum_n\{\delta_{mn}e^{-ik_ny}-S_{mn}e^{ik_ny}\}\chi_n(x), \tag{23}$$

where $k_n=\sqrt{E-\epsilon_n}$ are the reduced channels momenta.

Substitution (23) into boundary conditions (18) yields

the following set of equations:

$$\sum_n \{\delta_{mn} e^{-ip_n x}(\alpha \partial_x + gA_n^- + ik_n) - S_{mn} e^{ip_n x} \times (\alpha \partial_x + gA_n^+ - ik_n)|\chi_n(x)\rangle\} = 0, \quad (24)$$

where $p_n = \alpha k_n$ are renormalized channels momenta and $g = \beta^2\sqrt{1+\alpha^2}$. To determine the operators A_n^+ one needs to calculate the action of the integral operator $Q(E + d^2/dy^2)$ on the elements $\psi_n^\pm(x,y) = e^{\pm k_n y}\chi_n(x)$ considered as the functions of the only variable $y = \alpha x$,

$$Q\left(E + \frac{d^2}{dy^2}\right)\psi_n^\pm(y/\alpha,\gamma) = e^{\pm ip_n x} Q\left(E + \left(\frac{1}{\alpha}\partial_x \pm ik_n\right)^2\right)\chi_n(x) = e^{\pm ip_n x}\frac{1}{2\pi i}\oint_\Gamma \Delta(\xi)\varphi_n^\pm(x, E-\xi)\,d\xi \equiv -e^{\pm ip_n x} A_n^\pm|\chi_n\rangle. \quad (25)$$

Here the functions $\varphi_n^\pm(x,\lambda)$ are the L^2 solutions of the boundary-value problem

$$[(1/\alpha)\partial_x \pm ik_n]^2\varphi_n^\pm + \lambda\varphi_n^\pm = \chi_n, \quad \varphi_n^\pm|_{x=0} = 0. \quad (26)$$

If the channel n is opened, $\operatorname{Im} k_n = 0$, then $\varphi_n^+ = \overline{\varphi_n^-}$.

Formally, the S matrix can be determined by projecting the system (24) on the basis $\{\chi_n\}$. In this case the system (24) can be written in terms of the matrices $W^\pm$, playing the role of the wave operators

$$W_{mn}^\pm = \langle\chi_m|e^{\pm ip_n x}(\alpha\partial_x + gA_n^\pm \mp ik_n)|\chi_n\rangle \quad (27)$$

in the following way:

$$W^+ S = W^-. \quad (28)$$

On the opened channels we obviously have $(W_{mn}^+)^* = W_{mn}^-$.

IV. RESUME

The next step is the studying of analytical properties of the suggested S matrix

$$S = (W^+)^{-1}W^-. \quad (29)$$

It will be done anywhere. Let us note here the following circumstances.

(1) The scheme of the exclusion of the channels $\mathscr{H}^b$ and $\mathscr{H}^q$ described above demonstrates that the scattering on DQB can be reduced to some effective matrix-many-channel problem in the space $\mathscr{H}_2$,

$$\left(-\frac{d^2}{dy^2}\otimes I + B - E\right)\Phi = 0, \quad (30)$$

with energy-dependent boundary conditions

$$\Phi^{-1}\Phi'|_{y=0} = \mathscr{P}(E). \quad (31)$$

Here B is a diagonal matrix of thresholds

$$B = \operatorname{diag}\{\epsilon_n\}, \quad (32)$$

$$P(E) = -iK^{1/2}(W^+ + W^-)^{-1}(W^+ - W^-)K^{1/2} \quad (33)$$

is a P matrix[7] of the problem,

$$K = \operatorname{diag}\{k_n\}. \quad (34)$$

(2) In the opened-channel approximation we can use the finite set of equations

$$\sum_{n=1}^N W_{mn}^+ S_{nl} = W_{ml}^-. \quad (35)$$

Denoting by Δ^+ the determinate of the matrix W^+,

$$\Delta^+ = \det W^+ = \Delta^+(k_1,k_2,\ldots,k_n),$$

we find

$$S_{ml} = (1/\Delta^+)\Delta^+(k_1,k_2,\ldots,k_{l-1},k_m,k_{l+1},\ldots,k_n).$$

Here we treat the channel momenta k_n as independent variables and consider matrix element S_{ml} as the function of these variables.

The studying of analytical properties of the S matrix can be reduced to the investigation of corresponding Fredholm determinants Δ^+ (Ref. 8).

Nevertheless the most interesting question is to study the limit $N \to \infty$ and analytical behavior of the the total S matrix.

In conclusion, we have demonstrated that the new method of coupling of different degrees of freedom in the system proposed, leads to a rather interesting scattering problem here. For more realistic simulation, the peripheral interaction in the Hamiltonian H_2, as well as higher partial waves in the corresponding channel, should be taken into account.

ACKNOWLEDGMENTS

We wish to thank F. Lenz for fruitful discussions. One of the authors (K. A. M.) is indebted to the Fachbereich Physik for hospitality received during his stay at the Freie Universität Berlin where this work was finished.

[1] B. S. Pavlov, Teor. Mat. Fiz. **59**, 346 (1984).
[2] Yu. A. Kuperin, K. A. Makarov, and B. S. Pavlov, Teor. Mat. Fiz. **63**, 78 (1985); **69**, 100 (1986).
[3] Yu. A. Kuperin, K. A. Makarov, and Yu. B. Mel'nikov, Teor. Mat. Fiz. **74**, 103 (1988).
[4] Yu. A. Kuperin, K. A. Makarov, S. P. Merkuriev, B. S. Pavlov, and A. K. Motovilov, ITP Budapest Report No. 441, 1986.
[5] Y. Nogami and L. Tomio, Can. J. Phys. **62**, 260 (1984); G. E. Brown, J. W. Durso, and M. B. Johnson, Nucl. Phys. A **397**, 447 (1983).
[6] F. Lenz, J. T. Londergan, E. J. Moniz, R. Rosenfelder, M. Stingl, and K. Yazaki, Ann. Phys. **170**, 65 (1986).
[7] I. M. Narodetskii, in *Few-Body Problem in Physics*, edited by L. D. Faddeev and T. I. Kapaleishvili (World Scientific, Singapore, 1985), Vol. 55.
[8] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

Kepler-type dynamical symmetries of long-range monopole interactions

B. Cordani
Instituto Matematico dell' Università, Via Saldini 50, I-20133 Milano, Italy

L. Gy. Fehér
Bolyai Institute, Jate, H-6720 Szeged, Hungary

P. A. Horváthy
Departement de Mathématiques, Fac. de Science, Université, F-84 Avignon, France and Centre de Physique Théorique Laboratoire Propre da CNRS Luminy, Case 907, F-13288 Marseille Cedex 9, France

(Received 25 April 1989; accepted for publication 16 August 1989)

A general framework for understanding Kepler-type dynamical symmetries is presented. The main concern is the geodesic motion in Euclidean Taub–NUT space, which approximates the scattering of self-dual monopoles for long distances. Other examples include a test particle moving in the asymptotic field of a self-dual monopole and two other related metrics.

I. INTRODUCTION

In this paper we present a general method for understanding Kepler-type dynamical symmetries. Our main interest lies in explaining those symmetries found recently in the long-distance limit of monopole–monopole scattering,[1–4] as well as for a test particle in the asymptotic field of a self-dual monopole.[5]

In the long-distance limit, the relative motion of two monopoles is approximately described in fact by the geodesics of the Euclidean Taub–NUT space of parameter $m = -\frac{1}{2}$, with the line element

$$ds^2 = (1+4m/r)(dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)) + [(4m)^2/(1+4m/r)](d\psi + \cos\theta\, d\theta)^2. \quad (1.1)$$

For $m>0$, (1.1) is just (the space part of) the line element of the celebrated Kaluza–Klein monopole of Gross and Perry and Sorkin.[6] The problem of geodesic motion in this metric therefore has its own interest independent of monopole scattering. The relativistic aspects of such metrics have been studied recently by Gibbons and Ruback[3] in great detail. Here we explore instead the relation to *dynamical symmetries*.

In the Taub–NUT limit the electric charge q (the Nœther quantity conjugate to $\partial/\partial\psi$) is conserved and, for a fixed q, symplectic reduction[7] leads to the three-dimensional Hamiltonian

$$h = \tfrac{1}{2}(\mathbf{p}^2/(1+4m/r) + (1+4m/r)(q/4m)^2), \quad (1.2)$$

where $\mathbf{p} = (1+4m/r)\mathbf{v} = \partial\mathfrak{L}/\partial\mathbf{v} - q\mathbf{A}$, where $\mathbf{A}$ is a Dirac monopole vector potential. The fundamental Poisson brackets $\{p_i,p_j\} = -q\epsilon_{ijk}(r^k/r^3)$, $\{r_i,p_j\} = \delta_{ij}$ correspond to the symplectic form

$$\Omega = d\mathbf{r}\wedge d\mathbf{p} + (q/2r^3)\epsilon_{ijk}r^i\, dr^j\wedge dr^k \quad (1.3)$$

on the phase space, according to $\{f,g\} = \Omega(X_f,X_g)$, $df = \Omega(X_f,\cdot)$.

The second system we study here is a *spinless test particle* moving *outside the core of a self-dual monopole*. The Higgs field Φ can be identified with the extra space component of a pure Yang–Mills field in four dimensions. The equations of motion can be obtained from the Kerner–Wong[8] equations in $1+4$ dimensions by dimensional reduction.[5] For large distances, the only surviving gauge field component is the one parallel to Φ. The particle's isospin projects into the conserved electric charge q.[9] This leaves us with an effective Dirac monopole and a long-range scalar field $\Phi\sim 1-1/r$. Our particle is described by the same symplectic structure as in (1.3) and the Hamiltonian

$$h = \tfrac{1}{2}(\mathbf{p}^2 + q^2[1-1/r]^2), \quad (1.4)$$

where $\mathbf{p}$ is the ordinary momentum. This system was studied previously by McIntosh and Cisneros and Zwanziger[10] because of its remarkable symmetries, but without its present physical interpretation. Such a role has been hinted at by Schönfeld.[11] See, also, Refs. 5 and 12.

For both systems, the clue of the solutions is provided by a conserved Runge–Lenz-type vector, which allows one[2,3] to prove that the trajectories are conic sections. We shall mainly consider the bound motions. We mainly concentrate on the more recent and less explored Taub–NUT problem.

Observing that the conserved angular momentum vector $\mathbf{j}$ and the rescaled Runge–Lenz vector $\mathbf{k}$ [(2.8)] form an $o(4)$ algebra for the bound motions and an $o(3,1)$ algebra for the scattered motions, the Pauli method[13] allows one to recover the bound-state spectrum and the Zwanziger method[10] allows one to derive the S matrix.[2]

The $o(4)/o(3,1)$ symmetry can be extended into $o(4,2)$. For example,[3] application of the so-called "Kustaanheimo–Stiefel"[14] transformation carries the Taub–NUT system into a *harmonic oscillator*. The latter admits an $sp(8,\mathbf{R})$ dynamical symmetry; those transformations that preserve the charge constraint form an $su(2,2)\approx o(4,2)$.

In Barut's method[15] (for the Taub–NUT system, for example[4]), one starts instead with the *time-independent* Schrödinger equation $\hat{h}\Psi = e\Psi$. Assuming that $e<q^2/32m^2$, one multiplies the Schrödinger equation by $(r+4m)$ and redefines position and momenta as

$$\mathbf{R} = \sqrt{(q/4m)^2 - 2e}\,\mathbf{r}, \quad \mathbf{P} = \mathbf{p}/\sqrt{(q/4m)^2 - 2e}. \quad (1.5)$$

After rearrangement, the Schrödinger equation takes the form

$$\{\tfrac{1}{2}R(\mathbf{P}^2+1) + q^2/2R\}\Psi = (4m[e-(q/4m)^2]/\sqrt{(q/4m)^2-2e})\Psi. \quad (1.6)$$

 0022-2488/90/010202-10$03.00

On the lhs of (1.6) one recognizes Γ_0, the generator of an o(2,1) group, to which one can add[4,15] 14 more operators, cf. (3.11 a–g) which generate an o(4,2) operator algebra *independent* of the energy constraint. Therefore, the solution of the eigenvalue equation (1.6) can be deduced from the spectrum of Γ_0. The same procedure works in the other cases.[15]

The method we present here consists in completing $f:(\mathbf{r},\mathbf{p})\to(\mathbf{R},\mathbf{P})$ into a *canonical transformation*. We do this by unfolding the system into Souriau's[16] "espace d'évolution" (evolution space) $\mathscr{E}=M\times\mathbf{R}$, which is endowed with the presymplectic structure $\sigma=\Omega+dh\wedge dt$. [$\mathscr{E}$ can also be viewed as the seven-dimensional "energy surface" lying in the eight-dimensional extended phase space $T^*(\mathbf{R}^3\times\mathbf{R})$[17].] The classical motions are the characteristic curves of σ. This is basically a generalized variational formalism[16]: If θ is a potential for σ, $d\theta=-\sigma$, then the classical action is $\int\mathcal{L}=\int\theta$.

The quotient $(\mathscr{N},\omega)$ of $(\mathscr{E},\sigma)$ by the characteristic foliation of σ is Souriau's "espace des mouvements" (space of motions).[16] In this framework, a symmetry is a transformation of $\mathscr{E}$ which preserves σ and thus permutes the classical motions: it projects into a symplectomorphism of $(\mathscr{N},\omega)$.

The information on the global structure is encoded into the topology of $\mathscr{N}$. A fixed $t=t_0$ section N_0 of the evolution space is the "phase space at t_0"[16]; the restriction of σ to N_0 is symplectic. The mapping $N_0\to\mathscr{N}$ (obtained by composing with the projection $\mathscr{E}\to\mathscr{N}$) is injective and symplectic, but may *not* be *onto*. In the Kepler problem, for example, the phase space N_0 does not intersect those motions that hit the center at $t=t_0$.[18] Therefore, N_0 may not reflect the global structure of the space of motions.

The situation is similar for the Taub–NUT system. The metric (1.1) is singular for $r=4|m|$, which should be excluded. The energy is positive for $r>4|m|$ and negative for $r<4|m|$ and, by energy conservation, no motion can cross the singular sphere $S=\{r=4|m|\}$. Hence the space of bound motions has two connected components. The negative-energy part $\mathscr{E}_-$ of the Taub–NUT evolution space $\mathscr{E}$ contains the *tightly-bound motions* ($\mathscr{N}_-$) and the positive-energy part $\mathscr{E}_+$ contains the *lightly-bound motions* ($\mathscr{N}_+$). For us, $\mathscr{E}_+$ is more interesting since the Taub–NUT approximation is justified only for large r.

In both components, the radial motions leave their regions and hit the singularity. In other words, for $m<0$ the Taub–NUT space is not a complete Riemann manifold. Consequently, the spaces of motions $\mathscr{N}_\pm=\mathscr{E}_\pm/\mathrm{Ker}\,\sigma$ are not Hausdorff.

A regular system is one whose presymplectic form defines a foliation with one-dimensional, infinite curves: Its space of motions *is* a Hausdorff manifold. Regularizing the Taub–NUT problem requires imbedding it into a regular "unphysical" one by an injective, symplectic mapping f whose image is a dense, open subset. Those "unphysical" motions, that correspond to the Taub–NUT motions that leave the evolution space can be made infinite by restoring their points not in $\mathrm{Im}\,f$. Identifying the preimages in the Taub–NUT space of motions, we obtain a smooth Hausdorff manifold, namely the "unphysical" motion space.

We choose the following regular "unphysical" system $(\mathscr{M}_s,\Sigma_s)$: We consider in fact those zero-mass, helicity-s, coadjoint orbits (M_s,ϖ_s) associated with the action of SU(2,2) on twistor space.[19–21] The $\mathrm{su}(2,2)\simeq\mathrm{o}(4,2)$ generators are the classical counterparts of those operators in Refs. 4, 15, 21, and 22. Choosing the generator Γ_0 [(3.12a)] as Hamiltonian and adding a "fake time" T, we obtain an "unphysical" evolution space $\mathscr{M}_s=M_s\times\mathbf{R}$ endowed with $\Sigma_s=\varpi_s+dH\wedge dT$. The o(4,2) generators are extended to $\mathscr{M}_s$ so as to remain constant along the trajectories. The space of "unphysical" motions, $\mathscr{M}_s/\mathrm{Ker}\,\Sigma_s$, is *globally symplectomorphic* to the $T=0$ phase space which is (M_s,ϖ_s). This system has a *manifest* $C^1_+(3,1)\simeq\mathrm{SU}(2,2)/(\text{center})$ *symmetry*.

To summarize, our canonical transformation f allows us to *regularize* the "physical" problem as well as exhibit its "hidden" conformal symmetry. Our method is particularly useful in discussing global problems.

This transformation is found by completing (1.5) with the rule of transforming the time,

$$T=\left[\sqrt{(q/4m)^2-2h}\,/4mh\right]\times\{-\mathbf{p}\cdot\mathbf{r}-((q/4m)^2-2h)t\}. \tag{1.7}$$

Equation (1.7) is chosen to compensate for the noninvariance of $d\mathbf{r}\wedge d\mathbf{p}$ under (1.5), due to the energy being a function rather than a constant. The pullback of the "unphysical" presymplectic form is the Taub–NUT presymplectic form. (The Lagrangians differ by a total derivative.)

The regularized lightly-bound Taub–NUT motion spaces $\mathscr{N}_+$ are thus shown to be symplectomorphic to (M_s,ϖ_s). The same is true for the tightly-bound motions $\mathscr{N}_-$. Therefore, both carry a $C^1_+(3,1)$ conformal symmetry.

In the McIntosh–Cisneros (MIC)–Zwanziger case no regularization is necessary and M_s is thus also the space of bound test-particle motions in the asymptotic monopole field. (M_0 is the space of regularized motions of the Kepler problem.[21] Our method also yields the $C^1_+(3,1)$ symmetry.[12]

The space of twistors can also be viewed as the phase space of a four-dimensional harmonic oscillator from which the "unphysical" system is obtained by reduction.[3,15,22]

In Sec. V we study the scattered motions. We show that the space of regularized hyperbolic motions is symplectomorphic to the orbit (M_s,ϖ_s) and hence carries (unlike in the Kepler case[18]) an action of the conformal *group*.

We end this paper by a short discussion of two other (closely related) metrics whose geodesics are also o(4,2) symmetric. The first metric (which is new) can be viewed as a curved-space model for a particle in a self-dual monopole field and the other has been found recently in describing some special motions of a closed string in the Taub–NUT background.

Applied to the Kepler problem, our method would yield an imbedding into M_0, which is the standard regularization,[14,18,23] since $M_0\simeq T^+(\mathrm{S}^3)=T^*(\mathrm{S}^3)\setminus(\text{zero section})$. The conformal symmetry is obtained for free.

II. CLASSICAL MOTIONS IN TAUB–NUT

Neglecting radiation, the motion of two self-dual monopoles is approximately described by the geodesics in the space of solutions of the Bogomolny equation, called the moduli space.[1] The moduli space is the product of $\mathbb{R}^3\times S^1$, the manifold of the center-of-mass motion, with a curved four-manifold whose metric was found by Atiyah and Hitchin.[1] The latter describes the relative motion of the monopoles. In the long-distance limit exponential terms can be neglected and we obtain a Euclidean Taub–NUT space of parameter $m=-\frac{1}{2}$, with the line element (1.1). The geodesic motion of a spinless particle of unit mass in (1.1) is described by the Lagrangian

$$\mathcal{L}=\tfrac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu=\tfrac{1}{2}((1+4m/r)\mathbf{v}^2 + [(4m)^2/(1+4m/r)](\dot{\psi}+\cos\theta\dot{\phi})^2), \tag{2.1}$$

where $\mathbf{v}=\dot{\mathbf{r}}$. Here $r>0$ and the angles θ, ϕ, ψ ($0\leqslant\theta\leqslant\pi$, $0\leqslant\phi<2\pi$, $0\leqslant\psi<4\pi$) parametrize S^3. The points $r=4|m|$, where the metric (1.1) is singular, are excluded. The conserved Noether quantity

$$q=(4m)^2[(\dot{\psi}+\cos\theta\dot{\phi})/(1+4m/r)] \tag{2.2}$$

associated to the cyclic variable ψ is the *relative electric charge*. From now on we choose and fix a nonzero value for q. It is convenient to introduce the "mechanical momentum" $\mathbf{p}=(1+4m/r)\mathbf{v}$. The equation of motion is then

$$\frac{d\mathbf{p}}{dt}=-2m\frac{\mathbf{v}^2}{r^3}\mathbf{r}+\frac{q^2}{8m}\frac{\mathbf{r}}{r^3}-q\frac{\mathbf{v}\times\mathbf{r}}{r^3}. \tag{2.3}$$

We have the following conserved quantities. First, the energy,

$$e=\tfrac{1}{2}(1+4m/r)[\mathbf{v}^2+(q/4m)^2], \tag{2.4}$$

and next the monopole angular momentum

$$\mathbf{j}=\mathbf{r}\times\mathbf{p}+q(\mathbf{r}/r). \tag{2.5}$$

Finally, we have the Runge–Lenz-type vector

$$\mathbf{a}=\mathbf{p}\times\mathbf{j}-\frac{\mathbf{r}}{r}(4m(e-(q/4m)^2)). \tag{2.6}$$

Hence the trajectories lie simultaneously on the cone $\mathbf{j}\cdot\mathbf{r}/r=q$ and also in the plane perpendicular to

$$\mathbf{n}=q\mathbf{a}+4m[e-(q/4m)^2]\mathbf{j}, \tag{2.7}$$

because of the relation $\mathbf{n}\cdot\mathbf{r}=q(j^2-q^2)$. They are thus conic sections.[1,2]

For energies smaller than $(q/4m)^2/2$ (which is only possible for $m<0$) the motions are bound. We assume henceforth that $m<0$.

Under the Poisson bracket and for $e<(q/4m)^2/2$, the angular momentum $\mathbf{j}$ closes, with the rescaled Runge–Lenz vector

$$\mathbf{k}=\frac{\mathbf{a}}{\sqrt{|(q/4m)^2-2e|}} =\frac{\mathbf{p}\times\mathbf{j}-(\mathbf{r}/r)(4m(e-(q/4m)^2))}{\sqrt{|(q/4m)^2-2e|}}, \tag{2.8}$$

into an o(4) dynamical symmetry algebra. For $e>(q/4m)^2/2$ we instead obtain an o(3,1) algebra.[2]

Via (2.4), the sign of the energy depends on r being smaller or larger than $4|m|$. The excluded points form a singular sphere

$$S=\{\mathbf{r}\in\mathbb{R}^3|r=4|m|\}, \tag{2.9}$$

which divides the space into two regions. Energy conservation implies that a particle cannot go from one region into the other (although it can hit the boundary S, see below). If a finite-energy motion approaches S, its velocity $|v|$ goes, via (2.4), to infinity as $(1-4|m|/r)^{-1/2}$; its momentum $\mathbf{p}$ hence goes to zero as $(1-4|m|/r)^{1/2}$. Those motions in the interior of S have negative energy; they are the *tightly-bound motions* $\mathcal{N}_-$. Those motions in the exterior and having energy $0<e<q^2/32m^2$ are the *lightly bound motions* $\mathcal{N}_+$. We shall focus our attention on $\mathcal{N}_+$.

In the generic case the orbital angular momentum $\mathbf{r}\times\mathbf{p}$ is nonzero, and the cone has opening angle $\alpha(\cos\alpha=|q|/|\mathbf{j}|)$. Such motions avoid S. Indeed, we see from (2.5) that for nonvanishing orbital angular momentum, $\mathbf{j}$ cannot be radial. However, when hitting the singular sphere S, the orbital part necessarily vanishes requiring $\mathbf{j}$ to be radial.

Consider now the radial motions. Fixing a direction, we work with r, p. If the initial velocity is inward, the particle reaches the singularity in finite time, and leaves the "physical" space. If the velocity is outward (but sufficiently low as to remain bound),

$$v=\sqrt{((2e-(q/4m)^2)r+q^2/4|m|)/(r+4m)} \tag{2.10}$$

shows that there will be a unique turning point $r_1>0$ where v vanishes, namely at

$$r_1=4|m|\{(q/4m)^2/[(q/4m)^2-2e]\}>4|m|. \tag{2.11}$$

After reaching r_1, the particle returns and falls inward until it disappears in S. At this very moment, another radial motion leaves the singularity and follows the same phase-space trajectory backward. When passing to the quotient, any two neighborhoods of these two motions intersect. In order to obtain a Hausdorff topology, such motions should—and will—be identified. All motions then become periodic.

The set of outer turning points of radial motions is

$$B^0\times\mathbb{R}=\{(\mathbf{r},\mathbf{p},t)|\ |\mathbf{r}|>4|m|,\mathbf{p}=0\} \tag{2.12}$$

and the set of inner turning points is

$$S\times\mathbb{R}=\{(\mathbf{r},\mathbf{p},t)|\ |\mathbf{r}|=4|m|,\mathbf{p}=0\}. \tag{2.13}$$

The situation is basically the same for the tightly-bound motions. The nonradial motions are ellipses which avoid the origin as well as S. A radial motion has an internal turning point at $0<r_1<4|m|$, according to (2.11). All radial motions fall into S in finite time from the inside, with infinite velocity and zero momentum: Such a motion should be identified with the motion that leaves S in the opposite direction along the same trajectory.

III. SOME MANIFESTLY o(4,2)-SYMMETRIC SYSTEMS

A twistor[19,20] can be represented by a pair of spinors $Z^\alpha=(\omega^A,\pi_{A'})$ in $\mathbb{T}=(\mathbb{C}^2\times\mathbb{C}^2)\setminus\{0\}$. (Here $\pi_{A'}$ plays the role of a generalized coordinate and ω^A plays that of a generalized momentum.) The conjugate of $Z^\alpha(\alpha=0,1,2,3)$ is $Z^*_\alpha=(\pi^*_A,(\omega^*)^{A'})=((\pi_{A'})^*,(\omega^A)^*)$ (the asterisk means

complex conjugate). The space of twistors is endowed with a Hermitian quadratic form of signature (2,2) given by

$$Z^\alpha Z^*_\alpha = \omega^A \pi^*_A + \pi_{A'}(\omega^*)^{A'}, \quad A = 0,1; A' = 0',1'. \quad (3.1)$$

To each real number s we associate a (real) seven-dimensional manifold $\mathbf{T}_s$, namely the level surface

$$Q(Z^\alpha, Z^*_\alpha) = \tfrac{1}{2} Z^\alpha Z^*_\alpha = s. \quad (3.2)$$

Here $\mathbf{T}$ carries the (by construction) U(2,2) invariant one-form

$$\theta = -(i/2)(Z^\alpha dZ^*_\alpha - Z^*_\alpha dZ^\alpha) \quad (3.3)$$

whose exterior derivative

$$-d\theta = i\, dZ^\alpha \wedge dZ^*_\alpha \quad (3.4)$$

is a symplectic form on $\mathbf{T}$. The Poisson brackets are thus $\{Z^\alpha, Z^*_\beta\} = -i\delta^\alpha_\beta$. The restriction ϖ_s of $(-d\theta)$ to the level surface $\mathbf{T}_S$ defines a one-dimensional integrable foliation and ϖ_s descends to M_s, the quotient of $\mathbf{T}_s$ by the characteristic foliation of ϖ_s. In this way M_s becomes a six-dimensional symplectic manifold. Explicitly, the characteristic curves of ϖ_s (the Hamiltonian flow of Q) are circles,

$$Z^\alpha \to e^{-ip/2} Z^\alpha, \quad Z^*_\alpha \to e^{ip/2} Z^*_\alpha, \quad 0 \leqslant p \leqslant 4\pi, \quad (3.5)$$

which identifies M_s as $\mathbf{T}_s/\mathrm{U}(1)$.

The unitary group U(2,2) leaves invariant the quadratic form (3.1) and thus, also, the level surfaces $\mathbf{T}_s$. The action of U(2,2) on $\mathbf{T}_s$ is clearly transitive. The action of the diagonal U(1) subgroup of U(2,2) on $\mathbf{T}_s$ coincides with the flow (3.5). Therefore, it is only SU(2,2) that acts on the quotient. In this way we obtain a transitive, symplectic action of SU(2,2) on (M_s, ϖ_s). Souriau's moment map[16] therefore identifies (M_s, ϖ_s) with a coadjoint orbit of SU(2,2), endowed with its canonical symplectic structure. M_s can also be viewed as a U(2,2) coadjoint orbit, where s is an element in the center of the Lie algebra.

For $s \neq 0$ the Poincaré subgroup of SU(2,2) already acts transitively, so that M_s is actually sumplectomorphic to the Poincaré orbit $(\mathcal{Q}_{0,s,+})$, the space of motions of a relativistic, zero-mass, helicity-s, elementary particle.

For $s = 0$ the action of the Poincaré subgroup on M_0 is no longer transitive and M_0 is rather the space of motions of a helicity-zero, mass-zero particle in *compactified* Minkowski space $\mathbf{S}^1 \times \mathbf{S}^3$. M_0 is obtained from the zero-mass Poincaré orbit $(\mathcal{Q}_{0,0,+})$ by adding those motions that lie along the generators of the light cone at infinity.

As will be clear from the parametrization below, all zero-mass Poincaré orbits are diffeomorphic to $\mathbf{R}^3 \times (\mathbf{R}^3 \setminus \{0\})$. This is thus the topology of M_s for $s \neq 0$. The topology of M_0 is, in turn, $\mathbf{S}^3 \times (\mathbf{R}^3 \setminus \{0\})$. Indeed,

$$R^\mu = \sigma^\mu_{AA'} \pi^*_A \pi_{A'}, \quad (3.6)$$

where the σ_μ are the Pauli matrices, determines, for any $\pi_{A'} \in \mathbf{C}^2 \setminus \{0\}$, a unique future-pointing light-like vector $(R^\mu) = (R, \mathbf{R})$ $(R = |\mathbf{R}|)$ in Minkowski space. Conversely, those π's that solve Eq. (3.6) for a given $\mathbf{R}$ belong to a circle. This is clear from the following:

$$\pi_{A'} = \sqrt{R} \begin{pmatrix} \cos(\theta/2) e^{-i(\psi+\phi)/2} \\ \sin(\theta/2) e^{i(-\psi+\phi)/2} \end{pmatrix}. \quad (3.7)$$

The vector $\mathbf{R}$ has the polar coordinates R, θ, ϕ. The map $\pi_{A'} \to \mathbf{R}$ is thus essentially the projection of the Hopf fibering $\mathbf{S}^3 \to \mathbf{S}^2$; its (multivalued) inverse is the Kustaanheimo–Stiefel[14] transformation.

Choosing $\pi_{A'}$ to a pair $(\mathbf{P}, \mathbf{R})$ in $\mathbf{R}^3 \times (\mathbf{R}^3 \setminus \{0\})$ we can associate a twistor $Z^\alpha = (\omega^A, \pi_{A'})$ by setting

$$\omega^A = i(P^k \sigma_k^{AA'} - i(s/R)\sigma_0^{AA'})\pi_{A'}. \quad (3.8)$$

For any choice of $\pi_{A'}$ (i.e., of the phase ψ) Z^α belongs to $\mathbf{T}_s$ and the whole of $\mathbf{T}_s$ is obtained. Thus the pairs $(\mathbf{P}, \mathbf{R})$ parametrize those circles in Eq. (3.5) and thus the quotient manifold M_s.

For $s = 0$, $\mathbf{T}_0 = \mathbf{T}^0_0 \cup \mathbf{T}^0_\infty$, $(\mathbf{T}^0_0 \cap \mathbf{T}^0_\infty = 0)$, where $\mathbf{T}^0_0 = \{(\omega^A, \pi_{A'}) \in \mathbf{T}_0 | \pi_{A'} \neq 0\}$ and $\mathbf{T}^0_\infty = \{(\omega^A, \pi_{A'}) | \omega^A \neq 0, \pi_{A'} = 0\}$. The complex projective lines in $\mathbf{PT}$ meeting $\mathbf{PT}^0_\infty$ corresponds to points at infinity in (compactified and complexified) Minkowski space. Therefore, the orbit M_0 is decomposed as

$$M_0 = \mathcal{O}_{0,0,+} \cup M^0_\infty, \quad \mathcal{O}_{0,0,+} = \mathbf{T}^0_0/\mathrm{U}(1),$$
$$M^0_\infty = \mathbf{T}^0_\infty/\mathrm{U}(1). \quad (3.9)$$

As anticipated by the notation, $\mathcal{O}_{0,0,+}$ is a zero-mass, zero-helicity Poincaré orbit because the Poincaré subgroup of SU(2,2), leaves the constraint $\pi_{A'} \neq 0$ invariant. Here M^0_∞ describes those motions that lie along the generators of the light cone at infinity. The decomposition (3.9) also shows that M_0 is symplectomorphic to $T^+\mathbf{S}^3$, the cotangent bundle of the three-sphere with its zero section deleted. Indeed, $\mathbf{R}^3$ is (by stereographic projection) $\mathbf{S}^3$ without its north pole; the Poincaré orbit $\mathcal{O}_{0,0,+}$ is $\mathbf{R}^3 \times (\mathbf{R}^3 \setminus \{0\}) \simeq T^+(\mathbf{S}^3 \setminus \{N\})$ and $M^0_\infty \simeq T^+_N(\mathbf{S}^3)$.

The action of $\mathrm{su}(2,2) \approx \mathrm{o}(2,4)$ on $\mathbf{T}$ is generated by the matrices

$$\gamma_{0k} = -\frac{i}{2}\begin{pmatrix} \sigma_k & 0 \\ 0 & -\sigma_k \end{pmatrix}, \quad \gamma_{jk} = \frac{1}{2}\epsilon_{jkn}\begin{pmatrix} \sigma_n & 0 \\ 0 & \sigma_n \end{pmatrix},$$
$$\gamma_{06} = \frac{1}{2}\begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}, \quad \gamma_{k6} = \frac{1}{2}\begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix},$$
$$\gamma_{05} = \frac{1}{2}\begin{pmatrix} 0 & \sigma_0 \\ -\sigma_0 & 0 \end{pmatrix}, \quad \gamma_{k5} = \frac{1}{2}\begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix},$$
$$\gamma_{56} = \frac{i}{2}\begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix},$$
$$(\gamma_{LK} = -\gamma_{KL}, \quad K,L = 0,\ldots,3; 5, 6). \quad (3.10)$$

[Our convention for the metric on $\mathbf{R}^{2,4}$ is $g_{KL} = \mathrm{diag}(g_{\mu\mu}, g_{55}, g_{66}) = (+1, -1, -1, -1; -1, +1)$.] The matrices in (3.10) leave invariant the quadratic form (3.1) and the symplectic form (3.4). The components of the moment map are $J_{KL} = Z^*_\alpha (\gamma_{KL})^\alpha_\beta Z^\beta$. In dynamical group notations,[15] on each orbit we have the 15 generators

$$J_{06} \to \Gamma_0 = R[(\mathbf{P}^2 + 1)/2] + s^2/2R, \quad (3.11a)$$
$$\tfrac{1}{2}\epsilon_{ijk} J_{jk} \to \mathbf{J} = \mathbf{R} \times \mathbf{P} + s(\mathbf{R}/R), \quad (3.11b)$$
$$J_{5i} \to \mathbf{K} = \mathbf{R}\frac{\mathbf{P}^2 - 1}{2} - \mathbf{P}(\mathbf{R}\cdot\mathbf{P}) - \frac{s}{R}\mathbf{J} + s^2\frac{\mathbf{R}}{2R^2}$$
$$= \mathbf{P} \times \mathbf{J} - (\mathbf{R}/R)\Gamma_0, \quad (3.11c)$$

$$J_{i6} \to \mathbf{U} = \mathbf{R}\frac{\mathbf{P}^2+1}{2} - \mathbf{P}(\mathbf{R}\cdot\mathbf{P}) - \frac{s}{R}\mathbf{J} + s^2\frac{\mathbf{R}}{2R^2}$$
$$= \mathbf{P}\times\mathbf{J} - (\mathbf{R}/R)\Gamma_4, \tag{3.11d}$$
$$J_{56} \to D = -\mathbf{R}\cdot\mathbf{P}, \tag{3.11e}$$
$$J_{i0} \to \mathbf{V} = -R\cdot\mathbf{P}, \tag{3.11f}$$
$$J_{50} \to \Gamma_4 = R\frac{\mathbf{P}^2-1}{2} + \frac{s^2}{2R}. \tag{3.11g}$$

In particular, Γ_0, Γ_4, and D generate an o(2,1) subalgebra; those generators that commute with Γ_0 are $\mathbf{J}$ and $\mathbf{K}$, which form an o(4) subalgebra—the "invariance algebra" of the Hamiltonian Γ_0. The remaining o(4,2) generators are sometimes called "noninvariance" generators.

From (3.11) we see that

$$\mathbf{R} = \mathbf{U} - \mathbf{K}, \quad \mathbf{P} = -\mathbf{V}/R; \tag{3.12}$$

thus from the o(2,4) relations $\{J_{KL},J_{MN}\} = g_{KN}J_{LM} + g_{LM}J_{KN} - g_{KM}J_{KL} - g_{LN}J_{KM}$ we derive the symplectic form ϖ_s of $\mathcal{O}_{0,s,+}$:

$$\varpi_s = dR_i \wedge dP_i + (s/2R^3)\epsilon_{ijk}R^i dR^j \wedge dR^k. \tag{3.13}$$

Now we construct a classical dynamical system which has a *manifest* SU(2,2) symmetry. Consider, in fact, the evolution space

$$\mathcal{M}_s = M_s \times \mathbb{R} = \{\mathbf{R},\mathbf{P},T\}, \quad \Sigma_s = \varpi_s + dH\wedge dT, \tag{3.14}$$

where ϖ_s is the symplectic form of the orbit M_s and the Hamiltonian is

$$H(\mathbf{R},\mathbf{P}) = \Gamma_0(\mathbf{R},\mathbf{P}). \tag{3.15}$$

Let us extend the 15 generators of o(4,2) in (3.11) to $\mathcal{M}_s$ such that they remain conserved along the trajectories:

$$H^{\sim} = H = \Gamma_0, \tag{3.16a}$$
$$\mathbf{J}^{\sim} = \mathbf{J}, \tag{3.16b}$$
$$\mathbf{K}^{\sim} = \mathbf{K}, \tag{3.16c}$$
$$U_\alpha^{\sim} = U_\alpha \cos T + V_\alpha \sin T, \quad \alpha = 1,2,3,5, \tag{3.16d}$$
$$V_\alpha^{\sim} = -U_\alpha \sin T + V_\alpha \cos T, \quad \alpha = 1,2,3,5, \tag{3.16e}$$

where we have introduced the "Bacry–Györgyi"[24] four-vectors $(U_\alpha) = (\mathbf{U},D)$ and $(V_\alpha) = (\mathbf{V},\Gamma_4)$.

Combining (3.12) with (3.16) yields an explicit integration of the equations of motion:

$$\mathbf{R}(T) = \mathbf{U}^{\sim}\cos T - \mathbf{V}^{\sim}\sin T - \mathbf{K},$$
$$\mathbf{P}(T) = -(\mathbf{U}^{\sim}\sin T + \mathbf{V}^{\sim}\cos T)/R(T). \tag{3.17}$$

Equations (3.17) show that the orbits are ellipses, with period $\Delta T = 2\pi$. The Runge–Lenz vector $\mathbf{K}$ points from the origin into the center of the ellipse. The orbit is the intersection of the cone $\mathbf{R}\cdot\mathbf{J} = s$, with the plane normal to the vector

$$\mathbf{N} = -\mathbf{U}^{\sim}\times\mathbf{V}^{\sim} = s\mathbf{K} + \Gamma_0\mathbf{J}. \tag{3.18}$$

The quotient of $(\mathcal{M}_s, \Sigma_s)$ by the characteristic foliation of Σ_s is the space of "unphysical" motions. Since every motion is infinite and depends regularly on the initial conditions, this quotient is *globally symplectomorphic* to the $T = 0$ phase space, which is the SU(2,2) orbit (M_s,ϖ_s). The projection $\pi:\mathcal{M}_s \to M_s$ maps a point $(\mathbf{R}_0, \mathbf{P}_0, T_0)$ into the point at $T = 0$ on the unique classical motion through $(\mathbf{R}_0, \mathbf{P}_0, T_0)$.

The group SU(2,2) acts on $\mathcal{M}_s = M_s \times \mathbf{R}$ by acting on the first factor alone, without changing T. The center of SU(2,2) acts on the coadjoint orbit M_s trivially. Therefore, the Lie algebra action (3.16) integrates into a global symplectic action of the adjoint group of SU(2,2), which is the *conformal group* $C^{\uparrow}_{+}(3,1)$.

Notice that the "unphysical" energy function $H = \Gamma_0$ *satisfies* $H \geqslant |s|$ and equality is only achieved for $R = |\mathbf{R}| = |s|$ and $\mathbf{P} = 0$. A particle with the initial conditions $|\mathbf{R}(0)| = |s|$, $\mathbf{P}(0) = 0$ is in equilibrium.

Let us consider a motion with the initial conditions $\mathbf{R}(0) = \mathbf{R}_0$, $\mathbf{P}(0) = 0$ at $T = 0$. Since now $\mathbf{U}^{\sim} = (\mathbf{R}_0/R_0)\Gamma 4^{\sim}$, $\mathbf{V}^{\sim} = 0$, $-\mathbf{K}^{\sim} = (\mathbf{R}_0/R_0)\Gamma_0$ the motion oscillates on a line segment between the turning points $\Gamma_0 \pm \Gamma_4^{\sim}$ according to (3.17):

$$\mathbf{R}(T) = (\mathbf{R}_0/R_0)(\Gamma_0 - \Gamma_4^{\sim}\cos T),$$
$$\mathbf{P}(T) = -[\mathbf{U}^{\sim}/|\mathbf{R}(T)|]\sin T. \tag{3.19}$$

Notice that

$$\Sigma\times\mathbb{R} = \{0 < R \leqslant |s|, P = 0, T\}, \quad \text{resp.},$$
$$\Xi\times\mathbb{R} = \{R > |s|, P = 0, T\} \tag{3.20}$$

are the sets of the inner, resp., outer turning points.

IV. REGULARIZATION

The classical flow of the Taub–NUT evolution space is not complete: The radial motions leave it. Simply adding $S\times\mathbb{R}$ would not solve the problem since from the points of $S\times\mathbb{R}$ infinitely many motions start, all with zero momentum. Therefore, we regularize by relating the Taub–NUT problem to the regular "unphysical" dynamical system of Sec. III. Let us first study the lightly bound case.

Our guiding principle is that the "hidden" o(4) symmetry generators $\mathbf{j}$ and $\mathbf{k}$ of Taub–NUT should go into the manifest o(4) symmetry generators $\mathbf{J}$ and $\mathbf{K}$ of the "unphysical" problem. This is achieved by setting $s = q$ $(\neq 0)$ and defining $f(\mathbf{r},\mathbf{p},t) = (\mathbf{R},\mathbf{P},T)$, where

$$\mathbf{R} = \sqrt{(q/4m)^2 - 2h}\,\mathbf{r}, \quad \mathbf{P} = \mathbf{p}/\sqrt{(q/4m)^2 - 2h},$$
$$T = [\sqrt{(q/4m)^2 - 2h}/4mh]$$
$$\times(-\mathbf{p}\cdot\mathbf{r} - ((q/4m)^2 - 2h)t), \tag{4.1}$$

and h is the Taub–NUT Hamiltonian (1.2). The first two of Eqs. (4.1) ensure that f intertwines the o(4) generators and the last makes f canonical:

$$f^*\Sigma_s = f^*(\varpi_s + dH\wedge dT) = \Omega + dh\wedge dt = \sigma, \tag{4.2}$$

where

$$H = 4m\{[h - (q/4m)^2]/\sqrt{(q/4m)^2 - 2h}\}. \tag{4.3}$$

Expressing through the new variables $\mathbf{R}$ and $\mathbf{P}$ shows that H is the generator Γ_0 in (3.11a), which we have chosen for the "unphysical" Hamiltonian.

Now $f:(\mathbf{r},\mathbf{p},t) \to (\mathbf{R},\mathbf{P},T)$ maps the positive-energy Taub–NUT evolution space $\mathcal{E}_+$ into the "unphysical" evolution space $\mathcal{M}_s = M_s\times\mathbb{R}$. The formal inverses are

$(\mathbf{r},\mathbf{p},t)=f^{-1}(\mathbf{R},\mathbf{P},T)$, with

$$\mathbf{r}=4|m|\frac{\mathbf{R}}{H\mp\sqrt{H^2-s^2}},\quad \mathbf{p}=\frac{H\mp\sqrt{H^2-s^2}}{4|m|}\mathbf{P},$$

$$t=\left(\frac{4m}{H\mp\sqrt{H^2-s^2}}\right)^2\left\{\left(H+\frac{s^2}{H\mp\sqrt{H^2-s^2}}\right)T-\mathbf{P}\cdot\mathbf{R}\right\}, \tag{4.4}$$

where $q=s$ and the energy transforms according to

$$h=[\sqrt{H^2-s^2}/(4m)^2](\pm H-\sqrt{H^2-s^2}). \tag{4.5}$$

In order to obtain a positive sign for h, we have to choose the upper signs.

Clearly, f cannot be a symplectomorphism because $(\mathscr{E}_+,\sigma)$ is not complete, while $(\mathscr{M}_s,\Sigma_s)$ is complete. In fact, $f=f|\mathscr{E}_+$ is not onto—but this is what we need. Denote $M_s\setminus\Sigma$ by M_s^0, where Σ is given in (3.20).

Proposition: Consider the dense, open subset $\mathscr{M}_s^0=(M_s\setminus\Sigma)\times\mathbf{R}$ of the unphysical evolution space. Then (i) $f:(\mathscr{E}_+,\sigma)\to(\mathscr{M}_s^0,\Sigma_s)$ is a (pre)symplectic bijection and (ii) the inverse (4.4) extends naturally into a continuous mapping $\mathscr{M}_s\to\mathscr{E}_+\cup(S\times\mathbf{R})$. Here f^{-1} carries the whole $\Sigma\times\mathbf{R}$ into the singularity $S\times\mathbf{R}$.

Proof: Since f preserves directions, it is sufficient to work with the absolute values $r=|\mathbf{r}|$, $p=|\mathbf{p}|$, $R=|\mathbf{R}|$, and $P=|\mathbf{P}|$. Also, since T depended on t linearly, we can—and will—drop the time variables when studying the global properties of f.

We first show that $\mathrm{Im}\,f$ does not contain those points $\{0<R\leqslant|q|,P=0\}$, i.e., the subset $\Sigma\subset M_s$. Indeed, $P=0$ requires $p=0$. Then, by (4.1), $R=|q|(r/4|m|)^{1/2}>|q|$ since $r>4|m|$. Thus $\mathrm{Im}\,f\subset\mathscr{M}^0$. In order to prove that $\mathrm{Im}\,f$ fills $\mathscr{M}^0$, it is convenient to introduce some more points (cf. Fig. 1):

$$A'=(r=\infty,p=0),\quad A=(R=\infty,P=0),$$

$$C=(R=0,P=\infty),$$

and we set $B^0=\{(r>4|m|,\ p=0\}$ [cf. (2.11)] and $B^1=\{(r,p(r))\}$, where $p(r)=(|q|/\sqrt{4|m|})\sqrt{1/r+4m/r^2}$. Here B^0 belongs to $\mathscr{E}_+$, but B^1 does not. As we have seen, the interior points of the region whose boundary is $\{S\}\cup\{A'\}\cup B^1$ are carried by f into $\mathscr{M}^0$ and $f(A')=A$ $f(B^0)=\Xi$, $f(B^1)=C$.

Let a be an arbitrary non-negative number and let us consider the hyperbolas $\mathscr{H}'_a=\{r\cdot p=a\}$ and $\mathscr{H}_a=\{R\cdot P=a\}$. The Taub–NUT evolution space is clearly the boundary B^0 plus the union of its intersections with the hyperbolas $\mathscr{H}'_a$. In turn, $\mathscr{M}^0$ is Ξ plus the union of the hyperbolas $\mathscr{H}^a$. Each hyperbola intersects the "upper boundary" B^1 at exactly one point, which is sent into C. Furthermore, the hyperbola $\mathscr{H}'_a$ is carried into $\mathscr{H}_a$ since $R\cdot P=r\cdot p$. It follows that $f(\mathscr{H}'_a\cap\mathscr{E}_+)=\mathscr{H}_a$. Adding the bottom line B^0 whose image is Ξ, we conclude that the image of f is the *entire* $\mathscr{M}^0$.

Finally, f is injective: A point (R,P) in Ξ is the image of $(r=4|m|(R/s)^2,\ p=0)$ from B^0; otherwise it lies on a unique hyperbola $\mathscr{H}_a$ and thus has a unique preimage in $\mathscr{H}'_a\cap\mathscr{E}_+$. This proves (i) of the proposition.

To prove (ii) of the proposition, observe that (4.4) is naturally defined for any point of $\mathscr{M}$. However, for a point in $\Sigma\times\mathbf{R}$, $H=R/2+s^2/2R$, so that $H-(H^2-s^2)^{1/2}=R$ since $R/2-s^2/2R$ is negative for $R\leqslant|s|$. From (4.4) we infer that

$$f^{-1}(\Sigma\times\mathbf{R})=(4|m|,0,\mathbf{R}),$$

which is in the boundary $S\times\mathbf{R}=\{r=4|m|,p=0\}\times\mathbf{R}$ and does not belong to $\mathscr{E}_+$. The extension f^{-1} is clearly many-to-one. Q.E.D.

Regularizing $\mathrm{Im}\,f\subset\mathscr{M}_s$ is trivial: It is sufficient to add those turning points that we have excluded, i.e., $\Sigma\times\mathbf{R}$ [cf. (3.20)]. For Taub–NUT this amounts to gluing together the branches of the radial motions. When passing to the space of motions, this means identifying those points that thus far represent different (not infinite) motions and whose neighborhoods are not separated. This procedure yields a smooth, Hausdorff topology, namely that of M_s. To summarize, we present the following theorem.

Theorem: The map f: $(\mathscr{E}_+,\sigma)\to(\mathscr{M}_s,\Sigma_s)$ regularizes the Taub–NUT problem: It intertwines the time-independent o(4) symmetries. Here $(\mathscr{N}_+,\omega)$, the space of regularized lightly-bound Taub–NUT motions, is symplectomorphic to the SU(2,2) orbit (M_s,ϖ_s) and hence carries a symplectic action of the conformal group $C^\uparrow_+(3,1)$.

The results in Sec. II are consequences of what we have found in Sec. III and the properties of the canonical transformation f. For example, it follows from (4.1) and (3.17) that the trajectories are ellipses in the plane perpendicular to the vector n in (2.7),

$$\mathbf{n}=q\mathbf{k}+4m\{[e-(q/4m)^2]/\sqrt{(/4m)^2-2e}\}\mathbf{j}, \tag{4.6}$$

which is (up to normalization) the image of $\mathbf{N}$ in (3.18).

The pullbacks of the 15 generators in (3.16) by f yield o(4,) symmetry generators of the Taub–NUT system: They coincide with the classical counterparts of the quantum operators written in Ref. 4. *Without* regularization, this would only yield an o(4,2) *algebra.*

The period of a Taub–NUT motion could be obtained as the image under f of the "unphysical" period 2π. This would yield a "generalized third Kepler law."

Essentially the same argument works for the tightly

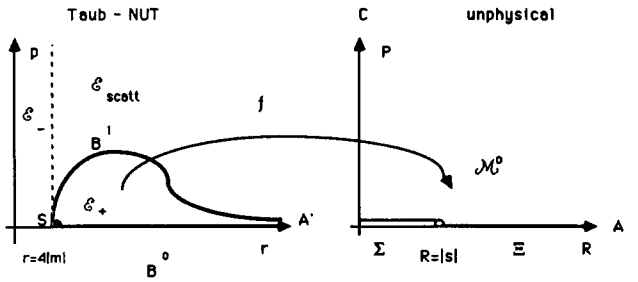


FIG. 1. The canonical transformation f in (4.1) takes the lightly-bound Taub–NUT evolution space $\mathscr{E}_+$ symplectically onto $\mathscr{M}^0=(M_s\setminus\Sigma)\times\mathbf{R}$. The image of the entire $\Sigma\times\mathbf{R}$ by the inverse f^{-1} is the singularity $S\times\mathbf{R}$. Similarly, the tightly-bound evolution space $\mathscr{E}_-$ is carried into $(M_s\setminus\Xi)\times\mathbf{R}$ and now $f^{-1}(\Xi\times\mathbf{R})=S\times\mathbf{R}$. The unbound part $\mathscr{E}_{scatt}$ is symplectomorphic to the full $\mathscr{M}$. Only the absolute values are shown and the time variables are dropped.

bound motions. The restriction of f to E_- is injective, but not onto: The inverse (4.4) (with the lower sign now) maps the $\Xi^* \times \mathbf{R} = \{^-R > |s|,\ P = 0,\ T \in \mathbf{R}\}$ into $S \times \mathbf{R}$, again in a many-to-one manner. The space of regularized negative-energy motions is thus once more the orbit M_s and therefore carries a symplectic action of the conformal group $C^{\uparrow}_{+}(3,1)$.

Let us now consider the *MIC–Zwanziger system* (1.3) and (1.4). In the bound motion region $0 < e < q^2/2$ we apply the transformation similar to the one used in the Kepler problem,[17] namely

$$\mathbf{R} = \mathbf{r}\sqrt{q^2 - 2h(\mathbf{r},\mathbf{p})}, \quad \mathbf{P} = \mathbf{p}/\sqrt{q^2 - 2h(\mathbf{r},\mathbf{p})}, \tag{4.7}$$

$$T = \left[\sqrt{q^2 - 2h(\mathbf{r},\mathbf{p})}/q^2\right]((q^2 - 2h(\mathbf{r},\mathbf{p}))t + \mathbf{r}\cdot\mathbf{p}).$$

The transformation (4.7) maps the MIC–Zwanziger system into the "unphysical" one and the pullback by (4.7) of the "unphysical" presymplectic form Σ_s is $\Omega + dh \wedge dt$, where h is the Hamiltonian (1.4). Therefore, Eqs. (4.7) are canonical. The energy transforms as

$$H = q^2/\sqrt{q^2 - 2h}. \tag{4.8}$$

Since $q \neq 0$ by assumption, (4.7) can be inverted:

$$r_1 = [H(R_i,P_i)/q^2]R_i, \quad P_i = [q^2/H(R_i,P_i)]P_i,$$
$$t = (H(R_i,P_i)/q^2)^2[H(R_i,P_i)T - R_iP_i]. \tag{4.9}$$

No regularization is needed in this case because the "MIC–Zwanziger" system is itself regular: No motion reaches the center. This is clear from $r = (H/|q|)(R/|q|) \geq R/|q|$.

The point $r = 1$, $p = 0$ (the image of $R = |s|$, $P = 0$) is now a regular equilibrium point. It is just where $V(r) = q^2(1 - 1/r)^2$ takes its minimum. It has no physical role, however, because the "MIC–Zwanziger" approximation to test particle motion in a self-dual monopole field already breaks down for much larger distances.

We conclude that for the MIC–Zwanziger system, (4.7) is a *global symplectomorphism*. The interpretation of the symmetry generators is analogous: For example, $\mathbf{K}$ corresponds to the rescaled Runge–Lenz vector

$$\mathbf{k} = (1/\sqrt{q^2 - 2h})(\mathbf{p} \times \mathbf{j} - q^2(\mathbf{r}/r)), \tag{4.10}$$

etc. The trajectories are ellipses in the plane perpendicular to

$$\mathbf{n} = q\mathbf{k} + (q^2/\sqrt{q^2 - 2h})\mathbf{j}. \tag{4.11}$$

This proves the $C^{\uparrow}_{+}(3,1)$ dynamical symmetry for the MIC–Zwanziger system, with generators given in (3.11), cf. Refs. 12 and 15. As a secondary result, we also obtain the equivalence between the regularized Taub–NUT and MIC–Zwanziger systems, cf. Ref. 25.

V. UNBOUND MOTIONS

Now we give a brief account of the unbound motions. We start with another "unphysical" system described by $\mathcal{M}_s = M_s \times \mathbf{R}$ and $\sigma_s = \varpi_s + dH \wedge dT$ [thus far identical to (3.15)], but instead choose the Hamiltonian

$$H = \Gamma_4 = \tfrac{1}{2}\{R(\mathbf{P}^2 - 1) + s^2/R\}. \tag{5.1}$$

All motions of this system are infinite and thus the space of motions $\mathcal{M}_s/\mathrm{Ker}\,\sigma_s$ is globally symplectomorphic to the $T = 0$ phase space, which is again (M_s, ϖ_s).

The generators (3.11) of the action of the conformal group are extended into $\mathcal{M}_s$ as

$$\tilde{\Gamma}_4 = \Gamma_4, \tag{5.2a}$$
$$\tilde{\mathbf{J}} = \mathbf{J}, \tag{5.2b}$$
$$\tilde{\mathbf{U}} = \mathbf{U}, \tag{5.2c}$$
$$\tilde{\mathbf{K}} = \mathbf{K}\,\mathrm{ch}\,T + \mathbf{V}\,\mathrm{sh}\,T, \tag{5.2d}$$
$$\tilde{D} = D\,\mathrm{ch}\,T + \Gamma_0\,\mathrm{sh}\,T, \tag{5.2e}$$
$$\tilde{\mathbf{V}} = \mathbf{K}\,\mathrm{sh}\,T + \mathbf{V}\,\mathrm{ch}\,T, \tag{5.2f}$$
$$\tilde{\Gamma}_0 = D\,\mathrm{sh}\,T + \Gamma_0\,\mathrm{ch}\,T. \tag{5.2g}$$

Combing with (3.12) we deduce that the trajectories are

$$\mathbf{R}(T) = \mathbf{U}(T) - \mathbf{K}(T)$$
$$= \mathbf{U} - \tilde{\mathbf{K}}\,\mathrm{ch}\,T + \tilde{\mathbf{V}}\,\mathrm{sh}\,T, \tag{5.3}$$

which are hyperbolas with the center at $\mathbf{U}$ and perpendicular to

$$\mathbf{N} = \tilde{\mathbf{K}} \times \tilde{\mathbf{V}} = s\mathbf{K} + \Gamma_4\mathbf{J}. \tag{5.4}$$

For the initial condition $\mathbf{R}(0) = \mathbf{R}_0$, $(0 < |\mathbf{R}_0| < \infty)$, $\mathbf{P}(0) = 0$, we obtain a semi-infinite radial motion

$$\mathbf{R}(T) = (\mathbf{R}_0/R_0)(\tilde{\Gamma}_0\,\mathrm{ch}\,T - \tilde{\Gamma}_4),$$
$$\mathbf{P}(T) = \left(\frac{\mathbf{R}_0}{R}\right)\frac{1}{1 - (\tilde{\Gamma}_4/\tilde{\Gamma}_0)(1/\mathrm{ch}\,T)}\,\mathrm{th}\,T, \tag{5.5}$$

whose (unique) turning point is at $R(0) = R_0$. The set of turning points is

$$\Delta \times \mathbf{R} = \{(\mathbf{R},\mathbf{P},T) | 0 < R < \infty, \mathbf{P} = \mathbf{0}\}. \tag{5.6}$$

Now we turn to the "physical" systems. Let us first assume that we are working with the $m < 0$ Taub–NUT case and with the energy $e > (q/4m)^2/2$, so that the motions are hyperbolas. As for the bound case, nonradial motions avoid the singular sphere S. All radial motions hit S in finite time, with infinite velocity and zero momentum. Such a motion should be identified with the one bouncing off at the same moment along the same phase-space trajectory. An unbound motion has a single turning point, which lies in $S \times \mathbb{R}$.

Let us now relate these two systems by an appropriately modified version of (4.1), $f(\mathbf{r},\mathbf{p},t) = (\mathbf{R},\mathbf{P},T)$, with

$$\mathbf{R} = \sqrt{2h - (q/4m)^2}\,\mathbf{r}, \quad \mathbf{P} = \mathbf{p}/\sqrt{2h - (q/4m)^2},$$
$$T = \frac{\sqrt{2h - (q/4m)^2}}{-4mh}(\mathbf{p}\cdot\mathbf{r} - (2h - (q/4m))^2 t). \tag{5.7}$$

Again, (5.7) is canonical and $f^*(\varpi_s + dH \wedge dT) = \sigma$ for

$$H = 4m[h - (q/4m)^2]/\sqrt{2h - (q/4m)^2} = \Gamma_4. \tag{5.8}$$

The same argument as for the bound motions shows that f is injective, but not surjective: $\mathrm{Im}(f) = \mathcal{M}_s \setminus (\Delta \times \mathbf{R})$. The formal inverse of (5.7) carries $\Delta \times \mathbf{R}$ into $\{r = 4|m|, p = 0\} \times \mathbf{R}$. In this case, the regularization amounts to restoring the turning-point set $\Delta \times \mathbf{R}$. The space of regularized, hyperbolic Taub–NUT motions hence becomes *globally symplectomorphic* to (M_s, ϖ_s). Therefore, it carries an action of the conformal *group* $C^{\uparrow}_{+}(3,1)$. This is in contrast with what occurs for the Kepler problem, where the scattered motions only have a Lie algebra symmetry, which does

TABLE I. Regularization and group action in various cases.

	Bound	Unbound			
		Regularization	Group action	Regularization	Group action
Taub–NUT	$m>0$	no bound motions		no	yes
	$m<0$	yes	yes	yes	yes
Asymptotic BPS		no	yes	no	no

not integrate into a group action.[18] Those generators commuting with the Hamiltonian (5.1), namely **J** and **U**, form an o(3,1) subalgebra. It is now **U** (rather than **K**) that goes into the rescaled Runge–Lenz vector **k** under f. This is not a surprise since (5.7) could have been obtained by requiring (besides canonicity) that the time-independent o(3,1) algebras go into each other.

The remaining cases are analogous: For $m>0$, the original Kaluza–Klein monopole situation, the metric is everywhere regular, including at the origin.[26] All motions are hyperbolic and none reaches the center, but rather has a turning point [still given by (2.11)]. The transformation (5.7) yields a *global symplectomorphism* between the phase space N_0 (which is now a global chart of the space of motions for $q\neq 0$) and M_s. Therefore, we have a global $C^{\uparrow}_{+}$ (3,1) conformal symmetry.

For a test particle in the long-range self-dual background unbound motions arise for $e\geqslant q^2/2$. No motion reaches the center and thus no regularization is necessary. Equation (4.12) (with a sign change under the root) is again an injective symplectic mapping: Its image is, however, only the *positive-energy part* $H>0$ of $\mathcal{M}_s$. Therefore, there is only a Lie *algebra* action, which does not integrate into a *group* action because the group trajectories leave the positive-energy part. The situation is summarized in Table I.

VI. OTHER o(2,4)-SYMMETRIC GEODESICS

Curiously enough, the same type of o(2,4) symmetry is encountered for the geodesics of some other metrics. Let us first consider the metric obtained from the Taub–NUT line element (1.1) by "rescaling":

$$ds^2=\{dr^2+r^2(d\theta^2+\sin^2\theta\,d\phi^2)\}+\frac{(d\psi+\cos\theta\,d\phi)^2}{(1-1/r)^2}. \tag{6.1}$$

Here $\partial/\partial\psi$ is a Killing vector, the Kaluza–Klein analog of an internal symmetry. The associated conserved quantity $q=(1-1/r)^2(\dot\psi+\cos\theta\dot\phi)$ is again an electric charge. The geodesics of (6.1) satisfy

$$\frac{d^2\mathbf{r}_i}{dt^2}=\alpha\frac{\mathbf{r}_i}{r^3}+q\epsilon_{ijk}\frac{\mathbf{r}_j\mathbf{v}_k}{r^3}+q^2\frac{\mathbf{r}_i}{r^4}, \tag{6.2}$$

where $\mathbf{v}=\dot{\mathbf{r}}$ and $\alpha=q^2/4m$. However, this is exactly the equation of motion one obtains for a test particle in the asymptotic field of a self-dual monopole whose electric charge is q.[5] Observe that (6.2) is the equation of motion for the MIC–Zwanziger system (1.3) and (1.4) (with the Coulomb coefficient replaced by α); thus it admits a $C^{\uparrow}_{+}$ (3,1) conformal dynamical symmetry with all its aforementioned consequences.

The metric (6.1) has the Kaluza–Klein form

$$g_{\mu\nu}=\begin{pmatrix} g_{ij}+A_iA_j/V & A_i/V \\ A_j/V & 1/V \end{pmatrix}, \tag{6.3}$$

where $g_{ij}=\delta_{ij}$ $(i,j=1,2,3)$ is the flat Euclidean metric, A_i is a Dirac monopole vector potential, and the "Brans–Dicke" scalar $V=(1-1/r)^2$ is the square of the asymptotic Higgs field of a BPS monopole: It can therefore be considered as a *curved-space model* for a test particle in *the long-range self-dual monopole field*. The metric (6.1) is singular at $r=1$, yielding a singularity in the definition of the electric charge q. This is consistent with the behavior of a test particle in the monopole field. (Both the "MIC–Zwanziger" approximation and the definition of the electric charge are only valid for $r\gg 1$.)

Yet another example was found very recently by Gibbons and Ruback,[27] who consider a closed string (a "winder") in a five-dimensional static Kaluza–Klein space-time g_{AB} $[A,B=0,1,2,3,5]$,

$$g_{AB}=\begin{pmatrix} -g_{00} & & \\ & g_{ij}+A_iA_j/V & A_i/V \\ & A_j/V & 1/V \end{pmatrix} \tag{6.4}$$

$(g_{00}=1)$. The string motion is governed by the Nambu–Goto action

$$S=-\frac{1}{2\pi\alpha'}\int\sqrt{-\det g_{AB}\frac{\partial x^A}{\partial u^a}\frac{\partial x^B}{\partial u^b}}\,du^1\,du^2, \tag{6.5}$$

where $u^1=\sigma$ is periodic with period 2π since the string is closed and $u^2=\tau$. Gibbons and Ruback[27] assume that the string moves entirely in the internal space, winding m times around the internal circle: More precisely, they assume that $x^5=mR_K\sigma, x^\alpha=x^\alpha(\tau)$, where R_K is the radius of the internal circle at infinity. Substituting this ansatz into (6.5) and integrating over s reduces the Nambu–Goto action into that of a relativistic particle with rest mass $m=mR_K/\alpha'$:

$$S=-\left(\frac{mR_K}{\alpha'}\right)\int\sqrt{-h_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}}\,d\tau, \tag{6.6}$$

where the new metric $h_{\mu\nu}$ is $h_{\mu\nu}=g_{\mu\nu}/V$. If, in particular, the original K–K metric is that of a Kaluza–Klein monopole $V=1+R_k/2r$, the new metric is simply

$$ds^2=-\frac{dt^2}{1+R_k/2r}+d\mathbf{r}^2. \tag{6.7}$$

The geodesics correspond to the Hamiltonian

$h=\sqrt{(\mathbf{p}^2+\mu^2)/V}$ and phase space symplectic structure $d\mathbf{r}\wedge d\mathbf{p}$, where $\mathbf{p}=\mu\, d\mathbf{x}/dt$. Gibbons and Ruback[27] then point out that the geodesics of (6.7) lie in the plane perpendicular to the conserved angular momentum $\mathbf{j}=\mathbf{r}\times\mathbf{p}$ and are ellipses, parabolas, or hyperbolas depending on the energy *square* e^2 being smaller, equal, or larger than the rest-mass square μ^2. This is explained by the conservation of a "Runge–Lenz" vector

$$\mathbf{a}=\mathbf{p}\times\mathbf{j}-(\mathbf{r}/r)(e^2R_k/4). \tag{6.8}$$

Furthermore, $\mathbf{j}$ and $\mathbf{k}=\mathbf{a}/\sqrt{(\mu^2-h^2)}$ generate a *Kepler*-type (in contrast to the "MIC–Zwanziger-type") o(4)/o(3,1) dynamical symmetry. The energy levels

$$e^2=\mu^2(8n^2/\mu^2R_K^2)\,(\sqrt{1+\mu^2R_K^2/4n^2}-1) \tag{6.9}$$

($n=1,2,\ldots$) are n^2 degenerate.

Our method provides an insight into the above statements. One inverts the energy relation

$$n=(R_k/4)(e^2/\sqrt{\mu^2-e^2}). \tag{6.10}$$

Define now a transformation $(\mathbf{R},\mathbf{P},T)=f(\mathbf{r},\mathbf{p},t)$:

$$\mathbf{R}=\sqrt{\mu^2-h^2}\,\mathbf{r},\quad \mathbf{P}=\mathbf{p}/\sqrt{\mu^2-\eta^2},$$
$$T=\frac{4}{R_k}\frac{\sqrt{\mu^2-h^2}}{2\mu^2-h^2}\left[\frac{\mu^2-h^2}{h}t+\mathbf{r}\cdot\mathbf{p}\right]. \tag{6.11}$$

It is easy to see that (6.11) is canonical, $d\mathbf{R}\wedge d\mathbf{P}+dH\wedge dT=d\mathbf{r}\wedge d\mathbf{p}+dh\wedge dt$, if the new Hamiltonian is

$$H=(R_k/4)(h^2/\sqrt{\mu^2-h^2}). \tag{6.12}$$

Substituting h, expressed by $\mathbf{R}$ and $\mathbf{P}$, into (6.12), we obtain

$$H=\Gamma_0=\tfrac{1}{2}R(\mathbf{P}^2+1), \tag{6.13}$$

which is the SU(2,2) generator (3.11a) for helicity $s=0$, i.e., the Hamiltonian of the geodesic flow on S^3 expressed in stereographic coordinates.

We conclude that f in (6.11) is an (injective) symplectic mapping from the "reduced string system" into the mass-zero helicity-zero SU(2,2) orbit $\mathcal{O}_0\simeq T^+S^3$, which is[16,18] the space of regularized motions of the *Kepler problem*. In this case $\mathbf{R}$ and $\mathbf{P}$ are only local coordinates obtained by stereographic projection. Now f is not onto; those points not in $\mathrm{Im}(f)$ can be used to regularize the problem along the same lines as before.

It follows that the geodesics of the metric (6.7) have an o(4,2) conformal symmetry, with the generators (3.11) (for $s=0$).

VII. CONCLUSION

In this paper we have only studied the classical mechanics. Quantum aspects are found in Refs. 1–4 and could (in principle) be obtained from implementing the canonical transformation (4.1) at the quantum level.

The complications arise because of the collisions, which require regularization. The quantum motions actually behave better than the classical ones: Intuitively, the Heisenberg uncertainty relations make the collisions irrelevant. Remarkably, it is for the radial motions that the "Atiyah–Hitchin" and "Taub–NUT" motions differ the most.[1]

The Taub–NUT approximation is only valid for large distances, when the exponential terms are small with respect to those in r^{-1}. In the "real" (Atiyah–Hitchin) case the relative electric charge may not be conserved; the trajectories may not stay in a plane, etc.[1,28] However, numerical as well as theoretical calculations[28] show that the system still admits bound motions; for large angular momentum the real spectrum is very close to the one in the Taub–NUT limit.

An isospinor test particle in the long-range field of a monopole[5,11] has similar properties. In particular, for large angular momentum, the "real" bound motions peak far outside the monopole core and the spectrum quickly converges to the "MIC–Zwanziger" one.[29]

Finally, notice that the evolution space formalism has been useful in the past in understanding the symmetries of the Dirac monopole.[30]

[1]N. S. Manton, Phys. Lett. **110B**, 54 (1982); Phys. Lett. **154B**, 397 (1985); Phys. Lett. **157B**, 475 (1985); M. F. Atiyah and N. Hitchin, Phys. Lett. **107A**, 21 (1985); *The Geometry and Dynamics of Magnetic Monopoles* (Princeton U. P., Princeton, 1988); G. W. Gibbons and N. S. Manton, Nucl. Phys. B **274**, 183 (1986).

[2]L. Gy. Fehér and P. A. Horváthy Phys. Lett. **182B**, 183 (1987); **188B**, 512 (1987).

[3]G. W. Gibbons and P. Ruback, Phys. Lett. **188B**, 226 (1987); Commun. Math. Phys. **115**, 267 (1988); L. Gy. Fehér, in *Relativity Today, Proceedings of the* 2nd *Hungarian Relativity Workshop, Budapest,1987*, edited by Z. Perjés (World Scientific, Singapore, 1988).

[4]B. Cordani, L. Gy. Fehér, and P. A. Horváthy, Phys. Lett. **201B**, 481 (1988).

[5]L. Gy. Fehér, J. Phys. A **19**, 1259 (1986); in *Proceedings of the Siofok Conference on Non-Perturbative Methods in Quantum Field Theory*, edited by Z. Horváth, L. Palla, and A. Patkos (World Scientific, Singapore, 1987); L. Gy. Fehér and P. A. Horváthy, Mod. Phys. Lett. A **3**, 1451 (1988).

[6]D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 29 (1983); R. Sorkin, Phys. Rev. Lett. **51**, 87 (1983).

[7]J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).

[8]R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968); S. K. Wong, Nuovo Cimento A **65**, 689 (1970); C. Duval and P. A. Horváthy, Ann. Phys. (NY) **142**, 10 (1982).

[9]P. A. Horváthy and J. H. Rawnsley, Phys. Rev. D **32**, 968 (1985); C. H. Oh, C. P. Soo, and C. H. Lai, Phys. Rev. D **36**, 2533 (1987).

[10]H. V. McIntosh and A. Cisneros, J. Math. Phys. **11**, 896 (1970); D. Zwanziger, Phys. Rev. **176**, 1480 (1968).

[11]J. F. Schönfeld, J. Math. Phys. **21**, 2528 (1980).

[12]L. Bates, Calgary Univ. preprint 639 (1987).

[13]W. Pauli, Z. Phys. **36**, 336 (1926).

[14]P. Kustaanheimo and E. Stiefel, E. J. Reine Angew. Math. **218**, 204 (1965); E. Stiefel and G. Schiefele, *Linear and Regular Celestial Mechanics* (Springer, Berlin, 1971); F. A. E. Pirani, Nuovo Cimento B **19**, 189 (1974).

[15]A. O. Barut and G. L. Bornzin, J. Math. Phys. **12**, 841 (1971); E. D'Hoker and L. Vinet, Nucl. Phys. B **260**, 79 (1985).

[16]J.-M. Souriau, *Structure des Systèmes Dynamiques* (Dunod, Paris, 1969); R. Abraham and J. E. Marsden *Foundations of Mechanics* (Benjamin, Reading, MA, 1978).

[17]B. Cordani, Commun. Math. Phys. **103**, 403 (1986).

[18]J.-M. Souriau, Symp. Math. **14**, 343 (1974); J.-M. Souriau, in *Proceedings of the IUTAM–ISSIM Symposium on Modern Developments in Analytical Mechanics*, edited by M. Francaviglia, Torino, 1982, Atti Acad. Sci. Torino **117**, 369 (1983).

[19]R. Penrose and M. A. H. McCallum, Phys. Rep. C **6**, 241 (1972); R. Penrose, Rep. Math. Phys. **12**, 65 (1977); S. A. Huggett and K. P. Tod, *An Introduction to Twistor Theory* (Cambridge U. P., Cambridge, 1985).

[20]N. M. J. Woodhouse, in *Conference in Group Theoretical Methods in Physics*, Nijmegen, 1975, edited by A. Janner, T. Jansen, and M. Boon, Springer Lecture Notes in Physics Vol. 50 (Springer, Berlin, 1976).

[21]B. Cordani, J. Math. Phys. **27**, 2920 (1986).

[22]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN-Polish Scientific, Warsaw, 1977); I. T. Todorov, *Conformal Description of Spinning Particles* (Springer, Berlin, 1986).
[23]V. A. Fock, Z. Phys. **98**, 145 (1935); J. Moser, Commun. Pure. Appl. Math. **23**, 609 (1970).
[24]H. Bacry, Nuovo Cimento A **41**, 222 (1967); G. Györgyi, Nuovo Cimento A **53**, 717 (1968).
[25]G. Clément, Alger Univ. preprint (1988).
[26]C. W. Misner, J. Math. Phys. **4**, 924 (1963).
[27]G. W. Gibbons and P. Ruback, Cambridge Univ. preprint (1988).
[28]N. S. Manton, Phys. Lett. **198B**, 226 (1987); M. Woitkowski, Bull. Am. Math. Soc. **xx**, L. Bates and R. Montgomery, Commun. Math. Phys. **118**, 635 (1988); M. Temple-Raston, Cambridge Univ. preprint DAMTP/88-15 (1988).
[29]L. Gy. Fehér, I. Gyémát, and P. A. Horváthy, (1987) (unpublished).
[30]R. Jackiw, Ann. Phys. (NY) **129**, 183 (1980); P. A. Horváthy, J. Geom. Phys. **1**(3) 39 (1984); L. Gy. Fehér, J. Math. Phys. **28**, 234 (1987).

Stochastic field theory, holomorphic quantum mechanics, and supersymmetry

Pratul Bandyopadhyay and Kohinur Hajra
Indian Statistical Institute, Calcutta-700035, India

Pradip Ghosh
Maharaja Manindra Chandra College, Calcutta-700003, India

(Received 28 June 1988; accepted for publication 3 May 1989)

Holomorphic quantum mechanics are studied from the point of view of stochastic quantization in Minkowski space which involves the introduction of two stochastic fields, one in the external space and the other in the internal space. The equilibrium condition is given by Z_2 symmetry between the external and internal fields. In the nonequilibrium case, $N=2$ Wess–Zumino quantum fields are arrived at giving rise to supersymmetry. This helps to define the supercharge operator Q when the Hamiltonian is given by $H=Q^2$ and an index theorem is derived for an interacting case when the superpotential is given by $V(\phi)=\lambda\phi^n$, ϕ being complex with $n>2$. It is found that the vacuum is degenerate and is in conformity with the result obtained by Jaffe, Lesniewski, and Lewenstein [Ann. Phys. **178**, 313 (1987)] in the two-dimensional $N=2$ Wess–Zumino quantum field model.

I. INTRODUCTION

Recently Jaffe, Lesniewski, and Lewenstein[1] have considered the ground state structure of the two-dimensional $N=1$ and $N=2$ Wess–Zumino quantum field models and have pointed out that the $N=2$ quantum mechanics has degenerate vacua. The space of vacuum states is found to be bosonic and its dimension is determined by the topological properties of the superpotential. The physical interpretation of $N=2$ Wess–Zumino quantum mechanics has been discussed and the feasibility of realizing holomorphic quantum mechanics has been pointed out with special reference to a spin $\frac{1}{2}$ particle in an external SU(2) gauge field and in the study of nuclear matter interacting with a pion condensate. Here we shall show that holomorphic quantum mechanics is realized in stochastic field theory also when stochastic quantization is achieved in Minkowski space, introducing a doublet of fields corresponding to the fields in the external and internal space. This can also be generalized to finite temperature when the formalism of thermofield dynamics is utilized identifying the internal field with the fictitious tilde field introduced by Takahashi and Umezawa.[2] We shall study here the supersymmetric properties of such fields when the equilibrium condition of Z_2 symmetry between the external and internal field is destroyed and shall show that we can uniquely define a supercharge for such a system. The index theorem for such a system representing holomorphic quantum mechanics is then discussed and it is found that the space of vacuum states has its dimension determined by the topological properties of the superpotential in conformity with the results obtained by Jaffe, Lesniewski, and Lewenstein[1] in the two-dimensional Wess–Zumino quantum field model.

In a recent paper,[3] it has been pointed out that stochastic quantization in Minkowski space as well as its generalization at finite temperature leading to the formalism of thermofield dynamics necessitates the introduction of a doublet of stochastic fields. This doublet can be interpreted as comprising two fields, one corresponding to the field in the external space and the other corresponding to the field in the internal space. This internal field is also necessary to have a relativistic generalization of Nelson's formalism of stochastic quantization and the quantization of a Fermi field.[4] The equilibrium condition for such a doublet of stochastic fields is given by the Z_2 symmetry corresponding to the time reversal symmetry of the two fields. The nonequilibrium condition gives rise to supersymmetric quantum mechanics.

Here we shall point out that the doublet of stochastic fields may be taken to give rise to holmorphic quantum mechanics in four dimensions and the break down of Z_2 symmetry gives rise to the $N=2$ Wess–Zumino quantum field model. Moreover it is found that the two-dimensional result of Jaffe *et al.* regarding the degeneracy of the ground state except when the superpotential is quadratic is also valid here and the index $i(Q_t)$ is found to be identical with the degree of ∂V, where V is a polynomial of degree $n\geqslant 2$.

In Sec. II we shall recapitulate the main features of stochastic quantization in Minkowski space and its generalization to finite temperature utilizing the formalism of thermofield dynamics. In Sec. III we shall formulate holomorphic quantum mechanics from stochastic field theory and shall derive supercharge for such a system. In Sec. IV we shall discuss the index theorem.

II. STOCHASTIC QUANTIZATION IN MINKOWSKI SPACE

Nelson's stochastic quantization procedure is based on the assumption that the configuration variable $q(t)$ is promoted to a Markov process $q(t)$.[5] The process $q(t)$ is determined by two conditions; the first is the hypothesis of universal Brownian motion and the second is the validity of the Euler–Lagrange equation. In a recent paper,[4] it has been shown that in Nelson's formalism, the relativistic generalization as well as the quantization of a Fermi field is achieved

 0022-2488/90/010212-09$03.00

when an anisotropy in the internal space of a particle is introduced and it is taken into account that there are universal Brownian motions both in the external and internal space. In this formalism, the configuration variables are denoted as $Q(t,\xi_0)$, where ξ_0 is the fourth component (real) of the internal four-vector ξ_μ which is considered to be the attached vector to the space-time point x_μ. We assume that $Q(t,\xi_0)$ is a separable function and can be denoted as

$$Q(t,\xi_0) = q(t)q(\xi_0). \tag{1}$$

The process $Q(t,\xi_0)$ is assumed to satisfy the stochastic differential equations,

$$dQ_i(t,\xi_0) = b_i(Q(t,\xi_0),t,\xi_0)dt + dw_i(t), \tag{2}$$

$$dQ_i(t,\xi_0) = b'_i(Q(t,\xi_0),t,\xi_0)d\xi_0 + dw_i(\xi_0), \tag{3}$$

where $b_i(Q(t,\xi_0),t,\xi_0)$ and $b'_i(Q(t,\xi_0),t,\xi_0)$ correspond to certain velocity fields and dw_i are independent Brownian motions. It is assumed that $dw_i(t)(dw_i(\xi_0))$ does not depend on $Q(s,s')$ for $s\leqslant t$ $(s'\leqslant\xi_0)$ and the expectations have the following values at $T=0$ and $T\neq 0$:

$$\langle dw_i(t)\rangle_{T=0} = 0, \tag{4}$$

$$\langle dw_i(t)dw_j(t')\rangle_{T=0} = (\hbar/m)\delta_{ij}\delta(t-t')dt\,dt', \tag{5}$$

$$\langle dw_i(\xi_0)\rangle_{T=0} = 0, \tag{6}$$

$$\langle dw_i(\xi_0)dw_j(\xi'_0)\rangle_{T=0} = (\hbar/\pi^0)\delta_{ij}\delta(\xi_0-\xi'_0)d\xi_0\,d\xi'_0, \tag{7}$$

$$\langle dw_i(t)\rangle_{T\neq 0} = 0, \tag{8}$$

$$\langle dw_i(r)dw_j(t')\rangle_{T\neq 0} = \frac{\delta_{ij}}{\beta m}\sum_{n=\alpha}^{-\alpha} e^{iw_n(t-t')}dt\,dt^1, \tag{9}$$

$$\langle dw_i(\xi_0)\rangle_{T\neq 0} = 0, \tag{10}$$

$$\langle dw_i(\xi_0)dw_j(\xi'_0)\rangle_{T\neq 0} = \frac{\delta_{ij}}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha} e^{iw_n(\xi_0-\xi'_0)}d\xi_0\,d\xi'_0, \tag{11}$$

with $w_n = 2\pi n/\beta\hbar$.

It is easily seen that in the limit $\beta\to\alpha$ Eqs. (9) and (11) give Eqs. (5) and (7), respectively. The form of Eqs. (9) and (11) is dictated by the KMS condition. To make the description time symmetrical in both "external" and "internal" time we also write

$$dQ_i(t,\xi_0) = b^*_i(Q(t,\xi_0),t,\xi_0)dt + dw^*_i(t), \tag{12}$$

$$dQ_i(t,\xi_0) = b'^*_i(Q(t,\xi_0),t,\xi_0)d\xi_o + dw^*_i(\xi_0), \tag{13}$$

where dw^* has the same properties as dw except that $dw^*_i(t)(dw^*_i(\xi_0))$ are independent of $Q(s,s')$ with $s>t$ $(s'\geqslant\xi_0)$.

From the stochastic differential equations considered here, the following moments can be derived.

$$\langle Q_i(t,\xi_0)\rangle_{T=0} = \langle Q_i(t,\xi_0)\rangle_{>T\neq 0} = 0, \tag{14}$$

$$\langle Q_i(t,\xi_0)Q_i(t',\xi'_0)\rangle_{T=0} = \frac{\hbar}{2mw}\frac{\hbar}{2\pi^0 w'}\delta_{ij}e^{-w(t-t')}e^{-w'(\xi_0-\xi'_0)} \quad (t>t',\xi_0>\xi'_0), \tag{15}$$

$$\langle Q_i(t,\xi_0)Q_j(t',\xi'_0)\rangle_{T\neq 0} = \frac{\delta_{ij}}{\beta m}\sum_{n=-\alpha}^{\alpha}\frac{e^{iw_n(t-t')}}{w^2 t w_n^2} \times\frac{1}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha}\frac{e^{iw_n(\xi_0-\xi'_0)}}{w'^2 t w_n^2}, \tag{16}$$

with $w_n = 2\pi n/\beta\hbar$.

This follows from the fact that $Q_i(t,\xi_0)$ can be written in a separable way $q_i(t)q_i(\xi_0)$ and we can utilize the results for the moments of $q_i(t)$ as derived by Moore[6]

$$\langle q_i(t)\rangle_{T=0} = \langle q_i(t)\rangle_{T\neq 0} = 0, \tag{17}$$

$$\langle q_i(t)q_i(t')\rangle_{T=0} = \frac{\hbar}{2mw}\delta_{ij}e^{-w(t-t')} \quad (t>t'), \tag{18}$$

$$\langle q_i(t)q_i(t')\rangle_{T\neq 0} = \frac{\delta_{ij}}{\beta m}\sum_{n=-\alpha}^{\alpha}\frac{e^{iw_n(t-t')}}{w^2+w_n^2}. \tag{19}$$

These results can be extended to the variable $q_i(\xi_0)$ in an analogous way:

$$\langle q_i(\xi_0)\rangle_{T=0} = \langle q_i(\xi_0)\rangle_{T\neq 0} = 0, \tag{20}$$

$$\langle q_i(\xi_0)q_j(\xi'_0)\rangle_{T=0} = \frac{\hbar}{2\pi^0 w'}\delta_{ij}e^{-w'(\xi_0-\xi'_0)}(\xi_0>\xi'_0), \tag{21}$$

$$\langle q_i(\xi_0)q_j(\xi'_0)\rangle_{T\neq 0} = \frac{\delta_{ij}}{B\pi^0}\sum_{n=-\alpha}^{\alpha}\frac{e^{iw_n(t-t')}}{w'^2+w_n^2}. \tag{22}$$

Let $\{e_i(\mathbf{x})\}$ denote the complete orthonormal set of eigenfunctions of the three-dimensional Laplacian $-\Delta$:

$$\Delta e_i(\mathbf{x}) = -\mathbf{k}_i^2 e_i(\mathbf{x}). \tag{23}$$

Also we denote $\{e_j(\xi)\}$ as the set of complete orthonormal set of eigenfunctions of the three-dimensional Laplacian $-\Delta'$ in terms of the variables ξ_i

$$\left(\Delta' = \frac{\partial^2}{\partial\xi_1^2}+\frac{\partial^2}{\partial\xi_2^2}+\frac{\partial^2}{\partial\xi_3^2}\right)$$

so that

$$\Delta' e_j(\xi) = -\pi_j^2 e_j(\xi). \tag{24}$$

Now we can construct a stochastic field ϕ which can be expressed as an orthonormal expansion in terms of $q_i(t),e_i(\mathbf{x}), q_j(\xi_0), e_j(\xi)$ and write

$$\phi(x,t,\xi) = \sum_{i,j} q_i(t)e_i(\mathbf{x})q_j(\xi_0)e_j(\xi). \tag{25}$$

Now from the moments of $q_i(t),q_j(\xi_0)$ we can determine the moments of $\phi(x,t,\xi)$,

$$\langle\phi(x,t,\xi)\rangle_{T=0} = \langle\phi(x,t,\xi)\rangle_{T=0}\neq 0, \tag{26}$$

$$\langle\phi(x,t,\xi)\phi(x',t',\xi')\rangle = \frac{1}{(2\pi)^3}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}g(t-t') \otimes\frac{1}{(2\pi)}{}^3\int d^3\boldsymbol{\pi}\,e^{i\boldsymbol{\pi}(\xi-\xi')}g(\xi_0-\xi'_0), \tag{27}$$

where $g(t-t')$ and $g(\xi_0-\xi'_0)$ are given by Eqs. (18) and (21) for $T=0$ and by Eqs. (19) and (22) for $T\neq 0$. Substituting these relations, we find

$$\langle\phi(x,t,\xi)\phi(x',t',\xi')\rangle_{T=0} = \frac{1}{(2\pi)^4}\int \frac{d^4x\, e^{i(k,(x-x'))}}{(k,k)+m^2} \times \frac{1}{(2\pi)^4}\int \frac{d^4\pi\, e^{i(\pi,(\xi-\xi'))}}{(\pi,\pi)+\pi^{02}}. \tag{28}$$

Here (A,B) denotes a Euclidean product and the units have been chosen to be $\hbar = m = \pi^0 = 1$. In the case of $T \neq 0$, we find

$$\langle\phi(x,t,\xi)\phi(x',t',\xi')\rangle_{T\neq 0} = \frac{1}{(2\pi)^3}\frac{1}{\beta m}\int d^3\mathbf{k}\, e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\sum_{n=-\alpha}^{\alpha}\frac{e^{iw_n}(t-t')}{w^2+w_n^2} \times \frac{1}{(2\pi)}{}^3\frac{1}{\beta\pi^0}\int d^3\boldsymbol{\pi}\, e^{i\boldsymbol{\pi}(\xi-\xi')}\sum_{n=-\alpha}^{\alpha}\frac{e^{iw_n}(\xi_0-\xi_0')}{w_1{}^2+w_n^2}. \tag{29}$$

Now for a particular mode $n = 1$, we find that the expression becomes

$$\frac{1}{(2\pi)^3}\frac{1}{\beta m}\int d^3\mathbf{k}\, e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\frac{e^{ik_0(t-t')}}{k_0^2+w^2}\delta(k_0-w_1)dk_0 \times \frac{1}{(2\pi)^3}\frac{1}{\beta\pi^0}\int d^3\boldsymbol{\pi}\, e^{i\boldsymbol{\pi}(\xi-\xi')} \times \frac{e^{i\pi_0}(\xi_0-\xi_0')}{\pi_0^2+w'^2}\delta(\pi_0-w_1)d\pi_0 = \frac{1}{\beta m}\frac{1}{(2\pi)^4}\int \frac{d^4k\, e^{i(k,(x-x'))}}{(k,k)+m^2}2\pi\delta(k_0-w_1) \times \frac{1}{\beta\pi^0}\frac{1}{(2\pi)^4}\int \frac{d^4\pi\, e^{i(\pi,(\xi-\xi'))}}{(\pi,\pi)+\pi^{02}}2\pi\delta(\pi_0-w_1), \tag{30}$$

where π^0 corresponds to the quantity in the internal space analogous to the mass of the system.

Now from the relations (21) and (27) we note that for $\xi_0 = \xi_0' = 0$ and integrating over the internal space variable $\boldsymbol{\xi}$ the correlation function just reduces to that of the scalar field in Euclidean space,

$$\langle\phi(x,t)\phi(x',t')\rangle_{T=0} = \frac{1}{(2\pi)^4}\int \frac{d^4k\, e^{i(k,(x-x_1))}}{(k,k)+m^2}. \tag{31}$$

In a similar way for $T \neq 0$, we find from (22) and (27) considering one particular mode $n = 1$,

$$\langle\phi(x,t)\phi(x',t')\rangle_{T=0} = \frac{1}{(2\pi)^4}\int \frac{d^4k\, e^{i(k,(x-x'))}}{(k,k)+m^2}2\pi\delta(k_0-w_1) \tag{32}$$

normalizing $\beta = \pi^0 = m = 1$.

This is the Euclidean Markov field result which has been obtained from Nelson's real time formalism of Brownian motion and in this sense gives rise to the equivalence of these two formalisms as advocated by Guerra and Ruggiro.[7]

Now if we introduce an anisotropic feature of the internal space-time corresponding to the variable ξ_μ, we can obtain the fermionic propagator in Euclidean space-time. To this end, we introduce the anisotropy by having two opposite orientations of the internal variable ξ_μ (and hence of $\pi_\mu = i\delta/\delta\xi_\mu$) and take into account that each orientation denotes a separate field and the two opposite orientations depict two separate fields having two internal helicities corresponding to particle and antiparticle configurations. From Eq. (28), we note that it is effectively a correlation function in eight-dimensional space-time, four dimensional in the external space-time variable and four dimensional in the internal space-time variable. To make it an effective four-dimensional expression in the external space-time variable we may take into account that $k(x)$ is an implicit function of $\pi(\xi)$. For simplicity and dimensional reasons we take the form $k^2 = (k',\pi)$, $m^2 = m'\pi^0$, where (k',π) is the Euclidean product and each component of k is given by $k_i = \sqrt{k_i'}\pi_i$. So from the new field variable $\bar{\phi}(x,t,\xi)$ where this mapping is taken into account, we find from Eq. (31) the correlation function for $T = 0$,

$$\langle\bar{\phi}(x,t,\xi)\bar{\phi}(x',t',\xi')\rangle_{T=0} = \int \frac{e^{i(\sqrt{k'}\pi)(x(\xi)-x'(\xi'))}}{(k',\pi)+m'\pi^0}d^4\sqrt{k'}\pi = \int \frac{e^{i(\sqrt{k'}\pi)(x(\xi)-x'(\xi'))}}{(i\sqrt{k'},\pi)+\sqrt{m'}\pi^0)(-i\sqrt{(k'},\pi)+\sqrt{m'}\pi^0)} \times d^4\sqrt{k'}\pi. \tag{33}$$

Now taking into account that $i\sqrt{\pi}$ and $-i\sqrt{\pi}$ correspond to two different internal helicity states and denote two separate fields and particle and antiparticle states, for a single particle state with a specific internal helicity, we should take $-i\sqrt{\pi}$ (or $i\sqrt{\pi}$) as a vanishing term. Taking $-i\sqrt{\pi} = 0$, we see that the expression (33) reduces to the form

$$\langle\bar{\phi}(x,t,\xi)\bar{\phi}(x',t',\xi')\rangle_{T=0} = \frac{1}{(2\pi)^4}\int \frac{e^{i(k,(x-x'))}}{i\sqrt{k^2}+m}d^4k, \tag{34}$$

where we have chosen the unit $m = \pi^0 = 1$.

Now we can choose a matrix $\gamma_\mu k_\mu + m = \not{k} + m$ with two degenerate eigenvalues $\pm i\sqrt{k^2}+m$, which can be diagonalized by a unitary matrix U:

$$\not{k}+m = U^{-1}\begin{pmatrix} i\sqrt{k^2}+m & 0 & 0 & 0 \\ 0 & i\sqrt{k^2}+m & 0 & 0 \\ 0 & 0 & -i\sqrt{k^2}+m & 0 \\ 0 & 0 & 0 & -i\sqrt{k^2}+m \end{pmatrix}U \tag{35}$$

Thus we just get the fermionic propagator in Euclidean space-time

$$\langle\bar{\phi}(x,t,\xi)\bar{\phi}(x',t',\xi')\rangle_{T=0} = \frac{1}{(2\pi)^4}\int \frac{d^4k\, e^{i(k(x-x')}}{\not{k}+m}. \tag{36}$$

This shows that when a direction vector giving rise to an internal helicity in an anisotropic microlocal space-time is taken into account, we can have the quantized fermionic field from a Brownian motion process. This result will be valid for $T \neq 0$ also. Indeed, in a similar way we find from Eq. (32) for a particular mode $n = 1$,

$$\langle\bar{\phi}(x,t,\xi)\bar{\phi}(x',t',\xi')\rangle_{T=0} = \frac{1}{(2\pi)^4}\int \frac{d^4k\, e^{i(k,(x-x'))}}{\not{k}+m} 2\pi\delta(k_0 - w_1) \tag{37}$$

normalizing $\beta = \pi^0 = m = 1$.

From this analysis, it is noted that the statistics of the particle depend on the internal space-time variable ξ_μ. That is, when ξ_μ appears as a direction vector with a fixed orientation in the structure of the particle so that it gives rise to two opposite internal helicities which correspond to particle and antiparticle states, we get a Fermi field. Indeed, the fermion number is associated with this internal helicity. Again when there is no anisotropy in the internal space so that there is no manifestation of ξ_μ in the external space we get a boson. Now the effect of temperature should definitely affect the internal motion and as such it may happen that at high temperature the anisotropic feature of the internal space will be destroyed and the fermion will be transformed into a boson. This is, a massive extended body depicting a fermion can have such a phase transition. However, this does not mean that fermion number conservation will be violated as Lorentz invariance and the equilibrium condition will not allow such a process to occur. The only effect of such a phase transition will be that a thermal pair of opposite statistics will emerge as zero energy modes at the critical temperature possibly leading to a nonequilibrium state corresponding to a supersymmetric phase.[8] Indeed the stochastic nonlocal field $\phi(x_\mu,\xi_\mu)$ which is assumed to satisfy the condition of separability $\phi(x_\mu,\xi_\mu) = \phi(x_\mu)\phi(\xi_\mu)$ can be written as a thermal doublet $\binom{\phi(x_\mu)}{\phi(\xi_\mu)}$ as the thermal effect on ξ_μ may be such that it may alter the statistics of the particle. However, though x and ξ represent two different spaces, yet as the external motion may be thought to be a manifestation of the internal motion, a mapping of x and ξ is possible. In that case x may be represented in the functional form $x(\xi)$ and the simplest form of the mapping can be taken to be $x = c\xi$, where c is a suitable constant. In view of this, there should be a mapping of $\phi(x)$ and $\phi(\xi)$ also. We can assume that $\phi(\xi) = \lambda\phi^\dagger(x) = \tilde{\phi}^\dagger(x)$, where λ is a suitable parameter. Thus the thermal doublet $\binom{\phi(x)}{\phi(\xi)}$ can be written as $\binom{\phi(x)}{\phi^\dagger(x)}$. This helps us to consider that there exists a conjugate Hilbert space $\tilde{H}$ associated with the Hilbert space H such that $\tilde{H}$ is the set H with the scalar multiplication $\lambda,\xi \to \bar{\lambda}\xi$, where $\lambda \in C$ and $\xi \in H$ and with scalar product $(\xi,\eta) \to (\overline{\xi,\eta})$ with $\xi,\eta \in H$ and $(\xi,\eta) \to (\xi,\eta)$ is the scalar product of H. In effect H is the Hilbert space associated with the external space and $\tilde{H}$ is the conjugate Hilbert space associated with the internal space.

Now we write the bosonic field function in terms of the thermal doublets

$$\phi(x) = \begin{pmatrix}\phi(x)\\ \phi(\xi)\end{pmatrix} = \begin{pmatrix}\phi(x)\\ \tilde{\phi}^\dagger(x)\end{pmatrix} = \int \frac{d^4p}{(2\pi)^3}\theta(p_0)\delta(p^2-m^2) \times\left\{\begin{pmatrix}a_+(\mathbf{p})\\ \tilde{a}^\dagger_+(\mathbf{p})\end{pmatrix}e^{-ipx} + \begin{pmatrix}a^\dagger_-(\mathbf{p})\\ \tilde{a}_-(\mathbf{p})\end{pmatrix}e^{ipx}\right\}, \tag{38}$$

$$\phi'(x) = \begin{pmatrix}\phi^\dagger(x)\\ \phi^\dagger(\xi)\end{pmatrix} = \begin{pmatrix}\phi^\dagger(x)\\ \tilde{\phi}(x)\end{pmatrix} = \int \frac{d^4p}{(2\pi)^3}\theta(p_0)\delta(p^2-m^2) \times\left\{\begin{pmatrix}a_-(\mathbf{p})\\ \tilde{a}^\dagger_-(\mathbf{p})\end{pmatrix}e^{-ipx} + \begin{pmatrix}a^\dagger_+(\mathbf{p})\\ \tilde{a}_+(\mathbf{p})\end{pmatrix}e^{ipx}\right\}. \tag{39}$$

In the case of fermions, we have to introduce the anisotropic feature in the internal space so that it can generate two internal helicities, corresponding to particle and antiparticle, and in view of this we can obtain the Dirac propagator when the external space-time variable x_μ is considered to be a function of the internal space-time variable ξ_μ. But it may be remarked that we may do the opposite also, i.e., the internal space-time variable ξ_μ may be taken to be a function of the external space-time variable x_μ, and we may obtain the Dirac propagator in the internal variable ξ_μ and conjugate π_μ when an anisotropy is introduced in its attached vector x_μ. That is, we can write Dirac functions $\psi(x)$ and $\psi(\xi)$ in Hilbert spaces H and $\tilde{H}$, respectively, in a symmetric way. It may be noted that when at high temperature the anisotropy of the internal ξ space is destroyed, the spinorial characteristic of the field $\psi(x_\mu)$ which is acquired through the anisotropy of the attached vector ξ_μ will be changed to a bosonic one, but the spinorial characteristic of the conjugate field $\psi(\xi_\mu)$ which is acquired through the anisotropy of the attached vector x_μ will not be altered and as such we will have a thermal doublet of opposite statistics. This indicates that at that critical temperature, we will have a nonequilibrium state corresponding to a supersymmetric phase and as such fermion number conservation will not be violated due to such a phase transition.[9] This is similar to the features of Z_2 symmetry which arises in the finite temperature formalism of quantum field theory in Minkowski space as proposed by Niemi and Semenoff.[8] Indeed, the field function in the internal space here corresponds to the ghost field introduced by these authors and the corresponding Z_2 symmetry is manifested in the anisotropic feature of the internal space leading to the generation of two opposite internal helicities. As argued by Niemi and Semenoff, the broken Z_2 symmetry leads to a nonelquilibrium state. Our present formalism also leads to a similar situation when at the critical temperature, the internal helicity is destroyed leading to a nonequilibrium state. Now introducing the mapping $\psi(\xi) = \lambda\psi^\dagger(x) = i\tilde{\psi}^\dagger(x)$ we can write the Fermi field $\psi(x)$ in terms of the thermal doublets as follows.

$$\psi(x)=\begin{pmatrix}\psi\ (x)\\ i\psi^{\dagger}(x)\end{pmatrix}=\int\frac{d^4p}{(2\pi)^3}\theta(p_0)\delta(p^2-m^2)\sum_{\lambda=1,2}\left\{\begin{pmatrix}b_+\ (\mathbf{p},\lambda)\\ i\tilde{b}^{\dagger}_+\ (\mathbf{p},\lambda)\end{pmatrix}V(\mathbf{p},\lambda)e^{-ipx}+\begin{pmatrix}b^{\dagger}_-\ (\mathbf{p},\lambda)\\ i\tilde{b}_-\ (\mathbf{p},\lambda)\end{pmatrix}V(\mathbf{p},\lambda)e^{ipx}\right\},\tag{40}$$

$$\bar{\psi}(x)=\begin{pmatrix}\bar{\psi}(x)\\ -i\tilde{\psi}(x)\gamma_0\end{pmatrix}=\int\frac{d^4p}{(2\pi)}{}^3\theta(p_0)\delta(p^2-m^2)\sum_{\lambda=1,2}\left\{\begin{pmatrix}b_-(\mathbf{p},\lambda)\\ -i\tilde{b}^{\dagger}_-\ (\mathbf{p},\lambda)\end{pmatrix}\bar{V}(\mathbf{p},\lambda)e^{-ipx}+\begin{pmatrix}b'_+\ (\mathbf{p},\lambda)\\ -i\tilde{b}_+(\mathbf{p},\lambda)\end{pmatrix}\tilde{V}(\mathbf{p},\lambda)e^{ipx}\right\}.\tag{41}$$

This doubling of field may be suitably represented through the complexification of space-time variables. Indeed, if we write the doublet as

$$\begin{pmatrix}\phi(x)\\ \phi(\xi)\end{pmatrix}=\phi(x)+i\phi(\xi)\tag{42}$$

it implies that the space-time coordinate is given by $z=x+i\xi$. As we have constructed the stochastic fields from the stochastic variables $q(t)$ and $q(\xi_0)$ we can write for the configuration variable in complexified space-time

$$q(z_0)=q(t)+iq(\xi_0).\tag{43}$$

From this we have the correlations at $T=0$

$$\langle q(z_0)\rangle=\langle q(t)\rangle+i\langle q(\xi_0)\rangle=0,\tag{44}$$

$$\begin{aligned}\langle q(z_0)q(z_0')\rangle&=\langle q(t+i\xi_0)q(t'+i\xi_0')\rangle\\&=\langle q(t)q(t')\rangle-\langle q(\xi_0)q(\xi_0')\rangle\\&\quad+i[\langle q(t)q(\xi_0')\rangle+\langle q(\xi_0)q(t')\rangle].\end{aligned}\tag{45}$$

Now introducing the mapping $\lambda q(t)=q(\xi_0)$, we find

$$\langle q(z_0)q(z_0')\rangle=\langle q(t)q(t')\rangle[1-\lambda^2+2i\lambda\ ].\tag{46}$$

Noting that

$$\langle q(t)q(t')\rangle=(\hbar/2mw)e^{-w(t-t')},\tag{47}$$

we finally have

$$\langle q(z_0)q(z_0')\rangle=(\hbar/2mw)e^{-w(t-t')}[1+\lambda^2-2i\lambda\ ].\tag{48}$$

Now we can choose

$$\begin{aligned}(1-\lambda^2)e^{-w(t-t')}&=\cos w(t-t'),\\ 2\lambda e^{-w(t-t')}&=\sin w(t-t'),\end{aligned}\tag{49}$$

which implies that λ is a suitable function of the dimensionless variable $w(t-t')$ with the constraint

$$(1+\lambda^2)e^{-w(t-t')}=1.\tag{50}$$

So we can write

$$\langle q(z_0)q(z_0')\rangle=(\hbar/2mw)e^{iw(t-t')}.\tag{51}$$

As we have constructed the stochastic fields from the configuration variables through the relations

$$\begin{aligned}\phi(x)&=\sum_i e_i(\mathbf{x})q_i(t),\\ \phi(\xi)&=\sum_j e_i(\xi)q_j(\xi_0),\end{aligned}\tag{52}$$

where $e_i(\mathbf{x})\,[e_j(\xi)]$ are the set of orthonormal eigenfunctions of the Laplacian

$$\begin{aligned}-\Delta e_i(\mathbf{x})&=\mathbf{k}_i^2e_i(\mathbf{x}),\\ -\Delta' e_j(\xi)&=\pi_j^2e_i(\xi).\end{aligned}\tag{53}$$

We can now derive the two-point correlation using the mapping $x=c\xi$, $k=(1/c)\pi$, and $\lambda q(t)=q(\xi_0)$ for the complex field

$$\begin{aligned}\phi(z)&=\phi(x)+i\phi(\xi)=\phi(x)+i\lambda\phi^{\dagger}(x)\\&=\phi_R(x)+i\phi_I(x).\end{aligned}$$

Indeed, we find

$$\begin{aligned}\langle\phi(z)\phi(z')\rangle&=\langle(\phi_R(x)+i\phi_g(x))(Q_R(x')+i\phi_I(x'))\rangle\\&=\langle\phi_R(x)\phi_R(x')\rangle-\langle\phi_I(x)\phi_I(x')\rangle\\&\quad+i[\langle\phi_R(x)\phi_I(x')\rangle+\langle\phi_I(x)\phi_R(x')\rangle].\end{aligned}\tag{54}$$

Now using relation $\phi_I(x)=\lambda\phi^{\dagger}(x)$, we find from Eq. (54)

$$\langle q(z)\phi(z')\rangle=\langle\phi(x)\phi'(x')\rangle[1-\lambda^2+2i\lambda\ ].\tag{55}$$

At $T=0$, $\langle\phi(x)\phi^{\dagger}(x')\rangle$ is given by

$$\begin{aligned}&\langle\phi(x)\phi^{\dagger}(x')\rangle\\&=\frac{1}{(2\pi)^3}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\langle q(t)q(t')\rangle\\&=\frac{1}{(2\pi)^3}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\frac{1}{2w}e^{-w(t-t')}(\hbar=m=1).\end{aligned}\tag{56}$$

Now utilizing the relations (49) and (51), we can finally write from Eq. (55)

$$\begin{aligned}&\langle\phi(z)\phi(z')\rangle\\&=\frac{1}{(2\pi)^3}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\frac{1}{2w}e^{iw(t-t')}\\&=\frac{i}{(2\pi)^4}\int d^4k\,\frac{e^{ik(x-x')}}{k_0^2-w^2+i\epsilon}.\end{aligned}\tag{57}$$

When we write $\phi(z)=\phi_R(x)+i\phi_I(x)$ as the doublet, we find from Eqs. (54) and (57)

$$\begin{aligned}\left\langle\begin{pmatrix}\phi_R(x)\\ \phi_I(x)\end{pmatrix}(\phi_R(x'),\phi_I(x'))\right\rangle&=\begin{bmatrix}\langle\phi_R(x)\phi_R(x')\rangle & \langle\phi_R(x)\phi_I(x')\rangle\\ \langle\phi_I(x)\phi_R(x')\rangle & \langle\phi_I(x)\phi_I(x')\rangle\end{bmatrix}\\&=\frac{1}{(2\pi)^3}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\begin{bmatrix}\cos w(t-t')/2w & \sin w(t-t')/2w\\ \sin w(t-t')/2w & -\cos w(t-t')/2w\end{bmatrix}\\&=\frac{i}{(2\pi)^4}\int\frac{d^4k\,e^{ik(x-x')}}{k_0^2-w^2+iE}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}.\end{aligned}\tag{58}$$

Thus we get stochastic quantization in Minkowski space at $T=0$. This is identical with the result obtained in the path integral formulation at $T=0$ excepting the matrix $\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$. The matrix corresponds to reflection invariance representing Z_2 symmetry, which is the criterion for equilibrium condition. When this Z_2 symmetry is destroyed, we have nonequilibrium statistical mechanics and it corresponds to supersymmetric quantum mechanics.[9] This is in conformity with the idea that the current velocity in the internal space is related to the osmotic velocity in the external space which helps us to interpret the Heisenberg uncertainty relation from the inherent stochastic nature of the internal space-time variable.[10]

This analysis can be generalized to the fermionic case also. Indeed writing

$$\psi(z)=\psi(x)+i\psi(\xi)=\psi(x)+i_\lambda\psi^\dagger(x)=\psi_R(x)+i\psi_I(x)$$

and taking the doublet

$$\psi(z)=\begin{pmatrix}\psi(x)\\ \psi(\xi)\end{pmatrix}=\begin{pmatrix}\psi_R(x)\\ \psi_I(x)\end{pmatrix}$$

we will have correlations at $T=0$

$$\langle\psi(z)\psi(z')\rangle=\left\langle\begin{pmatrix}\psi_R(x)\\ \psi_I(x)\end{pmatrix}(\psi_R(x'),\psi_I(x'))\right\rangle=\frac{i}{(2\pi)^4}\int d^4p\,e^{ip(x-x')}\frac{1}{\not p-m+iE}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{59}$$

These results can now be generalized to finite temperature using the formalism of thermofield dynamics when we identify $\phi(\xi)=\phi_I(x)\ =\tilde\phi^\dagger(x)$ and $\psi(\xi)=\psi_I(x)\ =i\tilde\psi^\dagger(x)$.
From these we can write

$$\langle T(\phi(x)\phi^t(y))\rangle=\langle 0(\beta)|T\begin{pmatrix}\phi(x)\\ \tilde\phi^\dagger(x)\end{pmatrix}(\phi^\dagger(y),\tilde\phi(y))|0(\beta)\rangle$$

$$=\int\frac{d^4p}{(2\pi)^4}e^{-ip(x-y)}V_B(|\mathbf{p}|,\beta)\begin{pmatrix}1/(p^2-m^2+i0) & 0\\ 0 & -1/(p^2-m^2-i0)\end{pmatrix}V_B^\dagger(|\mathbf{p}|,\beta) \tag{60}$$

$$\langle\theta(x_0-y_0)[\phi(x),\phi^\dagger(y)]\rangle=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}i\theta(x^0-y^0)\Delta(x-y)=-i\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\Delta_{\rm ret}(x-y), \tag{61}$$

$$\langle T\psi(x)\bar\psi(y)\rangle=\langle 0(\beta)\left[T\begin{pmatrix}\psi(x)\\ i\tilde\psi^\dagger(x)\end{pmatrix}(\bar\psi(y),-i\tilde\psi(y)\gamma_0)\right]|0(\beta)\rangle=i\int\frac{d^4p}{(2\pi)^4}e^{-ip(x-y)}V_F(p,\beta)$$

$$\times\begin{pmatrix}1/(\not p-m+i0) & 0\\ 0 & 1/(\not p-m-i0)\end{pmatrix}V_F^\dagger(p,\beta), \tag{62}$$

$$\langle\theta(x^0-y^0)\{\psi(x),\bar\psi(y)\}\rangle=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}(i\gamma_\mu\gamma_\mu+m)i\theta(x^0-y^0)\Delta(x-y)=-\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}iS_{\rm ret}(x-y). \tag{63}$$

The matrices V_B and V_F are the coefficients of Bogoliubov transformations given by

$$V_B(|\mathbf{p}|,\beta)=\begin{pmatrix}\cosh\theta(|\mathbf{p}|,\beta) & \sin\theta(|\mathbf{p}|,\beta)\\ \sinh\theta(|\mathbf{p}|,\beta) & \cosh\theta(|\mathbf{p}|,\beta)\end{pmatrix}=\begin{pmatrix}1/\sqrt{1-e^{-\beta\epsilon(\mathbf{p})}} & e^{-\beta\epsilon(\mathbf{p})/2}/\sqrt{1-e^{-\beta\epsilon(\mathbf{p})}}\\ e^{-\beta\epsilon(\mathbf{p})/2}/\sqrt{1-e^{-\beta\epsilon(\mathbf{p})}} & 1/\sqrt{1-e^{-\beta\epsilon(\mathbf{p})}}\end{pmatrix}, \tag{64}$$

$$V_F(p,\beta)=\begin{pmatrix}\cos\nu(|\mathbf{p}|),\beta) & -\epsilon(p^0)\sin\nu(|\mathbf{p}|,\beta)\\ \epsilon(p_0)\sin\nu(|\mathbf{p}|,\beta) & \cos\nu(|\mathbf{p}|,\beta)\end{pmatrix}$$

$$=\begin{pmatrix}1/\sqrt{1+e^{-\beta\epsilon(p)}} & -\epsilon(p^0)e^{-\beta\epsilon(\mathbf{p})/2}/\sqrt{1+e^{-\beta\epsilon(\mathbf{p})}}\\ \epsilon(p^0)e^{-\beta\epsilon(\mathbf{p})/2}/\sqrt{1+e^{-\beta\epsilon(\mathbf{p})}} & 1/\sqrt{1+e^{-\beta\epsilon(p)}}\end{pmatrix} \tag{65}$$

with

$$\epsilon(\mathbf{p})=\sqrt{\mathbf{p}^2+m^2},\quad \epsilon(p_0)=\theta(p_0)-\theta(-p_0). \tag{66}$$

This suggests that in the case of free charged scalar field and free Dirac spinor field ψ, the total Lagrangians are given by

$$\bar L_\varphi=\alpha_\varphi-\tilde\alpha_\varphi$$

$$=\partial_\mu\phi^+\partial_\mu\phi-m^2\phi^\dagger\phi-\partial_\mu\tilde\phi^+\partial_\mu\tilde\phi+m^2\tilde\phi^2\phi^2$$

$$=\partial_\mu\phi^+\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\partial_\mu\phi-m^2\phi^+\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\phi, \tag{67}$$

$$\bar L_\varphi=\alpha_\psi-\tilde\alpha_\psi$$

$$=\bar\psi((i/2)\gamma_\mu\overleftrightarrow{\partial}_\mu-m)\psi-\tilde\psi((-i/2)\gamma_\mu\overleftrightarrow{\partial}_\mu-m)\tilde\psi$$

$$=\bar\psi\left(\frac{i}{2}\gamma_\mu\overleftrightarrow{\partial}_\mu-m\right)\psi. \tag{68}$$

It is noted that the vacuum is now temperature dependent and satisfies the relation

$$(H-\tilde H)|0(\beta)\rangle=0 \tag{69}$$

and the total Hamiltonian is given by

$$\bar H=H-\tilde H. \tag{70}$$

III. STOCHASTIC FIELD THEORY, HOLOMORPHIC QUANTUM MECHANICS, AND SUPERSYMMETRY

From our above analysis, we note that we can construct holomorphic quantum mechanics when for the configuration variable we take the complexified space-time $z_\mu = x_\mu + i\xi_\mu$. In fact, we can now depict the fields $\phi^{\pm}(z) = \phi(x) \pm i\phi(\xi)$ and can consider that $\phi^{\pm}$ is holomorphic in z. Now defining the operators

$$\partial = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial \xi}\right), \quad \bar{\partial} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial \xi}\right)$$

we can write for a free field, the Hamiltonian

$$H = -2\partial\bar{\partial} + m^2\phi^-\phi^+ . \tag{71}$$

Identifying $\phi(\xi)$ as $\tilde{\phi}^+(x)$ as discussed in the previous section, we note that the Hamiltonian H corresponds to the system of free fields where the Lagrangian is given by $\bar{L}_\phi$ in Eq. (67). Now if we identify $\phi^{\pm}(z) = \mp i\sqrt{2}\,\partial V$ we can construct two operators Q_+ and Q_- such that

$$Q_- = \begin{pmatrix} \partial V & i\partial \\ i\bar{\partial} & -(\partial V)^* \end{pmatrix},$$
$$Q_+ = \begin{pmatrix} (\partial V) & i\partial \\ i\bar{\partial} & -\partial V \end{pmatrix}, \tag{72}$$

and the Hamiltonian H given by Eq. (71) can be expressed as

$$H = \mathrm{Tr}\, Q_+Q_- \quad (m=1), \tag{73}$$

where

$$Q_+Q_- = (-\partial\bar{\partial} + |\partial V|^2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{74}$$

Since Q_+Q_- is the sum of two positive operators, it has no zero mode. Besides, we note that this maintains the equilibrium condition of the Z_2 symmetry between the external and internal fields as the expression is invariant under the transformations, $\phi(x) \to -\phi(\xi), x \to -\xi$.

However, from expression (72), we note that we can construct another operator Q_-Q_+ which is given by

$$Q_-Q_+ = Q_+Q_- + \begin{pmatrix} 0 & -i(\partial^2 V) \\ i(\partial^2 V)^* & 0 \end{pmatrix}. \tag{75}$$

This expression for Q_-Q_+ contains nondiagonal elements and the presence of the term $\partial^2 V$ breaks the reflection invariance $\phi(x) \to -\phi(\xi)$, $x \to -\xi$. Thus the system describes the nonequilibrium condition and corresponds to supersymmetric quantum mechanics. Indeed, we can now define an operator Q such that

$$Q = \begin{pmatrix} 0 & Q_- \\ Q_+ & 0 \end{pmatrix} \tag{76}$$

and we can construct the Hamiltonian H_S given by

$$H_S = Q^2 = \begin{pmatrix} Q_-Q_+ & 0 \\ 0 & Q_+Q_- \end{pmatrix}. \tag{77}$$

Evidently, the system given by the Hamiltonian H_S breaks down the Z_2 symmetry of reflection invariance of the external and internal fields. In fact, due to the presence of the operator Q_-Q_+ in H_S, it possesses zero modes as has been explicitly shown by Jaffe *et al.*[1] Thus we can define the supercharge Q such that the Hamiltonian is given by $H_S = Q^2$ when the stochastic field theory involving external and internal fields is described in terms of holomorphic quantum mechanics.

Moreover, following the procedure of Jaffe *et al.*,[1] we can show that this formalism of holomorphic quantum mechanics for stochastic field theory gives rise to $N=2$ Wess–Zumino quantum mechanics. In fact, we can also choose for the Hamiltonian H_S the following expression:

$$H_S = Q^2 = -\partial\bar{\partial} + |\partial V|^2 - \bar{\psi}_1\psi_1\,\partial^2 V - \bar{\psi}_2\psi_2(\partial^2 V)^*, \tag{78}$$

where

$$\psi_1 = \tfrac{1}{2}(\gamma_0 - i\gamma_3), \quad \bar{\psi}_1 = \tfrac{1}{2}(\gamma_1 + i\gamma_2),$$
$$\psi_2 = \tfrac{1}{2}(\gamma_1 - i\gamma_2), \quad \bar{\psi}_2 = \tfrac{1}{2}(\gamma_0 + i\gamma_3), \tag{79}$$

with

$$\gamma_0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \quad \gamma_j = \begin{pmatrix} 0 & i\sigma_j \\ -i\sigma_j & 0 \end{pmatrix}, \quad j = 1,2,3. \tag{80}$$

These fermionic degrees of freedom satisfy the following anticommutation relations at equal time:

$$\{\bar{\psi}_1,\psi_2\} = \{\bar{\psi}_2,\psi_1\} = 1,$$
$$\{\psi_i,\psi_j\} = \{\bar{\psi}_i,\bar{\psi}_j\} = 0. \tag{81}$$

We can now define two conserved charges given by

$$Q_1 = i\bar{\psi}_1\partial - i\bar{\psi}_2(\partial V)^*,$$
$$Q_2 = i\bar{\psi}\partial + i\psi_1\,\partial V, \tag{82}$$

so that the supercharge Q is given by

$$Q = Q_1 + Q_2. \tag{83}$$

The Lagrangian for such a system can be taken to be[1]

$$L = |\dot{z}|^2 + i(\bar{\psi}_1\dot{\psi}_2 + \bar{\psi}_2\dot{\psi}_1) + \bar{\psi}_1\psi_1\,\partial^2 V + \bar{\psi}_2\psi_2(\partial^2 V)^* - |\partial V|^2, \tag{84}$$

where $V = V(z)$ is a polynomial of degree n. The action $\int L\,dt$ is invariant under the following infinitesimal transformations:

$$\delta z = \bar{\psi}_1\epsilon, \quad \delta\bar{z} = \bar{\epsilon}\psi_2, \quad \delta\psi_1 = -(\partial V)^*\epsilon,$$
$$\delta\bar{\psi}_1 = i\dot{\bar{z}}\bar{\epsilon}, \quad \delta\psi_2 + i\dot{z}\epsilon, \quad \delta\bar{\psi}_2 = (\partial V)\epsilon. \tag{85}$$

Thus we find that we can derive $N=2$ Wess–Zumino quantum mechanics from stochastic fields in a complexified space-time. The supersymmetric quantum mechanics arises through the introduction of nondiagonal elements which breaks down the reflection invariance between the external and internal fields which is necessary for equilibrium condition. Also we note that through this formalism of holomorphic quantum mechanics we can derive a supercharge Q such that the supersymmetric Hamiltonian is given by $H_S = Q^2$. This links up the inherent supersymmeric feature in the stochastic quantization procedure as we first pointed out by Parisi and Sourlas[11] with the conventional formalism of supersymmetric quantum mechanics.

IV. STOCHASTIC FIELD THEORY AND INDEX THEOREM

Jaffe *et al.*[1] have shown that the $N=2$ Wess–Zumino quantum mechanics has degenerate vacua. The space of

vacuum states is bosonic and its dimension is determined by the topological properties of the superpotential. The same result can be derived from the stochastic field theoretical formalism using the formalism of thermofield dynamics.

The index theorem for a free field theory has been derived from the stochastic field theory using the formalism of thermofield dynamics in a recent paper.[9] Here we shall generalize this result in the presence of a superpotential. In the free field case, we can define two Klein operators

$$\theta = (-1)^F, \tilde{\theta} = (-1)^{\tilde{F}}, \tag{86}$$

where $\tilde{\theta} = J\theta J$, J being an involution operator with the property $J^2 = 1$. As discussed in Sec. II, the tilde function is associated with the internal field and the total Hamiltonian of the system is given by

$$\bar{H} = H - \tilde{H},$$

where

$$\tilde{H} = JHJ. \tag{87}$$

In this formalism, the vacuum is temperature dependent and we have the relations

$$\bar{H}\,|0(\beta)\rangle = 0$$
$$\langle A \rangle = \langle 0(\beta)|A\,|0(\beta)\rangle. \tag{88}$$

Now it is noted that for the Klein operators θ and $\tilde{\theta}$ we have

$$\theta\,|0(\beta)\rangle = \tilde{\theta}\,|0(\beta)\rangle \neq |0(\beta)\rangle,$$
$$\bar{\theta}\,|0(\beta)\rangle = \theta\tilde{\theta}\,|0(\beta)\rangle = |0(\beta)\rangle. \tag{89}$$

We can define an index for the ground state given by

$$i(Q_+) = \mathrm{Tr}(-1)^{\bar{F}} e^{-\beta\bar{H}}|_{\beta\to\alpha}, \tag{90}$$

where $\bar{F} = F + \tilde{F}$, $\bar{H} = H - \tilde{H}$. This can also written as

$$i(Q_+) = \langle 0(\beta)|\theta\tilde{\theta}\,|0(\beta)\rangle_{\beta\to\alpha}. \tag{91}$$

As we have mentioned in Sec. II, the thermodynamic equilibrium is maintained as long as Z_2 symmetry (time reversal symmetry) is operative in nontilde and tilde objects corresponding to the external and internal space. That is, the orientation in the external space should be opposite to that in the internal space. However, this formalism of stochastic field theory suggests that there may exist a critical temperature T_c when the orientation of the internal space is changed leading to a nonequilibrium state. Indeed for the thermal doublet of a bosonic field $\phi(x) = \binom{\phi(x)}{\phi(\xi)} = \binom{\phi(x)}{\tilde{\phi}'(x)}$ it may so happen that the isotropic feature of the bosonic field $\phi(\xi) = \tilde{\phi}^{\dagger}(x)$ is lost at this critical temperture and an internal helicity is generated for this field giving rise to an anisotropic feature leading to the generation of a fermion. Thus beyond this temperature T_c we have the supersymmetric feature due to the breakdown of the Z_2 symmetry which will then give rise to thermal doublets of different statistics which will appear as zero energy modes as suggested by Matsumoto *et al.*[12]

Now to find out the index theorem in the supersymmetric phase we note that the equilibrium condition demands

$$\langle 0(\beta)|\theta\tilde{\theta}\,|0(\beta)\rangle = 1 = \int \delta(F + \tilde{F})dF. \tag{92}$$

In the nonsupersymmetric case, for a bosonic thermal doublet $F = \tilde{F} = 0$. However, for a supersymmetric phase, we may have $\tilde{F} = \pm 1$ depending on the orientation of the internal helicity developed leading to an anisotropy in the internal space. Thus we will have the index

$$i(Q_+) = \langle 0(\beta)|\tilde{\theta}\,|0(\beta)\rangle_{\beta\to\alpha}$$
$$= \int \delta(F \pm 1)dF = 1. \tag{93}$$

However, in an interacting case with a superpotential given by

$$V(\phi)\tfrac{1}{2}m\phi^2 + \sum_{j=3}^{n} a_j\phi^j$$

with complex ϕ, $a_n \neq 0$ $n \geqslant 3$ as we have identified $\phi^{\pm} = \mp i\sqrt{2}\,\partial V$, we note that we will have $(n-1)$ images of such thermal doublets. So for an interacting case, with superpotential $V(\phi) = \lambda\phi^n$ we will have

$$i(Q_+) = (n-1)\langle 0(\beta)|\theta\tilde{\theta}\,|0(\beta)\rangle_{\beta\to\alpha}$$
$$= (n-1)\int \delta(F \pm 1)dF = n - 1 = \deg \partial V.$$

this is identical with the result obtained by Jaffe *et al.*[1] in the two-dimensional $N = 2$ Wess–Zumino field model and we can conclude that the holomorphic quantum mechanics constructred from the stochastic field theory will also lead to a degenerate vacua in the interacing case when the superpotential is given by $V(\phi) = \lambda\phi^n$ $n > 2$.

V. DISCUSSION

We have shown here that the relativistic generalization of Nelson's stochastic mechanics as well as stochastic quantization in Minkowski space helps us to construct holomorphic quantum mechanics when in the nonequilibrium condition we can realize $N = 2$ Wess–Zumino quantum mechanics and supersymmetry. In the equilibrium condition, we get stochastic quantization in Minkowski space and we have Z_2 symmetry between the external and internal fields which form a doublet. When this reflection invariance is broken, we get supersymmetric quantum mechanics which imply that supersymmetry gets broken in a multiply connected space.

The inherent supersymmeric feature in stochastic quantization leading to Euclidean quantum field theory from a Langevin equation incorporating a fictitious time was first pointed out by Parisi and Sourlas.[11] However, it was not clear whether this supersymmeric feature which involves invariance of the action under certain supersymmetric transformations is equivalent to the conventional supersymmetric quantum mechanics which defines a supercharge Q such that the Hamiltonian is given by $H = Q^2$. Indeed in that case the action involves fermionic variables only through the determinant which arises in the averaging procedure and hence the invariance of the action under supersymmetric transformation in this case does not imply the existence of the supercharge Q as well as the existence of the grading operator γ such that $H = Q^2$ and $\gamma Q + Q\gamma = 0$. However, here we have pointed out that stochastic quantization in Minkowski space introduces two stochastic fields, one in the external space and the other in the internal space; in the nonequilibrium case we can construct holomorphic quantum mechanics out

of these two fields, which becomes equivalent to $N=2$ Wess–Zumino quantum mechanics and gives rise to the supercharge operator.

Finally, it may be pointed out that this formalism helps us to study finite temperature field theory as well as finite temperature supersymmetry through the methodology of thermofield dynamics when we identify the tilde field with the stochastic field in the internal space. Indeed in a recent paper,[9] we have pointed out that there exists a critical temperature T_c below in which supersymmetry is broken and the zero energy mode is given by a thermal doublet of opposite statistics and in the case of a free field theory the index $i(Q_+) = n_+ - n_-$, where $n_\pm = \dim \text{kernel}\, Q_\pm$ is found to be 1. It may take the value $\frac{1}{2}$ also when the zero energy mode is given by a thermal doublet of the same fermionic statistics.[9] Here we have pointed out that in the interacting case when the potential is given by $V(\phi) = \lambda\phi^n, \phi$ being a complex scalar field, and $n > 2$, the vacuum is degenerate and the index takes the value $n - 1 = \deg \partial V$. This is identical with the result obtained by Jaffe *et al.* in the two-dimensional Wess–Zumino quantum field model.

[1] A. Jaffe, A. Lesniewski, and M. Lewenstein, Ann. Phys. **178**, 313 (1987).
[2] Y. Takahashi and H. Umezawa, Collect. Phenom. **2**, 151 (1974).
[3] P. Bandyopadhyay and K. Hajra, "Stochastic quantization in Minkowski space," to be published.
[4] P. Bandyopadhyay and K. Hajra, J. Math. Phys. **28**, 711 (1987).
[5] E. Nelson, Phys. Rev. **150** 1079 (1966).
[6] S. M. Moore, J. Math. Phys. **21**, 2102 (1980).
[7] F. Guerra and P. Ruggiero, Phys. Rev. Lett. **31**, 1022 (1973).
[8] A. J. Niemi and G. W. Semenoff, Ann. Phys. **152**, 105 (1983).
[9] P. Ghosh and P. Bandyopadhyay, Ann. Phys. **179**, 1 (1987).
[10] K. Hajra, Mod. Phys. Lett. A **4**, 1469 (1989).
[11] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[12] H. Matsumoto, N. Nakahara, Y. Nakano, and H. Umezawa, Phys. Lett. B **104**, 53 (1984).

New class of nonlinear realizations of extended graded anti-de-Sitter supersymmetry OSP (N,4), $N \geqslant 2$

M. Mozrzymas
Institute of Theoretical Physics, University of Wroclaw, ul. Cybulskiego, 36, 50-205 Wroclaw, Poland

(Received 20 January 1988; accepted for publication 26 April 1989)

The nonlinear realizations of N-extended graded anti-de-Sitter supersymmetry with three different ways of splitting into linearly and nonlinearly realized sectors of OSP (N,4) superalgebra are considered. Nonlinear transformation laws are derived for Goldstone fermions and Goldstone scalars that describe a spontaneously broken sector in extended anti-de-Sitter supersymmetry. Cartan forms on the supercosets are derived and corresponding nonlinear Lagrangians for Goldstone fermions and scalars in anti-de-Sitter space are constructed. The contraction of OSP ($2N$,4) supersymmetry to Poincaré supersymmetry with central changes is discussed.

I. INTRODUCTION

A geometric description of spontaneously broken symmetries and supersymmetries may be achieved within the framework of nonlinear realizations. In the late sixties it was realized that the interactions of Goldstone fields can be geometrized by introducing the framework of σ models and nonlinear realizations of internal symmetries.[1,2] On the other hand, it is interesting to recall that the first field-theoretic example of supersymmetry[3,4] was described by nonlinear realizations on the supercoset parametrized by spin-$\frac{1}{2}$ fields called Goldstone fermions.

In Ref. 4 Volkov and Akulov considered the following supercoset:

$$K_N = \mathrm{SP}_N/\mathrm{O}(3,1)\times\mathrm{U}(N), \tag{1.1}$$

where SP_N is the N-extended super-Poincaré group and $\mathrm{U}(N)$ is the internal symmetry group. The supercoset K_N is parametrized by the four space-time translations and $4N$ anticommuting fermionic degrees of freedom describing supertranslations. These $4N$ fermionic degrees of freedom describe the fermionic Goldstone fields. On the supercoset (1.1), which is a homogeneous space, the supergroup SP_N is realized nonlinearly, only with its subgroup $\mathrm{O}(3,1)\times\mathrm{U}(N)$ realized in a linear way. Using standard procedures in nonlinear realizations theory[1,3,5] Volkov and Akulov introduced Cartan forms on K_N and by taking their fourfold outer product they obtained the invariants under the transformations of the supergroup SP_N. These invariants can be interpreted as the Lagrangians of the interacting Goldstone fields. Later the nonlinear realizations of N-extended Poincaré supergroup SP'_N with $\frac{1}{2}N(N-1)$ central charges were also considered.[6,7] In Ref. 6 Ferrara did show that their coupling to N-extended supergravity provides the simultaneous occurance of Higgs and super-Higgs effects. In this case with a maximal number of central charges the whole internal symmetry subgroup $\mathrm{U}(N)$ is spontaneously broken, and the supercoset under consideration has the following form:

$$K'_N = \mathrm{SP}'_N/\mathrm{O}(3,1). \tag{1.2}$$

In all the cases discussed above the Goldstone fields are conventional $D=4$ Minkowski fermionic fields.

In order to have a better analogy with nonlinear realization of internal symmetry group, it is appropriate to consider anti-de-Sitter extended supersymmetry described by semisimple supergroup OSP(N,4). Zumino,[8] by applying the standard theory of nonlinear realizations to supergroup OSP(1,4), derived the nonlinear transformation laws for the four-component Goldstone–Majorana spinor implemented by spontaneous breaking of $N=1$ global supersymmetry in anti-de-Sitter space. In the anti-de-Sitter case the action for the Goldstone field contains the mass term proportional to the inverse of the anti-de-Sitter radius R. There exists also the formulation of anti-de-Sitter supersymmetry with superfield realization on the superspace described by the supercoset OSP(1,4)/SO(3,1).[9,10] Keck[11] has constructed the corresponding superspace and has studied superfields, which were reduced to its irreducible parts. Further, Ivanov and Sorin[10] did show explicitly how to construct OSP(1,4) invariants from superfields and used them for the construction of the nonlinear Lagrangian densities. Recently, Azcaraga and Lukierski[12] proposed a superfield extension of the Volkov–Akulov method of constructing nonlinear realizations of extended Poincaré supergroup. Their formulation describes partial supersymmetry breaking ($N=1$ supersymmetry remaining unbroken), but it was shown that the model contain ghosts.

It has been shown that there is a connection between the superfield approach to the supersymmetry and nonlinear realization of the Volkov–Akulov type. Ivanov and Kapustnikov in Ref. 13 and Sammuel and Wess in Ref. 14 gave the formulas that describe the transition from nonlinear realizations to the linear ones, with Goldstone fermions described by constrained linear Goldstone superfield.

In this paper we provide a group theoretical framework for the model invariant under extended anti-de-Sitter supersymmetry with spontaneously broken generators by considering nonlinear realizations of OSP(N,4) supergroup on the following supercosets:

(a) $\mathrm{OSP}(N,4)/\mathrm{SO}(3,1)\times\mathrm{O}(N)$;
(b) $\mathrm{OSP}(N,4)/\mathrm{SO}(3,1)$;
(c) $\mathrm{OSP}(N,4)/\mathrm{SO}(3,1)\times\mathrm{U}(N)$.

We are reminded of the nonlinear transformation laws and deduce the nonlinear Lagrangians constructed from

 0022-2488/90/010221-15$03.00

Cartan forms. After the discussion of basic formulas of nonlinear realizations (Sec. II) we investigate the supercoset a (Sec. III). The formalism describes the generalization of the Zumino model[8] to arbitrary N. Using Cartan forms we obtain a nonlinear Lagrangian of N interacting Goldstone fermions. In Sec. IV the nonlinear realization on a product of symmetric supercosets is discussed. Such a formalism is applied to the second supercoset b, with the spontaneously broken sector described by N Goldstone fermions and $\frac{1}{2}N(N-1)$ scalar Goldstone fields. We propose the nonlinear Lagrangian which describes the interactions of all these Goldstone fields in anti-de-Sitter space. The last supercoset c describes the nonlinear realization of the OSP$(2N,4)$ supergroup. By contraction to the OSP$(2N,4)$ superalgebra, proposed by Lukierski and Rytel,[15] one can obtain the nonlinear realization of a N-extended Poincaré supergroup with $\frac{1}{2}N(N-1)$ central charges, given by Ferrara in Ref. 6. In Sec. V we construct the nonlinear realization on supercoset c and perform the contraction to SP'_N.

II. BASIC FORMULAS DESCRIBING NONLINEAR REALIZATIONS ON COSET SPACES

A. Global nonlinear realization on symmetric spaces

Following Ref. 16 we consider the Lie algebra (superalgebra) $\hat{G}$ in the form

$$\hat{G}=\hat{H}\oplus\hat{S}$$

and

$$[\hat{H}_i,\hat{H}_j]=a_{ij}^k\hat{H}_k;\quad [\hat{H}_i,\hat{S}_j]=b_{ij}^k\hat{S}_k;$$
$$[\hat{S}_i,\hat{S}_j]=c_{ij}^k\hat{H}_k, \tag{2.1}$$

where $\hat{H}_j$ are the generators of subalgebra $\hat{H}$ called symmetric subalgebra. We denote the group elements by

$$s=\exp\xi\cdot\hat{S}=\exp\xi^j\hat{S}_j,$$
$$h=\exp\lambda\cdot\hat{H}=\exp\lambda^j\hat{H}_j, \tag{2.2}$$

where ξ^i and λ^i are the Lie group parameters. By S we denote coordinates, parametrizing the coset, and subalgebra $\hat{H}$ generates the subgroup H. If H is a connected subgroup and if the relations (2.1) hold, then the coset G/H is an example of a symmetric space.[17]

Every element g of a Lie group G can be written in the form

$$g=s\cdot h,$$

where s belongs to the coset and h belongs to H.

We now consider the transformation of s under the action of the group element g. We evaluate first the case in which $g_0=h_0\in H_0$. Then $g_0=\exp\lambda^i\hat{H}_i$ and we have

$$S'=g_0sA(g_0^{-1})=h_0sA(h_0^{-1})=h_0sh_0^{-1}$$
$$=\exp(\lambda^i\hat{H}_i)\exp(\xi^i\hat{S}_i)\exp(\lambda^i\hat{H}_i)=\exp\xi'^i\hat{S}_i.$$

In this case the transformation is linear and has the form

$$\xi'^i=(\exp B)^i_j\xi^j, \tag{2.3}$$

where $B^i_j=\lambda^k b^i_{kj}$ and b^i_{kj} are given by the relations (2.1). Consider now the case when $g_0\in S$, so we have $g_0=\exp\xi_0^i\hat{S}_i$. Then, because $A(s)=s^{-1}$, we have

$$s'=g_0sA(g_0^{-1})=\exp\xi_0^i\hat{S}_i\exp\xi^i\hat{S}_i\exp\xi_0^i\hat{S}_i.$$

Using the Baker–Haussdorff formula we find

$$\exp 2\xi'^j\hat{S}_j=\exp\Big[2\xi_1^i\hat{S}_i+2\xi\cdot\hat{S}-\tfrac{2}{3}\Big[\Big[\xi\cdot S,\xi_0\cdot S\Big],\xi\cdot S\Big]$$
$$+\text{higher commutators}$$
$$+O(\xi_0S)^2+\cdots\Big]$$
$$=\exp 2\xi'^j\hat{S}_j.$$

So finally using the relations (2.1) we obtain

$$2\delta\xi^i=2\xi_0^i-\tfrac{2}{3}c_{ij}^kb_{km}^l\xi^m\xi^i\xi_0^j+\text{higher-order terms.} \tag{2.4}$$

We see that the transformation is nonlinear.

B. Nonlinear realization on a homogenous space

In many cases of physical interests the algebra $\hat{G}$ has not the form given by (2.1) and hence the formula (2.4) are not valid. However in infinitesimal case under some assumptions it is possible to derive the formulas of nonlinear group action on a coset space.

We assume now that $\hat{G}$ has the form

$$[\hat{H}_i,\hat{H}_j]=a_{ij}^k\hat{H}_k;\ [\hat{H}_i,\hat{S}_j]=b_{ij}^i\hat{S}_i;$$
$$[\hat{S}_i,\hat{S}_j]=c_{ij}^m\hat{H}_m+d_{ij}^k\hat{S}_k, \tag{2.5}$$

As in the previous case every element g of the group G can be represented uniquely in the following way[7,16]:

$$g=s\cdot h=[\exp\xi\cdot\hat{S}\,]h;\quad h\in H.$$

The action of the group element g_0 on the coset is the following:

$$g_0e^{\xi\cdot\hat{S}}=e^{\xi\cdot\hat{S}}h_1. \tag{2.6}$$

Now h_1 is a function of g_0 and ξ. As before, if $g_0=h_0\in H$ we have

$$h_0e^{\xi\cdot\hat{S}}=h_0e^{\xi\cdot\hat{S}}h_0^{-1}h_0=e^{\xi\cdot\hat{S}}h_1.$$

The condition $[\hat{H},\hat{S}]\subset\hat{S}$ implies

$$e^{\xi\cdot\hat{S}}=h_0e^{\xi\cdot\hat{S}}h_0^{-1}$$

and the transformation law of ξ is linear.

Let us consider the case in which $g_0\in s$, i.e., $g_0=e^{\xi\cdot\hat{S}}$. Equation (2.6), in general, is not solvable, but if g_0 is infinitesimal then h_1 is also infinitesimal, and using the notation introduced in Appendix A, Eq. (2.6) can be written in the form

$$e^{-\xi\cdot\hat{S}}\wedge\xi_0\hat{S}-[(1-e^{\xi\cdot\hat{S}})/(\xi\cdot\hat{S})\,]\wedge\delta\xi\cdot\hat{S}=h_1-1,$$
$$g=1+\xi\cdot\hat{S};\quad \xi_0^i\ \text{is infinitesimal.} \tag{2.7}$$

On the right-hand side of the equation we have an element of the subalgebra $\hat{H}$. This implies that the coefficients multiplying the generators $\hat{S}_i$ on the left-hand side of Eq. (2.7) must be equal to zero. In such a way we are able to find $\delta\xi^i$ and $h-1$. The condition that the coefficients of generators $\hat{S}_i$ must be equal to zero gives the following relation:

$$\delta\xi^l=\xi_0^l-\tfrac{1}{3}[c_{ij}^lb_{km}^l\xi_0^j\xi^i\xi^m+\xi^i\xi_0^j\xi^md_{ij}^kd_{km}^l]$$
$$+\text{higher-order terms.}$$

We see that if we put $d_{im}^l=0$ we obtain in the case of symmetric coset space the formula (2.4).

C. Cartan forms on the coset space

Cartan form ω is the Lie algebra valued form defined on the coset[8]

$$\omega = e^{-\xi\cdot\hat{S}} d e^{\xi\cdot\hat{S}} = \mu\cdot\hat{S} + \nu\cdot\hat{H},$$

where $\mu^i(\xi^i, d\xi^i)$ and $\nu^j(\xi^i, d\xi^i)$ are the ordinary differential one-forms. The transformation law of ω is easy to derive if $g_0 = h_0 = H$. We get

$$\omega' = h_0 e^{-\xi\cdot\hat{S}} h_0^{-1} h_0 d e^{\xi\cdot\hat{S}} h_0^{-1} = h_0 \omega h_0^{-1}. \tag{2.8}$$

On the other hand, if $g_0 = e^{\xi_i \hat{S}}$ one can check that

$$\omega' = h_1 \omega h_1^{-1} + h_1\, dh_1^{-1} \tag{2.9}$$

or more explicitly

$$\mu\cdot\hat{S} = h_1 \mu\cdot\hat{S} h_1^{-1},$$
$$\nu\cdot\hat{H} = h_1 \nu\cdot\hat{H} h_1^{-1} + h_1\, dh_1^{-1}. \tag{2.10}$$

So we see that ω, μ^i, and ν^i transform in both cases by the element of subgroup H, but in the second case h_1 depends nonlinearly on the coset coordinates ξ^i and on the parameters of g. Because G acts on the form ω by the transformation belonging to H, therefore every H-invariant constructed with μ^i will be also invariant under entire group G. In these invariants the coordinates ξ^i will transform linearly under action of subgroup H and nonlinearly under remaining group elements. If H is the Lorentz group such invariants can be used as nonlinear physical Lagrangina densities. This will be discussed in next paragraphs.

III. NONLINEAR REALIZATION OF SUPERGROUP OSP (N,4) ON THE COSET OSP (N,4)/SO(3,1)×O(N)

A. Algebraic considerations

The superlattice OSP(N,4) has the form[15]

$$\{Q^i_\alpha, Q^j_\beta\} = \delta^{ij}(\gamma^a\gamma^0)_{\alpha\beta} P_a + m[\delta^{ij}(\sigma^{ab}\gamma^0\gamma^5)_{\alpha\beta} M_{ab} + (\gamma^0\gamma^5)_{\alpha\beta} T^{ij}];$$
$$[P_a, Q^i_\alpha] = \tfrac{1}{2} im(\gamma_a\gamma_5) Q^i_\alpha; \quad [M_{ab}, Q^i_\alpha] = i(\delta_{ab} Q^i)_\alpha;$$
$$[T_r, Q^i_\alpha] = i J^{ik}_T (Q^i_k)_\alpha;$$
$$-i[M_{ab}, M_{cd}] = \eta_{ac} M_{bd} + \eta_{bd} M_{ac} + \eta_{ad} M_{cb} + \eta_{bc} M_{dQ}; \tag{3.1}$$
$$[P_a, P_b] = -im^2 M_{ab};$$
$$-i[M_{ab}, P_c] = \eta_{ac} P_b - \eta_{bc} P_a;$$
$$[T_{ij}, T_{kl}] = \delta_{ik} T_{jl} + \delta_{jl} T_{ik} - \delta_{il} T_{jk} - \delta_{jl} T_{il};$$
$$[T_{ij}, P_a] = [T_{ij}, M_{ab}] = 0; \quad m = 1/R,$$

where $i,j = 1\cdots N$, and $\beta,\alpha = 1\cdots 4$ (we use the four-component Majorana formalism), so we have $4N$ real supercharges Q^i_α. Besides $\sigma_{ab} = \frac{1}{4}[\gamma_a, \gamma_b]$, where $a,b = 0,1,2,3$ and γ^a are the real Dirac matrices satisfying

$$\{\gamma_a, \gamma_b\} = 2\eta_{ab}; \quad \eta_{ab} = \mathrm{diag}(-1,1,1,1)$$

and $T^{ij} = -T^{ji} = J^{ij}_r T^r$ describe $O(N)$ generators expanded in the $N\times N$ matrix basis denoted by $J^{ij}_r = -J^{ji}_r$ $[r = 1\cdots N(N-1)/2]$. The Lie algebra $O(N)$ describe the internal symmetry group. Real supercharges extend the Lie algebra $SO(3,2)\times O(N)$ to superalgebra $OSP(N,4)$. The extension is possible because the fundamental spinor representation of SO(3,2) Lie algebra is a real four-component Majorana spinor. One can introduce five 4×4 real matrices Γ_A, which describe real spinor representation of the SO(3,2) Clifford algebra as follows:

$$\{\Gamma_A, \Gamma_B\} = 2\eta_{AB},$$
$$\eta_{AB} = \mathrm{diag}(-1,1,1,1,-1); \quad A,B = 0,\ldots,4;$$

and choose $\Gamma_A = (\gamma_a, \gamma_5)$.

Zumino has derived the nonlinear realization for OSP(1,4) supergroup.[8] It is easy to see that SO(3,2) Lie algebra is a symmetric subalgebra of OSP(1,4). On the other hand, SO(3,1) Lie algebra is a symmetric subalgebra of SO(3,2) Lie algebra [but is not symmetric subalgebra of OSP (1,4) !]. Thus according to the general scheme 2a, OSP(1,4)/SO(3,2) and SO(3,2)/SO(3,1) are symmetric spaces. This implies the following parametrization of the coset[8]:

$$\exp(\bar{\theta} Q)\exp(-ix^a P_a), \tag{3.2}$$

where θ is the Grassmann parameter describing fermionic coset OSP(1,4)/SO(3,2) and x^a are the parameters describing the coset SO(3,2)/SO(3,1) (Ref. 8). We shall proceed similarly for general OSP(N,4) case. Now the reducible Lie algebra $SO(3,2)\times O(N)$ form the symmetric subalgebra of OSP(N,4). Generalizing the previous particular OSP(1,4) case we can write

$$\frac{\mathrm{OSP}(N,4)}{\mathrm{SO}(3,1)\times\mathrm{O}(N)} = \frac{\mathrm{OSP}(N,4)}{\mathrm{SO}(3,2)\times\mathrm{O}(N)} \times \frac{\mathrm{SO}(3,2)}{\mathrm{SO}(3,1)}. \tag{3.3}$$

So the supercoset (3.3) is the product of two symmetric spaces and has the following parametrization:

$$\exp(\bar{\theta}^i Q_i)\exp(-ix^a P_a). \tag{3.4}$$

Now we can apply the general scheme from Sec. II B.

B. Nonlinear realization for arbitrary N

Following Sec. II B we shall study only infinitesimal transformations. Because the considered coset is the product of two cosets, the supergroup OSP(N,4) acts in the following way:

$$g_0 e^{\bar{\theta}^i Q_i} = e^{\bar{\theta}'^i Q} h_1(\theta^i, g_0); \tag{3.5}$$
$$h_1 e^{-ix^a P_a} = e^{-ix'^a P_a} l_1(g_0, \theta^i, x^a); \tag{3.6}$$
$$g_0 \in \mathrm{OSP}(N,4); \quad h_1 \in \mathrm{SO}(3,2)\times\mathrm{O}(N);$$
$$l_1 \in \mathrm{SO}(3,1)\times\mathrm{O}(N) = L.$$

First let us study the case when the group element g_0 belongs to subgroup L. So we have

$$g_0 = l_0 = 1 + \tfrac{1}{2} i\nu^{ab} M_{ab} + \tfrac{1}{2}\lambda^{ij} T_{ij},$$

where ν^{ab} and λ^{ij} are infinitesimal parameters. The formula (2.7) gives

$$e^{\bar{\theta}'^i Q_i} = l_0 e^{\bar{\theta}^i Q_i} l_0^{-1}$$
$$e^{-ix'^a P_a} = l_0 e^{-ix^a P_a} l_0^{-1} \qquad h_1 = l_1 = l_0. \tag{3.7}$$

The transformation law of θ^i and x^a is linear

$$\theta'^i_\alpha = \theta^i_\alpha - \lambda^{kl}(J_r)_{kl}(T^r)^{im}(\theta_m)_\alpha - \tfrac{1}{2}\nu^{ab}(\sigma_{ab}\theta^i)_\alpha$$
$$x'^d = \exp[\tfrac{1}{2} i\nu^{ab} M_{ab}]^d_c x^c.$$

Next case is given by $g_0 = h_0 = \exp(-is^a P_a)$, s^a is infinitesimal. The $SO(3,2)\times O(N)$ Lie algebra is symmetric subalgebra of $OSP(N,4)$ superalgebra, so we have

$$e^{\bar\theta_i Q^i} = h_0 e^{\bar\theta_i Q^i} h_0^{-1} \tag{3.8}$$

and this implies

$$\theta'^1_\alpha = \theta^1_\alpha + \tfrac{1}{2} m s_a (\gamma^a\gamma^5\theta^i)_\alpha .$$

The next step gives

$$h_0 e^{-ix^a P_a} = e^{-ix^a P_a} l_1(s^a, x^a). \tag{3.9}$$

Here we should deal with the nonlinear realization of $SO(3,2)$ group on the coset $SO(3,2)/SO(3,1)$. This example of nonlinear realization is treated in Ref. 8. The transformation law of x^a under the action of group element $h_0 = \exp - is^a P_a$ is nonlinear and inhomogenous

$$x^a = s^a + \left(\sqrt{m^2x^2}\coth\sqrt{m^2x^2} - 1\right)(s^a - x^b s_b x^a/x^2) . \tag{3.10}$$

Now we shall consider the case $g_0 = \exp \bar\alpha_i Q^i$. We should use Eqs. (3.5) and (3.6). Equation (3.6) has the same form as (3.9), but now h_0 depends on Grassmann parameters θ^i and α^i. Therefore x^a has the same form given by (3.10), but now s^a is the function of θ^i_α and α^i_β. If α^i are infinitesimal then using (3.5) one can write

$$\begin{aligned} &e^{-\bar\theta_i Q^i}\alpha_1 Q^i e^{\bar\theta^i Q_i} - e^{\bar\theta_i Q^i}\delta e^{\bar\theta_i Q^i} \\ &\quad = -is(\bar\theta^i,\bar\alpha^i)P_a + \tfrac{1}{2} i\nu^{ab}(\theta^i,\alpha^i)M_{ab} \\ &\qquad + \lambda^{ij}(\theta^i,\alpha^i)T_{ij} = h_1(\alpha^i,\theta^i). \end{aligned} \tag{3.11}$$

As it was mentioned in Sec. II B we can obtain from this equation the transformation law of θ^i and the parameters $s^a(\theta^i,\alpha^i)$, $\nu^{ab}(\theta^i,\alpha^i)$, and $\lambda^{il}(\theta^i,\alpha^i)$.

The result is the following (see Appendix B):

$$\delta\theta^i = \alpha^i + \sum_{k=1}^{k=4N+1} \frac{(-1)^{k+1}2^{2k}}{(2k)!} B_{2k-1} \overline{\square}^i_{\square 2k}, \tag{3.12}$$

where B_{2k-1} are the Bernoulli numbers and

$$\begin{aligned} \overline{\square}^i_{\square 2k} = (im)^k [&-3\overline{\square}^i_{\square \underline{2k-2}}\bar\theta^i\theta_i + 2\bar\theta^i\theta_m \overline{\square}^m_{\square \underline{2k-2}} \\ &+ \overline{\square}^i_{\square \underline{2k-2}}\bar\theta^i\theta_l], \end{aligned} \tag{3.13}$$

where $\overline{\square}^i_{\square \underline{0}} = \alpha^i$. This the recurrent formula. We have also

$$s^a = \bar\theta_i\gamma^a\gamma^5\eta^i_1; \quad V^{ab} = m\bar\theta_i\sigma^{ab}\eta_1; \quad \lambda^{ij} = m\bar\theta^i\eta^j_1; \tag{3.14}$$

and

$$\begin{aligned} \eta^i_1 = \sum_{k=0}^{k=4N+1} &\left[\frac{2(2^{2k-1}-1)}{(2k)!} B_{2k}\right. \\ &\left. - \frac{(-1)^{k+1}2^{2k}}{(2k)!} B_{2k-1}\right] \overline{\square}^i_{\square \underline{2k}}. \end{aligned} \tag{3.15}$$

In fact the sums in (3.12) and (3.15) are not infinite. The term $\overline{\square}^i_{\square 2k}$ when $2k > 4N+1$ will vanish identically due to the anticommutating properties of the spinor components (see also Appendix B). For example, in the (OSP 1,4) case we have only $\overline{\square}^i_{\square \underline{2}}$ and $\overline{\square}^i_{\square \underline{4}}$. If we shall makes now the use of (3.14) and (3.10) we get the transformation law of x^a under action of $g_0 = \exp \bar\alpha_i Q^i$.

C. Cartan form on the supercoset OSP(N,4)/SO(3,1) ×O(M)

According to the general scheme in Sec. III C, the Cartan form is given by the formula

$$\begin{aligned} \omega &= e^{ix^a P_a} e^{-\bar\theta_i\omega^i} d\left(e^{\bar\theta_i Q^i} e^{-ix^a P_a}\right) \\ &= \omega_i Q^i - i\mu^a P_a + \tfrac{1}{2} i\nu^{ab} M_{ab} + \rho^{ij} T_{ij}, \end{aligned}$$

where $\omega_i, \mu^a, \nu^{ab}$, and δ^{ij} are the forms that depends on x^a, θ^i, dx^a, $d\theta^i$. Using the method given in Appendix B we can find the components of Cartan form ω. The result is the following:

$$\begin{aligned} \overline{\omega}^i &= \operatorname{ch}\tfrac{1}{2}\sqrt{m^2x^2}\,\bar\eta_2 - (1/\sqrt{x^2})\operatorname{sh}\tfrac{1}{2}\sqrt{m^2x^2}\,x^a\eta_2\gamma_a\gamma_5, \\ \mu^a &= s^a + (\operatorname{ch}\sqrt{m^2x^2} - 1)\left(s^a + \frac{x^b s_b x^a}{x^2}\right) \\ &\quad + \frac{2}{\sqrt{m^2x^2}} \operatorname{sh}\sqrt{m^2x^2}\, t^{ab}x_b \\ &\quad + \left[dx^a + \left(\frac{\operatorname{sh}\sqrt{m^2x^2}}{\sqrt{m^2x^2}} - 1\right)\left(dx^a + \frac{x^b dx_b x^b}{x^2}\right)\right], \\ \rho^{ij} &= m\bar\theta^i\eta^j_3, \\ V^{ab} &= \sqrt{\frac{m^2x^2}{x^2}} \operatorname{sh}\sqrt{m^2x^2}\,x^a s^b \\ &\quad - 2\frac{m^2}{x^2}\operatorname{ch}\sqrt{m^2x^2}\, t^{ca}x_c x^b, \end{aligned}$$

where

$$\begin{aligned} &s^a = i\bar\theta_i\gamma^a\gamma^5\eta^i_3; \quad t^{ab} = im\bar\theta_i\sigma^{ab}\eta^i_3; \\ &\eta^i_2 = \sum_{k=0}^{k=4N+1} \frac{1}{(2k+1)!} \overline{\square}^i_{\square \underline{2k}}; \\ &\eta^i_3 = -\sum_{k=0}^{k=4N+1} \frac{1}{(2k+2)!} \overline{\square}^i_{\square \underline{2k}}; \end{aligned}$$

$\overline{\square}^i_{\square \underline{2k}}$ is defined by (3.13) and now $\overline{\square}^i_{\square \underline{0}} = d\theta^i$.

The generators P_a belong to the coset $OSP(N,4)/SO(3,1)\times O(N)$ and therefore the transformation law for the one-form μ^a has the form

$$\mu'^a P_a = \mu^a P_a,$$

where $\epsilon SO(3,1)\times O(N)$ and it is defined by the relation

$$h_1 e^{-ix^a P_a} = e^{-ix'^a P_a}(\theta^i, x^a, \alpha^i)$$

and h_1 is determined by relations (3.14).

The forms μ^a transform only under $SO(3,1)$ part of L, because of $[T_{ij}, P_a] = 0$. We deal here with the local $SO(3,1)$ transformations, because $L_1 = L_1(\theta^i, x^a, \alpha^i)$ depends on the parameters x^a that we interpret as the space-time coordinates in anti-de-Sitter space.[8] We shall interpret θ^i as the spinor fields $\theta^i(x^a)$, with the following transformation law:

$$\Delta\theta^i = \delta\theta^i + \delta x^a \partial_a \theta^i.$$

Following Zumino[8] we define

$$\mu^0\mu^1\mu^2\mu^3 = -a^2\mathscr{L}\, dx^0\, dx^1 dx^2\, dx^3.$$

It is easy to check that

$$\mu'^0\mu'^1\mu'^2\mu'^3 = \det(L_1)\mu^0\mu^1\mu^2\mu^3,$$

where $\mu'^a = (l_1)^a{}_b \mu^b$ and $l_1 \in SO(3,1)$. We see that the product $\mu^0\mu^1\mu^2\mu^3$ is invariant under the SO(3,1) group and therefore is invariant under the entire OSP(N,4) supergroup.

The first terms of μ^a are the following:

$$\mu^a = dx^2 + \frac{1}{2} i\bar{\theta}_i\gamma^a\gamma^5 d\theta^i + im\bar{\theta}_i\sigma^{ab} d\theta^i x_b - \frac{1}{4!} m\bar{\theta}_i\gamma^a\gamma^5 d\theta^i\bar{\theta}^b\theta_b - \frac{2}{4!} m\bar{\theta}_i\gamma^a\gamma^5 d\theta^i\bar{\theta}^i\theta_i + \frac{3}{4!} m\bar{\theta}_i\gamma^a\gamma^5\theta^k d\bar{\theta}_k\theta^i + O(m^2).$$

This gives the following action density:

$$\mathcal{L} = -\frac{1}{a^2}\det\Big(\delta^a_b + \frac{1}{2} i\bar{\theta}_i\gamma^a\gamma^5\partial_b\theta^i + im\bar{\theta}_i\sigma^{ac}\delta_a\theta^i x_c - \frac{m}{4!}\bar{\theta}_i\gamma^a\gamma^5\partial_6\theta^i\bar{\theta}^i\theta_l - \frac{2}{4!} m\bar{\theta}_i\gamma^a\gamma^5\partial_6\theta^i\bar{\theta}^i\theta_l + \frac{3m}{4!}\bar{\theta}_i\gamma^a\gamma^5\theta^k\partial_6\bar{\theta}_k\theta^i + O(m^2)\Big).$$

The first few terms are the following:

$$\mathcal{L} = -\frac{1}{a^2} - \frac{1}{2} i\bar{\psi}_i\gamma^a\gamma^5\partial_a\psi^i - im\bar{\psi}^i\sigma^{ab}\partial_b\psi^i x_a + \frac{m}{4}\bar{\psi}_i\gamma^a\gamma^5\partial_b\psi^i\bar{\psi}^i\psi_l - \frac{2}{4!} m\bar{\psi}_i\gamma^a\gamma^5\partial_a\psi^i\psi^i\psi_l + \frac{3m}{4!}\psi_i\gamma^a\gamma^5\psi^k\partial_a\bar{\psi}_k\psi^i - O(m^2),$$

where ψ^i are correctly normalized Goldstone fermion fields $\psi^i = (1/a)\theta^i$, $[a] = 1/x^2$. Using the Lee–Gursey transformation[18,19]

$$\lambda^i = (1 + \tfrac{1}{2} m\gamma^a x^a)\psi^i + \cdots$$

and setting $\gamma'^a = \gamma^a\gamma^5$ we obtain finally

$$\mathcal{L} = -\frac{1}{a^2} - \frac{1}{2} i\bar{\lambda}_i\gamma'\partial\lambda^i - im\bar{\lambda}_i\lambda^i + \frac{m}{4!}\bar{\lambda}_i\gamma'\partial\lambda^i\bar{\lambda}^i\lambda_l + \frac{2m}{4!}\bar{\lambda}_i\gamma'\partial\lambda^i\lambda^i\lambda_l - \frac{3m}{k!}\bar{\lambda}_i\lambda^k\gamma'\partial\lambda_\epsilon\lambda^i - \frac{1}{4} i\bar{\lambda}_i\gamma'\partial\lambda^i m\gamma^a x_a + O(m^2).$$

The Lagrangian density describes interacting Goldstone fermions in anti-de-Sitter space. The mass term is proportional to $m = 1/R$, where R is the anti-de-Sitter radius. The Lagrangian $L = -1/a^2 \int \mathcal{L}\, dx$ is nonlinear and has global OSP(N,4) supersymmetry. In a particular OSP(1,4) case all formulas describing the transformation laws and the components of Cartan form have the simplest form given by Zumino.[8] Our Lagrangian density differs in principle from the Zumino one only by the summation over index j, and extends the results of Refs. 3 and 4 for N-extended Poincaré supergroup to the case of nonvanishing anti-de-Sitter radius R.

IV. NONLINEAR REALIZATION ON PRODUCT OF SYMMETRIC SUPERCOSETS APPLIED TO THE SUPERCOSET OSP(N,4)/SO(3,1)

A. Algebraic considerations

The superalgebra OSP(N,4) contains the generators Q^i, T^{ij}, P_a, M_{ab}, where $i, = 1...N$, $a, b = 0...3$. Using the commutation relation (3.1) one can check that the generators Q^i and T^{ij}, where $i,j = 1...N-1$, belong to the symmetric subalgebra OSP($N-1$,4) and Q^N, T^{Ni}, $i = 1...N-1$, are generators of the supercoset OSP(N,4)/OSP($N-1$,4). Using (3.1) we have

$$\{Q^N_\alpha, Q^i_\beta\} = m(\gamma_0\gamma_5)_{\alpha\beta} T^{Ni}.$$

Expanding T^{ij} in matrix basis J^{ij}_r and defining

$$T^{Nk} = J^{Nk}_r T^r = T^k; \quad J^{Nr}_r = 1 = -J^{rN}_r;$$

we obtain

$$[T_k, Q^N] = iQ_k; \quad [T_k, Q^i] = -\delta^i_k Q^N;$$
$$[Q^N, M_{ab}] = i\sigma_{ab} Q^N; \quad [P_a, Q^N] = i\tfrac{1}{2} m\gamma_a\gamma_5 Q^N;$$
$$\{Q^N_\alpha, Q^N_\beta\} = (\gamma^a\gamma^0)_{\alpha\beta} P_Q + m(\sigma^{ab}\gamma^0\gamma^5)_{\alpha\beta} M_{ab}.$$

Commutation relations for Q^i, T^{ij} are the same as in (3.1) but now $i,j = 1...N-1$. We see therefore that using notation from Sec. II A,

$$\hat{H} = (Q^i, T^{ij}, M_{ab}, P_a), \quad \hat{S} = (Q^N, T^{Ni}),$$

we get the relation (2.1), with the supercoset OSP(N,4)/OSPN-4/OSP($N-1$,4) as a symmetric space. The supercoset OSP(N,4)/SO(3,1) can be represented as the following product of symmetric supercosets

$$K = \frac{\mathrm{OSP}(N,4)}{\mathrm{SO}(3,1)} = \frac{\mathrm{OSP}(N,4)}{\mathrm{OSP}(N-1,4)} \times \frac{\mathrm{OSP}(N-1,4)}{\mathrm{OSP}(N-2,4)} \times \cdots \times \frac{\mathrm{OSP}(2,4)}{\mathrm{OSP}(1,4)} \times \frac{\mathrm{OSP}(1,4)}{\mathrm{SO}(3,2)} \times \frac{\mathrm{SO}(3,2)}{\mathrm{SO}(3,1)} = K_N \times K_{N-1} \times \cdots K_1 \times K_0.$$

At the end of this product we have the supercoset

$$\frac{\mathrm{OSP}(1,4)}{\mathrm{SO}(3,2)} \times \frac{\mathrm{SO}(3,2)}{\mathrm{SO}(3,1)}$$

considered by Zumino.[8] Supercoset OSP(K,4)/OSP($K-1$,4) can be parametrized in the following way ($1 < k < N$):

$$K_k = \exp(\theta^k Q^k + \xi^{ki} T^k_i); \quad i = 1 \cdots k-1$$

(without summation over k!).

B. Nonlinear realization on the product of symmetric supercosets

Let us consider the action of an arbitrary element $g(\rho) \in$ OSP(N,4) (ρ denotes parameters of g) on the supercoset K. According to the general scheme from Sec. II C we have (also see Ref. 20)

$$g(\rho)K_N = K'_N g_{N-1}(\rho,\theta^N,\xi^{Ni}),$$
$$g_{N-1}K_{N-1} = K'_{N-1}g_{N-2}(\rho,\theta^N,\theta^{N-1},\xi^{Ni},\xi^{N-1i}), \quad (4.1)$$
$$\cdots$$
$$g_0 e^{-ix^a P_a} = e^{-ix'^a P_a} l(\rho,\theta^i,\xi^{ij},x^a), \quad ij = 1\cdots N,$$

where g_k belongs to OSP(K,4) and l belongs to SO(3,1) but depends on ρ and the parameters of entire coset. Thus the right transformation on K is represented by the left local SO(3,1) transformation $l = l(\rho,x^a,\theta^i,\xi^{ij})$, $ij = 1\cdots N$. In Sec. III C it was shown that the components of the Cartan form ω,

$$\omega = K^{-1}\,dK = \bar{\omega}_i Q^i + \omega_{ij}T^{ij} - i\omega^a P_a + \tfrac{1}{2} i\omega^{ab}M_{ab},$$

transform under the action of an arbitrary element $g\in$OSP(N,4) by an element $l\in$SO(3,1) which depends on parameters ρ of g and on parameters of the supercoset, i.e., $l = l(\rho,x^a,\theta^i,\xi^{ij})$,

$$\omega' = l\omega l^{-1} + l\,dl^{-1}. \quad (4.2)$$

Therefore any expression constructed with $\bar{\omega}^i$, ω^{ij}, ω^a, ω^{ab} which is invariant under local SO(3,1) transformations, will also be invariant globally under the entire supergroup OSP(N,4). The first equation in (4.1) describes the nonlinear realization of OSP(N,4) on the coset OSP(N,4)/OSP($N-1$,4), the second one, describes the nonlinear realization of OSP(N-1,4) on the supercoset OSP(N-1,4)/OSP(N-2,4), and so on. We see that to find the nonlinear realization of OSP(N,4) on the supercoset OSP(N,4)/SO(3,1) it is necessary to use N times, step by step, the nonlinear realization of OSP(K,4) on the coset OSP(K,4)/OSO($K-1$,4) $1\leqslant K\leqslant N$, accordingly with formula (4.1). The formulas for nonlinear realization of OSP(K,4) are the following. (All expressions not defined here are given in Appendix D.) If $g_0\in$OSP(K,4), i.e., $g = 1 + \bar{\alpha}^{-1}Q_i + \lambda^{ij}T_{ij} - is^a P_a + \frac{1}{2}\rho^{ab}M_{ab}$, $i,j = 1\cdots k-1$ and g_0 is infinitesimal then

$$\delta\bar{\theta}^k = i\xi^j\bar{\alpha}_j + \tfrac{1}{2}ms^a\bar{\theta}^k\gamma_a\gamma_5 - \tfrac{1}{2}\rho^{ab}\bar{\theta}^k\sigma_{ab},$$
$$\delta\xi^{kj} = -m\bar{\alpha}^j\theta^k + \lambda^{ij}\xi^k_i,$$

and the transformation is linear. On the other hand, if $g_0 = \bar{\alpha}^k Q^k + \lambda^{kj}T^k_D; \lambda^{kj},\bar{\alpha}^k$ are infinitesimal, we get

$$\delta\bar{\theta}^k = \bar{\alpha}^k + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}2^{2n}}{(2n)!}B_{2n}\bar{Y}_n(\alpha^k,\lambda^{ki}),$$
$$\delta\xi^{ki} = \lambda^{ki} + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}2^{2n}}{(2n)!}B_{2n}Z^{ki}_n(\alpha^k,\lambda^{ki}).$$

The element g_{k-1} has the following form:

$$g_{k-1} = \bar{\beta}^i Q_i + V^{AB}M_{AB} + \rho^{ik-1}T_{ik-1}$$

where $A,B = 0\cdots 4$, $i,j = 1\cdots K-1$, and

$$\bar{\beta}^i = \sum_{n=4}^{\infty}\frac{(1+(-1)^{n+2})2^{2n}-2}{(2n)!}B_{2n-1}\times\bar{I}_n(\alpha^k,\lambda^{ki}),$$
$$V^{AB} = \sum_{n=1}^{\infty}\frac{(1+(-1)^{n-2})2^{2n}-2}{(2n)!}B_{2n-1}\times J^{AB}_n(\alpha^k,\lambda^{ki}),$$
$$\rho^{ik} = \sum_{n=1}^{\infty}\frac{(1+(-1)^{n-2})2^{2n}-2}{(2n)!}B_{2n-1}\times R^{ik}_n(\alpha^k,\lambda^{ki}).$$

C. Cartan form

The components of the Cartan form ω are the following:

$$\omega^k Q^k = \mathrm{ch}(\tfrac{1}{2}mx)(\bar{A}_1 + \bar{C}_1)Q^k - (1/x)(\tfrac{1}{2}mx)\times x^a(\bar{A}_1 + \bar{C}_1)\gamma_a\gamma_5 Q,$$
$$\omega^{ki}T_{ki} = (B^i_1 + D^i_1)T_{ki}; \quad i = 1\cdots K-1,$$

where A_l, C_i, B^i_l, D^i_l are defined by the recurrent formulas

$$\bar{A}_l = (\bar{A}_{l+1} + \bar{C}_{l+1} - [m/(1+x_l)]\bar{\theta}^l(\bar{A}_{l+1} + \bar{C}_{l+1})\theta^l,$$
$$\bar{C}_l = -i[B^i_{l+1} + D^i_{l+1} + (B^p_{l+1} + D^p_{l+1})\xi_{lp}]\times[1/(1+x_l)]\bar{\theta}_l,$$
$$B^i_l = [m/(1+x_l)]\bar{\theta}^l(A_{l+1} + C_{l+1})\xi^i_l,$$

and

$$\xi^l_l = 1, \quad i>1 \Rightarrow B^i_l = 0,$$
$$D^i_l T_{l+1i} = D^i_{l+1}T_{l+1i} + [1/(1+x_l)][-(B^l_{l+1} + D^l_{l+1}) + (B^p_{l+1} + D^p_{l+1})\xi^p_l T_{lp} + (B^p_{l+1} + D^p_{l+1})\xi_{lp}T_{l+1i} - (B^l_{l+1} + D^l_{l+1})x_l T_{l+1i},$$

and

$$x_l = im\,\bar{\theta}^l\theta^l + \xi^p_l\xi_{lp}, \quad p = 1\cdots l-1.$$

For $l>k$ we set

$$\bar{C}_{k+1} = 0; \quad D^i_{k+1} = 0;$$
$$\bar{A}_{k+1} = \bar{E}_{k+1} + \bar{H}_{k+1} + \bar{G}_{k+1};$$
$$B^i_{k+1} = E^i_{k+1} + F^i_{k+1} + G^i_{k+1},$$

where

$$\bar{E}_{l-1} = \bar{E}_{l-1}(\bar{E}_l + \bar{H}_l + \bar{G}_l);$$
$$\bar{H}_{l-1} = \bar{H}_{l-1}(E^{AB}_l + G^{AB}_l + H^{AB}_l);$$
$$\bar{G}_{l-1} = \bar{G}_{l-1}(\bar{E}_l + \bar{E}^k_l + F^k_l + \bar{H}_l + G^k_k + G^{kl}_k);$$
$$E^i_{l-1} = E^i_{l-1}(\bar{E}_l + G^k_l + H^k_l + F^k_l);$$
$$i,k = 1\cdots K-1;$$
$$F^k_{l-1} = F^k_{l-1}(G^{ki}_k + E^i_l + F^{ij}_k + F^{ij}_k);$$
$$F^{ik}_{l-1} = F^{ik}_{l-1}(G^{mn}_l + F^{mn}_l);$$
$$i,k = 1\cdots k-2, \quad m,n = 1\cdots k-2.$$

We use these recurrent formulas up to $l = K-1$ and finally we put

$$\bar{E}_k = \bar{E}_k(u^p); \quad H_k = H_k(u^{Ab});$$
$$\bar{F}_k = \bar{F}_k(u^{kl}); \quad G_k = G_k(\bar{u}^k,u^{kl}).$$

The remaining components are the following:

$$\omega^{AB}M_{AB}=\frac{1}{x}2\,\mathrm{ch}(mx)(x^c x_a t^{ab}M_{bc}+\frac{1}{mx}\times\mathrm{sh}(mx)\frac{1}{x^2}x^a t^{b4}M_{ab},$$

$$\omega^a=2t^{ab}x_b\frac{\mathrm{sh}(mx)}{mx}+t^{b4}(\delta^a_b+\mathrm{ch}(mx)-1)\times\left(\delta^a_b+\frac{x^a x_b}{x^2}\right),$$

where

$$t^{AB}=E_1^{AB}+H_1^{AB}+G_1^{AB},$$
$$E^{AB}_{l-1}=E^{AB}_{l-1}(\bar{E}^p_l+F^p_l+\bar{G}^p_k+H^p_k),$$
$$H^{AB}_{l-1}=H^{AB}_{l-1}(E^{AB}_l+H^{AB}_l+G^{AB}_l),$$
$$G^{AB}_{l-1}=G^{AB}_{l1}(E^k_k+H^k_k+G^k_k+F^{kl}_k+E^{kl}_k+G^{kl}_k).$$

Finally we put

$$E^{AB}_k=E^{AB}_k(u^r);\quad H^{AB}_k(u^r);$$

Explicit forms of above formulas are given in Appendix D. As it was mentioned above the supergroup OPS(N,4) acts on the components ω^a and ω^{ij} of the Cartan form by a transformation belonging to SO(3,1) subgroup [see formula (4.2)]. We shall construct, using these components of ω, the SO(3,1) invariants, which will be also OSP(N,4) invariants. More precisely, the forms ω^a transform as follows:

$$\omega'^a P_a=l_1\omega^a P_a l_1^{-1};\quad l_1\in\mathrm{SO}(3,1),$$

and because we have $[M_{ab},T_{ij}]=0$, ω^{ij} are the Lorentz scalars

$$-\omega'^{lk}T_{lk}=l_1\omega^{lk}T_{kl}l_1^{-1}=\omega^{lk}T_{lk}.$$

Let us write the first few terms of ω^a and ω^{ij}:

$$\omega^a=dx^a-\tfrac{1}{2}\bar{\theta}^i\gamma^a\gamma^5\,d\theta_i-im\bar{\theta}_i\sigma^{ab}\,d\theta^i x_b+\tfrac{1}{4}(\xi^2_{12}+2imx)\bar{\theta}^i\gamma^a\gamma^5\,d\theta^i+\tfrac{1}{2}\bar{\theta}^1\gamma^a\gamma^5\,\theta^2+\cdots,$$

$$\omega^{ij}=d\xi^{ij}+\xi^{mi}\,d\xi^j{}_m+\xi^{mj}\,d\xi^i{}_m+m\theta^i\,d\theta^j+2im^2\,d\theta^k\,\theta^i\theta^j\theta_k+\cdots.$$

The forms ω^{ij} can be written

$$\omega^{ij}=\omega^{ij\mu}e_\mu{}^a\omega_a=\omega^{ij\mu}e^a_\mu e^\nu_a\,dx_\nu$$
$$=\left(\frac{\partial\xi^{ij}}{\partial x^\mu}+\xi^{mi}\frac{\partial\xi^{mj}}{\partial x^\mu}+\xi^{mj}\frac{\partial\xi^i{}_m}{\partial x^\mu}+\cdots\right)\times e^a_\mu e^\nu_a\,dx_\nu,$$

where e^a_μ are the vierbeins defined by the relation

$$g_{\mu\nu}=e^a_\mu e_{a\nu}=\eta_{ab}e^a_\mu e^b_\nu,$$

$g_{\mu\nu}$ are the metrios in anti-de-Sitter space, and η_{ab} are in Minkowski space.

We define the scalar product

$$\langle\omega_a,\omega_b\rangle=\eta_{ab}. \tag{4.3}$$

Using (4.3) we get

$$\langle\omega^{ij},\omega_{ij}\rangle=\left(\frac{\partial\xi^{ij}}{\partial x^\mu}+\xi^{mi}\partial\xi^j{}_m+\xi^{mj}\frac{\partial\xi^i{}_m}{\partial x^\mu}+\cdots\right)\times\left(\frac{\partial\xi_{ij}}{\partial x^\nu}+\xi_{mi}\frac{\partial\xi^m{}_j}{\partial x^\nu}+\cdots\right)\times e^a_\mu e^b_\nu\langle\omega_a,\omega_b\rangle$$
$$=\frac{\partial\xi^{ij}}{\partial x^\mu}\frac{\partial\xi_{ij}}{\partial x^\nu}g^{\mu\nu}+\frac{\partial\xi^{ij}}{\partial x^\mu}\xi_{mi}\frac{\partial\xi^m{}_j}{\partial x^\nu}g^{\mu\nu}+\cdots. \tag{4.4}$$

The formula (4.4) defines a Lorentz scalar. We see that the following action:

$$L=\int\mu^0\mu^1\mu^2\mu^3(1+\langle\omega^{ij},\omega_{ij}\rangle)$$
$$=\int d^4x\left(-\frac{1}{a^2}-\frac{1}{2}i\lambda^i\gamma\cdot\partial\lambda_i-im\bar{\lambda}_i\lambda^i_1+\cdots\right)\times\left(1+\frac{\partial\xi^{ij}}{\partial x^\mu}\frac{\partial\xi_{ij}}{\partial x^\nu}g_{\mu\nu}+\frac{\partial\xi^{ij}}{\partial x^\mu}\xi_{mi}\frac{\partial\xi^m{}_j}{\partial x^\nu}g_{\mu\nu}+\cdots\right)$$
$$=\int d^4x\left(-\frac{1}{a^2}-\frac{\partial\xi'^{ij}}{\partial x^\mu}\frac{\partial\xi_{ij}}{\partial x^\nu}g^{\mu\nu}-\frac{1}{2}i\bar{\lambda}^i\gamma\cdot\partial\lambda_i-im\bar{\lambda}_i\lambda^i+\cdots\right)+\cdots,$$

where $\xi'^{ij}=(1/a)\xi^{ij}$ and $\bar{\lambda}^i=(1/a)\times(1+\frac{1}{2}m\gamma^a x_a+\cdots)\theta^i_{21}$ is Lorentz invariant and therefore OSP(N,4) invariant.[21]

This is a nonlinear action describing N-Goldstone fermions interacting with $N(N-1)/2$ Goldstone scalar fields $\xi^{ij}(x)$ in anti-de-Sitter space. The action is globally OSP(N,4) supersymmetric. If we take the OSP(K,4) subgroup $K\leqslant N$ then all fields $\theta^i(x)$ and $\xi^{ij}(x)$, where $l>K$ will transform linearly under OSP(K,4) and remaining fields with index $l\leqslant K$ in a nonlinear way.

D. Nonlinear realization of OSP(2,4) supergroup on the supercoset OSP(2,4)/OSP(1,4)

The supergroup OSP(2,4) can be parametrized in the following way:

$$\exp(\bar{\theta}^2Q^2+\xi T)\exp(\bar{\theta}^1Q^1)\exp-ix^aP_a l_0,$$

where l_0 is an element of SO(3,1) group. The supercoset parametrized by the exponential

$$\exp(\bar{\theta}^2Q^2+\xi T) \tag{4.5}$$

is a symmetric space. The transformation laws of the parameters θ^2 and ξ belonging to the supercoset (4.5) are the following:

(i) $g_0\in\mathrm{OSP}(1,4)$,

i.e., $g_0=1+\alpha^1Q^1-is^aP_a+\frac{1}{2}iv^{ab}M_{ab}$

and g_0 is infinitesimal then

$$\delta\bar{\theta}^2=i\xi\bar{\alpha}^1+\tfrac{1}{2}ms^a\bar{\theta}^2\gamma^a\gamma^5-\tfrac{1}{2}v^{ab}\bar{\theta}^2\sigma_{ab},$$
$$\delta\xi=-m\bar{\alpha}^1\theta^2. \tag{4.6}$$

Both parameters transform in a linear way.

(ii) $g = 1 + \bar{\alpha}^2 Q^2 + \lambda T$, θ^2, λ are infinitesimal then

$$\delta\bar{\theta}^2 = \bar{\alpha}^2 + \sum_{n=1}^{\infty} \frac{(-\lambda)^{n+1} 2^{2n}}{(2n)!} B_{2n-1} \bar{Y}_n,$$

$$\delta\xi = \lambda + \sum_{n=1}^{\infty} \frac{(-1)^{n+1} 2^{2n}}{(2n)!} B_{n-1} z_n,$$

where

$$\begin{aligned} Y_n = \{ & \xi^{2n} \bar{\alpha}^2 + \tfrac{1}{2} im \xi^{2u-2} [-5(n-1)^2 + \tfrac{1}{4}(n-2)^2 \bar{\theta}^2 \theta^2 \bar{\theta}^2 \bar{\alpha}^2 \\ & + n + \tfrac{1}{2}(n-1) \bar{\theta}^2 \Gamma_A \theta^2 \bar{\alpha}^2 \Gamma^A] - m^2 \xi^{2u-4} \Box(2n-4) (-5(n-1)^2 + \tfrac{1}{4}(n-2)^2 (\bar{\theta}^2 \theta^2)^2 \alpha^2 \\ & - \lambda \xi^{2n-1} \bar{\theta}^2 + im(n-1) \xi^{2u-3} \Box(2n-3) \lambda \bar{\theta}^2 \bar{\theta}^2 \bar{\theta}^2 \}, \\ Z_n = \{ & im \xi^{2k-1} \bar{\theta}^2 \alpha^2 + m(n-1) \xi^{2u-3} \Box(2n-3) \bar{\theta}^2 \theta^2 \bar{\theta}^2 \alpha^2 \\ & - im\lambda \xi^{2n-3} \bar{\theta}^2 \theta^2 - m^2 \lambda (n-1) \xi^{2u-4} \Box(2n-4) (\bar{\theta}^2 \theta^2)^2 \}. \end{aligned}$$

The function $\Box(k)$ is defined in the following way:

$$\Box(k) = \begin{cases} 1, & \text{if } k \geqslant 0, \\ 0, & \text{if } k < 0. \end{cases}$$

The Cartan form on the coset OSP(2,4)/SO(3,1) parametrized by the exponentials

$$\exp(\bar{\theta}^2 Q^2 + \xi T) \exp(\bar{\theta}^1 Q) \exp(-i x^a P_a)$$

take the form

$$\begin{aligned} \omega^a = & dx^a + \left(\frac{\operatorname{sh}\sqrt{m^2 x^2}}{mx} - 1 \right) \left(dx^a - \frac{x^b \, dx_b \, x^a}{x^2} \right) + 2t^{ab} x_b \frac{\operatorname{sh}\sqrt{mx^2}}{mx} \\ & + s^b \left[\delta_b^a - (\operatorname{ch}\sqrt{m^2 x^2} - 1) \left(\delta_b^a - \frac{x^b s_b x^a}{x^2} \right) \right], \\ \omega^{ab} = & -\frac{1}{x^2} \operatorname{ch}\sqrt{m^2 x^2} x^a \, dx^b + \frac{1}{x^2} \operatorname{ch}\sqrt{m^2 x^2} t^{ab} x_c x^b - \operatorname{sh}(mx) \left(\frac{m}{x} \right) x^a s^b, \\ \omega = & z(1 - \tfrac{1}{2} im \bar{\theta}^1 \theta^1 + \tfrac{1}{4} m^2 (\bar{\theta}^1 \theta^1)^2 - m \bar{\theta}^1 Y) + im^2 \bar{\theta}^1 Y \bar{\theta}^1 \theta^1, \\ \bar{\omega}^2 = & \operatorname{ch}\left(\frac{1}{2} mx \right) \bar{P}_2 + \frac{1}{x} \operatorname{sh}\left(\frac{1}{2} mx \right) x^a \bar{P}_2 \gamma_a \gamma_5, \\ \omega^1 = & \operatorname{ch}\left(\frac{1}{2} mx \right) \bar{P}_1 + \frac{1}{x} \operatorname{sh}\left(\frac{1}{2} mx \right) x^a \bar{P}_1 \gamma_a \gamma_5, \end{aligned}$$

where

$$\begin{aligned} s^a = & \tfrac{1}{2}(1 + im \bar{\theta}^1 \theta^1)(\bar{\theta}^1 \gamma^a \gamma^5 \, d\theta^1) + x \theta^2 \gamma^a \gamma^5 \, d\theta^2 + A \theta^1 \gamma^a \gamma^5 \theta^2 \\ & + B \bar{\theta}^1 \gamma^a \gamma^5 \, d\theta^2 - (1 + \tfrac{1}{3} im \bar{\theta}^1 \theta^1 \gamma^a \gamma^5 V), \\ t^{ab} = & m(1 + im \bar{\theta}^1 \theta^1) \bar{\theta}^1 \sigma^{ab} \, d\theta^1 + mx \bar{\theta}^2 \sigma^{ab} \, d\theta^2 + mA \bar{\theta}^1 \sigma^{ab} \theta^2 + mB \bar{\theta}^1 \sigma^{ab} \, d\theta^1 \\ & - m(1 + \tfrac{1}{3} m \bar{\theta}^1 \theta^1) \bar{\theta}^1 \sigma^{ab} V, \end{aligned}$$

and

$$\begin{aligned} A = & X(-2im\theta^1 \, d\theta^2 - 4m \bar{\theta}^1 \, d\theta^2 \bar{\theta}^1 \theta^1), \\ B = & X(2im\theta^2 \, d\theta^1 + 4m \bar{\theta}^2 \theta^1 \bar{\theta}^1 \theta^1), \\ X = & \sum_{k=0}^{\infty} \frac{1}{(2k+2)!} [\xi^{2k} - 2kim \xi^{2k-2} \Box[2k-2]], \\ \bar{V} = & \sum_{k=0}^{\infty} \frac{1}{(2k+2)!} \bar{V}_k \\ = & \sum_{k=0}^{\infty} \frac{1}{(2k+2)!} [i \xi^{2k+1} \, d\bar{\theta}^2 + \xi^{2k} \, d\xi \, \bar{\theta}^2 \\ & + 3m \xi^{2k-1} \Box(2k-1) k \bar{\theta}^2 \, d\theta^2 \bar{\theta}^2 \\ & + 2m \xi^{2k-1} \Box(2k-1) k \bar{\theta}^2 \theta^2 \, d\bar{\theta}^2 \end{aligned}$$

$$+ im^2\xi^{2k-3}\Box(2k-3)(k-1)^2(\tfrac{1}{4}-5)(\bar\theta^2\theta^2)^2 d\theta^2$$
$$- mk\xi^{2k-2}\Box(2k-2)d\xi\,\bar\theta^2\theta^2\theta^2],$$
$$Z = d\xi + \sum_{k=1}^{\infty}\frac{1}{(2k+1)}Z_k,$$
$$\overline{Y} = d\theta^2 + \sum_{k=1}^{\infty}\frac{1}{(2k+1)}\overline{Y}_k.$$

Here $\overline{Y}_k$ and Z_k are given by (4.7) provided that we put $\alpha^2 = d\theta^2$ and finally

$$\overline{P}_2 = \overline{Y} + \frac{1}{2}im\bar\theta^1 Y\bar\theta^1 + \frac{1}{u!}m^2\bar\theta^1 Y\bar\theta^1\theta^1\bar\theta^1 + iZ\bar\theta^1 - \frac{1}{3!}zm\bar\theta^1\theta^1\bar\theta^1,$$
$$\overline{P}_1 = d\bar\theta^1 + \frac{1}{3}im\bar\theta^1 d\theta^1\theta^1 + \frac{1}{3}im\bar\theta^1\theta^1 d\bar\theta^1 - \frac{1}{4!}m^2(\bar\theta^1\theta^1)^2 d\theta^1 + V$$
$$- im\bar\theta^1 V\bar\theta^1 + im\bar\theta^1\theta^1\overline{V} - \frac{5}{4!}m(\bar\theta^1\theta^1)^2\overline{V} + X(2im\bar\theta^1 d\theta^2\,\bar\theta^2 - 2im\bar\theta^1\theta^2 d\theta^2$$
$$- 4m^2\bar\theta^1\theta^1\bar\theta^2\theta^1 d\theta^2 + 4m^2\bar\theta^1\theta^1\bar\theta^1 d\theta^2\theta^2).$$

V. NONLINEAR REALIZATION OF OSP (2N,4) SUPERGROUP ON THE COSET OSP (2N,4)/ SO (3,1)×U(N)

A. Algebraic considerations

Following Lukierski and Rytel[15] we introduce for OSP (2N,4) the following pair of N-component Majorana spinor:

$$Q_\alpha^{i\pm} = \frac{1}{2}(Q_\alpha^i \pm \gamma_5 Q_\alpha^{i+N}); \quad i = 1\cdots N.$$

The internal symmetry algebra O(2N) has the symmetric subalgebra U (N). The generators of this subalgebra are A^{ij} and S^{ij}, so any elment of the group U(N) can be written as

$$\exp(a_{ij}A^{ij} + t_{ij}S^{ij}), \quad i,j = 1\cdots N;$$

and the coset O(2N)/U(N) can be described as

$$\exp(u_{ij}X^{ij} + v_{ij}Y^{ij}), \quad i,j = 1\cdots N,$$

where X^{ij} and Y^{ij} are the generators of the coset [see Appendix C—nonlinear realization of O (2N)].

The commutation relations for internal symmetry sector O(2N), i.e., for A_{ij}, S_{ij}, X_{ij}, Y_{ij} are given in Appendix C. Anticommutation relations for $Q^{i\pm}$ are the following[15]:

$$\{Q_\alpha^{i+},Q_\beta^{j+}\} = \delta^{ij}\gamma^c\gamma^0_{\alpha\beta}P_c + m[\gamma^0_{\alpha\beta}X^{ij} + (\gamma_0\gamma_5)_{\alpha\beta}Y^{ij}], \tag{5.1}$$
$$\{Q_\alpha^{i+},Q_\beta^{j-}\} = m\delta^{ij}(\sigma^{ab}\gamma^0\gamma^5)_{\alpha\beta}M_{ab} + m[\gamma^0_{\alpha\beta}A^{ij} + (\gamma^0\gamma^5)_{\alpha\beta}S^{ij}], \tag{5.2}$$
$$\{Q_\alpha^{i-},Q_\beta^{i-}\} = \delta^{ij}(\gamma^a\gamma^0)_{\alpha\beta}P_a + m[\gamma^0_{\alpha\beta}X^{ij} - (\gamma^0\gamma^5)_{\alpha\beta}Y^{ij}]. \tag{5.3}$$

The covariance relations for supercharges are as follows:

$$[M_{ab},Q^{i\pm}] = i\sigma^{ab}Q^{i\pm}; \quad [P_a,Q^{i\pm}] = i\tfrac{1}{2}m(\gamma_a\gamma_5)\,Q^{i\pm}; \tag{5.4a}$$
$$[A_r,Q^{i\pm}] = iJ_r^{ij}Q^{j\pm}; \quad [S_v,Q^{i\pm}] = iJ_r^{ij}Q^{j\pm}; \tag{5.4b}$$

where $A^{ij} = A_rJ_r^{ij}$. Additionally we have

$$[X_r,Q^{i\pm}] = J_r^{ij}Q^{j\mp}; \quad [Y_r,Q^{i\pm}] = J_r^{ij}\gamma^5Q^{j\mp}. \tag{5.5}$$

The OSP (2N,4) superalgebra admits Z_4 grading,[15,22] with the following four sectors:

L_0	L_1	L_2	L_3
M_{ab}	Q_α^{i+}	P_a	Q_α^{i-}
A_{ij}		FX^{ij}	
S_{ij}		Y^{ij}	

which satisfy the grading relation

$$[L_m,L_n] \subset L_{m+n}, \quad n+m \bmod 4.$$

This gives the possibility of performing the following rescaling[15]:

$$L_0\to L_0; \quad L_1\to\frac{1}{\sqrt{r}}L; \quad L_2\to\frac{1}{r}L_2; \quad L_3\to\frac{1}{\sqrt{r}}L_3.$$

We see that the relations (5.1), (5.3), and (5.4) are unchanged, but the relations (5.2) and (5.5) are rescaled, and we get

$$[X_r,Q^{i\pm}]\sim\frac{1}{r}; \quad [Y_r,Q^{i\pm}]\sim\frac{1}{r}, \tag{5.6}$$
$$[Q^{i+};Q^{j-}]\sim\frac{1}{r}; \quad [P_a,P_b]\sim\frac{1}{r^2}. \tag{5.7}$$

For the internal symmetry group we obtain

$$[X^{ij},X^{kl}]\sim\frac{1}{r^2}U(N); \quad [X^{ij},Y^{kl}]\sim\frac{1}{r^2}U(N);$$
$$[Y^{ij},Y^{kl}]\sim\frac{1}{r^2}U(N). \tag{5.8}$$

From (5.7) we see that in the contraction limit $r\to\infty$ sectors Q^{i+} and Q^{i-} decouple, so there remain only three sectors L_0,L_1,L_2. The relation (5.8) implies that Y^{ij} and X^{ij} become Abelian generators, and describes central charges, because they commute with all generators of superalgebra.

The generators P_a, in the contraction limit also became Abelian. So setting $r\to\infty$ we obtain N-extended Poincaré superalgebra, with $N(N-1)$ central charges.[15] Group

$U(N)$ is now the internal symmetry group. The entire group $U(N)$ appears as unbroken symmetry only when central charges vanish. When central charges are present, the internal symmetry group must be broken to its subgroup commuting with central charges. In our case the number of central charges $N(N-1)$ does not admit internal symmetry group $U(N)$. So taking the limit $\bar{r}\to\infty$ we shall set also a^{ij} and t^{ij} equal to zero. Such a reduction of internal symmetry group was considered in Refs. 23 and 24.

B. Nonlinear realization of OSP (2N,4)

Now we use the rescaled commutation relations to obtain nonlinear realization of the OSP $(2N,4)$ supergroup on the supercoset $OSP(2N,4)/SO(3,1)\times U(N)$. We use the following parametrization of the supercoset:

$$\exp(\bar{\theta}_i^{\,+}Q^{i+}+\bar{\theta}_i^{\,-}Q^{i-})\exp(u_{ij}X^{ij}+v_{ij}Y^{ij})\exp(-ix^aP_a).$$

The parameters θ^{+i} and $\bar{\theta}^{-i}$ transform under the action of the group element $g_0\in SO(3,2)\times O(2N)$, i.e.,

$$g_0=1+u_{ij}X^{ij}+v_{ij}Y^{ij}+a_{ij}A^{ij}+t_{ij}S^{ij}-is^aP_a+\tfrac{1}{2}\rho^{ab}M_{ab},$$

as follows

$$\delta\bar{\theta}_l^{\,+}=\frac{2}{r}u_{kl}\bar{\theta}^{k-}-\frac{2}{r}iv_{kl}\bar{\theta}^k\gamma^5+2ia_{kl}\bar{\theta}^{k+}+2it_{kl}\bar{\theta}^k+\frac{1}{2r}ms^a\bar{\theta}_a^{\,+}\gamma^5\gamma_a-\tfrac{1}{2}\rho^{ab}\bar{\theta}_l^{\,+}\sigma_{ab},$$

$$\delta\bar{\theta}_l^{\,-}=2ia_{kl}\bar{\theta}^k+2it_{kl}\bar{\theta}^{k}_{+}+\frac{2i}{r}u_{kl}\bar{\theta}^{k}_{+}+\frac{1}{r}2iv_{kl}\bar{\theta}^k\gamma^5+\frac{1}{r}ms^a\bar{\theta}_l^{\,-}\gamma_a\gamma_5-\tfrac{1}{2}\rho^{ab}\bar{\theta}_l^{\,-}\sigma_{ab}.$$

The transformations are linear.

The nonlinear transformations are obtained if $g_c=1+\bar{\alpha}_i^{\,+}Q^{i}_{\,+}+\bar{\alpha}_i^{\,-}Q^{i}_{\,-}$, and $\alpha_i^{\pm}$ are infinitesimal. In such a case

$$\delta\bar{\theta}_l^{\,+}=\bar{\alpha}_l^{\,+}+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}2^{2n}}{(2n)!}B_{2n-1}\bar{\theta}_i^{\,+(2n)},\tag{5.9}$$

$$\delta\bar{\theta}_l^{\,-}=\bar{\alpha}_l^{\,-}=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}2^{2n}}{(2n)!}B_{2n-1}\bar{\theta}_i^{\,-(2n)},\tag{5.10}$$

where

$$\bar{\theta}_l^{\,+(2n)}=\frac{1}{2r}mq^a_{2n-1}\bar{\theta}_l^{\,+}\gamma_a\gamma_t-w^{ab}_{(2n-1)}\bar{\theta}_l^{\,+}\sigma_{ab}-2i(a_{(2n-1)})_{ab}\bar{\theta}^{k}_{+}-2i(t_{2n-1})_{kl}\bar{\theta}^{k}_{+}-\frac{1}{r}2i(u_{2n-1})_{kl}\bar{\theta}^{k}_{-}+\frac{2i}{r}(v_{(2n-1)})_{kl}\bar{\theta}^{k}_{-}\gamma^5,\tag{5.11}$$

$$\bar{\theta}_l^{\,(2n)}=\frac{1}{2r}mq^a_{(2n-1)}\bar{\theta}_i\gamma_a\gamma_5-w^{ab}_{(2n-1)}\bar{\theta}_l^{\,-}\sigma_{ab}-2i(a_{(2n-1)})_{kl}\bar{\theta}^k-2i(t_{2n-1})_{kl}\bar{\theta}^k-2i\frac{1}{r}(u_{(2n-1)})_{kl}\bar{\theta}^{k}_{+}\frac{2}{r}i(v_{(2n-1)})_{kl}\bar{\theta}^{b}_{+}\gamma^5,$$

and

$$q^a_{(2n-1)}=-\left[\bar{\theta}^{\,i}_{+}\gamma^a\gamma^5\bar{\theta}_i^{\,+(2n-2)}+\bar{\theta}^{\,i}_{+}\gamma^a\gamma^5\bar{\theta}_i^{\,-(2n-2)}\right],$$

$$w^{ab}_{(2n-1)}=\frac{1}{r}m\left[\bar{\theta}^{\,i}_{+}\sigma^{ab}\bar{\theta}_i^{\,+(2n-2)}+\bar{\theta}^{\,i}\sigma^{ab}\theta_i^{(2n-2)}\right],$$

$$(u_{2n})_{ij}=-\tfrac{1}{2}m\left[(\bar{\theta}_i^{\,+}\gamma_5\theta_j^{\,+(2n-2)}-\bar{\theta}_j^{\,+}\gamma^5\theta_i^{\,+(2n-2)})+(\bar{\theta}_i^{\,-}\gamma^5\theta_j^{\,-(2n-2)}-i\leftrightarrow j)\right],\tag{5.12}$$

$$v_{(2n-1)}=\tfrac{1}{2}m\left[(\bar{\theta}_l^{\,+}\theta_j^{\,+(2n-1)}-\bar{\theta}_j^{\,+}\theta_l^{\,+(2n-2)})-(\bar{\theta}_i^{\,-}\theta_j^{\,-(2n-2)}-\bar{\theta}_j^{\,-}\theta_i^{\,+(2n-2)})\right],$$

$$(a_{2u-1})_{ij}=\frac{1}{2r}m\left[(\bar{\theta}_i^{\,-}\gamma^5\theta_j^{\,+(2u-2)}-\bar{\theta}_j^{\,-}\gamma^5\theta_i^{\,+(2u-2)})-(\bar{\theta}_i^{\,+}\gamma^5\bar{\theta}_j^{\,-(2u-2)}-i\leftrightarrow j)\right],$$

$$(t_{2u-1})_{ij}=\frac{1}{2r}m\left[(\bar{\theta}_i^{\,+}\theta_j^{\,-(2u-2)}+\bar{\theta}_j^{\,+}\theta_i^{\,-(2u-2)})-(\bar{\theta}_i^{\,-}\theta_j^{\,+(2u-2)}+\bar{\theta}_j^{\,-}\bar{\theta}_i^{\,+(2u-2)})\right],$$

$$\bar{\theta}_i^{\,+(0)}=\bar{\alpha}_i^{\,+},\quad\text{and}\quad\bar{\theta}_l^{\,-(0)}=\bar{\alpha}_i^{\,-}.$$

The action of the element g_0 on the supercoset $\exp\bar{\alpha}_i^{\,+}Q^{+i}+\bar{\alpha}_i^{\,-}Q^{-i}$ defines the element $h_1\in SO(3,2)\times O(2N)$ (see Sec. II B), via the formula

$$h_1=q^aP_a+w^{ab}M_{ab}+u^{ij}X_{ij}+q^{ij}Y_{ij}+E_{ij}S^{ij}+a^{ij}A_{ij},$$

where

$$q^a=-\sum_{k=1}\frac{(1+(-1)^{k+2})2^{2k}-2}{(2k)!}B_{2k-1}q^a_{(2k-1)},$$

$$w^{ab}=-\sum_{k=1}B_{2k-1}w^{ab}_{(2k-1)},$$

$$u^{ij}=-\sum_{k=1}B_{2k-1}u^{ij}_{(2k-1)},$$

$$a^{ij}=-\sum_{k=1}B_{2k-1}a^{ij}_{(2k-1)},$$

$$t^{ij}=-\sum_{k=1}B_{2k-1}t^{ij}_{(2k-1)}.$$

This gives

$$\delta x^a=q^a+\left(\sqrt{\frac{m^2x^2}{r^3}}\,\mathrm{ch}\sqrt{\frac{m^2x^2}{r^3}}-1\right)\times\left(q^a-\frac{x^bq_bx^a}{x^2}\right).\tag{5.13}$$

Using the results of Appendix C we obtain

$$\delta u^{ij}=p^{ij}+\sum_{k=1}^{\infty}\frac{(-1)^{u+1}2^{2u}}{(2n)!}B_{2u-1}\hat{u}^{ij}_{2u},\tag{5.14}$$

$$\delta v^{ij}=q^{ij}+\sum_{k=1}^{\infty}\frac{(-1)^{u+1}2^{2u}}{(2u)!}B_{2u-1}\hat{v}^{ij}_{2u},\tag{5.15}$$

where $\hat{u}^{ij}_{2u}$ and $\hat{v}^{ij}_{2u}$ are defined in Appendix C.

C. Cartan form

The Cartan form for the coset $OSP(2N,4)/SO(3,1)\times U(N)$ is the following:

$$\omega = e^{ik^a P_a} e^{(u_{ij}x^{ij} + v_{ij}y^{ij})} e^{(\bar\theta_i{}^+ Q^i{}_+ + \bar\theta_i{}^- Q^i}$$
$$\times d(e^{\bar\theta_i Q^i{}_+ - \bar\theta_i{}^- \bar Q^i}) e^{(u_{ij}x^{ij} + v_{ij}y^{ij})} e^{-ix^a P_a}$$
$$= \omega_a^{ij} A_{ij} \pm \bar\omega_i^+ Q_i + \bar\omega_i{}^- \bar Q^{i-} + \omega_x^{ij} X_{ij}$$
$$+ \omega_y^{ij} Y_{ij} - i\omega^a P_a + \tfrac{1}{2}\omega^{ab} M_{ab} + \omega^{ij}{}_+ \delta_{ij}.$$

Using the same methods as in previous cases we obtain

$$\bar\omega_l^+ = \mathrm{ch}\left(\frac{1}{2r} mx\right)\left[\bar\xi^{+k} \exp(-a_{kl} + t_{kl}) - \frac{1}{x}\mathrm{sh}\left(\frac{1}{2r} mx\right)\right]$$
$$\times [\bar\xi^{+k} \exp(-a_{kl} + t_{kl})\gamma_a\gamma_5 x^a],$$
$$\bar\omega_l^- = \mathrm{ch}\left(\frac{1}{2r} mx\right)\left[\bar\xi^{-k} \exp(-a_{kl} + t_{kl}) - \frac{1}{x}\mathrm{sh}\left(\frac{1}{2r} mx\right)\right]$$
$$\times [\bar\xi^{-k} \exp(-a_{kl} + t_{kl})\gamma^a\gamma^5 x_a],$$

where

$$\bar\xi_l{}^+ = \bar\lambda_l{}^+ + \frac{2i}{r}\bar\lambda^{k-} D_{k6} + \frac{1}{2!}\left(\frac{2i}{r}\right)^2 \lambda^{k+} B_k{}^m D_{ml}$$
$$+ \frac{1}{3!}\left(\frac{2i}{r}\right)^3 \lambda^{k-} D_k{}^m B_m{}^k D_{kl}$$
$$+ \frac{1}{n!}\left(\frac{2i}{r}\right)^4 \lambda^{k+} B_m^k D_m{}^n B_n{}^p D_{pl} + \cdots,$$
$$\bar\xi_l{}^- = \bar\lambda_l{}^- + \left(\frac{2i}{r}\right)\bar\lambda^{k+} B_{kl} + \frac{1}{2!}\bar\lambda^{k-}\left(\frac{2i}{r}\right)^2 D_k{}^m B_{ml}$$
$$+ \frac{1}{3!}\left(\frac{2i}{r}\right)^3 \lambda^{k-} B_k{}^m D_m{}^n B_{kl}^+,$$
$$B_{kl} = u_{kl} + v_{kl}\gamma^5; \quad D_{kl} = u_{kl} - v_{kl}\gamma^5,$$

and

$$\bar\lambda_l{}^+ = \sum_{k=0}^{\infty} \frac{1}{(2k+1)!}\theta_l{}^{+(2k)};$$
$$\bar\lambda_l{}^- = \sum_{k=0}^{\infty} \frac{1}{(2k+1)}\bar\theta_l{}^{-(2k)},$$

where $\bar\theta_l{}^{+(2k)}$ and $\bar\theta_l{}^{-(2k)}$ were defined in (5.11). For $k=0$ we have

$$\bar\theta_l{}^{+(0)} = d\bar\theta_l{}^+ \quad \text{and} \quad \bar\theta_l{}^{-(0)} = d\bar\theta_l{}^-.$$

The forms ω_x^{il} and ω_y^{il} are the following:

$$\omega_x^{il} = \omega_x^{1il} + \omega_x^{2il}; \quad \omega_y^{1il} = \omega_y^{il} + \omega_y^{2il}$$

and we have

$$\omega_x^{1il} = \sum_{k=0}^{\infty} \frac{1}{(k!)}\omega_x^{il}(k); \quad \omega_y^{1il} = \sum_{k=0}^{\infty} \frac{1}{(k!)}\omega_y^{il}(k),$$

where

$$\omega_{x(k)}^{il} = -2(t^i_k \omega^{kl}_{y(k-1)} - a^i_k \omega^{kl}_{y(k+1)}); \quad \omega_{x(0)}^{il} = \delta u^{il},$$
$$\omega_{y(k)}^{il} = 2(a^i_k \omega^{il}_{y(k-1)} + t^i_k \omega^{kl}_{y(k-1)}); \quad \omega_{y(0)}^{il} = \delta v^{il},$$

δu^{il} and δv^{il} are defined as follows:

$$\delta u^{ij} = \sum_{k=0}^{\infty} \frac{1}{(2k)!}\hat u_{(2k)}^{ij} \quad \text{and} \quad \hat u_{(0)}^{ij} = m^{ij},$$
$$\delta v^{ij} = \sum_{k=0}^{\infty} \frac{1}{(2k)!}\hat v_{(2k)}^{ij} \quad \text{and} \quad \hat v_{(0)}^{ij} = n^{ij}.$$

Finally,

$$m^{ij} = -\sum_{k=0}^{\infty} \frac{1}{(2k+2)!} u_{(2k+1)}^{ij} \quad \text{and} \quad \bar\theta_{(0)}^{i+} = d\bar\theta^{i+}$$
$$n^{ij} = \sum_{k=0}^{\infty} \frac{1}{(2k+2)!} v_{(2k+1)}^{ij} \quad \text{and} \quad \theta_{(0)}^{l} = d\bar\theta^{i}$$

The quantities $v_{(2k+1)}^{ij}$ and $u_{(2k+1)}^{ij}$ were defined in (5.12). We shall define also

$$\omega_x^{2ij} = \sum_{k=0}^{\infty} \frac{1}{(k!)}\omega_{x(k)}^{ij}; \quad \omega_y^{2ij} = \sum_{k=0}^{\infty} \frac{1}{(k!)}\omega_{y(k)}^{ij}$$

but now

$$\omega_{x(0)}^{ij} = m'^{ij} \quad \text{and} \quad \omega_{y(0)}^{ij} = n'^{ij},$$

where

$$m'^{ij} = \sum_{k=0}^{\infty} \frac{1}{(2k+1)!}\hat u_{(2k)}^{ij}, \quad \text{and} \quad u_{(0)}^{ij} = du^{ij},$$
$$n'^{ij} = \sum_{k=0}^{\infty} \frac{1}{(2k+1)}\hat v_{(2k)}^{ij} \quad \text{and} \quad v_{(0)}^{ij} = dv^{ij},$$

with u^{ij} and v^{ij} describing the parameters of the coset $O(2N)/U(N)$. The remaining components of the Cartan form ω are the following:

$$\omega^a = q'^a + \left(\mathrm{ch}\left(\frac{mx}{r}\right) - 1\right)\left(q'^a + \frac{x_b q'^b x^a}{x^2}\right)$$
$$+ \frac{2}{mx}\mathrm{sh}\left(\frac{mx}{r}\right) w'^{ab} x_b,$$
$$\omega^{ab} = \left(\frac{m}{x}\right)\mathrm{sh}\left(\frac{mx}{r}\right) x^a q'^b - 2\left(\frac{m^2}{x^2}\right)\mathrm{ch}\left(\frac{mx}{r}\right) w'^{ca} x_c x^b,$$

where

$$q'^a = -\sum_{k=0}^{\infty} q^a_{(2k+1)} \frac{1}{(2k+2)} \quad \text{and} \quad \bar\theta_{(0)}^{i+} = d\bar\theta^{i+},$$
$$w'^a = \sum_{k=0}^{\infty} \frac{1}{(2k+2)} w^{ab}_{(2k+1)} \quad \text{and} \quad \bar\theta^{i} = d\bar\theta^{i+},$$

$q^a_{(2k+1)}$ and $w^{ab}_{(2k+1)}$ are defined in (5.12).

The forms ω^a depend on θ', $d\theta'$, x^a, dx^a only. Then, similarly as in Sec. III we can construct purely fermionic nonlinear Lagrangian density $\mathcal{L}$ defined as follows:

$$\omega^0\omega^1\omega^2\omega^3 = -a^2\,\mathcal{L}\,dx^0\,dx^1\,dx^2\,dx^3, \quad [a] = 1/x^2.$$

Using correctly normalized Goldstone fermion field $\psi^i = (1/a)\theta^i$ and using the Lee–Gursey transformation given in Sec. III we can write the first few terms of $\mathcal{L}$

$$\mathcal{L} = -\frac{1}{Q^2} - \frac{1}{2} i\bar\lambda_i{}^- \gamma\cdot\partial\lambda^{-i} - im\bar\lambda_i{}^-\lambda^{-i}$$
$$-\frac{1}{2}(\bar\lambda_i{}^+\gamma\cdot\partial\lambda_i{}^+ - im\bar\lambda_i^1\lambda^{i1}$$
$$-\frac{m}{r4!}\bar\lambda_i{}^+\gamma\gamma\lambda_k\bar\lambda^{k+i}\partial\lambda^{k+}$$
$$+\frac{m}{r4!}\bar\lambda_i{}^+\gamma\gamma^5\lambda_k\bar\lambda^{k+}\partial\lambda^{i} + \cdots.$$

D. The contraction limit

If we take the contraction limit $r\to\infty$ in formulas from Secs. V B and V C, and if we interpret u^{ij}, v^{ij}, $\theta^i = \theta^{i+}$ as the

fields in anti-de-Sitter space with coordinates x^a, we obtain the following transformation laws:

$$u^{ij} = -\tfrac{1}{2}m[\bar\theta^i\gamma^5\alpha^j - \bar\theta^j\gamma^5\alpha^i] - \tfrac{1}{2}\bar\theta_i\gamma^a\gamma^5\alpha^i\partial_a u^{ij},$$
$$v^{ij} = -\tfrac{1}{2}m[\bar\theta^i\alpha^j - \bar\theta^j\alpha^i] - \tfrac{1}{2}\theta_i\gamma^a\gamma^5\alpha^i\partial_a v^{ij},$$
$$\delta\theta^i = \alpha^i - \tfrac{1}{2}\bar\theta_j\gamma^a\gamma^5\alpha^j\partial_a\theta^i, \quad \delta x^a = -\tfrac{1}{2}\bar\theta_i\gamma^a\gamma^5\alpha^i.$$

Similarly in the contraction limit the Cartan form became

$$\bar\omega^l = \bar\omega^l = d\bar\theta^l; \quad \omega^a = dx^a + \bar\theta_i\gamma^a\gamma^5\, d\theta^i,$$
$$\omega_y^{ij} = dv^{ij} - \tfrac{1}{4}m[\bar\theta^i\, d\theta^j - \bar\theta^j\, d\theta^i],$$
$$\omega_x^{ij} = du^{ij} + \tfrac{1}{4}m[\bar\theta^i\gamma^5\, d\theta^j - \bar\theta^j\gamma^5\, d\theta^i].$$

Identical formulas were obtained earlier by Ferrara.[6]

ACKNOWLEDGMENTS

I would like to thank Professor J. Lukierski for helpful conversations, suggestions, and criticism. I thank Dr. L. Rytel for helpful discussions.

APPENDIX A

Following Zumino[8] we define the operation $\wedge$

$$X\wedge Y = [X,Y]$$
$$X^2\wedge Y = [X,[X,Y]]\ldots,$$

where X, Y are arbitrary generators belonging to Lie algebra.

For every function $f(x)$ the expression

$$f(x)\wedge y$$

can be derived by expanding the function $f(x)$ as a power series in X. One can easily check the following formulas[8]:

$$e^x Y e^{-x} = e^x\wedge Y, \tag{A1}$$
$$e^x\delta e^{-x} = [(1-e)^x/X]\wedge\delta X, \tag{A2}$$

where δ is an arbitrary variation.

APPENDIX B

Using the notation introduced in Appendix A, Eq. (3.11) can be written in the form[8]

$$e^{-\bar\theta_i Q^i}\wedge\bar\alpha_i Q^i - \frac{1-e^{-\bar\theta_i Q^i}}{\bar\theta_i Q^i}\wedge\delta\bar\theta_i Q^i = h_1. \tag{B1}$$

Anticommutation relations for spinor generators Q^i can be written in the form of commutation relations if we use the Majorana spinor parameters θ^i. We get

$$\{\bar\theta_i Q^i, \theta'_i Q'^i\} = m\bar\theta_i\Sigma^{ABi}\theta'^i M_{AB} + m\bar\theta^i{}_1\theta'^j T_{ij}, \tag{B2}$$

where $A,B = 0,1,2,3,4$ and

$$m\bar\theta_i\Sigma^{AB}\theta'^i M_{AB} = m\bar\theta_i\gamma^a\gamma^j\theta'^i P_a + m\bar\theta_i\sigma^{ab}\theta'^i M_{ab},$$
$$M_{ab} = (1/m)P_a.$$

We have also

$$[M_{AB},\bar\theta_i Q^i] = i\bar\theta_i\Sigma_{AB}Q^i.$$

It can be written as

$$[M_{ab},\bar\theta^i Q_i] = i\bar\theta^i\sigma_{ab}Q_i,$$
$$[P_a,\bar\theta_i Q^i] = im\bar\theta_i\gamma_a\gamma_b Q^i.$$

Using these commutation relations we can find that

$$(\bar\theta_i Q^i)^{2k}\wedge\bar\alpha_k Q^k \sim Q^i,$$
$$(\bar\theta_i Q^i)^{2k+1}\wedge\bar\alpha_k Q^k \sim M_{AB} + T_{ij}\in \mathrm{SO}(3,2)\times\mathrm{O}(N).$$

On the right-hand side of Eq. (B1) we have only generators M_{ab}, P_a, and T_{ij}, so we must set equal to zero in left-hand side the part with even powers of $\bar\theta_i Q^i$, because even powers are proportional to Q^i. This condition gives

$$\delta\bar\theta_i Q^i = \bar\theta_i Q^i\coth(\bar\theta_i Q^i)\wedge\bar\theta_i Q^i$$
$$= \sum_{k=1}^{\infty}\frac{(-1)^{k+1}2^{2k}}{(2k)!}B_{2k-1}(\bar\theta_i Q^i)^{2k}\wedge\bar\alpha_i Q^i, \tag{B3}$$

where B_{2k-1} are Bernoulli numbers. Using (B2) we can also find the element $h_a\in\mathrm{SO}(3,2)\times\mathrm{O}(N)$,

$$h_1 = \frac{1-\mathrm{ch}(\bar\theta_i Q^i)}{\mathrm{sh}\,\bar\theta_i Q^i}\wedge\bar\alpha_i Q^i$$
$$= \sum_{k=1}^{\infty}\left[\frac{2(2^{2k}-1)}{(2k)!}B_{2k} - \frac{(-1)^{k+1}2^{2k}}{(2k)!}B_{2k-1}\right]$$
$$\times(\bar\theta_i Q^i)^{2k+1}\wedge\bar\theta_i Q^i. \tag{B4}$$

Because of anticommuting property of spinor components any power of $\bar\theta_i Q^i$, $i = 1\cdots N$, higher then $4N$ vanishes identically. Therefore power series (B3) and (B4) are not infinite. Using the commutation relation (B2) we find

$$(\bar\theta_i Q^i)^2\wedge\bar\alpha_i Q^i = im[3\bar\alpha^k\theta^i\bar\theta_k + 2\bar\theta_i\theta^l\bar\alpha_k + \bar\alpha^i\theta^k\bar\theta_k]Q_i,$$

generally

$$(\bar\theta_i Q^i)^{2k}\wedge\bar\alpha_i Q^i = {}^{\bar\square\, i}_{\square\, 2k}Q_i,$$

where

$${}^{\bar\square\, i}_{\square\, 2k} = (im)^k[-3^{\square\,(b)}_{\square\, 2k-2}\theta^i\theta_i + 2\bar\theta^i\theta_l{}^{\bar\square\, l}_{\square\, 2k-2} + {}^{\bar\square\, i}_{\square\, 2k-2}\bar\theta^i\theta_l].$$

We find also

$$(\theta_i Q^i)^{2k+1}\wedge\alpha_i Q^i = m\bar\theta_l\Sigma^{AB}{}^{\square\, i}_{\square\, k}M_{AB} + m\bar\theta_i{}^{\square\, j}_{\square\, k}T_{ij},$$

where

$${}^{\square\, i}_{\square\, h} = \sum_{k=1}^{\infty}\left[\frac{2(2^{2k-1}-1)}{(2k)!}B_{2k} - \frac{(-1)^{k+1}2^{2k}}{(2k)!}B_{2k-1}\right]{}^{\square\, i}_{\square\, 2k}.$$

Therefore we can write

$$\delta\bar\theta_i = \bar\alpha^i + \sum_{k=1}^{\infty}\frac{(-1)^{k+1}2^{2k}}{(2k)!}B_{2k-1}{}^{\bar\square\, i}_{\square\, 2k},$$
$$h_1 = m\bar\theta_i\Sigma^{AB}{}^{\square\, i}_{\square\, h}M_{AB} + m\bar\theta{}^{\square\, j}_{\square\, h}T_{ij}.$$

APPENDIX C

The commutation relations for generators T^{ij} of internal symmetry group $\mathrm{O}(2N)$ are given in (3.1). We define[15,24]

$$A_{ij} = \tfrac{1}{2}(T_{ij} + T_{N+iN+j}); \quad S_{ij} = \tfrac{1}{2}(T_{N+ij} - T_{iN+j});$$
$$X_{ij} = \tfrac{1}{2}(T_{ij} - T_{N+iN+j}); \quad Y_{ij} = \tfrac{1}{2}(T_{N+ij} + T_{iN+j}),$$

where $i,j = 1\cdots N$. In this notation and in rescaling given in Sec. V, the commutation relations for $\mathrm{O}(2N)$ are the following:

$$[A_{ij},Y_{kl}] = \tfrac{1}{2}(\delta_{jk}Y_{ik} + \delta_{ik}Y_{jl} - \delta_{il}Y_{jk} - \delta_{jk}Y_{il}),$$
$$[S_{ij},Y_{kl}] = \tfrac{1}{2}(\delta_{ik}X_{jl} - \delta_{jl}X_{ik} - \delta_{il}X_{jk} + \delta_{jk}X_{il}),$$

$$[S_{ij},X_{kl}]=\tfrac{1}{2}(\delta_{jl}Y_{ik}-\delta_{jk}Y_{il}-\delta_{il}Y_{jl}+\delta_{il}Y_{jk}),$$
$$[A_{ij},X_{kl}]=\tfrac{1}{2}(\delta_{ik}X_{jl}+\delta_{jl}X_{ik}-\delta_{il}X_{jk}-\delta_{jk}X_{il}),$$
$$[Y_{ij},Y_{kl}]=(1/2r^2)(\delta_{ik}A_{jl}+\delta_{jl}A_{ik}-\delta_{il}A_{jk}-\delta_{jk}A_{il}),$$
$$[S_{ij},S_{kl}]=\tfrac{1}{2}(\delta_{ik}A_{jl}+\delta_{jl}A_{ik}+\delta_{il}A_{jk}+\delta_{jk}\mathrm{A}_{il}),$$
$$[A_{ij},A_{kl}]=\tfrac{1}{2}(\delta_{ik}A_{jl}+\delta_{jl}A_{ik}-\delta_{il}A_{jk}-\delta_{jk}A_{il}),$$
$$[X_{ij},X_{kl}]=(1/2r^2)(\delta_{ik}A_{jl}+\delta_{jl}A_{ik}-\delta_{il}A_{jk}-\delta_{jk}A_{il}),$$
$$[A_{ij},S_{kl}]=\tfrac{1}{2}(-\delta_{jk}S_{ik}+\delta_{il}S_{jk}+\delta_{ik}S_{jl}-\delta_{jk}S_{il}),$$
$$[X_{ij},Y_{kl}]=(1/2r^2)(-\delta_{jl}S_{ik}+\delta_{il}S_{jk}-\delta_{ik}S_{jl}+\delta_{jk}S_{il}).$$

The generators A^{ij} and S^{ij} give us the real representation of the subalgebra $\mathrm{U}(N)$. This is the symmetric subalgebra, and therefore the coset $\mathrm{O}(2N)/\mathrm{U}(N)$ is a symmetric space. We use the following parametrization of the group $\mathrm{O}(2N)$:

$$\exp(u_{ij}X^{ij}+v_{ij}Y^{ij})\exp(a_{ij}A^{ij}+t_{ij}S^{ij}).$$

The first exponent describes the coset $\mathrm{O}(2N)/\mathrm{U}(N)$ and the second the subgroup $\mathrm{U}(N)$. The nonlinear action of $\mathrm{O}(2N)$ group on the coset $\mathrm{O}(2N)/\mathrm{U}(N)$ is the following.

If $g_0\mathrm{U}(N)$ and $g_0=1+a_{ij}A^{ij}+t_{ij}S^{ij}$, g_0 is infinitesimal then

$$\delta u^{il}=2(-t^{i}{}_{k}v^{kl}+a^{i}{}_{k}u^{kl}),$$
$$\delta v^{il}=-2(t^{i}{}_{k}u^{kl}+a^{i}{}_{k}v^{kl}).$$

On the other hand, if $g_0=1+p^{ij}X_{ij}+q^{ij}Y_{ij}$ and g_0 is infinitesimal we get

$$\delta u^{il}=p^{il}+\sum_{u=1}^{a}\frac{(-1)^{u+1}2^{2u}}{(2u)!}B_{2u-1}\hat{u}^{il}_{(2u)},$$
$$\delta v^{il}=q^{il}+\sum_{u=1}^{a}\frac{(-1)^{u+1}2^{2u}}{(2u)!}B_{2u-1}\hat{v}^{il}_{(2u)},$$

where

$$\hat{u}^{il}_{(2u)}=-2(u^{i}_{k}a^{kl}_{(2u-1)}+v^{i}_{k}t^{kl}_{(2u-1)}),$$
$$\hat{v}^{il}_{2u}=2(u^{i}_{k}t^{kl}_{(2u-1)}-v^{i}_{k}a^{kl}_{(2u-1)}),$$

and

$$a^{il}_{(2u+1)}=-\frac{2}{v^2}(u^{i}_{k}\hat{u}^{kl}_{(2u)}+v^{i}_{k}\hat{v}^{kl}_{(2u)}),$$
$$t^{il}_{(2u+1)}=\frac{2}{v^2}(u^{i}_{k}\hat{u}^{kl}_{(2u)}-v^{i}_{k}\hat{v}^{kl}_{(2u)}),$$

with $u^{il}_{(0)}=p^{il}$ and $v^{il}_{(0)}=q^{il}$.

APPENDIX D

We give here explicit forms of expressions used in Sec. IV. The expressions used in Sec. IV B are the following:

$$\begin{aligned}\overline{Y}_u(\alpha^k,\lambda^{ki})=&\{(\xi^{ik}\xi^{k}_{i})^{u}\bar{\alpha}^{k}+\tfrac{1}{2}im(\xi^{ik}\xi^{k}_{i})^{u-1}[(5u+\tfrac{1}{2}(u-1))\bar{\theta}^{k}\theta^{k}\alpha^{k}\\&+(u+\tfrac{1}{2}(u-1))\theta^{k}\Gamma_{A}\theta^{k}\bar{\alpha}^{k}\Gamma^{A}]-(-\xi^{lk}\xi^{k}_{l})^{u-2\square}_{\square}(u-2)m^2(-s(u-1)^2+\tfrac{1}{2}(u-2)^2)(\bar{\theta}^{k}\theta^{k})^2\bar{\alpha}^{k}\\&+(\xi^{ik}\lambda^{k}_{i}+i(u-1)m\bar{\theta}^{k}\theta^{k})(\xi^{lk}\xi^{k}_{l})^{u-2\square}_{\square}(u-2)\xi^{lk}\lambda^{k}_{l}\bar{\theta}^{k}\},\\ Z^{ki}_{u}(\alpha^k,\lambda^{ki})=&\{(\xi^{lk}\xi^{k}_{i})^{u-1}im\bar{\theta}^{k}\alpha^{k}\xi^{ki}-(u-1)m^2(\xi^{lk}\xi^{k})^{u-2\square}_{\square}(u-2)\\&\times\bar{\theta}^{k}\theta^{k}\alpha^{k}\xi^{ki}-im(\xi^{ik}\xi^{k}_{i})^{u-2\square}_{\square}(u-2)\bar{\theta}^{k}\theta^{k}\xi^{lk}\lambda^{k}_{i}\xi^{ki}\\&+[(-\xi^{lk}\xi^{k}_{i})^{u-1}-imu\bar{\theta}^{k}\theta^{k}(-\xi^{lk}\xi^{k}_{l})^{u-2}-(n-2)^2m^2(\bar{\theta}^{k}\theta^{k})^2(-\xi^{lk}\xi^{k}_{l})^{u-2}(u-2)]\\&\times\xi^{lk}(\lambda^{k}_{l}\xi^{ki}-\xi^{k}_{l}\lambda^{ki})-(u-1)m^2(\bar{\theta}^{k}\theta^{k})^2\xi^{lk}\lambda^{k}_{l}(\xi^{lk}\xi^{k}_{l})^{u-3\square}_{\square}(u-2)\xi^{ki}\},\\ \overline{I}^{i}_{u}(\alpha^k,\lambda^{ki})=&\{i(\xi^{lk}\xi^{k}_{l})^{u}\xi^{ik}\alpha^{k}+3um(\xi^{lk}\xi^{k}_{l})^{u-1}\xi^{ik}\bar{\theta}^{k}\theta^{k}\alpha^{k}\\&-2um^2\xi^{ik}(\xi^{lk}\xi^{k}_{l})^{u-1}\bar{\theta}^{k}\theta^{k}\bar{\alpha}^{k}+\xi^{ik}(-s+\tfrac{1}{4})im^2(u-1)^2(\xi^{ik}\xi^{k}_{l})^{u-2}\\&\times\square_{\square}(u-2)(\bar{\theta}^{k}\theta^{k})^2\alpha^{k}-i(\xi^{lk}\xi^{k}_{l})^{u-1}\xi^{lk}\lambda^{k}_{l}\xi^{ik}\bar{\theta}^{k}\\&-mu(\xi^{lk}\xi^{k}_{l})^{u-2\square}_{\square}(n-2)\xi^{lk}\lambda^{k}_{l}\xi^{ik}\bar{\theta}^{k}\theta^{k}\bar{\theta}^{k}\},\\ J^{AB}_{n}(\alpha^k,\lambda^{ki})=&\{(\xi^{lk}\xi^{k}_{l})^{u}m\bar{\theta}^{k}\Sigma^{AB}\alpha^{k}+2im^2u(\xi^{lk}\xi^{k}_{l})^{u-1}\\&\times\bar{\theta}^{k}\theta^{k}\bar{\theta}^{k}\Sigma^{AB}\alpha^{k}\},\end{aligned}$$

and

$$R^{im}_{u}(\alpha^k,\lambda^{ki})=\{[\tfrac{1}{2}(-\xi^{lk}\xi^{k}_{l})^{u}-imu\bar{\theta}^{k}\theta^{k}(-\xi^{lk}\xi^{k}_{l})^{u-1}+\tfrac{1}{2}(u-1)^2(-im\bar{\theta}^{k}\theta^{k})^2(-\xi^{lk}\xi^{k}_{l})^{u-2\square}_{\square}(u-2)](\xi^{ik}\lambda^{mk}-\xi^{mk}\lambda^{il})\}.$$

In Sec. IV C the following expressions were used:

$$\begin{aligned}&\overline{E}^{i}_{l}(u^k)Q_i+\overline{E}_{lk}(u^i)Q_k\\&=\overline{U}^{i}Q_i+\sum_{k=1}^{\infty}\frac{1}{(2u)!}[(\xi^{i}_{l}\xi_{li})^{u-1}\xi^{p}_{l}\xi^{i}_{l}\overline{U}_{p}Q_i+im(3u+1)(\xi^{i}_{l}\xi_{li})^{u-2\square}_{\square}(u-2)]\\&\times\xi^{p}_{l}\xi^{i}_{l}\overline{U}_{p}\theta^{l}\bar{\theta}^{l}Q_i+2imu(\xi^{i}_{l}\xi_{li})^{u-3\square}_{\square}(u-3)\xi^{p}_{l}\xi^{i}_{l}\bar{\theta}^{l}\theta^{l}\overline{U}_{p}Q_i\\&+m^2(\xi^{i}_{l}\xi_{li})^{u-3\square}_{\square}(u-3)\xi^{i}_{l}\xi^{k}_{l}[\tfrac{1}{2}(u-1)^2-5(u-2)^2]\bar{\theta}^{l}\theta^{l}\overline{U}_{k}Q_i\end{aligned}$$

$$+\sum_{u=0}^{\infty}\frac{1}{(2u-1)!}\Big[-i(\xi^i_l\xi_{ki})^u\xi^i_k\bar{U}_iQ_k-3um(\xi_{ki}\xi^i_k)^{u-1}$$
$$\times\xi^i_k\bar{\theta}^kU_i\bar{\theta}^k\theta^k+2um(\xi^i_k\xi_{ki})^{u-1}\xi^i_k\bar{\theta}_k\theta_k\bar{U}_l\bar{\theta}_k$$
$$+\frac{1\rho}{u}im^2(\xi^i_k\xi_{ki})^{u-2\square}_{\square}(u-2)\xi^i_k(\bar{\theta}^k\theta^k)^2\bar{U}_iQ_k\Big],$$

where $p,l,i=1\cdots k-1$,

$$F^p_l(u^{ul})Q_p=\sum_{u=1}^{\infty}\frac{1}{(2u)!}\left[-2i-(\xi^i_l\xi_{li})^{n-1}\xi_{km}U^{mp}\theta_k+2m(u-1)\bar{\theta}^k\theta^k(-\xi^i_l\xi_{li})^{u-2\square}_{\square}(u-2)\xi_{lm}U^{mp}\theta^l\right]Q_p.$$

Here also $p=1\cdots k-1$,

$$F^{mi}_l(u^{ul})T_{mi}=\sum_{u=0}^{\infty}\frac{1}{(2u+1)!}\left[-2(\xi^i_l\xi_{li})^n+2imu\bar{\theta}^l\theta^l(\xi^i_l\xi_{li})^{u-1}\right.$$
$$\left.+2m^2(\theta^l\theta^l)^2(u-1)^2(\xi^i_l\xi_{li})^{u-2\square}_{\square}(u-2)\right]\xi^m_lu^i_lT_{mi}\quad m,i=1\ldots k-1,$$
$$H^i_l(u^{AB})Q_l+H^k_l(u^{AB})Q_k=u\cdot\sum_{u=1}^{\infty}\frac{1}{(2u)!}\left[(\xi^i_l\xi_{li})^{u-1}+2im(u-1)(\xi^i_l\xi_{li})^{u-2\square}_{\square}(u-2)\bar{\theta}^l\theta^l\right]$$
$$\times\xi^i_l\bar{\theta}^k\Sigma^{AB}d\theta^k\theta^l\Sigma_{AB}Q_i+u\cdot\sum_{n=0}^{\infty}\frac{1}{(2+1)!}\left[-i(\xi^i_l\xi_{li})^n\right]$$
$$+2mu\bar{\theta}^k\theta^k(\xi^i_l\xi_{ii})^{u-1}\bar{\theta}^k\Sigma^{AB}d\theta^k\theta^k\Sigma_{AB}Q_k$$

and also

$$\bar{G}^k_k(u^k,u^{kl})=u^k+\sum_{u=1}^{\infty}\frac{1}{(2u)!}\bar{Y}_u(u^k,u^{kl}),$$
$$\bar{G}^i_k(u^k,u^{kl})=\sum_{u=0}^{\infty}\frac{1}{(2u+1)!}\bar{I}^i_u(u^k,u^{kl}),$$
$$G^{li}_k(u^k,u^{kl})=\sum_{u=1}^{\infty}\frac{1}{(2u)!}Z^{li}_u(u^k,u^{kl}),$$
$$G^{mn}_k(u^k,u^{kl})=\sum_{n=0}^{\infty}\frac{1}{(2n+1)!}R^{mn}_n(u^k,u^{kl}),\quad m,n<k$$
$$G^{AB}_k(u^k,u^{kl})=\sum_{u=0}^{\infty}\frac{1}{(2u+1)!}J^{AB}_u(u^k,u^{kl}).$$

The remaining expressions are the following:

$$E^{AB}_l(u_i)=\sum_{n=1}^{\infty}\frac{1}{(2u)!}\left[-im(\xi^i_l\xi_{li})^{u-1}+2m^2(\xi^i_l\xi_{li})^{u-2\square}_{\square}(u-2)\bar{\theta}^l\theta^l\right]\xi^i_l\theta^k_l\Sigma^{AB}u_i,$$
$$H^{AB}_l(u^{AB})=u^{AB}+u\cdot\left\{\sum_{u=1}^{\infty}\frac{1}{(2u)!}\left[2im(\xi^i_l\xi_{ii})^{u-1}+um^2(\xi^i_l\xi_{li})^{u-2\square}_{\square}(u-2)\theta^l\theta^l\right]\right.$$
$$\left.\times(d\theta^{l+1}\theta^l\theta^l\Sigma^{AB}\theta^{l+1}-\theta^{l+1}\theta^l\theta^l\Sigma^{AB}d\theta^l)\right\},\quad\text{where } i=1\ldots k-1.$$

We must define now the expressions u, u^{AB}, u^i, u^{ik}, u^k:

$$\bar{u}^k=d\theta^k+\sum_{u=1}^{\infty}\frac{1}{(2u-1)}\bar{\gamma}_u(d\theta^k,d\xi^{ik}),$$
$$u^i=\sum_{n=1}^{\infty}\frac{1}{(2n+2)!}I^i_n(d\theta^k,d\xi^{ik}),$$
$$u_k{}^uT_{kn}=d\xi_k{}^uT_{ku}+\sum_{k=1}^{\infty}\frac{1}{(2k+1)!}Z^n_{uk}(d\theta^k,d\xi^i_k)T_{uk},$$
$$u^{ul}=\sum_{m=0}\frac{1}{(2m+2)!}R^{ul}_m(d\theta^k,d\xi^{ik}),\quad n,l<k,$$
$$u^{AB}=u\cdot\theta^k\Sigma^{AB}d\theta^k=\sum_{u=0}\frac{1}{(2u+1)!}J^{AB}_u(d\theta^k,d\xi^{ik}).$$

[1]S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969).
[2]C. G. Callan, S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2247 (1969).
[3]D. V. Volkov, Sov. J. Part. Nucl. **4**, 3 (1973).
[4]D. V. Volkov, and V. P. Akulov, Phys. Lett. B **46**, 109 (1973).
[5]J. Lukierski, Czech. J. Phys. B **32**, 504 (1982).
[6]S. Ferrara, in *Quantum Theory of Particles and Fields—Birthday Volume dedicated to Jan Lopuszański*, edited by B. Jancewicz and J. Lukierski (Word Scientific, Singapore, 1983), p. 49.
[7]S. Ferrara, L. Maiani, and P. C. West, CERN preprint T. H. 3513, 1983.
[8]B. Zumino, Nucl. Phys. B **127**, 189 (1977); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[9]A. Salam and J. Stranthde, Nucl. Phys. B **76**, 477 (1974).
[10]E. A. Ivanov and A. S. Sorin, J. Phys. A **13**, 1159 (1980).
[11]B. W. Keck, J. Phys. A **8**, 1819 (1975).
[12]J. A. de Azcaraga and J. Lukierski, Chess. Quantum Grav. **2**, 683 (1985).
[13]E. A. Ivanov and A. A. Kapustnikov, Phys. Lett. B **143**, 379 (1984).
[14]S. Sammuel and J. Wess, Nucl. Phys. B **221**, 153 (1983).
[15]J. Lukierski and L. Rytel, J. Phys. A. Math. Gen **15**, 215 (1982).
[16]C. J. Isham, Nuovo Cimento A **59** (3), 356 (1969).
[17]H. Tilgner, in *Group Theory in Nonlinear Problems*, edited by A. O. Barut (Reidel, New York, 1974), p. 142.
[18]D. Z. Freedman and A. Das, Nucl. Phys. B **120**, 221 (1977).
[19]S. W. McDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[20]F. Gursey and T. D. Lee, Proc. Natl. Acad. Sci. (USA) **49**, 179 (1963).
[21]J. Lukierski, M. Mozrzymas, and L. Rytel, Mod. Phys. Lett. A **2** (4), 261 (1987).
[22]J. Lukierski, in *Proceedings of the XIX Winter School in Karpacz, February 1986* (World Scientific, Singapore, 1986).
[23]Z. Hasiewicz and J. Lukierski, Phys. Lett. B **155**, 347 (1985).
[24]J. Lopuszański, M. Wolf, Nucl. Phys. B **194**, 251 (1982).

Cayley–Klein parameters and evolution of two- and three-level systems and squeezed states

G. Dattoli and A. Torre
ENEA, Dip. TIB, U.S. Fisica Applicata, CRE Frascati, C.P. 65-00044—Frascati, Rome, Italy

(Received 1 March 1989; accepted for publication 9 August 1989)

In this paper the time behavior of quantum states ruled by Hamiltonians linear in the SU(2), SU(1,1), and SU(3) generators in terms of the Cayley–Klein parameters, originally introduced in classical mechanics is analyzed. Also pointed out is the link between the Cayley–Klein parameters and the Wei–Norman ordering functions, exploited in the context of the Schrödinger representation.

I. INTRODUCTION

Optics and quantum optics have largely benefited from group theoretic methods.[1,2] Just to quote a few examples we recall that SU(n) is now a widespread mathematical tool to treat the evolution of n-level systems interacting with intense em fields.[3]

On the other hand, the SU(1,1) group has been exploited both in quantum and classical optics to analyze, for instance, the evolution of squeezed states[4] or the propagation of Gaussian beams in selfoc fibers.[5,6] Finally, Lie algebraic methods have provided a useful mathematical framework to construct numerical codes for the design of the magnetic optics for transport channels in particle accelerators.[7] The interest that has risen up around these techniques, already exploited with astonishing success in atomic, nuclear, and particle spectroscopy,[8] is therefore fully justified.

Within the above mathematical context, specific difficulties have been encountered and solution techniques have been discovered, or rediscovered. This is indeed the case of the ordering theorems of the Wei–Norman (WN) type,[9] which play a significant role in the analysis of the time behavior of quantum states ruled by Hamiltonians linear combination of the generators of Lie groups, as SU(2), SU(1,1) or SU(3).[10] Although forgotten for many years, this technique became increasingly popular, owing to the renewed interest in the algebraic treatment of problems in optics.

The WN theorems may be viewed, from a historical perspective, as the completion of the program originally undertaken by Magnus[11] and Fer[12] of developing, whenever possible, a systematic and rigorous treatment of time-ordering problems as opposed to the perturbative Feynman–Dyson technique[13] (also see Refs. 2 and 14 for further comments).

According to the WN technique, the evolution operator relevant to a Hamiltonian linear combination of Lie group generators can be written as an ordered product of exponentials, whose arguments are the product of a time-dependent function and a generator of the group. It can be shown that the time-dependent functions appearing in the exponentials, namely, the characteristic functions of the ordering procedure, are related to the coefficients of the linear combination in the Hamiltonian by a nonlinear system of first-order differential equations.[2,9] For groups of particular physical interest the WN characteristic equations have been further elaborated and, by introducing appropriate functions, cast in the form of generalized Bloch equations.[10,15]

The newly introduced functions deserve a particular comment. In fact, according to whether one is dealing with SU(2), SU(1,1), or SU(3) Hamiltonians, they obey a set of equations identical to that of two or three coupled harmonic oscillators, respectively. It is also worth stressing that all the physical quantities are directly linked to these functions, whose meaning we will try to further clarify within the context of the present paper, where a different method to treat ordering problems is proposed, when the above quoted three groups are involved. Taking advantage from the spinorial representation of the group generators[16,17] we will show that the characteristic functions of the ordering procedure can be identified with the Cayley–Klein (CK) parameters[18] introduced in classical mechanics to treat rotation problems. As a consequence, the WN equations can be deduced from those defining the time dependence of the CK parameters. The remarkable practical advantage of the method we present in the paper lies on its simplicity and on the consequent possibility of avoiding the amazing amount of calculations implicit in other proposed methods, employing, e.g., the Cayley–Hamilton theorem.[19]

II. THE SU(2) AND SU(1,1) CASE

Let us consider, as first example, the Hamiltonian

$$\hat{H} = \omega(t)\hat{J}_3 + \Omega^*(t)\hat{J}_+ + \Omega(t)\hat{J}_- \quad (\hbar = 1)\,, \tag{2.1}$$

with the operator $\hat{J}_3$, $\hat{J}_\pm$ obeying the rules of commutation of SU(2)

$$[\hat{J}_3, \hat{J}_\pm] = \pm \hat{J}_\pm\,, \quad [\hat{J}_+, \hat{J}_-] = 2\hat{J}_3\,. \tag{2.2}$$

The Hamiltonian (2.1) may describe, e.g., the interaction of a chirped classical em field with a two-level atomic system. Introducing the spinorial representation of the $\hat{J}$ operators [the fact that we restrict ourselves to the lowest dimensionality representation of the group by no means affects the generality of our results (also see Ref. 16)],

$$\hat{J}_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \hat{J}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \hat{J}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{2.3}$$

we can rewrite the Hamiltonian (2.1) in the form of a 2×2 Hermitian matrix as

$$\hat{H} = \begin{pmatrix} \omega(t)/2 & \Omega^*(t) \\ \Omega(t) & -\omega(t)/2 \end{pmatrix}. \tag{2.4}$$

 0022-2488/90/010236-05$03.00

Denoting with

$$\Psi=\begin{pmatrix}\Psi_+\\ \Psi_-\end{pmatrix}, \tag{2.5}$$

the wave function representing the system, one immediately gets the following system of coupled equations:

$$\begin{aligned} i\dot{\Psi}_+ &= \tfrac{1}{2}\omega(t)\Psi_+ + \Omega^*(t)\Psi_-\,,\\ i\dot{\Psi}_- &= -\tfrac{1}{2}\omega(t)\Psi_- + \Omega(t)\Psi_+\,. \end{aligned} \tag{2.6}$$

The components Ψ_- and Ψ_+ can be understood from the physical point of view as the probability amplitudes for the system of being in the lowest or excited state, respectively. Since Eq. (2.6) represents a rotation of the vector Ψ in the complex space, we can write its solution using the CK matrix[18] as

$$\begin{pmatrix}\Psi_+\\ \Psi_-\end{pmatrix}=\begin{pmatrix}H^* & G\\ -G^* & H\end{pmatrix}\begin{pmatrix}\Psi_+(0)\\ \Psi_-(0)\end{pmatrix}, \tag{2.7}$$

where $H(0)=1$, $G(0)=0$, and

$$|H|^2+|G|^2=1\,. \tag{2.8}$$

Inserting (2.7) into Eq. (2.6) we immediately find that the CK parameters H and G obey the following system of first-order differential equations:

$$i\dot{H}=-(\omega/2)H+\Omega G,\quad i\dot{G}=(\omega/2)G+\Omega^*H\,. \tag{2.9}$$

The above equations resemble the ordering equations derived in Ref. 16 and, in fact, within the present framework, the CK matrix in (2.7) can be regarded as the matrix representation of the time-evolution operator $\hat{U}$ relevant to the Hamiltonian (2.4), or equivalently (2.1). In order to point out the connection with the ordering method of Ref. 9, let us write the evolution operator as the WN ordered product

$$\hat{U}(t)=e^{2h(t)\hat{J}_3}e^{g(t)\hat{J}_+}e^{-f(t)\hat{J}_-}, \tag{2.10}$$

which, using the spinorial representation (2.3), can be cast in the simple matrix form[16]

$$\hat{U}(t)=\begin{pmatrix}(1-fg)e^h & ge^h\\ -fe^{-h} & e^{-h}\end{pmatrix}. \tag{2.11}$$

As a consequence of the unitarity of $\hat{U}$ we readily obtain the relations

$$ge^h=(fe^{-h})^*,\quad (1-fg)e^h=e^{-h^*}\,. \tag{2.12}$$

Therefore, comparing (2.11) with the CK matrix (2.7), it is easy to recognize the CK parameters H and G as

$$H=e^{-h},\quad G=ge^h,\quad F=fe^{-h}=G^*\,, \tag{2.13}$$

which are the characteristic functions of the ordering procedure introduced in Ref. 10.

The above relations, together with (2.9), yield the system of differential equations specifying the WN functions (h,g,f), namely,

$$\begin{aligned} \dot{h} &= -i(\omega/2)h+i\Omega g e^{2h}\,,\\ \dot{g} &= -i(\omega/2)g-i\Omega^* e^{-2h}-\dot{h}g\,,\\ \dot{f} &= i\Omega e^{2h}\,. \end{aligned} \tag{2.14}$$

It is worth noticing that the CK parameters, rather than the WN functions, specify measurable physical quantities relevant to the system under study. In fact, it can be easily shown that the evolution of the Bloch vector as well as the wave function is entirely specified by H and G.

The evolution of the pseudospin of the system can be immediately obtained, noticing that

$$\begin{aligned} \langle\hat{J}_3\rangle &= \langle\Psi|\hat{J}_3|\Psi\rangle=\tfrac{1}{2}[|\Psi_+|^2-|\Psi_-|^2]\,,\\ \langle\hat{J}_+\rangle &= \langle\Psi|\hat{J}_+|\Psi\rangle=\Psi_+^*\Psi_-\,,\\ \langle\hat{J}_-\rangle &= \langle\Psi|\hat{J}_-|\Psi\rangle=\Psi_-^*\Psi_+\,. \end{aligned} \tag{2.15}$$

Furthermore, introducing the Bloch vector $\mathbf{s}$,

$$\mathbf{s}=\begin{pmatrix}\langle\hat{J}_1\rangle\\ \langle\hat{J}_2\rangle\\ \langle\hat{J}_3\rangle\end{pmatrix}, \tag{2.16}$$

where

$$\hat{J}_1=\tfrac{1}{2}[\hat{J}_+ +\hat{J}_-],\quad \hat{J}_2=(1/2i)[\hat{J}_+-\hat{J}_-]\,, \tag{2.17}$$

it is readily seen that it changes with time according to

$$\mathbf{s}(t)=\mathsf{R}(t,t_0)\mathbf{s}(t_0) \tag{2.18}$$

with $\mathbf{s}(t_0)$ denoting the Bloch vector at the initial time t_0. The 3×3 matrix $\mathsf{R}(t)$, which, as noticed elsewhere,[16] is a generalization of the Rabi matrix,[20] can be written in terms of the CK parameters as

$$\mathsf{R}(t)=\begin{pmatrix}\mathrm{Re}(H^{*2}-G^2) & \mathrm{Im}(H^{*2}+G^2) & -2\mathrm{Re}(H^*G)\\ -\mathrm{Im}(H^{*2}-G^2) & \mathrm{Re}(H^{*2}+G^2) & 2\mathrm{Im}(H^*G)\\ 2\mathrm{Re}(HG) & -2\mathrm{Im}(HG) & |H|^2-|G|^2\end{pmatrix}. \tag{2.19}$$

The vector (2.16) describes the evolution of the expectation values of the angular momentum-type operators J, but no information can be inferred on the relevant fluctuations. To this aim, it is convenient to introduce the tensor

$$F_{ij}=\langle\hat{J}_i\hat{J}_j\rangle-\langle\hat{J}_i\rangle\langle\hat{J}_j\rangle,\quad i,j=1,2,3, \tag{2.20}$$

which for $i=j$ yields the variance of the Bloch vector components. The time behavior of the fluctuation tensor can be easily calculated in terms of the H and G functions; the explicit expression are omitted for the sake of conciseness.

As a further support of the importance of the CK parameters, with respect to the WN functions, let us briefly discuss the evolution of the wave function.

Labeling with $|J,m\rangle$ a generic angular momentum state, we can express $\Psi(t)$ as

$$\Psi(t)=\sum_{m=-J}^{J}C_m(t)|J,m\rangle\,. \tag{2.21}$$

The time-dependent coefficients C_m depend on both the initial value of the wavefunction and the "scattering matrix" $\mathsf{S}(t,t_0)$, whose elements defined as

$$S_{m,n}=\langle J,n|\hat{U}|J,m\rangle\,, \tag{2.22}$$

are related to $C_m(t)$, according to

$$C_m(t)=\sum_{n=-J}^{J}S_{m,n}(t,t_0)C_n(t_0)\,. \tag{2.23}$$

The matrix elements $S_{m,n}$, whose explicit expression can be inferred from (2.22) and (2.10) as[15]

$$S_{m,n}(t,t_0) = \left[\binom{J+n_>}{J+n_<}\binom{J-n_<}{J-n_>}\right]^{1/2} H^{-(n+m)} \cdot [\mathrm{sgn}(m-n)|G|]^{n_>-n_<} \cdot \exp\{i\chi(m-n)\} \cdot \times {}_2F_1(-J-n_<, J-n_<+1; n_>-n_<+1; |G|^2), \tag{2.24}$$

with $\chi = \arg(G)$, $n_< = \min(m,n)$, $n_> = \max(m,n)$, and ${}_2F_1$ hypergeometric function, are fully defined by the functions H and G.

With regard to the time dependence of the CK parameters, let us stress that H and G satisfy the same second-order differential equation, easily inferred from the system (2.9) as[15]

$$\ddot{y} - \frac{\dot{\Omega}}{\Omega}\dot{y} + \left[|\Omega|^2 + \frac{\omega^2}{4} - i\frac{\dot{\omega}}{2} + i\frac{\omega}{2}\frac{\dot{\Omega}}{\Omega}\right]y = 0, \tag{2.25a}$$

with initial conditions

$$H(0) = 1, \quad \dot{H}(0) = i(\omega/2), \qquad G(0) = 0, \quad \dot{G}(0) = i\Omega(0). \tag{2.25b}$$

Equation (2.24), as it stands, cannot be solved exactly for any time dependence of ω and Ω. Bambini and Berman[21] found a class of solitary pulses that allow the solution of (2.24) in terms of hypergeometric functions. Later on the method of Ref. 21 was generalized to chirped pulses.[15,22]

In this paper, for the sake of completeness, we discuss the case of time-decaying pulses, which leads to exact solutions of (2.24) in terms of Bessel functions.

Assuming for Ω the time dependence

$$\Omega(t) = \Omega_0 e^{-\alpha t} \tag{2.26}$$

and using the Liouville transformation

$$y = e^{-(1/2)\alpha t}\delta(t), \tag{2.27}$$

we can turn (2.25a) in the following harmonic oscillator-type equation for δ:

$$\ddot{\delta} + [\Omega_0^2 e^{-2\alpha t} - \tfrac{1}{4}(\alpha + i\omega)^2]\delta = 0. \tag{2.28}$$

Redefining the variable according to

$$x = e^{-\alpha t}, \tag{2.29}$$

(2.28) reduces to a Bessel-type equation

$$x^2\frac{d^2}{dx^2}\bar{\delta} + x\frac{d}{dx}\bar{\delta} + \left[\frac{\Omega_0^2}{\alpha^2}x^2 - \frac{1}{4}\left(1 + i\frac{\omega}{\alpha}\right)^2\right]\bar{\delta} = 0. \tag{2.30}$$

Consequently, the function H can be written as

$$H = e^{-(\alpha/2)t}\left[aJ_\nu\left(\frac{\Omega_0}{\alpha}e^{-\alpha t}\right) + bY_\nu\left(\frac{\Omega_0}{\alpha}e^{-\alpha t}\right)\right], \tag{2.31}$$

the constants a and b being evaluated from the initial conditions (2.25b), so that we finally obtain

$$H = \frac{\pi}{2\alpha}\Omega_0 e^{-\alpha t}\left[q_\nu\left(\frac{\Omega_0}{\alpha}e^{-\alpha t}, \frac{\Omega_0}{\alpha}\right) + \frac{\alpha}{\Omega_0}\nu p_\nu\left(\frac{\Omega_0}{\alpha}e^{-\alpha t}, \frac{\Omega_0}{\alpha}\right)\right], \qquad \nu = \tfrac{1}{2}(1 + i(\omega/\alpha)) \tag{2.32}$$

where the functions p_ν and q_ν are defined as[23]

$$p_\nu(a,b) = J_\nu(a)Y_\nu(b) - J_\nu(b)Y_\nu(a), \qquad q_\nu(a,b) = J_\nu(a)Y'_\nu(b) - J'_\nu(b)Y_\nu(a). \tag{2.33}$$

A similar expression can be obtained for G, namely,

$$G = e^{-(\alpha/2)t}\frac{\pi\Omega_0}{2\alpha}\left[\frac{\alpha}{2\Omega_0} - i\right]p_\nu\left(\frac{\Omega_0}{\alpha}e^{-\alpha t}, \frac{\Omega_0}{\alpha}\right). \tag{2.34}$$

At this point we can find a very simple result concerning the long-time behavior of a two-level system driven by a time-decaying pulse. Assuming, e.g., that initially the only nonzero component of the Bloch vector is the third one, we can prove, according to (2.17) and (2.18), that the population inversion behaves like

$$\langle \hat{J}_3 \rangle \propto |H|^2 - |G|^2 = 1 - 2|G|^2. \tag{2.35}$$

Finally, using the expression of J_ν and Y_ν for small arguments,[23] we obtain the asymptotic behavior of $\langle \hat{J}_3 \rangle$ or equivalently of the cross section of the process, namely,

$$\langle \hat{J}_3 \rangle \sim 1 - \frac{\pi\Omega_0}{\alpha\cosh(\pi\omega/2\alpha)}\left(1 + \frac{\alpha^2}{4\Omega_0^2}\right)\cdot\left|J_\nu\left(\frac{\Omega_0}{\alpha}\right)\right|^2. \tag{2.36}$$

It is important to emphasize that the crucial parameter of the scaling (2.36) is Ω_0/α, which is the total area of the pulse integrated in time.

We will finally spend a few words to comment on the SU(1,1) case. The Hamiltonian we consider is now

$$\hat{H} = \omega(t)\hat{k}_3 + \Omega^*(t)\hat{k}_+ + \Omega(t)\hat{k}_-, \tag{2.37}$$

where $\hat{k}_3$, $\hat{k}_\pm$ are the generators of the SU(1,1) group obeying the rules of commutation,

$$[\hat{k}_3, \hat{k}_\pm] = \pm\hat{k}_\pm, \quad [\hat{k}_+, \hat{k}_-] = -2\hat{k}_3. \tag{2.38}$$

The operator (2.37) is a model Hamiltonian for parametric amplification in a nonlinear medium and furthermore, as noticed elsewhere,[4] is the most general SU(1,1) coherence preserving Hamiltonian.

The group SU(1,1) consists of the set of all two-dimensional pseudounitary unimodular matrices that leave invariant the quadratic form

$$|Z_1|^2 - |Z_2|^2. \tag{2.39}$$

To keep the discussion as close as possible to the SU(2) case, we therefore use the non-Hermitian realization of the SU(1,1) algebra,[16] i.e.,

$$\hat{k}_3 = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad \hat{k}_+ = \begin{pmatrix}0 & -1\\ 0 & 0\end{pmatrix}, \quad \hat{k}_- = \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}. \tag{2.40}$$

In this case the Hamiltonian (2.37) writes in the following 2×2 non-Hermitian form [this apparently surprising fact has been discussed in Ref. 16, where the non-Hermitian representation (2.40) is used. The operator H is "Hermitian" according to $\hat{H}\hat{M} = \hat{M}\hat{H}^+$ with M being the metric matrix $\hat{M} = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$]

$$\hat{H} = \begin{pmatrix}\frac{1}{2}\omega(t) & -\Omega^*(t)\\ \Omega(t) & -\frac{1}{2}\omega(t)\end{pmatrix}. \tag{2.41}$$

It is now easy to prove that the SU(1,1) CK matrix is

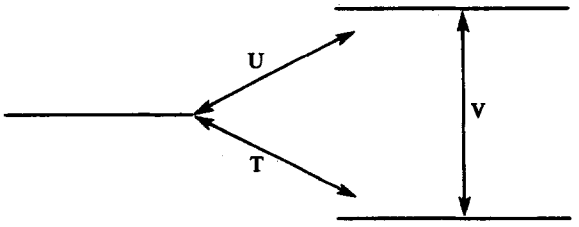


FIG. 1. The interaction of a three-level system with three intense em fields.

$$\hat{U}=\begin{pmatrix} H^* & -G \\ -G^* & H \end{pmatrix}, \tag{2.42}$$

where $H(0)=1$, $G(0)=0$, and

$$|H|^2-|G|^2=1\,. \tag{2.43}$$

Correspondingly, the time behavior of H and G is specified by

$$\begin{aligned} i\dot{H} &= -(\omega/2)H-\Omega G\,,\\ i\dot{G} &= +(\omega/2)G+\Omega^* H\,. \end{aligned} \tag{2.44}$$

The derivation of the WN equations is also straightforward along the lines previously outlined, and will not be discussed for the sake of conciseness. For further comments the reader is addressed to Ref. 16.

III. SU(3)-TYPE HAMILTONIANS

The most general Hamiltonian, allowing SU(3) coupling, is

$$\begin{aligned} \hat{H} &= \omega_T\hat{T}_3+\Omega_T^*\hat{T}_+ +\Omega_T\hat{T}_- +\omega_U\hat{U}_3+\Omega_U^*\hat{U}_+ +\Omega_U\hat{U}_-\\ &\quad +\omega_V\hat{V}_3+\Omega_V^*\hat{V}_+ +\Omega_V\hat{V}_-\,, \end{aligned} \tag{3.1}$$

where the nonsingular time-dependent functions ω's and Ω's are real and complex, respectively. From a physical point of view, the operator (3.1) describes the interaction of a three-level system with three intense electromagnetic (em) fields according to the scheme of Fig. 1. (For the case of three-level atomic or molecular systems one coupling is forbidden according to the Laporte rule. We discuss, however, the case of three nonvanishing couplings in order to treat a more general situation, which can be encountered in the treatment of three coupled harmonic oscillators.)

The Hamiltonian (3.1) exhibits an SU(2) $\otimes$ SU(2) $\otimes$ SU(2) group structure rather than SU(3), which can be immediately recovered defining the "hypercharge" operator

$$\hat{Y}=(1/\sqrt{3})(\hat{U}_3+\hat{V}_3)\,. \tag{3.2}$$

Representing the $(\hat{T},\hat{U},\hat{V})$ operators in terms of the Gell-Mann and Ne'eman matrices as[8]

$$\hat{T}_3=\frac{1}{2}\begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix},\quad \hat{T}_+=\begin{pmatrix}0&1&0\\0&0&0\\0&0&0\end{pmatrix},$$

$$\hat{T}_-=\begin{pmatrix}0&0&0\\1&0&0\\0&0&0\end{pmatrix},$$

$$\hat{U}_3=\frac{1}{2}\begin{pmatrix}1&0&0\\0&0&0\\0&0&-1\end{pmatrix},\quad \hat{U}_+=\begin{pmatrix}0&0&1\\0&0&0\\0&0&0\end{pmatrix},$$

$$\hat{U}_-=\begin{pmatrix}0&0&0\\0&0&0\\1&0&0\end{pmatrix},$$

$$\hat{V}_3=\frac{1}{2}\begin{pmatrix}0&0&0\\0&1&0\\0&0&-1\end{pmatrix},\quad \hat{V}_+=\begin{pmatrix}0&0&0\\0&0&-1\\0&0&0\end{pmatrix},$$

$$\hat{V}_-=\begin{pmatrix}0&0&0\\0&0&0\\0&-1&0\end{pmatrix}, \tag{3.3}$$

the Hamiltonian (3.1) can be written in the following 3×3 matrix form:

$$\hat{H}=\begin{pmatrix} \frac{1}{2}(\omega_U+\omega_T) & \Omega_T^* & \Omega_U^* \\ \Omega_T & -\frac{1}{2}(\omega_T-\omega_V) & -\Omega_V^* \\ \Omega_U & -\Omega_V & -\frac{1}{2}(\omega_U+\omega_V) \end{pmatrix}. \tag{3.4}$$

Now we proceed in analogy to the SU(2) case and introduce the following column vector for the system wave function:

$$\Psi=\begin{pmatrix}\langle\Psi_u\rangle\\ \langle\Psi_d\rangle\\ \langle\Psi_s\rangle\end{pmatrix}, \tag{3.5}$$

where Ψ_α represents the probability amplitude for the system of being in one of the levels of Fig. 1. From the Schrödinger equation we obtain

$$\begin{aligned} i\dot{\Psi}_u &= \tfrac{1}{2}(\omega_U+\omega_T)\Psi_u+\Omega_T^*\Psi_d+\Omega_U^*\Psi_s\,,\\ i\dot{\Psi}_d &= \Omega_T\Psi_u-\tfrac{1}{2}(\omega_T-\omega_V)\Psi_d-\Omega_V^*\Psi_s\,,\\ i\dot{\Psi}_s &= \Omega_U\Psi_u-\Omega_V\Psi_d-\tfrac{1}{2}(\omega_U+\omega_V)\Psi_s\,. \end{aligned} \tag{3.6}$$

As before, the CK matrix allows the solution to the system (3.6) in the form. (At the initial time the only nonvanishing terms in the CK matrix are the diagonal elements, all set equal to unity.)

$$\begin{pmatrix}\Psi_u\\ \Psi_d\\ \Psi_s\end{pmatrix}=\begin{pmatrix}A&B&C\\D&E&F\\G&H&I\end{pmatrix}\begin{pmatrix}\Psi_u(0)\\ \Psi_d(0)\\ \Psi_s(0)\end{pmatrix}. \tag{3.7}$$

Inserting (3.7) into Eqs. (3.6) we find the important result that the column elements in the SU(3) CK matrix (A,D,G), (B,E,H), and (C,F,I) obey the same system of first-order differential equations

$$i\frac{d}{dt}\begin{pmatrix}m\\ n\\ p\end{pmatrix}=\begin{pmatrix} \frac{1}{2}(\omega_U+\omega_V) & \Omega_T^* & \Omega_U^* \\ \Omega_T & -\frac{1}{2}(\omega_T-\omega_V) & -\Omega_V^* \\ \Omega_U & -\Omega_V & -\frac{1}{2}(\omega_U+\omega_V) \end{pmatrix}\begin{pmatrix}m\\ n\\ p\end{pmatrix}. \tag{3.8}$$

The CK parameters in (3.7) are the auxiliary functions introduced by the authors in Ref. 10 to treat SU(3) time-ordering problems and the result (3.8) was obtained after a tedious large amount of algebra.

To better clarify this point we write the evolution operator à la Wei–Norman, thus obtaining

$$\hat{U} = \hat{U}_T \hat{U}_U \hat{U}_V , \tag{3.9}$$

where each $\hat{U}$ is given by

$$\hat{U}_\alpha = e^{2h_\alpha(t)\hat{J}_3^\alpha} e^{g_\alpha(t)\hat{J}_+^\alpha} e^{-f_\alpha(t)\hat{J}_-^\alpha} . \tag{3.10}$$

Therefore, using the representation (3.3), we can express $\hat{U}$ as a 3×3 matrix with elements

$$\begin{aligned}
U_{11} &= (1+G_TF_T)(1+G_UF_U)/H_TH_U , \\
U_{12} &= (G_T/H_V)(1+G_VF_V) \\
&\quad -(G_UF_V/H_T)(1+G_TF_T) , \\
U_{13} &= (G_UH_V/H_T)(1+G_TF_T) - G_TG_V , \\
U_{21} &= (F_T/H_U)(1+G_UF_U) , \\
U_{22} &= (H_T/H_V)(1+G_VF_V) - F_TG_UF_V , \\
U_{23} &= F_TG_UH_V - H_TG_V , \\
U_{31} &= F_U , \\
U_{32} &= H_UF_V , \\
U_{33} &= H_UH_V ,
\end{aligned} \tag{3.11}$$

where

$$H_\alpha = e^{-h_\alpha}, \quad G_\alpha = g_\alpha e^{h_\alpha}, \quad F_\alpha = f_\alpha e^{-h_\alpha} . \tag{3.12}$$

Comparing (3.11) to (3.7) we immediately obtain the relation between CK parameters and WN ordering functions. In analogy to the procedure of the previous section we could deduce from (3.8) the WN ordering equations; this is straightforward but rather tedious; the interested reader is therefore addressed to Ref. 17, where those equations have been obtained within a different context.

Again, we stress that there is no need for the knowledge of the functions (3.12). Indeed, as in the SU(2) case, the CK parameters have relevance from the physical point of view. Furthermore, introducing the eight-dimensional vector s, whose components are defined according to

$$\begin{aligned}
&s_1 = \langle\Psi|(\hat{T}_+ + \hat{T}_-)/2|\Psi\rangle, \quad s_2 = \langle\Psi|(\hat{T}_+ - \hat{T}_-)/2i|\Psi\rangle , \\
&s_3 = \langle\Psi|\hat{T}_3|\Psi\rangle, \quad s_4 = \langle\Psi|(\hat{U}_+ + \hat{U}_-)/2|\Psi\rangle , \\
&s_5 = \langle\Psi|(\hat{U}_+ - \hat{U}_-)/2i|\Psi\rangle, \quad s_6 = \langle\Psi|(\hat{V}_+ + \hat{V}_-)/2|\Psi\rangle , \\
&s_7 = \langle\Psi|(\hat{V}_+ - \hat{V}_-)/2i|\Psi\rangle, \quad s_8 = \langle\Psi|(\hat{U}_3 + \hat{V}_3)/\sqrt{3}|\Psi\rangle ,
\end{aligned} \tag{3.13}$$

we can write the time evolution of the generalized Bloch vector s in the form

$$s_\alpha(t) = R_{\alpha,\beta}(t)s_\beta(o) , \tag{3.14}$$

where $R_{\alpha\beta}$ are the elements of an 8×8 matrix, which is the direct generalization of (2.18). Needless to say, the matrix $\hat{R}$ can be expressed in terms of the CK parameters and its explicit expression is reported in Ref. 17.

It is also important to emphasize that a number of conservation laws follow directly from the unitarity of the CK matrix and from Eq. (3.8), namely,

$$\begin{aligned}
&|A|^2 + |B|^2 + |C|^2 = 1, \quad |G|^2 + |H|^2 + |I|^2 = 1 , \\
&|A|^2 + |D|^2 + |G|^2 = 1, \quad |C|^2 + |F|^2 + |I|^2 = 1 , \\
&|D|^2 + |E|^2 + |F|^2 = 1, \quad |B|^2 + |E|^2 + |H|^2 = 1 .
\end{aligned} \tag{3.15}$$

The above laws of conservation are in some sense "intrinsic," since they are contained implicitly in the ordering procedure. They acquire physical meaning in the context of the particular problem under study. As a final comment, we notice that the procedure we have developed is independent of the particular SU(3) representation one chooses. The Gell-Mann and Ne'eman representation is only one of the possibilities, but, e.g., the Morris realization[24] could be straightforwardly used as well.

[1] J. Sanchez-Mandragon and K. B. Wolf, *Lie Methods in Optics* (Springer, Berlin, 1986).
[2] G. Dattoli, A. Torre, and J. C. Gallardo, "An Algebraic View to the Operatorial Ordering and Applications to Optics," Nuovo Cimento **11** (1988).
[3] J. N. Elgin, Phys. Lett. A **80**, 140 (1980); F. T. Hioe and J. H. Eberly, Phys. Rev. Lett. **47**, 838 (1981).
[4] C. C. Gerry, Phys. Rev. A **31**, 1271 (1985).
[5] G. Dattoli, S. Solimeno, and A. Torre, Phys. Rev. A **35**, 1668 (1987).
[6] See, e.g., V. I. Mankó in Ref. 1, p. 193.
[7] A. J. Dragt, *Lectures in Nonlinear Orbit Dynamics*, American Institute of Physics, Conference Proceedings (A.I.P., New York, 1982), Vol. 87.
[8] M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1967).
[9] J. Wei and E. Norman, J. Math. Phys. **4**, 575 (1963).
[10] G. Dattoli, P. Di Lazzaro, and A. Torre, Phys. Rev. A **35**, 1582 (1987).
[11] W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[12] M. Fer, Bull. Class. Sci. Acad. R. Belgium **44**, 818 (1958).
[13] R. P. Feynman, *Quantum Electrodynamics* (Benjamin, New York, 1961); F. J. Dyson, Phys. Rev. **75**, 486, 1736 (1949).
[14] R. M. Wilcox, J. Math. Phys. **4**, 962 (1967).
[15] G. Dattoli and A. Torre, J. Math. Phys. **28**, 619 (1987).
[16] G. Dattoli, M. Richetta, and A. Torre, J. Math. Phys. **29**, 2586 (1988).
[17] G. Dattoli and A. Torre, submitted to J. Opt. Soc. Am. B.
[18] H. Goldstein, *Classical Mechanics* (Addison, Reading, MA, 1959).
[19] J. H. Wilkinson, *The Algebraic Eigenvalues Problem* (Claredon, Oxford, 1965).
[20] L. Allen and J. H. Eberly, *Optical Resonance and Two-level Atoms* (Wiley, New York, 1975).
[21] A. Bambini and P. R. Berman, Phys. Rev. A **23**, 2496 (1981).
[22] F. T. Hioe and C. E. Carroll, Phys. Rev. A **32**, 1541 (1985).
[23] M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[24] R. J. Morris, Phys. Rev. **133**, A740 (1964).

Two explicit solutions in a reducible relativistic system

P. Carbonaro
Dipartimento di Matematica dell'Università di Catania, via A. Doria 6, I-95125 Catania, Italy and Istituto di Fisica dell'Università di Palermo, via Archirafi 36, I-90123 Palermo, Italy

(Received 15 March 1989; accepted for publication 9 August 1989)

By reconsidering the linear equation which describes in the hodograph plane the motion of a relativistic fluid, a significant difference with respect to the analogous equation obtained in the classical fluid dynamic theory is found; while the latter satisfies a condition which greatly simplifies the determination of the Riemann function (required for the integration), the former does not fulfill in general a similar condition except when adopting specific pressure laws. Implications and properties of the relativistic system in these extreme cases are discussed.

I. INTRODUCTION

The hodograph transformation in combination with the Riemann method of integration of a linear hyperbolic equation has been used to obtain explicit solutions to the equations describing the motion of a classical perfect fluid as well as to many other systems of physical interest.

This technique, in the form which we refer to, has been developed, on the basis of previous works, by Jeffrey,[1] who has shown that it can be applied to all 2×2 "reducible" quasilinear hyperbolic systems, i.e., to systems which in general have the form

$$U_t + AU_x = 0, \tag{1}$$

where U is a two-component vector, and A is a 2×2 matrix.

The term "reducible" means that the independent variables x and t do not appear in the entries of the matrix A, so we have simply $A = A(U)$. The system (1) is supposed to be totally hyperbolic, i.e., the matrix A possesses a pair of distinct real eigenvalues μ and ν, and in correspondence with each eigenvalue, left and right eigenvectors that will be indicated by $l^{(\lambda)}$ and $r^{(\lambda)}$ (for $\lambda = \mu,\nu$ and $\mu > \nu$). The eigenvectors $l^{(\mu)}, l^{(\nu)}$ and $r^{(\mu)}, r^{(\nu)}$ are linearly independent and satisfy the equations

$$l^{(\mu)}A = \mu l^{(\mu)}, \quad l^{(\nu)}A = \nu l^{(\nu)},$$

$$Ar^{(\mu)} = \mu r^{(\mu)}, \quad Ar^{(\nu)} = \nu r^{(\nu)},$$

together with the orthogonality conditions

$$l^{(\mu)}r^{(\nu)} = l^{(\nu)}r^{(\mu)} = 0, \quad l^{(\mu)}r^{(\mu)} \neq 0, \quad l^{(\nu)}r^{(\nu)} \neq 0.$$

Following Refs. 1 and 2 we left-multiply the system (1) by $l^{(\mu)}$ and $l^{(\nu)}$, and then we obtain

$$l^{(\mu)}(U_t + \mu U_x) = 0 \quad \text{and} \quad l^{(\nu)}(U_t + \nu U_x) = 0, \tag{2}$$

which may be written as

$$l^{(\mu)}\frac{dU}{d\sigma} = 0 \quad \text{and} \quad l^{(\nu)}\frac{dU}{d\tau} = 0, \tag{3}$$

where $d/d\sigma$ and $d/d\tau$ denote the derivatives along the characteristics $dx/dt = \mu$ and $dx/dt = \nu$.

The expressions for the Riemann invariants are then

$$\int q^{(\mu)}l^{(\mu)}\,dU = r(\tau) \quad \text{and} \quad \int q^{(\nu)}l^{(\nu)}\,dU = s(\sigma), \tag{4}$$

where $q^{(\mu)}$ and $q^{(\nu)}$ are two integrant factors. From (4) it follows that

$$l^{(\mu)} = \frac{1}{q^{(\mu)}}\nabla_U r \quad \text{and} \quad l^{(\nu)} = \frac{1}{q^{(\nu)}}\nabla_U s,$$

$$\left(\nabla_U = \frac{\partial}{\partial U_1}, \frac{\partial}{\partial U_2}\right),$$

then, Eqs. (2) become

$$\nabla_U r(U_t + \mu U_x) = 0 \quad \text{and} \quad \nabla_U s(U_t + \nu U_x) = 0,$$

which give, by the chain rule,

$$r_t + \mu r_x = 0 \quad \text{and} \quad s_t + \nu s_x = 0. \tag{5}$$

Now if it is possible, under the usual conditions, to interchange the roles of dependent and independent variables, we can get

$$x = x(r,s) \quad \text{and} \quad t = t(r,s),$$

with x and t satisfying the two linear equations

$$x_r = \nu t_r \tag{6a}$$

and

$$x_s = \mu t_s. \tag{6b}$$

By differentiating (6a) with respect to s and (6b) with respect to r, and equating the cross derivatives, one gets the following second-order equation for t:

$$t_{rs} - [\nu_s/(\mu-\nu)]t_r + [\mu_r/(\mu-\nu)]t_s = 0. \tag{7}$$

This equation can be solved, at least in principle, by standard methods.

Up to this point, we have given an account of the technique expounded in Refs. 1 and 2. What we are going to do now is apply the Jeffrey method to the system of equations describing the motion of a perfect fluid in special relativity. We will firstly make such a system fit the general form (1), whereupon we shall specialize (7) in the case of a relativistic fluid. We shall show finally that the integration of (7) in such a case presents significant differences with respect to the nonrelativistic fluid.

II. THE RELATIVISTIC FLUID

A perfect relativistic fluid is described by the energy-momentum tensor[3]

$$T^{ab} = (w+p)u^a u^b + pg^{ab},$$

 0022-2488/90/010241-04$03.00

where w and p are, respectively, the total energy density and the pressure measured in a frame in which the fluid is at rest, u^a is the four-velocity ($u^a u_a = -1$), and g^{ab} is the metric tensor.

The conservation laws for the matter and for the energy-momentum are written

$$\nabla_a(\rho u^a) = 0 \tag{8a}$$

and

$$\nabla_a T^{ab} = 0, \tag{8b}$$

where ρ is the proper matter density. We shall confine our interest in what follows to the one-dimensional motion in a Minkowski space. Then we shall introduce the inertial coordinates (x,y,z,t) and the relative velocity v so that

$$u^a \equiv \gamma(v)\{c,v,0,0\}, \quad \gamma(u) = (1-u^2/c^2)^{-1/2},$$

while ∇_a denotes the usual partial derivative. Furthermore we assume that the flow is isentropic, and then (8b) reduces to the sole equation of conservation of momentum and the Gibbs relation writes $de = -p(\rho)\, d(1/\rho)$, with e the density of internal energy.

With these hypotheses the system (8) in matricial form is

$$\mathscr{A}^0 U_t + \mathscr{A}^1 U_x = 0,$$

where

$$\mathscr{A}^0 = \begin{pmatrix} \gamma^{-2}(v) & \rho v c^{-2} \\ (c\gamma(v))^{-2} v p'_\rho & \rho f \end{pmatrix}, \qquad \mathscr{A}^1 = \begin{pmatrix} v\gamma^{-2}(v) & \rho \\ \gamma^{-2}(v) p'_\rho & \rho f v \end{pmatrix}. \tag{9}$$

Note that $U^T = (\rho, v)$, $f = 1 + i/c^2$ is the so-called "index" of the fluid,[4] and i is the specific enthalpy.

If we multiply (9) by the matrix $(\mathscr{A}^0)^{-1}$ we get a system in the form (1) with $A = (\mathscr{A}^0)^{-1}\mathscr{A}^1$ given by

$$A = \begin{pmatrix} \dfrac{v}{\gamma^2(a)(1-a^2v^2/c^4)} & \dfrac{\rho}{(1-a^2v^2/c^4)} \\ \dfrac{a^2/\rho}{\gamma^4(v)(1-a^2v^2/c^4)} & \dfrac{v}{\gamma^2(a)(1-a^2v^2/c^4)} \end{pmatrix}, \tag{10}$$

where $a^2 = p'_\rho/f$. The zeros of the determinant $\det(A - \lambda I)$ are

$$\lambda_1 = \mu = \frac{v+a}{1+av/c^2} \quad \text{and} \quad \lambda_2 = \nu = \frac{v-a}{1-av/c^2}, \tag{11}$$

μ and ν are the expressions for the Doppler-shifted velocities of sound in a fluid moving with velocity v, while $a(\rho) = c(p'_w)^{1/2} = (p'_\rho/f)^{1/2}$ is the relativistic sound speed in a reference frame in which the fluid is at rest.[4]

The set of left and right orthonormal eigenvectors spanning the space (ρ,v) are then

$$l^{(\mu)} = (a/\rho, \gamma^2(v)), \quad l^{(\nu)} = (-a/\rho, \gamma^2(v)), \tag{12}$$

$$r^{(\mu)} = \frac{1}{2}\begin{pmatrix} \rho/a \\ \gamma^{-2}(v) \end{pmatrix}, \quad r^{(\nu)} = \frac{1}{2}\begin{pmatrix} -\rho/a \\ \gamma^{-2}(v) \end{pmatrix}. \tag{13}$$

One easily recognizes in (12) and (13) the structure relevant to the system in which the Riemann invariants are additively separable.[2] In fact, if we take in (4) $q^{(\mu)} = q^{(\nu)} = 1$, the expressions for the Riemann invariants are

$$r = c\,\mathrm{arctgh}(v/c) + \int \frac{a(\rho)}{\rho}\, d\rho, \qquad s = c\,\mathrm{arctgh}(v/c) - \int \frac{a(\rho)}{\rho}\, d\rho, \tag{14}$$

which coincide with those deduced by A. H. Taub.[3] The form of $a(\rho)$ depends on a suitable choice of the state equation. Once the function $p = p(\rho)$ has been given, the problem reduces to integrate Eq. (7). In several applications of Jeffrey's method[2,5,6] it has been shown that the determination of the Riemann function is simplified if

$$\delta h = [\mu_r/(\mu-\nu)]\,\mathrm{dr} + [\nu_s/(\nu-\mu)]\,\mathrm{ds} \tag{15}$$

is an exact differential; in such a case one has

$$\frac{\partial}{\partial s}\left(\frac{\mu_r}{\mu-\nu}\right) = \frac{\partial}{\partial r}\left(\frac{\nu_s}{\nu-\mu}\right). \tag{16}$$

In the above cited works[5,6] the fact has been pointed out that many systems of physical interest satisfy this condition. In particular, in the case of the classical gas dynamics, the relation (16) holds without restrictions on the form of the equation of state $p = p(\rho)$. So as a first approach, we wish to know if this circumstance is still valid in our case.

III. THE CONDITION FOR δh BEING AN EXACT DIFFERENTIAL

It is convenient for our purposes to write the relation (16) in the form

$$R = \nu_{rs} + \mu_{rs} + (\mu_r\mu_s - \nu_r\nu_s)/(\nu-\mu) = 0. \tag{17}$$

By setting

$$\alpha(\rho) = \gamma^2(a)\,\frac{\rho}{a}\,a'_\rho, \tag{18}$$

simple calculations give

$$R = (1/c^2)(\mu^2-\nu^2)\{(\rho/a)\alpha'_\rho + (1/c^2)(1-\alpha^2)(\mu-\nu)\}. \tag{19}$$

This expression vanishes when $\alpha = \pm 1$, i.e., if the function $a(\rho)$ satisfies the following equation:

$$\gamma^2(a)\frac{\rho}{a}a'_\rho \mp 1 = 0. \tag{20}$$

Thus at variance with the nonrelativistic case (when $c \to \infty$, $R \to 0$), here the condition for δh being an exact differential is not in general verified, except when the pressure law is such that (20) holds. In order to see what this condition implies, we calculate the gradient of the eigenvalues μ and ν, namely

$$\nabla_U \mu = \gamma^{-2}(\mu)(\gamma^2(a)a'_\rho, \ \gamma^2(v)), \tag{21}$$

$$\nabla_U \nu = \gamma^{-2}(\nu)(-\gamma^2(a)a'_\rho, \ \gamma^2(v)). \tag{22}$$

By combining these results with (13) to form the products $(\nabla_U\lambda)\cdot r^{(k)}$ $(k,\lambda = \mu,\nu)$, one obtains

$$(\nabla_U\mu)\cdot r^{(\mu)} = \tfrac{1}{2}\mu'_r = ([1+\alpha(\rho)]/2)\gamma^{-2}(\mu), \tag{23}$$

$$(\nabla_U\nu)\cdot r^{(\nu)} = \tfrac{1}{2}\nu'_s = ([1+\alpha(\rho)]/2)\gamma^{-2}(\nu), \tag{24}$$

$$(\nabla_U \mu)\cdot r^{(\nu)} = \tfrac{1}{2}\mu_s' = (\{1-\alpha(\rho)\}/2)\gamma^{-2}(\mu), \quad (25)$$

$$(\nabla_U \nu)\cdot r^{(\mu)} = \tfrac{1}{2}\nu_r' = (\{1-\alpha(\rho)\}/2)\gamma^{-2}(\nu). \quad (26)$$

The condition $\alpha = -1$ yields $(\nabla_U \mu)\cdot r^{(\mu)} = (\nabla_U \nu)\cdot r^{(\nu)} = 0$ and corresponds to a system which after P. Lax is said to be completely "exceptional." As pointed out in Ref. 2, in this case, because of the orthogonality property of vectors (12) and (13), one has $l^{(\nu)}\alpha\nabla_U\mu$ and $l^{(\mu)}\alpha\nabla_U\nu$. From (25) and (26) one also gets $\gamma^2(\mu)(\nabla_U\mu)\cdot r^{(\nu)} = \gamma^2(\nu)(\nabla_U\nu)\cdot r^{(\mu)} = 1$, which precisely yield

$$l^{(\nu)} = \gamma^2(\mu)\nabla_U\mu \quad \text{and} \quad l^{(\mu)} = \gamma^2(\nu)\nabla_U\nu.$$

By using these relations in system (2) one obtains two equations in diagonal form[2]:

$$\nu_t + \mu\nu_x = 0 \quad \text{along the characteristics} \quad \frac{dx}{dt} = \mu,$$

$$\mu_t + \nu\mu_x = 0 \quad \text{along the characteristics} \quad \frac{dx}{dt} = \nu.$$

As far as the Riemann invariants are concerned, for $\alpha = -1$ they take the simple form

$$r = \int \gamma^2(\nu)\nabla_U\nu\, dU = c\,\mathrm{arctgh}(\nu/c),$$
$$s = \int \gamma^2(\mu)\nabla_U\mu\, dU = c\,\mathrm{arctgh}(\mu/c) \quad (27)$$

and, by putting $\theta(u) = c\,\mathrm{arctgh}(u/c)$, we get the relations

$$r = \theta(\nu) = \theta(v) - \theta(a), \quad s = \theta(\mu) = \theta(v) + \theta(a),$$

i.e., r and s are nothing more than the hyperbolic representation of the sound speeds ν and μ.

The hodograph system writes

$$x_\mu = \mu t_\mu \quad \text{and} \quad x_\nu = \nu t_\nu,$$

by eliminating x one sees that for any exceptional system the equation for t has the canonical form[2]

$$t_{\mu\nu} = 0. \quad (28)$$

When $\alpha = 1$ one has $(\nabla_U\mu)\cdot r^{(\nu)} = (\nabla_U\nu)\cdot r^{(\mu)} = 0$ and $\gamma^2(\mu)(\nabla_U\mu)\cdot r^{(\mu)} = \gamma^2(\nu)(\nabla_U\nu)\cdot r^{\nu} = 1$, which implies

$$l^{(\mu)} = \gamma^2(\mu)\nabla_U\mu \quad \text{and} \quad l^{(\nu)} = \gamma^2(\nu)\nabla_U\nu,$$

with these eigenvectors the system (2) splits into two inviscid Burgers equations:

$$\mu_t + \mu\mu_x = 0 \quad \text{along the characteristics} \quad \frac{dx}{dt} = \mu,$$
$$\nu_t + \nu\nu_x = 0 \quad \text{along the characteristics} \quad \frac{dx}{dt} = \nu, \quad (29)$$

while r and s are given by

$$r = \int \gamma^2(\mu)\nabla_U\mu\, dU = c\,\mathrm{arctgh}\left(\frac{\mu}{c}\right),$$
$$s = \int \gamma^2(\nu)\nabla_U\nu\, dU = c\,\mathrm{arctgh}\left(\frac{\nu}{c}\right), \quad (30)$$

with

$$r = \theta(\mu) = \theta(v) + \theta(a), \quad s = \theta(\nu) = \theta(v) - \theta(a).$$

Once more the Riemann invariants are equal to the angles that in the complex plane represent the two eigenvalues, but in an interchanged way with respect to (27). We wish to mention that Eqs. (29) are known in classical context as the "Staniukovich system" and have been used in the theory of detonation.[7]

When $\alpha = 1$ the hodograph system takes the form

$$x_\mu = \nu t_\mu \quad \text{and} \quad x_\nu = \mu t_\nu.$$

The condition for the compatibility of these equations leads to a degenerate form of the Euler–Poisson–Darboux equation

$$t_\nu - t_\mu + (\mu - \nu)t_{\mu\nu} = 0,$$

which can be written as

$$X_{\mu\nu} = 0 \quad (31)$$

with $X = (\mu - \nu)t$.

IV. DISCUSSION ON THE PRESSURE LAWS CORRESPONDING TO $\alpha = \pm 1$ AND CONCLUSIONS

Bearing in mind that $a = (p_\rho'/f)^{1/2}$, a simple integration of Eq. (20) gives the pressure law $p = p(\rho)$. This has already been discussed in both cases $\alpha = \pm 1$ in Ref. 8. In relativity an important role is played by the barotropic fluids which are characterized by an equation of state relating the pressure to the total energy density $w = \rho(c^2 + e)$. It is a trivial matter to write Eqs. (20) in terms of w instead of ρ, so we have

$$(w+p)p_{ww}'' + 2p_w'(1 - p_w') = 0 \quad \text{for } \alpha = -1 \quad (32)$$

and

$$(w+p)p_{ww}'' - 2p_w'(1 - p_w') = 0 \quad \text{for } \alpha = 1, \quad (33)$$

whose solutions, in order to be consistent with the relativistic causality, must satisfy the inequality

$$p_w' \le 1, \quad (34)$$

while the further condition

$$p_{ww}'' > 0 \quad (35)$$

is required for the shock to be compressive. The latter inequality can be however violated in several nuclear matter fluids.[9]

Equation (32) is well known in the theory of relativistic discontinuity waves[9]; its general solution reads[10]

$$p = B - A^2/(w + B/c^2) \quad (A \text{ and } B \text{ constants}),$$

which does not satisfy (35). The only solution of (32) compatible with (34) and (35) is the pressure law characterizing the incompressible relativistic fluid, i.e.,

$$p = w + \text{const}.$$

Equation (33) admits the parametric solution [satisfying both (34) and (35)]:

$$p(x) = k(\sinh x - x), \quad w(x) = k(\sinh x + x),$$

and the trivial one, too:

$$p = w + \text{const}.$$

So when the fluid is incompressible ($p = w + \text{const}$) one has $\mu_r = \nu_r = \mu_s = \nu_s = 0$, then μ and ν do not depend either on r or s. In this case in fact $a = c, \mu = c$, and $\nu = -c$, and (5) reduces to two linear and uncoupled equations,

$$r_t + cr_x = 0 \quad \text{and} \quad s_t - cs_x = 0.$$

Now we wish to correct an error which makes some formulae of this paper inconsistent with those of my previous one.[8] The correct forms of Eqs. (1b) and (11) in Ref. 8 are

$$a_t \pm \alpha v_t + (v \pm a)(a_x \pm \alpha v_x) = 0,$$
$$Aa_t \pm v_t + [(v \pm a)/(1 \pm av/c^2)](Aa_x \pm v_x) = 0,$$

i.e., the sign behind the fraction is $+$ (in Ref. 8 it erroneously is $\pm$). This leads to some equations which are incorrect, as far as a sign is concerned, but does not affect in any way the substance of the paper.

In conclusion, the main result of the present work is to point out a significant difference between the classical and the relativistic gas dynamics in Minkowski space, as far as the integrability of the equation in the hodograph plane is concerned; the property of δh in (15) being an exact differential is not in general shared by the relativistic fluid. It is recovered only for fluids obeying certain pressure laws. In these cases, the equation in the hodograph plane assumes the canonical form of the linear homogeneous wave equation, so the problem is integrable in an elementary way. This circumstance might be exploited to test numerical codes.

ACKNOWLEDGMENTS

This work was written during a stay at the Courant Institute of Mathematical Sciences (New York). The visit was made possible by a grant from the Consiglio Nazionale delle Ricerche.

[1] A. Jeffrey, "Quasilinear hyperbolic systems and waves," *Research Notes in Mathematics*, Vol. 5 (Pitman, London, 1976).

[2] A. Jeffrey and A. Donato, "The occurrence of singularities in solutions of homogeneous systems of two first order quasilinear hyperbolic equations with smooth initial data," Wave Motion **1**, 11 (1979).

[3] A. H. Taub, "Relativistic fluid mechanics," Ann. Rev. Fluid Mech. **10**, 301 (1978).

[4] A. Lichnerowicz, "Shock waves in relativistic magnetohydrodynamics under general assumptions," J. Math. Phys. **17**, 2135 (1976).

[5] A. Donato and D. Fusco, "Some applications of the Riemann method to electromagnetic wave propagation in nonlinear media," Z.A.M.M. **60**, 539 (1980).

[6] D. Fusco, A. Donato, and W. F. Ames, "Group analysis for a linear hyperbolic equation arising from a quasilinear reducible system," Wave Motion **6**, 517 (1984).

[7] I. B. Zel'dovich and A. S. Kompaneets, *Theory of Detonation* (Academic, New York, 1960).

[8] P. Carbonaro, "Exceptional relativistic gasdynamics," Phys. Lett. A **129**, 372 (1988).

[9] A. M. Anile, *Relativistic Fluids and Magnetofluids* (Cambridge U.P., Cambridge, 1989).

[10] G. Boillat, "Covariant disturbances and exceptional waves," J. Math. Phys. **14**, 973 (1973).

The discrete Boltzmann equation with multiple collisions: Global existence and stability for the initial value problem

N. Bellomo
Department of Mathematics, Politecnico, Torino, Italy

S. Kawashima
Department of Applied Science, Faculty of Engineering 36, Kyushu University, Fukuoka 812, Japan

(Received 26 January 1989; accepted for publication 23 August 1989)

This paper deals with the analysis of the initial value problem in all space $\mathbf{R}^d$, $d \geqslant 2$, for the discrete Boltzmann equation with multiple collisions. After a preliminary analysis of the mathematical modeling of the evolution equations, global existence and asymptotic behavior are proven using suitable stability criteria. Some applications verify the analysis for specific models.

I. INTRODUCTION

The discrete Boltzmann equation is a mathematical model in the discrete kinetic theory of gases,[1] which defines the time-space evolution of a system of gas particles with a finite number of velocities $\mathbf{v}_i$, $i = 1,...,n$.

As known,[1] this model has the structure of a system of semilinear partial differential equations of hyperbolic type that defines the evolution of the number densities $N_i = N_i(t,\mathbf{x})$, where t is time and $\mathbf{x}$ is space, joined to the velocities $\mathbf{v}_i$.

Classically, the study of the initial value problem in all space essentially consists of the analysis of the existence, uniqueness, and asymptotic behavior of the solutions to the discrete Boltzmann equation with given initial conditions $N_{i0} = N_i(0,\mathbf{x})$ in all space.

It is well known, after the results of Tartar,[2] Beale,[3] Bony,[4] Cabannes and Kawashima,[5] and Toscani,[6] that the initial value problem in one space dimension has always a global solution for bounded initial conditions. Then one may derive information on the asymptotic behavior of the solutions under suitable assumptions on the initial data.

On the other hand, global existence and uniqueness in more than one space dimension can be proven only under suitable smallness assumptions of the initial data. In other words, the distance between the initial conditions and either the zero solution or the Maxwell solution must usually be small.

Several papers provide mathematical results on the initial value problem in discrete kinetic theory for initial data close to vacuum in more than one space dimension, whereas the mathematical theory for initial conditions near equilibrium has been essentially developed through other papers.[7,8]

In addition to the papers that have already been cited, the reader is addressed to the review paper by Platkowski and Illner[9] for the mathematical aspects of discrete kinetic theory and to Ref. 10 for the more general theory of the initial value problem referred to as the full Boltzmann equation.

All papers that have been cited above essentially deal with mathematical models derived on the basis of simple binary collisions. On the other hand, very little is known about the analogous results for models that may include multiple collisions. These collisions, in particular, may hopefully provide a more accurate description of nondiluted gases.

Therefore, it seems interesting that the development of a mathematical theory for the analysis of the initial value problem for the discrete Boltzmann equation with multiple collisions is suitable to show, in particular, the influence of multiple collisions upon existence and stability results.

It will be shown, indeed, that models with multiple collisions have better, in a sense to be specified afterward, stability properties than the corresponding models with binary collisions only, referring to the Cauchy problem with initial conditions near equilibrium.

More details on the content of this paper are as follows: Sec. II provides the general framework for the mathematical modeling of kinetic equations with multiple collisions as well as the general definitions of the equilibrium Maxwellian state. Section III deals with the analysis of the initial value problem for the discrete Boltzmann equation with multiple collisions: A global existence and stability result is proven for an initial data close to equilibrium. Finally, Sec. IV contains some applications: Two discrete velocity models are proposed according to the analysis developed in Sec. II and their stability properties are verified on the basis of the analysis developed in Sec. III.

II. THE DISCRETE BOLTZMANN EQUATION WITH MULTIPLE COLLISIONS

As already mentioned, discrete velocity models in kinetic theory define the time-space evolution of the number densities $N_i = N_i(t,\mathbf{x}):[0,T]\times\mathbf{R}^d \to \mathbf{R}_+$, $d \geqslant 2$, of a gas particle system with a finite number of velocities $\mathbf{v}_i$, $i \in I = \{i,...,n\}$.

If one considers simple binary collisions between particles with velocities $(\mathbf{v}_i,\mathbf{v}_j)\leftrightarrow(\mathbf{v}_h,\mathbf{v}_k)$, the discrete kinetic theory[1] leads, in the case of simple collision preserving momentum and energy, to equations of the type

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_i\cdot\nabla_{\mathbf{x}}\right)N_i = \sum_{j=1}^{n}\sum_{(hk)} (A_{hk}^{ij}N_hN_k - A_{ij}^{hk}N_iN_j), \tag{2.1}$$

where the transition rates A_{ij}^{hk} are non-negative constants joined to the transition probability densities a_{ij}^{hk} in a fashion that

 0022-2488/90/010245-09$03.00

$$A_{ij}^{hk} = S\,|v_j - v_i| a_{ij}^{hk}, \quad \sum_{(h,k)} a_{ij}^{hk} = 1, \tag{2.2}$$

where S is the cross-sectional area of the gas particles.

Classically, the transition probability densities a_{ij}^{hk} satisfy the indistinguishability and reversibility properties

$$a_{ij}^{hk} = a_{ij}^{kh} = a_{ji}^{hk} = a_{ji}^{kh}, \tag{2.3a}$$

$$a_{ij}^{hk} = a_{hk}^{ij}. \tag{2.3b}$$

Moreover, the same properties are fulfilled, owing to Eq. (2.2), by the transition rates.

Using these properties, Eq. (2.1) can also be written in the following form:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_i \cdot \nabla_{\mathbf{x}}\right) N_i = \tfrac{1}{2} \sum_{jhk} A_{ij}^{hk} (N_h N_k - N_i N_j). \tag{2.1'}$$

Consider now triple collisions between triplets of particles with velocities $(\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_l) \leftrightarrow (\mathbf{v}_g, \mathbf{v}_h, \mathbf{v}_k)$, and analogously for higher-order multiple collisions. Then, the evolution equation takes the form

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_i \cdot \nabla_{\mathbf{x}}\right) N_i = J_i[N] = \sum_{p=2}^{P} J_i^{(p)}[N], \tag{2.4}$$

where the term $J_i^{(2)}$ is the one already shown in Eq. (2.1) and the additional terms have to be computed on the basis of pertinent physical assumptions. In particular, the mathematical theory should provide models which indicate that for a rarefied gas, multiple collisions are less probable than binary collisions. The said probability may become of the same order only for a dense gas.

Keeping this in mind and to avoid being too formal, let us limit our attention to triple collisions. Then, also taking into account the analysis of Refs. 11 and 12, the following assumptions are here proposed.

(*i*) Collisions preserve momentum and energy;

$$\mathbf{v}_i + \mathbf{v}_j + \mathbf{v}_l = \mathbf{v}_g + \mathbf{v}_h + \mathbf{v}_k, \tag{2.5}$$

$$v_i^2 + v_j^2 + v_l^2 = v_g^2 + v_h^2 + v_k^2. \tag{2.6}$$

(ii) The gas particles are not distinguishable and undergo reversible collisions;

$$a_{ijl}^{ghk} = a_{ilj}^{ghk} = \cdots a_{lji}^{ghk} = a_{ijl}^{gkh} = \cdots = a_{ijl}^{khg}, \qquad a_{ijl}^{ghk} = a_{ghk}^{ijl}. \tag{2.7}$$

(iii) The transition rates A are related, for symmetric collisions, to the transition probability densities by the relation

$$A_{ijl}^{ghk} = (2/\sqrt{\pi}) S^{3/2} (s/3) (|\mathbf{v}_i - \mathbf{v}_j| + |\mathbf{v}_i - \mathbf{v}_l| + |\mathbf{v}_j - \mathbf{v}_l|) a_{ijl}^{ghk}, \tag{2.8}$$

which express the probability that a particle, say i, j, or l is in the action volume of the other two colliding particles. All three possible encounters being equally probable.

Remark 2.1: The indistinguishability and reversibility properties of the terms a is transferred, by Eq. (2.8), to the terms A.

Taking into account this property as well as all previous statements, the evolution equations for the densities N_i, when both binary and triple collisions are considered, can be written as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_i \cdot \nabla_{\mathbf{x}}\right) N_i = \tfrac{1}{2} \sum_{jhk} A_{ij}^{hk} (N_h N_k - N_i N_j) + \frac{1}{3!} \sum_{jlghk} A_{ijl}^{ghk} (N_g N_h N_k - N_i N_j N_e), \tag{2.9}$$

where the terms A_{ij}^{hk} and A_{ijl}^{ghk} are given by (2.2) and (2.8), respectively.

Remark 2.2: The modelization (i)–(iii) provides a result formally analogous to the one of Ref. 11 as far as the permutability of the indexes of the terms A is concerned. However, it introduces, as in Ref. 12, the concept that multiple collisions are less probable than binary collisions.

In order to point out the difference of collision frequency between binary and multiple collisions, it is convenient writing Eqs. (2.9) in a dimensionless form obtained normalizing time, space, velocity, and density with respect to suitable reference quantities, say $t_c, l_c, l_c/t_c$, and N_c and also normalizing the cross-sectional area with respect to σ^2, where σ is the radius of the particles, $\pi\sigma^2 = S$. This yields the dimensionless equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_i \cdot \nabla_{\mathbf{x}}\right) N_i = \frac{1}{\epsilon} (J_i^{(2)}[N] + \eta J_i^{(3)}[N]), \tag{2.10}$$

where ϵ and η are dimensionless constants

$$1/\epsilon = \sigma^2 l_c N_c, \quad \eta = \sigma^3 N_c, \tag{2.11}$$

and where $J_i^{(2)}$ and $J_i^{(3)}$ are the operators indicated in Eq. (2.9) referred to the new dimensionless variables.

Remark 2.3: Here, ϵ is proportional to the Knudsen number referred to l_c, i.e., $\epsilon \propto \mathrm{Kn}$, and η has physical meaning if of a smaller order with respect to the unity.

The analysis proposed to model the evolution equations when triple collisions are considered can be straightforwardly extended to model an evolution equation in the case of multiple collisions. Some additional notations are useful for this purpose. Accordingly, let us first write the term A, corresponding to encounters with a number p of particles, as

$$\overset{p}{A}{}_{\beta}^{\alpha} = A_{\beta_1 \cdots \beta_p}^{\alpha_1 \cdots \alpha_p}; \quad \alpha = \{\alpha_1, \ldots, \alpha_p\}, \quad \beta = \{\beta_1, \ldots, \beta_p\} \in I^p, \tag{2.12}$$

where $I^p = I \times \cdots \times I$ with p terms.

Consequently, the general expression of the collision operator $J_{\alpha_i}^{(p)}$, $\alpha_i \in I$, defined in Eq. (2.4), takes the form

$$J_{\alpha_i}^{(p)}[N] = \frac{1}{p!} \sum_{\alpha_i^*} \sum_{\beta} \overset{p}{A}{}_{\beta}^{\alpha} \left(\prod_{q=1}^{p} N_{\beta_q} - \prod_{q=1}^{p} N_{\alpha_q} \right), \tag{2.13}$$

where $\alpha_i^* = \{\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_p\} \in I^{p-1}$ and where

$$\overset{p}{A}{}_{\beta}^{\alpha} = \frac{2}{p(p-1)} \sum_{\alpha_i \neq \alpha_j} |\mathbf{v}_{\alpha_i} - \mathbf{v}_{\alpha_j}| s \left(\frac{2}{\sqrt{\pi}} s^{3/2} \right) \eta^{(p-2)} \overset{p}{a}{}_{\beta}^{\alpha}. \tag{2.14}$$

The general expression of the corresponding evolution equation can finally be written in a dimensionless form as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_i \cdot \nabla_{\mathbf{x}}\right) N_i = \frac{1}{\epsilon} \sum_{p=2}^{P} \eta^{(p-2)} J_i^{(p)}[N], \tag{2.15}$$

where $J_i^{(p)}$ are given by Eq. (2.13) and are referred to as the new dimensionless variables.

Remark 2.4: The indistinguishability and reversibility properties, referred for every $p \geqslant 2$ to the terms A, can also formally be written as

$$A_\beta^{\sigma(\alpha)} = A_\beta^\alpha = A_{\sigma(\beta)}^\alpha, \quad \forall \sigma \in s_p, \quad \forall \alpha, \beta \in I^p, \tag{2.16a}$$

$$A_\beta^\alpha = A_\alpha^\beta, \quad \forall \alpha, \beta \in I^p, \tag{2.16b}$$

where s_p is the permutation group over $\{1,...,p\}$ and

$$\sigma(\alpha) = \{\alpha_{\sigma(1)},...,\alpha_{\sigma(p)}\}, \quad \sigma \in s_p, \quad \alpha \in I^p. \tag{2.17}$$

The notations of Remark 2.4 will be used in the analysis developed in what follows.

The definitions of collision invariants and Maxwellian state can now be given.

Definition 1: A vector $\phi \in \mathbb{R}^n$ is defined "collision invariant" if for each $p = 2,...,P$ the following equality holds:

$$\sum_\beta^p A_\beta^\alpha \left(\sum_{q=1}^p \phi_{\alpha_q} - \sum_{q=1}^p \phi_{\beta_q} \right) = 0, \quad \forall \alpha, \beta \in I^p. \tag{2.18}$$

The space of collision invariants is denoted by $\mathcal{M}$.

Definition 2: A vector function $N > 0$, $N \in \mathbb{R}^n$, is a Maxwellian if, for each $p = 2,...,P$, the following equality holds:

$$\sum_\beta^p A_\beta^\alpha \left(\prod_{q=1}^p N_{\alpha_q} - \prod_{q=1}^p N_{\beta_q} \right) = 0, \quad \forall \alpha, \beta \in I^p. \tag{2.19}$$

It is now a matter of straightforward calculations, also after Ref. 1 so that the proof is not repeated, proving the following propositions.

Proposition 2.1: For a vector $\phi \in \mathbb{R}^n$, the following relations are equivalent:

(i) $\phi \in \mathcal{M}$;

(ii) for each p: $\langle \phi, J^{(p)}[N] \rangle = 0, \forall N \in \mathbb{R}^n$;

(iii) $\langle \phi, J[N] \rangle = 0, \forall N \in \mathbb{R}^n$.

Proposition 2.2: For a vector $N \in \mathbb{R}^n$, with $N > 0$, the following relations are equivalent;

(i) N is a Maxwellian;

(ii) $\log N \in \mathcal{M}$;

(iii) for each $p = 2,...,P$: $J^{(p)}[N] = 0$;

(iv) $J[N] = 0$.

Here, J and N denote, in the statements of Propositions 2.1–2, the vectors with components J_i and N_i, respectively. Analogous definitions and properties are given in Ref. 11.

III. STABILITY CONDITIONS AND GLOBAL EXISTENCE NEAR EQUILIBRIUM

Global existence and asymptotic behavior for the solutions to the initial value problem for initial conditions close to equilibrium have been studied in Refs. 7 and 8 for discrete velocity models with binary collisions only. This section will develop the analysis for general discrete velocity models with multiple collisions.

The line developed throughout this section is the following: We first derive an evolution equation for the perturbation of the Maxwellian state, then a global existence theorem is proven under suitable stability conditions, and finally, some detailed criteria are provided in order to verify the stability of specific models.

Keeping this in mind and putting $N = \{N_1,...,N_n\}^T$, $J[N] = \{J_1[N],...,J_n[N]\}^T$, then Eq. (2.15) can be rewritten in the vector form

$$\frac{\partial N}{\partial t} + \sum_{j=1}^d v^j \frac{\partial N}{\partial x_j} = J[N] = \sum_{p=2}^P J^{(p)}[N], \tag{3.1}$$

where in Eq. (3.1) the dimensionless constants, say $\epsilon, \eta,...$, have been put (as well as in the equations which follow), for simplicity, equal to one. Moreover,

$$v^j = \text{diag}\{v_{1j},...,v_{nj}\}, \quad 1 \leqslant j \leqslant d, \tag{3.2}$$

where the jth components of the vectors $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{v}_i \in R^d$ have been denoted by x_j and v_{ij}.

We deal with the initial value problem referred to Eq. (3.1) in all space $\mathbb{R}^d$ when the initial conditions are close to a constant Maxwellian $M = \{M_1,...,M_n\}^T > 0$ chosen arbitrarily. Letting $N = N(t,\mathbf{x})$ be the solution of such a problem, it is convenient, as usual, to express N as a perturbation of M, by means of a suitable vector function $f = f(t,\mathbf{x})$, and derive an evolution equation for f. Thus let

$$N = M + \Lambda_M f, \quad \Lambda_M = \text{diag}\{M_1,...,M_n\}, \tag{3.3}$$

then, substituting (3.3) into (3.1) and recalling that $J[M] = 0$, yields

$$\Lambda_M \frac{\partial f}{\partial t} + \sum_{j=1}^d v^j \Lambda_M \frac{\partial f}{\partial x_j} + L_M f = T_M[f], \tag{3.4}$$

where, with obvious meaning of symbols, $J[N]$ has been decomposed into the linear and nonlinear terms $-L_M f$ and $T_M[f]$, respectively. That is, $J[M + \Lambda_M f] = -L_M f + T_M[f]$.

The term L_M will be called, in what follows, the linearized collision operator. Recalling that J is the sum of the $J^{(p)}$ terms, one has, based on the fact that L_M is linear, the following:

$$L_M f = \sum_{p=2}^P L_M^{(p)} f. \tag{3.5}$$

In addition, using (2.13), an explicit expression of each $L_M^{(p)} f$ can be provided. In details the α_ith component of $L_M^{(p)} f$ is, for any $\alpha_i \in I$, the following:

$$\frac{1/2}{p!} \sum_{\alpha_i^*} \sum_\beta^p A_\beta^\alpha \left(\prod_{q=1}^p M_{\alpha_q} + \prod_{q=1}^p M_{\beta_q} \right) \times \left(\sum_{q=1}^p f_{\alpha_q} - \sum_{q=1}^p f_{\beta_q} \right), \tag{3.6}$$

where the double summation is taken over all $\alpha_i^* \in I^{p-1}$ and $\beta \in I^p$.

Remark 3.1: The fact that M is a Maxwellian and consequently it satisfies the equality $A_\beta^\alpha \Pi_q M_{\alpha_q} = A_\beta^\alpha \Pi_q M_{\beta_q}$, for any $\alpha, \beta \in I^p$, has been used in deriving Eq. (3.6).

By virtue of (3.6), it is simple to prove the following.

Lemma: L_M is an $n \times n$ real symmetric matrix with constant entries. Moreover, L_M is non-negative definite and its null space $\mathcal{N}(L_M)$ coincides with the space $\mathcal{M}$ of the collision invariants.

Proof: Let f and g be arbitrary vectors in $\mathbb{R}^n$. Simple calculations, using (3.6), give, besides constants put equal to one, the following:

$$\langle L_M^{(p)} f,g\rangle = \sum_\alpha \sum_\beta A_\beta^\alpha \left(\prod_{q=1}^p M_{\alpha_q} + \prod_{q=1}^p M_{\beta_q}\right) \times \left(\sum_{q=1}^p f_{\alpha_q} - \sum_{q=1}^p f_{\beta_q}\right)\left(\sum_{q=1}^p g_{\alpha_q} - \sum_{q=1}^p g_{\beta_q}\right), \tag{3.7}$$

for each $p = 2,...,P$.

The right-hand side of (3.7) is symmetric with respect to f and g. Therefore, we get

$$\langle L_M^{(p)} f,g\rangle = \langle L_M^{(p)} g,f\rangle;$$

on the other hand, as an elementary property of the inner product of $\mathbb{R}^n$, we have

$$\langle L_M^{(p)} g,f\rangle = \langle f,L_M^{(p)} g\rangle,$$

then

$$\langle L_M^{(p)} f,g\rangle = \langle f,L_M^{(p)} g\rangle.$$

This shows that each $L_M^{(p)}$ is real symmetric and hence the same is true also for L_M. Moreover, putting $f = g$ in Eq. (3.7), shows that $\langle L_M^{(p)} f,f\rangle \geqslant 0$, which means that each $L_M^{(p)}$ is non-negative. Therefore, also, L_M is non-negative. In addition, $L_M f = 0$ holds if and only if $L_M^{(p)} f = 0$ for each $p = 2,...,P$. This property, referred to in (3.7) with $f = g$ is equivalent to the following one:

$$\text{for each } p = 2,...,P: A_\beta^\alpha \left(\sum_{q=1}^p f_{\alpha_q} - \sum_{q=1}^p f_{\beta_q}\right) = 0, \quad \forall \alpha,\beta \in I^p.$$

This means that $f \in \mathcal{M}$. Then we have proved that $\mathcal{N}(L_M) = \mathcal{M}$. The proof of the Lemma is then complete. ■

Remark 3.2: The multiple-collision linearized operator L_M satisfies the same properties of the corresponding operator in the case of binary collisions. This equivalence was stated, without proof, in Ref. 11.

According to Remark 3.2, the same techniques applied in Ref. 7 can now be straightforwardly applied in the multicollisional case in order to provide global existence results for the solutions to Eq. (3.1) for initial data in a neighborhood of a Maxwellian.

In general, global existence does not hold for all discrete velocity models, but only for models that satisfy suitable stability conditions, which assure the energy decay of the solutions to the linearized problem.

Keeping this in mind, we recall the stability conditions formulated in Ref. 13 as an improved version of the conditions previously given in Ref. 7. Then let $\omega = \{\omega_1,...,\omega_d\} \in s^{d-1}$ and let

$$V(\omega) = \sum_{j=1}^d V^j \omega_j = \text{diag}(\mathbf{v}_1 \cdot \omega,...,\mathbf{v}_n \cdot \omega), \tag{3.8}$$

then the two equivalent stability conditions, which follow, can be stated:

Stability condition 3.1: Let $\phi \in \mathcal{M}$ and let $v(\omega)\phi = \lambda\phi$ for $\omega \in s^{d-1}$ and $\lambda \in \mathbb{R}$, then $\phi = 0$.

Stability condition 3.2: There exists a matrix $K(\omega)$ smoothly depending upon $\omega \in s^{d-1}$ with the following properties:

(i) $K(-\omega) = -K(\omega)$ for $\omega \in s^{d-1}$;

(ii) $K(\omega)$ is, for each $\omega \in s^{d-1}$, a real skew-symmetric matrix;

(iii) the symmetric part of the matrix $K(\omega)$ $V(\omega) + L_M$ is positive definite for any $\omega \in s^{d-1}$.

Under Stability condition 3.1, which is equivalent to 3.2 (according to Theorem 1.1 of Ref. 13, see also Sec. 4.3 of Ref. 4), global existence of the solution to the initial value problem for initial conditions near equilibrium can be proven according to the following theorem.

Theorem: Let $d \geqslant 2$. Assume that stability condition 3.1 holds and consider the initial value problem defined by Eq. (3.4) with initial data $f(0,\mathbf{x}) = f_0(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d$.

(a) *Global existence and stability.* Suppose that $f_0 \in H^s(\mathbb{R}^d)$ for $s \geqslant [d/2] + 1$ such that the norm of f_0, $\|f_0\|_s$ is sufficiently small. Then, the initial value problem has a unique global solution $f = f(t,\mathbf{x})$ which satisfies the following:

$$f \in B^0([0,\infty);H^s(\mathbb{R}^d)) \cap B^1([0,\infty);H^{s-1}(\mathbb{R}^d)), \quad \partial_x f \in L^2([0,\infty);H^{s-1}(\mathbb{R}^d)), \tag{3.9}$$

with obvious meaning of symbols. In addition, the following estimate,

$$\|f(t)\|_s^2 + \int_0^t \|\partial_x f(\tau)\|_{s-1}^2 \, d\tau \leqslant c\|f_0\|_s^2, \tag{3.10}$$

holds for any $t \geqslant 0$, where c is a constant. Moreover $f(t,\mathbf{x})$ converges to zero uniformly in $x \in \mathbb{R}^d$ as $t \to \infty$.

(b) *Quantitative asymptotic behavior.* Suppose, in addition, that $f_0 \in H^s(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ for $s \geqslant [d/2] + 1$, $1 \leqslant p < 2$ and that the norm $\|f_0\|_{s,p}$ is sufficiently small, then the solution obtained according to step (a) satisfies the following inequality:

$$\|f(t)\|_s \leqslant C(1+t)^{-\gamma}\|f_0\|_{s,p}, \tag{3.11}$$

for any $t \geqslant 0$, where C is a constant and $\gamma = (d/2)(1/p - 1/2)$.

Proof: Considering that the statement of the theorem refers first to global existence and stability and then to quantitative asymptotic behavior, then also the proof, which follows the same logic line as in Ref. 7, is in two steps.

Step 1: Global existence is proven applying the classical fixed point theorem. Consider then the Banach space X^s defined as

$$X^s = \{f \in B^0([0,\infty);H^s(\mathbb{R}^d)); \partial_x f \in L^2([0,\infty); \times H^{s-1}(\mathbb{R}^d))\},$$

equipped with the norm $|||\cdot|||_s$:

$$|||f|||_s^2 = \sup_{t>0}\|f(t)\|_s^2 + \int_0^\infty \|\partial_x f(t)\|_{s-1}^2 \, dt. \tag{3.12}$$

For a given function $g \in X^s$, let us consider the linearized equation

$$\Lambda_M \frac{\partial f}{\partial t} + \sum_{j=1}^d V^j \Lambda_M \frac{\partial f}{\partial x_j} + L_M f = T_M[g], \tag{3.13}$$

with initial conditions $f(0,\mathbf{x}) = f_0(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d$. This linearized problem has a unique solution:

$f \in C^0([0,\infty), H^s(\mathbb{R}^d)) \cap C^1([0,\infty); H^{s-1}(\mathbb{R}^d))$.

An estimate on the solution will now be derived in order to show that $f \in X^s$. To this aim, we take the Fourier transform of (3.13) with respect to the variable $\mathbf{x}$, obtaining

$$\Lambda_M \frac{\partial \hat{f}}{\partial t} + (i|\xi| V(\omega)\Lambda_M)\hat{f} = \hat{T}_M[g], \tag{3.14}$$

where $\hat{f}(t,\xi)$ is the Fourier image of $f(t,\mathbf{x})$, and $\omega = \xi/|\xi|$. Since Stability condition 3.1 is equivalent to condition 3.2, we can find a matrix $K(\omega)$ characterized by the properties defined in the said condition. Then, $E[\hat{f}]$ can be defined, with the aid of such a matrix, as

$$E[\hat{f}] = \langle \Lambda_M \hat{f}\hat{f}\rangle - \alpha\tilde{\rho}(|\xi|)\langle iK(\omega)\hat{f}\hat{f}\rangle, \tag{3.15}$$

where $\alpha > 0$ is a constant, $\tilde{\rho}(r)$ is a function of $r \geqslant 0$ such that $\tilde{\rho}(r) = r/(1+r^2)$, and $\langle \cdots \rangle$ is the inner product of C^n.

It can now be easily verified that if $\alpha > 0$ is chosen sufficiently small, then repeating calculations technically analogous to the ones contained in the proof of Proposition 4.1 of Ref. 7, one has

$$c|\hat{f}|^2 \leqslant E[\hat{f}] \leqslant c|\hat{f}|^2 \tag{3.16}$$

and

$$\frac{\partial E[\hat{f}]}{\partial t} + c\rho(|\xi|)|\hat{f}|^2 \leqslant C|\hat{T}_M[g]|^2, \qquad \rho(r) = r^2/(1+r^2), \tag{3.17}$$

and c and C are positive constants. In particular, inequalities (3.16) and (3.17) can be derived on the basis of properties (ii), (iii) of $K(\omega)$, on the ones of L_M stated in the lemma as well as the fact that $T_M[g]$ is orthogonal to $\mathcal{M}$. If now (3.17) is multiplied by $(1+|\xi|^2)^s$ and the integration over t and $\xi \in \mathbb{R}^d$ is executed, then also using (3.16) together with Plancharel's theorem, one can prove the following estimate:

$$\|f(t)\|_s^2 + \int_0^t \|\partial_x f(\tau)\|_{s-1}^2 d\tau \leqslant C_0^2 \|f_0\|_s^2 + C\int_0^t \|T_M g(\tau)\|_s^2 \, d\tau\|, \tag{3.18}$$

where $C_0 > 1$ and C are constants and $t \geqslant 0$.

On the other hand, by virtue of Nirenberg's inequality, one has for some constant C and using the fact $d \geqslant 2$ (the estimate, in fact, is not true for $d = 1$),

$$\|T_M g\|_s \leqslant C\left(\sum_{k=0}^{P-2} \|g\|_x^k\right)\|g\|_s \|\partial_x g\|_{s-1}.$$

Then one obtains

$$\int_0^t \|T_M g(\tau)\|_s^2 d\tau \leqslant C\left(\sum_{k=0}^{P-2} |||g|||_s^k\right)^2 |||g|||_s^4. \tag{3.19}$$

The estimate (3.18) combined with (3.19) shows that if $g \in X^s$ then also $f \in X^s$. In addition, the same inequalities [(3.18) and (3.19)] show that the mapping $g \to f$ defined by (3.13) has an invariant subset $W \in X^s$ defined by the closure

$$W = \{f \in X^s; |||f|||_s \leqslant 2C_0|f_0|_s\}, \tag{3.20}$$

where C_0 is the constant already defined in (3.18), provided that the norm of f_0 is sufficiently small. Moreover, if the norm of f_0 is sufficiently small, the mapping $g \to f$ is a contractive mapping with respect to the norm $|||\cdot|||_s$. Consequently, a unique fixed point of such a mapping exists in W and is the solution of the initial value problem. This proves the first step of the theorem's proof.

Step 2: The solution f of (3.4) satisfies (3.16) and (3.17) with $f = g$, therefore we have

$$|f(t,\xi)|^2 \leqslant C\exp(-c\rho(|\xi|)t)|\hat{f}_0(\xi)|^2 + C\int_0^t \exp(-c\rho(|\xi|)(t-\tau)) \times |\hat{T}_M[f](\tau,\xi)|^2 \, d\tau, \tag{3.21}$$

where c and C are positive constants. By straightforward calculations, (3.21) yields

$$\|f(t)\|_s^2 \leqslant C(1+t)^{-2\gamma}\|f_0\|_{s,p}^2 + C\int_0^t (1+t-\tau)^{-d/2} \times \|T_M f(\tau)\|_{s,1}^2 \, d\tau, \tag{3.22}$$

where C is a constant and $\gamma = (d/2)(1/p - 1/2)$. On the other hand, for some constant C one has

$$\|T_M f\|_{s,1} \leqslant C\left(\sum_{k=0}^{P-2} \|f\|_s^k\right)\|f\|_s^2.$$

Substituting this estimate into (3.22), yields the estimate (3.11) required by the statement of the theorem, if $\|f_0\|_{s,p}$ is sufficiently small. The proof of the theorem is completed. ■

Remark 3.3: The method employed in the theorem is not valid for $d = 1$. Nevertheless, it seems simply a technical problem deriving a mathematical result analogous to the one proved above still in the line followed in Ref. 7. Nevertheless, this aspect is not developed in this paper.

The application of the theorem to the analysis of the intial value problem near equilibrium referred to specific models requires the analysis of the space $\mathcal{M}$ of collision invariants and the verification of Stability condition 3.1. Nevertheless, it is reasonable to expect that several specific models with multiple collisions are characterized by only classical collision invariants defined by mass, momentum, and energy:

$$\begin{aligned} &\phi^{(0)} = (1,\ldots,1)^T; \\ &\phi^{(j)} = (v_{1J},\ldots,v_{nJ})^T, \quad 1 \leqslant j \leqslant d; \\ &\phi^{(d+1)} = (v_1^2,\ldots,v_n^2)^T. \end{aligned} \tag{3.23}$$

When this situation is verified, simple stability criteria can be provided in order to verify technically Stability condition 3.1. The first one is essentially the same as the one proposed by Cercignani in Ref. 15 referred to models with binary collisions only, and can be formulated as the following.

Criterion 3.1 (Ref. 15)*:* Let $d \geqslant 2$ and consider a d-dimensional model such that

$$\dim \mathcal{M} = d+2, \quad \mathcal{M} = \operatorname{span}\{\phi^{(0)}, \phi^{(j)}, 1 \leqslant j \leqslant d, \phi^{(d+1)}\}. \tag{3.24}$$

Then the model satisfies Stability conditions 3.1 if the following $(d^2+5d+2)/2$ vectors are linearly independent:

$$\begin{aligned} &\phi^{(0)}, \phi^{(j)}, 1 \leqslant j \leqslant d; \quad v^j\phi^{(k)}, \quad 1 \leqslant j \leqslant k \leqslant d; \\ &v^j\phi^{(d+1)}, \quad 1 \leqslant j \leqslant d. \end{aligned} \tag{3.25}$$

Remark 3.4: Criterion 3.1 requires that the number n of velocities in the model satisfies the condition

$n \geqslant (d^2 + 5d + 2)/2$ and that the model involves different velocity moduli.

When the discrete model is a relatively simpler one and involves only one velocity modulus, the collision invariants $\phi^{(0)}$ and $\phi^{(d+1)}$ are not linearly independent. In this case, a similar criterion can be formulated

Criterion 3.2: Let $d \geqslant 2$ and consider a d-dimensional model with one only velocity modulus and such that

$$\dim \mathscr{M} = d + 1, \quad \mathscr{M} = \operatorname{span}\{\phi^{(0)}, \phi^{(j)}, 1 \leqslant j \leqslant d\} \tag{3.26}$$

then such a model satisfies Stability condition 3.1 if the following $(d^2 + 3d)/2$ vectors are linearly independent:

$$\begin{aligned} &\phi^{(0)}, \phi^{(j)}, 1 \leqslant j \leqslant d, v^j \phi^{(k)}, \quad 1 \leqslant j \leqslant k \leqslant d, \\ &v^j \phi^{(j)}, \quad 1 \leqslant j \leqslant d-1. \end{aligned} \tag{3.27}$$

Proof: Let $\phi \in \mathscr{M}$ and let $V(\omega)\phi = \lambda\phi$ for $\omega \in S^{d-1}$ and $\lambda \in \mathbf{R}$. The relation $\phi \in \mathscr{M}$ means, according to (3.26), that

$$\phi = \alpha_0 \phi^{(0)} + \sum_{j=1}^{d} \alpha_j \phi^{(j)}, \tag{3.28}$$

for some real constants α_0 and α_j, $1 \leqslant j \leqslant d$. If this expression is substituted into $V(\omega)\phi = \lambda\phi$, then a simple calculation that uses (3.8) as well as the relations

$$V^j \phi^{(0)} = \phi^{(j)}, \quad 1 \leqslant j \leqslant d;$$

$$V^j \phi^{(k)} = V^k \phi^{(j)}, \quad 1 \leqslant j,k \leqslant d;$$

$$\sum_{j=1}^{d} V^j \phi^j = \phi^{(d+1)} = v^2 \phi^{(0)};$$

where v is the modulus of the velocities $\mathbf{v}_i$, shows that

$$\begin{aligned} &\alpha_d \omega_d v^2 \phi^{(0)} + \sum_{j=1}^{d} \alpha_0 \omega_j \phi^{(j)} + \sum\sum_{1 \leqslant j < k \leqslant d} (\alpha_k \omega_j + \alpha_j \omega_k) V^j \\ &\times \phi^{(k)} + \sum_{j=1}^{d-1} (\alpha_j \omega_j - \alpha_d \omega_d) V^j \phi^{(j)} \\ &= \alpha_0 \lambda \phi^{(0)} + \sum_{j=1}^{d} \alpha_j \lambda \phi^{(j)}. \end{aligned}$$

Considering now that the vectors listed in (3.27) are assumed to be linearly independent, we can conclude that

$$\begin{aligned} &\alpha_d \omega_d v^2 = \alpha_0 \lambda, \quad \alpha_0 \omega_j = \alpha_j \lambda, \quad 1 \leqslant j \leqslant d, \\ &\alpha_k \omega_j + \alpha_j \omega_k = 0, \quad 1 \leqslant j < k \leqslant d, \\ &\alpha_j \omega_j = \alpha_d \omega_d, \quad 1 \leqslant j \leqslant d-1. \end{aligned}$$

These relations show that $\alpha_0 = 0$ and $\alpha_j = 0$, $1 \leqslant j \leqslant d$, which imply, according to (3.28), that $\phi = 0$. This shows that the model which has been taken into account satisfies Stability condition 3.1. The proof is then complete. ■

IV. APPLICATION

As an application of the theory developed in the preceding sections, we shall consider here two specific models and verify, for these models, the stability criteria developed in Sec. III. It is interesting, indeed, verifying if the stability conditions are fulfilled by model with multiple collisions when the same conditions are not verified for the corresponding model with binary collisions only.

In details, we shall consider here a plane regular discrete velocity model such that all velocities have the same modulus but six directions in the plane and a second model with 12 velocities in the plane, six with a modulus and six with a different one.

A. Model 1

Consider the following six-velocity discretization:

$$\begin{aligned} i = 1,\ldots,6: \mathbf{v}_i = c\mathbf{e}_i, \quad \mathbf{e}_i &= \cos(\pi/6 + \pi(i-1)/3)\boldsymbol{\omega}_1 \\ &+ \sin(\pi/6 + \pi(i-1)/3)\boldsymbol{\omega}_2; \end{aligned} \tag{4.1}$$

if one restricts the attention to the so-called "nontrivial" collisions, namely, collisions which modify the fluxes, that is,

$$(\mathbf{v}_i, \mathbf{v}_j) \neq (\mathbf{v}_g, \mathbf{v}_h), \quad (\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_l) \neq (\mathbf{v}_g, \mathbf{v}_h, \mathbf{v}_k). \tag{4.2}$$

It is simple to verify that condition (4.2) holds true, in the case of both of binary and triple collision, when the resultant momentum before and after the collision is equal to zero.

In detail, the nontrivial collisions consistent with the velocity discretization (4.1) are the following.

(a) Binary head-on collisions with equally probable scattering in all radial directions:

$$(N_i, N_{i+3}) \leftrightarrow (N_h, N_{h+3}), \quad h = 1,\ldots,6.$$

(b) Triple collisions between particles forming an equilateral triangle with equally probable scattering into three equally angularly spaced directions:

$$(N_i, N_{i+2}, N_{i+4}) \leftrightarrow (N_h, N_{h+2}, N_{h+4}), \quad h = 1,\ldots,6.$$

The application of the method developed in Sec. II provides the following model equation:

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \mathbf{v}_i \cdot \nabla_{\mathbf{x}}\right) N_i = \frac{1}{\epsilon} \Bigg\{ &\frac{cS}{3} \sum_{h=1}^{6} (N_h N_{h+3} - N_i N_{i+3}) \\ &+ \eta \frac{\sqrt{6/\pi}}{3} cS \cdot S^{3/2} \sum_{h=1}^{6} (N_h N_{h+2} N_{h+4} \\ &- N_i N_{i+2} N_{i+4}) \Bigg\}, \end{aligned} \tag{4.3}$$

where all quantities have been normalized as indicated in Eqs. (2.10) and (2.11).

Remark 4.1: If all velocities in the plane are considered, say $\mathbf{v}(\theta) = c(\cos\theta\boldsymbol{\omega}_1 + \sin\theta\boldsymbol{\omega}_2)$, instead of six velocities only, then one obtains the so-called semidiscrete Boltzmann equation with triple (symmetrical) collisions. Such a model was proposed by Cabannes[16] and studied in Refs. 17 and 18 in the case of binary collisions only.

This equation defines the time-space evolution of the density $N = N(t,\mathbf{x};\theta)$ by means of the integrodifferential equation

$$\left(\frac{\partial}{\partial t}+\mathbf{v}(\theta)\cdot\nabla_{\mathbf{x}}\right)N(t,\mathbf{x};\theta)$$
$$=\frac{1}{\epsilon}\left\{\frac{2cS}{\pi}\int_0^{\pi}(N(t,\mathbf{x};\phi)N(t,\mathbf{x};\phi+\pi)-N(t,\mathbf{x};\theta)N(t,\mathbf{x};\theta+\pi))d\phi+\eta\frac{\sqrt{6/\pi}}{(2\pi)}cS\cdot S^{3/2}\right.$$
$$\left.\times\int_0^{2\pi}\left(N(t,\mathbf{x};\phi)N\left(t,\mathbf{x};\phi+\frac{2\pi}{3}\right)N\left(t,\mathbf{x};\phi+\frac{4\pi}{3}\right)-N(t,\mathbf{x};\theta)N\left(t,x;\theta+\frac{2\pi}{3}\right)N\left(t,\mathbf{x};\theta+\frac{4\pi}{3}\right)\,d\phi\right)\right\}. \tag{4.4}$$

B. Model 2

Model 1 is characterized by only one velocity modulus. This implies that only two fluid-dynamic parameters can be regarded as independent variable or, in other words, the temperature is not an independent variable.

On the other hand, using the same direction discretization, as in model 1, it is possible to obtain a 12 velocity model with 2 velocity moduli by means of the following discretization:

$$i=1,\ldots,6:\mathbf{x}_i=c\mathbf{e}_i,\quad \mathbf{v}_i'=2c\mathbf{e}_i\,. \tag{4.5}$$

If each set of velocities is joined to the densities N_i and M_i, respectively, the following nontrivial collisions are consistent with such a discretization.

(a) Binary collisions between N and M particles, separately, as in (a) of model 1.

(b) Binary collisions mixing N and M particles:

$$(N_i,M_{i+2})\leftrightarrow(N_{i+3},M_{i+1});(N_i,M_{i+4})\leftrightarrow(N_{i+3},M_{i+5})\,. \tag{4.6}$$

(d) Triple collisions among N and M particles, separately, as in (b) of model 1.

(e) Triple collisions mixing N and M particles:

$$(N_i,N_i,M_{i+3})\leftrightarrow(N_h,N_h,M_{h+3});$$
$$(M_i,N_{i+3},N_{i+3})\leftrightarrow(M_h,N_{h+3},N_{h+3}),$$

for $h=1,\ldots,6$ and

$$(N_i,M_{i+1},M_{i+3})\leftrightarrow(N_{i+3},M_{i+2},M_i);$$
$$(N_i,M_{i+5},M_{i+3})\leftrightarrow(N_{i+3},M_{i+1},M_i)\,.$$

Consequently, the discrete velocity model can be written as

$$\left(\frac{\partial}{\partial t}+c\mathbf{e}_i\cdot\nabla_{\mathbf{x}}\right)N_i$$
$$=\frac{1}{\epsilon}\left\{\frac{cS}{3}\sum_{h=1}^{6}(N_hN_{h+3}-N_iN_{i+3})+2cS(N_{i+3}M_{i+1}+N_{i+3}M_{i+5}-N_iM_{i+2}-N_iM_{i+4})+\eta\left(\frac{\sqrt{6/\pi}}{3}\right)cS\cdot S^{3/2}\right.$$
$$\times\sum_{h=1}^{6}(N_hN_{h+2}N_{h+4}-N_iN_{i+2}N_{i+4})+\eta\left(\frac{2}{3\sqrt{\pi}}\right)cS\cdot S^{3/2}\left[\sum_{h=1}^{6}(N_hN_hM_{h+3}-N_iN_iM_{i+3})\right.$$
$$\left.\left.+6(N_{i+3}M_i(M_{i+1}+M_{i+2})-N_iM_{i+3}(M_{i+1}+M_{i+5}))\right]\right\}, \tag{4.7a}$$

$$\left(\frac{\partial}{\partial t}+2c\mathbf{e}_i\cdot\nabla_{\mathbf{x}}\right)M_i=\frac{1}{\epsilon}\left\{\frac{2cS}{3}\sum_{h=1}^{6}(M_hM_{h+3}-M_iM_{i+3})+4cS(M_{i+3}N_{i+1}+M_{i+3}N_{i+5}-M_iN_{i+2}-M_iN_{i+4})\right.$$
$$+\eta\;2\cdot\left(\frac{\sqrt{6/\pi}}{3}\right)cS\cdot S^{3/2}\sum_{h=1}^{6}(M_hM_{h+2}M_{h+4}-M_iM_{i+2}M_{i+4})+\eta\frac{2}{3\sqrt{\pi}}cS\cdot S^{3/2}$$
$$\times\left[\sum_{h=1}^{6}(M_{h+3}N_hN_h-M_iN_{i+3}N_{i+3})\right.$$
$$\left.\left.+6(M_{i+3}N_i(N_{i+1}+N_{i+2})-M_iN_{i+3}(N_{i+1}+N_{i+5}))\right]\right\}. \tag{4.7b}$$

C. Global existence and stability

Section III has shown that global existence of the solutions to the initial value problem near equilibrium for discrete velocity models with multiple collisions can be proven under suitable stability conditions. In particular, we refer to conditions 3.1 and 3.2 as well as to the sufficient stability criteria provided in order to verify the above conditions.

Model 1 will be studied in details, whereas only some indications will be given referring to model 2 in order to repeat technically the analysis. In particular, condition 3.1 is verified for model 1 by using criterion 3.2. While the same condition can be verified for model 2 by using criterion 3.1. Referring now to model 1, the following can be proven.

Proposition: Consider model 1 when both binary and triple collisions are taken into account, then

(i) Dim $\mathcal{M}=3$, $\mathcal{M}=\text{span}\{\phi^{(0)},\phi^{(1)},\phi^{(2)}\}$, and the stability condition 3.1 is verified through criterion 3.2. On the other hand, when binary collisions only are taken into account, then

(ii) Dim $\mathcal{M}=4$, $\mathcal{M}=\text{span}\{\phi^{(0)},\phi^{(1)},\phi^{(2)},\chi\}$ where χ is

an artificial invariant, specified later, and Stability condition 3.1 is not verified.

Proof: Let $c = 2$, then the six velocities of the model are

$$\mathbf{v}_1 = -\mathbf{v}_4 = (\sqrt{3},1),$$

$$\mathbf{v}_2 = -\mathbf{v}_5 = (0,2), \quad \mathbf{v}_3 = -\mathbf{v}_6 = (-\sqrt{3},1).$$

Therefore, in two space dimensions

$$V^1 = \sqrt{3}\,\mathrm{diag}(1,0,-1,-1,0,1),$$

$$V^2 = \mathrm{diag}(1,2,1,-1,-2,-1).$$

The space of collision invariants has dimension 3 and is spanned by mass, and momentum, in the directions ω_1 and ω_2, as vectors of $\mathbf{R}^6$:

$$\dim \mathcal{M} = 3, \quad \mathcal{M} = \mathrm{span}\{\phi^{(1)},\phi^{(1)},\phi^{(2)}\},$$

where

$$\phi^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \phi^{(1)} = \sqrt{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \quad \phi^{(2)} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \\ -2 \\ -1 \end{bmatrix}.$$

A simple calculation shows that $V^1\phi^{(2)} = \phi^{(3)}$ and $V^1\phi^{(1)} = 2\phi^{(0)} + \phi^{(4)}$, where

$$\phi^{(3)} = 3\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \quad \phi^{(4)} = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 1 \\ -2 \\ 1 \end{bmatrix},$$

the vectors $\phi^{(0)},\phi^{(1)},\phi^{(2)},\phi^{(3)}$, and $\phi^{(4)}$ form an orthogonal system in $\mathbf{R}^6$ and therefore $\phi^{(0)},\phi^{(1)},\phi^{(2)}$, $V^1\phi^{(1)}$, $V^1\phi^{(2)}$ are linearly independent vectors so that the stability condition 3.1 is verified. (Note that Stability condition 3.1 can be directly verified by simple technical calculations.)

On the other hand, model 1 does not satisfy condition 3.1 when binary collisions only are considered. To see this, we first note that in this case,

$$\dim \mathcal{M} = 4, \quad \mathcal{M} = \mathrm{span}\{\phi^{(0)},\phi^{(1)},\phi^{(2)},\chi\},$$

where χ is an artificial invariant and is given by

$$\chi = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

Consider now the set Ω consisting of six unit vectors in $\mathbf{R}^2$:

$$\Omega = \{(\pm 1,0), \pm\tfrac{1}{2}, \pm\tfrac{1}{2}\}, \tag{4.8}$$

then, it is easy to see that for each $\omega \in \Omega$, there exist real numbers α_1, α_2, and β satisfying $(\alpha_1,\alpha_2,\beta) \neq 0$ such that

$$V(\omega)\phi = 0 \text{ for } \phi = \alpha_1\phi^{(1)} + \alpha_2\phi^{(2)} + \beta\chi,$$

which means that condition 3.1 is not verified in this case. The proposition is then proven. ■

Remark 4.1: Even in the case of binary collisions only, if one considers a one-dimensional flow in ω directions, it is possible to verify Stability condition 3.1 frozen at ω, provided that $\omega \neq \Omega$, where Ω is given by Eq. (4.8). However, as it has been pointed out above, the stability condition does not hold for all ω.

The stability analysis can be applied, in the same fashion as for model 1, also in the case of the second model thus arriving to the same conclusions, which essentially confirm that including triple collisions improves the stability properties of the model itself. We can acknowledge these conclusions without repeating all calculations, which can be regarded as a repetition.

Some conclusions can now be worked out from the analysis developed throughout the paper. Referring to Ref. 7 where the stability problem was originally posed in the style of this paper, the author conjectured that some models do not show stability properties toward equilibrium as their intrinsic structure was too simple to be physically consistent with a reasonable physical behavior (in our case: stability of small perturbations of equilibrium). In the light of the analysis developed in this paper, the said conjecture appears essentially correct.

In fact, the introduction of triple collisions exploits more completely the applicability of a certain discretization and may, in the largest part of cases, stabilize unstable models. In any case, even if this cannot be regarded as a general rule, the paper provides a rigorous method to establish stability properties of the mathematical models of the discrete kinetic theory with multiple collisions.

On a physical ground, it can be conjectured that introducing multiple collisions increases the gas viscosity so that trend to equilibrium is improved. This conjecture, however, still has to be put in mathematical terms. The main conclusion still remains that a careful analysis of the stability properties of the discrete Boltzmann equation is one of the main steps toward the validation of specific discrete velocity models.

ACKNOWLEDGMENTS

This work was partially supported by the Italian Minister for Education, and the National Council for the Research, Project MMAI. This paper has been concluded during the visit of both authors at the "Laboratoire de Modelisation en Mecanique," University of Paris VI, France. The authors are indebted to Professor H. Cabannes and Professor R. Gatignol for their support and for their encouraging this research line.

[1] R. Gatignol, *Théorie Cinétique des Gas a Répartition Discrete des Vitesses, Lecture Notes in Physics*, Vol. 36 (Springer, Berlin 1985).
[2] L. Tartar, Internal Report, Univ. Wisconsin–Madison (1980).
[3] T. Beale, Comm. Math. Phys. **106**, 569 (1986).
[4] J. M. Bony, Journées Equations aux Derivées Partialles, Centre Ecòle Polytecnique, Paris (1987).
[5] H. Cabannes and S. Kawashima, Comp. Rend. Acad. Sci. Paris **307**, 507 (1988).
[6] G. Toscani, to be published in Commun. Math. Phys. (1988).

[7]S. Kawashima, in *Recent Topics in Nonlinear PDE*, Lect. Notes in Numer. Appl. Anal. 6, Kinokuniya, Japan, 59 (1983).
[8]S. Kawashima, Commun. Math. Phys. **109**, 563 (1987).
[9]T. Platkowski and R. Illner, SIAM Rev. **30**, 213 (1988).
[10]N. Bellomo, A. Palczewski, and G. Toscani, *Mathematical Topics in Nonlinear Kinetic Theory* (World Scientific, Singapore, 1988).
[11]R. Gatignol and F. Coulouvrat, in *Discrete Kinetic Theory, Lattice Gas Dynamics and Foundations of Fluid Dynamics*, edited by R. Monaco (World Scientific, Singapore, 1989).
[12]E. Longo and R. Monaco, in *Rarefied Gas Dynamics, AIAA Progress in Astronautics and Aeronautics* n. 118, edited by E. P. Muntz, D. P. Weaver, and B. H. Campbell (AIAA, Washington, 1989).
[13]Y. Shizuta and S. Kawashima, Hokkaido Math. J. **14**, 249 (1985).
[14]S. Kawashima and Y. Shizuta, J. Mec. Théor. Appl. **7**, 597 (1988).
[15]C. Cercignani, Comp. Rend. Acad. Sci. Paris **301**, 89 (1985).
[16]H. Cabannes, in *Mathematical Methods in the Kinetic Theory of Gases*, edited by H. Neunzert and D. Pack (Land-Verlag, Frankfurt, 1979).
[17]N. Bellomo, R. Illner, and G. Toscani, Comp. Rend. Acad. Sci. Paris **279**, 835 (1984).
[18]E. Longo, R. Monaco, and T. Platkowski, J. Mec. Théor. Appl. **7**, 233 (1988).

Addendum: Static spherically symmetric space-times with six Killing vectors [J. Math. Phys. 29, 2473 (1988)]

Asghar Qadir and M. Ziad
Department of Mathematics, Quaid-i-Azam University, Islamabad, Pakistan

(Received 29 June 1989; accepted for publication 19 July 1989)

In a previous paper[1] static spherically symmetric space-times with six Killing vectors were studied. They gave the symmetry group $SO(1,2) \otimes SO(3)$. Of those space-times one metric was a generalization of a metric given by Petrov.[2] The Killing vectors for that space-time were not given separately. They are given explicitly by Petrov and satisfy the algebra $G \otimes SO(3)$, where G has the following commutation relations:

$$[X_1,X_2] = -X_3\,,$$
$$[X_3,X_1] = X_2,$$
$$[X_2,X_3] = 0\,,$$

and is not an $SO(1,2)$, but a solvable Lie algebra.

[1] A. Qadir and M. Ziad, J. Math. Phys. **29**, 2473 (1988).
[2] A. Z. Petroz, *Einstein Spaces* (Pergamon, New York, 1969).

 0022-2488/90/010254-01$03.00